

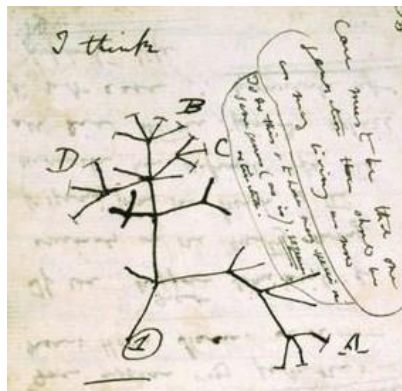
Grao en Bioloxía

Memoria del Trabajo de Fin de Grao

Revisión bibliográfica / Investigación: Pegadas de selección sobre o xene da supresina, recrutado dun retrovirus endógeno e implicado na inhibición da fusión celular nos mamíferos.

Revisión bibliográfica / Investigación: Huellas de selección sobre el gen de la supresina, reclutado de un retrovirus endógeno e implicado en la inhibición de la fusión celular en los mamíferos.

Literature review / Research: Footprints of selection on the *suppressyn* gene, a recruited sequence from an endogenous retrovirus involved in cell fusion inhibition in mammals.



Marta Rubira Pérez

Septiembre, 2018

Director Académico: Horacio Naveira Fachal

Ilustración de la portada: árbol filogenético de los organismos hecho por Darwin (1937).
Google Imágenes.

ÍNDICE

RESUMEN:	4
RESUMO:	4
SUMMARY:	4
PALABRAS CLAVE	4
KEYWORDS:	4
INTRODUCCIÓN:	5
OBJETIVOS:	8
MATERIAL Y MÉTODOS:	8
RESULTADOS:	10
Obtención de la supresina en humanos:	10
Obtención de ortólogos:	10
Estudio de la funcionalidad de la región de la supresina en las distintas especies de primates:	13
Evidencias sobre el modo de acción de la selección natural sobre el gen de la supresina:	15
DISCUSIÓN:	18
CONCLUSIONES:	20
Conclusiones:	20
Conclusions:	20
BIBLIOGRAFÍA:	21
ANEXOS:	23
Anexo 1: alineamiento de las secuencias nucleotídicas de las especies de primates analizadas en este trabajo.	23

RESUMEN:

Los retrovirus endógenos conforman aproximadamente el 8% del genoma humano. Hay 450.000 copias procedentes de infecciones que se han venido produciendo desde hace aproximadamente 45 MA en las líneas germinales de nuestros antepasados. La mayoría de las copias no son funcionales, lo que no evita que puedan tener una importante influencia reguladora. Algunas de ellas se han asociado con distintas patologías. Sin embargo, existe una creciente evidencia que señala el reclutamiento de algunas de estas secuencias para servir beneficiosamente al anfitrión. Un ejemplo de este efecto beneficioso podría ser el del gen llamado supresina, que parece jugar, junto con la sincitina, un importante papel en la regulación de la fusión celular en los tejidos humanos. En este trabajo se estudia la historia evolutiva de este gen en los primates y se pretende determinar el momento aproximado en el que fue reclutado para esta función celular.

RESUMO:

Os retrovirus endóxenos conforman aproximadamente o 8% do xenoma humano. Existen 450.000 copias que procedentes de infeccións que viñéronse producindo dende hai aproximadamente 45 MA nas liñas xerminais dos nosos antepasados. A maioría das copias non son funcionais, o que non evita que poidan ter unha importante influencia reguladora. Algunhas delas asociáronse con distintas patoloxías. Sen embargo, existe unha crecente evidencia que sinala o recrutamento dalgunhas destas secuencias para servir beneficiosamente ó anfitrión. Un exemplo diste efecto podería ser o xene denominado supresina, que parece xogar, xunto ca sincitina, un importante papel na regulación da fusión celular nos tecidos humanos. Neste traballo vaise estudar a historia evolutiva desde xene nos primates e preténdese determinar o momento aproximado no que foi recrutado para esta función celular.

SUMMARY:

The endogenous retroviruses form approximately 8% of the human genome. There are 450,000 copies derived from infections of the germ lines of our ancestors that have been taking place during the last 45 MA. Most of the copies are not functional, which does not prevent their being able to have an important regulating influence. Some of them have been associated with different pathologies. However, there is an increasing evidence that indicates that some of these sequences have been recruited to serve beneficially to the host. An example of this beneficial effect could be that of the gene called suppressyn, which it seems to play, together with the sincitina, an important paper in the regulation of the cellular melting in the human tissues. In this work the evolutionary history of this gene is studied in the primates and it is intended to determine the approximate moment in which it was recruited for this cellular function.

PALABRAS CLAVE:

Primates, placenta, sincitina, trofoblasto, HERVH48, HERVHFb-1, Hominoidea.

KEYWORDS:

Primates, placenta, sincityn, trofoblast, HERVH48, HERVHFb, Hominoidea.

INTRODUCCIÓN:

Los retrovirus son aquellos virus que poseen RNA como material genético, este RNA es de cadena simple. Los retrovirus también se caracterizan por poseer una transcriptasa inversa, que es la encargada de sintetizar el DNA a partir del RNA.

En los genomas de los mamíferos se encuentran dispersos, con miles de copias, los retrovirus endógenos (ERVs), elementos genéticos móviles que son reliquias de antiguas infecciones retrovirales (Meyer, Rosenkrantz, Carbone, & Chavez, 2017). La endogenización de los retrovirus se produce cuando la infección afecta a las células de la línea germinal. De este modo la infección se transmite de forma mendeliana a través de las generaciones (Ilustración 1).

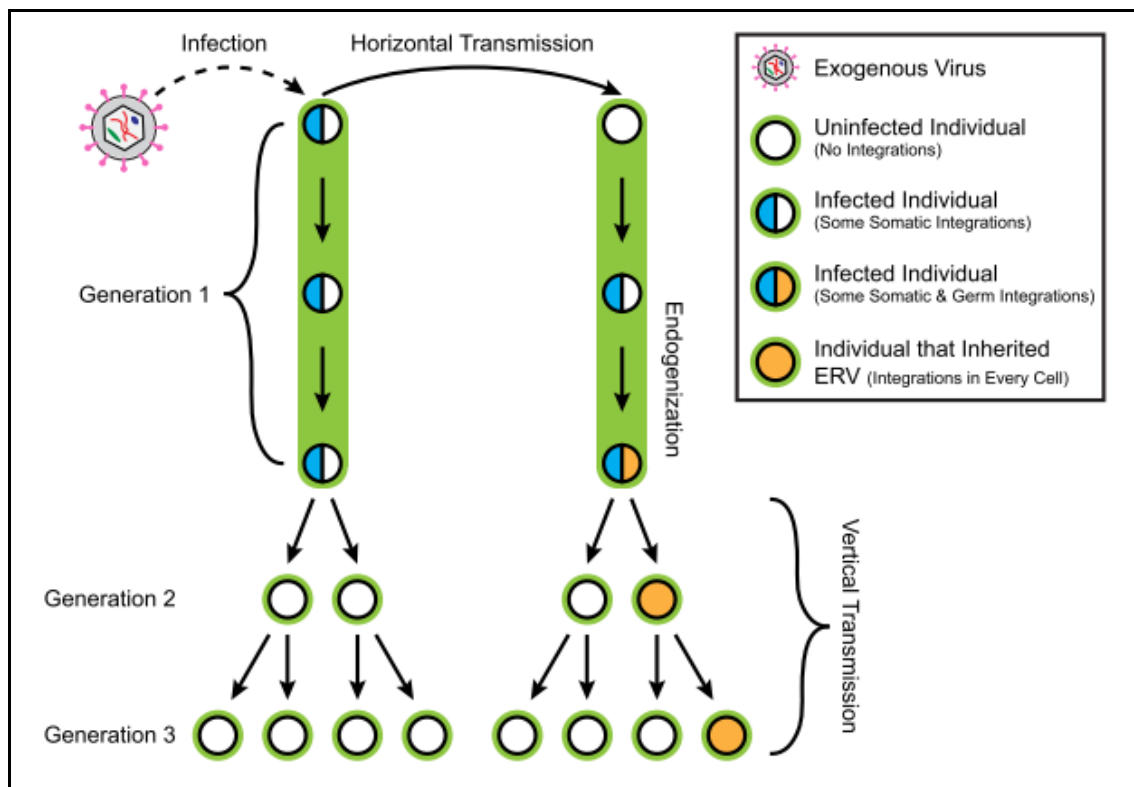


Ilustración 1: Transmisión horizontal, endogenización y transmisión vertical de los retrovirus (Meyer et al., 2017). Un retrovirus exógeno infecta a un primer individuo lo que supone su integración en las células somáticas. Se produce una transmisión horizontal hacia un segundo individuo, donde vuelve a haber una integración en las células somáticas. Sin embargo, en este segundo individuo también se produce una integración en las células de la línea germinal. Por lo tanto, la descendencia del segundo individuo, al contrario que la del primero, hereda la integración del retrovirus de las células de la línea germinativa.

Los retrovirus endógenos se agrupan en diferentes familias, las cuales están bien definidas. Cada familia, generalmente representa una única infección seguida de una expansión del número de copias en el genoma del hospedador (Magiorkinis, Belshaw, & Katzourakis, 2013). Para finalmente fijarse en la

población. Los retrovirus que han quedado integrados en el genoma de los humanos se denominan HERVs. Estos retrovirus endógenos conforman aproximadamente el 8% del genoma total de los humanos (Magiorkinis et al., 2013). Cerca de 450.000 copias de estos retrovirus en nuestro DNA que infectaron hace millones de años nuestra línea germinal (Goñi, 2015). En los últimos años se han descubierto, en humanos, decenas familias distintas de HERVs. Sin embargo, cada especie posee su propio repertorio en cantidad, localización y tipo de ERVs, lo que las diferencia de las demás.

Los ERVs completos codifican para varias proteínas (Gag, Pol y Env) y están flanqueados por dos repeticiones terminales largas (LTRs) (Goñi, 2015) (Ilustración 2). A pesar de la gran cantidad de copias que hay en los genomas de los hospedadores la gran mayoría de estas copias no son funcionales, es decir, pierden su capacidad de proliferación. Esto se debe a la acumulación de mutaciones con el paso del tiempo producidas en las divisiones de las líneas germinativas. A pesar de que esto puede producir codones de paro prematuros (Magiorkinis et al., 2013) esto no evita que algunas secuencias retrovíricas tengan una importante influencia reguladora en el genoma de los hospedadores en los que están presentes. Por un lado, las LTRs que están en los extremos contienen secuencias reguladoras y promotoras que podrían unirse a distintos factores transcripcionales del hospedador (Goñi, 2015). Por otro, muchas de las inserciones se transcriben intensamente en distintos tejidos, frecuentemente en condiciones de estrés o asociada a importantes patologías.

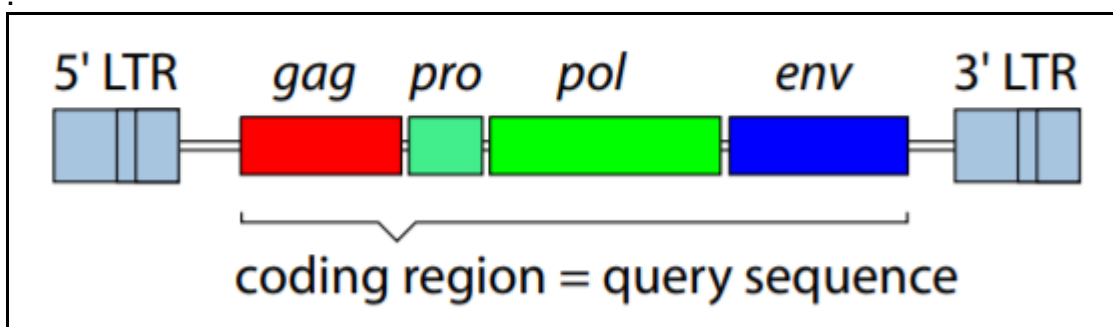


Ilustración 2: Organización simple de un retrovirus presente como provirus (DNA) integrado en el genoma del hospedador (Villesen, Aagaard, Wiuf, & Pedersen, 2004). se observa como las repeticiones terminales largas flanquean a las proteínas que codifican.

Por lo tanto, hay una gran preocupación por la relación entre los retrovirus endógenos y distintas enfermedades en los hospedadores. Sin embargo, existe una reciente evidencia de que una parte de este material genético es adoptado por el hospedador para servir beneficiosamente al anfitrión (Malfavon-Borja & Feschotte, 2015). Es decir, se produce una domesticación de algunos genes de los ERVs por parte del hospedador.

En este trabajo nos vamos a centrar en el gen de la supresina, que forma parte de la inserción proviral HERVH48-1, que se haya en el cromosoma 21 (<https://www.ncbi.nlm.nih.gov/gene/90625>). Esta inserción corresponde a una de las copias del elemento HERV-Fb1, comprendido en el supergrupo

HERVHF de la clase I, gamma-like, el más numeroso entre los retrovirus endógenos humanos, aunque este elemento en particular tiene tan sólo 16 representantes en el genoma humano (Vargiu et al., 2016). La invasión de este elemento en la línea germinal tuvo lugar poco antes de que el linaje de los Platyrrinos (monos del nuevo mundo) se separase del de los Catarrinos (monos del viejo mundo) hace aproximadamente unos 45 MA (Greenwood, Stengel, Erfle, Seifarth, & Leib-Mösch, 2005).

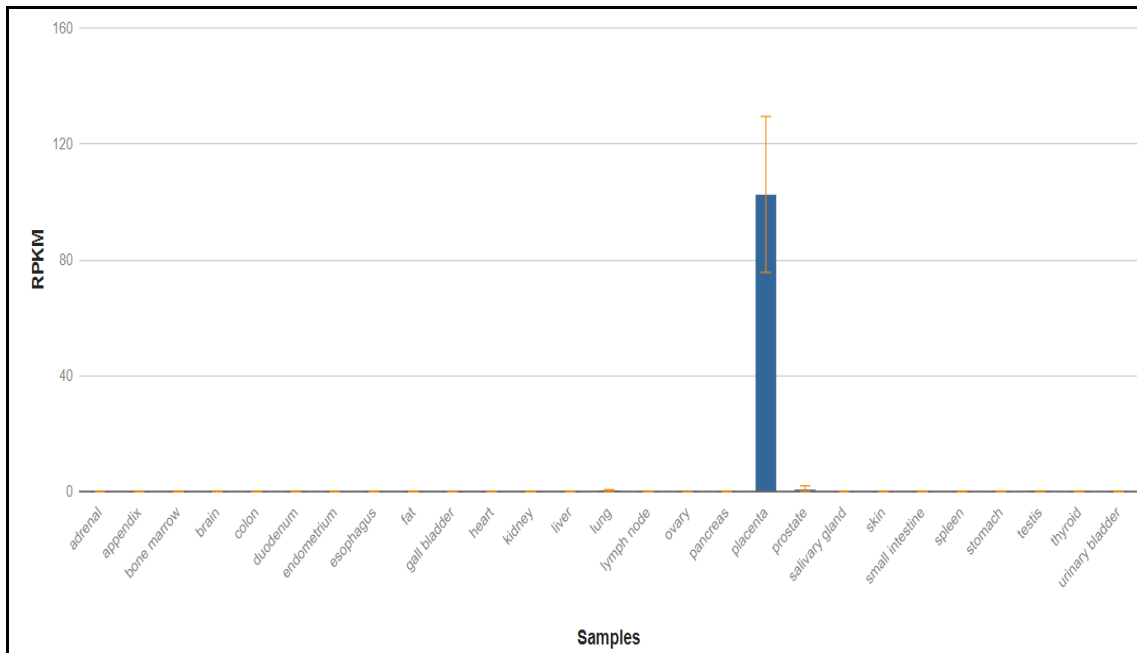


Ilustración 3: Gráfico de los tejidos en los que hay expresión de la supresina en humanos. Esta imagen, procedente del NCBI nos muestra los tejidos en los que se ha hallado que se expresa la supresina en tejidos *in vivo* en humanos.

Se transcribe solo en la placenta *in vivo* y su secuencia codificante de la supresina deriva de la región env codificante (Sugimoto, Sugimoto, Bernstein, Jinno, & Schust, 2013). Recientemente se ha visto que este gen está implicación en la inhibición de la fusión de tejidos en el trofoblasto (grupo de células de la parte externa del blastocisto) de la placenta de algunos mamíferos, entre los que se encuentra la especie humana. Para poder hablar sobre la función de inhibición de la fusión de tejidos primero debemos hablar sobre las sincitinas. Las sincitinas son genes de ERVs y que se han cooptado para desempeñar una función en la fusión de tejidos durante la placentación (Esnault, Cornelis, Heidmann, & Heidmann, 2013). La unión, entre sí, de los trofoblastos da como resultado el sincitiotrofoblasto, que es una precursor de la placenta (Goñi, 2015). Por lo que la actuación de las sincitinas en el desarrollo embrionario es de gran importancia para la supervivencia y el correcto desarrollo del feto. Se conocen dos tipos de sincitina, la sincitina-1 y la sincitina-2. La inhibición de la fusión celular inducida por la supresina se produce por su interferencia con la sincitina-1, puesto que se une a su receptor el ASCT2 (Sugimoto et al., 2013). En individuos sanos la fusión espontanea de células o sincitización está muy restringida a unos pocos tejidos (músculo,

hueso y placenta). Es en esta restricción donde la supresina desempeña un papel determinante (Sugimoto et al., 2013).

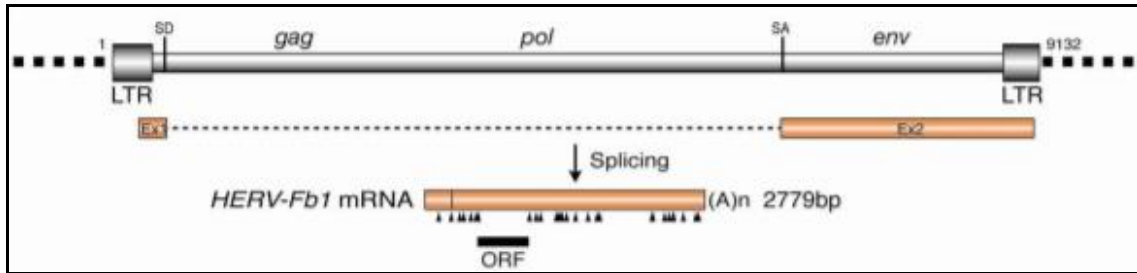


Ilustración 4: patrón de la unión del gen HERV-Fb1 y el producto de la transcripción (Sugimoto et al., 2013).

OBJETIVOS:

Los objetivos de este trabajo los podemos presentar en cuatro puntos distintos:

1. Obtener secuencias ortólogas de la inserción del HERVH48 en tantas especies de primates como sea posible a partir de los repositorios públicos de secuencias nucleotídicas.
2. Identificar la región de la supresina en las secuencias ortólogas de HERVH48 de aquellos primates en los que encontramos dicha inserción.
3. Determinar en qué momento de la historia evolutiva de los primates se produjo el reclutamiento de la supresina.
4. Cuantificar la acción de la selección natural sobre la secuencia de la supresina.

MATERIAL Y MÉTODOS:

Se usaron los accesos a información biomédica y genómica, así como distintas herramientas que ofrece el NCBI, el Centro Nacional para la Información Biotecnológica de los Estados Unidos de Norteamérica (<https://www.ncbi.nlm.nih.gov/>). Todas las secuencias de interés se exportaron a un ordenador personal en formato FASTA (.fas).

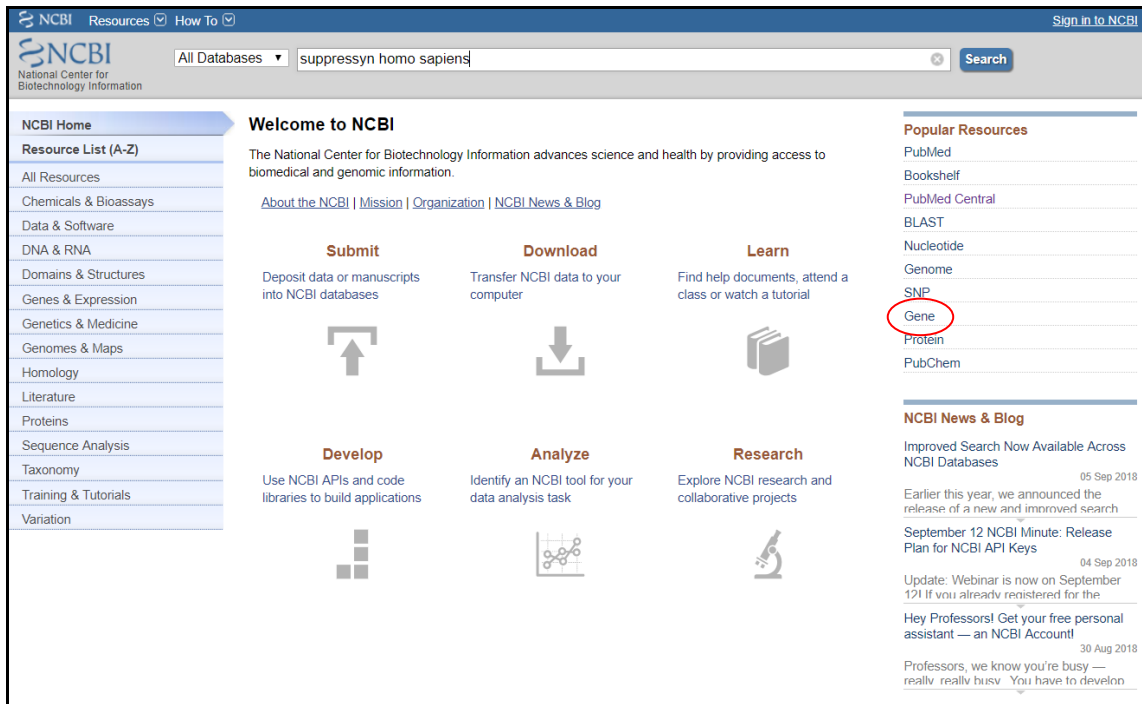


Ilustración 5: Página principal del NCBI. Usado para la obtención de la secuencia de la supresina humana y de algunos de los ortólogos de la misma. Gene es la base de datos en la que buscaremos la secuencia de la supresina en humanos y sus correspondientes ortólogos.

Se utilizó también el navegador genómico Ensembl para el estudio comparativo más detallado de las secuencias ortólogas de los distintos primates (<https://www.ensembl.org/index.html>) utilizando la herramienta BLAST/BLAT para el minado de las bases de datos públicas y la exportación de las secuencias de interés.

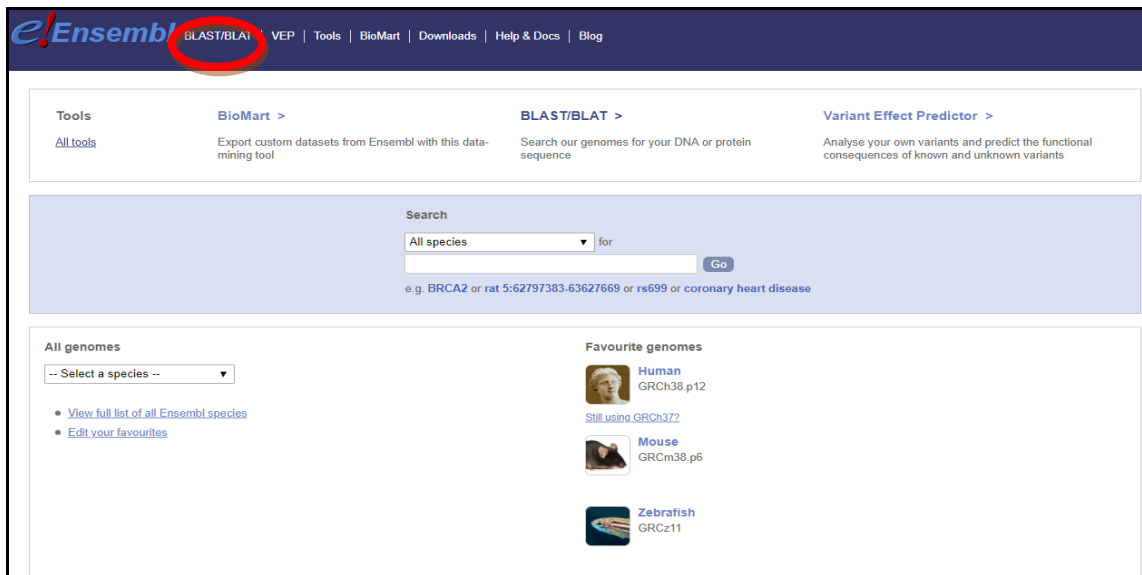


Ilustración 6: Página de inicio del navegador Ensembl. Usado para la obtención de ortólogos de la supresina. Señalada con un círculo la herramienta que vamos a usar para obtener dichas secuencias.

Para el alineamiento de las secuencias, su edición y el estudio de las pautas abiertas de lectura contenidas en ellas se usó el programa BioEdit (Hall, 1999). Entre otras usaremos la herramienta ClustalW para realizar el alineamiento de las secuencias ortólogas de la supresina en primates.

El programa DnaSp (Librado & Rozas, 2005) se utilizó para el análisis de las tasas de sustitución sinónima y no sinónima correspondientes al gen de la supresina en las comparaciones dos a dos de las distintas especies estudiadas. En este programa vamos a usar la herramienta de análisis sinónimos y no sinónimos.

Usamos también el programa Mendeley como gestor de la bibliografía (<https://www.mendeley.com/>)

RESULTADOS:

Obtención de la supresina en humanos:

Tras la búsqueda del término “suppressyn” en el NCBI, la base de datos Gene, nos devuelve 60 resultados, seleccionando el correspondiente a humanos: ID: 90625, HERVH48-1

Official Symbol	ERVH48-1 provided by HGNC
Official Full Name	endogenous retrovirus group 48 member 1 provided by HGNC
Primary source	HGNC:HGNC:17216
Gene type	protein coding
RefSeq status	REVIEWED
Organism	Homo sapiens
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as	C21orf105; NDUFV3-AS1
Summary	Many different endogenous retrovirus families are expressed in normal placental tissue at high levels, suggesting that endogenous retroviruses are functionally important in reproduction. This gene is part of an endogenous retrovirus provirus that has placenta specific expression. The protein encoded has the characteristics of a retroviral envelope protein but is truncated and lacks the transmembrane domain. The protein inhibits cell fusion by competing with syncytin-1 for binding to a cell receptor. [provided by RefSeq, May 2015]
Expression	Restricted expression toward placenta (RPKM 102.6) See more
Orthologs	all

Ilustración 7: Información sobre la supresina humana en la base de datos Gene del NCBI. Desde esta información de la supresina en *Homo sapiens* ya podemos acceder a los ortólogos de otras especies.

Accedemos a las secuencias nucleotídicas y aminoácidas a través del GenBank y las descargamos en formato FASTA. Y analizamos las secuencias en el BioEdit.

Obtención de ortólogos:

Los ortólogos los hemos obtenido de distintas bases de datos. Los primeros que extraemos y analizamos son los que obtenemos del link de la información de la supresina en humanos (Ilustración 7).

Name/Gene ID	Description	Location	Aliases
<input type="checkbox"/> ERVH48-1 ID: 90625	endogenous retrovirus group 48 member 1 [<i>Homo sapiens</i> (human)]	Chromosome 21, NC_000021.9 (42917251..42925598, complement)	C21orf105, NDUFV3-AS1
<input type="checkbox"/> ERVH48-1 ID: 101148698	endogenous retrovirus group 48 member 1 [<i>Gorilla gorilla</i> (western gorilla)]	Chromosome 21, NC_018445.2 (34075225..34083138, complement)	
<input type="checkbox"/> ERVH48-1 ID: 100981918	endogenous retrovirus group 48 member 1 [<i>Pan paniscus</i> (pygmy chimpanzee)]	Chromosome 21, NC_027889.1 (42539734..42547631, complement)	
<input type="checkbox"/> ERVH48-1 ID: 100938401	endogenous retrovirus group 48 member 1 [<i>Pongo abelii</i> (Sumatran orangutan)]	Chromosome 21, NC_036924.1 (30752446..30760368, complement)	
<input type="checkbox"/> ERVH48-1 ID: 100602270	endogenous retrovirus group 48 member 1 [<i>Nomascus leucogenys</i> (northern white-cheeked gibbon)]	Chromosome 25, NC_019840.1 (29164071..29176172, complement)	
<input type="checkbox"/> ERVH48-1 ID: 736930	endogenous retrovirus group 48 member 1 [<i>Pan troglodytes</i> (chimpanzee)]	Chromosome 21, NC_036900.1 (29433282..29441181, complement)	
<input type="checkbox"/> ERVH48-1 ID: 112621062	endogenous retrovirus group 48 member 1 [<i>Theropithecus gelada</i> (gelada)]	Chromosome 3, NC_037670.1 (20106795..20107475)	
<input type="checkbox"/> ERVH48-1 ID: 111527277	endogenous retrovirus group 48 member 1 [<i>Ptilocolobus tephrosceles</i> (Ugandan red Colobus)]		
<input type="checkbox"/> ERVH48-1 ID: 110567290	endogenous retrovirus group 48 member 1 [<i>Aotus nancymaae</i> (Ma's night monkey)]		
<input type="checkbox"/> ERVH48-1 ID: 108640503	endogenous retrovirus group 48 member 1 [<i>Manacus vitellinus</i> (golden-collared manakin)]		
<input type="checkbox"/> ERVH48-1 ID: 108588116	endogenous retrovirus group 48 member 1 [<i>Callithrix jacchus</i> (white-tufted-ear marmoset)]	Chromosome 14, NC_013909.1 (65703767..65711224, complement)	
<input type="checkbox"/> ERVH48-1 ID: 108283534	endogenous retrovirus group 48 member 1 [<i>Cebus capucinus imitator</i>]		
<input type="checkbox"/> ERVH48-1 ID: 105575492	endogenous retrovirus group 48 member 1 [<i>Cercocebus atys</i> (sooty mangabey)]		
<input type="checkbox"/> ERVH48-1 ID: 105555557	endogenous retrovirus group 48 member 1 [<i>Mandrillus leucophaeus</i> (drill)]		
<input type="checkbox"/> ERVH48-1 ID: 105472573	endogenous retrovirus group 48 member 1 [<i>Macaca nemestrina</i> (pig-tailed macaque)]		
<input type="checkbox"/> ERVH48-1 ID: 103882563	endogenous retrovirus group 48 member 1 [<i>Papio anubis</i> (olive baboon)]	Chromosome 3, NC_018154.2 (3728658..3734618)	

Ilustración 8: Resultados de la búsqueda de ortólogos en el NCBI. El NCBI nos devuelve 16 resultados de tras la búsqueda, sin embargo, uno de ellos es de un ave por lo que debemos descartarlo, es el *Macacus vitellinus*.

Descartamos todas aquellas secuencias que no pertenecen al linaje de los primates. Es decir, *Macacus vitellinus* (Tucán pechiblanco). Las demás especies si son de interés para este estudio por lo que descargamos sus secuencias, tanto la aminoacídica como la nucleotídica, son incluidas en el BioEdit formando parte de la misma salida para realizar su correspondiente alineamiento.

También realizamos la correspondiente búsqueda en el navegador Ensembl para la obtención de más ortólogos que se unirán a la salida de BioEdit anterior. En este caso haremos dos búsquedas, primero una con la secuencia nucleotídica de la supresina de *Homo sapiens* y otra posterior con la secuencia aminoacídica de *Mandrillus leucophaeus* (Mandrill), esta segunda búsqueda se hace para obtener los ortólogos de Cercopithecoidea, que no aparecen al usar la del humano.

Esta base de datos nos permite buscar los ortólogos en 24 especies de primates de las cuales hay tres especies que no muestran coincidencia (*Otolemur garnettii*, *Propithecus coquereli* y *Microcebus murinus*) por lo cual nos quedamos con 21 resultados de los cuales también ignoramos el que el Mandrill ha realizado sobre sí mismo. Dándonos un total de 20 secuencias nucleotídicas para la supresina en los primates. Entre los que se encuentran especies tanto de Platiirinos como de Catarrinos.

Nombre científico	Nombre común	Linaje
<i>Homo sapiens</i>	Humano	Catarrinos
<i>Gorilla gorilla</i>	Gorila	Catarrinos
<i>Pan paniscus</i>	Bonobo	Catarrinos
<i>Pan troglodytes</i>	Chimpancé	Catarrinos
<i>Pongo abelii</i>	Orangután	Catarrinos
<i>Nomascus leucogenis</i>	Gibón	Catarrinos
<i>Theropithecus gelada</i>	Gelada	Catarrinos
<i>Ptilocolobus tephrosceles</i>	Colobo rojo	Catarrinos
<i>Papio anubis</i>	Papión oliva	Catarrinos
<i>Chlorocebus sabaues</i>	Mono verde	Catarrinos
<i>Macaca fascicularis</i>	Macaco cangrejero	Catarrinos
<i>Macaca nemestrina</i>	Macaco cola de cerdo	Catarrinos
<i>Macaca mulatta</i>	Macaco rhesus	Catarrinos
<i>Colobus angolensis</i>	Colobo angoleño	Catarrinos
<i>Rhinopithecus roxellana</i>	Langur chato dorado	Catarrinos
<i>Rhinopithecus bieti</i>	Langur chato negro	Catarrinos
<i>Mandrillus leucophaeus</i>	Mandrill	Catarrinos
<i>Aotus nancymaae</i>	Mono nocturno	Platirrinios
<i>Cebus capucinos</i>	Mono carablanca	Platirrinios
<i>Callithrix jacchus</i>	Tití común	Platirrinios

Tabla 1: Especies de primates que poseen secuencias acordes con la búsqueda de ortólogos de la supresina. En esta tabla mostramos las especies que poseen secuencias que podrían ser ortólogas a la supresina del Mandril. Vemos que entre los resultados hay especies de linajes de Catarrinos y Platirrinios.

Descargamos las secuencias nucleotídicas de estos primates y los introducimos en una misma salida del BioEdit tras lo cual los alineamos.

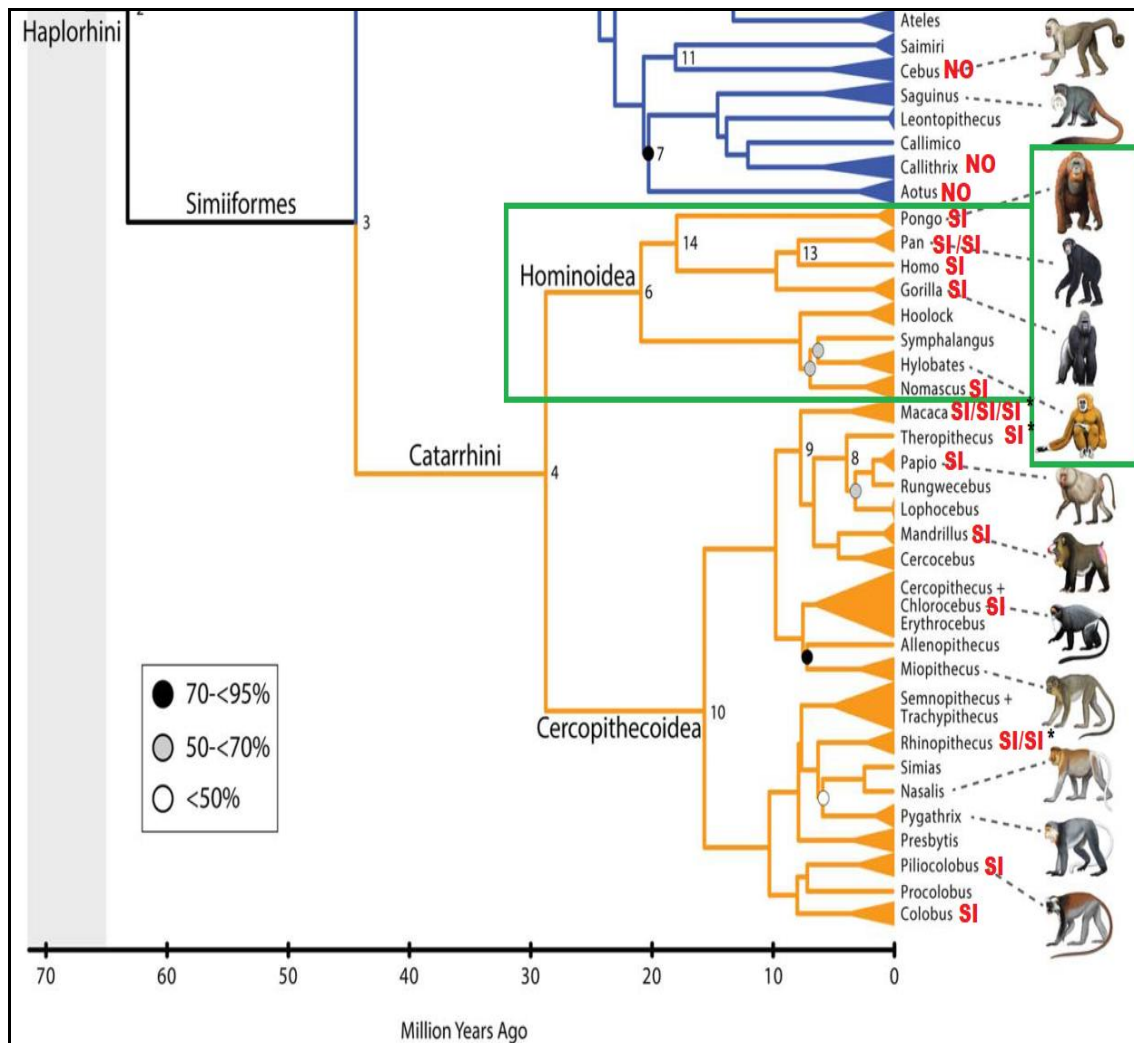


Ilustración 9: Árbol filogenético de primates. (Springer et al., 2012). En este árbol están señalizados los géneros de los primates en los cuales hemos encontrados resultados de la secuencia de la supresina. Al lado de los nombres de los géneros hay un “SI” o un “NO” indicando si cumplen los requisitos establecidos para este estudio (tamaño de secuencia e identidad). Está indicando con asteriscos “*” si hay algún tipo de truncamiento o indel en esas secuencias. En *Theropithecus* hay un truncamiento debido a un codón de stop prematuro, al igual que en *Macaca nemestrina*. En *Macaca mulatta* y en *Rhinopithecus roxellana* aparecen indels. *Rhinopithecus biete* tiene su codón de iniciación aguas arriba. Todos estos factores pueden hacer que la supresina presente en el genoma de estas especies no sea funcional. Por último, enmarcamos la familia Hominoidea puesto que en ella todas las secuencias de supresina parecen estar funcionales.

Estudio de la funcionalidad de la región de la supresina en las distintas especies de primates:

De todas las secuencias que hemos obtenido, para este apartado solo nos quedamos con aquellas que cumplan los requisitos de tamaño (mínimo 100 aminoácidos o 300 nucleótidos) e identidad (que sea igual o superior al 80%).

Genomic Location	Overlapping Gene(s)	Orientation	Query start	Query end	Length	Score	E-val	%ID
25_83674864-83675412 [Sequence]	CNST	Reverse	6	179	190 [Sequence]	84.7	1e-16	39.47 [Alignment]
2_86771379-86771930 [Sequence]		Reverse	1	184	184 [Sequence]	332	4e-102	97.28 [Alignment]
21_103085283-103085822 [Sequence]		Forward	1	179	181 [Sequence]	132	9e-33	53.04 [Alignment]
11_83602962-83603468 [Sequence]		Forward	1	171	171 [Sequence]	219	9e-63	74.27 [Alignment]
1_85117948-85118424 [Sequence]		Reverse	1	159	159 [Sequence]	222	7e-64	78.62 [Alignment]
32_12866986-12867390 [Sequence]		Reverse	44	179	136 [Sequence]	48.1	1e-05	30.15 [Alignment]

Ilustración 10: Resultados de la búsqueda de secuencias ortólogas usando el Mandril como “query”, estas secuencias pertenecen a *Chlorocebus sabaues*. Debemos escoger aquellas secuencias cuyos “query start” y “query end” sean los señalados en la ilustración, es decir, 1 y 184 respectivamente, también debe poseer un %ID alto (mayor del 80%).

Por lo tanto, tras aplicar estas condiciones descartamos las secuencias de los primates que pertenecen al linaje de los Platyrrinos ya que las secuencias conservadas en ellas tienen un tamaño mucho menor al de los Catarrinos y su identidad también es baja (es menor al 80%). Lo que nos da un total de 17 secuencias ortólogas a la supresina (Ilustración 9).

Nombre científico	Nombre común	Linaje
<i>Homo sapiens</i>	Humano	Catarrinos
<i>Gorilla gorilla</i>	Gorila	Catarrinos
<i>Pan paniscus</i>	Bonobo	Catarrinos
<i>Pan troglodytes</i>	Chimpancé	Catarrinos
<i>Pongo abelii</i>	Orangután	Catarrinos
<i>Nomascus leucogenis</i>	Gibón	Catarrinos
<i>Theropithecus gelada</i>	Gelada	Catarrinos
<i>Ptilocolobus tephrosceles</i>	Colobo rojo	Catarrinos
<i>Papio anubis</i>	Papión oliva	Catarrinos
<i>Chlorocebus sabaues</i>	Mono verde	Catarrinos
<i>Macaca fascicularis</i>	Macaco cangrejero	Catarrinos
<i>Macaca nemestrina</i>	Macaco cola de cerdo	Catarrinos
<i>Macaca mulatta</i>	Macaco Rhesus	Catarrinos
<i>Colobus angolensis</i>	Colobo angoleño	Catarrinos
<i>Rhinopithecus roxellana</i>	Langur chato dorado	Catarrinos
<i>Rhinopithecus bieti</i>	Langur chato negro	Catarrinos
<i>Mandrillus leucophaeus</i>	Mandrill	Catarrinos

Tabla 2: Especies de primates del linaje de los Catarrinos que poseen secuencias ortólogas a la supresina.

Tras el debido alineamiento en BioEdit de estas secuencias, añadimos una pauta de lectura (123123123123123...). Este análisis de pautas de lectura en las 17 secuencias nos revela que varias de ellas contienen codones de paro prematuros, como es el caso de *Macaca nemestrina* y *Theropithecus gelada*, también encontramos indels en *Macaca mulatta* o *Rhinopithecus roxellana*. Incluso encontramos que el codón de inicio en *Rhinopithecus bieti* aguas arriba de la posición en la que, se supone, debería estar. Por lo que parece que estas secuencias no son funcionales. Mientras que en la familia de Hominoidea todas las secuencias parecen ser funcionales (Ilustración 9) (Anexo 1). A continuación, eliminamos todas aquellas posiciones del alineamiento que contengan “gaps” teniendo especial cuidado en no alterar la pauta de lectura,

es decir, en ocasiones también debemos eliminar las posiciones adyacentes para evitar la alteración de dicha pauta de lectura.

Para la comprobación de este punto las secuencias son introducidas en el DnaSp y se realiza un análisis de sustituciones sinónimas y no sinónimas dos a dos. Para que los resultados sean más concretos este análisis se realiza en distintos grupos cuya relación evolutiva sea mayor, en nuestro caso obtenemos tres grupos (para la formación de los grupos nos valemos de un árbol filogénico de primates). Estos grupos son: Hominoidea, Cercopithecinae y Colobinae.

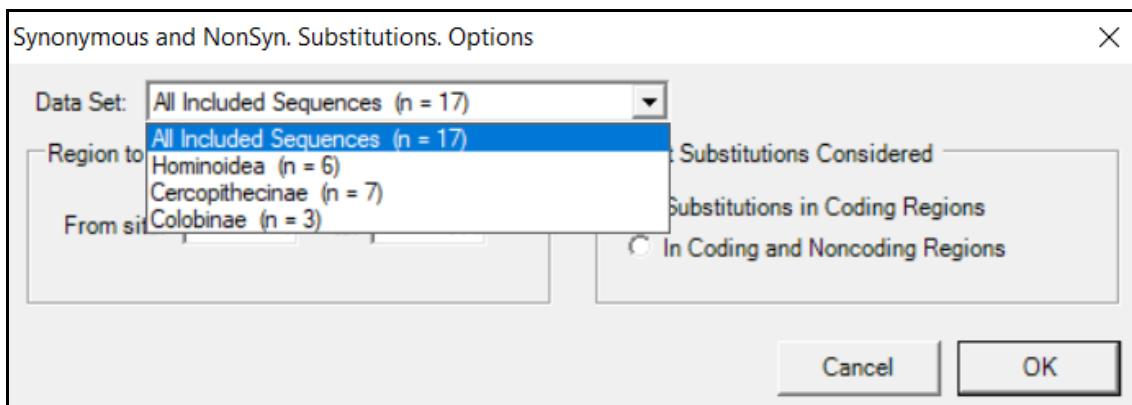


Ilustración 11: Grupos de primates usados para el análisis de sustituciones sinónimas y no sinónimas en el DnaSp. Nos muestra los grupos que tenemos y cuantas secuencias hay en cada grupo.

Cercopithecinae	Hominoidea	Colobinae
<i>Theropithecus gelada</i>	<i>Homo sapiens</i>	<i>Colobus angolensis</i>
<i>Mandrillus leucopaeus</i>	<i>Gorilla gorilla</i>	<i>Rhinopithecus roxellana</i>
<i>Papio anubis</i>	<i>Pan paniscus</i>	<i>Rhinopithecus bieti</i>
<i>Chlorocebus sabaeus</i>	<i>Pan troglodytes</i>	
<i>Macaca fascicularis</i>	<i>Pongo abelii</i>	
<i>Macaca nemestrina</i>	<i>Nomascus leucogenis</i>	
<i>Macaca mulatta</i>		

Tabla 3: Grupos de primates según la proximidad evolutiva para el análisis con el DnaSp y las especies que hay en cada uno de ellos.

Evidencias sobre el modo de acción de la selección natural sobre el gen de la supresina:

La acción de la selección natural sobre la supresina se puede ver gracias a las sustituciones sinónimas y no sinónimas que sufre la secuencia. Al hacer este análisis en el DnaSp, de los grupos establecidos en la Tabla 3, obtenemos la relación de los cambios sinónimos y no sinónimos dos a dos de las secuencias (Ilustración 12).

Seq 1	Seq 2	SynDif	SynPos	Ks	NSynDif	NSynPos	Ka
Col ancole	'Rhi bieti	4,50	97,50	0,0476	17,50	316,50	0,0574
Col ancole	Rhi roxell	23,00	99,25	0,2772	53,00	314,75	0,1907
'Rhi bieti	Rhi roxell	23,50	99,25	0,2845	54,50	314,75	0,1969

Ilustración 12: Tabla de sustituciones sinónimas y no sinónimas realizada al grupo de Colobinae usando el DnaSp. SynDif: diferencias sinónimas, SynPos: posiciones sinónimas, NSynDif: diferencias no sinónimas, NSynPos: posiciones no sinónimas, Ks: cambios sinónimos y Ka: cambios no sinónimos.

Estas tablas las pasamos a Excel para poder calcular Ka/Ks (Ka: cambios no sinónimos, es decir, hay reemplazamiento aminoacídico en la secuencia, Ks: cambios sinónimos, no hay reemplazamiento aminoacídico en la secuencia). Una vez tenemos todos los análisis en el Excel sacamos el cociente Ka/Ks y realizamos un gráfico para poder observar si la sección ha actuado y, en caso de hacerlo, si esa selección ha sido positiva o purificadora. Para ello debemos fijarnos en el cociente de Ka/Ks que nos da, si este cociente es mayor que 1 es que hay una selección adaptativa, si es igual o muy cercano a 1 es que no hay selección y si es menor que 1 es que hay selección purificadora.

La comparación de *Pan paniscus* con *Pan troglodytes* nos da unos valores de Ka y Ks iguales a 0 dado que son absolutamente idénticas, por lo que no se incluyen en la formación de la tabla y el gráfico.

Seq 1	Seq 2	SynDif	SynPos	Ks	NSynDif	NSynPos	Ka
'Hom sapie	Gor qorill	3,00	100,33	0,0305	1,00	313,67	0,0032
'Hom sapie	'Pan panis	1,00	101,33	0,0099	2,00	312,67	0,0064
'Hom sapie	'Pan trogl	1,00	101,33	0,0099	2,00	312,67	0,0064
'Hom sapie	'Pon abell	3,50	100,58	0,0356	9,50	313,42	0,0309
'Hom sapie	'Nom leuco	3,50	102,33	0,0350	9,50	311,67	0,0311
Gor qorill	'Pan panis	4,00	101,00	0,0407	3,00	313,00	0,0096
Gor qorill	'Pan trogl	4,00	101,00	0,0407	3,00	313,00	0,0096
Gor qorill	'Pon abell	6,50	100,25	0,0678	8,50	313,75	0,0276
Gor qorill	'Nom leuco	6,50	102,00	0,0666	8,50	312,00	0,0278
'Pan panis	'Pan trogl	0,00	102,00	0,0000	0,00	312,00	0,0000
'Pan panis	'Pon abell	5,00	101,25	0,0511	9,00	312,75	0,0293
'Pan panis	'Nom leuco	5,00	103,00	0,0502	9,00	311,00	0,0295
'Pan trogl	'Pon abell	5,00	101,25	0,0511	9,00	312,75	0,0293
'Pan trogl	'Nom leuco	5,00	103,00	0,0502	9,00	311,00	0,0295
'Pon abell	'Nom leuco	6,00	102,25	0,0611	8,00	311,75	0,0261

Ilustración 13: Tabla de salida de los resultados del análisis de sinónimos y no sinónimos del grupo de Hominoidea. En este cuadro podemos ver como los valores de Ka y Ks para la comparación de secuencias entre las dos especies de Pan son iguales a 0. Es decir, no hay cambios ni sinónimos ni no sinónimos en la secuencia.

En el grupo de los Cercopithecinae tenemos varias secuencias truncadas y con indels, por lo que el análisis se hace por separado para algunas de las especies, puesto que debemos eliminar algunas de las posiciones y el análisis de sustituciones sinónimas y no sinónimas no es tan fiable sin las secuencias no poseen la misma longitud. En la comparación del género de Macaca, que realizamos por separado, descartamos la comparación de *Macaca nemestrina* contra *Macaca mulatta* puesto que su Ks es igual a 0.

Seq 1	Seq 2	SilentDif	SilentPos	Ks	NSynDif	NSynPos	Ka
'Macaca fa	'Macaca ne	1,00	139,33	0,0072	3,00	409,67	0,0074
'Macaca fa	'Macaca mu	1,00	140,33	0,0072	4,00	408,67	0,0099
'Macaca ne	'Macaca mu	0,00	139,33	0,0000	5,00	409,67	0,0123

Ilustración 14: Tabla de salida de los resultados del análisis de sinónimos y no sinónimos del género *Macaca*. podemos ver como en la comparación entre *Macaca nemestrina* y *Macaca mulatta* la Ks es igual a 0 esto se debe a que la secuencia de *Macaca nemestrina* está truncada y que la secuencia de *Macaca mulatta* posee varios indels. (Anexo 1)

Seq 1	Seq 2	SynDif	SynPos	Ks	NSynDif	NSynPos	Ka	Ka/Ks	Tipo selección
Hom_sapien	Gor gorill	3,00	100,33	0,0305	1,00	313,67	0,0032	0,10491803	S. Purificadora
Hom_sapien	Pan_panisc	1,00	101,33	0,0099	2,00	312,67	0,0064	0,64646465	S. Purificadora
Hom_sapien	Pan_troglo	1,00	101,33	0,0099	2,00	312,67	0,0064	0,64646465	S. Purificadora
Hom_sapien	Pon_abelli	3,50	100,58	0,0356	9,50	313,42	0,0309	0,86797753	S. Purificadora
Hom_sapien	Nom_leucog	3,50	102,33	0,035	9,50	311,67	0,0311	0,88857143	S. Purificadora
Gor gorill	Pan_panisc	4,00	101	0,0407	3,00	313	0,0096	0,23587224	S. Purificadora
Gor gorill	Pan_troglo	4,00	101	0,0407	3,00	313	0,0096	0,23587224	S. Purificadora
Gor gorill	Pon_abelli	6,50	100,25	0,0678	8,50	313,75	0,0276	0,40707965	S. Purificadora
Gor gorill	Nom_leucog	6,50	102	0,0666	8,50	312	0,0278	0,41741742	S. Purificadora
Pan_panisc	Pon_abelli	5,00	101,25	0,0511	9,00	312,75	0,0293	0,57338552	S. Purificadora
Pan_panisc	Nom_leucog	5,00	103	0,0502	9,00	311	0,0295	0,5876494	S. Purificadora
Pan_troglo	Pon_abelli	5,00	101,25	0,0511	9,00	312,75	0,0293	0,57338552	S. Purificadora
Pan_troglo	Nom_leucog	5,00	103	0,0502	9,00	311	0,0295	0,5876494	S. Purificadora
Pon_abelli	Nom_leucog	6,00	102,25	0,0611	8,00	311,75	0,0261	0,42716858	S. Purificadora
Macaca_fas	Macaca_nem	1,00	139,33	0,0072	3,00	409,67	0,0074	1,02777778	No selección
Macaca_fas	Macaca_mul	1,00	140,33	0,0072	4,00	408,67	0,0099	1,375	S. Adpatativa
Theropithe	Mandrillus	1,00	117,67	0,0085	6,00	362,33	0,0167	1,96470588	S. Adpatativa
Col_angole	'Rhi_bieti	4,50	97,50	0,0476	17,50	316,5	0,0574	1,20588235	S. Adpatativa
Col_angole	'Rhi_roxell	23,00	99,25	0,2772	53,00	314,75	0,1907	0,68795094	S. Purificadora
'Rhi_bieti	'Rhi_roxell	23,50	99,25	0,2845	54,50	314,75	0,1969	0,69209139	S. Purificadora
'The_gelad	'Pap_anubi	2,00	102,17	0,0198	6,00	311,83	0,0195	0,98484848	No selección
'The_gelad	'Chl_sabae	3,00	102,17	0,03	7,00	311,83	0,0228	0,76	S. Purificadora
'The_gelad	'Mac_fasci	3,00	102,17	0,03	7,00	311,83	0,0228	0,76	S. Purificadora
'The_gelad	'Mac_nemes	3,00	101,92	0,03	7,00	312,08	0,0228	0,76	S. Purificadora
'The_gelad	'Mac_mulat	3,00	102,5	0,0299	9,00	311,5	0,0295	0,98662207	No selección
'Man_leuco	'Pap_anubi	2,00	103,33	0,0196	3,00	310,67	0,0097	0,49489796	S. purificadora
'Man_leuco	'Chl_sabae	3,00	103,33	0,0296	4,00	310,67	0,013	0,43918919	S. purificadora
'Man_leuco	'Mac_fasci	3,00	103,33	0,0296	4,00	310,67	0,013	0,43918919	S. purificadora
'Man_leuco	'Mac_nemes	3,00	103,08	0,0297	4,00	310,92	0,013	0,43771044	S. purificadora
'Man_leuco	'Mac_mulat	3,00	103,67	0,0295	6,00	310,33	0,0196	0,66440678	S. purificadora
'Pap_anubi	'Chl_sabae	3,00	103,17	0,0297	3,00	310,83	0,0097	0,32659933	S. purificadora
'Pap_anubi	'Mac_fasci	3,00	103,17	0,0297	3,00	310,83	0,0097	0,32659933	S. purificadora
'Pap_anubi	'Mac_nemes	3,00	102,92	0,0297	3,00	311,08	0,0097	0,32659933	S. purificadora
'Pap_anubi	'Mac_mulat	3,00	103,5	0,0296	5,00	310,5	0,0163	0,55067568	S. purificadora
'Chl_sabae	'Mac_fasci	4,00	103,17	0,0398	2,00	310,83	0,0065	0,16331658	S. purificadora
'Chl_sabae	'Mac_nemes	4,00	102,92	0,0399	2,00	311,08	0,0065	0,16290727	S. purificadora
'Chl_sabae	'Mac_mulat	4,00	103,5	0,0397	4,00	310,5	0,013	0,32745592	S. purificadora

Ilustración 15: Tabla de Excel con los resultados de la comparación de sinónimos y no sinónimos del DnaSp. Nos muestra los resultados de las comparaciones dos a dos de las secuencias de primates. También nos indica el tipo de selección que sufre la secuencias según el valor de Ka/Ks.



Gráfico 1: Gráfico de dispersión de los valores de Ka y Ks. En este gráfico se muestran la relación de Ka y Ks en el espacio. Están marcadas con un círculo los puntos que corresponden al grupo de Hominoidea.

DISCUSIÓN:

Tras la búsqueda y recopilación de todas las secuencias ortólogas de la secuencia de supresina tanto del Humano como del Drill vemos que los tamaños y las identidades de dichas secuencias en Platininos son muy bajos por lo que, a pesar de no incluirlas en los análisis, podemos decir que no son funcionales.

También es interesante decir que en la visualización de las secuencias alineadas de Catarrinos en el BioEdit se aprecia la delección de dos codones seguidos en tres especies distintas (*Pan paniscus*, *Pan troglodytes* y *Pongo abelii*). *Pan paniscus* y *Pan troglodytes* están muy relacionados evolutivamente, pero no lo están tanto con *Pongo abelii*, por lo que es curioso que haya la misma delección en los tres. Esto podría deberse a recombinación dentro de la partícula retroviral entre distintas copias de la familia HERVf-1 (Vargiu et al., 2016).

Homo sapiens	AGCCAAAGCCAAAGCCTCAAAAAC
Gorilla gorilla	AGCCAAAGCCAAAGCCTCAAAAAC
Pan paniscus	AGCCAAAGCC-----TCAAAAAC
Pan troglodytes	AGCCAAAGCC-----TCAAAAAC
Pongo abelli	AGCCAAAGCC-----TCAAAAAC
Nomascus leucog	AGCCAAAGCCAAATG CCTCAAAAAC
Theropithecus g	AGCCAAAGCCAAAGCCTCAAAAAC

Ilustración 16: Fragmento de las secuencias alineadas de la supresina en primates (Anexo 1). En ella podemos ver cómo están ausentes los mismos dos codones en las especies *Pongo abelli*, *Pan paniscus* y *Pan troglodytes*.

También podemos observar que las secuencias de *Pan troglodytes* y *Pan paniscus* son iguales, lo que nos indica que no ha pasado tiempo suficiente para que las secuencias de ambos hayan cambiado desde que se produjo la separación evolutiva de ambas especies, hace entre 2,3 y 3,8 MA (Pozzi et al., 2014). Esto también nos lo confirma el análisis de los cambios sinónimos y no sinónimos del DnaSp (Ilustración 13). Aunque al ser iguales las secuencias no las usemos para la construcción del gráfico de dispersión, puesto que ese dato no es fiable. Las secuencias pertenecientes a Hominoidea están bien conservadas y los valores de su Ka/Ks son siempre menores de 1 (Ilustración 15), por lo que se está llevando a cabo una selección purificadora, se produce la eliminación de las mutaciones que producen cambios en la secuencia aminoacídica. Lo que nos indica que la supresina es funcional en este grupo.

En las secuencias de las especies pertenecientes al grupo de Colobinae las secuencias son de tamaño adecuado para poder realizar la pauta de lectura y la identidad con respecto a secuencia “query” también es alta. Sin embargo, vemos que los resultados de la tabla de Excel que los valores de Ka/Ks son superiores de 1 (Ilustración 15), por lo que se está produciendo una selección adaptativa, las mutaciones que se producen en las secuencias no se eliminan y por lo tanto la supresina no ejerce ningún tipo de función en las especies que hemos analizado.

Por último, vamos a analizar los resultados que hemos obtenido en el grupo de Cercopithecinae. En este grupo tenemos las secuencias de dos especies (*Macaca nemestrina* y *Theropithecus gelada*) que poseen un codón de paro antes de lo normal, lo que significa que la proteína está truncada, esto no significa necesariamente que la supresina carezca de función en el organismo, pero sí que existen posibilidades de que lo esté. También vemos que hay un número importante de indels en la secuencia de la supresina de *Macaca mulatta* lo que nos indica que esa supresina no es funcional. En la relación de Ka/Ks vemos que existe una selección purificadora y en otros casos no hay selección. A pesar de que selección esté eliminando los cambios en las secuencias es muy probable que en este grupo la supresina tampoco sea funcional debido a los truncamientos que hay en las proteínas de algunas de las especies.

Por lo tanto, solo podemos afirmar la funcionalidad de la supresina en el grupo de Hominoidea, mientras que en Cercopithecinae no se puede saber solo con estos análisis y en Colobinae no es funcional. Esta última cuestión nos lleva a preguntarnos cuál podría ser entonces la proteína que ejerce la función de la supresina en estos dos últimos grupos.

CONCLUSIONES:

Conclusiones:

En relación con el primer objetivo se obtuvieron secuencias ortólogas de la inserción del HERVH48 humano en otras 19 especies de primates. De las cuales aparecen 4 en Platirrinos y 15 en Catarrinos (6 en Hominoidea y 9 en Cercopitheoidea).

Se ha conseguido caracterizar pautas abiertas de lectura de la supresina de al menos 300 nucleótidos tan solo en secuencias procedentes de los Catarrinos. Todas las secuencias ortólogas del HERFH48 obtenidas en los Platirrinos presentan severas alteraciones en la región que correspondería a la supresina, por lo que deben considerarse no funcionales.

El análisis detallado del grado de conservación de la secuencia de la supresina en los Catarrinos indica que fue reclutada antes de la separación de los linajes de Homonoidea y Cercapitheoidea.

La comparación de las tasas de sustitución sinónima y no sinónima para el gen de la supresina indican, no obstante, que tan solo está actuando la selección purificadora sobre las secuencias de los Hominoidea.

Conclusions:

In relation to the first goal orthologs sequences of the insertion of the human HERVH48 were obtained in another 19 species of primates. Of which 4 in Platirrine and 15 in Catarrine (6 in Hominoidea and 9 in Cercopitheoidea) appear.

It has been managed to characterize open guidelines of reading of the suppressyn of at least 300 nucleotides only in sequences proceeding from the Catarrine. All the orthologs sequences of the HERFH48 obtained in the Platirrine present severe alterations in the region that would correspond to the suppressyn, therefore they must be considered not functional.

The detailed analysis of the degree of conservation of the sequence of the suppressyn in the Catarrine indicates that it was recruited before the separation of the lineages of Homonoidea and Cercapitheoidea.

The comparison of the rates of synonymous and not synonymous substitution for the gene of the suppressyn indicate, however, that only the purifying selection is acting on the sequences of the Hominoidea.

BIBLIOGRAFÍA:

- Esnault, C., Cornelis, G., Heidmann, O., & Heidmann, T. (2013). Differential Evolutionary Fate of an Ancestral Primate Endogenous Retrovirus Envelope Gene, the EnvV Syncytin, Captured for a Function in Placentation. *PLoS Genetics*, 9(3), e1003400. doi:10.1371/journal.pgen.1003400
- Goñi, I. L. (2015). *Somos lo que somos porque somos virus y bacterias: el impacto de los microorganismos endógenos en la biología del huésped. Nova Acta Científica Compostelana* (Vol. 22). Universidade de Santiago de Compostela, Servicio de Publicacións e Intercambio Científico. Retrieved from <http://www.usc.es/revistas/index.php/nacc/article/view/2512>
- Greenwood, A. D., Stengel, A., Erfle, V., Seifarth, W., & Leib-Mösch, C. (2005). The distribution of pol containing human endogenous retroviruses in non-human primates. *Virology*, 334(2), 203–213. doi:10.1016/J.VIROL.2005.01.045
- Hall. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41.
- Librado, P., & Rozas, J. (2005). *DnaSP v5: A software for comprehensive analysis of DNA poly-morphism data*. Retrieved from <http://www.ub.edu/dnasp>
- Magiorkinis, G., Belshaw, R., & Katzourakis, A. (2013). “There and back again”: revisiting the pathophysiological roles of human endogenous retroviruses in the post-genomic era. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 368(1626), 20120504. doi:10.1098/rstb.2012.0504
- Malfavon-Borja, R., & Feschotte, C. (2015). Fighting fire with fire: endogenous retrovirus envelopes as restriction factors. *Journal of Virology*, 89(8), 4047–50. doi:10.1128/JVI.03653-14
- Meyer, T. J., Rosenkrantz, J. L., Carbone, L., & Chavez, S. L. (2017). Endogenous Retroviruses: With Us and against Us. *Frontiers in Chemistry*, 5, 23. doi:10.3389/fchem.2017.00023

- Pozzi, L., Hodgson, J. A., Burrell, A. S., Sterner, K. N., Raaum, R. L., & Disotell, T. R. (2014). Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. *MOLECULAR PHYLOGENETICS AND EVOLUTION*, *75*, 165–183. doi:10.1016/j.ympev.2014.02.023
- Springer, M. S., Meredith, R. W., Gatesy, J., Emerling, C. A., Park, J., Rabosky, D. L., ... Murphy, W. J. (2012). Macroevolutionary Dynamics and Historical Biogeography of Primate Diversification Inferred from a Species Supermatrix. *PLoS ONE*, *7*(11), e49521. doi:10.1371/journal.pone.0049521
- Sugimoto, J., Sugimoto, M., Bernstein, H., Jinno, Y., & Schust, D. (2013). A novel human endogenous retroviral protein inhibits cell-cell fusion. *Scientific Reports*, *3*, 1462. doi:10.1038/srep01462
- Vargiu, L., Rodriguez-Tomé, P., Sperber, G. O., Cadeddu, M., Grandi, N., Blikstad, V., ... Blomberg, J. (2016). Classification and characterization of human endogenous retroviruses; mosaic forms are common. *Retrovirology*, *13*(1), 7. doi:10.1186/s12977-015-0232-y
- Villesen, P., Aagaard, L., Wiuf, C., & Pedersen, F. (2004). Identification of endogenous retroviral reading frames in the human genome. *Retrovirology*, *1*(1), 32. doi:10.1186/1742-4690-1-32

ANEXOS:

Anexo 1: alineamiento de las secuencias nucleotídicas de las especies de primates analizadas en este trabajo.

```

.....10.....20.....30.....40.....50.....60.....70
Homo sapiens ATGGCCTG-TATCTACCCAACCACTTTCTATACCTCTCTTCCAACCAAAAGTCTTAATATGGGAATATCC
Gorilla gorilla ATGGCCTG-TATCTACCCAACCACTTTCTATACCTCTCTTCCAACCAAAAGTCTTAATATGGGAATATCC
Pan paniscus ATGGCCTG-TATCTACCCAACCACTTTCTATACCTCTCTTCCAACCAAAAGTCTTAATACGGGAATATCC
Pan troglodytes ATGGCCTG-TATCTACCCAACCACTTTCTATACCTCTCTTCCAACCAAAAGTCTTAATACGGGAATATCC
Pongo abelli ATGGCCTG-TATCTACCCAACCACTTTCTATACCTCTCTTCCAACCAAAAGTCTTAATACAGGAATATCT
Nomascus leucogenys ATGGCCTG-TATCTACCCAACCACTTTCTATACCTCTCTTCCAACCAAAAGTCTTAATACAGGAATATCC
Theropithecus gelada full ATGGCCTG-TACCTACCCAACC---TGGTATACCTCTCTTCCAACCAAAAGTCTTAATACAGGAATATCC
Ptilocobolus tephrosceles ATGGCCTG-TACCTACCCAACCGCTTGTATACCTCTCTTCCAACCAAAAGTCTTCAATACAGGAATATCC
Mandrillus leucophaeus ATGGCCTG-TACCTACCCAACCGCTTGTATACCTCTCTTCCAACCAAAAGTCTTAATACAGGAATATCC
Papio anubis ATGGCCTG-TACCTACCCAACCGCTTGTATACCTCTCTTCCAACCAAAAGTCTTAATACAGGAATATCC
Chlorocebus sabaeus ATGGCCTG-TACCTACCCAACCGCTTGTATACCTCTCTTCCAACCAAAAGTCTTAATACAGGAATATCC
Macaca fascicularis ATGGCCTG-TACCTACCCAACCGCTTGTATACCTCTCTTCCAACCAAAAGTCTTAATACAGGAATATCC
Macaca nemestrina ATGGCCTG-TACCTACCCAACCGCTTGTATACCTCTCTTCCAACCAAAAGTCTTAATACAGGAATATCC
Macaca mulata ATGGCCTG-TACCTACCCAACCGCTTGTATACCTCTCTTCCAACCAAAAGTCTTAATACAGGAATATCC
Colobus angolensis palliatus ATGGCCTA-TATCTACCCAACCTGCTGTATACCTCTCTTCCAACCAAAAGTCTTAATACAGGAATATCC
Rhinopithecus bieti ATTGACTA-TATCTACCCAACCTGCTGTATACCTCTCTTCCAACCAAAAGTCTTAATACAGGAATATCC
Rhinopithecus roxellana ATGGCCTATATCTACCCAACCGGATGCTGTATACCTCTCTTCCAACCAAAAGTCTTCAATCAGGAACATAC
Aotus nancymae TCAACCAAAATCTTAATACAGGAACATCC
Cebus capucinus ATGGCCTGCTATCTACGCAACCACTTGTGTCTCTCACTTCCAACCAAAATCTTAATACAGGAACATCC
Callithrix jacchus ATGACGGCTATCTACGCAACCACTTGTGTCTCTCACTTCCAACCAAAATCTTAATACAGGAACATCC

.....00.....90.....100.....110.....120.....130.....140
Homo sapiens CTCACCCGATCCTAACTGTCTAGTAGCTGTCTCTGCTGTCCAACGAGCCCTCCGAGCTGCCGTGAGT
Gorilla gorilla CTCACCCGATCCTAACTGTCTAGTAGCTGTCTCTGCTGTCCAACGAGCCCTCCGAGCTGCCGTGAGT
Pan paniscus CTCACCCGATCCTAACTGTCTAGTAGCTGTCTCTGCTGTCCAACGAGCCCTCCGAGCTGCCGTGAGT
Pan troglodytes CTCACCCGATCCTAACTGTCTAGTAGCTGTCTCTGCTGTCCAACGAGCCCTCCGAGCTGCCGTGAGT
Pongo abelli CTCACCCGATCCTAACTGTCTAGTAGCTGTCTCTGCTGTCCAACGAGCCCTCCGAGCTGCCGTGAGT
Nomascus leucogenys CTCACCCGATCCTAACTGTCTAGTAGCTGTCTCTGCTGTCCAACGAGCCCTCCGAGCTGCCGTGAGT
Theropithecus gelada full CTCACCCGATCCTAACTGTCTAGTAGCTGTCTCTGCTGTCCAACGAGCCCTCCGAGCTGCCGTGAGT
Ptilocobolus tephrosceles CTCACCCGATCCTAACTGTCTAGTAGCTGTCTCTGCTGTCCAACGAGCCCTCCGAGCTGCCGTGAGT
Mandrillus leucophaeus CTCACCCGATCCTAACTGTCTAGTAGCTGTCTCTGCTGTCCAACGAGCCCTCCGAGCTGCCGTGAGT
Papio anubis CTCACCCGATCCTAACTGTCTAGTAGCTGTCTCTGCTGTCCAACGAGCCCTCCGAGCTGCCGTGAGT
Chlorocebus sabaeus CTCACCCGATCCTAACTGTCTAGTAGCTGTCTCTGCTGTCCAACGAGCCCTCCGAGCTGCCGTGAGT
Macaca fascicularis CTCACCCGATCCTAACTGTCTAGTAGCTGTCTCTGCTGTCCAACGAGCCCTCCGAGCTGCCGTGAGT
Macaca nemestrina CTCACCCGATCCTAACTGTCTAGTAGCTGTCTCTGCTGTCCAACGAGCCCTCCGAGCTGCCGTGAGT
Macaca mulata CTCACCCGATCCTAACTGTCTAGTAGCTGTCTCTGCTGTCCAACGAGCCCTCCGAGCTGCCGTGAGT
Colobus angolensis palliatus CTCACCCGATCCTAACTGTCTAGTAGCTGTCTCTGCTGTCCAACGAGCCCTCCGAGCTGCCGTGAGT
Rhinopithecus bieti CTCACCCGATCCTAACTGTCTAGTAGCTGTCTCTGCTGTCCAACGAGCCCTCCGAGCTGCCGTGAGT
Rhinopithecus roxellana CTCACCCGATCCTAACTGTCTAGTAGCTGTCTCTGCTGTCCAACGAGCCCTCCGAGCTGCCGTGAGT
Aotus nancymae CTCACCCGATCCTAACTGTCTAGTAGCTGTCTCTGCTGTCCAACGAGCCCTCCGAGCTGCCGTGAGT
Cebus capucinus CTCACCCGATCCTAACTGTCTAGTAGCTGTCTCTGCTGTCCAACGAGCCCTCCGAGCTGCCGTGAGT
Callithrix jacchus TTCACCCATCCTAACTGTCTAGTAGCTGTCTCTGCTGTCCAACGAGCCCTCCGAGCTGCCAT----

.....150.....160.....170.....180.....190.....200.....210
Homo sapiens GTT-AT-CAGTCTTTGCACCTACGAGGGGAGATGCAACAATCTTTACTTACCATACTCA-----
Gorilla gorilla GTT-AT-CAGTCTTTGCACCTACGAGGGGAGATGCAACAATCTTTACTTACCATACGCA-----
Pan paniscus GTT-AT-CAGTCTTTGCACCTACGAGGGGAGATGCAACAATCTTTACTTACCATACTCA-----
Pan troglodytes GTT-AT-CAGTCTTTGCACCTACGAGGGGAGATGCAACAATCTTTACTTACCATACTCA-----
Pongo abelli GTT-AT-CAGTCTTTGCACCTACGAGGGGAGATGCAACAATCTTTACTTACTATACTCA-----
Nomascus leucogenys GTT-AT-CAGTCTTTGACTACGAGGGGAGAGGCAACAATCTTTACTTACCATACTCA-----
Theropithecus gelada full GTT-AT-CAGTCTTTCCACTACGAGGAAAGATTCAACAATCCTTTACTTACCATACTCA-----
Ptilocobolus tephrosceles GTT-AT-CAGTCTTTCCACTACGAGGAAAGATTCAACAATCCTTTACTTACCATACTCA-----
Mandrillus leucophaeus GTT-AT-CAGTCTTTCCACTACGAGGAAAGATTCAACAATCCTTTACTTACCATACTCA-----
Papio anubis GTT-AT-CAGTCTTTCCACTACGAGGAAAGATTCAACAATCCTTTACTTACCATACTTA-----
Chlorocebus sabaeus GTT-AT-CAGTCTTTCCACTACGAGGAAAGATTCAACAATCCTTTACTTACCATACTCA-----
Macaca fascicularis GTT-AT-CAGTCTTTCCACTACGAGGAAAGATTCAACAATCCTTTACTTACCATACTCA-----
Macaca nemestrina GTT-AT-CAGTCTTTCCACTACGAGGAAAGATTCAACAATCCTTTACTTACCATACTCA-----
Macaca mulata GTT-AT-CAGTCTTTCCACTACGAGGAAAGATTCAACAATCCTTTACTTACCATACTCA-----
Colobus angolensis palliatus GTT-AT-CAATTTTCCACTGTGGAGGAAATGCAACAATCCTTTACTTACCATACTCA-----

Rhinopithecus bieti GTT-AT-CAGTCTTTCCACTGTGGAGGAAATGCAACAATCCTTTACTTACCATACTCA-----
Rhinopithecus roxellana GCT-ATTCAGTCTTTCCACCAGGAAAGCTGGCAATCCTTCACTCAGCATACTCATGTCAACAGT
Aotus nancymae GTT-AT-CAGTCTTTCCACTGTGGAGGAAATGCAATAATCCTGTACTTACCATACTCA-----
Cebus capucinus ATTAAT-TGGTCTTTCCATTATGGAGGAAATGCAACAATCCTTTACTTACCATCTTCA-----
Callithrix jacchus GTTAAT-CAGTCTTTCCATTATGGAGGAAATGCAACAATCCTTTACTTACCATCTTCA-----

```

220 230 240 250 260 270 280

Homo sapiens TATGAAAGATCCTGTTATGGAAACCTTAATCGAGGAATGTCTTGAATCAGG
 Gorilla_gorilla TATGAAAGATCCTGTTATGGAAACCTTAATCGAGGAATGTCTTGAATCAGG
 Pan_paniscus TATGAAAGATCCTGTTATGGAAACCTTAATCGAGGAATGTCTTGAATCAGG
 Pan_troglodytes TATGAAAGATCCTGTTATGGAAACCTTAATCGAGGAATGTCTTGAATCAGG
 Pongo_abelii TATGAAAGATCCTGTTATGGAAACCTTAATCGAGGAATGTCTTGAATCAGG
 Nomascus_leucogenys TATGAAAGATCCTGTTATGGAAACCTTAATCGAGGAATGTCTTGAATCAGG
 Theropithecus_gelada_full TATGAAAGATCCTGTTATGGAAACCTTAATCGAGGAATGTCTTGAATCAGG
 Ptilocolobus_LephrosCeles CATCCTGTTATGGAAACCTTAATCGAGGAATGTCTTGAATCAGG
 Mandrillus_Leucophaeus TATGAAAGATCCTGTTATGGAAACCTTAATCGAGGAATGTCTTGAATCAGG
 Papio_anubis CATGAAAGATCCTGTTATGGAAACCTTAATCGAGGAATGTCTTGAATCAGG
 Chlorocebus_sabaeus CATGAAAGATCCTGTTATGGAAACCTTAATCGAGGAATGTCTTGAATCAGG
 Macaca_fascicularis CATGAAAGATCCTGTTATGGAAACCTTAATCGAGGAATGTCTTGAATCAGG
 Macaca_nemestrina CATGAAAGATCCTGTTATGGAAACCTTAATCGAGGAATGTCTTGAATCAGG
 Macaca_mulata CATGAAAGATCCTGTTATGGAAACCTTAATCGAGGAATGTCTTGAATCAGG
 Colobus_angolensis_palliatu TATGAAAAACCTGTTATGGAAACCTTAATCGAGGAATGTCTTGAATCAGG
 Rhinopithecus_bieti TATGAAAAACCTGTTATGGAAACCTTAATCGAGGAATGTCTTGAATCAGG
 Rhinopithecus_roxellana CCTTCACTCACTATACTCATACGAAAGATCCTGTTGGGAAACCTCAGTCCAGGAATGTCTTGAATCAGG
 Aotus_nancymaae TAGGAAAAAGCCCTACTATGGAACTTTAGTCAAGGAATCGGGTGAATCAGG
 Cebus_capucinus TATGAAAAACCTGTTATGGAAACCTTAATCGAGGAATGTCTTGAATCAGG
 Callithrix_jacchus TGTGAAAAACCTGTTATGGAAACCTTAATCGAGGAATGTCTTGAATCAGG

290 300 310 320 330 340 350

Homo sapiens AAAGAGTTATTATAAAGTAAAGAA-TCTAGGAGTATGTGGCAGTCGTAATGGGGCTATTTGCCCCAGAGG
 Gorilla_gorilla AAAGAGTTATTATAAAGTAAAGAA-TCTAGGAGTATGTGGCAGTCGTAATGGGGCTATTTGCCCCAGAGG
 Pan_paniscus AAAGAGTTATTATAAAGTAAAGAA-TCTAGGAGTATGTGGCAGTCGTAATGGGGCTATTTGCCCCAGAGG
 Pan_troglodytes AAAGAGTTATTATAAAGTAAAGAA-TCTAGGAGTATGTGGCAGTCGTAATGGGGCTATTTGCCCCAGAGG
 Pongo_abelii AAAGAGTTATTATAAAGTAAAGAA-TCTAGGAGTATGTGGCAGTCGTAATGGGGCTATTTGCCCCAGAGG
 Nomascus_leucogenys AAAGAGTTATTATAAAGTAAAGAA-TCTAGGAGTATGTGGCAGTCGTAATGGGGCTATTTGCCCCAGAGG
 Theropithecus_gelada_full AAAGAGTTATTATAAAGTAAAGAA-TCCAGGAGTATCTGGCAGTCGTAATGGGGCTATATGCCCCAGAGG
 Ptilocolobus_LephrosCeles AAAGAGTTATTATAAAGTAAAGAA-TCTAGGAGTATCTGGCAGTCGTAATGGGGCTATATGCCCCAGAGG
 Mandrillus_Leucophaeus AAAGAGTTATTATAAAGTAAAGAA-TCCAGGAGTATCTGGCAGTCGTAATGGGGCTATATGCCCCAGAGG
 Papio_anubis AAAGAGTTATTATAAAGTAAAGAA-TCCAGGAGTATCTGGCAGTCGTAATGGGGCTATATGCCCCAGAGG
 Chlorocebus_sabaeus AAAGAGTTATTATAAAGTAAAGAA-TCTAGGAGTATCTGGCAGTCGTAATGGGGCTATATGCCCCAGAGG
 Macaca_fascicularis AAAGAGTTATTATAAAGTAAAGAA-TCTAGGAGTATCTGGCAGTCGTAATGGGGCTATATGCCCCAGAGG
 Macaca_nemestrina AAAGAGTTATTATAAAGTAAAGAA-TCTAGGAGTATCTGGCAGTCGTAATGGGGCTATATGCCCCAGAGG
 Macaca_mulata AAAGAGTTATTATAAAGTAAAGAA-TCTAGGAGTATCTGGCAGTCGTAATGGGGCTATATGCCCCAGAGG
 Colobus_angolensis_palliatu GAAAGGTTATTATAAAGTAAAAAA-TCCAGGAGTATCCGGCAGTCGTAATGAGGCTATATGTTCAAAGG
 Rhinopithecus_bieti GAAAGGTTATTATAAAGTAAAAAA-TCCAGGAGTATCCGGCAGTCGTAATGAGGCTATATGTTCAAAGG
 Rhinopithecus_roxellana AAAGAGTTATTATAAAGTAAAAAA-TCTAGGAGTATCTGGCAGTCGTAATGGGGCTATATGCCCCAGAGG
 Aotus_nancymaae GAAAGGTTATTATAAAGTAAAAAA-TCTAGGAGTATCTGGCAGTCGTAATGGGGCTATATGCCCCAGAGG
 Cebus_capucinus GAAAGGTTATTATAAAGTAAAAAA-TCTAGGAGTATCTGGCAGTCGTAATGGGGCTATATGCCCCAGAGG
 Callithrix_jacchus GAAAGGTTATTATAAAGTAAAAAA-TCTAGGAGTATCTGGCAGTCGTAATGAGGCTGTACGTCGCCAAGG

360 370 380 390 400 410 420

Homo sapiens GAAGCAGTGGCTTTGCTTCAACCAAAATTGG-ACAATGGGGA-GTAAACACTCAGGTGCTTGGGGACATA
 Gorilla_gorilla GAAGCAGTGGCTTTGCTTCAACCAAAATTGG-ACAATGGGGA-GTAAACACTCAGGTGCTTGGGGACATA
 Pan_paniscus GAAGCAGTGGCTTTGCTTCAACCAAAATTGG-ACAATGGGGA-GTAAACACTCAGGTGCTTGGGGACATA
 Pan_troglodytes GAAGCAGTGGCTTTGCTTCAACCAAAATTGG-ACAATGGGGA-GTAAACACTCAGGTGCTTGGGGACATA
 Pongo_abelii GAAGCAGTGGCTTTGCTTCAACCAAAATTGG-ACAATGGGGA-GTAAACACTCAGGTGCTTGGGGACATA
 Nomascus_leucogenys GAAGCAGTGGCTTTGCTTCAACCAAAATTGG-ACAATGGGGA-GTAAACACTCAGGTGCTTGGGGACATA
 Theropithecus_gelada_full GAAGCAGTGGCTTTGCTTCAACCAAAATTGG-ACAATGGGGA-GTAAACACTCAAGTACTCGAGGACATA
 Ptilocolobus_LephrosCeles GAAGCAGTGGCTTTGCTTCAACCAAAATTGG-ACAATGGGGA-GTAAACACTCAAGTACTCAAGGACATA
 Mandrillus_Leucophaeus GAAGCAGTGGCTTTGCTTCAACCAAAATTGG-ACAATGGGGA-GTAAACACTCAAGTACTCGAGGACATA
 Papio_anubis GAAGCAGTGGCTTTGCTTCAACCAAAATTGG-ACAATGGGGA-GTAAACACTCAAGTACTCGAGGACATA
 Chlorocebus_sabaeus GAAGCAGTGGCTTTGCTTCAACCAAAATTGG-ACAATGGGGA-GTAAACACTCAAGTACTCGAGGACATA
 Macaca_fascicularis GAAGCAGTGGCTTTGCTTCAACCAAAATTGG-ACAATGGGGA-GTAAACACTCAAGTACTCGAGGACATA
 Macaca_nemestrina GAAGCAGTGGCTTTGCTTCAACCAAAATTGG-ACAATGGGGA-GTAAACACTCAAGTACTCGAGGACATA
 Macaca_mulata GAAGCAGTGGCTTTGCTTCAACCAAAATTGGGAAACCCATGGGGAATGAACACTCAAGTACTCGAGGACATA
 Colobus_angolensis_palliatu TAGGCAGTGGCTTTGCTTCAACCAAAATTGG-ACAATGGGGA-GTAAACACTCAAGTACTCAAGGACATA
 Rhinopithecus_bieti TAGGCAGTGGCTTTGCTTCAACCAAAATTGG-ACAATGGGGA-GTAAACACTCAAGTACTAGAGACATA
 Rhinopithecus_roxellana GAAGCAGTGGCTTTGCTTCAACCAAAATTGG-ACAATGGGGA-GTAAACACTCAAGTACTCGAGGACATA
 Aotus_nancymaae TAAGCAGCAGCTTTGCTTCACTAAAATTGG-ACAATGGGGT-GTAAACACTCAAGTACTAGAGACATA
 Cebus_capucinus TAAGCAGTGGCTTTGCTTCAACCAAAATTGG-ACAAT--AA-GTAAACACTCAAGTACTCGAGGACATA
 Callithrix_jacchus TAAGCAGTGGCTTTGCTTCAACCAAAATTGG-ACAAT--GA-GTAAACTCAAGTACTCGAGGACATA

