


## AUTORIZACIÓN DE PUBLICACIÓN

Quien suscribe, **Nicolás Edmundo Gárate Guerrero, 19.089.717-6** alumno de la carrera de Biología, de la Facultad de Cs. Naturales y Oceanográficas, de la Universidad de Concepción, declara ser autor de la Tesis: **Reconstrucción filogenética del sistema endocannabinoide y estudio de sectores proteicos del receptor CB1** y concede derecho de publicación, comunicación al público y reproducción de esa obra, en forma total o parcial en cualquier medio y bajo cualquier forma del mismo, a la Universidad de Concepción, Chile, para formar parte de la colección material o digital de cualquiera de las bibliotecas de la Universidad de Concepción y del Repositorio UDEC. Esta autorización es de forma libre y gratuita, y considera la reproducción de la obra con fines académicos y de difusión tanto nacional como internacionalmente.

Asimismo, quien suscribe declara que dicha obra no infringe derechos de autor de terceros.



(FIRMA)

CONCEPCIÓN, 29 de septiembre de 2021



Tesis para optar al título profesional de Biólogo

# **Reconstrucción filogenética del sistema endocannabinoide y estudio de sectores proteicos del receptor CB1**

Por Nicolás Gárate Guerrero



Laboratorio de Bioquímica y Biología Celular

Departamento de Biología Celular, Facultad de Ciencias Biológicas

Facultad de Ciencias Naturales y Oceanográficas

Universidad de Concepción

2021

## HOJA DE FIRMAS

Esta tesis ha sido desarrollada en el Laboratorio de Bioquímica y Biología Celular (LBBC), Departamento de Biología Celular, Facultad de Ciencias Biológicas, Universidad de Concepción.

Dr. Fernando Sepúlveda – Profesor Guía

Dr. Cristian Hernández – Comisión Evaluadora.



Dr. Alexis Salas – Comisión Evaluadora.

Dr. Víctor Hernández – Encargado de la asignatura Seminario de Título – Biólogo.

## AGRADECIMIENTOS

Solo quiero agradecer al equipo docente que me orientó, me enseñó y me disciplinó para sacar adelante esta tesis.

Al profesor, y ahora futuro colega, Fernando Sepúlveda por adoptarme desde mi tercer año de carrera y fomentando siempre el interés y la formulación de preguntas, y también por hacer la pega de docente, psicólogo y amigo a todas horas del día. Calidad de mentor te diré.

A Sergio "Tato" Alarcón por mejorar el diseño de las figuras pertenecientes al aspecto filogenético de la tesis.

A Diego Ortiz por ayudarme con los modelos de los receptores CB1 utilizando su cuenta en el servidor Rosetta, responsable de que esta tesis se vea atractiva.

A los miembros de la comisión, Cristian por nuevamente tener que estar revisando un trabajo mío (previamente comisión de mi unidad de investigación). Y Alexis, por invitarme a sus reuniones y así tomar apuntes y sacar adelante la parte bioinformática de la tesis.

Agradecer a Felipe Aguilera, por ser mi tutor desde el año 2019 y enseñarme lo esencial de las técnicas de bioinformática, recopilación y análisis de datos. Y por seguir aportando a mi formación profesional.

Al profesor encargado del ramo y vicedecano de la facultad de la ballena, Víctor Hernández por ser tan comprensivo y dar todas las facilidades para que los alumnos saquen adelante su tesis.

A la familia y amigos les agradeceré de forma presencial ya que dudo que alguna vez lean esto, pero si por alguna razón lo hacen: gracias también.

Y me agradezco y felicito a mí mismo porque no hay mejor sensación que saber que puedes terminar algo y hacerlo.

## ÍNDICE DE CONTENIDOS

<b>1. RESUMEN</b> .....	6
<b>2. ABSTRACT</b> .....	7
<b>3. INTRODUCCIÓN</b> .....	8
3.1 Sistema de endocannabinoides .....	9
3.1.1 Síntesis de endocannabinoides .....	10
3.1.2 Degradación de endocannabinoides .....	10
3.1.3 Receptores del SEC y vías de señalización .....	10
3.2 Origen evolutivo de los componentes del SEC .....	13
3.3 Actividad de los componentes del SEC en distintos organismos .....	15
3.4 Residuos funcionales del receptor cannabinoide CB1 .....	15
<b>4. PLANTEAMIENTO DE PROBLEMA</b> .....	17
<b>5. HIPÓTESIS DE TRABAJO</b> .....	18
<b>6. OBJETIVOS</b> .....	18
6.1 Objetivos específicos .....	18
<b>7. METODOLOGÍA</b> .....	18
7.1 Recopilación de proteínas del sistema endocannabinoide .....	18
7.2 Búsqueda y confirmación de proteínas del sistema ECB en especies representativas de animales, plantas, hongos, y bacterias .....	19
7.3 Análisis filogenético de los receptores del SEC asociados a proteína G .....	20
7.4 Alineamiento múltiple de secuencias para el receptor CB1R en vertebrados.....	21
7.5 Posiciones conservadas en alineamiento múltiple de secuencias .....	21
7.6 Sectores proteicos y regiones que co-evolucionan .....	22
7.7 Modelamiento del receptor CB1 .....	23
<b>8. RESULTADOS</b> .....	24
8.1 Distribución filogenética y origen evolutivo de los componentes del sistema endocannabinoide .....	24
8.2 Receptores acoplados a proteína G: CB1R, CB2R y receptor huérfano GPR55 .....	24
8.3 Receptor cannabinoide 1 en vertebrados .....	25
8.4 Sectores proteicos del receptor CB1 .....	26
<b>9. DISCUSIÓN</b> .....	42
9.1 RAPGs del sistema endocannabinoide: CB1R, CB2R Y GPR55 .....	43
9.2 Análisis estructural e inferencia de sectores para el receptor CB1 .....	44
<b>10. CONCLUSIÓN</b> .....	46
<b>11. BIBLIOGRAFÍA</b> .....	47
<b>12. ANEXO</b> .....	60
12.1 Alineamiento múltiple de secuencias del receptor CB1 para las 200 especies de vertebrados ocupadas en este estudio .....	xx
12.2 Análisis de entropía de Shannon .....	xx

## ÍNDICE DE FIGURAS

<b>Figura 1:</b> Representación estructural del receptor CB1 .....	11
<b>Figura 2:</b> Cladograma para los componentes del SEC .....	14
<b>Figura 3:</b> Clasificación de residuos funcionales para el receptor CB1 .....	16
<b>Figura 4:</b> Cladograma para enzimas del SEC .....	28
<b>Figura 5:</b> Cladograma para receptores del SEC .....	29
<b>Figura 6:</b> Hipótesis filogenética para RAPG del SEC .....	30
<b>Figura 7:</b> Dominios transmembrana del receptor CB1 .....	31
<b>Figura 8:</b> Alineamiento múltiple de secuencias y análisis de entropía de Shannon para CB1 .....	32
<b>Figura 9:</b> Gráfico de valores de entropía de Shannon .....	33
<b>Figura 10:</b> Posición con alto valor de entropía .....	34
<b>Figura 11:</b> Sectores inferidos con el programa RocaSec .....	35
<b>Figura 12:</b> Valores de superposición entre sectores .....	36
<b>Figura 13:</b> Correlación entre sectores inferidos y dominios funcionales .....	37
<b>Figura 14:</b> Modelos tridimensionales del receptor CB1 .....	38

## ÍNDICE DE TABLAS

<b>Tabla 1:</b> Proteínas del sistema endocannabinoide mencionadas en esta investigación .....	13
<b>Tabla 2:</b> Dominios funcionales de las proteínas del SEC caracterizadas en humano .....	20
<b>Tabla 3:</b> Residuos del receptor CB1 con función conocida .....	39
<b>Tabla 4:</b> Comparación del origen evolutivo de los componentes del sistema de endocannabinoides .....	42

## 1. RESUMEN

El sistema de endocannabinoides (SEC) presenta componentes que se distribuyen en los distintos reinos en los que se organiza la naturaleza. Estudios previos sugieren un origen putativo de los distintos componentes del SEC, pero no presentan un número de especies suficiente para que sus resultados sean concluyentes. Con el fin de esclarecer esta interrogante y proponer un origen para los componentes del SEC, en este estudio se robustece el número de especies para así obtener una mejor visión del origen de las enzimas encargadas de la síntesis de endocannabinoides (DAGL $\alpha$  y NAPE-PLD), las enzimas encargadas de la degradación de endocannabinoides (MGL y FAAH) y los receptores CB1, CB2, GPR55 y TRPV1, junto con el análisis de las relaciones filogenéticas para los receptores asociados a proteína G (RAPG): CB1, CB2 Y GPR55. Mediante análisis de similitud de secuencias y presencia de dominios funcionales, se pudo determinar el origen evolutivo de los receptores del sistema de endocannabinoides (CB1, CB2, GPR55, y TRPV1) ocurrió en el último ancestro común de los sarcopterigios (Sarcopterygii). Mientras que las enzimas encargadas de la degradación (MGLL y FAAH) y síntesis (DAGL $\alpha$  y NAPE-PLD) de los endocannabinoides tiene un data más antigua que abarca el Reino Protista e incluso Bacteria. Los análisis filogenéticos de Máxima Verosimilitud e Inferencia Bayesiana indican que los receptores CB1 y CB2 están más estrechamente relacionados, mientras que los receptores GPR55 son evolutivamente más divergentes dentro de la familia de RAPG. A partir de nuestros resultados quisimos evaluar si las posiciones de residuos críticos han sufrido modificaciones que comprometan la relación estructura-función para el receptor CB1 en distintas especies de vertebrados. Bajo el principio de modularidad de las proteínas y a través del programa RocaSec pudimos dividir al receptor CB1 en sectores proteicos, que corresponden a residuos que no se encuentran en contacto en la estructura primaria, pero que estarían interaccionando en la estructura tridimensional y que están asociados a una función específica y que, además, evolucionan de manera correlacionada. Uno de los sectores inferidos contiene a aquellos residuos que están involucrados en la unión con ligandos, en mayor instancia a agonistas para el receptor CB1. Otro sector inferido contiene a aquellos residuos que están involucrados en la activación del heterotrímico proteína G. Y, por último, se obtiene un sector de solapamiento que contiene a los residuos que poseen funciones compartidas. Concluimos que las proteínas del sistema endocannabinoide a lo largo del tiempo han experimentado tanto procesos de expansión como de pérdida de genes que resultan ser linaje-específicos. Sin embargo, la distribución de las enzimas indica que éstas no están limitadas a un grupo específico de organismos, a diferencia de los receptores. Y que CB1 presenta aminoácidos que han evolucionado de manera correlacionada, asociados a la unión con ligandos y a la activación de proteína G.

**Palabras clave:** Sistema endocannabinoide, Máxima Verosimilitud, Inferencia Bayesiana, Residuos funcionales, Principio de modularidad, Sectores proteicos.

## 2. ABSTRACT

The endocannabinoid system (ECS) has components that are distributed in the different kingdoms in which nature is organized. Previous studies suggest a putative origin of the different components of the ECS, but they do not present a sufficient number of species for their results to be conclusive. In order to clarify this question and propose an origin for the components of the ECS, in this study the number of species is strengthened in order to obtain a better vision of the origin of the enzymes responsible for the synthesis of endocannabinoids (DAGL $\alpha$  and NAPE-PLD), the enzymes responsible for the degradation of endocannabinoids (MGL and FAAH) and the CB1, CB2, GPR55 and TRPV1 receptors, along with the analysis of the phylogenetic relationships for the G protein-associated receptors (GPCRs): CB1, CB2 and GPR55. Through sequence similarity analysis and the presence of functional domains, the evolutionary origin of the receptors of the endocannabinoid system (CB1, CB2, GPR55, and TRPV1) occurred in the last common ancestor of sarcopterygii (Sarcopterygii). While the enzymes responsible for the degradation (MGLL and FAAH) and synthesis (DAGL $\alpha$  and NAPE-PLD) of endocannabinoids have an older date that includes the Protista Kingdom and even Bacteria. Maximum Likelihood and Bayesian Inference phylogenetic analyzes indicate that CB1 and CB2 receptors are more closely related, while GPR55 receptors are evolutionarily more divergent within the GPCR family. Based on our results, we wanted to evaluate whether the positions of critical residues have undergone modifications that compromise the structure-function relationship for the CB1 receptor in different vertebrate species. Under the principle of modularity of proteins and through the RocaSec software, we were able to divide the CB1 receptor into protein sectors, which correspond to residues that are not in contact in the primary structure, but that would be interacting in the three-dimensional structure and that are associated with a specific function and that, moreover, evolve in a correlated way. One of the inferred sectors contains those residues that are involved in ligand binding, mostly agonists for the CB1 receptor. Another inferred sector contains those residues that are involved in the activation of the protein G heterotrimer. And, finally, an overlapping sector is obtained that contains the residues that have shared functions. We conclude that the proteins of the endocannabinoid system over time have undergone both gene expansion and loss processes that turn out to be lineage-specific. However, the distribution of enzymes indicates that they are not limited to a specific group of organisms, unlike receptors. And that CB1 has amino acids that have evolved in a correlated way, associated with ligand binding and G protein activation.

**Keywords:** Endocannabinoid system, Maximum Likelihood, Bayesian Inference, Functional residues, Principle of modularity, Protein sectors.



### 3. INTRODUCCIÓN

La especie *Cannabis sativa*, conocida comúnmente como marihuana, tiene un historial de haber sido utilizada con fines medicinales a lo largo de la historia humana (Mechoulam, 1986). No fue sino hasta medio siglo atrás que el primer indicio de la versatilidad de la marihuana como sustancia recreacional y terapéutica se dio a conocer gracias al descubrimiento del  $\Delta^9$ -tetrahidrocannabinol (THC), el principal componente psicoactivo entre aproximadamente 70 fitocannabinoides identificados en la planta (Pacher et al. 2006). Este hallazgo trajo como consecuencia la síntesis de una variedad de cannabinoides sintéticos con estructuras similares y diferentes a los fitocannabinoides ya descritos, lo que conllevó a la identificación y posterior clonamiento del receptor cannabinoide 1 (CB1R) (Matsuda et al. 1990). No mucho después un segundo receptor cannabinoide sería identificado y clonado, denominado receptor cannabinoide 2 (CB2R) (Munro et al. 1993). A pesar del hecho de que CB1R y CB2R son los principales receptores cannabinoides, una variedad de receptores, desde otros receptores asociados a proteína G (RAPG) hasta receptores de canal iónico han sido descritos como blancos para cannabinoides (Howlett et al. 2002, Kano et al. 2009). Por otra parte, las sustancias N-araquidonil-etanolamina (AEA, Anandamida) y 2-araquidonil glicerol (2-AG) han sido descritas como agonistas endógenos para los receptores de cannabinoides, denominados endocannabinoides (Sugiura et al. 1995). Estos componentes presentan distintas vías de síntesis y degradación en el organismo, pero ambos derivan de la estructura del ácido araquidónico (Pacher et al. 2006).

Las enzimas encargadas de la síntesis y degradación de los endocannabinoides, en conjunto con los receptores de cannabinoides corresponden a los principales componentes del sistema de endocannabinoides (SEC) en el organismo. La activación de los receptores del SEC juega un papel importante en el desarrollo cognitivo (Howlett et al. 2002), la señalización neuronal y función sináptica (Castillo et al. 2012), junto con la ingesta alimenticia y el balance energético

(Koch et al. 2015), por lo cual resulta llamativo comprender los mecanismos de activación de estos receptores.

Un estudio de McPartland y colaboradores el año 2006 sugiere la presencia de los componentes del SEC en distintos organismos de los reinos en que se organiza la naturaleza. Dicha investigación se llevó a cabo utilizando 12 especies pertenecientes a distintos grupos taxonómicos correspondientes a: primates, tunicados, apicomplexos, ciliados, murinos, peces, insectos, nemátodos, fungi, plantas y procariotas. Si bien este trabajo nos entrega una visión de los orígenes evolutivos de los componentes del SEC, no abarca un amplio espectro de especies representativas de las grandes subdivisiones taxonómicas en las que se clasifican los seres vivos. Por lo tanto, el origen evolutivo y específicamente la historia evolutiva de las proteínas y enzimas asociadas al SEC se encuentra en debate.

### **3.1. Sistema de endocannabinoides**

Los endocannabinoides son ligandos de origen lipídico que ejercen una amplia variedad de funciones en el organismo. Los efectos producidos por este tipo de moléculas resultan ser parcialmente análogos con aquellos producidos por los componentes psicoactivos de la planta de cannabis, principalmente el THC (Luk et al. 2004). Los primeros y mejor caracterizados endocannabinoides son anandamida y 2-AG, cuyos precursores se encuentran en membranas lipídicas, y ante demanda (ya sea por activación de RAPG o despolarización) son liberados al espacio extracelular mediante 1 o 2 pasos enzimáticos. Lo anterior resulta contrastante con la manera en que los neurotransmisores generalmente son sintetizados antes de tiempo y almacenados en vesículas sinápticas. La eficacia intrínseca de los endocannabinoides varía de una manera en que 2-AG es un agonista de alta eficacia tanto para CB1 como para CB2, sin embargo, anandamida es un agonista de baja eficacia para el receptor CB1, y mucho más baja para el receptor CB2 (Gonsiorek et al. 2000; Luk et al. 2004).

### **3.1.1. Síntesis de endocannabinoides**

A pesar de que tanto anandamida como 2-AG contienen la estructura del ácido araquidónico, sus rutas de síntesis y degradación *in vivo* difieren en algunos pasos enzimáticos, y son mediadas por diferentes moléculas (Pacher et al. 2006). La mayor cantidad de anandamida es sintetizada por N-araquidonil fosfatidil etanol (NAPE), y la primera vía de síntesis identificada para anandamida es a partir de la hidrólisis de NAPE por una fosfolipasa D (NAPE-PLD) (Di Marzo et al. 1994).

Por otra parte 2-AG es producido por fosfolípidos que contengan 2-araquidonil en su estructura, como es el caso de fosfatidil inositol bis-fosfato (PIP2). Gran parte de 2-AG es sintetizado por una serie de hidrolisis de PIP2 por una beta-fosfolipasa C ( $\beta$ PLC), seguido por una hidrólisis del diacilglicerol resultante por acción de un alfa-diacilglicerol lipasa (DAGL $\alpha$ ) (Murataeva et al. 2014).

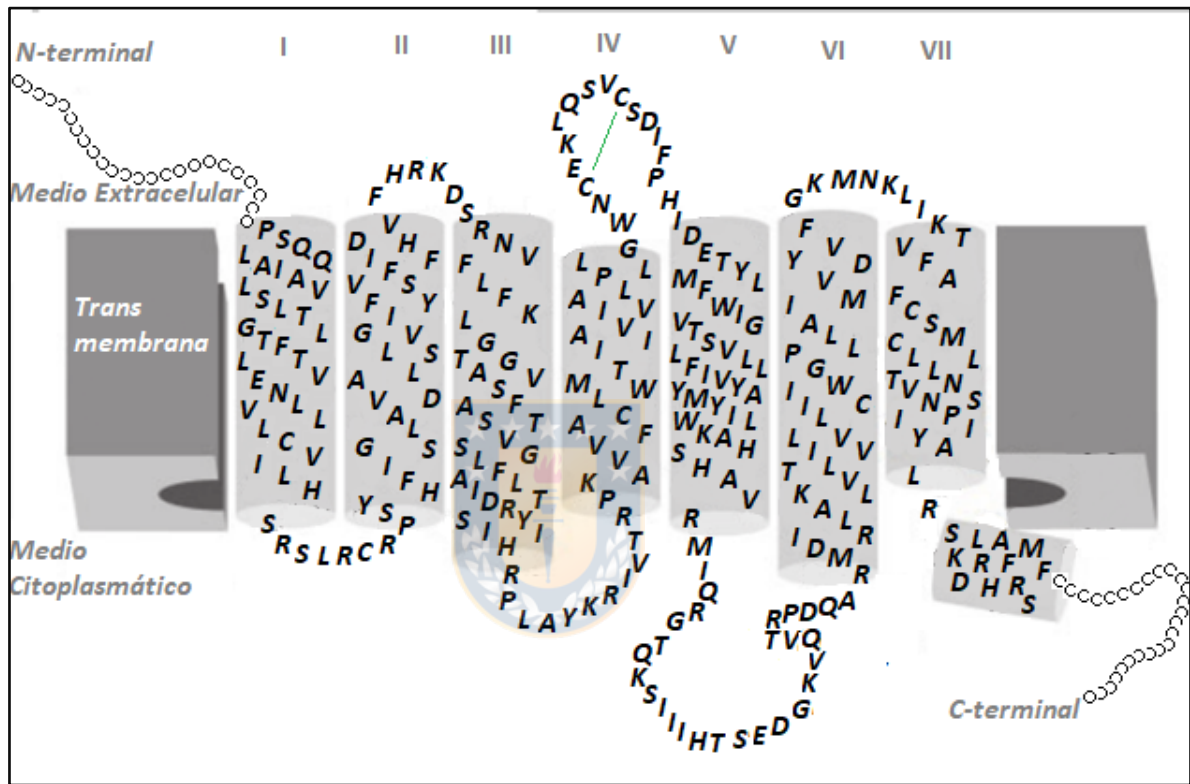
### **3.1.2 Degradación de endocannabinoides**

La degradación de la anandamida en el sistema nervioso central se lleva a cabo principalmente por la enzima amino hidrolasa de ácidos grasos (por sus siglas en inglés, FAAH) que, precisamente, degrada múltiples ácidos grasos tales como palmitoil y oleoil etalonamida (Cravatt et al. 1996). Por otra parte, la degradación de 2-AG ocurre gracias a la acción de 3 enzimas hidrolíticas, principalmente monoacil glicerol lipasa (MGL), e hidrolasas 6 y 12 con sus dominios alfa/beta (ABHD6 y 12) (Blankman et al. 2007).

### **3.1.3. Receptores del SEC y vías de señalización**

Los efectos de los endocannabinoides son mediados principalmente por los receptores cannabinoides CB1R y CB2R, ambos corresponden a RAPGs pertenecientes a la familia de la rodopsina (Fredriksson et al. 2003), que generan complejos con proteína G de clases Gi y Go (Howlett et al. 2002). Estos receptores comparten una característica estructural común que corresponde a los 7 dominios transmembrana alfa hélice (7TM), conectados por 3 loops extracelulares (ECL1-3) y 3 loops intracelulares (ICL1-3) (Figura 1) (McPartland

& Pruitt 2002). La región 7TM de los RAPG forman el núcleo o “core” estructural, uniendo ligandos y realizando transducción de señales desde el medio extracelular hasta el medio intracelular a través de cambios conformacionales, y los componentes de la región intracelular interactúan con proteína G, arrestinas, quinasas intracelulares asociadas a proteína G (GIRK) y otros efectores de señalización (Zhang et al. 2015).



**Figura 1. Representación estructural del receptor CB1.** Ilustración representativa del receptor cannabinoide CB1R. Indica dominios transmembrana, y loops extra- e intracelulares.

La activación de estos receptores cannabinoide está involucrada en procesos cognitivos y asociada a funciones neuroprotectoras entre otros aspectos neurofisiológicos (Howlett et al. 2002). Respecto a la función neuroprotectora, se sabe que los receptores cannabinoide están asociados a las quinasas de señal extracelular regulada 1 y 2 (ERK1 y ERK2) (Asimaki & Mangoura 2011), y numerosos estudios indican que CB1 activa ERK (Bouaboula et al. 1995, Galve-Roperh et al. 2002, Davis et al. 2003, Derkinderen et al. 2003, Rubovitch et al.

2004). Procesos como la proliferación celular y la apoptosis son consecuencia de la actividad de ERK activado por CB1 (Bouaboula et al. 1995, Galve-Roperh et al. 2002).

CB1 y CB2 interactúan con múltiples ligandos para llevar a cabo procesos de transducción de señales (Pertwee 2008). La activación de CB1 conlleva a interacciones de señalización con proteína G y B-arrestina, presumiblemente comprometiendo la conformación del receptor (principio del agonismo sesgado) (Laprairie et al. 2017). Al mismo tiempo estas interacciones están involucradas en la regulación de los niveles de AMP cíclico intracelular (Eldeeb et al. 2016), como también la activación de los canales de potasio asociados a proteína G (GIRK) (Guo & Ikeda 2003).

Por otra parte, la activación de CB2 inhibe la acción de las adenilil ciclasas disminuyendo la cantidad de AMPc intracelular (Mukherjee et al. 2004), y activa MAPK involucrado en procesos de crecimiento celular y la expresión de genes (Bouaboula et al. 1996).

En la última década, estudios dirigidos a la selectividad de los ligandos de CB1 y CB2 en ratones knockout de estos receptores, sugieren la existencia de receptores cannabinoides distintos a los ya mencionados CB1 y CB2 (Begg et al. 2005). Recientemente el receptor huérfano GPR55 demostró afinidad con ligandos cannabinoides por lo que se presenta como un posible “tercer” receptor cannabinoide (Ryberg et al. 2007). El hecho de referirse a este receptor como proteína huérfana significa que no tiene afinidad filogenética con otras RAPGs pero conserva el dominio transmembranal 7tm\_1, característica que la hace pertenecer a esta familia de proteínas receptoras (Sawzdargo et al. 1999; Tang et al. 2012).

Por último, los canales de potencial de receptor transciente (o por sus siglas en inglés, TRP) también resultan ser mediadores de la acción endocannabinoide, particularmente de las aciletanolamidas. Estos canales, especialmente TRPV1 son activados por anandamida bajo ciertas condiciones de vasodilatación (Zygmunt et al. 1999).

Tabla 1

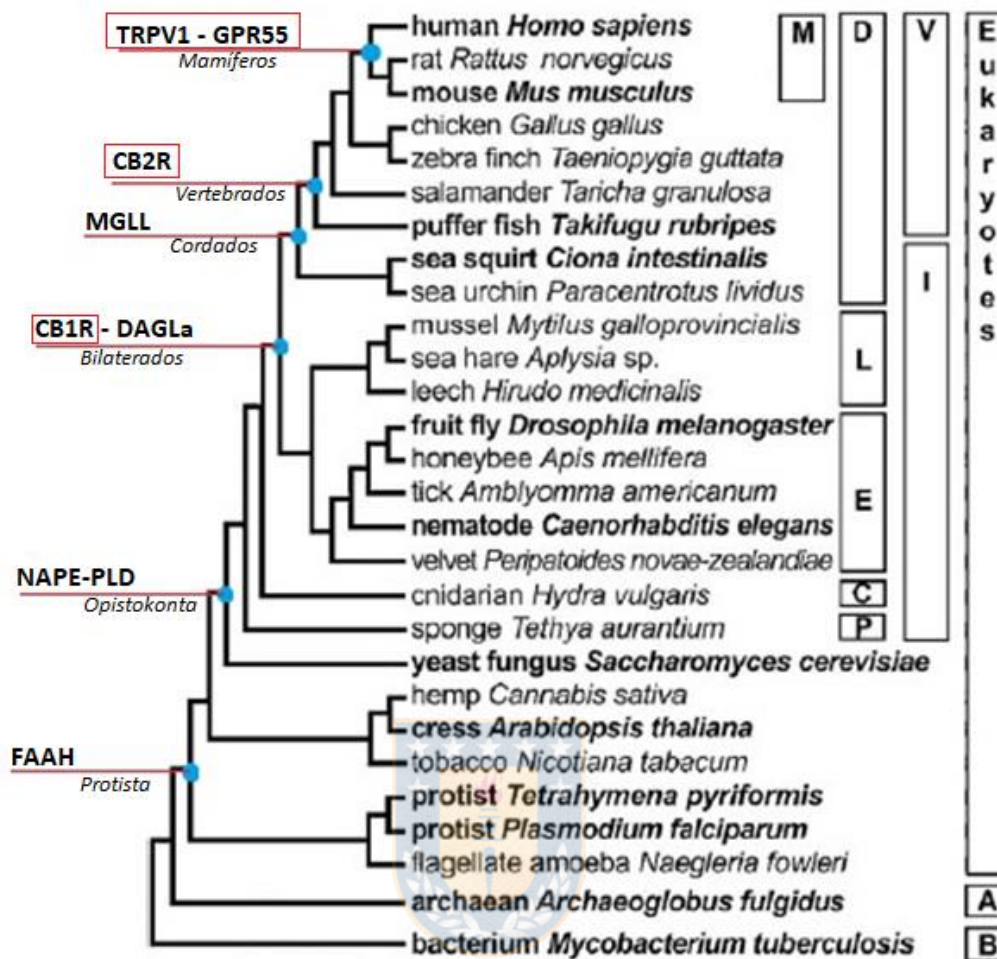
Proteínas del sistema de endocannabinoides mencionadas en esta investigación

Receptores	Enzimas para la síntesis	Enzimas para la degradación
CB1R	DAGL $\alpha$ (2-AG)	MGL (2-AG)
CB2R	NAPE-PLD (AEA)	FAAH (AEA)
GPR55		
TRPV1		

### 3.2. Origen evolutivo de los componentes del SEC

El sistema de endocannabinoides es un complejo de proteínas que se expresa de forma extensa en la naturaleza. En el año 2006, McPartland y colaboradores llevaron a cabo una investigación acerca de los componentes del SEC que hasta esa fecha se conocían, correspondientes a las ya mencionadas proteínas receptoras de endocannabinoides, y las enzimas encargadas de la degradación y síntesis de estas sustancias (Tabla 1). Mediante estrategias *in silico*, principalmente el uso de algoritmos que predicen dominios proteicos o estructuras especializadas, desarrollaron una metodología de la cual se desprenden los primeros indicios para determinar el origen e historia evolutiva de los componentes del SEC. Los resultados que arrojó este estudio indican que los receptores funcionales TRPV1 y GPR55 estaban presentes de manera exclusiva en el grupo de los mamíferos, mientras que el receptor cannabinoide CB2R pareciera tener un origen evolutivo a nivel de los vertebrados y la enzima encargada de la degradación de 2-AG (MGL) a nivel de los cordados.

Los autores también determinan el origen evolutivo de CB1R y DAGL $\alpha$  en el último ancestro común de los bilaterados, el de la enzima NAPE-PLD en el último ancestro común de los opisthokonta (Metazoa + Fungi), y la enzima FAAH en el último ancestro común de los eucariotas (Figura 2).



**Figura 2. Cladograma para componentes del SEC.** Cladograma que indica los potenciales orígenes evolutivos de las proteínas del SEC en los distintos reinos en los que se organiza la naturaleza. Encerrados en un rectángulo rojo se indican los receptores de cannabinoides (Modificado de McPartland y colaboradores el 2006.)

El estudio abordado por el grupo de investigadores encabezado por McPartland el 2006, nos entrega un origen putativo de los componentes del SEC en distintas especies. Sin embargo, esta investigación consideró un número acotado de especies lo cual no refleja de forma precisa la distribución filogenética de estos receptores.

### 3.3. Actividad de las proteínas del SEC en distintos organismos

Destacando los principales resultados del estudio de McPartland y colegas el 2006, llama la atención si las proteínas del SEC se comportarán de la misma manera en todas las especies. Estudios previos se han realizado en los receptores cannabinoides, principalmente en el grupo de los mamíferos tales como ratones y ratas (Navarro et al. 2001). Respecto a la función fisiológica, existe evidencia de que los cannabinoides pueden inducir signos de alteración de los aspectos sensoriales de la intensidad del dolor en modelos de dolor agudo a través de los receptores CB1 ubicados en las neuronas tanto dentro como fuera del cerebro y la médula espinal (Pertwee 2001).

La activación de los receptores del SEC está involucrada tanto en el desarrollo cognitivo (Howlett et al. 2002), la señalización neuronal y función sináptica (Castillo et al. 2012), como en la ingesta alimenticia y el balance energético (Koch et al. 2015). Sin embargo, de estar presentes estos receptores en distintos organismos, a la fecha se desconocen los potenciales efectos de la activación de estas proteínas en diferentes especies. Una manera de obtener una mirada preliminar es estudiar los mecanismos de activación de estos receptores, desde el punto de vista estructural, a partir de la secuencia de aminoácidos.

### 3.4. Residuos funcionales del receptor cannabinoide CB1

Para comprender mecanismos de activación a partir de la secuencia de aminoácidos de una proteína es necesario considerar que si la afinidad que tiene un ligando se define por la fuerza con la se une a un receptor, y la eficacia del receptor se define por la efectividad con la que un receptor activa la proteína G, un ligando como tal debería mostrar una alta afinidad de unión y, lo que es más importante, una eficacia de unión deseable (Jäger et al. 2007).

El investigador Joon-Youn Shim el 2010 intenta dividir los residuos funcionales del receptor CB1R clasificándolos en 3 grupos diferentes: unión o afinidad con ligando (**U**), estabilización estructural del receptor (**E**), y activación del receptor (**A**). Es interesante como todos los residuos en conjunto pueden afectar de



distintas maneras la activación del receptor CB1R, por lo que resulta complejo definir la función de un residuo específico como única, ya que sus roles están asociados entre ellos (Figura 3). Si tuviéramos que agregar una nueva nomenclatura a los residuos encargados de la unión del ligando (**U**) que están en contacto directo (**d**) con el ligando, las mutaciones de tipo **U<sub>d</sub>** resultan en la alteración de la unión con ligando, junto con la alteración de sitio de unión del ligando. Una mutación a un residuo asociado a la función de estabilizador estructural (**E**) trae como consecuencia un mal plegamiento del receptor, y si esta mutación al residuo de tipo **E** indirectamente (**i**) afecta la topología del sitio de unión, la podemos clasificar como una mutación **U<sub>iE</sub>**. Los residuos de tipo **A** están directamente involucrados a la activación del receptor, y cualquier mutación comprometería la señalización con proteína G. Mutaciones a residuos de tipo A han demostrado afectar indirectamente la topología del sitio de unión, alterando la afinidad con ligando (Song & Feng 2002), lo anterior puede clasificarse como una mutación de tipo **U<sub>iA</sub>**.



**Figura 3. Clasificación de residuos funcionales para el receptor CB1.** El diagrama refleja la manera en que están asociados los residuos funcionales del

receptor CB1R, y cómo todos estarían afectando la señalización con proteína G ante mutagénesis (Modificado de la revisión de Joon-Youn Shim el 2010).

Teniendo acceso a la información de los residuos funcionales del receptor CB1R con su respectiva función, posición y ubicación espacial, podríamos realizar análisis comparativos entre especies que pudieran tener también este receptor y de esta manera obtener las primeras diferencias estructurales para, posteriormente identificar cuáles son los residuos más importantes para la funcionalidad del receptor. Lo anterior, bajo el principio de secuencia-estructura-función descrito por Anfinsen el año 1973, que señala que la secuencia de aminoácidos de la proteína determina su estructura tridimensional y su función bioquímica.

Con respecto a lo anterior, grupos diferentes de residuos se asocian a funciones o propiedades moleculares de una proteína. Propiedades como la estabilidad estructural, afinidad o unión con ligando y activación de proteína G podrían estar bajo presión selectiva, esto corresponde al principio de modularidad. Un grupo de investigadores encabezado por Halabi el 2009 realizó un estudio que incluye a la familia de Serina Proteasas S1A, que integra análisis bioinformáticos de los cuales se determinó una descomposición de las proteínas en tres grupos cuasi-independientes de aminoácidos correlacionados a los cuales denominó sectores proteicos. Cada sector está físicamente conectado en la estructura terciaria, tiene un rol funcional distintivo y constituye un modo independiente de divergencia de secuencias.

#### **4. PLANTEAMIENTO DE PROBLEMA**

Estudios previos no cubren un amplio espectro de organismos respecto a la filogenia de los componentes del sistema de endocannabinoides. Si bien el estudio realizado por McPartland y colaboradores el año 2006 nos entrega un indicio del origen evolutivo de las proteínas del SEC, esta información no refleja de forma precisa la manera en que se distribuyen estas proteínas en los organismos vivos.

A la fecha información acerca de los aspectos estructurales del receptor cannabinoide 1 no resulta suficiente, y no existe información que indique la actividad biológica del receptor CB1 en distintos organismos, como tampoco de su relación con endocannabinoides.

## 5. HIPOTESIS DE TRABAJO

A lo largo de la historia evolutiva, posiciones de residuos críticos han sufrido modificaciones que comprometen la relación estructura-función para el receptor cannabinoide 1 en distintas especies de vertebrados.

## 6. OBJETIVOS

Reconstrucción del origen evolutivo de los componentes del SEC.

Caracterización de componentes estructurales que influyen en la función del receptor CB1 en distintas especies.

### 6.1. *Objetivos específicos*

- Determinar el origen evolutivo de los componentes del SEC.
- Identificación de residuos funcionales de del receptor CB1R en distintas especies de vertebrados.
- Evaluar la conservación de la estructura primaria del receptor CB1.

## 7. METODOLOGÍA

### 7.1. **Recopilación de proteínas del sistema endocannabinoide**

Se realizó una recopilación de las secuencias de aminoácidos de las proteínas del sistema endocannabinoide descritas en Humano (*Homo sapiens*), Ratón (*Mus musculus*), y Rata (*Rattus norvegicus*). Estas secuencias fueron descargadas desde la base de datos UniProtK (Uniprot Consortium, 2008) y ocupadas como referencias para la posterior caracterización de los dominios funcionales para cada una utilizando las bases de datos InterPro (Hunter et al. 2009) y PANTHER (Thomas et al. 2003).

### 7.2. **Búsqueda y confirmación de proteínas del sistema ECB en especies representativas de animales, plantas, hongos, y bacterias.**

Una vez caracterizadas las proteínas del sistema ECB en humano, ratón y rata acudimos a repositorios que contienen los genomas completos de animales, plantas, hongos y bacterias, con el fin de abarcar un amplio rango de grupos taxonómicos de organismos vivos. Estos repositorios corresponden a Ensembl Genome Browser 97, OIST Marine Genomics Unit, OIST Hemichordata Genomics, Echinobase, WormBase ParaSite, NCBI, Genoscope, Compagen, Joint Genome Institute, UCSC Genome Bioinformatics, Amphien Code, JGI Genome Portal, Axolotl-Omics, Xenbase, y Srivastava Lab Site.

La búsqueda de secuencias de aminoácidos similares a las proteínas del sistema endocannabinoide de *H.sapiens*, *M.musculus* y *R.norvegicus* en las 144 especies reunidas de animales, plantas, hongos y bacterias se realizó mediante análisis de similitud local con BLASTp. Si bien un parámetro para determinar la homología de una proteína corresponde a un índice de expectación de valor  $1 \times 10^{-6}$  (Camacho et al. 2009), nuestra investigación incluye resultados obtenidos con un valor umbral de  $1 \times 10^{-272}$ .

Secuencias similares a las proteínas del SEC fueron sometidas a análisis de dominios funcionales con el programa InterProScan (Jones et al. 2014) con el fin de corroborar la presencia de dominios previamente caracterizados de ser fundamentales en la función de las proteínas de referencia. La presencia de los dominios funcionales que se muestran en la Tabla 2 permitirá discriminar entre proteínas del sistema ECB y otras con alto nivel de similitud.

**Tabla 2**

Dominios funcionales de las proteínas del SEC caracterizadas en humano.

<b>Proteína</b>	<b>Dominio</b>	<b>Código de acceso (InterPro)</b>	<b>Código de acceso (Panther)</b>
CB1R	Cannabinoid receptor type 1	IPR000810	PTHR22750:SF47

CB2R	Cannabinoid receptor type 2	IPR001551	PTHR22750:SF10
GPR55	GPCR 55 Orphan	IPR028334	PTHR24232:SF56
TRPV1	TRPC channel subfamily V member 1	IPR024863	PTHR10582:SF17
DAGL	Lipase_3	IPR002921	PTHR45792:SF3
NAPE-PLD	Lactamase_B2	IPR001279	PTHR15032:SF4
MGL	Hydrolase_4	IPR022742	PTHR11614:SF87
FAAH	Amidase	IPR023631	PTHR45847:SF3

### 7.3. Análisis filogenético de los receptores del SEC asociados a proteína G.

Las proteínas confirmadas como parte del SEC fueron usadas para análisis filogenéticos del tipo Máxima Verosimilitud e inferencia Bayesiana.

Para ello y para cada RAPG se realizó un alineamiento múltiple de secuencias con el programa MAFFT (Kato et al. 2019), los alineamientos de cada proteína del sistema endocannabinoide fueron posteriormente sometidos a un refinamiento de las secuencias con el programa trimAL (Capella-Gutierrez et al. 2009), lo que implica una remoción de regiones débilmente alineadas y, a partir de parámetros como el número de aperturas y similitud de aminoácidos, de forma automática optimiza la obtención de datos con alta señal biológica.

Posteriormente se seleccionó el mejor modelo evolutivo para el alineamiento de los receptores del SEC con la ayuda del programa ProtTest (Darriba et al. 2019), luego realizamos los análisis filogenéticos de Máxima Verosimilitud e Inferencia Bayesiana. Para Máxima Verosimilitud la reconstrucción filogenética se realizó con el software RaxML (Stamatakis 2009), y un valor de bootstrap de 1000, mientras que los análisis de inferencia Bayesiana fueron realizados con MrBayes (Ronquist et al. 2012), en conjunto con cuatro cadenas de Markov Monte Carlo

(MCMC) simultáneas y un número de generaciones de 10.000.000 para los receptores CB1R, CB2R y GPR55. La convergencia de las cadenas MCMC para cada una de los análisis de inferencia bayesiana fue realizado con el programa Tracer (<http://tree.bio.ed.ac.uk/software/tracer/>)

Finalmente, los árboles filogenéticos fueron visualizados con el programa FigTree ([www.tree.bio.ed.ac.uk](http://www.tree.bio.ed.ac.uk)).

#### **7.4. Alineamiento múltiple de secuencias para el receptor CB1R en vertebrados**

Una vez confirmada la presencia del receptor CB1R en el clado de los vertebrados, procedimos a recopilar un total de 200 secuencias descritas para el receptor cannabinoide 1 en vertebrados desde la base de datos Ensembl. Posterior a ello, repetimos la metodología correspondiente al alineamiento de las secuencias con el programa MAFFT (Kato et al. 2019), y luego un posterior refinamiento del alineamiento con el programa TrimAL (Capella-Gutierrez et al. 2009) lo que nos arrojó un largo de alineamiento de 467 aminoácidos. La visualización del alineamiento se hizo a través del programa Esript3 (Robert & Gouet 2014), proporcionando la secuencia de la estructura cristalizada de CB1R en humano (PDB:5TGZ) (Hua et al. 2016), para así observar la estructura secundaria del receptor en regiones del alineamiento.

#### **7.5. Posiciones conservadas en alineamiento múltiple de secuencias**

Las posiciones totalmente conservadas tienden a formar parte del núcleo estructural de la proteína y también se encuentran en regiones funcionales, como en sitios de interacción y sitios catalíticos, lo que se puede discriminar mediante análisis topológicos-estructurales. Suelen identificarse mediante análisis de entropía de Shannon (Shannon, 1948), y nosotros lo realizamos con la ayuda del servidor HIV Databases, perteneciente a Los Alamos National Laboratory (<https://www.hiv.lanl.gov/content/sequence/ENTROPY>).

El análisis de entropía de Shannon (H) para cada posición  $i$  de un MSA se define según la siguiente ecuación (Cover & Thomas 1991):

$$H_i = - \sum_x p_i(x) \log_2 p_i(x)$$

Donde  $P_i(x)$  es la fracción de residuos del aminoácido tipo  $x$  presentes en la posición  $i$ , y  $x$  es el número de tipos de aminoácidos (20).  $H$  varía de 0 (solo un residuo presente en esa posición) a 4.322 (los 20 residuos están igualmente representados en esa posición). Las posiciones con  $H$  mayor a 2.0 se suelen considerar variables, mientras que aquellas con  $H$  menor a 2 se consideran conservadas. Las posiciones altamente conservadas son aquellas con  $H$  menor a 1.0 (Litwin & Jores 1992).

### **7.6. Sectores proteicos y regiones que co-evolucionan**

Los análisis de entropía pueden ayudar a caracterizar mejor un sitio activo definido mediante criterios estructurales y a centrar el análisis en los residuos que mayormente puedan contribuir a la diversidad funcional del receptor en estudio. Sin embargo, estos análisis no son capaces de identificar redes de interacciones de residuos correlacionados.

Patrones de mutaciones correlacionadas de secuencias de proteínas revelan sectores que co-evolucionan y que estarían ligados a propiedades funcionales y/o estructurales de las proteínas. Autores (Quadeer et al. 2020) han desarrollado un método de inferencia estadística con el fin de predecir sectores que co-evolucionan en estructuras proteicas (y controlando cualquier error estadístico por insuficiencia de datos). Este análisis fue puesto a prueba, infiriendo sectores proteicos de proteínas virales que concordaban con dominios bioquímicos ya conocidos. Este método se resume en un programa cuya interfaz nos permite realizar análisis robustos de co-evolución (RoCA, por sus siglas en inglés), denominado RocaSec (Quadeer et al. 2020).

Se requiere un alineamiento múltiple de secuencia de una proteína para inferir sectores que co-evolucionan. Adicionalmente, si proporcionamos información de

dominios bioquímicos de la proteína el programa nos entrega una asociación estadística entre los sectores inferidos y los dominios bioquímicos.

El estudio de los patrones de correlaciones mutacionales a partir de secuencias resulta útil ya que estos patrones han demostrado ser informativos en cuanto a sectores de residuos que co-evolucionan, que trabajan en conjunto para mediar de forma específica funciones de la proteína.

Para esta investigación ejecutamos el programa RocaSec proporcionando el alineamiento múltiple de secuencias del receptor cannabinoide 1, junto con la información de dominios funcionales para el receptor CB1R en vertebrados.

Para obtener la información de los residuos funcionales del receptor CB1R, se realizó una búsqueda exhaustiva en literatura donde se indiquen aminoácidos de la estructura primaria con su respectiva función. La información a obtenida se resumirá en una tabla indicando la función principal del residuo, su posición en la secuencia de aminoácidos, su ubicación espacial en la estructura terciaria y, finalmente, su respectiva cita bibliográfica.

### **7.7. Modelamiento del receptor CB1**

El modelo del receptor CB1 se obtuvo por homología mediante el uso de múltiples *templados* en el servidor RosettaCM (<https://www.rosettacommons.org>) que hibrida las porciones más homólogas de múltiples *templados* en un solo modelo mientras modela los residuos faltantes *de novo* (Song et al. 2013).

## **8. RESULTADOS**

### **8.1. Distribución filogenética y origen evolutivo de los componentes del sistema endocannabinoide**

La búsqueda sistemática de genomas secuenciados reveló la presencia de proteínas de los diferentes componentes del SEC, tanto para receptores como enzimas, en representantes de los diversos reinos de los seres vivos (Figuras 4



y 5). La distribución filogenética de las enzimas DAGL $\alpha$  y NAPE-PLD se caracteriza por poseer un origen evolutivo en el Reino Protista, específicamente en el subreino Chromista (*Ectocarpus siliculosus*). La enzima DAGL $\alpha$  responsable de la síntesis de 2-AG se ausenta en los grupos Plantae, Fungi, Ctenophora y Rotifera, mientras que la enzima NAPE-PLD que sintetiza el AEA se encuentra ausente en los representantes de los grupos Porífera, Orthonectida, Urochordata y Cephalochordata.

El origen evolutivo de la enzima MGLL que degrada 2-AG parece ser en el Dominio Bacteria, encontrándose ausente en los grupos Plantae, Fungi, y Nematoda, mientras que la enzima FAAH, responsable de la degradación de AEA, tiene un origen evolutivo en el reino Protista, encontrándose ausente en los grupos Plantae, Ctenophora, Xenacoelomorpha, Tardigrada, Echinoderma, Hemichordata, Urochordata y Cephalochordata.

La distribución filogenética de los receptores asociados al sistema endocannabinoide (CB1R, CB2R, GPR55, y TRPV1) se caracteriza por poseer un origen evolutivo en la división evolutiva Actinistia-Rhipidistia (celacanto-peces pulmonados+tetrápodos) (Figura 5), con algunas pérdidas en linajes específicos, tales como CB2R en osteíctios y GRP55 en reptiles.

## **8.2. Receptores acoplados a proteína G: CB1R, CB2R y receptor huérfano GPR55**

Mediante los análisis de búsqueda de dominios funcionales, se corroboró que los receptores CB1R, CB2R, y GPR55 pertenecen a la familia de receptores acoplados a proteínas G (RAPGs), y podemos observar que GPR55 se ubica muy divergentemente de los receptores CB1R y CB2R en el árbol filogenético obtenido mediante análisis de Máxima Verosimilitud (Figura 6).

Los receptores acoplados a proteína G y que pertenecen al SEC forman dos grupos claramente diferentes donde se puede observar la estrecha relación de los receptores CB1R y CB2R, en comparación a la proteína huérfana GRP55 (Figura 6). A partir de la figura 5, se puede observar que CB1R estaba presente

en el último ancestro común de los sarcopterigios (Sarcopterygii), ya que se encuentra presente en el celacanto *Latimeria chalumnae* (estrella roja en Figura 6).

La distribución de CB1R a lo largo de las diferentes especies de sarcopterigios sugiere que este receptor no ha sufrido considerables modificaciones, lo que puede llevar a concluir que la función de este receptor se mantendría conservada en las diferentes especies de sarcopterigios. Interesantemente, la ausencia del receptor CB2R en las especies de peces óseos y la pérdida del receptor GPR55 en las especies de reptiles analizadas en este estudio (Figura 5), sugiere que este receptor habría seguido una evolución convergente en estas especies de sarcopterigios.

### **8.3. Receptor cannabinoide 1 en vertebrados**

Una de las principales regiones conservadas que podemos apreciar en las distintas secuencias de CB1R en vertebrados corresponde a los dominios transmembrana. Con la ayuda de la estructura cristalizada de CB1R (PDB: 5TGZ) podemos ilustrar estas regiones en las distintas secuencias del grupo de los sarcopterigios (Figura 7). Por otra parte, las regiones variables de la secuencia corresponden principalmente a las regiones N-terminal (desde la posición 1 a la 146) y C-terminal (desde la posición 408 a la 467), según indica el alineamiento múltiple de secuencia y posterior análisis de entropía de Shannon (Figura 8).

El análisis de entropía de Shannon indica una mayor fracción de residuos conservados con valores de entropía (H) menores a 1 (Figura 9, bajo línea verde), mientras que para valores sobre 1 representa un menor nivel de conservación en contra de aquellos que se encuentran bajo la línea verde. Sin embargo, existe una posición en el alineamiento donde una mayor cantidad de aminoácidos se ven representados (Figura 9, sobre línea roja), que corresponde a la posición 410 del alineamiento (S414 en la secuencia de CB1 en humano). La posición se indica de manera simplificada en la figura 10, que consiste en un

alineamiento de la secuencia consenso del análisis de entropía, junto con secuencias de especies representativas de cada clado.

#### **8.4. Sectores proteicos en el receptor CB1**

Como se menciona en párrafos anteriores, las 7 regiones transmembrana parecieran ser la región más conservada del receptor CB1R. Una vez identificados los aminoácidos que componen estas regiones podemos definir un dominio de característica estructural para ingresar al programa RocaSec y así potencialmente obtener un sector que esté asociado a este complejo transmembranal.

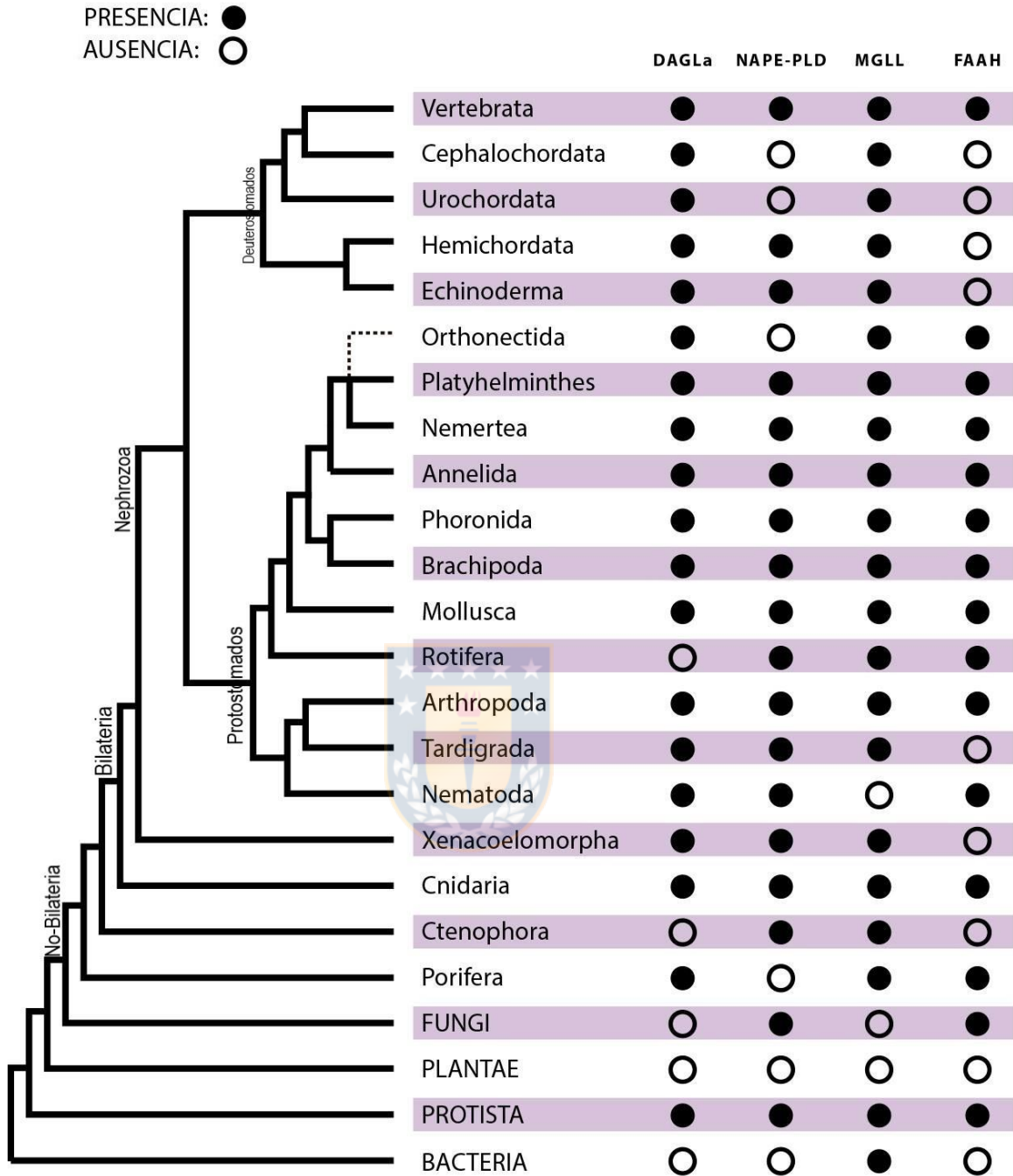
Luego, a partir de la información reunida en la tabla 3, se establecieron dominios bioquímicos que se pueden clasificar en 3 categorías. La primera categoría corresponde a la unión con antagonistas e inhibidores, que incluye a todo aquel residuo que esté interaccionando de manera directa la unión con inhibidor o antagonista, ya sea por interacciones hidrofóbicas, puente de hidrogeno o fuerzas de Van de Waals, o cuyas mutaciones pudieran provocar una alteración a la afinidad con ligandos antagonistas o inhibidores. La segunda categoría corresponde a la unión con sustancias agonistas y se aplica el mismo criterio que la categoría anterior, es decir, se incluye a todo residuo que esté interaccionando de manera directa o indirecta con ligando agonista, y cuyas mutaciones alteren la afinidad del receptor por ligando agonista. Por último, definimos un último dominio que corresponde a la activación del heterotrímico proteína G, que incluye a todos los aminoácidos que interaccionen directamente con las subunidades de la proteína G, así como también incluye aminoácidos cuyas mutaciones afectan directamente la activación de proteína G y actividad del receptor CB1R (Figura 11, región superior).

A partir de los dominios bioquímicos y alineamiento múltiple de secuencias proporcionados, el programa RocaSec arrojó como resultado 3 sectores definidos independientes entre sí (Figura 11, región inferior) que se pueden distinguir con los colores Rojo, Azul y Verde.

La independencia de los sectores mencionados no es absoluta ya que, si bien se distinguen uno del otro, al momento de realizar un análisis de superposición de sectores podemos observar semejanzas en el sector 3 y el sector 1, y viceversa (Figura 12). A partir de un análisis de correlación de los sectores inferidos con los dominios bioquímicos proporcionados (Figura 13), podemos deducir que el sector 1 (rojo) contiene a todos aquellos residuos que, de alguna manera, tienen un rol fundamental en la unión con ligandos, principalmente de los ligandos agonistas con un valor estadístico de  $3.864e-06$ . Por otra parte, se infiere que el sector 3 (verde) incluye a todos esos residuos que estarían jugando un rol clave en la activación del heterotrímero proteína G, con un valor estadístico de  $1.317e-06$ .

En esta oportunidad no existe un apoyo estadístico que indique correlación entre el sector 2 (azul) y los dominios bioquímicos proporcionados. Así como tampoco existe un valor de superposición significativo del sector 2 en contra de los otros sectores inferidos (1 y 3).

Para una mejor visualización de los sectores se obtuvo un modelo correspondiente a la secuencia del receptor CB1 en humano a través del servidor Rosetta, en el cual se ilustran los 3 sectores inferidos en esta investigación (Figura 14), y además se realizó una representación gráfica para indicar aquellos aminoácidos que pertenecen tanto al sector 1 como al sector 3, acompañado de una tabla indicando nombre, posición y ubicación espacial de aquellos residuos en la estructura terciaria. (Figura 14.D)

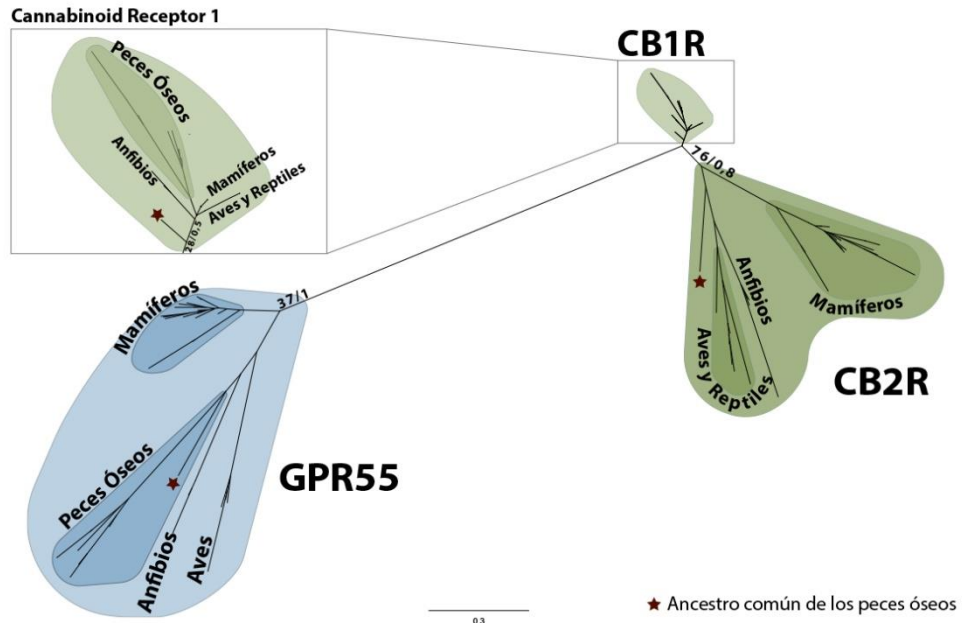


**Figura 4. Cladograma para enzimas del SEC.** Distribución filogenética de las enzimas encargadas de la síntesis y degradación de endocannabinoides, en los distintos reinos de la naturaleza.



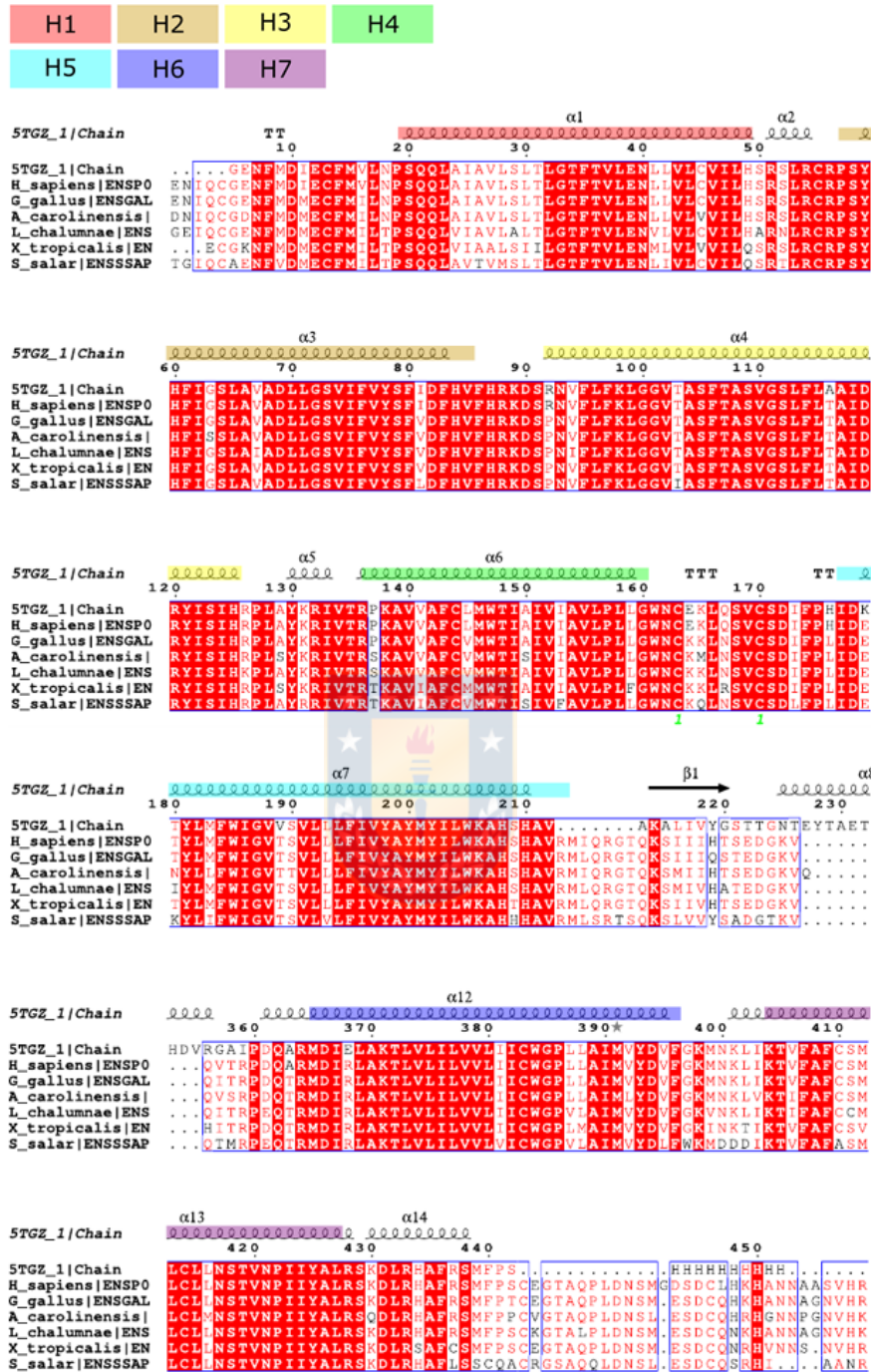
**Figura 5. Cladograma para receptores del SEC.** Distribución filogenética de los receptores asociados al sistema de endocannabinoides en vertebrados.





**Figura 6. Relaciones filogenéticas de los RAPG pertenecientes al SEC.** Análisis filogenético de los receptores acoplados a proteína G (RAPG) pertenecientes al SEC: CB1R, CB2R, y GRP55. Árbol consenso basado en la topología obtenida del análisis de Máxima Verosimilitud. En nodos particulares se muestran los valores de los porcentajes de bootstrap obtenidos con RAxML y los de probabilidad posterior del análisis de Inferencia Bayesiana realizado con MrBayes.

## Dominios Transmembrana

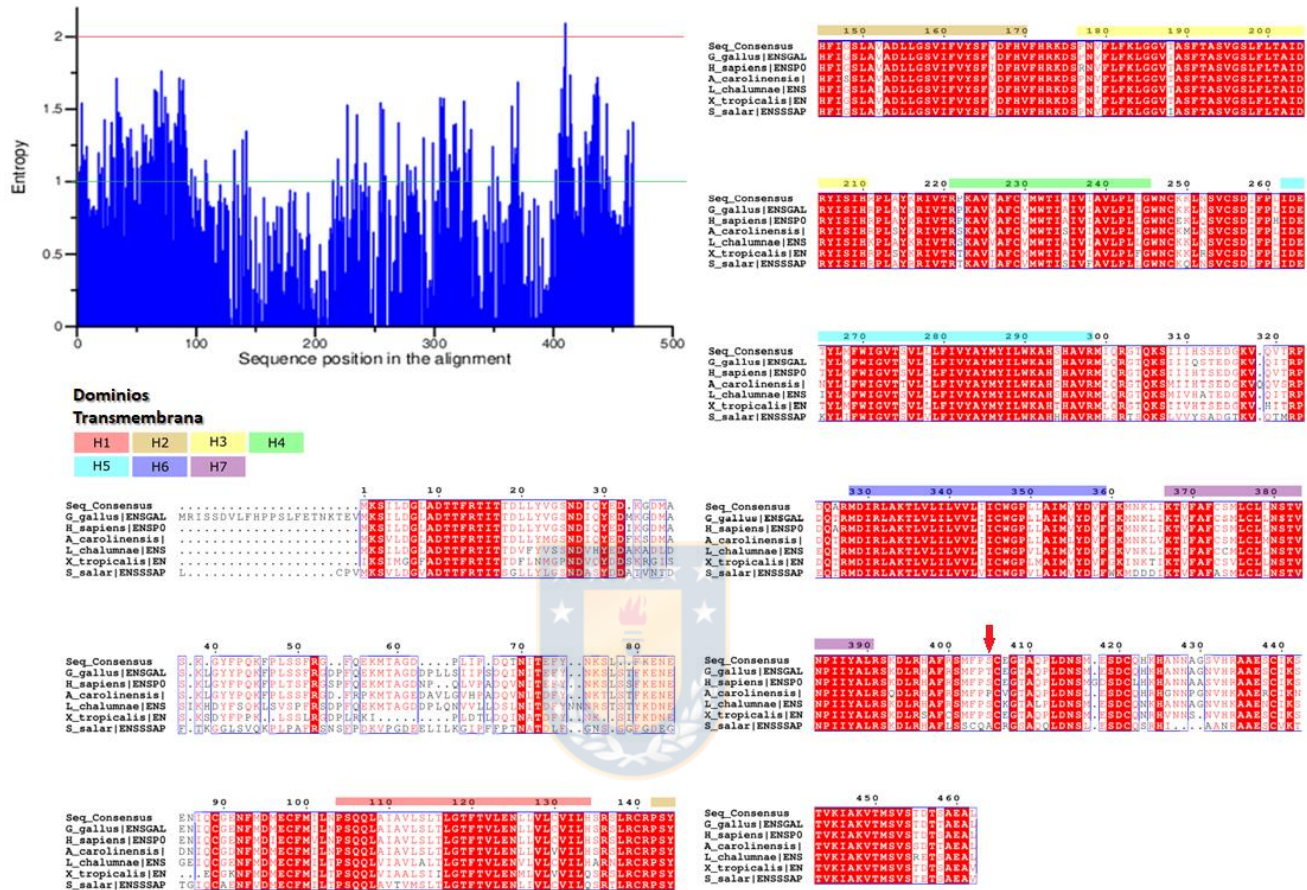


**Figura 7. Dominios transmembrana del receptor CB1.** Versión simplificada del alineamiento múltiple de secuencias de la estructura cristalizada del receptor CB1 (PDB: 5TGZ), junto con las secuencias del receptor CB1 de una (1) especie



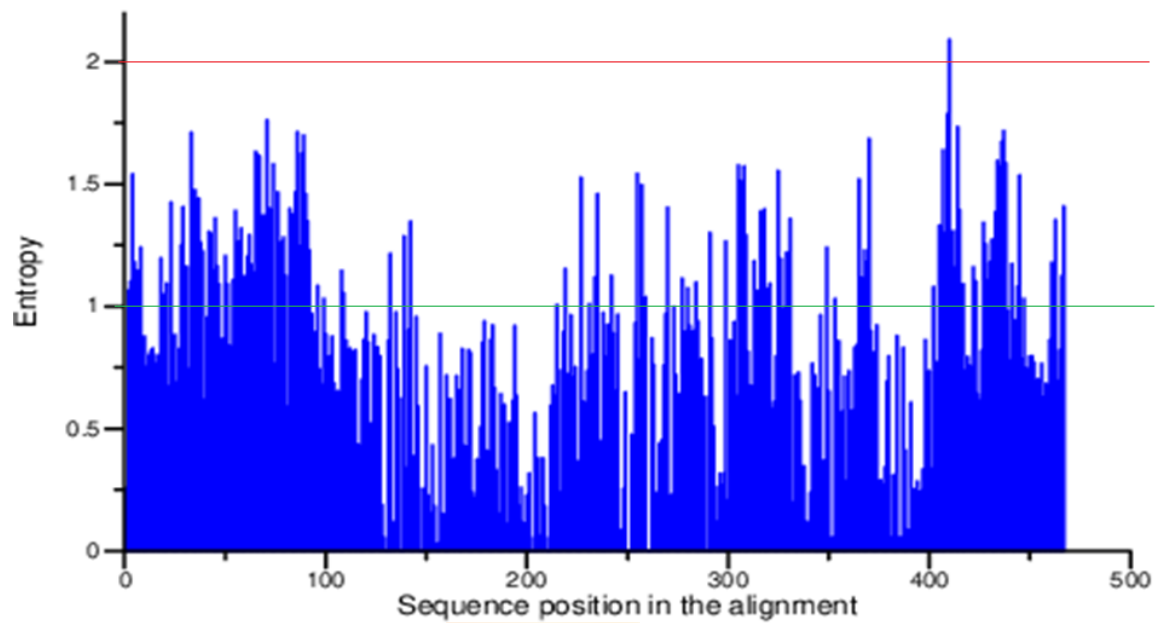
representativa de cada clado. Se indican dominios transmembrana para referencia.

## Análisis de Entropía de Shannon

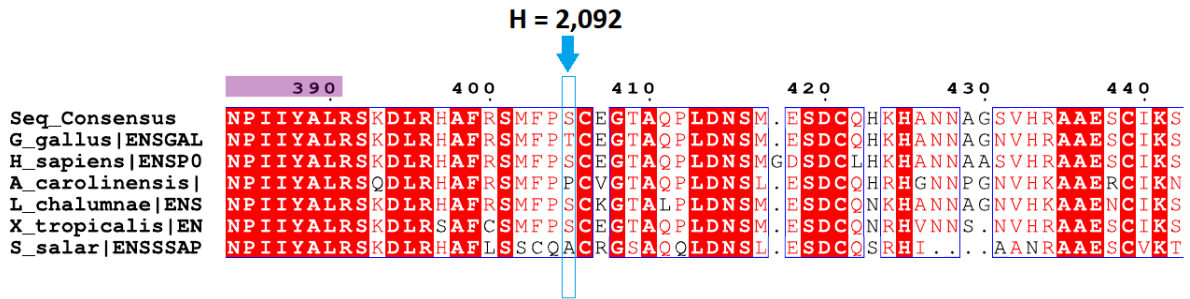


**Figura 8. Alineamiento múltiple de secuencias y análisis de entropía de Shannon para CB1.** Análisis de entropía de Shannon para el receptor CB1 en vertebrados y versión simplificada del alineamiento múltiple de secuencias utilizando la secuencia consenso del análisis de entropía junto con secuencias del receptor CB1 para especies representativas de cada clado.

# Análisis de Entropía de Shannon

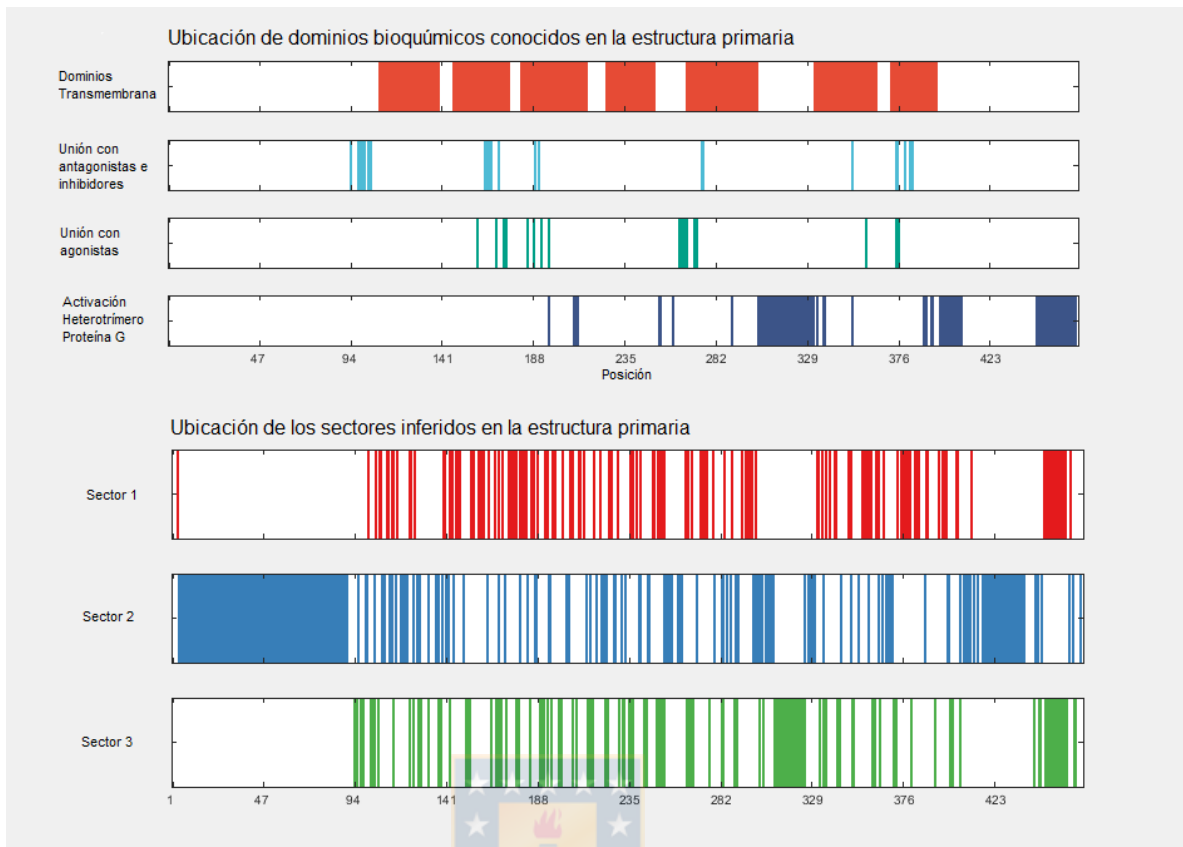


**Figura 9. Gráfico de valores de entropía de Shannon.** Análisis de entropía de Shannon para las secuencias del receptor CB1 en vertebrados.

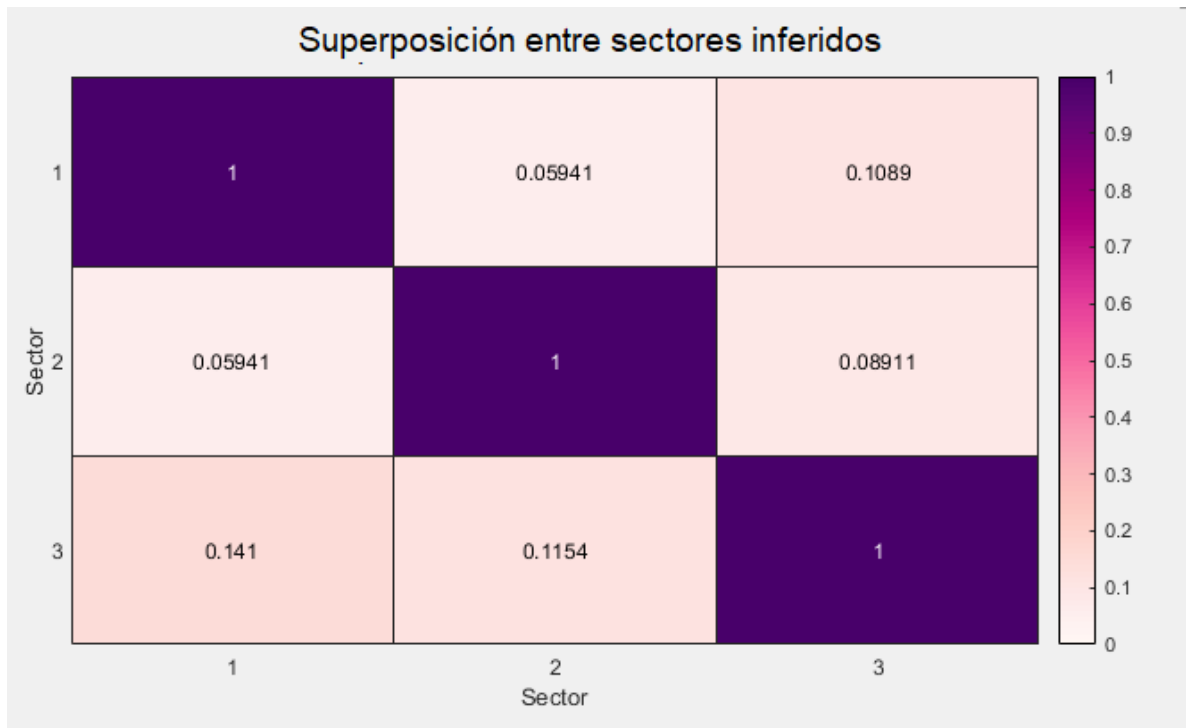


**Figura 10. Posición con alto valor de entropía.** Porción del alineamiento múltiple de secuencias correspondiente a la región proximal del extremo c-terminal de receptor CB1. Se indica la posición del aminoácido con valor de entropía  $H = 2.092$ . Versión simplificada del alineamiento múltiple de secuencias con la secuencia consenso del análisis de entropía con un (1) representante de cada clado

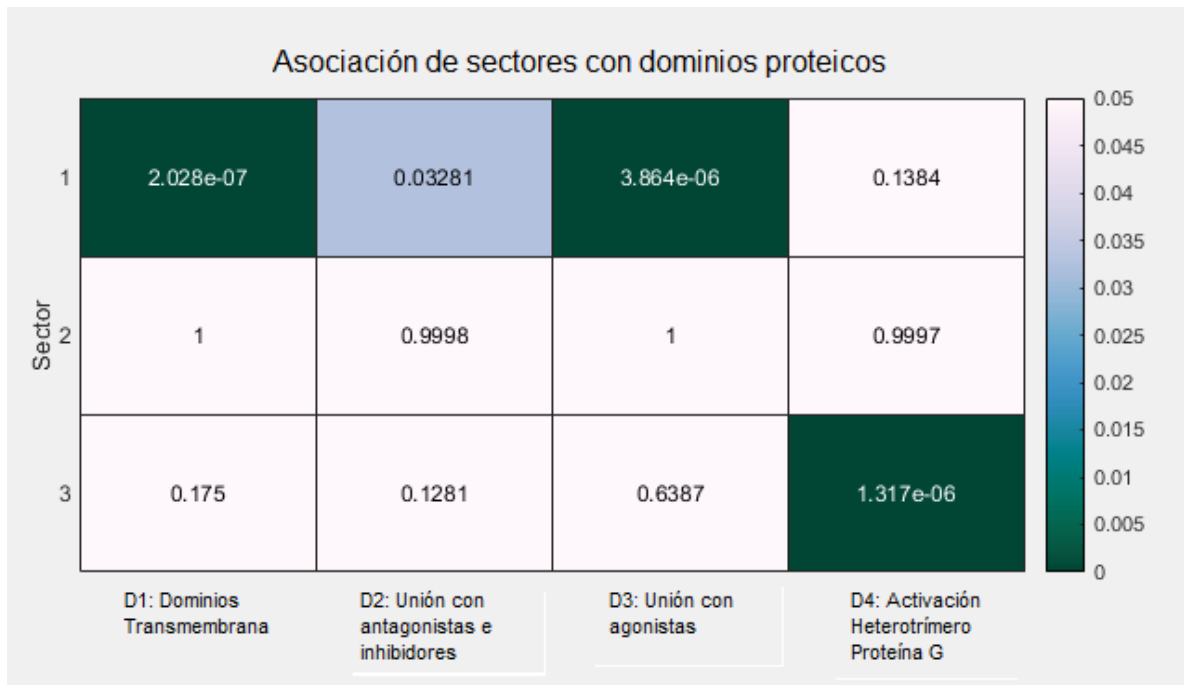




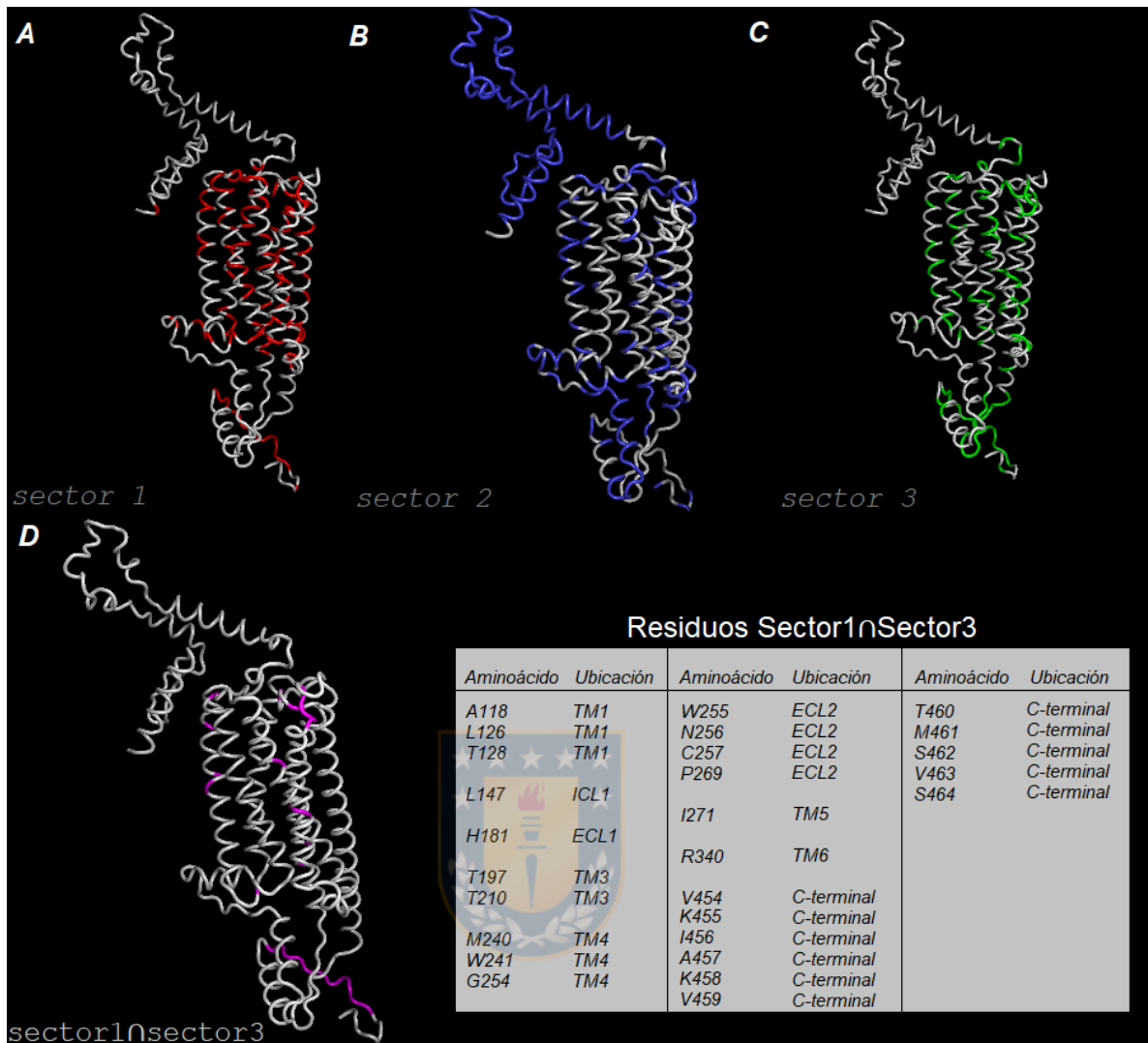
**Figura 11. Sectores inferidos con el programa RocaSec.** Interfaz del programa RocaSec indicando en la parte superior dominios bioquímicos conocidos ingresados para ejecutar el programa, e indicando en la parte inferior la ubicación de los sectores inferidos en la estructura primaria.



**Figura 12. Valores de superposición entre sectores.** Superposición entre sectores inferidos donde el valor 1 corresponde al mismo sector y el valor 0 corresponde a sectores completamente independientes uno del otro.



**Figura 13. Correlación entre sectores inferidos y dominios funcionales.** Asociación estadística de los sectores proteicos inferidos, y los dominios bioquímicos proporcionados para ejecutar RocaSec. Las casillas en color verde oscuro representan un valor estadísticamente significativo, mientras que las casillas de color más claro demuestran una menor significancia.



**Figura 14. Modelos tridimensionales del receptor CB1.** Modelo representativo del receptor CB1 obtenido con el servidor Rosetta. A, representa el sector 1 coloreado en rojo. B, representa el sector 2 coloreado en azul. C, representa el sector 3 coloreado en verde. D, representa a los residuos que se repiten tanto en el sector 1 como en el sector 3, indicando en una tabla los aminoácidos y su respectiva ubicación en la estructura tridimensional.


**Tabla 3**

Residuos del receptor CB1 con función conocida.

Residuo	Ubicación	Rol funcional	Referencias
W255	ECL2	Estabilizador estructural	McAllister et al. 2003 Wess et al. 1993
F200* C257 C264 Y294 D338 L341 A342 W356*	TM3 ECL2 ECL2 TM5 TM6 TM6 TM6 TM6	Activación/Señalización con proteína G. *Activación del receptor CB1 mediante distintas interacciones conformacionales entre F200/W356	Fay et al. 2005 Singh et al. 2002 Shire et al. 1996
L193 F268 P269 H270 I271 C355	TM6 ECL2 ECL2 ECL2 ECL2 TM6	Parte del sitio de unión de cannabinoides clásicos y no clásicos.	Bertalovitz et al. 2010 Ahn et al. 2009 Hurst et al. 2002
K192	TM3	Principal sitio de interacción con el grupo –OH de los cannabinoides, indirectamente modificando la estructura del sitio de unión.	Hurst et al. 2002 Chin et al. 1998
Toda la porción del loop intracelular 3	ICL3	Interacción con subunidades Gi1 y Gi2 de proteína G acoplada. Inhibición de AMPc, y activación de MAPK	Hart et al. 2004 Ulfers et al. 2002 Mukhopadhyay et al. 2001 Abadji et al. 1999



			Howlett et al. 1998 Bouaboula et al. 1995
S401 K402 D403 L404 R405 H406 A407 F408 R409 S410 M411 F412	TM7/H8 (C-terminal)	Interacción con subunidades Gi3 y Go de proteína G acoplada. Inhibición de los canales de calcio.	Pan et al. 1998 Hampson et al. 1998
Últimos 20 aminoácidos del extremo C-terminal	Región distal C-terminal	Interaccionan con arrestina para desensibilización de GIRK	Daigle et al. 2008 Jin et al. 1999
H181 R182 K183 D184 F189 S383	ECL1 ECL1 ECL1 ECL1 TM3 TM7	Mutaciones de Ala en residuos del ECL1 resultan en cambio estructural del sitio de unión para agonista CP55940. Mutación F189A disminuye afinidad para CP55940 y AEA. Mutación S383A disminuye afinidad para CP55940.	Kapur et al. 2007 McAllister et al. 2004 McAllister et al. 2003 Murphy et al. 2003

D163 F200 W279	TM2 TM3 TM5	Mutaciones D163N, F200A, W279A disminuyen afinidad para agonista WIN55212-2	Shen et al. 2006 McAllister et al. 2003
C98 C107	N-terminal N-terminal	Modula unión ortostérica del inhibidor Taranabant	Shao et al. 2016
F170 F174	TM2 TM2	Mutaciones F170A/F174A disminuyen afinidad con Rimonabant y AM6538	Hua et al. 2016
M103 G166 S167 F170 V196 W356 F379 S383 C386 L387	N-terminal TM2 TM2 TM2 TM3 TM6 TM7 TM7 TM7 TM7	Interacciones hidrofóbicas con antagonista AM6538  	Hua et al. 2016
D213 R214 Y215 N393 P394 Y297	TM3 TM3 TM3 TM7 TM7 TM7	Residuos críticos para la activación de RAPG pertenecientes a la familia de la rodopsina	Fritze et al. 2003

## 9. DISCUSION

Los resultados presentados en este trabajo en relación al origen evolutivo de los componentes del SEC presentan algunas diferencias a los publicados por McPartland y colaboradores en el 2006.

**Tabla 4.** Comparación del origen evolutivo de los componentes del sistema de endocannabinoides.

Proteína	McPartland <i>et al.</i> 2006	ESTA INVESTIGACION
CB1R	BILATERADOS	VERTEBRADOS
CB2R	VERTEBRADOS	VERTEBRADOS
GRP55	MAMIFEROS	VERTEBRADOS
TRPV1	MAMIFEROS	VERTEBRADOS
DAGLa	BILATERADOS	PROTISTA
NAPE-PLD	OPISTOKONTA	PROTISTA
MGLL	CORDADOS	BACTERIAS
FAAH	PROTISTA	PROTISTA

Sin embargo, debido al amplio espectro de especies utilizadas en esta investigación, fue posible determinar con una mayor veracidad el origen evolutivo de los componentes del SEC. Los orígenes evolutivos de cada proteína abarcan desde el grupo de las bacterias hasta los cordados superiores. Esta diversidad, junto con la amplia variedad de funciones que cumplen algunos de sus componentes, ha dificultado determinar la evolución temprana de estos, específicamente de las enzimas encargadas de la síntesis y degradación de 2-AG y AEA. Para poder explicar la ausencia de enzimas en plantas, Shrestha y colaboradores en el 2003 sugirieron que excluir la enzima FAAH del grupo de las “FAAH-like” amidasas sería muy estricto ya que se ha demostrado que estas enzimas “FAAH-like” en plantas (*Arabidopsis thaliana*) metabolizan AEA de

manera similar a la enzima FAAH. Lo anterior, podría explicar que otras enzimas con capacidad amidasa podrían estar cumpliendo funciones asociadas al SEC. En el caso de los receptores asociados al SEC, los resultados indican claramente que representan innovaciones más recientes ya que se encuentran exclusivamente en cordados superiores (sarcopterigios).

### **9.1. RAPGs del sistema endocannabinoide: CB1R, CB2R Y GPR55**

Los receptores acoplados a proteínas G y que participan en el SEC están presentes exclusivamente en las especies de sarcopterigios utilizadas en esta investigación. Pese a encontrar el receptor CB2R en el ancestro común de los sarcopterigios (Figura 5), se observa la pérdida del este receptor en peces óseos, lo cual podría estar explicado por el estilo de vida estrictamente acuático de estas especies y que probablemente no necesiten tener este receptor ya que el CB1R es capaz de cumplir su función. Otra posible explicación a la pérdida de CB2R en peces óseos podría ser la acumulación de mutaciones en las que pudieron verse comprometidos los aminoácidos que componen el sitio de unión o binding pocket del receptor, siendo una consecuencia la posible pérdida de su función biológica y posterior desaparición de los genomas de las especies de peces óseos. Si analizamos la filogenia de estos receptores en la figura 6, CB2R pareciera presentar una conservación más notoria, respecto a aquella de CB1R. CB2R pudiera mantener más íntegra su estructura debido a que su región N-terminal resulta ser mucho más corta que la región N-terminal de CB1R, disminuyendo el número de aminoácidos de su secuencia, por lo que cualquier variación significativa de la secuencia de CB2R sería evidente. CB1R es uno de los receptores acoplados a proteína G con la región N-terminal más larga dentro de las RAPGs (Jakowiecki et al. 2021) y las principales variaciones para el receptor CB1R ocurren a nivel de los extremos (región N- y C-terminal) (Anexo 1.), lo que pudiera ser causa de la divergencia que se puede apreciar en clados del receptor CB1R (Figura 6).

En reptiles, para el receptor huérfano GPR55, el escenario podría ser similar a lo ocurrido con el receptor CB2R en peces óseos. El receptor GPR55 a pesar de ser activado por sustancias cannabinoides, no pareciera ser imprescindible para el sistema endocannabioide (Kapur et al. 2009). En mamíferos, más específicamente en murino y humano, GPR55 se expresa altamente en el bazo y los leucocitos, y tiene un papel fundamental en la modulación de las respuestas inmunes innatas y adaptativas (Staton et al. 2008; Lin et al. 2011; Schicho & Storr 2012; Stančić et al. 2015). Al observar la Figura 6 podemos observar a través del largo de las ramas que la proteína GPR55 es más conservada en mamíferos, mientras que en aves, anfibios, osteíctios y celacanto han sufrido un mayor número de mutaciones y por tanto ha evolucionado de forma diferente a lo largo del tiempo para estos clados. Con esto podríamos especular que GPR55 en aves, anfibios, osteíctios y el celacanto poseen una función diferente a la que se conoce en mamíferos; sin embargo, lo anterior necesita ser evaluado en futuras investigaciones para ver si es una realidad.

## **9.2. Análisis estructural e inferencia de sectores para el receptor CB1**

Los sectores proteicos corresponden a residuos que no se encuentran en contacto en la estructura primaria, pero que estarían interaccionando en la estructura tridimensional. Si bien en un principio la intención era encontrar aminoácidos críticos para la funcionalidad del receptor CB1, de esta investigación se desprenden sectores proteicos asociados a una función específica dentro de la proteína, y se presume que estos aminoácidos han evolucionado de manera correlacionada en distintas especies de vertebrados.

Si consideramos la nomenclatura de residuos ocupada por el autor Shim el 2010 (Figura 3), los sectores inferidos estarían cubriendo 2 de las 3 clasificaciones propuestas por este autor correspondientes a la unión con ligando (**U**) y activación del receptor (**A**), descartando a la tercera clasificación que incluye a los aminoácidos encargados de la estabilización del receptor (**E**). El sector 2 pudiera pertenecer a esta última clasificación, pero para corroborar lo anterior se necesita un análisis más acabado en cuanto a análisis de residuos. Esta

incógnita se podría resolver con una metodología dirigida hacia la simulación molecular u otra estrategia experimental. Al mismo tiempo, llama la atención que el sector 2 cubra una vasta región de la porción N-terminal del receptor CB1. A partir de este fenómeno podríamos especular que el sector 2 representa una función exclusiva del receptor CB1, ya que esta proteína posee la región N-terminal más larga dentro de los receptores asociados a proteína G pertenecientes a la familia de la rodopsina (Jakowiecki et al. 2021).

Por otra parte, la porción C-terminal del receptor CB1 presenta aminoácidos que, según nuestra investigación, encontramos en los sectores proteicos inferidos 1 y 3, asignándole a esta región una funcionalidad asociada a la afinidad con ligando y a la activación de proteína G. Lo anterior coincide con lo indicado en literatura, ya que la porción C-terminal de los receptores asociados a proteína G se encuentra hacia la región intracelular, interaccionando directamente con el heterotrímero proteína G (Rosenbaum et al. 2009) y otros componentes intracelulares (Jin et al. 1999, Daigle et al. 2008).

Estos aminoácidos de la región C-terminal, no resultan ser los únicos pertenecientes a los sectores 1 y 3, de hecho, en esta investigación realizamos un modelo indicando el solapamiento de los sectores 1 y 3 y los aminoácidos que están presentes en ambos sectores (Figura 14, D). Podríamos proponer a este “sector de solapamiento” como responsable de la unión a ligando y activación de proteína G, rescatando lo propuesto por Shim el 2010 que indica la existencia de residuos que comparten funcionalidad. En este sector de solapamiento se encuentran residuos responsables para la activación del receptor CB1 como C257, que en conjunto a C264 forman un puente disulfuro de cuya estabilidad depende el correcto funcionamiento del receptor CB1 (Fay et al. 2005). Otros aminoácidos presentes en este sector de solapamiento corresponden a la unión con ligandos, principalmente ligandos agonistas, como es el caso de H181 (Murphy & Kendall 2003), y P269 y I271, que en conjunto a F268 y H270 resultan ser la porción del loop extracelular 2 (ECL2) responsable

para la unión de todos los ligandos agonistas para el receptor CB1 (Ahn et al. 2009, Bertalovitz et al. 2010).

Resulta curiosa la presencia del residuo W255 en el sector de solapamiento, ya que este aminoácido cumple un rol de estabilizador estructural en el receptor CB1 (Wess et al. 1993, McAllister et al. 2003). Nuevamente podríamos recurrir al criterio de Shim el 2010 y proponer a este residuo como un aminoácido multifuncional responsable de la estabilización estructural, en conjunto con la activación de proteína G o con la unión de ligando. Incluso podríamos definir a este aminoácido como un residuo que cumple las 3 funciones bajo el criterio de Shim (Figura 3), pero lo anterior debe estar respaldado por análisis y experimentación funcional dirigida al residuo W255 en el receptor CB1.

## 10. CONCLUSION

De los análisis realizados en este estudio, se desprende que las proteínas del sistema endocannabinoide han evolucionado de manera dinámica y caracterizada por la pérdida de algunos componentes en ciertos linajes o grupos taxonómicos. Sin embargo, la distribución de las enzimas indica que éstas no están limitadas a un grupo específico de organismos, a diferencia de los receptores. Variaciones en la forma en la que se distribuyen las proteínas del SEC a lo largo del tiempo se podría atribuir al estilo de vida e interacciones que los organismos tienen con el medio que los rodea.

Podemos concluir que existen aminoácidos que evolucionan de manera correlacionada y en conjunto están encargados de una función específica, responsables del correcto funcionamiento del receptor CB1, y resulta ser un patrón que se mantiene en las distintas especies de vertebrados analizadas en este estudio.

Conocer en profundidad la relación estructura función del receptor cannabinoide 1 resulta útil en el estudio de variantes de genes que pudieran generar una estructura diferente. Un alcance biomédico considerando lo anterior favorecería al diseño y elaboración de fármacos.

## 11. BIBLIOGRAFÍA

- ABADJI, V., LUCAS-LENARD, J.M., CHIN, C., & KENDALL, D.A. 1999. Involvement of the carboxyl terminus of the third intracellular loop of the cannabinoid CB1 receptor in constitutive activation of Gs. *Journal of Neurochemistry* 72:2032–2038.
- AHN K.H., BERTALOVITZ A.C., MIERKE D.F., & KENDALL D.A. 2009. Dual Role of the Second Extracellular Loop of the Cannabinoid Receptor One: Ligand Binding and Receptor Localization. *Molecular Pharmacology* 76:833–842.
- ANFINSEN, C. B. 1973. Principles that govern the folding of protein chains. *Science* 181(4096), 223–230.
- ASIMAKI, O. & MANGOURA, D. 2011. Cannabinoid receptor 1 induces a biphasic ERK activation via multiprotein signaling complex formation of proximal kinases PKCepsilon, Src, and Fyn in primary neurons. *Neurochemistry International* 58: 135-144.
- BEGG, M., PACHER, P., BÁTKAI, S., OSEI-HYIAMAN, D., OFFERTÁLER, L., MO, F. M., LIU, J., & KUNOS, G. 2005. Evidence for novel cannabinoid receptors. *Pharmacology & therapeutics* 106(2), 133–145.
- BERTALOVITZ, A. C., AHN, K. H., & KENDALL, D. A. 2010. Ligand Binding Sensitivity of the Extracellular Loop Two of the Cannabinoid Receptor 1. *Drug development research* 71(7), 404–411.
- BLANKMAN, J. L., SIMON, G. M., & CRAVATT, B. F. 2007. A comprehensive profile of brain enzymes that hydrolyze the endocannabinoid 2-arachidonoylglycerol. *Chemistry & biology* 14(12), 1347–1356.
- BOUABOULA, M., POINOT-CHAZEL, C., BOURRIE, B., CANAT, X., CALANDRA, B., RINALDI-CARMONA, M., LE FUR, G., & CASELLAS, P. 1995. Activation of mitogen-activated protein kinases by stimulation



of the central cannabinoid receptor CB1. *Biochemical journal* 312(Pt 2): 637-641.

BOUABOULA M., POINOT-CHAZEL, C., MARCHAND, J., CANAT, X., BOURRIÉ, B., RINALDI-CARMONA, M., CALANDRA, B., LE FUR, G., & CASELLAS, P. 1996. Signaling pathway associated with stimulation of CB2 peripheral cannabinoid receptor. Involvement of both mitogenactivated protein kinase and induction of Krox-24 expression. *European Journal of Biochemistry* 237: 704–711.

CAMACHO, C., COULOURIS, G., AVAGYAN, V., MA, N., PAPADOPOULOS, J., BEALER, K., & MADDEN, T. L. 2009. BLAST+: architecture and applications. *BMC bioinformatics*, 10, 421.

CAPELLA-GUTIÉRREZ, S., SILLA-MARTÍNEZ, J. M., & GABALDÓN, T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics (Oxford, England)*, 25(15), 1972–1973.

CASTILLO, P. E., YOUNTS, T. J., CHÁVEZ, A. E., & HASHIMOTODANI, Y. 2012. Endocannabinoid signaling and synaptic function. *Neuron* 76(1), 70–81.

CHIN C.N., LUCAS-LENARD J., ABADJI V. & KENDALL D.A. 1998. Ligand binding and modulation of cyclic AMP levels depend on the chemical nature of residue 192 of the human cannabinoid receptor 1. *Journal of Neurochemistry* 70:366–373.

COVER, T. M., & THOMAS, J. A. 1991. Entropy, relative entropy and mutual information. *Elements of Information Theory*, 2, 1–55.

CRAVATT, B. F., GIANG, D. K., MAYFIELD, S. P., BOGER, D. L., LERNER, R. A., & GILULA, N. B. 1996. Molecular characterization of an enzyme that degrades neuromodulatory fatty-acid amides. *Nature* 384(6604), 83–87.

- DAIGLE T.L., KEARN C.S. & MACKIE K. Rapid CB1 cannabinoid receptor desensitization defines the time course of ERK1/2 MAP kinase signaling. *Neuropharmacology*. 2008; 54:36–44.
- DARRIBA, D., POSADA, D., KOZLOV, A., STAMATAKIS, A., MOREL, B., & FLOURI, T. (2019) ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Biorxiv*
- DAVIS M. I., RONESI, J., & LOVINGER, D. M. 2003. A predominant role for inhibition of the adenylate cyclase/protein kinase A pathway in ERK activation by cannabinoid receptor 1 in N1E-115 neuroblastoma cells. *Journal of Biological Chemistry* 278: 48973-48980.
- DI MARZO, V., FONTANA, A., CADAS, H., SCHINELLI, S., CIMINO, G., SCHWARTZ, J. C., & PIOMELLI D. 1994. Formation and inactivation of endogenous cannabinoid anandamide in central neurons. *Nature* 372(6507), 686–691.
- ELDEEB, K., LEONE-KABLER, S., & HOWLETT A. C. 2016. CB1 cannabinoid receptor-mediated increases in cyclic AMP accumulation are correlated with reduced Gi/o function. *Journal of basic and clinical physiology and pharmacology* 27(3), 311–322.
- FAY J.F., DUNHAM, T.D., & FARRENS, D.L. 2005. Cysteine residues in the human cannabinoid receptor: only C257 and C264 are required for a functional receptor, and steric bulk at C386 impairs antagonist SR141716A binding. *Biochemistry* 44:8757–8769.
- FREDRIKSSON, R., LAGERSTROM, M. C., LUNDIN, L. G. & SCHIOTH, H. B. 2003. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Molecular Pharmacology* 63, 1256–1272.
- FRITZE, O., FILIPEK, S., KUKSA, V., PALCZEWSKI, K., HOFMANN, K.P., & ERNST, O.P. 2003. Role of the conserved NPxxY (x)5,6F motif in the rhodopsin ground state and during activation. *Proceedings of the*

National Academy of Sciences of the United States of America  
100:2290–2295.

GALVE-ROPERH I., RUEDA, D., GOMEZ DEL PULGAR, T., VELASCO, G.,  
& GUZMAN, M. 2002. Mechanism of extracellular signal-regulated  
kinase activation by the CB(1) cannabinoid receptor. *Molecular  
Pharmacology* 62: 1385-1392.

GONSIORREK, W., LUNN, C., FAN, X., NARULA, S., LUNDELL, D., & HIPKIN,  
R.W. 2000. Endocannabinoid 2-Arachidonyl Glycerol Is a Full Agonist  
through Human Type 2 Cannabinoid Receptor: Antagonism by  
Anandamide. *Molecular Pharmacology* 57:1045–1050

GUO, J., & IKEDA, S.R. 2004. Endocannabinoids modulate N-type calcium  
channels and G-protein-coupled inwardly rectifying potassium channels  
via CB1 cannabinoid receptors heterologously expressed in mammalian  
neurons. *Molecular pharmacology* 65(3), 665–674.

HALABI, N., RIVOIRE, O., LEIBLER, S., & RANGANATHAN, R. 2009. Protein  
sectors: Evolutionary units of three-dimensional structure. *Cell* 138(4),  
774–786.

HAMPSON, A.J., BORNHEIM, L.M., SCANZIANI, M., YOST, C.S., GRAY,  
A.T., HANSEN, B.M., LEONOUDAKIS, D.J., & BICKLER, P.E. 1998.  
Dual effects of anandamide on NMDA receptor-mediated responses  
and neurotransmission. *Journal of Neurochemistry* 70:671–676.

HART, S., FISCHER, O.M., & ULLRICH, A. 2004. Cannabinoids induce cancer  
cell proliferation via tumor necrosis factor alpha-converting enzyme  
(TACE/ADAM17)-mediated transactivation of the epidermal growth  
factor receptor. *Cancer Research* 64:1943–1950.

HOWLETT, A.C., SONG, C., BERGLUND, B.A., WILKEN, G.H., & PIGG, J.J.  
1998. Characterization of CB1 cannabinoid receptors using receptor  
peptide fragments and site-directed antibodies. *Molecular  
Pharmacology* 53:504–510.

- HOWLETT, A. C., BARTH, F., BONNER, T. I., CABRAL, G., CASELLAS, P., DEVANE W. A., FELDER C. C., HERKENHAM, M., MACKIE, K., MARTIN, B. R., MECHOULAM, R., & PERTWEE, R. G. 2002. International Union of Pharmacology. XXVII. Classification of cannabinoid receptors. *Pharmacological Reviews* 54: 161-202.
- HUA, T., VEMURI, K., PU, M., QU, L., HAN, G. W., WU, Y., ZHAO, S., SHUI, W., LI, S., KORDE, A., LAPRAIRIE, R. B., STAHL, E. L., HO, J. H., ZVONOK, N., ZHOU, H., KUFAREVA, I., WU, B., ZHAO, Q., HANSON, M. A., BOHN, L. M., & LIU, Z. J. 2016. Crystal Structure of the Human Cannabinoid Receptor CB1. *Cell* 167(3), 750–762.e14.
- HUNTER, S., APWEILER, R., ATTWOOD, T. K., BAIROCH, A., BATEMAN, A., BINNS, D., BORK, P., DAS, U., DAUGHERTY, L., DUQUENNE, L., FINN, R. D., GOUGH, J., HAFT, D., HULO, N., KAHN, D., KELLY, E., LAUGRAUD, A., LETUNIC, I., LONSDALE, D., LOPEZ, R., & YEATS, C. 2009. InterPro: the integrative protein signature database. *Nucleic acids research* 37(Database issue), D211–D215.
- HURST, D.P., LYNCH, D.L., BARNETT-NORRIS, J., HYATT, S.M., SELTZMAN, H.H., ZHONG, M., SONG, Z.H., NIE, J., LEWIS, D., & REGGIO, P.H. 2002. N-(piperidin-1-yl)-5-(4-chlorophenyl)-1-(2,4-dichlorophenyl)-4-methyl-1H-pyrazole-3-carboxamide (SR141716A) interaction with LYS 3.28(192) is crucial for its inverse agonism at the cannabinoid CB1 receptor. *Molecular Pharmacology* 62:1274–1287.
- JÄGER, D., SCHMALENBACH, C., PRILLA, S., SCHROBANG, J., KEBIG, A., SENNWITZ, M., HELLER, E., TRÄNKLE, C., HOLZGRABE, U., & HÖLTJE, H-D. 2007. Allosteric small molecules unveil a role of an extracellular E2/transmembrane helix 7 junction for G protein-coupled receptor activation. *The Journal of biological chemistry* 282(48), 34968–34976.

- JAKOWIECKI, J., ABEL, R., ORZEŁ, U., PASZNIK, P., PREISSNER, R., & FILIPEK, S. 2021. Allosteric Modulation of the CB1 Cannabinoid Receptor by Cannabidiol-A Molecular Modeling Study of the N-Terminal Domain and the Allosteric-Orthosteric Coupling. *Molecules* (Basel, Switzerland), 26(9), 2456.
- JIN, W., BROWN, S., ROCHE, J.P., HSIEH, C., CELVER, J.P., KOVOOR, A., CHAVKIN, C., & MACKIE, K. 1999. Distinct domains of the CB1 cannabinoid receptor mediate desensitization and internalization. *Journal of Neuroscience* 19:3773–3780.
- JONES, P., BINNS, D., CHANG, H. Y., FRASER, M., LI, W., MCANULLA, C., MCWILLIAM, H., MASLEN, J., MITCHELL, A., NUKA, G., PESSEAT, S., QUINN, A. F., SANGRADOR-VEGAS, A., SCHEREMETJEW, M., YONG, S. Y., LOPEZ, R., & HUNTER, S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* (Oxford, England), 30(9), 1236–1240.
- KANO, M., OHNO-SHOSAKU, T., HASHIMOTODANI, Y., UCHIGASHIMA, M., & WATANABE, M. 2009. Endocannabinoid-mediated control of synaptic transmission. *Physiological reviews* 89(1), 309–380.
- KAPUR, A., HURST, D.P., FLEISCHER, D., WHITNELL, R., THAKUR, G.A., MAKRIYANNIS, A., REGGIO, P.H., & ABOOD, M.E. 2007. Mutation studies of Ser7.39 and Ser2.60 in the human CB1 cannabinoid receptor: evidence for a serine-induced bend in CB1 transmembrane helix 7. *Molecular Pharmacology* 71:1512–1524.
- KAPUR, A., ZHAO, P., SHARIR, H., BAI, Y., CARON, M. G., BARAK, L. S., & ABOOD, M. E. 2009. Atypical responsiveness of the orphan receptor GPR55 to cannabinoid ligands. *The Journal of biological chemistry*, 284(43), 29817–29827.

- KATOH, K., ROZEWICKI, J., & YAMADA, K. D. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in bioinformatics*, 20(4), 1160–1166.
- KOCH, M., VARELA, L., KIM, J. G., KIM, J. D., HERNÁNDEZ-NUÑO, F., SIMONDS, S. E., CASTORENA, C. M., VIANNA, C. R., ELMQUIST, J. K., MOROZOV, Y. M., RAKIC, P., BECHMANN, I., COWLEY, M. A., SZIGETI-BUCK, K., DIETRICH, M. O., GAO, X. B., DIANO, S., & HORVATH, T. L. 2015. Hypothalamic POMC neurons promote cannabinoid-induced feeding. *Nature* 519(7541), 45–50.
- LAPRAIRIE, R. B., STAHL, E. L., & BOHN, L. M. 2017. Approaches to Assess Biased Signaling at the CB1R Receptor. *Methods in enzymology* 593, 259–279.
- LIN, X. H., YUECE, B., LI, Y. Y., FENG, Y. J., FENG, J. Y., YU, L. Y., LI, K., LI, Y. N., & STORR, M. 2011. A novel CB receptor GPR55 and its ligands are involved in regulation of gut movement in rodents. *Neurogastroenterology and motility : the official journal of the European Gastrointestinal Motility Society*, 23(9), 862–e342.
- LITWIN, S., & JORES, R. 1992. Shannon information as a measure of amino acid diversity. In *Theoretical and experimental insights into immunology* (pp. 279–287). Springer.
- LUK, T., JIN, W., ZVONOK, A., LU, D., LIN, X.Z., & CHAVKIN, C. 2004. Identification of a potent and highly efficacious, yet slowly desensitizing CB1 cannabinoid receptor agonist. *British Journal of Pharmacology* 142:495–500.
- MATSUDA, L.A., LOLAIT, S.J., BROWNSTEIN, M.J., YOUNG, A.C., & BONNER, T.I. 1990. Structure of a cannabinoid receptor and functional expression of the cloned cDNA. *Nature* 346, 561–564.
- MCALLISTER, S.D., HURST, D.P., BARNETT-NORRIS, J., LYNCH, D., REGGIO, P.H., & ABOOD, M.E. 2004. Structural mimicry in class A

GPCR rotamer toggle switches: The importance of the F3.36(201)/W6.48(357) interaction in cannabinoid CB1 receptor activation. *Journal of Biological Chemistry* 279:48024–48037.

MCALLISTER, S.D., RIZVI, G., ANAVI-GOFFER, S., HURST, D.P., BARNETT-NORRIS, J., LYNCH, D.L., REGGIO, P.H., & ABOOD, M.E. 2003. An aromatic microdomain at the cannabinoid CB(1) receptor constitutes an agonist/inverse agonist binding region. *Journal of Medicinal Chemistry* 46:5139–5152.

MCPARTLAND, J. & PRUITT, P. 2002. Sourcing the Code: Searching for the Evolutionary Origins of Cannabinoid Receptors, Vanilloid Receptors, and Anandamide. *Journal of Cannabis Therapeutics*, 2(1), 74 – 76.

MCPARTLAND, J. M., MATIAS, I., DI MARZO, V., & GLASS, M. 2006. Evolutionary origins of the endocannabinoid system. *Gene*, 370, 64–74.

MECHOULAM, R. 1986. *The Pharmacohistory of Cannabis sativa*, in *Cannabis as Therapeutic Agent*; CRC Press: Boca Raton, FL, USA,

MUKHERJEE, S., ADAMS, M., WHITEAKER, K., DAZA, A., KAGE, K., CASSAR, S., MEYER, M., & YAO, B. B. 2004. Species comparison and pharmacological characterization of rat and human CB2 cannabinoid receptors. *European journal of pharmacology*, 505(1-3), 1–9.

MUKHOPADHYAY, S. & HOWLETT, A.C. 2001. CB1 receptor-G protein association. Subtype selectivity is determined by distinct intracellular domains. *European Journal of Biochemistry* 268:499–505.

MUNRO, S., THOMAS, K.L., & ABU-SHAAR, M. 1993. Molecular characterization of a peripheral receptor for cannabinoids. *Nature* 365, 61–65.

MURATAEVA, N., STRAIKER, A., & MACKIE, K. 2014. 2-AG synthesis and degradation in the CNS. *British Journal of Pharmacology* 171: 1379-1391.

- MURPHY, J.W., & KENDALL, D.A. 2003. Integrity of extracellular loop 1 of the human cannabinoid receptor 1 is critical for high-affinity binding of the ligand CP 55,940 but not SR 141716A. *Biochemical Pharmacology* 65:1623–1631.
- NAVARRO, M., CARRERA, M. R., FRATTA, W., VALVERDE, O., COSSU, G., FATTORE, L., CHOWEN, J. A., GOMEZ, R., DEL ARCO, I., VILLANUA, M. A., MALDONADO, R., KOOB, G. F., & RODRIGUEZ DE FONSECA, F. 2001. Functional interaction between opioid and cannabinoid receptors in drug self-administration. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 21(14), 5344–5350.
- PACHER, P., BATKAI, S., & KUNOS, G. 2006. The endocannabinoid system as an emerging target of pharmacotherapy. *Pharmacological Reviews* 58, 389–462.
- PAN, X., IKEDA, S.R., & LEWIS, D.L. 1998. SR 141716A acts as an inverse agonist to increase neuronal voltage-dependent Ca<sup>2+</sup> currents by reversal of tonic CB1 cannabinoid receptor activity. *Molecular Pharmacology* 54:1064–1072.
- PERTWEE, R. G. 2001. Cannabinoid receptors and pain. *Progress in neurobiology*, 63(5), 569–611.
- PERTWEE, R. G. 2008. The diverse CB1 and CB2 receptor pharmacology of three plant cannabinoids: delta9-tetrahydrocannabinol, cannabidiol and delta9-tetrahydrocannabivarin. *British journal of pharmacology*, 153(2), 199–215.
- QUADEER, A. A., MORALES-JIMENEZ, D., & MCKAY, M. R. 2020. RocaSec: a standalone GUI-based package for robust co-evolutionary analysis of proteins. *Bioinformatics (Oxford, England)*, 36(7), 2262–2263.
- ROBERT, X. & GOUET, P. 2014 Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* 42(W1), W320-W324.



- RONQUIST, F., TESLENKO, M., VAN DER MARK, P., AYRES, D. L., DARLING, A., HÖHNA, S., LARGET, B., LIU, L., SUCHARD, M. A., & HUELSENBECK, J. P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology*, 61(3), 539–542.
- RUBOVITCH V., M. GAFNI, & Y. SARNE. 2004. The involvement of VEGF receptors and MAPK in the cannabinoid potentiation of Ca<sup>2+</sup> flux into N18TG2 neuroblastoma cells. *Brain Research; Molecular Brain Research* 120: 138-144.
- RYBERG, E., LARSSON, N., SJÖGREN, S., HJORTH, S., HERMANSSON, N.O., & LEONOVA, J. 2007. The orphan receptor GPR55 is a novel cannabinoid receptor. *British Journal of Pharmacology* 152:1092–101.
- SAWZDARGO, M., NGUYEN, T., LEE, D. K., LYNCH, K. R., CHENG, R., HENG, H. H., GEORGE, S. R., & O'DOWD, B. F. 1999. Identification and cloning of three novel human G protein-coupled receptor genes GPR52, PsiGPR53 and GPR55: GPR55 is extensively expressed in human brain. *Brain research. Molecular brain research*, 64(2), 193–198.
- SCHICHO R., & STORR M. 2012. A potential role for GPR55 in gastrointestinal functions. *Current Opinion in Pharmacology* 12 653–658.
- SHANNON, C.E. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27, 379-423, 623-656.
- SHAO, Z., YIN, J., CHAPMAN, K., GRZEMSKA, M., CLARK, L., WANG, J., & ROSENBAUM, D. M. (2016). High-resolution crystal structure of the human CB1 cannabinoid receptor. *Nature*, 540(7634), 602–606.
- SHEN, C.P., XIAO, J.C., ARMSTRONG, H., HAGMANN, W., & FONG, T.M. 2006. F200A substitution in the third transmembrane helix of human cannabinoid CB1 receptor converts AM2233 from receptor agonist to inverse agonist. *European Journal of Pharmacology* 531:41–46.

- SHIM J. Y. 2010. Understanding functional residues of the cannabinoid CB1. *Current topics in medicinal chemistry*, 10(8), 779–798.
- SHIRE D, CALANDRA B, DELPECH M, DUMONT X, KAGHAD M, LE FUR G, CAPUT D, & FERRARA P. 1996. Structural features of the central cannabinoid CB1 receptor involved in the binding of the specific CB1 antagonist SR 141716A. *Journal of Biological Chemistry* 271:6941–6946.
- SHRESTHA, R., DIXON, R.A. & CHAPMAN, K.D. 2003. Molecular identification of a functional homologue of the mammalian fatty acid amide hydrolase in *Arabidopsis thaliana*. *Journal of Biological Chemistry* 278:34990–34997.
- SINGH, R., HURST, D.P., BARNETT-NORRIS, J., LYNCH, D.L., REGGIO, P.H., & GUARNIERI, F. 2002. Activation of the cannabinoid CB1 receptor may involve a W648/F336 rotamer toggle switch. *Journal of peptide research* 60:357–370.
- SONG, Z.H., & FENG, W. 2002. Absence of a conserved proline and presence of a conserved tyrosine in the CB2 cannabinoid receptor are crucial for its function. *FEBS Letters* 531:290–294
- SONG, Y., DIMAIO, F., WANG, R. Y., KIM, D., MILES, C., BRUNETTE, T., THOMPSON, J., & BAKER, D. 2013. High-resolution comparative modeling with RosettaCM. *Structure (London, England : 1993)*, 21(10), 1735–1742.
- STAMATAKIS, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)*, 30(9), 1312–1313.
- STANČIĆ A., JANDL K., HASENÖHRL C., REICHMANN F., MARSCHE G., & SCHULIGOI R. 2015. The GPR55 antagonist CID16020046 protects against intestinal inflammation. *Neurogastroenterology and motility* 27 1432–1445.

- STATON P. C., HATCHER J. P., WALKER D. J., MORRISON A. D., SHAPLAND E. M., & HUGHES J. P. 2008. The putative cannabinoid receptor GPR55 plays a role in mechanical hyperalgesia associated with inflammatory and neuropathic pain. *Pain* 139 225–236.
- SUGIURA, T., KONDO, S., SUKAGAWA, A., NAKANE, S., SHINODA, A., ITOH, K., YAMASHITA, A., & WAKU, K. 1995. 2-arachidonoylglycerol— A possible endogenous cannabinoid receptor-ligand in brain. *Biochemical and Biophysical Research in Communications* 215, 89–97.
- TANG, X. L., WANG, Y., LI, D. L., LUO, J., & LIU, M. Y. 2012. Orphan G protein-coupled receptors (GPCRs): biological functions and potential drug targets. *Acta pharmacologica Sinica*, 33(3), 363–371.
- THOMAS, P. D., CAMPBELL, M. J., KEJARIWAL, A., MI, H., KARLAK, B., DAVERMAN, R., & NARECHANIA, A. 2003. PANTHER: a library of protein families and subfamilies indexed by function. *Genome research*, 13(9), 2129–2141.
- ULFERS, A.L., MCMURRY, J.L., MILLER, A., WANG, L., KENDALL, D.A., & MIERKE, D.F. 2002. Cannabinoid receptor-G protein interactions: G( $\alpha$ 1)-bound structures of IC3 and a mutant with altered G protein specificity. *Protein Science* 11:2526–2531.
- UNIPROT CONSORTIUM. 2008. The universal protein resource (UniProt). *Nucleic acids research*, 36(Database issue), D190–D195.
- WESS, J., NANAVATI, S., VOGEL, Z., & MAGGIO, R. 1993. Functional role of proline and tryptophan residues highly conserved among G protein-coupled receptors studied by mutational analysis of the m3 muscarinic receptor. *EMBO Journal* 12:331–338.
- ZHANG, D., ZHAO, Q., & WU, B. 2015. Structural Studies of G Protein-Coupled Receptors. *Molecules and cells*, 38(10), 836–842.

ZYGMUNT, P.M., PETERSSON, J., ANDERSSON, D.A., CHUANG, H.,  
SORGARD, M., & DI MARZO, V. 1999. Vanilloid receptors on sensory  
nerves mediate the vasodilator action of anandamide. *Nature* 400:452–  
457.



## 12. ANEXO

### 12.1. Alineamiento múltiple de secuencias del receptor CB1 para las 200 especies de vertebrados ocupadas en este estudio.





E_lucius ENSELU	MKSVLDGVADTTFRITIS	GLQYLGSDASY	DDATINADFKGRLSVQKPLSAFRGNFQD	KV
T_rubripes ENST	MKSVLDGVADTTFRITIS	GLQYLGSDANY	DDPLNDAAFKTGFSLQKPLSAFRSNFPN	KV
T_nigroviridis	MKSVLDGVADTTFRITIS	GLQYLGSDADY	DDPLGGADFKTGFSQKPLSAFRSNFPN	KL
L_bergylta ENSL	MKSVLDGVADTTFRITIS	GLQYLGSDASY	EGPLNDVNFKGGFSMQKPLSAFRSNFPD	KV
A_percula ENSAP	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPN	KV
S_partitus ENSS	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	EDPINDVNFKGSFSLQKPLSAFRSNFPN	KV
P_ranga ENSRNP	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFALQKPLSAFRSNFPD	KV
L_calcarifer EN	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPTNNVDFKGSFSLQKPLSAFRSNFPN	KV
L_crocea ENSLCR	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPLNDADFKAGFSLQKPLSAFRSNFPD	KV
S_aurata ENSSAU	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPLNDVDFKAGFSLQKPLSAFRSNFPD	KV
D_labrax ENSDLA	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPLNDVDFKAGFSLQKPLSAFRSNFPN	KV
S_lucioperca EN	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPLNDVDFKAGFSLQKPLSAFRSNFPN	KV
O_aureus ENSOAB	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPISDVDFKGSFSLQKPLSAFRSNFPN	KV
M_zebra ENSMZEP	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPLSDVDFKGSFSLQKPLSAFRSNFPN	KV
A_citrinellus E	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDAVSNVDFKGSFSLQKPLSAFRSNFPN	KV
S_dumerili ENSS	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPLNDVDFKGTFSLQKPLSAFRSNFPN	KV
S_lalandi ENSSLD	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPLNDVDFKGTFSLQKPLSAFRSNFPN	KV
M_mola ENSMOP0	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPLSNVDFKAGFSLQKPLSAFRSNFPN	KV
C_lumpus ENSCLM	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	EDPLNDVDFKAGFSLQKPLSAFRSNFPN	KV
M_armatus ENSMA	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPN	KV
E_naucrates ENS	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPTNDVDFKGSFSLQKPLSAFRSNFPN	KV
S_fasciatus ENS	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPTSDIDFKGSFPLQKPLSAFRSNFPD	KV
M_murdjan ENSM	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPN	KV
O_melastigma EN	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPN	KV
O_javanicus ENS	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPN	KV
K_marmoratus EN	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPN	KV
P_formosa ENSPF	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPN	KV
P_reticulata EN	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPN	KV
X_maculatus ENS	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPN	KV
G_affinis ENSGA	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPN	KV
G_affinis ENSGA	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPN	KV
C_variegatus EN	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KI
F_heteroclitus	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPN	KV
M_albus ENSMALP	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPN	KV
C_harengus ENSC	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
E_electricus ENS	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
I_punctatus EN	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
G_morhua ENSGMO	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
D_rerio ENDARP	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
C_auratus ENSCA	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
C_carpio ENSCCR	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
S_formosus ENSS	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
C_gobio ENSCGOP	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
P_paniscus ENSP	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
P_troglodytes E	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
G_gorila ENSGGO	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
N_leucogenys EN	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
P_tephrosceles	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
C_angolensis EN	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
R_bieti ENSRBIP	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
C_atys ENSCATP0	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
M_leucophaeus E	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
P_anubis ENSPAN	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
T_gelada ENSTGE	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
M_nemestrina EN	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
P_simus ENSPSMP	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
C_hoffmanni ENS	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
R_ferrumequinum	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
N_vison ENSNVIP	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
P_leo ENSPLOP00	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
P_pardus ENSPPR	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
F_catus ENSFCAP	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
L_canadensis EN	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
S_suricatta ENS	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
Z_californianus	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
U_americanus EN	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
P_catodon ENSPC	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
O_cuniculus ENS	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
S_vulgaris ENSS	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
M_domestica ENS	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
V_ursinus ENSVU	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
P_cinereus ENSP	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
O_anatinus ENSO	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
P_textilis ENSP	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
N_scutatus ENSN	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
N_naja ENSNNAPO	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
P_castaneus ENS	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
P_castaneus ENS	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
C_serpentina EN	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
T_carolina ENST	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
C_abingsdonii E	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
S_punctatus ENS	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
B_japonicus ENS	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
C_ustulatus ENS	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
M_cyaneus ENSMC	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
A_mexicanus ENS	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
C_syrichtha ENST	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
U_maritimus ENS	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
P_tigris ENSPTI	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
P_coquereli ENS	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
B_mutus ENSBMUP	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
C_capucinus ENS	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
S_boliviensis E	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
A_collaria ENSA	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
A_nancymaae ENS	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV
M_murinus ENSMI	MKSVLDGVADTTFRITIS	GLQYLGSDNANY	DDPINDVDFKGSFSLQKPLSAFRSNFPD	KV





E\_lucius|ENSELU PGDEELIFIPFFPT NATDLFNRS GF . . . . . ED DGA GI QCGENFMDMECFMILTPSQ  
T\_rubripes|ENST PAD EELIFIPFFPT NSTDL FNRS TT . . . . . R DE NSI QCGENFMDMECFMILTPSQ  
T\_nigroviridis| PAE EELIFISFFFPT NATDL FNRS TF . . . . . R DE NN I QCGENFMDMECFMILTPSQ  
L\_bergylla|ENSL PGDEELILIP.FPT NGTDL FNRS TF . . . . . R DES N T I QCGENFMDMECFMILTPSQ  
A\_percula|ENSAP PGDEELILIPFFYPT NATDL FNRS TF . . . . . R DET NN I QCGENFMDMECFMILTPSQ  
S\_partitus|ENSS PAD EELILIPFFYPT NATDL FNRS TF . . . . . R DET NN I QCGENFMDMECFMILTPSQ  
P\_ranga|ENSRNP PAD EELIIPFFYPT NATDL FNRS TF . . . . . R DET NN I QCGENFMDMECFMILTPSQ  
L\_calcarifer|ENL PPDEELILIPFFYPT NATDL FNRS TF . . . . . R DE TE NI QCGENFMDMECFMILTPSQ  
L\_crocea|ENSLCR PPDEELILIPFFPT NATDL FNRS TF . . . . . R DET N T I QCGENFMDMECFMILTPSQ  
S\_aurata|ENSSAU PSD EELILIPFFPT NATDL FNRS TF . . . . . R DET N T I QCGENFMDMECFMILTPSQ  
D\_labrax|ENSDLA PAD EELILIPFFPT NATDL FNRS TF . . . . . I DET N V QCGENFMDMECFMILTPSQ  
S\_lucioperca|EN PAD EELILIPFFYPT NATDL FNRS TF . . . . . R DET N NI QCGENFMDMECFMILTPSQ  
O\_aureus|ENSOAB PPDEELILIPFFYPT NATDL FNRS TF . . . . . K DET N NI QCGENFMDMECFMILTPSQ  
M\_zebra|ENSMZEP PPDEELILIPFFYPT NATDL FNRS TF . . . . . K DET N NI QCGENFMDMECFMILTPSQ  
A\_citrinellus|E PPDEELILIQFYPT NATDL FNRS TL . . . . . R DET NI QCGENFMDMECFMILTPSQ  
S\_dumerili|ENSS PAE EELILIPFFPT NATDL FNRS TF . . . . . R DET D NI QCGENFMDMECFMILTPSQ  
S\_lalandi|ENSSLD PAE EELILIPFFPT NATDL FNRS TF . . . . . R DET D NI QCGENFMDMECFMILTPSQ  
M\_mola|ENSMMP0 PD EELILIPFFPT NATDL FNRS TL . . . . . R D D T N NI QCGENFMDMECFMILTPSQ  
C\_lumpus|ENSCLM PAD EELILIPFFPT NATDL FNRS TF . . . . . R DET N T I QCGENFMDMECFMILTPSQ  
M\_armatus|EN SMA PAE EELILIPFFYPT NATDL FNRS TL . . . . . R DET N NI QCGENFMDMECFMILTPSQ  
E\_naucrates|ENS PAE DELVLI PFFPT NATDL FNRS TF . . . . . K D T G NI K CGENFMDMECFMILTPSQ  
S\_fasciatus|ENS . DELIL . . . . . DET N T I QCGENFMDMECFMILTPSQ  
M\_murdjan|ENSM PAD EQLI LIPLYPT NATDL FNRS MF . . . . . G ED N NI QCGENFMDMECFMILTPSQ  
O\_melastigma|EN PAD GELIFIPFPT NASDF FNRS TF . . . . . R DET N NI QCGENFMDMECFMILTPSQ  
O\_javanicus|ENS PAD GELIFIPFPT NASDF FNRS TF . . . . . R DET N NI QCGENFMDMECFMILTPSQ  
K\_marmoratus|EN HPG EELT MIP FPT NATDL FNRS TT . . . . . R DET N NI QCGENFMDMECFMILTPSQ  
P\_formosa|ENSPF PT D EELILIPFFPT NATDL FNRS TF . . . . . R DET N NI QCGENFMDMECFMILTPSQ  
P\_reticulata|EN PT D EELILIPFFPT NATDL FNRS TF . . . . . R DET N NI QCGENFMDMECFMILTPSQ  
X\_maculatus|ENS PAD EELILIPFFPT NATDL FNRS TF . . . . . R DET N NI QCGENFMDMECFMILTPSQ  
G\_affinis|ENSGA PT D EELILIPFFPT NATDL FNRS TF . . . . . R DET N NI QCGENFMDMECFMILTPSQ  
G\_affinis|ENSGA PT D EELILIPFFPT NATDL FNRS TF . . . . . R DET N NI QCGENFMDMECFMILTPSQ  
C\_variegatus|EN PAD EELILIPFFPT NATDL FNRS TF . . . . . R DET N NI QCGENFMDMECFMILTPSQ  
F\_heteroclitus| PPDEELILIPFFPT NATDL FNRS TF . . . . . R DET N T I QCGENFMDMECFMILTPSQ  
M\_albus|ENSMALP PAE EELIFIPLYPT NATDL FNRS TF . . . . . R DET N NI QCGENFMDMECFMVLNPSQ  
C\_harengus|ENSC APNEELILLSFYPT NSTDL FNKS SF . . . . . AGE GS T I E CGENFMDMECFMILTPSQ  
E\_electricus|EN GPDEELILLPFFPT NGTDV FNRS LA . . . . . E AG D NM K CGENFMDMECFMILTPGQ  
I\_punctatus|ENS PAD EELIVLFFYPT NGTDL FNQS .L . . . . . G EGA SV QCGENFMDMECFMILTPSQ  
G\_morhua|ENSGMO PAD EDLILMLP YPT NATEL LNRS AF . . . . . G EDGD TMRCS ENFMDMECFMILTPSQ  
D\_rerio|ENSDARP APDEELIVLFFYPT NNSDV FNRS . . . . . HA EDG SL QCGENFMDMECFMILTPSQ  
C\_auratus|ENSCA GPDEELIVIPFYP NTGSDV FNFS .L . . . . . G EGS SV QCGENFMDMECFMILTPSQ  
C\_carpio|ENSCCR GPDEELIVFFPYPT NNSDV FNFS .L . . . . . G EGS SL QCGENFMDMECFMILTPSQ  
S\_formosus|ENSS LAGEELILIPFYPT NVTELLNRS SD . . . . . VA EEG SI QCGENFMDMECFMILTPSQ  
C\_gobio|ENSCGOP .GNKEVIYLAIFPT NVSDFLNGT SV . . . . . D SGG AV QCGED FVDMECFMILTPGQ  
P\_panicus|ENSP . . . . . MEECWV . . . . . TE IANGSKDGLDSNP MKDYMILSGPQ  
P\_troglodytes|E . . . . . MEECWV . . . . . TE IANGSKDGLDSNP MKDYMLISGPQ  
G\_gorila|ENSGGO . . . . . MEECWV . . . . . TE IANGSKDGLDSNP MKDYMLISGPQ  
N\_leucogenys|EN . . . . . MEECWV . . . . . KE IANGSKDGLDSNP MKDYMLISGPQ  
T\_tephrosceles| . . . . . MEECWV . . . . . TE LANGSKNGLDFNP MKDYMLISGPQ  
C\_angolensis|EN . . . . . MEECWV . . . . . TE LANGSKNGLDFNP MKDYMLISGPQ  
R\_bieti|ENSRBIP . . . . . MEECWV . . . . . TE LANGSKNGLDFNP MKDYMLISGPQ  
C\_atys|ENSCATP0 . . . . . MEECWV . . . . . TE IANGSKNGLDFNP MKDYMLISGPQ  
M\_leucophaeus|E . . . . . MEECWV . . . . . TE IANGSKNGLDFNP MKDYMLISGPQ  
P\_anubis|ENSPAN . . . . . MEECWV . . . . . TE IANGSKNGLDFNP MKDYMLISGPQ  
T\_gelada|ENSTGE . . . . . MEECWV . . . . . TE IANGSKNGLDFNP MKDYMLISGPQ  
M\_nemestrina|EN . . . . . MEECWV . . . . . TE IANGSKNGLDFNP MKDYMLISGPQ  
P\_simus|ENSPSMP . . . . . MEECWV . . . . . ME TANASKDGLDFNP MKDYMLISGPQ  
C\_hoffmanni|ENS . . . . . MERCVW . . . . . ME AANS SKVGLDFNP MKDYMLIS GPQ  
R\_ferrumequinum . . . . . MERCPV . . . . . ME AANGSKDGLDFNP MKDYMLIS SHQ  
N\_vison|ENSNVIP . . . . . MERRVW . . . . . ME AANGCKDGLDFNP MKEYMVLN SSQ  
P\_leo|ENS PLOP0 . . . . . MERRVW . . . . . ME AANGCKDGLDFNP MKEYMVLN ASQ  
P\_pardus|ENSPPR . . . . . MERRVW . . . . . ME AANGCKDGLDFNP MKEYMVLN ASQ  
F\_catus|ENSFCAP . . . . . MERRVW . . . . . ME AANGCKDGLDFNP MKEYMVLN ASQ  
L\_canadensis|EN . . . . . MERRVW . . . . . ME AANGCKDGLDFNP MKEYMVLN ASQ  
M\_suricatta|ENS . . . . . MERRVW . . . . . ME AANGCKDGLDFNP MKEYMVLN ASQ  
Z\_californianus . . . . . MERRVW . . . . . ME AANGCKDGLDFNP MKEYMVLN ASQ  
U\_americanus|EN . . . . . . . . . . RE GERES QDGLDFNP MKEYMVLN SSQ  
P\_catodon|ENSPC . . . . . MEICVK . . . . . I QAANGS TDGLDFNP MKDYMLIS DPQ  
O\_cuniculus|ENS . . . . . MEPCGW . . . . . TE AANGS TEGLDFNP MKEYMVLN SDQ  
S\_vulgaris|ENSS . . . . . MEVC .L . . . . . ER AVNGSKQGLNSNP MEAYMVLN SPQ  
M\_domestica|ENS . . . . . MEGCWN . . . . . SG VANASHGDLT FSI FRNYT VLN I P  
V\_ursinus|ENSVU . . . . . MDTCWN . . . . . . STANTS QGHLDSDALRAYMVDFFPQ  
P\_cinereus|ENSP . . . . . MDACNV . . . . . . SANT S QGHLDSEALRDYMVDFFPQ  
O\_anatinus|ENSO . . . . . MEHRPG . . . . . SQ KVMGNSG . . . . . SAN YMVLS ASQ  
P\_textilis|ENSP . . . . . MEECNP . . . . . TE IFTNLQCNSTMLYLKCFMVLNQSE  
N\_scutatus|ENSN . . . . . MEECNP . . . . . TE IFTNLQCNSTMLYLKCFMVLNQSE  
N\_naja|ENSNNAPO . . . . . MEECNP . . . . . TE IFTNLQCNSTMLYLKCFMVLNQSE  
P\_castaneus|ENS . . . . . MDRSQM . . . . . NH TAOITN CNMNS SP MKGYMVLS PSL  
P\_castaneus|ENS . . . . . MDRSQM . . . . . NH TAOITN CNMNS SP MKGYMVLS PSL  
C\_serpentina|EN . . . . . MERCQM . . . . . GE TAKVSNCSMNAMP MPECYMLIS LAQ  
T\_carolina|ENST . . . . . MERCQ . . . . . VSNCSMNTMDLKYMVLNLTQ  
C\_abingsdonii|E . . . . . METCQ . . . . . ISNCSINTINLTCYMVLSHTQ  
S\_punctatus|ENS . . . . . MEKCKT . . . . . NGTANTSECSHHVAMECYMVLS PAQ  
B\_japonicus|ENS . . . . . MNSCKI . . . . . H . . . . . ENASKCSVN . . . . . TMECFMVLSQAQ  
C\_ustulatus|ENS . . . . . MDTCRV . . . . . P . . . . . ANTSKCSGN . . . . . TMECFMVLSQAQ  
M\_cyaneus|ENSMC . . . . . MTQTEVWGWCWLLSPLQG . . . . . GA AF A E CCGN . . . . . TMECFMVLSQTQ  
A\_mexicanus|ENS . . . . . YNKN TP T SY FPEAAM . . . . . TE EQS VVSAGFDG EYLMRYMVLN EPE  
C\_syricha|ENST . . . . . DPKVVV LPRDYL DAT AA . . . . . . . . . . ENVSATSQVPAVEPEP  
U\_maritimus|ENS . . . . . DIIVRH YNYTGL . . . . . NI SADK . . . . . EN  
P\_tigris|ENSPTI . . . . . DIIVRH YNYTGL . . . . . NVSAD . . . . . EN  
P\_coquereli|ENS . . . . . DIIVRH YNYTGL . . . . . NI SADK . . . . . EN  
B\_mutus|ENSBMUP . . . . . DIIVRH YNYTGL . . . . . KI SADK . . . . . DN  
C\_capucinus|ENS . . . . . DIIVLH YNYTGL . . . . . NTSADK . . . . .  
S\_boliviensis|E . . . . . DIIVLH YNYTGL . . . . . NTSADK . . . . .  
A\_collaria|ENSA . . . . . VIKHE YNYTGL . . . . . NENAD . . . . . S  
A\_nancymaae|ENS . . . . . FYNE SIAFF YNRS GL . . . . . AT EWN TV . . . . .  
M\_murinus|ENSMI . . . . . FYNE SIAFF YNRS GL . . . . . AT EWN TV . . . . .

	120	130	140	150	160	170
H_sapiens ENSP0	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
M_mulatta ENSM1	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
M_abelii ENSPY1	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
M_fascicularis	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
C_sabaeus ENSCS	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
R_roxellana ENS	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
O_aries ENSOAR1	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
M_moschiferus E	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
T_belangeri ENS	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
D_novemcinctus	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
C_dromedarius E	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
V_vulpes ENSVV1	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
L_africana ENSL	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
E_asinus ENSEAS	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
S_scrofa ENSSSC	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
A_melanoleuca E	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
B_taurus ENSBTA	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
C_hircus ENSCHI	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
T_truncatus ENS	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
D_leucas ENSDLE	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
P_vampyrus ENSP	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
V_pacos ENSVPA1	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
I_tridecemlinea	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
U_parryii ENSPU	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
M_marmota ENSM1	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
H_glaber ENSHGL	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
C_porcellus ENS	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
C_aperea ENSCAP	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
F_damarensis EN	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
C_lanigera ENSC	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
O_degus ENSODE1	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
M_auratus ENSM1	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
M_unguiculatus	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
M_ochrogaster E	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
M_musculus ENSM	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
M_spicilegus EN	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
M_caroli MGP_CA	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
M_spretus MGP_S	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
R_norvegicus EN	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
M_pahari MGP_Pa	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
C_canadensis EN	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
S_araneus ENSSA	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
C_lupus ENSCAFP	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
F_capensis ENSP	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
E_europaeus ENS	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
O_garnettii ENS	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
D_ordii ENSDOR1	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
J_jaculus ENSJJ1	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
S_harrisii ENSS	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
N_eugenii ENSM1	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
S_canaria ENSSC	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
L_striata ENSL1	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
G_guttata ENSTG	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
C_caerulesus ENS	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
Z_albicolis ENS	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
J_hyemalis ENSJ	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
G_fortis ENSGF1	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
C_parvulus ENSC	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
P_major ENSPMJ1	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
F_albicollis EN	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
Z_lateralis ENS	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
C_moneduloides	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
M_gallopavo ENS	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
P_cristatus ENS	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
G_gallus ENSGAL	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
C_japonica ENSC	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
N_meleagris ENS	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
C_pugnax ENSCPU	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
C_pygmaea ENSCP	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
F_tinnunculus E	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
B_bubo ENSOBPO	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
A_chryseatus EN	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
O_sunia ENSO1	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
M_undulatus ENS	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
A_cunicularia E	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
S_habroptila EN	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
A_platyrhynchos	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
A_zonorhyncha E	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
D_novahollandia	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
S_camelus ENSSC	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
A_owenii ENSAO1	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
A_haastii ENSAH	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
C_porosus ENSCP	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
N_perdicaria EN	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
L_coronata ENSL	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
P_sinensis ENSP	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
C_picta ENSCPB1	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
G_agassizii ENS	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
C_pictus ENSCPI	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
C_jacchus ENSCJ	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
M_putorius ENSM	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
A_carolinensis	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
P_muralis ENSPM	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
V_komodoensis E	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
P_vitticeps ENS	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
L_chalumnae ENS	QLVIAALSILGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
X_tropicalis EN	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
S_dauricus ENSS	QLAIAVLSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
S_trutta ENSS1	QLAVTMSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
S_salar ENSSAP	QLAVTMSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
O_mykiss ENSO1	QLAVTMSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
H_hucho ENSHHUP	QLAVTMSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD
O_tshawyscha E	QLAVAVMSLTLGTF	VLENI	LLVLCV	ILHSRSLR	CRPSYHFI	GSGLAVAD

E\_lucius|ENSELU QLA VAVMSLT LGTFTVLE NLVVLVFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
T\_rubripes|ENST QLA VAVLSLT LGTFTVLE NLVVLVCFI FQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
T\_nigroviridis|ENST QLA VAVLSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
L\_bergylla|ENSL QLA VAVLSLT LGTFTVLE NVVLCVFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
A\_percula|ENSAP QLA VAVMSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
S\_partitus|ENSS QLA VAVMSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
P\_ranga|ENSRNP QLA VAVLSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
L\_calcarifer|EN QLA VAVMSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
L\_crocea|ENSLCR QLA VAVMSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
S\_aurata|ENSSAU QLA VAVLSLT LGTFTVLE NLVVLVCFI FQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
D\_labrax|ENSDLA QLA VAVMSLT LGTFTVLE NLVVLVCFI FQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
S\_lucioerca|EN QLA VAVMSLT LGTFTVLE NLVVLVCFI FQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
O\_aureus|ENSOAB QLA VAVMSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
M\_zebra|ENSMZEP QLA VAVMSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
A\_citrinellus|E QLA VAVMSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
S\_dumerili|ENSS QLA VAVMSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
S\_lalandi|ENSSLD QLA VAVMSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
M\_mola|ENSMOP0 QLA VAVLSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
C\_lumpus|ENSCLM QLA VAVMSLT LGTFTVLE NLVVLVCFI FQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
M\_armatus|ENSM QLA VAVMSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
E\_naucratus|ENS QLA VAVMSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
S\_fasciatus|ENS QLA VAVMSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
M\_murdjan|ENSM QLA VAVMSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
O\_melastigma|EN QLA VAVLSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
O\_javanicus|ENS QLA VAVMSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
K\_marmoratus|EN QLA VAVMSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
P\_formosa|ENSPF QLA VAVMSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
P\_reticulata|EN QLA VAVMSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
X\_maculatus|ENS QLA VAVMSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
G\_affinis|ENSGA QLA VAVMSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
G\_affinis|ENSGA QLA VAVMSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
C\_variegatus|EN QLA VAVMSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
F\_heteroclitus|EN QLA VAVMSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
M\_albus|ENSMALP QLA VAVVCLT LGTFTVLE NLVVLVWFI LQ SRTLRSRPSYHFIGSLAVAD LLGSVIFVYSFL  
C\_harengus|ENSC QLA VAVMSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
E\_electricus|EN QLA IAVMSLT LGTFTVLE NLVVLVCFI LH SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
I\_punctatus|ENS QLA VAVMSLT LGTFTVLE NLVVLVCFI LH SRTLRCRPSYHFIGSLAMAD LLGSVIFVYSFL  
G\_morhua|ENSGMO QLA VAVMSLT LGTFTVLE NLVVLVCFI FQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
D\_rerio|ENDARP QLA IAVLSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
C\_auratus|ENSCA QLA IAVLSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
C\_carpio|ENSCCR QLA IALLSLT LGTFTVLE NLVVLVCFI LQ SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
S\_formosus|ENSS QLA VAVMALT LGTFTVLE NLVVLVCFI LH SRTLRCRPSYHFIGSLAVAD LLGSVIFVYSFL  
C\_gobio|ENSCGOP QLA VAVILALT LGTFTVLE NLVVLVCFI LH SRTLRSRPSYHFIGSLAVAD LIGSIFVYSFL  
P\_paniscus|ENSP QTA VAVLCTLLGLLSALENVAVLYLILS SHRLRRKPSYLFIFGSLAGADFLASVVFACSFV  
P\_trogodytes|E QTA VAVLCTLLGLLSALENVAVLYLILS SHRLRRKPSYLFIFGSLAGADFLASVVFACSFV  
G\_gorila|ENSGGO QTA VAVLCTLLGLLSALENVAVLYLILS SHRLRRKPSYLFIFGSLAGADFLASVVFACSFV  
N\_leucogenys|EN QTA VAVLCTLLGLLSALENMAVLYLILS SHRLRRKPSYLFIFGSLAGADFLASVVFACSFV  
P\_tephrrosceles| QIA VAVLCTLLGLLSALENVAVLYLILS SHRLRRKPSYLFIFGSLAGADFLASVVFACSFV  
C\_angolensis|EN QIA VAVLCTLLGLLSALENVAVLYLILS SHRLRRKPSYLFIFGSLAGADFLASVVFACSFV  
R\_bieti|ENSRBP QIA VAVLCTLLGLLSALENVAVLYLILS SHQLRRKPSYLFIFGSLAGADFLASVVFACSFV  
C\_atys|ENSCATP QIA IAVLCTLLGLLSALENVAVLYLILS SHRLRQKPSYLFIFGSLAGADFLASVVFACSFV  
M\_leucophaeus|E QIA IAVLCTLLGLLSALENVAVLYLILS SHRLRQKPSYLFIFGSLAGADFLASVVFACSFV  
P\_anubis|ENSPAN QIA IAVLCTLLGLLSALENVAVLYLILS SHRLRQKPSYLFIFGSLAGADFLASVVFACSFV  
T\_gelada|ENSTGE QIA IAVLCTLLGLLSALENVAVLYLILS SHRLRQKPSYLFIFGSLAGADFLASVVFACSFV  
M\_nemestrina|EN QIA IAVLCTLLGLLSALENVAVLYLILS SHRLRQKPSYLFIFGSLAGADFLASVVFACSFV  
P\_simus|ENSPSMP QIA IAVMCTLLGLLTALENLAVLYLILS SHRLRRKPSYLFIFGSLAGADFLASVVFACSFV  
C\_hoffmanni|ENS QIA VAVLCTLLGLLSALENVAVLYLILS SHRLRRKPSYLFIFGSLAGADFLASVVFACSFV  
R\_ferrumequinum QIA VAVLCTLLGLLSALENVAVLYLILS SHRLRRKPSYLFIFGSLAGADFLASVVFACSFV  
N\_vison|ENSNVIP RIA IAVLCTSLALLSALENLAVLYLILS SHRLRRKPSYLFIFGSLAVADFLASVIFACNFV  
P\_leo|ENSPLOP0 RIA IAVLCTSLALLSALENLAVLYLILS SHRLRRKPSYLFIFGSLAVADFLASVIFACNFV  
P\_pardus|ENSPPR RIA IAVLCTSLALLSALENLAVLYLILS SHRLRRKPSYLFIFGSLAVADFLASVIFACNFV  
F\_catus|ENSCFAP RIA IAVLCTSLALLSALENLAVLYLILS SHRLRRKPSYLFIFGSLAVADFLASVIFACNFV  
L\_canadensis|EN RIA IAVLCTSLALLSALENLAVLYLILS SHRLRRKPSYLFIFGSLAVADFLASVIFACNFV  
S\_suricatta|ENS RIA IAVLCTSLALLSALENLAVLYLILS YHRLRRKPSYLFIFGSLAVADFLASVIFACNFV  
Z\_californianus RIA VAVLCTSLALLSALENLAVLYLILS SHRLRRKPSYLFIFGSLAVADFLASVIFACNFV  
U\_americanus|EN RIA IAVLCTALALLSALENLAVLYLILS SHRLRRKPSYLFIFGSLAVADFLASVIFACNFV  
P\_catodon|ENSPC QIA IAVLCTSLGLLSALENLVVLVCFI LQ SRTLRRKPSYLFIFGSLAGADFLASVVFALNFV  
O\_cuniculus|ENS QIA VAVLCTLLGLLSALENLAVLYLILS SHWLRRKPSYLFIFGSLAVADFLASVVFACNFV  
S\_vulgaris|ENSS QAAVTVLCCTLLGLLSALENLTVLYLILS SHRLRRRPSYLFIFGSLADLLASVIFACNFV  
M\_domestica|ENS QTA IAVLCSLLGVLCALENAIVFLILS SPRLRRKPSYLFIFGSLAGADFLASLFFVIFSV  
V\_ursinus|ENSVU QTA IAVLCSLLGILCVLENAIVFFILS SPRLRRKPSYLFIFGSLAGADFLASLFFVIFSV  
P\_cinereus|ENSP QTA IAVLCSLLGVLCVLENAIVFFILS SPRLRRKPSYLFIFGSLAGADFLASLFFVIFSV  
O\_anatinus|ENS QTA VAVLCSVAGALCVLENAIVFFILS CRHLRRKPSYLFIFGSLAGADLLASPVFVCSFL  
P\_textilis|ENSP KPF IAVLCSIGGLCIENSLVLFILS SSKLRKKPSYILLSS LALADTLATVTFVSSFL  
N\_scutatus|ENSN KPF IAVLCSIGGLCIENSLVLFILS SSKLRKKPSYILLSS LALADTLATVTFVSSFL  
N\_naja|ENSNNAPO KPF IAVLCSIGGLCIENSLVLYLILS SSKLRKKPSYILLSS LALADTLATVTFVSSFL  
P\_castaneus|ENS KIS IAVLFSIFGILCIENCLVLYLILS SPKIRKKPSYLLISS LALADVLVSIIFVCSFV  
P\_castaneus|ENS KIS IAVLFSIFGILCIENCLVLYLILS SPKIRKKPSYLLISS LALADVLVSIIFVCSFV  
C\_serpentina|EN KTG IAVLCSIVGILCIENSLVLYLILS FASPKIRKKPSYLLISS LALADVLASIVFVCSFL  
T\_carolina|ENST KTG IAVLCSIVGILCVLENSLVLYLILS FASPKIRKKPSYLLISS LALADVLASIVFVCSFL  
C\_abingsdonii|E KTGITVLCIVGILCVLENSLVLLILFASPKIRKKPSYLLISS LALADVLASIVFVCSFL  
S\_punctatus|ENS KMA VAVLCSIVGTLCILENLVLYLILS SPRIHKKPSYLLISS LALADLLASIVFVCSFL  
B\_japonicus|ENS KIS IATLCLGFLGTCIFENSLVLYLILS SPRTIRKKPSYLFIFGSLALADILASIVFVCSFV  
C\_ustulatus|ENSC KISIGILCGLFGTMCVFENSLVLYLILS SPGIRKKPSYLFIFGSLALADILASIVFVCSFV  
M\_cyaneus|ENSMC KISIGILCGLFGTMCVFENSLVLYLILS SPGTRKKPSYLFIFGSLALADILASIVFVCSFV  
A\_mexicanus|ENS KKLISLIVLCTSGTFLISENAIVVLI LHNPSLR.AP MFLIGSLALADLLAGVGLIHFV  
S\_syrichtha|ENST ELAVNIVLCTSGTFLISENAIVVLI LHNPSLR.AP MFLIGSLALADLLAGVGLIHFV  
U\_maritimus|ENST GIKLSVVFI LICCFI ILENIFVLLT IWKTKKFH.RPMYFIGNLALSDLLAGVAYTANLL  
P\_tigris|ENSPTI GIKLSVVFI LICCFI ILENIFVLLT IWKTKKFH.RPMYFIGNLALSDLLAGVAYTANLL  
P\_coquereli|ENS AIKPTVVFI LICCFI ILENIFVLLT IWKTKKFH.RPMYFIGNLALSDLLAGVAYTANLL  
B\_mutus|ENSBMUP GIKLIVVFI LICCFI ILENIFVLLT IWKTKKFH.RPMYFIGNLALSDLLAGVAYTANLL  
C\_capucinus|ENS DIKLTVVFI LICCFI ILENIFVLLT IWKTKKFH.RPMYFIGNLALSDLLAGVAYTANLL  
S\_boliviensis|E DIKMTVVFI LICCFI ILENIFVLLT IWKTKKFH.RPMYFIGNLALSDLLAGVAYTANLL  
A\_collaria|ENSA GIKVTVVFI VICCFI ILENIFVLLT IWKTKKFH.RPMYFIGNLALSDLLAGVAYTANLL  
A\_nancymaae|ENS SKLVMGLITVCFIFIMLANLLVMVAITYVNRREF.FPIYYLMANLAAADFFAGLAYFYLMF  
M\_murinus|ENSMI SKLVMGLITVCFIFIMLANLLVMVAITYVNRREF.FPIYYLMANLAAADFFAGLAYFYLMF

	180	190	200	210	220	230
H_sapiens ENSP0	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
M_mulatta ENSM	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
M_abelii ENSPY	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
M_fascicularis	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
C_sabaeus ENSC	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
R_roxellana ENS	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
O_aries ENSOAR	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
M_moschiferus E	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
T_belangeri ENS	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
D_novemcinctus	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
C_dromedarius E	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
V_vulpes ENSVV	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
L_africana ENSL	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
E_asinus ENSEAS	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
S_scrofa ENSSC	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
A_melanoleuca E	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
B_taurus ENSBTA	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
C_hircus ENSCHI	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
T_truncatus ENS	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
D_leucas ENSDLE	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
P_vampyrus ENSP	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
V_pacos ENSVPA	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
I_tridecemlinea	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
U_parryii ENSPU	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
M_marmota ENSM	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
H_glaber ENSHGL	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
C_porcellus ENS	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
C_aperea ENSCAP	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
F_damarensis EN	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
C_lanigera ENSC	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
O_degus ENSODEP	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
M_auratus ENSMA	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
M_unguiculatus	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
M_ochrogaster E	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
M_musculus ENSM	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
M_spicilegus EN	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
M_caroli MGP_CA	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
M_spretus MGP_S	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
R_norvegicus EN	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
M_pahari MGP_Pa	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
C_canadensis EN	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
S_araneus ENSSA	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
C_lupus ENSCAFP	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
P_capensis ENSP	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
E_europaeus ENS	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
O_garnettii ENS	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
D_ordii ENSODOR	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
J_jaculus ENSJJ	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
S_harrisii ENSS	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
N_eugenii ENSME	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
S_canaria ENSSC	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
L_striata ENSL	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
T_guttata ENSTG	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
C_caeruleus ENS	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
Z_albicolis ENS	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
J_hyemalis ENSJ	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
G_fortis ENSGFO	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
C_parvulus ENSC	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
P_major ENSPMPJ	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
F_albicollis EN	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
Z_lateralis ENS	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
C_moneduloides	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
M_gallopavo ENS	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
P_cristatus ENS	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
G_gallus ENSGAL	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
C_japonica ENSC	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
N_meleagris ENS	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
C_pugnax ENSCPU	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
C_pygmaea ENSCP	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
F_tinnunculus E	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
B_bubo ENSBOBP	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
A_chryseatus EN	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
O_sunia ENSOSUP	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
M_undulatus ENS	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
A_cunicularia E	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
S_habroptila EN	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
A_platyrrhynchos	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
A_zonorhyncha E	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
D_novahollandia	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
S_camelus ENSSC	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
A_owenii ENSAOW	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
A_haastii ENSAH	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
C_porosus ENSCP	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
N_perdicaria EN	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
L_coronata ENSL	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
P_sinensis ENSP	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
C_picta ENSCPBP	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
G_agassizii ENS	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
C_pictus ENSCPI	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
C_jacchus ENSCJ	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
M_putorius ENSM	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
A_carolinensis	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
P_muralis ENSPM	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
V_komodoensis E	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
P_vitticeps ENS	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
L_chalumnae ENS	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
X_tropicalis EN	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
S_dauricus ENSS	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
S_trutta ENSSTU	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
S_salar ENSSSAP	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
O_mykiss ENSOXY	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
H_hucho ENSHHUP	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID
O_tshawyscha E	DFHVFH	HRKDSR	NVFLFK	LGCVTAS	FTASVGS	LFLTAID

E\_lucius|ENSELU DFHVFHRRKDTPSVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTRAVID  
T\_rubripes|ENST DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
T\_nigroviridis|EN DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
L\_bergyllta|ENSL DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
A\_percula|ENSAP DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
S\_partitus|ENSS DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
P\_ranga|ENSRNP DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
L\_calcarifer|EN DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
L\_crocea|ENSLCR DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
S\_aurata|ENSSAU DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
D\_labrax|ENSDLA DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
S\_lucioerca|EN DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
O\_aureus|ENSOAB DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
M\_zebra|ENSMZEP DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
A\_citrinellus|E DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
S\_dumerili|ENSS DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
S\_lalandi|ENSSLD DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
M\_mola|ENSMOP0 DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
C\_lumpus|ENSCLM DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
M\_armatus|ENSMMA DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
E\_naucrates|ENS DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
S\_fasciatus|ENS DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
M\_murdjan|ENSM DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
O\_melastigma|EN DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
O\_javanicus|ENS DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
K\_marmoratus|EN DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
P\_formosa|ENSNFP DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
P\_reticulata|EN DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
X\_maculatus|ENS DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
G\_affinis|ENSGA DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
G\_affinis|ENSGA DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
C\_variegatus|EN DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
F\_heteroclitus|EN DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
M\_albus|ENSMALP DFHVFHRRKDTPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
C\_harengus|ENSC DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
E\_electricus|EN DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
I\_punctatus|ENS DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
G\_morhua|ENSGMO DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
D\_rerio|ENDSARP DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
C\_auratus|ENSCA DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
C\_carpio|ENSCCR DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
S\_formosus|ENSS DFHVFHRRKDTPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
C\_gobio|ENSCGOF DFHVFHRRKDSPNVFLFKLGGVTASFTASVGSLLFLTAIDRYISIHRRLAYRRIVTRTKAVI  
P\_panicus|ENSP NFHVFHGVDSKAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLRYPPSYKALLTRGRALV  
P\_trogodytes|E NFHVFHGVDSKAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLRYPPSYKALLTRGRALV  
G\_gorila|ENSGGO NFHVFHGVDSKAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLRYPPSYKALLTRGRALV  
N\_leucogenys|EN NFHVFHGMDSKAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLRYPPSYKALLTRGRALV  
P\_tephrosceles| NFHVFHGMDSKAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLRYPPSYKALLTRGRALV  
C\_angolensis|EN NFHVFHGMDSKAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLRYPPSYKALLTRGRALV  
R\_bieti|ENSRBIP NFHVFHGMDSKAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLRYPPSYKALLTRGRALV  
C\_atys|ENSCATP0 NFHVFHGMDSKAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLRYPPSYKALLTRGRALV  
M\_leucophaeus|E NFHVFHGMDSKAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLRYPPSYKALLTRGRALV  
P\_anubis|ENSPAN NFHVFHGMDSKAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLRYPPSYKALLTRGRALV  
T\_gelada|ENSTGE NFHVFHGMDSKAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLRYPPSYKALLTRGRALV  
M\_nemestrina|EN NFHVFHGMDSKAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLRYPPSYKALLTRGRALV  
P\_simus|ENSPSMP NFHVFHGVDSKAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLRYPPSYKALLTRGRALV  
C\_hoffmanni|ENS NFHVFHGVDSKAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLRHPPAYKALLTRGRALA  
R\_ferrumequinum|NFHVFHGVDSKAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLRYPPAYKALLTRGRALV  
N\_vision|ENSNVIP EFHVFHDMDSKAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLCHPPTYKALLTRGRALA  
P\_leo|ENSPLOP0 NFHVFHGVDSKAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLCYPPPTYKALLTRGRALA  
P\_pardus|ENSPPR NFHVFHGVDSKAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLCYPPPTYKALLTRGRALA  
F\_catus|ENSFCAP NFHVFHGVDSQAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLCYPPPTYKALLTRGRALA  
L\_canadensis|EN NFHVFHGVDSQAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLCYPPPTYKALLTRGRALA  
S\_saricatta|ENS NFHVFHGVDSKAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLCYPPPTYKALLTRGRALA  
Z\_californianus|NFHVFHGMDSKAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLCYPPPTYKALLTRGRALA  
N\_americanus|EN NFHVFHGVDSKAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLCYPPPTYKALLTRGRALA  
P\_catodon|ENSPC NFHVFHGVDSKAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLRYPPPTYKALLTRGRALV  
O\_cuniculus|ENS NFHVFQGVDSRAVFLMKIGSVTMTFTASVGSLLFLTAIDRYLCLRYPPPTYKALLTRGRALV  
S\_vulgaris|ENSS SFHVFHNMDSEVIFLLKIGSVTMTFTASVGSLLFLTAIDRYLCLRYPPPTYKALLTRGRALV  
M\_domestica|ENS NFHVFHQEDSRDIFLLKIGSVTMTFTASVGSLLFLTAIDRYLCLRHAPARYKALFTRGRALV  
V\_ursinus|ENSVU KFHVHFQEEESRTAFLLKIGSVTMTFTASVGSLLFLTAIDRYLCLCHPARKYKALLTRGRALA  
P\_cinereus|ENSP KFHVHFQEEESRSAPLLKIGSVTMTFTASVGSLLFLTAIDRYLCLCHPARYKALLTRGRALA  
O\_anatinus|ENSO DFHVLNRVDSRETYLLKIGSVTMTFTASVGSLLFLTAGDRLYCLGHPARYKALLTRGRALA  
P\_textilis|ENSP SFHVFNRDTSKEIYLLKIGSVTMTFTASVGSLLFLMADFDRYICILRPSHYKQLVTRRAVV  
N\_scutatus|ENSN SFHVFNRDTSKEIYLLKIGSVTMTFTASVGSLLFLMADFDRYICILRPSHYKQLVTRRAVV  
N\_naja|ENSNNAPO SFHVFNRDTSKEIYLLKIGSVTMTFTASVGSLLFLMADFDRYICILRPSHYKQLVTRRAVV  
P\_castaneus|ENS DFHVFNTDIFKEIFLLKIGSVTMTFTASVGSLLFLMADFDRYVICIQRSEYKIVTRKRALV  
P\_castaneus|ENS DFHVFNTDIFKEIFLLKIGSVTMTFTASVGSLLFLMADFDRYVICIQRSEYKIVTRKRALV  
C\_serpentina|EN DFHVFNRDTSKEIYLLKIGSVTMTFTASVGSLLFLMADFDRYVICIQRSEYKIVTRKRALV  
T\_carolina|ENST DFHVFNRDTSKEIYLLKIGSVTMTFTASVGSLLFLMADFDRYVICIQRSEYKIVTRKRALV  
C\_abingsdonii|E DFHVLNRDTSKEIYLLKIGSVTMTFTASVGSLLFLMADFDRYVICIHRPSEYKIVTRERAVM  
S\_punctatus|ENS HFHVFNGDTSKDFLLKIGSVTMTFTASVGSLLFLMADFDRYICIHRRPSEYKIVTRKRALV  
B\_japonicus|ENS NFHVFNETNSKEMFLLQGGVNTSFSASLSLSTLALDRYISISRPSSEYKLLMTRKRAWI  
C\_ustulatus|ENS NFHVFNDTDSKEVFLQGGVNTSFSASLSLSTLALDRYISISRPSSEYKLLMTRKRAWI  
M\_cyaneus|ENSMC NFHVFNDTDSKEVFLQGGVNTSFSASLSLSTLALDRYISISRPSSEYKLLMTRKRAWI  
A\_mexicanus|ENS DFHFLKRSDSPEIYLLKIGSVTMTFTASVGSLLFLTSMDRVLICIQAPRYKSLMTCRRALF  
C\_syrichtha|ENST FAYLLQ...SEATKLVITGLIVASFSASVCSLLAITVDRLYLSLYALTYHSERTVFTYV  
U\_maritimus|ENS LSGATTYKLTPAQWFLREGSMFVALSASVFLSLLAIAIERYITMLK.MKLHNGSNSEFRSFL  
P\_tigris|ENSPTI LSGATTYKLTPAQWFLREGSMFVALSASVFLSLLAIAIERYITMLK.MKLHNGSNSEFRSFL  
P\_coquereli|ENS LSGATTYKLTPAQWFLREGSMFVALSASVFLSLLAIAIERYITMLK.MKLHNGSNSEFRSFL  
B\_mutus|ENSMUP LSGATTYKLTPAQWFLREGSMFVALSASVFLSLLAIAIERYITMLK.MKLHNGSNSEFRSFL  
C\_capucinus|ENS LSGATTYKLTPAQWFLREGSMFVALSASVFLSLLAIAIERYITMLK.MKLHNGSNSEFRSFL  
S\_boliviensis|E LSGATTYKLTPAQWFLREGSMFVALSASVFLSLLAIAIERYITMLK.MKLHNGSNSEFRSFL  
A\_collaria|ENSA LSGHKTYSLTPSQWFLREGSMFVALSASVFLSLLAIAIERYITMLK.MKLHNGSNSEFRSFL  
A\_nancymaae|ENS NTGPNTRRLLTVSTWLLRQGLIDTSLTASVANLLAIAIERHITVFR.MQLHTRMSNRVVV  
M\_murinus|ENSMI NTGPNTRRLLTVSTWLLRQGLIDTSLTASVANLLAIAIERHITVFR.MQLHTRMSNRVVV

	240	250	260	270	280	290
H_sapiens ENSP0	AFCLMWTIA	IVIAVLP	LLGWNCEK	LQSVCS	DIFPHI	DET
M_mulatta ENSM	AFCLMWTIA	IVIAVLP	LLGWNCEK	LQSVCS	DIFPHI	DET
P_abelii ENSPY	AFCLMWTIA	IVIAVLP	LLGWNCEK	LQSVCS	DIFPHI	DET
M_fascicularis	AFCLMWTIA	IVIAVLP	LLGWNCEK	LQSVCS	DIFPHI	DET
C_sabaeus ENSC	AFCLMWTIA	IVIAVLP	LLGWNCEK	LQSVCS	DIFPHI	DET
R_roxellana ENS	AFCLMWTIA	IVIAVLP	LLGWNCEK	LQSVCS	DIFPHI	DET
O_aries ENSOAR	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
M_moschiferus E	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
T_belangeri ENS	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
D_novemcinctus	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
C_dromedarius E	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
V_vulpes ENSVU	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
L_africana ENSL	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
E_asinus ENSEAS	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
S_scrofa ENSSC	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
A_melanoleuca E	AFCLMWTIA	IVIAVLP	LLGWNCKK	LRSVCS	DIFPHI	DET
B_taurus ENSBT	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
C_hircus ENSC	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
T_truncatus ENS	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
D_leucas ENSDLE	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
P_vampyrus ENSP	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
V_pacos ENSVPA	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
I_tridecemlinea	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
U_parryii ENSP	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
M_marmota ENSM	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
H_glaber ENSHGL	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
C_porcellus ENS	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
C_aperea ENSCAP	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
F_damarensis EN	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
C_lanigera ENSC	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
O_degus ENSODEP	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
M_auratus ENSM	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
M_unguiculatus	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
M_ochrogaster E	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
M_musculus ENSM	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
M_spicilegus EN	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
M_caroli MGP_CA	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
M_spretus MGP_S	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
R_norvegicus EN	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
M_pahari MGP_Pa	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
C_canadensis EN	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
S_araneus ENSSA	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
C_lupus ENSCAFP	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
P_capensis ENSP	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
E_europaeus ENS	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
O_garnettii ENS	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
D_ordii ENSDORP	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
J_jaculus ENSJJ	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
S_harrisii ENSS	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
N_eugenii ENSME	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
S_canaria ENSSC	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
L_striata ENSL	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
T_guttata ENSTG	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
C_caeruleus ENS	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
Z_albicolis ENS	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
H_hyemalis ENSJ	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
G_fortis ENSGFO	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
C_parvulus ENSC	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
P_major ENSPMJP	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
F_albicollis EN	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
Z_lateralis ENS	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
C_moneduloides	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
M_gallopavo ENS	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
P_cristatus ENS	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
G_gallus ENSGAL	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
C_japonica ENSC	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
N_meleagris ENS	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
C_pugnax ENSCPU	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
C_pygmaea ENSCP	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
F_tinnunculus E	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
B_bubo ENSSOBB	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
A_chryseatus EN	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
O_sunia ENSOSUP	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
M_undulatus ENS	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
A_cunicularia E	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
S_habroptila EN	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
A_platyrrhynchos	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
A_zonorhyncha E	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
D_novahollandia	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
S_camelus ENSSC	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
A_owenii ENSAOW	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
A_haastii ENSAH	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
C_porosus ENSCP	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
N_perdicaria EN	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
L_coronata ENSL	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
P_sinensis ENSP	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
C_picta ENSCBP	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
G_agassizii ENS	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
C_pictus ENSCPI	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
C_jacchus ENSCJ	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
M_putorius ENSM	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
A_carolinensis	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
P_muralis ENSPM	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
V_komodoensis E	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
P_vitticeps ENS	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
L_chalumnae ENS	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
X_tropicalis EN	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
S_dauricus ENSS	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
S_trutta ENSSTU	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
S_salar ENSSSAP	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
O_mykiss ENSSOMY	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
H_hucho ENSHHUP	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET
O_tshawytscha E	AFCLMWTIA	IVIAVLP	LLGWNCKK	LQSVCS	DIFPHI	DET

E\_lucius|ENSELU AF CVMWTTISIVFAVLP LLGWNC KQLKSV CS DIF PLIDEKYLFWIG VTSVLVLF I IYAYM  
T\_rubripes|ENST AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
T\_nigroviridis| AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
L\_bergyta|ENSL AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
A\_percula|ENSAP AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
S\_partitus|ENSS AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDEKYLFWIG VTSVLVLF I IYAYI  
P\_ranga|ENSRNP AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
L\_calcarifer|EN AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
L\_crocea|ENSLCR AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
S\_aurata|ENSSAU AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
D\_labrax|ENSDLA AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
S\_lucioerca|EN AF CVMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
O\_aureus|ENSOAB AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
M\_zebra|ENSMZEP AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
A\_citrinellus|E AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
S\_dumerili|ENSS AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
S\_lalandi|ENSSLD AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
M\_mola|ENSMOP0 AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
C\_lumpus|ENSCLM AF CVMWTTISIIIAVLP LLGWNC KRLNSI CS DIF PLIDVNYLFWIG VTSVLVLF I IYAYI  
M\_armatus|ENSM AF CMMWTTISIIIAVLP LLGWNC KYLNSV CS DIF PLIDKNYLFWIG VTSVLVLF I IYAYI  
E\_naucrates|ENS AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
S\_fasciatus|ENS AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
M\_murdjan|ENSM AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYM  
O\_melastigma|EN AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDETYLFWIG VTSVLVLF I IYAYI  
O\_javanicus|EN AF CMMWTTISIIIAVLP LLGWNC KRLNSI CS DIF PLIDETYLFWIG VTSVLVLF I IYAYI  
K\_marmoratus|EN AF IMMWAISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
P\_formosa|ENSPF AF SVMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
P\_reticulata|EN AF SVMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
X\_maculatus|ENS AF SVMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
G\_affinis|ENSGA AF SVMWTTISIIIAVLP LLGWNC KRLNSA CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
G\_affinis|ENSGA AF SVMWTTISIIIAVLP LLGWNC KRLNSA CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
C\_variegatus|EN AF SVMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
F\_heteroclitus| AF SVMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
M\_albus|ENSMALP AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYI  
C\_harengus|ENSC AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYM  
E\_electricus|EN AF SVMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYM  
I\_punctatus|ENS AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYM  
G\_morhua|ENSGMO AF CVMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDKNYLFWIG VTSVLVLF I IYAYM  
D\_rerio|ENDARP AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYM  
C\_auratus|ENSCA AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYM  
C\_carpio|ENSCCR AF CMMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYM  
S\_formosus|ENSS AF CLMWAISIIIAVLP LLGWNC KRLNSV CS DIF PLIDENYLMFWIG VTSVLVLF I IYAYV  
C\_gobio|ENSCGOP AF SVMWTTISIIIAVLP LLGWNC KRLNSV CS DIF PLIDQKYLFWIG MTIILVLF I IYAYM  
P\_panicus|ENSP TLGIMWVLSALVSYLPLMGWTC...PRP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
P\_troglodytes|E TLGIMWVLSALVSYLPLMGWTC...PRP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
G\_gorila|ENSGGO TLGIMWVLSALVSYLPLMGWTC...PRP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
N\_leucogenys|EN TLGIMWVLSALVSYLPLMGWTC...PRP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
P\_tephrosceles| TLGIMWVLSALVSYLPLMGWTC...PRP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
C\_angolensis|EN TLGIMWVLSALVSYLPLMGWTC...PRP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
R\_bieti|ENSRBIP TLGIMWVLSALVSYLPLMGWTC...PRP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
C\_atys|ENSCATP TLGIMWVLSALVSYLPLMGWTC...PRP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
M\_leucophaeus|E TLGIMWVLSALVSYLPLMGWTC...PRP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
P\_anubis|ENSPAN TLGIMWVLSALVSYLPLMGWTC...PRP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
T\_gelada|ENSTGE TLGIMWVLSALVSYLPLMGWTC...PRP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
M\_nemestrina|EN TLGIMWVLSALVSYLPLMGWTC...PRP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
P\_simus|ENSPSMP LGITWVLSALVSYLPLMGWTC...PSP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
C\_hoffmanni|ENS TLVIMWVLSALVSYLPLMGWTC...PRP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
R\_ferrumequinum TLAIMWVLSALVSYLPLMGWTC...PSP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
N\_vision|ENSNVIP TLGIMWVLSALVSYLPLMGWTC...PSP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
P\_leo|ENSPLOP0 TLGIMWVLSALVSYLPLMGWTC...PSP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
P\_pardus|ENSPPR TLGIMWVLSALVSYLPLMGWTC...PSP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
F\_catus|ENSFCA TLGIMWVLSALVSYLPLMGWTC...PSP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
L\_canadensis|EN TLGIMWVLSALVSYLPLMGWTC...PSP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
S\_suricatta|ENS TLGVMWVLSALVSYLPLMGWTC...PSP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
Z\_californianus TLGIMWVLSALVSYLPLMGWTC...PSP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
U\_americanus|EN TLGIMWVLSALVSYLPLMGWTC...PSP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
P\_catodon|ENSPC TLGIMWVLSALVSYLPLMGWTC...PSP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
O\_cuniculus|ENS TLGVMWVLSALVSYLPLMGWTC...PSP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
S\_vulgaris|ENSS TLGLMWVLSALVSYLPLMGWTC...PSP CSELP FLIPNDYLLSWL FTI AFLFSGI IYTYG  
M\_domestica|ENS MLTAMWVLSALVSYLPLMGWTC...PRL CSELP FLIPNDYLLSWL FTI AFLFVA IYTYG  
V\_ursinus|ENSVU ALAAMWTTISVVVSYLPLMGWTC...PRP CSELP FLIPNDYLLSWL FTI ILLVA IYTYT  
P\_cinereus|ENSP TLAGMWVLSMVSYLPLMGWTC...PRP CSELP FLISDDYLLSWL FTI ILLVA IYTYT  
O\_anatinus|ENSO ALAAMWVLSALVSYLPLMGWTC...GSP CSELP FLIPNDYLLSWL FTI ILLVA IYTYG  
P\_textilis|ENSP ALLVWVLSALVSYLPLMGWTC...PSP CSELP FLIPNDYLLSWL FTI ILLVA IYTYG  
N\_scutatus|ENSN ALLVWVLSALVSYLPLMGWTC...PSP CSELP FLIPNDYLLSWL FTI ILLVA IYTYG  
N\_naja|ENSNNAPO ALLVWVLSALVSYLPLMGWTC...PSP CSELP FLIPNDYLLSWL FTI ILLVA IYTYG  
P\_castaneus|ENS AMAIMWTTITMIAFPLMGWTC...KA CSELP FLIPNDYLLSWL FTI ILLVA IYTYG  
P\_castaneus|ENS AMAIMWTTITMIAFPLMGWTC...KA CSELP FLIPNDYLLSWL FTI ILLVA IYTYG  
C\_serpentina|EN AMAVMWIATMITAFPLMGWTC...CNSA CSELP FLIPNDYLLSWL FTI ILLVA IYTYG  
T\_carolina|ENST AMAVMWIATMITAFPLMGWTC...CNSA CSELP FLIPNDYLLSWL FTI ILLVA IYTYG  
C\_abingsdonii|E AMAVMWIATMITAFPLMGWTC...CNSA CSELP FLIPNDYLLSWL FTI ILLVA IYTYG  
S\_punctatus|ENS ALAVLWVLSALVSYLPLMGWTC...CNSA CSELP FLIPNDYLLSWL FTI ILLVA IYTYG  
B\_japonicus|ENS ALGVWVLSALVSYLPLMGWTC...CNSA CSELP FLIPNDYLLSWL FTI ILLVA IYTYG  
C\_ustulatus|ENS ALAVLWVLSALVSYLPLMGWTC...CNSA CSELP FLIPNDYLLSWL FTI ILLVA IYTYG  
M\_cyaneus|ENSMC ALAVLWVLSALVSYLPLMGWTC...CNSA CSELP FLIPNDYLLSWL FTI ILLVA IYTYG  
A\_mexicanus|ENS GIVALWVLSALVSYLPLMGWTC...CNSA CSELP FLIPNDYLLSWL FTI ILLVA IYTYG  
C\_syrichtha|ENST MLVLLWVLSALVSYLPLMGWTC...CNSA CSELP FLIPNDYLLSWL FTI ILLVA IYTYG  
U\_maritimus|ENS ...ACVWVLSALVSYLPLMGWTC...CNSA CSELP FLIPNDYLLSWL FTI ILLVA IYTYG  
P\_tigris|ENSPPT LISACVWVLSALVSYLPLMGWTC...CNSA CSELP FLIPNDYLLSWL FTI ILLVA IYTYG  
P\_coquereli|ENS LISACVWVLSALVSYLPLMGWTC...CNSA CSELP FLIPNDYLLSWL FTI ILLVA IYTYG  
B\_mutus|ENSMUP LISACVWVLSALVSYLPLMGWTC...CNSA CSELP FLIPNDYLLSWL FTI ILLVA IYTYG  
C\_capucinus|ENS LISACVWVLSALVSYLPLMGWTC...CNSA CSELP FLIPNDYLLSWL FTI ILLVA IYTYG  
S\_boliviensis|E LISACVWVLSALVSYLPLMGWTC...CNSA CSELP FLIPNDYLLSWL FTI ILLVA IYTYG  
A\_collaria|ENSA LISACVWVLSALVSYLPLMGWTC...CNSA CSELP FLIPNDYLLSWL FTI ILLVA IYTYG  
A\_nancymaae|ENS IVVIVWVLSALVSYLPLMGWTC...CNSA CSELP FLIPNDYLLSWL FTI ILLVA IYTYG  
M\_murinus|ENSMI IVVIVWVLSALVSYLPLMGWTC...CNSA CSELP FLIPNDYLLSWL FTI ILLVA IYTYG

	300	310	320	330	340	350
H_sapiens ENSP0	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
M_mulatta ENSM0	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
P_abelii ENSPPY	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
M_fascicularis E	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
C_sabaeus ENSSC	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
R_roxellana ENS	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
O_aries ENSOARP	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
M_moschiferus E	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
T_belangeri ENS	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
D_novemcinctus	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
C_dromedarius E	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
V_vulpes ENSVU	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
L_africana ENSL	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
E_asinus ENSEAS	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
S_scrofa ENSSSC	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
A_melanoleuca E	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
B_taurus ENSBTA	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
C_hircus ENSMCH	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
T_truncatus ENS	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
L_leucas ENSDLE	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
P_vampyrus ENSP	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
V_pacos ENSVVAP	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
I_tridecemlinea	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
U_parryii ENSPU	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
M_marмота ENSM	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
H_glaber ENSHGL	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
C_porcellus ENS	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
C_aperea ENSCAP	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
F_damarensis EN	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
C_lanigera ENSC	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
O_degus ENSOSEP	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
M_auratus ENSM	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
M_unguiculatus	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
M_ochrogaster E	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
M_musculus ENSM	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
M_spicilegus EN	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
M_caroli MGP_CA	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
M_spretus MGP_S	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
R_norvegicus EN	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
M_pahari MGP_Pa	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
C_canadensis EN	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
S_araneus ENSSA	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
C_lupus ENSCAFP	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
F_capensis ENSP	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
E_europaeus ENS	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
O_garnettii ENS	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
D_ordii ENSDORP	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
J_jaculus ENSJJ	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
S_harrisii ENSS	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
N_eugenii ENSME	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
S_canaria ENSSC	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
L_striata ENSL	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
T_guttata ENSTG	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
C caeruleus ENS	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
Z_albicolis ENS	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
J_hyemalis ENSJ	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
G_fortis ENSGFO	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
C_parvulus ENSC	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
P_major ENSPMJP	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
F_albicollis EN	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
Z_lateralis ENS	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
C_monedulaoides	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
M_gallopavo ENS	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
P_cristatus ENS	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
G_gallus ENSGAL	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
C_japonica ENSC	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
N_meleagris ENS	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
C_pugnax ENSCPU	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
C_pygmaea ENSCP	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
E_tinnunculus E	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
B_bubo ENSBOBP	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
A_chryseatus EN	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
O_sunia ENSOSUP	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
M_undulatus ENS	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
A_cunicularia E	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
S_habroptila EN	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
A_platyrynchos	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
A_zonorhyncha E	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
D_novahollandia	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
S_camelus ENSSC	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
A_owenii ENSAOW	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
A_haastii ENSAH	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
C_porosus ENSCP	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
N_perdicaria EN	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
L coronata ENSL	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
P_sinensis ENSP	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
C_picta ENSCPBP	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
G_agassizii ENS	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
C_pictus ENSCPI	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
C_jacchus ENSCJ	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
M_putorius ENSM	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
A_carolinensis	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	SRPDQ	TRMDIRLAKTLVLILVLLIIC	
P_muralis ENSPM	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPEQ	TRMDIRLAKTLVLILVLLIIC	
V_komodoensis E	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPEQ	TRMDIRLAKTLVLILVLLIIC	
P_vitticeps ENS	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	SRPEQ	TRMDIRLAKTLVLILVLLIIC	
L_chalumnae ENS	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPEQ	TRMDIRLAKTLVLILVLLIIC	
X_tropicalis EN	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPEQ	TRMDIRLAKTLVLILVLLIIC	
S_dauricus ENSS	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPDQ	ARMDIRLAKTLVLILVLLIIC	
S_trutta ENSSU	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPEQ	TRMDIRLAKTLVLILVLLIIC	
S_salar ENSSSAP	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPEQ	TRMDIRLAKTLVLILVLLIIC	
O_mykiss ENSSOMY	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPEQ	TRMDIRLAKTLVLILVLLIIC	
H_hucho ENSHHUP	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPEQ	TRMDIRLAKTLVLILVLLIIC	
O_tshawyscha E	YILWKAHSHAVRMIQR	GTQKSIITHT	SEDDGKVQV	TRPEQ	TRMDIRLAKTLVLILVLLIIC	



E\_lucius|ENSELU YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
T\_rubripes|ENST YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
T\_nigroviridis| YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
L\_bergylta|ENSL YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
A\_percula|ENSAP YILWRAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
S\_partitus|ENSS YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
P\_ranga|ENSRNP YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
L\_calcarifer|EN YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
L\_crocea|ENSLCR YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
S\_aurata|ENSSAU YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
D\_labrax|ENSDLA YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
S\_lucioerca|EN YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
O\_aureus|ENSOAB YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
M\_zebra|ENSMZEP YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
A\_citrinellus|E YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
S\_dumerili|ENSS YILWKAHHHVRMLSR TSQKSLVYYSADGT KVLT TRPEQ TRMDIRLAKTLVLILVVLVIC  
S\_lalandi|ENSSLD YILWKAHHHVRMLSR TSQKSLVYYSADGT KVLT TRPEQ TRMDIRLAKTLVLILVVLVIC  
M\_mola|ENSMOP0 YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
C\_lumpus|ENSCLM YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
M\_armatus|ENSMA YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
E\_naucrates|ENS YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
S\_fasciatus|ENS YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
M\_murdjan|ENSM YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
O\_melastigma|EN YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
O\_javanicus|ENS YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
K\_marmoratus|EN YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
P\_formosa|ENSPF YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
P\_reticulata|EN YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
X\_maculatus|ENS YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
G\_affinis|ENSGA YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
G\_affinis|ENSGA YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
C\_variegatus|EN YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
F\_heteroclitus| YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
M\_albus|ENSMALP YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
C\_harengus|ENSC YILWRAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
E\_electricus|EN YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
I\_punctatus|ENS YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
G\_morhua|ENSGMO YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
D\_rerio|ENDARP YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
C\_aureus|ENSCA YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
C\_carpio|ENSCCR YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
S\_formosus|ENSS YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
C\_gobio|ENSCGOF YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
P\_panicus|ENSP YILWKAHHHVRMLSR TSQKSLVYYSADGT KVQSTRPEQ TRMDIRLAKTLVLILVVLVIC  
P\_trogodytes|E HVLWKAHQHVASLSGH QDRQV . . . . . PGMARMRLDVRLAKTLGLVLAFLVIC  
G\_gorila|ENSGGO HVLWKAHQHVASLSGH QDRQV . . . . . PGMARMRLDVRLAKTLGLVLAFLVIC  
N\_leucogenys|EN HVLWKAHQHVASLSGH QDRQV . . . . . PGMARMRLDVRLAKTLGLVLAFLVIC  
P\_tephrosceles| HVLWKAHQHVASLSGH QDRQV . . . . . PGMARMRLDVRLAKTLGLVLAFLVIC  
C\_angolensis|EN HVLWKAHQHVASLSGH QDRQV . . . . . PGMARMRLDVRLAKTLGLVLAFLVIC  
R\_bieti|ENSRBIP HVLWKAHQHVASLSGH QDRQV . . . . . PGMARMRLDVRLAKTLGLVLAFLVIC  
C\_atys|ENSCATP0 HVLWKAHQHVASLSGH QDRQV . . . . . PGMARMRLDVRLAKTLGLVLAFLVIC  
M\_leucophaeus|E HVLWKAHQHVASLSGH QDRQV . . . . . PGMARMRLDVRLAKTLGLVLAFLVIC  
P\_anubis|ENSPAN HVLWKAHQHVASLSGH QDRQV . . . . . PGMARMRLDVRLAKTLGLVLAFLVIC  
T\_gelada|ENSTGE HVLWKAHQHVASLSGH QDRQV . . . . . PGMARMRLDVRLAKTLGLVLAFLVIC  
M\_nemestrina|EN HVLWKAHQHVASLSGH QDRQV . . . . . PGMARMRLDVRLAKTLGLVLAFLVIC  
P\_simus|ENSPSMP HVLWKAHQHVASLAEHRRHV . . . . . PGMARMRLDVRLAKTLGLVLAFLVIC  
C\_hoffmanni|ENS HVLWKAHQHVASLAEH QDRQV . . . . . PGLARMRLDVRLAKTLGVVLAFLVIC  
R\_ferrumequinum HVLWKAHQHVVSLAEHQGRV . . . . . PGMARMRLDVRLAKTLGVVLAFLVIC  
N\_vison|ENSNVIP HVLWKAHQHVASLAEHQDRLM . . . . . PGMVRLDVRLAKTLGVVLAFLVIC  
P\_leo|ENSPLOP0 HVLWKAHQHVASLAEHQDRQV . . . . . PGMARMRLDVRLAKTLGVVLAFLVIC  
P\_pardus|ENSPPR HVLWKAHQHVASLAEHQDRQV . . . . . PGMARMRLDVRLAKTLGVVLAFLVIC  
F\_catus|ENSFCAH HVLWKAHQHVASLAEHQDRQV . . . . . PGMARMRLDVRLAKTLGVVLAFLVIC  
L\_canadensis|EN HVLWKAHQHVASLAEHQDRQV . . . . . PGMARMRLDVRLAKTLGVVLAFLVIC  
S\_suricatta|ENS HVLWKAHQHVASLAEHQDSQV . . . . . PGMARMRLDVRLAKTLGVVLAFLVIC  
Z\_californianus HVLWKAHQHVVSLAEHQDRQV . . . . . PGMARMRLDVRLAKTLGVVLAFLVIC  
U\_americanus|EN HVLWKAHQHVASLAEHRDRQV . . . . . PGMARMRLDVRLAKTLGVVLAFLVIC  
P\_catodon|ENSPC HVLWKAHQHVASLAEHRDRQV . . . . . LGMAQMRDLVRLAKTLGVVLAFLVIC  
O\_cuniculus|ENS HVLWKAHQHVASLAEHQDQV . . . . . PGVAWMLDVRLAKTLGLVLAFLVIC  
S\_vulgaris|ENSS HVLWKAHQHVASLAEHQDRQV . . . . . PGMIRMRLDVRLAKTLGLLLAVLFMC  
M\_domestica|ENS YVLWKAHQHTSTLVYMQKRL . . . . . AGMSRMLDVRLAKTLGVVLMVLLVC  
V\_ursinus|ENSVU YVLWKAHYRHTSNLVYQRRLL . . . . . TGVSRIRLDVRLAKTLGVVLMVLLLC  
P\_cinereus|ENSP YVLWKAHYRHTSNLVYQRRLL . . . . . TGMSSRMLDVRLAKTLGVVLMVLLLC  
O\_anatinus|ENSO YILRLARRHAACLAGRRAQPG . . . . . PGQARLRLDVRLAKTLGLVLAFLVIC  
P\_textilis|ENSP HVLWKAHQHTANMKKHHLEES . . . . . QQNPKRRIIDIKLAKTLISVLMVLLVIC  
N\_scutatus|ENSN HVLWKAHQHTANMKKHHLEES . . . . . QQNPKRRIIDIKLAKTLISVLMVLLVIC  
N\_naja|ENSNNAPO HVLWKAHQHTANMKKHHLEES . . . . . QQNPKRRIIDIKLAKTLISVLMVLLVIC  
P\_castaneus|ENS KVLWKAHKHTVYMEKHQSQAQ . . . . . RQNAARMRDVTLTKTLVMVLAFLVIC  
P\_castaneus|ENS KVLWKAHKHTVYMEKHQSQAQ . . . . . RQNAARMRDVTLTKTLVMVLAFLVIC  
C\_serpentina|EN RVLWKAHKHTVYMEKRTQAG . . . . . QQNAKMRMDVTLAKTLVMVLAFLVIC  
T\_carolina|ENST RVLWKAHKHTVYMEKRTQAG . . . . . QRNAKMRMDVTLAKTLVMVLAFLVIC  
C\_abingsdonii|E QVLWKAHKHTVYMEKHQTQAG . . . . . QRNAKMRMDITLAKTLVMVLAFLVIC  
S\_punctatus|ENS HVLYKAHKHVAYMQTH . . . . . QQKAKMRMDVTLTKTLMLVLAFLVIC  
B\_japonicus|ENS HVLWRAHQHVAYMEKHQAQAG . . . . . KQNTMRMDVLAFLTKTLVMVLAFLVIC  
C\_ustulatus|ENS HVLWRAHQHVAYMEKHQAQAG . . . . . KQNTMRMDVLAFLTKTLVMVLAFLVIC  
M\_cyaneus|ENSMC HVLWRAHQHVAYMEKHQAQAG . . . . . KQNSRMRMDVLAFLTKTLVMVLAFLVIC  
A\_mexicanus|ENS LILWKAHQHAEAMEAPTAGRSTM . . . . . GRHARMRVDIHLAQTLGLILMLVLC  
S\_syrichtha|ENST QILVMRHAHQIALQHHFLATS . . . . . HVTTRKGVSTLAIILGTFAAC  
U\_maritimus|ENS RILVTRRSRRLTFRKNI . . . . . RLSSEKSLALLKTVIIVLSVFIAC  
P\_tigris|ENSPTI RILVTRRSRRLTFRKNI . . . . . RLSSEKSLALLKTVIIVLSVFIAC  
P\_coquereli|ENS RILVTRRSRRLTFRKNI . . . . . RLSSEKSLALLKTVIIVLSVFIAC  
B\_mutus|ENSMUP RILVTRRSRRLTFRKNI . . . . . RLSSEKSLALLKTVIIVLSVFIAC  
C\_capucinus|ENS RILVTRRSRRLTFRKNI . . . . . RLSSEKSLALLKTVIIVLSVFIAC  
S\_boliviensis|E RILVTRRSRRLTFRKNI . . . . . RLSSEKSLALLKTVIIVLSVFIAC  
A\_collaria|ENSA RIMVTRRSRRLTFRKNI . . . . . RLSSEKSLALLKTVIIVLSVFIAC  
A\_nancymaae|ENS HIYVRQRTMRMS . . . . . RNRDMMMSLKTVVIVLGAFLVIC  
M\_murinus|ENSMI HIYVRQRTMRMS . . . . . RNRDMMMSLKTVVIVLGAFLVIC

	360	370	380	390	400	410
H_sapiens ENSP0	WG	KL	IK	IV	FA	CS
M_mulatta ENSM0	WG	KL	IK	IV	FA	CS
M_abelii ENSPY	WG	KL	IK	IV	FA	CS
M_fascicularis	WG	KL	IK	IV	FA	CS
C_sabaeus ENSCS	WG	KL	IK	IV	FA	CS
R_roxellana ENS	WG	KL	IK	IV	FA	CS
O_aries ENSOAR	WG	KL	IK	IV	FA	CS
M_moschiferus E	WG	KL	IK	IV	FA	CS
T_belangeri ENS	WG	KL	IK	IV	FA	CS
D_novemcinctus	WG	KL	IK	IV	FA	CS
C_dromedarius E	WG	KL	IK	IV	FA	CS
V_vulpes ENSVV	WG	KL	IK	IV	FA	CS
L_africana ENSL	WG	KL	IK	IV	FA	CS
E_asinus ENSEAS	WG	KL	IK	IV	FA	CS
S_scrofa ENSSC	WG	KL	IK	IV	FA	CS
A_melanoleuca E	WG	KL	IK	IV	FA	CS
B_taurus ENSBTA	WG	KL	IK	IV	FA	CS
C_hircus ENSCHI	WG	KL	IK	IV	FA	CS
T_truncatus ENS	WG	KL	IK	IV	FA	CS
D_leucas ENSDLE	WG	KL	IK	IV	FA	CS
P_vampyrus ENSP	WG	KL	IK	IV	FA	CS
V_pacos ENSVPA	WG	KL	IK	IV	FA	CS
I_tridecemlinea	WG	KL	IK	IV	FA	CS
U_parryii ENSPU	WG	KL	IK	IV	FA	CS
M_marmota ENSM	WG	KL	IK	IV	FA	CS
H_glaber ENSHGL	WG	KL	IK	IV	FA	CS
C_porcellus ENS	WG	KL	IK	IV	FA	CS
C_aperea ENSCAP	WG	KL	IK	IV	FA	CS
F_damarensis EN	WG	KL	IK	IV	FA	CS
C_lanigera ENSC	WG	KL	IK	IV	FA	CS
O_degus ENSODEP	WG	KL	IK	IV	FA	CS
M_auratus ENSM	WG	KL	IK	IV	FA	CS
M_unquiculatus	WG	KL	IK	IV	FA	CS
M_ochrogaster E	WG	KL	IK	IV	FA	CS
M_musculus ENSM	WG	KL	IK	IV	FA	CS
M_spicilegus EN	WG	KL	IK	IV	FA	CS
M_caroli MGP_CA	WG	KL	IK	IV	FA	CS
M_spretus MGP_S	WG	KL	IK	IV	FA	CS
R_norvegicus EN	WG	KL	IK	IV	FA	CS
M_pahari MGP_Pa	WG	KL	IK	IV	FA	CS
C_canadensis EN	WG	KL	IK	IV	FA	CS
S_araneus ENSSA	WG	KL	IK	IV	FA	CS
C_lupus ENSCAFP	WG	KL	IK	IV	FA	CS
P_capensis ENSP	WG	KL	IK	IV	FA	CS
E_europaeus ENS	WG	KL	IK	IV	FA	CS
O_garnettii ENS	WG	KL	IK	IV	FA	CS
D_ordii ENSODOP	WG	KL	IK	IV	FA	CS
J_jaculus ENSJJ	WG	KL	IK	IV	FA	CS
S_harrisii ENSS	WG	KL	IK	IV	FA	CS
N_eugenii ENSME	WG	KL	IK	IV	FA	CS
S_canaria ENSSC	WG	KL	IK	IV	FA	CS
L_striata ENSL	WG	KL	IK	IV	FA	CS
T_guttata ENSTG	WG	KL	IK	IV	FA	CS
C caeruleus ENS	WG	KL	IK	IV	FA	CS
Z_albicolis ENS	WG	KL	IK	IV	FA	CS
H_hyemalis ENSJ	WG	KL	IK	IV	FA	CS
G_fortis ENSGFO	WG	KL	IK	IV	FA	CS
C_parvulus ENSC	WG	KL	IK	IV	FA	CS
P_major ENSPMPJ	WG	KL	IK	IV	FA	CS
F_albicollis EN	WG	KL	IK	IV	FA	CS
Z_lateralis ENS	WG	KL	IK	IV	FA	CS
C_moneduloides	WG	KL	IK	IV	FA	CS
M_gallopavo ENS	WG	KL	IK	IV	FA	CS
P_cristatus ENS	WG	KL	IK	IV	FA	CS
G_gallus ENSGAL	WG	KL	IK	IV	FA	CS
C_japonica ENSC	WG	KL	IK	IV	FA	CS
N_meleagris ENS	WG	KL	IK	IV	FA	CS
C_pugnax ENSCPU	WG	KL	IK	IV	FA	CS
C_pygmaea ENSCP	WG	KL	IK	IV	FA	CS
E_tinnunculus E	WG	KL	IK	IV	FA	CS
B_bubo ENSBOP	WG	KL	IK	IV	FA	CS
A_chryseatus EN	WG	KL	IK	IV	FA	CS
O_sunia ENSOSUP	WG	KL	IK	IV	FA	CS
M_undulatus ENS	WG	KL	IK	IV	FA	CS
A_cunicularia E	WG	KL	IK	IV	FA	CS
S_habroptila EN	WG	KL	IK	IV	FA	CS
A_platyrrhynchos	WG	KL	IK	IV	FA	CS
A_zonorhyncha E	WG	KL	IK	IV	FA	CS
D_novahollandia	WG	KL	IK	IV	FA	CS
S_camelus ENSSC	WG	KL	IK	IV	FA	CS
A_owenii ENSAOW	WG	KL	IK	IV	FA	CS
A_haastii ENSAH	WG	KL	IK	IV	FA	CS
C_porosus ENSCP	WG	KL	IK	IV	FA	CS
N_perdicaria EN	WG	KL	IK	IV	FA	CS
L_coronata ENSL	WG	KL	IK	IV	FA	CS
P_sinensis ENSP	WG	KL	IK	IV	FA	CS
C_picta ENSCPBP	WG	KL	IK	IV	FA	CS
G_agassizii ENS	WG	KL	IK	IV	FA	CS
C_pictus ENSCPI	WG	KL	IK	IV	FA	CS
C_jacchus ENSCJ	WG	KL	IK	IV	FA	CS
M_putorius ENSM	WG	KL	IK	IV	FA	CS
A_carolinensis	WG	KL	IK	IV	FA	CS
P_muralis ENSPM	WG	KL	IK	IV	FA	CS
V_komodoensis E	WG	KL	IK	IV	FA	CS
P_vitticeps ENS	WG	KL	IK	IV	FA	CS
L_chalumnae ENS	WG	KL	IK	IV	FA	CS
X_tropicalis EN	WG	KL	IK	IV	FA	CS
S_dauricus ENSS	WG	KL	IK	IV	FA	CS
S_trutta ENSSTU	WG	KL	IK	IV	FA	CS
S_salar ENSSSAP	WG	KL	IK	IV	FA	CS
O_mykiss ENSOXY	WG	KL	IK	IV	FA	CS
H_hucho ENSHHUP	WG	KL	IK	IV	FA	CS
O_tshawyscha E	WG	KL	IK	IV	FA	CS

E\_lucius|ENSELU WGPVLAIMVYDLF WKMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF MS SQOAC  
T\_rubripes|ENST WGPVLAIMVYDLF WKMDDN IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
T\_nigroviridis| WGPVLAIMVYDLF WKMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
L\_bergylla|ENSL WGPVLAIMVYDLF WRMDDN IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
A\_percula|ENSAP WGPVLAIMVYDLF WRMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
S\_partitus|ENSS WGPVLAIMVYDLF WRMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
P\_ranga|ENSRNP WGPVLAIMVYDLF WKMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
L\_calcarifer|EN WGPVLAIMVYDLF WKMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
L\_crocea|ENSLCR WGPVLAIMVYDLF WKMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
S\_aurata|ENSSAU WGPVLAIMVYDLF WRMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
D\_labrax|ENSDLA WGPVLAIMVYDLF WRMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
S\_lucioerca|EN WGPVLAIMVYDLF WKMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
O\_aureus|ENSOAB WGPVLAIMVYDLF WRMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
M\_zebra|ENSMZEP WGPVLAIMVYDLF WRMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
A\_citrinellus|E WGPVLAIMVYDLF WKMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
S\_dumerili|ENSS WGPVLAIMVYDLF WKMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
S\_lalandi|ENSSLD WGPVLAIMVYDLF WKMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
M\_mola|ENSMOP0 WGPVLAIMVYDLF WKMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
C\_lumpus|ENSCLM WGPVLAIMVYDLF WRMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
M\_armatus|ENSM WGPVLAIMVYDLF WRMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
E\_naucrates|ENS WGPVLAIMVYDLF WRMDDN IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
S\_fasciatus|ENS WGPVLAIMVYDLF WRMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
M\_murdjan|ENSM WGPVLAIMVYDLF WKMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
O\_melastigma|EN WGPVLAIMVYDLF WRMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
O\_javanicus|ENS WGPVLAIMVYDLF WRMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
K\_marmoratus|EN WGPVLAIMVYDLF WRMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
P\_formosa|ENSPF WGPVLAIMVYDLF WKMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
P\_reticulata|EN WGPVLAIMVYDLF WKMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
X\_maculatus|ENS WGPVLAIMVYDLF WKMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
G\_affinis|ENSGA WGPVLAIMVYDLF WKMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
G\_affinis|ENSGA WGPVLAIMVYDLF WKMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
C\_variegatus|EN WGPVLAIMVYDLF WKMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
F\_heteroclitus| WGPVLAIMVYDLF WKMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
M\_albus|ENSMALP WGPVLAIMVYDLF WKMDDN IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
C\_harengus|ENSC WGPVLAIMVYDLF WKMDDA IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
E\_electricus|EN WGPVLAIMVYDLF WKMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
I\_punctatus|ENS WGPVLAIMVYDLF WKMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
G\_morhua|ENSGMO WGPVLAIMVYDLF WKMDDD IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
D\_rerio|ENDARP WGPVLAIMVYDLF WRMDDN IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
C\_auratus|ENSCA WGPVLAIMVYDLF WRMDDN IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
C\_carpio|ENSCCR WGPVLAIMVYDLF WRMDDN IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
S\_formosus|ENSS WGPVLAIMVYDLF WKMDDN IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
C\_gobio|ENSCGOF WGPVLAIMVYDLF WKMDDN IKTVFACMCLLNSTVNP IYALRSKDLRHAF LS SCHA  
P\_panicus|ENSP WFPVLAALMAHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
P\_troglodytes|E WFPVLAALMAHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
G\_gorila|ENSGGO WFPVLAALMAHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
N\_leucogenys|EN WFPVLAALMAHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
P\_tephrosceles| WFPVLAALMVHSLA T L S D E V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
C\_angolensis|EN WFPVLAALMVHSLA T L S D E V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
R\_bieti|ENSRBIP WFPVLAALMVHSLA T L S D E V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
C\_atys|ENSCATP WFPVLAALMAHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
M\_leucophaeus|E WFPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
P\_anubis|ENSPAN WFPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
T\_gelada|ENSTGE WFPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
M\_nemestrina|EN WFPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
P\_simus|ENSPSMP WFPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
C\_hoffmanni|ENS WFPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
R\_ferrumequinum WFPVLAALMAYSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
N\_vision|ENSNVIP WLPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
P\_leo|ENSPLOPO WFPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
P\_pardus|ENSPPR WFPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
F\_catus|ENSFCA WFPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
L\_canadensis|EN WFPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
S\_suricatta|ENS WFPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
Z\_californianus WLPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
U\_americanus|EN WLPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
P\_catonodon|ENSPC WLPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
O\_cuniculus|ENS WFPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
S\_vulgaris|ENSS WFPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
M\_domestica|ENS WLPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
V\_ursinus|ENSVU WLPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
P\_cinereus|ENSP WLPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
O\_anatinus|ENSO WSPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
P\_textilis|ENSP WLPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
N\_scutatus|ENSN WLPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
N\_naja|ENSNNAPO WLPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
P\_castaneus|ENS WFPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
P\_castaneus|ENS WFPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
C\_serpentina|EN WSPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
T\_carolina|ENST WSPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
C\_abingsdonii|E WSPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
S\_punctatus|ENS WSPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
B\_japonicus|ENS WSPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
C\_ustulatus|ENS WSPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
M\_cyaneus|ENSMC WSPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
A\_mexicanus|ENS WWPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
C\_syrichtha|ENST WWPVLAALMVHSLA T L S D Q V K K A F A F C M S L C L I N S M V N P V I Y A L R S G E I R S S A H H C L A H W  
U\_maritimus|ENS WAPLFIALLLDVGC KVKT . . . . . CDILAVLNSGTNP IYTLTKEMRRRAF IRLISC  
P\_tigris|ENSPTI WAPLFIALLLDVGC KVKT . . . . . CDILAVLNSGTNP IYTLTKEMRRRAF IRLISC  
P\_coquereli|ENS WAPLFIALLLDVGC KVKT . . . . . CDILAVLNSGTNP IYTLTKEMRRRAF IRLISC  
B\_mutus|ENSMUP WAPLFIALLLDVGC KVKT . . . . . CDILAVLNSGTNP IYTLTKEMRRRAF IRLISC  
C\_capucinus|ENS WAPLFIALLLDVGC KVKT . . . . . CDILAVLNSGTNP IYTLTKEMRRRAF IRLISC  
S\_boliviensis|E WAPLFIALLLDVGC KVKT . . . . . CDILAVLNSGTNP IYTLTKEMRRRAF IRLISC  
A\_collaria|ENSA WAPLFIALLLDVGC KVKS . . . . . CDILAVLNSGTNP IYTLTKEMRRRAF IRLISC  
A\_nancymaae|ENS WTPGLVLLLDVGC CPO . . . . . CNVLAEFNSAMNP IYSYRDKEMSATFR QILCC  
M\_murinus|ENSMI WTPGLVLLLDVGC CPO . . . . . CDVLAEFNSAMNP IYSYRDKEMSATFR QILCC

420 430 440 450 460

H_sapiens ENSP0	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
M_mulatta ENSM0	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
P_abelii ENSPPY	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
M_fascicularis	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
C_sabaeus ENSSC	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
R_roxellana ENSC	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
O_aries ENSOARP	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
M_moschiferus E	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
T_belangeri ENS	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
D_novemcinctus	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
C_dromedarius E	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
V_vulpes ENSVVU	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
L_africana ENSVL	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
E_asinus ENSEAS	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
S_scrofa ENSSSC	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
A_melanoleuca E	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
B_taurus ENSBTA	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
C_hircus ENSCHE	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
T_truncatus ENS	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
D_leucas ENSDLE	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
P_vampyrus ENSP	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
V_pacos ENSPVAP	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
I_tridecemlinea	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
U_parryii ENSPU	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
M_marmota ENSM	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
H_glaber ENSHGL	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
C_porcellus ENS	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
C_aperea ENSCAP	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
F_damarensis EN	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
C_lanigera ENSC	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
O_degus ENSODEP	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
M_auratus ENSMA	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
M_unquiculatus	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
M_ochrogaster E	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
M_musculus ENSM	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
M_spicilegus EN	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
M_caroli MGP_CA	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
M_spretus MGP_S	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
R_norvegicus EN	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
M_pahari MGP_Pa	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
C_canadensis EN	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
S_araneus ENSSA	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
C_lupus ENSCAFP	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVPKTS	SDAY
F_capensis ENSP	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
E_europaeus ENS	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
O_garnettii ENS	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
D_ordii ENSODRP	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
J_jaculus ENSJJ	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
S_harrisii ENSS	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
N_eugenii ENSME	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
S_canaria ENSSC	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
L_striata ENSL	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
T_guttata ENSTG	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
C_caeruleus ENS	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
Z_albicolis ENS	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
J_hyemalis ENSJ	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
G_fortis ENSGFO	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
C_parvulus ENSC	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
P_major ENSPMPJ	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
F_albicollis EN	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
Z_lateralis ENS	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
C_moneduloides	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
M_gallopavo ENS	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
P_cristatus ENS	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
G_gallus ENSGAL	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
C_japonica ENSC	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
N_meleagris ENS	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
C_pugnax ENSCPU	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
C_pygmaea ENSCP	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
F_tinnunculus E	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
B_bubo ENSOBBO	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
A_chryseatus EN	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
O_sunia ENSOUP	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
M_undulatus ENS	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
A_cunicularia E	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
S_habroptila EN	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
A_platyrhynchos	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
A_zonorhyncha E	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
D_novahollandia	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
S_camelus ENSSC	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
A_owenii ENSAOW	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
A_haastii ENSAH	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
C_porusus ENSCP	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
N_perdicaria EN	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
L_coronata ENSL	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
P_sinensis ENSP	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
G_picta ENSPBP	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
G_agassizii ENS	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
C_pictus ENSCPI	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
C_jacchus ENSCJ	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
M_putorius ENSM	EG	TAQ	... P	LDNSMDSDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
A_carolinensis	VG	TAQ	... P	LDNSLESDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
P_muralis ENSPM	NG	TAQ	... P	LDNSLESDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
V_komodoensis E	EG	TAQ	... P	LDNSLESDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
P_vitticeps E	AG	TAQ	... P	LDNSLESDC	LHK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
L_chalunnae ENS	KG	TAL	... P	LDNSMESDC	QNK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
X_tropicalis EN	EG	TAQ	... P	LDNSMESDC	QNK	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
S_dauricus ENSS	EAKKG	... RV	QGHISR	P	LNFI	Y	I	HAHL	...	AVNT	Q...
S_trutta ENSSTU	RG	SAQ	... Q	LDNSLESDC	QSR	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
S_salar ENSSSAP	RG	SAQ	... Q	LDNSLESDC	QSR	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
O_mykiss ENSSOMY	RG	SAQ	... Q	LDNSLESDC	QSR	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
H_hucho ENSHHUP	RG	SAQ	... Q	LDNSLESDC	QSR	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS
O_tshawyscha E	RG	SAQ	... Q	LDNSLESDC	QSR	.....	HANNAAS	VHRAAE	ESC	IKSTVKIAKVTMSVST	DTTS

E_lucius ENSELU	RG	SAQ	LDNSLESDC	QNR	HIT	SHRAAES	SCVKTTVKIAKVTMSVST	ETS
T_rubripes ENST	RG	SAQ	LDNSLESDC	QNR	NVNIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
L_nigroviridis	RG	SSQ	LDNSLESDC	QNR	NVNIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
L_bergylta ENSL	RG	SAQ	LDNSLESDC	QNR	NANIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
A_percula ENSAP	RG	SAQ	LDNSLESDC	QNR	NANIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
S_partitus ENSS	RG	SAQ	LDNSLESDC	QNR	NANIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
P_ranga ENSPRNP	RG	SAQ	LDNSLESDC	QNR	NANIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
L_calcarifer EN	RG	SAQ	LDNSLESDC	QNR	HANIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
L_crocea ENSLCR	RG	SAQ	LDNSLESDC	QNR	NANIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
S_aurata ENSSAU	RG	SAQ	LDNSLESDC	QNR	NANIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
D_labrax ENSDLA	RG	SAQ	LDNSLESDC	QNR	NANIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
S_lucioerca EN	RG	SAQ	LDNSLESDC	QNR	NANIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
O_aureus ENSOAB	RG	SGQ	LDNSLESDC	QNR	NTHIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
M_zebra ENSMZEP	RG	SGQ	LDNSLESDC	QNR	NTHIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
A_citrinellus E	RG	SAQ	LDNSLESDC	QNR	NIHIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
S_dumerili ENSS	RG	SAQ	LDNSLESDC	QNR	NANIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
S_lalandi ENSSLD	RG	SAQ	LDNSLESDC	QNR	NANIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
M_mola ENSMOP0	RG	SAQ	LDNSLESDC	QNR	NVNIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
C_lumpus ENSCLM	RG	SVQ	LDNSLESDC	QNR	NANIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
M_armatus ENSM	RG	SAQ	LDNSLESDC	QNR	NANIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
E_naucrates ENS	RG	SAQ	LDNSLESDC	QNR	NTHIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
S_fasciatus ENS	RG	SAQ	LDNSLESDC	QNR	NANIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
M_murdjan ENSM	RG	SAQ	LDNSLESDC	QNR	NVTIA	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
O_melastigma EN	RG	SGQ	LDNSLESDC	QNR	NIHIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
O_javanicus ENS	RG	SGQ	LDNSLESDC	QNR	NIHIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
K_marmoratus EN	RS	SSQ	LDNSLESDC	PNR	NAHIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
P_formosa ENSPF	RS	SSQ	LDNSLESDC	QNR	NANIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
P_reticulata EN	RS	SSQ	LDNSLESDC	QNR	NANIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
X_maculatus ENS	RS	SSQ	LDNSLESDC	QNR	NANIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
G_affinis ENSGA	QSTQ	LDNSLESDC	QNR	NANIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS	
G_affinis ENSGA	QSTQ	LDNSLESDC	QNR	NANIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS	
C_variegatus EN	RS	SAQ	LDNSLESDC	QNR	HTNIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
F_heteroclitus	RS	SAQ	LDNSLESDC	HNR	NATIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
M_albus ENSMALP	RS	SAL	LDNSLESDC	QNR	NANIS	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
C_harengus ENSC	RG	SSH	LDNSLESDY	QSR	HAHIA	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
E_electricus EN	RG	SSQ	LDNSLESDC	QNR	NLHIH	ANRAAES	SCVKTTVKIAKVTMSVST	ETS
I_punctatus ENS	RS	SSQ	LDNSLESDC	QNR	ILQIH	AKRAAES	SCVKTTVKIAKVTMSVST	ETS
G_morhua ENSGMO	RG	TGQ	LDNSLESDC	QSR	NTHLA	AHRAAES	SCVKTTVKIAKVTMSVST	ETS
D_rerio ENDARP	RG	TSL	LDNSLESDC	H.R	NQH	RAAES	SCVKTTVKIAKVTMSVST	ETS
C_auratus ENSCA	RG	TSL	LDNSLESDC	Q.R	NQHIA	AKRAAES	SCVKTTVKIAKVTMSVST	ETS
C_carpio ENSCCR	RS	TSL	LDNSLESDC	Q.R	NQHIA	AHRAAES	SCVKTTVKIAKVTMSVST	ETS
S_formosus ENSS	TGASQ	P.LDNSLESDY	QSR	N		ANKPGES	SCVKTTVKIAKVTMSVST	ETS
C_gobio ENSCGOF	RVATQ	P.LDNSAESDWSN		SVRGTA	AGRAGKD	CGKTRVKVAQVTVSVT	ETS	
P_paniscus ENSP	KKCVR	GLGSEAK.EEAPR		S			SVT.ETE	
P_trogodytes E	KKCVR	GLGSEAK.EEAPR		S			SVT.ETE	
G_gorila ENSGGO	KKCVR	GLGSEAK.EEAPR		S			SVT.ETE	
N_leucogenys EN	KKRVR	GLGSEAK.EEAPR		S			SVT.ETE	
P_tephrosceles	RKCVR	GLGSEAK.EEAPR		S			SVT.ETE	
C_angolensis EN	RKCVR	GLGSEAK.EEAPR		S			SVT.ETE	
R_bieti ENSRBIP	RKCVR	GLGSEAK.EEAPR		S			SVT.ETE	
C_atys ENSCATPO	RKCVR	GLGSEAK.EEAPR		S			SVT.ETE	
M_leucophaeus E	RKCVR	GLGSEAK.EEAPR		S			SVT.ETE	
P_anubis ENSPAN	RKCVR	GLGSEAK.EEAPR		S			SVT.ETE	
T_gelada ENSTGE	RKCVR	GLGSEAK.EEAPR		S			SVT.ETE	
M_nemestrina EN	RKCVR	GLGSEAK.EEAPR		S			SVT.ETE	
P_simus ENSPSMP	RRCLR	DIGPEGK.NEAPR		S			SVT.ETE	
C_hoffmanni ENS	KKDLQ	GYGHEGK.EEALR		T			SVT.ETE	
R_ferrumequinum	KKHLR	GLGPEGK.EDVPR		S			SVS.ETE	
N_vison ENSNVIP	RKRLR	RLGLEGN.KEVPR		S			SVT.ETE	
P_leo ENSPLOPO	RKHLR	RRGLEGS.KEVPR		S			SVT.ETE	
P_pardus ENSPPR	RKHLR	RRGLEGS.KEVPR		S			SVT.ETE	
F_catus ENSFCAF	RKHLR	RRGLEGS.KEVPR		S			SVT.ETE	
L_canadensis EN	RKHLR	RRGLEGS.KEVPR		S			SVT.ETE	
S_suricatta ENS	RKHLR	RRGLEGN.KEVPR		S			SVT.ETE	
Z_californianus	RKHLR	RLGLEGN.KEVPR		S			SVT.ETE	
U_americanus EN	RKHLR	RRLRLEGN.KEVPR		S			SVT.ETE	
P_catodon ENSPC	QKCLR	GGPEGK.EVPR		S			SVT.ETE	
O_cuniculus ENS	KKRLQ	GLGREEK.EEAPR		S			RSS.ISE	
S_vulgaris ENSS	KSYLR	GLGPEGR.EVPR		S			SVT.ETE	
M_domestica ENS	KNMLR	GTRTDSM.EDAQK		S			SVT.ETE	
V_ursinus ENSVU	KNMLR	GIGAEGT.EDTQK		S			SVT.ETE	
P_cinereus ENSP	KNMLR	GTGAEGA.EDTQK		S			SVT.ETE	
O_anatinus ENSO	RR.	GPQEPDEARK		T			SAT.ETE	
P_textilis ENSP	GKVLG	NSESNQEAESS.						
N_scutatus ENSN	GKVLG	NSESNQEAERPWG		S			SRVCVCV	
N_naja ENSNNAPO	GKVLG	NSESNQEAESS.						
P_castaneus ENS	RKKLR	SSESPAETDSTQK		A			LPI.KTT	
P_castaneus ENS	RKKLR	SSESPAETDSTQK		A			LPI.KTT	
C_serpentina EN	RKKMS	IESFPEDSTQK		S			SPI.ETI	
T_carolina ENST	RKKVS	IESSPEDSTQK		S			SPI.ETI	
C_abingsdonii E	RKKVS	IEITPEARDT.						
S_punctatus ENS	RRKLT	WLGISASNPEADSTQK		S			STL.GTI	
B_japonicus ENS	RRQLK	ASEESPEAESTRCK		S			SVT.ETV	
C_ustulatus ENS	RKQLK	VTEESPEAESTHCK		S			SVT.ETV	
M_cyaneus ENSMC	RRQLK	VSEESPEAESTHCK		S			SVT.ETV	
A_mexicanus ENS	CGVCK	PVCVCLNKDGDVK		A			ETV	
C_syrichtha ENST	IP	SLL		AQ			ARSPSD.	
U_maritimus ENS	KCPSR	PIIAGME.FSRSK		SDNSSH	PQKDDGDN		PETIMSSV.NSS	
P_tigris ENSPTI	KCPSR	PIIAGME.FSRSK		SDNSSH	PQKDDGDN		PETIMSSV.NSS	
P_coquereli ENS	KCPSR	PIVAGVE.FSRSK		SDNSSH	PQKDEADN		PETIMSSV.NSS	
B_mutus ENSBMUP	KCPSR	PIIAGME.FSRSK		SDNSSH	PQKDDGDN		PETIMSSV.NSS	
C_capucinus ENS	KCPSR	PIIAGME.FSRSK		SDNSSH	PQKDDGDN		PETIMSSV.NSS	
S_bolivianensis E	KCPSR	PIIAGME.FSRSK		SDNSSH	PQKDDGDN		PETIMSSV.NSS	
A_collaria ENSA	KCPFR	PIIGGME.FSRSK		SDNSSH	PQKEGDH		PETIMSSV.TSS	
A_nancymaae ENS				Q			PTEGSDRSASS.LNHTILAGV.HSN	
M_murinus ENSMI				Q			SENPSG.PTEGSDRSASS.LNHTILAGV.HSN	

H\_sapiens|ENSP0 AEAL  
M\_mulatta|ENSM0 AEAL  
P\_abelii|ENSPPY AEAL  
M\_fascicularis| AEAL  
C\_sabaeus|ENSCS AEAL  
R\_roxellana|ENS AEAL  
O\_aries|ENSOARF AEAL  
M\_moschiferus|E AEAL  
T\_belangeri|ENS AEAL  
D\_novemcinctus| AEAL  
C\_dromedarius|E AEAL  
V\_vulpes|ENSVVU AEAL  
L\_africana|ENSL AEAL  
E\_asinus|ENSEAS AEAL  
S\_scrofa|ENSSSC AEAL  
A\_melanoleuca|E AEAL  
B\_taurus|ENSBTA AEAL  
C\_hircus|ENSCHI AEAL  
T\_truncatus|ENS AEAL  
D\_leucas|ENSDLE AEAL  
P\_vampyrus|ENSP . . . .  
V\_pacos|ENSVPAP AEAL  
I\_tridecemlinea AEAL  
U\_parryii|ENSUP AEAL  
M\_marmota|ENSM0 AEAL  
H\_glaber|ENSHGL AEAL  
C\_porcellus|ENS AEAL  
C\_aperea|ENSCAP AEAL  
F\_damarensis|EN AEAL  
C\_lanigera|ENSC AEAL  
O\_degus|ENSODEP AEAL  
M\_auratus|ENSMA AEAL  
M\_unguiculatus| AEAL  
M\_ochrogaster|E AEAL  
M\_musculus|ENSM AEAL  
M\_spicilegus|EN AEAL  
M\_caroli|MGP\_CA AEAL  
M\_spretus|MGP\_S AEAL  
R\_norvegicus|EN AEAL  
M\_pahari|MGP\_Pa AEAL  
C\_canadensis|EN AEAL  
S\_araneus|ENSSA AEAL  
C\_lupus|ENSCAFP ASST  
P\_capensis|ENSP AEAL  
E\_europaeus|ENS AEAL  
O\_garnettii|ENS . . . .  
D\_ordii|ENSDORP AEAL  
J\_jaculus|ENSJJ AEAL  
S\_harrisii|ENSS AEAL  
N\_eugenii|ENSME AEAL  
S\_canaria|ENSSC AEAL  
L\_striata|ENSLS AEAL  
T\_guttata|ENSTG AEAL  
C\_caeruleus|ENS AEAL  
Z\_albicolis|ENS AEAL  
J\_hyemalis|ENSJ AEAL  
G\_fortis|ENSGFO AEAL  
C\_parvulus|ENSC AEAL  
P\_major|ENSPMJP AEAL  
F\_albicollis|EN AEAL  
Z\_lateralis|ENS AEAL  
C\_moneduloides| AEAL  
M\_gallopavo|ENS AEAL  
P\_cristatus|ENS AEAL  
G\_gallus|ENSGAL AEAL  
C\_japonica|ENSC AEAL  
N\_meleagris|ENS AEAL  
C\_pugnax|ENSCPU AEAL  
C\_pygmaea|ENSCP AEAL  
F\_tinnunculus|E AEAL  
B\_bubo|ENSOBPO AEAL  
A\_chryseatos|EN AEAL  
O\_sunia|ENSOSUP AEAL  
M\_undulatus|ENS AEAL  
A\_cunicularia|E AEAL  
S\_habroptila|EN AEAL  
A\_platyrrhynchus AEAL  
A\_zonorhyncha|E AEAL  
D\_novahollandia AEAL  
S\_camelus|ENSSC AEAL  
A\_owenii|ENSAOW AEAL  
A\_haastii|ENSAH AEAL  
C\_porosus|ENSCP AEAL  
N\_perdicaria|EN AEAL  
L\_coronata|ENSL AEAL  
P\_sinensis|ENSP AEAL  
C\_picta|ENSCPBP AEAL  
G\_agassizii|ENS AEAL  
C\_pictus|ENSCPI AEAL  
C\_jacchus|ENSCJ AEAL  
M\_putorius|ENSM AEAL  
A\_carolinensis| AEAL  
P\_muralis|ENSPM AEAL  
V\_komodoensis|E AEAL  
P\_vitticeps|ENS . . E I  
L\_chalumnae|ENS AEAL  
X\_tropicalis|EN AEAV  
S\_dauricus|ENSS . . . .  
S\_trutta|ENSSTU AEAV  
S\_salar|ENSSSAP AEAV  
O\_mykiss|ENSSOMY AEAV  
H\_hucho|ENSSHUP AEAV  
O\_tshawytscha|E AEAV



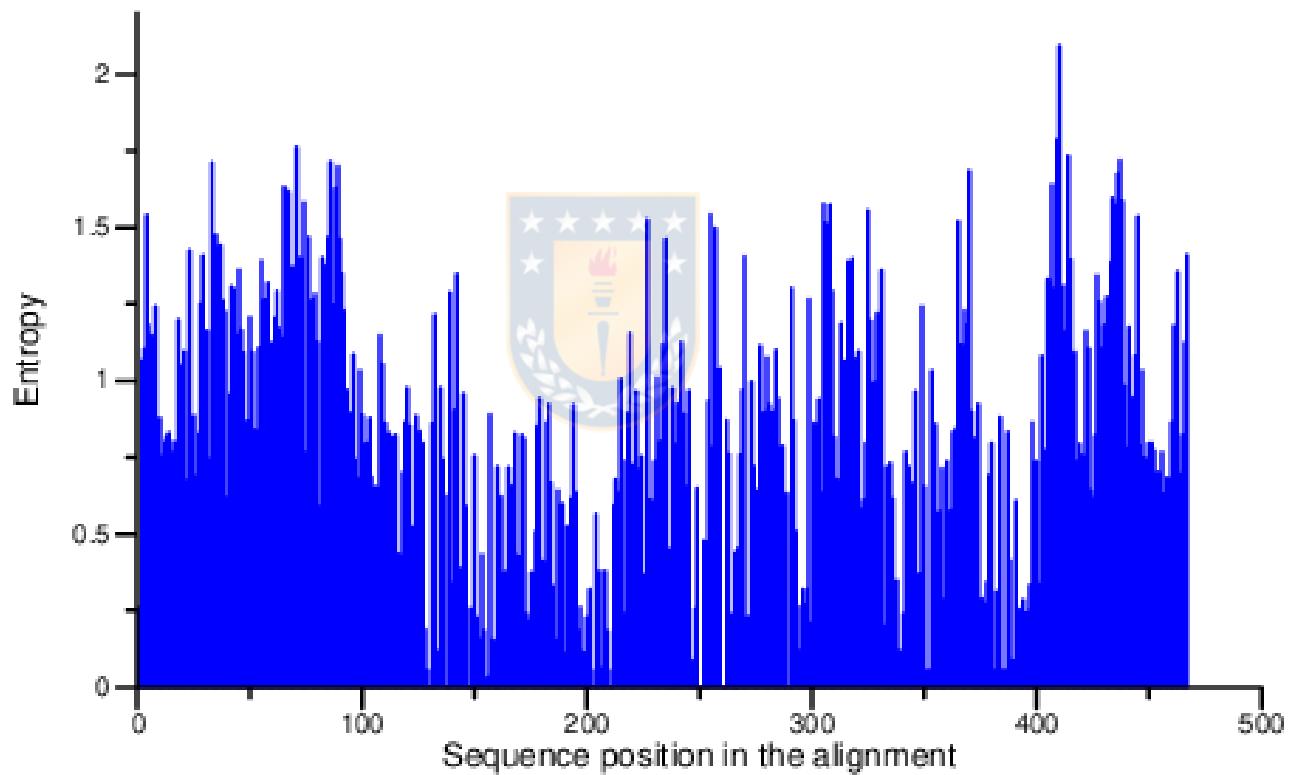
E_lucius ENSELU	AEAV
T_rubripes ENST	AEAV
T_nigroviridis	AEAV
L_bergylta ENSL	AEAV
A_percula ENSAP	AEAV
S_partitus ENSS	AEAV
P_ranga ENSFRNP	AEAV
L_calcarifer EN	AEAV
L_crocea ENSLCR	AEAV
S_aurata ENSSAU	AEAV
D_labrax ENSDLA	AEAV
S_lucioperca EN	AEAV
O_aureus ENSOAB	AEAV
M_zebra ENSMZEP	AEAV
A_citrinellus E	AEAV
S_dumerili ENSS	AEAV
S_lalandi ENSSLD	AEAV
M_mola ENSMOPO	AEAV
C_lumpus ENSCLM	AEAV
M_armatus ENSMA	AEAV
E_naucrates ENS	AEAV
S_fasciatus ENS	AEAV
M_murdjan ENSM	AEAV
O_melastigma EN	AEAV
O_javanicus ENS	AEAV
K_marmoratus EN	AEAV
P_formosa ENSFP	AEAV
P_reticulata EN	AEAV
X_maculatus ENS	AEAV
G_affinis ENSGA	AEAV
G_affinis ENSGA	AEAV
C_variegatus EN	AEAV
F_heteroclitus	AEAV
M_albus ENSMALP	AEAV
C_harengus ENSC	AEAV
E_electricus EN	AEAV
I_punctatus ENS	AEAV
G_morhua ENSGMO	APV
D_rerio ENDARP	AEAV
C_auratus ENSCA	AEAV
C_carpio ENSCCR	AEAV
S_formosus ENSS	AEAV
C_gobio ENSCGOP	APV
P_paniscus ENSP	ADGK
P_troglodytes E	ADGK
G_gorila ENSGGO	ADGK
N_leucogenys EN	ADGK
P_tephrosceles	AEVK
C_angolensis EN	ADGK
R_bieti ENSRBIP	ADGK
C_atys ENSCATPO	ADGK
M_leucophaeus E	ADGK
P_anubis ENSPAN	ADGK
T_gelada ENSTGE	ADGK
M_nemestrina EN	ADGK
P_simus ENSPSMP	ADVK
C_hoffmanni ENS	ADVK
R_ferrumequinum	ADVK
N_vison ENSNVIP	ADVK
P_leo ENSPLOPO	ADVK
P_pardus ENSPPR	ADVK
F_catus ENSFCAP	ADVK
L_canadensis EN	ADVK
S_suricatta ENS	ADVK
Z_californianus	ADVK
U_americanus EN	ADVK
P_catodon ENSPC	ADVK
O_cuniculus ENS	TDNR
S_vulgaris ENSS	ADAK
M_domestica ENS	GETR
V_ursinus ENSVU	GETK
P_cinereus ENSP	GETK
O_anatinus ENSO	G...
P_textilis ENSP	...
N_scutatus ENSN	C...
N_naja ENSNNAPO	...
P_castaneus ENS	CDSI
P_castaneus ENS	CDSI
C_serpentina EN	CDDI
T_carolina ENST	CDDI
C_abingsdonii E	..TV
S_punctatus ENS	CEDA
B_japonicus ENS	CE..
C_ustulatus ENS	CE..
M_cyaneus ENSMC	CE..
A_mexicanus ENS	SESL
C_syrichtha ENST	...V
U_maritimus ENS	S...
P_tigris ENSPTI	S...
P_coquereli ENS	S...
B_mutus ENSEMUP	S...
C_capucinus ENS	S...
S_boliviensis E	S...
A_collaria ENSA	S...
A_nancymaae ENS	DHSV
M_murinus ENSMI	DHSV



## 12.2. Análisis de entropía de Shannon

### The Consensus of the Input set

```
1      MKSILDGLAD  TTFRTITTDL  LYVGSNDIQY  ED-KGDMASK  GYFPQKFPLS
51     SFRG-FQEKM  TAGD-P-LIP  -DQTNITEFY  NKSLFKENEE  NIQCGENFMD
101    MECFMILNPS  QQLAIAVLSL  TLGTFTVLEN  LLVLCVILHS  RSLRCRPSYH
151    FIGSLAVADL  LGSVIFVYSF  VDFHVFHRKD  SPNVFLFKLG  GVTASFTASV
201    GSLFLTADR  YISIHRPLAY  KRIVTRPKAV  VAFVMTWIA  IVIAVLPPLG
251    WNCKKLNSVC  SDIFPLIDET  YLMFWIGVTS  VLLLFIVYAY  MYILWKAHSH
301    AVRMIQRGTQ  KSIIIHSSD  GKVQVTRPDQ  ARMDIRLAKT  LVLILVVLII
351    CWGPLLAIMV  YDVFGKMNKL  IKTVFAFCSM  LCLLNSTVNP  IYYALRSKDL
401    RHAFRSMFPS  CEGTAQPLDN  SMESDCQHKH  ANNAGSVHRA  AESCIKSTVK
451    IAKVTMSVST  DTSAEAL
```





Position	1st seq in input set	Entropy
1	M	0.258
2	K	1.066
3	S	1.104
4	I	1.542
5	L	1.182
6	D	1.147
7	G	0.824
8	L	1.242
9	A	0.774
10	D	0.877
11	T	0.754
12	T	0.802
13	F	0.820
14	R	0.830
15	T	0.765
16	I	0.802
17	T	0.704
18	T	1.198
19	D	1.052
20	L	0.682
21	L	1.093
22	Y	0.672
23	V	1.426
24	G	0.739
25	S	0.884
26	N	0.690
27	D	0.826
28	I	1.250
29	Q	1.408
30	Y	0.574
31	E	1.164
32	D	0.743
33	I	1.711
34	K	1.395
35	G	1.478

36	D	0.993
37	M	1.443
38	A	1.263
39	S	1.225
40	K	0.616
41	G	0.955
42	Y	1.306
43	F	1.300
44	P	1.148
45	Q	1.361
46	K	1.165
47	F	1.091
48	P	0.699
49	L	0.868
50	T	1.207
51	S	1.091
52	F	0.842
53	R	0.676
54	G	1.107
55	S	1.390
56	F	0.650
57	Q	1.267
58	E	1.320
59	K	0.603
60	M	1.124
61	T	1.204
62	A	1.292
63	G	1.172
64	D	1.139
65	N	1.633
66	P	1.621
67	Q	1.615
68	L	0.959
69	V	1.373
70	P	0.773
71	A	1.762
72	D	1.401
73	Q	1.296

74	V	1.582
75	N	0.766
76	I	1.468
77	T	0.889
78	E	1.266
79	F	1.284
80	Y	1.128
81	N	0.589
82	K	1.400
83	S	0.867
84	L	1.378
85	F	1.469
86	K	1.715
87	E	1.247
88	N	1.626
89	E	1.700
90	E	1.459
91	N	1.351
92	I	1.231
93	Q	0.969
94	C	0.896
95	G	0.677
96	E	1.084
97	N	0.744
98	F	0.679
99	M	1.031
100	D	0.886
101	I	0.593
102	E	0.797
103	C	0.877
104	F	0.685
105	M	0.247
106	V	0.655
107	L	0.216
108	N	1.147
109	P	1.055
110	S	0.861
111	Q	0.368

112	Q	0.828
113	L	0.818
114	A	0.551
115	I	0.822
116	A	0.436
117	V	0.322
118	L	0.701
119	S	0.864
120	L	0.977
121	T	0.856
122	L	0.524
123	G	0.351
124	T	0.885
125	F	0.593
126	T	0.834
127	V	0.798
128	L	0.188
129	E	0.056
130	N	0.000
131	L	0.862
132	L	1.216
133	V	0.032
134	L	0.119
135	C	0.978
136	V	0.743
137	I	0.000
138	L	0.624
139	H	1.287
140	S	0.340
141	R	0.907
142	S	1.349
143	L	0.390
144	R	0.199
145	C	0.958
146	R	0.593
147	P	0.000
148	S	0.256
149	Y	0.032

150	H	0.756
151	F	0.228
152	I	0.154
153	G	0.433
154	S	0.184
155	L	0.000
156	A	0.032
157	V	0.890
158	A	0.152
159	D	0.000
160	L	0.718
161	L	0.088
162	G	0.622
163	S	0.277
164	V	0.378
165	I	0.716
166	F	0.216
167	V	0.658
168	Y	0.826
169	S	0.432
170	F	0.208
171	I	0.822
172	D	0.810
173	F	0.240
174	H	0.216
175	V	0.376
176	F	0.339
177	H	0.508
178	R	0.852
179	K	0.940
180	D	0.411
181	S	0.283
182	R	0.864
183	N	0.924
184	V	0.668
185	F	0.332
186	L	0.152
187	F	0.642

188	K	0.293
189	L	0.598
190	G	0.110
191	G	0.525
192	V	0.230
193	T	0.614
194	A	0.921
195	S	0.636
196	F	0.184
197	T	0.259
198	A	0.117
199	S	0.000
200	V	0.228
201	G	0.317
202	S	0.056
203	L	0.000
204	F	0.564
205	L	0.199
206	T	0.381
207	A	0.063
208	I	0.380
209	D	0.184
210	R	0.000
211	Y	0.056
212	I	0.595
213	S	0.676
214	I	0.633
215	H	1.006
216	R	0.739
217	P	0.240
218	L	0.896
219	A	1.155
220	Y	0.216
221	K	0.726
222	R	0.965
223	I	0.709
224	V	0.752
225	T	0.208

226	R	0.368
227	P	1.527
228	K	0.610
229	A	0.295
230	V	0.738
231	V	1.011
232	A	0.719
233	F	0.803
234	C	1.121
235	L	1.460
236	M	0.452
237	W	0.000
238	T	0.976
239	I	0.792
240	A	0.926
241	I	0.827
242	V	1.128
243	I	0.892
244	A	0.654
245	V	0.967
246	L	0.088
247	P	0.000
248	L	0.256
249	L	0.650
250	G	0.000
251	W	0.000
252	N	0.477
253	C	0.000
254	E	0.933
255	K	1.543
256	L	0.781
257	Q	1.498
258	S	0.617
259	V	1.038
260	C	0.000
261	S	0.000
262	D	0.870
263	I	0.763

264	F	0.240
265	P	0.000
266	H	0.440
267	I	0.453
268	D	0.761
269	E	0.970
270	T	1.406
271	Y	0.063
272	L	0.231
273	M	0.998
274	F	0.721
275	W	0.184
276	I	0.644
277	G	1.115
278	V	0.861
279	T	0.899
280	S	1.077
281	V	0.922
282	L	0.088
283	L	0.900
284	L	1.099
285	F	0.939
286	I	0.078
287	V	0.785
288	Y	0.370
289	A	0.631
290	Y	0.000
291	M	1.301
292	Y	0.871
293	I	0.509
294	L	0.119
295	W	0.262
296	K	0.317
297	A	0.271
298	H	0.319
299	S	1.267
300	H	0.208
301	A	0.862



302	V	0.819
303	R	0.938
304	M	0.633
305	I	1.577
306	Q	1.513
307	R	0.774
308	G	1.574
309	T	1.292
310	Q	0.816
311	K	0.678
312	S	0.605
313	I	1.187
314	I	1.064
315	I	1.065
316	H	1.389
317	T	1.134
318	S	1.397
319	E	1.061
320	D	1.073
321	G	1.093
322	K	0.580
323	V	0.610
324	Q	0.797
325	V	1.555
326	T	1.196
327	R	0.843
328	P	0.996
329	D	1.221
330	Q	0.812
331	A	1.360
332	R	0.199
333	M	0.720
334	D	0.199
335	I	0.731
336	R	0.615
337	L	0.078
338	A	0.349
339	K	0.119

340	T	0.056
341	L	0.240
342	V	0.768
343	L	0.720
344	I	0.665
345	L	0.088
346	V	0.965
347	V	0.282
348	L	0.371
349	I	1.242
350	I	0.655
351	C	0.056
352	W	0.056
353	G	1.032
354	P	0.056
355	L	0.859
356	L	0.261
357	A	0.572
358	I	0.715
359	M	0.286
360	V	0.737
361	Y	0.521
362	D	0.576
363	V	0.833
364	F	0.845
365	G	1.520
366	K	1.121
367	M	0.846
368	N	1.229
369	K	1.181
370	L	1.686
371	I	0.901
372	K	0.416
373	T	0.813
374	V	0.923
375	F	0.259
376	A	0.290
377	F	0.273

378	C	0.343
379	S	0.693
380	M	0.798
381	L	0.056
382	C	0.311
383	L	0.295
384	L	0.881
385	N	0.056
386	S	0.056
387	T	0.832
388	V	0.413
389	N	0.088
390	P	0.088
391	I	0.608
392	I	0.254
393	Y	0.088
394	A	0.286
395	L	0.231
396	R	0.247
397	S	0.333
398	K	0.865
399	D	0.673
400	L	0.738
401	R	0.338
402	H	1.080
403	A	0.637
404	F	0.775
405	R	1.333
406	S	1.261
407	M	1.641
408	F	1.300
409	P	1.788
410	S	2.092
411	C	0.861
412	E	1.309
413	G	1.159
414	T	1.734
415	A	1.396

416	Q	1.032
417	P	1.092
418	L	0.737
419	D	0.794
420	N	0.757
421	S	0.749
422	M	1.161
423	D	1.103
424	S	0.645
425	D	0.615
426	C	0.820
427	L	1.344
428	H	1.257
429	K	1.106
430	H	1.181
431	A	1.275
432	N	1.134
433	N	1.387
434	A	1.597
435	A	1.572
436	S	1.675
437	V	1.719
438	H	1.586
439	R	0.984
440	A	0.780
441	A	1.174
442	E	0.640
443	S	0.944
444	C	1.083
445	I	1.535
446	K	0.786
447	S	1.032
448	T	0.748
449	V	0.617
450	K	0.797
451	I	0.797
452	A	0.772
453	K	0.629

454	V	0.702
455	T	0.629
456	M	0.767
457	S	0.629
458	V	0.684
459	S	0.654
460	T	0.865
461	D	1.179
462	T	0.423
463	S	1.354
464	A	0.696
465	E	0.823
466	A	1.124
467	L	1.411

