



UNIVERSIDAD AUTÓNOMA DE QUERÉTARO
FACULTAD DE QUÍMICA
MAESTRÍA EN CIENCIAS QUÍMICO BIOLÓGICAS

Caracterización genética de los receptores de antígenos de los linfocitos T y B del
lobo marino de California, *Zalophus californianus*

TESIS

Que como parte de los requisitos para obtener el Grado de
MAESTRO EN CIENCIAS QUÍMICO BIOLÓGICAS

PRESENTA:
ARANZAZU ARIAS ROJAS

DIRIGIDA POR:
Dra. Karina Alethya Acevedo Whitehouse

CO-DIRIGIDA POR:
Dra. Laura Cristina Berumen Segura

SINODALES

Dra. Karina Alethya Acevedo Whitehouse
Presidente

Dra. Laura Cristina Berumen Segura
Secretario

Dra. María Guadalupe García Alcocer
Vocal

Dra. Jesica Esther Escobar Cabrera
Suplente

Dra. Angelina Rodríguez Torres
Suplente

Centro Universitario, Querétaro, Qro.
Noviembre, 2019
México

Declaratoria de autoría

Declaro que el trabajo que se presenta en esta tesis fue realizado por mí, a excepción de aquellas secciones que formaron parte de colaboraciones, mismas que se indican explícitamente en el texto.

Confirmando que las ideas originales aquí presentadas son de mi autoría y que a lo largo de la tesis le he dado crédito al trabajo de otros mediante el uso adecuado de las referencias.

Esta copia se ha sometido a la Universidad Autónoma de Querétaro bajo el entendido de que constituye material con derechos de autor y que no puede citarse ningún pasaje de esta tesis sin darle el crédito adecuado.

© 2019. Universidad Autónoma de Querétaro y Aranzazu Arias Rojas

AGRADECIMIENTOS

Agradezco el financiamiento al Proyecto 446 de *Fronteras de la Ciencia* SEP-CONACYT otorgado a Karina A. Acevedo-Whitehouse el cual fue fundamental para realizar esta tesis y sus productos derivados.

También quiero agradecer a CONACYT por otorgarme la beca de maestría (627934), la cual me permitió dedicarme por completo a este proyecto.

A los miembros del laboratorio de Plasticidad Inmune y Eco Epidemiología Molecular (GMEE) de la Facultad de Ciencias Naturales de la Universidad Autónoma de Querétaro que recolectaron las muestras incluidas en este trabajo.

Al Posgrado en Ciencias Químico Biológicas de la facultad de Química por contribuir a mi formación.

A mi codirectora, Laura C. Berumen Segura por su apoyo durante mi estancia en el posgrado.

A mis asesoras, Jesica E. Escobar Cabrera, Angelina Rodríguez Torres y Guadalupe García Alcocer.

A la M.C. Alejandra Herbert Mainero y al Dr. Rolando Bárcenas Luna por su apoyo durante el análisis filogenético.

Al Dr. Orlando Bruno Giorgetti por las sugerencias sobre el análisis.

Al la Unidad de Proteogenómica, del Instituto de Neurobiología de la Universidad Nacional Autónoma de México, especialmente a la M.C Adriana González Gallardo, por las veces que me proporcionó recombinantes.

Finalmente, al Laboratorio Nacional de Visualización Científica Avanzada (LAVIS), de la Universidad Nacional Autónoma de México, por el uso del Cluster DNA para los análisis bioinformáticos de este trabajo.

DEDICATORIA

Para mi familia, especialmente a mi Mamá y a mis abuelas Amparo y Conchita, gracias por su apoyo en todo momento y por transmitirme siempre su entusiasmo por aprender.

A Karina, por desbloquear mis ideas absolutistas y permitirme conocer muchas más cosas sobre la Inmunología, pero sobre todo por confiar en mí en todo momento y en estas manos que no sabían que podían ser verdes.

Dirección General de Bibliotecas UAG

ÍNDICE

I. Introducción	16
II. Antecedentes	18
1.1 <i>El sistema inmune y la evolución del reconocimiento antigénico</i>	18
1.2 <i>Células del sistema inmune adaptativo de mamíferos</i>	21
1.2.1 <i>Linfocitos T</i>	21
a) <i>Receptor de linfocitos T</i>	24
1.2.2 <i>Linfocitos B</i>	27
a) <i>Receptor de linfocitos B</i>	29
1.3 <i>Los segmentos V(D)J a través de los taxones de mamíferos</i>	31
1.4 <i>Evolución de los segmentos V(D) en mamíferos</i>	37
1.4.1 <i>Limitadores en la plasticidad del repertorio de los receptores TCR y BCR</i>	38
1.5 <i>El lobo marino de California (Z. californianus)</i>	40
1.5.1 <i>Inmunología del lobo marino de California (Z. californianus)</i>	42
1.6 <i>Técnicas genómicas de nueva generación en inmunología</i>	44
1.7 <i>La caracterización de los receptores TCR y BCR en especies no modelo mediante técnicas Bioinformáticas</i>	45
Justificación	47
Preguntas de investigación	49
III. Objetivos	50
IV. Metodología	51
2.1 <i>Muestras utilizadas</i>	51
2.2 <i>Extracción de RNA de las muestras</i>	51
2.3 <i>Retrotranscripción de cDNA de muestras de RNA de Z. californianus</i>	52
2.4 <i>Diseño racional de oligonucleótidos cebadores para especies no modelo</i>	52
2.5 <i>Estandarización de la amplificación por PCR de punto final en cDNA de Z. californianus</i>	54
2.6 <i>Preparación del ARN para la rápida amplificación del extremo 5'/3' del cDNA</i>	55
2.6.1 <i>Desfosforilación del ARN para la rápida amplificación de los extremo 5'/3' del cDNA</i>	55
2.6.2 <i>Precipitación del ARN para la rápida amplificación de los extremo 5'/3' del cDNA</i>	56

2.6.3 Decapitación del ARN para la rápida amplificación de los extremo 5'/3' del cDNA	56
2.6.4 Ligación de Oligo 5' al ARN para la rápida amplificación del extremo 5'/3' del cDNA	57
2.6.5 Retrotranscripción del ARN para la rápida amplificación de los extremo 5'/3' del cDNA	57
2.7 Rápida amplificación de los extremo 5'/3' del cDNA.	58
2.7.1 PCR Touchdown para las cadenas heavy y lambda del BCR, y Beta TCR.	58
2.7.2 PCR Touchdown para la cadena Alpha del TCR.....	60
2.8 PCR anidada para las cadenas Heavy y Lambda del BCR, y alpha y beta del TCR60	
2.9 Obtención de bandas únicas por PCR anidada para la secuenciación de los productos del protocolo de 5'/3' RACE	62
2.9.1 PCR punto final para obtener bandas únicas de las cadenas heavy y lambda del BCR, y alpha y beta del TCR	62
2.10 Clonación para obtener bandas únicas de las cadenas Heavy y Lambda del BCR, y Alpha y Beta del TCR.....	64
2.11 Secuenciación tipo Sanger de los productos del protocolo de 5'/3' RACE	64
2.12 Análisis bioinformático de las secuencias obtenidas por secuenciación tipo Sanger de los productos del protocolo de 5'/3' RACE.....	64
2.13 Análisis estadístico de los tamaños de los receptores.....	65
2.14 Secuenciación de nueva generación de los productos del protocolo de 5'/3' RACE	65
2.15 Análisis bioinformático de las secuencias obtenidas por secuenciación de nueva generación de los productos del protocolo de 5'/3' RACE.....	66
2.16 Análisis estadístico de los valores E del análisis de secuencias de nueva generación y de frecuencias de genes V(D)J.....	68
2.17 Identificación de los segmentos génicos V(D)J de los receptores TCR y BCR (Ig)69	
2.18 Análisis filogenético.....	69
2.19 Distribución geográfica	70
2.20 Caracterización de las secuencias de aminoácidos de las cadenas de los receptores TCR y BCR.....	70
2.21 Modelado de los receptores TCR y BCR	70
2.22 Análisis de calidad de las estructuras tridimensionales	71
V. Resultados	73

3.1 Secuencias de las cadenas alpha/beta y heavy del TCR y BCR por secuenciación Sanger.....	73
3.2 Caracterización de secuencias alpha/beta,heavy/lambda del TCR y BCR(Ig) por secuenciación de nueva generación.....	76
3.3 Caracterización de los segmentos génicos V(D)J de los receptores TCR y BCR (Ig) de <i>Z. californianus</i>	80
3.3.1 Frecuencias de los segmentos V, J de la cadena alpha del TCR <i>Z. californianus</i>	81
3.3.2 Frecuencias de los segmentos V(D)J de la cadena beta del TCR de <i>Z. californianus</i>	83
3.3.4 Frecuencias de las configuraciones VJ de la cadena alpha y V(D)J de la cadena beta del TCR de <i>Z. californianus</i>	84
3.3.5 Frecuencias de los segmentos V(D)J de la cadena Heavy del BCR de <i>Z. californianus</i>	86
3.3.6 Frecuencias de los segmentos VJ de la cadena lambda del BCR <i>Z. californianus</i>	89
3.3.7 Frecuencias de las configuraciones VJ de la cadena Lambda y V(D)J de la cadena heavy del BCR de <i>Z. californianus</i>	90
3.3.8 Número de segmentos V(D)J de los receptores TCR y BCR(Ig) a través de los taxones.....	92
3.4 Evolución de los segmentos génicos V(D)J de las cadenas del TCR y BCR de lobo marino de California (<i>Zalophus californianus</i>).....	95
3.4.1 Evolución de los segmentos V de las cadenas alpha (TRAV) de carnívoros y mamíferos marinos	95
3.4.2 Evolución de los segmentos V de las cadenas beta (TRBV) de carnívoros y mamíferos marinos	98
3.4.3 Evolución de los segmentos V de las cadenas heavy (IGHV) de carnívoros y mamíferos marinos	101
3.4.4 Evolución de los segmentos V de las cadenas heavy (IGLV) de carnívoros y mamíferos marinos	104
3.4.5 Evolución de los segmentos J de las cadenas alpha (TRAJ) de carnívoros y mamíferos marinos	107
3.4.6 Evolución de los segmentos DJ de las cadenas beta (TRBJD) de carnívoros y mamíferos marinos	110
3.4.7 Evolución de los segmentos DJ de las cadenas heavy (IGHDJ) de carnívoros y mamíferos marinos	114

3.4.8 Evolución de los segmentos J de las cadenas lambda (IGLJ) de carnívoros y mamíferos marinos	118
3.5 Evolución y ecología de los receptores TCR y BCR de <i>Z. californianus</i>	120
3.5.1 Relación ecológica y evolución de los segmentos V en las cadenas del TCR alpha/beta del lobo marino de California.....	120
3.5.2 Relación ecológica y evolución de los segmentos V en las cadenas del BCR (Ig) heavy/lambda del lobo marino de California.....	123
3.6 Evolución de las familias de los segmentos V(D)J de las cadenas del TCR de <i>Z. californianus</i> y BCR de <i>Z. californianus</i>	126
3.7 Modelado 3D de los receptores TCR, BCR y MHC II de <i>Z. californianus</i>	127
Discusión	130
VI. Conclusiones	144
Referencias	145
Anexos	161

ÍNDICE DE FIGURAS

Figura 1: Sistemas análogos al TCR y BCR con dominios de la super familia de Inmunoglobulinas (IgSF).	19
Figura 2: Evolución de los órganos linfoides.	22
Figura 3: Estructura del receptor de linfocitos T.	25
Figura 4: Formación del receptor de linfocitos T.	26
Figura 5: Estructura del receptor de linfocitos B.	29
Figura 6: Formación del receptor de linfocitos B.	30
Figura 7: Repertorio de genes V y J de la cadena Alpha en los mamíferos.	32
Figura 8: Repertorio de genes V (D) J de la cadena Beta en los mamíferos.	33
Figura 9: Repertorio de genes V (D) J y pseudogenes V de la cadena Heavy en los mamíferos	34
Figura 10: Repertorio de genes V y J de la cadena kappa en los mamíferos.	35
Figura 11: Repertorio de genes V y J de la cadena Lambda en los mamíferos.	37
Figura 12: Dimorfismo sexual entre hembras y machos de <i>Z. californianus</i>	41
Figura 13: Zona de la toma de muestra de <i>Z. californianus</i> en este estudio.	42
Figura 14: Esquematización de la selección de las regiones conservadas para el diseño de oligonucleótidos cebadores.	53
Figura 15: Esquematización del protocolo general del protocolo de RACE.	58
Figura 16: Esquematización del protocolo general de la PCR Touchdown.	59
Figura 17: Esquematización del protocolo general de la PCR Anidada.	61
Figura 18: Esquematización del control de calidad para las lecturas derivadas de la secuenciación de nueva generación.	66
Figura 19: Diagrama de flujo del funcionamiento de Trinity (Beta).	67
Figura 20: Identidad en la secuencia nucleotídica de las secuencias caracterizadas por secuenciación Sanger de la cadena beta.	74
Figura 21: Identidad en la secuencia nucleotídica de las secuencias caracterizadas por secuenciación Sanger de la cadena heavy.	75
Figura 22: Esquema de la secuencia caracterizada por secuenciación Sanger de la cadena lambda.	76
Figura 23: Análisis de componentes principales de la cadena Alpha muestra AA.	77
Figura 24: Análisis de componentes principales de la cadena Beta muestra BB.	78
Figura 25: Análisis de componentes principales de la cadena heavy de la muestra HH.	79

Figura 26: Análisis de componentes principales de la cadena heavy de la muestra LL.	80
Figura 27: Caracterización de los segmentos génicos V de la cadena alpha del TCR de <i>Z. californianus</i>	81
Figura 28: Caracterización de los segmentos génicos J de la cadena alpha del TCR de <i>Z. californianus</i>	82
Figura 29: Caracterización de los segmentos génicos V de la cadena beta del TCR de <i>Z. californianus</i>	83
Figura 30: Caracterización de los segmentos génicos D y J de la cadena beta del TCR de <i>Z. californianus</i>	84
Figura 31: Configuraciones VJ de la cadena alpha y configuraciones VDJ de la cadena beta del TCR de <i>Z. californianus</i>	85
Figura 32: Caracterización de los segmentos génicos V de la cadena heavy del BCR (IG) de <i>Z. californianus</i>	87
Figura 33: Caracterización de los segmentos génicos D y J de la cadena heavy del BCR (IG) de <i>Z. californianus</i>	88
Figura 34: Caracterización de los segmentos génicos V de la cadena lambda del BCR (IG) de <i>Z. californianus</i>	89
Figura 35: Caracterización de los segmentos génicos J de la cadena lambda del BCR (IG) de <i>Z. californianus</i>	90
Figura 36: Configuraciones VJ de la cadena lambda del BCR (IG) <i>Z. californianus</i>	91
Figura 37: Configuraciones VDJ de la cadena heavy del BCR (IG) <i>Z. californianus</i>	92
Figura 38: Número de segmentos génicos VJ de la cadena alpha del TCR de mamíferos incluidos en este estudio.	93
Figura 39: Número de segmentos génicos V(D)J de la cadena beta del TCR de mamíferos incluidos en este estudio.	93
Figura 40: Numero de segmentos génicos V(D)J de la cadena Heavy del BCR de mamíferos incluidos en este estudio.	94
Figura 41: Configuraciones VDJ de la cadena lambda del BCR (IG) <i>Z. californianus</i>	94
Figura 42: Análisis filogenético molecular de los segmentos TRAV de mamíferos carnívoros.	96
Figura 43: Análisis filogenético molecular de los segmentos TRAV de mamíferos marinos.	97
Figura 44: Análisis Filogenético Molecular de los segmentos TRBV de mamíferos carnívoros.	99
Figura 45: Análisis Fflogenético molecular de los segmentos TRBV de mamíferos marinos.	100
Figura 46: Análisis filogenético molecular de los segmentos IGHV de mamíferos carnívoros.	102
Figura 47: Análisis Filogenético Molecular de los segmentos IGHV de mamíferos marinos.	103
Figura 48: Análisis filogenético molecular de los segmentos IGLV de mamíferos carnívoros.	105

Figura 49: Análisis filogenético molecular de los segmentos IGLV de mamíferos marinos.	106
Figura 50: Análisis Filogenético Molecular de los segmentos TRAJ de mamíferos carnívoros.	108
Figura 51: Análisis Filogenético Molecular de los segmentos TRAJ de mamíferos marinos.	109
Figura 52: Análisis filogenético molecular de los segmentos TRBD de mamíferos carnívoros.	110
Figura 53: Análisis filogenético molecular de los segmentos TRBD de mamíferos marinos.	111
Figura 54: Análisis filogenético molecular de los segmentos TRBJ de mamíferos carnívoros.	112
Figura 55: Análisis Filogenético Molecular de los segmentos TRBJ de mamíferos marinos.	113
Figura 56: Análisis filogenético molecular de los segmentos IGHD de mamíferos carnívoros.	114
Figura 57: Análisis filogenético molecular de los segmentos IGHD de mamíferos marinos.	115
Figura 58: Análisis Filogenético Molecular de los segmentos IGHJ de mamíferos carnívoros.	116
Figura 59: Análisis filogenético molecular de los segmentos IGHJ de mamíferos marinos.	117
Figura 60: Análisis filogenético molecular de los segmentos IGLJ de mamíferos carnívoros.	118
Figura 61: Análisis filogenético molecular de los segmentos IGLJ de mamíferos marinos.	119
Figura 62: Distribución geográfica de las especies de mamíferos del estudio.	120
Figura 63: Análisis filogenético de los segmentos TRAV de las especies ecológicamente relativas a <i>Z. californianus</i>	121
Figura 64: Análisis filogenético de los segmentos TRBV de las especies ecológicamente relativas a <i>Z. californianus</i>	122
Figura 65: Análisis filogenético de los segmentos IGHV de las especies ecológicamente relativas a <i>Z. californianus</i>	124
Figura 66: Análisis filogenético de los segmentos IGLV de las especies ecológicamente relativas a <i>Z. californianus</i>	125
Figura 67: Evolución de las familias de los segmentos V(D)J de los receptores TCR y BCR de <i>Z. californianus</i>	126
Figura 68: Estructura 3D del receptor TCR $\alpha\beta$ de <i>Z. californianus</i>	127
Figura 69: Estructura 3D del receptor IGH λ de <i>Z. californianus</i>	128
Figura 70: Estructura 3D del receptor MHC II de <i>Z. californianus</i>	129

ACRÓNIMOS

APCs	Células presentadoras de antígenos, por sus siglas en inglés
B1	Subtipo de linfocito B que expresa en su membrana el cúmulo de diferenciación 5
B2	Subtipo de linfocito B que no expresa en su membrana el cúmulo de diferenciación 5
BCR	Receptor de linfocitos B
CD25+	Cúmulo de diferenciación 25
CD28	Cúmulo de diferenciación 28
CD4+	Cúmulo de diferenciación 4
CD40L	Ligando del cúmulo de diferenciación 40
CD8+	Cúmulo de diferenciación 8
CD80	Cúmulo de diferenciación 80
CD86	Cúmulo de diferenciación 86
cDNA	ADN que es una copia del ácido ribonucleico mensajero (ARNm)
DLL4	Ligando número 4 de Notch
DNA-PK	Proteína cinasa dependiente de ácido desoxirribonucleico
dNTPs	2'-desoxinucleósido'trifosfato
DSCAM	Molécula de adhesión celular del Síndrome de Down
DTT	Ditiotreitol
EDTA	Ácido etilendiaminotetraacético
FREP	Proteínas relacionadas con fibrinógeno con dominios de inmunoglobulina
Ig	Inmunoglobulina
IgA	Inmunoglobulina isotipo alpha
IgA1	Inmunoglobulina isotipo alpha subclase 1
IgA2	Inmunoglobulina isotipo alpha subclase 2
IgD	Inmunoglobulina isotipo delta
IgE	Inmunoglobulina isotipo épsilon
IgG	Inmunoglobulina isotipo gamma
IgG1	Inmunoglobulina isotipo gamma subclase 1
IgG2	Inmunoglobulina isotipo gamma subclase 2
IgM	Inmunoglobulina isotipo Mu
IL-7	Interleucina 7
IL-7R	Receptor de Interleucina 7
Ku70	Reparador del daño por rayos X al ADN 70
Ku80	Reparador del daño por rayos X al ADN 80

Ligasa DNA IV	Ligasa del ADN IV
LRR	Dominios ricos en repeticiones de Leucina
MHC	Complejo mayor de histocompatibilidad
MHC I	Complejo mayor de histocompatibilidad clase I
MHC II	Complejo mayor de histocompatibilidad clase II
mTECs	Células epiteliales tímicas medulares
PAMPs	Patrones moleculares asociados a patógenos
PHA	Fitohematoaglutinina
RACE	Rápida amplificación de terminación de ADN complementario (ADNc)
RAG1	Activador de la recombinación VDJ 1
RAG2	Activador de la recombinación VDJ 2
RNA	Ácido ribonucleico
RSS	Sitio de reconocimiento de recombinación
Tc	Linfocito T citotóxico
TCR	Receptor de linfocitos T
TCR $\alpha\beta$	Receptor de linfocitos T cadenas alpha y beta
TCR $\gamma\delta$	Receptor de linfocitos T cadenas gamma y delta
TD	Activación de linfocitos B dependiente de linfocitos T
TdT	Desoxinucleotidil transferasa terminal
Th	Linfocito T cooperador
Th1	Linfocito T cooperador subtipo 1
Th17	Linfocito T cooperador subtipo 17
Th2	Linfocito T cooperador subtipo 2
Th22	Linfocito T cooperador subtipo 22
Th9	Linfocito T cooperador subtipo 9
TI	Activación de linfocitos B independiente de linfocitos T
TNF	Factor de necrosis tumoral
Treg	Linfocitos T reguladores
VCBP	Familia de receptores que codifica dominios IgV unidos a quitina

RESUMEN

La efectividad del reconocimiento antigénico y generación de memoria inmunológica está coordinada por los receptores antigénicos TCR, BCR y MHC. Los primeros sufren modificaciones durante el desarrollo de linfocitos T y B, en las regiones génicas de variabilidad, diversidad y unión (V(D)J). Estos segmentos se diversifican mediante mecanismos de origen y muerte a través de las cadenas que forman estos receptores, por lo que el estudio de su evolución genera conocimiento importante sobre la historia de vida antigénica y los canjes evolutivos inmunológicos de las especies. En esta tesis se caracterizaron a los receptores antigénicos TCR y BCR en una especie modelo no tradicional, el lobo marino de California, *Zalophus californianus*. Este es un mamífero carnívoro semiacuático que durante su reinserción a la vida en el océano, estuvo sometido a un amplio espectro de presiones patogénicas y ambientales. Con base en los resultados obtenidos en esta investigación, se propone que las presiones selectivas llevaron a canjes evolutivos y una tendencia a la expansión del repertorio de las cadenas que forman el receptor de linfocitos T. A su vez, estos hechos evolutivos condujeron a un sesgo hacia respuestas celulares en vez de respuestas humorales, debido a que su repertorio BCR (Ig) en comparación con el repertorio del TCR, esta reducido a una familia de segmentos de variabilidad. Los receptores TCR y BCR de *Z. californianus* conservan sitios de unión antigénica con otros mamíferos, lo que vuelve a este pinnípedo una especie de interés biomédico, para el estudio del desarrollo de enfermedades infecciosas, cáncer y tolerancia inmune. Esta es la primera investigación que examina la conformación y explora la historia evolutiva de los receptores linfocitarios en un pinnípedo.

Palabras clave: TCR, BCR, evolución, V(D)J, receptor antigénico

ABSTRACT

The effectiveness of antigen recognition and the generation of immunological memory is coordinated by the TCR, BCR and MHC antigenic receptors. The former experience genetic modifications during early development of T and B lymphocytes, particularly in the variability, diversity and union regions (V (D) J). These segments are diversified through mechanisms of birth and death that occur in the chains of these receptors. The study of their evolution can generate interesting insights about the antigenic life-history and the immunological evolutionary trade-offs of a given species. This thesis characterized the TCR and BCR receptors of a non-model species, the California sea lion, *Zalophus californianus*. The species is a semi-aquatic marine carnivore that since its return to the aquatic environment became exposed to a wide spectrum of pathogenic and environmental pressures. Based on the results found in this study, it appears that the species underwent evolutionary trade-offs and an expansion of its T lymphocyte repertoire, which resulted in a bias towards cellular responses instead of humoral responses, as its BCR (Ig) repertoire is reduced compared to the TCR repertoire, consisting in a single family of variability segments. The TCR and BCR receptors of the California sea lion conserve antigenic binding sites with other mammals, which makes this pinniped a species of biomedical interest, particularly for the study of infectious disease ontogeny, cancer and immune tolerance. This is the first study to examine the configuration and evolutionary history of the TCR and BCR receptors for any pinniped species.

Keywords: TCR, BCR, evolution, V(D)J, antigenic receptor

I. Introducción

La principal tarea del sistema inmune es distinguir entre lo propio y lo extraño. De manera general, este sistema está estratificado en dos tipos; innata y adaptativa (Abbas *et al.*, 2017). El sistema inmune adaptativo, presente únicamente en los vertebrados gnatoostomados, orquesta las respuestas por medio de linfocitos T y B, que reconocen antígenos propios y extraños mediante sus receptores TCR y BCR, respectivamente (Murphy & Weaver, 2017).

Los receptores TCR y BCR son altamente variables debido a que durante su desarrollo, los segmentos que los conforman son reordenados somáticamente (Murphy & Weaver, 2017). Además, son adicionados nucleótidos a los receptores re-arreglados, un fenómeno conocido como 'arreglo del receptor', lo que genera todavía más diversidad. Los segmentos génicos que conforman a estos receptores son de tres tipos: Variabilidad (V), Unión (J) y Diversidad (D), genéricamente referidos como V(D)J (Outters *et al.*, 2015). Los genes V y J están en las cadenas ligeras y pesadas, como lo son la cadena alpha y lambda, mientras que los genes D están únicamente en las cadenas pesadas como la cadena beta y heavy (Abbas *et al.*, 2017).

La diversidad de los segmentos génicos V(D)J varía en cada especie (Guo *et al.*, 2011; Olivieri *et al.*, 2014; Qin *et al.*, 2014; Ujvari & Belov, 2015; Qin *et al.*, 2015; Murphy & Weaver, 2017), lo que refleja su historia evolutiva en determinado hábitat (Olivieri *et al.*, 2014; Olivieri *et al.*, 2015). Por lo tanto, los loci de las cadenas de los receptores muestran tendencias de expansión en una cadena o más (Olivieri *et al.*, 2014) o la pérdida de una cadena por el desuso. Este proceso es coordinado por mecanismos de origen y muerte, lo cual da origen a genes funcionales, pseudogenes y genes de marco abierto de lectura (Nei *et al.*, 1997).

En este trabajo se caracterizarán molecularmente los segmentos que forman a las cadenas alpha y beta del TCR y las cadenas heavy y lambda del BCR del lobo

marino de California (*Zalophus californianus*), una especie modelo no tradicional que es considerada centinela del ecosistema costero marino y para el que existe buena información sobre los patógenos que lo afectan en vida libre. Dado que no es una especie modelo para la que exista información genómica completa, se utilizó la técnica de 5'/3' RACE y la secuenciación de nueva generación, para tener una mayor cobertura de los segmentos V(D)J y poder comparar la evolución de su repertorio, con respecto a la de otros mamíferos para los que ya está disponible esta información. También fue modelada la estructura putativa de ambos receptores y se buscaron secuencias de aminoácidos que podrían reflejar los procesos co-evolutivos entre el receptor TCR y el MHC II en esta especie.

II. Antecedentes

1.1 El sistema inmune y la evolución del reconocimiento antigénico

El sistema inmunológico es el conjunto de respuestas que permiten a un organismo el discernir entre lo propio y lo extraño. De manera convencional se divide en respuestas innatas y adaptativas, de las que las innatas se encuentran presentes en todos los seres vivos, mientras que las respuestas adaptativas sólo evolucionaron en los vertebrados mandibulados. El sistema inmune innato es rápido e inespecífico, ya que depende del reconocimiento de patrones moleculares generales asociados a patógenos o a daño celular (Abbas *et al.*, 2017). En contraste, las respuestas adaptativas pueden tomar desde días hasta meses en implementarse, pero son específicas contra antígenos foráneos o propios alterados y generan memoria (Murphy & Weaver, 2017; Cooper & Alder, 2006).

Los productos génicos que se expresan en la superficie celular de células efectoras del sistema inmune adaptativo, los linfocitos T y B, son los principales mediadores del reconocimiento antigénico, debido a que son la interfaz con otras células en el medio extracelular, además de ser claves para diversos procesos metabólicos y de tolerancia inmune (Eason *et al.*, 2004).

Por evolución convergente, en el genoma de la mayoría de los seres vivos se encuentran agrupados los genes que codifican para productos relacionados con el reconocimiento de lo foráneo (Cooper & Alder, 2006). Estos sistemas sofisticados orquestan las respuestas inmunes innatas y adaptativas entre taxones (Figura 1) (Boehm, 2006). Aunque los componentes estructurales de estos sistemas varían en su mayoría (Cannon *et al.*, 2002; Zhang *et al.*, 2004; Holland *et al.*, 2014), cada sistema se ha sofisticado de diversas formas para reconocer un amplio espectro de epítomos foráneos o propios alterados, y patrones moleculares asociados a patógenos (PAMPs) o daño (DAMPs).

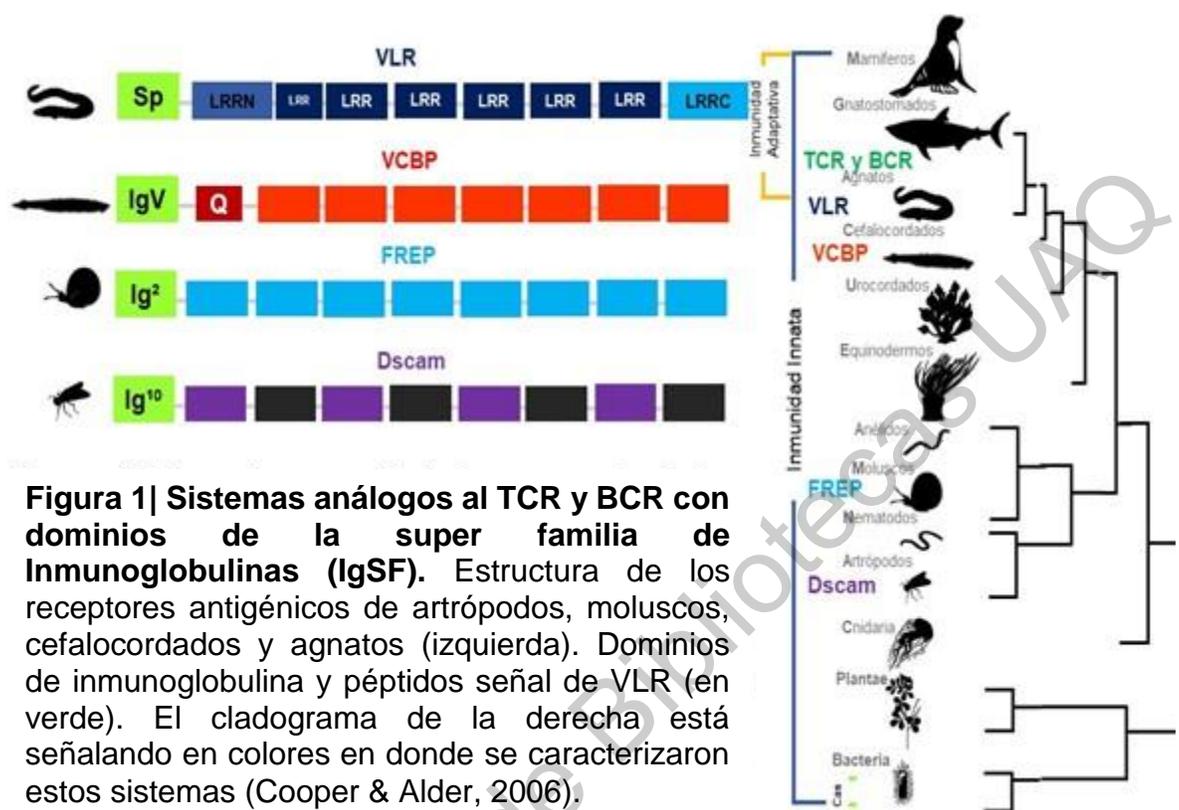


Figura 1| Sistemas análogos al TCR y BCR con dominios de la super familia de Inmunoglobulinas (IgSF). Estructura de los receptores antigénicos de artrópodos, moluscos, cefalocordados y agnatos (izquierda). Dominios de inmunoglobulina y péptidos señal de VLR (en verde). El cladograma de la derecha está señalando en colores en donde se caracterizaron estos sistemas (Cooper & Alder, 2006).

Interesantemente, aunque carecen de los mecanismos de generación de diversidad y memoria, los sistemas VCBP, FREP y Dscam de los invertebrados tienen cierta especificidad inmunológica al adaptar los dominios de la súper familia de las inmunoglobulinas (IgSF) (Dermody *et al.*, 2009). Esta súper familia también está presente en los receptores de linfocitos T (TCR) y B (BCR), por lo cual son catalogados como sus posibles ancestros (Cooper & Alder, 2006). Sin embargo, a pesar de que comparten algunas características estructurales, estos sistemas generan diversidad inmunológica de maneras distintas. Por una parte, el Dscam debe su diversidad al uso de exones variables y no variables para generar un repertorio amplio (Watson *et al.*, 2005). Los VCBP y FREP por otro lado, usan dominios de quitina y fibrinógeno respectivamente adicional a los dominios de IgSF (Cannon *et al.*, 2002; Zhang, 2004). A pesar de que los receptores mencionados arriba, participan en la respuesta inmune de diferentes organismos con alta especificidad, reconociendo PAMPs y DAMPs, al no generar memoria

inmunológica, ni están sometidos a restricciones de tolerancia central a diferencia de los TCRs y BCRs, son reconocidos como integrantes de la respuesta innata (Boehm *et al.*, 2012).

Debido a la sofisticación del sistema de reconocimiento antigénico de los TCRs y BCRs, ha sido difícil reconstruir su historia evolutiva antes de su diversificación de otros sistemas (Boehm & Swann, 2014). Sin embargo, debido a que los sistemas que se mencionan anteriormente usan a los miembros de la IgSF se reconoce a esta familia como posible precursora. Esta familia, puede adaptar sus pliegues a distintos procesos y duplicarlos fácilmente, por lo que hace aproximadamente 500 millones de años (Laird *et al.*, 2000), eventos de duplicación en los dominios variables (V) de esta familia dieron lugar a las regiones determinantes de complementariedad (CDR), en el ancestro común de los vertebrados (Dermody *et al.*, 2009). Estas regiones, conformadas por los segmentos génicos V(D)J, son un punto clave para entender la evolución del sistema de reconocimiento antigénico de los linfocitos T y B, debido a que el evento que originó esta sofisticación tuvo lugar entre los segmentos V y J, lo que posteriormente dio lugar a los segmentos de diversidad (D) (Laird *et al.*, 2000; Outters *et al.*, 2015). Estos procesos explican la generación de nuevas especies génicas especializadas provenientes de un sistema ancestral (Boehm & Swann, 2014). Otro de los mecanismos que han permitido la super especialización de los sistemas TCR y BCR, es la acción catalítica de las enzimas de los genes de activación de la recombinación (RAG1 & RAG2, por sus siglas en inglés), las cuales se encargan de la recombinación de los segmentos V(D)J (Abbas *et al.*, 2017). Estas enzimas se encuentran conservadas en vertebrados, pero el lugar en el que actúan es diferente entre mamíferos, aves, peces óseos y peces cartilaginosos y, en consecuencia, la contribución de los segmentos V(D)J a lo largo de los taxones es diversa, particularmente en los BCRs (Litman *et al.*, 2005).

1.2 Células del sistema inmune adaptativo de mamíferos

La diversificación entre los vertebrados superiores trajo consigo innovaciones morfológicas importantes, las cuales dieron lugar a nuevas funciones inmunológicas, como la alta especificidad y la memoria inmunológica, a través del origen de órganos super especializados que permitieron controlar mecanismos de tolerancia central y expansión clonal (Figura 2) (Boehm & Bleul, 2007). A partir de estas nuevas funciones, se especializaron dos tipos de respuestas adaptativas, las celulares y las humorales, que son llevadas a cabo por los linfocitos T y B, respectivamente (Boehm & Swann, 2014).

1.2.1 Linfocitos T

Los linfocitos T son células mononucleares del sistema inmune adaptativo de los vertebrados que coordinan la respuesta celular. Estas células reconocen a los antígenos que les son presentados por células presentadoras de antígenos (APCs: células dendríticas, macrófagos y linfocitos B) mediante el complejo mayor de histocompatibilidad (MHC). La arquitectura genética del MHC se divide en cuatro regiones, conocidas como clases; MHC I, II, III y IV (Robinson *et al.*, 2012). Los péptidos del MHC I, son expresados en la superficie celular de todas las células nucleadas de los vertebrados, y mediante su reconocimiento, activan a los linfocitos T citotóxicos (Tc). Los péptidos del MHC II son expresados en la superficie de las APCs y activan a los linfocitos T cooperadores (Th). Estas células contribuyen a la generación de memoria inmunológica, cuando sus subtipos se diversifican en células de memoria y efectoras (Abbas *et al.*, 2017).

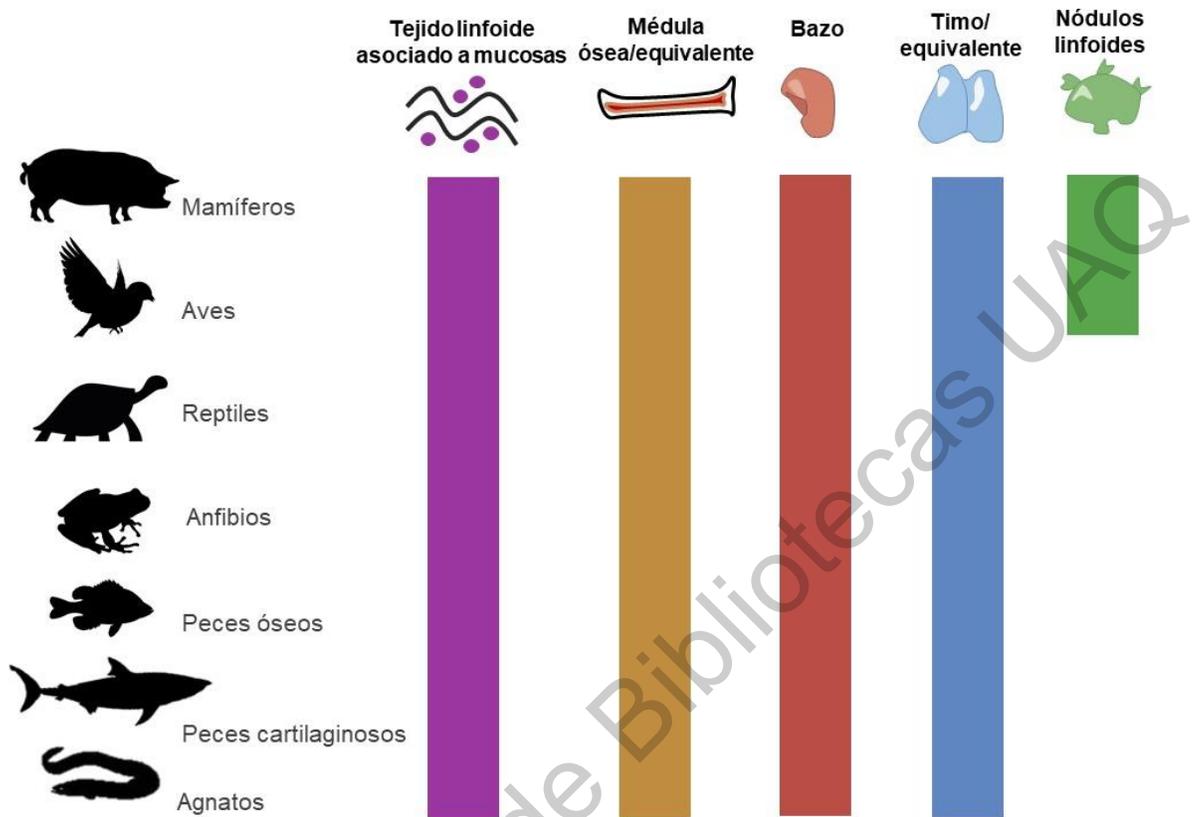


Figura 2| Evolución de los órganos linfoides. El sistema inmunológico de los mamíferos y aves se caracteriza por la presencia de órganos linfoides secundarios, como los nódulos linfoides (en verde), estos tejidos sirven para que las células del sistema inmune interactúen. Dado que en los vertebrados más basales están presentes los genes TNF e IL-7R, que han sido asociados a la aparición de estos órganos en los mamíferos, se ha sugerido una convergencia evolutiva en estos genes. El timo y la médula ósea (azul y marrón) son órganos primarios que están presentes en la mayoría de los vertebrados debido a que son órganos precursores de poblaciones celulares inmunes. El bazo y el tejido linfoide asociado a mucosas (rojo y morado), son órganos linfoides secundarios que evolutivamente han permitido restricciones de tolerancia inmune. Imagen adaptada de: (Boehm & Bleul, 2007).

Los linfocitos T derivan del progenitor linfoide en la médula ósea y migran al timo para continuar su desarrollo durante las etapas tempranas del desarrollo (Abbas *et al.*, 2017). En el timo, los progenitores de los linfocitos T son sometidos a filtros de selección positiva y negativa (Murphy & Weaver, 2017). La selección positiva tiene lugar en la corteza tímica, en donde se seleccionan aquellos linfocitos con un TCR re-arreglado y maduro. La selección negativa es mediada por las células epiteliales tímicas medulares (mTECs); durante este proceso son eliminados los linfocitos cuyos receptores reconocen antígenos propios (Kyewski & Feuerer, 2014).

Una vez finalizado su desarrollo en timo, los linfocitos T que cuentan con un receptor aún virgen (*naïve*), migran a los nódulos linfoides, en donde las APCs les presentan antígenos derivados de organismos patógenos o de células transformadas (Abbas *et al.*, 2017). Esta presentación, en conjunto con la síntesis de citocinas por parte de diversos efectores inmunes, da lugar a la activación de los subtipos de linfocitos T (Murphy & Weaver, 2017). El subtipo cooperador engloba a las células que expresan en la membrana a la glicoproteína CD4 (Th1, Th2, Th9, Th17, Th22 y T reguladoras) (Araujo-Pires *et al.*, 2014). Mientras que el subtipo citotóxico engloba a los linfocitos que expresan en su membrana a la glicoproteína CD8 (Abbas *et al.*, 2017).

Los subtipos de linfocitos T son los encargados de dirigir los distintos tipos de respuestas por medio de la liberación al espacio extracelular de enzimas por parte de las CD8+ y de citocinas por parte de las CD4+ (Dustin & Choudhuri, 2016). Esta respuesta inmunológica se clasifica como tipo 1, ya que es dirigida por los linfocitos Th1, y se relaciona a las respuestas de reconocimiento de antígenos derivados de bacterias intracelulares, virus o autoinmunidad (Bretscher., 2014), a esta respuesta también se le denomina pro-inflamatoria, por lo tanto promueve la síntesis de las citocinas IL-1 β , TNF- α , IL-6 e INF- γ (Murphy & Weaver, 2017). Por otro lado, la respuesta coordinada por linfocitos Th2 coordina la respuesta tipo 2, conocida también como anti-inflamatoria, la cual en respuesta a alérgenos o antígenos

derivados de parásitos helmintos, produce las citocinas IL-4, IL-10, TGF- β (Paul & Zhu, 2010). El tiempo en el que estas dos respuestas se activan es determinante para la protección del hospedero. Por esto, la respuesta pro-inflamatoria (Th1) es la primera en activarse para remover al patógeno, seguida por la respuesta (Th2) encargada de modular a la Th1 y no generar una respuesta exacerbada que perjudique al hospedero. De esta manera, las respuestas Th1 y Th2 actúan en sinergia y sin excluirse, inclusive en la defensa frente a un mismo patógeno (Cicchese *et al.*, 2018).

Existen algunas respuestas que son dirigidas por otros subtipos linfocitarios como las células Th17, que principalmente se relacionan con respuestas contra bacterias extracelulares (Piccirillo, 2008); los Tregs, involucrados en el control de tolerancia y supresión tumoral (Louten *et al.*, 2009) y las Th9, que también participan en las respuestas a helmintos (Kaplan *et al.*, 2015). De manera general, estas respuestas también son clasificadas como tipo 1 y tipo 2.

a) Receptor de linfocitos T

El receptor de linfocitos T es un heterodímero de dominios de la súper familia de inmunoglobulinas, que forman las cadenas constantes y variables, además de un segmento transmembranal y una cola citoplasmática, la cual es el paso inicial en la señalización intracelular de linfocitos T, permitiéndoles activarse durante el reconocimiento del complejo p-MHC (Figura 3) (Murphy & Weaver, 2017).

Receptor de linfocitos T (TCR)

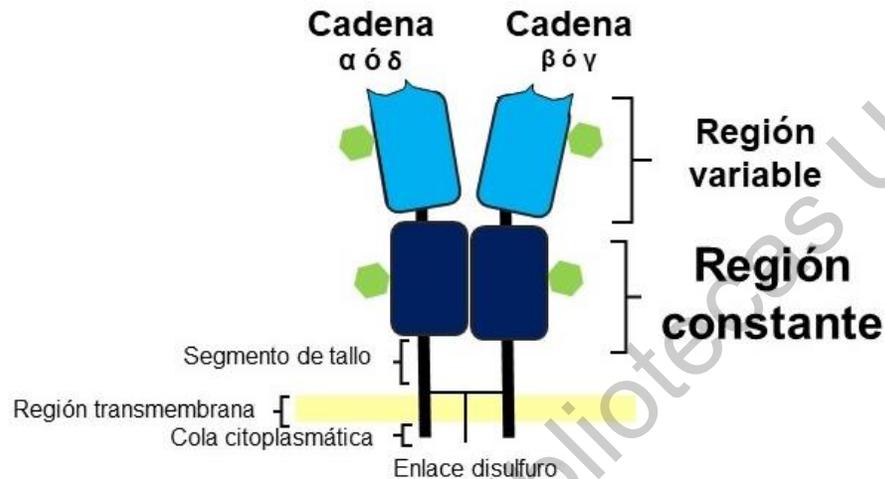


Figura 3| Estructura del receptor de linfocitos T. En la imagen se muestra la estructura del receptor de linfocitos T (TCR), cadenas $\alpha\beta$ y $\gamma\delta$ unidas por un enlace disulfuro. Las cadenas constan de una región constante y variable, a las cuales están unidos carbohidratos tipo hexosas (en verde). Además del segmento del tallo que se dividen en la región transmembranal y cola citoplásmica. Imagen adaptada de: (Murphy & Weaver, 2017).

La configuración genética del receptor de linfocitos T está compuesta por cuatro tipos de segmentos génicos: Los segmentos variables (V), unión (J), diversidad (D) y constante (C). Los segmentos V, J y C forman las cadenas alpha (α) y gamma (γ), mientras que las cadenas beta (β) y gamma (δ), además de los V, J y C, tienen segmentos D. Cuando los segmentos génicos que forman las cadenas α , β , γ y δ se reorganizan para formar el receptor generan la alta especificidad y diversidad clonal. De esta manera, además de generar variabilidad en los segmentos de cada cadena, la súper-especificidad también es posible, debido a que los segmentos V(D)J, provienen de distintos locus, de otra manera estos receptores serían homodímeros de una cadena no variables. (Figura 4) (Abbas *et al.*, 2017).

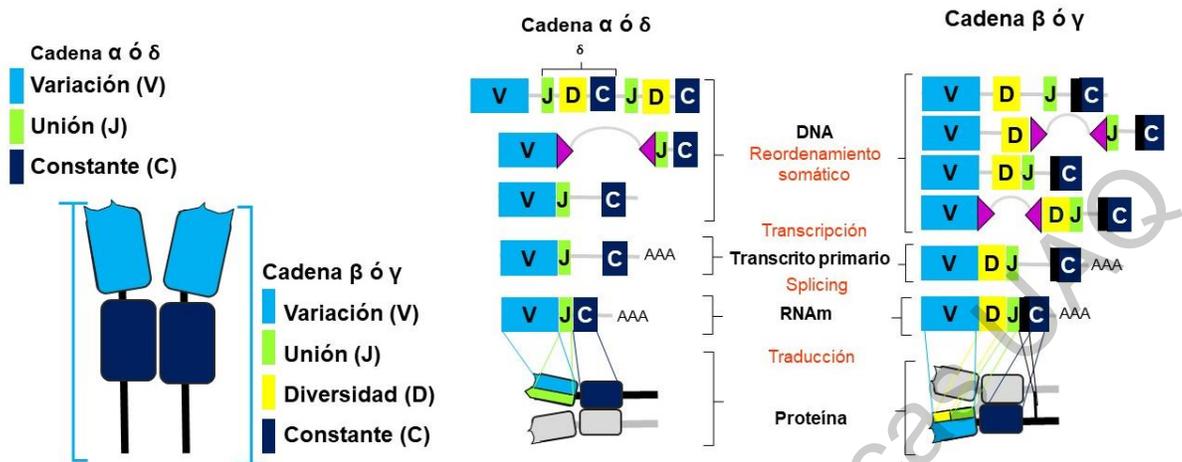


Figura 4| Formación del receptor de linfocitos T. Las cadenas del receptor se forman por la unión de los segmentos J y D, al V, la unión ocurre cuando los sitios RSS (en morado), de los V y J, se reconocen. Esta unión da lugar al reordenamiento somático para que se transcriba. Una vez formado el transcrito primario se eliminan los intrones por corte-empalme (splicing), lo cual permite que los segmentos VDJ se unan a los segmentos C, para que se traduzcan las proteínas de las cadenas que se ensamblaran para formar el receptor. Imagen modificada de Murphy & Weaver, 2017.

El reordenamiento de los genes V(D)J, tiene lugar en el timo y es mediado por las enzimas RAG1 y RAG2 cuando reconocen sitios palindrómicos en los extremos 3' y 5', denominados sitios de reconocimiento de reordenamiento (RSS). RAG1/2 se unen a estos sitios promoviendo la escisión de segmentos V(D)J en el locus de las cadenas del receptor. Los sitios que quedan separados por la ruptura son reparados y ligados por las enzimas Ku70 y 80. Posteriormente el complejo DNA-PK y Artemis dan lugar a sitios monocatenarios para unir los extremos que RAG1/2 dejó expuestos. Finalmente, los extremos son unidos por la ligasa de DNA IV con ayuda de la desoxinucleotidil-transferasa terminal (TdT) (Rossjohn *et al.*, 2015).

Aunque inicialmente se pensaba que el proceso estaba regulado únicamente por los efectores mencionados arriba. Estudios posteriores demostraron que existen procesos transcripcionales adicionales que regulan la selección de un segmento

V(D)J, como su cercanía a sitios de unión a factores de transcripción como CTCF (Proudhon *et al.*, 2015).

Los segmentos génicos que codifican para las cadenas del TCR, se encuentran en diferentes loci, a excepción de las cadenas α y δ debido a que, los segmentos V y J de la cadena δ , se encuentran entre los de la cadena α (Abbas *et al.*, 2017). Sin embargo, debido a que estos segmentos génicos sufren exclusión alélica entre sí, sólo pueden expresarse dos tipos de receptores, TCR $\alpha\beta$ o TCR $\gamma\delta$; el primer receptor es el más abundante (Zarin *et al.*, 2015), mientras que el receptor $\gamma\delta$ está diferencialmente expresado en células residentes de epitelio. Además, la activación del TCR $\gamma\delta$, a diferencia del TCR $\alpha\beta$, es independiente del MHC II (Holland *et al.*, 2014).

Una vez que la versión reordenada del receptor se transcribe y traduce, este adquiere una nueva clasificación, regiones CDR (regiones altamente variables) y regiones Framework (FR; regiones más conservadas). Los CDR, se subclasifican en CDR1, CDR2, que comprenden sólo segmentos V, que provienen de línea germinal, y la región CDR3. La región CDR3 comprende la unión de los segmentos V(D)J en las cadenas pesadas y VJ de las cadenas ligeras. De la misma forma, las regiones FR también se subclasifican en FR1, FR2, FR3 y FR4 (Abbas *et al.*, 2017).

1.2.2 Linfocitos B

La respuesta humoral adaptativa de los mamíferos es orquestada por los linfocitos B, los cuales contribuyen a la activación de células del sistema inmune innato y efectores del sistema de complemento, en su versión soluble conocida como anticuerpos o inmunoglobulinas (Ig) (Abbas *et al.*, 2017). Los linfocitos B, a diferencia de los linfocitos T, no son completamente dependientes de la activación por el MHC, debido a que pueden reconocer epítomos libres sin restricción por parte de otra molécula (LeBien & Tedder, 2008). Además, las inmunoglobulinas activan

vías de señalización del complemento y a células del sistema inmune innato por medio de sus regiones constantes (Fc), sin restricción más que la complementariedad molecular (Murphy & Weaver, 2017). Adicional a esta idea los linfocitos B también fungen un papel como APC no especializada activando a linfocitos T (Preciado-Llanes *et al.*, 2014).

Aunque, los linfocitos B derivan del progenitor linfoide común en la médula ósea, al igual que los linfocitos T, no salen de la médula ósea para continuar su desarrollo, sino que permanecen en ella hasta que madura su receptor (Dorshkind & Montecino-Rodriguez, 2007). En el proceso de maduración de los linfocitos B las células estromales de la médula ósea cumplen un papel similar al de las mTec con las células T (Sellars, 2011). Después de concluir su desarrollo, estas células migran a los nódulos linfoides y al bazo como linfocito B virgen, expresando en su membrana BCR/IgM (inmunoglobulina M) e Ig D (Hoernes *et al.*, 2011).

En los nódulos linfoides se activan al reconocer un patógeno y escinden segmentos de la cadena Heavy del BCR/IgM, para generar diferentes isotipos de inmunoglobulinas (IgM, IgD, IgG1, IgG2, IgA1, IgA2 e IgE) (Hoffman *et al.*, 2015). Similar a lo que ocurre con los linfocitos T, existen subpoblaciones de linfocitos B, que son categorizadas por la expresión de glicoproteínas en la superficie celular en B1 CD5+ y B2 o linfocitos B convencionales CD5- (Sellars *et al.*, 2011).

La diferenciación de estos subtipos también requiere de citocinas en el medio extracelular para su diferenciación y maduración, como linfocitos B de memoria o células plasmáticas (Batista & Harwood, 2009). Existen dos mecanismos para generar estos dos tipos de linfocitos B. La activación dependiente de linfocitos T (TD) ocurre cuando el linfocito B reconoce al TCR y CD40 del linfocito T por medio de su MHC II y CD40L (Martin *et al.*, 2001), o por activación independientemente de linfocitos T (TI), en la que la célula B entra en contacto directo con un antígeno mediante su BCR (Preciado-Llanes *et al.*, 2014).

a) *Receptor de linfocitos B*

El receptor de linfocitos B (BCR) es un receptor heterodimérico con dominios de inmunoglobulinas (Yang & Reth, 2010). Formado por una cadena pesada denominada μ y dos ligeras, κ y λ . éstas a su vez están constituidas por los segmentos génicos V(D)J y C para la cadena μ y V, J y C para las cadenas κ y λ (Figura 5) (Treanor, 2012).

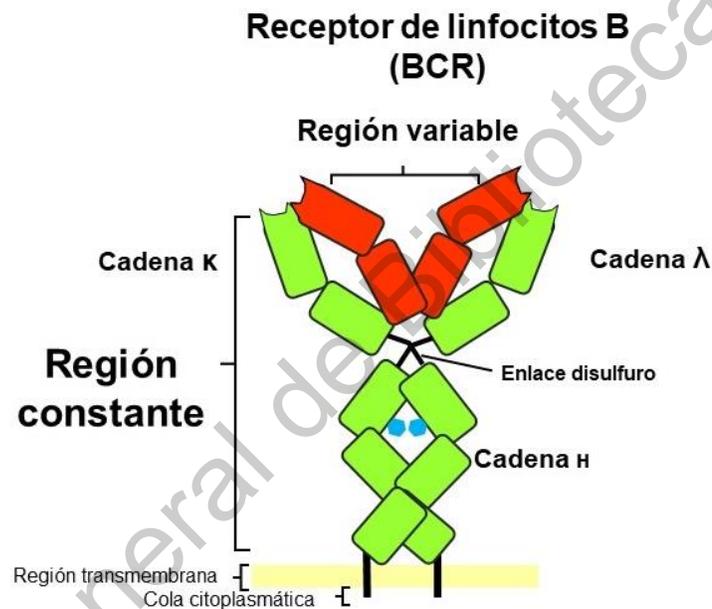


Figura 5| Estructura del receptor de linfocitos B. En la imagen se muestra la estructura proteica del receptor de linfocitos B (BCR), la región pesada formada por la cadena μ y la región ligera por las cadenas κ y λ . Indicado por amarillo, el segmento de cruce transmembranal conformado únicamente por segmentos constantes, denominada cola citoplásmica la cual permite al linfocito B activarse después del reconocimiento de un epítipo. Imagen modificada de Murphy & Weaver, 2017.

El BCR y el TCR comparten la maquinaria molecular para re-arreglar los segmentos V(D)J que conforman sus cadenas (Figura 6). Sin embargo, a diferencia del TCR, el BCR adicional al re-arreglo de sus genes, sufre mutación somática en todas sus cadenas, lo que genera una amplia gama de inmunoglobulinas (Abbas *et al.*, 2017).

Este evento tiene lugar en la zona clara de los nódulos linfoides (Defranco *et al.*, 1993). Una de las diferencias más notables entre los receptores de linfocitos T y B, es que los linfocitos B pueden secretar su receptor, inclusive en pentámeros de IgM (Murphy & Weaver, 2017).

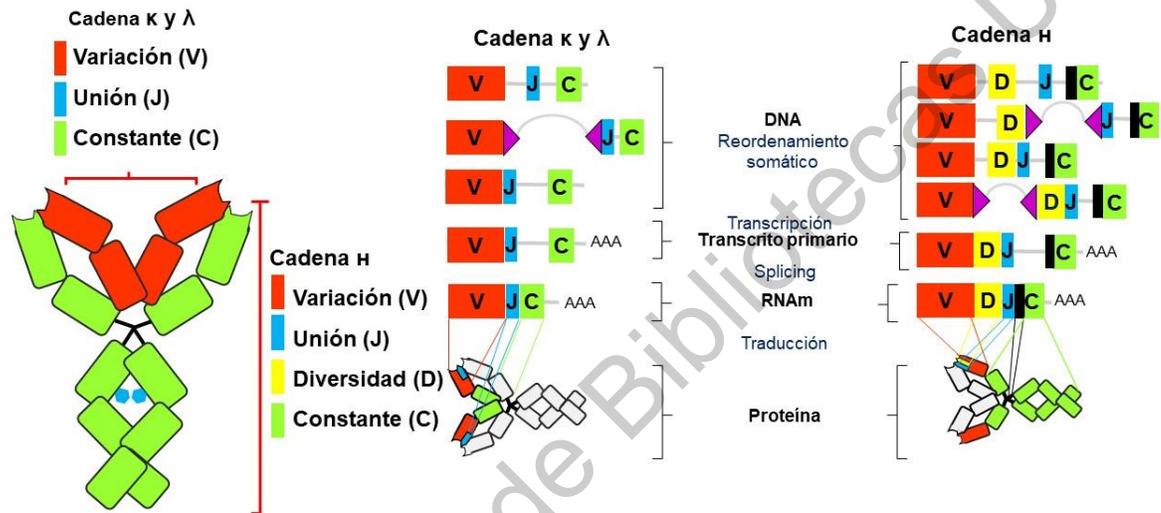


Figura 6| Formación del receptor de linfocitos B. Las cadenas del receptor se forman por la unión de los segmentos J y D, al V, la unión ocurre cuando los sitios RSS (en morado), de los V y J, se reconocen. Esta unión da lugar al reordenamiento somático para que se transcriba. Una vez formado el transcrita primario se eliminan los intrones por corte y empalme, lo cual permite que los segmentos V(D)J se unan a los segmentos C, para que se traduzca el receptor. Posterior a la reorganización, en centros germinales el receptor BCR madura su afinidad por mutación somática. Imagen modificada de Murphy & Weaver, 2017.

En comparación con el receptor de células T, el BCR parece haber adquirido más innovaciones durante su evolución. La más importante de estas es su capacidad para generar versiones secretables de sí mismo; (IgM, IgD, IgG, IgE e IgA), las cuales pueden desencadenar respuestas específicas (Yang & Reth, 2010). De manera interesante, aunque se conoce la maquinaria que permite el corte y ensamblaje de la cadena H, no se conoce hasta ahora el principio que dio origen a este proceso, se hipotetiza que, en el caso de la IgA, la disbiosis intestinal fue el evento que generó

el corte de esta cadena cuando hubo necesidad de recubrir las mucosas (Mantis *et al.*, 2011). Adicionalmente, algunos mamíferos han perdido la capacidad de generar estos subtipos de inmunoglobulinas, específicamente los relacionados a activación de efectores de la respuesta innata y adaptativa (Li *et al.*, 2018).

1.3 Los segmentos V(D)J a través de los taxones de mamíferos

La aportación de los genes V, D y J en los loci de las cadenas del TCR y BCR(Ig), varía entre taxones. Sin embargo, las cadenas que conforman a estos receptores permanecen en la mayoría de los vertebrados (Abbas *et al.*, 2017).

En algunas especies como monotremas, marsupiales y peces existen *loci* que forman a los receptores únicos en esas especies, que surgen como estrategia de adaptación a su hábitat (Parra *et al.*, 2007; Wang *et al.*, 2011; Mashoof & Criscitiello, 2016). Por otro lado, el origen de una cadena extra en los receptores TCR y BCR (Ig), no es la única forma en la que los vertebrados adaptan sus requerimientos inmunes de hecho, es más común la expansión del repertorio de una cadena como mecanismo de adaptación a determinado hábitat en este sentido, especies cercanas pueden tener repertorios muy diferentes entre sí (Olivieri *et al.*, 2014) (Figura 7-11).

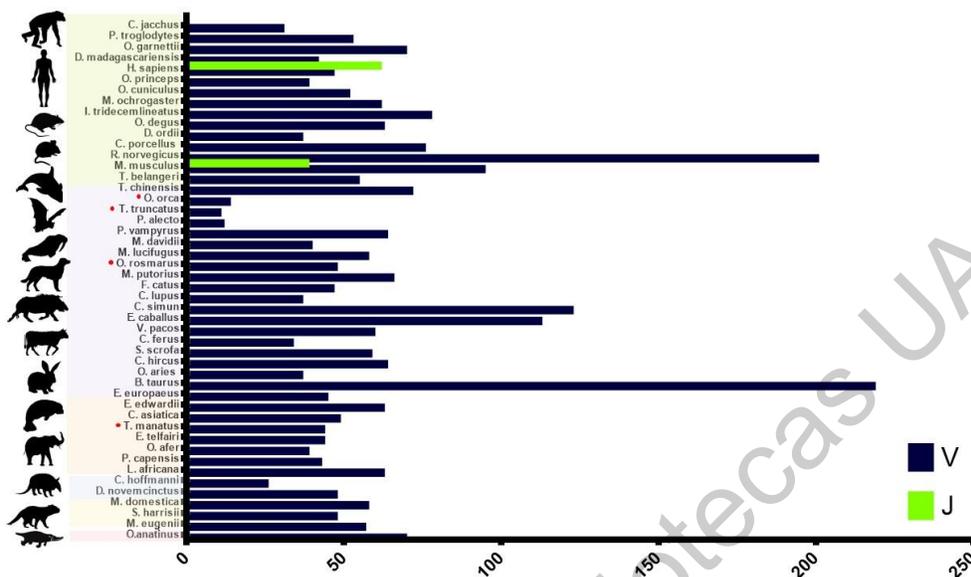


Figura 7| Repertorio de genes V y J de la cadena alpha en los mamíferos. En azul marino, se muestra el repertorio de los genes V. En verde, se muestra el repertorio de genes J. Los puntos rojos indican a especies de mamíferos marinos. Imagen adaptada de Olivieri *et al.*, 2014 y de Murphy & Weaver, 2017.

Aunque no se conoce cuáles sean los hábitats o las condiciones que favorecen la expansión de una cadena determinada, la expansión está regulada por el mecanismo de origen y muerte, en el cual se mantienen los segmentos que representan ventaja como genes funcionales y se descartan en forma de pseudogenes los que no representan una ventaja inmunológica, para después ser eliminados por cruce desigual (Nei & Sitnikova, 1997; Das *et al.*, 2008).

La hipótesis que mantiene que el repertorio de genes V(D)J en las cadenas de los receptores está sujeta a selección positiva por presión patogénica, también refiere que en las cadenas del TCR, en especial las cadenas alpha y beta, el repertorio está comprometido a la selección de los alelos del MHC clase II, por lo que el repertorio de determinada especie no sólo tiene que ser reactivo a patógenos extrínsecos, sino que también mantener grados de tolerancia a los propios, para poder permanecer (Nikolich-Žugich *et al.*, 2004).

Si bien el número de genes V(D)J, está relacionado con la diversidad de patógenos que puede reconocer el sistema inmune de una especie, existen reportes en los que la reducción en la diversidad de los receptores T, no compromete la reactividad (Bouso *et al.*, 2000). Por otro lado, otros estudios demuestran que reducir el repertorio de las cadenas del TCR, reduce la capacidad de montar resistencia hasta 50% menos e impacta en la reactividad hacia el MHC II, lo que posiblemente demuestra que la reactividad, está comprometida con el contexto de una especie (Nanda *et al.*, 1991).

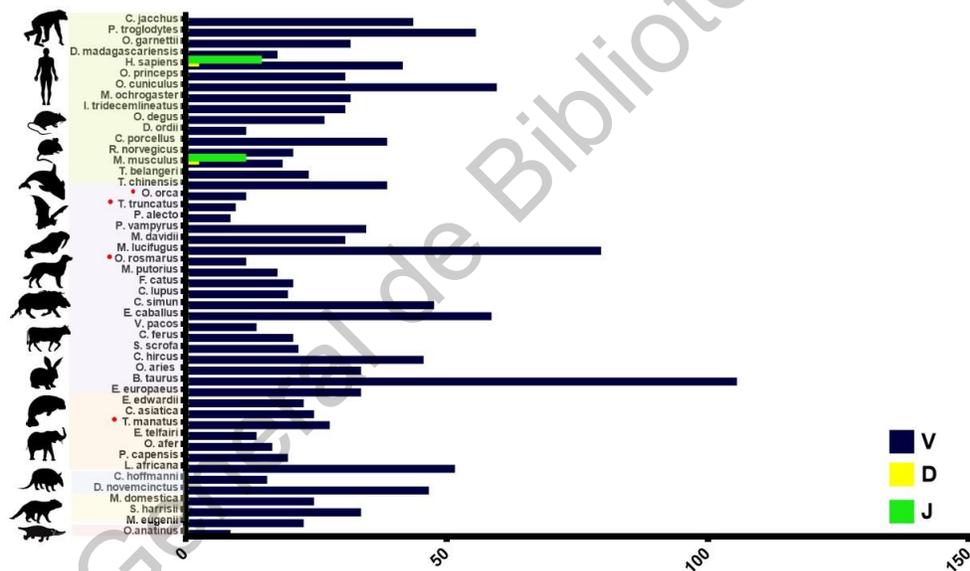


Figura 8| Repertorio de genes V (D) J de la cadena beta en los mamíferos. En azul marino, se muestra el repertorio de los genes V. En amarillo, se muestra el repertorio de genes D y en verde, se muestra el repertorio de genes J. Los puntos rojos indican a los mamíferos marinos. Imagen adaptada de Olivieri *et al.*, 2014 y Murphy & Weaver, 2017.

En mamíferos, los segmentos de variabilidad de las cadenas del TCR provienen de nueve familias ancestrales TRAV1-5 en la cadena alpha y TRBV1-4 en la cadena beta (Olivieri *et al.*, 2015) y se diversifican de maneras distintas en todas las

especies, proporcionándoles mayor inmunocompetencia, de manera que existen hasta 43 familias conocidas en algunos carnívoros en la cadena alpha (Figura 7) y 30 en la beta (Martin *et al.*, 2017) (Figura 8). Por otro lado, en comparación de los segmentos TRBD en los cuales se conocen dos familias TRBD1-2 y los segmentos de unión (J) en los que en la cadena alpha se reconocen hasta 64 familias y 2 en la cadena beta, los segmentos V parecen ser los principales aportadores en la generación de diversidad en el TCR (Lefranc, 2001). Sin embargo, a pesar de que el contexto determina el requerimiento del repertorio, como se explicó con anterioridad, se mantiene la idea de que la reducción del repertorio es una señal de presión patogénica menor en una especie.

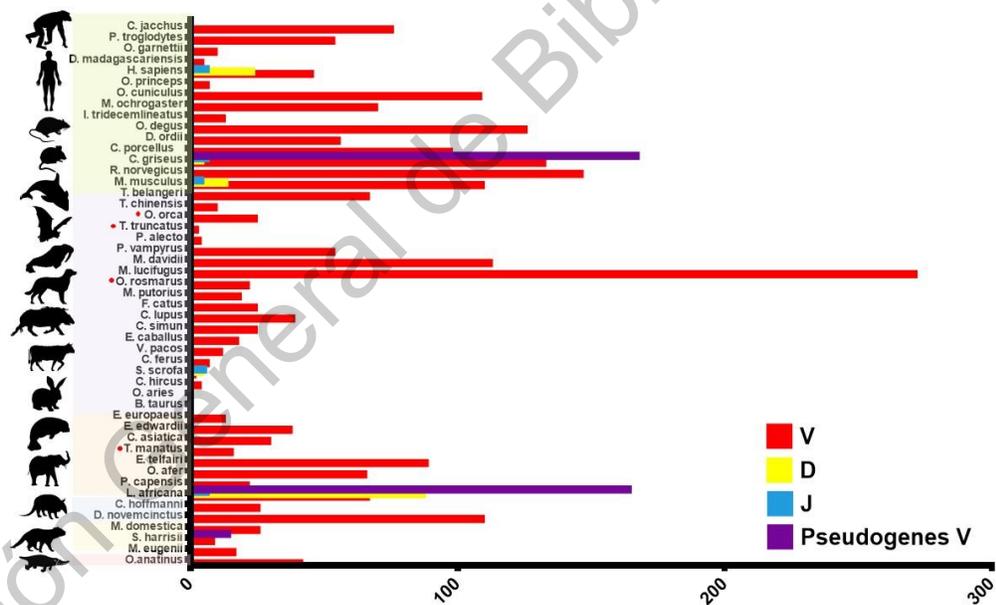


Figura 9| Repertorio de genes V (D) J y pseudogenes V de la cadena Heavy en los mamíferos. En rojo, se muestra el repertorio de los genes V. En amarillo, se muestra el repertorio de genes D y en azul, se muestra el repertorio de genes J. Los pseudogenes V, se muestran en color morado. Los puntos rojos indican a los mamíferos marinos. Imagen adaptada de: Guo *et al.*, 2011; Qin *et al.*, 2014; Olivieri *et al.*, 2014; Qin *et al.*, 2015; Ujvari & Belov, 2015 y Murphy *et al.*, 2017.

El repertorio de las cadenas del BCR (Ig) al igual que las del TCR tiende a expandirse, por lo tanto, también varía entre taxones (Figura 9-11). Las cadenas del BCR, al no estar restringidas por el MHC directamente, exhiben repertorios más amplios en todas las especies; sin embargo, la diversificación y expansión de las familias que forman al receptor, también es resultado de su evolución, debido a que se hipotetiza que este receptor surgió antes que el TCR (Boehm & Bleul, 2007).

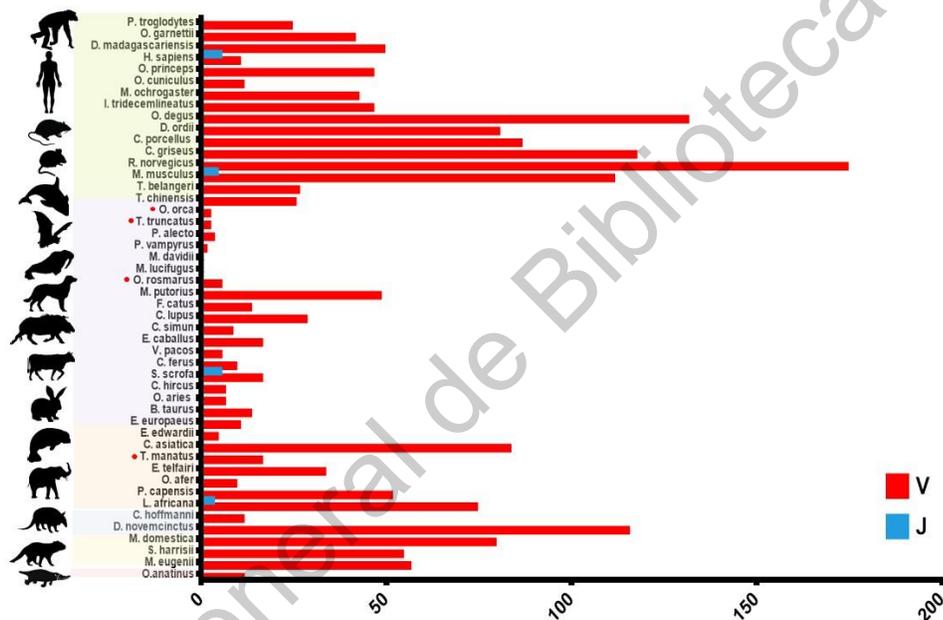


Figura 10| Repertorio de genes V y J de la cadena kappa en los mamíferos. En rojo, se muestra el repertorio de los genes V. En azul, se muestra el repertorio de genes J. Los puntos rojos indican a los mamíferos marinos. Imagen adaptada de: Guo *et al.*, 2011; Qin *et al.*, 2014; Olivieri *et al.*, 2014; Qin *et al.*, 2015; Ujvari & Belov, 2015 y Murphy *et al.*, 2017

Las familias de las cuales se diversifican los segmentos V en mamíferos difiere entre las órdenes y estas se extienden desde 1-16 IGHV en algunos roedores, 1-7 IGHV en humanos (Pallares *et al.*, 1999), IGHV1-4 en algunos carnívoros (Olivieri *et al.*, 2014) e IGHV1-2 en ungulados. En contraste, las familias que agrupan a los segmentos V de las cadenas ligeras van desde IGKV1-4 a IGKV1-16 en la cadena

kappa y en la cadena lambda 1-9IGLV en humanos, IGLV1-8 en algunos roedores e IGLV1-3 en ganado. Por otra parte, las familias de los segmentos de diversidad se agrupan en general en las familias IGHD1-8 e IGHDQ52, en tanto que para los de unión (J) se conocen las familias IGHJ1-4, IGKJ1-5 e IGLJ1-3 (Kaushik & Pasman, 2014).

En ambos receptores, el proceso de origen y muerte diversifica las familias de las cadenas de los receptores, el cual genera un sesgo en la expansión de un locus o lo deshabilita, este fenómeno se ha observado principalmente en las cadenas ligeras del BCR (Ig). Mientras que, en el mismo receptor en otras especies la expansión de una cadena ligera, se ha asociado con la compensación en la disminución del repertorio de las cadenas del TCR y viceversa (Olivieri *et al.*, 2014). Lo anterior refuerza la idea de que el proceso de origen y muerte, está regulado por el contexto ecológico de una especie (Nei & Sitnikova, 1997). Sin embargo, que el incremento del repertorio en los receptores TCR y BCR esta únicamente relacionado con la exposición patogénica en determinado hábitat es dual, debido a que el reconocimiento a antígenos no propios, solo es uno de los mecanismos que los receptores regulan (Nikolich-Zugich *et al.*, 2004). Además, se sabe que el uso del repertorio está sujeto a la proximidad de los segmentos a los promotores y en este sentido tener un mayor número de segmentos no implica que sean utilizados al montar una respuesta inmune (Proudhon *et al.*, 2015). Por lo tanto, establecer la competitividad inmunológica basada únicamente en el número de segmentos génicos V(D)J, es una aproximación deficiente.

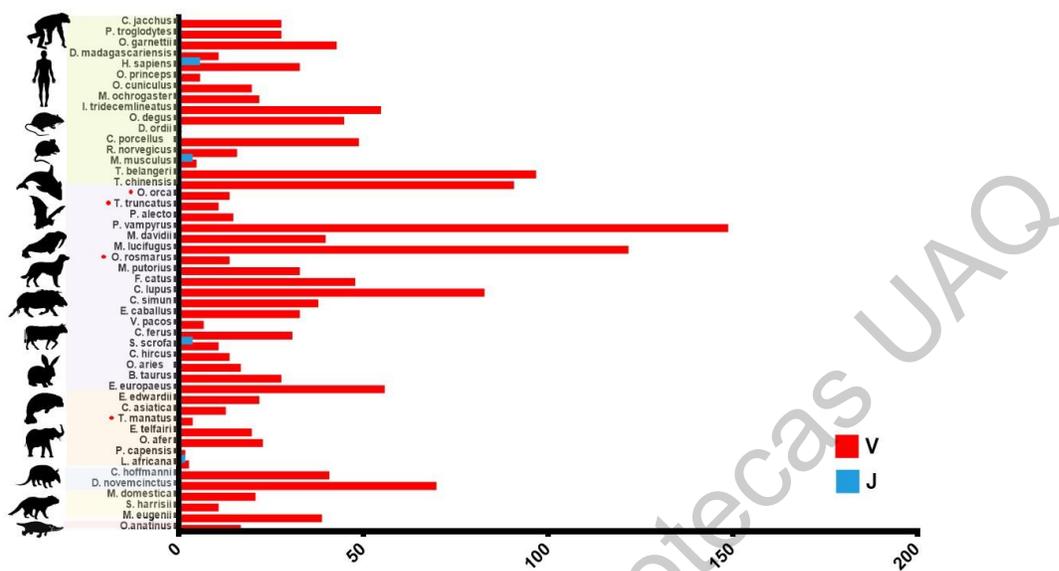


Figura 11| Repertorio de genes V y J de la cadena lambda en los mamíferos. En rojo, se muestra el repertorio de los genes V. En azul, se muestra el repertorio de genes J. Los puntos rojos indican a los mamíferos marinos. Imagen adaptada de: Guo *et al.*, 2011; Qin *et al.*, 2014; Olivieri *et al.*, 2014; Qin *et al.*, 2015; Ujvari & Belov, 2015 y Murphy *et al.*, 2017.

1.4 Evolución de los segmentos V(D) en mamíferos

Los segmentos V(D)J como se mencionó con anterioridad, son parte de la super familia de Inmunoglobulinas. Por si mismos, los miembros de esta familia son capaces de adaptar sus pliegues para cumplir diferentes funciones por lo normal asociadas al reconocimiento antigénico (Dermody *et al.*, 2009). Sin embargo, es poco probable que esta familia haya tenido un papel inmune en un inicio, y se propone que hasta que ocurrieron eventos de duplicación y la inserción de un trasposon de la familia Transib (Kapitonov *et al.*, 2005) entre los segmentos proto-V y proto-J generó la recombinación de los sitios de reconocimiento y posteriormente emergió su función antigénica (Flajnik & Kasahara, 2010). Los eventos de duplicación subsecuentes dieron origen al proto-D (Proudhon *et al.*, 2015). Adicionalmente, estos segmentos adquirieron distintas características ontogénicas propias de las cadenas en las que residían (Boehm, 2006).

Este sistema rompe los paradigmas genéticos de la selección Darwiniana, por lo que en cada individuo se establecen configuraciones distintas de estos genes. De manera que su aportación individual entre mamíferos es distinta. Además, a través de los mamíferos estos genes están regulados por el mecanismo de origen y muerte, que permite desechar genes que no son adaptables a una función específica o representan desventaja en el hábitat en donde se desarrolla un organismo (Nei & Sitnikova, 1997).

1.4.1 Limitadores en la plasticidad del repertorio de los receptores TCR y BCR

La aportación del repertorio entre las especies de mamíferos es distinta debido a su hábitat. Por lo tanto, la expansión y el ensanchamiento de los loci que forman a los receptores, está limitado por el repertorio V(D)J que durante la historia de vida un individuo ha generado. En este sentido los factores que pueden limitar el repertorio de una especie pueden ser en consecuencia de dos procesos, la generación de resistencia y tolerancia (Nikolich-Zugich *et al.*, 2004).

Cuando un organismo se desenvuelve en un hábitat que no representa estrés patogénico, segmentos de sus cadenas se pierden, volviéndolo incapaz a reconocer antígenos en vida libre (Nanda *et al.*, 1991). Lo anterior apoya la idea de que ciertos hábitats pueden generar reducción del repertorio, por ser menos inmunopatogénicos (Olivieri *et al.*, 2014). Sin embargo, esta idea genera dos vertientes que no son excluyentes una de la otra.

El repertorio de genes V(D)J, como se mencionó arriba evoluciona en cada individuo y como toda función fisiológica, la respuesta inmune adaptativa requiere de la alocación de recursos, los cuales están en competencia con otras funciones fisiológicas. Aunque en términos energéticos generales montar la respuesta

adaptativa es menos costoso que montar la respuesta innata, existen procesos post la activación de la respuesta adaptativa que son costosos (Demas & Nelson, 2012).

En general se acepta la idea de la heterosis, en la que ser más diverso genéticamente es más efectivo para el sistema inmune de las especies, al permitir reconocer una mayor cantidad de antígenos. Sin embargo, tener un mayor número de segmentos V(D)J conlleva una amplia variedad de receptores, que necesitan regulación. En el caso del BCR la diversidad serológica está directamente asociada a la coevolución con patógenos (Mantis *et al.*, 2011). Por otro lado, la diversidad del TCR, además de estar asociada a la diversidad patogénica como se mencionó antes está restringida por el MHC (Abbas *et al.*, 2017). En términos energéticos, la restricción por MHC necesita la asignación de recursos para la eliminación de células T autorreactivas, que a su vez están en competencia con los procesos inmunopatógenos, por lo cual aumentar la diversidad del MHC promueve la eliminación de células T y por lo tanto la diversidad del repertorio de los genes que forman al receptor (Demas & Nelson, 2012).

Las especies necesitan establecer niveles de competencia frente a sus depredadores. Los patógenos como depredadores establecen niveles de oscilación del uso de genes, por lo que los hospederos deben adaptarse y generar alelos, a los que durante su historia de vida los patógenos no hayan generado resistencia. Este proceso es costoso, por lo que establecer procesos de tolerancia entre huésped y hospedero resulta más eficiente para ambos (Demas & Nelson, 2012). Desde el punto de vista del repertorio de los TCR y BCR, establecer niveles de tolerancia antes que resistencia puede llevar la reducción del repertorio también, lo cual descarta la posibilidad de generar receptores autorreactivos (Collins & Watson, 2018).

1.5 El lobo marino de California (*Z. californianus*)

Los mamíferos marinos evolucionaron a partir de diferentes mamíferos terrestres, los cuales regresaron al hábitat acuático debido a una falta de alimento en el hábitat terrestre. Existen cinco órdenes taxonómicos que contienen mamíferos marinos: los cetáceos y los sirenios que son completamente acuáticos, y dentro del orden carnívora los osos polares, las nutrias y los pinnípedos (Uhen, 2007). Los tiempos de evolución de estos grupos varían, siendo más antiguos para los cetáceos, y más recientes para los osos polares. Específicamente los pinnípedos regresaron hace 28 millones de años al medio ambiente marino (Uhen *et al.*, 2007).

El lobo marino de California (*Zalophus californianus*) es un pinnípedo de la familia Otariidae. *Zalophus californianus*, es una especie sexualmente dimórfica, en la que los machos son más grandes que las hembras y presentan una cresta sagital muy pronunciada. La estructura anatómica de esta especie y de los otros miembros de su familia le permite tener locomoción terrestre y acuática (Berta *et al.*, 2015) (Figura 12).

El hábitat de *Z. californianus* se extiende en las costas del Pacífico oriental, desde Vancouver, Canadá, hasta el noreste de la península de Baja California, México, en donde se localizan en ambas costas de la Península de Baja California (Figura 13) (Szteren, 2011).



Figura 12| Dimorfismo sexual entre hembras y machos de *Z. californianus*. Del lado derecho se muestra a la hembra y de lado izquierdo al macho. Imagen tomada de: What We Thought: Mating system of the California Sea Lion. (2019).

El genoma del *Z. californianus* no se conoce por completo, pero por la cantidad de regiones codificantes se estima que la longitud de su genoma es similar a la de otros mamíferos incluyendo al humano (Chinta, 2013) y perro (*Canis lupus familiaris*) (Arnason *et al.*, 2006). Sin embargo, el nivel de compactación de su genoma es diferente, quizá por el porcentaje de regiones espaciadoras en el genoma de este pinnípedo lo que da a la especie 36 cromosomas (Riedman, 1990). Sin embargo, se ha propuesto el uso del genoma del perro doméstico, (*Canis lupus familiaris*), para investigar a especies de carnívoros silvestres como *Z. californianus*, ya que el perro es la divergencia más temprana en la superfamilia Canoidae que incluye a los pinnípedos (Breen *et al.*, 2004; Ostrander & Wayne, 2005).

Se han caracterizado 379 regiones codificantes en el genoma de *Z. californianus*, las cuales son similares a las de otros carnívoros, 42 al panda gigante, *Ailuropoda melanoleuca*, 22 a *Canis lupus familiaris* y 30 al gato doméstico, *Felis catus* (Chinta, 2013).

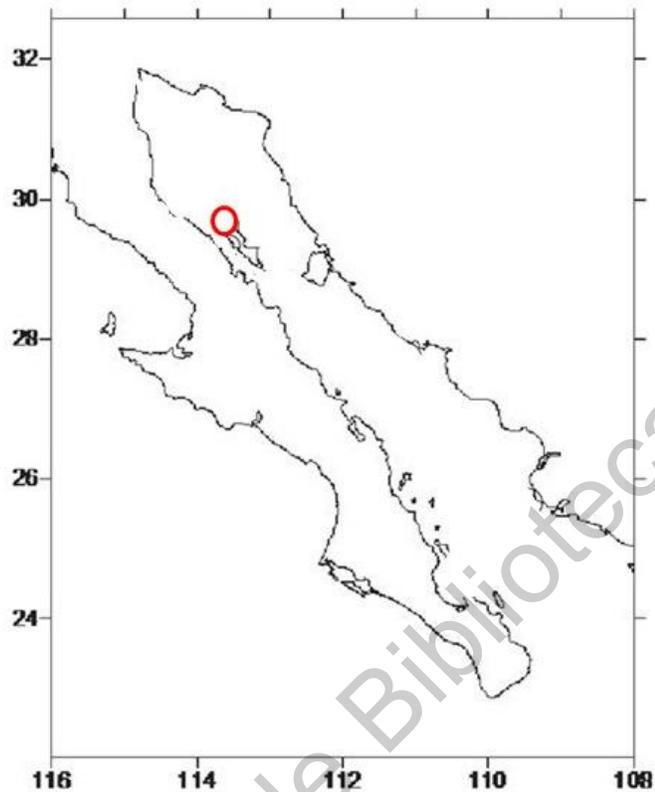


Figura 13| Zona de la toma de muestra de *Z. californianus* en este estudio. El mapa muestra la zona de recolección de las muestras de este trabajo en el círculo rojo, Isla granito en donde originalmente se caracterizaron todos los alelos del MHC de *Z. californianus*. Imagen tomada de: Szteren *et al.*, 2006.

1.5.1 Inmunología del lobo marino de California (*Z. californianus*)

Durante la transición del hábitat terrestre al semiacuático, la respuesta inmune de los pinnípedos, estuvo sometida a diversas presiones patogénicas tanto del medio terrestre como del acuático (Olivieri *et al.*, 2014). Por tanto, el estudio de la respuesta inmunológica de los pinnípedos resulta interesante en términos comparativos con especies terrestres. Sin embargo, a la fecha, es relativamente poco lo que se conoce sobre la inmunología de *Z. californianus*.

En términos anatómicos, los órganos linfoides primarios y secundarios de *Z. californianus* son iguales a los de otros mamíferos marinos y terrestres. Sin embargo, sus requerimientos son distintos debido a que están sometidos a distintos niveles de presión durante su buceo, lo que deriva en la oscilación constante de niveles de oxígeno en el torrente sanguíneo, lo cual sugiere que la fisiología inmune puede ser distinta (Ponganis *et al.*, 1992).

En cuanto al sistema inmune de *Z. californianus* se han determinado muchos de los genes que moldean la respuesta inmunológica de los mismos, y también se ha investigado la relación de varios polimorfismos genéticos con la progresión de enfermedades infecciosas y carcinoma (e.g. Acevedo-Whitehouse *et al.*, 2003; Bowen *et al.*, 2005; Browning *et al.*, 2014).

Una de las regiones genéticas con relevancia inmune que se han estudiado más en esta especie es el complejo mayor de histocompatibilidad clase II (ZacaDRB) lo que ha proporcionado conocimiento sobre la respuesta inmune adaptativa de este pinnípedo, debido a la alta variabilidad del MHC clase II en esta especie y su mecanismo atípico para generar diversidad (Bowen *et al.*, 2004). Variantes genéticas como el ZacaDRB.A se han asociado con el desarrollo con carcinoma genitourinario (Bowen *et al.*, 2005), así como con la disminución en la respuesta inflamatoria inespecífica (Montano-Frías *et al.*, 2016). También se ha reportado que, contrario a lo que predice la teoría evolutiva de la diversidad de los alelos del MHC, la alta variabilidad en ZacaDRB, tiene como consecuencia respuestas exacerbadas a infecciones por la espiroqueta *Leptospira interrogans*, lo que resulta en inmunopatología (Acevedo-Whitehouse *et al.*, 2018).

1.6 Técnicas genómicas de nueva generación en inmunología

El estudio de los receptores antigénicos de *Z. californianus* se ha enfocado principalmente en la cuantificación de la respuesta humoral durante su desarrollo (Espinosa-de Aquino *et al.*, 2017), y bajo condiciones de estrés ambiental (Banuet-Martinez *et al.*, 2017). Sin embargo, la estructura y organización génica del receptor BCR en membrana y del TCR, permanecen aún sin explorar.

La mayoría de las moléculas que orquestan las funciones inmunes de diversos organismos se descubrieron debido al alcance de técnicas como el radioinmunoanálisis, la inmunoprecipitación y la hibridación *in situ*. Sin embargo, a medida que el entendimiento de la respuesta inmune avanzó, las técnicas proteicas y moleculares mostraron ser más prometedoras, por lo tanto se comenzaron a utilizar para ampliar nuestro entendimiento sobre el estudio del sistema inmune (Murphy & Weaver, 2017).

Las herramientas genómicas han contribuido al progreso en general de la ciencia y por consiguiente al estudio sistema inmune (Mori *et al.*, 2013). En general las características genómicas del sistema inmune son de gran relevancia, porque cambios en la secuencia de los genes derivan en variaciones fenotípicas que impactan el reconocimiento antigénico (Schlötterer *et al.*, 2016).

Actualmente, las técnicas genómicas de nueva generación permiten desde el estudio de la expresión génica de una célula individual hasta el estudio de dominios de asociación topológica en una célula única (TADs) (Proudhon *et al.*, 2015). Las técnicas de secuenciación masiva pueden reflejar todo aquello que se transcribe en un sistema en determinado momento y bajo ciertas condiciones o la regulación transcripcional bajo diferentes condiciones (Mori *et al.*, 2013). Por otra parte, la era de la genómica, le ha permitido a la inmunología, conocer el repertorio de genes involucrado, en el desarrollo, ecología y evolución de la respuesta inmunológica

(Longo *et al.*, 2014). Sin embargo, existen otras técnicas como la rápida amplificación de terminación de cDNA (RACE), que permite conocer la organización molecular de distintos genes, incluso en especies de las que no se tiene anotación genómica (Yeku & Frohman, 2011) y con base a esto, también permite la comparación filogenética de las características de un gen (Guo *et al.*, 2011).

1.7 La caracterización de los receptores TCR y BCR en especies no modelo mediante técnicas Bioinformáticas

Los receptores de los linfocitos T y B son los principales orquestadores de la respuesta adaptativa de vertebrados. Durante su desarrollo su configuración genética cambia, lo cual les confiere especificidad clonal y a su vez genera memoria inmunológica, a través de la cual promueven la activación de efectores innatos (Abbas *et al.*, 2017). Aunque los receptores TCR y BCR (estos últimos llamados Ig) no sólo reconocen antígenos foráneos sino también propios, el estudio de su diversidad está directamente asociado a la diversidad antigénica a la que un organismo se ha enfrentado. Por otra parte, al ser contexto dependiente (Wong *et al.*, 2016; Holder *et al.*, 2018) la diversidad antigénica varía de individuo a individuo.

El repertorio de los receptores antigénicos es un reflejo directo del estado inmunológico de un individuo. Sin embargo, debido a la alta variabilidad de los segmentos génicos que los conforman, su estudio se ha centrado en las zonas donde los genes V, D y J se unen; la zona de complementariedad CDR3, principalmente de las cadenas pesadas (Rosati *et al.*, 2017). Previamente se había estudiado y caracterizado el repertorio V(D)J en especies modelo para las que el genoma está anotado, como el ratón y el humano, debido a que se conocían las secuencias de los sitios de recombinación (RSS; por sus siglas en inglés), y por lo tanto se podía estudiar la funcionalidad de los transcritos después del reordenamiento somático (Murphy & Weaver, 2017). Sin embargo, también se estudió la evolución de las multifamilias de segmentos V(D)J en estas especies (Nei

& Sitnikova, 1997; Balakirev & Ayala, 2003), lo que permitió determinar cuáles eran las regiones conservadas en los receptores para ambas especies. A partir de estas regiones se definieron las reglas para determinar el repertorio en otras especies (Giudicelli *et al.*, 1999; Giudicelli *et al.*, 2005; Madi *et al.*, 2017; Migalska *et al.*, 2018).

El estudio de los TCRs y BCRs ha cambiado con el tiempo, y el desarrollo de nuevas herramientas moleculares y de análisis bioinformático, ahora permite la caracterización de estos receptores en otras especies, así como el modificar las bases de datos existentes para establecer reglas de repertorio y análisis de ontogenia más robustos (Lefranc *et al.*, 2014). Por otro lado, otros programas fueron propuestos para adaptar los análisis de ontogenia actuales a otras especies, sin generar un sesgo hacia los repertorios existentes (Olivieri *et al.*, 2013; Hou *et al.*, 2016; Rosati *et al.*, 2017; Migalska *et al.*, 2018).

Por otra parte, aunque algunas de las técnicas moleculares son estándares, como el 5' RACE, la secuenciación de nueva generación ha permitido nuevas aproximaciones para el estudio del repertorio, como la PCR Multiplex y el enriquecimiento de flancos (Rosati *et al.*, 2017). Normalmente estas técnicas permiten caracterizar los receptores de especies no anotadas, , sin embargo, el diseño experimental por lo general está basado en las regiones menos variables de los receptores; las regiones constantes, las cuales también mantienen sitios conservados con otros miembros de la super familia de las inmunoglobulinas (Marchler-Bauer *et al.*, 2016), lo cual en general puede generar desajustes al analizar los datos, generando falsos positivos en las lecturas caracterizadas, particularmente en especies que carecen de anotación.

En el presente capítulo se presenta el diseño experimental que permitió la caracterización de las cadenas de los receptores TCR y BCR en especies sin genoma de referencia. Mediante la técnica 5' RACE se optimizó el número de las lecturas obtenidas por secuenciación multiplex, y se evitaron desajustes en el análisis de los datos, mediante una clusterización previa al análisis de las lecturas con las de otras especies.

Justificación

El sistema inmunológico ha evolucionado en distintos organismos para poder cubrir los requerimientos patogénicos y de tolerancia en cada hábitat, por lo cual se considera un carácter de historia de vida (Demas & Nelson, 2012). La evolución del sistema de sofisticación en los TCR y BCR ha permanecido como el *Big Bang* inmunológico, no sólo por la presencia de sitios hipervariables en estos genes que permiten un polimorfismo sin precedentes en el resto del genoma, sino además porque su ontogenia sigue rigurosos controles de calidad, mediados en un principio por la historia evolutiva del hospedero, coevolución con patógenos en el medio, que derivaron en canjes evolutivos de auto-tolerancia y resistencia en cada vertebrado (Guo *et al.*, 2011; Qin *et al.*, 2014; Qin *et al.*, 2015; Ujvari & Belov, 2015; Murphy & Weaver, 2017).

La expansión de ciertos locus de los receptores TCR y BCR, sigue la tendencia hacia el reconocimiento de epítomos específico, en donde la expansión en los loci del BCR representa una inclinación hacia las respuestas humorales, mientras que la expansión de los loci del TCR indica una inclinación hacia las respuestas celulares (Olivieri *et al.*, 2014). Es por ésto que estudiar la evolución de los receptores linfocitarios puede proveer de conocimiento sobre la selección positiva de genes V, D, J e, incluso, C, y su relación con la susceptibilidad a patógenos y la progresión de enfermedades (Proudhon *et al.*, 2015). La alta diversidad de estos receptores en los vertebrados, ha llevado al estudio de su asociación en perfiles de patologías. Sin embargo, a la fecha nuestro entendimiento sobre la evolución del TCR y BCR es limitado debido a que está centrado en humanos y modelos animales experimentales, lo cual no permite tomar en cuenta ni el contexto evolutivo ni el contexto ecológico que dieron lugar a los receptores en diferentes grupos taxonómicos (Boehm *et al.*, 2006).

En el contexto del cambio ambiental global, y el incremento en el número de enfermedades infecciosas (Jones *et al.*, 2008), el estudio de los receptores linfocitarios en organismos de vida libre es cada vez más relevante, sobre todo en aquellos que por sí solos son un modelo inmunológico, como *Z. californianus*, que recientemente se ha propuesto como modelo de estudio del desarrollo de carcinogénesis (Browning *et al.*, 2015). Además de su alta tolerancia a parásitos intestinales (Acevedo-Whitehouse *et al.*, 2006) y su susceptibilidad genética a *Leptospira interrogans* (Acevedo-Whitehouse *et al.*, 2018) y a otros patógenos (e.g. Lyons *et al.*, 1997; King *et al.*, 2002; Johnson *et al.*, 2006; Lloyd-Smith *et al.*, 2007; Ng *et al.*, 2009; Rivera *et al.*, 2012; Kuzmina *et al.*, 2018). Lo anterior podría tener relevancia biomédica, debido a que los receptores TCR y BCR son efectores directos o reguladores de segundos mensajeros inmunes durante la transformación oncogénica y la exposición a helmintos y bacterias (Murphy *et al.*, 2017).

Estudiar los receptores antigénicos de los linfocitos de *Z. californianus* es importante desde una perspectiva ecológica y evolutiva, debido a que en el caso del TCR, éste co-evolucionará con el MHC II (Castro *et al.*, 2015), el cual, en *Z. californianus*, es altamente polimórfico, con respecto a otros mamíferos, y tiene un mecanismo muy poco común para generar dicha diversidad (Bowen *et al.*, 2004). Por este motivo es que tenemos razones para creer que surgieron eventos de duplicación en estos receptores, proporcionales a la expansión del MHC II, como se ha reportado para otros mamíferos (Olivieri *et al.*, 2014; Li *et al.*, 2018).

Preguntas de investigación

¿Cuáles segmentos génicos V(D)J en los receptores TCR y BCR de *Zalophus californianus* son únicos de esta especie y cuáles segmentos comparte con otros mamíferos carnívoros?

¿Cómo es la estructura tridimensional de los receptores TCR y cuál es su relación estructural con las moléculas del MHC de *Zalophus californianus*?

Dirección General de Bibliotecas UAQ

III. Objetivos

Objetivo general

Caracterizar genéticamente los receptores TCR y BCR de *Zalophus californianus*.

Objetivos específicos

1. Describir la composición genética de los receptores TCR y BCR de *Zalophus californianus*.
2. Comparar los segmentos génicos V(D)J del TCR y BCR de *Zalophus californianus* con los receptores de otros mamíferos carnívoros.
3. Definir la estructura tridimensional putativa de los receptores TCR y BCR de *Zalophus californianus*.

IV. Metodología

2.1 Muestras utilizadas

Las dos muestras que se utilizaron fueron obtenidas en el 2016 durante un muestreo realizado como parte de un proyecto de investigación “Patrones espacio-temporales de la diversidad de receptores antigénicos y efectores inmunes de *Z. californianus* una aproximación eco-inmunogenómica” (Proyecto 446 del programa Fronteras de la Ciencia). Estas muestras consistían en capa leucocitaria colectada después de la centrifugación (3200 rpm por 10 minutos) de 8 ml de sangre periférica preservada con heparina de sodio que había sido colectada de la vena glútea caudal de crías de *Z. californianus* nacidas en la Isla Granito, en el Golfo de California, México. Todas las muestras fueran colectadas bajo los permisos de SGPA/DGVS/09004/15 y de acuerdo al visto bueno del Comité de Bioética de la Facultad de Ciencias Naturales de la Universidad Autónoma de Querétaro. El trabajo de colecta de muestras fue realizado por miembros del Laboratorio de Plasticidad Inmune y Ecoepidemiología Molecular, y las muestras forman parte del banco de muestras de ese laboratorio.

2.2 Extracción de RNA de las muestras

Para cada una de las dos muestras, la capa leucocitaria, almacenada en un criotubo con RNAlater (76104; QIAGEN, EEUUA) y guardada en criocongelación hasta su análisis, fue centrifugada para obtener el pellet de leucocitos. El pellet fue resuspendido en TRIzol Reagent y lisado con una jeringa de insulina para después obtener la fase acuosa mediante cloroformo en una proporción de 500µl/1ml de Trizol. La muestra se precipitó con isopropanol, se lavó con etanol al 70% y se eluyó en 20µl de H₂O DEPC (10977035; Invitrogen, EEUUA). La integridad del RNA (RIN, por sus siglas en inglés) se determinó mediante espectrometría en un Nanodrop (ND-2000; ThermoFisher, EEUUA) y mediante visualización directa en un gel de

agarosa al 1.2% (Figura A2; Material Anexo) de acuerdo con protocolos publicados (Pederson, 2015).

2.3 Retrotranscripción de cDNA de muestras de RNA de Z. californianus

Se utilizó 1µg de RNA total para realizar la retrotranscripción. Para ésto, fueron adicionados 2µl de gDNAWipeout Buffer 7x (205310; QIAGEN, EEUUA) y se llevó con H₂O DEPC (10977035; Invitrogen, EEUUA) a un volumen final de 14 µl antes de incubar la reacción por 2 min a 42°C.

Posteriormente, fueron agregados a la reacción 1µl de Quantiscript Reverse Transcriptase (205310; QIAGEN, EEUUA), 4 µl de Quantiscript RT Buffer 5x (205310; QIAGEN, EEUUA), 1µl de RT Primer Mix (205310; QIAGEN, EEUUA) e incubamos los 20 µl de la reacción durante 15 min a 42°C. Finalmente inactivamos la transcriptasa a 95°C durante 3 min. Todo se realizó conforme a las instrucciones del producto (205310; QIAGEN, EEUUA).

Para determinar la cantidad de cDNA que se utilizaría en cada reacción de PCR de punto final, fue determinada la cantidad y calidad del ADNc en el Nanodrop (ND-2000; ThermoFisher, EEUUA), y se generaron alícuotas para usar 1µg/µl en cada reacción.

2.4 Diseño racional de oligonucleótidos cebadores para especies no modelo

Se realizó una búsqueda en la base de secuencias de GenBank (<https://www.ncbi.nlm.nih.gov/>) reportadas para TCR-alpha, TCH-beta, IGH e IGL de mamíferos carnívoros, para realizar el diseño de oligonucleótidos cebadores de NCBI. Se eligieron las secuencias de las regiones constantes de cada cadena de

Canis lupus familiaris TCR-alpha (Accession code 607937), TCR-beta (480788), IGH (102154760), IGL (607368).

Posteriormente fueron seleccionadas las coordenadas de las secuencias codificantes más grandes para buscar los sitios más conservados entre los mamíferos disponibles utilizando el programa de la Universidad de Santa Cruz de California (<https://genome.ucsc.edu/cgi-bin/hgGateway>). Se utilizó el algoritmo blastz y encadenamiento de redes (Schwartz, 2003). Las alineaciones de pares se realizaron de forma múltiple utilizando el programa var_multiz (Blanchette, 2004) de acuerdo con la topología: ((canFam2 hg17) (mm6 rn3)). La determinación del grado de conservación entre especies se asignó con phastCons, basado en el modelo probabilístico de phylo-HMM (Siepel & Haussler, 2005; Siepel *et al.*, 2005), el cual describe el proceso de sustitución de DNA en cada sitio en un genoma como la forma en que este proceso cambia de un sitio a otro (Figura 14).

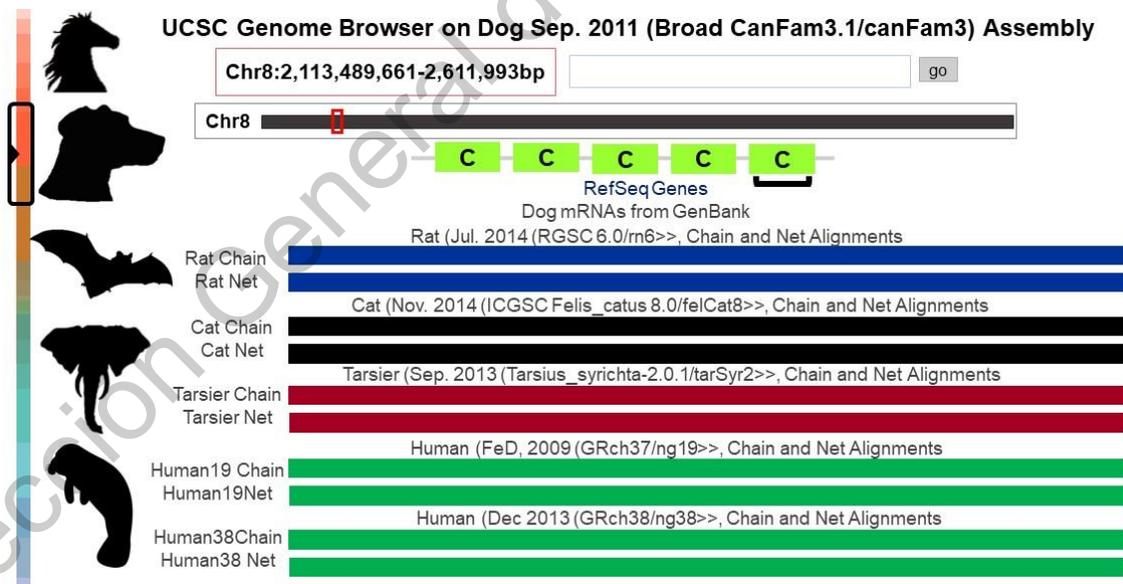


Figura 14| Esquemización de la selección de las regiones conservadas para el diseño de oligonucleótidos cebadores. La imagen ejemplifica la selección de las zonas para el diseño de los oligonucleótidos, basados en las regiones contantes más conservadas de cada cadena del genoma de *Canis lupus familiaris* 3.1.

Se seleccionaron las coordenadas de los sitios más conservados para cada cadena entre especies, y con base en ello se diseñaron los oligonucleótidos cebadores Forward y Reverse (Tabla A1; Material Anexo) con la herramienta Primer-Blast de NCBI (Untergasser *et al.*, 2012). Finalmente fueron seleccionados aquellos oligonucleótidos que amplificaran secuencias específicas y que eran complementarios a las secuencias de más de una especie de mamífero carnívoro.

2.5 Estandarización de la amplificación por PCR de punto final en cDNA de Z. californianus

Se estandarizó la PCR de punto final en las muestras de ADNc. La temperatura de fusión (T_m , por sus siglas en inglés) óptima para cada par de oligonucleótidos cebadores diseñados fue determinada por PCR de punto final de gradiente (Tabla A1; Material Anexo). Las reacciones contenían a 1 µg/µl de cDNA, 1X buffer ADN polimerasa (10342053; Invitrogen, EEUUA), MgCl₂ (50mM; 10342053; Invitrogen, EEUUA), dNTPs 10mM (10 mM dATP, 10 mM dGTP, 10 mM dCTP, 10 mM dTTP; 10342053; Invitrogen, EEUUA), cebadores forward y reverse (10 µM; sintetizados por OligoT4, México), (Tabla A1; Material Anexo), H₂O bidestilada (10977035; Invitrogen, EEUUA) y ADN Taq Polimerasa (5 U/µl; 10342053; Invitrogen, EEUUA). Los productos finales fueron analizados en un gel de agarosa al 2%.

Mediante PCR de gradiente se amplificaron fragmentos de las secuencias constantes para las cadenas alpha (Figura A3; Material Anexo), beta (Figura A4; Material Anexo), heavy (Figura A4; Material Anexo) y lambda (Figura A5; Material Anexo). Las bandas de pesos esperados y únicas se secuenciaron y se caracterizaron por alineamiento. De esta manera se identificaron las temperaturas óptimas de hibridación de los cebadores diseñados (Tabla A1; Material Anexo).

Los cebadores seleccionados para usarse como cebadores específicos de la cadena alpha fueron los Alpha-A y Alpha B. El par de cebadores Alpha-A, para realizar la PCR touchdown y el par de cebadores alpha-B para realizar la PCR anidada (Figura A4; Material Anexo).

Para la cadena beta los pares de cebadores seleccionados fueron, Beta-B para RACE y Beta-C para PCR anidado (Figura A4; Material Anexo). En el caso de la cadena heavy se seleccionaron los cebadores Heavy-A para el protocolo de RACE y los cebadores Heavy-B, Heavy-C y Heavy-D para realizar la PCR anidada (Figura A4; Material Anexo).

En la cadena lambda se seleccionó el par de cebadores Lambda-A para realizar el protocolo RACE, mientras que Lambda-B y Lambda-C para ser cebadores específicos en la PCR anidada (Figura A5; Material Anexo).

2.6 Preparación del ARN para la rápida amplificación del extremo 5'/3' del cDNA

2.6.1 Desfosforilación del ARN para la rápida amplificación de los extremo 5'/3' del cDNA

El ARN total usado fue obtenido con el protocolo descrito en el punto 2.2 Desfosforilamos 2 µg de ARN adicionándole 1 µl de Fosfatasa Alcalina de intestino de ternero (CIP 10 U/µl; L1500-01; Invitrogen, EEUUA), 1µl de 10X Buffer CIP (L1500-01; Invitrogen, EEUUA), con H₂O DEPC se llevó a 9 µl y se protegió la reacción de la acción de ARNasas con 1µl RNaseOut (40 U/µl; L1500-01; Invitrogen, EEUUA). La reacción se incubó a 50°C durante 1 hora en baño María (Figura 15).

2.6.2 Precipitación del ARN para la rápida amplificación de los extremo 5'/3' del cDNA

Para precipitar el ARN posterior a la desfosforilación se agregaron 90 μ l de H₂O DEPC y 100 μ l de Phenol:Chloroform:Alcohol-Isoamílico (25:24:1) 8-hidroxiquinoleína 0.1% y se obtuvo la fase acuosa después de centrifugar a temperatura ambiente durante 5 min a máxima velocidad. La fase acuosa se trasladó a un microtubo de 1.5, se le adicionaron 2 μ l de glicógeno de mejillón (10 mg/ml), 10 μ l de acetato de sodio (3M; pH 5.2) y etanol al 95% y se obtuvo el pellet de ARN al centrifugar a máxima velocidad por 20 min a 4°C. Por último, el pellet de ARN se limpió con etanol al 70%, se centrifugó a máxima velocidad durante 5 min a 4°C y se dejó secar a temperatura ambiente durante 2 min para finalmente ser eluido en 7 μ l de H₂O DEPC, (Figura 15). Todos los reactivos utilizados fueron del kit comercial Gene Racer (L1500-01; Invitrogen, EEUUA).

2.6.3 Decapitación del ARN para la rápida amplificación de los extremo 5'/3' del cDNA

Al ARN obtenido después de la desfosforilación y precipitación, se le descapuchó con 1 μ l de pirofosfatasa ácida de tabaco (TAP, 0.5 U/ μ l), 1 μ l de buffer 10X TAP), y se protegió la reacción de las ARNasas con 1 μ l de RNaseOut (40 U/ μ l). La reacción se incubó a 37°C durante 1 h en baño María. El ARN descapuchado se precipitó siguiendo el protocolo indicado previamente en el punto 2.6.2 (Figura 15). Todos los reactivos utilizados fueron del kit comercial Gene Racer (L1500-01; Invitrogen, EEUUA).

2.6.4 Ligación de Oligo 5' al ARN para la rápida amplificación del extremo 5'/3' del cDNA

Se adicionó el ARN descapuchado a un tubo con cebador de ARN liofilizado [5'-CGACUGGAGCACGAGGACACUGACAUGGACUGAAGGAGUAGAAA-3'], 0.25µg y la reacción fue homogenizada mediante pipeteo antes de incubar a 65°C durante 5 min en baño María para prevenir la formación de estructuras secundarias. Posterior a la incubación adicionamos 1µl buffer de ligasa 10X, 1 µl de ATP 10 mM, 1 µl de RNA ligasa T4 (5 U/µl) y se protegió la reacción de RNAsas con 1 µl RNaseOut (40 U/µl). La reacción se incubó a 37°C durante 1 h en baño María. Finalmente, se obtuvo el ARN mediante la precipitación descrita anteriormente en el punto 2.6.2 (Figura 15) Todo el protocolo fue basado en el método descrito por Scotto-Lavino *et al.* (2006). Todos los reactivos utilizados fueron del kit comercial Gene Racer (L1500-01; Invitrogen, EEUUA).

2.6.5 Retrotranscripción del ARN para la rápida amplificación de los extremo 5'/3' del cDNA

Después de obtener el RNA ligado al oligo de RNA, se procedió a retrotranscribirlo para obtener el RNA mensajero, para lo cual adicionamos en el primer paso 1µl de Oligo dT [5'-GCTGTCAACGATACGCTACGTAACGGCATGACAGTG(T)24-3'] (50 µM; L1500-01; Invitrogen, EEUUA), 1µl de mix de dNTPs 10mM (10 mM dATP, 10 mM dGTP, 10 mM dCTP, 10 mM dTTP; L1500-01; Invitrogen, EEUUA) y 1µl de H₂O DEPC (L1500-01; Invitrogen, EEUUA) . Se incubaron las reacciones a 65°C durante 5 min en el termociclador; para evitar la formación de estructuras secundarias.

En el segundo paso se adicionaron al homogenizado del primer paso 4 µl de buffer SuperScript 5X (L1500-01; Invitrogen, EEUUA), 1 µl de DTT (1M Ditioneitol; L1500-01; Invitrogen, EEUUA), 1 µl de RT SuperScript III (200 U/µl; L1500-01; Invitrogen, EEUUA) y se protegió la reacción de RNAsas con la adición 1 µl de RNaseOut (40

U/μl; L1500-01; Invitrogen, EEUA). La reacción fue homogenizada por pipeteo y retrotranscrita en el termociclador usando el programa de 50°C durante 1 hora y 70°C por 15 minutos. Finalmente se agregó 1 μl de RNasa H (2 U; L1500-01; Invitrogen, EEUA) y se incubó durante 20 min a 37°C para hidrolizar y eliminar el ARN de una sola cadena, no sintetizado a ADN complementario (Figura 15).

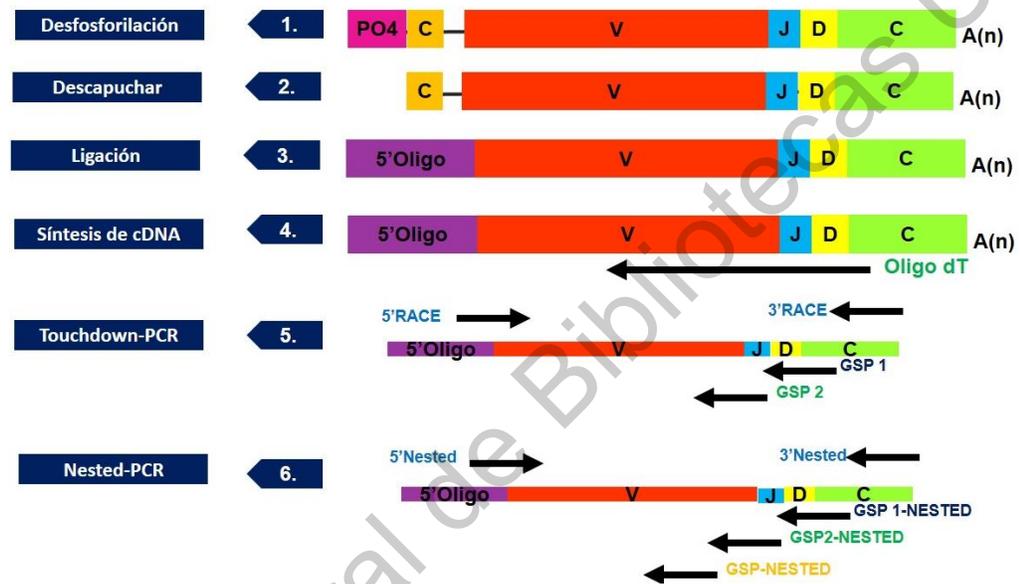


Figura 15| Esquemización del protocolo general del protocolo de RACE. El punto 1 hace referencia al proceso de desfosforilación descrito en 2.2.6. El segundo punto hace referencia al proceso de desfosforilación del inciso 2.6.2. Los siguientes puntos son descritos en los incisos 2.6.3, 2.6.4 y 2.6.5 respectivamente.

2.7 Rápida amplificación de los extremos 5'/3' del cDNA.

2.7.1 PCR Touchdown para las cadenas heavy y lambda del BCR, y Beta TCR.

Se llevó a cabo la amplificación de los extremos de la cadena complementaria basados en el protocolo de PCR Touchdown (Figura 16) (Korbie & Mattick, 2008). Fue utilizado como template el ADNc generado en el punto 2.6 y se agregaron 2.25 μl del cebador GeneRacer 5'/3' 5'GeneRacer [5'-CGACTGGAGCACGAGGACACTGA-3'], (10 μM; N150-50; Invitrogen, EEUA), 3'GeneRacer [5'-GCTGTCAACGATACGCTACGTAACG-3'], (10 μM; N150-30;

Invitrogen, EEUUA) 2.25 µl cebador gen-especifico Reverse/Forward (10µM; OligoT4, México) (Tabla A1; Material Anexo), buffer Pfu 10X (EP0501; Fermentas, EEUUA), 0.75 µl de mix de dNTPs 10mM (10 mM dATP, 10 mM dGTP, 10 mM dCTP, 10 mM dTTP; 18427013; Invitrogen, EEUUA), MgSO₄ (25Mm; EP0501; Fermentas, EEUUA), ADN polimerasa recombinante Platinum Pfu; (2.5 U/µl; EP0501; Fermentas, EEUUA) y se llevó a 25 µl con H₂O bidestilada (10977035; Invitrogen, EEUUA).

Se programaron las reacciones de PCR acorde a las Tms específicas de cada par de cebadores (Figura 16) y se corrieron los productos amplificados en un gel de agarosa 2% y borato de sodio (SB) 1X. Fueron purificadas las bandas resultantes, modificando el protocolo de *Freeze and Squeeze* (Tautz & Renz, 1983). Brevemente, esto significó el corte de la banda de interés que era recubierta con papel Parafilm antes de congelarla a -70 °C por 10 min. Posteriormente, era aplastada la bandas con una superficie plana y se decantaba el líquido resultante por una esquina del papel parafilm.

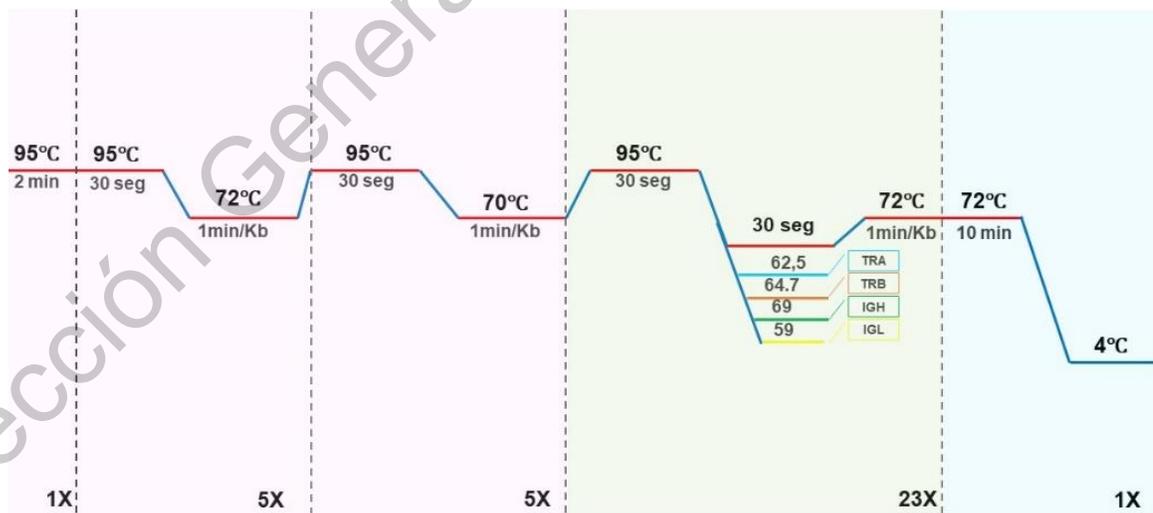


Figura 16| Esquematización del protocolo general de la PCR Touchdown. En la imagen se muestra el programa de Touchdown que se realizó durante el presente estudio ajustado a las enzimas Pfu 2.5 U/µl; EP0501; Fermentas, EEUUA Fermentas, Phusion High-Fidelity ADN polimerasa (2 U/µl; M0530S; BioLabs, UK) y a los cebadores específicos para cada cadena, ejemplificado en la imagen, TRA (azul), TRB (anaranjado), IGL (verde) e IGL (Amarillo).

2.7.2 PCR Touchdown para la cadena Alpha del TCR

Se realizó la amplificación de los extremos de la cadena complementaria con base en el protocolo PCR Touchdown (Figura 16) (Korbie & Mattick, 2008). Se adicionaron al ADNc generado en el punto 2.6 1.5 µl del cebador GeneRacer 5'/3' 5'GeneRacer (N150-50; Invitrogen, EEUUA), [5'-CGACTGGAGCACGAGGACACTGA-3'] 3'GeneRacer (N150-30; Invitrogen, EEUUA), [5'-GCTGTCAACGATACGCTACGTAACG-3']. 1.5µl del cebador gen-especifico Reverse/Forward (10µM; OligoT4, México) (Tabla A1 ; Material Anexo), 5µl de buffer Phusion HF 5X con MgCl₂ (7.5mM; 2 U/µl; M0530S; BioLabs, UK), 0.5 µl de mix de dNTPs 10mM (10 mM dATP, 10 mM dGTP, 10 mM dCTP, 10 mM dTTP; 18427013; Invitrogen, EEUUA), 0.5µl DMSO (100X; M0530S; BioLabs, UK), 0.5 µl de Phusion High-Fidelity ADN polimerasa (2 U/µl; M0530S; BioLabs, UK) y se llevó a 25 µl con H₂Odd (10977035; Invitrogen, EEUUA).

Se ajustó el programa de PCR acorde a la T_m específica de cada par de cebadores (Figura 16) y se corrieron las muestras en un gel de agarosa 2% y borato de sodio (SB) 1X para visualizarlas.

2.8 PCR anidada para las cadenas Heavy y Lambda del BCR, y alpha y beta del TCR

Se utilizó 1 µg de la muestra de la PCR Touchdown para re-amplificar específicamente las cadenas Heavy, Lambda y Beta. Se modificó el protocolo de PCR de punto final para cada par de cebadores (Tablas A2; ; Material Anexo). Fueron adicionados a la muestra 2.5 µl de de buffer de ADN polimerasa 10X(10342053; Invitrogen, EEUUA), 0.5 µl de mix de dNTPs 10mM (10 mM dATP, 10 mM dGTP, 10 mM dCTP, 10 mM dTTP; Invitrogen), 0.75 µl de MgCl₂(10342053; Invitrogen, EEUUA), 0.5 µl de 5'/3'Nested cebador 5'Nested (N150-51; Invitrogen, EEUUA), [5'-GGACACTGACATGGACTGAAGGAGTA-3'], 3'Nested (N150-31;

Invitrogen, EEUUA), [5'-CGCTACGTAACGGCATGACAGTG-3'], 0.5 µl del cebador anidado gen-especifico Reverse/Forward (10µM; OligoT4, México), 0.25 µl de ADN polimerasa recombinante (10342053; Invitrogen, EEUUA) y se llevó a 25 µl con H₂O bidestilada(10977035; Invitrogen, EEUUA).

Se modificó el programa de PCR de acuerdo a la T_m específica de cada par de cebadores (Figura 17) y se corrieron las las muestras en un gel de agarosa 2% y borato de sodio (SB) 1X. Fueron purificadas las bandas resultantes, modificando el protocolo de *Freeze and Squeeze* (Tautz & Renz, 1983), como se explica en el punto 2.7.1.

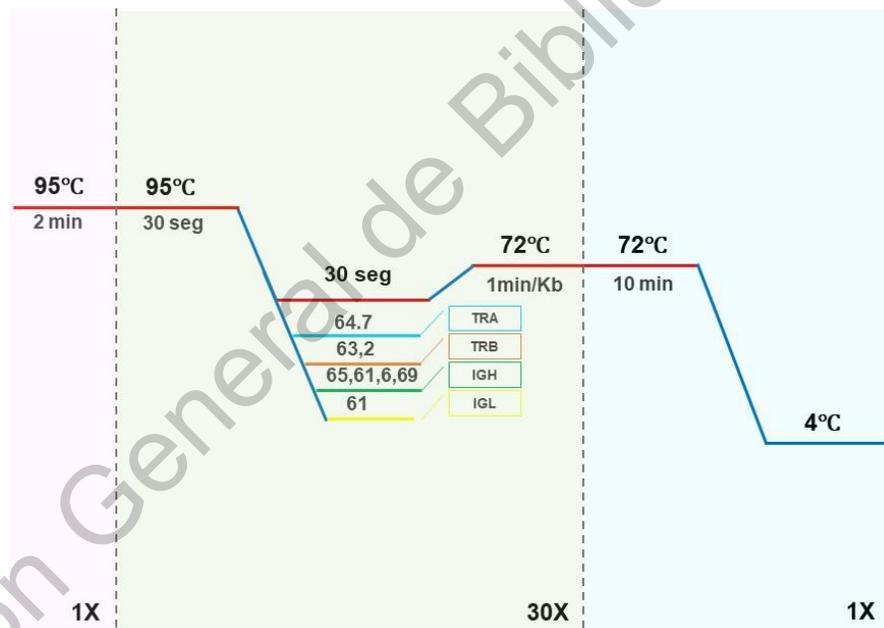


Figura 17| Esquematización del protocolo general de la PCR Anidada. En la imagen se muestra el programa de Touchdown que se realizó durante el presente estudio ajustado a la enzima ADN polimerasa recombinante (Invitrogen) y a los cebadores específicos para cada cadena, ejemplificado en la imagen, TRA (azul), TRB (Anaranjado), IGH (verde) e IGL (amarillo).

Se obtuvieron bandas de pesos entre los 100pb-1100pb con el protocolo RACE-Alpha (Figura A6-A; Material Anexo). Para enriquecer las especificidad de los amplicones se realizó una PCR anidada con lo que se obtuvieron bandas de con mayor concentración entre los 120-900 pb (Figura A6-B; Material Anexo).

Del mismo modo se obtuvieron bandas de pesos diversos para cada cadena con el protocolo RACE, para la cadena beta entre los 150pb-1300 pb (Figura A7-A; Material Anexo), heavy; 75pb-1500 pb (Figura A7-A; Material Anexo) y lambda 100pb-800 pb (Figura A8-A; Material Anexo). Se enriqueció la especificidad de las bandas obtenidas para cada caso obteniendo bandas de pesos entre 100 -500 pb para la cadena beta (Figura A8-B; Material Anexo), 100 pb-900 pb en la heavy (Figura A8-B; ; Material Anexo) y 90 pb-650 pb en la Lambda (Figura A9-B; Material Anexo). En las cadenas heavy y lambda se obtuvieron algunos segmentos de pesos similares debido a que se utilizó más de un par de cebadores para realizar PCR anidada, aunque las reacciones se hicieron individualmente los cebadores hibridaron en sitios muy cercanos.

2.9 Obtención de bandas únicas por PCR anidada para la secuenciación de los productos del protocolo de 5'/3' RACE

2.9.1 PCR punto final para obtener bandas únicas de las cadenas heavy y lambda del BCR, y alpha y beta del TCR

Se utilizó como templete el purificado de *Freeze and Squeeze*, para realizar PCR de punto final de cada banda purificada. Se adicionaron a la muestra 1.5 µl de 10X buffer de ADN polimerasa (10342053; Invitrogen, EEUUA), 0.5 µl de mix de dNTPs 10mM (10 mM dATP, 10 mM dGTP, 10 mM dCTP, 10 mM dTTP; Invitrogen), 0.45 µl de MgCl₂ (10342053; Invitrogen, EEUUA), 0.25 µl de 5'/3'Nested cebador 5'Nested (N150-51; Invitrogen, EEUUA), [5'-GGACTGACATGGACTGAAGGAGTA-3'], 3'Nested (N150-31; Invitrogen,

EEUUA), [5'-CGCTACGTAACGGCATGACAGTG-3'], 0.25µl del cebador anidado gen-especifico Reverse/Forward (10µM; OligoT4, México), 0.15µl de ADN polimerasa recombinante (10342053; Invitrogen, EEUUA) y se llevó a 15 µl con H₂O bidestilada (10977035; Invitrogen, EEUUA).

Se analizaron las reacciones visualmente mediante un gel 2% de agarosa y borato de sodio (SB) 1X y se aislaron bandas únicas por el protocolo de *Freeze and Squeeze* (Tautz & Renz., 1983).

Se obtuvieron bandas únicas, que fueron re-amplificadas para cada cadena y secuenciadas, debido a que no se conocían los pesos de cada transcrito de las cadenas de *Zalophus californianus*. En la cadena beta se obtuvieron 18 amplicones (Figura A10; Material Anexo), para la cadena heavy 34 (Figuras S12; Material Anexo) y en la lambda 13 (Figura A12; Material Anexo).

Fueron secuenciadas 47 muestras; 18 pertenecientes a la cadena beta, 20 a la cadena heavy y nueve a la lambda. Se caracterizaron ocho ampliaciones de la cadena beta mediante alineamiento de secuencias (Figura A17A-H; Material Anexo), pertenecientes a las muestras B12, B3, B8, B17, B10, B3 y B1 (Figura A10; Material Anexo). Finalmente se determinó que el peso medio del transcrito de la cadena beta de *Z. californianus* es de 276 pb (Figura A13; Material Anexo).

Por otro lado, en las 20 muestras de la cadena heavy caracterizamos cinco de las muestras H23, H17, H8, H24 y H9 (Figura A22I-M; Material Anexo), (Figura A11; Material Anexo), de las cuales el peso medio de esta cadena para *Z. californianus* es de 561 pb (Figura A13; Material Anexo), mientras que de las nueve muestras para la cadena lambda, fue caracterizada una única secuencia de la muestra L11 (Figura A17-N; Material Anexo), con el peso de 424 pb (Figura A13; Material Anexo).

2.10 Clonación para obtener bandas únicas de las cadenas Heavy y Lambda del BCR, y Alpha y Beta del TCR

Las bandas menores a 200 pb fueron clonadas en el vector pCR[®]4TOPO (K457502; Invitrogen, EEUUA) y se confirmó la inserción del producto mediante PCR de colonia específica, con los cebadores M13 del Kit de pCR[®]4TOPO (K457502; Invitrogen, EEUUA) y restricción con 10X RE Buffer (FD0275; Invitrogen, EEUUA) y EcoRI (10 U/μl; FD0275; Invitrogen, EEUUA) y se purificaron las colonias por miniprep (Pederson, 2015) (Figura A18; Material Anexo).

2.11 Secuenciación tipo Sanger de los productos del protocolo de 5'/'3' RACE

Las muestras derivadas de la purificación de bandas, para prueba de cebadores, extracción de plásmidos y RACE se secuenciaron en MacroGen Inc. (*Beotkkot-ro, Geumcheon-gu*). Se generaron las secuencias consenso con la herramienta de secuencia consenso del menú de alineamiento del program BioEdit, versión 7.0.5.3 y se alinearon las secuencias para determinar las regiones conservadas de cada cadena y el record de identidad entre ellas en el panel de alineamiento de secuencias en el programa Jalview versión 2.11.

2.12 Análisis bioinformático de las secuencias obtenidas por secuenciación tipo Sanger de los productos del protocolo de 5'/'3' RACE

Se alinearon las secuencias de nucleótidos en la herramienta básica de búsqueda de alineación local BLAST del Centro Nacional de Información Biotecnología NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Fueron seleccionadas aquellas secuencias que alineaban con secuencias de cadenas alpha, beta, heavy o lambda de otros mamíferos y se verificó de nuevo la calidad de la secuencia en *BioEdit*, versión

7.0.5.3 para analizarla. Finalmente se determinó el tamaño de los amplicones específicos para cada cadena (Figura A13; Material Anexo).

2.13 Análisis estadístico de los tamaños de los receptores

Todo el análisis estadístico fue realizado en R versión 3.6. Específicamente, se utilizó una ANOVA para comparar el tamaño de la media de los pesos moleculares (en pb), de las cadenas alpha/beta y heavy de los receptores TCR y BCER del *Z. californianus*.

2.14 Secuenciación de nueva generación de los productos del protocolo de 5'/3' RACE

Basados en los fragmentos del protocolo de RACE, se generaron bibliotecas para realizar secuenciación de nueva generación con ellas. Previo a esto fue evaluada una muestra de la cadena alpha y por fluorescencia se identificaron fragmentos de tamaños que no se habían observado en el gel de agarosa o en la secuenciación Sanger (Figura A14-A; Material Anexo).

Previamente a la secuenciación de nueva generación, las muestras se validaron por fluorescencia en un bioanalizador (Agilent 2100 Bioanalyzer) y se realizaron bibliotecas de cuatro muestras las cuales representaban un pool de seis muestras provenientes de seis individuos de *Z. californianus*, de esta manera durante este estudio se trabajó con las muestras AA: TRA, BB: TRB, HH: IGL, LL: IGL (Figura A15; Material Anexo), a las cuales se les ligó un código de barras de 120 pares de bases (Figura A14-S15; Material Anexo).

Las muestras totales, obtenidas por el protocolo de 5'/3' RACE, se secuenciaron en la plataforma de Secuenciación por Síntesis (SBS), del sistema Illumina MiSeq tipo

nano, con la profundidad de 2x250, en la unidad de secuenciación del Laboratorio Nacional de Genómica Laboratorio Nacional de Genómica CINVESTAV (Irapuato, Guanajuato). Se caracterizaron en promedio 1,325,091 lecturas entre 35 pb y 10,380 pb por muestra con secuenciación de nueva generación, de las cuales la distribución de puntuación de calidad fue en promedio de 20 Phreds por muestra (Figura A14C-D; Material Anexo).

2.15 Análisis bioinformático de las secuencias obtenidas por secuenciación de nueva generación de los productos del protocolo de 5'3' RACE

Las lecturas derivadas de la secuenciación se sometieron a un análisis de calidad con el programa Trimmomatic versión 0.36.6 (Bolger *et al.*, 2014), con el cual se recortaron los nucleótidos de baja calidad (Figura 18). Posteriormente las lecturas se ensamblaron usando el programa Trinity (Beta) versión 0.0.1 (Haas *et al.*, 2013), (Figura 19).

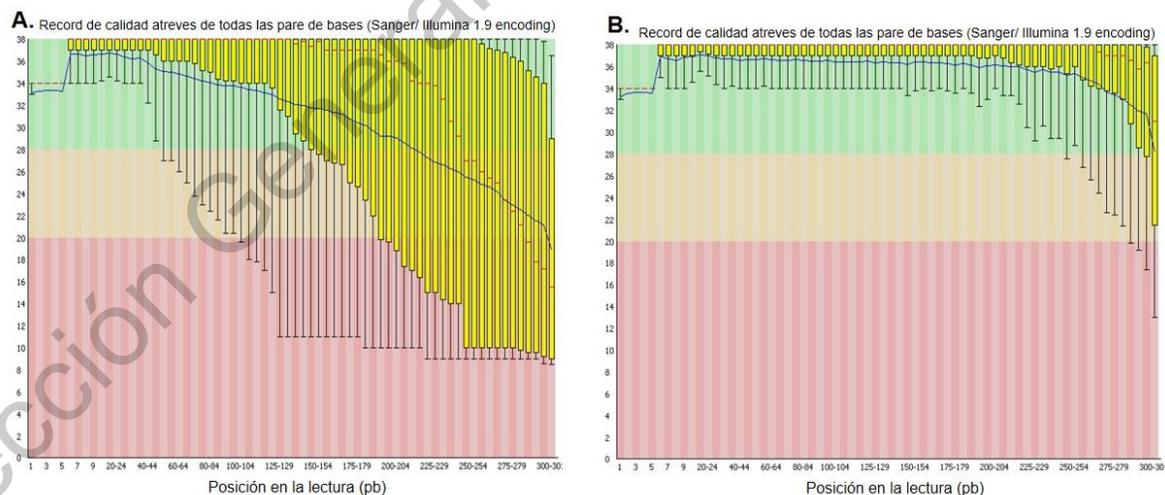


Figura 18| Esquematzación del control de calidad para las lecturas derivadas de la secuenciación de nueva generación. A) la gráfica de las lecturas antes de eliminar las pb de poca calidad. B) la gráfica de las lecturas después de la eliminación de las pb de poca calidad. En las gráficas los colores se muestran como una escala; por lo tanto, el color rojo ejemplifica la zona de poca calidad, el amarillo la zona aceptable y el verde la zona óptima. Los gráficos se generaron mediante archivos Fastq, en el programa FastQC versión 0.11.3.

Para establecer cuales lecturas eran secuencias TCR e Ig, realizamos un análisis de componentes para las lecturas AA, BB, HH y LL en conjunto con secuencias TCR e Ig de otros carnívoros y mamíferos marinos (Texto A24; Material Anexo). Adicionalmente se agregamos otras secuencias con dominios de la súper familia de inmunoglobulinas con las cuales los receptores TCR e Ig comparten dominios conservado en las regiones constantes y variables (Figura A23; Material Anexo).

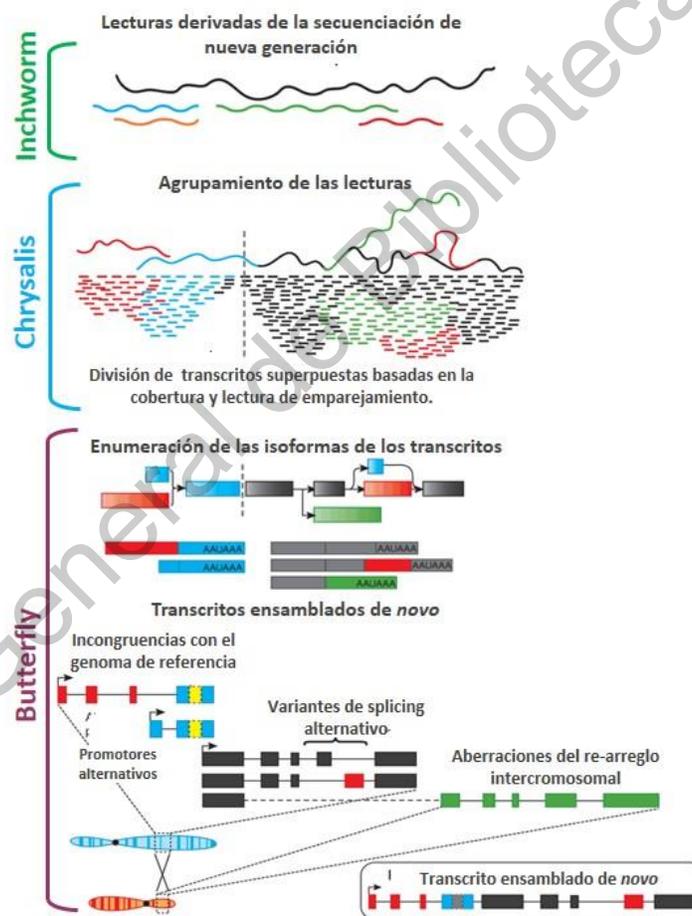


Figura 19| Diagrama de flujo del funcionamiento de Trinity (Beta). El diagrama muestra las tres etapas de análisis de Trinity (Beta). En Inchworm las lecturas son reconocidas y extendidas complementariamente mediante valores de k-mer. En Chrysalis, las lecturas se agrupan por sobreposición y emparejamientos por valores de k-mer. Por último, en Butterfly los transcritos se ensamblan y se identifican las posibles variantes de splicing.

Para el análisis anterior, fueron alineados el conjunto de secuencias (Kato *et al.*, 2013) y generada una matriz de distancias (Schloss, 2010), posteriormente fueron producidas las coordenadas para el análisis de componentes principales de la matriz de distancia con el programa Mothur (Schloss *et al.*, 2009). Para generar el gráfico de componentes principales se usaron los valores obtenidos de la matriz de distancias y coordenadas de componentes principales, los cuales fueron gráficos en dos planos (PC1) y (PC2) con el paquete *ggplot2* en RStudio versión 3.6.0 y se seleccionaron las secuencias que se trasponían a los puntos de las secuencias TCR e Ig de otras especies (Script A25; Material Anexo). El método de clusterización que se utilizó fue denominado Eco-Ig.

Las secuencias fueron filtradas a partir de una lista en Samtools versión 1.9 construido en conda (Python 3) (Script A26; Material Anexo). Finalmente, después de procesar con IgDiscover (Script A26; Material Anexo) fue determinado el tamaño en pares de bases de cada cadena, ahora por secuenciación de nueva generación (Figura A16; Material Anexo).

2.16 Análisis estadístico de los valores E del análisis de secuencias de nueva generación y de frecuencias de genes V(D)J

Se alinearon las secuencias contra los genes V de perro (*Canis lupus familiaris*), reportados en IMGT y se obtuvieron los valores E con el programa BLASTn para Ubuntu, desarrollado en conda. Fueron unificados por el tamaño de las muestras mediante random seed de R y se compararon las medias de los valores E de los grupos previos al filtro y después de filtrarlos en R versión 3.6.0, por prueba T de student.

Las frecuencias de cada gen V(D)J se calcularon con base en el porcentaje de contribución de cada gen, al número total del repertorio. Finalmente estas frecuencias fueron gráficas en el programa R 3.6.0, usando el paquete *ggplot2*.

2.17 Identificación de los segmentos génicos V(D)J de los receptores TCR y BCR (Ig)

Se analizaron las secuencias de *Zalophus californianus* obtenidas mediante secuenciación tipo Sanger y de nueva generación así como de otras 18 especies registradas en GenBank en el programa IgDiscover 0.11, (Texto A26; Material Anexo) (Corcoran *et al.*, 2016), siguiendo las reglas para la nomenclatura establecidas por IMGT (Lefranc, 2001).

2.18 Análisis filogenético

Para evaluar la evolución de los segmentos V(D)J de los receptores antigénicos de *Z. californianus*, fueron comparados los segmentos con los correspondientes de 13 especies de carnívoros terrestres y marinos (*C. lupus familiaris*, *F. catus*, *A. jubatus*, *A. melanoleuca*, *C. ursinus*, *E. lutris*, *M. putorius*, *N. schauinslandi*, *O. rosmarus*, *P. concolor*, *U. arctos*, *U. maritimus*, *L. weddelli*) y seis mamíferos marinos de las ordenes Cetacea y Sirenia (*D. leucas*, *L. vexillifer*, *T. truncatus*, *O. orca*, *T. manatus*, *L. acutus*). Los árboles se construyeron únicamente con secuencias funcionales de los segmentos V, los pseudogenes y ORFs se eliminaron en este análisis. Mientras que en el análisis entre el repertorio de los segmentos DJ, de los receptores TCR y BCR de *Z. californianus*, los pseudogenes y ORFs se incluyeron en la construcción de los árboles.

La historia evolutiva de los segmentos V(D)J de *Z. californianus* fue inferida usando el método de Maximum Likelihood, basado en el modelo de Kimura 2 con 1000 réplicas de bootstrap (Kimura, 1980; Felsenstein, 1985). La búsqueda heurística se obtuvo automáticamente por los algoritmos de Neighbor-Join y BioNJ, a partir de una matriz de distancias pares, usando la aproximación Maximum Composite

Likelihood (MCL). Los análisis fueron realizados en el programa Mega X (Kumar *et al.*, 2018).

2.19 Distribución geográfica

Se obtuvo la distribución geográfica de las especies *C. lupus familiaris*, *F. catus*, *A. jubatus*, *A. melanoleuca*, *C. ursinus*, *E. lutris*, *M. putorius*, *N. schauinslandi*, *O. rosmarus*, *P. concolor*, *U. arctos*, *U. maritimus*, *L. weddelli*) usando los datos reportados en la base de datos de Maldonado *et al.*, 2015 y se gráfcaron las coordenadas usando los paquetes ggplot2, dplyr, rgbif, sp, countrycode y CoordinateCleaner en R versión 3.6.

2.20 Caracterización de las secuencias de aminoácidos de las cadenas de los receptores TCR y BCR

De las secuencias de nucleótidos caracterizadas por el programa IgDiscover, fueron seleccionadas aquéllas que contaban con las zonas Framework 1, 2, 3 y CDR 1, 2 y 3. Posteriormente fueron alineadas y se generó una secuencia consenso con el programa BioEdit, versión 7.0.5.3, para después correrla de nuevo en el programa IgDiscover y generar una secuencia de aminoácidos única para cada cadena.

2.21 Modelado de los receptores TCR y BCR

Se obtuvo el consenso de secuencias de aminoácidos de cada cadena de los receptores de *Z. californianus* para después parearlas y generar secuencias de aminoácidos lineal de los receptores TCR (alpha/beta), BCR (heavy/lambda) y MHC

II (DRBA). Posteriormente fueron predichas las estructuras tridimensionales de los receptores usando un enfoque jerárquico, basado en plantillas estructurales del repertorio de Protein Data Bank (<http://www ww pdb.org/>), mediado por la aproximación LOMETS de modelos atómicos de longitud completa construidos por simulaciones de ensamblajes interactivos, los cuales mediante enhebración logran proponer modelos 3D funcionales. Finalmente, de los modelos resultantes se eligió el modelo que tuviera una puntuación con un rango entre -5 y -2, la cual permitió estimar la calidad de cada modelo (Figura A20A, A21A y A22A; Material Anexo). Para el uso de los subprocesos anteriormente desglosados se utilizó el programa I-TASSER (Yang *et al.*, 2015).

2.22 Análisis de calidad de las estructuras tridimensionales

Se evaluó la contribución energética entre los ángulos diédricos contra los residuos de aminoácidos en la estructura de cada proteína. Para determinar los contornos de las regiones favorecidas se extrajeron los datos de 12,521 estructuras experimentales con un corte de identidad del 30%, el cual es un corte espectral que proporciona una resolución de 2.5 Å de rayos X. Las ocurrencias en los ángulos Phi y Psi para todas las categorías (Prolina, Pre-Prolina y Glicina), se determinaron por el número de pares en los espectros Phi/Psi observados en las líneas de contorno (Figura A20B, A21B y A22B; Material Anexo).

Complementario al análisis de arriba, se estimó también la calidad con base en las diferencias geométricas de calidad total de la estructura y de cada residuo. Fue estimada la calidad global del modelo (GMQE) y la puntuación Z de QMEAN, para estimar la concordancia entre la estructura del modelo y las estructuras experimentales de tamaños similares y su potencial de solvatación y potencial de ángulo de torsión (Figura A20D, A21D, A22D; Material Anexo). Finalmente, se comparó la calidad de los componentes individuales de los modelos obtenidos por

la comparación con otras estructuras similares a nivel de residuos con los valores normalizados QMEAN (Figura A20C, A21C, A22C; Material Anexo). Los subprocesos listados en los párrafos anteriores se realizaron en el programa SWISS-MODEL (Waterhouse *et al.*, 2018).

Dirección General de Bibliotecas UAQ

V. Resultados

3.1 Secuencias de las cadenas alpha/beta y heavy del TCR y BCR por secuenciación Sanger

Las secuencias de la cadena beta mostraron el 85% de identidad nucleotídica entre sí (Figura 20). Las secuencias de la cadena heavy tuvieron un porcentaje de identidad ligeramente más alto 88% (Figura 21). Por otro lado, la identidad entre secuencias nucleotídicas de la cadena lambda no pudo llevarse a cabo, debido a la falta de información obtenida para esta cadena, como se menciona en la sección de Metodología de la tesis (Figura 22).

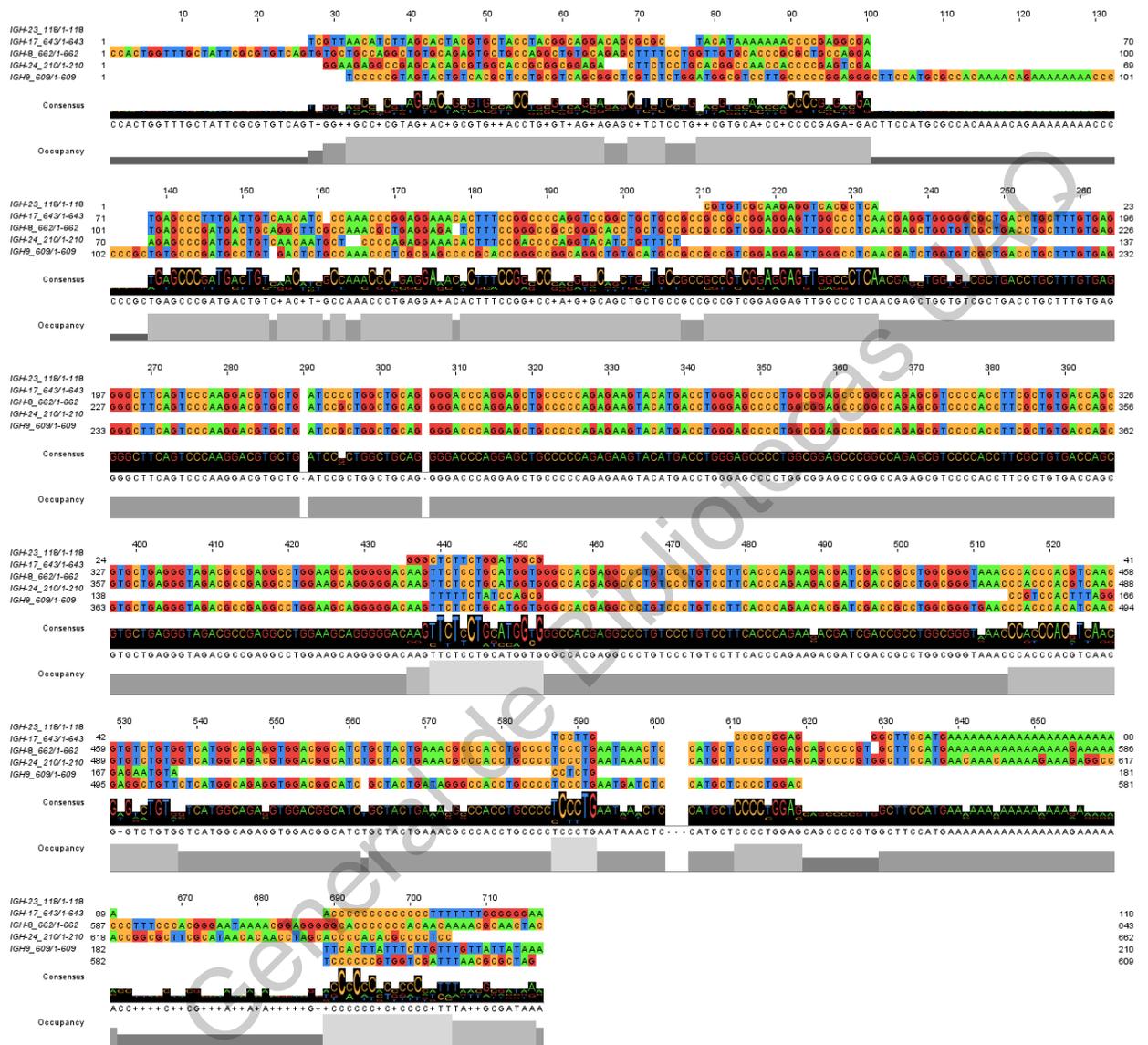


Figura 21| Identidad en la secuencia nucleotídica de las secuencias caracterizadas por secuenciación Sanger de la cadena heavy. Las secuencias se alinearon por el método de alineamiento de secuencias múltiples (MUSCLE) y se visualizó en el programa Jalview versión 2.11.0, los colores en el gráfico indican la conservación de los nucleótidos entre secuencias, el apartado de consenso muestra la secuencia consenso. Las abreviaciones IGH seguidas de un número, corresponden a los identificadores de las muestras de las cuales se caracterizó la secuencia.

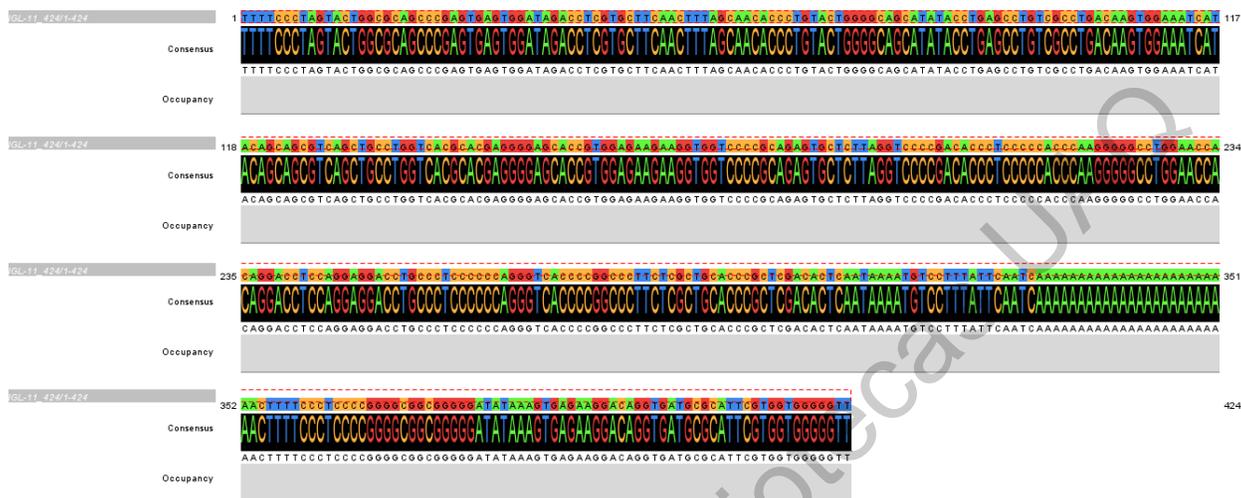


Figura 22| Esquema de la secuencia caracterizada por secuenciación Sanger de la cadena lambda. La secuencia se visualizó en el programa Jalview versión 2.11.0, las abreviaciones IGL-11 corresponden al identificador de la muestra de la cual se caracterizó la secuencia.

3.2 Caracterización de secuencias alpha/beta,heavy/lambda del TCR y BCR(IG) por secuenciación de nueva generación

De las bibliotecas generadas por secuenciación de nueva generación obtuvimos 1,500,349 lecturas (Figura A15-C; Material Anexo) entre 35 y los 10,380 pb en la muestra AA (Figura A19A-B; Material Anexo), de las cuales 6,479 lecturas eran correspondientes a la cadena alpha (Figura 23). De la muestra BB obtuvimos 1,042,004 lecturas (Figura A15-C; Material Anexo) entre 35 y 6,708 pb (Figura A19C-D; Material Anexo) de las cuales 1,003 lecturas correspondían a la cadena beta del TCR (Figura 24).

Para las muestras del BCR, en la muestra HH obtuvimos un total de 1,262,949 lecturas (Figura A15-C; Material Anexo) entre 35 y 8,418 pb, de las cuales identificamos 5,120 pertenecientes a la cadena heavy (Figura 25). Mientras que de la muestra LL obtuvimos 1,612,936 lecturas (Figura A15-C; Material Anexo), entre

los 35 y 6,554 pb (Figura A19G-H; Material Anexo), entre estas 6621 correspondían a la cadena lambda (Figura 26).

En la comprobación del método utilizado se obtuvieron valores E cercanos a los genes correspondientes en perro (*Canis lupus familiaris*). En la cadena alpha las diferencias entre el análisis pre y post de nuestro filtro dieron como resultado valores no significativos, TRAV10, $p=0.6$ (Figura 23B), TRAV9-3, $p=0.8$ (Figura 23C) y TRAV37, $p=0.48$ (Figura 23D).

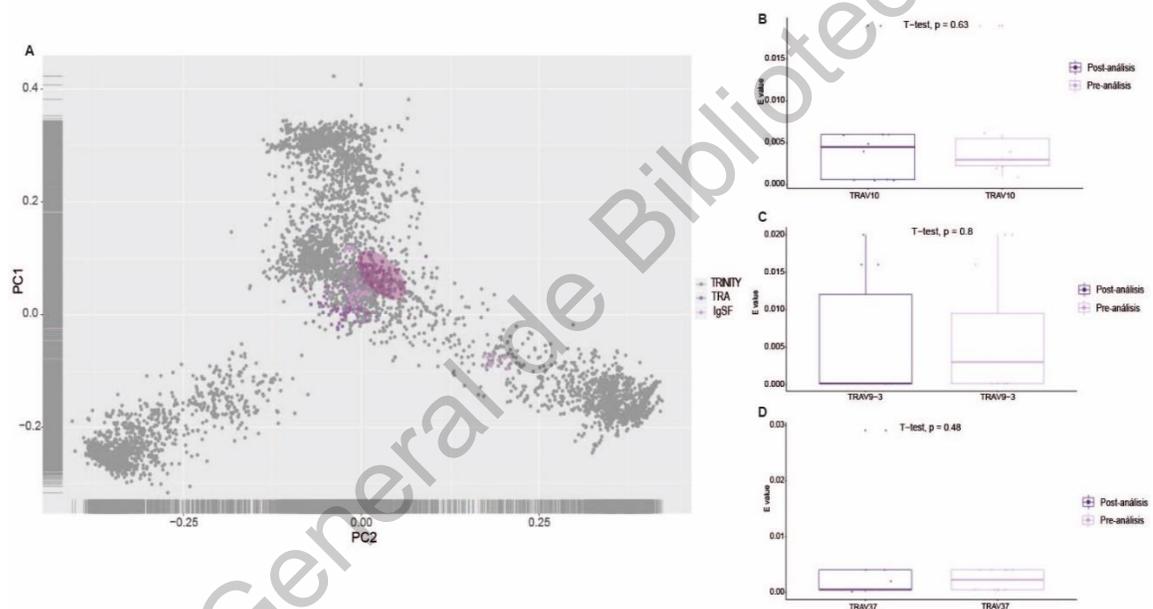


Figura 23| Análisis de componentes principales de la cadena alpha de la muestra AA. A, El gráfico de puntos, muestra en gris las lecturas derivadas de la secuenciación NGS, ensamblada con el programa Trinity; los puntos lila son lecturas de miembros de la super familia de inmunoglobulinas. En morado, lecturas de la cadena alpha de otras especies de mamíferos, el óvalo color morado delimita las secuencias que seleccionamos para analizar. B, C y D, muestran las diferencias entre los genes V de la cadena alpha antes de filtrar las secuencias y después de filtrarlas. TRAV10, $p=0.6$, TRAV9-3, $p=0.8$ y TRAV37, $p=0.48$.

Para la cadena beta, se observaron diferencias significativas en el análisis pre y post para el gen TRBV12, $p= 0.036$ (Figura 24B). Mientras que para los genes TRBV29 (Figura 24C) y TRBV4-1 (Figura 24D), no se detectaron diferencias significativas ($p=0.46$ y $p=0.25$, respectivamente).

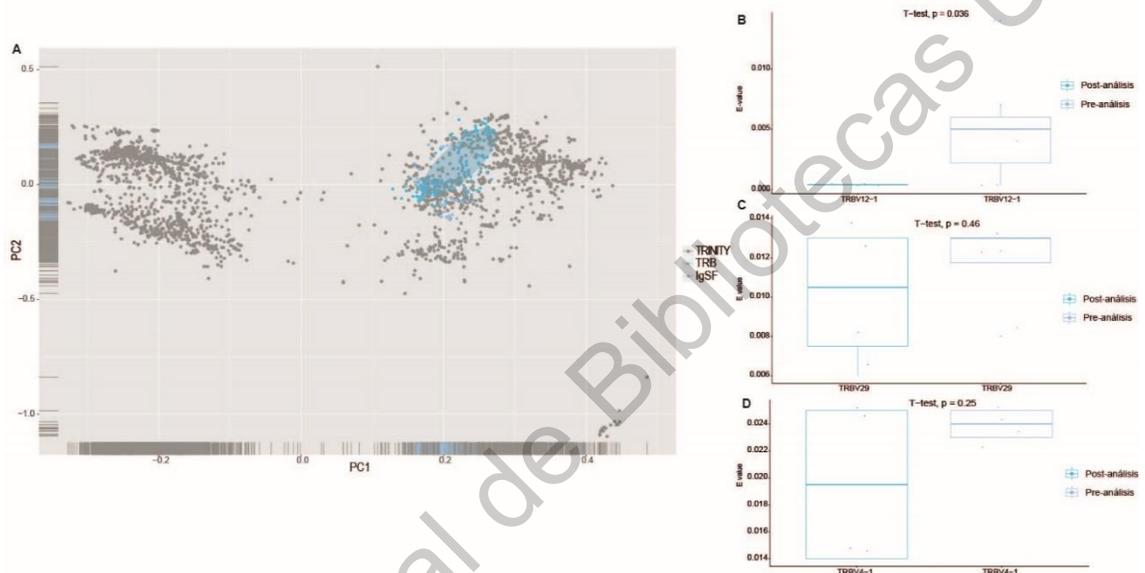


Figura 24| Análisis de componentes principales de la cadena beta de la muestra BB. El gráfico de puntos, muestra en gris las lecturas derivadas de la secuenciación NGS, ensamblada con el programa Trinity; los puntos azul pastel son lecturas de miembros de la super familia de inmunoglobulinas. En azul celeste, lecturas de la cadena beta de otras especies de mamíferos, el óvalo color azul celeste delimita las secuencias que seleccionamos para analizar. B, C y D, muestran las diferencias entre los genes V de la cadena beta antes de filtrar las secuencias y después de filtrarlas. TRBV12, $p= 0.036^*$, TRBV29, $p=0.46$ y TRBV4-1, $p=0.25$.

En la comparación de los valores E del análisis pre y post de la cadena heavy del BCR no se detectaron diferencias significativas para los genes IGHV1-15 ($p= 0.11$; Figura 25B), IGHV3-10 ($p=0.18$; Figura 25C) y IGHV4-1 ($p=0.95$; Figura 25D), pre y post clusterización. De la misma forma, en la cadena lambda tampoco se detectaron diferencias significativas en los valores E para los genes IGLV1-48 ($p= 0.15$; Figura 26B), IGLV1-58 ($p=0.73$; Figura 26C) y IGHV5-64 ($p=0.54$; Figura 26D).

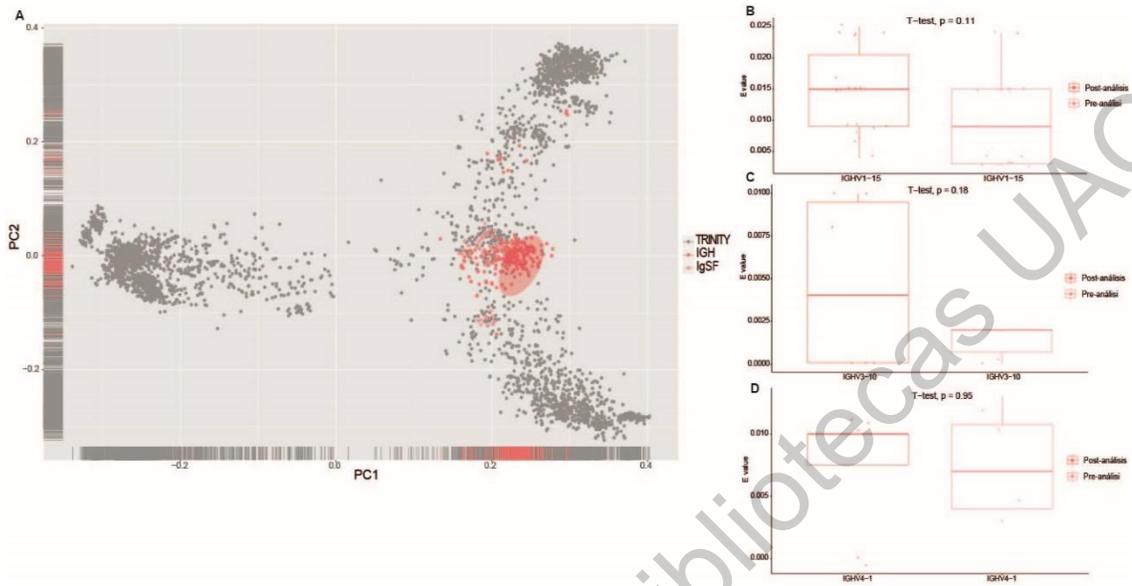


Figura 25| Análisis de componentes principales de la cadena heavy de la muestra HH. El gráfico de puntos, muestra en gris las lecturas derivadas de la secuenciación NGS, ensamblada con el programa Trinity; los puntos rosa pastel son lecturas de miembros de la super familia de inmunoglobulinas. En salmón, lecturas de la cadena heavy de otras especies de mamíferos, el óvalo color salmón delimita las secuencias que seleccionamos para analizar. B, C y D, muestran las diferencias entre los genes V de la cadena heavy antes de filtrar las secuencias y después de filtrarlas. IGHV1-15, $p=0.11$, IGHV3-10, $p=0.18$ y IGHV4-1, $p=0.95$.

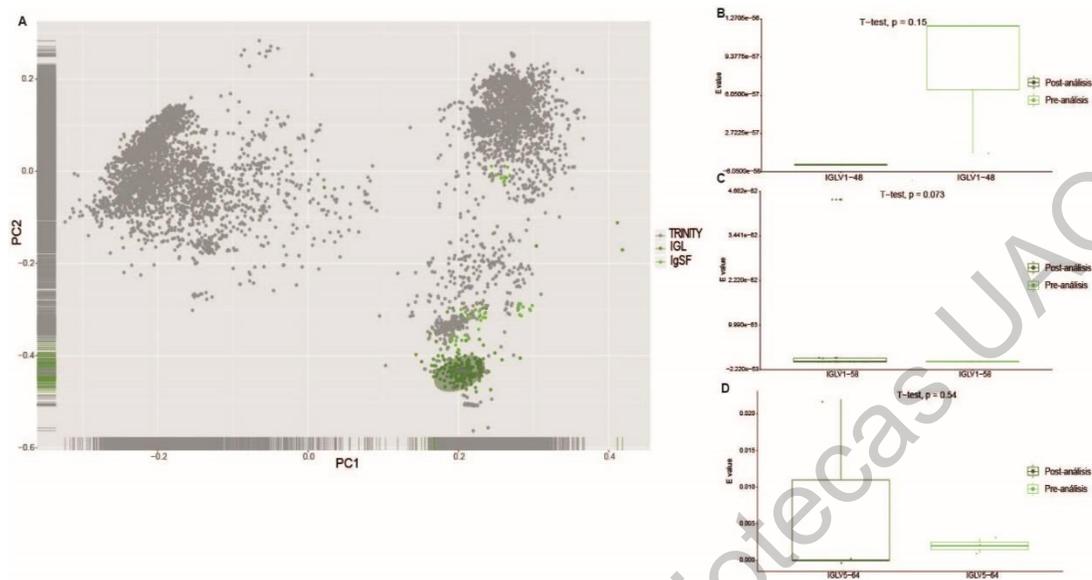


Figura 26| Análisis de componentes principales de la cadena lambda de la muestra LL. El gráfico de puntos, muestra en gris las lecturas derivada de la secuenciación NGS, ensamblada con el programa Trinity; los puntos verde pastel son lecturas de miembros de la super familia de inmunoglobulinas. En verde oscuro, lecturas de la cadena lambda de otras especies de mamíferos, el óvalo color verde oscuro, delimita las secuencias que seleccionamos para analizar. B, C y D, muestran las diferencias entre los genes V de la cadena lambda antes de filtrar las secuencias y después de filtrarlas. IGLV1-48, $p=0.15$, IGLV1-58, $p=0.73$ y IGHV5-64, $p=0.54$.

3.3 Caracterización de los segmentos génicos V(D)J de los receptores TCR y BCR (Ig) de *Z. californianus*

El objetivo principal de esta tesis fue la caracterización genética de los receptores TCR y BCR del lobo marino de California. Se logró caracterizar individualmente los segmentos génicos V, D y J que forman las cadenas de estos receptores. Las lecturas totales de los segmentos V, D y J se dividieron en genes funcionales, pseudogenes y ORF (por sus siglas en inglés: Open Reading Frame).

Las frecuencias de los segmentos individuales se analizaron con base en el número total de lecturas de genes V, D y J funcionales, pseudogenes y ORF de cada

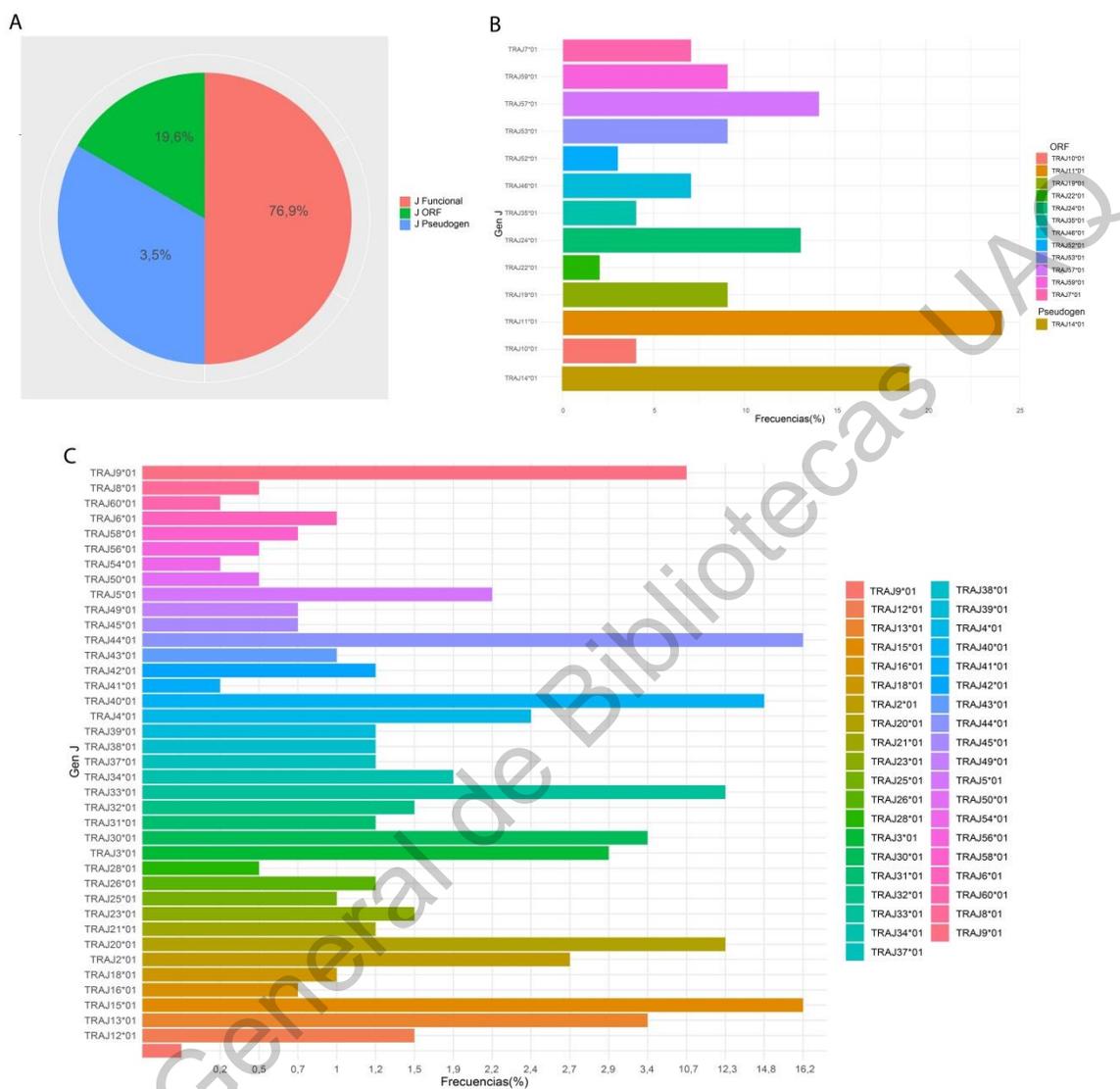


Figura 28| Caracterización de los segmentos génicos J de la cadena alpha del TCR de *Z. californianus*. A) Porcentajes de genes J funcionales, pseudogenes y ORF. B) Frecuencias individuales de los pseudogenes y ORF TRAJ. C, frecuencias individuales de genes TRAJ funcionales. Los valores de frecuencia están expresados en porcentajes, basados en el número total de lecturas para cada grupo de gén o pseudogén.

3.3.2 Frecuencias de los segmentos V(D)J de la cadena beta del TCR de *Z. californianus*

La cadena beta se conformó de 19 genes V funcionales, cuatro pseudogenes y un ORF, con las lecturas de 48, 17 y 1 totales, respectivamente (Figura 29). Para los genes D, se encontraron dos genes, con ocho lecturas totales (Figura 30-D). En cuanto a los genes J de esta cadena se encontraron cuatro genes J funcionales, un ORF y un pseudogén con las lecturas de 9, 2 y 2, respectivamente (Figura 30A-C).



Figura 29| Caracterización de los segmentos génicos V de la cadena beta del TCR de *Z. californianus*. A) Porcentajes de genes V funcionales, pseudogenes y ORF. B) Frecuencias individuales de los pseudogenes y ORF TRBV. C, frecuencias individuales de genes TRBV funcionales. Los valores de frecuencia están expresados en porcentajes, basados en el número total de lecturas para cada grupo de gén o pseudogén.

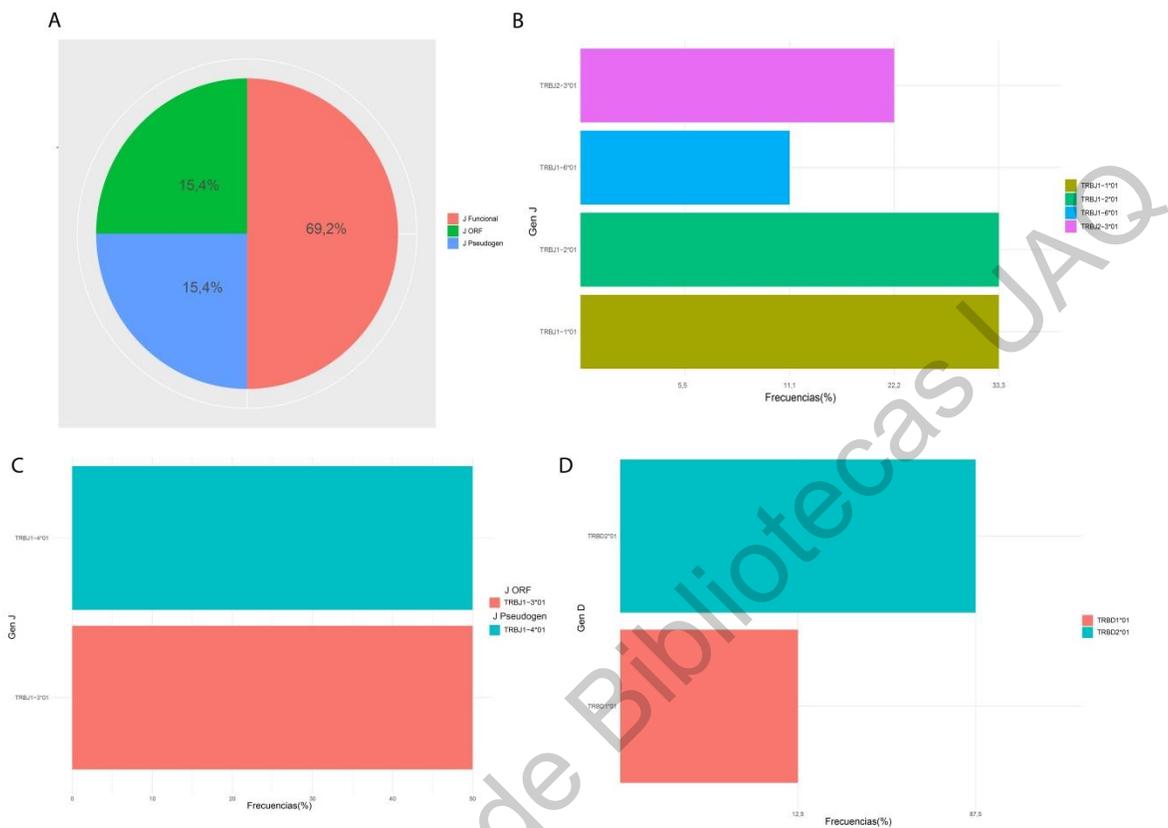


Figura 30| Caracterización de los segmentos génicos D y J de la cadena beta del TCR de *Z. californianus*. A) Porcentajes de genes J funcionales, pseudogenes y ORF. B) Frecuencias individuales de genes TRBJ funcionales. C, frecuencias individuales de los pseudogenes y ORF TRBJ. D, frecuencias individuales de genes D. Los valores de frecuencia están expresados en porcentajes, basados en el número total de lecturas para cada grupo de gén o pseudogén.

3.3.4 Frecuencias de las configuraciones VJ de la cadena alpha y V(D)J de la cadena beta del TCR de *Z. californianus*.

La cadena alpha tuvo 85 configuraciones VJ (Figura 31-A), con 160 lecturas y tres configuraciones VDJ de la cadena Beta, y 100 lecturas totales (Figura 31-B).

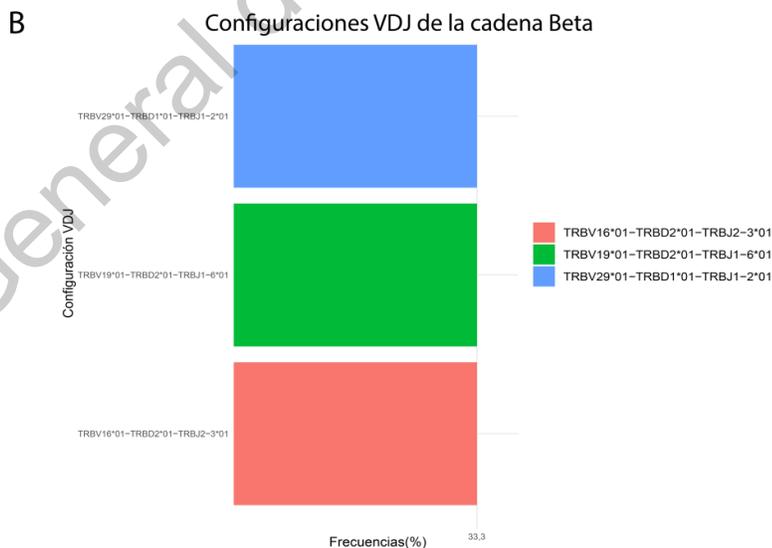
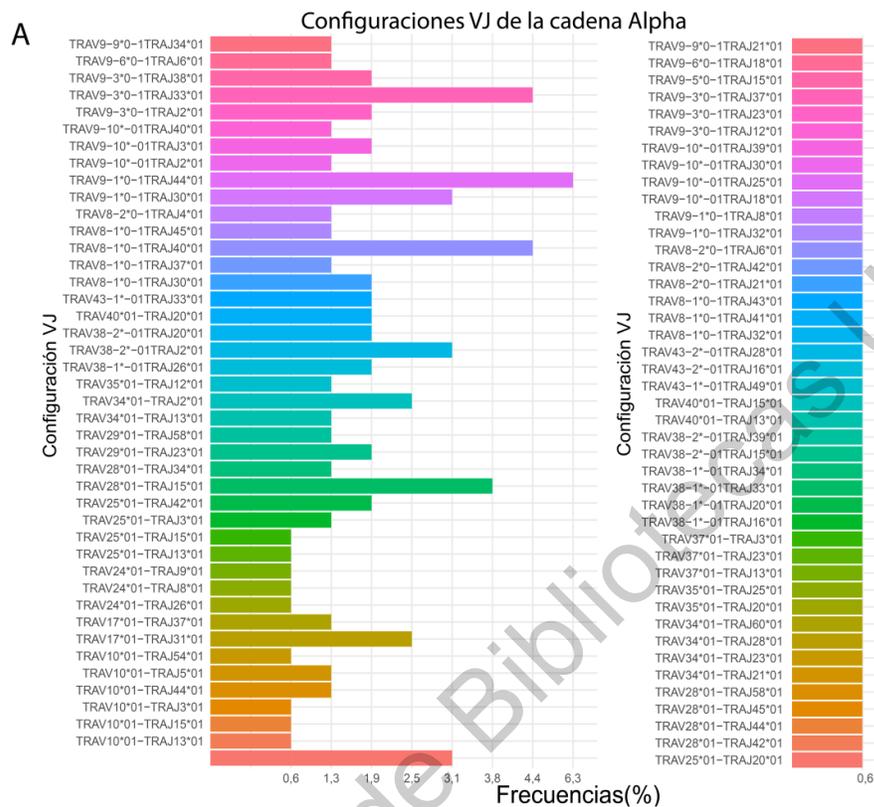


Figura 31| Configuraciones VJ de la cadena alpha y configuraciones VDJ de la cadena beta del TCR de *Z. californianus*. En A, frecuencias individuales de configuraciones VJ. B, frecuencias individuales de configuraciones VDJ. Los valores de frecuencia están expresados en porcentajes, basados en el número total de lecturas para el total de configuraciones VJ y VDJ.

3.3.5 Frecuencias de los segmentos V(D)J de la cadena Heavy del BCR de *Z. californianus*

Para la cadena heavy, se registraron 37 genes V funcionales, siete pseudogenes y dos ORFs, con un total de lecturas de 214, 72 y 3, respectivamente (Figura 32). De los genes D cinco fueron funcionales y se encontró un ORF, con un total de 90 lecturas para los genes funcionales y 10 para el ORF (Figura 33-C). En esta cadena se caracterizó un ORF y cinco genes J funcionales con 23 y 13 lecturas, respectivamente (Figura 33 A-B).

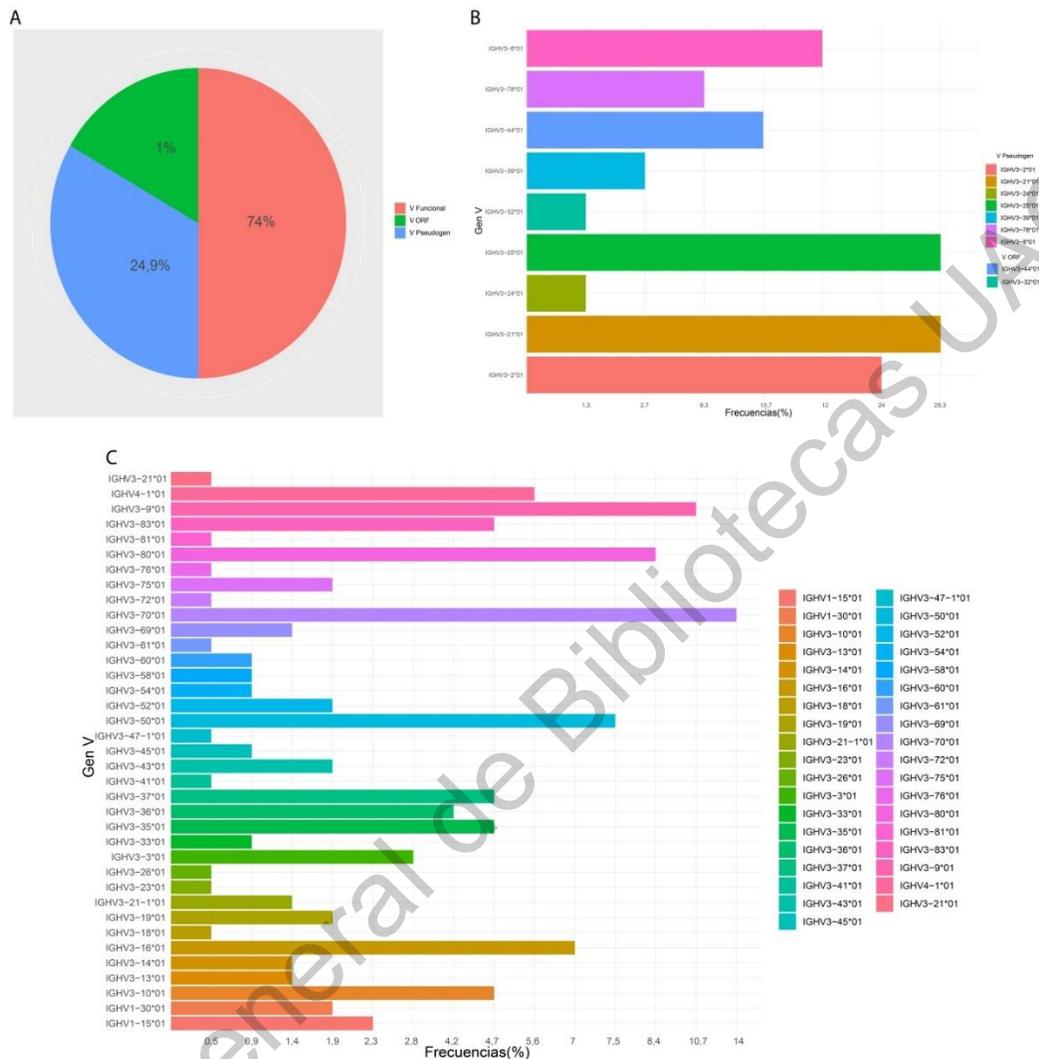


Figura 32| Caracterización de los segmentos génicos V de la cadena heavy del BCR (IG) de *Z. californianus*. A) Porcentajes de genes V funcionales, pseudogenes y ORF. B) Frecuencias individuales de los pseudogenes y ORF IGHV. C, frecuencias individuales de genes IGHV funcionales. Los valores de frecuencia están expresados en porcentajes, basados en el número total de lecturas para cada grupo de gén o pseudogén.

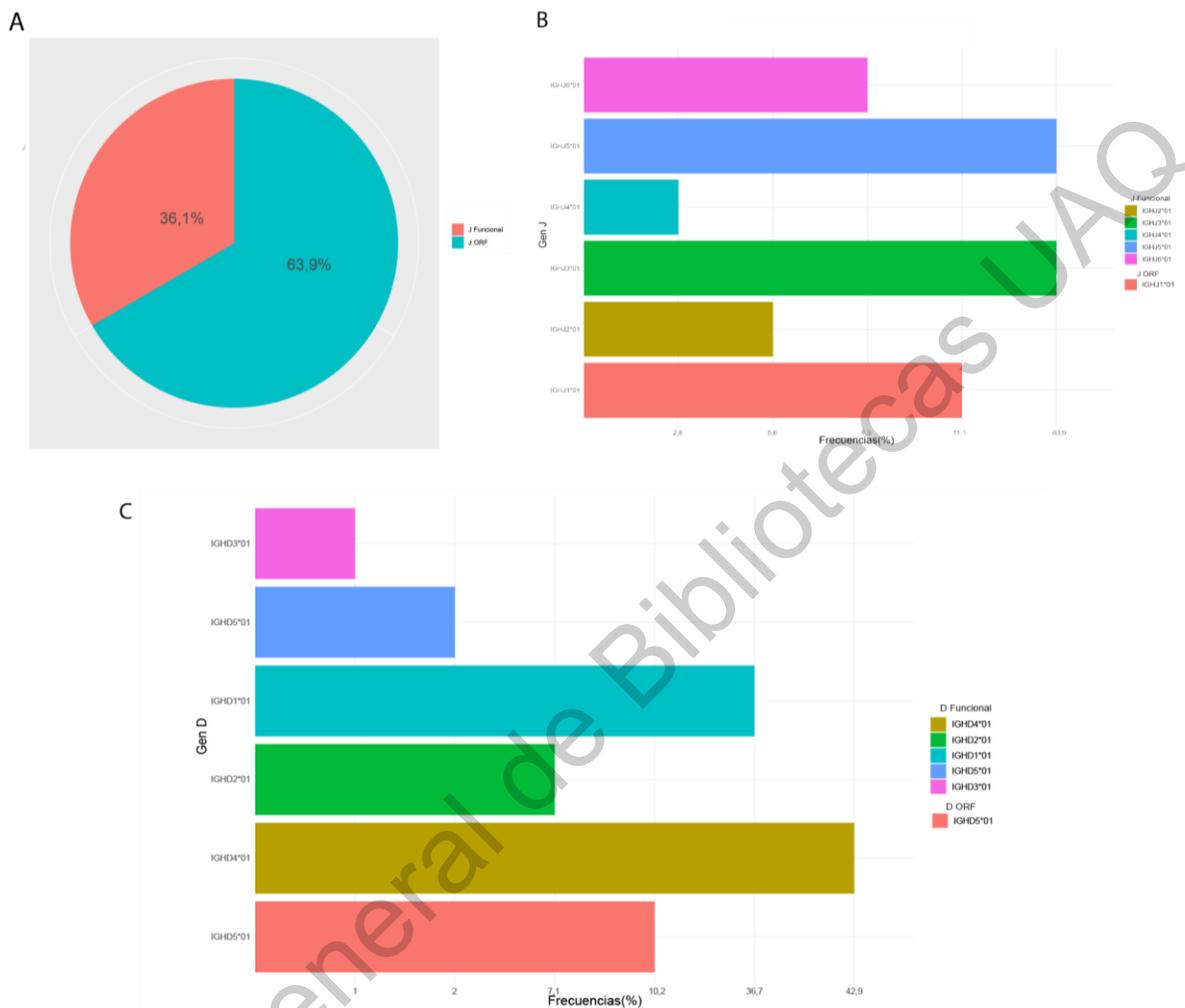


Figura 33| Caracterización de los segmentos génicos D y J de la cadena heavy del BCR (IG) de *Z. californianus*. A) Porcentajes de genes J funcionales y pseudogenes. B) Frecuencias individuales de los genes funcionales y ORF IGHJ. C, frecuencias individuales de genes funcionales y ORF IGHD. Los valores de frecuencia están expresados en porcentajes, basados en el número total de lecturas para cada grupo de gén o pseudogén.

3.3.6 Frecuencias de los segmentos VJ de la cadena lambda del BCR Z. californianus.

En la cadena lambda del BCR, se encontraron 28 genes V funcionales, cuatro pseudogenes y cuatro ORFs con las lecturas correspondientes de 79, 6 y 4, respectivamente (Figura 34). En los genes J se encontraron seis genes funcionales con 40 lecturas (Figura 35).

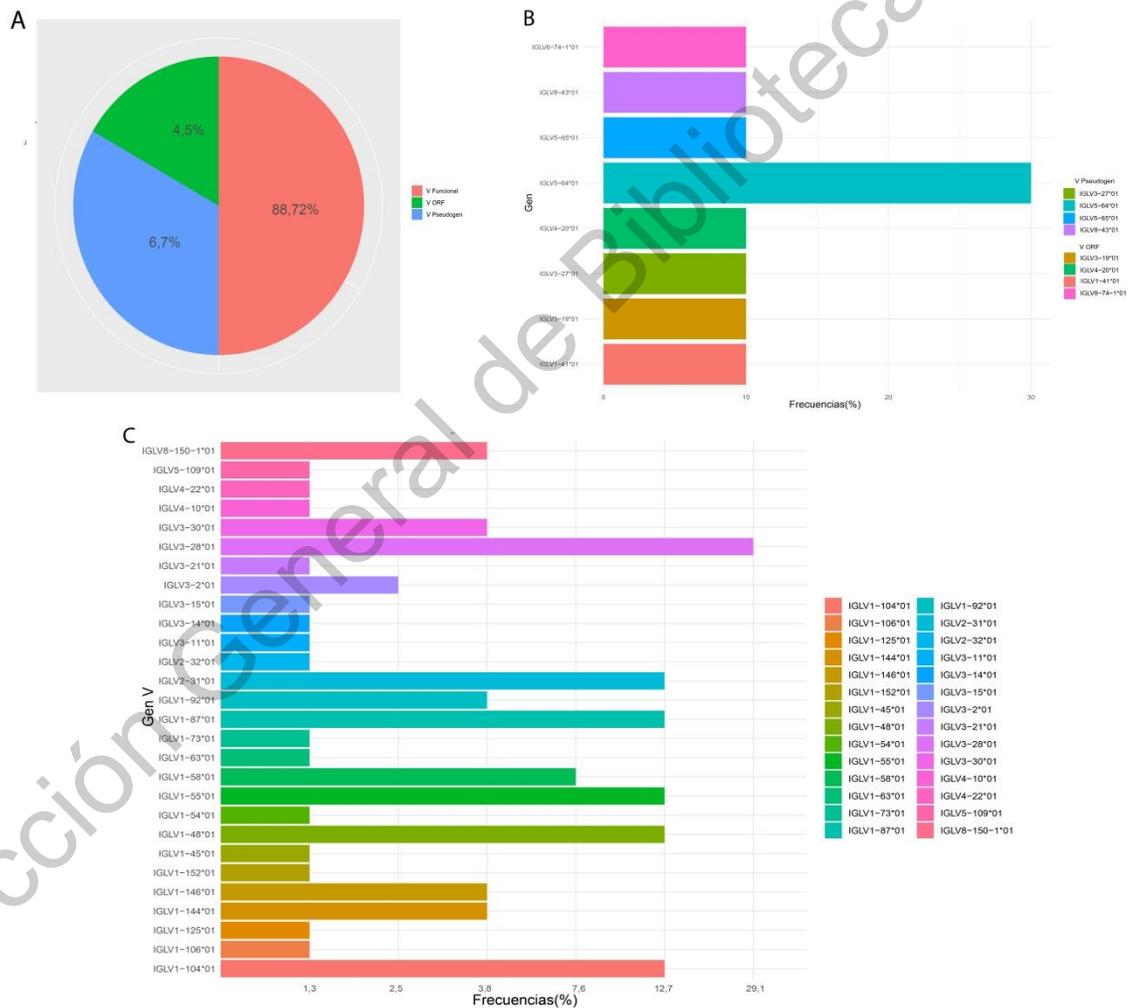


Figura 34| Caracterización de los segmentos génicos V de la cadena lambda del BCR (IG) de *Z. californianus*. A) Porcentajes de genes V funcionales, pseudogenes y ORF. B) Frecuencias individuales de los pseudogenes y ORF IGL. C, frecuencias individuales de genes IGLV funcionales. Los valores de frecuencia están expresados en porcentajes, basados en el número total de lecturas para cada grupo de gén o pseudogén.

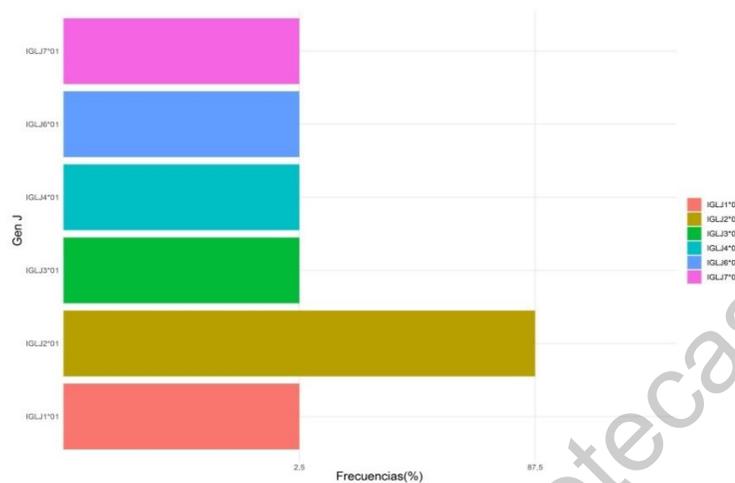


Figura 35| Caracterización de los segmentos génicos J de la cadena lambda del BCR (IG) de *Z. californianus*. Frecuencias de genes J funcionales de la cadena Lambda. Los valores de frecuencia están expresados en porcentajes, basados en el número total de lecturas para cada grupo de gen o pseudogen.

3.3.7 Frecuencias de las configuraciones VJ de la cadena Lambda y V(D)J de la cadena heavy del BCR de *Z. californianus*.

En la cadena lambda, se encontraron nueve configuraciones VJ, con 38 lecturas (Figura 36), mientras que en la cadena heavy, se obtuvieron 15 configuraciones V(D)J con 36 lecturas (Figura 37).

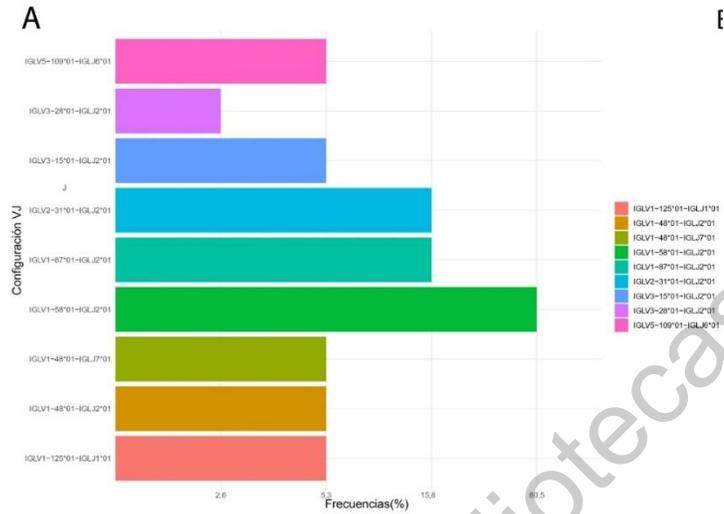


Figura 36| Configuraciones VJ de la cadena lambda del BCR (IG) de *Z. californianus*. Frecuencia de las configuraciones VJ de la cadena Lambda. Los valores de frecuencia están expresados en porcentajes, basados en el número total de lecturas para el total de configuraciones VJ.

Dirección General de Bibliotecas UAQ

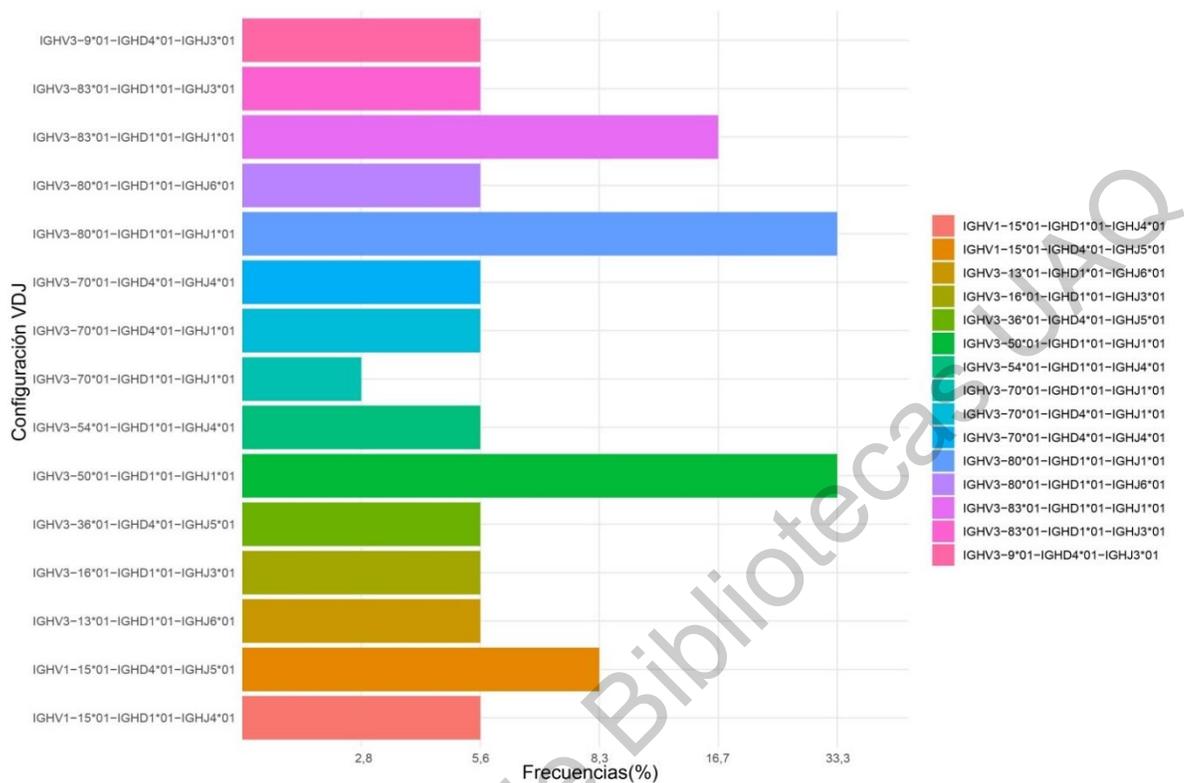


Figura 37| Configuraciones VDJ de la cadena heavy del BCR (IG) de *Z. californianus*. Frecuencia de las configuraciones VDJ de la cadena heavy. Los valores de frecuencia están expresados en porcentajes, basados en el número total de lecturas para el total de configuraciones VDJ.

3.3.8 Número de segmentos V(D)J de los receptores TCR y BCR(Ig) a través de los taxones

Con el objetivo de comparar el repertorio de los segmentos génicos V(D)J de los receptores del lobo marino de California, se analizó el repertorio de un amplio número de especies pertenecientes a la orden Carnívora y de otros mamíferos marinos.

Se analizaron 19 especies, incluyendo a *Z. californianus* y basados en el análisis de ontogenia descrito previamente se determinaron los segmentos de las cadenas del TCR y BCR (Figura 38-41).

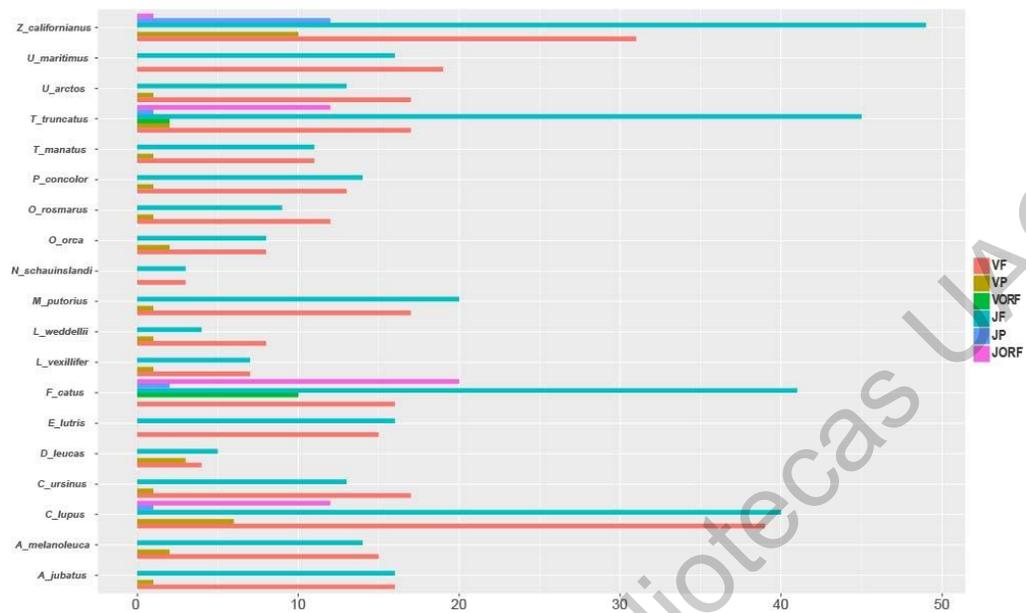


Figura 38| Número de segmentos génicos VJ de la cadena alpha del TCR de mamíferos incluidos en este estudio. En color salmón, se muestran los genes V funcionales en mostaza, los pseudogenes V y en verde los ORFs V. En azul celeste se muestran los genes J funcionales, en azul rey los pseudogenes J y en rosa los ORFs J.



Figura 39| Número de segmentos génicos V(D)J de la cadena beta del TCR de mamíferos incluidos en este estudio. En color salmón, se muestran los genes V funcionales en mostaza, los pseudogenes V y en verde los ORFs V. En azul celeste se muestran los genes J funcionales, en azul rey los pseudogenes J, en lila los ORFs J y en rosa los genes D.

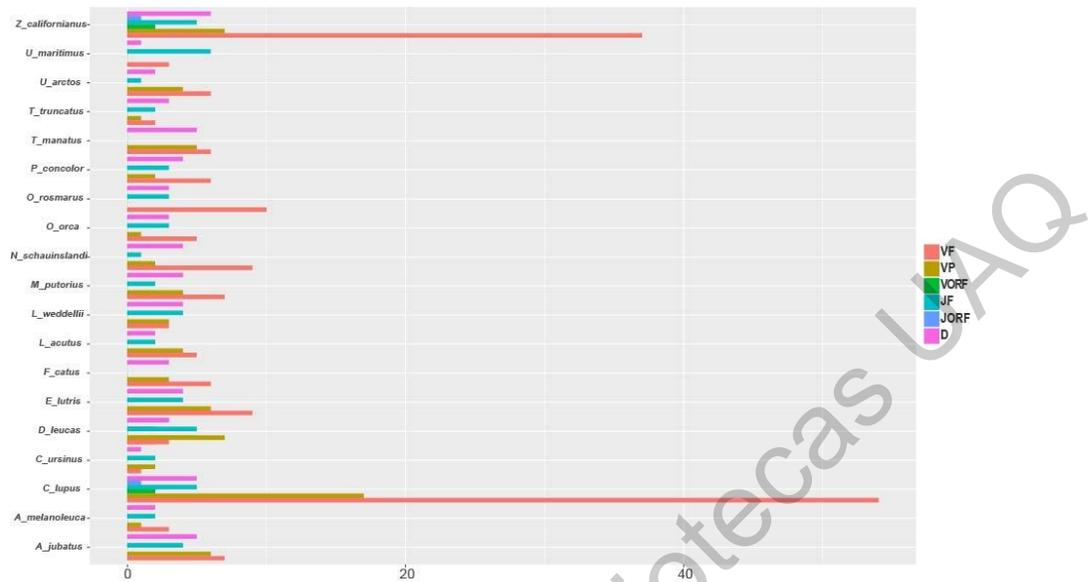


Figura 40| Numero de segmentos génicos V(D)J de la cadena Heavy del BCR de mamíferos incluidos en este estudio. En salmón, se muestran los genes V funcionales en mostaza, los pseudogenes V y en verde los ORFs V. En azul celeste los genes J funcionales, en lila los ORFs J y en rosa los genes D.

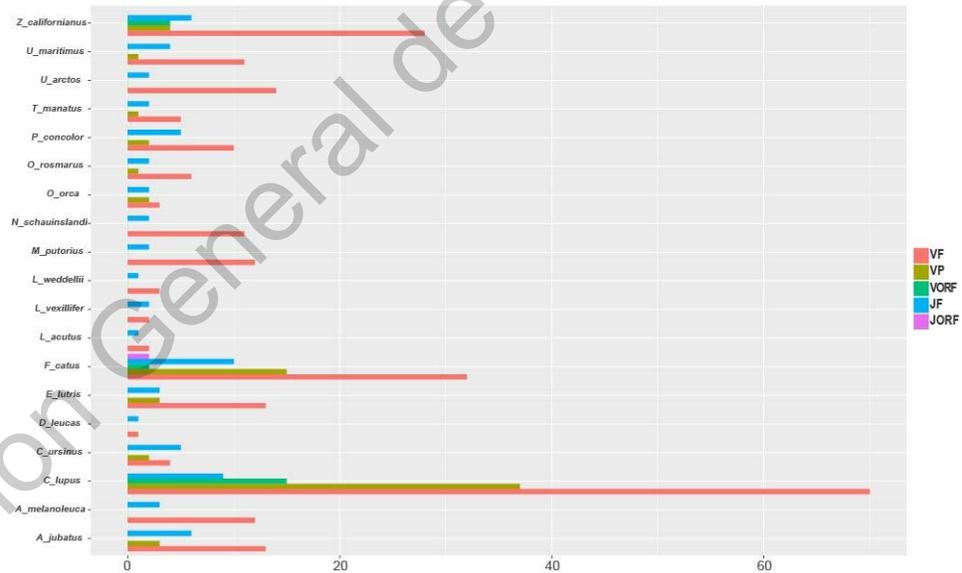


Figura 41| Configuraciones VDJ de la cadena lambda del BCR (IG) de *Z. californianus*. En color salmón se muestran los genes V funcionales en mostaza, los pseudogenes V y en verde los ORFs V. En azul celeste se muestran los genes J funcionales y en rosa los ORFs J.

3.4 Evolución de los segmentos génicos V(D)J de las cadenas del TCR y BCR de lobo marino de California (*Zalophus californianus*)

3.4.1 Evolución de los segmentos V de las cadenas alpha (TRAV) de carnívoros y mamíferos marinos

El árbol de mamíferos carnívoros incluyó 239 secuencias nucleótidas. El análisis filogenético de los segmentos variables de la cadena alpha, revela que estos segmentos de *Z. californianus* en su mayoría se agrupan individualmente a excepción de los grupos TRAV29, TRAV35, TRAV37, TRAV38 y TRAV40 (Figura 42), los cuales están conservados entre las especies de mamíferos indistintamente de su orden filogenético de procedencia (Figura 42-43).

El árbol filogenético de las secuencias de diversos mamíferos marinos incluyó 140 secuencias nucleótidas. A partir de este análisis se observó que los genes TRAV de *Z. californianus* se agruparon monofiléticamente a diferencia de lo observado en el árbol filogenético de carnívoros (Figura 43).

En ambos árboles los segmentos TRAV43, TRAV9, TRAV3 y TRAV8 de *Z. californianus* se agruparon en clanes individuales (Figura 42-43).

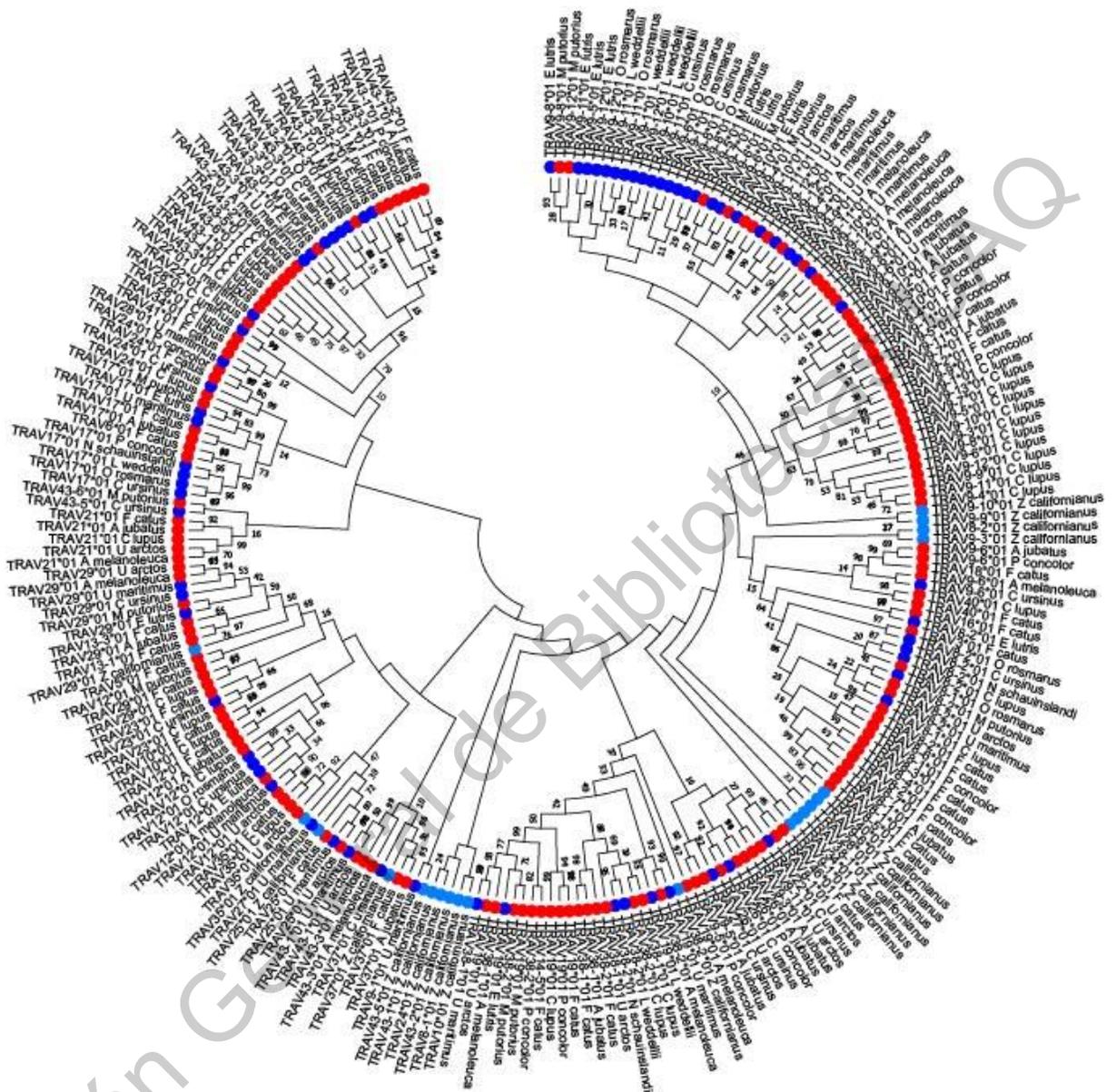


Figura 42| Análisis filogenético molecular de los segmentos TRAV de mamíferos carnívoros. Se muestra el árbol consenso de 1000 réplicas de Bootstrap. Las posiciones incluidas corresponden a los codones 1st+2nd+3rd+Noncoding. Todas las posiciones con menos del 95% de cobertura se eliminaron. Menos del 5% de espacios y bases ambiguas fueron permitidos. Los puntos rojos, están indicando a los mamíferos carnívoros terrestres, los puntos azules están indicando a los mamíferos carnívoros marinos y los de color azul cielo están indicando secuencias de *Z. californianus*.

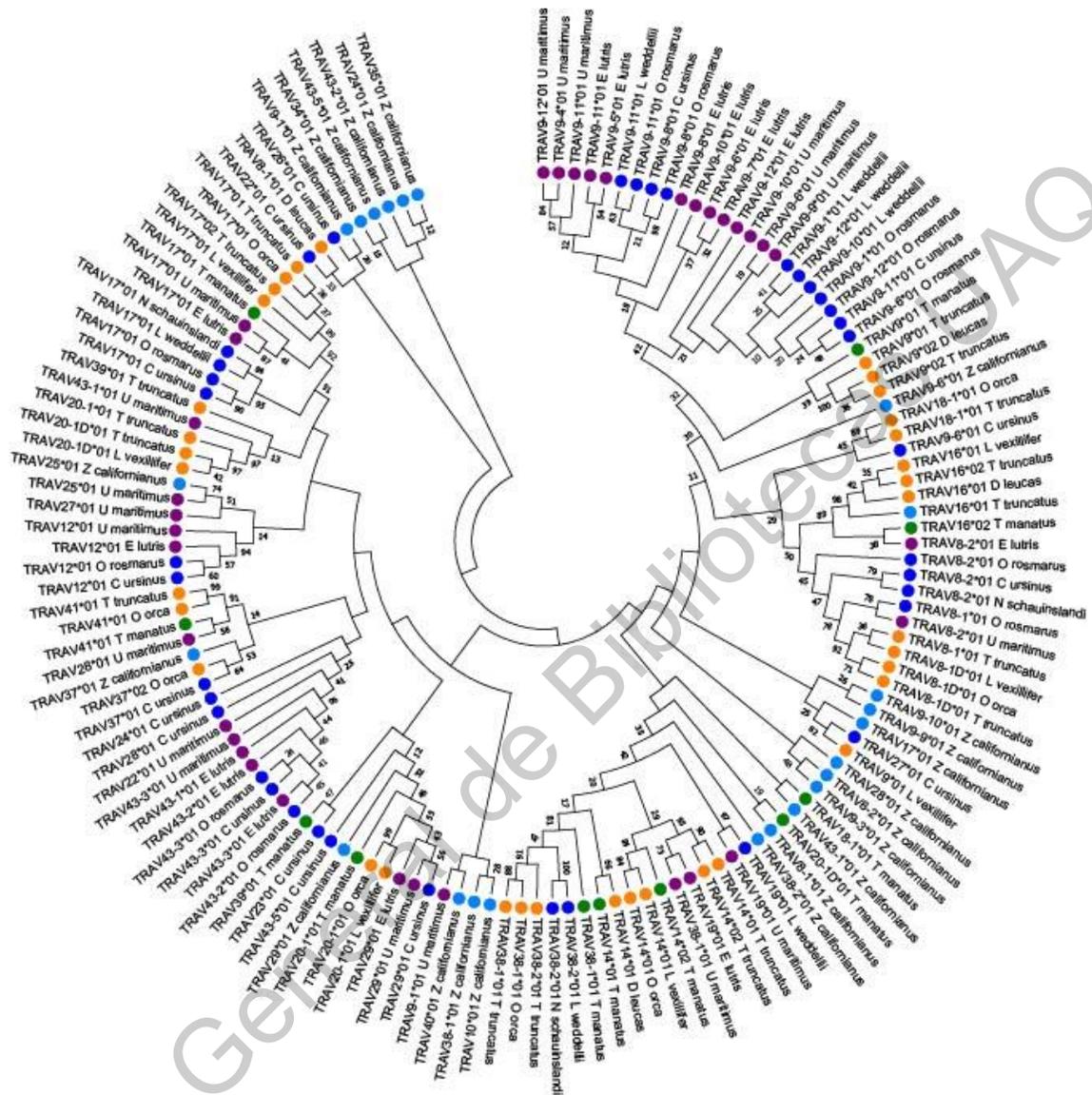


Figura 43| Análisis filogenético molecular de los segmentos TRAV de mamíferos marinos. Se muestra el árbol consenso de 1000 réplicas de Bootstrap. Las posiciones incluidas corresponden a los codones 1st+2nd+3rd+Noncoding. Todas las posiciones con menos del 95% de cobertura se eliminaron. Menos del 5% de espacios y bases ambiguas fueron permitidos. Los puntos azules, están indicando a los pinnípedos, los naranjas a cetáceos, los verdes a los sirenios, los morados al grupo de (otros mamíferos marinos) y los de color azul cielo están indicando secuencias de *Z. californianus*.

3.4.2 Evolución de los segmentos V de las cadenas beta (TRBV) de carnívoros y mamíferos marinos

El árbol filogenético de mamíferos carnívoros incluyó 160 secuencias nucleótidas, y en este análisis el gen TRBV19 de *Z. californianus* se mostró altamente conservado con respecto a los de otros carnívoros (Figura 44). En contraste, en el árbol filogenético de mamíferos marinos, el cual incluyó 80 secuencias nucleótidas, los clanes TRBV20, TRBV10, TRBV7 y TRBV4-1 mostraron más homología (Figura 45).

Dirección General de Bibliotecas UAO

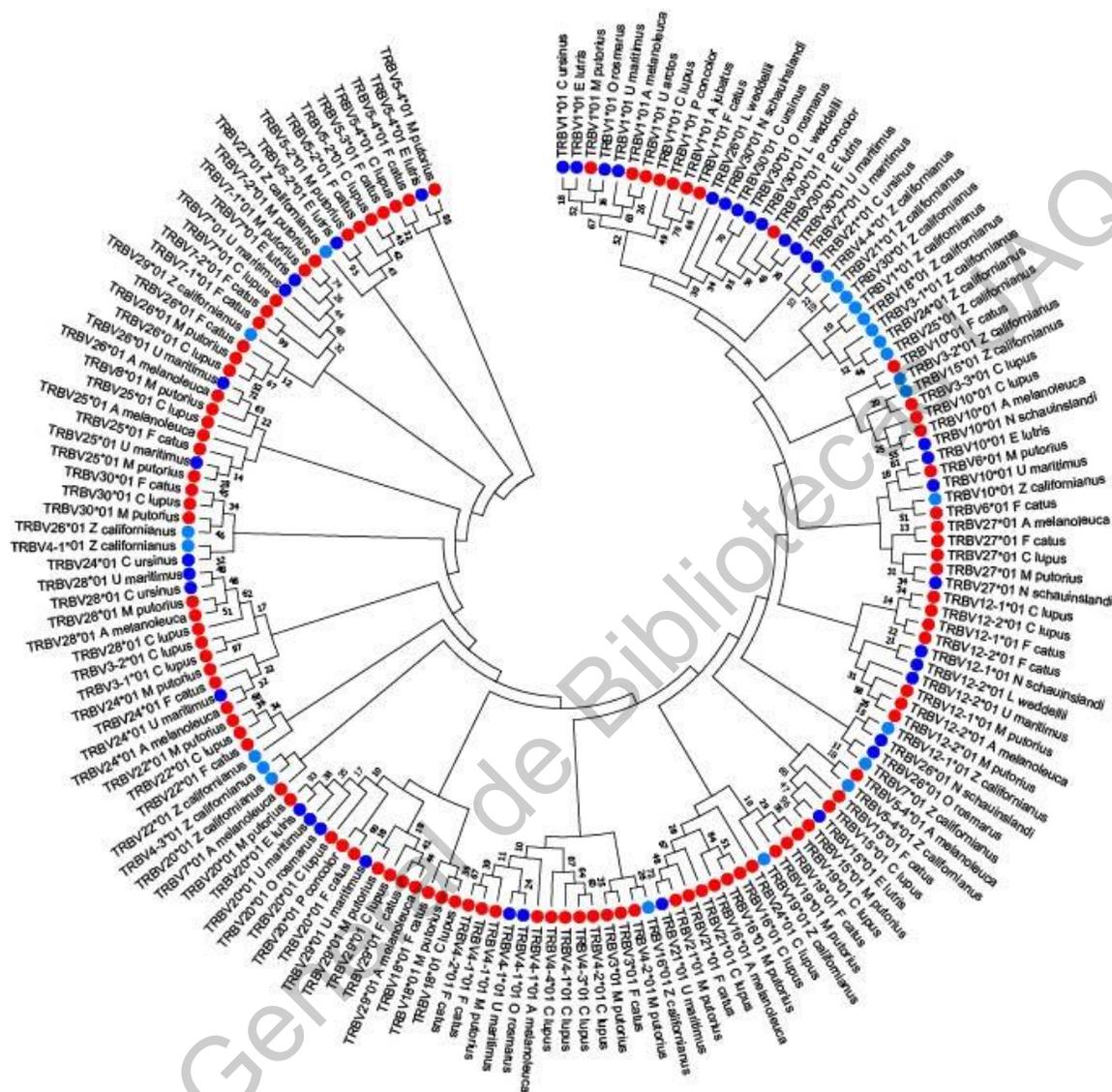


Figura 44| Análisis filogenético molecular de los segmentos TRBV de mamíferos carnívoros. Se muestra el árbol consenso de 1000 réplicas de Bootstrap. Las posiciones incluidas corresponden a los codones 1st+2nd+3rd+Noncoding. Todas las posiciones con menos del 95% de cobertura se eliminaron. Menos del 5% de espacios y bases ambiguas fueron permitidos. Los puntos rojos están indicando a los mamíferos carnívoros terrestres, los puntos azules están indicando a los mamíferos carnívoros marinos y los de color azul cielo están indicando secuencias de *Z. californianus*.

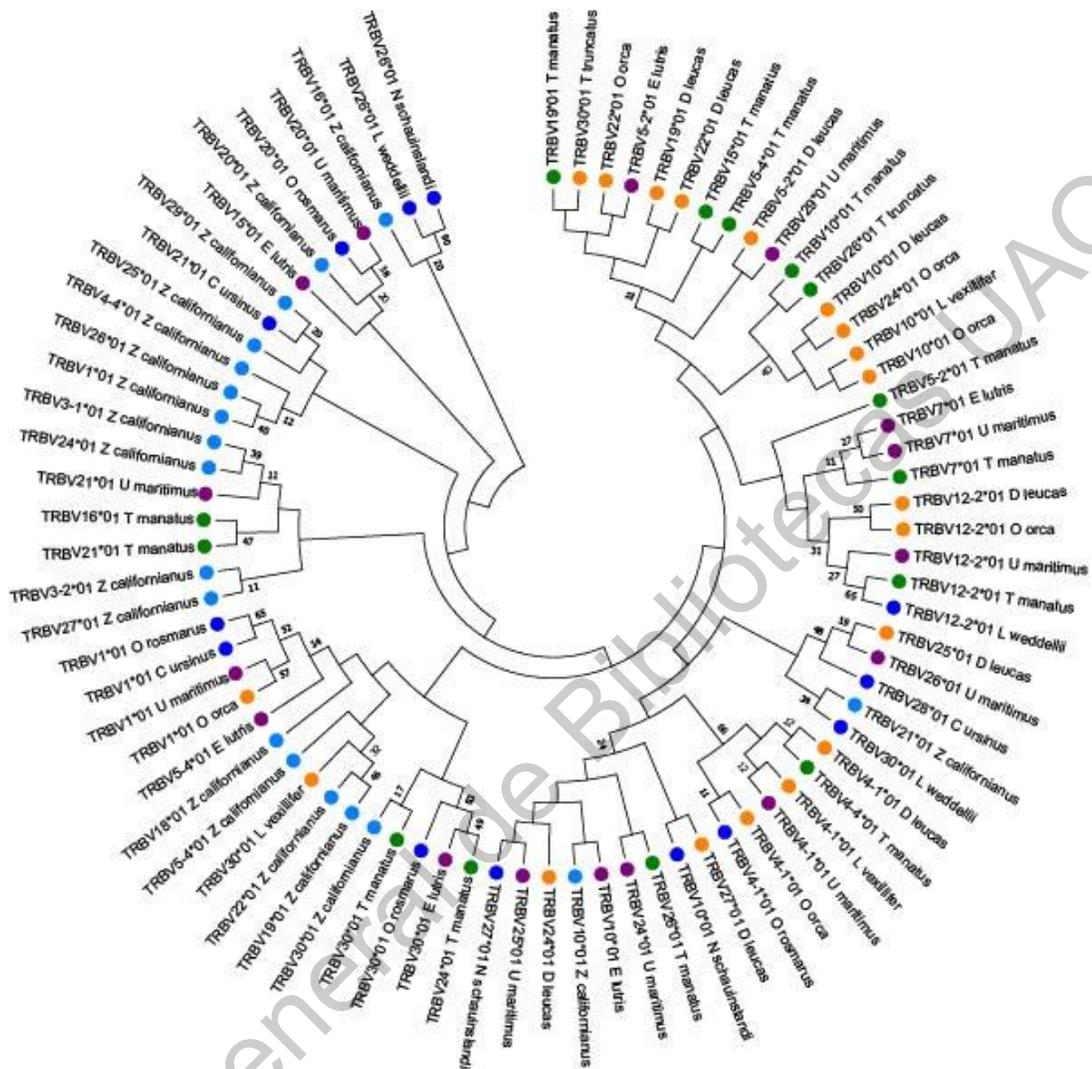


Figura 45| Análisis filogenético molecular de los segmentos TRBV de mamíferos marinos. Se muestra el árbol consenso de 1000 réplicas de Bootstrap. Las posiciones incluidas corresponden a los codones 1st+2nd+3rd+Noncoding. Todas las posiciones con menos del 95% de cobertura se eliminaron. Menos del 5% de espacios y bases ambiguas fueron permitidos. Los puntos azules están indicando a los pinnípedos, los naranjas a cetáceos, los verdes a los sirenios, los morados al grupo de (otros mamíferos marinos) y los de color azul cielo están indicando secuencias de *Z. californianus*.

3.4.3 Evolución de los segmentos V de las cadenas heavy (IGHV) de carnívoros y mamíferos marinos

El árbol filogenético de los mamíferos carnívoros incluyó 159 secuencias nucleótidas, y la mayoría de los segmentos IGHV de *Z. californianus* no mostraron homología con miembros del repertorio IGHV de carnívoros marinos o terrestres, a excepción de los genes IGHV4-1, IGHV3-81 e IGHV1-15 (Figura 46). Por otro lado, el árbol filogenético de mamíferos marinos que incluyó 53 secuencias nucleótidas mostró que los miembros IGHV de *Z. californianus* son más homólogos a genes IGHV de otras especies y tienen más ortólogos entre las secuencias de otros mamíferos marinos (Figura 47).

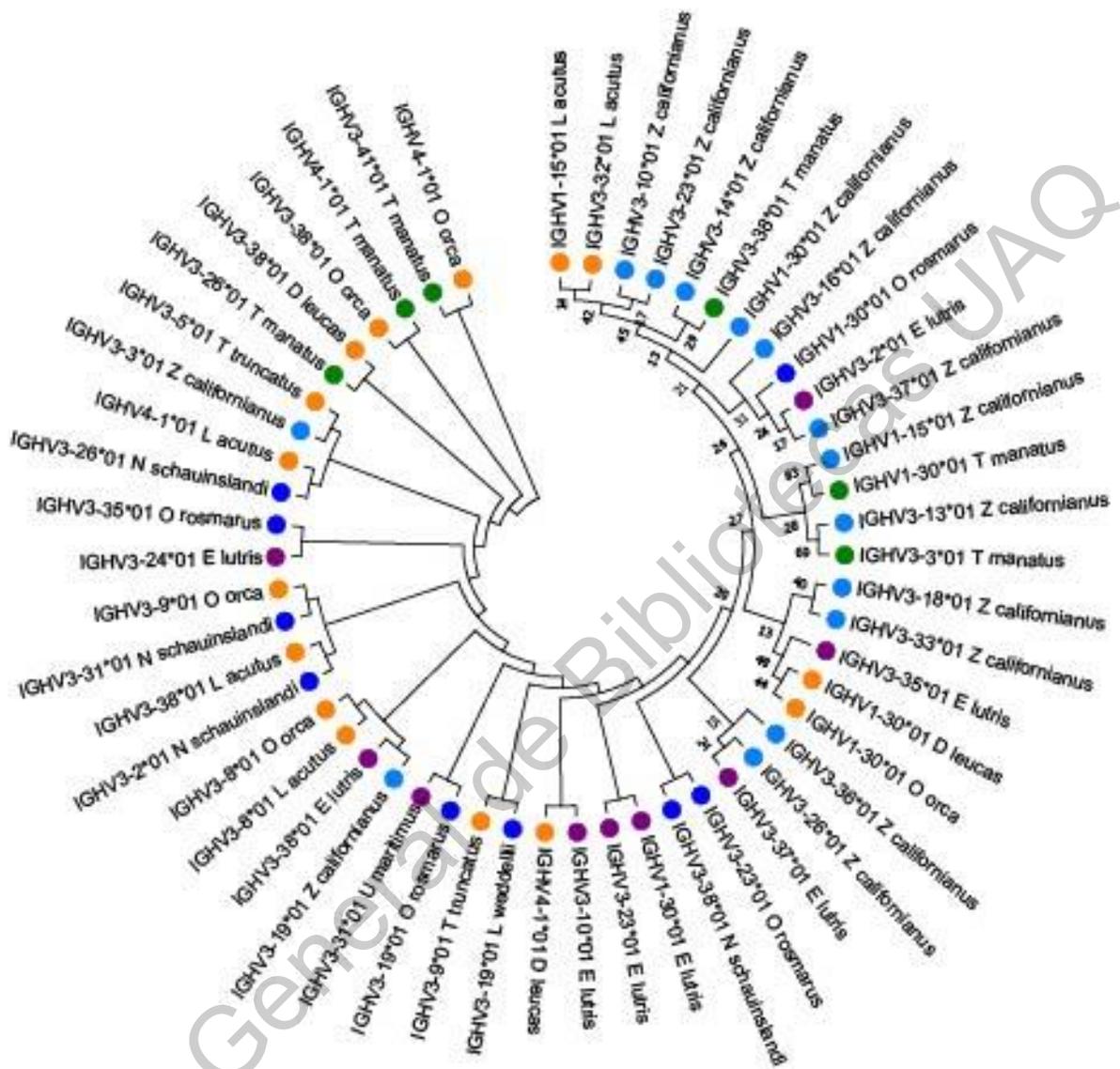


Figura 47 | Análisis filogenético molecular de los segmentos IGHV de mamíferos marinos. Se muestra el árbol consenso de 1000 réplicas de Bootstrap. Las posiciones incluidas corresponden a los codones 1st+2nd+3rd+Noncoding. Todas las posiciones con menos del 95% de cobertura se eliminaron. Menos del 5% de espacios y bases ambiguas fueron permitidos. Los puntos azules están indicando a los pinnípedos, los naranjas a cetáceos, los verdes a los sirenios, los morados al grupo de (otros mamíferos marinos) y los de color azul cielo están indicando secuencias de *Z. californianus*.

3.4.4 Evolución de los segmentos V de las cadenas heavy (IGLV) de carnívoros y mamíferos marinos

En el árbol filogenético de los mamíferos carnívoros, que incluyó 239 secuencias nucleótídicas, se observó que los segmentos IGLV de *Z. californianus* en su mayoría pertenecían a grupos polifiléticos de carnívoros terrestres y acuáticos, aunque principalmente acuáticos (Figura 48). En el árbol filogenético de mamíferos marinos que incluyó 88 secuencias nucleótídicas, los segmentos IGLV de *Z. californianus* se agruparon de nuevo con ortólogos de distintas especies. Además, se observó un grupo monofilético de estos segmentos con respecto a otros mamíferos marinos (Figura 49).

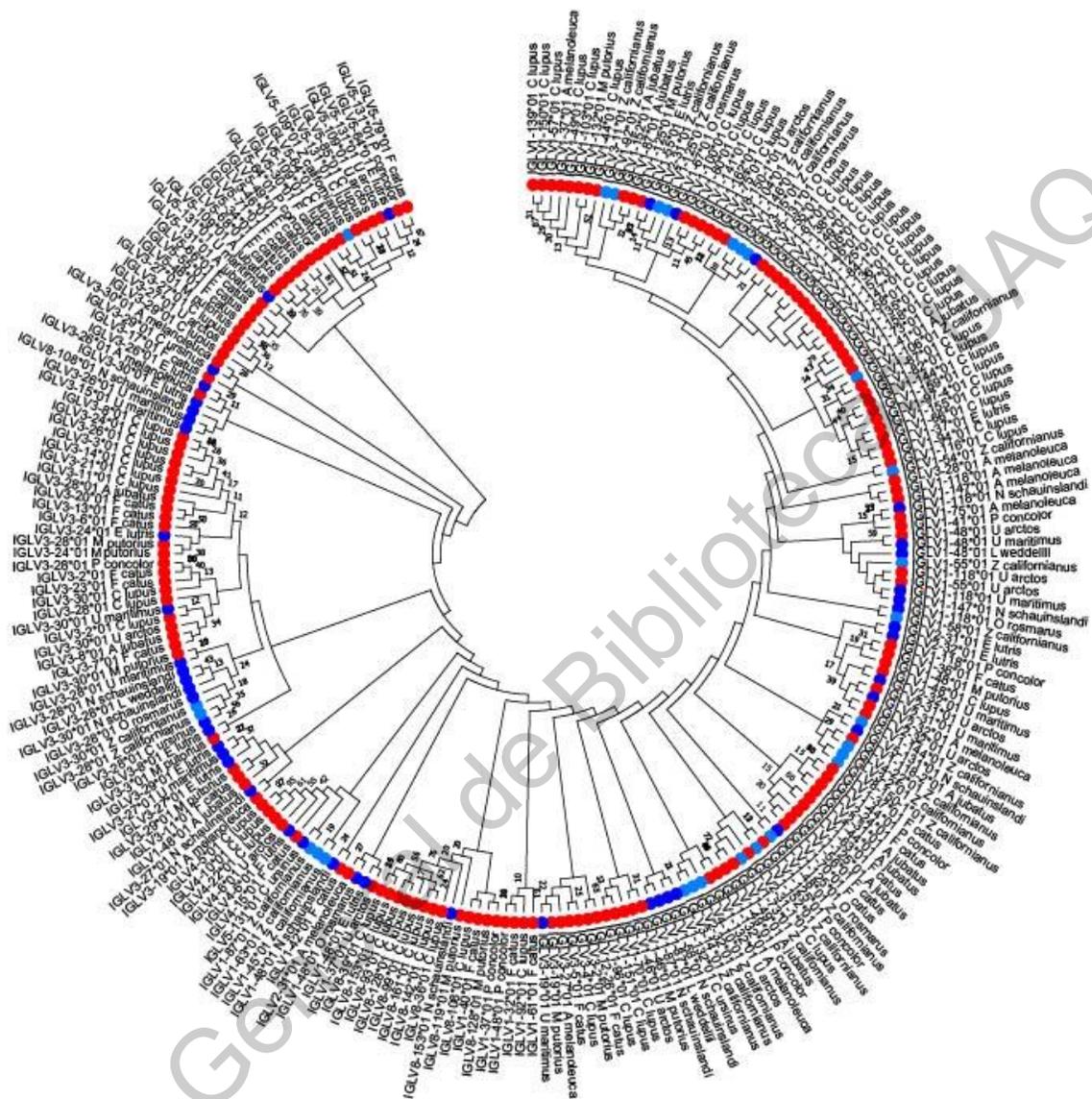


Figura 48| Análisis filogenético molecular de los segmentos IGLV de mamíferos carnívoros. Se muestra el árbol consenso de 1000 réplicas de Bootstrap. Las posiciones incluidas corresponden a los codones 1st+2nd+3rd+Noncoding. Todas las posiciones con menos del 95% de cobertura se eliminaron. Menos del 5% de espacios y bases ambiguas fueron permitidos. Los puntos rojos están indicando a los mamíferos carnívoros terrestres, los puntos azules están indicando a los mamíferos carnívoros marinos y los de color azul cielo están indicando secuencias de *Z. californianus*.

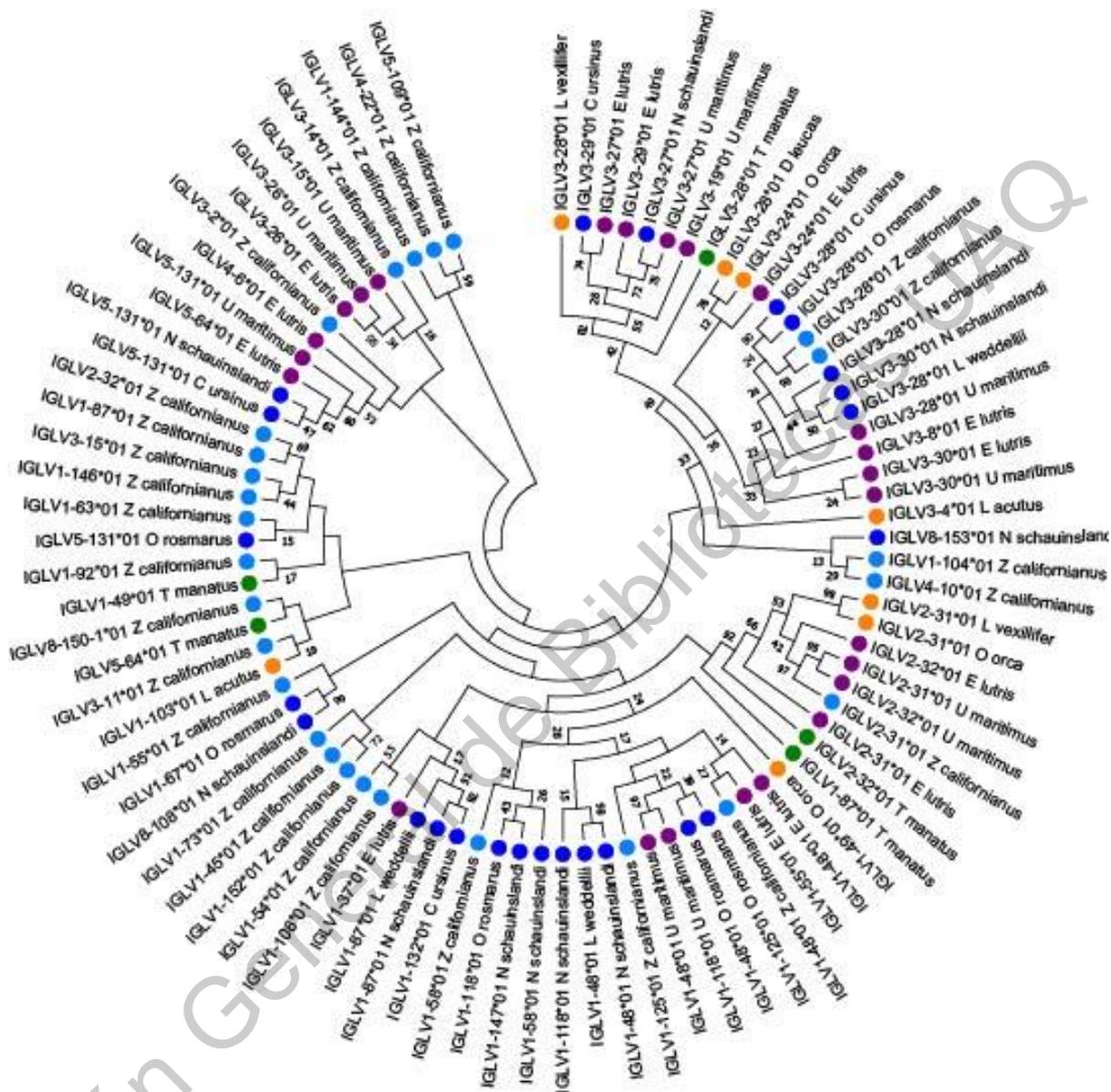


Figura 49| Análisis filogenético molecular de los segmentos IGLV de mamíferos marinos. Se muestra el árbol consenso de 1000 réplicas de Bootstrap. Las posiciones incluídas corresponden a los codones 1st+2nd+3rd+Noncoding. Todas las posiciones con menos del 95% de cobertura se eliminaron. Menos del 5% de espacios y bases ambiguas fueron permitidos. Los puntos azules están indicando a los pinnípedos, los naranjas a cetáceos, los verdes a los sirenios, los morados al grupo de (otros mamíferos marinos) y los de color azul cielo están indicando secuencias de *Z. californianus*.

3.4.5 Evolución de los segmentos J de las cadenas alpha (TRAJ) de carnívoros y mamíferos marinos

El árbol filogenético de los mamíferos carnívoros, que incluyó 263 secuencias nucleotídicas y el de los mamíferos marinos, que incluyó 88 secuencias nucleotídicas mostraron que los segmentos TRAJ de *Z. californianus* se agruparon con un amplio grado de homología en ambas reconstrucciones filogenéticas (Figura 50-51).

Dirección General de Bibliotecas UAG

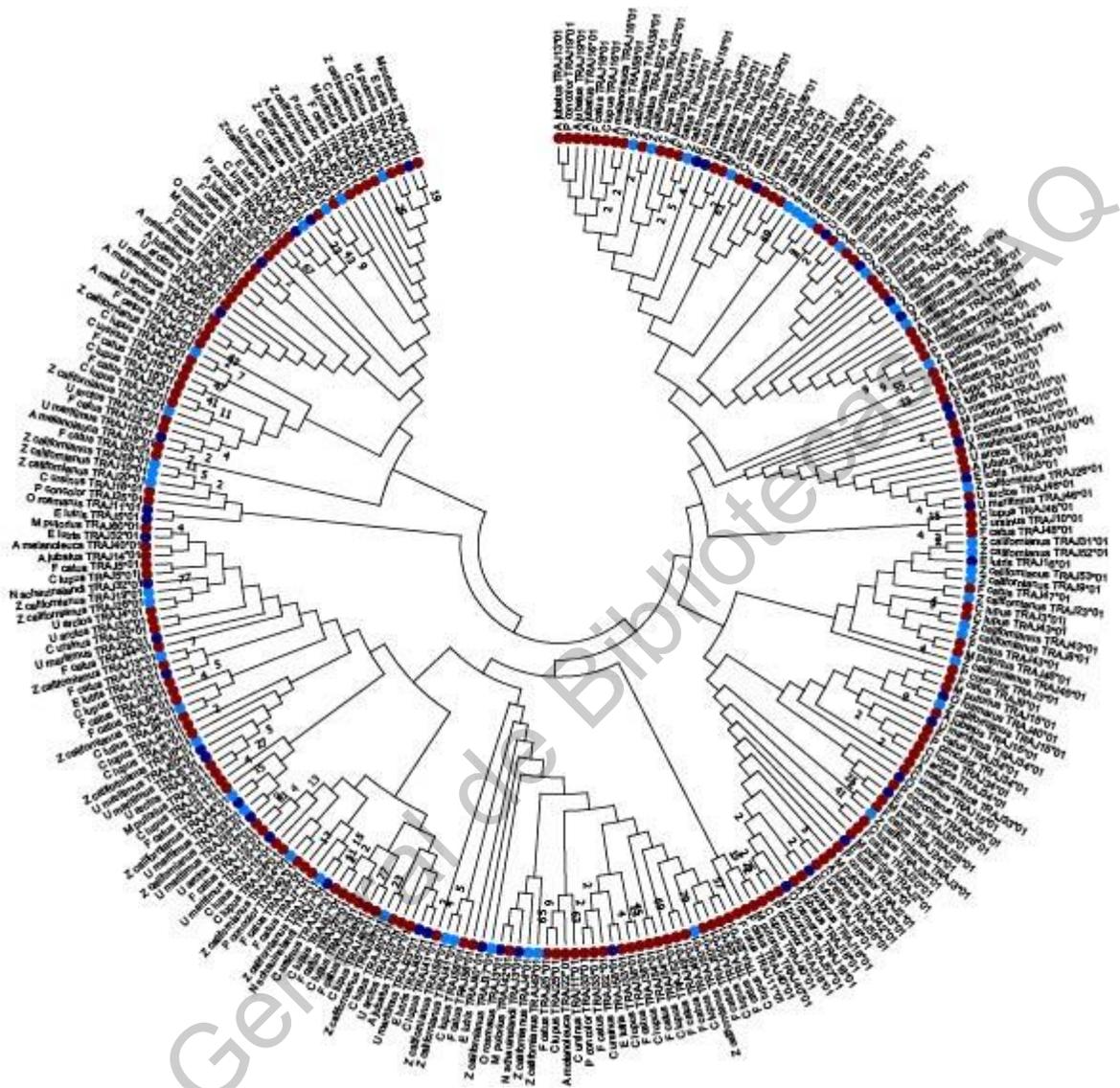


Figura 50| Análisis filogenético molecular de los segmentos TRAJ de mamíferos carnívoros. Se muestra el árbol consenso de 1000 réplicas de Bootstrap. Todas las posiciones con menos del 95% de cobertura se eliminaron. Menos del 5% de espacios y bases ambiguas fueron permitidos. Los puntos rojos están indicando a los mamíferos carnívoros terrestres, los puntos azules están indicando a los mamíferos carnívoros marinos y los de color azul cielo están indicando secuencias de *Z. californianus*.

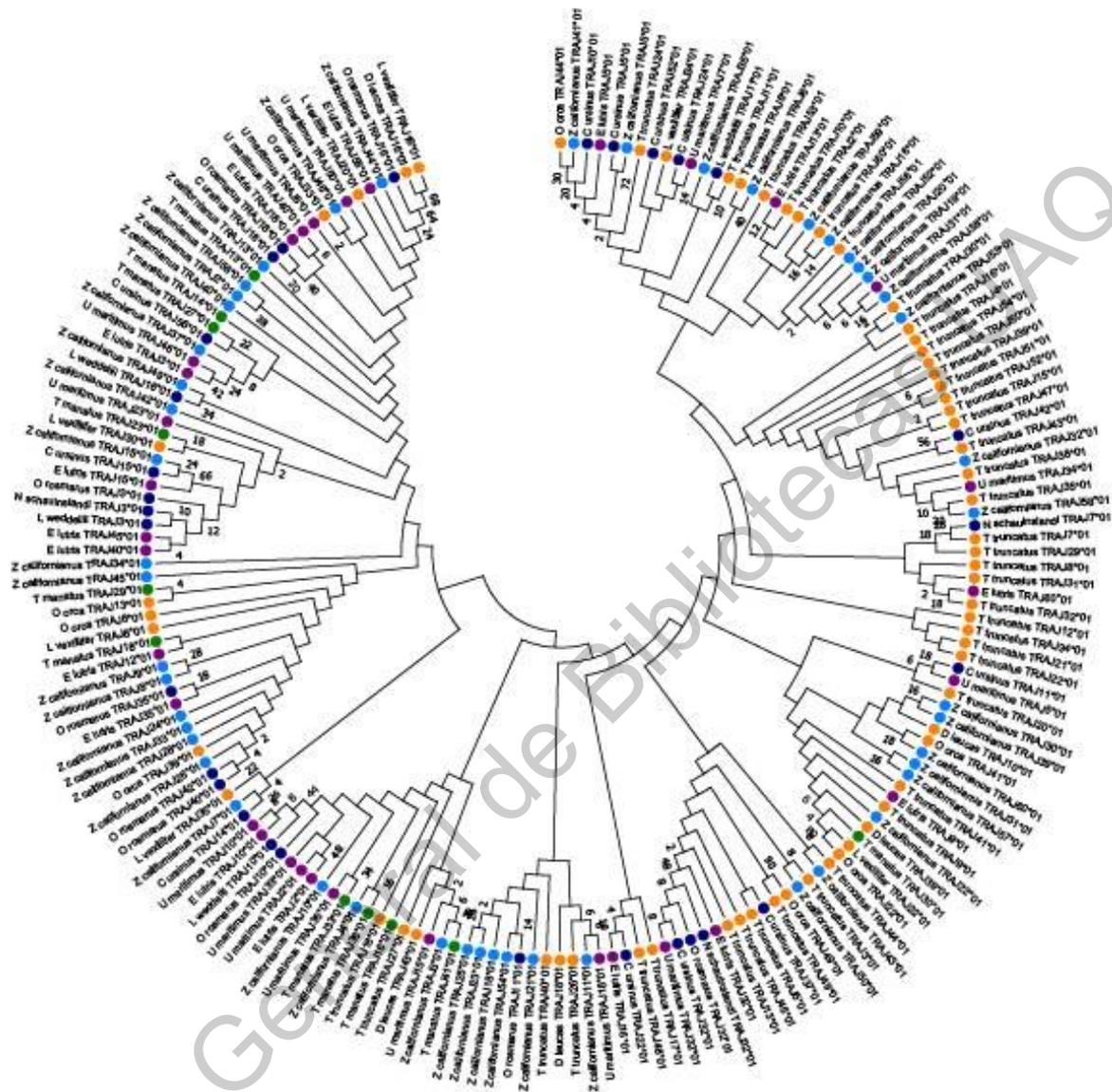


Figura 51| Análisis filogenético molecular de los segmentos TRAJ de mamíferos marinos. Se muestra el árbol consenso de 1000 réplicas de Bootstrap. Todas las posiciones con menos del 95% de cobertura se eliminaron. Menos del 5% de espacios y bases ambiguas fueron permitidos. Los puntos azules están indicando a los pinnípedos, los naranjas a cetáceos, los verdes a los sirenios, los morados al grupo de (otros mamíferos marinos) y los de color azul cielo están indicando secuencias de *Z. californianus*.

3.4.6 Evolución de los segmentos DJ de las cadenas beta (TRBJD) de carnívoros y mamíferos marinos

Durante la reconstrucción filogenética de los genes TRBD de *Z. californianus* mostraron tener dos posibles orígenes, con base a nuestro análisis entre carnívoros y mamíferos marinos (Figura 52-53). Este análisis incluyó 60 (Figura 52) y 48 secuencias (Figura 53).

Al reconstruir la historia evolutiva de los genes TRBJ, mostraron que estos tienen homólogos entre las especies de mamíferos carnívoros, mientras encontramos menos homólogos de las familias TRBJ de *Z. californianus* con respecto a mamíferos marinos (Figura 54-55). En este análisis incluimos 21 (Figura 54) y 20 secuencias (Figura 55).

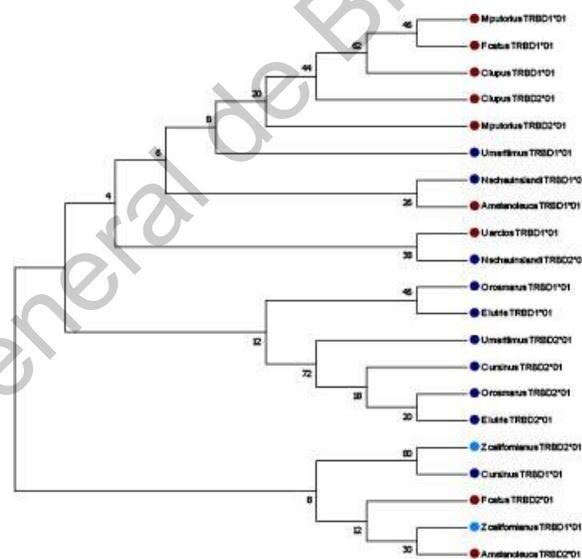


Figura 52| Análisis filogenético molecular de los segmentos TRBD de mamíferos carnívoros. Se muestra el árbol consenso de 1000 réplicas de Bootstrap. Todas las posiciones con menos del 95% de cobertura se eliminaron. Menos del 5% de espacios y bases ambiguas fueron permitidos. Los puntos rojos están indicando a los mamíferos carnívoros terrestres, los puntos azules están indicando a los mamíferos carnívoros marinos y los de color azul cielo están indicando secuencias de *Z. californianus*.

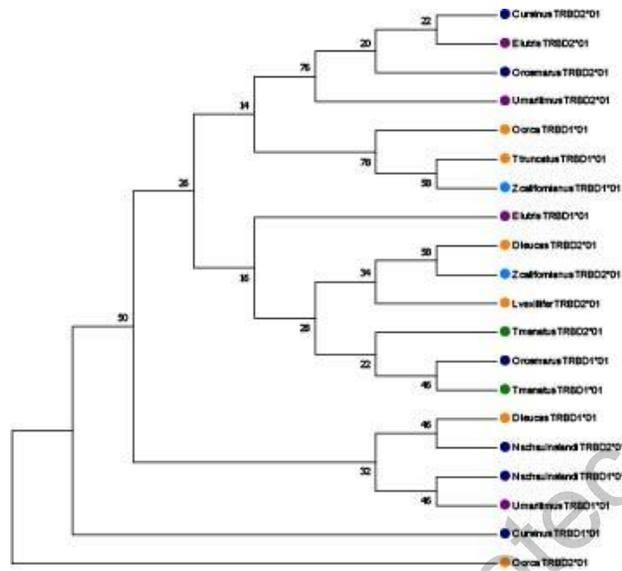


Figura 53| Análisis filogenético molecular de los segmentos TRBD de mamíferos marinos. Se muestra el árbol consenso de 1000 réplicas de Bootstrap. Todas las posiciones con menos del 95% de cobertura se eliminaron. Menos del 5% de espacios y bases ambiguas fueron permitidos. Los puntos azules están indicando a los pinnípedos, los naranjas a cetáceos, los verdes a los sirenios, los morados al grupo de (otros mamíferos marinos) y los de color azul cielo están indicando secuencias de *Z. californianus*.

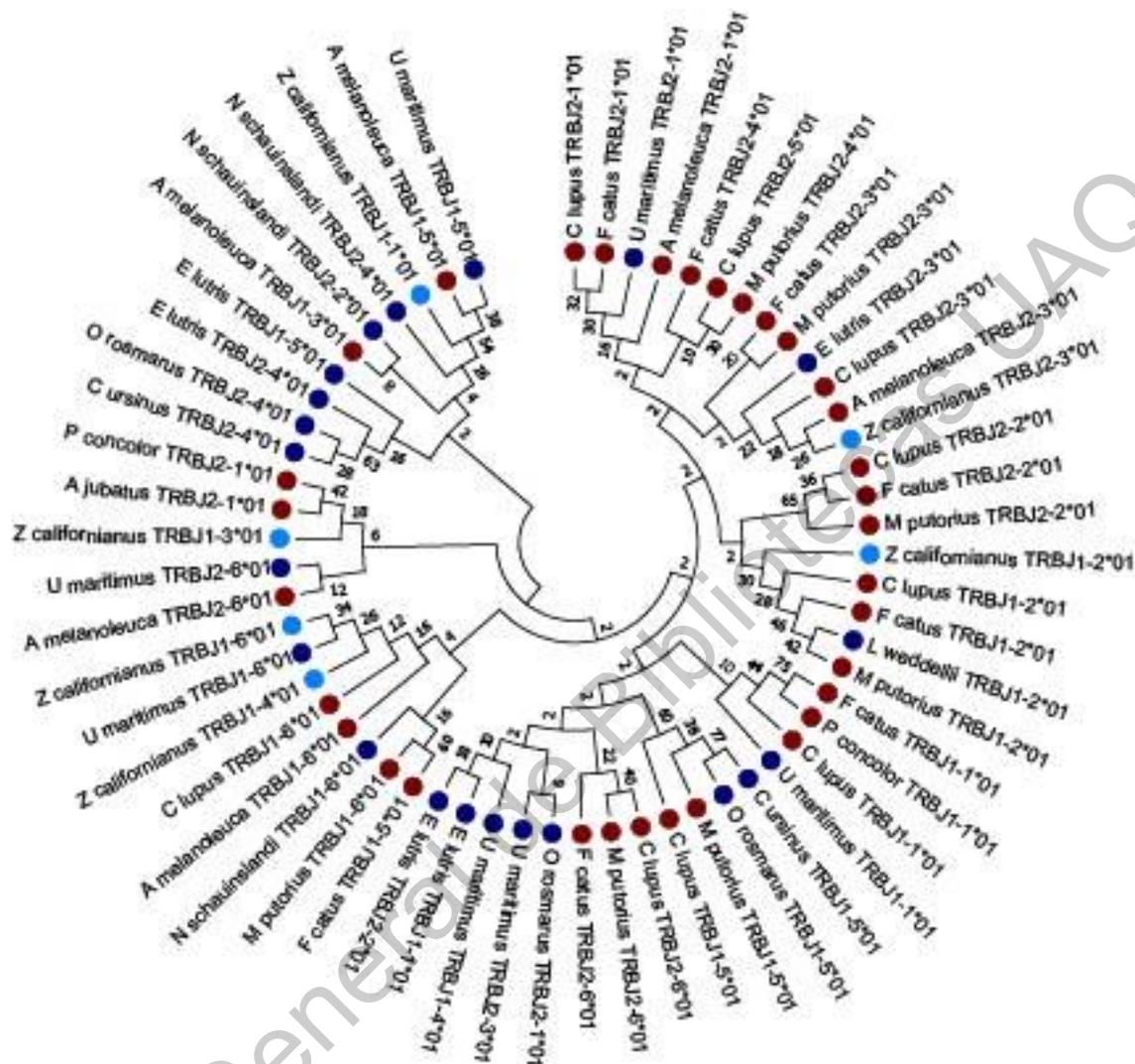


Figura 54| Análisis filogenético molecular de los segmentos TRBJ de mamíferos carnívoros. Se muestra el árbol consenso de 1000 réplicas de Bootstrap. Todas las posiciones con menos del 95% de cobertura se eliminaron. Menos del 5% de espacios y bases ambiguas fueron permitidos. Los puntos rojos están indicando a los mamíferos carnívoros terrestres, los puntos azules están indicando a los mamíferos carnívoros marinos y los de color azul cielo están indicando secuencias de *Z. californianus*.

3.4.7 Evolución de los segmentos DJ de las cadenas heavy (IGHDJ) de carnívoros y mamíferos marinos

En este análisis la reconstrucción filogenética de carnívoros incluyeron 48 secuencias nucleotídicas de los segmentos IGHD (Figura 56) y la de mamíferos marinos, 41 secuencias (Figura 57). Para el análisis de los segmentos IGHJ fueron 46 (Figura 58) y 27 (Figura 59) secuencias, respectivamente. En el primer análisis los segmentos IGHD de *Z. californianus* mostraron homología con distintos grupos filogenéticos (Figura 56-57). Los genes IGHJ de *Z. californianus* por otro lado, parecieron tener un origen único entre especies de pinnípedos (Figura 58-59).

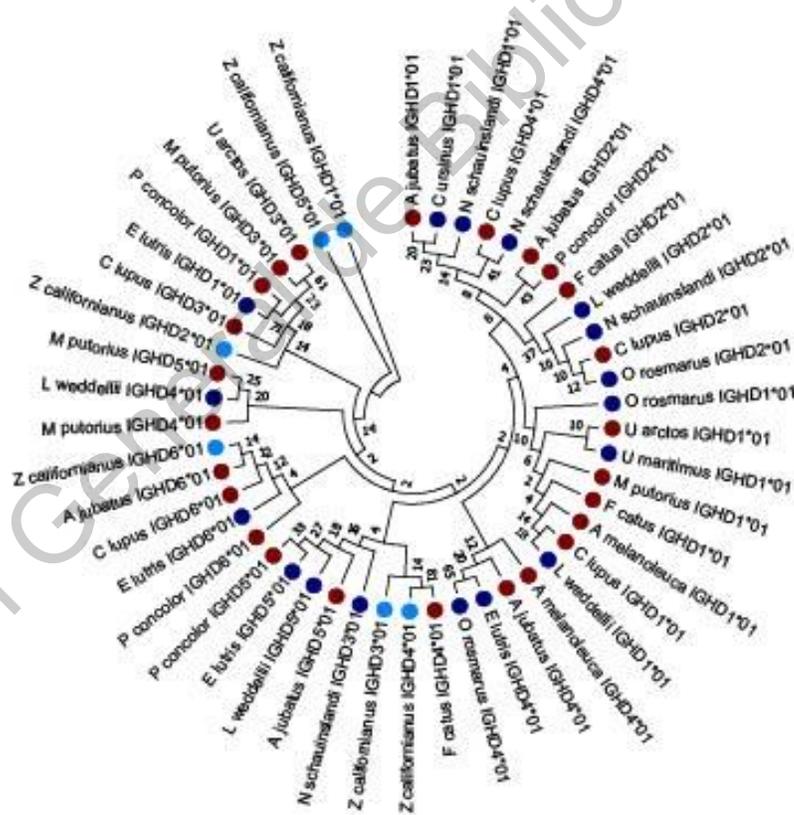


Figura 56| Análisis filogenético molecular de los segmentos IGHD de mamíferos carnívoros. Se muestra el árbol consenso de 1000 réplicas de Bootstrap. Todas las posiciones con menos del 95% de cobertura se eliminaron. Menos del 5% de espacios y bases ambiguas fueron permitidos. Los puntos rojos están indicando a los mamíferos carnívoros terrestres, los puntos azules están indicando a los mamíferos carnívoros marinos y los de color azul cielo están indicando secuencias de *Z. californianus*.

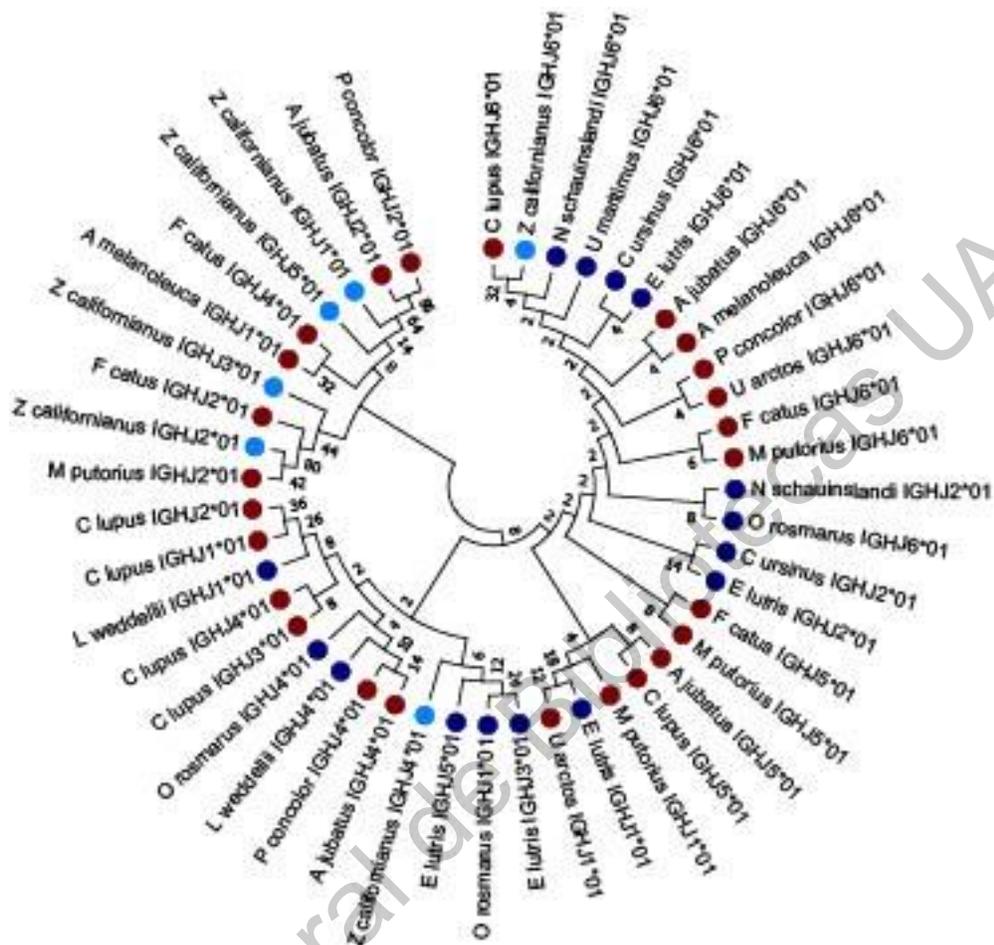


Figura 58| Análisis filogenético molecular de los segmentos IGHJ de mamíferos carnívoros. Se muestra el árbol consenso de 1000 réplicas de Bootstrap. Todas las posiciones con menos del 95% de cobertura se eliminaron. Menos del 5% de espacios y bases ambiguas fueron permitidos. Los puntos rojos están indicando a los mamíferos carnívoros terrestres, los puntos azules están indicando a los mamíferos carnívoros marinos y los de color azul cielo están indicando secuencias de *Z. californianus*.

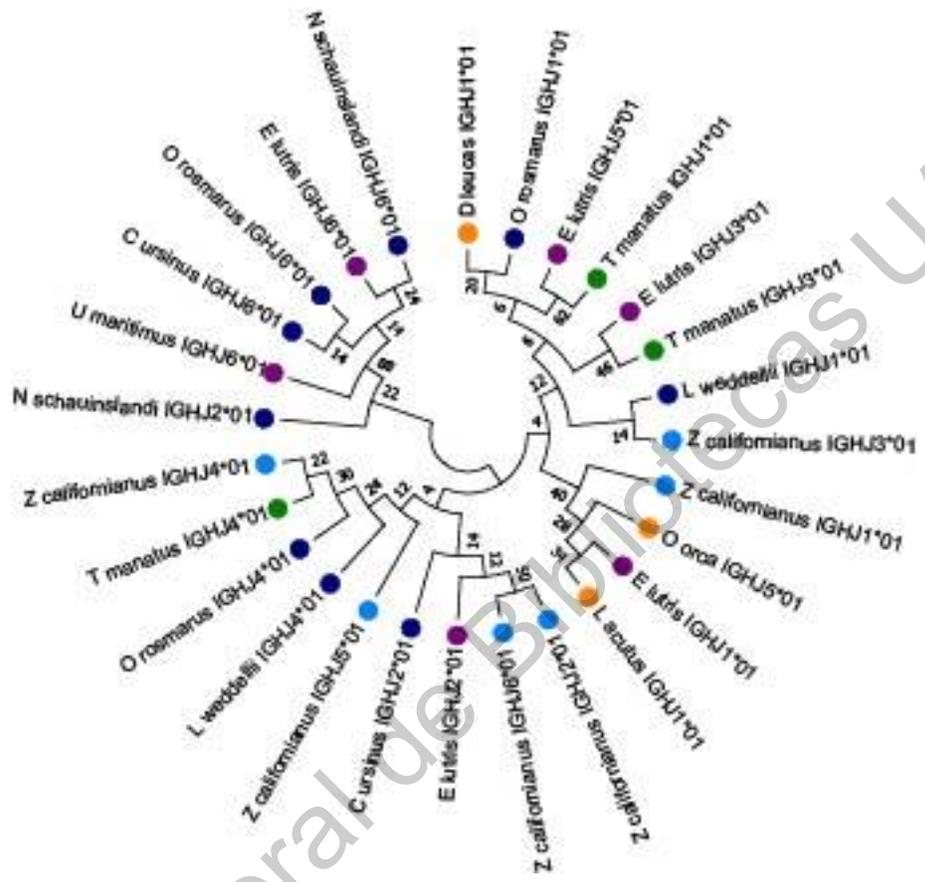


Figura 59| Análisis filogenético molecular de los segmentos IGHJ de mamíferos marinos. Se muestra el árbol consenso de 1000 réplicas de Bootstrap. Las posiciones incluidas corresponden a los codones 1st+2nd+3rd+Noncoding. Todas las posiciones con menos del 95% de cobertura se eliminaron. Menos del 5% de espacios y bases ambiguas fueron permitidos. Los puntos azules están indicando a los pinnípedos, los naranjas a cetáceos, los verdes a los sirenios, los morados al grupo de (otros mamíferos marinos) y los de color azul cielo están indicando secuencias de *Z. californianus*.

3.4.8 Evolución de los segmentos J de las cadenas lambda (IGLJ) de carnívoros y mamíferos marinos

Al reconstruir la historia evolutiva de los segmentos IGLJ de *Z. californianus*, el análisis incluyó 58 y 37 secuencias en la inferencia con otros carnívoros y mamíferos marinos, respectivamente. Estos segmentos mostraron homología entre los segmentos de otros carnívoros, a excepción de los segmentos IGLJ1 e IGLJ7 en ambas reconstrucciones (Figura 60-61). Por otro lado, el segmento IGL2 mostró tener homología entre los segmentos IGL2 de distintas especies de carnívoros y mamíferos marinos (Figura 60-61).

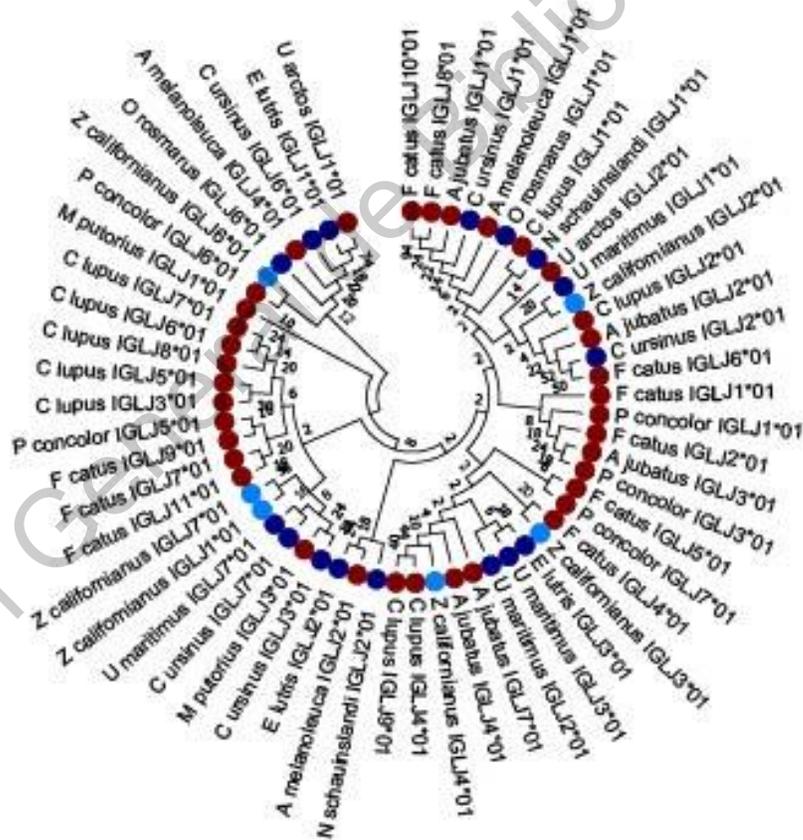


Figura 60| Análisis filogenético molecular de los segmentos IGLJ de mamíferos carnívoros. Se muestra el árbol consenso de 1000 réplicas de Bootstrap. Todas las posiciones con menos del 95% de cobertura se eliminaron. Menos del 5% de espacios y bases ambiguas fueron permitidos. Los puntos rojos están indicando a los mamíferos carnívoros terrestres, los puntos azules están indicando a los mamíferos carnívoros marinos y los de color azul cielo están indicando secuencias de *Z. californianus*.

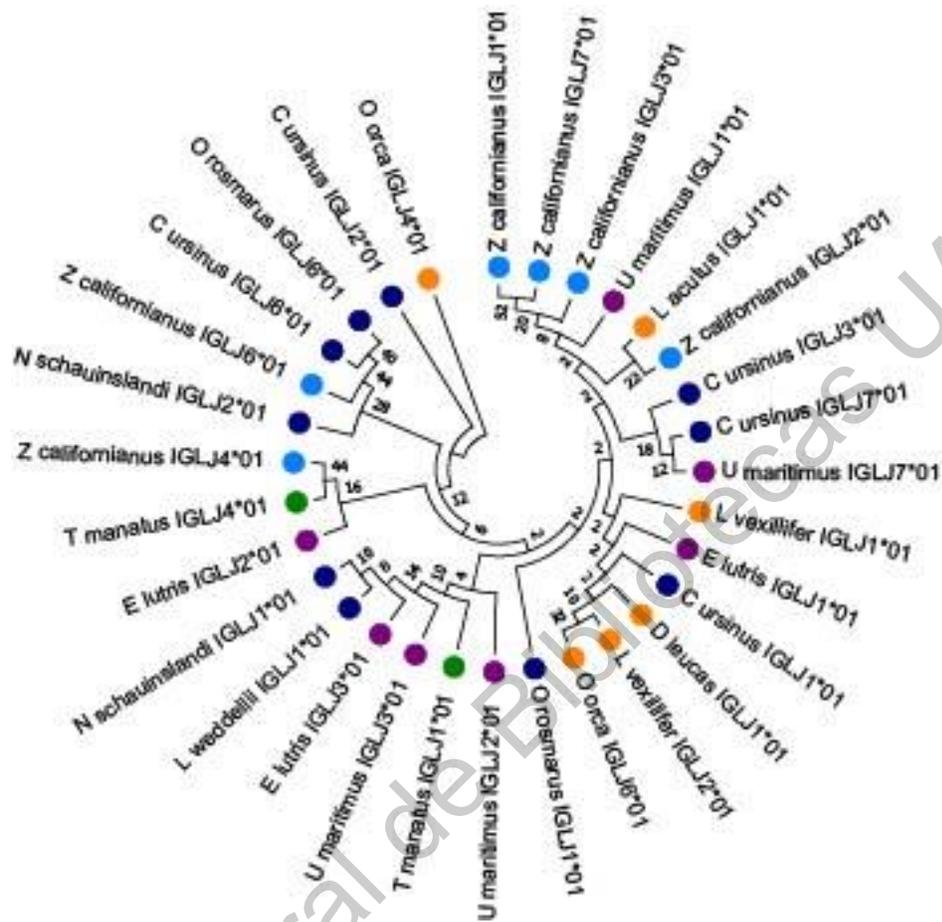


Figura 61| Análisis filogenético molecular de los segmentos IGLJ de mamíferos marinos. Se muestra el árbol consenso de 1000 réplicas de Bootstrap. Las posiciones incluidas corresponden a los codones 1st+2nd+3rd+Noncoding. Todas las posiciones con menos del 95% de cobertura se eliminaron. Menos del 5% de espacios y bases ambiguas fueron permitidos. Los puntos azules están indicando a los pinnípedos, los naranjas a cetáceos, los verdes a los sirenios, los morados al grupo de otros mamíferos marinos y los de color azul cielo están indicando secuencias de *Z. californianus*.

3.5 Evolución y ecología de los receptores TCR y BCR de *Z. californianus*

3.5.1 Relación ecológica y evolución de los segmentos V en las cadenas del TCR alpha/beta del lobo marino de California

La reconstrucción filogenética con respecto a las especies con relaciones ecológica con *Z. californianus*, incluyó 86 secuencias nucleotídicas de los segmentos TRAV (Figura 63) y 44 de los segmentos TRBV (Figura 64). En este análisis se observó que los segmentos TRBV tienen homólogos y ortólogos a través de las especies incluidas en el análisis, principalmente los segmentos (Figura 63-64).

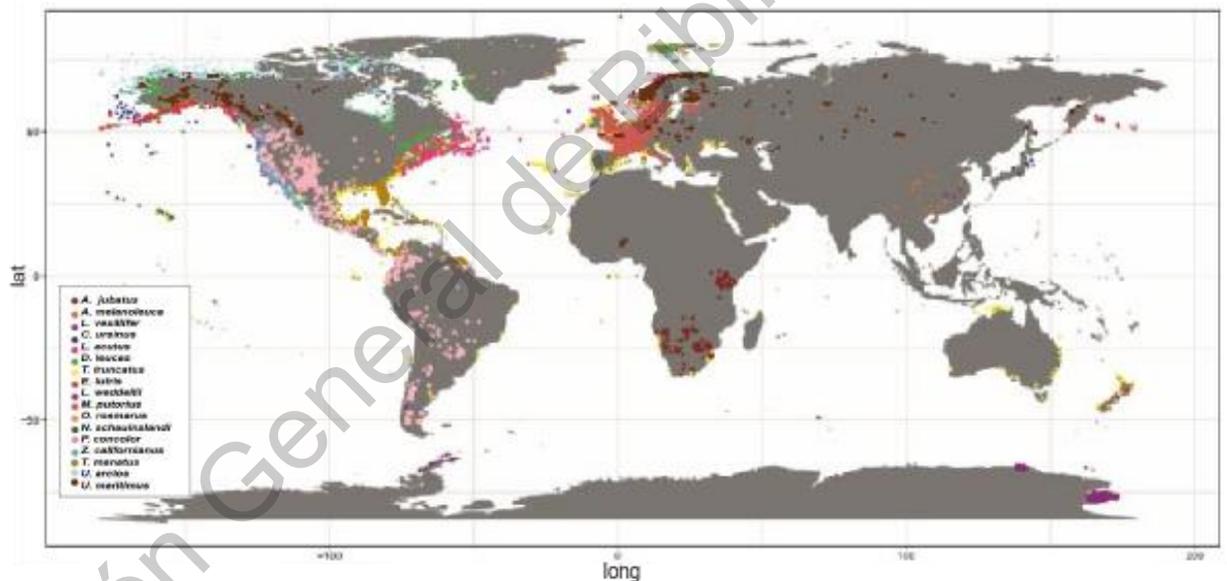
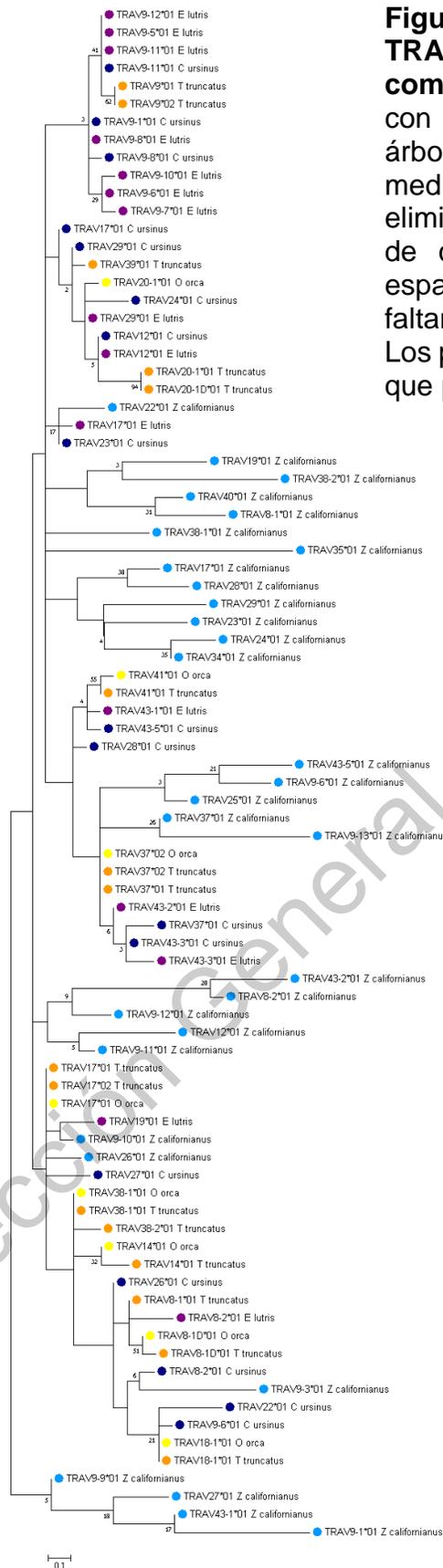


Figura 62| Distribución geográfica de las especies de mamíferos incluidas en el estudio. Los puntos en el mapa representan la distribución de las especies analizadas en este estudio. El recuadro indica el color que ejemplifica a cada especie.

Figura 63| Análisis filogenético de los segmentos TRAV de especies con las que *Z. californianus* comparte nichos ecológicos. Se muestra el árbol con la mayor probabilidad de registro (-838.50). El árbol se dibujó a escala, con longitudes de rama medidas en el número de sustituciones por sitio. Se eliminaron todas las posiciones con menos del 95% de cobertura del sitio, es decir, se permitieron espacios de alineación de menos del 5%, datos faltantes y bases ambiguas en cualquier posición. Los puntos de colores están indicando la especie a la que pertenece las secuencias nucleotídicas.



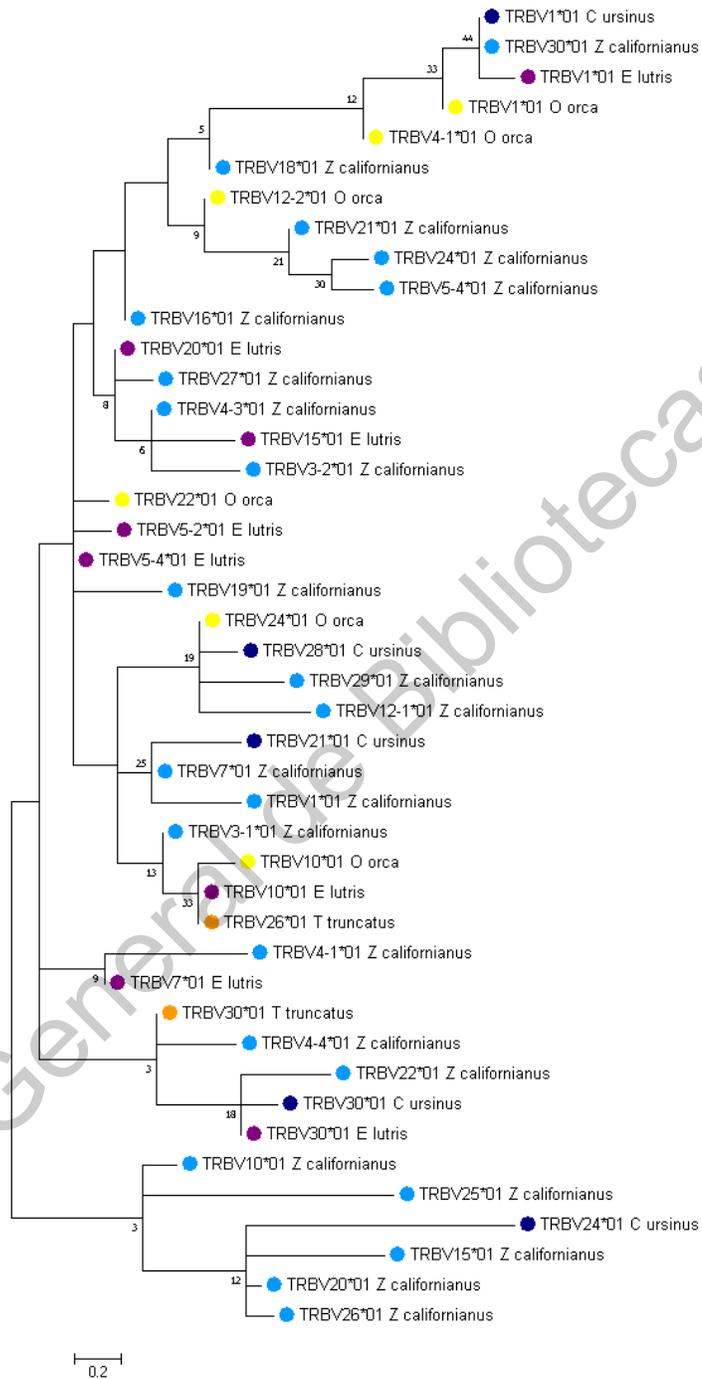


Figura 64| Análisis filogenético de los segmentos TRBV de especies con las que *Z. californianus* comparte nichos ecológicos. Se muestra el árbol con la mayor probabilidad de registro (-79793.00). El árbol se dibuja a escala, con longitudes de rama medidas en el número de sustituciones por sitio. Se eliminaron todas las posiciones con menos del 95% de cobertura del sitio, es decir, se permitieron espacios de alineación de menos del 5%, datos faltantes y bases ambiguas en cualquier posición. Los puntos de colores están indicando la especie a la que pertenece las secuencias nucleotídicas.

3.5.2 Relación ecológica y evolución de los segmentos V en las cadenas del BCR (Ig) heavy/lambda del lobo marino de California

En la reconstrucción de la historia evolutiva de estos segmentos en *Z. californianus* se observó que IGHV e IGLV mantienen alto grado de homología con *E. lutris* (Figura 65-66). En este análisis incluimos 65 (Figura 65) y 62 secuencias nucleotídicas (Figura 66).

Dirección General de Bibliotecas UAO

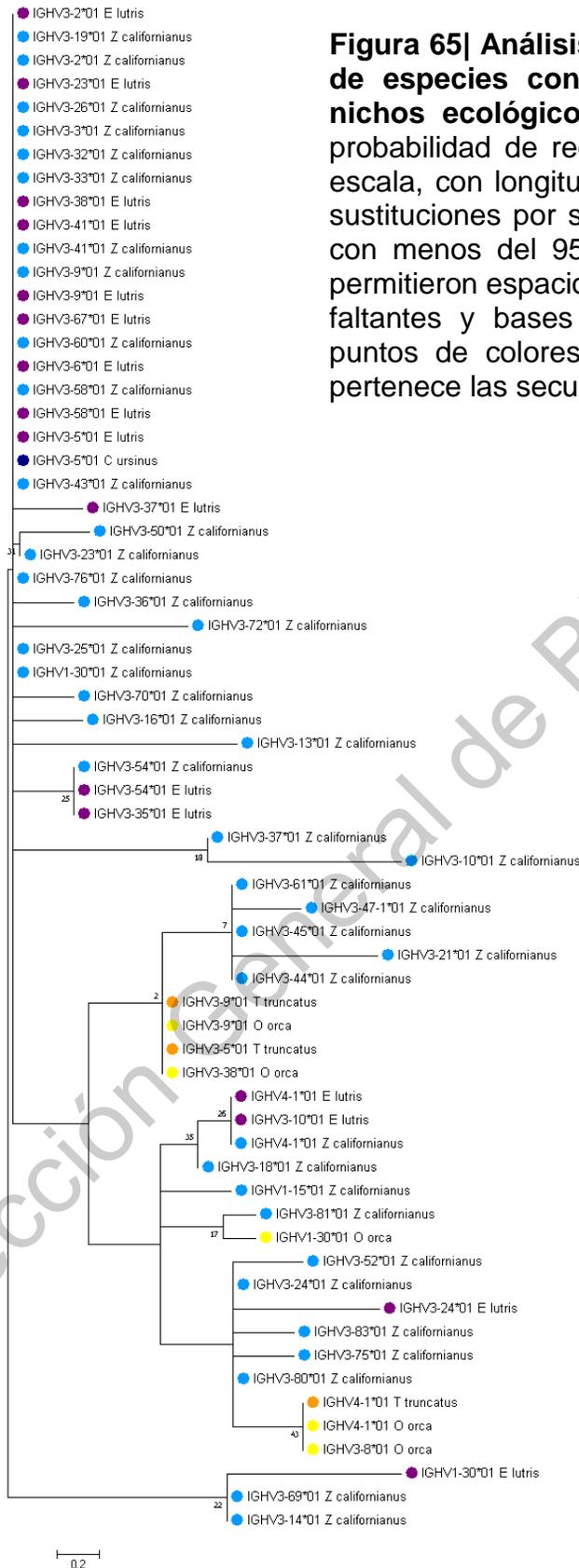


Figura 65| Análisis filogenético de los segmentos IGHV de especies con las que *Z. californianus* comparte nichos ecológicos. Se muestra el árbol con la mayor probabilidad de registro (-57016.47). El árbol se dibuja a escala, con longitudes de rama medidas en el número de sustituciones por sitio. Se eliminaron todas las posiciones con menos del 95% de cobertura del sitio, es decir, se permitieron espacios de alineación de menos del 5%, datos faltantes y bases ambiguas en cualquier posición. Los puntos de colores están indicando la especie a la que pertenece las secuencias nucleotídicas.

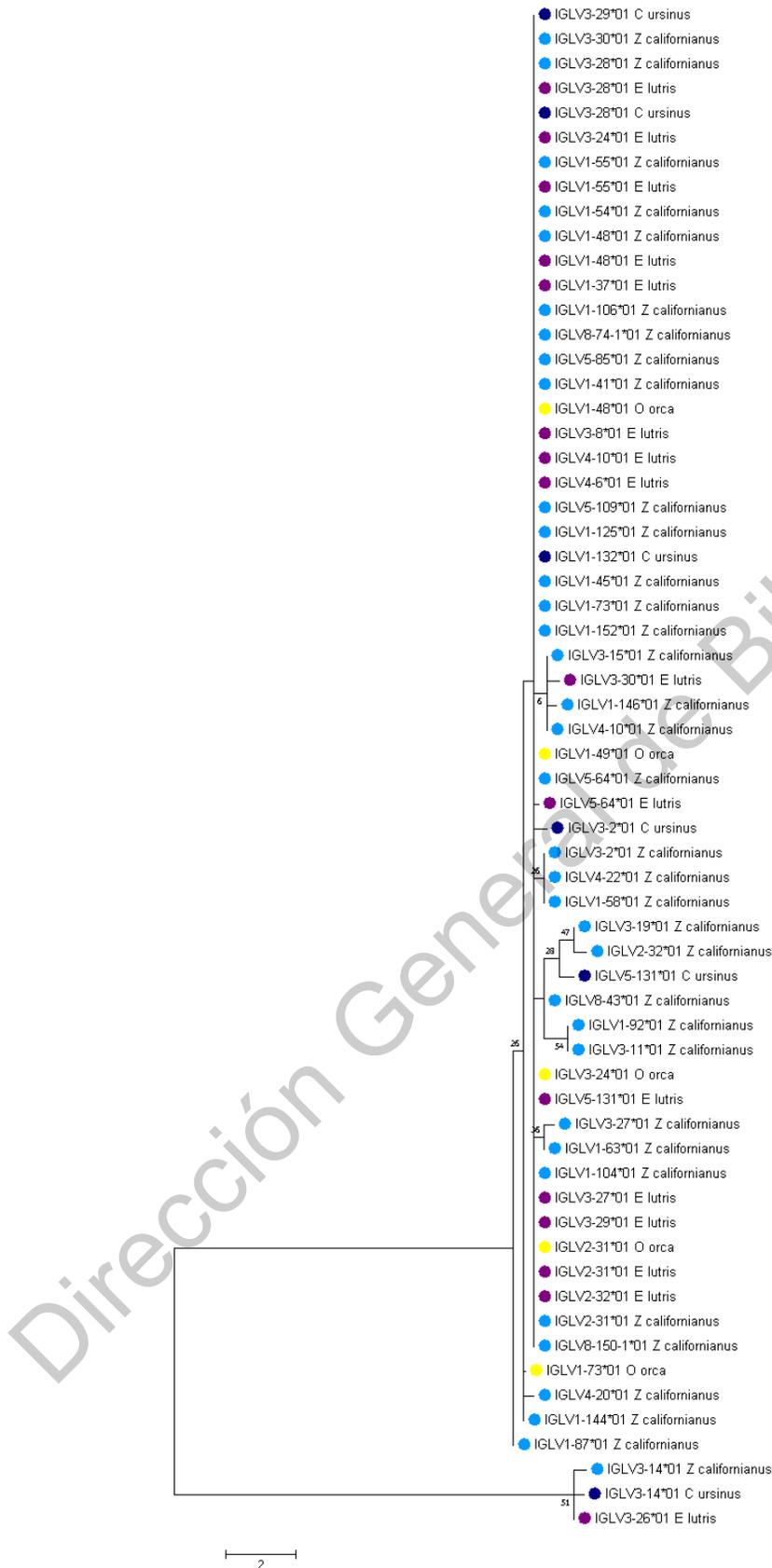


Figura 66| Análisis filogenético de los segmentos IGLV de especies con las que *Z. californianus* comparte nichos ecológicos. Se muestra el árbol con la mayor probabilidad de registro (-45652.11). El árbol se dibuja a escala, con longitudes de rama medidas en el número de sustituciones por sitio. Se eliminaron todas las posiciones con menos del 95% de cobertura del sitio, es decir, se permitieron espacios de alineación de menos del 5%, datos faltantes y bases ambiguas en cualquier posición. Los puntos de colores están indicando la especie a la que pertenece las secuencias nucleotídicas.

3.6 Evolución de las familias de los segmentos V(D)J de las cadenas del TCR de *Z. californianus* y BCR de *Z. californianus*

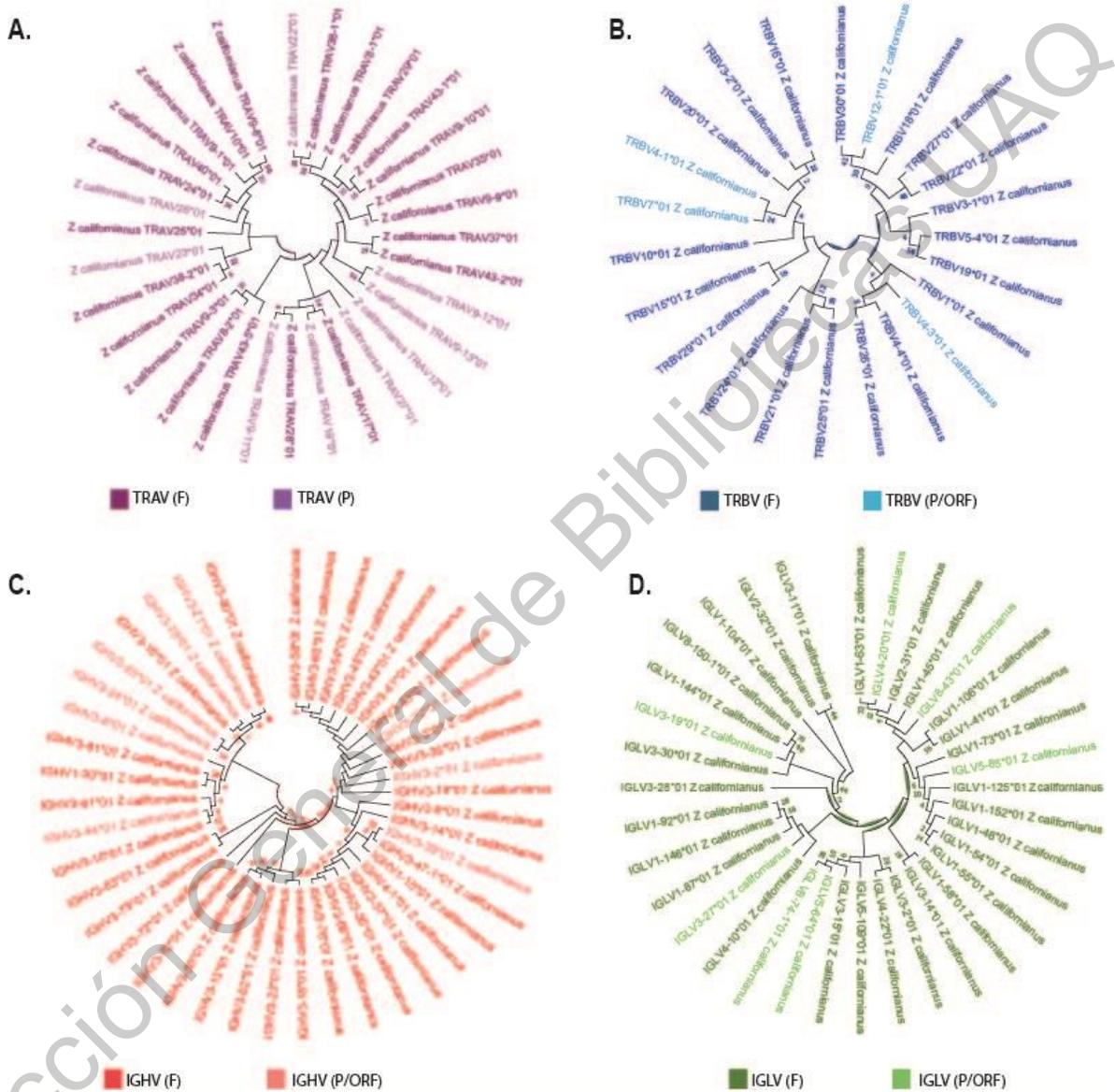


Figura 67| Evolución de las familias de los segmentos V(D)J de los receptores TCR y BCR de *Z. californianus*. A) Árbol de los segmentos TRAV (-2071.21). B) Árbol de los segmentos TRBV (-2115.88).C) Árbol de los segmentos IGHV (-45604.65).D) Árbol de los segmentos IGLV (-237.79). Los árboles se contruyeron por el método de Maximum likelihood y el método de Kimura de 2 parámetros. Los cuadros de colores indican el tipo de segmento en cada cadena.

fue de -20.75 (Figura A22D; Material Anexo) y un valor menor a 0.6 de la puntuación Z (Figura A22C; Material Anexo). En el análisis QMEAN individual la mayoría de los aminoácidos de la estructura tenían calidad alta (Figura A22E-F; Material Anexo). El receptor putativo 3D del MHC II de *Z. californianus* se muestra en la figura 70.

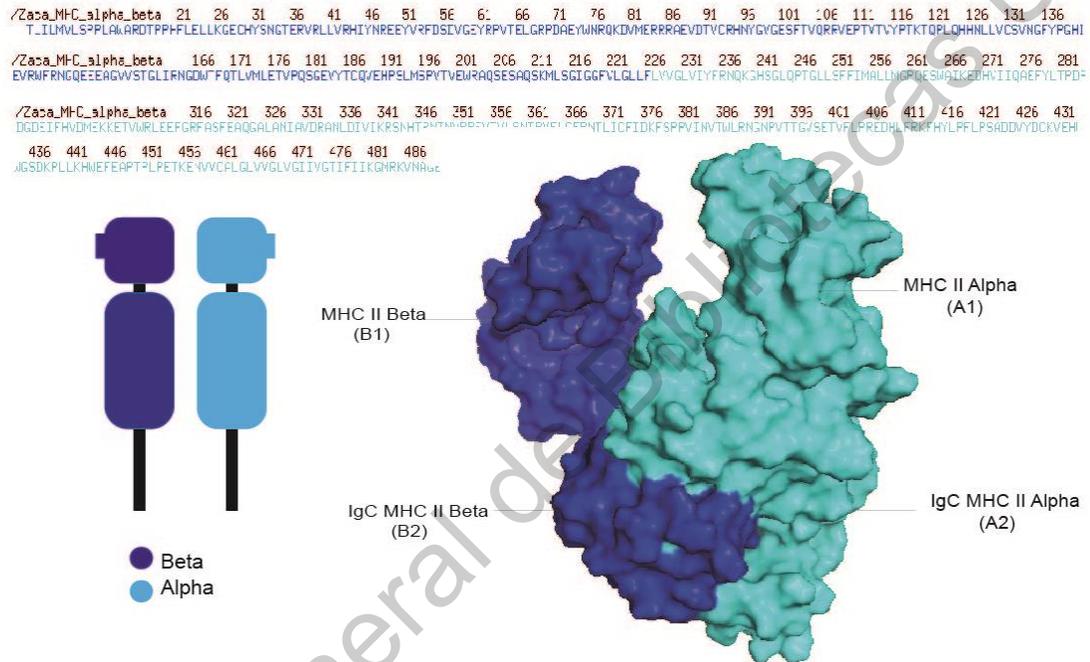


Figura 70| Estructura 3D del receptor MHC II de *Z. californianus*. El diagrama muestra la estructura tridimensional del receptor MHC II (DRBA), los colores están ejemplificando las cadenas del receptor.

Discusión

En esta tesis se caracterizó el repertorio de los receptores antigénicos de *Z. californianus*, y dado que no es una especie tradicional, fue desarrollado y optimizado un método que permitiera trabajar con una amplia cobertura, sin sesgar los datos. Aunque existen múltiples programas que permiten extraer los segmentos génicos V(D)J, la mayoría de estos utilizan especies modelo para demostrar su exactitud (e.g. Olivieri *et al.*, 2013; Hou *et al.*, 2016; Rosati *et al.*, 2017; Migalska *et al.*, 2018). Por lo tanto, suelen sesgar el repertorio únicamente a genes funcionales debido a que basan su caracterización en regiones conservadas previamente reportadas, lo cual en ecología y evolución representa un problema, debido a que los pseudogenes y los ORF son huellas genéticas del pasado patogénico de una especie (Nei & Sitnikova, 1997; Balakirev & Ayala, 2003).

En esta tesis se demostró que no está siendo sesgado el repertorio con el filtro de las secuencias, sino que se favorece la asignación del repertorio de los receptores TCR y BCR debido a que las secuencias con las cuales se generó el filtrado provienen de receptores caracterizados y anotados de una gran variedad de especies.

Por otro lado, esta tesis tomó en cuenta que la mayoría de las características que delimitan el que una secuencia sea un receptor TCR o BCR (Ig), están establecidas en las secuencias aminoacídicas (Giudicelli, 2004; Olivieri *et al.*, 2013), lo cual en especies sin genoma de referencia resulta inconveniente, por lo tanto la metodología aquí desarrollada incluyó secuencias nucleotídicas, y se establecieron relaciones cuantitativas para su caracterización.

La exactitud del método utilizado se comprobó cuantitativamente y en la mayoría de los segmentos V de las cadenas fueron optimizadas las similitudes entre los genes V de las especies filogenéticamente cercanas. En comparación con los diagramas de trabajo de otros programas el alineamiento en esta etapa del análisis resultó ser de suma importancia, porque demostró que, aunque se conservan similitudes con

otras especies, los genes V(D)J son representativos de cada especie, lo que en su mayoría se resuelve mediante filogenia con el conjunto de datos de los genes y no individualmente (Olivieri *et al.*, 2013; Hou *et al.*, 2016; Rosati *et al.*, 2017; Migalska *et al.*, 2018). Fue interesante encontrar que los valores de similitud de estos segmentos eran más cercanos filogenéticamente para las cadenas pesadas, específicamente la cadena beta del TCR. En contraste, en las cadenas ligeras de los receptores, los valores de similitud en el pre y post análisis no mostrarán diferencias significativas debido a que existían lecturas compartidas entre las dos muestras.

El método de clusterización aquí empleado logró discernir entre miembros de la superfamilia de las inmunoglobulinas y receptores TCR y BCR (Ig) de forma cuantitativa y permitió diferenciar inclusive entre genes con características muy conservadas entre los receptores (Marchler-Bauer *et al.*, 2016). Otros programas se esfuerzan en mantener los criterios de ontología génica de los receptores y olvidan que en otras especies estos varían significativamente y así terminan sesgando la caracterización potencial de nuevos segmentos (Hou *et al.*, 2016; Rosati *et al.*, 2017). El método utilizado, en contraste, permitió que el programa IgDiscover caracterizara un amplio repertorio en todas las cadenas de los receptores.

Por otro lado, existen programas que prueban su funcionalidad en repertorios previamente repostados y pierden de vista las similitudes genéticas con otros miembros de la superfamilia de las inmunoglobulinas (Olivieri *et al.*, 2013; Migalska *et al.*, 2018), por lo que el método aquí empleado también es una potencial extensión para evitar este tipo de errores en estos programas.

Durante el presente estudio, se encontraron caracteres conservados en las cadenas del TCR de *Z. californianus*. La primera es la relación que se mantiene entre los genes funcionales V, pseudogenes y ORFs, en donde los genes funcionales predominan. También se observó que los genes J funcionales de *Z. californianus* son tan diversos como en otras especies (Lefranc, 2001; Linguiti *et al.*, 2016; Martin

et al., 2017). Por lo anterior, es posible afirmar que la cadena alpha del TCR de *Z. californianus* tiende a la expansión, como ocurre en otras especies (Olivieri *et al.*, 2014).

Por otro lado, el uso de los genes V de la cadena alpha de *Z. californianus* es marcado en los segmentos TRAV8-1, TRAV38, TRAV24, TRAV9-1 y TRAV28. Por otro lado, en especies que comparten nichos ecológicos con *Z. californianus* el uso es marcado en las familias TRAV18 y TRAV20 (Linguiti *et al.*, 2016). Lo anterior además de ser un indicio en la organización del TRA de *Z. californianus*, (Proudhon *et al.*, 2015), también contradice la idea de que el repertorio está restringido a la presión patogénica en un hábitat determinado (Olivieri *et al.*, 2014).

Fueron caracterizados pseudogenes en la cadena alpha: TRAV12, TRAV19, TRAV27 TRAV21, TRAV26, TRAV22, TRAV23, TRAV9-13, TRAV9-11, TRAV9-12 y beta: TRBV4-1, TRBV12-1, TRBV4-3, TRBV7, además de un potencial ORF (TRBV15), los cuales en otras especies tienen un rol funcional. En ambas cadenas también se caracterizaron genes funcionales TRAV25 y TRAV37, TRBV19, TRBV21 y TRBV27, que en otras especies, como el perro (*Canis lupus familiaris*), gato (*Felis catus*) y hurón (*Mustela putorius furo*), se encuentran pseudogenizados o en forma de ORFs (Lefranc, 2001; Linguiti *et al.*, 2016; Martin *et al.*, 2017). Ejemplos similares se han observado entre el humano y el chimpancé, *Pan troglodytes*, en los que los pseudogenes de una especie son funcionales en otra, e inclusive participan en las respuestas contra virus que son potencialmente inmunogénicos en ambas especies (Meyer-Olson *et al.*, 2003).

Los segmentos de variabilidad de la cadena alpha y beta de *Z. californianus* pertenecieron a una gran diversidad de familias y la relación entre genes funcionales, pseudogenes y ORFs, fue similar a lo reportado para el perro (*Canis lupus familiaris*), gato (*Felis catus*) y hurón (*Mustela putorius furo*). De la misma manera la relación de genes J funcionales, pseudogenes, ORF, y genes D fue consistente con la de otros carnívoros (Martin *et al.*, 2017; Lefranc, 2001). Asimismo, se encontró que en la cadena beta el uso de los segmentos es marcado hacia los

miembros de las familias TRBV19, TRBV16 y TRBV29. Lo anterior concuerda con lo publicado por Proudhon *et al.* (2015), quienes propusieron que el uso de genes V(D)J refleja cercanía a las zonas reguladoras, y ofrece una idea sobre como se encuentra organizado el locus TRB en *Z. californianus*.

Las frecuencias y la diversidad de las familias V en la cadena alpha fueron más altas que las de la cadena beta. Una explicación podría ser la limitación que el MHC II ejerce sobre esta cadena durante el establecimiento de la sinapsis inmunológica (Nikolich-Zugich *et al.*, 2004). Adicionalmente, se observó que las cadenas del TCR de *Z. californianus* mantienen la tendencia en el ratio de genes V observada para otros mamíferos (Das *et al.*, 2008), en el que el número en la cadena alpha es en general más alto que el de la cadena Beta (Olivieri *et al.*, 2014).

Los segmentos TRAV y TRBV con altas frecuencias encontradas en este estudio se han asociado al desarrollo de memoria inmune periférica durante la patogénesis del carcinoma renal en el humano (Wang *et al.*, 2012). Interesantemente, el carcinoma urogenital es un proceso patológico frecuente en *Z. californianus* (Browning *et al.*, 2015), por lo que los hallazgos de esta tesis sugieren que la alta prevalencia de este cuadro neoplásico en el lobo marino de California podría tener un origen evolutivo, en el que estos genes que fueron ventajosos para la regulación de la respuesta de los linfocitos T en el pasado, ahora predisponen a esta especie al cáncer debido a la configuración genética de los loci del TCR (Proudhon *et al.*, 2015). Adicional a lo anterior, algunos de esos segmentos TRBV en *Z. californianus* están pseudogenizados, lo que podría indicar procesos diferentes para establecer mecanismos de memoria periférica o procesos de tolerogenia durante el desarrollo del carcinoma.

A diferencia de lo observado para el TCR, en donde se encontró un amplio número de familias constituyendo en repertorio del receptor, en el BCR se observó un sesgo en las familias de este receptor. Específicamente, en la cadena heavy, encontramos un marcado uso de genes de la familia IGHV3, en comparación con las otras dos familias encontradas (IGHV1 e IGHV4). Aunque en general la familia III se encuentra

enriquecida en el repertorio de otros carnívoros (Das *et al.*, 2008), el sesgo hacia el uso de esta familia no se había reportado. Observamos lo mismo en los segmentos IGHD, en donde el uso de los segmentos IGHD1 e IGHD4, es marcado en comparación con las otros. En contraste, este sesgo en el uso no fue encontrado en los segmentos J, lo cual podría deberse al número reducido de genes funcionales encontrados en *Z. californianus*. De la misma manera, como se planteó para los loci TRB, las frecuencias de los genes V(D)J de la cadena heavy también podrían ofrecer indicios sobre la organización y posible expansión del locus IGH de esta especie (Proudhon *et al.*, 2015). De manera interesante los pseudogenes V de esta cadena también pertenecen únicamente a la familia III, contrario a lo observado en otras especies (Olivieri *et al.*, 2014; Martin *et al.*, 2017).

Por otro lado, en la cadena lambda del receptor, no se observó un sesgo de uso hacia las familias IGLV ni hacia las familias IGLJ. Además se encontraron frecuencias similares entre los segmentos V de esta cadena. En general los segmentos de ambas cadenas mantienen una relación entre genes funcionales, pseudogenes y ORFs, que se ha caracterizado previamente en otras especies de carnívoros (Das *et al.*, 2008).

Similar a lo observado para el TCR, se detectaron segmentos IGHV e IGLV que son funcionales en otras especies (IGHV3-2, IGHV3-24, IGHV3-39, IGHV3-8, IGLV5-64, IGLV3-27, IGLV5-85) (Martin *et al.*, 2017). Sin embargo, fue interesante observar que el número de pseudogenes encontrados en este receptor, era menor que el número de pseudogenes encontrados en el receptor de linfocitos T, lo cual podría indicar que el proceso de generación, diversificación y muerte de las cadenas de este receptor es menor en comparación a las cadenas del receptor TCR, lo que es apoyado por el número reducido de familias que forman el repertorio de heavy y lambda. Hallazgos similares en la cadena lambda de *O. rosmarus* soportan esta idea (Olivieri *et al.*, 2014). Sin embargo, a diferencia del estudio de Olivieri de 2014, en el cual se asocia el número de genes V de la cadena lambda con el entorno patogénico del hábitat acuático, con base en los resultados aquí obtenidos, se

sugiere que podría deberse a que las cadenas ligeras cumplen un papel importante en el establecimiento de auto tolerancia (Collins & Watson, 2018), por lo que esto representa un cange evolutivo entre la pérdida de diversidad y la auto reactividad.

En cuanto a la cadena heavy, se ha observado la pérdida de segmentos constantes en esta cadena en dos especies de fócidos, *O. rosmarus* y *L. weddellii*, similar a lo que nosotros observamos en los segmentos de variación donde predomina solamente una familia. En conjunto, estos hallazgos plantean un escenario interesante sobre la adaptación de este loci en mamíferos semi acuáticos, en donde la pérdida de estos segmentos llevó a un sesgo hacia la respuesta celular en lugar de la respuesta humoral, por un proceso de segregación durante el regreso al medio acuático del ancestro de los pinnípedos (Li *et al.*, 2018). Por otro lado, vale la pena tomar en cuenta que miembros de la familia 4 de este locus se han relacionado con la progresión a desarrollo de autoinmunidad en humanos (Bashford-Rogers *et al.*, 2018) en *Z. californianus* el uso del único miembro caracterizado de esta familia (IGHV4-1) exhibe frecuencias bajas, lo que apoya la explicación propuesta de la disminución en las familias de este locus IGH pudo ser en consecuencia de la reducción evolutiva de genes que representaban riesgos de auto-reactividad.

Además de caracterizar los segmentos V(D)J de los receptores de *Z. californianus*, se generó un repositorio de los segmentos de otras especies de mamíferos. Eso permitió detectar que *Z. californianus* se relaciona, en cuanto al número de segmentos funcionales, ORFs y pseudogenes de las distintas cadenas, con otros vertebrados de distintas órdenes. En las cadenas alpha y beta, el número de segmentos V(D)J fue muy cercano al de otros mamíferos terrestres, tanto en los datos de esta tesis, como en los de repositorios publicados previamente (Olivieri *et al.*, 2015). Por otro lado, los segmentos V(D)J de la cadena lambda en número también resultaron ser cercanos a los de otras especies aún sin compartir el mismo hábitat, tanto en nuestro repertorio como en reportes publicados (Olivieri *et al.*, 2014). Por lo tanto, la hipótesis anteriormente propuesta de que el número de segmentos está relacionado a la pobre o amplia restricción patogénica en un hábitat

determinado (Olivieri *et al.* 2014) no se sostiene, ya que los resultados de esta tesis muestran que el repertorio de los genes del TCR y BCR de mamíferos marinos y terrestres es similar en número. La discrepancia entre los resultados del estudio previo y esta tesis podrían deberse a que, como se indicó con antelación, el repertorio de una especie está estrechamente relacionado al microambiente y canjes energéticos con otras características de historia de vida, además de la presión patogénica (Das *et al.*, 2008).

A partir del análisis comparativo de los segmentos de la cadena alpha de *Z. californianus*, se propone que los segmentos TRAV de esta especie evolucionaron en distintas etapas. Los subgrupos TRAV29, TRAV35, TRAV37, TRAV38 y TRAV40 de *Z. californianus* están más conservados entre especies de carnívoros y mamíferos marinos lo que indica que estos clanes evolucionaron en el ancestro común antes de que se diversificaran hace 28 millones de años (Berta, 1991). Ejemplos similares para TRAV37 y TRAV40 se han reportado, en los que se observó que mantienen homología con segmentos con otras especies de mamíferos (Martin *et al.*, 2017; Linguiti *et al.*, 2016). Sin embargo, estos segmentos no mantienen homología con los ungulados (Reinink & Van Rhijn, 2009), por lo que pudieron haberse diversificado antes de la separación entre perisodáctilos y carnívoros, ocurrida hace aproximadamente 87.6 millones de años (Waddell *et al.*, 1999).

Otro amplio repertorio de segmentos TRAV resultaron ser ortólogos o parálogos a segmentos de pinnípedos, úrsidos y mustélidos acuáticos, principalmente. Esto podría deberse a la especiación de estos segmentos en los carnívoros ancestros de los pinnípedos, que durante su regreso al hábitat acuático evolucionaron de manera neutral (Nei, 2007). Por otro lado, familias de los segmentos TRAV, como TRAV43, TRAV9, TRAV3 y TRAV8 podrían haber evolucionado adaptativamente y posteriormente diversificarse en *Z. californianus* (Olivieri *et al.*, 2014), debido a que estos clanes se agruparon monofiléticamente durante la reconstrucción filogenética, mientras que para los mismos segmentos se observaron algunos ortólogos en el grupo de sirenios y cetáceos, lo cual podría indicar evolución convergente entre los

TRAV de estas especies, debido a que comparten presiones patogénicas similares en el océano por parte de los virus (Lang *et al.*, 2009).

Los segmentos TRAJ de *Z. californianus* se agruparon indistintamente a través de los clados de mamíferos. Lo anterior es consistente con estudios previos que reportaron que los genes TRAJ mantienen niveles de homología entre especies filogenéticamente lejanas (Martin *et al.*, 2017) y su uso en general parece aleatorio (Linguiti *et al.*, 2016), posiblemente porque los clanes TRAJ mantienen restricciones evolutivas coordinadas por el proceso de origen y muerte (Nei *et al.*, 1997) de los clústeres V cercanos.

Resultó interesante observar que las familias TRAV que parecen haberse diversificado en *Z. californianus*, también muestran selección positiva durante la configuración del receptor y mantienen una alta tasa de segmentos pseudogenizados, lo cual sugiere que estos segmentos confieren una ventaja inmunológica a este pinnípedo en su hábitat. Sin embargo, a pesar del sesgo en el uso y diversificación de determinadas familias en *Z. californianus*, la tasa de evolución de este locus es consistente con otras especies (Lefranc, 2001; Linguiti *et al.*, 2016; Martin *et al.*, 2017). Por otro lado, estudios sobre las secciones constantes del TCR de otros pinnípedos apoyan la idea del aumento de repertorio en las cadenas del TCR en organismos de nichos semiacuáticos (Zhang *et al.*, 2019).

En el análisis comparativo de los segmentos de la cadena beta, se encontraron características conservadas con el locus TRAV. Se encontró que el clan TRAV19 estaba altamente conservado entre los mamíferos incluidos en este estudio y estudios previos en *C. lupus* y *A. lemurinus* (Martin *et al.*, 2017; Moncada *et al.*, 2005). En contraste con los segmentos TRAV que parecen haber emergido antes de la diversificación las órdenes de Laurasiatheria, la homología de este segmento se extiende a Perissodactyla (Connelley *et al.*, 2009), lo que lo postula como un segmento ancestral de los vertebrados. A lo largo de los clanes del TRBV, se encontraron homólogos y ortólogos, principalmente con mustélidos y pinnípedos.

De manera interesante, cuando se infiere la historia evolutiva de estos segmentos en el grupo de mamíferos marinos, la homología de los clanes TRBV20, TRBV10, TRBV7 y TRBV4-1 fue más clara entre pinnípedos, indicando que estos clanes evolucionaron adaptativamente en el ancestro común de pinnípedos durante su regreso al medio acuático (Nei, 2007). Por otro lado, en algunos clanes TRBV observamos indicios de evolución adaptativa en *Z. californianus*, cuando de manera similar a lo observado en la cadena alpha estos se clusterizaban en clados independientes. En este sentido, la adaptación de las familias TRBV en *Z. californianus* podría deberse a una co-evolución con la región DRB del MHC, el cual es altamente diverso en este pinnípedo (Bowen *et al.*, 2004). Mecanísticamente, es necesario establecer comunicación entre ambos durante el reconocimiento antigénico, y estas interacciones se establecen principalmente con los segmentos TRBV (Nikolich-Zugich *et al.*, 2004). Además, un estudio reciente encontró que el aumento en la expresión de los alelos del MHC II están asociados con la progresión de inmunopatologías infecciosas (Acevedo-Whitehouse *et al.*, 2018). En este sentido, el que *Z. californianus* use los genes TRBV más conservados en lugar de los que han evolucionado adaptativamente podría ser una señal sobre la restricción que ejercen los alelos del MHC sobre este locus en esta especie (Castro *et al.*, 2015).

La reconstrucción filogenética de los segmentos TRBD de *Z. californianus* fue muy interesante, ya que TRBD1 parece haber aparecido en el ancestro común de los mamíferos, mientras que TRBD2 podría haber aparecido más adelante en este mismo ancestro durante su reincorporación al medio acuático para después evolucionar adaptativamente en *Z. californianus*. TRBD2, además, muestra homología y ortología con los segmentos de mustélidos acuáticos y cetáceos, respectivamente. Este resultado es consistente con la diversificación de los cetáceos con respecto a los pinnípedos, en la cual los cetáceos pudieron haber adaptado el segmento TRBD2 (Kumar *et al.*, 2017) hace 50 millones de años (Uhen,

2010), en contraste a los pinnípedos que lo hicieron más tarde, hace 28 millones de años (Berta, 1991).

Por otro lado, los segmentos TRBJ, mostraron ortólogos y homólogos a lo largo de su historia evolutiva, y fue interesante notar que las homologías más frecuentes eran con el grupo de los úrsidos, lo cual implica que los clanes TRBJ de *Z. californianus* no han sufrido cambios importantes durante su diversificación del ancestro común, como se ha reportado en *Canis lupus familiaris* (Mineccia *et al.*, 2012).

Como se mencionó previamente, el uso de los segmentos IGHV muestra un sesgo hacia la familia IGHV III, interesantemente al reconstruir la historia evolutiva las familias IGHV. Algunos miembros de esta familia fueron homólogos a los de otras especies, lo cual indica que posiblemente aparecieron en el ancestro común, antes de la diversificación de mamíferos. Por otro lado, la mayoría de los miembros de esta familia parecen haber evolucionado de forma adaptativa, durante la reinserción de los pinnípedos al medio acuático y después haber surgido *Z. californianus*. Lo anterior es sustentado por los hallazgos que se comentaron anteriormente, en donde los pseudogenes IGHV provienen únicamente de esta familia, lo que en general implica una tendencia hacia el uso de los segmentos de una familia (Nei *et al.*, 1997). Adicional a lo anterior, se ha demostrado que los ungulados construyen su repertorio principalmente a partir de la familia III, mientras que los roedores lo hacen a partir de la familia II (Das *et al.*, 2008). El enriquecimiento de la familia III en *Z. californianus*, podría haber tenido su origen durante el declive del locus heavy en el ancestro común de mamíferos, en el que se perdieron familias IGH (Olivieri *et al.*, 2014), entonces el precursor de la familia III debió haberse diversificado en el ancestro de pinnípedos, para después expandirse en una única familia en *Z. californianus*. Lo anterior quizá, a que enriquecer otras familias implicaba coordinar mecanismos de auto-tolerancia (Bashford-Rogers *et al.*, 2018), que son evolutivamente más costosos (Demas & Nelson, 2012) .

En las otras familias génicas aquí caracterizadas para *Z. californianus* la homología con otras especies fue muy alta, principalmente para la familia IV. En el análisis comparativo no se observaron ortólogos a la familia II en esta especie lo que se ha reportado también en el cerdo (Aitken *et al.*, 1997).

El origen de los segmentos IGHD en *Z. californianus*, según los resultados aquí obtenidos, es dual. IGHD6, IGH3 y IGH4 posiblemente se originaron en el ancestro común antes de su separación, debido a que son evolutivamente cercanos a los de otros carnívoros (Martin *et al.*, 2017). Mientras que IGHD1 e IGH6 parecen estar relacionados al ancestro común de mamíferos. Resulta interesante que el uso de estos segmentos muestra una tendencia hacia ambas las familias IGHD1 e IGHD4, lo cual sugiere procesos de selección natural sobre estos clanes y una habituación del sistema inmune de *Z. californianus* al hábitat semi acuático, como en otras especies en donde su evolución se ha relacionado con procesos de segregación que conduce a una mayor homocigosis (Martin *et al.*, 2017).

Por otra parte, los segmentos de unión, a diferencia de los del TCR, mostraron más homología con las familias IGHD de los pinnípedos que con otros mamíferos, lo cual también podría ser un indicio de cómo han evolucionado las cadenas IGH a través de los taxones para conferir mayor habituación al medio acuático, y compensar la pérdida de genes IGH, como se ha postulado en trabajos anteriores (Olivieri *et al.*, 2014). Sin embargo, a diferencia de lo observado por estos autores, aquí se propone que la reducción en el repertorio antigénico, no es sinónimo de reducción en la capacidad inmune de una especie, sino una estrategia evolutiva que podría ayudar al desarrollo de mecanismos de tolerancia, que a su vez, podrían incrementar su susceptibilidad a desarrollar cáncer (Browning *et al.*, 2015).

Además, resulta interesante que la pérdida en la expansión y uso de otras familias IGHD fue consistente con lo reportado para las regiones constantes de otros pinnípedos, en donde este grupo parece no haber duplicado algunos segmentos para la generación de inmunoglobulinas solubles IgG y en donde IgD parece estar incompleta (Li *et al.*, 2018).

Durante la reconstrucción evolutiva de los segmentos VJ de la cadena lambda, se encontró que la mayoría de los segmentos tenían ortólogos y homólogos en otras especies de pinnípedos principalmente, seguido de su relación con úrsidos. Dentro de estos, IGLV1-58, IGLV2-31, IGLV1-125, IGLJ2 también fueron de los más usados y consistentemente los que guardaban mayor relación con otros pinnípedos. Esto podría ser indicios de su habituación durante el regreso al ecosistema acuático, además debido a su cercanía evolutiva también podrían hablarnos sobre su historia patogénica. Por otro lado, en esta cadena como en IGH encontramos un sesgo hacia el uso de los miembros de la familia IGL1 e IGLJ2 que, aunque en general podemos encontrar homólogos en pinnípedos y úrsidos, su uso en otras especies no podría ser el mismo. En este sentido el número de los segmentos reportados en este trabajo también es menor al reportado para otras especies como *Canis lupus familiaris*, que tiende a la expansión de esta cadena (Olivieri *et al.*, 2014; Martin *et al.*, 2017). Aunque en este contexto *Z. californianus* podría parecer contar con menos restricciones patogénicas, esto podría ser compensado por la hipermutación somática del BCR (Abbas *et al.*, 2017). En contraste, la reducción de las familias VJ que conforman el repertorio en *Z. californianus*, también podría deberse a controles estrictos de auto-reactividad durante la evolución de sus segmentos, como se ha observado en la cadena kappa de otras especies (Collins & Watson, 2018).

Cuando comparamos los segmentos V del TCR y BCR de *Z. californianus* con especies con las que comparten nichos ecológicos, encontramos que en la cadena alpha la evolución de la mayoría de sus segmentos, no compartían homología con otros mamíferos, mientras que la cadena beta, comparte homología con los segmentos de otras especies con las que comparten nichos ecológicos, pero siguen existiendo clústeres de segmentos V de *Z. californianus* que se agrupan solos.

Una posible explicación a lo anterior podría ser que las especies como *E. lutris* y *C. ursinus*, quienes a pesar de compartir nichos ecológicos similares y ser semi-acuáticas, no son especies sociales (Silverstein *et al.*, 1995; Gentry, 2014), lo no favorece la interacción entre los miembros de un ecosistema y no permite que los

patógenos entre las poblaciones de estas especies y llevar a la generación de alelos nuevos, lo que se ha asociado al número de secuencias constantes en el TCR (Zhang *et al.*, 2019). Esto sería un factor importante para determinar la respuesta inmunológica ya que en especies sociales como *Z. californianus* (Schusterman *et al.*, 1966), es más común observar una alta prevalencia y diversidad de patógenos (e.g. Lyons *et al.*, 1997; King *et al.*, 2002; Johnson *et al.*, 2006; Lloyd-Smith *et al.*, 2007; Ng *et al.*, 2009; Rivera *et al.*, 2012; Kuzmina *et al.*, 2018)

Por lo anterior, la evolución de los segmentos TRAV en *Z. californianus*, parece haber seguido una línea evolutiva adaptativa, mientras que los segmentos TRBV parecen estar más restringidos por el MHC de esta especie debido, a que no se han diversificado tanto como el locus alpha en nichos ecológicos similares.

En el caso del BCR de *Z. californianus*, los segmentos V de las cadenas parecen ser más cercanos a los de especies con los que comparte nichos ecológicos similares, especialmente a *E. lutris* a pesar del comportamiento diferencial que exhiben las dos especies en las cuales la diseminación de patógenos y probabilidad de infección son distintas, parece que estas especies en el BCR comparten genes ortólogos y parálogos, una posible explicación de nuevo podría deberse a que comparten agentes infecciosos virales similares (Lang *et al.*, 2009). Aunque a diferencia del TCR en donde trabajos previos relacionan la genética de los segmentos y variables ecológicas, para el BCR no se han publicado estudios semejantes. Con base en los resultados aquí obtenidos, se sugiere que como en otros genes relacionados con la respuesta innata (Tian *et al.*, 2018), en donde no existen limitaciones al reconocimiento de otras molecular y que son capaces de reconocer patrones derivados de una infección libremente, como lo hacen los BCRs con antígenos libres en su forma soluble como inmunoglobulina (Murphy & Weaver, 2017).

Al modelar la estructura putativa del TCR Alpha/beta de *Z. californianus* se encontraron concordancias con los receptores de otras especies, principalmente el humano (Adams *et al.*, 2005; Gras *et al.*, 2009). Sin embargo, es probable que esta

relación refleje el que existen muy pocas especies dentro del repertorio de Protein Data Bank (PDB) para las que se cuente con estructuras cristalizadas a partir de las cuales I-TASSER pudiera modelar la estructura de este receptor. Ésto además explicaría el porcentaje de similitud con los receptores de estas especies. Por otro lado, tanto en el TCR Alpha/beta como en el MHC II se encontraron aminoácidos de contacto entre sus cadenas conservadas entre especies con estructuras bien reportadas (Marrack *et al.*, 2008).

En el caso del MHC II DRB de *Z. californianus* la similitud estructural de este receptor con otras estructuras también fue principalmente con el humano (Fremont *et al.*, 1997), de nuevo plausiblemente debido al número reducido de estructuras en el PDB. Fue interesante observar que para la cadena beta subunidad 1 de este receptor se encontraron aminoácidos conservados, útiles para el establecimiento de la sinapsis inmunológica (Marrack *et al.*, 2008), lo que habla de un patrón bien establecido a través de las especies para la co-evolución de estos dos receptores, el cual se extiende hasta *Z. californianus*. Serán necesarios más análisis a futuro para determinar, en las regiones contantes, cómo evolucionaron estos patrones en los mamíferos.

Para el BCR heavy/lambda (Ig), a pesar de que existe un gran número de receptores de diversas especies en el repertorio de PDB, su principal asociación por interacciones y secuencia aminoacídica fue de nuevo con humano (Sato *et al.*, 2019). Durante el modelado de la estructura del receptor se asoció al reconocimiento de antígenos propios. Sin embargo, no es posible relacionar este receptor a un antígeno específico porque su estructura fue modelada a partir de una secuencia de aminoácidos proveniente de una concatenación de secuencias.

VI. Conclusiones

Esta tesis generó conocimiento primerio novedoso sobre cómo ha evolucionado el repertorio antigénico en distintas especies, a partir de cuatro hallazgos principales:

Primero, a través de este estudio se conoce ahora el repertorio de los segmentos génicos V(D)J del TCR y BCR de *Z. californianus*. En términos de número y diversidad, este repertorio es similar al de otros mamíferos carnívoros, lo que descarta la idea de que el medio acuático contiene menos patógenos.

En este sentido, la metodología (Eco-Ig) desarrollada que permitió caracterizar el repertorio o puede aplicarse para estudiar a otros organismos no modelo, debido a que no sesga la caracterización del repertorio en especies no anotadas. Además, puede aplicarse al estudio de otros genes, independientemente de su función.

Segundo, el repertorio de los segmentos V(D)J del TCR de *Z. californianus* exhibe tendencias hacia la expansión, como ocurre en otras especies de pinnípedos, mientras que repertorio del BCR (Ig) está constituido por un número reducido de familias, indicando tendencias hacia la reducción de la diversidad. Este hallazgo podría ser evidencia de que existe un sesgo hacia la respuesta inmune celular y no humoral en esta especie.

Tercero, la evolución del repertorio de *Z. californianus* podría estar indicando mecanismos de tolerancia inmune en especies semiacuáticas del ambiente marino, debido al amplio espectro al cual son expuestos .

Cuarto, a nivel estructura, los receptores conservan patrones de reconocimiento similar a las de otras especies, incluyendo al humano.

Finalmente, no se puede descartar que pueda existir expansión de las cadenas gamma/delta o kappa, las cuales no fueron exploradas en esta tesis. Sin embargo, con base en los resultados obtenidos, se hipotetiza que estas cadenas mantienen el patrón de uso aquí observado. Estudios subsecuentes podrán desafiar esta hipótesis

Referencias

- Abbas, A., Lichtman, A. & Pillai, S. (2017). Cellular and molecular immunology. 9th ed. Philadelphia: Elsevier.
- Acevedo-Whitehouse, K., Gulland, F. M. D., & Bowen, L. (2018). MHC class II DRB diversity predicts antigen recognition and is associated with disease severity in California sea lions naturally infected with *Leptospira interrogans*. *Infection, Genetics and Evolution*, 57, 158–165. <https://doi.org/10.1016/j.meegid.2017.11.023>
- Acevedo-Whitehouse, K., Gulland, F., Greig, D., & Amos, W. (2003). Disease susceptibility in California sea lions. *Nature*, 422(6927), 35–35. <https://doi.org/10.1038/422035a>
- Adams, E. J., & Garcia, K. C. (2005). Immune receptor. Protein Data Bank, Rutgers University. <https://doi.org/10.2210/pdb1ypz/pdb>
- Aitken, R., Gilchrist, J., & Sinclair, M. C. (1997). A single diversified VHgene family dominates the bovine immunoglobulin repertoire. *Biochemical Society Transactions*, 25(2), 326S–326S. <https://doi.org/10.1042/bst025326s>
- Araujo-pires, A. C., Francisconi, C. F., Bigueti, C. C., Cavalla, F., Aranha, A. M. F., Letra, A., ... Garlet, G. P. (2014). Simultaneous analysis of T helper subsets (Th1, Th2, Th9, Th17, Th22, Tfh, Tr1 and Tregs) markers expression in periapical lesions reveals multiple cytokine clusters accountable for lesions activity and inactivity status. *Journal of Applied Oral Science*, 22(4), 336–346. <https://doi.org/10.1590/1678-775720140140>
- Arnason, U., Gullberg, A., Janke, A., Kullberg, M., Lehman, N., Petrov, E. A., & Väinölä, R. (2006). Pinniped phylogeny and a new hypothesis for their origin and dispersal. *Molecular Phylogenetics and Evolution*, 41(2), 345–354. <https://doi.org/10.1016/j.ympev.2006.05.022>
- Balakirev, E. S., & Ayala, F. J. (2003). Pseudogenes: Are They “Junk” or Functional DNA? *Annual Review of Genetics*, 37(1), 123–151. <https://doi.org/10.1146/annurev.genet.37.040103.103949>
- Banuet-Martínez, M., Espinosa-de Aquino, W., Elorriaga-Verplancken, F. R., Flores-Morán, A., García, O. P., Camacho, M., & Acevedo-Whitehouse, K. (2017). Climatic anomaly affects the immune competence of California sea lions. *PLOS ONE*, 12(6), e0179359. <https://doi.org/10.1371/journal.pone.0179359>

- Bashford-Rogers, R. J. M., Smith, K. G. C., & Thomas, D. C. (2018). Antibody repertoire analysis in polygenic autoimmune diseases. *Immunology*, 155(1), 3–17. <https://doi.org/10.1111/imm.12927>
- Batista, F. D., & Harwood, N. E. (2009). The who, how and where of antigen presentation to B cells. *Nature Reviews Immunology*, 9(1), 15–27. <https://doi.org/10.1038/nri2454>
- Berta, A. (1991). New Enaliarctos* (Pinnipedimorpha) from the Oligocene and Miocene of Oregon and the Role of “Enaliarctids” in Pinniped Phylogeny. *Smithsonian Contributions to Paleobiology*, (69), 1–33. <https://doi.org/10.5479/si.00810266.69.1>
- Berta, A., Sumach, J. and Kovacs, K. (2015). *Marine mammals: evolutionary biology* (Third edition). *Polar Research*. 26(1).
- Blanchette, M. (2004). Aligning Multiple Genomic Sequences with the Threaded Blockset Aligner. *Genome Research*, 14(4), 708-715. doi: 10.1101/gr.1933104
- Boehm, T. (2006). Co-evolution of a primordial peptide-presentation system and cellular immunity. *Nature Reviews Immunology*, 6(1), 79–84. <https://doi.org/10.1038/nri1749>
- Boehm, T., & Bleul, C. C. (2007). The evolutionary history of lymphoid organs. *Nature Immunology*, 8(2), 131–135. <https://doi.org/10.1038/ni1435>
- Boehm, T., & Swann, J. B. (2014). Origin and Evolution of Adaptive Immunity. *Annual Review of Animal Biosciences*, 2(1), 259–283. <https://doi.org/10.1146/annurev-animal-022513-114201>
- Boehm, T., Iwanami, N., & Hess, I. (2012). Evolution of the Immune System in the Lower Vertebrates. *Annual Review of Genomics and Human Genetics*, 13(1), 127–149. <https://doi.org/10.1146/annurev-genom-090711-163747>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bousso, P., Wahn, V., Douagi, I., Horneff, G., Pannetier, C., Le Deist, F., ... de Saint Basile, G. (2000). Diversity, functionality, and stability of the T cell repertoire derived in vivo from a single human T cell precursor. *Proceedings of the*

National Academy of Sciences, 97(1), 274–278.
<https://doi.org/10.1073/pnas.97.1.274>

Bowen, L., Aldridge, B. M., DeLong, R., Melin, S., Buckles, E. L., Gulland, F., ... Johnson, M. L. (2005). An immunogenetic basis for the high prevalence of urogenital cancer in a free-ranging population of California sea lions (*Zalophus californianus*). *Immunogenetics*, 56(11), 846–848.
<https://doi.org/10.1007/s00251-004-0757-z>

Bowen, L., Aldridge, B. M., Gulland, F., Van Bonn, W., DeLong, R., Melin, S., ... Johnson, M. L. (2004). Class II multiformity generated by variable MHC-DRB region configurations in the California sea lion (*Zalophus californianus*). *Immunogenetics*, 56(1), 12–27. <https://doi.org/10.1007/s00251-004-0655-4>

Breen, M., Hitte, C., Lorentzen, T. D., Thomas, R., Cadieu, E., Sabacan, L., ... Ostrander, E. A. (2004). *BMC Genomics*, 5(1), 65.
<https://doi.org/10.1186/1471-2164-5-65>

Bretscher, P. A. (2014). On the Mechanism Determining the Th1/Th2 Phenotype of an Immune Response, and its Pertinence to Strategies for the Prevention, and Treatment, of Certain Infectious Diseases. *Scandinavian Journal of Immunology*, 79(6), 361–376. <https://doi.org/10.1111/sji.12175>

Browning, H. M., Acevedo-Whitehouse, K., Gulland, F. M. D., Hall, A. J., Finlayson, J., Dagleish, M. P., Hammond, J. A. (2014). Evidence for a genetic basis of urogenital carcinoma in the wild California sea lion. *Proceedings of the Royal Society B: Biological Sciences*, 281(1796), 20140240.
<https://doi.org/10.1098/rspb.2014.0240>

Browning, H. M., Gulland, F. M. D., Hammond, J. A., Colegrove, K. M., & Hall, A. J. (2015). Common cancer in a wild animal: the California sea lion (*Zalophus californianus*) as an emerging model for carcinogenesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1673), 20140228. <https://doi.org/10.1098/rstb.2014.0228>

Cannon, J. P., Haire, R. N., & Litman, G. W. (2002). Identification of diversified genes that contain immunoglobulin-like variable regions in a protochordate. *Nature Immunology*, 3(12), 1200–1207. <https://doi.org/10.1038/ni849>

Castro, C. D., Luoma, A. M., & Adams, E. J. (2015). Coevolution of T-cell receptors with MHC and non-MHC ligands. *Immunological Reviews*, 267(1), 30–55.
<https://doi.org/10.1111/imr.12327>

- Chinta, S. (2013) Sequencing and annotation of sea lion genome (Master of Science of Bioinformatics and Medical Informatics). San Diego State University. San Diego, California
- Cicchese, J. M., Evans, S., Hult, C., Joslyn, L. R., Wessler, T., Millar, J. A., ... Kirschner, D. E. (2018). Dynamic balance of pro- and anti-inflammatory signals controls disease and limits pathology. *Immunological Reviews*, 285(1), 147–167. <https://doi.org/10.1111/imr.12671>
- Collins, A. M., & Watson, C. T. (2018). Immunoglobulin Light Chain Gene Rearrangements, Receptor Editing and the Development of a Self-Tolerant Antibody Repertoire. *Frontiers in Immunology*, 9. <https://doi.org/10.3389/fimmu.2018.02249>
- Connelley, T., Aerts, J., Law, A., & Morrison, W. I. (2009). Genomic analysis reveals extensive gene duplication within the bovine TRB locus. *BMC Genomics*, 10(1), 192. <https://doi.org/10.1186/1471-2164-10-192>
- Cooper, M. D., & Alder, M. N. (2006). The Evolution of Adaptive Immune Systems. *Cell*, 124(4), 815–822. <https://doi.org/10.1016/j.cell.2006.02.001>
- Corcoran, M. M., Phad, G. E., Bernat, N. V., Stahl-Hennig, C., Sumida, N., Persson, M. A. A., Hedestam, G. B. K. (2016). Production of individualized V gene databases reveals high levels of immunoglobulin genetic diversity. *Nature Communications*, 7(1). <https://doi.org/10.1038/ncomms13642>
- Das, S., Nozawa, M., Klein, J., & Nei, M. (2008). Evolutionary dynamics of the immunoglobulin heavy chain variable region genes in vertebrates. *Immunogenetics*, 60(1), 47–55. <https://doi.org/10.1007/s00251-007-0270-2>
- DeFranco, A. L. (1993). Structure and Function of the B Cell Antigen Receptor. *Annual Review of Cell Biology*, 9(1), 377–410. <https://doi.org/10.1146/annurev.cb.09.110193.002113>
- Demas, G. and Nelson, R. (2012). *Ecoimmunology*. Oxford: Oxford University Press.
- Dermody, T. S., Kirchner, E., Guglielmi, K. M., & Stehle, T. (2009). Immunoglobulin Superfamily Virus Receptors and the Evolution of Adaptive Immunity. *PLoS Pathogens*, 5(11), e1000481. <https://doi.org/10.1371/journal.ppat.1000481>
- Dorshkind, K., & Montecino-Rodriguez, E. (2007). Fetal B-cell lymphopoiesis and the emergence of B-1-cell potential. *Nature Reviews Immunology*, 7(3), 213–219. <https://doi.org/10.1038/nri2019>

- Dustin, M. L., & Choudhuri, K. (2016). Signaling and Polarized Communication Across the T Cell Immunological Synapse. *Annual Review of Cell and Developmental Biology*, 32(1), 303–325. <https://doi.org/10.1146/annurev-cellbio-100814-125330>
- Eason, D. D., Cannon, J. P., Haire, R. N., Rast, J. P., Ostrov, D. A., & Litman, G. W. (2004). Mechanisms of antigen receptor evolution. *Seminars in Immunology*, 16(4), 215–226. <https://doi.org/10.1016/j.smim.2004.08.001>
- Espinosa-de Aquino, W., Bailey, M., & Acevedo-Whitehouse, K. (2017). Changes in mucosal and serum immunoglobulin levels of California sea lions during early development. *Developmental & Comparative Immunology*, 77, 166–173. <https://doi.org/10.1016/j.dci.2017.08.006>
- Felsenstein, J. (1985). Confidence Limits On Phylogenies: An approach Using The Bootstrap. *Evolution*, 39(4), 783–791. <https://doi.org/10.1111/j.1558-5646.1985.tb00420.x>
- Flajnik, M. F., & Kasahara, M. (2010). Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nature Reviews Genetics*, 11(1), 47–59. <https://doi.org/10.1038/nrg2703>
- Fremont, D. H., Hendrickson, W. A., Marrack, P., & Kappler, J. (1997). HISTOCOMPATIBILITY ANTIGEN. Protein Data Bank, Rutgers University. <https://doi.org/10.2210/pdb1ieb/pdb>
- Gentry, R. (2014). *Behavior and Ecology of the Northern Fur Seal*. Princeton: Princeton University Press.
- Giudicelli, V. (2004). IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Research*, 33(Database issue), D256–D261. <https://doi.org/10.1093/nar/gki010>
- Gras, S., Burrows, S. R., Kjer-Nielsen, L., Clements, C. S., Liu, Y. C., Sullivan, L. C., ... Rossjohn, J. (2009). Crystal Structure of CF34 TCR in complex with HLA-B8/FLR. Protein Data Bank, Rutgers University. <https://doi.org/10.2210/pdb3ffc/pdb>
- Guo, Y., Bao, Y., Wang, H., Hu, X., Zhao, Z., Li, N., & Zhao, Y. (2011). A Preliminary Analysis of the Immunoglobulin Genes in the African Elephant (*Loxodonta africana*). *PLoS ONE*, 6(2), e16889. <https://doi.org/10.1371/journal.pone.0016889>

- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8(8), 1494–1512. <https://doi.org/10.1038/nprot.2013.084>
- Hoernes, M., Seger, R., & Reichenbach, J. (2011). Modern management of primary B-cell immunodeficiencies. *Pediatric Allergy and Immunology*, 22(8), 758–769. <https://doi.org/10.1111/j.1399-3038.2011.01236.x>
- Hoffman, W., Lakkis, F. G., & Chalasani, G. (2015). B Cells, Antibodies, and More. *Clinical Journal of the American Society of Nephrology*, 11(1), 137–154. <https://doi.org/10.2215/cjn.09430915>
- Holder, A., Mirczuk, S. M., Fowkes, R. C., Palmer, D. B., Aspinall, R., & Catchpole, B. (2018). Perturbation of the T cell receptor repertoire occurs with increasing age in dogs. *Developmental & Comparative Immunology*, 79, 150–157. <https://doi.org/10.1016/j.dci.2017.10.020>
- Holland, S. J., Gao, M., Hirano, M., Iyer, L. M., Luo, M., Schorpp, M., ... Boehm, T. (2014). Selection of the lamprey VLRC antigen receptor repertoire. *Proceedings of the National Academy of Sciences*, 111(41), 14834–14839. <https://doi.org/10.1073/pnas.1415655111>
- Hou, D., Chen, C., Seely, E. J., Chen, S., & Song, Y. (2016). High-Throughput Sequencing-Based Immune Repertoire Study during Infectious Disease. *Frontiers in Immunology*, 7. <https://doi.org/10.3389/fimmu.2016.00336>
- Johnson, S., Lowenstine, L., Gulland, F., Jang, S., Imai, D., Almy, F., Gardner, I. (2006). Aerobic bacterial flora of the vagina and prepuce of California sea lions (*Zalophus californianus*) and investigation of associations with urogenital carcinoma. *Veterinary Microbiology*, 114(1–2), 94–103. <https://doi.org/10.1016/j.vetmic.2005.11.045>
- Kapitonov, V. V., & Jurka, J. (2005). RAG1 Core and V(D)J Recombination Signal Sequences Were Derived from Transib Transposons. *PLoS Biology*, 3(6), e181. <https://doi.org/10.1371/journal.pbio.0030181>
- Kaplan, M. H., Hufford, M. M., & Olson, M. R. (2015). The development and in vivo function of T helper 9 cells. *Nature Reviews Immunology*, 15(5), 295–307. <https://doi.org/10.1038/nri3824>

- Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, 30(4), 772–780. <https://doi.org/10.1093/molbev/mst010>
- Kaushik, A. & Pasman, Y. (2014). *Comparative immunoglobulin genetics*. Toronto: Apple Academic Press.
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, 16(2), 111–120. <https://doi.org/10.1007/bf01731581>
- King, D. P., Hure, M. C., Goldstein, T., Aldridge, B. M., Gulland, F. M. ., Saliki, J. T., Stott, J. L. (2002). Otarine herpesvirus-1: a novel gammaherpesvirus associated with urogenital carcinoma in California sea lions (*Zalophus californianus*). *Veterinary Microbiology*, 86(1–2), 131–137. [https://doi.org/10.1016/s0378-1135\(01\)00497-7](https://doi.org/10.1016/s0378-1135(01)00497-7)
- Korbie, D. J., & Mattick, J. S. (2008). Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nature Protocols*, 3(9), 1452–1456. <https://doi.org/10.1038/nprot.2008.133>
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, 35(6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Kumar, S., Stecher, G., Suleski, M., & Hedges, S. B. (2017). TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Molecular Biology and Evolution*, 34(7), 1812–1819. <https://doi.org/10.1093/molbev/msx116>
- Kuzmina, T. A., Spraker, T. R., Kudlai, O., Lisitsyna, O. I., Zabludovskaja, S. O., Karbowski, G., Kuchta, R. (2018). Metazoan parasites of California sea lions (*Zalophus californianus*): A new data and review. *International Journal for Parasitology: Parasites and Wildlife*, 7(3), 326–334. <https://doi.org/10.1016/j.ijppaw.2018.09.001>
- Kyewski, B., & Feuerer, M. (2014). Love Is in the Aire: mTECs Share Their Assets. *Immunity*, 41(3), 343–345. <https://doi.org/10.1016/j.immuni.2014.08.013>
- Laird, D., De Tomaso, A., Cooper, M., & Weissman, I. (2000). 50 million years of chordate evolution: Seeking the origins of adaptive immunity. *Proceedings of*

the National Academy of Sciences, 97(13), 6924-6926. doi: 10.1073/pnas.97.13.6924

- Lang, A. S., Rise, M. L., Culley, A. I., & Steward, G. F. (2009). RNA viruses in the sea. *FEMS Microbiology Reviews*, 33(2), 295–323. <https://doi.org/10.1111/j.1574-6976.2008.00132.x>
- LeBien, T. W., & Tedder, T. F. (2008). B lymphocytes: how they develop and function. *Blood*, 112(5), 1570–1580. <https://doi.org/10.1182/blood-2008-02-078071>
- Lefranc, M.-P. (2001). IMGT, the international ImMunoGeneTics database. *Nucleic Acids Research*, 29(1), 207–209. <https://doi.org/10.1093/nar/29.1.207>
- Lefranc, M.P. (2001). Nomenclature of the Human Immunoglobulin Genes. *Current Protocols in Immunology*. John Wiley & Sons, Inc. <https://doi.org/10.1002/0471142735.ima01ps40>
- Lefranc, M.P., Giudicelli, V., Duroux, P., Jabado-Michaloud, J., Folch, G., Aouinti, S., ... Kossida, S. (2014). IMGT®, the international ImMunoGeneTics information system® 25 years on. *Nucleic Acids Research*, 43(D1), D413–D422. <https://doi.org/10.1093/nar/gku1056>
- Lefranc, M.-P., Giudicelli, V., Ginestoux, C., Bodmer, J., Muller, W., Bontrop, R., ... Chaume, D. (1999). IMGT, the international ImMunoGeneTics database. *Nucleic Acids Research*, 27(1), 209–212. <https://doi.org/10.1093/nar/27.1.209>
- Li, L., Rong, X., Li, G., Wang, Y., Chen, B., Ren, W., ... Xu, S. (2018). Genomic organization and adaptive evolution of IGHC genes in marine mammals. *Molecular Immunology*, 99, 75–81. <https://doi.org/10.1016/j.molimm.2018.04.011>
- Linguiti, G., Antonacci, R., Tasco, G., Grande, F., Casadio, R., Massari, S., ... Ciccacese, S. (2016). Genomic and expression analyses of *Tursiops truncatus* T cell receptor gamma (TRG) and alpha/delta (TRA/TRD) loci reveal a similar basic public $\gamma\delta$ repertoire in dolphin and human. *BMC Genomics*, 17(1). <https://doi.org/10.1186/s12864-016-2841-9>
- Litman, G. W., Cannon, J. P., & Dishaw, L. J. (2005). Reconstructing immune phylogeny: new perspectives. *Nature Reviews Immunology*, 5(11), 866–879. <https://doi.org/10.1038/nri1712>

- Lloyd-Smith, J. O., Greig, D. J., Hietala, S., Ghneim, G. S., Palmer, L., St Leger, J., Gulland, F. M. (2007). Cyclical changes in seroprevalence of leptospirosis in California sea lions: endemic and epidemic disease in one host species? *BMC Infectious Diseases*, 7(1). <https://doi.org/10.1186/1471-2334-7-125>
- Longo, A. V., Burrowes, P. A., & Zamudio, K. R. (2014). Genomic Studies of Disease-Outcome in Host-Pathogen Dynamics. *Integrative and Comparative Biology*, 54(3), 427–438. <https://doi.org/10.1093/icb/ ICU073>
- Louten, J., Boniface, K., & de Waal Malefyt, R. (2009). Development and function of TH17 cells in health and disease. *Journal of Allergy and Clinical Immunology*, 123(5), 1004–1011. <https://doi.org/10.1016/j.jaci.2009.04.003>
- Lyons, E. T., DeLong, R. L., Melin, S. R., & Tolliver, S. C. (1997). Uncinariasis in Northern fur seal and California sea lion pups from California. *Journal of Wildlife Diseases*, 33(4), 848–852. <https://doi.org/10.7589/0090-3558-33.4.848>
- Madi, A., Poran, A., Shifrut, E., Reich-Zeliger, S., Greenstein, E., Zaretsky, I., ... Friedman, N. (2017). T cell receptor repertoires of mice and humans are clustered in similarity networks around conserved public CDR3 sequences. *eLife*, 6. <https://doi.org/10.7554/elife.22057>
- Maldonado, C., Molina, C. I., Zizka, A., Persson, C., Taylor, C. M., Albán, J., ... Antonelli, A. (2015). Estimating species diversity and distribution in the era of Big Data: to what extent can we trust public databases? *Global Ecology and Biogeography*, 24(8), 973–984. <https://doi.org/10.1111/geb.12326>
- Mantis, N. J., Rol, N., & Corthésy, B. (2011). Secretory IgA's complex roles in immunity and mucosal homeostasis in the gut. *Mucosal Immunology*, 4(6), 603–611. <https://doi.org/10.1038/mi.2011.41>
- Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., ... Bryant, S. H. (2016). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research*, 45(D1), D200–D203. <https://doi.org/10.1093/nar/gkw1129>
- Marrack, P., Scott-Browne, J. P., Dai, S., Gapin, L., & Kappler, J. W. (2008). Evolutionarily Conserved Amino Acids That Control TCR-MHC Interaction. *Annual Review of Immunology*, 26(1), 171–203. <https://doi.org/10.1146/annurev.immunol.26.021607.090421>

- Martin, F., Oliver, A. M., & Kearney, J. F. (2001). Marginal Zone and B1 B Cells Unite in the Early Response against T-Independent Blood-Borne Particulate Antigens. *Immunity*, 14(5), 617–629. [https://doi.org/10.1016/s1074-7613\(01\)00129-7](https://doi.org/10.1016/s1074-7613(01)00129-7)
- Martin, J., Ponstingl, H., Lefranc, M.-P., Archer, J., Sargan, D., & Bradley, A. (2017). Comprehensive annotation and evolutionary insights into the canine (*Canis lupus familiaris*) antigen receptor loci. *Immunogenetics*, 70(4), 223–236. <https://doi.org/10.1007/s00251-017-1028-0>
- Mashoof, S., & Criscitiello, M. (2016). Fish Immunoglobulins. *Biology*, 5(4), 45. <https://doi.org/10.3390/biology5040045>
- Meyer-Olson, D., Brady, K. W., Blackard, J. T., Allen, T. M., Islam, S., Shoukry, N. H., Kalams, S. A. (2003). Analysis of the TCR β Variable Gene Repertoire in Chimpanzees: Identification of Functional Homologs to Human Pseudogenes. *The Journal of Immunology*, 170(8), 4161–4169. <https://doi.org/10.4049/jimmunol.170.8.4161>
- Migalska, M., Sebastian, A., & Radwan, J. (2018). Profiling of the TCR β repertoire in non-model species using high-throughput sequencing. *Scientific Reports*, 8(1). <https://doi.org/10.1038/s41598-018-30037-0>
- Mineccia, M., Massari, S., Linguiti, G., Ceci, L., Ciccarese, S., & Antonacci, R. (2012). New insight into the genomic structure of dog T cell receptor beta (TRB) locus inferred from expression analysis. *Developmental & Comparative Immunology*, 37(2), 279–293. <https://doi.org/10.1016/j.dci.2012.03.010>
- Moncada, C. A., Guerrero, E., Cardenas, P., Suarez, C. F., Patarroyo, M. E., & Patarroyo, M. A. (2005). The T-cell receptor in primates: identifying and sequencing new owl monkey TRBV gene sub-groups. *Immunogenetics*, 57(1–2), 42–52. <https://doi.org/10.1007/s00251-004-0758-y>
- Montano-Frías, J. E., Vera-Massieu, C., Álvarez-Martínez, R., Flores-Morán, A., & Acevedo-Whitehouse, K. (2016). MHC class II transcription is associated with inflammatory responses in a wild marine mammal. *Infection, Genetics and Evolution*, 42, 77–82. <https://doi.org/10.1016/j.meegid.2016.04.022>
- Mori, A., Deola, S., Xumerle, L., Mijatovic, V., Malerba, G., & Monsurrò, V. (2013). Next generation sequencing: new tools in immunology and hematology. *Blood Research*, 48(4), 242. <https://doi.org/10.5045/br.2013.48.4.242>

- Murphy, K. & Weaver, C. (2017). *Janeway's immunobiology*. (9th ed.). New York: Garland Science, 93(1), 59–59. <https://doi.org/10.1086/696793>
- Nanda, N. K., Apple, R., & Sercarz, E. (1991). Limitations in plasticity of the T-cell receptor repertoire. *Proceedings of the National Academy of Sciences*, 88(21), 9503–9507. <https://doi.org/10.1073/pnas.88.21.9503>
- Nei, M. (2007). The new mutation theory of phenotypic evolution. *Proceedings of the National Academy of Sciences*, 104(30), 12235–12242. <https://doi.org/10.1073/pnas.0703349104>
- Nei, M., Gu, X., & Sitnikova, T. (1997). Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proceedings of the National Academy of Sciences*, 94(15), 7799–7806. <https://doi.org/10.1073/pnas.94.15.7799>
- Ng, T. F. F., Suedmeyer, W. K., Wheeler, E., Gulland, F., & Breitbart, M. (2009). Novel anellovirus discovered from a mortality event of captive California sea lions. *Journal of General Virology*, 90(5), 1256–1261. <https://doi.org/10.1099/vir.0.008987-0>
- Nikolich-Žugich, J., Slifka, M. K., & Messaoudi, I. (2004). The many important facets of T-cell repertoire diversity. *Nature Reviews Immunology*, 4(2), 123–132. <https://doi.org/10.1038/nri1292>
- Olivieri, D. N., Gambón-Cerdá, S., & Gambón-Deza, F. (2015). Evolution of V genes from the TRV loci of mammals. *Immunogenetics*, 67(7), 371–384. <https://doi.org/10.1007/s00251-015-0850-5>
- Olivieri, D. N., von Haefen, B., Sánchez-Espinel, C., & Gambón-Deza, F. (2014). The immunologic V-gene repertoire in mammals. Cold Spring Harbor Laboratory. <https://doi.org/10.1101/002667>
- Olivieri, D., Faro, J., von Haefen, B., Sánchez-Espinel, C., & Gambón-Deza, F. (2013). An automated algorithm for extracting functional immunologic V-genes from genomes in jawed vertebrates. *Immunogenetics*, 65(9), 691–702. <https://doi.org/10.1007/s00251-013-0715-8>
- Ostrander, E. A. (2005). The canine genome. *Genome Research*, 15(12), 1706–1716. <https://doi.org/10.1101/gr.3736605>

- Outters, P., Jaeger, S., Zaarour, N., & Ferrier, P. (2015). Long-Range Control of V(D)J Recombination & Allelic Exclusion. In *Advances in Immunology* (pp. 363–413). Elsevier. <https://doi.org/10.1016/bs.ai.2015.08.002>
- Pallarès, N., Lefebvre, S., Contet, V., Matsuda, F., & Lefranc, M.-P. (1999). The Human Immunoglobulin Heavy Variable Genes. *Experimental and Clinical Immunogenetics*, 16(1), 36–60. <https://doi.org/10.1159/000019095>
- Parra, Z. E., Baker, M. L., Schwarz, R. S., Deakin, J. E., Lindblad-Toh, K., & Miller, R. D. (2007). A unique T cell receptor discovered in marsupials. *Proceedings of the National Academy of Sciences*, 104(23), 9776–9781. <https://doi.org/10.1073/pnas.060910610>
- Paul, W. E., & Zhu, J. (2010). How are TH2-type immune responses initiated and amplified? *Nature Reviews Immunology*, 10(4), 225–235. <https://doi.org/10.1038/nri2735>
- Pederson, T. (2015). Molecular Biology of the Gene. *The FASEB Journal*, 29(11), 4399–4401. <https://doi.org/10.1096/fj.15-1101ufm>
- Piccirillo, C. A. (2008). Regulatory T cells in health and disease. *Cytokine*, 43(3), 395–401. <https://doi.org/10.1016/j.cyto.2008.07.469>
- Ponganis, P. J., Kooyman, G. L., Sartoris, D., & Jobsis, P. (1992). Pinniped splenic volumes. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 262(2), R322–R325. <https://doi.org/10.1152/ajpregu.1992.262.2.r322>
- Preciado-Llanes, L., Wing, J. B., Foster, R. A., Carling, J., Lees, A., Read, R. C., & Heath, A. W. (2014). Contact-dependent suppression of CD4 T-cell activation and proliferation by B cells activated through IgD cross-linking. *Immunology*, 144(3), 444–452. <https://doi.org/10.1111/imm.12389>
- Proudhon, C., Hao, B., Raviram, R., Chaumeil, J., & Skok, J. A. (2015). Long-Range Regulation of V(D)J Recombination. In *Advances in Immunology* (pp. 123–182). Elsevier. <https://doi.org/10.1016/bs.ai.2015.07.003>
- Qin, T., Zhao, H., Zhu, H., Wang, D., Du, W., & Hao, H. (2015). Immunoglobulin genomics in the prairie vole (*Microtus ochrogaster*). *Immunology Letters*, 166(2), 79–86. <https://doi.org/10.1016/j.imlet.2015.06.001>
- Qin, T., Zhu, H., Wang, D., Hao, H., & Du, W. (2014). Genomic Organization and Expression of Immunoglobulin Genes in the Chinese Hamster (*Cricetulus*

- griseus*). Scandinavian Journal of Immunology, 81(1), 11–22. <https://doi.org/10.1111/sji.12243>
- Reinink, P., & Van Rhijn, I. (2009). The bovine T cell receptor alpha/delta locus contains over 400 V genes and encodes V genes without CDR2. Immunogenetics, 61(7), 541–549. <https://doi.org/10.1007/s00251-009-0384-9>
- Riedman, M. (1990). The pinnipeds: seals, sea lions, and walruses. University of California Press, Berkeley, CA.
- Rivera, R., Robles-Sikisaka, R., Hoffman, E. M., Stacy, B. A., Jensen, E. D., Nollens, H. H., & Wellehan, J. F. X., Jr. (2012). Characterization of a novel papillomavirus species (ZcPV1) from two California sea lions (*Zalophus californianus*). Veterinary Microbiology, 155(2–4), 257–266. <https://doi.org/10.1016/j.vetmic.2011.09.027>
- Robinson, J., Halliwell, J., McWilliam, H., Lopez, R., Parham, P., & Marsh, S. (2012). The IMGT/HLA database. Nucleic Acids Research, 41(D1), D1222–D1227. doi: 10.1093/nar/gks949
- Rosati, E., Dowds, C. M., Liaskou, E., Henriksen, E. K. K., Karlsen, T. H., & Franke, A. (2017). Overview of methodologies for T-cell receptor repertoire analysis. BMC Biotechnology, 17(1). <https://doi.org/10.1186/s12896-017-0379-9>
- Rossjohn, J., Gras, S., Miles, J. J., Turner, S. J., Godfrey, D. I., & McCluskey, J. (2015). T Cell Antigen Receptor Recognition of Antigen-Presenting Molecules. Annual Review of Immunology, 33(1), 169–200. <https://doi.org/10.1146/annurev-immunol-032414-112334>
- Sato, Y., Weiss, E., & Rochel, N. (2014). Structure of human Gankyrin in complex to the single chain antibody F5. Protein Data Bank, Rutgers University. <https://doi.org/10.2210/pdb4nik/pdb>
- Schloss, P. D. (2010). The Effects of Alignment Quality, Distance Calculation Method, Sequence Filtering, and Region on the Analysis of 16S rRNA Gene-Based Studies. PLoS Computational Biology, 6(7), e1000844. <https://doi.org/10.1371/journal.pcbi.1000844>
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Weber, C. F. (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing

- Microbial Communities. *Applied and Environmental Microbiology*, 75(23), 7537–7541. <https://doi.org/10.1128/aem.01541-09>
- Schlötterer, C., Kofler, R., Versace, E., Tobler, R., & Franssen, S. U. (2016). Erratum: Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. *Heredity*, 116(2), 248–248. <https://doi.org/10.1038/hdy.2015.85>
- Schusterman, R. J., Gentry, R., & Schmook, J. (1966). Underwater Vocalization by Sea Lions: Social and Mirror Stimuli. *Science*, 154(3748), 540–542. <https://doi.org/10.1126/science.154.3748.540>
- Schwartz, S. (2003). Human-Mouse Alignments with BLASTZ. *Genome Research*, 13(1), 103-107. doi: 10.1101/gr.809403
- Scotto–Lavino, E., Du, G., & Frohman, M. A. (2006). 5' end cDNA amplification using classic RACE. *Nature Protocols*, 1(6), 2555–2562. <https://doi.org/10.1038/nprot.2006.480>
- Sellars, M. (2011). Ikaros in B cell development and function. *World Journal of Biological Chemistry*, 2(6), 132. <https://doi.org/10.4331/wjbc.v2.i6.132>
- Siepel, A. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, 15(8), 1034-1050. doi: 10.1101/gr.3715005
- Siepel, A., & Haussler, D (2005). Phylogenetic Hidden Markov Models. *Statistical Methods In Molecular Evolution*, 325-351. doi: 10.1007/0-387-27733-1_12
- Silverstein, A., Silverstein, V., & Silverstein, R. (1995). *The sea otter*. Brookfield, Conn.: Millbrook Press.
- Szteren, D. (2011). Ecological regionalization of *Zalophus californianus* rookeries, as a tool for conservation in the Gulf of California. *Ciencias Marinas*, 37(3):349-368.
- Szteren, D., Aurioles, D., & Gerber, L. (2006). Population status and trends of the California sea lion (*Zalophus californianus californianus*) in the Gulf of California, Mexico. *Sea Lions of the World*. doi: 10.4027/slw.2006.25
- Tautz, D., & Renz, M. (1983). An optimized freeze-squeeze method for the recovery of DNA fragments from agarose gels. *Analytical Biochemistry*, 132(1), 14–19. [https://doi.org/10.1016/0003-2697\(83\)90419-0](https://doi.org/10.1016/0003-2697(83)90419-0)

- Tian, R., Chen, M., Chai, S., Rong, X., Chen, B., Ren, W., Yang, G. (2018). Divergent Selection of Pattern Recognition Receptors in Mammals with Different Ecological Characteristics. *Journal of Molecular Evolution*, 86(2), 138–149. <https://doi.org/10.1007/s00239-018-9832-1>
- Treanor, B. (2012). B-cell receptor: from resting state to activate. *Immunology*, 136(1), 21–27. <https://doi.org/10.1111/j.1365-2567.2012.03564.x>
- Uhen, M. D. (2007). Evolution of marine mammals: Back to the sea after 300 million years. *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology*, 290(6), 514–522. <https://doi.org/10.1002/ar.20545>
- Uhen, M. D. (2010). The Origin(s) of Whales. *Annual Review of Earth and Planetary Sciences*, 38(1), 189–219. <https://doi.org/10.1146/annurev-earth-040809-152453>
- Ujvari, B., & Belov, K. (2015). Characterization of antibody V segment diversity in the Tasmanian devil (*Sarcophilus harrisii*). *Veterinary Immunology and Immunopathology*, 167(3–4), 156–165. <https://doi.org/10.1016/j.vetimm.2015.08.001>
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Research*, 40(15), e115–e115. <https://doi.org/10.1093/nar/gks59>
- Waddell, P. J., Cao, Y., Hasegawa, M., & Mindell, D. P. (1999). Assessing the Cretaceous Superordinal Divergence Times within Birds and Placental Mammals by Using Whole Mitochondrial Protein Sequences and an Extended Statistical Framework. *Systematic Biology*, 48(1), 119–137. <https://doi.org/10.1080/106351599260481>
- Wang, Q. J., Hanada, K., Robbins, P. F., Li, Y. F., & Yang, J. C. (2012). Distinctive Features of the Differentiated Phenotype and Infiltration of Tumor-Reactive Lymphocytes in Clear Cell Renal Cell Carcinoma. *Cancer Research*, 72(23), 6119–6129. <https://doi.org/10.1158/0008-5472.can-12-0588>
- Wang, X., Parra, Z. E., & Miller, R. D. (2011). Platypus TCR μ Provides Insight into the Origins and Evolution of a Uniquely Mammalian TCR Locus. *The Journal of Immunology*, 187(10), 5246–5254. <https://doi.org/10.4049/jimmunol.1101113>

- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46(W1), W296–W303. <https://doi.org/10.1093/nar/gky427>
- Watson, F. (2005). Extensive Diversity of Ig-Superfamily Proteins in the Immune System of Insects. *Science*, 309(5742), 1874-1878. doi: 10.1126/science.1116887
- What We Thought: Mating system of the California Sea Lion. (2019). Retrieved 12 October 2019, from https://www.reed.edu/biology/courses/BIO342/2014_syllabus_old/2014_WEBSITES/Spaeth_SeaLion_Final/what_we_thought.html
- Wong, G. K., Millar, D., Penny, S., Heather, J. M., Mistry, P., Buettner, N., ... Cobbold, M. (2016). Accelerated Loss of TCR Repertoire Diversity in Common Variable Immunodeficiency. *The Journal of Immunology*, 197(5), 1642–1649. <https://doi.org/10.4049/jimmunol.1600526>
- Yang, J., & Reth, M. (2010). Oligomeric organization of the B-cell antigen receptor on resting cells. *Nature*, 467(7314), 465–469. <https://doi.org/10.1038/nature09357>
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2014). The I-TASSER Suite: protein structure and function prediction. *Nature Methods*, 12(1), 7–8. <https://doi.org/10.1038/nmeth.3213>
- Yeku, O., & Frohman, M. A. (2011). Rapid Amplification of cDNA Ends (RACE). In *RNA* (pp. 107–122). Humana Press. https://doi.org/10.1007/978-1-59745-248-9_8
- Zarin, P., Chen, E. L. Y., In, T. S. H., Anderson, M. K., & Zúñiga-Pflücker, J. C. (2015). Gamma delta T-cell differentiation and effector function programming, TCR signal strength, when and how much? *Cellular Immunology*, 296(1), 70–75. <https://doi.org/10.1016/j.cellimm.2015.03.007>
- Zhang, S.M. (2004). Diversification of Ig Superfamily Genes in an Invertebrate. *Science*, 305(5681), 251–254. <https://doi.org/10.1126/science.1088069>
- Zhang, Z., Mu, Y., Shan, L., Sun, D., Guo, W., Yu, Z., Yang, G. (2019). Divergent Evolution of TRC Genes in Mammalian Niche Adaptation. *Frontiers in Immunology*, 10. <https://doi.org/10.3389/fimmu.2019.00871>

Anexos

Tabla A1

Lista de primers usados en el estudio

Cadena	Nombre de primer	Secuencia del primer	Primer pb	Tm	Amplicon pb
TCR alpha	AlphaAForward	[5'-ATCCCCTTTTCCCCCAAGATGTGG-3']	24	62,5	65
	AlphaAReverse	[5'-GCTCAGATAACAGCATAGCAGGGC-3']	24	62,5	65
	AlphaBForward	[5'-TGGGTCTCTTGGGCTCTAGCTCC-3']	24	64.7	108
	AlphaBReverse	[5'-CCCCACATCTTGGGGGAAAAGGG-3']	24	64.7	108
TCR beta	BetaCForward	[5'-GCGAGGGTTGTGCCAGAAGG-3']	20	64.7	214
	BetaCReverse	[5'-GATCTCCCGCACCCAGAAGG-3']	20	64.7	214
	BetaC2Forward	[5'-CCCCTGGAAGTGGACTTGGC-3']	20	63,2	130
	BetaC2Reverse	[5'-AGACCCGCAGCCCTACAAGG-3']	20	63,2	130
BCR heavy	HeavyHA-2Forward	[5'-TCAGGGTGGGTGGCTGTGCAGG-3']	22	69	179
	HeavyHA-2Reverse	[5'-GGCCTGAAAGACCCCAAGGGTGC-3']	24	69	179
	HeavyEForward	[5'-GCACAGCCTGGTAGGACTGG-3']	22	65	108
	HeavyEReverse	[5'-TGCCACCTTCACCTGGAACCC-3']	21	65	108
	HeavyG1Forward	[5'-CCATCCAGAAGAGCCCTGAGCG-3']	22	61.6	71
	HeavyG1Reverse	[5'-TGCACAGCCTGGCAGCACAC-3']	20	61.6	71
	HeavyG2Forward	[5'-CCAGAAGAGCCCTGAGCGTGA-3']	21	69	97
	HeavyG2Reverse-A2	[5'-GGAGAAGGTCTCCCCACGGT-3']	20	69	97
BCR lambda	LambdaAForward	[5'-TACCCAGCGGCGTGACGGT-3']	20	59	219
	LambdaAReverse	[5'-AGAGCACTCTGCGGGGGCCA-3']	20	59	219
	LambdaCForward	[5'-CCCAGCGGCGTGACAGTGGC-3']	20	61	184
	LambdaCReverse	[5'-CGGTGCTCCCCTCGTGCCTG-3']	20	61	184
	LambdaB2Forward	[5'-GGGCGCCAACAAGGCCACCC-3']	20	61	264
	LambdaAeverseA	[5'-AGAGCACTCTGCGGGGGCCA-3']	20	61	264

Tabla A1| Lista de primers usados en el estudio. La columna izquierda en colores muestra el nombre de la cadena que amplifica el primer, la segunda columna de izquierda a derecha muestra el nombre del primer, seguido por la secuencia del primer y el tamaño en pares de bases. La columna de Tm muestra la Temperatura de melting determinada por PCR de gradiente. La última columna de la derecha muestra el tamaño del amplicon en pares de bases determinado por PCR por punto final.

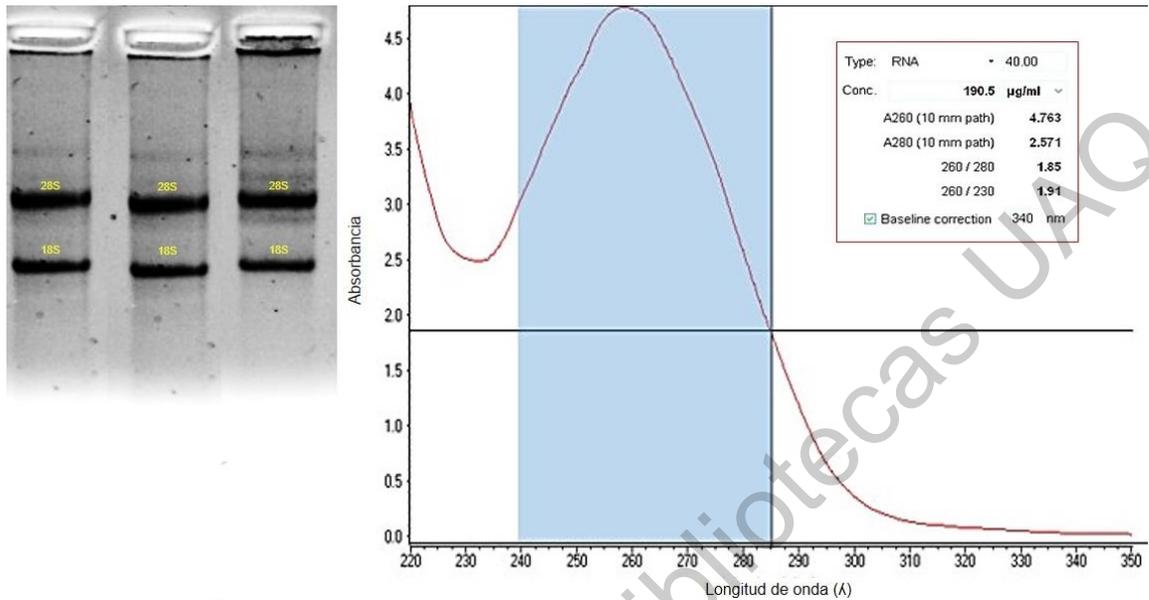


Figura A2| Ejemplificación del número de integridad del RNA. En la imagen se muestra un gel de agarosa al 1.2%, donde se observan las fracciones 28s y 18s de la extracción de RNA. También se muestra la lectura en el nanodrop de RNA. El cuadro azul en la gráfica muestra el pico en 240 y 280nm para determinar la lectura óptima del RNA que se utilizó durante todo el trabajo, además el cuadro color rojo ejemplifica las lecturas que se determinaron como óptimas con cada muestra de RNA.

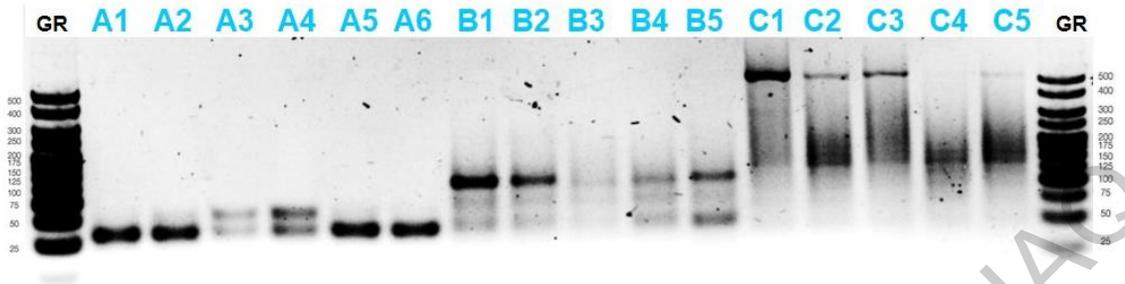


Figura A3| Gel del PCR de gradiente para optimizar la Tm de los primers diseñados para la cadena Alpha. El gel al 2% de agarosa, muestra las reacciones de PCR para los pares de primers Alpha-A (1-6), Alpha-B (1-5) y Alpha-C (1-5) para la cadena alpha del TCR de *Z. californianus* El marcador de peso molecular está indicado por (GR).

Dirección General de Bibliotecas UJA

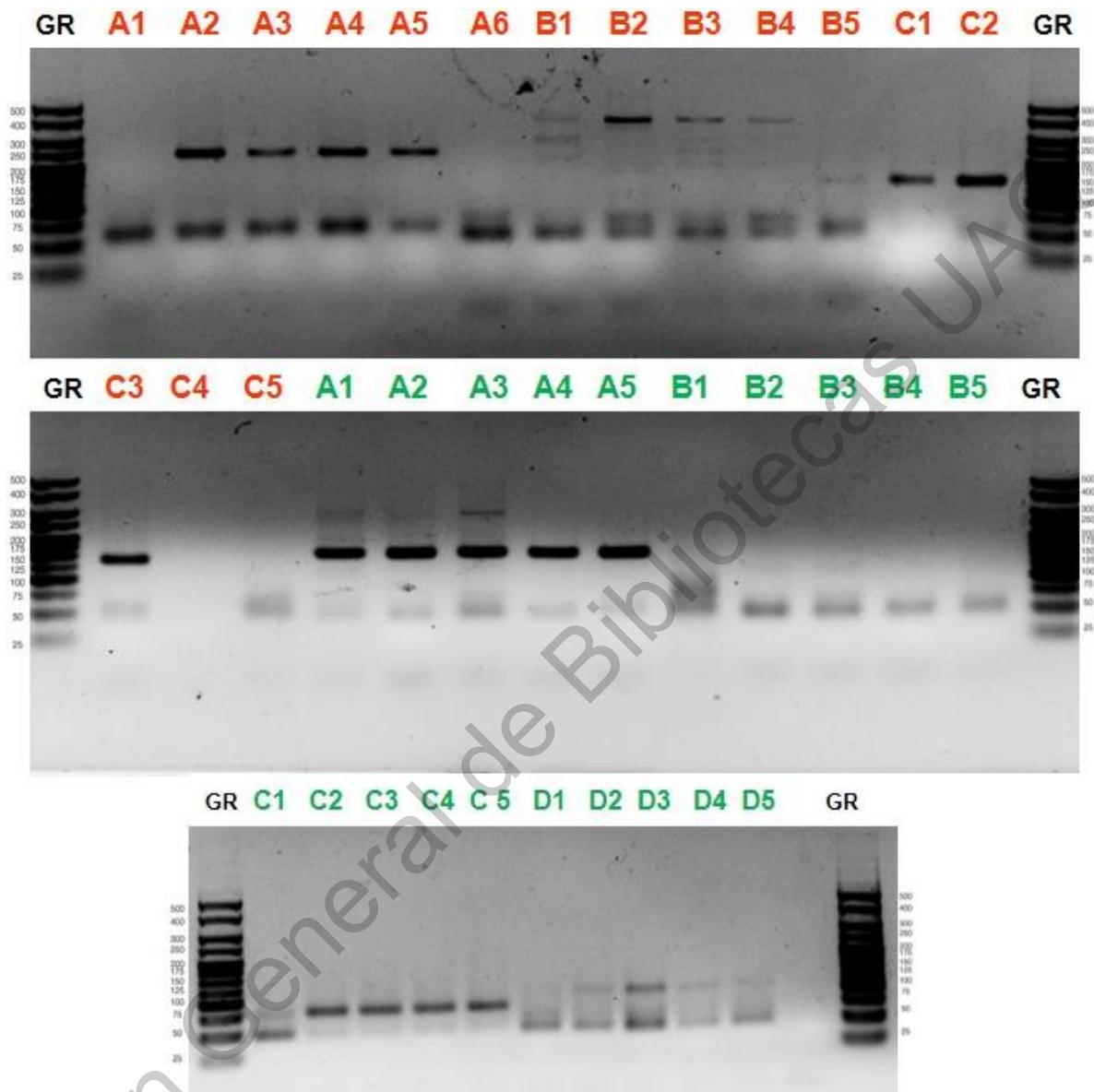


Figura A4| Gel del PCR de gradiente para optimizar la T_m de los primers diseñados para las cadenas Beta y Heavy. El gel al 2% de agarosa enumera en anaranjado las reacciones de PCR para los pares de primers Beta-A (1-6), Beta-B(1-5) y Beta-C (1-5), para la cadena beta del TCR de *Z. californianus*. En color verde se enumeran las reacciones de PCR para los pares de primers Heavy-A (1-5), Heavy-B (1-5), Heavy-C (1-5) y Heavy-D (1-5), para la cadena Heavy del BCR de *Z. californianus*. El marcador de peso molecular está indicado por (GR).

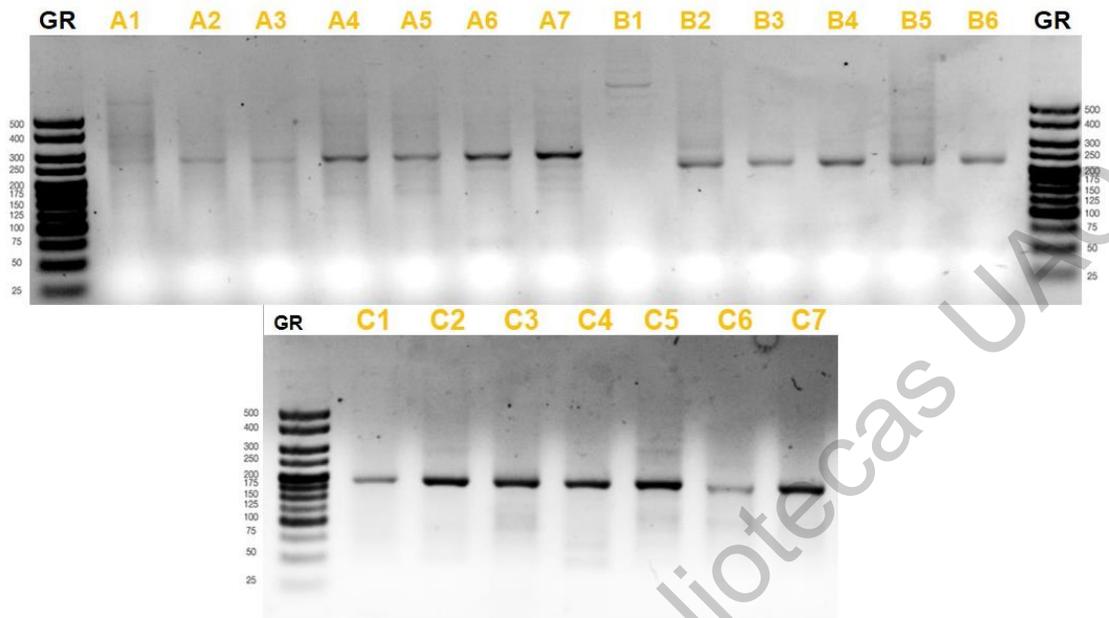


Figura A5| Gel del PCR de gradiente para optimizar la Tm de los primers diseñados para la cadena Lambda. El gel al 2% de agarosa, muestra las reacciones de PCR para los pares de primers Lambda-A (1-7), Lambda-B(1-6) y Lambda-C (1-7) para la cadena Lambda del BCR de *Z. californianus*. El marcador de peso molecular está indicado por (GR).

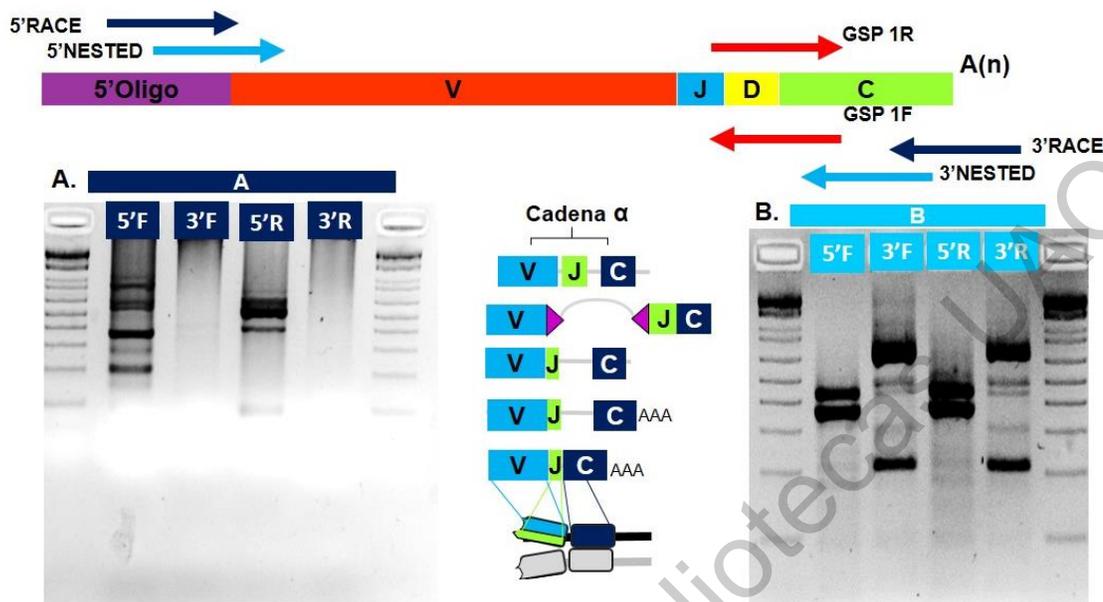


Figura A6| Geles de RACE y Nested de la cadena alpha. La imagen ejemplifica los sitios de hibridación de los primers 5'/3' Race, 5'/3' Nested y los primers específicos (GSP). En A se muestra el gel al 2% de agarosa de los productos de RACE (PCR Touchdown) con los primers 5'/3' Race y Forward y Reverse Alpha-A. En B se muestra el gel al 2% de agarosa de los productos de la PCR anidada (Nested) con los primers 5'/3' Nested y Forward y Reverse Alpha-B.

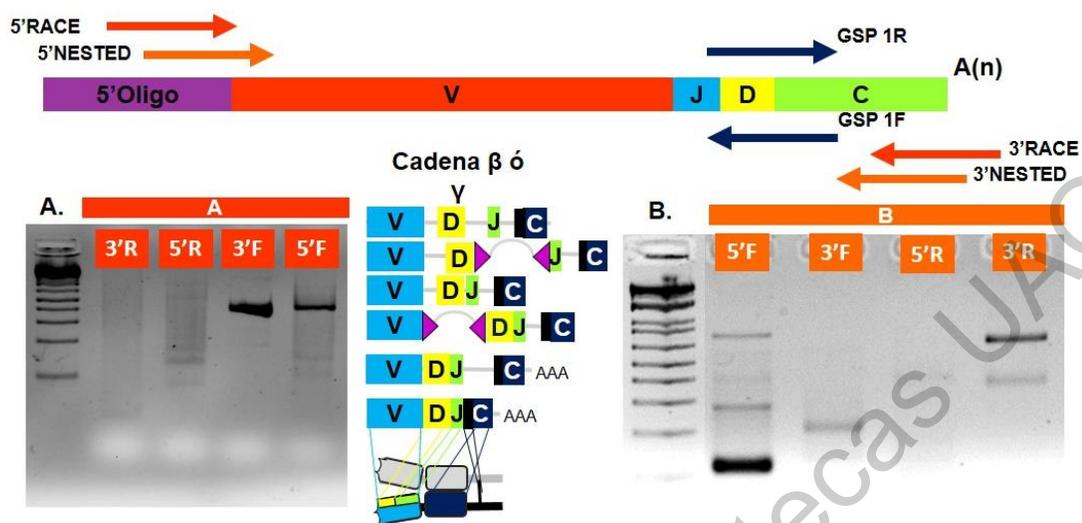


Figura A7| Geles de RACE y Nested de la cadena Beta. La imagen ejemplifica los sitios de hibridación de los primers 5'/3' Race, 5'/3' Nested y los primers específicos (GSP). En A se muestra el gel al 2% de agarosa de los productos de RACE (PCR Touchdown) con los primers 5'/3' Race y Forward y Reverse Beta-A. En B se muestra el gel al 2% de agarosa de los productos de la PCR anidada (Nested) con los primers 5'/3' Nested y Forward y Reverse Beta-B.

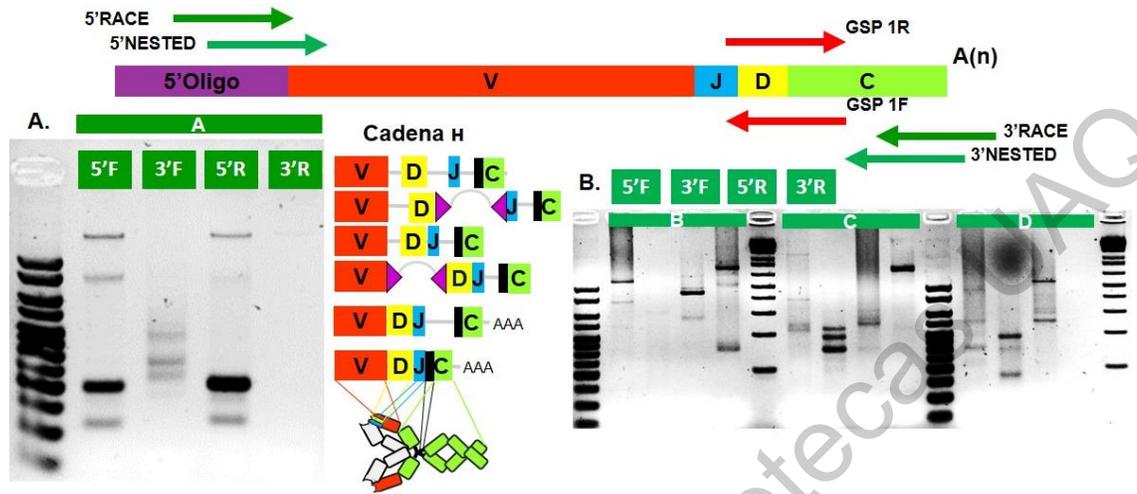


Figura A8| Geles de RACE y Nested de la cadena Heavy. La imagen ejemplifica los sitios de hibridación de los primers 5'/3' Race, 5'/3' Nested y los primers específicos (GSP). En A se muestra el gel al 2% de agarosa de los productos de RACE (PCR Touchdown) con los primers 5'/3' Race y Forward y Reverse Heavy-A. En B se muestra el gel al 2% de agarosa de los productos de la PCR anidada (Nested) con los primers 5'/3' Nested y Forward y Reverse Heavy-B, Heavy-C y Heavy-D.

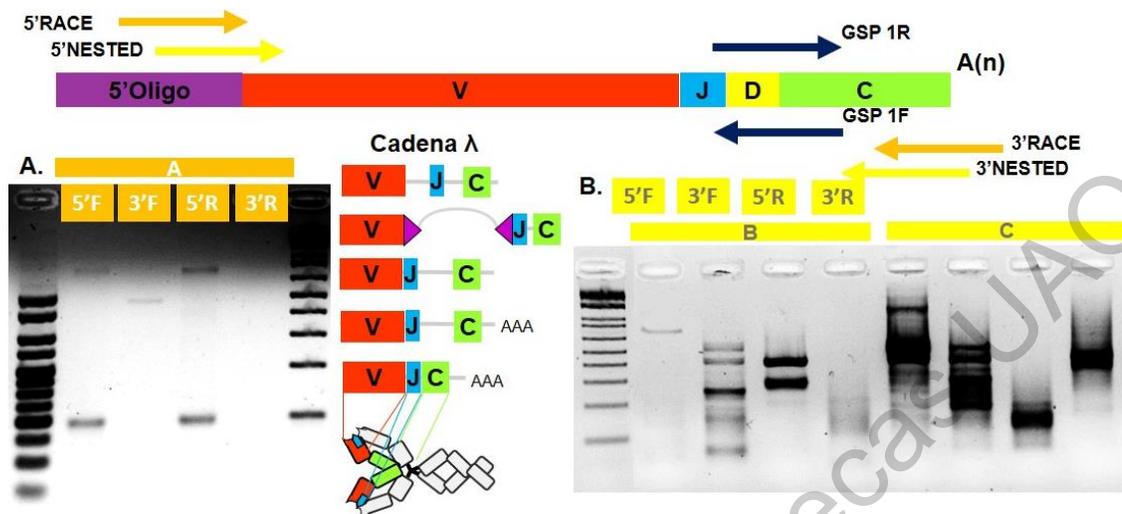


Figura A9| Geles de RACE y Nested de la cadena Lambda. La imagen ejemplifica los sitios de hibridación de los primers 5'/3' Race, 5'/3' Nested y los primers específicos (GSP). En A se muestra el gel al 2% de agarosa de los productos de RACE (PCR Touchdown) con los primers 5'/3' Race y Forward y Reverse Lambda-A. En B se muestra el gel al 2% de agarosa de los productos de la PCR anidada (Nested) con los primers 5'/3' Nested y Forward y Reverse Lambda-B y Lambda-C.

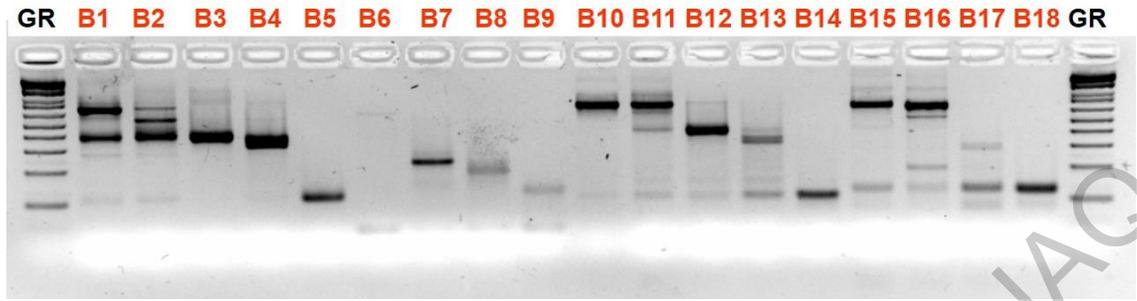


Figura A10| Gel del PCR punto final para aislar bandas únicas de la cadena Beta. Gel 2% de agarosa en la cual se muestran las bandas re-amplificadas provenientes del protocolo de Nested de la cadena Beta obtenidas por el protocolo Frezee and Squeeze (Tautz & Renz, 1983).

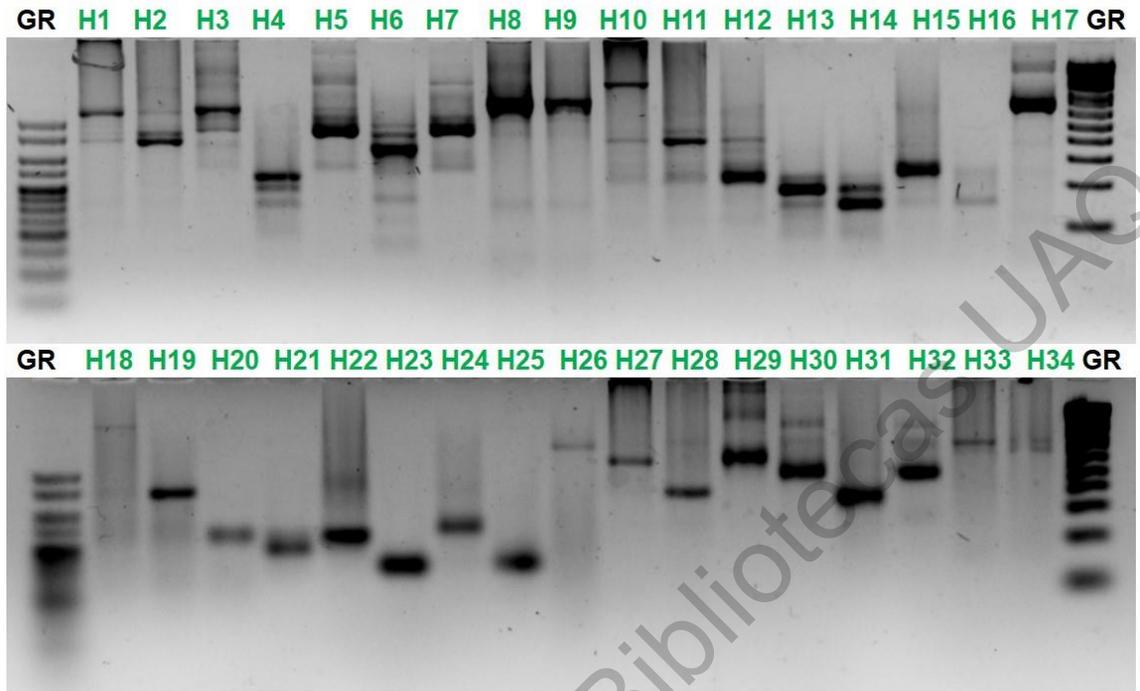


Figura A11| Gel del PCR punto final para aislar bandas únicas de la cadena Heavy. Gel 2% de agarosa en la cual se muestran las bandas reamplificadas provenientes del protocolo Nested de la cadena Heavy, obtenidas por el protocolo Freeze and Squeeze (Tautz & Renz, 1983).

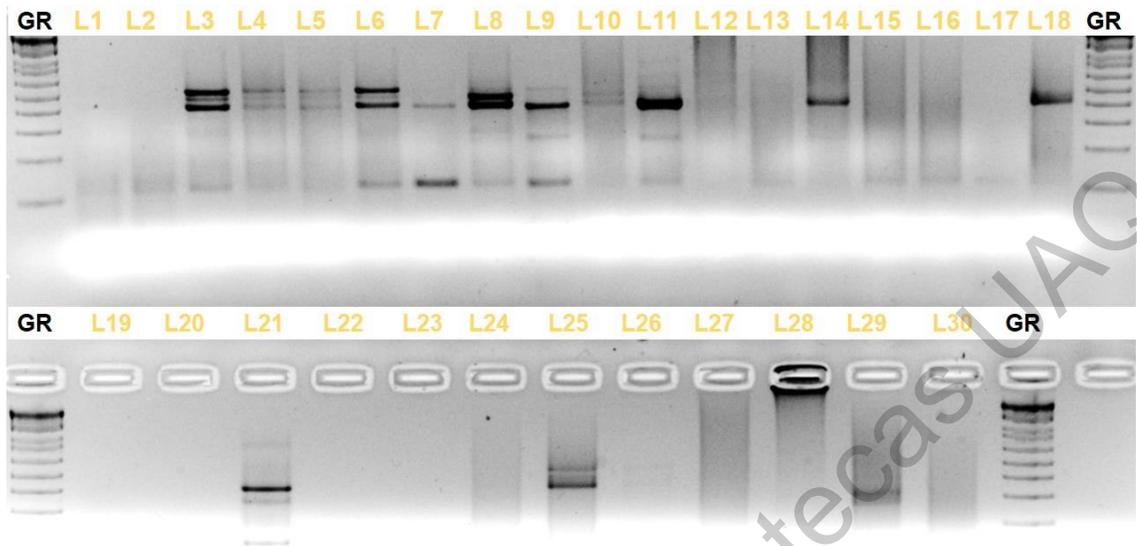


Figura A12| Gel del PCR punto final para aislar bandas únicas de la cadena Lambda. Gel 2% de agarosa en la cual se muestran las bandas re-amplificadas provenientes del protocolo Nested de la cadena Lambda, obtenidas por el protocolo Frezee and Squeeze (Tautz & Renz., 1983).

Dirección General de Bibliotecas UAQ

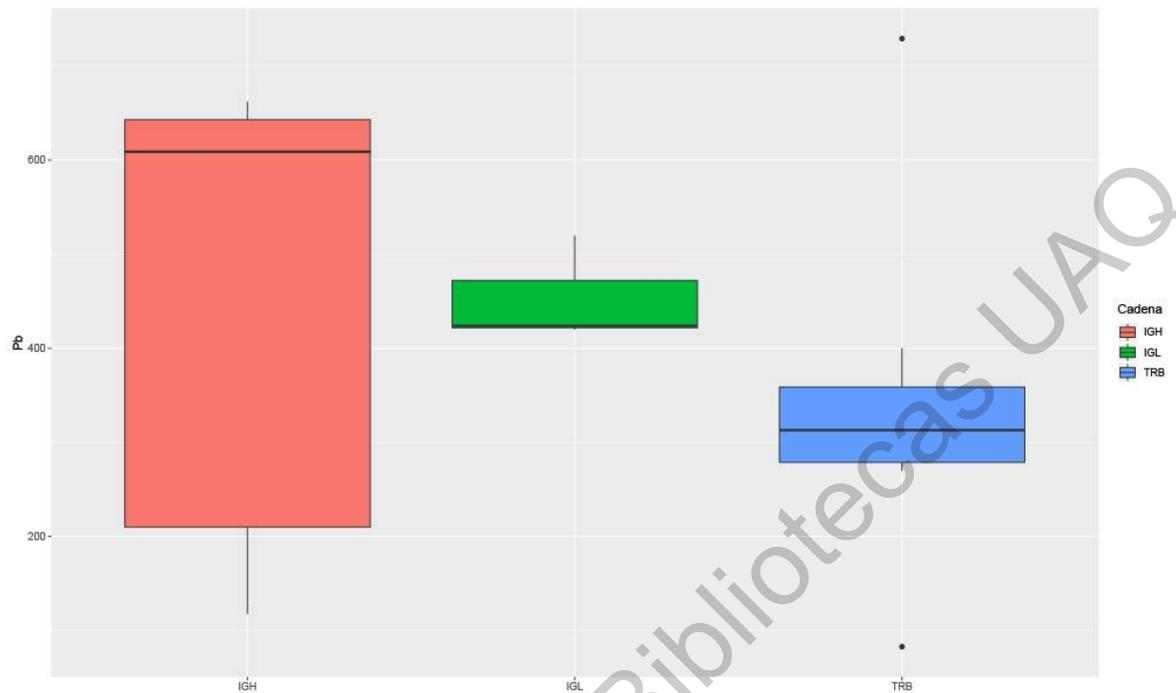


Figura A13| Largo medio de las secuencias caracterizadas por secuenciación Sanger. El gráfico muestra la media de tamaño promedio de las cadenas IGH, IGL y TRB caracterizadas por secuenciación tipo Sanger. IGH; 561pb, IGL; 424pb y TRB; 276pb.

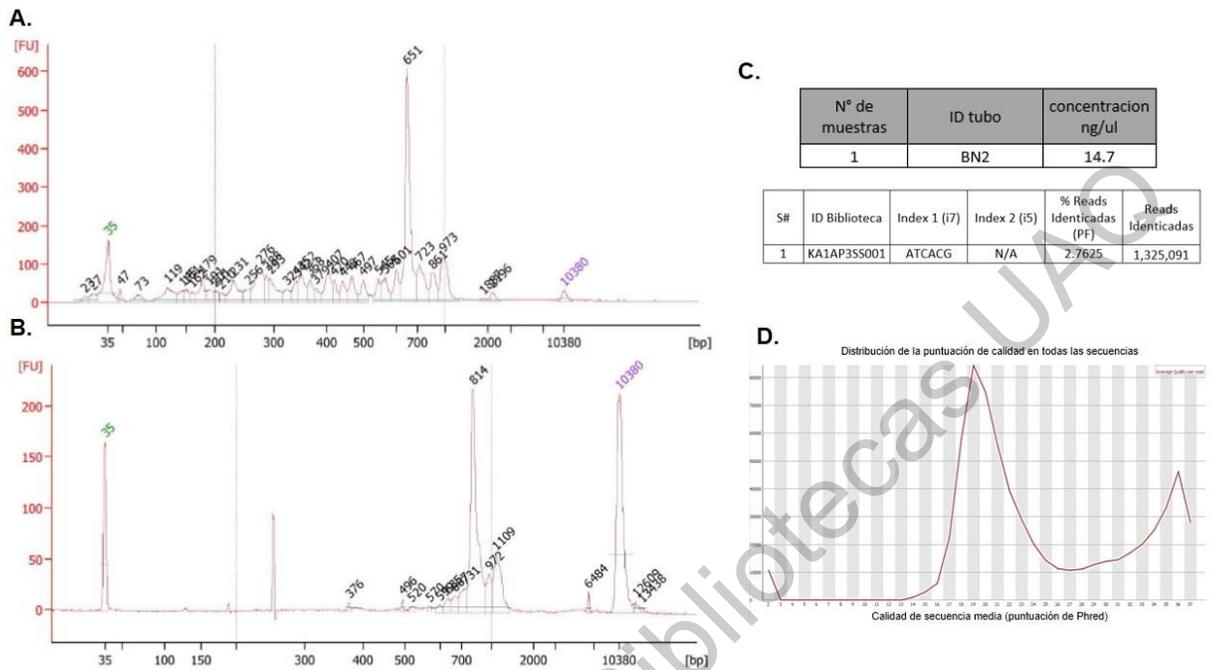
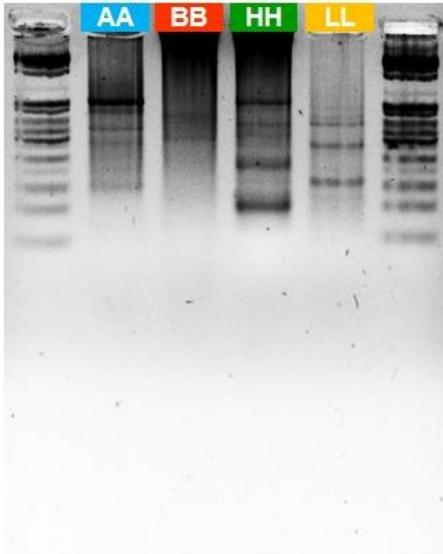


Figura A14| Electrograma de lecturas del Bioanalyzer y calidad de las lecturas por secuenciación de nueva generación de la muestra prueba. En A, el electrograma de los fragmentos identificados previo a la generación de bibliotecas. En B, el electrograma de los fragmentos identificados por el posterior a la generación de bibliotecas. En C, la concentración de la muestra y el número de lecturas identificadas con la secuenciación de nueva generación. En D, la distribución general de la puntuación de calidad de todas las secuencias.

A.



B.

Muestra original (cuantificación por fluorescencia)

N° de muestras	ID tubo	concentracion ng/ul	relacion 260/230	relacion 260/280
1	AA	350	1.3	1.59
2	BB	350	1.35	1.6
3	HH	350	1.33	1.17
4	LL	350	1.29	1.21

C.

CUANTIFICACION INTERNA POR FLUORESCENCIA Y CALCULOS PARA REACCION					
N° de muestras	ID tubo	concentracion ng/ul	ID BIBLIOTECA	INDEX HT	
1	AA	11.8	KA1AP3SS02	708	502
2	BB	45.2	KA1AP3SS03	708	503
3	HH	26.6	KA1AP3SS04	708	504
4	LL	22	KA1AP3SS05	708	505

D.

# Sample	ID biblioteca	Index 1 (I7)	Index 2 (I5)	% Reads identificadas	Reads Identificadas
1	KA1AP3SS05	TAATGCGC	AGGCGAAG	7.4396	1,612,936
2	KA1AP3SS02	TAATGCGC	ATAGAGGC	6.9203	1,500,349
3	KA1AP3SS03	TAATGCGC	CCTATCCT	4.8062	1,042,004
4	KA1AP3SS04	TAATGCGC	GGTCTGA	5.8253	1,262,949

Figura A15| Muestras analizadas por el protocolo de secuenciación de nueva generación. En A, las mandas del mix de muestras de TRA (AA), TRB (BB), IGH (HH) y IGL (LL). En B, las concentraciones y calidad por muestras. En C, las concentraciones usadas para la generación de las bibliotecas. En D, el identificador usado en las bibliotecas por muestra y las lecturas por muestra por secuenciación de nueva generación.

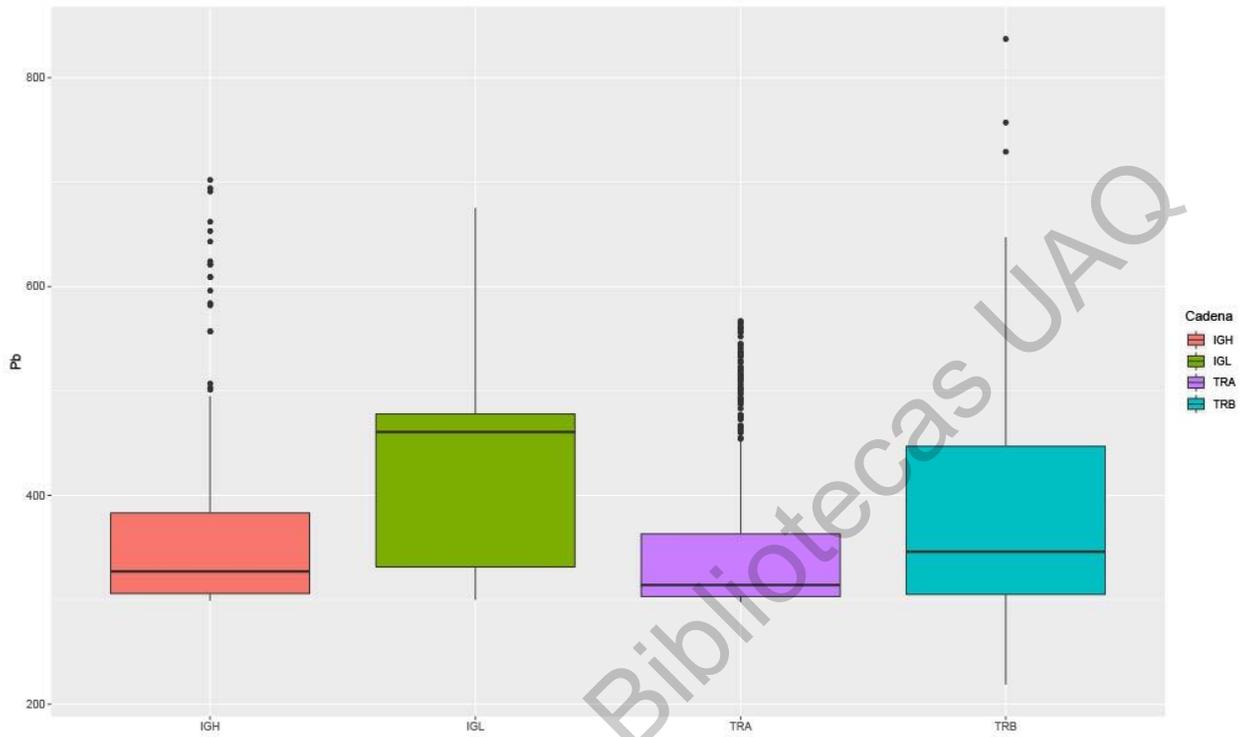
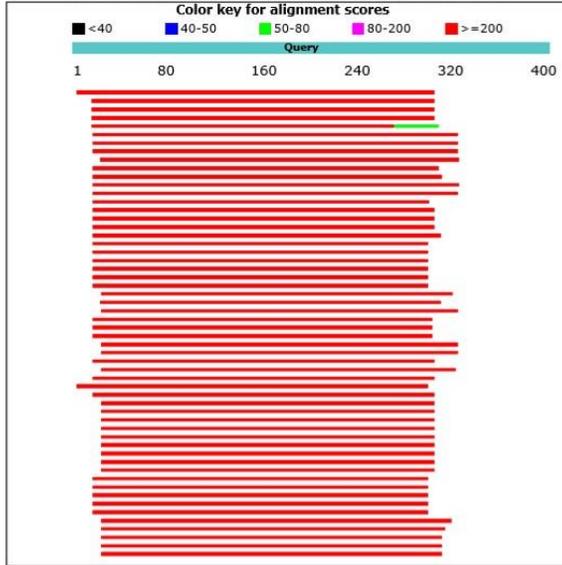


Figura A16| Largo medio de las secuencias caracterizadas por secuenciación de nueva generación. El gráfico muestra la media de tamaño promedio de las cadenas TRA; 314, TRB; 347, IGH; 327, IGL;460.

Job title: Nucleotide Sequence (400 letters)

RID 0NEGGSV9015 (Expires on 12-09 01:21 am)
 Query ID IcdQuery_214191 Database Name nr
 Description None Nucleotide collection (nt)
 Molecule type nucleic acid Program BLASTN 2.8.1+
 Query Length 400

Distribution of the top 103 Blast Hits on 100 subject sequences



Description	Max score	Total score	Query cover	E value	Ident	Accession
PREDICTED: Murine eotaxin-1a (LOC101668008), mRNA	433	433	76%	5e-117	92%	XM_012090448.1
Canis lupus familiaris mRNA for T cell receptor, partial cds	397	397	73%	4e-106	90%	D18411.1
Dog T cell receptor rearranged beta-chain mRNA V-D-J-C-region, partial cds	397	397	73%	4e-106	90%	M97519.1
Canis lupus familiaris mRNA for T cell receptor, partial cds	383	383	73%	2e-102	89%	D18408.1
Canis lupus familiaris genomic DNA region of TRB gene	374	366	73%	1e-99	92%	HE83129.1
Bus scrofa mRNA, clone PBIL1008E01, expressed in peripheral blood mononuclear cells	366	366	72%	6e-97	86%	AK549097.1
Bus scrofa mRNA, clone PBIL1004P06, expressed in peripheral blood mononuclear cells	366	366	72%	6e-97	86%	AK228822.1
Bus scrofa clone pTRB2 T cell receptor beta chain (TCR-β) mRNA, complete cds	359	359	73%	9e-95	86%	EJ84404.1
Feline leukemia virus (FeLV) transduced T cell receptor beta-chain (transduced gene = vtr)	355	355	76%	1e-93	86%	J02555.1
Bus scrofa clone pTRB4 T cell receptor beta chain (TCR-β) mRNA, complete cds	354	354	74%	1e-93	86%	EJ84406.1
Bus scrofa mRNA, clone THY1008B04, expressed in thymus	344	344	74%	2e-90	86%	AK511895.1
Bus scrofa mRNA, clone THY1008B11, expressed in thymus	344	344	73%	2e-90	84%	AK511453.1
Bus scrofa mRNA, clone THY10087C09, expressed in thymus	340	340	72%	2e-89	84%	AK511896.1
Bus scrofa T cell receptor beta-chain mRNA, partial sequence, clone PB	340	340	72%	2e-89	86%	AB079623.1
Bus scrofa clone pTRB2 T cell receptor beta chain (TCR-β) mRNA, complete cds	338	338	73%	8e-89	86%	EJ84404.1
Bus scrofa mRNA, clone THY10084J05, expressed in thymus	338	338	73%	8e-89	86%	AK511935.1
Bus scrofa mRNA, clone THY10114h11, expressed in thymus	338	338	73%	8e-89	86%	AK228694.1
Bus scrofa TCRβ mRNA for T cell receptor beta-chain, partial cds, clone T14	338	338	74%	8e-89	85%	AB079629.1
PREDICTED: Bus scrofa EPH receptor β6 (EPHβ6), transcript variantX28, mRNA	336	336	71%	3e-88	86%	XM_021272881.1

Figura A17-A| Resultados del Blast de NCBI de la muestra TRB12. Del lado izquierdo se muestra el alineamiento con secuencias del RefSeq. En el lado derecho se muestra la lista de los alineamientos con el Identificador de la lectura del RefSeq ordenados por orden descendente acorde al valor E.

Job title: Nucleotide Sequence (359 letters)

RID [ONEY77FS014](#) (Expires on 12-09 01:28 am)

Query ID [lcl|Query_34507](#)

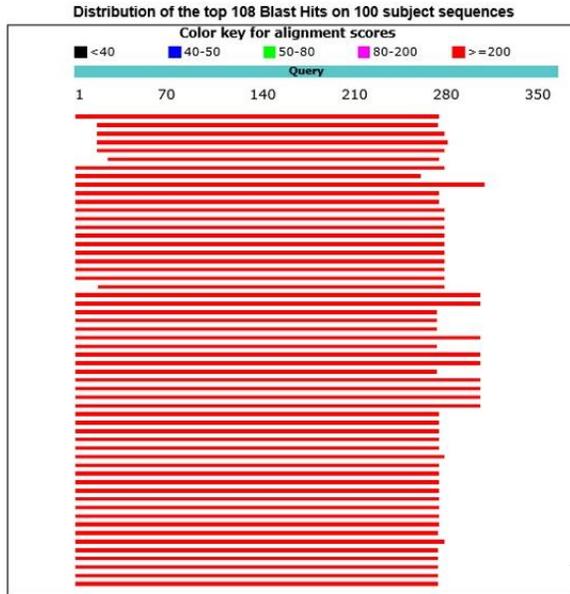
Database Name nr

Description Nucleotide collection (nt)

Molecule type nucleic acid

Program BLASTN 2.8.1+

Query Length 359



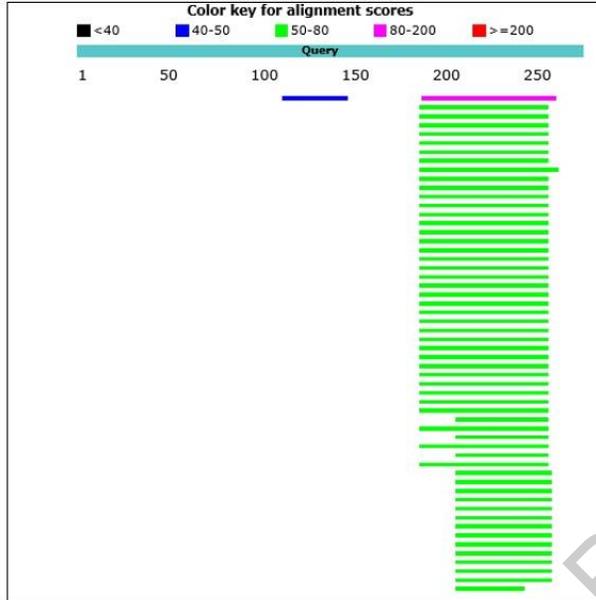
Description	Max score	Total score	Query cover	E value	Ident	Accession
PREDICTED: Mustela putorius furo uncharacterized LOC101836008 (LOC101836008), mRNA	404	404	76%	2e-108	93%	XM_013006468.1
Canis lupus familiaris genomic DNA region of TRAF gene	370	738	71%	4e-98	92%	HE663828.1
Dog T-cell receptor rearranged beta-chain mRNA V-D-J-C-region, partial cds	368	368	73%	8e-97	91%	M87819.1
Canis lupus familiaris mRNA for T-cell receptor, partial cds	360	360	73%	2e-98	90%	D18611.1
Canis lupus familiaris mRNA for T-cell receptor, partial cds	352	352	73%	1e-92	90%	D18408.1
Feline leukaemia virus (FeLV) transduced T-cell receptor beta-chain (transduced gene + vtr)	327	327	89%	1e-88	89%	J04156.1
PREDICTED: Eptesicus fuscus uncharacterized LOC102973008 (LOC102973008), mRNA	323	323	77%	8e-84	86%	XM_009163999.1
Camelus dromedarius partial T cell receptor beta locus	318	842	72%	7e-82	85%	L382787.1
Bufo scrofa mRNA, clone:THY010124A03, expressed in thymus	311	311	86%	1e-80	83%	AK239887.1
Bufo scrofa mRNA, clone:THY010090F11, expressed in thymus	309	309	76%	4e-80	85%	AK581816.1
Bufo scrofa mRNA, clone:THY010012B06, expressed in thymus	309	309	76%	4e-80	85%	AK581468.1
Bufo scrofa clone pTRB14 T cell receptor beta chain (TCR-β) mRNA, complete cds	307	307	77%	1e-79	85%	FJ846068.1
Bufo scrofa clone sTRB6 T cell receptor beta chain (TCR-β) mRNA, complete cds	307	307	77%	1e-79	85%	FJ844045.1
Bufo scrofa clone pTRB8 T cell receptor beta chain (TCR-β) mRNA, complete cds	307	307	77%	1e-79	85%	FJ846047.1
Bufo scrofa mRNA, clone:MLND10044002, expressed in mesenteric lymph nodes	307	307	77%	1e-79	85%	AK548728.1
Bufo scrofa mRNA, clone:THY010094Q08, expressed in thymus	307	307	77%	1e-79	85%	AK581826.1
Bufo scrofa mRNA, clone:THY010093Q06, expressed in thymus	307	307	77%	1e-79	85%	AK581808.1
Bufo scrofa mRNA, clone:PBLO0396D01, expressed in peripheral blood mononuclear cells	307	307	77%	1e-79	85%	AK5488007.1
Bufo scrofa mRNA, clone:PBLO10077D01, expressed in peripheral blood mononuclear cell	307	307	77%	1e-79	85%	AK537917.1

Figura A17-B| Resultados del Blast de NCBI de la muestra TRB3. Del lado izquierdo se muestra el alineamiento con secuencias del RefSeq. En el lado derecho se muestra la lista de los alineamientos con el Identificador de la lectura del RefSeq ordenados por orden descendente acorde al valor E.

Job title: Nucleotide Sequence (270 letters)

RID [ONEYUFDH015](#) (Expires on 12-09 01:45 am)
 Query ID [Id|Query_45445](#) Database Name nr
 Description None Description Nucleotide collection (nt)
 Molecule type nucleic acid Program BLASTN 2.8.1+
 Query Length 270

Distribution of the top 101 Blast Hits on 100 subject sequences



Description	Max score	Total score	Query cover	E value	Ident	Accession
PREDICTED: Mustela putorius furo unprocessed LOC101889008 (LOC101889008), mRNA	81.8	81.8	26%	2e-11	85%	XM_013060448.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (IG gene), clone SRC05	67.1	67.1	26%	9e-07	81%	HE862866.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (IG gene), clone SRC12	67.1	67.1	26%	9e-07	81%	HE862866.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (IG gene), clone SRC09	67.1	67.1	26%	9e-07	81%	HE862866.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (IG gene), clone SRC08	67.1	67.1	26%	9e-07	81%	HE862866.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (IG gene), clone SRC06	67.1	67.1	26%	9e-07	81%	HE862866.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (IG gene), clone SRC05	67.1	67.1	26%	9e-07	81%	HE862866.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (IG gene), clone SRC015	67.1	67.1	26%	9e-07	81%	HE862866.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (IG gene), clone SRC012	67.1	67.1	26%	9e-07	81%	HE862866.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (IG gene), clone SRC010	67.1	67.1	27%	9e-07	80%	HE862866.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (IG gene), clone SRC04	67.1	67.1	26%	9e-07	81%	HE862866.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (IG gene), clone SRC02	67.1	67.1	26%	9e-07	81%	HE862866.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (IG gene), clone SRm14	67.1	67.1	26%	9e-07	81%	HE862866.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (IG gene), clone SRm10	67.1	67.1	26%	9e-07	81%	HE862866.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (IG gene), clone SRm6	67.1	67.1	26%	9e-07	81%	HE862866.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (IG gene), clone SRm1	67.1	67.1	26%	9e-07	81%	HE862866.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (IG gene), clone SRm34	67.1	67.1	26%	9e-07	81%	HE862866.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (IG gene), clone SRm30	67.1	67.1	26%	9e-07	81%	HE862866.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (IG gene), clone SRm4	67.1	67.1	26%	9e-07	81%	HE862866.1

Figura A17-C| Resultados del Blast de NCBI de la muestra TRB8. Del lado Izquierdo se muestra el alineamiento con secuencias del RefSeq. En el lado derecho se muestra la lista de los alineamientos con el Identificador de la lectura del RefSeq ordenados por orden descendente acorde al valor E.

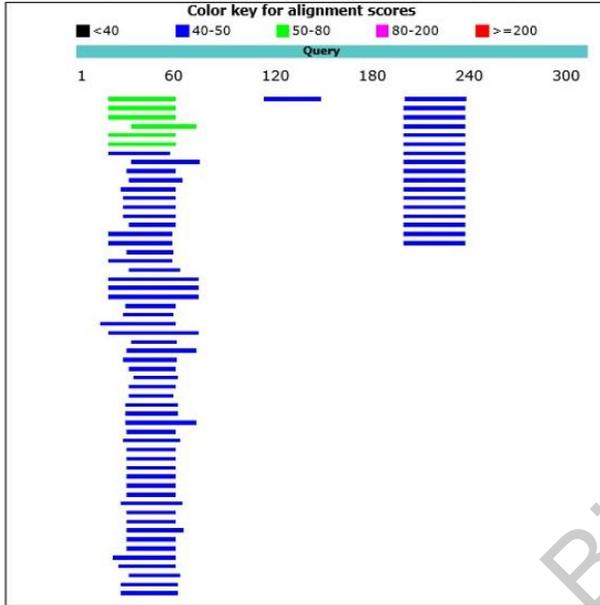
Job title: Nucleotide Sequence (309 letters)

RID [0NGP8DE01R](#) (Expires on 12-09 01:58 am)

Query ID |cl|Query_40097
 Description None
 Molecule type nucleic acid
 Query Length 309

Database Name nr
 Description Nucleotide collection (nt)
 Program BLASTN 2.8.1+

Distribution of the top 99 Blast Hits on 98 subject sequences



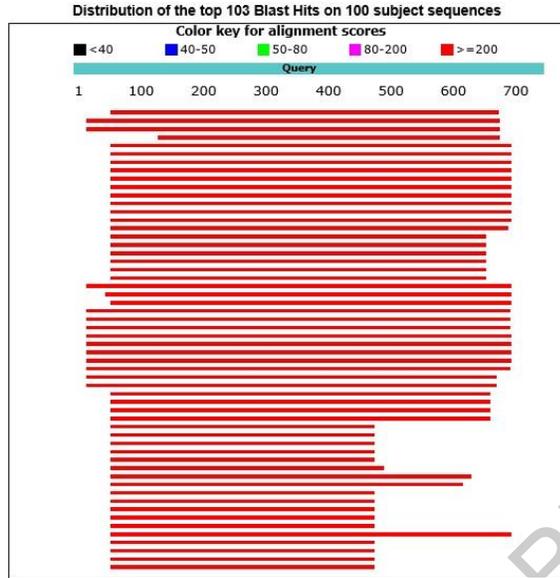
Description	Max score	Total score	Query cover	E value	Ident	Accession
Bos taurus hypothetical LOC609613, mRNA (cDNA clone MGC:159641 IMAGE:818796), complete cds	55.4	55.4	12%	0.001	90%	BC181268.1
Bos taurus T-cell receptor beta chain variable segment, mRNA (cDNA clone IMAGE:818240)	55.4	55.4	12%	0.001	90%	BC149662.1
Bos taurus uncharacterized LOC609613 (LOC609613), mRNA	55.4	55.4	12%	0.001	90%	NM_001008064.1
Plasmodium ovale genome assembly, chromosome: 13	51.8	51.8	12%	0.013	90%	LT684684.1
Bos taurus cDNA clone IMAGE:844763	50.9	50.9	12%	0.045	88%	BC148628.1
Bos taurus hypothetical LOC609613, mRNA (cDNA clone MGC:159636 IMAGE:844763), complete cds	50.9	50.9	12%	0.045	88%	BC142017.1
Bos taurus cDNA clone MGC:159664 IMAGE:843948, complete cds	50.0	50.0	11%	0.045	88%	BC142465.1
Homo sapiens 12a BAC RP11-482024 (Roswell Park Cancer Institute Human BAC Library) complete sequence	50.0	50.0	13%	0.045	84%	AC079355.11
Mus musculus acyl-CoA synthetase medium-chain family member 1, mRNA (cDNA clone MGC:23964 IMAGE:422896), complete cds	49.1	49.1	9%	0.16	97%	BC018416.1
Homo sapiens BAC clone RP11-1782 from 4, complete sequence	49.1	49.1	11%	0.16	91%	AC093726.2
PREDICTED: Cebus capuchus initiator heparin-associated protein 2-like (LOC108281578), mRNA	48.2	48.2	10%	0.16	94%	XM_017497204.1
Oryza sativa japonica Group cDNA, clone: J09028M23, full insert sequence	48.2	48.2	10%	0.16	94%	AK233382.1
Mus musculus transmembrane protein 119 (Tmem119), mRNA	48.2	48.2	10%	0.16	94%	NM_148182.2
Mus musculus transmembrane protein 119, mRNA (cDNA clone MGC:30086 IMAGE:526089), complete cds	48.2	48.2	10%	0.16	94%	BC026800.1
Meleca fascicularis testis cDNA clone: Q5A-12016, similar to human LOC132671 (LOC132671), mRNA, RefSeq: NM_145263.1	48.2	48.2	10%	0.16	94%	AB188420.1
Plasmodium malariae genome assembly, chromosome: 9	47.3	47.3	9%	0.55	95%	LT684620.1
PREDICTED: Panniplois hodgsonii Pflp3p3-related 16-like 1 (P. cerevisiae) (ATG16L1), transcript variant X2, mRNA	47.3	47.3	12%	0.55	87%	XM_006982252.1
PREDICTED: Panniplois hodgsonii Pflp3p3-related 16-like 1 (P. cerevisiae) (ATG16L1), transcript variant X1, mRNA	47.3	47.3	12%	0.55	87%	XM_006982251.1
Zea mays clone 1445266 mRNA sequence	47.3	47.3	9%	0.55	95%	EU845119.1

Figura A17-D| Resultados del Blast de NCBI de la muestra TRB17. Del lado izquierdo se muestra el alineamiento con secuencias del RefSeq. En el lado derecho se muestra la lista de los alineamientos con el Identificador de la lectura del RefSeq ordenados por orden descendente acorde al valor E.

Job title: Nucleotide Sequence (729 letters)

RID [ONGY1SVU014](#) (Expires on 12-09 02:02 am)
 Query ID [Ic|Query_56569](#)
 Description None
 Molecule type nucleic acid
 Query Length 729

Database Name nr
 Description Nucleotide collection (nt)
 Program BLASTN 2.8.1+



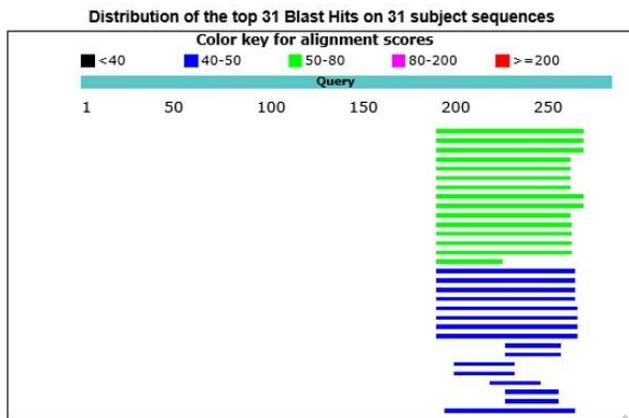
Description	Max score	Total score	Query cover	E value	Ident	Accession
PREDICTED: Museta putorius furc uncharacterized LOC101868008 (LOC101868008), mRNA	804	804	84%	0.0	89%	XM_013886448.1
Canis lupus familiaris mRNA for T-cell receptor, partial cds	724	724	89%	0.0	85%	D18411.1
Canis lupus familiaris mRNA for T-cell receptor, partial cds	724	724	89%	0.0	85%	D18409.1
Canis lupus familiaris mRNA for T-cell receptor, complete cds	626	626	74%	8e-175	96%	D18410.1
Bus scrofa mRNA, clone: THY010087C09, expressed in thymus	618	618	86%	1e-172	82%	AK331888.1
Bus scrofa mRNA, clone: PBLO1003804, expressed in peripheral blood mononuclear cells	618	618	86%	1e-172	82%	AK348124.1
Bus scrofa mRNA, clone: THY010038807, expressed in thymus	618	618	86%	1e-172	82%	AK328861.1
Bus scrofa mRNA, clone: THY010013A10, expressed in thymus	618	618	86%	1e-172	81%	AK351468.1
Bus scrofa mRNA, clone: THY01003A12, expressed in thymus	617	617	86%	4e-172	81%	AK388618.1
Bus scrofa mRNA, clone: THY010104F05, expressed in thymus	617	617	86%	4e-172	82%	AK328461.1
Bus scrofa mRNA, clone: THY010089004, expressed in thymus	614	614	86%	1e-171	82%	AK361203.1
Bus scrofa mRNA, clone: THY010087E11, expressed in thymus	612	612	86%	5e-171	81%	AK388468.1
Bus scrofa mRNA, clone: THY010084Q05, expressed in thymus	612	612	86%	5e-171	81%	AK351358.1
Bus scrofa mRNA, clone: THY010018C07, expressed in thymus	612	612	86%	5e-171	81%	AK328861.1
Bus scrofa T cell receptor beta-chain mRNA, partial sequence, clone #9	608	608	86%	2e-169	81%	AB078622.1
PREDICTED: Bus scrofa EPH receptor 86 (EPH86), transcript variant X32, misc_RNA	604	604	81%	2e-168	82%	XR_002340561.1
PREDICTED: Bus scrofa EPH receptor 86 (EPH86), transcript variant X31, misc_RNA	604	604	81%	2e-168	82%	XR_002340569.1
PREDICTED: Bus scrofa EPH receptor 86 (EPH86), transcript variant X31, misc_RNA	604	604	81%	2e-168	82%	XR_002340568.1
Bus scrofa T cell receptor beta-chain mRNA, partial sequence, clone.T11	604	604	81%	2e-168	82%	AB078622.1

Figura A17-E| Resultados del Blast de NCBI de la muestra TRB10. Del lado izquierdo se muestra el alineamiento con secuencias del RefSeq. En el lado derecho se muestra la lista de los alineamientos con el Identificador de la lectura del RefSeq ordenados por orden descendente acorde al valor E.

Dirección General de Biotecnología

Job title: Nucleotide Sequence (279 letters)

RID [QNHRCCHD015](#) (Expires on 12-09 02:15 am)
 Query ID [|c|Query_227799](#) Database Name nr
 Description None Description Nucleotide collection (nt)
 Molecule type nucleic acid Program BLASTN 2.8.1+
 Query Length 279

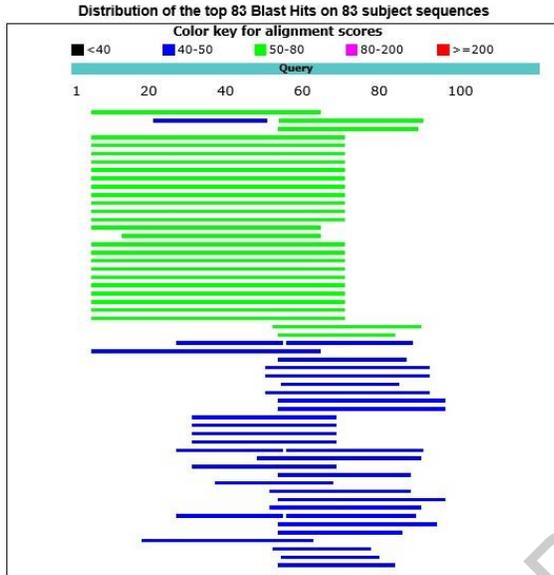


Description	Max score	Total score	Query cover	E value	Ident	Accession
Canis lupus familiaris partial mRNA for T cell receptor beta chain (no gene), clone SR01	66.2	66.2	27%	5e-07	80%	HE862862.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (no gene), clone SR002	66.2	66.2	27%	5e-07	80%	HE862864.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (no gene), clone SR01	66.2	66.2	27%	5e-07	80%	HE862864.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (no gene), clone SR000	62.6	62.6	25%	6e-06	81%	HE862860.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (no gene), clone SR01	62.6	62.6	25%	6e-06	81%	HE862860.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (no gene), clone SR01	62.6	62.6	25%	6e-06	81%	HE862864.1
Canis lupus familiaris mRNA for T cell receptor, partial cds	62.6	62.6	25%	6e-06	81%	D18411.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (no gene), clone SR0010	61.7	61.7	27%	2e-05	79%	HE862868.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (no gene), clone SR004	61.7	61.7	27%	2e-05	79%	HE862866.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (no gene), clone SR02	58.1	58.1	25%	3e-04	79%	HE862865.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (no gene), clone SR014	58.1	58.1	25%	3e-04	80%	HE862865.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (no gene), clone SR130	58.1	58.1	25%	3e-04	80%	HE862865.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (no gene), clone SR02	58.1	58.1	25%	3e-04	80%	HE862865.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (no gene), clone SR3	58.1	58.1	25%	3e-04	80%	HE862865.1
PREDICTED: Mustela putorius furo ultraconserved LOC101880008 (LOC101880008), mRNA	53.6	53.6	12%	0.003	94%	XM_011050448.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (no gene), clone SR010	50.0	50.0	26%	0.040	76%	HE862862.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (no gene), clone SR134	50.0	50.0	26%	0.040	76%	HE862868.1
Canis lupus familiaris partial mRNA for T cell receptor beta chain (no gene), clone SR02	50.0	50.0	26%	0.040	76%	HE862868.1

Figura A17-F| Resultados del Blast de NCBI de la muestra TRB7. Del lado izquierdo se muestra el alineamiento con secuencias del RefSeq. En el lado derecho se muestra la lista de los alineamientos con el Identificador de la lectura del RefSeq ordenados por orden descendente acorde al valor E.

Job title: Nucleotide Sequence (118 letters)

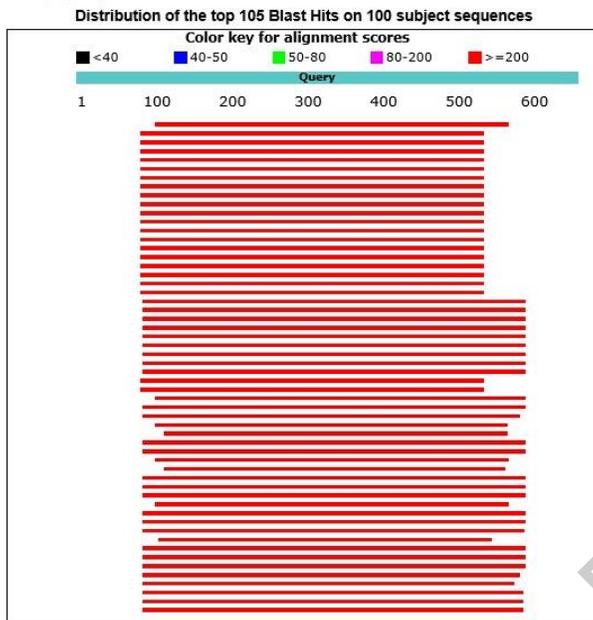
RID [QNK09MJS015](#) (Expires on 12-09 02:37 am)
 Query ID lc|Query_24223 Database Name nr
 Description None Description Nucleotide collection (nt)
 Molecule type nucleic acid Program BLASTN 2.8.1+
 Query Length 118



Description	Max score	Total score	Query cover	E value	Ident	Accession
PREDICTED: <i>Ursus arctos horribilis</i> uncharacterized LOC11287145 (LOC11287145), mRNA	68.0	68.0	90%	5e-06	99%	XM_028616113.1
<i>Asterix australis</i> mantell genome assembly Aglianti0, scaffold scaff007	57.2	57.2	31%	5e-05	95%	LJ084822.1
<i>Coccidioides immitis</i> RB HADH- <i>ubiquitin</i> cadorprotease 3.6 kDa subunit (DIA3_06717), mRNA	57.2	57.2	30%	5e-05	94%	XM_001242220.2
<i>Felis catus</i> clone cat 2-IgHA-11 IgA constant region mRNA, partial cds	56.3	56.3	95%	3e-04	80%	KY736046.1
<i>Felis catus</i> clone cat 2-IgHA-9 IgA constant region mRNA, partial cds	56.3	56.3	95%	3e-04	80%	KY736047.1
<i>Felis catus</i> clone cat 2-IgHA-4 IgA constant region mRNA, partial cds	56.3	56.3	95%	3e-04	80%	KY736042.1
<i>Felis catus</i> clone cat 2-IgHA-3 IgA constant region mRNA, partial cds	56.3	56.3	95%	3e-04	80%	KY736041.1
<i>Felis catus</i> clone cat 2-IgHA-7 IgA constant region mRNA, partial cds	56.3	56.3	95%	3e-04	80%	KY736038.1
<i>Felis catus</i> clone cat 1-IgHA-9 IgA constant region mRNA, partial cds	56.3	56.3	95%	3e-04	80%	KY736035.1
<i>Felis catus</i> clone cat 1-IgHA-7 IgA constant region mRNA, partial cds	56.3	56.3	95%	3e-04	80%	KY736036.1
<i>Felis catus</i> clone cat 1-IgHA-8 IgA constant region mRNA, partial cds	56.3	56.3	95%	3e-04	80%	KY736035.1
<i>Felis catus</i> clone cat 1-IgHA-3 IgA constant region mRNA, partial cds	56.3	56.3	95%	3e-04	80%	KY736032.1
<i>Felis catus</i> clone cat 1-IgHA-2 IgA constant region mRNA, partial cds	56.3	56.3	95%	3e-04	80%	KY736031.1
<i>Felis catus</i> clone cat 1-IgHA-1 IgA constant region mRNA, partial cds	56.3	56.3	95%	3e-04	80%	KY736030.1
<i>Aluropoda melanoleuca</i> immunoglobulin alpha heavy chain gene, partial cds	54.5	54.5	50%	0.001	81%	AF155327.1
<i>Canis familiaris</i> IgA heavy chain constant region gene, partial cds	52.7	52.7	43%	0.004	82%	U88271.1
<i>Felis catus</i> clone cat 2-IgHA-12 IgA constant region mRNA, partial cds	51.8	51.8	95%	0.004	79%	KY736050.1
<i>Felis catus</i> clone cat 2-IgHA-10 IgA constant region mRNA, partial cds	51.8	51.8	95%	0.004	79%	KY736048.1
<i>Felis catus</i> clone cat 2-IgHA-8 IgA constant region mRNA, partial cds	51.8	51.8	95%	0.004	79%	KY736046.1
<i>Felis catus</i> clone cat 2-IgHA-7 IgA constant region mRNA, partial cds	51.8	51.8	95%	0.004	79%	KY736045.1
<i>Felis catus</i> clone cat 2-IgHA-6 IgA constant region mRNA, partial cds	51.8	51.8	95%	0.004	79%	KY736044.1
<i>Felis catus</i> clone cat 2-IgHA-5 IgA constant region mRNA, partial cds	51.8	51.8	95%	0.004	79%	KY736043.1
<i>Felis catus</i> clone cat 2-IgHA-2 IgA constant region mRNA, partial cds	51.8	51.8	95%	0.004	79%	KY736040.1

Figura A17-I| Resultados del Blast de NCBI de la muestra IGH23. Del lado izquierdo se muestra el alineamiento con secuencias del RefSeq. En el lado derecho se muestra la lista de los alineamientos con el Identificador de la lectura del RefSeq ordenados por orden descendente acorde al valor E.

Job title: Nucleotide Sequence (643 letters)
 RID ON1RWUPR014 (Expires on 12-09 02:33 am)
 Query ID |cd|Query_76257 Database Name nr
 Description None Description Nucleotide collection (nt)
 Molecule type nucleic acid Program BLASTN 2.8.1+
 Query Length 643

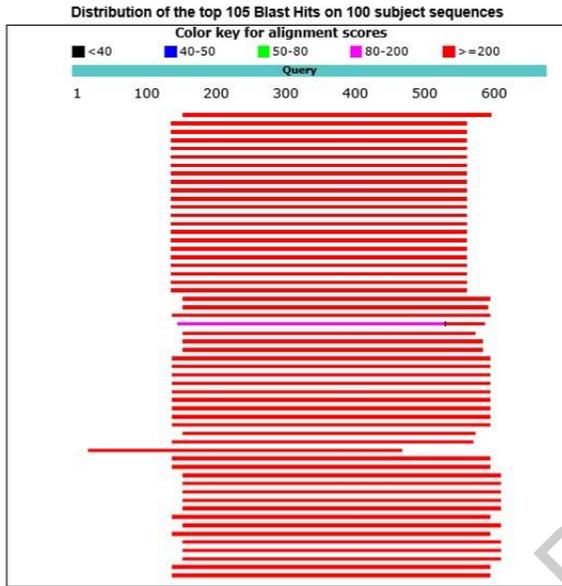


Description	Max score	Total score	Query cover	E value	Ident	Accession
AluYp808e melanocyte immunoglobulin alpha heavy chain part, partial cds	820	820	71%	0.0	82%	AY13287.1
Felis catus clone cat 2-IGHA-11 Iga constant region mRNA, partial cds	524	524	89%	2e-147	88%	KT22848.1
Felis catus clone cat 2-IGHA-9 Iga constant region mRNA, partial cds	524	524	89%	2e-147	88%	KT22847.1
Felis catus clone cat 2-IGHA-4 Iga constant region mRNA, partial cds	524	524	89%	2e-147	88%	KT22846.1
Felis catus clone cat 2-IGHA-2 Iga constant region mRNA, partial cds	524	524	89%	2e-147	88%	KT22845.1
Felis catus clone cat 1-IGHA-9 Iga constant region mRNA, partial cds	524	524	89%	2e-147	88%	KT22839.1
Felis catus clone cat 1-IGHA-7 Iga constant region mRNA, partial cds	524	524	89%	2e-147	88%	KT22838.1
Felis catus clone cat 1-IGHA-6 Iga constant region mRNA, partial cds	524	524	89%	2e-147	88%	KT22837.1
Felis catus clone cat 1-IGHA-1 Iga constant region mRNA, partial cds	524	524	89%	2e-147	88%	KT22836.1
Felis catus clone cat 2-IGHA-10 Iga constant region mRNA, partial cds	529	529	89%	4e-148	88%	KT22845.1
Felis catus clone cat 2-IGHA-8 Iga constant region mRNA, partial cds	529	529	89%	4e-148	88%	KT22844.1
Felis catus clone cat 2-IGHA-7 Iga constant region mRNA, partial cds	529	529	89%	4e-148	88%	KT22843.1
Felis catus clone cat 2-IGHA-6 Iga constant region mRNA, partial cds	529	529	89%	4e-148	88%	KT22842.1
Felis catus clone cat 2-IGHA-2 Iga constant region mRNA, partial cds	529	529	89%	4e-148	88%	KT22841.1
Felis catus clone cat 1-IGHA-8 Iga constant region mRNA, partial cds	529	529	89%	4e-148	88%	KT22837.1
Felis catus clone cat 1-IGHA-6 Iga constant region mRNA, partial cds	529	529	89%	4e-148	88%	KT22836.1
Felis catus clone cat 1-IGHA-4 Iga constant region mRNA, partial cds	529	529	89%	4e-148	88%	KT22835.1
Felis catus clone cat 1-IGHA-3 Iga constant region mRNA, partial cds	529	529	89%	4e-148	88%	KT22834.1
Sus scrofa mRNA, clone CU171028F05, expressed in colon	529	529	77%	4e-148	88%	AK321384.1
Sus scrofa mRNA, clone CU181028F05, expressed in colon	529	529	77%	4e-148	88%	AK32137.1
Sus scrofa mRNA, clone T1001028F05, expressed in trachea	529	529	77%	4e-148	88%	AK32131.1
Sus scrofa mRNA, clone T1001028A11, expressed in kidney	529	529	77%	4e-148	88%	AK32147.1

Figura A17-J| Resultados del Blast de NCBI de la muestra IGH17. Del lado izquierdo se muestra el alineamiento con secuencias del RefSeq. En el lado derecho se muestra la lista de los alineamientos con el Identificador de la lectura del RefSeq ordenados por orden descendente acorde al valor E.

Job title: Nucleotide Sequence (662 letters)

RID [0NJTMVBG014](#) (Expires on 12-09 02:34 am)
 Query ID Icl|Query_6871 Database Name nr
 Description None Nucleotide collection (nt)
 Molecule type nucleic acid Program BLASTN 2.8.1+
 Query Length 662



Description	Max score	Total score	Query cover	E value	Ident	Accession
Aluropoda melanotus Immunoglobulin alpha heavy chain gene, partial cds	670	670	88%	0.0	84%	AY818887.1
Felis catus clone cat 2-IGHA-11 IGA constant region mRNA, partial cds	530	530	63%	4e-146	88%	KY786048.1
Felis catus clone cat 2-IGHA-9 IGA constant region mRNA, partial cds	530	530	63%	4e-146	88%	KY786047.1
Felis catus clone cat 2-IGHA-4 IGA constant region mRNA, partial cds	530	530	63%	4e-146	88%	KY786042.1
Felis catus clone cat 2-IGHA-2 IGA constant region mRNA, partial cds	530	530	63%	4e-146	88%	KY786040.1
Felis catus clone cat 1-IGHA-9 IGA constant region mRNA, partial cds	530	530	63%	4e-146	88%	KY786058.1
Felis catus clone cat 1-IGHA-7 IGA constant region mRNA, partial cds	530	530	63%	4e-146	88%	KY786058.1
Felis catus clone cat 1-IGHA-6 IGA constant region mRNA, partial cds	530	530	63%	4e-146	88%	KY786056.1
Felis catus clone cat 1-IGHA-1 IGA constant region mRNA, partial cds	530	530	63%	4e-146	88%	KY786058.1
Felis catus clone cat 2-IGHA-10 IGA constant region mRNA, partial cds	526	526	63%	5e-145	88%	KY786045.1
Felis catus clone cat 2-IGHA-8 IGA constant region mRNA, partial cds	526	526	63%	5e-145	88%	KY786048.1
Felis catus clone cat 2-IGHA-7 IGA constant region mRNA, partial cds	526	526	63%	5e-145	88%	KY786046.1
Felis catus clone cat 2-IGHA-6 IGA constant region mRNA, partial cds	526	526	63%	5e-145	88%	KY786044.1
Felis catus clone cat 2-IGHA-5 IGA constant region mRNA, partial cds	526	526	63%	5e-145	88%	KY786041.1
Felis catus clone cat 2-IGHA-3 IGA constant region mRNA, partial cds	526	526	63%	5e-145	88%	KY786041.1
Felis catus clone cat 1-IGHA-6 IGA constant region mRNA, partial cds	526	526	63%	5e-145	88%	KY786047.1
Felis catus clone cat 1-IGHA-5 IGA constant region mRNA, partial cds	526	526	63%	5e-145	88%	KY786054.1
Felis catus clone cat 1-IGHA-4 IGA constant region mRNA, partial cds	526	526	63%	5e-145	88%	KY786051.1
Felis catus clone cat 1-IGHA-3 IGA constant region mRNA, partial cds	526	526	63%	5e-145	88%	KY786052.1
Felis catus clone cat 1-IGHA-2 IGA constant region mRNA, partial cds	526	526	63%	5e-145	88%	KY786051.1
Felis catus clone cat 2-IGHA-12 IGA constant region mRNA, partial cds	521	521	63%	2e-143	87%	KY786060.1
Felis catus clone cat 2-IGHA-1 IGA constant region mRNA, partial cds	521	521	63%	2e-143	87%	KY786058.1
Canis familiaris IGA heavy chain constant region gene, partial cds	510	510	68%	4e-140	86%	L88871.1

Figura A17-K| Resultados del Blast de NCBI de la muestra IGH8. Del lado izquierdo se muestra el alineamiento con secuencias del RefSeq. En el lado derecho se muestra la lista de los alineamientos con el Identificador de la lectura del RefSeq ordenados por orden descendente acorde al valor E.

Dirección General de Biotecnología y Recursos Genéticos

Job title: Nucleotide Sequence (210 letters)

RID [0NKSDM4D014](#) (Expires on 12-09 02:50 am)

Query ID Icl|Query_104467

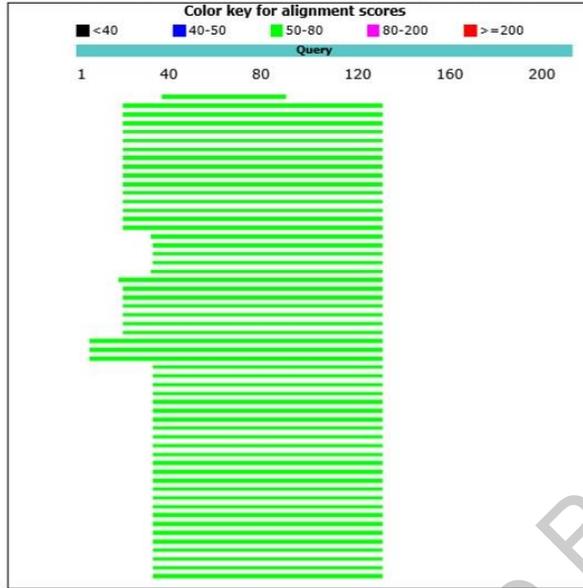
Description None

Molecule type nucleic acid

Query Length 210

Database Name nr
Description Nucleotide collection (nt)
Program BLASTN 2.8.1+

Distribution of the top 100 Blast Hits on 100 subject sequences

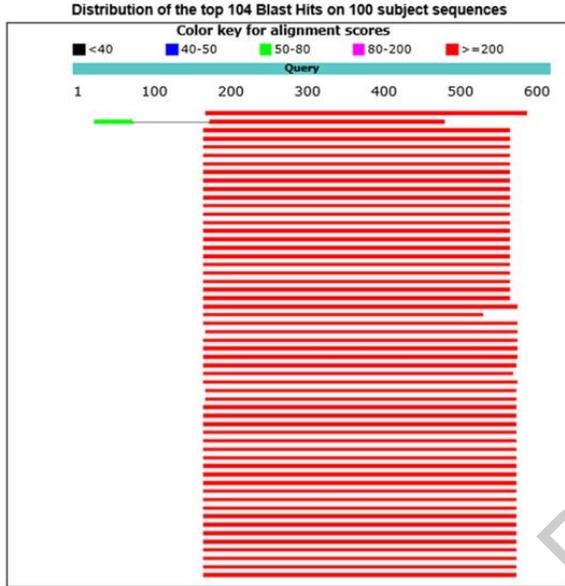


Description	Max score	Total score	Query cover	E value	Ident	Accession
Canis familiaris IgA heavy chain constant region gene, partial cds	65.3	65.3	25%	1e-06	87%	L8827.1
Felis catus clone cat 2-IgHA-12 IgA constant region mRNA, partial cds	61.7	61.7	53%	2e-05	74%	KY786055.1
Felis catus clone cat 2-IgHA-10 IgA constant region mRNA, partial cds	61.7	61.7	53%	2e-05	74%	KY786048.1
Felis catus clone cat 2-IgHA-8 IgA constant region mRNA, partial cds	61.7	61.7	53%	2e-05	74%	KY786046.1
Felis catus clone cat 2-IgHA-7 IgA constant region mRNA, partial cds	61.7	61.7	53%	2e-05	74%	KY786045.1
Felis catus clone cat 2-IgHA-6 IgA constant region mRNA, partial cds	61.7	61.7	53%	2e-05	74%	KY786044.1
Felis catus clone cat 2-IgHA-5 IgA constant region mRNA, partial cds	61.7	61.7	53%	2e-05	74%	KY786043.1
Felis catus clone cat 2-IgHA-4 IgA constant region mRNA, partial cds	61.7	61.7	53%	2e-05	74%	KY786042.1
Felis catus clone cat 1-IgHA-9 IgA constant region mRNA, partial cds	61.7	61.7	53%	2e-05	74%	KY786033.1
Felis catus clone cat 1-IgHA-8 IgA constant region mRNA, partial cds	61.7	61.7	53%	2e-05	74%	KY786032.1
Felis catus clone cat 1-IgHA-7 IgA constant region mRNA, partial cds	61.7	61.7	53%	2e-05	74%	KY786031.1
Felis catus clone cat 1-IgHA-6 IgA constant region mRNA, partial cds	61.7	61.7	53%	2e-05	74%	KY786030.1
Felis catus clone cat 1-IgHA-5 IgA constant region mRNA, partial cds	61.7	61.7	53%	2e-05	74%	KY786029.1
Felis catus clone cat 1-IgHA-4 IgA constant region mRNA, partial cds	61.7	61.7	53%	2e-05	74%	KY786028.1
Felis catus clone cat 1-IgHA-3 IgA constant region mRNA, partial cds	61.7	61.7	53%	2e-05	74%	KY786027.1
Felis catus clone cat 1-IgHA-2 IgA constant region mRNA, partial cds	61.7	61.7	53%	2e-05	74%	KY786026.1
Felis catus clone cat 1-IgHA-1 IgA constant region mRNA, partial cds	61.7	61.7	53%	2e-05	74%	KY786025.1
PREDICTED: Manis javanica uncharacterized LOC102390611 (LOC102390611), mRNA	61.7	61.7	47%	2e-05	74%	XM_017348728.1
Bus scrofa mRNA, clone: CLNT10024H05, expressed in colon	59.0	59.0	47%	6e-05	74%	AK380867.1
Bus scrofa mRNA, clone: CLNT10043D01, expressed in colon	59.0	59.0	47%	6e-05	74%	AK380868.1
Bus scrofa mRNA, clone: CLNT10021B02, expressed in colon	59.0	59.0	47%	6e-05	74%	AK380869.1
Bus scrofa mRNA, clone: CLNT10002F05, expressed in colon	59.0	59.0	47%	6e-05	74%	AK380870.1
Eptesicus fuscus clone EA1 immunoglobulin heavy chain mRNA, partial cds	57.2	57.2	53%	2e-04	73%	KY786049.1
Felis catus clone cat 2-IgHA-11 IgA constant region mRNA, partial cds	57.2	57.2	53%	2e-04	73%	KY786047.1

Figura A17-L| Resultados del Blast de NCBI de la muestra IGH24. Del lado izquierdo se muestra el alineamiento con secuencias del RefSeq. En el lado derecho se muestra la lista de los alineamientos con el Identificador de la lectura del RefSeq ordenados por orden descendente acorde al valor E.

Job title: Nucleotide Sequence (609 letters)

RID @NKSUVSV014 (Expires on 12-09 02:50 am)
 Query ID lcl|Query_226479 Database Name nr
 Description None Description Nucleotide collection (nt)
 Molecule type nucleic acid Program BLASTN 2.8.1+
 Query Length 609

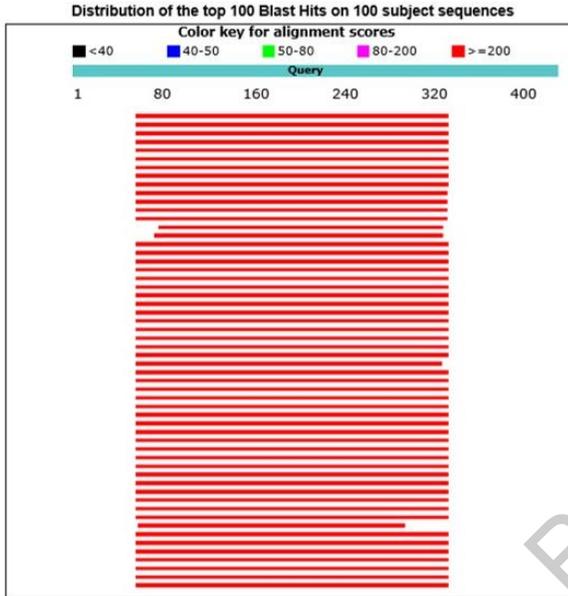


Description	Max score	Total score	Query cover	E value	Ident	Accession
<i>Aluroside melanosus</i> immunoglobulin alpha heavy chain gene, partial cds	563	563	66%	2e-166	90%	AY18327.1
PREDICTED: <i>Ursus arctos horribilis</i> uncharacterized LOC113267148 (LOC113267148), mRNA	486	509	57%	7e-124	93%	XM_028616115.1
<i>Felis catus</i> clone cat 2-IgHA-11 IgA constant region mRNA, partial cds	447	447	66%	3e-121	88%	KY784048.1
<i>Felis catus</i> clone cat 2-IgHA-9 IgA constant region mRNA, partial cds	447	447	66%	3e-121	88%	KY784047.1
<i>Felis catus</i> clone cat 2-IgHA-4 IgA constant region mRNA, partial cds	447	447	66%	3e-121	88%	KY784043.1
<i>Felis catus</i> clone cat 2-IgHA-2 IgA constant region mRNA, partial cds	447	447	66%	3e-121	88%	KY784040.1
<i>Felis catus</i> clone cat 1-IgHA-9 IgA constant region mRNA, partial cds	447	447	66%	3e-121	88%	KY784038.1
<i>Felis catus</i> clone cat 1-IgHA-7 IgA constant region mRNA, partial cds	447	447	66%	3e-121	88%	KY784036.1
<i>Felis catus</i> clone cat 1-IgHA-6 IgA constant region mRNA, partial cds	447	447	66%	3e-121	88%	KY784035.1
<i>Felis catus</i> clone cat 1-IgHA-1 IgA constant region mRNA, partial cds	447	447	66%	3e-121	88%	KY784030.1
<i>Felis catus</i> clone cat 2-IgHA-10 IgA constant region mRNA, partial cds	443	443	66%	4e-120	88%	KY784046.1
<i>Felis catus</i> clone cat 2-IgHA-8 IgA constant region mRNA, partial cds	443	443	66%	4e-120	88%	KY784045.1
<i>Felis catus</i> clone cat 2-IgHA-7 IgA constant region mRNA, partial cds	443	443	66%	4e-120	88%	KY784044.1
<i>Felis catus</i> clone cat 2-IgHA-6 IgA constant region mRNA, partial cds	443	443	66%	4e-120	88%	KY784043.1
<i>Felis catus</i> clone cat 2-IgHA-5 IgA constant region mRNA, partial cds	443	443	66%	4e-120	88%	KY784041.1
<i>Felis catus</i> clone cat 2-IgHA-3 IgA constant region mRNA, partial cds	443	443	66%	4e-120	88%	KY784037.1
<i>Felis catus</i> clone cat 1-IgHA-8 IgA constant region mRNA, partial cds	443	443	66%	4e-120	88%	KY784034.1
<i>Felis catus</i> clone cat 1-IgHA-5 IgA constant region mRNA, partial cds	443	443	66%	4e-120	88%	KY784033.1
<i>Felis catus</i> clone cat 1-IgHA-3 IgA constant region mRNA, partial cds	443	443	66%	4e-120	88%	KY784031.1
<i>Felis catus</i> clone cat 1-IgHA-2 IgA constant region mRNA, partial cds	443	443	66%	4e-120	88%	KY784031.1
<i>Felis catus</i> clone cat 2-IgHA-12 IgA constant region mRNA, partial cds	438	438	66%	2e-118	88%	KY784056.1
<i>Felis catus</i> clone cat 2-IgHA-1 IgA constant region mRNA, partial cds	438	438	66%	2e-118	88%	KY784028.1
<i>Ovis aries</i> immunoglobulin alpha heavy chain (IgA) mRNA, partial cds	434	434	66%	2e-117	84%	AF024846.1

Figura A17-M| Resultados del Blast de NCBI de la muestra IGH9. Del lado izquierdo se muestra el alineamiento con secuencias del RefSeq. En el lado derecho se muestra la lista de los alineamientos con el Identificador de la lectura del RefSeq ordenados por orden descendente acorde al valor E.

Job title: Nucleotide Sequence (424 letters)

RID [ONMF83BH014](#) (Expires on 12-09 03:02 am)
 Query ID [Icl|Query_30143](#) Database Name nr
 Description None Description Nucleotide collection (nr)
 Molecule type nucleic acid Program BLASTN 2.8.1+
 Query Length 424



Description	Max score	Total score	Query cover	E value	Ident	Accession
PREDICTED: <i>Calominus usinus</i> immunoglobulin lambda-1 light chain-like (LOC112512464), transcript variant X3, mRNA	444	444	65%	3e-120	95%	XM_026867225.1
PREDICTED: <i>Calominus usinus</i> immunoglobulin lambda-1 light chain-like (LOC112512464), transcript variant X3, mRNA	444	444	65%	3e-120	95%	XM_026867225.1
PREDICTED: <i>Calominus usinus</i> immunoglobulin lambda-1 light chain-like (LOC112512464), transcript variant X7, mRNA	444	444	65%	3e-120	95%	XM_026867225.1
PREDICTED: <i>Calominus usinus</i> immunoglobulin lambda-1 light chain-like (LOC112512464), transcript variant X7, mRNA	444	444	65%	3e-120	95%	XM_026867225.1
PREDICTED: <i>Calominus usinus</i> immunoglobulin lambda-1 light chain-like (LOC112512464), transcript variant X4, mRNA	444	444	65%	3e-120	95%	XM_026867225.1
PREDICTED: <i>Calominus usinus</i> immunoglobulin lambda-1 light chain-like (LOC112512464), transcript variant X4, mRNA	444	444	65%	3e-120	95%	XM_026867225.1
PREDICTED: <i>Calominus usinus</i> immunoglobulin lambda-1 light chain-like (LOC112512464), transcript variant X2, mRNA	444	444	65%	3e-120	95%	XM_026867225.1
PREDICTED: <i>Calominus usinus</i> immunoglobulin lambda-1 light chain-like (LOC112512464), transcript variant X1, mRNA	444	444	65%	3e-120	95%	XM_026867225.1
PREDICTED: <i>Odobenus rosmarus divergens</i> ig lambda chain V _H region BL2 (LOC101332618), mRNA	442	442	65%	1e-119	95%	XM_012682097.1
PREDICTED: <i>Leontideus weddellii</i> immunoglobulin lambda-like polypeptide S-like (LOC102738234), mRNA	419	419	65%	3e-113	94%	XM_008746575.1
PREDICTED: <i>Leontideus weddellii</i> immunoglobulin lambda-like polypeptide S (LOC101615251), mRNA	406	406	65%	7e-109	92%	XM_008741625.1
PREDICTED: <i>Neomacrus scabripinnis</i> immunoglobulin lambda-like polypeptide S (LOC110573369), mRNA	401	401	65%	5e-102	92%	XM_021827285.1
PREDICTED: <i>Ursus maritimus</i> immunoglobulin lambda-like polypeptide S (LOC103669269), mRNA	382	382	65%	8e-102	92%	XM_008697695.1

Figura A18| Esquematización del proceso de clonación utilizado durante el estudio. En A, el mapa del vector pCR4-TOPO utilizado en todas las clonaciones del estudio. En B, el proceso de cribado por restricción con EcoRI, en azul en C y por PCR de colonia específica en amarillo en C.

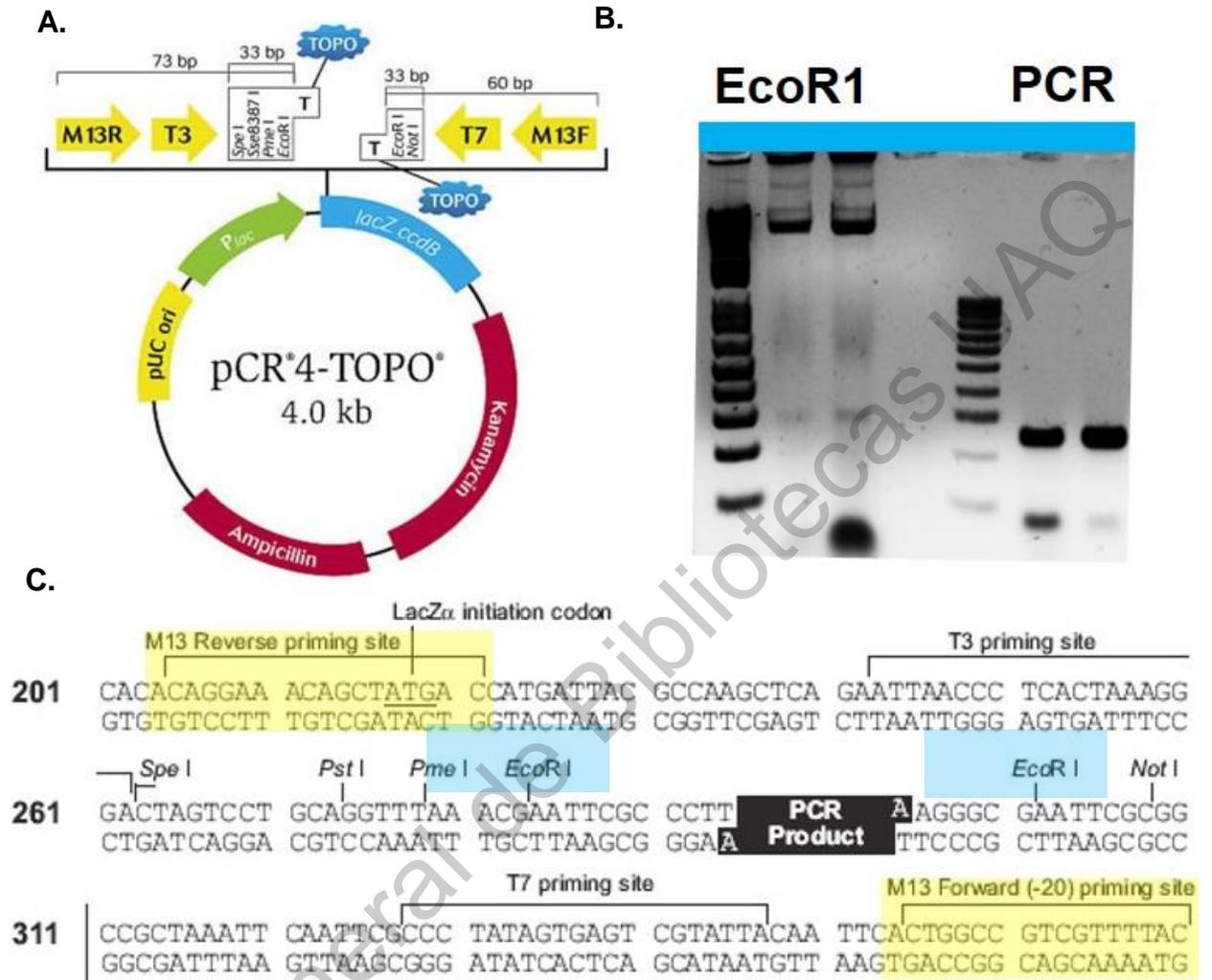


Figura A18| Esquematización del proceso de clonación utilizado durante el estudio. En A, el mapa del vector pCR4-TOPO utilizado en todas las clonaciones del estudio. En B, el proceso de cribado por restricción con EcoRI, en azul en C y por PCR de colonia específica en amarillo en C.

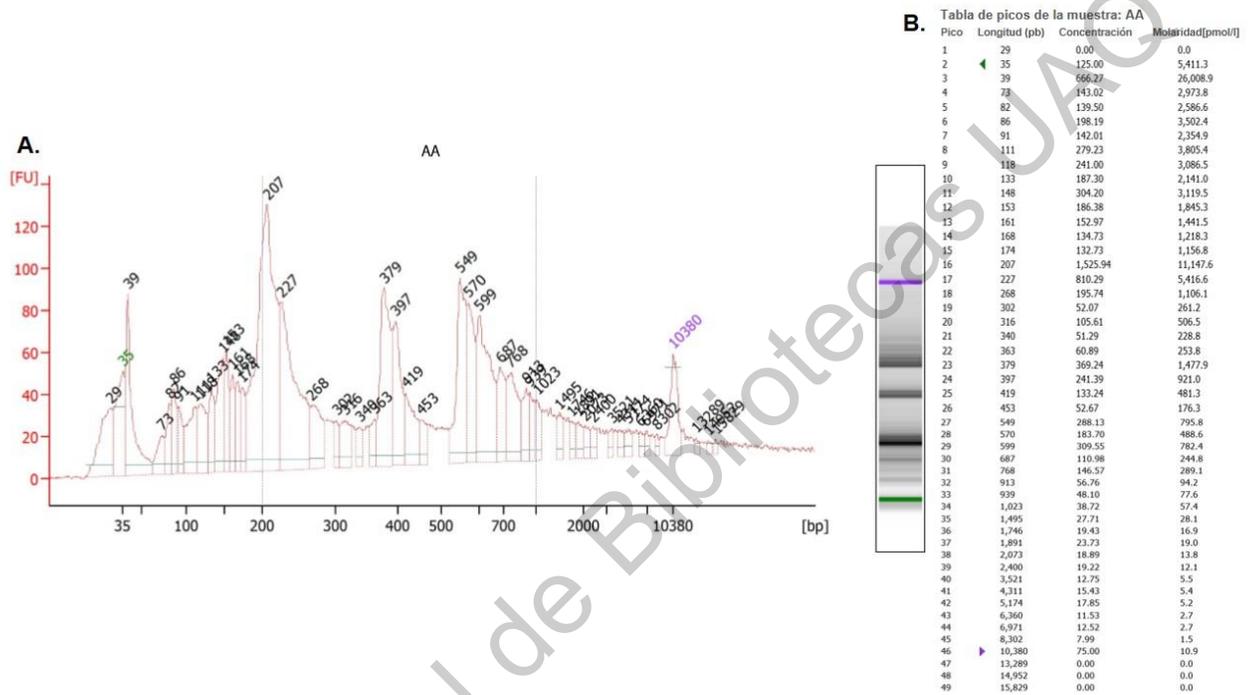


Figura A19-A| Electrograma y lecturas previas a la generación de biblioteca AA. En A, el electrograma de los fragmentos identificados previo a la generación de la biblioteca AA. En B la longitud en pb de cada pico identificado en el electrograma y la concentración de cada pico.

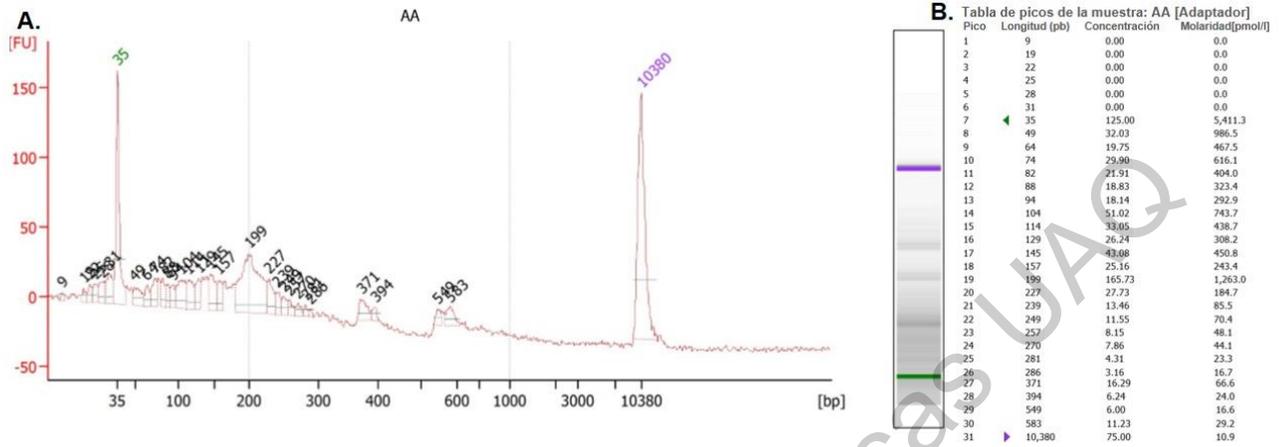


Figura A19-B| Electrograma y lecturas posteriores a la generación de biblioteca AA. En A, el electrograma de los fragmentos identificados posterior a la generación de la biblioteca AA. En B la longitud en pb de cada pico identificado en el electrograma y la concentración de cada pico.

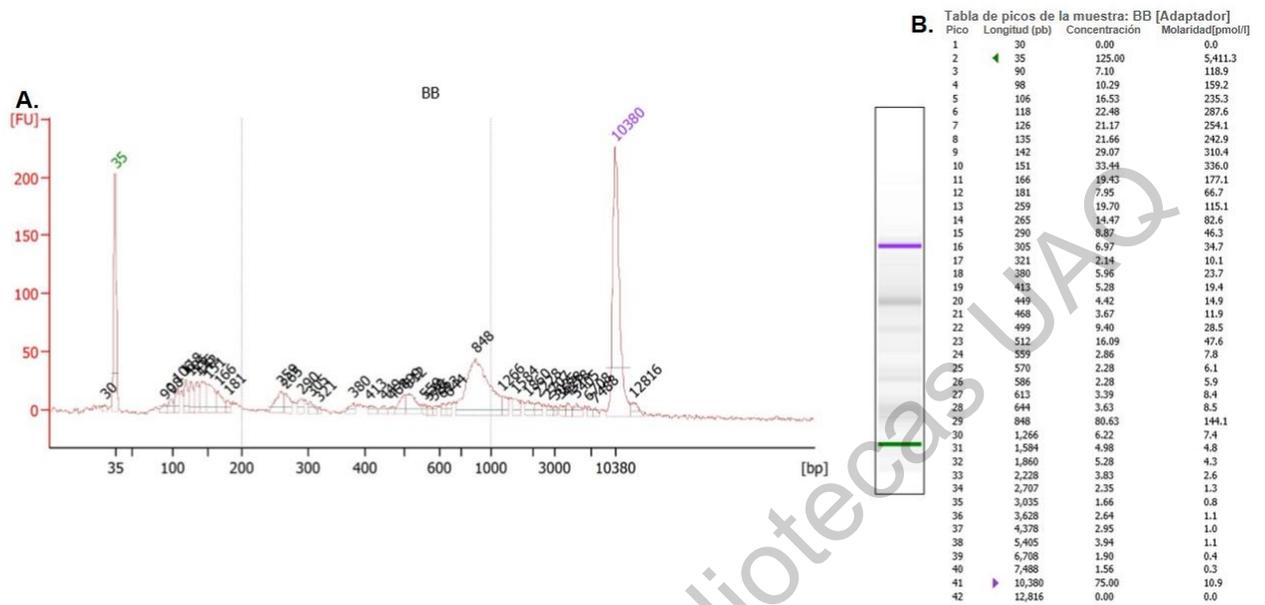
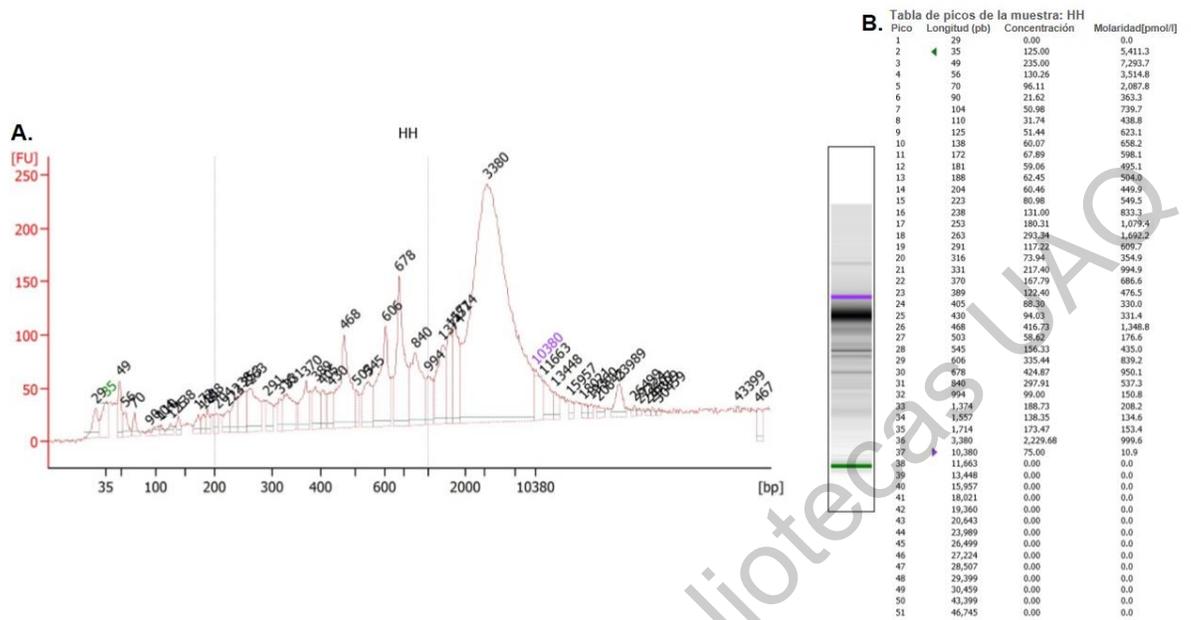


Figura A19-D| Electrograma y lecturas posteriores a la generación de biblioteca BB. En A, el electrograma de los fragmentos identificados posterior a la generación de la biblioteca BB. En B la longitud en pb de cada pico identificado en el electrograma y la concentración de cada pico.



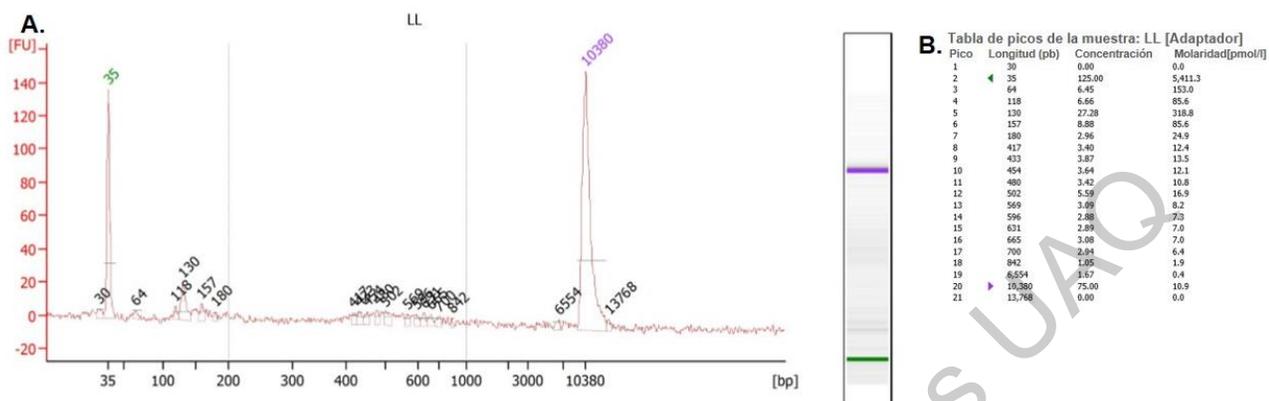


Figura A19-H| Electrograma y lecturas posteriores a la generación de biblioteca LL. En A, el electrograma de los fragmentos identificados posterior a la generación de la biblioteca LL. En B la longitud en pb de cada pico identificado en el electrograma y la concentración de cada pico.

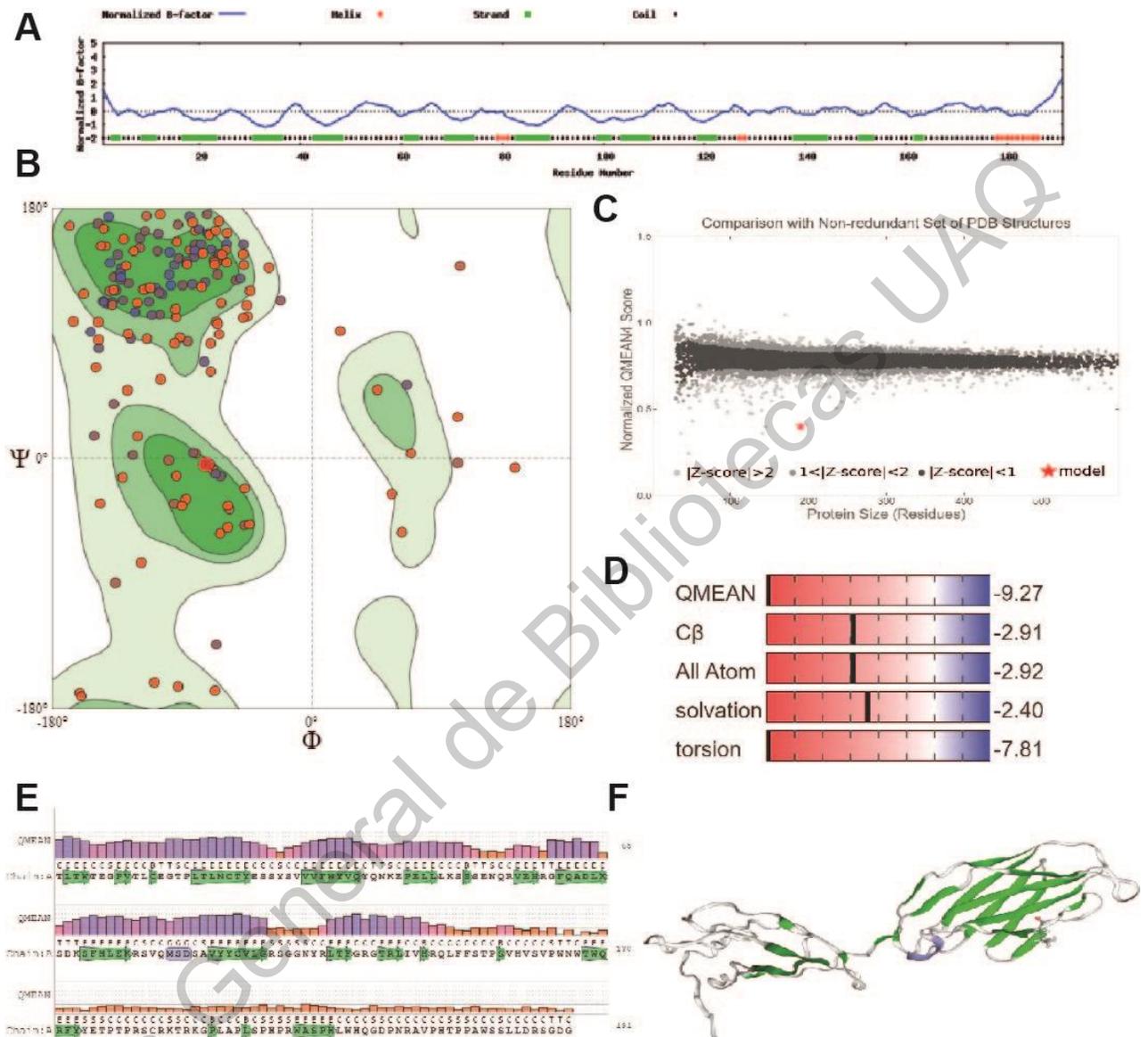


Figura A20| Análisis de calidad estructural y residual del TCRab. A, gráfico de factores B, (elasticidad de la estructura). B, gráfico de Ramachandran. C, factor Z de los residuos de la estructura TCRab. D, estimación de calidad por estructuras y características fisicoquímicas. E y F, valores de factor Z de cada residuo en la estructura TCRab

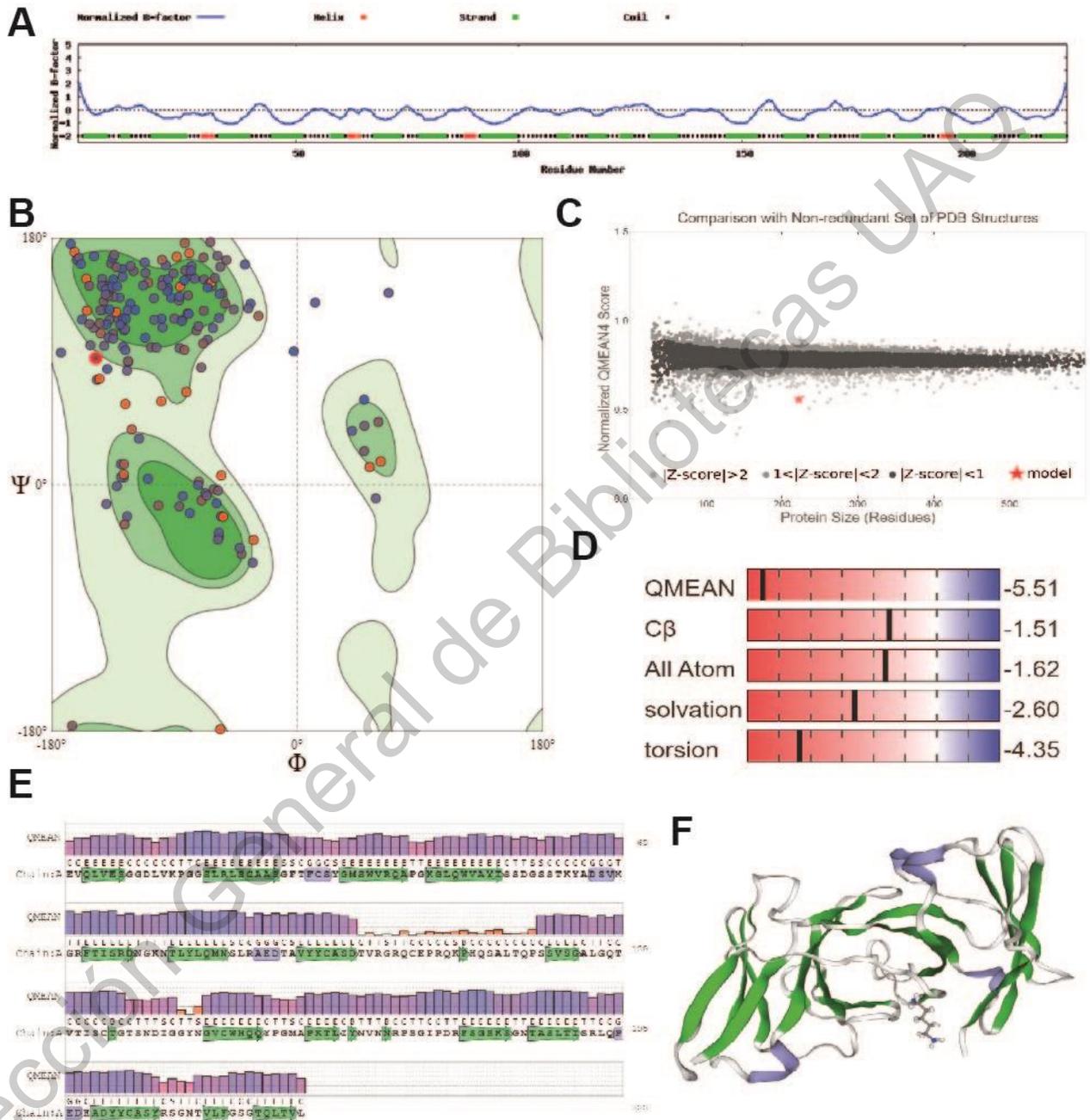


Figura A21| Análisis de calidad estructural y residual del BCRhI. A, gráfico de factores B, (elasticidad de la estructura). B, grafico de Rammachandran. C, factor Z de los residuos de la estructura BCRhI. D, estimación de calidad por estructuras y características fisicoquímicas. E y F, valores de factor Z de cada residuo en la estructura BCRhI.

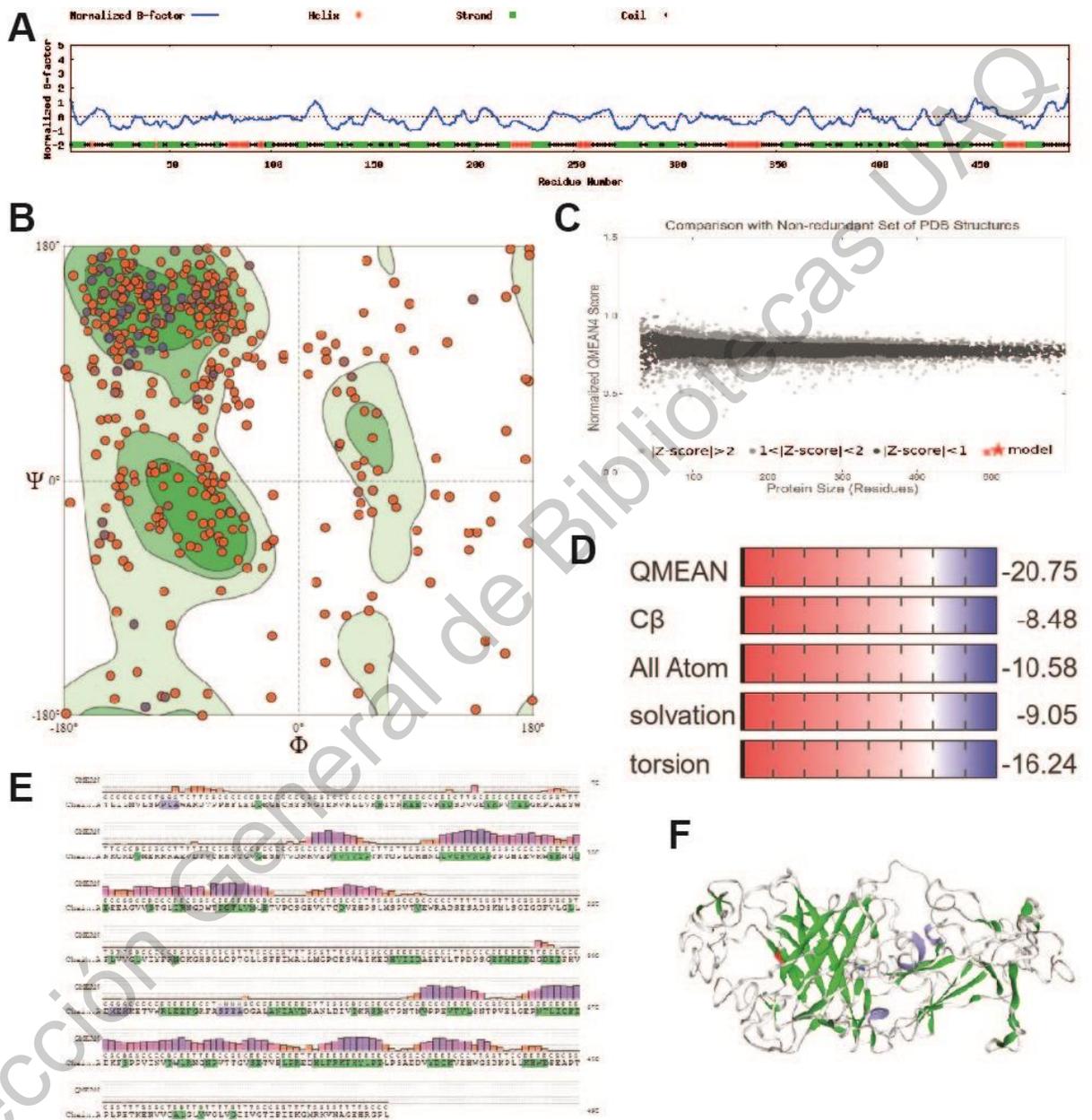


Figura A22| Análisis de calidad estructural y residual del MHC II (DRB). A, gráfico de factores B, (elasticidad de la estructura). B, gráfico de Ramachandran. C, factor Z de los residuos de la estructura MHC II. D, estimación de calidad por estructuras y características fisicoquímicas. E y F, valores de factor Z de cada residuo en la estructura MHC II.

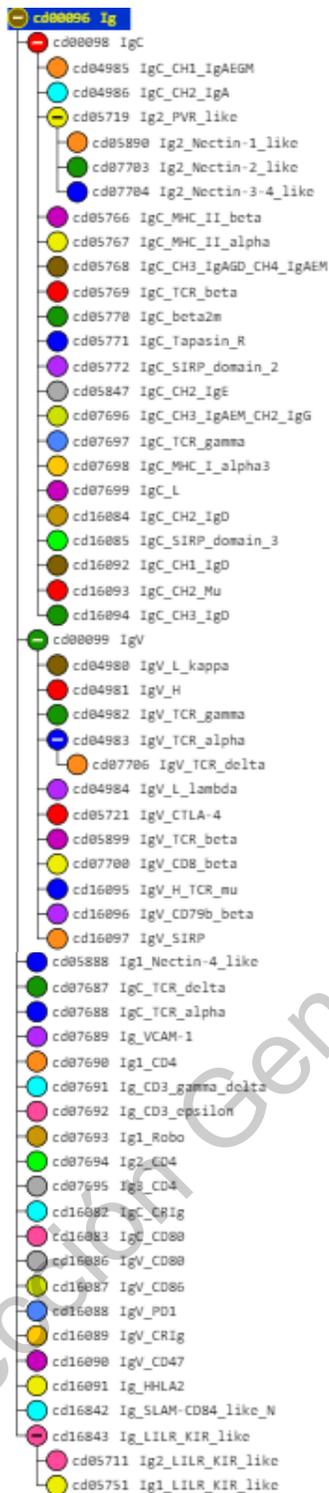


Figura A23| Identificadores de NCBI para las secuencias con dominios conservados con el TCR e Ig. Dendrograma de las regiones conservadas entre los dominios de inmunoglobulina del TCR e Ig compartidos con otros miembros de la IgSF.

Directorio General de Bibliotecas UAG

Anexo 25

Script para la selección de secuencias TCR e Ig (RStudio)

```
library(ggplot2)

#Importar la matriz
TRA <- read.csv("/AA.csv")

#Etiquetado de todas las filas que están etiquetadas como TRINITY
A <- grep("TRINITY", TRA$group)
Names <- TRA[A,]$group
TRA$group2 <- ifelse(TRA$group %in% Names, TRUE, FALSE)
table(ifelse(TRA$group %in% Names, TRUE, FALSE))

#Graficación del PCA
Ts <- ggplot(TRA, aes(x=axis1, y=axis2, color=group2)) +
geom_point() + geom_rug() + scale_color_manual(values =
c('#0099FF', '#FF9900'))
Ts

#Selección del grupo de interes
h + coord_cartesian(xlim = c(-0.2, 0.3), ylim = c(-0.4, -0.1))
TRASub <- TRA[TRA$axis1>0 & TRA$axis1<0.2,]
TRASub <- TRASub[TRASub$axis2<0.1 ,]
TrinityRows <- grep("TRINITY", TRASub$group)
TRASub_trinity <- TRASub[TrinityRows,]
TRASub_trinity <- TRASub[-TrinityRows,]

#Agrupamiento del subconjuntos como archivo csv
write.csv(TRASub, file = "TRAZaca.csv")
```

Anexo 26

Script para la selección de secuencias TCR e Ig (Samtools e IgDiscover)

```
$ conda install -c bioconda samtools

#Lista de TRINITY
$ grep "TRINITY" TRAZaca.csv > AatoSplit.txt

#Selección de secuencias TRINITY
$ xargs samtools faidx AA.fasta < AatoSplit.txt > SpS1AA.fasta
$ cat SpS1AA.fasta > TRAZaca.fasta
```

```
$ conda install -c bioconda samtools

#Iniciar el programa
$ activate source igdiscover

#Correr el programa
(igdiscover)$ igdiscover init --single-reads=TRAZaca.fasta --database=mydatabase/ myZacaTRA

#Edición del archivo igdiscover.yaml
(igdiscover)$ nano igdiscover.yaml

## configuración IgDiscover

# Cuantas iteraciones de descubrimiento ejecutar.

iterations: 1

# Tipo de secuencia: Escoge 'Ig' or 'TCR'.
#
sequence_type: TCR

## Barcoding settings

barcode_length_5prime: 0
barcode_length_3prime: 0
barcode_consensus: false

# detección de locaciones CDR3 funciona sólo con secuencias VH.
cdr3_location: 'false'

# Eliminación se oligos ligados de RACE
race_g: true
stranded: true

# List of 5' primers
#
forward_primers:
# - CGACTGGAGCACGAGGACTGA
# - GGAACTGACATGGACTGAAGGAGTA

# List of 3' primers
#
reverse_primers:
# - GCTGTCAACGATACGCTACGTAACG # TCR
# - CGCTACGTAACGGCATGACAGTG # TCR

# Trabajar solo en este número de lecturas (para ejecuciones de pruebas rápidas) Establecer en
```

```

# Filtre las lecturas combinadas que sean más cortas que esta longitud

minimum_merged_read_length: 300
merge_program: pear
flash_maximum_overlap: 300
rename: false
debug: false
seed: 1

# El filtro de preprocesamiento siempre se aplica directamente después de ejecutar IgBLAST,
# incluso si no se solicita el descubrimiento de un gen.
#
preprocessing_filter:
  v_coverage: 90 # Match must cover V gene by at least this percentage
  j_coverage: 60 # Match must cover J gene by at least this percentage
  v_value: 0.001 # Highest allowed V gene match E-value

## Configuración de descubrimiento de candidatos

# Al descubrir nuevos genes V, ignore si un gen J ha sido asignado
#y también ignorar su% SHM. verdadero: sí, ignora la J falso: no ignore la asignación de
# J, no ignore su% SHM
ignore_j: false

# Al agrupar secuencias para descubrir nuevos genes, haga una submuestra a este número de
# secuencias. Más alto es más lento.
#
subsample: 1000

# Al calcular la columna Ds_exact, considere solo D hits que
# cubren la secuencia del gen de referencia D al menos en este porcentaje.
#
d_coverage: 70

## V configuración de filtrado de candidatos (filtrado de línea germinal)

# Criterios de filtrado aplicados a secuencias candidatas en todas las iteraciones excepto la
última.
#
pre_germline_filter:
  unique_cdr3s: 2 # Minimum number of unique CDR3s (within exact matches)
  unique_js: 2 # Minimum number of unique J genes (within exact matches)
  whitelist: true # Add database sequences to the whitelist
  cluster_size: 0 # Minimum number of sequences assigned to cluster
  allow_stop: true # Whether to allow non-productive sequences containing stop codons
  cross_mapping_ratio: 0.02 # Threshold for removal of cross-mapping artifacts (set to 0 to
disable)
  clonotype_ratio: 0.12 # Required minimum ratio of clonotype counts between alleles of the
same gene
  exact_ratio: 0.12 # Required minimum ratio of "exact" counts between alleles of the
same gene
  cdr3_shared_ratio: 0.8 # Maximum allowed CDR3_shared_ratio
  unique_d_ratio: 0.3 # Minimum Ds_exact ratio between alleles
  unique_d_threshold: 10 # Check Ds_exact ratio only if highest-expressed allele has at
least this Ds_exact count

```

```

# Criterios de filtrado aplicados a secuencias candidatas en la última iteración.
# Estos deben ser más estrictos que los criterios pre_germline_filter.
#
germline_filter:
  unique_cdr3s: 5           # Minimum number of unique CDR3s (within exact matches)
  unique_js: 3             # Minimum number of unique J genes (within exact matches)
  whitelist: true          # Add database sequences to the whitelist
  cluster_size: 100       # Minimum number of sequences assigned to cluster
  allow_stop: false       # Whether to allow non-productive sequences containing stop codons
  cross_mapping_ratio: 0.02 # Threshold for removal of cross-mapping artifacts (set to 0 to
  disable)
  clonotype_ratio: 0.12   # Required minimum ratio of clonotype counts between alleles of the
  same gene
  exact_ratio: 0.12       # Required minimum ratio of "exact" counts between alleles of the
  same gene
  cdr3_shared_ratio: 0.8   # Maximum allowed CDR3_shared_ratio
  unique_d_ratio: 0.3      # Minimum Ds_exact ratio between alleles
  unique_d_threshold: 10   # Check Ds_exact ratio only if highest-expressed allele has at
  least this Ds_exact count

## configuración de descubrimiento J

j_discovery:
  allele_ratio: 0.2        # Required minimum ratio between alleles of a single gene
  cross_mapping_ratio: 0.1 # Threshold for removal of cross-mapping artifacts.
  propagate: true          # Use J genes discovered in iteration 1 in subsequent ones

(igdiscover)$ igdiscover run

```

Dirección General