



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO

MAESTRÍA Y DOCTORADO EN CIENCIAS AGROPECUARIAS
Y RECURSOS NATURALES

“PRODUCCIÓN DE VACUNAS CONTRA GARRAPATAS
ELABORADAS CON ANTÍGENOS BM86 ESPECÍFICOS DE LAS
CEPAS DE *R. microplus* DE MÉXICO”

TESIS

QUE PARA OBTENER EL GRADO DE DOCTOR EN CIENCIAS
AGROPECUARIAS Y RECURSOS NATURALES

P R E S E N T A

M. en C. A. R. N. SAÚL GABRIEL MARTÍNEZ ARZATE



El Cerrillo Piedras Blancas, Toluca Estado de México, Junio 2018



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO

MAESTRÍA Y DOCTORADO EN CIENCIAS AGROPECUARIAS
Y RECURSOS NATURALES

“PRODUCCIÓN DE VACUNAS CONTRA GARRAPATAS
ELABORADAS CON ANTÍGENOS BM86 ESPECÍFICOS DE LAS
CEPAS DE *R. microplus* DE MÉXICO”

TESIS

QUE PARA OBTENER EL GRADO DE DOCTOR EN CIENCIAS
AGROPECUARIAS Y RECURSOS NATURALES

PRESENTA:

M. en C. A. R. N. SAÚL GABRIEL MARTÍNEZ ARZATE

COMITÉ TUTORIAL:

Dr. JUAN CARLOS VÁZQUEZ CHAGOYÁN

Dr. ALBERTO BARBABOSA PLIEGO

Dr. ESVIETA TENORIO BORROTO



El Cerrillo Piedras Blancas, Toluca Estado de México, Junio 2018

AGRADECIMIENTOS

A la UAEM y FESE

Por el apoyo y financiamiento al proyecto de investigación No. 3378/2013E “**Producción de vacunas contra garrapatas elaboradas con antígenos Bm86 específicos de las cepas de *Rhipicephalus microplus* de México.**” y No. 3775/2014/CIC "Análisis de la variabilidad genética del gen Bm86 de poblaciones de *Rhipicephalus microplus* de ganado bovino de la República Mexicana para encontrar los epítopes más comunes para el desarrollo de una vacuna efectiva contra los diversos fenotipos existentes en nuestro país." financiados por la Fundación Educación Superior Empresa (FESE) y la Universidad Autónoma del Estado de México, respectivamente, para el desarrollo y culminación del presente trabajo.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT)

Por la beca proporcionada durante el periodo de estudios del Doctorado en Ciencias Agropecuarias y Recursos Naturales.

A la UAEM y la Secretaría de Investigación

Por brindar un espacio académico donde se desarrollaron habilidades y adquirieron conocimientos y donde se logró alcanzar las metas propuestas para la culminación del proyecto descrito en este documento.

Al Centro de Investigación y Estudios Avanzados en Salud Animal (CIESA)

Por todo el apoyo brindado en cuanto a la infraestructura y equipo necesarios para la conclusión de este trabajo

A los compañeros del departamento de Biología Molecular CIESA-UAEM

Mirna Faz, Viridiana Camacho, Laucel Ochoa, Eloy Contreras, Wael Hegazi, José A. Zepeda.
Por su apoyo brindado durante el proceso del Doctorado.

A los profesores

Porque en muchos momentos de este proceso, sus sabios consejos fueron muy valiosos para la culminación de este proyecto.

A la Universidad del País Vasco (UPV/EHU).

En especial la Dr. Humberto González-Díaz, responsable del departamento de Química Orgánica II de la Facultad de Ciencia y Tecnología, por recibirme y brindarme un espacio académico donde se desarrollaron habilidades y adquirieron conocimientos durante mi estancia en España.

A mis asesores:

Dr. Alberto Barbabosa Pliego.

Dr. Esvieta Tenorio Borroto.

Por el apoyo brindado en todos los aspectos para la realización del proyecto y por sus consejos que fueron tan importantes.

A mi tutor académico:

PhD: Juan Carlos Vázquez Chagoyán.

Por la guía, consejo y apoyo en cada etapa del proceso del posgrado en general y que llevo a la conclusión del mismo.

DEDICATORIAS

Para mi familia por el apoyo moral y estímulos brindados con infinito amor y confianza, pero sobre todo por infundir en mí el aliento necesario para concluir este trabajo de investigación, sin ellos simplemente no lo hubiera logrado.

Para mis hermanos en la fe, mi segunda familia, mi familia guerrera, porque con nadie he reído tanto como con ellos, aun en medio del dolor. Gracias por sus oraciones.

Para ti Dios que siempre me has sostenido con tu mano poderosa. "He peleado la buena batalla, he acabado la carrera, he guardado la fe." 2 Timoteo 4:7.

Para mi madre que es Luz, Paz, Apoyo, Torre fuerte, Refugio, etc. Gracias por aceptarme como tu hijo, por tu amor y por tus oraciones y por tu intercesión.

**“Producción de vacunas contra garrapatas elaboradas con antígenos Bm86
específicos de las cepas de *Rhipicephalus microplus* de México.”**

LISTA DE ABREVIATURAS

2-P	2-Propanol
Adj	Adyuvante usado en la vacuna junto al péptido
ALN	Agua Libre de Nucleasas
Am	Amitraz
BCP	1-bromo-3-cloropropano
CIESA	Centro de Investigaciones y Estudios Avanzados en Salud Animal
DNAc	Polímero de nucleótidos de doble cadena una de las cuales fue sintetizada a partir del ARN mensajero
dNTP	Desoxinucleótidos
DTT	Ditiotreitol
EGF	Factor de Crecimiento Epitelial
Host	Organismo huésped expuesto al antígeno
IEDB	Immune Epitope Database and Analysis Resource
LDA	Análisis linear discriminante
MJ	Método Median Joining
MLs	Lactonas Macroclícas
MP	Algoritmo heurístico de <i>parsimonia máxima</i>
MST	Árbol de Expansión Mínima
NOM	Norma Oficial Mexicana
OFs	Organofosforados
Org	Organismo fuente del péptido
OTUs	Operational Taxonomic Units
PBS	Solución Salina Tamponada con Fosfato
PCR	Reacción en Cadena de la Polimerasa
Proc	Proceso inmunológico de sensibilización
PS	Piretroides Sintéticos

PTML	Perturbation Theory and Machine Learning/Teoría de las perturbaciones y máquina de aprendizaje
<i>R. microplus</i>	<i>Rhipicephalus microplus</i>
RNA _t	Cadenas simples de Ribonucleótidos de extracción total de algún tipo de tejido
SAGARPA	Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación
SAP	Polimorfismo de un Solo Aminoácido
SENASICA	Servicio Nacional de Sanidad, Inocuidad y Calidad Agroalimentaria
SNP	Polimorfismo de un Solo Nucleótido
SRN	Red de Secuencia Recurrencia
Tech	Método o técnica utilizada para determinar la inmunogenicidad de los péptidos

INDICE

AGRADECIMIENTOS	1
DEDICATORIAS.....	3
LISTA DE ABREVIATURAS	5
ÍNDICE DE FIGURAS.....	9
ÍNDICE DE TABLAS.....	10
1 RESUMEN.....	11
2 ABSTRACT	13
3 INTRODUCCIÓN.....	15
4 MARCO TEÓRICO.....	18
4.1 La garrapata del ganado bovino.....	18
4.1.1 Generalidades de la garrapata	18
4.1.2 Distribución de la garrapata	20
4.1.3 Especies a las que afecta la garrapata	20
4.1.4 Las garrapatas en México	20
4.1.5 Breve consideración taxonómica de la garrapata	21
4.1.6 Morfología de la familia Ixodidae	22
4.1.7 La garrapata <i>R. microplus</i>	23
4.2.2 La campaña nacional contra <i>R. microplus</i> en México	25
4.2.3 El control químico	27
4.2.4 Desarrollo de resistencia a los garrapaticidas químicos.....	27
4.2.5 Mecanismos de resistencia	28
4.2.7 Desarrollo de resistencia a los garrapaticidas disponibles en México	29
4.3 El método inmunológico para combatir a la garrapata.....	30
4.3.1 Evidencia de la factibilidad del método inmunológico	30
4.3.2 Vacunas comerciales basadas en el antígeno recombinante <i>Bm86</i>	31
4.3.3 Mecanismo de acción de las vacunas <i>Bm86</i>	32
4.3.4 La molécula <i>Bm86</i>	32
4.3.5 Polimorfismo del gen <i>Bm86</i>	33
4.3.6 Predicción de epítomos lineales de células <i>B</i>	34
5 PLANTEAMIENTO DEL PROBLEMA	37
6 HIPÓTESIS	38
7 JUSTIFICACIÓN	39
8 OBJETIVO GENERAL.....	40
8.1 Objetivos particulares.....	40

9 MATERIALES Y MÉTODOS	41
9.1 Métodos experimentales	41
<i>9.1.1 Recolección y disección de los especímenes de R. microplus</i>	41
<i>9.1.2 Aislamiento y secuenciación del gen Bm86</i>	44
9.2 Métodos computacionales	49
<i>9.2.1 Análisis de las secuencias in silico</i>	49
<i>9.2.2 Análisis de la variabilidad Genética</i>	53
<i>9.2.3 Análisis filogenético</i>	55
<i>9.2.4. Diseño de un modelo predictivo para epítomos de células B</i>	57
<i>9.2.5 Determinación de regiones antigénicas consenso</i>	66
<i>9.2.6 Contrastación de los resultados con otros estudios</i>	67
10 RESULTADOS	68
10.1 Artículo publicado	68
<i>10.1.1 Resumen</i>	68
10.2 Artículo enviado	99
<i>10.2.1 Resumen</i>	99
10.3 Determinación de regiones antigénicas	124
11 DISCUSIÓN	126
12 CONCLUSIONES	135
13 BIBLIOGRAFIA	136
14 ANEXOS	147

ÍNDICE DE FIGURAS

Figura 1	Ciclo de vida la garrapata de un solo hospedero.....	19
Figura 2.-	Taxonomía de la garrapata.....	21
Figura 3.-	Características morfológicas de la familia <i>Ixodidae</i>	23
Figura 4.-	Ejemplares de la garrapata <i>R. microplus</i>	24
Figura 5.-	Situación actual del control de la garrapata <i>R. microplus</i>	26
Figura 6.-	Esquema representativo de desarrollo de la resistencia a los garrapaticidas.....	28
Figura 7.-	Vacunas comerciales en base al antígeno Bm86.....	31
Figura 8.-	Mapa de los estados muestreados.....	42
Figura 9.-	Esquema representativo y foto real de <i>R. microplus</i>	43
Figura 10.-	Diseción de los ejemplares <i>R. microplus</i>	44
Figura 11.-	Representación esquemática del diseño de los oligonucleótidos.....	46
Figura 12.	Electroforesis en gel de agarosa.....	48
Figura 13.-	Ensamble de secuencias en el software CodonCode Aligner.....	50
Figura 14.-	Alineamiento múltiple ClustalW.....	51
Figura 15.-	Análisis de secuencias en el software DnaSP 5.10.....	54
Figura 16.-	Ejemplo de hoja de cálculo de los índices de información del modelo.....	62
Figura 17.-	Matriz por pares para la validación y entrenamiento del modelo.....	63
Figura 18.-	Flujo de trabajo general.....	65
Figura 19.-	Resumen de los epítomos predichos.....	124

ÍNDICE DE TABLAS

Tabla 1.- Condiciones de los oligonucleótidos en la reacción de PCR.....	47
Tabla 2.- Secuencias homologas Bm86.....	52
Tabla 3. Características de los aminoácidos.....	60
Tabla 4. Comparación de resultados con distintos estudios.....	125

1 RESUMEN

Rhipicephalus microplus (antes conocida como *Boophilus microplus*), ha sido combatida por más de 60 años en nuestro país mediante la aplicación de tratamientos garrapaticidas sobre el cuerpo de los bovinos, sin embargo, la aparición de garrapatas resistentes y multirresistentes a los distintos químicos es un problema creciente. Un Método alternativo para combatir a la garrapata es el inmunológico, tal es el caso de una vacuna comercial basado en antígeno del intestino medio de la garrapata denominado como Bm86. Esta vacuna ha sido probada en campo con un éxito aceptable. Sin embargo, variaciones en la secuencia de aminoácidos mayores al 3% en comparación con el antígeno vacunal, parecen afectar la efectividad de la misma. La que efectividad de esta vacuna parece estar sujeta a la prevalencia de los diferentes genotipos de Bm86 en las distintas regiones geográficas del mundo. No se conoce la variabilidad del gen Bm86 en México, tampoco se ha hecho un estudio sobre como estas variantes impactan en la antigenicidad de la molécula. Conocer dicha información, podría ayudar a proponer candidatos vacunales más específicos. En el presente trabajo aislamos y secuenciamos el gen Bm86 de ejemplares de *R. microplus* provenientes de 10 localidades distintas del país, las secuencias obtenidas fueron sometidas a análisis moleculares para identificar los haplotipos circulantes en México. Se encontraron un total de 64 haplotipos en un total de 125 secuencias mexicanas. Los promedios de identidad para dicha secuencia se mantuvieron entre 99 y 98.5% en sus cadenas de nucleótidos y aminoácidos respectivamente. Los porcentajes de divergencia en la cadena de aminoácidos entre la cepa vacunal de referencia Yeerongpilly y nuestros aislados, oscilo entre 3 y 7.4%, en tanto que, la comparación con secuencias reportadas para otras partes del mundo, los valores más bajos de divergencia con nuestras secuencias se observaron en los reportes texanos (1-2.2%). Adicionalmente, en este

trabajo, desarrollamos un modelo de la teoría de las perturbaciones capaz predecir la propensión de un péptido de ser más, o menos antigénico después de una perturbación. Dicho modelo fue capaz de clasificar más de 26,112 casos de perturbaciones en péptidos provenientes de 14 secuencias del estado de Colima, México. Posteriormente, el modelo se aplicó a las 125 secuencias mexicanas. El modelo tiene una sensibilidad del 71% y especificidad del 80% para las series de entrenamiento y validación respectivamente. Con este modelo se predijo el impacto en la inmunogenicidad de cada péptido debido a la variación de dichas secuencias, lo que nos permitió proponer 3 secuencias consenso que reúnen toda la información de los epítomos lineales para células B como los candidatos vacunales más específicos para México.

2 ABSTRACT

Rhipicephalus microplus (formerly known as *Boophilus microplus*), has been combated by more than 60 years in our country. The most widely method used, has been the application acaricide treatments on the body of cattle. However, the emergence of acaricide resistance and multi-resistance ticks to different chemicals is a growing problem in our country. An alternative to fight the ticks is the immunological method, such is the case of the midgut antigen of the tick called Bm86. This commercial vaccine based on this antigen has been field tested with acceptable success. However, the effectiveness of this vaccine seems to depend on prevalence of the different genotypes of Bm86 in distinct geographical regions of the world. Ticks with variations in the Bm86 amino acid sequence greater than 3% in comparison with the vaccine antigen, are less sensitive to the vaccine. The variability of the Bm86 gene in Mexico is not known, nor has a study been carried out on how these variants impact on the antigenicity of the molecule. This information could help to propose more specific vaccine candidates. In the present work, we isolated and sequenced the Bm86 gene from *R. microplus* specimens from 10 different localities of the country. The sequences obtained were analyzed to identify the circulating haplotypes in the country. A total of 64 haplotypes were found in a total of 125 Mexican sequences, nucleotide and amino acid sequence identity means for the Bm86 were between 99 and 98.5% respectively. The percentages of divergence between the Yeerongpilly reference vaccine strain and our isolates, ranged between 3.1 and 7.4% in the amino acids sequence, while the lowest divergence values were observed between our sequences and the Texan reports (1-2.2%). Additionally, another task of this work was developing a model of perturbation theory capable of predicting the propensity of a peptide to be more or less antigenic after a disturbance. This model was able to classify more than 26112 cases of perturbations in

peptides from 14 sequences from the state of Colima, Mexico. Subsequently, the model was applied to the 125 Mexican sequences. The model has a sensitivity of 71% and specificity of 80% for the training and validation series, respectively. We predicted with our model the impact on the immune response due to the variation of these sequences, which allowed us to propose 3 consensus sequences that gather all the information of the linear epitopes for B cells as vaccine candidates more specific for Mexico.

3 INTRODUCCIÓN

La garrapata *Rhipicephalus microplus* (anteriormente conocida como *Boophilus microplus*) es considerada uno de los principales ectoparásitos del bovino en países tropicales y subtropicales de todo el mundo. De hecho, *R. microplus* ha sido combatida por más de 60 años en nuestro país por el programa federal “ACUERDO por el que se establece la Campaña Nacional para el control de la garrapata *Boophilus spp.*”, antes “Campaña nacional contra la garrapata *Boophilus spp.* (NORMA Oficial Mexicana NOM-019-ZOO-1994)”, debido al alto impacto económico negativo que la garrapata tiene sobre la ganadería bovina de carne y leche. Los esfuerzos de la campaña han sido sumamente importantes ya que han logrado mantener zonas libres y evitar hasta cierto punto la reproducción de la garrapata en zonas bajo control (Róger Iván Rodríguez-Vivas *et al.*, 2006), sin embargo, aproximadamente el 66% del territorio nacional se encuentra infestado por cepas de garrapatas *R. microplus*. El principal método de control utilizado por la campaña es la aplicación de productos químicos, sin embargo, en la actualidad el uso de garrapaticidas tiene una eficacia limitada debido a la aparición de garrapatas resistentes o multiresistentes a las distintas familias de químicos hasta ahora conocidos (Alonso-Díaz *et al.*, 2006). La resistencia de la garrapata a los garrapaticidas químicos, ha impulsado el desarrollo de vacunas para combatir al ectoparásito, este es el caso de una glicoproteína de las células epiteliales del intestino medio de la garrapata que en la década de los 80s fue caracterizada, secuenciada y denominada como Bm86 (Rand *et al.*, 1989). El antígeno Bm86 ha demostrado ser una alternativa viable para combatir a la garrapata, y al mismo tiempo bloquear la transmisión de enfermedades como la anaplasmosis y babesiosis que las garrapatas transmiten a su hospedero. Diversos estudios han mostrado que existen variaciones en la secuencia de la proteína Bm86 en cepas presentes en diferentes partes del mundo con respecto a las secuencias

en las que se basan las vacunas comerciales, lo que representaría una posible explicación al rango de eficacia de la vacuna (Anbarasi *et al.*, 2014; Freeman *et al.*, 2010; Kaewmongkol *et al.*, 2015). Una revisión de la efectividad de la vacuna a mediados de la primera década de este siglo mostró que el uso combinado de la vacuna y las sustancias químicas en baños inmersión redujo el uso de acaricida hasta en un 87% en Cuba, 100% en algunas regiones de Australia, 67% en algunas regiones de México, 77% en Colombia y 50% en Brasil (Andreotti *et al.*, 2008; de la Fuente *et al.*, 2007; de La Fuente *et al.*, 1999; Odongo *et al.*, 2007). Hasta ahora no se ha estudiado el grado de diversidad en la secuencia del gen Bm86 en México, por lo que se decidió evaluar el grado de polimorfismo y la estructuración geográfica de los haplotipos circulantes. Por lo que para este estudio se aisló y secuenció el gen Bm86 de ejemplares de garrapata *R. microplus* colectadas en 10 distintas localidades de México.

En los últimos años, se han desarrollado modelos de Teoría de Perturbación para predecir los cambios en la actividad de epítomos de células B después de múltiples variaciones/perturbaciones en sus secuencias péptidicas además de las variaciones en las distintas condiciones experimentales en las que estos fueron probados como antigénicos. Existe una gran cantidad de información disponible en bases de datos públicas (Immune Epitope Database and Analysis Resource, IEDB <http://www.iedb.org/>) que puede ser de gran importancia para el diseño de vacunas y la predicción de péptidos. Dado que la eficacia de la vacuna parece depender de las variaciones alélicas del gen Bm86 (Freeman *et al.*, 2010; García-García *et al.*, 1999; Kaewmongkol *et al.*, 2015; Lima *et al.*, 2000; Sossai *et al.*, 2005), se decidió desarrollar un modelo que pudiera predecir: 1) El impacto en la respuesta inmunológica debida a la variación de dichas secuencias y 2) La capacidad de un péptido de comportarse como un epítomo de células B. basados en esta información, propusimos candidatos vacunales

hipotéticamente más específicos para combatir a los genotipos Bm86 de la garrapata *R. microplus* de México.

4 MARCO TEÓRICO

4.1 La garrapata del ganado bovino.

4.1.1 Generalidades de la garrapata

Las garrapatas son arácnidos, sistemáticamente próximas a las arañas, pero sobre todo a los ácaros, representan después de los mosquitos el segundo vector transmisor de enfermedades a humanos y animales, transmitiendo bacterias, protozoarios y virus. Son ectoparásitos hematófagos obligados ya que requieren de ingerir sangre de sus hospedadores para completar su desarrollo, su ciclo de vida es complejo presentando dos principales fases: *Fase parasitaria* que puede ocurrir sobre un mismo hospedador o tener dos o tres diferentes, según la especie que se trate, en esta fase se alimentan de sangre al penetrar la piel de su huésped con ayuda de sus piezas bucales, *Fase de vida libre* que incluye el periodo de ovoposición y etapas entre mudas (Botello *et al.*, 2011; Róger Iván Rodríguez-Vivas *et al.*, 2006; Walker *et al.*, 2003).

Cada hembra de garrapata puede poner entre 3 y 4 mil huevos, esto sucede una vez que se ha dejado caer del hospedador preferentemente en pastizales. Durante su ciclo biológico, todas las garrapatas cursan cuatro estadios de maduración: huevo, larva, ninfa y adulto (Figura 1). Las garrapatas en su fase parasítica representan un problema sanitario para sus hospedadores, ya que es reservorio potencial de agentes infecciosos que pueden desencadenar en el hospedero alguna enfermedad en cuestión de días o semanas (Róger Iván Rodríguez-Vivas *et al.*, 2006).

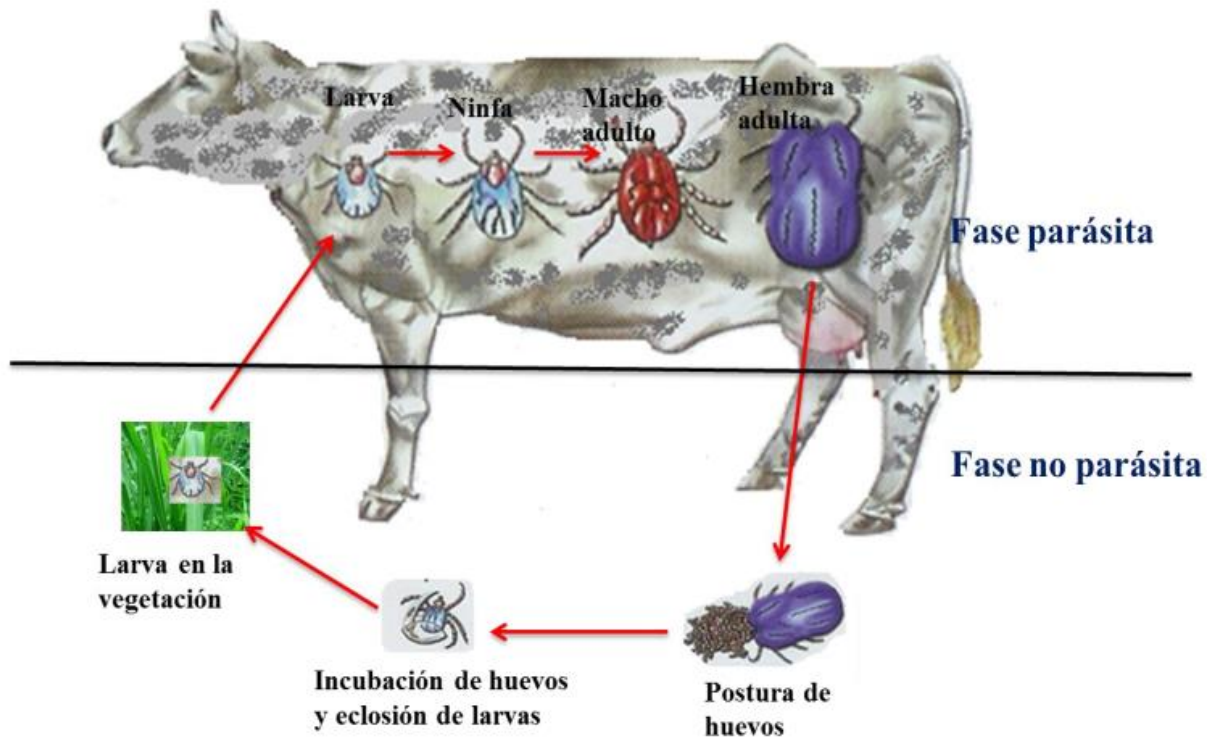


Figura 1.- Ciclo de vida la garrapata de un solo hospedero. En la imagen se puede observar el ciclo de vida de *R. microplus*, incluyendo su fase parasítica con el bovino como su principal hospedero y su fase no parasítica (Róger Iván Rodríguez-Vivas *et al.*, 2006).

Las garrapatas, causan pérdidas a la ganadería bovina, ya que afectan la ganancia de peso de los animales y la producción de leche, además reducen la fertilidad del ganado y provocan tiempos más prolongados de engorda, dichos daños pueden ser clasificados en directos o indirectos: *Daños directos*: lesiones y deterioro a las pieles por la acción de las piezas bucales de la garrapata; Efectos tóxicos debido a las enzimas y neurotoxinas presentes en la saliva que contribuyen a inhibir el apetito y la inmunidad del animal; Anemia y debilitamiento debido a la succión de sangre por parte del parásito (Nuttall *et al.*, 2006; Zivkovic *et al.*, 2010). *Daños indirectos*: desencadenados principalmente por la acción de agentes patógenos de suma importancia sanitaria transmitidos por la garrapata *R. microplus* como: *Babesia bovis*, *Babesia bigemina* y *Anaplasma marginale* (Domínguez-García *et al.*, 2010; Olaya Escobedo, 2007; Róger Iván Rodríguez-Vivas *et al.*, 2006).

4.1.2 Distribución de la garrapata

La garrapata puede encontrarse en las regiones tropicales y subtropicales de todo el mundo. Se han adaptado con éxito a la mayoría de los nichos terrestres, sin embargo, factores medioambientales como la altura (nivel del mar hasta los 2600msnm) temperatura (22-28°C) y la humedad relativa (80%) son determinantes en la distribución de las especies. Otros factores que considerar para la sobrevivencia de la garrapata son la presencia o ausencia de hospederos y vegetación, además de las prácticas de control y erradicación. En el caso específico de la garrapata *R. microplus*, es importante considerar que la presencia en zonas ganaderas de dicha garrapata podría deberse principalmente a los trasportes comerciales de ganado. La migración rutinaria de los animales silvestres en algunas regiones geográficas es un factor no menos importante ya que estos podrían estar participando en la distribución de la garrapata, lo que a su vez hace difícil su control ya que es prácticamente imposible capturar a las poblaciones de fauna silvestre para tratarlos por métodos químicos (Freeman *et al.*, 2010; Róger Iván Rodríguez-Vivas *et al.*, 2006).

4.1.3 Especies a las que afecta la garrapata

Las garrapatas pueden parasitar al hombre y a todos los animales vertebrados de sangre caliente, domésticos como: caballos, burros, cabras ovejitas, cerdos, perros; además de animales silvestres como ciervos y búfalos.

4.1.4 Las garrapatas en México

En México existen alrededor de 82 especies de garrapatas que afectan a los vertebrados domésticos, silvestres y al hombre. Las especies que se asocian al ganado bovino son los géneros *Boophilus*, *Amblyomma* y *Rhipicephalus*, sin embargo, la especie de mayor importancia para la ganadería bovina en nuestro país es *Rhipicephalus microplus*

(anteriormente conocida como *Boophilus microplus*), (Róger Iván Rodríguez-Vivas *et al.*, 2006; SAGARPA *et al.*, 2013).

4.1.5 Breve consideración taxonómica de la garrapata

Aproximadamente 899 especies de garrapatas han sido descritas en el mundo, todas ellas se agrupan en el suborden *Ixodoidea*, este da origen a tres familias (Ver Figura 2): *Ixodidae* a menudo llamadas garrapatas duras por la presencia de placas duras en su cuerpo, *Argasidae* llamadas garrapatas blandas debido a la ausencia de estas placas y *Nuttalliellidae* que comparte varias características de las familias anteriores, vive en África y está representada por el género mono-específico *Nuttalliella namaque* (Bazán Tene, 2002; Dwight, 2011; Róger Iván Rodríguez-Vivas *et al.*, 2006).

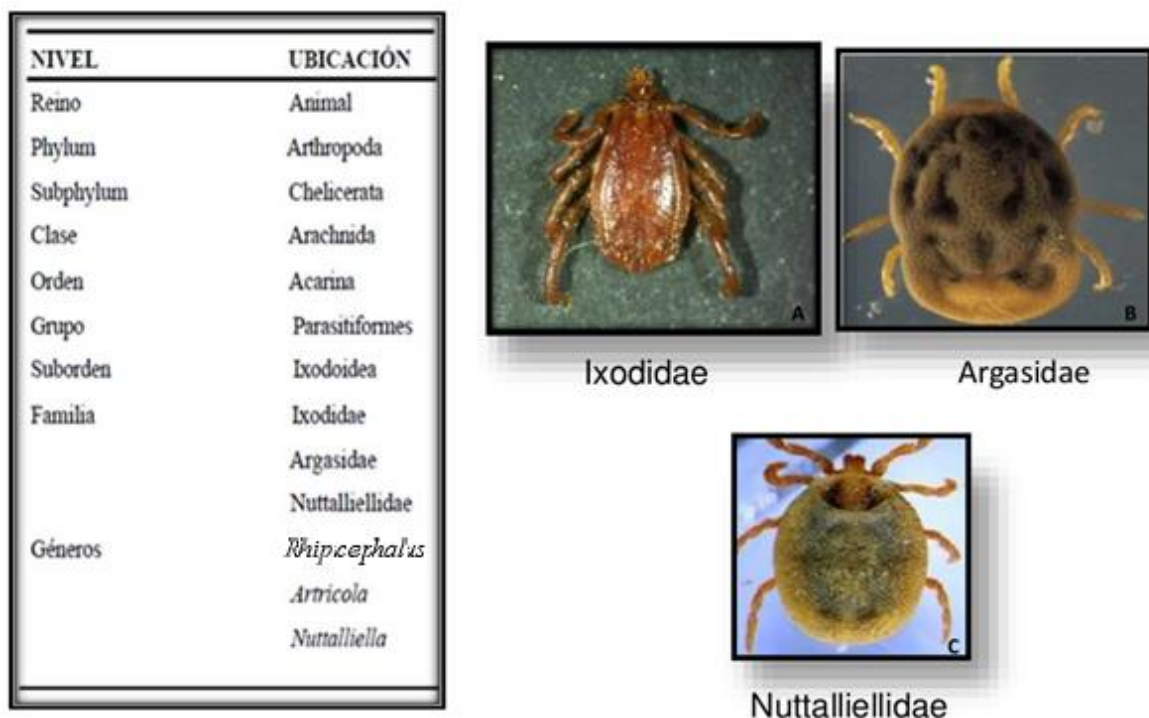


Figura 2.- Taxonomía de la garrapata. Se puede observar los niveles taxonómicos de la garrapata, además de ejemplares de las familias: **A** *Ixodidae* o garrapatas duras; **B** *Argasidae* o garrapatas blandas; **C** género mono-específico *Nuttalliellidae*; Tomado de www.unl.edu.ar/santafe/museocn.htm.

4.1.6 Morfología de la familia Ixodidae

Las garrapatas se pueden diferenciar morfológicamente hablando entre las fases de su ciclo de vida; por ejemplo, la larva presenta tres pares de patas mientras que en las fases ninfa y adulta se observan cuatro pares de patas, se pueden diferenciar también entre familia, género y especies, para fines prácticos de este trabajo nos dirigiremos a la familia *Ixodidae*. La familia *Ixodidae*, son las denominadas garrapatas duras, se caracterizan por tener un cuerpo redondeado, aplanado en estado de ayuno y globular cuando están repletas, pueden alcanzar dimensiones entre 1-12mm de longitud, presentan cuatro pares de patas con seis segmentos, su color puede variar de amarillento a café oscuro según la cantidad de sangre que ha ingerido (Alonso-Díaz *et al.*, 2006; Domínguez-García *et al.*, 2010). En su parte dorsal cuentan con la presencia de una gran placa esclerotizada. Las piezas bucales, están separadas del idiosoma y su posición es anterior a diferencia de argásidos que se encuentran en la cara ventral. Presentan dimorfismo sexual, se puede reconocer al macho por la presencia de un escudo dorsal quitinizado que lo cubre casi por completo, en las hembras dicho escudo dorsal se limita a la mitad anterior, además, el macho de menor tamaño que la hembra (Walker *et al.*, 2003). Un esquema más completo de las características morfológicas de la familia *Ixodidae*, se muestra en la Figura 3.

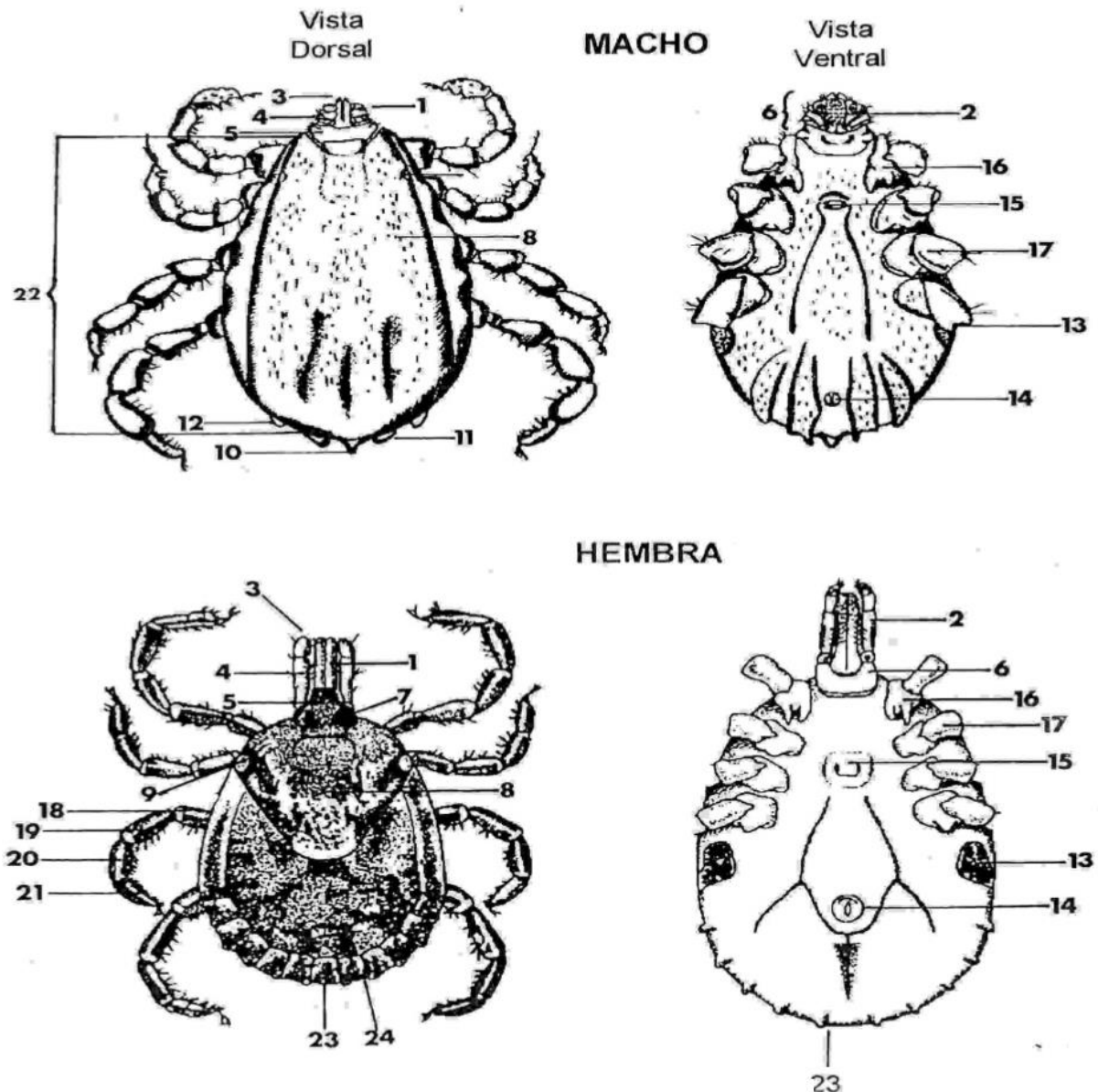


Figura 3.- Características morfológicas de la familia Ixodidae. Se puede observar un esquema representativo con las características morfológicas para la familia Ixodidae del macho y la hembra. 1.- Hipostoma, 2.- Dientes, 3.- Queliceros, 4.- Pedipalpos, 5.- Prosoma, 6.- Gnatosoma, 7.- Areas porosas, 8.- Escudo, 9.- Ocelos, 10.- Proceso caudal, 11.- Placas adanales, 12.- Placas accesorias, 13.- Peritrema, 14.- Orificio anal, 15.- Orificio genital, 16.- Coxas, 17.- Trocánter, 18.- Fémur, 19.- Tibia, 20.- Pretarso, 21.- Tarso, 22.- Idiosoma, 23.- Festones, 24.- Mamelones. Modificada de Walker (2003).

4.1.7 La garrapata *R. microplus*

R. microplus es considerada la garrapata más importante del ganado bovino en los países tropicales y subtropicales de todo el mundo, se ha extendido desde su origen en el subcontinente indio, a las principales áreas ganaderas como el nordeste de Australia, el sudeste de África, el caribe, México, varios países de América central, etc. En nuestro país *R. microplus* es uno de

los parásitos de mayor importancia, ya que causan pérdidas importantes a la ganadería mexicana, *R. microplus* actúan como reservorio del protozooario *Babesia* spp., y la bacteria *Anaplasma* spp. (Domínguez-García *et al.*, 2010). La garrapata *R. microplus* adulta, poseen un capitulum o prosoma corto y derecho, la hembra es considerablemente más grande que el macho, el cuerpo tiene forma entre ovalada y el escudo dorsal es ovalado ancho en la porción anterior y delgado en la posterior, las patas son de color amarillo y carecen de festones y ornamentos (Róger Iván Rodríguez-Vivas *et al.*, 2006), Figura 4.



Figura 4.- Ejemplares de la garrapata *R. microplus*. Se puede observar ejemplares vivos, aparentemente sin alimentar de la hembra (izquierda) y el macho (derecha) de *R. microplus*. Tomada de <http://enfermedades1987.blogspot.mx/2010/11/parasitos.html>.

4.2 Control de la garrapata *R. microplus* en México.

4.2.1 Impacto económico de la garrapata *R. microplus* a la ganadería mexicana

La importancia para combatir a *R. microplus*, radica en el impacto económico negativo. Se estima que en México las garrapatas y las enfermedades que transmiten producen pérdidas a la ganadería por aproximadamente 48 millones de dólares (USD) anuales (Rosario-Cruz *et al.*, 2009). Adicionalmente, el presupuesto asignado a la campaña para el control de la garrapata,

según la información hasta ahora disponible entre los años 2007-2015, oscila entre los 25 a 45 millones de pesos (SAGARPA and SENASICA, 2015). Adicionalmente, existe además un problema latente que afectan los tratados comerciales con Estados Unidos de Norte América, el principal socio comercial de México, ya que México exporta cerca de 1.3 millones de cabezas de animales en pie a los EE. UU, lo que genera divisas que fluctúan entre los 500 y 700 millones de dólares por año (Domínguez-García *et al.*, 2010; Rosario-Cruz *et al.*, 2009). Existe siempre el riesgo del cierre de la frontera debido a las infestaciones de garrapata en el ganado mexicano de exportación que representa un factor de riesgo para la ganadería de aquel país (Freeman *et al.*, 2010; Skaggs *et al.*, 2004).

4.2.2 La campaña nacional contra *R. microplus* en México

El control de la garrapata *R. microplus*, se inició en la segunda década del siglo XX, Cuando en los estados de Chihuahua (1927) y Sonora (1928), se llevaron a cabo acciones aisladas para combatir al ectoparásito. Fue hasta 1960 que el estado de Sonora inicio la campaña en forma técnica e intensiva, para 1969 la entonces Secretaría de Agricultura y Ganadería (SAG), instituyo un programa de carácter federal, dirigido a ganaderos, lo que hasta entonces significó el inicio de un proceso más eficaz y exitoso. A finales de 1975 el entonces Fideicomiso Campaña Nacional contra la Garrapata logró la canalización de créditos al campo con el fin de construir baños de inmersión, la distribución de ixodicidas y la construcción del Centro Nacional de Parasitología Animal. Como resultado de la extinción del Fideicomiso en 1984 las estrategias de la campaña fueron modificadas quedando la operatividad técnica y administrativa a cargo de la Dirección de Salud Animal dependiente de la Dirección General de Sanidad y Protección Agropecuaria y Forestal de la Secretaría de Agricultura y Recursos Hidráulicos (SARH). Hoy día el control de *R. microplus* se realiza de forma individualizada por los productores, con la participación de los gobiernos estatales y federales, las asociaciones

de productores ganaderos, a través de la Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación y del Servicio Nacional de Sanidad, Inocuidad, y Calidad Agroalimentaria (SAGARPA-SENASICA). *R. microplus* ha sido combatida por más de 60 años en nuestro país por el “ACUERDO por el que se establece la Campaña Nacional para el control de la garrapata *Boophilus spp.*”, antes “Campaña nacional contra la garrapata *Boophilus spp.* (NORMA Oficial Mexicana NOM-019-ZOO-1994)”. Los esfuerzos de la campaña han sido de suma importancia, ya que han logrado mantener zonas libres y limitar hasta cierto punto la reproducción de la garrapata en las zonas bajo control (Róger Iván Rodríguez-Vivas *et al.*, 2006), ver Figura 5.



Figura 5.- Situación actual del control de la garrapata *R. microplus*. Se puede observar en el mapa de la República Mexicana la superficie considerada como libre de *R. microplus* que comprende aproximadamente el 30.60% del territorio nacional, la superficie en fase de erradicación que comprende aproximadamente el 3.44% y la superficie en fase de control que corresponde aproximadamente al 65.96% del territorio nacional. Tomada de <https://www.gob.mx/senasica/documentos/situacion-actual-del-control-de-la-garrapata-boophilus-spp>.

4.2.3 El control químico

El control químico, consiste en romper el ciclo biológico de la garrapata por la aplicación de tratamientos con garrapaticidas (Rodríguez-Vivas *et al.*, 2006). Hasta ahora en México el método más eficiente para combatir a la garrapata es la aplicación de productos químicos a una frecuencia de tratamientos variable dependiendo la intensidad de la carga parasitaria en el ganado bovino.

La similitud en la estructura química de los garrapaticidas da origen a la agrupación de los químicos en familias. Las familias de productos químicos y productos químicos autorizados para emplearse en México son: organofosforados (OFs), piretroides sintéticos (PS), Amitraz (Am), fenilpirazolonas, inhibidores del desarrollo y lactonas macrocíclicas (MLs) (Taylor, 2001). Los métodos más comunes para la aplicación de los productos químicos son los baños garrapaticidas o la aplicación en forma de aerosol, por ello la presentación de los productos incluyen líquidos diluibles en agua, los emulsificables o los polvos humectables que faciliten el baño completo de los bovinos, la aplicación en forma de aerosol o el derrame dorsal (Pedroso *et al.*, 2007; R. I. Rodríguez-Vivas *et al.*, 2014; Rodríguez-Vivas *et al.*, 2014).

4.2.4 Desarrollo de resistencia a los garrapaticidas químicos.

Casi todo el siglo pasado, la investigación dirigió sus esfuerzos al desarrollo de estrategias de control y erradicación a través de insecticidas, ya fuera solos o en combinación con otros métodos tradicionales de control, pero tan solo lograron un éxito parcial. El desarrollo de resistencia en las poblaciones de garrapatas es un proceso de selección génica. La base del desarrollo de la resistencia es que, luego de la introducción de algún tipo de químico, al usarse este de manera intensiva ocasiona una fuerte presión de selección en un plazo variable de tiempo, dicha presión de selección elimina a los individuos susceptibles, los individuos

resistentes sobreviven, se reproducen y las características genéticas que les confieren resistencia son heredados, dando origen a una nueva generación de individuos también resistentes (Figura 6), es de esta forma, que de manera gradual aparecen dentro de las poblaciones tratadas garrapatas resistentes o multiresistentes a los distintos químicos usados (Kunz and Kemp, 1994; Cruz *et al.*, 2010). El problema parece hasta ahora no tener solución, ya que cada vez que se desarrolla un nuevo químico, después de un lapso de tiempo, se detectan casos de resistencia (Rosario-Cruz *et al.*, 2009). Esto ha ocasionado que los programas de control y erradicación incrementen considerablemente sus costos y aumenten las poblaciones de garrapatas que amenazan la estabilidad de la ganadería mexicana.

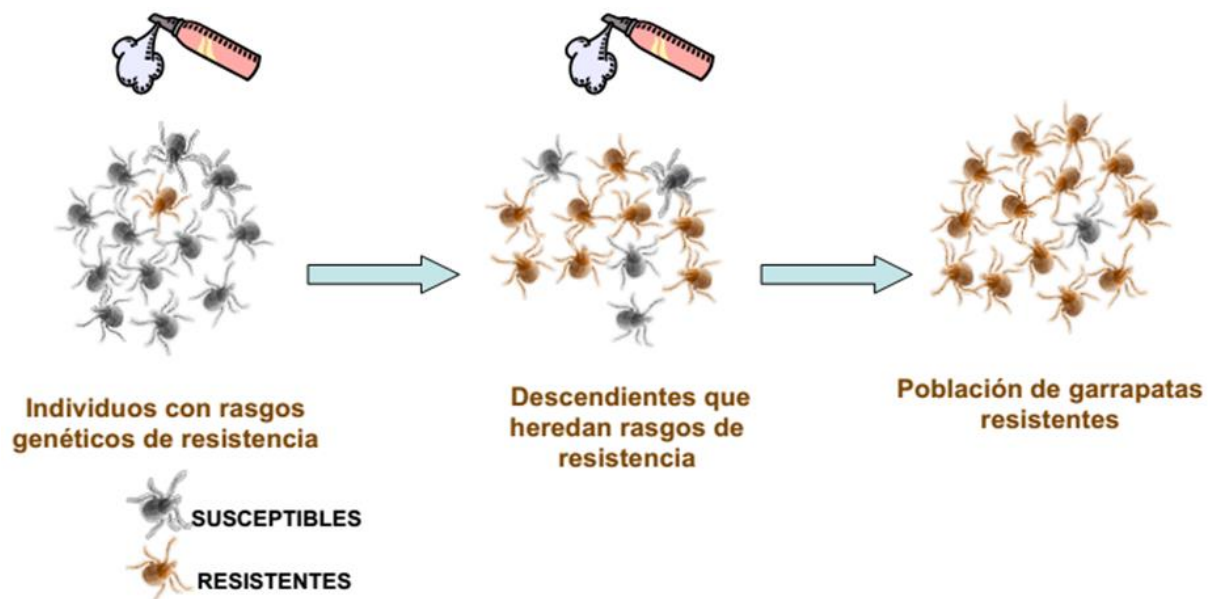


Figura 6.- Esquema representativo de desarrollo de la resistencia a los garrapaticidas. Tomada de (Cruz *et al.*, 2010).

4.2.5 Mecanismos de resistencia

Se reconoce que las garrapatas han desarrollado mecanismos de defensa a los garrapaticidas químicos por: 1.- *Modificación del comportamiento* para evitar el contacto con el insecticida; 2.- *Resistencia a la penetración del químico* por modificación del exoesqueleto; 3.- *Modificación del sitio blanco* que se da por la mutación genética en el sitio blanco del

garrapaticida, evitando así el efecto tóxico del ingrediente activo del producto químico; 4.- *Resistencia metabólica* por detoxificación del químico, mediado por enzimas que degradan y/o sequestran a las moléculas de los garrapaticidas.

4.2.7 Desarrollo de resistencia a los garrapaticidas disponibles en México

Los esfuerzos del “ACUERDO por el que se establece la Campaña Nacional para el control de la garrapata *Boophilus spp.*”, han sido sumamente importantes ya que han logrado mantener zonas libres y evitar hasta cierto punto la reproducción de la garrapata en zonas bajo control (Ver Figura 5), sin embargo en las últimas décadas debido al uso intensivo de los distintos garrapaticidas, *R. microplus* ha desarrollado resistencia a todas las principales clases de acaricidas, este es un grave problema para el control de la garrapata, ya que la disponibilidad de nuevos antiparasitarios es cada vez más escasa debido a que el desarrollo de nuevos acaricidas es un proceso costoso y prolongado (Alonso-Díaz *et al.*, 2006).

Los garrapaticidas OPs se usaron ampliamente en México entre 1974 y 1984, la resistencia por parte de las garrapatas a los OPs fue detectada aproximadamente en 1986, en este mismo año se introdujeron los PS y el Am, en el año 1993 surgieron los primeros casos de resistencia a PS (Pérez-Cogollo *et al.*, 2010). En lo que respecta al uso del Am en un principio era limitado debido al costo elevado, pero tras detectar la resistencia a PS se intensificó su uso, los primeros casos de resistencia al Am en México, se confirmaron en el año 2001 en un rancho en el estado de Tabasco (Rodríguez-Vivas *et al.*, 2006; Fernández-Salas, Rodríguez-Vivas and Alonso-Díaz, 2012; Róger Iván Rodríguez-Vivas *et al.*, 2014). Más recientemente en el año 2008 se informó del primer caso de resistencia al finopril en el norte del país (Miller *et al.*, 2008). En la década de los 80s las MLs o endectosidas comenzaron a comercializarse mostrando una excelente actividad frente a ectoparásitos y nematodos (Graf *et al.*, 2004; Lovis *et al.*, 2013;

Pérez-Cogollo *et al.*, 2010), sin embargo, diferentes niveles de resistencia por parte de *R. microplus* a la Ivermectina se detectaron por primera vez en el año 2010 en tres granjas de ganado en el estado de Yucatán (A. Fernández-Salas *et al.*, 2012; Pérez-Cogollo *et al.*, 2010; R. I. Rodríguez-Vivas *et al.*, 2014; Rodríguez-Vivas *et al.*, 2014; Rodríguez-Vivas *et al.*, 2010).

4.3 El método inmunológico para combatir a la garrapata

4.3.1 Evidencia de la factibilidad del método inmunológico

Existe abundante evidencia de los distintos mecanismos de resistencia de la garrapata a los acaricidas y del desarrollo de resistencia por parte del ectoparásito a las distintas familias de químicos, por lo que se han buscado métodos alternativos para combatir a la garrapata, este es el caso del método inmunológico. La factibilidad de controlar la infestación de garrapatas por el método inmunológico fue demostrada por primera vez por Allen y Humphreys (Allen and Humphreys, 1979) quienes demostraron la eficacia de la vacunación con extractos crudos de garrapatas, la identificación de los antígenos responsables de esta eficacia se dio a través de una serie compleja de fracciones proteicas, vacunación y evaluación de la respuesta antigénica. (Rand *et al.*, 1989; Willadsen, 2004). Fruto de estas investigaciones, en la segunda mitad de la década de los 80s, una glicoproteína situada en la superficie luminal de la membrana plasmática de las células epiteliales del intestino de la garrapata fue caracterizada, secuenciada y denominada como Bm86 (Rand *et al.*, 1989; Róger Iván Rodríguez-Vivas *et al.*, 2006). Esta molécula ha sido usada in diversos estudios en los que ha sido capaz de estimular una respuesta inmune que protege a los bovinos contra la posterior infestación (Andreotti, 2006; Canales *et al.*, 1997; de la Fuente *et al.*, 1998; García-García *et al.*, 2000; Willadsen *et al.*, 1995, 1992).

4.3.2 Vacunas comerciales basadas en el antígeno recombinante Bm86

Posterior al descubrimiento del antígeno Bm86, dos vacunas recombinantes (TickGARD y Gavac, Ver Figura 7) fueron desarrolladas y registradas: “La vacuna comercial TickGARD y posteriormente TickGARD^{plus}” desarrollada por la Organización de Investigación Científica e Industrial del Commonwealth (CSIRO) en colaboración con Biotech Australia y lanzada a mercado en Australia en el año de 1994 por Hoechst Animal Health. Por diversos factores comerciales y empresariales, TickGARD desapareció del mercado, sin embargo, después de varios años fue reintroducida y vendida por un corto periodo de tiempo por Intervet Australia Pty. Ltd. (Bendigo, Australia) específicamente a la industria láctea (Andreotti, 2006; Andreotti *et al.*, 2008; G. S. Cobon *et al.*, 1997; Cunha *et al.*, 2011; Odongo *et al.*, 2007); “La vacuna comercial Gavac y la posterior Gavac^{plus}” desarrollada en la Habana Cuba, por el Centro de Ingeniería Genética y Biotecnología y lanzada a mercado en 1993 por la empresa Heber Biotec (Figura 7). Gavac se ha comercializado en varios países de América Latina, en el caso específico de nuestro país, Gavac fue lanzada al mercado mexicano en 1997 por Revetmex S.A. de C.V (Canales *et al.*, 1997; Rodríguez Valle *et al.*, 2001; Hassan *et al.*, 2007; Vargas *et al.*, 2010).



Figura 7.- Vacunas comerciales en base al antígeno Bm86. Se puede observar las presentaciones comerciales de las vacunas: **TickGARD^{plus}** lanzada a mercado en Australia en el año de 1994 por Hoechst Animal Health (A) y **Gavac** lanzada a mercado en 1993 por la empresa Heber Biotec y posteriormente comercializada en México desde 1997 por Revetmex S.A. de C.V.

4.3.3 Mecanismo de acción de las vacunas Bm86

Las vacunas que contienen el antígeno Bm86, actúan a través de anticuerpos anti-Bm86 con la posible participación del complemento y otros mecanismos efectores, esta respuesta del huésped a la vacunación lisa las células epiteliales del intestino de la garrapata resultando en la reducción en el número de garrapatas que completan su ciclo de vida y en su fertilidad (Rodríguez *et al.*, 1994). Este principio para la inmunidad en el control de las infestaciones de garrapatas se basa en una respuesta de anticuerpos policlonales contra el antígeno objetivo, por lo que debería ser más difícil desarrollar resistencia por parte de las poblaciones de garrapatas (G. Cobon *et al.*, 1997; Rodríguez *et al.*, 1994).

4.3.4 La molécula Bm86

Bm86 es una glicoproteína que fue originalmente aislada por Willadsen y colaboradores en 1989 (Willadsen *et al.*, 1989) a partir de células epiteliales del intestino medio de la garrapata. La cadena traducida codifica para una molécula de 650 aminoácidos y un peso molecular predicho de aproximadamente 80 kDa. En lo que respecta al procesamiento de la molécula, se reconocen tres sitios de función: 1.- *El péptido señal N-Terminal* (1-19aa) que es la cadena que aparece primero cuando se sintetiza la cadena de polipéptidos y contiene la información sobre la eficiencia de secreción, ruta y destino de la proteína, 2.- *El antígeno o glicoproteína Bm86* (20-627aa) que es propiamente la cadena polipeptídica de la proteína madura después del procesamiento, 3.- *Un propéptido* (628-650aa), que es la parte de la proteína que se escinde durante la maduración o activación. Se sabe que la proteína madura contiene una región transmembrana hidrofóbica que abarca 23aa ubicada en la región del extremo C-terminal, el resto de la proteína presenta un alto potencial hidrofílico. La secuencia contiene también 7 dominios de tipo EGF (Epithelial Growth Factor) o factor de crecimiento epitelial (Azhahianambi *et al.*, 2009). Se sabe que el antígeno Bm86 es una glicoproteína de las células

epiteliales del intestino medio de las garrapatas, se presume que es muy plegada y que está relacionada con los procesos de endocitosis en las células del intestino. El antígeno está presente desde la fase de huevos, hasta la fase adulta (Willadsen, 2004; Willadsen *et al.*, 1989). La proteína Bm86 está extensamente glicosilada, lo cual, en su estado natural induce una respuesta fuerte de anticuerpos si esta se usa como antígeno (Willadsen, 2004).

4.3.5 Polimorfismo del gen Bm86

Las vacunas desarrolladas en Australia y Cuba basadas en el antígeno recombinante Bm86 tienen una eficacia variable de acuerdo con la región geográfica donde se usa, una posible explicación de esta inconsistencia en la eficacia podría ser el polimorfismo alélico que se deriva en cambios entre la secuencia de aminoácidos del antígeno Bm86 vacunal y la Bm86 nativa que se esté expresando en las garrapatas en alguna región en particular. En el pasado se ha demostrado que variaciones superiores del 2.8% en la secuencia de aminoácidos son suficientes para reducir la eficacia en la vacuna. (G. Cobon *et al.*, 1997; García-García *et al.*, 1999; Kaewmongkol *et al.*, 2015; Rodríguez *et al.*, 1994; Willadsen *et al.*, 1995). Es natural pensar que las variaciones genéticas se vean reflejadas en distintos haplotipos en distintas regiones geográficas, que permite generar un modelo de islas geográficas en el que cada población diverge al mismo tiempo en lugares separados, lo que permite que el polimorfismo del gen en cuestión muestre una estructura geoespacial. De ahí que las vacunas con base en el antígeno Bm86 hayan sido utilizadas con éxito en países como Cuba y Australia, siendo este último país de donde fue originalmente aislado el antígeno, pero que haya tenido un éxito considerablemente menor en otras regiones geográficas, presumiblemente debido a una reducida eficacia de la vacuna derivado del polimorfismo genético geoespacial de *R. microplus* (G. S. Cobon *et al.*, 1997; de la Fuente *et al.*, 2007; Valle *et al.*, 2004; Vargas *et al.*, 2010).

Se ha reportado variabilidad genética entre secuencias del gen Bm86 de distintas partes del mundo, sin embargo, hasta ahora solo dos trabajos han reportado la variación de Bm86 en regiones geográficas concretas. Fremman y colaboradores (Freeman *et al.*, 2010), reportaron 27 aislamientos del sur de Texas. Los reportes aislados fueron separados en 12 grupos distintos o haplogrupos basados en la identidad de la secuencia de aminoácidos; Kaewmongkol y colaboradores (Kaewmongkol *et al.*, 2015) recolectaron muestras de 13 localidades del norte, noreste, centro y sur de Tailandia, 29 secuencias fueron evaluadas para determinar la variabilidad del gen en ese país, en dicho estudio, según el autor, el análisis filogenético reveló patrones geográficos entre las secuencias Bm86 de Tailandia.

4.3.6 Predicción de epítomos lineales de células B

Los epítomos de linfocitos B pueden ser definidos como clusters de aminoácidos accesibles que son reconocidos por anticuerpos secretados o receptores en linfocitos B o T y que son capaces de producir una respuesta inmune humoral o celular (Getzoff *et al.*, 1988; Larsen *et al.*, 2006). Actualmente la predicción de epítomos lineales de células B, es una técnica computacional usada como parte del proceso del diseño de vacunas, en el diagnóstico de enfermedades y la investigación de alergias. Los métodos experimentales de laboratorio demandan gran cantidad de tiempo y de recursos económicos, lo que hace la labor del desarrollo de vacunas un proceso lento y costoso. Una alternativa que puede contribuir a reducir dichos costos es el uso de métodos *in silico* enfocados a la predicción de epítomos lineales de células B y la antigenicidad de dichas secuencias ahorrando así, tiempo y recursos.

El método de escala de propensión es la técnica mayormente utilizada en la predicción de epítomos lineales de células B, este método asigna un valor de propensión para cada aminoácido en la secuencia en cuestión, basándose en estudios sobre sus propiedades fisicoquímicas. Las

fluctuaciones en los valores de predicción de la secuencia son reducidas por la aplicación de modelos estadísticos (Larsen *et al.*, 2006).

Se han desarrollado varios métodos computacionales para predecir epítomos lineales de células B: a manera de ejemplo se describen a continuación una selección de ellos: *BepiPred* se basa en la combinación de dos métodos; las predicciones de modelo oculto de Markova y la escala de propensión de Parker. Este método asigna un valor a cada aminoácido en la cadena completa en cuestión, dependiendo de los parámetros de treshold en el que los valores de sensibilidad (verdaderos positivos) y especificidad (verdaderos negativos) pueden ser modificados. Disponible en línea en <http://www.cbs.dtu.dk/services/BepiPred/>. *Hopp and Woods*, asigna un valor numérico (hidrofilicidad) a cada aminoácido de la secuencia. Se muestran los valores en una gráfica, los valores positivos resaltan los péptidos en la secuencia con mayor tendencia a ser hidrofílicos, mientras que los valores negativos describen los péptidos con predicción a ser hidrofóbicos (Hopp and Woods, 1981), disponible en línea en <https://web.expasy.org/protscale/>. *SVMTriP* utiliza la máquina de vectores de soporte para integrar a la información la similitud y la propensión de los llamados tripéptidos, disponible en línea en <http://sysbio.unl.edu/SVMTriP/prediction.php>.

Sin embargo, el avance en la investigación inmunológica efectiva exige un rendimiento más robusto de los métodos de predicción que lo que los algoritmos actuales podrían proporcionar. En estudios recientes se ha propuesto el método de la teoría de las perturbaciones y máquina de aprendizaje o Perturbation Theory and Machine Learning (PTML), para predecir cambios en la actividad de un epítomo de células B, después de múltiples variaciones (Perturbaciones). Con estos modelos es posible tomar en cuenta las perturbaciones que pueden ocurrir por la variación en la secuencia de un péptido conocido y/o variaciones en las condiciones experimentales del

ensayo en las que es probado el epítipo (González-Díaz *et al.*, 2014; Vázquez-Prieto *et al.*, 2016).

5 PLANTEAMIENTO DEL PROBLEMA

Según la Campaña Nacional para el control de la garrapata, la superficie libre comprende 599,367.84 km², y representa el 30.60% del territorio nacional; la superficie en fase de erradicación abarca 67,472.76 km², y representa el 3.44%; La superficie en fase de control comprende 1, 292, 407.02 km², y representa 65.96% por lo que gran parte del territorio nacional se encuentra infestado de garrapatas. Dicha campaña no ha tenido el éxito deseado porque en México se han detectado poblaciones de garrapatas resistentes a prácticamente todos los ixodicidas. Un problema adicional que genera el uso de garrapaticidas químicos es la contaminación del medio ambiente y de los productos derivados de los bovinos como la carne y la leche, por lo que una estrategia de control de estos ectoparásitos con el estado actual del conocimiento debería incluir el uso de vacunas contra garrapata. La vacuna Gavac está disponible en nuestro país desde el año 1997 para ayudar a erradicar las infestaciones de garrapata, sin embargo, la especificidad de la vacuna es considerablemente menor en México que en otros países, posiblemente debido al polimorfismo genético entre las cepas de *R. microplus* que ya ha sido reportado por varios autores en distintas partes del mundo. Hasta ahora no se han estudiado las variantes del gen Bm86 en México, por lo que consideramos interesante evaluar el grado de polimorfismo a nivel nacional, estudiar la estructuración geográfica de los posibles haplotipos y analizar si estos pudieran ser candidatos vacunales más específicos para combatir a las poblaciones de garrapatas *R. microplus* de México.

6 HIPÓTESIS

Asumiendo que el gen Bm86 en nuestro país presenta variabilidad genética que ha dado origen a la presencia de haplotipos específicos y por otra parte que la variabilidad genética del gen Bm86 es la responsable de la menor eficacia de la vacuna Gavac en nuestro país, La identificación de los haplotipos circulantes permitirá proponer candidatos vacunales más específicos para combatir a las poblaciones de garrapata *R. microplus* de México.

7 JUSTIFICACIÓN

El presupuesto asignado a la campaña nacional para el control de la garrapata en los últimos años oscila entre los 25 a 45 millones de pesos. Los esfuerzos de la campaña nacional contra la garrapata *R. microplus*, han logrado mantener zonas libres y limitar hasta cierto punto la reproducción de la garrapata en las zonas bajo control, sin embargo, el uso intensivo de los garrapaticidas químicos ocasiona una fuerte presión selectiva que ha favorecido la aparición de poblaciones de garrapatas resistentes o multirresistentes a los distintos químicos hasta ahora conocidos.

El problema de resistencia afecta la economía de los productores ya que las garrapatas y las enfermedades que transmiten producen pérdidas a la ganadería de aproximadamente 48 millones de dólares (USD) anuales. La garrapata también afecta la competitividad internacional de la ganadería de nuestro país, por ejemplo, México exporta cerca de 1.3 millones de cabezas de animales en pie a los EE. UU, lo que genera divisas de que fluctúan entre los 500 y 700 millones de dólares por año. Las posibles infestaciones de garrapata en el ganado mexicano de exportación es un riesgo latente que podría causar el cierre de la frontera con EE. UU., el principal socio comercial de México. Una estrategia de control de estos ectoparásitos con el estado actual del conocimiento debería incluir el uso de vacunas contra garrapata más específicos para la región México.

8 OBJETIVO GENERAL

Proponer candidatos vacunales Bm86 específicos para combatir a las poblaciones de garrapatas *R. microplus* circulantes en diversas regiones de México.

8.1 Objetivos particulares

1. Aislar y secuenciar el gen Bm86 de distintas regiones geográficas de México.
2. Estudiar el polimorfismo genético en el gen Bm86 de *R. microplus* de México.
3. Verificar si los haplotipos observados se encuentran confinados a regiones geográficas específicas en México.
4. Determinar los haplotipos representativos de nuestro país.
5. Predecir epítomos lineales de células B.
6. Proponer uno o más candidatos vacunales contra *R. microplus* basados en el polimorfismo del gen Bm86

9 MATERIALES Y MÉTODOS

El presente trabajo se realizó en el laboratorio de biología molecular y bioterio del “Centro de Investigación y Estudios Avanzados en Salud Animal” (CIESA), de la “Facultad de Medicina Veterinaria y Zootecnia” (FMVZ) perteneciente a la “Universidad Autónoma de Estado de México” (UAEMEX). El CIESA se localiza en el kilómetro 15.5 de la carretera Panamericana en su tramo Toluca–Atlacomulco y en el “Departamento de Química Orgánica II” de la “Facultad de Ciencia y Tecnología” de la “Universidad del País Vasco” (UPV/EHU) con dirección en Barrio Sarriena, s/n, Lejona, Vizcaya, España.

La estrategia usada para esta investigación consta de dos partes: Métodos experimentales y Métodos computacionales.

9.1 Métodos experimentales.

9.1.1 Recolección y disección de los especímenes de *R. microplus*

Se recolectaron y seleccionaron ejemplares adultos de garrapatas *R. microplus* de bovinos infestados de 10 distintas localidades de la República Mexicana, específicamente del área en fase de control según el “ACUERDO por el que se establece la Campaña Nacional para el control de la garrapata *Boophilus spp.*”. Dichas localidades se citan a continuación: Rancho El palmar en el municipio de Sinaloa, *Sinaloa*; Establo sin nombre en la localidad de China, municipio de Campeche, *Campeche*; Rancho sin nombre, municipio de San Pedro Lagunillas, *Nayarit*; Rancho él solito, municipio Moyahua de Estrada en *Zacatecas*; Rancho Granja Santa Amalia, municipio de Colima, *Colima*; Establo sin nombre en Ciudad Altamirano, municipio de Pungabarato, *Guerrero*; hato de la Universidad Autónoma de Chiapas, municipio de Tuxtla Gutiérrez, *Chiapas*; Rancho sin nombre, Localidad de Luis Gil Pérez, municipio de centro,

Tabasco; carretera Boca del Rio-Córdoba Km. 43, Veracruz; Rancho sin nombre Carretera a Miacatlán, Morelos. Ver Figura 7. A cada individuo se le asigno la abreviatura del estado donde se recolecto y un número como identificador.

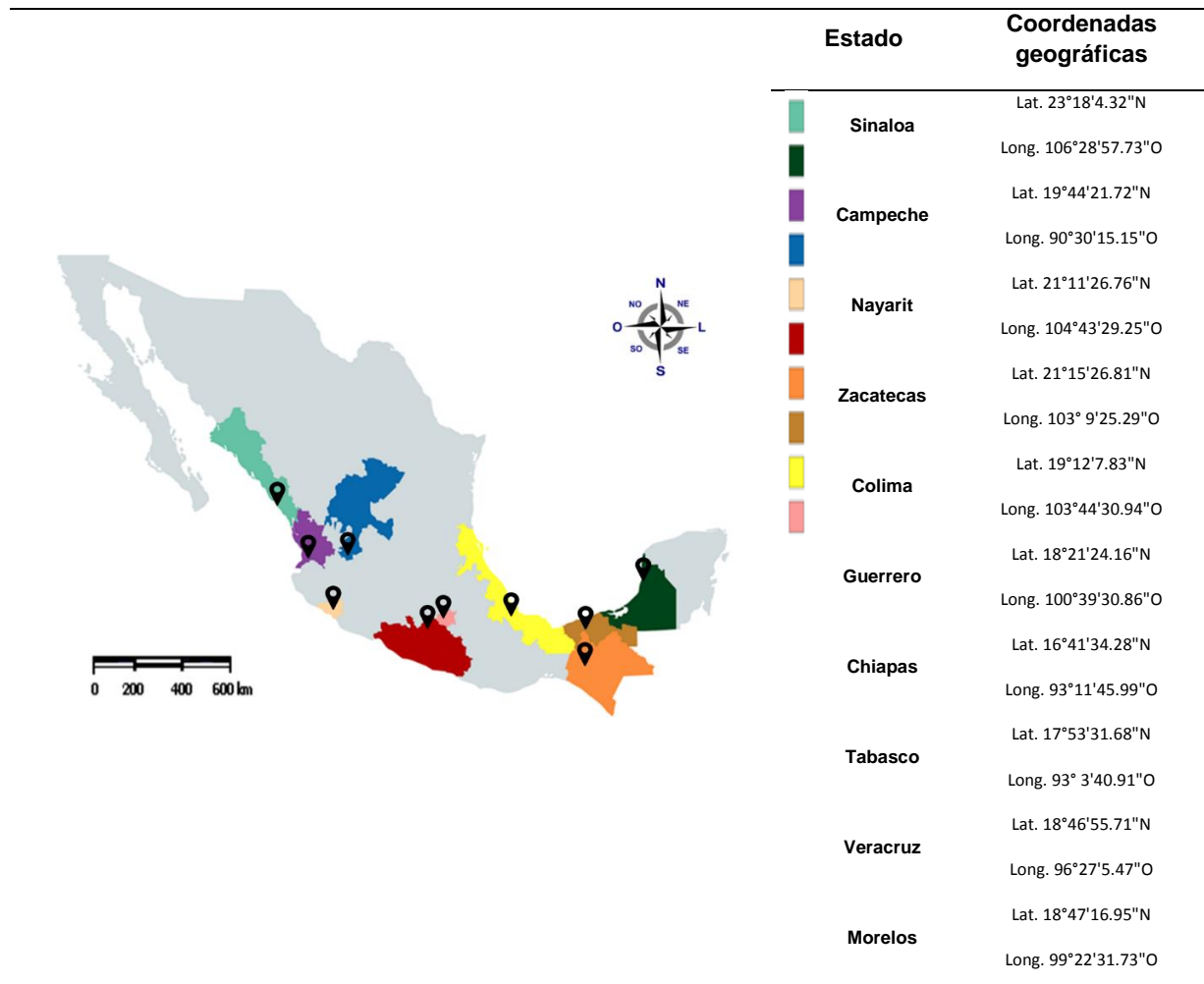


Figura 8.- Mapa de los estados muestreados. Se puede observar en el mapa la ubicación aproximada de las localidades muestreadas y las coordenadas geográficas exactas para cada localidad.

9.1.1.1 Determinación morfológica de *R. microplus*

Todas las muestras colectadas en las 10 localidades de la República Mexicana fueron observadas al microscopio estereoscópico con el fin de seleccionar únicamente ejemplares de la especie *R. microplus*. Utilizando el manual “Ticks of Domestic Animals in Africa: a Guide to Identification of Species” (Walker *et al.*, 2003), se buscaron las siguientes características: La presencia dorsal de una zona porosa en forma de ovalo sobre *basis capituli*; En la zona ventral

se buscó la presencia de un espolón en las coxas 2 y 3, dicho espolón está ausente en *R. annulatus*; Se observó por último la forma de los palpos, que en *R. microplus* deben ser cóncavos y cortos a diferencia de *R. annulatus* donde son cóncavos y largos (Figura 9).

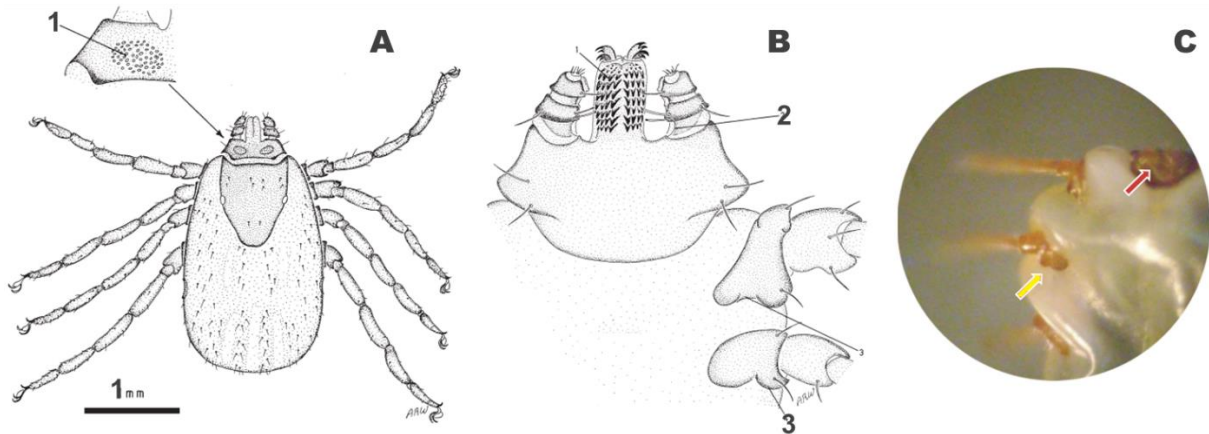


Figura 9.- A y B esquema representativo y C fotografía original de un espécimen visto al microscopio estereoscópico de *R. microplus*. (A #1) presencia dorsal de una zona porosa en forma de ovalo sobre *basis capituli*; (B #2 y C flecha roja) se pueden observar los palpos cóncavos y cortos; (B #3 y C flecha amarilla) presencia de un espolón en la coxa dos. Esquemas A y B, modificados de Walker (2003).

9.1.1.2 Disección de los ejemplares *R. microplus*

Para manipular a las garrapatas se usaron pinzas entomológicas de punta fina, los individuos fueron colocados en posición ventro-dorsal sobre la primera mitad de una laminilla estéril, sobre la segunda mitad se colocaron 100 μ L de solución salina tamponada con fosfato estéril (PBS). Con un bisturí estéril, se realizó un corte por encima del escudo dorsal a nivel de la coxa 1, para posteriormente ejercer una presión de barrido con el recazo del bisturí en dirección de caudal a craneal y así extraer los órganos internos de la garrapata, entre ellos el intestino (Figura 10). Los tejidos disecados se arrastraron inmediatamente hasta la solución PBS en la laminilla para evitar que los tejidos se deshidrataran y a la vez se pudieran manipular para separarlos los intestinos de otros tejidos.

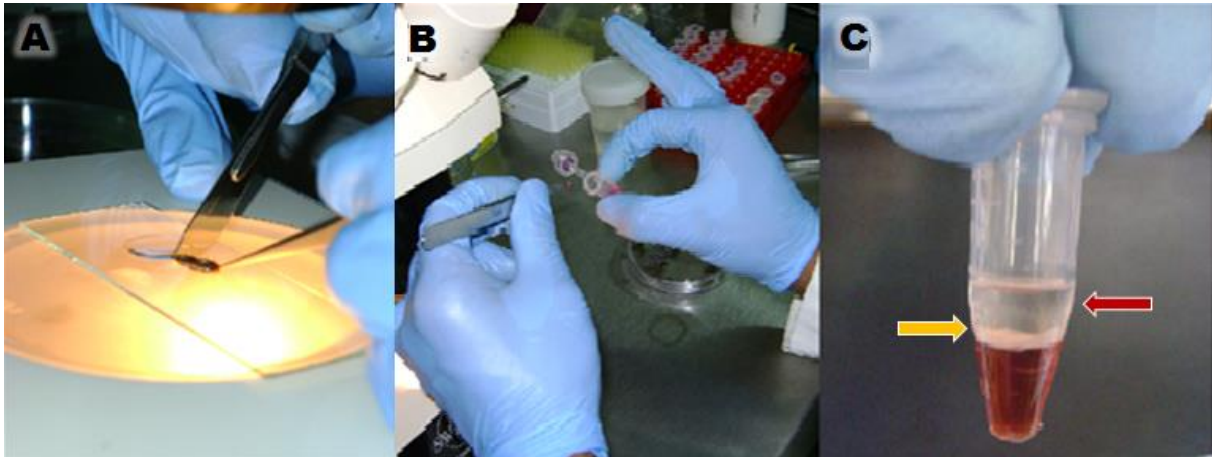


Figura 10.- Disección de los ejemplares *R. microplus*. Se puede observar: (A) Corte con bisturí estéril por encima del escudo dorsal a nivel de la coxa 1. (B) Inclusión de los intestinos en tubos de 1.5 ml con solución de Trizol. (C) Separación de fases: Fase acuosa flecha roja, Interfase flecha amarilla.

9.1.2 Aislamiento y secuenciación del gen Bm86.

9.1.2.1 Homogenización de las muestras

Entre 10 y 15mg de intestinos fueron recuperados por individuo, cada uno se homogenizo de manera individual en tubos de 1.5 ml con solución de Trizol (Invitrogen, USA., 1 ml por cada 50-100 miligramos de muestra). Se incubo el homogenizado durante 5 minutos a temperatura ambiente. Finalmente, todos los tejidos homogenizados se almacenaron a -80°C para su posterior análisis.

9.1.2.2 Extracción de RNA t

Se descongelaron las muestras de tejidos previamente homogenizadas, con el fin de eliminar el material insoluble se centrifugaron las muestras a $10,000\text{ g}/10\text{min}/4^{\circ}\text{C}$. Se recuperó el sobrenadante y se transfirió a un nuevo vial de 1.5 ml. Se agregaron $100\ \mu\text{L}$ de 1-bromo-3-cloropropano (BCP) por cada ml de Trizol utilizado en la homogenización. Posteriormente se centrifugo a $12,000\text{ g}/10-15\text{min}/4^{\circ}\text{C}$. En este paso se pudieron observar la separación de fases:

La fase fenólica al fondo del tubo, la interfase donde se mantiene el DNA y las proteínas y la fase acuosa donde permanece exclusivamente el RNAt. (Figura 10). Se recuperó la fase acuosa, teniendo cuidado de no tocar la interfase, se transfirió a un tubo nuevo de 1.5 ml. El RNAt se lavó, precipitó y resuspendió agregando 500 µl de 2-Propanol (2-P) por cada ml de Trizol utilizado en la homogenización, seguido de un paso de vortex por 5 segundos para después incubar las muestras a temperatura ambiente por 5 min. Posteriormente, para la precipitación del RNAt se centrifugó a 12,000 g/8 min/4°C, para después descartar el sobrenadante. El pellet obtenido se lavó agregando 1ml de Etanol al 75% por cada ml de Trizol utilizado en la homogenización. Finalmente se centrifugó a 7,500 g/5 min/4°C, se descartó el sobrenadante y el pellet obtenido se dejó secar a temperatura ambiente para permitir que el etanol residual se evaporara para posteriormente resuspender en 40 µl de agua libre de nucleasas (ALN).

9.1.2.3 Síntesis de DNAc

Previo a la síntesis de DNAc, se realizó una digestión enzimática con la finalidad de eliminar el DNA genómico posiblemente presente en las muestras que pudiera interferir en la síntesis de DNAc, para lo cual se usó el kit DNA Free de Ambion. En un tubo de PCR, se incubaron de manera individual en un termociclador (Techne TC-512 USA) a 37 °C, 20µl de RNAt de cada muestra en presencia de 0.5 µl de enzima DNAasa 2.2 µl de buffer 10x y 2.5 µl de reactivo inactivador de RNAsas.

Para la síntesis de DNAc, se siguió el protocolo SuperScript II Reverse Transcriptase (Invitrogen, by life technologies™). Se preparó una reacción inicial a un volumen total de 12 µl por muestra de la siguiente manera: en un tubo de PCR se agregaron, 1 µl de Oligo (dT)₁₂₋₁₈, 2 µL de RNAt, 1 µl de un mix de desoxiribonucleótidos (dNTP) y 8 µl de ALN. Las reacciones se incubaron en un termociclador (Techne TC-512 USA), a 65°C por 5 minutos,

seguido de 2 minutos a 4° C. Una vez concluida esta incubación, se agregaron a cada reacción 4 µl de 5x first strand buffer y 2 µl de ditiotreitól (DTT). Los tubos de PCR se incubaron nuevamente en el termociclador a una temperatura de 42 °C durante 2 minutos. Después se agregó 1 µl de enzima SuperScript II RT y se incubó nuevamente en el termociclador a 42 °C durante 50 minutos, finalmente se inactivó la enzima a 70 °C por 15 minutos.

9.1.2.4 Diseño de oligonucleótidos

Previo a la Reacción en Cadena de la Polimerasa (PCR), se diseñaron 4 pares de oligonucleótidos con el programa Primer 3 plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>), para lo cual se usó como plantilla la información de la secuencia publicada en Gen Bank para Bm86, (número de acceso M29321.1, cepa australiana Yeeronpilly).

Los oligonucleótidos fueron diseñados para producir 4 amplicones por muestra de DNAC. Los amplicones se diseñaron con un traslape de aproximadamente 200 pb entre ellos y un tamaño de entre 686 y 542 pb, esto, para mejorar el resultado de la secuenciación total del fragmento de ADN que codifica para la proteína Bm86 de 1824 pb (Figura 11). Los oligonucleótidos se mandaron a sintetizar a Uniparts, México, S.A. de C.V.

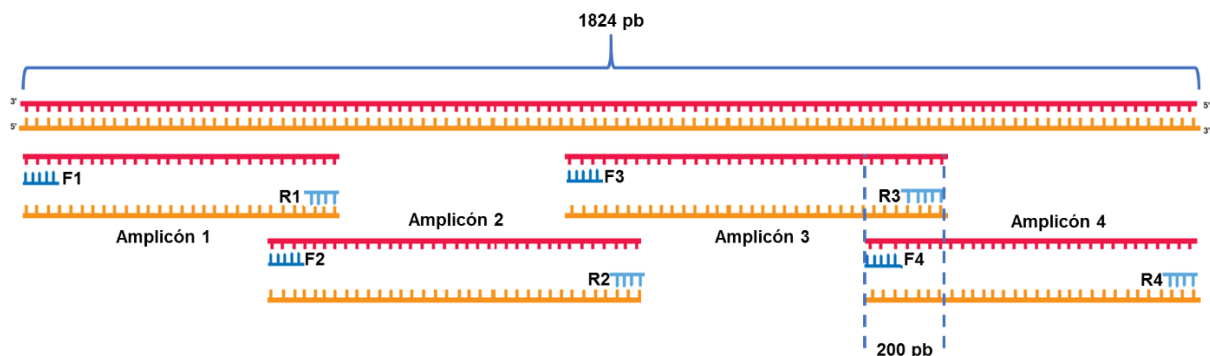


Figura 11.- Representación esquemática del diseño de los oligonucleotidos. Se puede observar una representación de la cadena completa del gen Bm86 de 1824pb y la posición de cada uno de los 4 amplicones forward y reverse en la cadena.

9.1.2.5 PCR punto final

Para establecer las condiciones de ciclado para la PCR, se tomaron en cuenta los parámetros indicados por el programa Primer3 plus en el diseño de los oligonucleótidos, además de las indicaciones del fabricante de la enzima Taq polimerasa usada en la reacción (kit GoTaq® DNA Polymerase, Promega). Los parámetros para la reacción con los que se programó el termociclador comenzaron con un primer paso de desnaturalización inicial a 94°C por 4 minutos seguidos de 30 ciclos que incluían: 1.- Temperatura de desnaturalización a 94°C por 1 minuto, 2.- Temperatura de alineamiento por 30 segundos que se ajustó para cada par de cebadores como se muestra en la Tabla 1, 3.- Temperatura de extensión a 72°C por 1 minuto. Finalmente, una temperatura de extensión final a 72°C por 5 minutos.

Tabla 1.- Condiciones de los oligonucleótidos en la reacción de PCR.

Oligo		Secuencia	Temperatura de alineamiento	Tamaño del amplicón pb	Posición en la secuencia
Bm86-1_	F	5'GTGGTTCGACGCAGTGAGAT-3'	60°/30s	680	15-34
	R	3'CCATGCTTGCAGACAAACCC-5'			675-694
Bm86-2_	F	5'CTGCAAAGACCTCTGCGAGA-3'	55°/30s	686	506-525
	R	3'TGTCCTGCGTGCAGTTAAGT-5'			1172-1191
Bm86-3_	F	5'CGGGCCCAAATGTCAACATC-3'	55°/30s	680	1006-1025
	R	3'AACGCACTCCAGCTTCTTGT-5'			1666-1685
Bm86-4_	F	5'AAAACGAGCAGTCGGAGTGT-3'	57°/30s	542	1513-1532
	R	3'GGTGTTCGATGTAAGCGTGA-5'			2035-2054

Se observan el nombre asignado a cada oligonucleótido, la secuencia, la temperatura de alineamiento usada para la reacción de PCR, el tamaño de los amplicones esperados después de la reacción y la posición en la secuencia de referencia.

9.1.2.6 Determinación de la presencia de los amplicones

Se determinó la presencia y tamaño de cada amplicón mediante electroforesis (Cámara Mudip-exu-AD140E) la corrida se hizo en gel al 1.5% de agarosa por 30 minutos a 100 Voltios. Se usó un marcador de peso molecular (GeneRuler DNA Ladders, Low range, Thermo Scientific), de 25 a 700 pb, un control negativo en el que se sustituyó el DNA por ALN durante la reacción de PCR y un control positivo conocido (Figura 12).

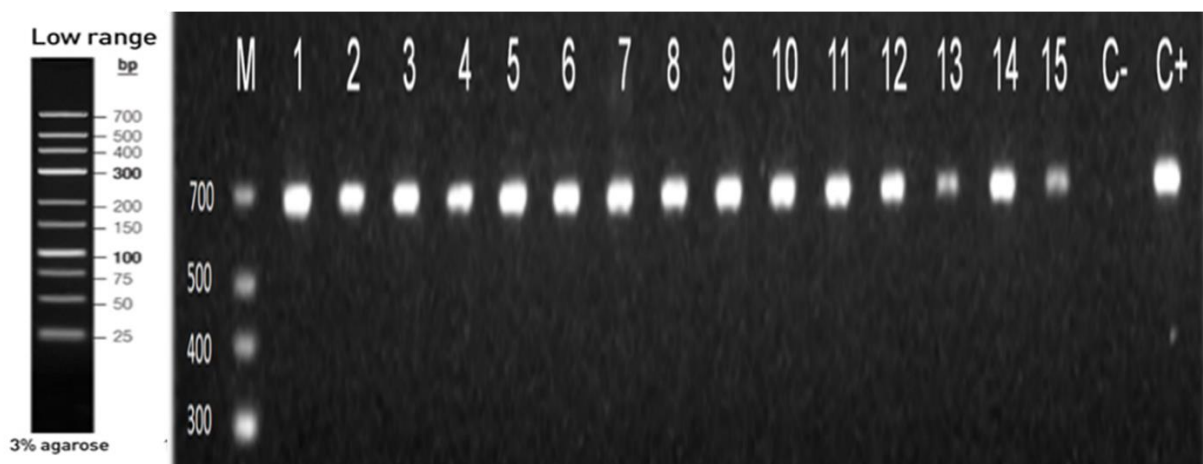


Figura 12. Electroforesis en gel de agarosa. En esta imagen se observan el tamaño esperado para los productos de la PCR correspondiente al par de oligonucleótidos Bm86-3_F y Bm86-3_R de aproximadamente 680pb. La corrida se hizo en un gel de agarosa al 1.5% teñido con bromuro de etidio. La letra M indica la posición del marcador de peso molecular de 25 a 700 pb; los números 1-15 el número identificador de cada amplicon; C-, el control negativo; C+, el control positivo (amplificado de cDNA, secuenciado y cotejado en GenBank, para demostrar que la secuencia corresponde a la reportada para el gen Bm86).

9.1.2.7 Purificación de amplicones

Con base al protocolo del kit Wizard® SV Gel y PCR Clean-up (Promega, EUA) se realizó la purificación de los productos de PCR. En una proporción de 1 a 1 se agregó la solución de unión a membrana a cada uno de los amplificadores de PCR, se transfirió la reacción a una mini-columna SV de sílice a la que el amplicón se unió, esta columna se centrifugó a 16,000 g/1 min. Posteriormente se hicieron dos lavados de la mini-columna con una solución de lavado de membrana primero con 700 µL y posteriormente con 500 µL centrifugando el primero a 16,000 g/1 min. y el segundo a 16,000 g/5 min. Finalmente, para recuperar el amplicón se centrifugó

la columna en un tubo recolector nuevo agregando 50µl de ALN y centrifugando a 16,000 g/1 min.

9.1.2.8 Secuenciación

Se determinó la concentración de los productos de PCR purificados en un espectrofotómetro (Quawell Q5000), se obtuvieron concentraciones mayores a 50ng/µl para todas las muestras. Las muestras fueron empacadas de manera individual en tubos de PCR previamente identificados para su envío a la empresa Genscript (Genscript Corporation, Piscatway, NJ, USA; www.genscript.com) donde fueron secuenciadas por el método de Sanger.

9.2 Métodos computacionales.

9.2.1 Análisis de las secuencias in silico

9.2.1.1 Ensamble de secuencias

Las cuatro secuencias de los amplicones (forward y reverse) del gen Bm86 de cada individuo fueron ensambladas con el software *CodonCode Aligner* (CodonCode Corporation, www.codoncode.com). La secuencia de la cepa de referencia Yeerongpilly (# M29321, Gen Bank), fue usada como plantilla para favorecer el ensamble de los amplicones, dicha secuencia y los cromatogramas para cada uno de los cuatro amplicones de cada individuo fueron cargadas en el software *CodonCode Aligner* y posteriormente ensambladas (Figura 13). Posteriormente se realizó la edición manual de los amplicones para eliminar los extremos con baja resolución.



Figura 13.- Ensamble de secuencias en el software CodonCode Aligner. Se puede observar tres ventanas de trabajo: (A) presenta un resumen esquemático del ensamble de las secuencias; (B) muestra la cadena de nucleótidos de las secuencias individuales y la secuencia consenso resultante del ensamble; (C) muestra los cromatogramas de cada secuencia. Las flechas verdes indican la secuencia usada como plantilla, las flechas rojas indican el amplicón #2 producto de la PCR, las flechas amarillas indican el amplicón 3, la flecha azul indica la secuencia consenso resultante.

Las secuencias consenso de cada individuo fueron recuperadas y posteriormente cargadas en el software MEGA6. La matriz de datos resultante incluía 125 secuencias de los aislados mexicanos, dichas secuencias fueron alineadas por el método de alineamiento múltiple ClustalW. Una vez alineadas, las secuencias de DNA se editaron para eliminar las regiones de las secuencias pro-péptido (región extremo NH₂) y el péptido señal (región extremo COOH), manteniendo solo el fragmento de secuencia de DNA que codifica para la glicoproteína madura Bm86.

9.2.1.2 Matriz de datos de las secuencias Bm86

Incluimos además en la matriz de datos, todos los informes de las secuencias Bm86 en GenBank de diferentes localidades del mundo hasta ahora conocidos. Para obtener todas las secuencias homólogas a Bm86, se realizó una búsqueda BLAST de la base de datos disponible en el sitio

web de NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Morgulis *et al.*, 2008; Tamura *et al.*, 2013; Zhang *et al.*, 2000). Estas secuencias habían sido aisladas a partir de garrapatas originarias del sur de Texas en los Estados Unidos de Norte América (15 secuencias), Australia, Argentina, India, Mozambique (2 secuencias), Brasil y Tailandia (29 secuencias), ver Tabla 2. La matriz de datos resultantes fue usada para los distintos análisis a continuación descritos (Figura 14).

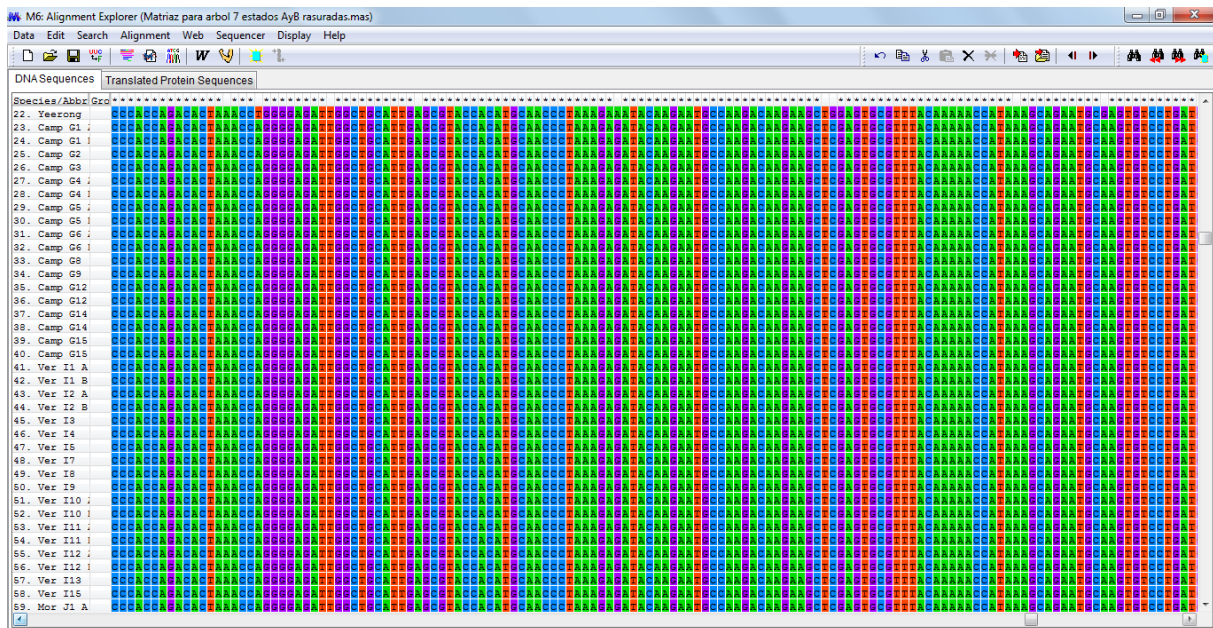


Figura 14.- Alineamiento múltiple ClustalW. Se muestra una captura de pantalla del programa Mega 6 después del alineamiento múltiple por el método ClustalW de los aislados mexicanos.

Tabla 2.- Secuencias homologas Bm86. Se muestran las secuencias reportadas en la base de datos de NCBI, GenBank usadas en la matriz de datos para los distintos análisis *in silico*.

Autores	Ciudad	Muestra	# Gen Bank	Nombre
(Rand <i>et al.</i> , 1989)	Australia		M29321.1	Yeerongpilly Bm86
(Canales <i>et al.</i> , 2009, 2008)	México		FJ456928.1	Susceptible Bm86
(García-García <i>et al.</i> , 2000)	Argentina	-	AF150891.2	Bm95
(Nijhof <i>et al.</i> , 2009)	Mozambique	-	FJ809946.1	Bm86
(Canales <i>et al.</i> , 2009)	Mozambique	-	EU191620	Bm86 (Rm)
		Zap. 11	HQ014398.1	
		Zap. 10	HQ014392.1	
		Zap. 05	HQ014391.1	
		Zap. 01	HQ014393.1	
		Zap. 03	HQ014389.1	
		Zap. 12	HQ014394.1	
(Freeman <i>et al.</i> , 2010)	Texas	Zap. 02	HQ014386.1	Bm86
		Starr 05	HQ014388.1	
		Starr 02	HQ014397.1	
		Starr 03	HQ014385.1	
		Starr 01	HQ014396.1	
		Hid 01	HQ014395.1	
		Kinne 01	HQ014401.1	
		Dimm.01	HQ014399.1	Ba86
(Andreotti <i>et al.</i> , 2008)	Brasil	Campo G.	EU352677.1	Bm86
Kumar,R., <i>et al.</i> No publicado	India	Bm	HQ166286	Bm
		S1	KJ995882	
		C1	KJ995883	
		C2	KJ995884	
		C3	KJ995885	
		N1	KJ995886	
		N2	KJ995887	
		N3	KJ995888	
		N4	KJ995889	
		N5	KJ9958890	
		N6	KJ9958891	
		N7	KJ9958892	
		N8	KJ9958893	
		N9	KJ9958894	
		N10	KJ9958895	
(Kaewmongkol <i>et al.</i> , 2015)	Tailandia	N11	KJ9958896	Bm86
		NE1	KJ9958897	
		NE2	KJ9958898	
		NE3	KJ9958899	
		NE4	KJ9958900	
		NE5	KJ9958901	
		NE6	KJ9958902	
		NE7	KJ9958903	
		NE8	KJ9958904	
		NE9	KJ9958905	
		NE10	KJ9958906	
		NE11	KJ9958907	
		NE12	KJ9958908	
		NE13	KJ9958909	
		NE14	KJ9958910	

9.2.2 Análisis de la variabilidad Genética

9.2.2.1 Diversidad del gen Bm86 en México

Para este análisis se usaron solo las 125 secuencias correspondientes a los aislados mexicanos. La variabilidad genética para los aislados de la proteína Bm86 de México, fue calculada para cada una de las poblaciones muestreadas. La variabilidad genética es una medida de la tendencia de los haplotipos dentro de cada población para diferenciarse entre ellos. Los valores de diversidad de nucleótidos indican la probabilidad de que dos sitios homólogos elegidos al azar sean diferentes, mientras que los valores de diversidad de haplotipos indican la probabilidad de que dos haplotipos elegidos al azar sean diferentes. Se calculó el número de haplotipos y valores de diversidad de nucleótidos y haplotipos para cada una de las localidades muestreadas de México con el software de polimorfismo de secuencia ADN DnaSP 5.10. (http://www.ub.edu/dnasp/DnaSP_OS.html), (Figura 15), (Librado and Rozas, 2009).

9.2.2.2 Constitución génica y frecuencia haplotípica

La constitución genética es la proporción o porcentaje de secuencias que pertenecen a cada haplogrupo. Para evaluar la constitución genética en las secuencias mexicanas de los aislados mexicanos de Bm86, primero se debe especificar los haplogrupos y la frecuencia con la que cada haplotipo aparece en un haplogrupo. Después se calculó con el software DnaSP 5.10, el grado de divergencia mantenido por cada haplotipo, con la cepa de referencia Yeroongpilly, (Figura 15), este grado de divergencia corresponde a la suma de polimorfismos de un solo nucleótido (SNP), a lo largo de cada secuencia, expresado como un porcentaje (Chen *et al.*, 2012; Zhao and Gong, 2015).

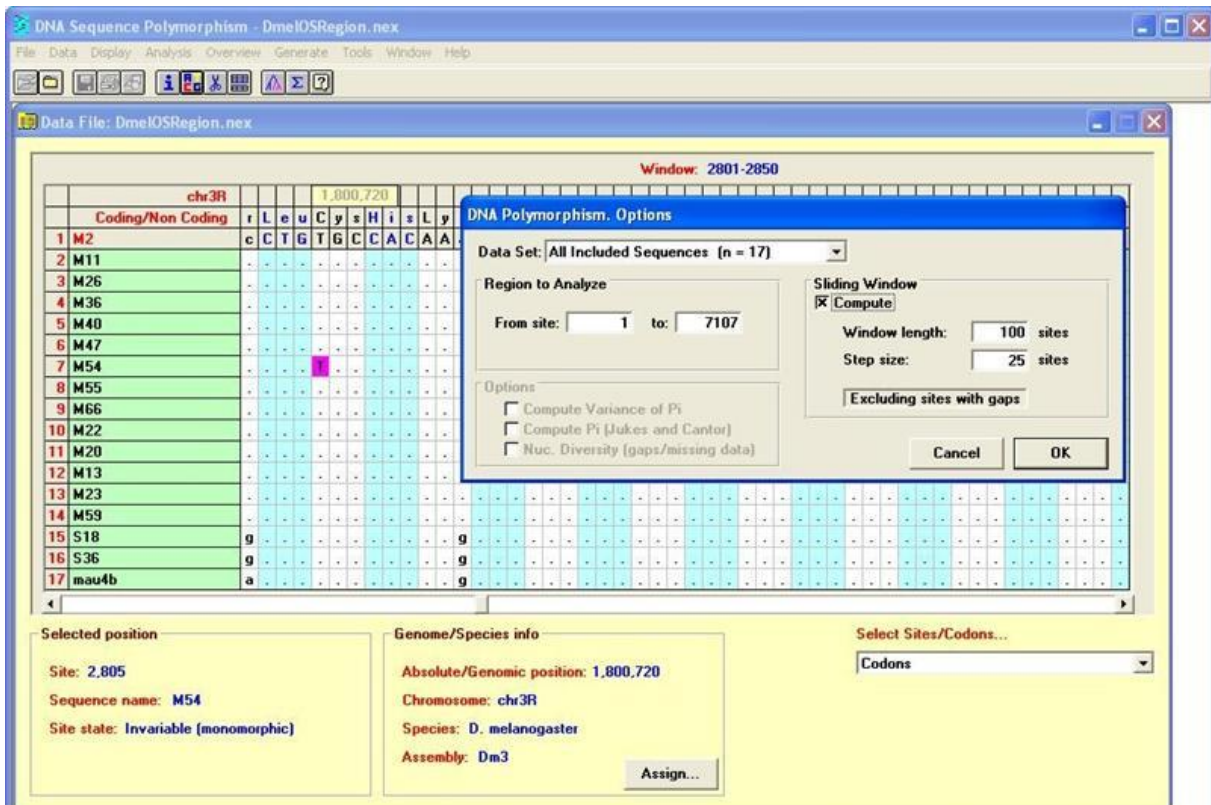


Figura 15.- Análisis de secuencias en el software DnaSP 5.10. Se muestra una captura de la ventana de trabajo después de cargar las secuencias de los aislados mexicanos Bm86 en el software DnaSP 5.10.

9.2.2.3 Prueba de Mantel

Se realizó una prueba de Mantel para explorar si existe una relación significativa entre las distancias genéticas y las distancias geográficas de cada una de las localidades que se muestrearon. El postulado es un modelo insular en el que cada población diverge al mismo tiempo. El propósito fue evaluar la cantidad de estructura espacial, específicamente si los haplotipos observados estaban confinados a una localidad o región geográfica en México o si se distribuían entre las localidades muestreadas.

Para la matriz de distancias genéticas, se utilizó un valor promedio de los porcentajes de identidad para todos los posibles pares de población, estos valores se calcularon con el software *MatGAT v2.01* (herramienta Matrix Global Alignment Tool), (Campanella *et al.*, 2003)

(www.angelfire.com/nj2/arabidopsis/MatGAT.html). El análisis de identidad incluyó, además de nuestros aislamientos los informes Bm86 en GenBank de diferentes localidades del mundo hasta ahora disponibles.

La matriz de distancias geográficas se construyó con el software *Google Earth* (<https://www.google.com/earth/download/ge/>), esta matriz representa la distancia en kilómetros que existe entre cada par de poblaciones muestreadas. Las matrices de datos se analizaron utilizando el software estadístico *XLSTAT* (<https://www.xlstat.com/es/>). Se utilizó la función de correlación de Pearson, que determina si existe una relación lineal entre dos variables. Los coeficientes de correlación entre los elementos de cada matriz se obtuvieron utilizando 10000 permutaciones y una significación del 95%.

9.2.3 Análisis filogenético

9.2.3.1 Algoritmo neighbor-joining

Para el análisis filogenético del gen Bm86 de México, se consideraron las 125 secuencias de las garrapatas recolectadas en 10 distintas localidades de México y las secuencias completas recuperadas de GenBank de todos los países arriba mencionados. Para ambos casos la cadena que codifica para la proteína madura, residuos 20-627 fue utilizada. El método para reconstruir los datos de distancias evolutivas entre las secuencias del gen Bm86 de México se generó utilizando el algoritmo *neighbor-joining* (Saitou and Nei, 1987). Este método se basa en agrupar pares de unidades taxonómicas operativas (Operational Taxonomic Units, por sus siglas en inglés. OTUs [= neighbors]), tratando siempre de minimizar la longitud total del árbol en cada etapa de agrupamiento de OTU (Bandelt *et al.*, 1999; Saitou and Nei, 1987). El análisis es básicamente una matriz de distancias genéticas que refleja el grado de similitud, sin embargo, el método genera árboles sin raíces. Para resolver esto, se incluyeron dos secuencias de

exogrupo para el análisis filogenético, Dimmi y Kinne de *R. annulatus* obtenidas de Genbank (HQ014399 y HQ014401 respectivamente), (Freeman *et al.*, 2010). Las distancias evolutivas se calcularon usando el método Kimura de 2 parámetros (Kimura, 1980), basado en el número de sustituciones de bases por sitio. Las posiciones del codón incluidas fueron 1° + 2° + 3° + no code (la 3ª posición de los codones acumula mutaciones mucho más rápidamente que la 2ª y 1ª). La variación de la tasa entre los sitios se modeló con una distribución gamma “G” (parámetro de forma = 1). El árbol se evaluó mediante *bootstrap* a 1000 permutaciones en el programa Mega 6 (Tamura *et al.*, 2013).

9.2.3.2 Red de expansión mínima

Se construyó también, un árbol de expansión mínima (MST), también conocida como red de expansión mínima, utilizando el software Network Publisher (<http://www.fluxus-engineering.com>). Para este análisis utilizamos el total de los haplotipos observados y los informes para el resto del mundo en GenBank. La coalescencia y las conexiones entre los haplotipos dada por el número mínimo de mutaciones entre ellos son la base del análisis para generar un árbol o red que muestra la frecuencia y la relación con las distancias en el número de mutaciones que separan cada haplotipo. El método *Median Joining* (MJ) fue elegido para la construcción de la red, este método combina características de los algoritmos de *Kruskal* y *Farris*, el primero favorece las conexiones cortas en la red para encontrar árboles de expansión mínima, mientras que el segundo es un algoritmo heurístico de *parsimonia máxima* (MP) que agrega los llamados vectores de mediana (Bandelt *et al.*, 1999).

9.2.4. Diseño de un modelo predictivo para epítomos de células B

9.2.4.1 Conjunto de datos

Descargamos todos los epítomos lineales de células B reportados en la base de datos IEDB (<http://www.iedb.org/>). El número de péptidos recuperados fue 83,683 que incluyo: 2,374 epítomos clasificados inmunogénicamente como positivos-altos, 76,310 clasificados como positivos, 5,325 epítomos clasificados como positivos-bajos, y 750 clasificados como positivos-intermedios. Para fines de clasificación, agrupamos todos estos epítomos en dos clases, de acuerdo con la intensidad de la respuesta inmunogénica relacionada con este epítomo (ϵ). Estas dos clases son epítomos positivos altos ($\epsilon = 1$) y epítomos de positivos a intermedios ($\epsilon = 0$). Los ensayos biológicos utilizados para determinar la intensidad de la respuesta inmunogénica se han llevado a cabo en condiciones experimentales diferentes c_j (ver las siguientes secciones). En consecuencia, se usó a partir de aquí ϵ como una función de estas condiciones $\epsilon(c_j)$.

9.2.4.2 Definición del modelo

Nuestro método se basa en la comparación de toda la información sobre la secuencia, las condiciones de ensayo e intensidad de la respuesta inmunogénica $\epsilon_q(c_j)$ para un péptido de consulta (q-péptido) con respecto a un péptido de referencia (péptido-r). En nuestro conjunto de datos, representamos cada péptido de consulta (q-péptido) como un vector \mathbf{c}_{qj} . Los componentes del vector cuantifican la información sobre diferentes factores experimentales o condiciones del ensayo $\mathbf{c}_{qj} = (c_{q1}, c_{q2}, c_{q3}, c_{q4}, c_{q5}, c_{q6}, c_{q7})$. Estas variables son, por ejemplo: $c_{q1} = \epsilon_q(c_j)$ la actividad del epítomo, $c_{q2} =$ Nueva secuencia, $c_{q3} =$ Adyuvante, etc. El péptido-r tiene una secuencia similar a la secuencia del péptido-q, pero se ha medido en las mismas o diferentes condiciones de ensayo. El vector para el péptido-r es similar a $\mathbf{c}_{rj} = (c_{r1}, c_{r2}, c_{r3}, c_{r4}, c_{r5}, c_{r6}, c_{r7})$. Los elementos son, por ejemplo: $c_{r1} = \epsilon_r(c_j)$ la actividad del péptido-r, $c_{r2} =$ información de la

secuencia del péptido-r, c_{r3} = adyuvante, etc. La fórmula general del tipo de modelo propuesto aquí es la siguiente:

$${}^iS_{qr}(j) = e_0 + e_1 \cdot \varepsilon_r + \sum_{j=0}^{j=5} a_k \cdot ({}^q\theta_j - {}^r\theta_j) \quad (1)$$

En este modelo, S_{qr} es el valor de puntuación para la actividad como epítipo del q-péptido. El coeficiente e_1 es el coeficiente de la secuencia-r, y e_0 es el término independiente. Los a_j son los coeficientes de las variables de entrada. La primera variable de entrada del modelo $c_{r1} = \varepsilon_r(c_j)$ es la actividad del epítipo del péptido-r; usado aquí como plantilla. El segundo tipo de condición variable o experimental $\Delta\theta_k = {}^q\theta_k - {}^r\theta_k$ cuantifica información sobre la variación o perturbación en la secuencia (estructura química) del péptido de consulta con respecto al péptido plantilla. Las variables ${}^q\theta_k$ y ${}^r\theta_k$ son los índices de información de entropía de Shannon sobre la estructura del péptido de consulta y el péptido de referencia. Los últimos tipos de variables o condiciones experimentales cuantifican la información sobre los factores no estructurales que condicionan la respuesta inmune. Estas últimas variables también son índices de información en valores de las entropías de Shannon.

9.2.4.3 Índices de información de la secuencia de los péptidos

Calculamos las medidas de información de entropía de Shannon de la secuencia de cada péptido con el fin de cuantificar la información estructural relevante. En primer lugar, utilizamos el software S2SNet para transformar cada secuencia peptídica en una *Sequence Recurrence Networks* (SRN) o Red de recurrencia de secuencia (Kowalski *et al.*, 1982; Munteanu *et al.*, 2013, 2009a, 2008). El SRN de una secuencia peptídica se define como una red donde los aminoácidos (residuos r_{aa} y r_{bb} en la secuencia) son los vértices/nodos y el par de nodos (nodos

n_{aa} y n_{bb} en la red) están conectados por un enlace ($\alpha_{ij} = 1$) si son nodos recurrentes adyacentes y/o vecinos. Esto significa que $\alpha_{ij} = 1$ si la distancia topológica es $d = 1$ (aminoácidos químicamente unidos) o que son los aminoácidos vecinos más cercanos del mismo tipo (A, R, N, D, C, Q, E, G, H, I, L, K, M, F, P, S, T, W, Y, V), (Ver Tabla 3), con $d = \min(d)$. El primer nodo en la secuencia (centro del gráfico estelar) es un vértice o vértice ficticio sin residuos.

Tabla 3.- Características de los aminoácidos.

Aminoácido	Abreviatura	Letra	Grupo	Tipo
Glicina	GLI	G	Hidrofóbicos	Neutros apolares
Alanina	ALA	A		
Valina	VAL	V		
Leucina	LEU	L		
Isoleucina	ILE	I		
Metionina	MET	M		
Fenilalanina	PHE	F	Hidrofílicos	Neutros aromáticos
Tirosina	TYR	Y		
Triptófano	TRP	W		
Serina	SER	S		Muy Polares
Cisteína	CYS	C		
Prolina	PRO	P		
Treonina	TRE	T		
Asparagina	ASG	N		
Glutamina	GLN	Q		
Ácido aspártico	ASP	D		
Ácido glutámico	GLU	E		
Lisina	LYS	K		
Histidina	HIS	H		
Arginina	ARG	R		

Se muestra la lista de los 20 aminoácidos que forman parte de las proteínas, con algunas de sus características fisicoquímicas.

En segundo lugar, necesitamos transformar la SRN de cada secuencia en una matriz estocástica ${}^1\Pi$. Los elementos de ${}^1\Pi$ son las probabilidades ${}^1p_{ij}$ de alcanzar una base o nodo n_i con peso w_j desde otro nodo n_j con peso w_j si el observador se mueve a través de una caminata de longitud $d_{ij} = k = 1$ (Aguero-Chapin *et al.*, 2009; González-Díaz *et al.*, 2007).

$$p_{ij} = \frac{w_j}{\sum_{n=1}^{n=L} \alpha_{ij} \cdot w_j} \quad (2)$$

En ambas ecuaciones, (1) y (2), $\alpha_{ij} = 1$ si los nodos n_i y n_j son adyacentes en el gráfico SRN o de otro modo igual a $\alpha_{ij} = 0$; de acuerdo con la función condicional anterior. El número de bases en las secuencias es igual al número de nodos (n) en el gráfico SRN y también es igual al número de filas y columnas en ${}^1\Pi$, la longitud de la secuencia (L) y la máxima distancia topológica en la secuencia max. (d_{ij}).

Se cuantificó el contenido de información de una secuencia de DNA de consulta (secuencia-q) con los valores de la entropía de Shannon (${}^q\theta_k$). El mismo procedimiento se usó para cuantificar la información (${}^r\theta_k$) para secuencias de referencia (secuencias-r) usadas como plantilla. Los valores de entropía ${}^q\theta_k$ o ${}^r\theta_k$ se han calculado a partir de las potencias naturales k-ésimas de la matriz de Markov ${}^1\Pi$:

$${}^q\theta_k(seq) = - \sum_{aa=0}^{aa=L} {}^k p_{aa} \cdot \log({}^k p_{aa}) \quad (3)$$

Donde, p_j son los elementos n_i del vector p , resultan de la multiplicación de la matriz potenciada de Markov ($n_i \times n_i$) y de un vector ($n_i \times 1$) con cada elemento igual a $1 / n_i$; $n = 0$ - límite de potencia, M = matriz de conectividad (dimensión i^*i); $ii = i^{\text{th}}$ elemento diagonal; d_{ij} son los elementos de la matriz de distancia y m_{ij} son los elementos de la matriz de conectividad M ; deg_i son los elementos de la matriz de grados.

9.2.4.4 Índices de información de condiciones experimentales

También calculamos los índices de información de entropía de Shannon $\theta_0(c_j)$ para cuantificar información útil sobre las otras condiciones experimentales (c_3, c_4, c_5, c_6, c_7). Estas variables son: $\theta_0(\text{Adj})$, $\theta_0(\text{Org})$, $\theta_0(\text{Host})$, $\theta_0(\text{Proc})$ y $\theta_0(\text{Tech})$. Las condiciones experimentales codificadas numéricamente por estas variables son: $c_3 = \text{Adj} = \text{Aditivos adyuvantes} = \text{Adyuvante usado en la vacuna junto al péptido}$, $c_4 = \text{Org} = \text{Nombre del organismo} = \text{Organismo fuente del péptido}$, $c_5 = \text{Host} = \text{Organismo huésped} = \text{Organismo huésped expuesto al antígeno}$, $c_6 = \text{Proc} = \text{Tipo de proceso } in vivo = \text{Proceso inmunológico de sensibilización}$, y $c_7 = \text{Tech} = \text{Técnicas experimentales} = \text{Método o técnica utilizada para determinar la inmunogenicidad de}$

los péptidos. Todas estas variables se calcularon como una entropía de Shannon de la probabilidad de ocurrencia de cada condición en el conjunto de datos. La probabilidad de ocurrencia p_j de cada condición c_j se determinó como una relación de la frecuencia relativa n_j de ocurrencia con respecto al número total n_{total} de entradas en el conjunto de datos original, ver la siguiente ecuación.

$${}^q\theta_0(c_j) = -\sum_{j=0}^{j=L} p_j \cdot \log(p_j) = -\sum_{j=0}^{j=L} \left(\frac{n_j}{n}\right) \cdot \log\left(\frac{n_j}{n}\right) \quad (4)$$

	A	B	C	D
1	Conteo total	84759		
2	Conjugate/Adjuvante	Conteo	Probabilidad	Sh0(Adj)
3	Función	-	'=B4/B1	'= -Logartimo (C)
4	FREE	82466	0.972946826	0.011910895
5	GLYC(N28)	3	3.53945E-05	4.451064569
6	MCM(C12)	1	1.17982E-05	4.928185824
7	METH(K4,K9,K10,K11)	1	1.17982E-05	4.928185824
8	SCM(E1)	1	1.17982E-05	4.928185824
9	GLYC(N20)	2	2.35963E-05	4.627155829
10	PYRE(E1)	30	0.000353945	3.451064569
11

Figura 16. Ejemplo de hoja de cálculo de los índices de información. Se puede observar la captura de pantalla de Excel para el cálculo de los índices de información, en este caso de $c3 = \text{Adj} = \text{Aditivos adyuvantes}$. En la celda B1, color naranja se observa el conteo total de todos los posibles adyuvantes en la base de datos. En la fila 3 se puede observar la función usada para calcular cada concepto. Los índices de información de entropía de Shannon $\theta_0(c_j)$, para la variable “Adj” se muestran en la columna D.

9.2.4.5 Entrenamiento y validación del modelo

Una vez que calculamos las variables de entrada, era necesario establecer y probar una hipótesis de modelo. Un modelo lineal fue seleccionado como la primera aproximación. Se eligió el análisis lineal discriminante (LDA) para encontrar la combinación lineal de variables independientes que mejor diferencian (discriminan) a ambos grupos (Beleites *et al.*, 2013; Min *et al.*, 2016). Al hacerlo, tenemos que crear una matriz de datos en bruto. Los datos se ordenaron

de la siguiente manera: se muestrearon 1,048,190 de pares de secuencias de la base de datos total de 83,683 secuencias. Cada par de secuencias contiene una secuencia de consulta (secuencia-q) y una secuencia de referencia (secuencia-r).

Estas secuencias fueron organizadas tanto numéricamente como alfabéticamente. El ordenamiento numérico fue de acuerdo con los valores de ϵ_{q1} de 1 a 0. La clasificación alfabética fue de A a Z de acuerdo con las letras de los aminoácidos en la secuencia (código de una letra). Se tomaron muestras de un total de 1,048,190 pares de secuencias al azar de acuerdo con el siguiente esquema. La etiqueta de la secuencia de consulta (n-query) se generó al azar entre 1 y 83683. La etiqueta de la secuencia de plantilla (n-template) se generó como un número aleatorio entre n-query y n-query + un número aleatorio entre 0 y 5. Este esquema de muestreo garantiza que tenemos pares de péptidos de consulta y de referencia con gran similitud en el conjunto de datos.

A	B	C	E	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V		
n-query	n-template	Eij	query	Seq-query	Seq-template	Val	y-templar	ShS(Seq&Rec)	query	ShQ(host)-quen	ShQ(Adju)-quen	ShQ(InviProT1)-quen	ShQ(EpOrgNa)-quen	ShQ(MTech)-quen	ShQ(MTech)-templar	DShS(Seq&Rec)	DShQ(host)	DShQ(Adju)	DShQ(InviProT1)	DShQ(EpOrgNa)	DShQ(MTech)
2	1	1	1	AAAIALRGMKADGK	AAAIALRGMKADGK	t	1	2.64	0.44	0.01	0.85	2.73	1.28	1.28	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	1	2	1	AAAIALRGMKADGK	AAAIALRGMKADGK	t	1	2.64	0.44	0.01	0.85	2.73	1.28	1.28	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	1	3	1	AAAIALRGMKADGK	AAEAFVLHVFEDAYL	t	1	2.64	0.44	0.01	0.85	2.73	1.28	1.28	-0.08	-2.35	0.00	-0.28	1.94	0.00	0.00
5	1	6	1	AAAIALRGMKADGK	AALCFGRAPYNKNE	v	1	2.64	0.44	0.01	0.85	2.73	1.28	0.85	-0.01	-1.59	0.00	0.63	0.17	0.42	0.00
6	1	5	1	AAAIALRGMKADGK	AAGLPGKCGVNVPIK	t	1	2.64	0.44	0.01	0.85	2.73	1.28	0.85	-0.05	-0.68	0.00	0.63	-0.62	0.42	0.00
7	1	4	1	AAAIALRGMKADGK	AAFYKTKTV	t	1	2.64	0.44	0.01	0.85	2.73	1.28	0.28	0.30	0.00	0.00	-0.28	1.94	1.00	0.00
8	2	3	1	AAAIALRGMKADGK	AAEAFVLHVFEDAYL	t	1	2.64	0.44	0.01	0.85	2.73	1.28	1.28	-0.08	-2.35	0.00	-0.28	1.94	0.00	0.00
9	2	4	1	AAAIALRGMKADGK	AAFYKTKTV	v	1	2.64	0.44	0.01	0.85	2.73	1.28	0.28	0.30	0.00	0.00	-0.28	1.94	1.00	0.00
10	2	5	1	AAAIALRGMKADGK	AAGLPGKCGVNVPIK	t	1	2.64	0.44	0.01	0.85	2.73	1.28	0.85	-0.05	-0.68	0.00	0.63	-0.62	0.42	0.00
11	2	2	1	AAAIALRGMKADGK	AAAIALRGMKADGK	t	1	2.64	0.44	0.01	0.85	2.73	1.28	1.28	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12	2	7	1	AAAIALRGMKADGK	AALCFGRAPYNKNE	t	1	2.64	0.44	0.01	0.85	2.73	1.28	0.85	-0.01	-0.77	0.00	0.63	0.17	0.42	0.00
13	2	6	1	AAAIALRGMKADGK	AALCFGRAPYNKNE	v	1	2.64	0.44	0.01	0.85	2.73	1.28	0.85	-0.01	-1.59	0.00	0.63	0.17	0.42	0.00
14	3	5	1	AAEAFVLHVFEDAYL	AAGLPGKCGVNVPIK	t	1	2.71	2.79	0.01	1.13	0.78	1.28	0.85	0.02	1.67	0.00	0.91	-2.57	0.42	0.00
15	3	6	1	AAEAFVLHVFEDAYL	AALCFGRAPYNKNE	t	1	2.71	2.79	0.01	1.13	0.78	1.28	0.85	0.06	0.76	0.00	0.91	-1.78	0.42	0.00
16	3	8	1	AAEAFVLHVFEDAYL	AALFYTHFNSSGCP	t	1	2.71	2.79	0.01	1.13	0.78	1.28	1.34	0.09	2.11	0.00	0.91	-1.18	-0.06	0.00
17	3	7	1	AAEAFVLHVFEDAYL	AALCFGRAPYNKNE	v	1	2.71	2.79	0.01	1.13	0.78	1.28	0.85	0.06	1.58	0.00	0.91	-1.78	0.42	0.00
18	3	3	1	AAEAFVLHVFEDAYL	AAEAFVLHVFEDAYL	t	1	2.71	2.79	0.01	1.13	0.78	1.28	1.28	0.00	0.00	0.00	0.00	0.00	0.00	0.00
19	3	4	1	AAEAFVLHVFEDAYL	AAFYKTKTV	t	1	2.71	2.79	0.01	1.13	0.78	1.28	0.28	0.37	2.35	0.00	0.00	0.00	0.00	1.00

Figura 17. Matriz por pares para la validación y entrenamiento del modelo. Se puede observar los pares de secuencias generadas al azar. En las columnas A y E, se observan los números identificadores para los epítomos de consulta (n-query) y en las columnas B y G las secuencias usadas como plantilla (n-template). Las columnas C e I, muestran la clasificación E_{ij} para la actividad de los epítomos. Además, en los pares de columnas: J y Q; K y R; L y S; M y T; N y U; O y V, n-query y n-template respectivamente, se pueden observar los índices de información de entropía de Shannon $\theta_0(c_j)$ calculados para cada condición experimental. Los índices de información de Shannon, cuantifican información de la medida de la distribución aleatoria de cada microsistema en un macrosistema.

El remanente 548,190 pares de secuencias se seleccionaron totalmente al azar para hacer un total de 1,048,190 pares de casos (este es el límite de nuestras capacidades de procesamiento actuales). Posteriormente, utilizamos LDA implementado en el software STATISTICA para buscar el modelo de clasificación (Aguero-Chapin *et al.*, 2009; Ran *et al.*, 2016; Xiao and Shen, 2015). La secuencia-r es la secuencia de referencia con una clasificación conocida. Tanto la actividad del péptido-q como el péptido-r tienen la misma codificación. Los valores $\varepsilon_{q1} = 1$ y/o $\varepsilon_{r1} = 1$ son para q-péptidos y/o p-péptidos con alta actividad de epítipo positivo en IEDB. Los valores $\varepsilon_{q1} = 0$ y/o $\varepsilon_{r1} = 0$ son para q-péptidos y/o p-péptidos con actividad igual a positivo, medio positivo o bajo positivo en el IEDB. La secuencia-q es la secuencia de consulta que el modelo clasificará como antigua ($\varepsilon_q = 1$) o nueva ($\varepsilon_q = 0$). También calculamos los valores de $\Delta\theta_j = ({}^q\theta_j - {}^r\theta_j)$ para todos estos 548,190 pares de secuencias. La cantidad de casos utilizados para entrenar el modelo incluye 83683 péptidos. Utilizamos el conjunto recuperado de la base de datos IEDB: incluyendo 1448 nombres de organismos origen de los epítipos, 323 nombres de hospederos, 15 tipos de procesos *in vivo*, 28 técnicas experimentales y 505 posibles aditivos o adyuvantes. los datos se organizaron en un archivo de Excel, consulte la Figura 13.

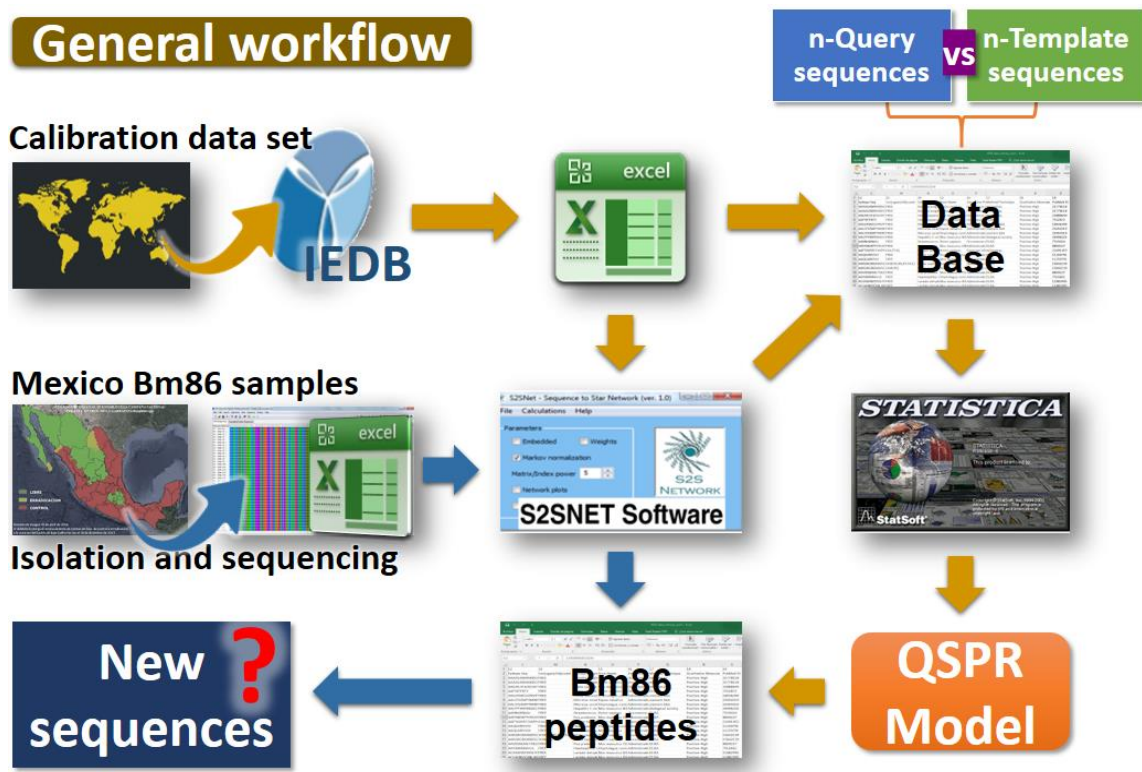


Figura 18.- Flujo de trabajo general del modelo. Conjunto de datos de calibración (flechas amarillas) conjunto de secuencias Bm86 (flechas azules).

9.2.4.6 Estudio predictivo computacional

El conjunto de secuencias Bm86 de DNA de México, se utilizó para ilustrar el funcionamiento del modelo en este trabajo. Estas secuencias de DNA se tradujeron *in silico* (Software Mega 6) a su secuencia proteica correspondiente (cadena aminoacídica 20-627) que codifica para la proteína madura. Usamos el nuevo modelo para predecir péptidos inmunogénicos en las secuencias mexicanas, ya que estas muestran cambios en la cadena de aminoácidos cuando se alinean. En un primer paso, definimos las secuencias de epítomos antigénicos probables de la siguiente manera. Seleccionamos todos los posibles fragmentos de péptidos con 15 a 35 aminoácidos, considerando todos los SAP's presentes en la cadena. La búsqueda encontró 26112 péptidos no redundantes. Usando S2SNet transformamos todas las secuencias de estos péptidos en SRN y calculamos los índices de información $q\theta^k$ (Seq). Por último, el modelo se usó para predecir péptidos inmunogénicos bajo las condiciones de frontera del ensayo. Las

condiciones seleccionadas para el péptido de consulta son: c_{q2} = Nueva secuencia; c_{q3} = Aditivos adyuvantes (González-Díaz *et al.*, 2014; Odongo *et al.*, 2007; Vázquez-Prieto *et al.*, 2016); c_{q4} = nombre del organismo del epítipo = *Rhipicephalus decoloratus*; c_{q5} = Nombre de host = *Bos indicus*; c_{q6} = Tipo de proceso *in vivo* = Administración *in vivo* y c_{q7} = Técnicas experimentales = ELISA. También seleccionamos un péptido de referencia con $c_{r1} = \epsilon_r = 1$ (el péptido de referencia tiene actividad alta positiva) de la literatura.

9.2.5 Determinación de regiones antigénicas consenso

El siguiente paso en el estudio fue consensar los epítipos lineares de células B predichos en este trabajo y los epítipos predichos en 2 trabajos recientes realizados por nuestro grupo de trabajo. Sánchez Bermúdez predijo en su estudio 81 péptidos no redundantes en 11 secuencias originarias del estado de Colima y la secuencia de referencia Yeerongpilly: 2 de ellos fueron predichos por el método SVMTriP, 32 por APP y 47 por BepiPred (Sánchez Bermúdez, 2017). En el segundo trabajo, a partir 10 secuencias del estado de Zacatecas y la secuencia de referencia Yeerongpilly, Sotelo Gómez predijo 21 péptidos con tendencia a ser antigénicas por el método Bepipred. (Sotelo Gómez, 2017).

Se alinearon todos los epítipos predichos para cada modelo en una matriz de datos. Es importante mencionar que solo se tomaron en cuenta aquellos péptidos que contenían una o varias mutaciones con respecto a la cepa de referencia Yeerongpilly, además, en muchos de los casos los péptidos se traslapaban por lo que se determinó tomar la longitud de dichas secuencias como una sola región antigénica a considerar. Las secuencias consenso se construyeron a partir de esta matriz, procurando incluir todas las regiones predichas como antigénicas por la predicción de epítipos lineares de células B en una sola secuencia.

9.2.6 Contratación de los resultados con otros estudios

Las secuencias consenso que contenían todas las regiones predichas como antigénicas, se compararon con los resultados de cuatro trabajos independientes en otras regiones geográficas del mundo. En los primeros dos trabajos buscaron epítomos de células B definidos por Ac después de la vacunación. El tercer caso consistió en un estudio de predicción de péptidos antigénicos por el método Kolaskar y Tongaonkar. En el último estudio el objetivo fue predecir regiones hidrofílicas mediante el algoritmo de Hopp & Woods.

10 RESULTADOS

10.1 Artículo publicado.

10.1.1 Resumen

En este trabajo, desarrollamos un método de la teoría general de las perturbaciones y el aprendizaje automático (PTML) para la extracción de datos de proteomas que nos permiten predecir cambios en la actividad de un epítipo de células B, después de múltiples variaciones o perturbaciones, estas perturbaciones pueden ocurrir por la variación en la secuencia de un péptido conocido y/o variaciones en las condiciones experimentales del ensayo o condiciones de contorno. El objetivo principal se centra en descubrir nuevos epítipos de células B útiles para el diseño de vacunas. El método predice la actividad del epítipo $\varepsilon_q(c_{qj})$ de un péptido de consulta (q-péptido) en un conjunto de condiciones experimentales de consulta (c_{qj}). El método usa como entrada la secuencia del q-péptido, además, también utiliza como información de entrada, la secuencia y la actividad del epítipo $\varepsilon_r(c_{rj})$ de un péptido de referencia (péptido-r) ensayada en condiciones experimentales similares (c_{rj}). El modelo propuesto aquí es capaz de clasificar 1,048,190 pares de secuencias de péptidos de consulta y referencia del proteoma de muchos organismos informados en la base de datos IEDB. Estos pares tienen variaciones (perturbaciones) en su secuencia o en las condiciones de ensayo. El modelo resultante para la serie de entrenamiento tiene una precisión de 73%, sensibilidad de 80.1 y especificidad de 70.6. Para la serie de validación tiene una precisión de 73%, sensibilidad de 79.7 y especificidad de 70.7. La información recuperada contiene cambios estructurales en 83683 secuencias de péptidos (Seq) determinadas en condiciones de ensayo experimentales o condiciones de contorno que implican 1448 Organismo fuente del péptido (Org), 323 Organismo huésped expuesto al antígeno (Host), 15 tipos de proceso inmunológico de sensibilización o proceso *in*

vivo (Proc), 28 Métodos o técnicas utilizada para determinar la inmunogenicidad de los péptidos o Técnicas experimentales (Tech) y 505 aditivos o Adyuvantes usados en la vacuna junto al péptido (Adj). También, informamos el muestreo aislamiento y secuenciación de 15 muestras completas del gen Bm86 del estado de Colima, México. Por último, utilizamos el modelo para predecir los puntajes inmunogénicos de un epítipo en diferentes condiciones experimentales para los 26112 péptidos no redundantes obtenidos a partir de las 15 secuencias Bm86 que se muestrearon en Colima, México. El modelo puede convertirse en una herramienta útil para la selección de epítipos hacia el diseño de una vacuna. Los resultados teórico-experimentales en la proteína Bm86 puede ayudar en el diseño futuro de una nueva vacuna basada en esta proteína.

Fw: Decision on Manuscript ID pr-2017-00477x.R1



S. Gabriel Martínez Arzate

mar 19/09/2017 08:39 a.m.

Para: Gabo Martínez (gabosaeta_@hotmail.com); jvch@yahoo.com; Esvieta Tenorio (esvieta@gmail.com) ↗



Responder | v

De: Journal of Proteome Research <onbehalfof+jlabaer+jpr.acs.org@manuscriptcentral.com>

Para: "gonzalezdiaz@yahoo.es" <gonzalezdiaz@yahoo.es>; "humberto.gonzalezdiaz@ehu.es" <humberto.gonzalezdiaz@ehu.es>

Enviado: martes, 19 de septiembre de 2017 0:14:16 CEST

Asunto: Decision on Manuscript ID pr-2017-00477x.R1

18-Sep-2017

Journal: Journal of Proteome Research

Manuscript ID: pr-2017-00477x.R1

Title: "PTML Model for Proteome Mining of B-cell Epitopes and Theoretic-Experimental Study of Bm86 Protein Sequences from Colima Mexico"

Author(s): Martínez, Saúl; Tenorio-Borroto, Esvieta; Barbabosa Pliego, Alberto; Diaz-Albiter, Hector; Vazquez-Chagoyan, Juan; Gonzalez-Diaz, Humbert

Dear Dr. Gonzalez-Diaz:

We are pleased to inform you that your manuscript has been accepted for publication in the Journal of Proteome Research. Your manuscript has been forwarded to the ACS Publications office. You will be contacted by Kelly Helton – khelton@acs.org, ACS Publications office, once the galley proof is ready for you to review. Please return your approval for galley proof promptly so as not to delay the publication of your manuscript. Mistakes not corrected at the galley proof stage can only be corrected through the submission of an "Addition/Correction" for approval by the Editor-in-Chief.

Note: An ACS ID is required to access proof materials, if you do not have an ACS ID you will be prompted to create one, or you may go to <http://www.acschemworx.org> to register.

(JAMs) Just Accepted Manuscript - upon acceptance, it will be posted within 24 hours as a PDF document accompanied by an HTML abstract. A DOI will be assigned for citations by readers and granting agencies. This is the unedited version, therefore any request for changes or corrections will not be made until the galley proof is returned, and major changes require the Editor-in-Chief's approval. The JAM version will have a Web-posting date as the official publication date. This version will be removed upon publication of the ASAP article.

(ASAP) As Soon As Publishable - once the author approves and returns the galley proof, the ASAP article will be posted within 48 hours, which has been technically edited and formatted, and represents the final scientific article of record.

** The DOI remains unchanged throughout for citations.

Thank you again for choosing the Journal of Proteome Research as a forum for the publication of your work.

With sincere regards,

Joshua LaBaer
Associate Editor
Journal of Proteome Research
Phone: 480-727-6995
Fax: 202-354-5449
Email: jlabaer@jpr.acs.org

Confirmation of your selection for "Just Accepted" publication:

YES, I agree that my accepted manuscript may be published on the ACS Web site as a "Just Accepted" manuscript.

PLEASE NOTE: This email message, including any attachments, contains confidential information related to peer review and is intended solely for the personal use of the recipient(s) named above. No part of this communication or any related attachments may be shared with or disclosed to any third party or organization without the explicit prior written consent of the journal Editor and ACS. If the reader of this message is not the intended recipient or is not responsible for delivering it to the intended recipient, you have received this communication in error. Please notify the sender immediately by e-mail, and delete the message. Thank you.

PTML Model for Proteome Mining of B-cell Epitopes and Theoretic- Experimental Study of Bm86 Protein Sequences from Colima Mexico

Saúl Gabriel Martínez, Esvieta Tenorio-Borroto, Alberto Barbabosa Pliego,
Hector Diaz-Albiter, Juan C. Vazquez-Chagoyan, and Humbert Gonzalez-Diaz

J. Proteome Res., **Just Accepted Manuscript** • DOI: 10.1021/acs.jproteome.7b00477 • Publication Date (Web): 18 Sep 2017

Downloaded from <http://pubs.acs.org> on September 19, 2017

Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.



ACS Publications

Journal of Proteome Research is published by the American Chemical Society, 1155
Sixteenth Street N.W., Washington, DC 20036
Published by American Chemical Society. Copyright © American Chemical Society.
However, no copyright claim is made to original U.S. Government works, or works
produced by employees of any Commonwealth realm Crown government in the course
of their duties.

**PTML Model for Proteome Mining of B-cell Epitopes and Theoretic-
Experimental Study of Bm86 Protein Sequences from Colima Mexico**

Saúl G. Martínez-Arzate[§], Esvieta Tenorio-Borroto[§], Alberto Barbabosa Pliego[§],

Héctor M. Díaz-Albiter^{#,Δ}, Juan C. Vázquez-Chagoyán^{§,*} and Humbert González-Díaz^{1,∇,*}

[§]Molecular Biology Laboratory, CIESA, FMVZ, Autonomous University of The State of Mexico (UAEM), 50200 Mexico State, Mexico.

[#] Laboratory of Biochemistry and Physiology of Insects, Oswaldo Cruz Institute, FIOCRUZ, 4365, Rio de Janeiro, Brazil.

^ΔWellcome Trust Centre for Molecular Parasitology, University of Glasgow, University Place, Glasgow G12 8TA, United Kingdom.

¹Department of Organic Chemistry II, University of the Basque Country (UPV/EHU), 48940, Biscay, Spain.

[∇]IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Biscay, Spain.

Abstract. In this work, we developed a general Perturbation Theory and Machine Learning (PTML) method for data mining of proteomes, in order to discover new B-cell epitopes useful for vaccine design. The method predicts the epitope activity $\epsilon_q(c_{qj})$ of one query peptide (q-peptide) in a set of experimental query conditions (c_{qj}). The method uses as input the sequence of the q-peptide. The method also uses as input information about the sequence and epitope activity $\epsilon_r(c_{rj})$ of a peptide of reference (r-peptide) assayed on similar experimental conditions (c_{rj}). The model proposed here is able to classify 1,048,190 pairs of query and reference peptide sequences from the proteome of many organisms reported on IEDB database. These pairs have variations (perturbations) in sequence or assay conditions. The model has accuracy, sensitivity, and specificity between 71% and 80% for training and external validation series. The retrieved information contains structural changes in 83683 peptides sequences (Seq) determined in experimental assays with boundary conditions involving 1448 Epitope Organisms (Org), 323 Host Organisms (Host), 15 types of *In vivo* Process (Proc), 28 Experimental Techniques (Tech), and 505 Adjuvant additives (Adj). Afterwards, we reported the experimental sampling, isolation, and sequencing of 15 complete sequences of Bm86 gene from state of Colima, Mexico. Last, we used the model to predict the epitope immunogenic scores in different experimental conditions for the 26112 peptides obtained from these sequences. The model may become a useful tool for epitope selection towards vaccine design. The theoretic-experimental results on Bm86 protein may help on the future design of a new vaccine based on this protein.

Keywords: Proteome mining; Epitope prediction; B-cell epitope; PCR; Bm86 protein; Machine Learning; Perturbation Theory.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCTION

There is a huge amount of information available on public proteome databases that may be of great importance for the prediction of immunogenic peptides that can be used for the design of new vaccines. For instance, the Immune Epitope Database (IEDB)¹ is an informational repository published, derived of infectious diseases, allergy and autoimmunity among other. IEDB (<http://www.iedb.org/>), captures epitopes structure and the experimental conditions in which were tested^{2,3}. The actual experimental assays are focused namely whether these molecules were immune or not, relating to T cell responses, B cell responses, Major Histocompatibility Complex (MHC) ligand elution, and MHC binding. This information is in the public domain and it can be recovered from IEDB.

However, a problem in experimental proteome research, connected to immunology and ultimately vaccine design, is the prediction of the impact over the immunological response of changes (perturbations) in the system. These perturbations may occur for variation in the sequence of a known peptide and/or variations in the experimental conditions of assay such as the technique of assay, the adjuvant used, the host organism, *etc.* It is possible to respond the question over the impact on the immune response after a perturbation through the use of bioinformatics tools⁴⁻⁶. In fact, there are different computational methods useful for predicting the function of proteins or peptides in proteome research. In particular, computational alignment-free methods based on Structure-Property Relationships use as input information about the spatial structure or the sequence of proteins in order to predict their biological activity. In this sense, we can use Machine Learning (ML) methods, like Linear Discriminant Analysis (LDA) or Artificial Neural Networks (ANN) to connect the structural information with biological properties⁶⁻¹⁰. Additionally, Perturbation Theory (PT) models use different operators to integrate information about the possible effect (perturbation) of multiple changes in one or more of the boundary conditions (experimental conditions). Then, we can define a PTML model as a PT model fitted using a ML algorithm. PTML-like models have been used successfully to predict biological properties not only on Proteome research but also in Nanoscience's and Medicinal Chemistry¹²⁻²³. In fact, a PTML model has been developed before to predict the changes on B-cell epitope activity after multiple variations / perturbations in factors like the sequences of the peptides, host organism, source organism, immunological process and experimental technique.³ In addition, Vázquez-Prieto *et al.*², developed a comparative study of PTML models of B-cell epitopes using different physicochemical molecular properties of peptide sequences. However, these models do not take into consideration very important factors like the use of different adjuvant additives in vaccine design and the dataset used was based on old versions of IEDB dataset with important data missing.

2

ACS Paragon Plus Environment

1 On the other hand, the universe of protein immunogenic molecules that can be studied with a
2 combination of experimental techniques of genomics/proteomics and computational models is huge.
3 We are interested in studying the Bm86 antigen, a tick mid gut epithelial protein used in commercial
4 vaccines to combat *R. microplus* in cattle²³⁻²⁶. The Australian strain used for the preparation of the
5 commercial vaccine is Yeerongpilly²⁴, and has been tested with relatively good success. However,
6 vaccine efficacy seems to be dependent of allelic variations of the Bm86 gen²⁶⁻³⁰. A single complete
7 allele sequence of the protein's gene has been reported so far in Mexico³¹, this information was not
8 enough to develop a vaccine against the product of a very polymorphic gene as Bm86.

9 In this work, we report for the first time a new computational method to predict B-cell epitopes using
10 Shannon's information measure indices of peptide sequence and experimental conditions. In order to
11 fit this model we processed 1,048,190 pairs of query and reference peptide sequences from the
12 proteome of many organisms reported on IEDB database. These pairs have variations (perturbations)
13 in sequence or assay conditions. The data include 1448 Epitope Organisms (Org), 323 Host
14 Organisms (Host), 15 types of *In vivo* Process (Proc), 28 Experimental Techniques (Tech), and 505
15 Adjuvant additives (Adj). It is important to note that this is probably the first model developed at this
16 level of complexity that places the main emphasis not only on the peptide but also on the adjuvant
17 together with other experimental conditions. In addition, we reported for the first time the
18 experimental isolation of 15 Bm86 complete sequences from tick specimens collected in the state of
19 Colima, Mexico. Last, we used the computational method to predict the epitope immunogenic scores
20 in different experimental conditions for the peptides obtained from these sequences.

21 MATERIALS AND METHODS

22 We know that for the model development it is vital to have four elements: a set of molecules called
23 training set, the choice of molecular descriptors to help us extract information from a molecule, the
24 choice of the mathematical function to help us establish the structure-property relationship, and a set
25 of sampling molecules to predict the model.

26 Experimental methods

27 **Sampling.** Specimens of tick *R. microplus* were collected of cattle of geographical region under
28 governmental tick control program "Agreement establishing the National Campaign is set to control
29 the tick *Boophilus spp*"^{32,33}. In total, 15 individual sequence were sampled from Colima, Mexico.

30 **RNA extraction.** The following experimental protocols were carried out for RNA extracted from
31 individual intestine samples obtained from *R. microplus* specimens. Ticks were dissected to remove
32 the intestine, where the epithelium expresses the protein Bm86, and the total RNA was extracted.
33 Samples (10-15 mg of intestine) were placed in Trizol solution (Invitrogen, USA)³⁴⁻³⁶. RNA was
34 extracted according to the protocol Solution TRI Reagent® (Ambion) in 4 steps: 1. Homogenization
35 tissue with 1mL TRI Reagent (The tissue should not exceed 10% of the total volume of Trizol

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

solution) followed by centrifugation at 12000 x g for 10 min at 4°C; 2. For RNA extraction 100 µl of 1-bromo-3-chloropropane (BCP) per 1mL of TRI Reagent solution were added followed by centrifugation at 12000 x g for 10-15 min at 4°C, aqueous phase was transferred to a fresh tube; 3. RNA was precipitated, washed and re-suspended by adding 500 µl of isopropanol per 1mL of TRI Reagent solution initially used, centrifuged at 12000 x g for 8 min at 4°C and the supernatant discarded. The pellet of RNA was washed with 1mL of 75% ethanol per 1mL of TRI Reagent solution initially used, centrifuged at 7500 x g for 5 min and finally the ethanol was removed and let the pellet dry. 4. RNA solubilization, the pellet of RNA was suspended in 40 µl of nuclease free double distilled water and stored at - 80 °C until use

cDNA synthesis. Total RNA was used to synthesize complementary DNA (cDNA) according to the Protocol of SuperScript II Reverse Transcriptase³⁷⁻³⁹ (Invitrogen, by life technologies)™. The reaction mixture was: 1 µL of Oligo (dT) 500µg/ml, 2 µL of total RNA (at 2000 ng/µl), 1 µL of dNTP (10mM), 8 µL of nuclease-free water, for a total volume of 12 µL. Reactions were incubated in a thermocycler (Techne TC-512 USA), at 65° C for 5 minutes and 4° C for 2 minutes, then 4 µL of 5 X First-Strand Buffer and 2 µL of DTT were added, incubated at 42° C for 2 minutes, then 1 µL reverse transcriptase SuperScript II added, and incubated at 42° C for 50 minutes. cDNA was used immediately after synthesis.

Amplification of the target sequences Bm86. Oligonucleotides were designed with the Primer3 software⁴⁰ to produce 4 overlapping amplicons (between 542-686 bp for each pair of primers) that included the whole Bm86 cDNA sequence (1950bp). PCR cycling conditions were set as follows: initial denaturation at 94°C per 4 min, followed by 30 cycles of denaturation at 94°C per 1 min, annealing temperatures as shown for each set of primers in **Table 1** per 30 sec and extension at 72°C per 1 min Finally, an extension incubation at 72°C per 5 min, and a cooling incubation at 10°C per 5 min were set^{39,41}. Bm86 amplicon sizes were verified in 1.5% agarose gel electrophoresis (Mudip-exu-AD140E) (30 min at 100V).

Table 1 comes about here

Sequencing. Samples' concentrations (ng/µL) of individual PCR amplicons were measured in a spectrophotometer (Quawell Q5000). Samples (50ng/µL in a 10µL) were packed individually in PCR tubes previously identified and shipped to Genscript (GenscriptCorporation, Piscatway, NJ, USA; www.genscript.com) where they were sequenced by the Sanger method.

Computational methods

Dataset. We downloaded all the linear B-cell epitopes reported on the IEDB database. The number of peptides retrieved was 83683, including: 2374 Positive-high, 76310 Positive, 5325 Positive-low, and 750 Positive-intermediate epitopes. For the purpose of classification, we grouped all these epitopes into two classes, according to the intensity of the immunogenic response related to this epitope (ϵ). These two classes are Positive-high epitopes ($\epsilon = 1$) and Positive-to-Intermediate epitopes

($\varepsilon = 0$). The biological assays used to determine the intensity of the immunogenic response have been carrying out in many different experimental conditions \mathbf{c}_j (see next sections). Consequently, from now on we depict ε as a function of these conditions $\varepsilon(\mathbf{c}_j)$.

Model definition. Our method is based on the comparison of all the information about sequence, assay conditions, and intensity of the immunogenic response $\varepsilon_q(\mathbf{c}_i)$ for a query peptide (q-peptide) with respect to the values of $\varepsilon_r(\mathbf{c}_j)$ and all the information for peptide of reference (r-peptide). In our dataset, we represented every query peptide (q-peptide) as a vector \mathbf{c}_{qj} . The components of the vector quantify information about different experimental conditions of assay $\mathbf{c}_{qj} = (c_{q1}, c_{q2}, c_{q3}, c_{q4}, c_{q5}, c_{q6}, c_{q7})$. These variables are for example: $c_{q1} = \varepsilon_q(\mathbf{c}_j)$ the epitope activity, $c_{q2} =$ New sequence, $c_{q3} =$ Adjuvant, *etc.* (see next sections). The r-peptide has a sequence similar to the sequence of the q-peptide but has been measured either on the same or different conditions of assay. The vector for the r-peptide is similar $\mathbf{c}_j = (c_{r1}, c_{r2}, c_{r3}, c_{r4}, c_{r5}, c_{r6}, c_{r7})$. The elements are for example: $c_{r1} = \varepsilon_r(\mathbf{c}_j)$ the r-peptide activity, $c_{r2} =$ sequence information of r-peptide, $c_{r3} =$ Adjuvant, *etc.* (see next sections). The general formula of the type of model proposed here is the following:

$$S_{qr}(\mathbf{c}_i, \mathbf{c}_j) = e_0 + e_1 \cdot \varepsilon_r(\mathbf{c}_j) + \sum_{j=3}^{j=7} c_k \cdot {}^q\theta_k + \sum_{k=0}^{k=5} d_k \cdot {}^q\theta_0(\mathbf{c}_i) + \sum_{k=0}^{k=5} b_k \cdot ({}^q\theta_k - {}^r\theta_k) + \sum_{i=j=3}^{i=j=7} a_j \cdot ({}^q\theta_i - {}^r\theta_j) \quad (1)$$

In this model, $S_{qr}(\mathbf{c}_i, \mathbf{c}_j)$ is the score value for the epitope activity of the q-peptide predicted under experimental conditions $\mathbf{c}_i = (c_{q1}, c_{q2}, c_{q3}, c_{q4}, c_{q5}, c_{q6}, c_{q7})$ using as reference the peptide r measured under a set of experimental conditions $\mathbf{c}_j = (c_{r1}, c_{r2}, c_{r3}, c_{r4}, c_{r5}, c_{r6}, c_{r7})$. The coefficient e_0 is the independent term and d_j is the coefficient of the observed value of epitope activity $\varepsilon_r(\mathbf{c}_j)$ of the r-sequence. The a_j are the coefficients of the input variables. The first input variable of the model $c_{r1} = \varepsilon_r(\mathbf{c}_j)$ is the epitope activity of the r-peptide; used here as template. The second type of variable or experimental condition $\Delta\theta_k = {}^q\theta_k - {}^r\theta_k$ quantifies information about the variation or perturbation on the sequence (chemical structure) of the query peptide with respect to the template. The variables ${}^q\theta_k$ and ${}^r\theta_k$, are Shannon's entropy information indices of the structure of the query and reference peptide, see next section. The last types of variables or experimental conditions quantify the information about the non-structural factors conditioning the immune response. These last variables are also information indices with the form of Shannon's entropies (see next sections).

Information Indices of Peptide Sequence. We calculated the Shannon's entropy information measures (${}^q\theta_k$ or ${}^r\theta_k$) of the sequence of each peptide in order to quantify relevant structural information. Firstly, we used the software S2SNet (<https://github.com/muntisa/S2SNet>) to transform each peptide sequence into one Sequence Recurrence Networks (SRN)⁴²⁻⁴⁵. The SRN of one peptide sequence is defined as a graph where the amino acids are the nodes and a pair of nodes (n_a and n_b) in

the network are connected by a link ($\alpha_{ab} = 1$) if they are adjacent and/or neighbor recurrent nodes.

This means that $\alpha_{ab} = 1$ if the topological distance between n_a and n_b is $d = 1$ (chemically bonded aminoacids) or they are the nearest neighbor aminoacids of the same type (A, R, N, D, C, Q, E, G, H, I, L, K, M, F, P, S, T, W, Y, V), with minimal topological distance $d_{ab} = \min(d_{ab})$ among them. The first node in the sequence (center of the star graph) is a bias or dummy non-residue vertex.

Secondly, we need to transform the SRN of each sequence into one stochastic matrix ${}^1\Pi$. The elements of ${}^1\Pi$ are the probabilities ${}^1p_{ab}$ of reaching an aminoacid (node n_j) walking from another aminoacid (node n_i) through a walk of length $d_{ij} = 1$ ^{17,18}:

$$p_{ab} = \frac{\alpha_{ab}}{\sum_{n=1}^n \alpha_{ab}} \quad (2)$$

Note that the number of amino acids in the sequences is equal to the number of nodes (n) in the SRN graph and also is equal to the number of rows and columns in ${}^1\Pi$, the length of the sequence (L), and the maximal topological distance in the sequence $\max(d_{ab})$. In this work, we quantified the information content of a peptide with the values of Shannon Entropy (θ_k) of the k -th natural powers of the Markov matrix ${}^1\Pi$. The same procedure was used to quantify the information of the q -peptides (${}^q\theta_k$) and r -peptides (${}^r\theta_k$). At follows we give the formula of Shannon entropy ${}^q\theta_k$, please see details in previous works.⁴²⁻⁴⁵

$${}^q\theta_k(seq) = -\sum_{a=0}^{a=L} {}^k p_a \cdot \log({}^k p_a) \quad (3)$$

Information Indices of Experimental Conditions. We also calculated Shannon's entropy information indices $\theta_0(c_j)$ to quantify useful information about the other experimental conditions $c_j = (c_3, c_4, c_5, c_6, c_7)$. These variables are: $\theta_0(\text{Adj.})$, $\theta_0(\text{Org.})$, $\theta_0(\text{Host})$, $\theta_0(\text{Proc.})$, and $\theta_0(\text{Tech.})$. The experimental conditions numerically encoded by these variables are: $c_3 = \text{AA} = \text{Adjuvant Additives} = \text{Vaccine Adjuvant with the peptide}$, $c_4 = \text{Org.} = \text{Epitope Organism} = \text{Organism source of the peptide}$, $c_5 = \text{Host} = \text{Host Organism exposed to the antigen}$, $c_6 = \text{Proc.} = \text{In vivo Process type} = \text{Immunological process of sensitization}$, and $c_7 = \text{Tech.} = \text{Techniques} = \text{Technique or Method used to determine the immunogenicity of the peptides}$. All these variables were calculated as a Shannon's entropy of the probability of occurrence of each condition on the dataset (see results and discussion section). The probability of occurrence p_j of each condition c_j was determined as a ratio of the relative frequency n_j of occurrence with respect to the total number n_{total} of entries in the original dataset, see the following equation.

$${}^q\theta_0(c_j) = -\sum_{j=0}^{j=L} p_j \cdot \log(p_j) = -\sum_{j=0}^{j=L} \left(\frac{n_j}{n}\right) \cdot \log\left(\frac{n_j}{n}\right) \quad (4)$$

Training and validation of the model. Once we calculated the input variables it is necessary to set and test a hypothesis of model. A linear model was selected as the first approximation. Linear

Discriminant Analysis was chosen to find the linear combination of independent variables that best differentiate (discriminate) both groups^{46,47}. In so doing, we need to create a raw data matrix. The data was ordered as follows: we sampled a 1,048,190 of pairs of sequences from the total database of 83683 sequences. Each pair of sequences contains a query sequence (q-sequence) and a sequence of reference (r-sequence). These sequences were organized both numerically and alphabetically. Numerical ordering was according to ϵ_{q1} values from 1 to 0. The alphabetical sorting was from A to Z according to the letters of the aminoacids in the sequence (one-letter code). A total of 1,048,190 pairs of sequences were sampled at random according to the following scheme. The label of the query sequence (n-query) was generated at random between 1 and 83683. The label of template sequence (n-template) was generated as a random number between n-query and n-query + a random number between 0 and 5. This sampling scheme guarantees that we have pairs of query and reference peptides with high similarity in the dataset. The remnant 548,190 pairs of sequences were selected totally at random to make a total of 1,048,190 pairs of cases (this is the limit of our current processing capabilities). Subsequently, we used LDA implemented in the software STATISTICA to seek for the classification model^{13,17,19}. The r-sequence is the sequence of reference with a known classification. Both, the activity of the q-peptide and r-peptide have the same codification. The values $\epsilon_r(\mathbf{c}_j) = 1$ and/or $\epsilon_q(\mathbf{c}_j) = 1$ are for q-peptides and/or r-peptides with High Positive Epitope activity in IEDB. The values $\epsilon_r(\mathbf{c}_j) = 0$ and/or $\epsilon_q(\mathbf{c}_j) = 0$ are for q-peptides and/or r-peptides with activity equal to Positive, Middle Positive, or Low Positive in the IEDB. The q-sequence is the query sequence to be classified by the model as High Positive Epitope ($\epsilon_q(\mathbf{c}_j) = 1$) or not ($\epsilon_q(\mathbf{c}_j) = 0$). We also calculated the values of $\Delta\theta_j = ({}^q\theta_j - {}^r\theta_j)$ for all these 548,190 pairs of sequences. The number of cases used to train the model includes 83683 peptides. We used the set retrieved from the database IEDB: involving 1448 EON, 323HN, 15 IVPT, 28 ET, and 505AA, the data were organized in an excel file, see **Figure 1**.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

7

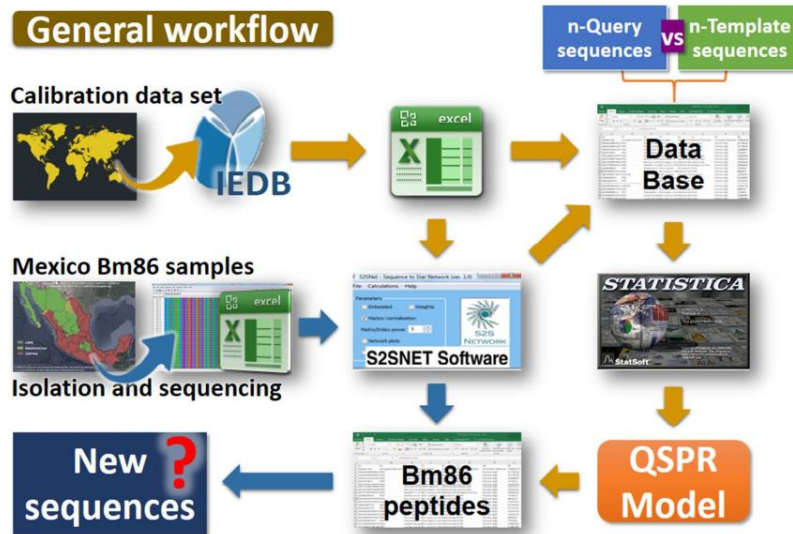


Figure 1. General workflow of model: calibration data set (brown arrows) set of sequences Bm86 (blue arrows).

Computational Predictive Study

Sequence assembly. Using the CodonCode Aligner software (CodonCode Corporation, www.codoncode.com), the 4 amplicon sequences from the Bm86 gene from each individual tick, were assembled into the complete Bm86 sequence. Consensus sequences of the ticks were aligned, resulting in a matrix that includes the 15 Bm86 Mexico tick sequences. DNA sequences were edited in MEGA6⁴⁸ to remove the low-resolution extremes of the chromatogram, the pro-peptide (NH₂ end region) and the signal peptide (COOH- end region) sequences, keeping only the DNA sequence fragment, corresponding to the glycoprotein region.

Table 2 comes about here

Alignment and phylogenetic analysis. A multiple sequence alignment was performed using the Clustal W method in Mega 6 Software. For the phylogenetic analysis, all the Bm86 gen, from tick collected in Colima, Mexico were used. Complete sequences retrieved from GenBank (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>), from the residues 20-627 of the following countries were considered for phylogenetic analysis: 14 sequences from Texas, United States, 2 from Mozambique and one each from Brazil, Argentina, Mexico and Australia^{23,27,31,37,49-51} (Table 2). Neighbor-joining⁵² method was used to study how the sequences are grouped, and the evolutionary distances were computed using the Kimura 2-parameter method⁵³. It is basically a matrix of genetic distances reflecting the degree of similarity; however, the method generates unrooted trees. To solve this, an out-group sequence was included in the model. The tree was assessed by bootstrapping 1000

repetitions, and the evolutionary distances were computed using the Kimura 2-parameter method based in the number of base substitutions per site. Codon positions included were 1st + 2nd + 3rd + Noncoding (3rd. position of the codons accumulate mutations much more quickly than the 2nd and 1st). The rate variation among sites was modeled with a gamma distribution (shape parameter = 1).

RESULTS AND DISCUSSION

Computational model for epitope prediction. In vaccine design, the ability to predict the properties of a molecule before its synthesis saves time and money in the area of research. It is possible to achieve this prediction through computational methods that, by mathematical models, demonstrate the relationship between structure and activity. General models to predict the results of multiple input-output perturbations in peptide sequence and experimental assay boundary conditions using ideas of perturbation analysis have been developed in the past^{12,14-16}. In this work, we reported a new model able to predict the propensity of a peptide to act as B-epitope after multiple perturbations in peptide sequence and boundary conditions of experimental assays. The equation of the model is the following^{2,3}.

$$S_{qr}(c_i, c_j) = -1.96618 \cdot \varepsilon_r - 0.00368 \cdot \theta_2(\text{Seq}) - 0.01354 \cdot \theta_0(\text{Org}) - 0.08383 \cdot \theta_0(\text{Tech}) \quad (5)$$

$$+ 0.00463 \cdot \Delta\theta_5(\text{Seq}) + 0.00404 \cdot \Delta\theta_0(\text{Adj}) + 0.00245 \cdot \Delta\theta_0(\text{Org})$$

$$+ 0.00089 \cdot \Delta\theta_0(\text{Host}) - 0.00095 \cdot \Delta\theta_0(\text{Proc})$$

$$+ 0.00438 \cdot \Delta\theta_0(\text{Tech}) + 7.95575$$

$$N_{\text{training}} = 786,142 \quad N_{\text{validation}} = 262,048 \quad N_{\text{total}} = 1,048,190$$

$$\chi^2 = 163065.0 \quad p < 0.005$$

The results of the model summarize the excellent predictive ability of the LDA function. The model discriminates both classes of peptides with a significance level ($p < 0.005$). The model needs only 7 variables to classify with accuracy higher 70%, in training and validation, more than one million of cases ($N_{\text{total}} = 1,048,190$). The model classified 786,142 cases for the training series (212,326 misclassifications) and 262,048 for the validation series (70,849 misclassifications), recording thus percentages of Specificity of 71% and Sensitivity of 80% for both series^{17,19,44-46}.

Table 3 comes about here

Comparing PTML model to others. Similar models have been developed in the past^{7-10,18,20}. However, the model proposed here is more robust. PTML model was able to classify 1,048,190 sequences, pairs of query and reference peptide, while with previous models only 200,000 had been classified. The best LDA model found in this work achieved a sensitivity and specificity of 80% and 71%. These percentages are relatively lower than previous proposals, it is possible to assume these results, since the data is larger and more heterogeneous than in previous studies, in addition include 505 different Adjuvant additives unlike previous studies, see **Table 3**. It should be noted that, comparatively to our study, previous reports that used the same data base, used a lower number of cases for each one of the boundary conditions, as follows: 1448/500 organisms; 323/50 host organisms; IVPT 15/10 *in vitro* process; and 30/28 Techniques. Additionally, in this work,

9

for the first time were included 505 different Adjuvant additives. It is essential to know that the adjuvants are non-specific heterogeneous agents; their importance is that, in association with vaccine antigens can induce a more potent and prolonged immune response. The use and choice of experimental conditions will depend on the nature of the experiment and expectations of each investigator. The model allows choosing between 505 possible options of adjuvants, however it is important to emphasize that 97.2% of the assays are free of adjuvants, so their use, at least in this database, are not very common. It is possible to choose the option “free of adjuvants” in this model.

Experimental-Theoretic study of Bm86 protein. The recombinant protein Bm86 has been used in the formulation of vaccines and tested in the field with an acceptable success for tick control^{24,35,36,38,49,53}. However, the genetic polymorphism found of Colima, Mexico, with respect to commercial antigens Gavac® and TickGARD^{24,54,55}, could explain the difference in effectiveness of the vaccine that was reported by several authors for several countries (Australia, Mexico, Cuba, Argentina, and Venezuela). For example, a review of the effectiveness of the vaccine in the middle of the first decade of this century, showed that the combined use of the vaccine with dip tanks reduced the use of acaricide up to 87% in Cuba, 100% in some regions of Australia, 67% in some regions of Mexico, 77% in Colombia and 50% in Brazil^{24,51,54,55}.

Bm86 has an impact on the immune activity, specifically for B cell epitopes. We successfully isolated, amplified and sequenced the cDNA coding for the putative Bm86 protein, from 15 tick specimens from the estate of Colima, Mexico. The presence of the amplicons for each primer used in the PCR reaction, was demonstrated by agarose gel electrophoresis, observed as UV light fluorescent bands; which corresponded to the presence and the size of expected DNA fragments, see **Figure 2**.

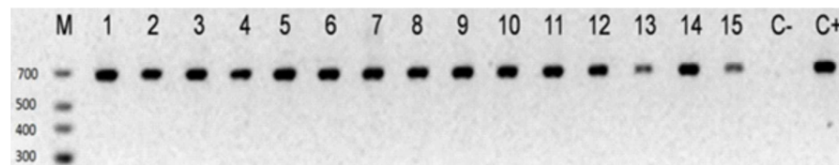


Figure 2. PCR amplicons: This image shows the 680pb amplicons corresponding to the expected cDNA band size for the Bm86-3_F and Bm86-3_R primer pair amplicon. PCR reaction was run in a 1.5% agarose gel and stained with ethidium bromide. M, molecular weight marker (25 -700 bp; GeneRuler Low Range DNA Ladder); Numbers 1-15 are individual samples; C-, negative control; C+, positive control (a previously sequenced cDNA matching the GenBank Bm86 gene sequence, access number M29321.1, was used as template).

When we aligned the sequences by the method ClustalW (software Mega6), we observed variations in the amino acid sequence between Mexico and the vaccine strain (Yeerongpilly), that may explain the reduced effectiveness of the vaccine in Mexican strains²⁴.

1 According to García-García *et al*⁴⁹, when the sequence of the vaccine has a variation higher than 3%,
2 with respect to the sequence of the target tick, the antibodies against the ectoparasite are less
3 effective. The variation observed in the amino acid sequence for all our isolates was greater than 3%
4 with respect to the Yeerongpilly Bm86 protein sequence, which helps to explain the low
5 effectiveness of the commercial vaccine observed in Mexico⁴⁹, **Table 6**. We predicted with our
6 model, the impact on the immunological response due to the variation of these sequences, see **Table**
7 **5**.
8

9
10
11 **Phylogenetic analysis.** In the phylogram (**Figure 3**), we can see the group of all sequences of Bm86
12 gene considered for this analysis. Additionally, two reference sequences of *R. annulatus*(identified as
13 DimmiTx and KinneTx), a close relative genus of *R. microplus* were included in the study as
14 outgroups. Phylogenetic relationship of Bm86 gene sequences was inferred using the Neighbor-
15 Joining method⁵². The evolutionary distances were computed using the Kimura 2-parameter
16 method⁵³, the phylogenetic tree yields a hypothesis of this relationship, through which the origins of
17 the strains found in different geographic regions can be inferred. Assuming that the tick has a
18 common ancestor between species *R. annulatus* and *R. microplus*, the phylogenetic tree was rooted in
19 the sequences of *R. annulatus* (DimmiTx and KinneTx), we found these sequences forming a
20 remarkably divergent clade, this clade includes the two strains *R. annulatus* of Texas, one strain from
21 Texas (Hid Tx) and one from Colima, which are closer to the *R. annulatus* strains than from those of
22 *R. microplus*. Suggesting that Hid Tx specimen was not classified correctly. Two loose branches or
23 polyphyletic are formed with sequences of Colima (F10 A and F10 B), which apparently diverged
24 from the same allele. Integrating another clade, we found the reports of Mozambique (Bm86 Moz
25 and Rm Moz) and Bm86 Yeroongpilly strains, with a significant degree of divergence.
26 Hypothetically, a common ancestor gave rise to most Colima's isolates analyzed in the present study,
27 which are grouped in a larger clade, together with the rest of the American reports. However, it is
28 interesting to note that Col F1 B, Col F12, 12 Tx and Star1 Tx sequences, integrate a clade, that
29 shows a greater degree of divergence. The rest of the sequences are less divergent from each other.
30 The phylogenetic tree gives us an idea of the genetic polymorphism (degree of divergence) present in
31 Colima Bm86 sequences and the phylogenetic relations that they keep with the reports of America,
32 Africa and Oceania. The variation in amino acid sequence of Bm86 from ticks from Colima, and the
33 sequence from the vaccine strain Yeerongpilly, was greater than 2.9% in all cases, see **Table 5**.
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50

51 **Table 5 comes about here**
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

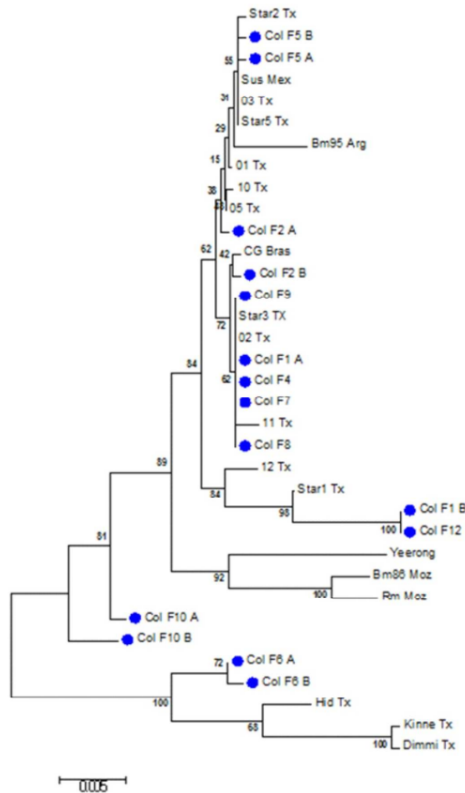


Figure 3. Phylogenetic tree

The phylogenetic tree gives us an idea of the phylogenetic relationships of the Bm86 gene, with respect to the reports for other regions of the world. Phylogenetic relationship of Bm86 gene of Mexico's isolates (blue points), and GenBank reports, was inferred using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The analysis involved 35 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 1757 positions in the final dataset. Evolutionary analyses were conducted in MEGA6. This distribution suggests the hypothesis that there has been a migration of ticks between Colima and Texas and that there is also migration between Brazil, Colima, and Texas, although there is insufficient information to support the migratory routes of the ecto-parasite. In addition, the phylogenetic tree suggests that a vaccine made from the Bm86 antigen from Colima or Texas strains could be more effective against strains circulating in these regions than the Yeroongpilly genotype-based vaccine, **Figure 3**.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Data mining study of sequences from Colima state. In this section, we illustrate the use of the model in a practical case. In so doing, we used the model to predict the activity scores S_{qj} for a very large set of 26112 non-redundant peptides from Bm86 protein sequences isolated from Colima, Mexico. Firstly, we calculated the information indices ${}^q\theta_5(\text{Seq})$ of the sequences of the q-peptides. Next, we substituted the values of the entropy information indices ${}^q\theta_5(\text{Seq})$ for each peptide into the equation of the model. Selected only ${}^q\theta_5(\text{Seq})$, the entropy ${}^q\theta_k(\text{Seq})$ with $k = 5$, because is the only one sequence entropy in the equation. We also had to substitute into the model, the entropy ${}^r\theta_5(\text{Seq})$ for the r-peptide used as template.

For input values r-peptide, we select a peptide of the sequence Bm86, but from a homologous strain to *R. microplus*, (*R. decoloratus*), which has been tested as a B-cell epitope, under experimental conditions c_{qj} , because so far, only Bm86 peptides from *R. decoloratus* were available in the IEDB database when this study was under course. Complete work on this peptide can be found at PubMed ID 17070625⁵⁴. Last, we had to substitute into the equation, the values of information indices ${}^q\theta_0(c_j)$ and ${}^r\theta_0(c_j)$ for the conditions of assay of both the q-peptide and the r-peptide. In **Table 5**, can be seen examples of the ${}^q\theta_0(c_j)$ values calculated for many different conditions of assay. We also reported a more detailed list of ${}^q\theta_0(c_j)$ values in supporting information, **Table SI02**.

In **Table SI05**, we report the Normalized values of $p_{qr}(c_i, c_j)$ for the score $S_{qr}(c_i, c_j)$ of the top-10 predicted peptides out of 26112 non-redundant peptides. These values are equal to $p_{qr}(c_i, c_j) = [S_{qr}(c_i, c_j) - \text{Min}(S_{qr}(c_i, c_j))] / [\text{Max}(S_{qr}(c_i, c_j)) - \text{Min}(S_{qr}(c_i, c_j))]$. In this min-max normalization equation $S_{qr}(c_i, c_j)$ is the score of activity predicted for the query peptide q under experimental conditions c_i using as reference the peptide r measured under experimental conditions c_j . In this equation $\text{Min}(S_{qr}(c_i, c_j))$ and $\text{Max}(S_{qr}(c_i, c_j))$ are the minimum and maximum values of $S_{qr}(c_i, c_j)$ for all the peptides studied. One of the more important advantages of this model is that one could select many different sets of conditions to do the predictions. In this example, for the sake of simplicity, we set the values of query conditions equal to the values of the conditions of reference $c_r = c_j = (c_3, c_4, c_5, c_6, c_7)$ which are according to the conditions of assay reported for this peptide in the literature⁵⁴. We used the normalized score to make easier the comparison because $p_{qr}(c_i, c_j)$ varies from 0 to 1. The normalized score $p_{qr}(c_i, c_j) = 0$ for $S_{qr}(c_i, c_j) = \text{Min}(S_{qr}(c_i, c_j))$ and $p_{qr}(c_i, c_j) = 1$ for $S_{qr}(c_i, c_j) = \text{Max}(S_{qr}(c_i, c_j))$.

Conclusions

We have proposed a model for prediction of new B-cell epitopes. Machine Learning methods are useful for predicting the function of attribute that ranks as antigenic peptide or not, in different experimental conditions, using as a basis the structural information (variations in peptide sequences) but with autonomy of sequence-sequence similarity. This model is able to classify more than 26112 cases of perturbations with 73% accuracy, 80% sensitivity and 71% specificity for training validation series.

Conflict of Interests

The authors declare no competing financial interest.

Acknowledgments

The present study was supported by Fundación Educación Superior-Empresa, A.C. (3378/2013e), Universidad Autónoma del Estado de México,3775/2014/CIC; Acknowledge financial support for graduate scholarship and student mobility granted by CONACYT. G.D.H personally acknowledges financial support from grants MINECO (CTQ2013-41229-P) - 2014 – 2016 and MINECO (CTQ2016-74881-P) - 2017 – 2019 of Spain government and grant (IT1045-16) - 2016 – 2021 of Basque Government.

Supporting Information Available: This material is available free of charge via the Internet at <http://pubs.acs.org>

REFERENCES

1. Vita, R.; Overton, J. A.; Greenbaum, J. A.; Ponomarenko, J.; Clark, J. D.; Cantrell, J. R.; Wheeler, D. K.; Gabbard, J. L.; Hix, D.; Sette, A.; Peters, B., The immune epitope database (IEDB) 3.0. *Nucleic acids research***2015**, 43 (Database issue), D405-12.
2. Vázquez-Prieto, S.; Paniagua, E.; Ubeira, F. M.; González-Díaz, H., QSPR-Perturbation Models for the Prediction of B-Epitopes from Immune Epitope Database: A Potentially Valuable Route for Predicting “In Silico” New Optimal Peptide Sequences and/or Boundary Conditions for Vaccine Development. *International Journal of Peptide Research and Therapeutics***2016**, 1-6.
3. Gonzalez-Diaz, H.; Perez-Montoto, L. G.; Ubeira, F. M., Model for vaccine design by prediction of B-epitopes of IEDB given perturbations in peptide sequence, in vivo process, experimental techniques, and source or host organisms. *Journal of immunology research***2014**, 2014, 768515.
4. Peters, B.; Sidney, J.; Bourne, P.; Bui, H. H.; Buus, S.; Doh, G.; Fleri, W.; Kronenberg, M.; Kubo, R.; Lund, O.; Nemazee, D.; Ponomarenko, J. V.; Sathiamurthy, M.; Schoenberger, S.; Stewart, S.; Surko, P.; Way, S.; Wilson, S.; Sette, A., The immune epitope database and analysis resource: from vision to blueprint. *PLoS biology***2005**, 3.
5. Sette, A., The immune epitope database and analysis resource: from vision to blueprint. *Genome informatics. International Conference on Genome Informatics***2004**, 15 (2), 299.
6. Salimi, N.; Fleri, W.; Peters, B.; Sette, A., The immune epitope database: a historical retrospective of the first decade. *Immunology***2012**, 137 (2), 117-23.
7. Gonzalez-Diaz, H.; Prado-Prado, F.; Garcia-Mera, X.; Alonso, N.; Abeijon, P.; Caamano, O.; Yanez, M.; Munteanu, C. R.; Pazos, A.; Dea-Ayuela, M. A.; Gomez-Munoz, M. T.; Garijo, M. M.; Sansano, J.; Ubeira, F. M., MIND-BEST: Web server for drugs and target discovery; design, synthesis, and assay of MAO-B inhibitors and theoretical-experimental study of G3PDH protein from *Trichomonas gallinae*. *Journal of proteome research***2011**, 10 (4), 1698-718.

- 1 8. Rodriguez-Soca, Y.; Munteanu, C. R.; Dorado, J.; Pazos, A.; Prado-Prado, F. J.; Gonzalez-Diaz, H.,
2 Trypano-PPI: a web server for prediction of unique targets in trypanosome proteome by using
3 electrostatic parameters of protein-protein interactions. *Journal of proteome research***2010**, 9 (2),
4 1182-90.
5
- 6 9. Munteanu, C. R.; Vazquez, J. M.; Dorado, J.; Sierra, A. P.; Sanchez-Gonzalez, A.; Prado-Prado, F. J.;
7 Gonzalez-Diaz, H., Complex network spectral moments for ATCUN motif DNA cleavage: first
8 predictive study on proteins of human pathogen parasites. *Journal of proteome research***2009**, 8 (11),
9 5219-28.
10
- 11 10. Concu, R.; Dea-Ayuela, M. A.; Perez-Montoto, L. G.; Bolas-Fernandez, F.; Prado-Prado, F. J.;
12 Podda, G.; Uriarte, E.; Ubeira, F. M.; Gonzalez-Diaz, H., Prediction of enzyme classes from 3D
13 structure: a general model and examples of experimental-theoretic scoring of peptide mass
14 fingerprints of Leishmania proteins. *Journal of proteome research***2009**, 8 (9), 4372-82.
15
- 16 11. Chou, K. C.; Shen, H. B., Euk-mPLoc: a fusion classifier for large-scale eukaryotic protein
17 subcellular location prediction by incorporating multiple sites. *Journal of proteome research***2007**, 6
18 (5), 1728-34.
19
- 20 12. Gonzalez-Diaz, H.; Arrasate, S.; Gomez-SanJuan, A.; Sotomayor, N.; Lete, E.; Besada-Porto, L.;
21 Ruso, J. M., General theory for multiple input-output perturbations in complex molecular systems. I.
22 Linear QSPR electronegativity models in physical, organic, and medicinal chemistry. *Curr Top Med*
23 *Chem***2013**, 13 (14), 1713-41.
24
- 25 13. Ran, T.; Liu, Y.; Li, H.; Tang, S.; He, Z.; Munteanu, C. R.; Gonzalez-Diaz, H.; Tan, Z.; Zhou, C.,
26 Gastrointestinal Spatiotemporal mRNA Expression of Ghrelin vs Growth Hormone Receptor and
27 New Growth Yield Machine Learning Model Based on Perturbation Theory. *Scientific Reports***2016**,
28 6, 30174.
29
- 30 14. Messina, P. V.; Besada-Porto, J. M.; Gonzalez-Diaz, H.; Ruso, J. M., Self-Assembled Binary
31 Nanoscale Systems: Multioutput Model with LFER-Covariance Perturbation Theory and an
32 Experimental-Computational Study of NaGDC-DDAB Micelles. *Langmuir***2015**, 31 (44), 12009-18.
33
- 34 15. Luan, F.; Kleandrova, V. V.; Gonzalez-Diaz, H.; Ruso, J. M.; Melo, A.; Speck-Planche, A.; Cordeiro,
35 M. N., Computer-aided nanotoxicology: assessing cytotoxicity of nanoparticles under diverse
36 experimental conditions by using a novel QSTR-perturbation approach. *Nanoscale***2014**, 6 (18),
37 10623-30.
38
- 39 16. Kleandrova, V. V.; Luan, F.; Gonzalez-Diaz, H.; Ruso, J. M.; Speck-Planche, A.; Cordeiro, M. N.,
40 Computational tool for risk assessment of nanomaterials: novel QSTR-perturbation model for
41 simultaneous prediction of ecotoxicity and cytotoxicity of uncoated and coated nanoparticles under
42 multiple experimental conditions. *Environmental Science & Technology***2014**, 48 (24), 14686-94.
43
- 44 17. Aguero-Chapin, G.; Varona-Santos, J.; de la Riva, G. A.; Antunes, A.; Gonzalez-Villa, T.; Uriarte, E.;
45 Gonzalez-Diaz, H., Alignment-free prediction of polygalacturonases with pseudofolding topological
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- indices: experimental isolation from *Coffea arabica* and prediction of a new sequence. *Journal of Proteome Research* **2009**, 8 (4), 2122-8.
18. Gonzalez-Diaz, H.; Saiz-Urra, L.; Molina, R.; Santana, L.; Uriarte, E., A model for the recognition of protein kinases based on the entropy of 3D van der Waals interactions. *Journal of Proteome Research* **2007**, 6 (2), 904-8.
19. Xiao, F.; Shen, H. B., Prediction Enhancement of Residue Real-Value Relative Accessible Surface Area in Transmembrane Helical Proteins by Solving the Output Preference Problem of Machine Learning-Based Predictors. *Journal of Chemical Information and Modeling* **2015**, 55 (11), 2464-74.
20. Pan, X. Y.; Zhang, Y. N.; Shen, H. B., Large-scale prediction of human protein-protein interactions from amino acid sequence based on latent topic features. *Journal of Proteome Research* **2010**, 9 (10), 4992-5001.
21. Zhu, L.; Yang, J.; Shen, H. B., Multi label learning for prediction of human protein subcellular localizations. *The Protein Journal* **2009**, 28 (9-10), 384-90.
22. Shen, H. B.; Chou, K. C., QuatIdent: a web server for identifying protein quaternary structural attribute by fusing functional domain and sequential evolution information. *Journal of Proteome Research* **2009**, 8 (3), 1577-84.
23. Rand, K. N.; Moore, T.; Sriskantha, A.; Spring, K.; Tellam, R.; Willadsen, P.; Cobon, G. S., Cloning and expression of a protective antigen from the cattle tick *Boophilus microplus*. *PNAS* **1989**, 86 (24), 9657-9661.
24. de la Fuente, J.; Almazan, C.; Canales, M.; Perez de la Lastra, J. M.; Kocan, K. M.; Willadsen, P., A ten-year review of commercial vaccine performance for control of tick infestations on cattle. *Conference of Research Workers in Animal Diseases* **2007**, 8 (1), 23-8.
25. Willadsen, P., Anti-tick vaccines. *Parasitology* **2004**, 129 Suppl, S367-87.
26. Sossai, S.; Peconick, A. P.; Sales-Junior, P. A.; Marcelino, F. C.; Vargas, M. I.; Neves, E. S.; Patarroyo, J. H., Polymorphism of the Bm86 gene in South American strains of the cattle tick *Boophilus microplus*. *Experimental & Applied Acarology* **2005**, 37 (3-4), 199-214.
27. Freeman, J. M.; Davey, R. B.; Kappmeyer, L. S.; Kammlah, D. M.; Olafson, P. U., Bm86 midgut protein sequence variation in South Texas cattle fever ticks. *Parasites & Vectors* **2010**, 3 (1), 1-8.
28. Garcia-Garcia, J. C.; Gonzalez, I. L.; Gonzalez, D. M.; Valdes, M.; Mendez, L.; Lamberti, J.; D'Agostino, B.; Citroni, D.; Fragoso, H.; Ortiz, M.; Rodriguez, M.; de la Fuente, J., Sequence variations in the *Boophilus microplus* Bm86 locus and implications for immunoprotection in cattle vaccinated with this antigen. *Experimental & applied acarology* **1999**, 23 (11), 883-95.
29. Lima, W. S.; Ribeiro, M. F.; Guimaraes, M. P., Seasonal variation of *Boophilus microplus* (Canestrini, 1887) (Acari: Ixodidae) in cattle in Minas Gerais State, Brazil. *Tropical Animal Health and Production* **2000**, 32 (6), 375-80.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
30. Kaewmongkol, S.; Kaewmongkol, G.; Inthong, N.; Lakkitjaroen, N.; Sirinarumitr, T.; Berry, C. M.; Jonsson, N. N.; Stich, R. W.; Jittapalpong, S., Variation among Bm86 sequences in *Rhipicephalus (Boophilus) microplus* ticks collected from cattle across Thailand. *Experimental and Applied Acarology* **2015**, 66 (2), 247-256.
 31. Canales, M.; Almazan, C.; Naranjo, V.; Jongejan, F.; de la Fuente, J., Vaccination with recombinant *Boophilusannulatus* Bm86 ortholog protein, Ba86, protects cattle against *B. annulatus* and *B. microplus* infestations. *BMC Biotechnology* **2009**, 9, 29.
 32. SAGARPA, Mexican Official Norm, Champaign against tick *Boophilus ssp.* Secretary of Agriculture, Animal husbandry, Rural Development, Fishing and Food, Ed. *DOF* **1994**.
 33. Agreement of National Campaign for the control of tick *Boophilus spp.* Secretary of Agriculture, Animal husbandry, Rural Development, Fishing and Food., Ed. *DOF*, NOM-019-ZOO-1994; **2012**.
 34. Nijhof, A. M.; Balk, J. A.; Postigo, M.; Rhebergen, A. M.; Taoufik, A.; Jongejan, F., Bm86 homologues and novel ATAQ proteins with multiple epidermal growth factor (EGF)-like domains from hard and soft ticks. *International Journal for Parasitology* **2010**, 40 (14), 1587-1597.
 35. Peconick, A. P.; Sossai, S.; Girão, F. A.; Rodrigues, M. Q. R. B.; Souza e Silva, C. H.; Guzman Q, F.; Patarroyo V, A. M.; Vargas, M. I.; Patarroyo, J. H., Synthetic vaccine (SBm7462) against the cattle tick *Rhipicephalus (Boophilus) microplus*: Preservation of immunogenic determinants in different strains from South America. *Experimental Parasitology* **2008**, 119 (1), 37-43.
 36. Ben Said, M.; Galai, Y.; Mhadhbi, M.; Jedidi, M.; de la Fuente, J.; Darghouth, M. A., Molecular characterization of Bm86 gene orthologs from *Hyalomma excavatum*, *Hyalommadromedarii* and *Hyalomma marginatum marginatum* and comparison with a vaccine candidate from *Hyalommascupense*. *Veterinary parasitology* **2012**, 190 (1-2), 230-40.
 37. Nijhof, A. M.; Balk, J. A.; Postigo, M.; Jongejan, F., Selection of reference genes for quantitative RT-PCR studies in *Rhipicephalus (Boophilus) microplus* and *Rhipicephalus appendiculatus* ticks and determination of the expression profile of Bm86. *BMC molecular biology* **2009**, 10, 112.
 38. Hope, M.; Jiang, X.; Gough, J.; Cadogan, L.; Josh, P.; Jonsson, N.; Willadsen, P., Experimental vaccination of sheep and cattle against tick infestation using recombinant 5'-nucleotidase. *Parasite Immunology* **2010**, 32 (2), 135-142.
 39. Almazan, C.; Lagunes, R.; Villar, M.; Canales, M.; Rosario-Cruz, R.; Jongejan, F.; de la Fuente, J., Identification and characterization of *Rhipicephalus (Boophilus) microplus* candidate protective antigens for the control of cattle tick infestations. *Parasitol Res* **2010**, 106 (2), 471-9.
 40. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B. C.; Remm, M.; Rozen, S. G., Primer3—new capabilities and interfaces. *Nucleic acids research* **2012**, 40 (15), e115-e115.
 41. van Zyl, W. A.; Stutzer, C.; Olivier, N. A.; Maritz-Olivier, C., Comparative microarray analyses of adult female midgut tissues from feeding *Rhipicephalus* species. *Ticks and tick-borne diseases* **2015**, 6 (1), 84-90.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
42. Munteanu, C. R.; Magalhaes, A. L.; Uriarte, E.; Gonzalez-Diaz, H., Multi-target QPDR classification model for human breast and colon cancer-related proteins using star graph topological indices. *Journal of theoretical biology***2009**, *257* (2), 303-11.
 43. Munteanu, C. R.; Gonzalez-Diaz, H.; Borges, F.; de Magalhaes, A. L., Natural/random protein classification models based on star network topological indices. *Journal of theoretical biology***2008**, *254* (4), 775-83.
 44. Munteanu, C.; Magalhaes, A.; Duardo-Sanchez, A.; Pazos, A.; Gonzalez-Diaz, H. S2SNet: A Tool for Transforming Characters and Numeric Sequences into Star Network Topological Indices in Chemoinformatics, Bioinformatics, Biomedical, and Social-Legal Sciences. *Curr. Bioinform.* **2013**, *8* (4), 429–437.
 45. Kowalski, R. D.; Wold, S.; In: Krishnaiah, P. R.; Kanal, L. N., (Eds.), Handbook of Statistic. *North Holland Publishing Company, Amsterdam*, **1982**, 673–697.
 46. Beleites, C.; Neugebauer, U.; Bocklitz, T.; Krafft, C.; Popp, J., Sample size planning for classification models. *Analytica Chimica Acta***2013**, *760*, 25-33.
 47. Min, B. K.; Dahne, S.; Ahn, M. H.; Noh, Y. K.; Muller, K. R., Decoding of top-down cognitive processing for SSVEP-controlled BMI. *Scientific reports***2016**, *6*, 36267.
 48. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S., MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution***2013**, *30* (12), 2725-9.
 49. Garcia-Garcia, J. C.; Montero, C.; Redondo, M.; Vargas, M.; Canales, M.; Boue, O.; Rodriguez, M.; Joglar, M.; Machado, H.; Gonzalez, I. L.; Valdes, M.; Mendez, L.; de la Fuente, J., Control of ticks resistant to immunization with Bm86 in cattle vaccinated with the recombinant antigen Bm95 isolated from the cattle tick, *Boophilus microplus*. *Vaccine***2000**, *18* (21), 2275-87.
 50. Canales, M.; de la Lastra, J. M.; Naranjo, V.; Nijhof, A. M.; Hope, M.; Jongejan, F.; de la Fuente, J., Expression of recombinant *Rhipicephalus (Boophilus) microplus*, *R. annulatus* and *R. decoloratus* Bm86 orthologs as secreted proteins in *Pichia pastoris*. *BMC biotechnology***2008**, *8*, 14.
 51. Andreotti, R.; Pedroso, M. S.; Caetano, A. R.; Martins, N. F., Comparison of predicted binders in *Rhipicephalus (Boophilus) microplus* intestine protein variants Bm86 Campo Grande strain, Bm86 and Bm95. *Brazilian Journal of Veterinary Parasitology***2008**, *17* (2), 93-8.
 52. Saitou, N.; Nei, M., The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution***1987**, *4* (4), 406-25.
 53. Kimura, M., A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of molecular evolution***1980**, *16* (2), 111-20.
 54. Odongo, D.; Kamau, L.; Skilton, R.; Mwaura, S.; Nitsch, C.; Musoke, A.; Taracha, E.; Daubenberger, C.; Bishop, R., Vaccination of cattle with TickGARD induces cross-reactive antibodies binding to conserved linear peptides of Bm86 homologues in *Boophilus decoloratus*. *Vaccine***2007**, *25* (7), 1287-96.

de la Fuente, J.; Rodriguez, M.; Montero, C.; Redondo, M.; Garcia-Garcia, J. C.; Mendez, L.;
Serrano, E.; Valdes, M.; Enriquez, A.; Canales, M.; Ramos, E.; Boue, O.; Machado, H.; Lleonart, R.,
Vaccination against ticks (*Boophilus* spp.): the experience with the Bm86-based vaccine Gavac. *Genetic
analysis: biomolecular engineering* **1999**, 15 (3-5), 143-8.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

List of figure legends

Figure 1. General workflow

Figure 2. PCR amplicons

Figure 3. Phylogenetic tree

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

TABLES TO BE INSERTED

Table 1. Bm86 gen PCR primers

Primer name	Sequence ^a	Position ^b	Tm/sec ^c	Size ^d (bp)
Bm86-1_F	5'GTGGTTCGACGCAGTGAGAT-3'	15-34	60°/30	680
Bm86-1_R	3'CCATGCTTGCAGACAAACCC-5'	675-694		
Bm86-2_F	5'CTGCAAAGACCTCTGCGAGA-3'	506-525	55°/30	686
Bm86-2_R	3'TGTCCTGCGTGCAGTTAAGT-5'	1172-1191		
Bm86-3_F	5'CGGGCCCAAATGTCAACATC-3'	1006-1025	55°/30	680
Bm86-3_R	3'AACGCACTCCAGCTTCTTGT-5'	1671-1690		
Bm86-4_F	5'AAAACGAGCAGTCGGAGTGT-3'	1513-1532	57°/30	542
Bm86-4_R	3'GGTGTTCGATGTAAGCGTGA-5'	2035-2054		

^a Primer sequence, can be observed directionality of forward and reverse primers.

^b Position of the primer in the Bm86 cDNA gene.

^c Melting temperature in Celsius degrees/incubation time in seconds for each pair of primers.

^d Amplicon size in bases are listed for each pair of primers.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 2. Bm86 sequences reported in GenBank

Autor's	Country	Cepa	# Gen Bank	Reference ^a
Rand KN., <i>et al.</i>	Australia	Yeerongpilly	M29321.1	23
Canales, M., <i>et al.</i>	Mexico	Susceptible	FJ456928.1	31
Garcia-Garcia, J.C., <i>et al.</i>	Argentina	-	AF150891.2	49
Nijhof, A.M., <i>et al.</i>	Mozambique	-	FJ809946.1	37
Canales, M., <i>et al.</i>	Mozambique	-	EU191620	50
		Zap. 11	HQ014398.1	
		Zap. 10	HQ014392.1	
		Zap. 05	HQ014391.1	
		Zap. 01	HQ014393.1	
		Zap. 03	HQ014389.1	
		Zap. 12	HQ014394.1	
Freeman, J.M., <i>et al.</i>	Texas	Zap. 02	HQ014386.1	27
		Starr 05	HQ014388.1	
		Starr 02	HQ014397.1	
		Starr 03	HQ014385.1	
		Starr 01	HQ014396.1	
		Hid 01	HQ014395.1	
		Kinne 01	HQ014401.1	
		Dimm.01	HQ014399.1	
Andreotti, R., <i>et al.</i>	Brazil	Campo G.	EU352677.1	51

^aReference number listed in this work

Table 3. Results of the LDA analysis

Method ^a	Set ^b	Statistics ^c	Symbol	%	Class ^d	$\epsilon_{qj\text{-pred}} = 0$	$\epsilon_{qj\text{-pred}} = 1$	Ref. ^e
Entropy	t	Specificity	Sp	70.6	$\epsilon_{qj\text{-obs}} = 0$	415299	172838	¹ This work
		Sensitivity	Sn	80.1	$\epsilon_{qj\text{-obs}} = 1$	39488	158517	
		Accuracy	Ac	73.0	Total			
	v	Specificity	Sp	70.7	$\epsilon_{qj\text{-obs}} = 0$	138578	57469	
		Sensitivity	Sn	79.7	$\epsilon_{qj\text{-obs}} = 1$	13380	52621	
		Accuracy	Ac	73.0	Total			
Polarizab.	t	Specificity	Sp	95.91	$\epsilon_{qj\text{-obs}} = 0$	83699	3568	²
		Sensitivity	Sn	91.92	$\epsilon_{qj\text{-obs}} = 1$	5486	62416	
		Accuracy	Ac	94.17	Total			
	v	Specificity	Sp	95.97	$\epsilon_{qj\text{-obs}} = 0$	27736	1164	
		Sensitivity	Sn	91.6	$\epsilon_{qj\text{-obs}} = 1$	1859	20267	
		Accuracy	Ac	94.08	Total			
Electroneg.	t	Specificity	Sp	97.0	$\epsilon_{qj\text{-obs}} = 0$	2660	84607	³
		Sensitivity	Sn	93.6	$\epsilon_{qj\text{-obs}} = 1$	63548	4354	
		Accuracy	Ac	95.5	Total			
	v	Specificity	Sp	97.1	$\epsilon_{qj\text{-obs}} = 0$	840	28060	
		Sensitivity	Sn	93.3	$\epsilon_{qj\text{-obs}} = 1$	20641	1485	
		Accuracy	Ac	95.4	Total			

^a Numerical descriptors used to quantify peptide structure, Entropy = Shannon entropy information indices, Polab. = Average atomic polarizability, Electroneg. = Average atomic electronegativity.

^b Set = Data set; t = training set, v = external validation series.

^c Specificity = $Sp (\%) = 100 \cdot / (NT0 + NF1)$;

Sensitivity = $Sn (\%) = 100 \cdot NT1 / (NT1 + NF0)$;

Accuracy = $Sn (\%) = 100 \cdot [NT1 + NT0] / (\text{Number of cases})$;

NT0 = Number of true 0 = Number of cases with $\epsilon_{qj\text{-obs}} = 0$ and $\epsilon_{qj\text{-pred}} = 0$

NT1 = Number of true 1 = Number of cases with $\epsilon_{qj\text{-obs}} = 1$ and $\epsilon_{qj\text{-pred}} = 1$

NF0 = Number of false 0 = Number of cases with $\epsilon_{qj\text{-obs}} = 0$ and $\epsilon_{qj\text{-pred}} = 1$

NF1 = Number of false 1 = Number of cases with $\epsilon_{qj\text{-obs}} = 1$ and $\epsilon_{qj\text{-pred}} = 0$

^d $\epsilon_{qj\text{-obs}} = 1$ when IEDB activity = Positive-High or Positive; $\epsilon_{qj\text{-obs}} = 0$ otherwise; $\epsilon_{qj\text{-pred}} = 0$ when model predict 0.

^e Reference number for the paper reporting this model according to list of references of this paper.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 4. Selected values of Shannon's Entropy information measures for different assay conditions

Condition (j)	Count	Prob	Entropy
Adjuvant additives	n_j	p_j	$\theta_0(\text{Adj.})$
FREE	82466	0.9729	0.0119
CITR(R9)	123	0.0015	2.8382
CITR(R5,R11)	98	0.0012	2.9369
CITR(R5,R10,R12)	82	0.0010	3.0143
CITR(R11)	49	0.0006	3.2379
CITR(R7)	40	0.0005	3.3261
CITR(R10)	36	0.0004	3.3718
CITR(R12)	34	0.0004	3.3967
PYRE(E1)	30	0.0004	3.4510
CITR(R6)	27	0.0003	3.4968
ACET(A1)	24	0.0003	3.5479
GLUC(N7)	24	0.0003	3.5479
PHOS(S2,S8,S14,S20,S26,S32)	20	0.0002	3.6271
PHOS(S10)	20	0.0002	3.6271
Epitope Organism Name	n_j	p_j	$\theta_0(\text{Org})$
NO DATA	10745	0.1268	0.8969
<i>Homo sapiens</i>	14003	0.1652	0.7819
<i>Bos Taurus</i>	2129	0.0251	1.6000
<i>Trypanosoma cruzi strain CL Brener</i>	2080	0.0245	1.6101
<i>Plasmodium falciparum</i>	1843	0.0217	1.6626
<i>Hepatitis C virus</i>	1614	0.0190	1.7202
<i>Mus musculus</i>	1522	0.0180	1.7457
<i>Triticumaestivum</i>	1179	0.0139	1.8566
<i>Human papillomavirus type 16</i>	1163	0.0137	1.8626
<i>Hepatitis C virus subtype 1b</i>	928	0.0109	1.9606
<i>Mycobacterium tuberculosis</i>	780	0.0092	2.0360
<i>Human T-lymphotropic virus 1</i>	766	0.0090	2.0439
<i>Hepatitis B virus</i>	700	0.0083	2.0830
<i>Clostridium botulinum</i>	699	0.0082	2.0837
<i>Bordetella pertussis</i>	619	0.0073	2.1364
<i>Arachishypogaea</i>	591	0.0070	2.1565
<i>Bacillus anthracis</i>	560	0.0066	2.1799
<i>Rattusnorvegicus</i>	551	0.0065	2.1870

<i>Plasmodium falciparum</i> 3D7			
Host organism	n_j	p_j	$\theta_0(\text{Host})$
NO DATA	113	0.0013	2.8751
<i>Homo sapiens</i>	31013	0.3659	0.4366
<i>Mus musculus</i> BALB/c	17791	0.2099	0.6779
<i>Oryctolagus cuniculus</i>	6444	0.0760	1.1190
<i>Mus musculus</i>	5950	0.0702	1.1536
<i>Oryctolagus cuniculus</i> New Zealand White	5314	0.0627	1.2027
<i>Mus musculus</i> C57BL/6	1319	0.0156	1.8079
<i>Ovisaries</i>	845	0.0100	2.0013
<i>Equus caballus</i>	803	0.0095	2.0234
<i>Rattusnorvegicus</i> Lewis	779	0.0092	2.0366
<i>Susscrofa</i>	629	0.0074	2.1295
<i>Mus musculus</i> A/J	590	0.0070	2.1573
<i>Mus musculus</i> PrP null	546	0.0064	2.1909
<i>Bos Taurus</i>	514	0.0061	2.2172
<i>Mus musculus</i> SJL	473	0.0056	2.2533
<i>Rattusnorvegicus</i>	368	0.0043	2.3623
In vivo Process Type	n_j	p_j	$\theta_0(\text{IVPT})$
NO DATA	0	0.0000	4.9281
Administration <i>in vivo</i>	51418	0.6066	0.2170
No immunization	3318	0.0391	1.4073
Occurrence of allergy	5016	0.0592	1.2278
Unknown	464	0.0055	2.2616
Occurrence of infectious disease	11939	0.1409	0.8512
Occurrence of autoimmune disease	6264	0.0739	1.1313
Administration <i>in vivo</i> to prevent or reduce disease	480	0.0057	2.2469
Occurrence of disease	988	0.0117	1.9334
Transplant/transfusion	381	0.0045	2.3472
Administration <i>in vivo</i> to cause disease	215	0.0025	2.5957
Environmental exposure to endemic/ubiquitous agent without evidence for disease	1531	0.0181	1.7432
Exposure with existing immune reactivity without evidence for disease	2183	0.0258	1.5891
Exposure without evidence for disease	346	0.0041	2.3891
Documented exposure without evidence for disease	88	0.0010	2.9837

	Occurrence of cancer	128	0.0015	2.8209
1	Experimental techniques	n_j	p_j	$\theta_0(ET)$
2				
3	NO DATA	0	0.0000	4.9281
4	ELISA	44611	0.5263	0.2787
5	Western Blot	11877	0.1401	0.8534
6	Antigen Inhibition	4503	0.0531	1.2746
7	Microarray	4491	0.0530	1.2758
8	Biological Activity	3918	0.0462	1.3351
9	Phage Display	3470	0.0409	1.3878
10	Immuno Staining	2802	0.0331	1.4807
11	Radio Immuno Assay (RIA)	2261	0.0267	1.5738
12	Surface Plasmon Resonance (SPR)	1835	0.0216	1.6645
13	Assay	1508	0.0178	1.7497
14	Immunoprecipitation	816	0.0096	2.0164
15	Flow Cytometry	639	0.0075	2.1226
16	Cross Blocking	381	0.0045	2.3472
17	X-Ray Crystallography	274	0.0032	2.4904

^aBoundary conditions: Epitope Organisms, Host Organisms, Types of *In vivo* Process, Experimental Techniques, and Adjuvant additives.

^bFrequency of each assay conditions.

^cProbability of the Frequency of each assay conditions.

^dEntropy values for each assay conditions.

18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 5. Experimental results of the genetic variants isolated from Colima

Var ^a	SAP's ^b	Identity (%) ^c	Num ^d
1	22/613	3.6	5
2	19/613	3.1	1
3	20/613	3.3	1
4	20/613	3.3	1
5	22/613	3.6	1
6	18/613	2.9	1
7	26/613	4.2	2
8	33/613	5.4	2
9	22/613	3.6	1

^aIdentification number assigned to the variant.

^bSingle Amino-ácid Polymorphism which differ from the strain of reference Yeroongpilly.

^cPercentage aminoacidic diversity vs. Yeroongpilly strain.

^dNumber of individuals where the variant was isolated.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

10.2 Artículo enviado.

10.2.1 Resumen

Las garrapatas son un problema mundial para la producción de ganado en regiones tropicales y subtropicales porque generan grandes pérdidas económicas. Ixodicidas y vacunas se han utilizado para tratar de mantener a las poblaciones de garrapatas bajo control. Sin embargo, el control se ha vuelto difícil debido a la resistencia de la garrapata a los ixodicidas y a la menor efectividad de la vacuna. La resistencia a las vacunas con proteína rBm86 se ha asociado con la diversidad genética de Bm86 entre las poblaciones de ectoparásitos. Hasta ahora no se ha estudiado la diversidad genética ni la distribución espacial de los haplotipos Bm86 circulantes dentro del territorio nacional. Aquí, exploramos la diversidad genética de 125 secuencias de cDNA Bm86 de *R. microplus* provenientes de 10 áreas endémicas de México mediante el análisis de los patrones de distribución de haplotipos para ayudar a comprender la estructura genética de la población de garrapatas mexicanas. Nuestros resultados mostraron una identidad de nucleótidos promedio entre los aislamientos mexicanos de 98.3%, variando de 91.1 a 100%. La divergencia entre las secuencias de México y Yeerongpilly (el antígeno Bm86 de referencia) varió de 3.1 a 7.4%. La distribución geográfica observada para los genotipos Bm86 en México, nuestros resultados sugieren la ocurrencia de flujo genético en diferentes regiones del territorio mexicano, e incluso en los Estados Unidos. El polimorfismo de Bm86 encontrado en las poblaciones incluidas en este estudio podría explicar la poca eficacia de la vacuna comercial actual basada en el antígeno Bm86 en muchas regiones de México. Nuestros datos pueden contribuir al diseño de nuevos candidatos vacunales antigénicos altamente específicos contra la Bm86 de *R. microplus* que circulan en México.

Haga un seguimiento de su presentación reciente en coautoría a IJPARA



Int J Parasitol <eesserver@eesmail.elsevier.com>
Hoy, 01:39 a.m.
Tú

Responder | v

Estimado Dr. Martínez-Arzate Saúl Gabriel,

Usted ha sido incluido como coautor de la siguiente comunicación:

Diario: Revista Internacional de Parasitología

Título: Diversidad genética de las secuencias Bm86 de *Rhipicephalus (Boophilus) microplus* garrapatas en México: Análisis de distribución de haplotipos patrones

Autor correspondiente: Juan Carlos Vázquez Chagoyan Coautores

: Martínez-Arzate Saúl Gabriel; Juan Carlos Sánchez-Bermúdez; Samuel Sotelo-Gómez; Héctor Díaz-Albiter; Wael Hegazy-Hassan; Esvieta Tenorio-Borroto; Alberto Barbabosa-Pliego

Para mantenerse informado sobre el estado de su presentación, regístrese o inicie sesión (si ya tiene un perfil de Elsevier).

Registrar aquí: <https://ees.elsevier.com/ijpara/default.asp?acw=&pg=preRegistration.asp&user=coauthor&fname=Martínez-Arzate> & Inname = Saúl Gabriel &email=gabosaeta_@hotmail.com

O inicie sesión: https://ees.elsevier.com/ijpara/default.asp?acw=&pg=login.asp&email=gabosaeta_@hotmail.com

Si no ha sido coautor de este envío, no siga el enlace anterior, sino que póngase en contacto con el autor correspondiente de este envío al jcvch@yahoo.com.

Gracias,

Revista Internacional de Parasitología

***Manuscript**

[Click here to view linked References](#)

1 Genetic Diversity of Bm86 sequences from *Rhipicephalus (Boophilus) microplus* ticks in Mexico:

2 Analysis of haplotype distribution patterns

3 Martínez-Arzate SG^a; Sánchez-Bermúdez JC^a; Sotelo-Gómez S^a; Díaz-Albiter HM^b; Hegazy-Hassan W^a; Tenorio-

4 Borroto E^a; Barbabosa-Pliego A^a; Vázquez-Chagoyán JC^a.

5 Centro de Investigación y Estudios Avanzados en Salud Animal, Facultad de Medicina Veterinaria y Zootecnia^a,

6 Universidad Autónoma del Estado de México, México, Kilometer 15.5 Pan-American Highway Toluca-Atlaconulco,

7 P.C. 50200.

8 Wellcome Centre for Molecular Parasitology, University of Glasgow, University Place, Glasgow G12 8TA, United

9 Kingdom^b.

10 E Mail : Martínez-Arzate SG gabosaeta_@hotmail.com; Sánchez-Bermúdez JC jcarlsanzberz@gmail.com; Sotelo-

11 Gómez S samsotelogom@gmail.com; Héctor M. Díaz-Albiter hector.diazalbiter@glasgow.ac.uk; Hegazy-Hassan W

12 hegazi_wael@yahoo.com; Tenorio-Borroto E esvieta@gmail.com; Barbabosa-Pliego A abarbabosa@yahoo.com.mx;

13 Vázquez-Chagoyán JC jcvch@yahoo.com.

14 **Corresponding author:** Vázquez Chagoyán Juan Carlos; e-mail address: jcvch@yahoo.com; Telephone number:

15 +5217225104075; Center for research and advanced studies in animal health, Faculty of Veterinary Medicine and

16 Zootechnics, autonomous Mexico State University, Mexico, Kilometer 15.5 Pan-American Highway Toluca-

17 Atlaconulco, P.C. 50200.

DNA and amino acid sequences reported in this paper were deposited in the GenBank repository. However, sequence data and accession numbers for submitted sequences are kept confidential.

Declarations of interest: none.

Note: Supplementary data associated with this article

1

18

19 **Abstract**

20 Ticks are a global problem for cattle production in tropical and subtropical regions because they generate great
21 economic losses. Ixodocides and vaccines have been used to try to keep tick populations under control. This has been
22 proven difficult given the resistance to ixodocides and vaccines observed in ticks. Resistance to protein rBm86-based
23 vaccines has been associated with the genetic diversity of Bm86 among the ectoparasite's populations. So far it not has
24 been studied genetic diversity neither spatial distribution of circulating Bm86 haplotypes within the national territory.
25 Here, we explored the genetic diversity of 125 Bm86 cDNA gene sequences from *R. microplus* from 10 endemic areas
26 of Mexico by analyzing haplotype distribution patterns to help in understanding the population genetic structure of
27 Mexican ticks. Our results showed an average nucleotide identity among the Mexican isolates of 98.3%, ranging from
28 91.1 to 100%. Divergence between the Mexican and Yeerongpilly (the Bm86 reference vaccine antigen) sequences
29 ranged from 3.1 to 7.4%. Based on the geographic distribution of Bm86 genotypes in Mexico, our results suggest gene
30 flow occurrence within different regions of the Mexican territory, and even the USA. The polymorphism of Bm86
31 found in the populations included in this study, could account for the poor efficacy of the current Bm86 antigen based
32 commercial vaccine in many regions of Mexico. Our data may contribute towards designing new, highly-specific,
33 Bm86 antigen vaccine candidates against *R. microplus* circulating in Mexico .

34 **Keywords:** Tick, *R. microplus*, Bm86, diversity, haplotype.

35 **1. Introduction**

36 Ticks are obligated hematophagous ectoparasites that feed on a wide variety of mammals, including humans and cattle.
37 *Rhipicephalus microplus* (Acari: Ixodidae) is a cattle one-host-tick, with a tropical and subtropical cosmopolitan
38 distribution. It is considered one of the most important ectoparasites of cattle in Mexico, negatively affecting weight
39 gain and milk production which results in substantial economic losses. Additionally, this species is able to transmit
40 dangerous diseases to cattle and humans, such as babesiosis and anaplasmosis (Almazán et al., 2010; Lew-Tabor et al.,
41 2014).

42 In Mexico, eradication campaigns against *R. microplus* have been conducted for over 60 years (Senasica, 2015).
43 Campaigns included strategies aiming to keep geographic regions free of ticks, as well as actions towards control and
44 eradication of the ectoparasite from endemic areas (Mexican Official Standard NOM-019-ZOO-1994). Such actions are
45 achieved mostly through application of acaricides which are currently presenting limited efficacy, as ticks have

2

46 developed resistance, an issue that has been widely discussed in previous researches (Cruz et al., 2010; Foil et al., 2004;
47 Lovis et al., 2013; Rodríguez-Vivas et al., 2006, 2006; Róger Iván Rodríguez-Vivas et al., 2006).

48 This is an escalating problem in Mexico for cattle production, due to the presence of various drug-resistant and multi-
49 resistant ticks (Alonso-Díaz et al., 2006; Domínguez-García et al., 2010; A. Fernández-Salas et al., 2012; Agustín
50 Fernández-Salas et al., 2012; Perez-Cogollo et al., 2010; R. I. Rodríguez-Vivas et al., 2014; Rodríguez-Vivas et al.,
51 2014; Róger Iván Rodríguez-Vivas et al., 2014). Furthermore, the continuous use of chemicals in tick control results in
52 environmental pollution, as well as contamination of dairy and meat products (Ben Said et al., 2012; Graf et al., 2004;
53 Kunz and Kemp, 1994).

54 Given the variability of resistance mechanisms to acaricides presented by ticks, alternative control methods have been
55 explored. These include pheromone decoys for attracting and killing ticks, selection of cattle with natural resistance to
56 ticks, biological control agents, as well as the administration of anti-tick vaccines (Kaewmongkol et al., 2015;
57 Sonenshine et al., 2006). The successful use of vaccines to combat ticks was first published by Allen and Humphreys
58 using tick midgut extracts (Allen and Humphreys, 1979). Bm86 is a tick midgut protein, which has been used since the
59 1990s as a vaccine antigen (Rand et al., 1989; Rodríguez et al., 1994; Willadsen, 2004). Three vaccines with protection
60 efficacy against *R. microplus* have been commercialized since then: Gavac® (Heber Biotec, Havana, Cuba),
61 TickGARD® (Hoechst, Animal Health; Australia), and TickGARD PLUS® (Intervet, Australia), (de la Fuente et al.,
62 2007, 1998; Guerrero et al., 2012; Kaewmongkol et al., 2015; Rand et al., 1989; Willadsen, 2004; Willadsen et al.,
63 1995).

64 The rBm86 protein-based vaccine, is marketed in several countries of Latin America, Asia and Oceania, showing a
65 variable efficacy. In Mexico, the commercial vaccine has exhibited a lower efficacy compared with countries such as
66 Australia and Cuba (Andreotti et al., 2008; de la Fuente et al., 2007; Freeman et al., 2010; García-García et al., 2000;
67 Nijhof et al., 2009; Pedroso et al., 2007; Rand et al., 1989). The effectiveness of the vaccine has recently been
68 associated with amino acid variations of the protein among different populations of *R. microplus* (Ben Said et al., 2012;
69 de la Fuente et al., 2007; de La Fuente and Kocan, 2006; Freeman et al., 2010; Kaewmongkol et al., 2015). This was
70 observed in a particular variant of the Bm86 antigen in Argentina. The antigen was renamed as Bm95, and it is
71 currently used as an alternative vaccine antigen (García-García et al., 2000).

72 Genetic polymorphism has resulted in diverse Bm86 haplotypes in different geographical regions of the world (G. S.
73 Cobon et al., 1997; García-García et al., 2000, 1999; Patarroyo et al., 2002; Willadsen, 2004), so a vaccine should be

74 designed considering regional antigenic variants. Exploring the haplotypes and haplogroups of the Bm86 sequences and
75 their distribution within the genetic phylogeographic structure, could help in designing better vaccines against *R.*
76 *microplus*. In this study, phylogeographic aspects such as the spatial distribution of haplotypes and the genetic
77 variability of a cDNA transcript sequence for the Bm86 gene from *R. microplus* collected from 10 Mexican states from
78 will be assessed.

79 2. Materials and methods

80 2.1 Experimental methods

81 2.1.1 Sampling

82 Cattle ticks (*R. microplus*), were collected from areas under official tick control, according to the National Campaign to
83 Control *Boophilus spp* ticks (Senasica, 2013). Samples from 10 states of Mexico were included (Fig 1).

84 Insert Fig. 1 here

85

86 2.1.2 RNA extraction

87 Total RNA was extracted and purified from midguts using TRI Reagent® (Ambion, USA), according to the
88 manufacturer's instructions, as follows: 1) Homogenization: tissue (10-15 mg midgut) from individual ticks was placed
89 in 1 ml of TRI Reagent solution, homogenized by vortex, followed by centrifugation at 12,000xg, at 4°C for 10
90 minutes, the supernatant was collected in a fresh microcentrifuge tube and stored at -80 °C until further use, or used
91 immediately, 2) RNA extraction: 100 µl of 1-bromo-3-chloropropane (BCP) per 1mL of TRI Reagent solution were
92 added, followed by centrifugation 12,000xg, at 4°C for 15 minutes and the aqueous phase was transferred to a fresh
93 tube, 3) RNA precipitation: RNA was washed and resuspended by adding 500 µl of isopropanol followed by
94 centrifugation at 12,000xg, at 4°C for 8 minutes, the supernatant was discarded, the RNA pellet washed with 1mL of
95 75% ethanol, followed by centrifugation 7,500xg, at 4°C for 5 minutes and the ethanol was discarded and the pellet left
96 to air dry 2-3 minutes, 4) RNA solubilization: the pellet, was resuspended in 40 µl of nuclease free water and stored at
97 -80 °C until further use (Andreotti et al., 2008; Ben Said et al., 2012; Nijhof et al., 2010).

98 2.1.3 cDNA synthesis

99 To synthesize complementary DNA (cDNA), SuperScript II Reverse Transcriptase (Invitrogen, USA) was used
100 according to the manufacturer's instructions. The reaction mixture for a total volume of 12 μ L was: 8 μ L of nuclease-
101 free water, 1 μ L of Oligo (dT) 500 μ g/ml, 2 μ L of total sample RNA (at 2000 ng/ μ l), 1 μ L of dNTPs (10mM). The
102 mixture was incubated in a thermocycler (Techne TC-512 USA) at 65° C for 5 min and then 4° C for 2 min. Then, 4 μ L
103 of 5 X First-Strand Buffer and 2 μ L of DTT were added and incubated at 42° C for 2 min. Finally, 1 μ L reverse
104 transcriptase SuperScript II was added, followed by a 42° C incubation for 50 minutes. (Martínez-Arzate et al., 2017;
105 Peconick et al., 2008; Temeyer et al., 2012). RNA samples were subjected to a DNA digestion according to the DNA-
106 free™ Kit protocol (Ambion).

107 2.1.4 Amplification of the target sequences and sequencing

108 To obtain the Bm86 cDNA target sequence, four pairs of primers (see Table 1) were designed using the Primer3
109 software (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>). Primers were designed using the
110 sequence of *R. microplus* reference strain Yeerongpilly, (NCBI access number M29321.1) as a template.

111 The primers were designed to produce 542-686bp amplicons with 200bp overlapping flanking regions to cover the
112 entire 1950bp cDNA sequence. The PCR thermal cycling reactions were performed in an automated thermal cycler
113 (Arktik, 96-well gradient; Thermo Scientific) as follows: an initial step of denaturation at 94°C for 4 min, followed by
114 30 cycles of denaturation at 94°C for 1 min, alignment temperature for 30 seconds (see alignment temperatures shown
115 for each set of primers in Table 1) and extension at 72°C for 1 min, and a final extension incubation at 72°C for 5 min.
116 Bm86 amplicon fragments were visually evaluated for size and integrity in a 1.5% agarose gel electrophoresis (30 min
117 at 100V) and stained with Ethidium Bromide (Ben Said et al., 2012; Martínez-Arzate et al., 2017).

118 **Table 1.** Primers sequences used to amplify the Bm86 cDNA by rt-PCR

Primer	^a Sequence	^b Tm/sec
F1	5'-GTGGTTCGACGCAGTGAGAT-3'	60°/30
F2	5'-CTGCAAAGACCTCTGCGAGA-3'	55°/30
F3	5'-CGGGCCCAAATGTCAACATC-3'	
Bm86_ F4	5'-AAAACGAGCAGTCGGAGTGT-3'	57°/30
R1	5'-CCATGCTTGCAGACAAACCC-3'	60°/30
R2	5'-TGTCCTGCGTGAGTAAAGT-3'	55°/30
R3	5'-AACGCACTCCAGCTTCTTGT-3'	

R4 5'-GGTGTTCGATGTAAGCGTGA-3' 57°/30

119 ^aPrimer sequence, directionality is indicated as forward (F) and reverse (R).

120 ^bMelting temperature in Celsius degrees/incubation time in seconds

121 PCR product concentration was estimated with a spectrophotometer (Quawell Q5000). Approved samples were packed
122 individually (50ng in a 10µL) in labeled PCR tubes and shipped to Genscript (Genscript Corporation, Piscataway, NJ,
123 USA; www.genscript.com), to be sequenced using the Sanger method.

124 2.2 Bioinformatics

125 2.2.1 *In silico* sequence analysis

126 DNA sequences were assembled and analyzed initially with the CodonCode Aligner software (CodonCode Corporation,
127 www.codoncode.com) to verify sequences reliability. Complete Bm86 cDNA sequences from each tick specimen
128 analyzed were retrieved and included in a data matrix in Mega6 software (Tamura et al., 2013) together with all the
129 cDNA Bm86 sequences available in GenBank at the time (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Morgulis et al.,
130 2008; Tamura et al., 2013; Zhang et al., 2000). Subsequently, a multiple alignment of all these sequences was
131 performed with Mega 6 software using the global ClustalW method. This data matrix was used for different analyses.

132 2.2.2 *Analysis of genetic variability*

133 Genetic variability of sequence Bm86 was analyzed in *R. microplus* collected from different localities in Mexico. The
134 trend of the haplotypes in the population was analyzed through nucleotide (the probability that two randomly-chosen
135 homologous sites are different), and haplotype (the probability that two randomly-chosen haplotypes are different)
136 diversity. The number of haplotypes and values of nucleotide and haplotype diversity were calculated for each of the
137 collection sites with the DNA sequence polymorphism software DnaSP 5.10.
138 (http://www.ub.edu/dnasp/DnaSP_OS.html) (Librado and Rozas, 2009).

139 2.2.3 *Haplotype frequency*

140 To evaluate the haplotype frequency in Mexican Bm86 sequences, we first specified the haplogroups and the frequency
141 in which a determined haplotype appears in each haplogroup. Then, we determined the genetic constitution, as the
142 proportion of sequences that belong to a haplogroup. Additionally, we calculated the degree of divergence of each
143 haplotype, with respect to the reference sequence (Yeerongpilly). This degree of divergence corresponds to the sum of

144 single nucleotide polymorphisms (SNPs), when comparing two aligned sequences, expressed as a percentage (Chen et
145 al., 2012; Zhao and Gong, 2015).

146 2.2.4 Mantel's test

147 A Mantel's test was performed to explore whether a significant relationship existed between genetic and geographic
148 distances among the sampled locations. The postulate is an island-like model, in which each population diverges at the
149 same time. This was carried out to evaluate the haplotypes' spatial structure, specifically if the observed haplotypes
150 were confined to a geographic region in Mexico.

151 To generate the genetic distances matrix, an average value of the identity percentage was used for all possible
152 population pairs. These values were calculated with MatGAT v2.01 (Matrix Global Alignment Tool) software
153 (Campanella et al., 2003) (<http://www.angelfire.com/nj2/arabidopsis/MatGAT.html>).

154 The matrix of geographic distances was constructed considering the geographic coordinates from each locality sampled
155 and data fidelity was confirmed using Google Earth software (<https://www.google.com/earth/download/ge/>). The matrix
156 represents the distance in km between each pair of tick populations analyzed. The data matrices were analyzed using the
157 statistical software XLSTAT (<https://www.xlstat.com/es/>). The Pearson correlation function was used, which
158 determines if there is a linear relationship between two variables. The correlation coefficients between the elements of
159 each matrix were obtained using 10,000 permutations and a significance of 95%. The hypotheses considered were: H_1
160 There is an isolation by distance for the sequence Bm86 in the different sampled populations; H_0 There is not an
161 isolation by distance for sequence Bm86 in the different sampled populations.

162 2.2.5 Phylogenetic analysis

163 Reconstruction of the evolutionary distances data between putative sequences of sequence Bm86 in Mexico was
164 performed using a neighbor-joining algorithm. This method is based on grouping pairs of operational taxonomic units
165 (OTUs [= neighbors]), always trying to minimize the total length of the tree branch at each stage of clustering of OTUs
166 (Bandelt et al., 1999; Saitou and Nei, 1987). The tree was built using the distances of Tamura-Nei, adding the parameter
167 G with 1000 permutations in the Mega 6 software (Tamura et al., 2013). In addition to our isolates, the analysis
168 included complete Bm86 sequences from different regions of the world retrieved from GenBank. These included
169 Australia (M29321), Argentina (AF150891), India (HQ166286), Mozambique (FJ809946, EU191620); Brazil,
170 (EU352677), 14 from the USA (HQ014385 to HQ014398) and 18 from Thailand (KJ995882 to KJ9958910). As an
171 external groups, two *R. annulatus* Bm86 homologous sequences, (HQ014399 and HQ014401) were included, (Freeman
172 et al., 2010).

173 A minimum spanning tree (MST) or Minimal Expansion Network, was built using the Network publisher software
174 (<http://www.fluxus-engineering.com>) (Bandelt et al., 1999). For this analysis, we used the total number of haplotypes
175 observed along with the global reports retrieved from GenBank. The coalescence and the connections between the
176 haplotypes, given by the minimum number of mutations among them, were the basis of the analysis to generate a tree
177 that shows the frequency and the relation with the distances in number of mutations that separate each haplotype from
178 others. The Median Joining (MJ) method was chosen for the construction of the network, this method combines
179 characteristics of the Kruskal and Farris algorithms. The first favors the short connections in the network for finding
180 minimum spanning trees, while the second is a heuristic algorithm of maximum parsimony (MP) that adds the so-called
181 median vectors.

182 3. Results

183 3.1 Genetic variability.

184 In the present study, 125 full-length cDNA strands, encoding Bm86 mature protein code from 10 different localities in
185 Mexico were analyzed. We found 145 single-nucleotide polymorphisms. Total haplotypic diversity was estimated in h
186 = 0.017 and total nucleotide diversity in $\pi = 0.012$. Diversity values were calculated for each sampled locality. Chiapas
187 was the locality with the highest haplotypic diversity ($h = 1$) from the 10 states sampled, while Zacatecas and Nayarit
188 had the lowest value ($h = 0.84$). Regarding nucleotide diversity, the lowest value was observed in the locality of
189 Veracruz ($\pi = 0.001$) and the highest value ($\pi = 0.018$) was observed for Zacatecas (**Table 2**).

190 **Table 2** Genetic variability of the Bm86 gene in the different sampled localities

Locality	n	n_s	n_{hs}	h	π
Campeche	18	11	7	0.94	0.007
Veracruz	18	8	4	0.90	0.001
Morelos	13	12	11	0.98	0.013
Colima	14	9	7	0.89	0.014
Sinaloa	11	10	9	0.98	0.017
Guerrero	11	9	8	0.97	0.015
Nayarit	10	6	4	0.84	0.003
Chiapas	9	9	9	1	0.015
Tabasco	11	9	7	0.96	0.005
Zacatecas	10	6	4	0.84	0.018

	<i>n</i> total	125
191	<i>n</i>	Size of population
192	<i>n_h</i>	Number of observed haplotypes
193	<i>n_{hu}</i>	Number of haplotypes unique
194	<i>h</i>	Haplotype diversity
195	π	Nucleotide diversity
196		
197	64 different haplotypes were identified from the genetic variance of sequence Bm86. Nine of them formed haplogroups	
198	(see Table 3) or groups of sequences that shared a 100% identity within each group, the remaining 55 sequences, were	
199	unique haplotypes. Haplogroups 1 and 2, represent respectively 22.2% and 14.2% of the total number of sequences	
200	analyzed, both were observed in 9 of the 10 sampled localities. The percentage of divergence for these 9 haplogroups	
201	remained between 3.1-7.4 % (Table 3).	

202

203 **Table 3** Haplogroups for sequence Bm86 from Mexican *R. microplus*

H ^a	F ^b	Gc (%) ^c	Fip ^d	SAP's ^e	Di (%) ^f
1	28	22.2	9/10	22/621	3.5
2	18	14.2	9/10	19/621	3.1
3	6	4.7	4/10	19/621	3.1
4	6	4.7	4/10	22/621	3.5
5	4	3.1	3/10	46/621	7.4
6	3	2.3	2/10	21/621	3.4
7	2	1.5	2/10	20/621	3.2
8	2	1.5	1/10	21/613	3.4
9	2	1.5	1/10	26/613	4.2
10-64	1	0.8	1/10	-	-

204 ^aHaplogroup number.

205 ^bFrequency of haplotype.

206 ^cGenetic constitution.

207 ^dFrequency of inter-population haplotype.

208 ^eSingle amino-acid polymorphism with respect to Yeerongpilly strain.

9

209 ^fPercentage of divergence with respect to Yeerongpilly strain.

210 3.2 Correlation between genetic and geographic distances

211 A Mantel test was performed to find out the level of correlation existing between the matrices of genetic distances and
212 geographic distances among all studied localities. The results of the test indicate that there is no correlation between
213 genetic ($r = 0.014$) and geographical ($p = 0.92$) matrices distance. Accepting the null hypothesis, there is no distance
214 isolation of the Bm86 gene, in the different sampled populations ($p > 0.05$).

215 A scatter plot (**Fig 2**) was used to visually evaluate the correlation or cause-effect relationship for geographic and
216 genetic distance matrices. The distribution of points in the diagram does not show a linear relationship between the
217 variables.

218 

219 The Pair-wise alignment in the identity matrix for amino acid sequences revealed that Texan sequences were similar to
220 some of the Mexican ones. Comparison of Identity values (See **Supplementary Table S1**, Comparison of Identity
221 values.) shows identities ranging between 93-100% among Mexican and Texan sequences, while identities were 91-
222 94% between sequences from Mexico and Thailand, and 86-88% between Mexican Bm86 and the Argentinian Bm95
223 sequences. Nucleotide sequences identity between Mexican and reference strain Yeerongpilly ranged from 96.9 to
224 92.6%.

225 3.3 Minimal Expansion Network

226 The MST diagram of the haplotypes does not show a defined phylogeographic structure. Fig 3 shows the phylogenetic
227 structure of the 64 haplotypes found in Mexican isolates and those retrieved from GenBank. It should be noted that the
228 size of the branches does not reflect the number of mutations of the derivatives. The diagram shows 9 haplogroups,
229 however, haplogroups 1 and 2 are notable, not only for containing the most abundant haplotypes, but also because they
230 show a pattern of dispersion within the localities sampled. Haplogroup 1 contains 28 sequences from Veracruz,
231 Chiapas, Sinaloa, Morelos, Campeche, Colima, Zacatecas, Tabasco, and Nayarit. In comparison, haplogroup 2 contains
232 18 sequences from the same localities of haplogroup 1, although it excludes sequences from Chiapas and includes
233 sequences from Guerrero. The rest of the haplotypes found in the present study are unique haplotypes.

234 Regarding the sequences found in GenBank, two of the of the Texan sequences are contained in haplogroup 1 and two
235 more in haplogroup 2, the rest of the Texan sequences and Latin American reports: Argentina, Brazil, USA in addition
236 to the only Mexican previous report show a random distribution. In a more prolonged expansion, by putative nodes of
237 divergence, it can be observed that the two sequences from Mozambique arise from haplogroup 1. Something similar
238 occurs with the Yeerongpilly reference sequence, which can be included in haplogroup 2. Lastly, in a monophyletic
239 cluster, we can see all the sequences of Thailand arising from haplogroup 5

240 Insert figure 3 here

241 3.4 Phylogenetic analysis

242 The phylogenetic tree constructed with the Bm86 complete cDNA sequences found in the present report, plus the
243 sequences available in GenBank, shows two main clades with a common possible ancestor: Clade **A** (the smaller) and
244 clade **B** with five subclades (B1-B5). Clade **A** shows an important degree of divergence, it is characterized by long
245 branches, and is mainly represented by sequences from Thailand (Thai) (N=16), Zacatecas (n=2), Guerrero (N=1),
246 Morelos (N=1), and Sinaloa (N=1). In this clade we can observe also the outer group, represented by three sequences
247 from Texas, two of them classified as *R. anulatus* (Kinne and Dimmi) and one reported as *R. microplus* (Hid Tx) that
248 was probably misclassified. Within clade **B**, we found most of the sequences of the present report as well as most of the
249 sequences reported for the Americas, two sequences from Mozambique, two from Thailand, one from India and one
250 from Australia. Subclade **B** has two homogeneous subclades and three clades formed by long branches with larger
251 levels of divergence. Subclade **B1** is formed by sequences from Campeche (N=9), Chiapas (N=6), Zacatecas (N=1),
252 Tabasco (N=1), Colima (N=3), Morelos (N=1), Sinaloa (N=1), and sequences from Texas (USA) (N=2), Australia
253 (Yeerongpilly) (N=1), Mozambique (N=2), India (N=1) and Thailand (N=2). Subclade **B2** is more homogeneous and
254 includes individuals from Colima (N=3), Veracruz (N=8), Tabasco (N=5), Campeche (N=3), Morelos (N=3), Nayarit
255 (N=3), Zacatecas (N=1), Sinaloa (N=1), Chiapas (N=2), Tamaulipas (N=1, the only previous report from Mexico; Sus
256 Mex) and also samples from Texas, USA (N=4). The small subclade **B3** includes sequences from Tabasco (N=2),
257 Veracruz (N=1) and Colima (N=1). The Bm85 sequence from Argentina stands on its own in a branch, it does not seem
258 to be part of any of the other subclades identified for the Americas. Subclade **B4**, represented by a group with divergent
259 branches of different sizes, is formed by sequences from Guerrero (N=10), Sinaloa (N=5), Morelos (N=4), Nayarit
260 (N=2), Colima (N=2) and Chiapas (N=1). Finally, subclade **B5**, is a homogeneous group formed with sequences from
261 Sinaloa (N=3), Morelos (N=2), Nayarit (N=5), Colima (N=6), Zacatecas (N=6), Veracruz (N=10), Campeche (N=6),
262 Tabasco (N=5) and also by sequences from Texas (USA) (N=4) and Brazil (N=1).

263 Insert figure 4 here

264 **4. Discussion**

265 There are two commercial vaccines based on the recombinant antigen Bm86: Tick-GARD™ (Australian antigen
266 Yeerongpilly) and GAVAC® (Cuban antigen Camcord, available in Mexico). However, the Camcord sequence is not
267 available in GenBank and considering that these two protein sequences are identical except for a single amino acid
268 (Freeman et al., 2010), we considered the Yeerongpilly as reference sequence for all analyses and comparisons purposes
269 in this work.

270 The lack of specificity of the commercial anti-tick vaccines, is probably due to the existing genetic polymorphism
271 among the *R. microplus* populations, as has already been reported (G. S. Cobon et al., 1997; Cunha Casquero et al.,
272 2012; de La Fuente et al., 2006; Freeman et al., 2010; García-García et al., 1999; Kaewmongkol et al., 2015; Lima et
273 al., 2000; Rodríguez et al., 1994; Willadsen, 2004). In Mexico, it was previously suggested that no significant
274 differences existed between Bm86 protein sequences from Mexican tick populations and the rBm86 protein derived
275 from Yeerongpilly, (Willadsen, 2004). This is inaccurate. The analysis carried out in that work, that correlated genetic
276 diversity and efficacy, was based in only two (partial) Bm86 protein sequences (35 amino acids, Tuxpan and Mora;
277 GenBank accession numbers AF150892 and AF150893, respectively) (García-García et al., 1999). Other than those two
278 sequences, only one single Mexican complete cDNA Bm86 sequence (1827pb and 608 aa, GenBank:FJ456928)
279 (Canales et al., 2009) had been reported at the time the present study was conducted. Our findings regarding the genetic
280 variability found within all Mexican populations sampled, challenge the previous knowledge on the influence of genetic
281 polymorphism over the effectiveness of the commercial vaccines in Mexico.

282 Here, 125 complete cDNA sequences of the Bm86 gene obtained from tick specimens collected from 10 different
283 regions of Mexico (Fig1) were analyzed. Table 2 shows the genetic variability of the Bm86 gene observed in the
284 studied localities and the number of samples studied per region is indicated. Not all samples analyzed by rtPCR
285 generated amplicons. This might be partially explained by sample handling and processing. However, this could also be
286 associated with primers specificity, particularly if one considers 1) the polymorphism of sequence Bm86 observed in
287 the present study, and 2) the presence of samples that yielded 2 or 3 of the 4 amplicons that covered the expected Bm86
288 sequence, but that were not included in the study as they were not complete (data not shown). These RNA samples
289 should be reanalyzed with new sets of primers to validate whether the lack of amplification was due to allele
290 polymorphism. Table 2 shows the total number of haplotypes and the number of unique haplotypes related to the
291 number of sequences studied per region. The polymorphism of Bm86 seems high considering that about 50% of the

12

292 haplotypes found are unique. However, this polymorphism is not extraordinarily high, as it is consistent with that
293 observed in similar studies developed in other regions of the world, such as those conducted in Texas and Thailand
294 (Freeman et al., 2010; Kaewmongkol et al., 2015).

295 The divergence of the commercial antigen compared against the sequences in this study was higher than 2.8% (Table
296 3). This value has been estimated as the higher limit of differences in amino acid sequences that allows efficient vaccine
297 performance, (García-García et al., 1999). This result also suggests that the low effectiveness of the commercial vaccine
298 in Mexico might be explained by Bm86 polymorphism. Although we expected some divergence between the sequences
299 from Mexican ticks and the one used in the commercial vaccine, it was somehow surprising not to find *R. microplus*
300 specimens in Mexico with the same haplotype as the one from commercial antigens. This could be explained by the fact
301 that the commercial vaccine has been used for over 15 years in Mexico, so it is plausible that such negative selection
302 could have eliminated a large amount of individuals with this genotype from the tick population. More extensive studies
303 covering other regions of Mexico and with higher sample numbers should be carried out to test this hypothesis. Table 3
304 also shows 9 haplogroups formed with the Bm86 sequence found in the samples collected in Mexico. Here, it can be
305 noted that haplotypes 1 and 2 are the most robust, and that they are present in 9 out of the 10 of the regions studied. The
306 seven other haplogroups are formed from at least two individuals and the remaining 55 sequences form unique
307 haplotypes. This information is relevant because it highlights the sequences of haplogroups that should be prioritized for
308 vaccine design. The divergence found for these 9 haplogroups ranged from 3.1 to 7.4%. This is an interesting finding
309 because, as mentioned before, it has been shown that when the rBm86 vaccine antigen sequence show differences
310 higher than 2.8% compared with the Bm86 target amino acid sequence, the vaccine will show a reduced effectiveness
311 (G. Cobon et al., 1997; García-García et al., 1999; Kaewmongkol et al., 2015; Rodríguez et al., 1994; Willadsen et al.,
312 1995). These data suggest that sequences from more than one haplogroup should be considered to design a vaccine that
313 aims to be effective against most *R. microplus* tick populations.

314 The Mantel test (Fig. 2) revealed no correlation between the genetic distances matrix of the Bm86 sequence and the
315 geographical distances matrix of the studied regions, so the observed haplotypes are not confined to a region. This data
316 suggest that a vaccine design should include the epitopes for most representative haplogroups.

317 The Minimum expansion network shows the frequency and the relation kept among haplotypes and haplogroups in a
318 diagrammatic way. This diagram exhibits the existence of 9 defined haplogroups with Mexican sequences found in all
319 of them. In this network, haplogroups 1, 4, 6 and 7 are closely related among them, haplogroups 2 and 3 are closely
320 related between them, and all other haplogroups remain somehow isolated. The distribution of the minimal expansion

321 network suggests the occurrence of a significant expansion event for the Bm86 sequence, and such distribution suggests
322 a high level of gene flow and Bm86 polymorphism for the analyzed locations. Since the phylogenetic tree and the
323 minimal expansion network showed a random distribution of the haplotypes, it is not possible to define a
324 phylogeographic pattern related to divergence.

325 The Neighbor-joining algorithm used for the reconstruction of the phylogenetic tree (Fig 4) shows two main clades: the
326 Americas and Thailand. However, five sequences from Mexico are present in Thailand's clade and two sequences from
327 Thailand are found in the Americas' clade. This sequence distribution could be attributed to commercialization of cattle
328 that was not appropriately inspected for tick control. The five Thailand-type genotypes found in Mexico were collected
329 from four states: Guerrero, Sinaloa, Morelos and Zacatecas. Considering that 1) sample size was not designed to detect
330 particular genotypes, 2) that it was not large enough to determine the real distribution of Thailand-type genotypes, and
331 3) that the dispersion of these haplotypes was observed in four distant regions, it can be hypothesized that this kind of
332 genotypes should have a larger distribution than the one observed in the present report, and that an additional study
333 should be conducted to find out a more precise distribution of Thailand-type genotypes in Mexico.

334 On the other hand, some genotypes found in Thailand were found to belong to the Americas' clade and are present in
335 subclade B1, together with sequences from Australia (1, strain Yeerongpilly), Mozambique (2), India (1) and Mexico
336 (5, sequences from three states). These are the genotypes more closely related to the vaccine sequence and it could be
337 inferred that the commercial vaccine should be effective against these tick populations. Subclade C is constituted
338 mainly by sequences from Campeche and Chiapas, two geographically-related states, and one sequence from the state
339 of Morelos, which, interestingly, is located ~600 km away from these states. The other two subclades (A and B),
340 include most of the Mexican sequences, several from Texas, one from Brazil, and one from Argentina. Subclade B
341 seems homogeneous. However, subclade A shows a strong divergence. Based on the phylogenetic tree information, it
342 seems that at least five different Bm86 sequences should be used to design a Mexican vaccine. However, epitope
343 prediction and protein modeling studies should be conducted before a sequence or series of sequences are chosen to
344 develop a Mexican vaccine.

345 The phylogenetic relationship between some Texan and Mexican Bm86 sequences, and the lack of phylogeographic
346 delimitation of the majority of the Mexican sequences could be explained by the movement of tick-infested livestock,
347 both within the Mexican territory and across the American border. The latter is supported by data from Skaggs *et al.*
348 (2004), showing Texas as one of the largest importer of Mexican cattle into the United States. Cattle commerce and
349 mobility is regulated by Mexican and American federal laws, and both parties are required to inspect cattle for ticks and

350 tick-borne diseases, such as babesiosis and anaplasmosis, (Senasica, 2013). Despite the inspection protocols followed to
351 prevent tick dissemination, phylogenetically-related ticks are found in Texas and different states of Mexico.
352 Epidemiologically, the tick issue at the border between Mexico and the United States might not be limited to cattle
353 commerce, since *R. microplus* is also found in wild Artiodactyla, such as deer and antelopes, and these animals are free
354 to move back and forth between both countries. In this scenario, the ticks could be dispersed by wild animals
355 throughout the Mexican and US territories despite cattle inspection performed by the tick control programs from both
356 countries. This cannot be evaluated at the moment, as there are no extensive databases recording the flow and
357 interaction of livestock and wild animals in Mexico, (Skaggs et al., 2004).

358 **Future considerations**

359 The high level of polymorphism of sequence Bm86 found in populations from 10 different states of Mexico and its
360 phylogenetic distance from the sequences used to produce the commercial vaccine currently administered to Mexican
361 cattle highlights the need to develop new anti-tick immunization agents. Future studies should aim to increase the
362 current database of Mexican Bm86 sequences. Sampling sites should include different areas from the states considered
363 for this study, as well as from other states that were not part of this work. Additionally, studies aiming to detect
364 antigenic regions and epitopes through computational analysis should be conducted to design appropriate antigens to
365 successfully develop a new, more effective, wide-spectrum anti-tick vaccine.

366

367 **Acknowledgements**

368 The authors would like to thank Universidad Autónoma de Estado de México (Grant No. UAEM 3775/2014CIC) and
369 Fundación Educación Superior Empresa (FESE, Grant number 3378/2013E) for their financial support awarded for this
370 project. Saúl Gabriel Martínez Arzate would like to thank Consejo Nacional de Ciencia y Tecnología (Scholarship No.
371 518232/291117) for their financial support during his postgraduate studies at the Universidad Autónoma del Estado de
372 México.

373

374 **References**

375 Allen, J.R., Humphreys, S.J., 1979. Immunisation of guinea pigs and cattle against ticks. *Nature* 280, 491–
376 493.

377 Almazán, C., Lagunes, R., Villar, M., Canales, M., Rosario-Cruz, R., Jongejan, F., de La Fuente, J., 2010.

- 378 Identification and characterization of *Rhipicephalus (Boophilus) microplus* candidate protective
379 antigens for the control of cattle tick infestations. *Parasitol. Res.* 106, 471–479.
- 380 Alonso-Díaz, M.A., Rodríguez-Vivas, R.I., Fragoso-Sánchez, H., Rosario-Cruz, R., 2006. Resistencia de la
381 garrapata *Boophilus microplus* a los ixodicidas. *Arch. Med. Vet.* 38, 105–113.
- 382 Andreotti, R., Pedroso, M.S., Caetano, A.R., Martins, N.F., 2008. Comparison of predicted binders in
383 *Rhipicephalus (boophilus) microplus* intestine protein variants Bm86 Campo Grande strain, Bm86 and
384 Bm95. *Rev. Bras. Parasitol. Vet.* 17, 93–98.
- 385 Bandelt, H.J., Forster, P., Rohl, A., 1999. Median-joining networks for inferring intraspecific phylogenies.
386 *Mol. Biol. Evol.* 16, 37–48.
- 387 Ben Said, M., Galai, Y., Mhadhbi, M., Jedidi, M., de la Fuente, J., Darghouth, M.A., 2012. Molecular
388 characterization of Bm86 gene orthologs from *Hyalomma excavatum*, *Hyalomma dromedarii* and
389 *Hyalomma marginatum marginatum* and comparison with a vaccine candidate from *Hyalomma*
390 *scupense*. *Vet. Parasitol.* 190, 230–240.
- 391 Campanella, J.J., Bitincka, L., Smalley, J., 2003. MatGAT: An application that generates similarity/identity
392 matrices using protein or DNA sequences. *BMC Bioinformatics* 4, 29.
- 393 Canales, M., Almazán, C., Naranjo, V., Jongejan, F., de la Fuente, J., 2009. Vaccination with recombinant
394 *Boophilus annulatus* Bm86 ortholog protein, Ba86, protects cattle against *B. annulatus* and *B.*
395 *microplus* infestations. *BMC Biotechnol.* 9, 29.
- 396 Chen, S., Xing, Y., Su, T., Zhou, Z., Dilcher, E.D.L., Soltis, D.E., 2012. Phylogeographic analysis reveals
397 significant spatial genetic structure of *Incarvillea sinensis* as a product of mountain building. *BMC*
398 *Plant Biol.* 12, 58.
- 399 Cobon, G., Moore, J.T., Johnston, L.A.Y., Willadsen, P., Kemp, D.H., Sriskantha, A., Riding, G.A., Rand,
400 K.N., 1997. DNA encoding a cell membrane glycoprotein of a tick gut. *Nucl. Phys. B* 15, 661–661.
- 401 Cobon, G.S., Levine, M.M., Woodrow, G.C., Kaper, J.B., 1997. An anti-arthropod vaccine: TickGARD – a
402 vaccine to prevent cattle tick infestations. In *New Generation Vaccines*, Marcel Dek. ed. New York.

- 403 Cruz, R.R., Inés, D., García, D., Ramírez, E.R., Estrada, M.O., Ibañez, F.M., 2010. Estrategias para el
404 control de la garrapata, 1ed. México.
- 405 Cunha Casquero, R., Pérez de León, A.A., Leivas Leite, F.P., Pinto, L. da S., dos Santos Júnior, A.G.,
406 Andreotti, R., 2012. Bovine immunoprotection against *Rhipicephalus (Boophilus) microplus* with
407 recombinant Bm86-Campo Grande antigen. Rev. Bras. Parasitol. Vet 21, 254–262.
- 408 de la Fuente, J., Almazán, C., Canales, M., Pérez de la Lastra, J.M., Kocan, K.M., Willadsen, P., 2007. A
409 ten-year review of commercial vaccine performance for control of tick infestations on cattle. Anim.
410 Heal. Res. Rev.
- 411 de La Fuente, J., Canales, M., Kocan, K.M., 2006. The importance of protein glycosylation in development
412 of novel tick vaccine strategies. Parasite Immunol. 28, 687–688.
- 413 de La Fuente, J., Kocan, K.M., 2006. Strategies for development of vaccines for control of ixodid tick
414 species. Parasite Immunol. 28, 275–283.
- 415 de la Fuente, J., Rodríguez, M., Redondo, M., Montero, C., García-García, J.C., Méndez, L., Serrano, E.,
416 Valdés, M., Enríquez, A., Canales, M., Ramos, E., Boué, O., Machado, H., Lleonart, R., de Armas, C.
417 a, Rey, S., Rodríguez, J.L., Artiles, M., García, L., 1998. Field studies and cost-effectiveness analysis
418 of vaccination with Gavac against the cattle tick *Boophilus microplus*. Vaccine 16, 366–373.
- 419 Domínguez-García, D.I., Rosario-Cruz, R., Almazán-García, C., Saltijeral Oaxaca, J.A., de la Fuente, J.,
420 2010. *Boophilus microplus*: Aspectos biológicos y moleculares de la resistencia a los acaricidas y su
421 impacto en la salud animal. Trop. Subtrop. Agroecosystems 12, 181–192.
- 422 Fernández-Salas, A., Rodríguez-Vivas, R.I., Alonso-Díaz, M.A., 2012. First report of a *Rhipicephalus*
423 *microplus* tick population multi-resistant to acaricides and ivermectin in the Mexican tropics. Vet.
424 Parasitol. 183, 338–342.
- 425 Fernández-Salas, A., Rodríguez-Vivas, R.I., Alonso-Díaz, M.Á., 2012. Resistance of *Rhipicephalus*
426 *microplus* to Amitraz and Cypermethrin in Tropical Cattle Farms in Veracruz, Mexico. J. Parasitol. 98,
427 1010–1014.

- 428 Foil, L.D., Coleman, P., Eisler, M., Fragoso-Sanchez, H., Garcia-Vazquez, Z., Guerrero, F.D., Jonsson, N.N.,
429 Langstaff, I.G., Li, A.Y., Machila, N., Miller, R.J., Morton, J., Pruett, J.H., Torr, S., 2004. Factors that
430 influence the prevalence of acaricide resistance and tick-borne diseases. *Vet. Parasitol.* 125, 163–181.
- 431 Freeman, J.M., Davey, R.B., Kappmeyer, L.S., Kammlah, D.M., Olafson, P.U., 2010. Bm86 midgut protein
432 sequence variation in South Texas cattle fever ticks. *Parasit. Vectors* 3, 101.
- 433 García-García, J.C., González, I.L., González, D.M., Valdés, M., Méndez, L., Lamberti, J., D'Agostino, B.,
434 Citroni, D., Fragoso, H., Ortiz, M., Rodríguez, M., de La Fuente, J., 1999. Sequence variations in the
435 *Boophilus microplus* Bm86 locus and implications for immunoprotection in cattle vaccinated with this
436 antigen. *Exp. Appl. Acarol.* 23, 883–895.
- 437 García-García, J.C., Montero, C., Redondo, M., Vargas, M., Canales, M., Boue, O., Rodríguez, M., Joglar,
438 M., MacHado, H., González, I.L., Valdés, M., Méndez, L., De La Fuente, J., 2000. Control of ticks
439 resistant to immunization with Bm86 in cattle vaccinated with the recombinant antigen Bm95 isolated
440 from the cattle tick, *Boophilus microplus*. *Vaccine* 18, 2275–2287.
- 441 Graf, J.F., Gogolewski, R., Leach-Bing, N., Sabatini, G. a, Molento, M.B., Bordin, E.L., Arantes, G.J., 2004.
442 Tick control: an industry point of view. *Parasitology* 129 Suppl, S427–S442.
- 443 Guerrero, F.D., Miller, R.J., Pérez de León, A.A., 2012. Cattle tick vaccines: Many candidate antigens, but
444 will a commercially viable product emerge? *Int. J. Parasitol.* 42, 421–427.
- 445 Kaewmongkol, S., Kaewmongkol, G., Inthong, N., Lakkitjaroen, N., Sirinarumit, T., Berry, C.M., Jonsson,
446 N.N., Stich, R.W., Jittapalapong, S., 2015. Variation among Bm86 sequences in *Rhipicephalus*
447 (*Boophilus*) *microplus* ticks collected from cattle across Thailand. *Exp. Appl. Acarol.* 66, 247–256.
- 448 Kunz, S.E., Kemp, D.H., 1994. Insecticides and acaricides: resistance and environmental impact. *Rev. Sci.*
449 *Tech.* 13, 1249–1286.
- 450 Lew-Tabor, A.E., Bruyeres, A.G., Zhang, B., Rodríguez Valle, M., 2014. *Rhipicephalus (Boophilus)*
451 *microplus* tick in vitro feeding methods for functional (dsRNA) and vaccine candidate (antibody)
452 screening. *Ticks Tick. Borne. Dis.* 5, 500–510.

- 453 Librado, P., Rozas, J., 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data.
454 *Bioinformatics* 25, 1451–1452.
- 455 Lima, W.S., Ribeiro, M.F., Guimaraes, M.P., 2000. Seasonal variation of *Boophilus microplus* (Canestrini,
456 1887) (Acari: Ixodidae) in cattle in Minas Gerais state, Brazil. *Trop. Anim. Health Prod.* 32, 375–380.
- 457 Lovis, L., Reggi, J., Berggoetz, M., Betschart, B., Sager, H., 2013. Determination of acaricide resistance in
458 *Rhipicephalus (Boophilus) microplus* (Acari: Ixodidae) field populations of Argentina, South Africa,
459 and Australia with the Larval Tarsal Test. *J. Med. Entomol.* 2, 326–335.
- 460 Martínez-Arzate, S.G., Tenorio-Borroto, E., Barbabosa Pliego, A., Díaz-Albiter, H.M., Vázquez-Chagoyán,
461 J.C., González-Díaz, H., 2017. PTML Model for Proteome Mining of B-Cell Epitopes and Theoretical-
462 Experimental Study of Bm86 Protein Sequences from Colima, Mexico. *J. Proteome Res.* 16, 4093–
463 4103.
- 464 Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T.L., Agarwala, R., Schäffer, A.A., 2008. Database
465 Indexing for Production MegaBLAST Searches. *Bioinformatics* 24, 1757–1764.
- 466 Nijhof, A.M., Balk, J.A., Postigo, M., Jongejan, F., 2009. Selection of reference genes for quantitative RT-
467 PCR studies in *Rhipicephalus (Boophilus) microplus* and *Rhipicephalus appendiculatus* ticks and
468 determination of the expression profile of Bm86. *BMC Mol. Biol.* 10, 112.
- 469 Nijhof, A.M., Balk, J.A., Postigo, M., Rhebergen, A.M., Taoufik, A., Jongejan, F., 2010. Bm86 homologues
470 and novel ATAQ proteins with multiple epidermal growth factor (EGF)-like domains from hard and
471 soft ticks. *Int. J. Parasitol.* 40, 1587–1597.
- 472 Patarroyo, J.H., Portela, R.W., De Castro, R.O., Couto Pimentel, J., Guzmán, F., Patarroyo, M.E., Vargas,
473 M.I., Prates, A.A., Días Mendes, M.A., 2002. Immunization of cattle with synthetic peptides derived
474 from the *Boophilus microplus* gut protein (Bm86). *Vet. Immunol. Immunopathol.* 88, 163–172.
- 475 Peconick, A.P., Sossai, S., Girão, F.A., Rodríguez, M.Q.R.B., Souza e Silva, C.H., Guzmán Q, F., Patarroyo
476 V, A.M., Vargas, M.I., Patarroyo, J.H., 2008. Synthetic vaccine (SBm7462) against the cattle tick
477 *Rhipicephalus (Boophilus) microplus*: Preservation of immunogenic determinants in different strains
478 from South America. *Exp. Parasitol.* 119, 37–43.

- 479 Pedroso, M.S., Mellor, L.M., Valdez, M., Souza, R.D.M., 2007. Control de las infestaciones de la garrapata
480 *Boophilus microplus* en la ganadería Cubana y en regiones de Latinoamérica con la aplicación del
481 inmunógeno Gavac dentro de un programa de lucha integral, in: Documento Sexta Conferencia
482 Electrónica Redectopar. p. 17.
- 483 Pérez-Cogollo, L.C., Rodríguez-Vivas, R.I., Ramírez-Cruz, G.T., Miller, R.J., 2010. First report of the cattle
484 tick *Rhipicephalus microplus* resistant to ivermectin in Mexico. *Vet. Parasitol.* 168, 165–169.
- 485 Rand, K.N., Moore, T., Sriskantha, A., Spring, K., Tellam, R., Willadsen, P., Cobon, G.S., 1989. Cloning
486 and expression of a protective antigen from the cattle tick *Boophilus microplus*. *Proc. Natl. Acad. Sci.*
487 86, 9657–9661.
- 488 Rodríguez-Vivas, R.I., Alonso-Díaz, M.A., Rodríguez-Arévalo, F., Fragoso-Sánchez, H., Santamaria, V.M.,
489 Rosario-Cruz, R., 2006. Prevalence and potential risk factors for organophosphate and pyrethroid
490 resistance in *Boophilus microplus* ticks on cattle ranches from the State of Yucatan, Mexico. *Vet.*
491 *Parasitol.* 136, 335–342.
- 492 Rodríguez-Vivas, R.I., Miller, R.J., Ojeda-Chi, M.M., Rosado-Aguilar, J.A., Trinidad-Martínez, I.C., Pérez
493 de León, A.A., 2014. Acaricide and ivermectin resistance in a field population of *Rhipicephalus*
494 *microplus* (Acari: Ixodidae) collected from red deer (*Cervus elaphus*) in the Mexican tropics. *Vet.*
495 *Parasitol.* 200, 179–188.
- 496 Rodríguez-Vivas, R.I., Pérez-Cogollo, L.C., Rosado-Aguilar, J.A., Ojeda-Chi, M.M., Trinidad-Martínez, I.,
497 Miller, R.J., Li, A.Y., Pérez de León, A., Guerrero, F., Klafke, G., 2014. *Rhipicephalus* (*Boophilus*)
498 *microplus* resistant to acaricides and ivermectin in cattle farms of Mexico. *Rev. Bras. Parasitol. Vet.* 23,
499 113–22.
- 500 Rodríguez-Vivas, R.I., Rodríguez-Arevalo, F., Alonso-Díaz, M.A., Fragoso-Sánchez, H., Santamaria, V.M.,
501 Rosario-Cruz, R., 2006. Prevalence and potential risk factors for amitraz resistance in *Boophilus*
502 *microplus* ticks in cattle farms in the State of Yucatan, Mexico. *Prev. Vet. Med.* 75, 280–286.
- 503 Rodríguez-Vivas, R.I., Rosado-Aguilar, J.A., Estrella Basto, G., García Vázquez, Z.S., Rosario-Cruz, R.,
504 Fragoso-Sánchez, H., 2006. Manual técnico para el control de garrapatas en el ganado bovino, INIFAP.

- 505 ed. Morelos, México.
- 506 Rodríguez-Vivas, R.I., Rosado-Aguilar, J.A., Ojeda-Chi, M.M., Pérez-Cogollo, L.C., Trinidad-Martínez, I.,
507 Bolio-González, M.E., 2014. Control Integrado De Garrapatas En La Ganadería Bovina. Ecosistemas
508 Recur. Agropecu. 1, 295–308.
- 509 Rodríguez, M., Rubiera, R., Penichet, M., Montesinos, R., Cremata, J., Falcón, V., Sánchez, G., Bringas, R.,
510 Cordovés, C., Valdés, M., Lleonart, R., Herrera, L., de la Fuente, J., 1994. High level expression of the
511 *B. microplus* Bm86 antigen in the yeast *Pichia pastoris* forming highly immunogenic particles for
512 cattle. J. Biotechnol. 33, 135–146.
- 513 Saitou, N., Nei, M., 1987. The neighbour-joining method: a new method for reconstructing phylogenetic
514 trees. Mol Biol Evo 4, 406–425.
- 515 Senasica, 2015. NOM-019-ZOO-1994 Campaña nacional contra la garrapata *Boophilus spp.*, Diario Oficial
516 de la Federación Mexicana. México.
- 517 Senasica, 2013. ACUERDO por el que se establece la Campaña Nacional para el control de la garrapata
518 *Boophilus spp.*, Diario Oficial de la Federación Mexicana. México.
- 519 Skaggs, R., Acuña, R., Torell, L.A., Southard, L.W., 2004. Exportaciones de ganado en pie de México hacia
520 los Estados Unidos: ¿De dónde viene el ganado y hacia dónde va? Rev. Mex. Agronegocios VIII, 212–
521 219.
- 522 Sonenshine, D.E., Kocan, K.M., de la Fuente, J., 2006. Tick control: further thoughts on a research agenda.
523 Trends Parasitol. 22, 550–551.
- 524 Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular evolutionary
525 genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.
- 526 Temeyer, K.B., Chen, A.C., Davey, R.B., Guerrero, F.D., Howell, J.M., Kammlah, D.M., Li, A.Y.,
527 Lohmeyer, K.H., Olafson, P.U., Perez De Leon, A.A., Phillips, P.L., Pound, J.M., Welcha, J.B., 2012.
528 Nuevos enfoques para el control de *Rhipicephalus (Boophilus) microplus*. Rev. Mex. Ciencias Pecu. 3,
529 25–40.

530 Willadsen, P., 2004. Anti-tick vaccines. *Parasitology* 129, S367–S387.

531 Willadsen, P., Bird, P., Cobon, G.S., Hungerford, J., 1995. Commercialization of a recombinant vaccine
532 against *Boophilus microplus*. *Parasitology* 110 Suppl, S43–S50.

533 Zhang, Z., Schwartz, S., Wagner, L., Miller, W., 2000. A greedy algorithm for aligning DNA sequences. *J*
534 *Comput Biol* 7, 203–14.

535 Zhao, Y.-J., Gong, X., 2015. Genetic divergence and phylogeographic history of two closely related species
536 (*Leucomeris decora* and *Nouelia insignis*) across the “Tanaka Line” in Southwest China. *BMC Evol.*
537 *Biol.* 15, 134.

538

539 Legends to Figures

540 **Figure 1.** Collection sites map. Samples from *R. microplus* specimens from 10 states of Mexico were collected. Pacific,
541 gulf, center and north center regions of the country were included. Each color corresponds to different collection state in
542 Mexico and each black pin represents the approximate geographical coordinates of the location in the map for each
543 collection site.

544 **Figure 1. Mantel test,** scatter plot. Real values for matrices are observed. A, gene distances in percentages of
545 similarity; B, geographical distances in kilometers.

546 **Figure 3.** Minimum expansion network. It shows the frequency and relationship among haplotypes and haplogroups.
547 Haplotypes and haplogroups are indicated with a colored label. Nine haplogroups can be observed. Haplogroups 1 and
548 2, containing the most abundant haplotypes. Haplotypes 10 to 64 are unique and have a random distribution in the
549 network. GenBank Bm86 sequences included in the network come from Thailand (TH), Argentina (ARG),
550 Mozambique (MOZ), United States of America (USA, TX), Mexico (MX), Brazil (BRA) and Australia (vaccine strain
551 Yeerongpilly; YEERO).

552 **Figure 4.** Neighbor-joining phylogenetic tree of complete Bm86 cDNA sequences. Tree reconstruction of evolutionary
553 distances among putative sequences of the Bm86 gene of Mexican tick populations by the algorithm neighbor-joining
554 and the method Kimura of 2 parameters. The main clades considered here are shown in brackets marked as A and B.
555 Clade B has 5 subclades labeled from B1 to B5. A total of 164 complete sequences, considering amino acid residues
556 from 20-627 from the Bm86 were used in the analysis, including 125 Mexican sequences (localities are marked with
557 circles of different colors) and GenBank sequences: 14 from Texas (USA, orange triangles); 18 from Thailand (brown

22

558 triangles); 1 from Australia (blue triangle); 1 from Argentina (light blue triangle); 1 from India (pink triangle); 1 from
559 Brazil (yellow triangle); the only previous Mexican (green triangle) and 2 Mozambique sequences used as outgroup
560 (black triangles). Bootstrap evaluation adjusted for 1000 permutation repetitions.

10.3 Determinación de regiones antigénicas

Se eligieron 9 regiones predichas como antigénicas de los 81 epítomos predichos por Sánchez Bermúdez y 10 regiones predichas como antigénicas de los 21 epítomos predichos por Sotelo Gómez (Sánchez Bermúdez, 2017; Sotelo Gómez, 2017), además 3 secuencias completas que contienen las regiones predichas como antigénicas en este trabajo. Los resultados arrojaron 3 secuencias consenso que contiene todas las regiones antigénicas predichas en los estudios antes mencionados (Ver figura 18), con dichas secuencias se producirán proteínas recombinantes que se probarán como vacunas.

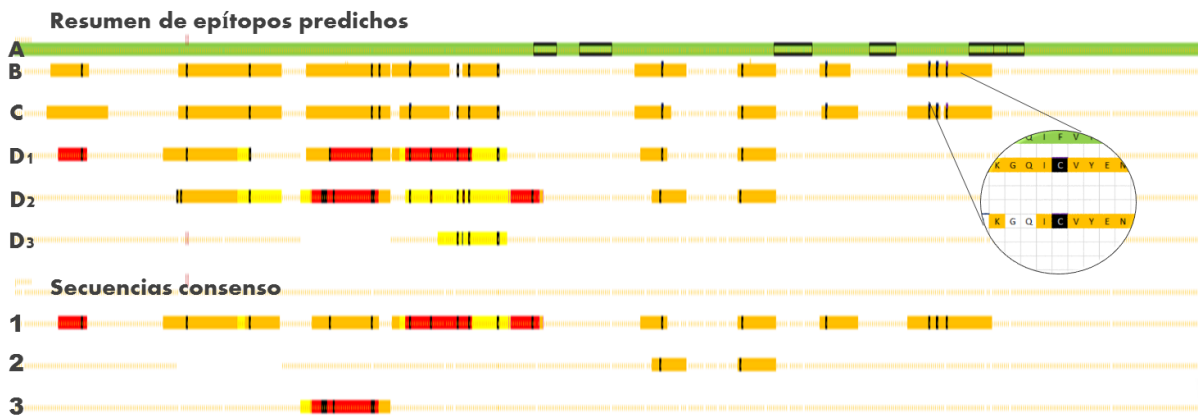


Figura 18.- Resumen de los epítomos predichos. La línea A de color verde, representa la cadena de 620 aa de longitud de la cepa de referencia vacunal Yeerongpilly. En colores rojo, naranja y amarillo, se muestra la posición de cada epítomo en la cadena. Las líneas negras muestran un cambio en la cadena de aa en relación con la cepa de referencia Yeerongpilly, ver ampliación de imagen en el círculo. En la secuencia B se muestra la posición de los epítomos predichos por Sánchez Bermúdez (Sánchez Bermúdez, 2017). En la secuencia C se muestra la posición de los epítomos predichos por Sotelo Gómez (Sotelo Gómez, 2017). En las secuencias D1, D2 y D3 se muestran los epítomos predichos en este trabajo, los colores de rojo a amarillo corresponden a los epítomos de mayor a menor valor de evaluación del modelo usado en este trabajo. **Secuencias consenso.** Se observa en la imagen la posición de los epítomos predichos en cada una de las secuencias consenso.

Finalmente, los epítomos lineares de células B predichos para las secuencias Bm86 mexicanas se compararon con otros estudios independientes para otras regiones del mundo. Por ejemplo, se encontró una coincidencia de mutación, pero con distinto aminoácido en el mismo sitio de la cadena entre uno de los dos epítomos definidos por Odongo *et al.*, (Ac. aspártico y Asparragina para *B. decoloratus*) y la región antigénica predicha en este trabajo (TYR para *R. microplus* de México). Una descripción más detallada se puede observar en la Tabla 4.

Tabla 4.- Comparación de resultados con distintos estudios. Comparación entre la predicción de epítomos lineales en este estudio y cuatro estudios independientes, se muestran solo los sitios que coincidían para una mutación. Mutaciones que se contenían en un solo péptido, se distinguen en distintos tonos de gris.

^a Estudios predictivos de...	^b Estudio	^c Yeerongpilly	^d México	^e S. aa
2 epítomos lineales conservados entre los homólogos de <i>Boophilus decoloratus</i> y <i>R. microplus</i> (epítomos de células B definidos por Ac.), (Odongo <i>et al.</i> , 2007).	Bd86 1 y 2; Ac. aspártico y Asparragina (D y N)	Ac aspártico (D)	Tirosina (Y)	35
	Metionina (M)	Metionina (M)	Lisina (K)	123
3 epítomos inmunogénicos sintéticos a partir de una secuencia inferida de la proteína Bm86, Colombia, (epítomos de células B definidos por Ac.), (Patarroyo <i>et al.</i> , 2002).	Histidina (H)	Histidina (H)	Tirosina (Y)	381
	Alanina (A)	Alanina (A)	Prolina (P)	162
	Asparragina (N)	Asparragina (N)	Lisina (K)	163
	Serina (S)	Serina (S)	Leucina (L)	165
	Alanina (A)	Alanina (A)	Prolina (P)	167
	Isoleucina (I)	Isoleucina (I)	Lisina (K)	187*
	Asparragina (N)	Asparragina (N)	Ac. aspártico (D)	188*
	Histidina (H)	Histidina (H)	Tirosina (Y)	232
	Serina (S)	Serina (S)	Glicina (G)	235
	Alanina (A)	Alanina (A)	Valina (V)	238
31 péptidos antigénicos predichos usando el método de Kolaskar y Tongaonkar, (Canales <i>et al.</i> , 2009).	Treonina (T)	Treonina (T)	Lisina (K)	271
	Lisina (K)	Lisina (K)	Asparragina (N)	380
	Histidina (H)	Histidina (H)	Tirosina (Y)	381
	Ac. aspártico (D)	Ac. aspártico (D)	Asparragina (N)	425
	Lisina (K)	Isoleucina (I)	Lisina (K)	187*
	Ac. aspártico (D)	Asparragina (N)	Ac. aspártico (D)	188*
	Arginina (R)	Lisina (K)	Arginina (R)	191
	Identificación de 2 regiones hidrofílicas mediante el algoritmo de Hopp & Woods en secuencias del sur de Texas, (Freeman <i>et al.</i> , 2010).			

^a Breve descripción del objetivo del estudio.

^b Aminoácido predicho en el estudio ^a

^c Aminoácido observado en secuencia de referencia.

^d Mutación en los aislados mexicanos observados en este trabajo.

^e Posición en la cadena de aminoácidos.

* mutaciones que coinciden en 2 estudios diferentes.

11 DISCUSIÓN

La vacuna comercial “Gavac” en base al antígeno Bm86 para combatir a la garrapata *R. microplus*, está disponible en México desde 1997, sin embargo la especificidad de la vacuna es menor en México que en Australia y Cuba, posiblemente debido al polimorfismo genético entre las cepas de *R. microplus* que ya han sido reportados por varios autores en diferentes partes del mundo (G. S. Cobon *et al.*, 1997; Cunha Casquero *et al.*, 2012; Cunha *et al.*, 2011; Freeman *et al.*, 2010; García-García *et al.*, 1999; Kaewmongkol *et al.*, 2015; Lima *et al.*, 2000; Rodríguez *et al.*, 1994; Sonenshine *et al.*, 2006; Willadsen, 2004).

En el pasado, se sugirió que estas variaciones eran responsables de la diferencia en la eficacia de las vacunas basadas en los antígenos Bm86/95 (García-García *et al.*, 2000; Parizi *et al.*, 2009). En el caso específico de México, para responder esta interrogante, en el pasado un modelo que estudia la correlación entre la heterogeneidad de la secuencia y la protección vacunal se usó para proponer que no hay diferencias significativas en la eficacia de la vacuna entre las moléculas mexicanas y Yeerongpilly Bm86 (Willadsen, 2004); sin embargo, en dicho ensayo, solo dos secuencias cortas de la molécula Bm86, identificadas como Tuxpan y Mora (número de acceso de GenBank AF150892 y AF150893 respectivamente) de 107 pb o 35 aminoácidos (de un total de 608 aa de la cadena de la proteína madura) se usaron para correlacionar la frecuencia de mutaciones en la secuencia, con la eficacia del antígeno (García-García *et al.*, 1999). Además de estas dos secuencias cortas, hasta ahora solo hay un informe mexicano de secuencia completa identificado como cepa *Susceptible* (número de acceso GenBank FJ456928) (Canales *et al.*, 2009) de 1827pb, que codifica la cadena de proteína de 608aa, y un 97% de identidad con la cepa Yeerongpilly. Esta información no es suficiente para explicar la efectividad de la vacuna en México.

Por otro lado, en la República Mexicana se desconoce si existe variabilidad genética de la proteína Bm86, su posible distribución, o si dicha variabilidad ha influido en la estructura genética del gen en el territorio nacional. Más específicamente se desconoce cuántos haplotipos existen y si están formando haplogrupos y si estos haplogrupos están confinados a una población o región del país o no. Conocer esta información es importante ya que la frecuencia con la que se presenta un haplotipo dentro de una localidad o entre las localidades muestreadas podría ser determinante para elegir en el futuro candidatos vacunales más específicos para la región geográfica de México. El análisis de las secuencias del gen Bm86 de *R. microplus* mostró un alto grado de polimorfismo entre las localidades muestreadas, pero también dentro de ellas. En el presente estudio, se analizaron 125 cadenas completas del cDNA del gen Bm86, para demostrar que existe variabilidad genética entre las poblaciones de garrapata que circulan en nuestro país con respecto a la cepa vacunal. Las muestras provienen de 10 localidades diferentes en México, Figura 8.

De manera global, en este estudio encontramos un total de 145 SNPs y 70 SAPs en las 125 secuencias de México. Dicha información nos condujo a identificar un total de 64 diferentes haplotipos, Ver Tablas 1 y 2 del apartado de ANEXOS. En este estudio, encontramos que, de los 10 estados muestreados, Chiapas fue la localidad con la mayor diversidad haplotípica ($h = 1$), mientras que en las localidades de Nayarit y Zacatecas tuvieron el valor más bajo ($h = 0.84$). Respecto a la diversidad de nucleótidos, el valor más bajo se observó en la localidad de Veracruz ($\pi = 0.001$) y el valor más alto ($\pi = 0.018$) para la localidad de Zacatecas. Los valores de diversidad para los 10 lugares muestreados se presentan en la Tabla 1 del apartado de ANEXOS. De los 64 diferentes haplotipos identificados, solo 9 formaron haplogrupos o grupos de secuencias con una identidad del 100% dentro de cada grupo, el resto (55 secuencias) son

haplotipos únicos. Los haplogrupos 1 y 2 representan respectivamente el 22.2% y el 14.2% del total de las secuencias analizadas, ambos se observaron en 9 de las 10 ubicaciones muestreadas. El porcentaje de divergencia para estos 9 haplogrupos se mantuvo entre 3.1-7.4%. una descripción de los 9 haplogrupos observados se presenta en la Tabla 2 del apartado de ANEXOS. Todos estos datos sugieren que existe un grado de polimorfismo importante para el gen Bm86 en la región México. La siguiente cuestión por resolver fue si dichos haplotipos estaban confinados a alguna localidad en particular. Para responder dicha pregunta, se realizó una prueba de Mantel para conocer el nivel de correlación que existe entre las matrices de distancias genéticas y distancias geográficas. Los resultados observados muestran que no existe relación entre las matrices ($r = 0.014$, $P = 0.92$). Estos resultados sugieren un alto nivel de flujo genético en el territorio nacional a partir del gen Bm86. La explicación más probable es el flujo de ganado en México (el origen y destino del ganado en el territorio nacional) y, en consecuencia, la garrapata que viaja con el bovino. Pero es difícil encontrar una base de datos formal o detallada sobre el flujo de ganado en México, posiblemente porque no está disponible para el interés público. Considerando la diversidad genética del gen Bm86 encontrada en *R. microplus* recolectada en nuestro territorio y que la vacuna Gavac no ha podido controlar las poblaciones de garrapatas en México se puede asumir que esta diversidad es la responsable de la relativamente baja eficiencia de dicha vacuna. De ahí que en el presente trabajo se propone la inclusión de al menos 3 secuencias diseñadas para incluir la mayor cantidad de epítomos derivados de distintas secuencias de la proteína Bm86 encontradas en nuestro país para la producción de una vacuna hipotéticamente más efectiva. Esta vacuna podría también servir para otras regiones con poblaciones de *R. microplus* que comparten haplotipos con las poblaciones de garrapatas en México. Tal es el caso de Texas, en este estudio, las secuencias mexicanas y los informes texanos tienen los valores más altos de identidad, (entre 93-100%). Dicho resultado puede deberse a la proximidad geográfica y al flujo génico de las poblaciones de

garrapatas entre ambos territorios. Es difícil explicar el flujo del gen Bm86 de la garrapata si no es con el flujo de ganado y la participación de los animales silvestres en la frontera de México y los Estados Unidos, sin embargo, bases de datos detalladas sobre el flujo de ganado entre ambos países no están disponibles tampoco para el interés público. Según un trabajo de investigación publicado por Skaggs y colaboradores (Skaggs *et al.*, 2004) donde proporciona información parcial sobre los registros de origen y destino del ganado vivo exportado desde México a los Estados Unidos, los funcionarios federales de ambos países inspeccionan estos bovinos para controlar la incidencia de enfermedades y de los vectores. Sin embargo, el problema de la garrapata en la frontera de México y los Estados Unidos no se limita al ganado bovino, sino que otras especies de animales silvestres, como ciervos y antílopes, podrían llevar la garrapata *R. microplus* a territorios de nuestro vecino del norte.

Por otro lado, los valores más bajos de identidad (86-88%) se observaron curiosamente entre nuestras secuencias y la secuencia Bm95 reportada en Argentina, incluso menor que con las secuencias de Tailandia (91-94%), que se encuentra geográficamente más distante. La variabilidad en la secuencia de aminoácidos entre los aislamientos mexicanos y el antígeno Bm95 se debe considerar como un factor importante, porque si el antígeno Bm95 se usara en México, hipotéticamente no mejoraría la eficacia de la vacuna.

El diagrama MST de los haplotipos observados no reveló una estructura filogeográfica definida. En la Figura 1 del apartado de ANEXOS, observamos la estructura filogenética de los 64 haplotipos encontrados en aislamientos mexicanos y los informes Bm86 para diferentes localidades del mundo. El diagrama muestra 9 haplogrupos, sin embargo, los haplogrupos 1 y 2 son notables por contener los haplotipos más abundantes, pero también muestran un patrón de dispersión filogenética muy similar en las localidades analizadas. El haplogrupo 2 contiene

15 secuencias de las localidades de: Veracruz, Morelos, Campeche, Zacatecas, Tabasco y Nayarit. El haplogrupo 1 contiene 29 secuencias de las mismas ubicaciones del haplogrupo 2, pero también secuencias de las localidades de Colima y Sinaloa que no se observaron en el haplogrupo 2. El resto de los haplotipos de aislados mexicanos (haplotipos 10-64) son haplotipos únicos, su distribución sugiere un evento de expansión muy significativo para el gen Bm86, dicha distribución sugiere un alto nivel de flujo génico y polimorfismo del gen Bm86 para las localidades analizadas. Estos hallazgos nos permiten pensar que en nuestro país no pueden diseñarse vacunas para aplicación regional dado que no se encontró una estructura filogeográfica para los diferentes haplotipos encontrados. De ahí que la vacuna que se diseñe para nuestro país debe contener todos los haplotipos encontrados, y deberá ser de aplicación universal.

Con respecto a los informes en la base de datos de Genbank, dos de las secuencias de los informes de Texas están contenidas en el haplogrupo 1 y dos más en el haplogrupo 2, el resto de los informes de Texas tienen una distribución aleatoria en la red. Informes de América Latina: Argentina, Brasil, Texas, además del único informe mexicano, se derivan de los haplogrupos 1 y 2. Por otro lado, en una expansión más prolongada, por putativos nodos de divergencia, se puede observar que los dos informes de Mozambique surgen del haplogrupo 1; algo similar ocurre con la cepa de referencia Yerongpilly en el haplogrupo 2. Por último, en un grupo monofilético, podemos ver todos los informes de Tailandia derivados del haplogrupo 5. Ver Figura 1 del apartado de Anexos.

La reconstrucción del árbol filogenético tampoco mostró un patrón filogeográfico evidente, pero si presentó una disposición politómica donde ninguna de las localidades está representada por la mayor abundancia en algunos de los clados, más bien, se caracterizó por una distribución

aleatoria.

Se observó que la mayoría de los aislados mexicanos y estadounidenses están separados unos de otros por ramas muy cortas que en la escala de porcentaje indican que han divergido en un tiempo relativamente reciente, sin embargo, también observamos algunos clados de aislamientos mexicanos que parecen divergir en mayor grado.

Hipotéticamente un ancestro común dio lugar, por un lado, a la mayoría de los aislamientos de América, los dos informes de Mozambique y dos de los informes de Tailandia y, por otro lado, la mayoría de las secuencias de Tailandia, están ligadas a cuatro aislamientos mexicanos de tres diferentes localidades y una secuencia de Texas que a pesar de ser clasificado como *R. microplus*, parece estar más emparentado con las secuencias que se utilizaron como secuencias externas de *R. anulatus* Kinne y Dimmi.

A pesar de que se pudieron observar varios haplogrupos, fue difícil proponer uno o varios haplotipos como candidatos vacunales, sin embargo, con la información obtenida en las secuencias de los aislados mexicanos, se puede predecir las propiedades, como la antigenicidad, de péptidos lineares de dichas secuencias antes de su síntesis. Esta estrategia ha tomado fuerza en los últimos años por su capacidad de predecir las propiedades de una molécula antes de su síntesis ya que ahorra tiempo y dinero en el área de la investigación.

Mediante distintas herramientas computacionales disponibles en la actualidad, es posible la predicción de epítopos de células B y su capacidad antigénica ya sea a partir de una secuencia conocida, o si se quiere predecir la capacidad antigénica después de una perturbación o cambio en la secuencia de aminoácidos. En este trabajo comparamos los resultados predictivos para las

secuencias mexicanas Bm86 de un método desarrollado por nosotros y dos trabajos recientes de nuestro grupo de investigación (Sánchez Bermúdez, 2017; Sotelo Gómez, 2017).

En este trabajo, desarrollamos un método general de Perturbation Theory and Machine Learning (PTML) para la extracción de datos de proteomas, con el fin de descubrir nuevos epítomos de células B útiles para el diseño de vacunas, aplicamos el uso de este modelo a las secuencias de los aislados mexicanos Bm86.

Para predecir las puntuaciones de actividad S_{qj} (la puntuación predice la propensión de un péptido de ser más antigénico o no después de una perturbación) para un conjunto muy grande (26112 péptidos no redundantes) de todos los posibles péptidos de las 14 secuencias de Colima, México. Los modelos generales para predecir los resultados de múltiples perturbaciones de entrada y salida en la secuencia de péptidos y las condiciones (condiciones de contorno) de ensayo experimental utilizando ideas de análisis de perturbación se han desarrollado en el pasado (González-Díaz *et al.*, 2013; Kleandrova *et al.*, 2014; Luan *et al.*, 2014; Messina *et al.*, 2015). En este trabajo, informamos un nuevo modelo capaz de predecir la propensión de un péptido a actuar como epítomo B después de múltiples perturbaciones en la secuencia de péptidos y en las condiciones experimentales de los ensayos.

Cabe señalar que los dos estudios previos utilizaron la misma base de datos. La diferencia de cada una de las condiciones experimentales entre estudios **previos/este trabajo**, se detalla a continuación: Organismo fuente del péptido **500/1448**; Organismo huésped expuesto al antígeno **50/323**; proceso inmunológico de sensibilización o proceso *in vivo* **10/15**; Métodos o técnicas utilizadas para determinar la inmunogenicidad de los péptidos o Técnicas experimentales **30/28**. Además, incluimos por primera vez en este trabajo, la condición

experimental "**aditivos adyuvantes**". Es importante saber que los adyuvantes son agentes heterogéneos no específicos, su importancia es que, en asociación con los antígenos de las vacunas, pueden inducir una respuesta inmune más potente y prolongada. El uso y la elección de las condiciones experimentales dependerán de la naturaleza del experimento y las expectativas de cada investigador. El modelo permite elegir entre 505 posibles opciones de adyuvantes, sin embargo, es importante destacar que el 97.2% de los ensayos están libres de adyuvantes, por lo que su uso al menos en esta base de datos no es muy común. Es posible elegir la opción "sin adyuvantes" en este modelo.

Los resultados del modelo resumen la excelente capacidad predictiva de la función LDA. El modelo discrimina ambas clases de péptidos (propensión de un péptido de ser más antigénico o no después de una perturbación) con un nivel de significancia ($p < 0.005$). Para la validación y entrenamiento del modelo, se clasificaron 1,048,190 pares de secuencias de péptidos de consulta y referencia, anteriormente solo se clasificaron 200,000. El modelo solo necesita 7 variables para clasificar con una precisión mayor al 70% en capacitación y validación en más de un millón de casos ($N_{\text{total}} = 1,048,190$). El modelo clasificó 786,142 casos para la serie de entrenamiento (212,326 errores de clasificación) y 262,048 para la serie de validación (70,849 errores de clasificación).

En resumen, el modelo LDA propuesto en este trabajo logró una sensibilidad y especificidad de 80% y 71%, respectivamente, Este modelo es más robusto que los trabajos anteriores, por lo que los porcentajes de sensibilidad y especificidad son más confiables, ver la Tabla 3 del apartado de ANEXOS. Estos resultados son buenos también en comparación con otros usos de técnica LDA en la investigación del proteoma y el análisis de datos en general (Aguero-Chapin *et al.*, 2009; Beleites *et al.*, 2013; Kowalski *et al.*, 1982; Munteanu *et al.*, 2013; Xiao and Shen,

2015). Las comparaciones con otros modelos similares han sido desarrolladas en el pasado (Concu *et al.*, 2009; González-Díaz *et al.*, 2011, 2007; Munteanu *et al.*, 2009b; Pan *et al.*, 2010; Rodríguez-Soca *et al.*, 2010).

Aproximadamente el 44% de las secuencias del gen Bm86 de los aislados mexicanos son haplotipos únicos. Debido al elevado polimorfismo de las secuencias mexicanas, fue difícil proponer una o varias secuencias como candidatos vacunales, por lo que se adoptó una estrategia que ha demostrado ser útil en el diseño racional de vacunas. El modelo PTML propuesto en este trabajo para la predicción de epítomos lineales de células B fue de gran utilidad para identificar regiones antigénicas relevantes en las secuencias Bm86 mexicanas. Como se puede observar en la Figura 8, los epítomos predichos por Sotelo Gómez, Sánchez Bermúdez y los epítomos predichos en este estudio, muestran un patrón de distribución bastante similar en la cadena de aminoácidos de la proteína Bm86. De la misma forma, 2 epítomos predichos por ambos métodos SVMTriP y BepiPred y que no se observaron por el método PTML, fueron considerados en las secuencias consenso. Los resultados obtenidos en la comparación de los distintos modelos se muestran alentadores camino a proponer candidatos vacunales más específicos para combatir a la garrapata en México, por lo que se puede formular la hipótesis de que la información de los epítomos con propensión a ser antigénicos puede reunirse en una o varias secuencias consenso que podrían provocar después de la vacunación, una respuesta inmune de anticuerpos anti-Bm86, que podría tener un efecto protector al repertorio de los diferentes haplotipos observados en este estudio.

12 CONCLUSIONES

Debido al polimorfismo observado en las secuencias Bm86 de México, no fue posible proponer candidatos vacunales derivados directamente de dichas secuencias, sin embargo, este estudio aumenta la base de datos actual de las secuencias mexicanas de Bm86 y contribuye a la comprensión global de la genética de poblaciones de garrapatas. Por otra parte, la comparación de la predicción de epítomos de células B en este trabajo fueron consistentes con las predicciones hechas por Sánchez Bermúdez y Sotelo Gómez, por lo que fue posible proponer 3 secuencias consenso que reúnen la información de todos los epítomos predichos en los trabajos antes mencionados, dichas secuencias servirán para sintetizar la proteína recombinante en la búsqueda de una vacuna Bm86 más específica para las garrapatas de México.

13 BIBLIOGRAFIA

1. Agüero-Chapin, G., Varona-Santos, J., de la Riva, G.A., Antunes, A., Gonzalez-Villa, T., Uriarte, E., González-Díaz, H., 2009. Alignment-free prediction of polygalacturonases with pseudofolding topological indices: experimental isolation from *Coffea arabica* and prediction of a new sequence. *J. Proteome Res.* 8, 2122–8.
2. Allen, J.R., Humphreys, S.J., 1979. Immunisation of guinea pigs and cattle against ticks. *Nature* 280, 491–493.
3. Alonso-Díaz, M.A., Rodríguez-Vivas, R.I., Fragosó-Sánchez, H., Rosario-Cruz, R., 2006. Resistencia de la garrapata *Boophilus microplus* a los ixodicidas. *Arch. Med. Vet.* 38, 105–113.
4. Anbarasi, P., Latha, B.R., Dhinakar raj, G., Sreekumar, C., Senthuran, S., 2014. Partial sequencing of Bm86 gene for studying the phylogeny of an Indian isolate of *Rhipicephalus(Boophilus) microplus* tick. *J. Parasit. Dis.* 38, 260–264.
5. Andreotti, R., 2006. Performance of two Bm86 antigen vaccine formulation against tick using crossbreed bovines in stall test. *Rev. Bras. Parasitol. Vet.* 15, 97–100.
6. Andreotti, R., Pedroso, M.S., Caetano, A.R., Martins, N.F., 2008. Comparison of predicted binders in *Rhipicephalus (boophilus) microplus* intestine protein variants Bm86 Campo Grande strain, Bm86 and Bm95. *Rev. Bras. Parasitol. Vet.* 17, 93–98.
7. Azhahianambi, P., de La Fuente, J., Suryanarayana, V.V.S., Ghosh, S., 2009. Cloning, expression and immunoprotective efficacy of rHaa86, the homologue of the Bm86 tick vaccine antigen, from *Hyalomma anatolicum anatolicum*. *Parasite Immunol.* 31, 111–122.
8. Bandelt, H.J., Forster, P., Rohl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16, 37–48.
9. Bazán Tene, M., 2002. Efecto de *Metarhizium anisopliae* (*Deuteromycotina: Hyphomycetes*) en el control biológico de *Boophilus microplus Canestrini* (Acari: Ixodidae). Universidad de Colima.
10. Beleites, C., Neugebauer, U., Bocklitz, T., Krafft, C., Popp, J., 2013. Sample size planning for classification models. *Anal. Acta* 760, 25–33.
11. Botello, A., Botello, A., Borroto, C., Suárez, M., Pérez, A., Rodríguez, Y., Fajardo, H., Pérez, C., González, A., Rodríguez, A., Linares, R., 2011. Control of

- Rhipicephalus (Boophilus) microplus* ticks in bovine with the Herber biogar inmunógeno. RedVet 12, 1–10.
12. Campanella, J.J., Bitincka, L., Smalley, J., 2003. MatGAT: An application that generates similarity/identity matrices using protein or DNA sequences. BMC Bioinformatics 4, 29.
 13. Canales, M., Almazán, C., Naranjo, V., Jongejan, F., de la Fuente, J., 2009. Vaccination with recombinant *Boophilus annulatus* Bm86 ortholog protein, Ba86, protects cattle against *B. annulatus* and *B. microplus* infestations. BMC Biotechnol. 9, 29.
 14. Canales, M., de la Lastra, J., Naranjo, V., Nijhof, A.M., Hope, M., Jongejan, F., de la Fuente, J., 2008. Expression of recombinant *Rhipicephalus (Boophilus) microplus*, *R. annulatus* and *R. decoloratus* Bm86 orthologs as secreted proteins in *Pichia pastoris*. BMC Biotechnol. 8, 14.
 15. Canales, M., Enríquez, A., Ramos, E., Cabrera, D., Dandie, H., Soto, A., Falcón, V., Rodríguez, M., de la Fuente, J., 1997. Large-scale production in *Pichia pastoris* of the recombinant vaccine Gavac™ against cattle tick. Vaccine 15, 414–422.
 16. Chen, S., Xing, Y., Su, T., Zhou, Z., Dilcher, E.D.L., Soltis, D.E., 2012. Phylogeographic analysis reveals significant spatial genetic structure of *Incarvillea sinensis* as a product of mountain building. BMC Plant Biol. 12, 58.
 17. Cobon, G., Moore, J.T., Johnston, L.A.Y., Willadsen, P., Kemp, D.H., Sriskantha, A., Riding, G.A., Rand, K.N., 1997. DNA encoding a cell membrane glycoprotein of a tick gut. Nucl. Phys. B 15, 661–661.
 18. Cobon, G.S., Levine, M.M., Woodrow, G.C., Kaper, J.B., 1997. An anti-arthropod vaccine: TickGARD – a vaccine to prevent cattle tick infestations. In New Generation Vaccines, Marcel Dek. ed. New York.
 19. Concu, R., Dea-Ayuela, M.A., Perez-Montoto, L.G., Bolas-Fernandez, F., Prado-Prado, F.J., Podda, G., Uriarte, E., Ubeira, F.M., Gonzalez-Diaz, H., 2009. Prediction of enzyme classes from 3D structure: a general model and examples of experimental-theoretic scoring of peptide mass fingerprints of Leishmania proteins. J. Proteome Res. 8, 4372–82.
 20. Cruz, R.R., Inés, D., García, D., Ramírez, E.R., Estrada, M.O., Ibañez, F.M., 2010. Estrategias para el control de la garrapata, 1st ed. México.
 21. Cunha Casquero, R., Pérez de León, A.A., Leivas Leite, F.P., Pinto, L. da S., dos Santos Júnior, A.G., Andreotti, R., 2012. Bovine immunoprotection against

- Rhipicephalus (Boophilus) microplus* with recombinant Bm86-Campo Grande antigen. Rev. Bras. Parasitol. Vet 21, 254–262.
22. Cunha, R.C., Andreotti, R., Leite, F.P.L., 2011. *Rhipicephalus (Boophilus) microplus*: expression and characterization of Bm86-CG in *Pichia pastoris*. Rev. Bras. Parasitol. Vet. = Brazilian J. Vet. Parasitol. 20, 103–10.
23. de la Fuente, J., Almazán, C., Canales, M., Pérez de la Lastra, J.M., Kocan, K.M., Willadsen, P., 2007. A ten-year review of commercial vaccine performance for control of tick infestations on cattle. Anim. Heal. Res. Rev.
24. de La Fuente, J., Rodríguez, M., Montero, C., Redondo, M., García-García, J.C., Méndez, L., Serrano, E., Valdés, M., Enríquez, A., Canales, M., Ramos, E., Boué, O., Machado, H., Leonart, R., 1999. Vaccination against ticks (*Boophilus spp.*): The experience with the Bm86-based vaccine Gavac(TM). Genet. Anal. - Biomol. Eng. 15, 143–148.
25. de la Fuente, J., Rodríguez, M., Redondo, M., Montero, C., García-García, J.C., Méndez, L., Serrano, E., Valdés, M., Enríquez, A., Canales, M., Ramos, E., Boué, O., Machado, H., Leonart, R., de Armas, C. a, Rey, S., Rodríguez, J.L., Artilés, M., García, L., 1998. Field studies and cost-effectiveness analysis of vaccination with Gavac against the cattle tick *Boophilus microplus*. Vaccine 16, 366–373.
26. Domínguez-García, D.I., Rosario-Cruz, R., Almazán-García, C., Saltijeral Oaxaca, J.A., de la Fuente, J., 2010. *Boophilus microplus*: Aspectos biológicos y moleculares de la resistencia a los acaricidas y su impacto en la salud animal. Trop. Subtrop. Agroecosystems 12, 181–192.
27. Dwight, D.B., 2011. Parasitología para Veterinarios, 9th ed. Elsevier, Barcelona, España.
28. Fernández-Salas, A., Rodríguez-Vivas, R.I., Alonso-Díaz, M.A., 2012. First report of a *Rhipicephalus microplus* tick population multi-resistant to acaricides and ivermectin in the Mexican tropics. Vet. Parasitol. 183, 338–342.
29. Fernández-Salas, A., Rodríguez-Vivas, R.I., Alonso-Díaz, M.Á., 2012. Resistance of *Rhipicephalus microplus* to Amitraz and Cypermethrin in Tropical Cattle Farms in Veracruz, Mexico. J. Parasitol. 98, 1010–1014.
30. Freeman, J.M., Davey, R.B., Kappmeyer, L.S., Kammlah, D.M., Olafson, P.U., 2010. Bm86 midgut protein sequence variation in South Texas cattle fever ticks. Parasit. Vectors 3, 101.

31. García-García, J.C., Gonzalez, I.L., González, D.M., Valdés, M., Méndez, L., Lamberti, J., D'Agostino, B., Citroni, D., Fragoso, H., Ortiz, M., Rodríguez, M., de La Fuente, J., 1999. Sequence variations in the *Boophilus microplus* Bm86 locus and implications for immunoprotection in cattle vaccinated with this antigen. *Exp. Appl. Acarol.* 23, 883–895.
32. García-García, J.C., Montero, C., Redondo, M., Vargas, M., Canales, M., Boue, O., Rodríguez, M., Joglar, M., MacHado, H., González, I.L., Valdés, M., Méndez, L., De La Fuente, J., 2000. Control of ticks resistant to immunization with Bm86 in cattle vaccinated with the recombinant antigen Bm95 isolated from the cattle tick, *Boophilus microplus*. *Vaccine* 18, 2275–2287.
33. Getzoff, E.D., Tainer, J. a, Lerner, R. a, Geysen, H.M., 1988. The chemistry and mechanism of antibody binding to protein antigens. *Adv. Immunol.* 43, 1–98.
34. González-Díaz, H., Arrasate, S., Gomez-SanJuan, A. Sotomayor, N., Lete, E., Besada-Porto, L., Ruso, J.M., 2013. General theory for multiple input-output perturbations in complex molecular systems. 1. Linear QSPR electronegativity models in physical, organic, and medicinal chemistry. *Curr Top Med Chem* 13, 1713–41.
35. Gonzalez-Díaz, H., Perez-Montoto, L.G., Ubeira, F.M., 2014. Model for vaccine design by prediction of B-epitopes of IEDB given perturbations in peptide sequence, in vivo process, experimental techniques, and source or host organisms. *J. Immunol. Res.* 2014, 15.
36. González-Díaz, H., Prado-Prado, F., García-Mera, X., Alonso, N., Abeijon, P., Caamano, O., Yanez, M., Munteanu, C.R., Pazos, A., Dea-Ayuela, M.A., Gómez-Muñoz, M.T., Garijo, M.M., Sansano, J., Ubeira, F.M., 2011. MIND-BEST: Web server for drugs and target discovery; design, synthesis, and assay of MAO-B inhibitors and theoretical experimental study of G3PDH protein from *Trichomonas gallinae*. *J. Proteome Res.* 10, 1698–718.
37. González-Díaz, H., Saiz-Urra, L., Molina, R., Santana, L., Uriarte, E., 2007. A model for the recognition of protein kinases based on the entropy of 3D van der Waals interactions. *J. Proteome Res.* 6, 904–8.
38. Graf, J.F., Gogolewski, R., Leach-Bing, N., Sabatini, G. a, Molento, M.B., Bordin, E.L., Arantes, G.J., 2004. Tick control: an industry point of view. *Parasitology* 129 Suppl, S427–S442.

39. Guerrero, F.D., Miller, R.J., Pérez de León, A.A., 2012. Cattle tick vaccines: Many candidate antigens, but will a commercially viable product emerge? *Int. J. Parasitol.* 42, 421–427.
40. Hassan, I.A., Hassan, S.M., El Hussein, A.M., Rodríguez, M., 2007. Evaluation of the recombinant Bm86 antigen (Gavac) against *Hyalomma dromedarii* and *Hyalomma anatolicum* (Acari: Ixodidae) in Sudan. *Biotecnol. Apl.* 24, 122–125.
41. Hopp, T.P., Woods, K.R., 1981. Prediction of protein antigenic determinants from amino acid sequences. *Immunology* 78, 3824–3828.
42. Kaewmongkol, S., Kaewmongkol, G., Inthong, N., Lakkitjaroen, N., Sirinarumitr, T., Berry, C.M., Jonsson, N.N., Stich, R.W., Jittapalapong, S., 2015. Variation among Bm86 sequences in *Rhipicephalus (Boophilus) microplus* ticks collected from cattle across Thailand. *Exp. Appl. Acarol.* 66, 247–256.
43. Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120.
44. Kleandrova, V. V., Luan, F., Gonzalez-Diaz, H., Ruso, J.M., Speck-Planche, A., Cordeiro, M.N., 2014. Computational tool for risk assessment of nanomaterials: novel QSTR-perturbation model for simultaneous prediction of ecotoxicity and cytotoxicity of uncoated and coated nanoparticles under multiple experimental conditions. *Environ. Sci. Technol.* 48, 14686–94.
45. Kowalski, R.D., Wold, S., In: Krishnaiah, P.R., Kanal, L.N., 1982. *Handbook of Statistic*, North Holl. ed. Amsterdam.
46. Kunz, S.E., Kemp, D.H., 1994. Insecticides and acaricides: resistance and environmental impact. *Rev. Sci. Tech.* 13, 1249–1286.
47. Larsen, J.E.P., Lund, O., Nielsen, M., 2006. Improved method for predicting linear B-cell epitopes. *Immunome Res.* 2, 2.
48. Librado, P., Rozas, J., 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452.
49. Lima, W.S., Ribeiro, M.F., Guimaraes, M.P., 2000. Seasonal variation of *Boophilus microplus* (Canestrini, 1887) (Acari: Ixodidae) in cattle in Minas Gerais state, Brazil. *Trop. Anim. Health Prod.* 32, 375–380.
50. Lovis, L., Reggi, J., Berggoetz, M., Betschart, B., Sager, H., 2013. Determination of acaricide resistance in *Rhipicephalus (Boophilus) microplus* (Acari: Ixodidae) field

- populations of Argentina, South Africa, and Australia with the Larval Tarsal Test. *J. Med. Entomol.* 2, 326–335.
51. Luan, F., Kleandrova, V. V., Gonzalez-Diaz, H., Ruso, J.M., Melo, A., Speck-Planche, A., Cordeiro N., M., 2014. Computer-aided nanotoxicology: assessing cytotoxicity of nanoparticles under diverse experimental conditions by using a novel QSTR perturbation approach. *Nanoscale* 6, 10623–30.
52. Messina, P. V., Besada-Porto, J.M., González-Diaz, H., Ruso, J.M., 2015. Self-Assembled Binary Nanoscale Systems: Multioutput Model with LFER-Covariance Perturbation Theory and an Experimental Computational Study of NaGDC-DDAB Micelles. *Langmuir* 31, 12009–18.
53. Miller, R.J., Almazán, G.C., Estrada, O.M., Davey, R.B., George, J.E., 2008. A survey of fipronil- and ivermectin-resistant *Rhipicephalus (Boophilus) microplus* collected in Northern Mexico and the options for the management of acaricide-resistant ticks with pesticides, in: VI Seminario Internacional de Parasitología. Impacto de Las Enfermedades Parasitarias Sobre La Ganadería Globalizada. Boca del Río, Veracruz, México.
54. Min, B.K., Dahne, S., Ahn, M.H., Noh, Y.K., Muller, K.R., 2016. Decoding of top-down cognitive processing for SSVEP-controlled BMI. *Sci. Rep.* 6, 362–67.
55. Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T.L., Agarwala, R., Schäffer, A.A., 2008. Database Indexing for Production MegaBLAST Searches. *Bioinformatics* 24, 1757–1764.
56. Munteanu, C., Magalhaes, A., Duardo-Sánchez, A., Pazos, A., González-Diaz, H., 2013. S2SNet: A Tool for Transforming Characters and Numeric Sequences into Star Network Topological Indices in Chemoinformatics, Bioinformatics, Biomedical, and Social-Legal Sciences. *Curr. Bioinform.* 8, 429–437.
57. Munteanu, C.R., González-Diaz, H., Borges, F., de Magalhaes, A.L., 2008. Natural/random protein classification models based on star network topological indices. *J. Theor. Biol.* 254, 775–83.
58. Munteanu, C.R., Magalhaes, A.L., Uriarte, E., González-Diaz, H., 2009a. Multi-target QPDR classification model for human breast and colon cancer-related proteins using star graph topological indices. *J. Theor. Biol.* 257, 303–11.
59. Munteanu, C.R., Vázquez, J.M., Dorado, J., Sierra, A.P., Sánchez-González, A., Prado-Prado, F.J., González-Diaz, H., 2009b. Complex network spectral moments for

- ATCUN motif DNA cleavage: first predictive study on proteins of human pathogen parasites. *J. Proteome Res.* 8, 5219–28.
60. Nijhof, A.M., Balk, J.A., Postigo, M., Jongejan, F., 2009. Selection of reference genes for quantitative RT-PCR studies in *Rhipicephalus (Boophilus) microplus* and *Rhipicephalus appendiculatus* ticks and determination of the expression profile of Bm86. *BMC Mol. Biol.* 10, 112.
61. Nuttall, P.A., Trimnell, A.R., Kazimirova, M., Labuda, M., 2006. Exposed and concealed antigens as vaccine targets for controlling ticks and tick-borne diseases. *Parasite Immunol.* 28, 155–163.
62. Odongo, D., Kamau, L., Skilton, R., Mwaura, S., Nitsch, C., Musoke, A., Taracha, E., Daubenberger, C., Bishop, R., 2007. Vaccination of cattle with TickGARD induces cross-reactive antibodies binding to conserved linear peptides of Bm86 homologues in *Boophilus decoloratus*. *Vaccine* 25, 1287–1296.
63. Olaya Escobedo, N., 2007. Piroplasmosis En El Ganado Bovino. Universidad Michoacana de San Nicolás de Hidalgo.
64. Pan, X.Y., Zhang, Y.N., Shen, H.B., 2010. Large-scale prediction of human protein-protein interactions from amino acid sequence based on latent topic features. *J. Proteome Res.* 9, 4992–5001.
65. Parizi, L.F., Pohl, P.C., Masuda, A., Vaz, I.D.S., 2009. New approaches toward anti-*Rhipicephalus (Boophilus) microplus* tick vaccine. *Rev. Bras. Parasitol. Veterinária* 18, 1–7.
66. Patarroyo, J.H., Portela, R.W., De Castro, R.O., Couto Pimentel, J., Guzman, F., Patarroyo, M.E., Vargas, M.I., Prates, A.A., Dias Mendes, M.A., 2002. Immunization of cattle with synthetic peptides derived from the *Boophilus microplus* gut protein (Bm86). *Vet. Immunol. Immunopathol.* 88, 163–172.
67. Pedroso, M.S., Mellor, L.M., Valdez, M., Souza, R.D.M., 2007. Control de las infestaciones de la garrapata *Boophilus microplus* en la ganadería Cubana y en regiones de Latinoamérica con la aplicación del inmunógeno Gavac dentro de un programa de lucha integral, in: Documento Sexta Conferencia Electrónica Redectopar. p. 17.
68. Pérez-Cogollo, L.C., Rodríguez-Vivas, R.I., Ramírez-Cruz, G.T., Miller, R.J., 2010. First report of the cattle tick *Rhipicephalus microplus* resistant to ivermectin in Mexico. *Vet. Parasitol.* 168, 165–169.

69. Ran, T., Liu, Y., Li, H., Tang, S., He, Z., Munteanu, C.R., González-Díaz, H., Tan, Z., Zhou, C., 2016. Gastrointestinal Spatiotemporal mRNA Expression of Ghrelin vs Growth Hormone Receptor and New Growth Yield Machine Learning Model Based on Perturbation Theory. *Sci. Rep.* 6, 301–74.
70. Rand, K.N., Moore, T., Sriskantha, A., Spring, K., Tellam, R., Willadsen, P., Cobon, G.S., 1989. Cloning and expression of a protective antigen from the cattle tick *Boophilus microplus*. *Proc. Natl. Acad. Sci.* 86, 9657–9661.
71. Rodríguez-Soca, Y., Munteanu, C.R., Dorado, J., Pazos, A., Prado-Prado, F. J. González-Díaz, H., 2010. Trypano-PPI: a web server for prediction of unique targets in trypanosome proteome by using electrostatic parameters of protein-protein interactions. *J. Proteome Res.* 9, 1182–90.
72. Rodríguez-Vivas, R.I., Arieta-Román, R.J., Pérez-Cogollo, L.C., Rosado-Aguilar, J.A., Ramírez-Cruz, G.T., Basto-Estrella, G., 2010. Uso de lactonas macrocíclicas para el control de la garrapata *Rhipicephalus (Boophilus) microplus* en el ganado bovino. *Arch. Med. Vet.* 42, 115–123.
73. Rodríguez-Vivas, R.I., Miller, R.J., Ojeda-Chi, M.M., Rosado-Aguilar, J.A., Trinidad-Martínez, I.C., Pérez de León, A.A., 2014. Acaricide and ivermectin resistance in a field population of *Rhipicephalus microplus* (Acari: Ixodidae) collected from red deer (*Cervus elaphus*) in the Mexican tropics. *Vet. Parasitol.* 200, 179–188.
74. Rodríguez-Vivas, R.I., Pérez-Cogollo, L.C., Rosado-Aguilar, J.A., Ojeda-Chi, M.M., Trinidad-Martínez, I., Miller, R.J., Li, A.Y., Pérez de León, A., Guerrero, F., Klafke, G., 2014. *Rhipicephalus (Boophilus) microplus* resistant to acaricides and ivermectin in cattle farms of Mexico. *Rev. Bras. Parasitol. Vet.* 23, 113–22.
75. Rodríguez-Vivas, R.I., Rodríguez-Arevalo, F., Alonso-Díaz, M.A., Fragoso-Sánchez, H., Santamaria, V.M., Rosario-Cruz, R., 2006. Prevalence and potential risk factors for amitraz resistance in *Boophilus microplus* ticks in cattle farms in the State of Yucatan, Mexico. *Prev. Vet. Med.* 75, 280–286.
76. Rodríguez-Vivas, R.I., Rosado-Aguilar, J.A., Estrella Basto, G., García Vázquez, Z.S., Rosario-Cruz, R., Fragoso-Sánchez, H., 2006. Manual técnico para el control de garrapatas en el ganado bovino, INIFAP. ed. Morelos, México.
77. Rodríguez-Vivas, R.I., Rosado-Aguilar, J.A., Ojeda-Chi, M.M., Perez-Cogollo, L.C., Trinidad-Martínez, I., Bolio-Gonzáles, M.E., 2014. Control Integrado De Garrapatas En La Ganadería Bovina. *Ecosistemas Recur. Agropecu.* 1, 295–308.

78. Rodríguez, M., Rubiera, R., Penichet, M., Montesinos, R., Cremata, J., Falcón, V., Sánchez, G., Bringas, R., Cordovés, C., Valdés, M., Leonart, R., Herrera, L., de la Fuente, J., 1994. High level expression of the *B. microplus* Bm86 antigen in the yeast *Pichia pastoris* forming highly immunogenic particles for cattle. *J. Biotechnol.* 33, 135–146.
79. Rodríguez Valle, M., Montero, C., Machado, H., Joglar, M., de la Fuente, J., García-García, J.C., 2001. The evaluation of yeast derivatives as adjuvants for the immune response to the Bm86 antigen in cattle. *BMC Biotechnol.* 1, 2.
80. Rosario-Cruz, R., Domínguez García, D.I., Ramírez Rojas, E., 2009. Estrategias para el control de la garrapata *Boophilus microplus* y la mitigación de la resistencia a los pesticidas. 22.
81. SAGARPA, S., SENASICA, S., 2015. Memorias 23a. Reunión Anual, CONASA, Parasitología y Parasiticidas.
82. SAGARPA, S.D.A.G.D.R., ALIMENTACIÓN, P.Y., Cantú Covarrubias, A., García Vázquez, Z., 2013. Estrategias Para El Control Integrado De Garrapata (*Boophilus spp.*) en la Producción de Bovinos de Carne en Pastoreo. *Foll. Técnico No. MX-0-310402-43-03-14-09-36* 1–37.
83. Saitou, N., Nei, M., 1987. The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evo* 4, 406–425.
84. Sánchez Bermúdez, J.C., 2017. Análisis de la Variación del Gen Bm86 de una Población de *Rhipicephalus microplus* de Ganado Bovino del Estado de Tabasco, México. Universidad Autónoma del Estado de México.
85. Skaggs, R., Acuña, R., Torell, L.A., Southard, L.W., 2004. Exportaciones de ganado en pie de México hacia los Estados Unidos: ¿De dónde viene el ganado y hacia dónde va? *Rev. Mex. Agronegocios VIII*, 212–219.
86. Sonenshine, D.E., Kocan, K.M., de la Fuente, J., 2006. Tick control: further thoughts on a research agenda. *Trends Parasitol.* 22, 550–551.
87. Sossai, S., Peconick, A.P., Sales, P.A., Marcelino, F.C., Vargas, M.I., Neves, E.S., Patarroyo, J.H., 2005. Polymorphism of the bm86 gene in South American strains of the cattle tick *Boophilus microplus*. *Exp. Appl. Acarol.* 37, 199–214.
88. Sotelo Gómez, S., 2017. Identificación de polimorfismo del gen Bm86 de especímenes de *Rhipicephalus microplus* recolectadas en el Estado de Zacatecas, México. Universidad Autónoma del Estado de México.

89. Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729.
90. Taylor, M.A., 2001. Recent developments in ectoparasiticides. *Vet. J.* 161, 253–268.
91. Valle, M.R., Mèndez, L., Valdez, M., Redondo, M., Espinosa, C.M., Vargas, M., Cruz, R.L., Barrios, H.P., Seoane, G., Ramírez, E.S., Boue, O., Vigil, J.L., Machado, H., Nordelo, C.B., Piñeiro, M.J., 2004. Integrated control of *Boophilus microplus* ticks in Cuba based on vaccination with the anti-tick vaccine Gavac™. *Exp. Appl. Acarol.* 34, 375–382.
92. Vargas, M., Montero, C., Sánchez, D., Pérez, D., Valdés, M., Alfonso, A., Joglar, M., Machado, H., Rodríguez, E., Méndez, L., Leonart, R., Suárez, M., Fernández, E., Estrada, M.P., Rodríguez-Mallón, A., Farnós, O., 2010. Two initial vaccinations with the Bm86-based Gavacplus vaccine against *Rhipicephalus (Boophilus) microplus* induce similar reproductive suppression to three initial vaccinations under production conditions. *BMC Vet. Res.* 6, 43.
93. Vázquez-Prieto, S., Paniagua, E., Ubeira, F.M., González-Díaz, H., 2016. QSPR-Perturbation Models for the Prediction of B Epitopes from Immune Epitope Database: A Potentially Valuable Route for Predicting “In Silico” New Optimal Peptide Sequences and/or Boundary Conditions for Vaccine Development. *Int. J. Pept. Res. Ther.* 1–6.
94. Walker, a., Bouattour, A., Camicas, J., Estrada-peña, A., Horak, I., Latif, A., Pegram, R., Preston, P., 2003. Ticks of domestic animals in Africa: a guide to identification of species, The University of Edinburgh.
95. Willadsen, P., 2004. Anti-tick vaccines. *Parasitology* 129, S367–S387.
96. Willadsen, P., Bird, P., Cobon, G.S., Hungerford, J., 1995. Commercialisation of a recombinant vaccine against *Boophilus microplus*. *Parasitology* 110 Suppl, S43–S50.
97. Willadsen, P., Kemp, D.H., Cobon, G.S., Wright, I.G., 1992. Successful vaccination against *Boophilus microplus* and *Babesia bovis* using recombinant antigens. *Mem. Inst. Oswaldo Cruz.*
98. Willadsen, P., Riding, G.A., Mckenna, R. V, Kemp, D.H., Tellam, R.L., Nielsen, J.N., Lahnstein, J., Cobon, G.S., Gough, J.M., Nielsen, J.N., Lahnstein, J., Cobon, G.S., Gough, J.M., 1989. Immunologic control of a parasitic arthropod. Identification of a protective antigen from *Boophilus microplus*. Information about subscribing to The Journal of Immunology is online at : IMMUNOLOGIC CONTROL OF A PARASITIC ARTHROPOD Identification of a Pro.

99. Xiao, F., Shen, H.B., 2015. Prediction Enhancement of Residue Real-Value Relative Accessible Surface Area in Transmembrane Helical Proteins by Solving the Output Preference Problem of Machine Learning-Based Predictors. *J. Chem. Inf. Model.* 55, 2464–74.
100. Zhang, Z., Schwartz, S., Wagner, L., Miller, W., 2000. A greedy algorithm for aligning DNA sequences. *J Comput Biol* 7, 203–14.
101. Zhao, Y.-J., Gong, X., 2015. Genetic divergence and phylogeographic history of two closely related species (*Leucomeris decora* and *Nouelia insignis*) across the “Tanaka Line” in Southwest China. *BMC Evol. Biol.* 15, 134.
102. Zivkovic, Z., Esteves, E., Almazán, C., Daffre, S., Nijhof, A.M., Kocan, K.M., Jongejan, F., de la Fuente, J., 2010. Differential expression of genes in salivary glands of male *Rhipicephalus (Boophilus) microplus* in response to infection with *Anaplasma marginale*. *BMC Genomics* 11, 186.

14 ANEXOS

Tabla 1.- Variabilidad genética del gen Bm86 de las diferentes localidades muestreadas en México.

Localidad	n	n_h	n_{hu}	h	π	% Id Tx
Campeche	18	11	7	0.94	0.007	98.7
Veracruz	18	8	4	0.90	0.001	99
Morelos	13	12	11	0.98	0.013	98.3
Colima	14	9	7	0.89	0.014	98.4
Sinaloa	11	10	9	0.98	0.017	97.8
Guerrero	11	9	8	0.97	0.015	97.9
Nayarit	10	6	4	0.84	0.003	99
Chiapas	9	9	9	1	0.015	98
Tabasco	11	9	7	0.96	0.005	99
Zacatecas	10	6	4	0.84	0.018	97.8

n	Tamaño de la muestra
n_h	Numero de haplotipos observados
n_{hu}	Numero de haplotipos únicos
h	Diversidad haplotípica
π	Diversidad nucleotídica
% Id Tx	Promedio de identidad vs reportes del Norte de Texas

Table 2.- Haplogrupos observados para las muestras del gen Bm86 de *R. microplus* de México.
Se muestran solo los 9 de los 64 haplotipos diferentes que formaron haplogrupos.

ID ^a	Frec ^b	Cg (%) ^c	Freci-p ^d	SAP's ^e	Div (%) ^f
1	28	22.2	9/10	22/621	3.5
2	18	14.2	9/10	19/621	3.1
3	6	4.7	4/10	19/621	3.1
4	6	4.7	4/10	22/621	3.5
5	4	3.1	3/10	46/621	7.4
6	3	2.3	2/10	21/621	3.4
7	2	1.5	2/10	20/621	3.2
8	2	1.5	1/10	21/613	3.4
9	2	1.5	1/10	26/613	4.2

^a Numero identificador del haplogrupo.

^b Frecuencia haplotípica.

^c Constitución genética.

^d Frecuencia de los haplotipos inter-poblaciones.

^e Polimorfismos de un solo aminoácido.

^f Porcentaje de divergencia con respecto a Yeerongpilly.

Table 3. Resultados de los análisis LDA

Method ^a	Set ^b	Statistics ^c	Symbol	%	Class ^d	$\epsilon_{qj-pred} = 0$	$\epsilon_{qj-pred} = 1$
Entropy	t	Specificity	Sp	70.6	$\epsilon_{qj-obs} = 0$	415299	172838
		Sensitivity	Sn	80.1	$\epsilon_{qj-obs} = 1$	39488	158517
		Accuracy	Ac	73.0	Total		
	v	Specificity	Sp	70.7	$\epsilon_{qj-obs} = 0$	138578	57469
		Sensitivity	Sn	79.7	$\epsilon_{qj-obs} = 1$	13380	52621
		Accuracy	Ac	73.0	Total		
Polarizab.	t	Specificity	Sp	95.91	$\epsilon_{qj-obs} = 0$	83699	3568
		Sensitivity	Sn	91.92	$\epsilon_{qj-obs} = 1$	5486	62416
		Accuracy	Ac	94.17	Total		
	v	Specificity	Sp	95.97	$\epsilon_{qj-obs} = 0$	27736	1164
		Sensitivity	Sn	91.6	$\epsilon_{qj-obs} = 1$	1859	20267
		Accuracy	Ac	94.08	Total		
Electroneg.	t	Specificity	Sp	97.0	$\epsilon_{qj-obs} = 0$	2660	84607
		Sensitivity	Sn	93.6	$\epsilon_{qj-obs} = 1$	63548	4354
		Accuracy	Ac	95.5	Total		
	v	Specificity	Sp	97.1	$\epsilon_{qj-obs} = 0$	840	28060
		Sensitivity	Sn	93.3	$\epsilon_{qj-obs} = 1$	20641	1485
		Accuracy	Ac	95.4	Total		

^a Descriptores numéricos utilizados para cuantificar la estructura del péptido, Entropy = Índices de información de entropía de Shannon, Polarizab. = Polarizabilidad atómica promedio, Electroneg. = Electronegatividad atómica promedio.

^b Set = Conjunto de datos; t = conjunto de entrenamiento, v = serie de validación externa.

^c Especificidad = $Sp (\%) = 100 \cdot / (NT0 + NF1)$;

Sensibilidad = $Sn (\%) = 100 \cdot NT1 / (NT1 + NF0)$;

Exactitud = $Sn (\%) = 100 \cdot [NT1 + NT0] / (\text{Número de casos})$;

NT0 = Número de verdaderos 0 = Número de casos con $\epsilon_{qj-obs} = 0$ y $\epsilon_{qj-pred} = 0$

NT1 = Número de verdaderos 1 = Número de casos con $\epsilon_{qj-obs} = 1$ y $\epsilon_{qj-pred} = 1$

NF0 = Número de falsos 0 = Número de casos con $\epsilon_{qj-obs} = 0$ y $\epsilon_{qj-pred} = 1$

NF1 = Número de falsos 1 = Número de casos con $\epsilon_{qj-obs} = 1$ y $\epsilon_{qj-pred} = 0$

^d $\epsilon_{qj-obs} = 1$ cuando la actividad IEDB = Positivo-Alto o Positivo; $\epsilon_{qj-obs} = 0$ de otra manera; $\epsilon_{qj-pred} = 0$ cuando el modelo predice 0.

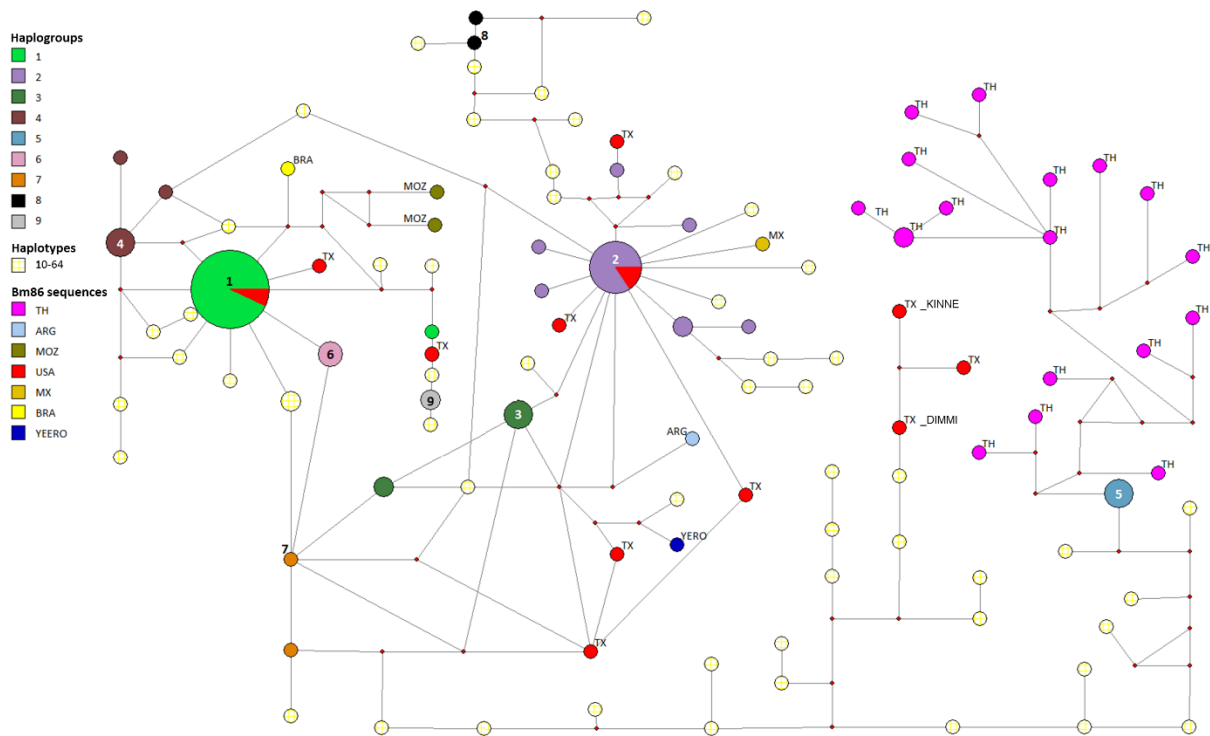


Figura 1.- Diagrama MST de los haplotipos observados. Red de expansión mínima que muestra la frecuencia y la relación de cada haplotipo. Los colores representan cada haplotipo y haplogrupo de acuerdo con la etiqueta. Se pueden observar 9 haplogrupos. Los haplogrupos 1 y 2, contienen los haplotipos más abundantes. Los haplotipos 10 a 64 son haplotipos únicos, la red muestra una distribución aleatoria en la red. Los informes incluidos en Genbank para el gen Bm86 son: Tailandia (TH), Argentina (ARG), Mozambique (MOZ), Estados Unidos de América (EE. UU., TX), México (MX), Brasil (BRA), cepa de la vacuna de Australia Yeerongpilly.