

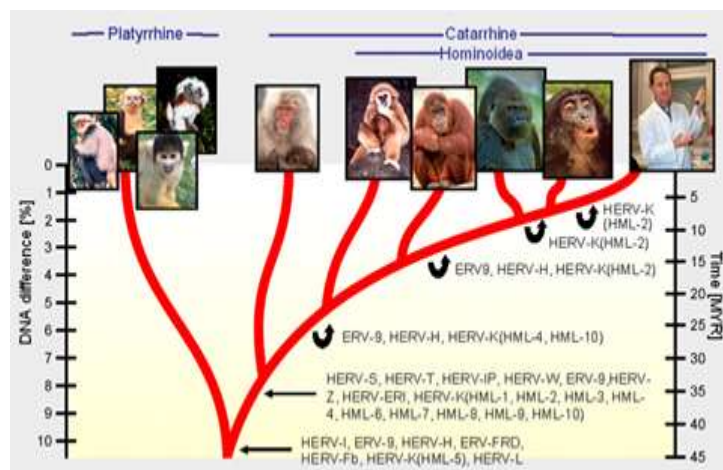
Grao en Bioloxía

Memoria do Traballo de Fin de Grao

Revisión bibliográfica: Análise da variación xenómica e da evolución do grupo Fb de retrovirus endóxenos humanos

Revisión bibliográfica: Análisis de la variación genómica y la evolución del grupo Fb de retrovirus endógenos humanos

Literature review: Analysis of genomic variation and evolution of Fb, a group of human endogenous retroviruses



María del Carmen Touzón Fernández

Julio, 2018

Director Académico: Horacio Naveira Fachal

Foto de la portada: *Integration and expansion of endogenous retroviruses in Primates*, VIRO
Institute of Virology, German Research Center for Environmental Health

<https://www.helmholtz-muenchen.de/viro/research/working-groups/human-endogenous-retroviruses/research-topics/index.html>

Análise da variación xenómica e da
evolución do grupo Fb de retrovirus
endóxenos humanos

Análisis de la variación genómica y de la
evolución del grupo Fb de retrovirus
endógenos humanos

Analysis of genomic variation and evolution
of Fb, a group of human endogenous
retroviruses

ÍNDICE

1. Introducción-----	1
2. Objetivos-----	5
3. Material y Métodos:	
3.1. Búsqueda de secuencias provirales de interés-----	5
3.2. Identificación de las TSD (<i>Target Site Duplications</i>)-----	6
3.3. Estudio de las LTR (<i>Long Terminal Repeat</i>)-----	6
3.4. Estudio de las regiones codificantes retrovirales-----	8
3.5. Reconstrucción de la secuencia activa-----	9
3.6. Reconstrucción filogenética-----	11
3.7. Gestor bibliográfico-----	12
4. Resultados:	
4.1. Búsqueda de secuencias provirales de interés-----	12
4.2. Identificación de las TSD (<i>Target Site Duplications</i>)-----	13
4.3. Estudio de las LTR (<i>Long Terminal Repeat</i>)-----	14
4.4. Estudio de las regiones codificantes retrovirales-----	15
4.5. Reconstrucción de la secuencia activa-----	16
4.6. Reconstrucción filogenética-----	17
5. Discusión:	
5.1. Alto grado de conservación del grupo HERV-Fb1-----	18
5.2. Funciones de la supresina-----	19
5.3. Funciones de los HERVs en el proceso de la placentación y patologías relacionadas-----	19
6. Conclusiones-----	20
7. Bibliografía-----	20

RESUMEN

El grupo de retrovirus endógenos humanos HERV-Fb1, perteneciente al grupo de los gammaretrovirus (grupo I) presenta algunas características que lo diferencian de los demás grupos. Este retrovirus es el precursor de la supresina. El gen de la supresina presenta una inserción muy conservada a lo largo del tiempo localizada en el cromosoma 21 en *Homo sapiens*. Para llegar a esta conclusión, fue necesario buscar posibles inserciones en los genomas. Por otra parte, fue necesario intentar de reconstruir la posible secuencia activa original y determinar el tiempo de inserción. Así fue posible reconstruir la filogenia que demuestra la antigüedad de la inserción y el alto grado de conservación de la misma tanto en *Homo sapiens* como en otras especies de primates catarrinos.

RESUMO

O grupo de retrovirus endógenos humanos HERV-Fb1, que pertence ao grupo dos gammaretrovirus (grupo I) presenta algunhas características que o diferencian dos demais grupos. Este retrovirus é o precursor da supresina. O xen da supresina presenta unha inserción moi conservada ao longo do tempo que se localiza no cromosoma 21 en *Homo sapiens*. Para chegar a esta conclusión, foi necesario buscar posibles insercións nos xenomas. Por outra parte, foi necesario tentar reconstruír a posible secuencia activa orixinal e determinar o tempo de inserción. Así foi posible reconstruír a filoxenia que demostra a antigüidade da inserción e o alto grado de conservación da mesma tanto en *Homo sapiens* como en outras especies de primates catarrinos.

SUMMARY

The group of endogenous human retroviruses HERV-Fb1, belonging to the group of gammaretrovirus (group I) has some characteristics that differentiate it from the other groups. This retrovirus is the precursor of suppressin. The gene of the supresin presents / displays a very conserved insertion throughout the time located in the chromosome 21 in *Homo sapiens*. To reach this conclusion, it was necessary to look for possible insertions in the genomes. On the other hand, it was necessary to try to reconstruct the possible original active sequence and determine the time of insertion. Thus it was possible to reconstruct the phylogeny that shows the antiquity of the insertion and the high degree of

preservation of it in both *Homo sapiens* and other species of catarrhine primates.

PALABRAS CLAVE

HERV-Fb1, Supresina, Sincitina-1, Región *env*, Catarrinos.

HERV-Fb1: Grupo de retrovirus endógenos humanos pertenecientes al grupo I o gammaretrovirus y que son precursores del gen de la supresina.

Supresina: Gen con alta expresión en la placenta in vivo de primates. Deriva del HERV-Fb1.

Sincitina-1: Proteína resultado de una infección viral hace 25 millones de años (Mya) de nuestros ancestros primates secretada en la superficie del embrión en desarrollo.

Región *env* : Región del retrovirus que se encarga de sintetizar las proteínas de la envuelta.

Catarrinos: Monos del viejo mundo que se caracterizan por presentar los orificios nasales abiertos hacia abajo y separados por un delgado tabique nasal.

1. INTRODUCCIÓN

Los Retrovirus Endógenos Humanos (HERVs) son secuencias de pasadas infecciones de células germinales por retrovirus exógenos y la posterior transmisión de los provirus integrados a los descendientes. Constituyen el 8-10% del genoma humano. Se componen de una secuencia interna formada por los cuatro genes retrovirales (*gag*, *pro*, *pol* y *env*) y de dos LTR (*Long Terminal Repeat*) en cada extremo de la secuencia interna (Vargiu et al., 2016). El gen *gag* codifica las proteínas estructurales internas; el gen *pol* codifica la transcriptasa inversa, la cual es esencial para el proceso de retrotranscripción y la integrasa que va a permitir que el retrovirus se integre en una región u otra del genoma; el gen *pro* que codifica la proteasa, encargada de ensamblar, dividir y cortar las cadenas largas de proteínas y el gen *env* que va a ser el encargado de codificar las proteínas de la envuelta (Cordeiro & Taroco, 1980).

El 45% del genoma se corresponde con diferentes tipos de retroelementos, es decir, elementos que se transcriben en el ARN, que tienen transcripción inversa en el ADN y que se insertan en un nuevo sitio del genoma. Estos retroelementos son: SINEs (*Short Interspersed Nuclear Elements*), que son elementos no autónomos, LINEs (*Long Interspersed Nuclear Elements*), que son elementos autónomos y LTR (*Long Terminal Repeat*) (Sverdlov, 2005). Los extremos LTR se generan durante el proceso de retrotranscripción e intervienen en la integración viral en el DNA genómico (Cordeiro & Taroco, 1980). En los extremos de las LTR se encuentran las *Target Site Duplications*. Las TSD (*Target Site Duplications*) son repeticiones idénticas que se producen en sitios de inserción de elementos transponibles, que llevan como función rellenar los extremos escalonados por las transposasas (Linheiro & Bergman, 2012).

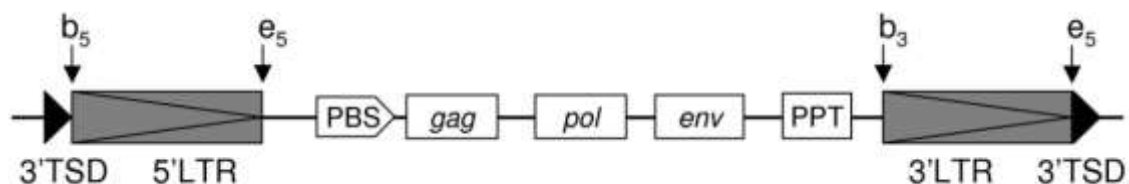


Figura 1. Esquema de retrovirus completo con la inserción de las TSD.

Los HERVs se han encontrado en todos los vertebrados, incluidos los seres humanos. La conservación de estos dentro del DNA humano a lo largo del tiempo, puede tener efectos en el organismo hospedador. La posible patogenicidad de los retrovirus estimuló el estudio de los mismos, encontrando correlaciones entre ellos y diferentes enfermedades humanas como el cáncer, la esclerosis múltiple, enfermedades autoinmunes, disfunciones placentarias, etc. (Vargiu et al., 2016)

Este trabajo se va a centrar en el estudio del grupo HERV-Fb1, perteneciente a la familia de HERV-F que se incluye en el grupo I de retrovirus o también llamados gammaretrovirus, que aparece representada en la tabla 1 (Nelson et al., 2003).

HERV family	Representative accession number
<i>Class I HERVs (type C related HERVs)</i>	
Group 1, HERV-HF	
HERV-H (RTVL-H, RGH)	AF108842
HERV-F	AF070684
Group 2, HERV-RW	
HERV-W	AF072506
HERV-R (ERV9)	X57147
HERV-P (HuERS-P, HuRRS-P)	X06279
Group 3, HERV-ER1	
HERV-E (4-1, ERVA, NP-2*)	S46403
51-1	J00273
HERV-R (ERV3)	M12140
RRHERV-I	M64936
Group 4, HERV-T	
HERV T (S71, CRTK1, CRTK6)	M32788
Group 5, HERV-IP	
HERV-I (RTVL-I)	X14953
HERV-IP-T47D (ERV-FTD)	U27241
Group 6, ERV-FRD	
ERV-FRD	U27240
<i>Class II HERVs (type A, B, and D related HERVs)</i>	
Group 1, HERV-K (HML-1)	
HERV-K (HML-1.1)	U35102
Group 2, HERV-K (HML-2)	
HERV-K10	M14123
HERV-K-HTDV	X8227
Group 3, HERV-K (HML-3)	
HERV K (HML3.1)	U35153
Group 4, HERV-K (HML-4)	
HERV-K-T47D	AF020092
Group 5, HERV-K (HML-5)	
HERV-K-NMWW2	AF015995
Group 6, HERV-K (HML-6)	
HERV K (HML-6p)	U86698
Group 7, HERV-K (HML-7)	
HERV-K-NMWW7	AF016000
Group 8, HERV-K (HML-8)	
HERV-K-NMWW3	AF015996
Group 9, HERV-K (HML-9)	
HERV-K-NMWW9	AF016001
Group 10, HERV-K (HML-10)	
HERV-KC4	U07856
<i>Class III Foamy virus related HERVs</i>	
HERV-L	X89211

Tabla 1. Clasificación de los diferentes grupos de HERVs.

HERV-Fb1 es el precursor de la supresina. Este grupo de retrovirus tiene una peculiaridad, y es que presenta una inserción en el cromosoma 21q22.3 de *Homo sapiens*. Por lo tanto este trabajo se va a centrar en estudiar el grado de conservación del gen de la supresina en esa inserción en la especie humana en comparación con otros primates.

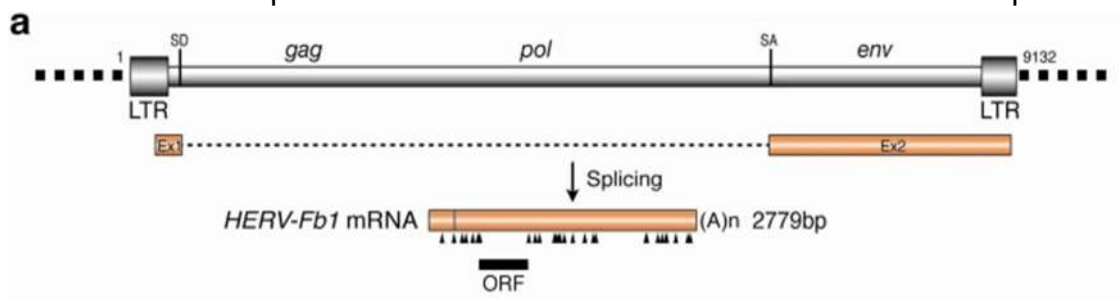


Figura 2. Patrón de empalme del gen HERV-Fb1 y su producto de transcripción.

La supresina es un gen derivado del HERV-Fb1 que se localiza en el cromosoma 21q22.3 en la región de la envuelta y sólo se transcribe en la placenta in vivo (Vargiu et al., 2016). Este gen está altamente conservado en la placenta humana (Sugimoto et al., 2016). La supresina se une a los receptores de la sincitina-1, que es otra proteína derivada de la envuelta de otro HERV. El resultado de esta unión va a provocar que no se produzca la fusión celular o también llamada sincitiazación, que da lugar a la formación de la placenta en primates. Este gen se encuentra situado en el dominio inmunosupresivo (ISU) localizado a su vez en la región transmembrana (TM) de la envuelta del retrovirus (Vargiu et al., 2016).

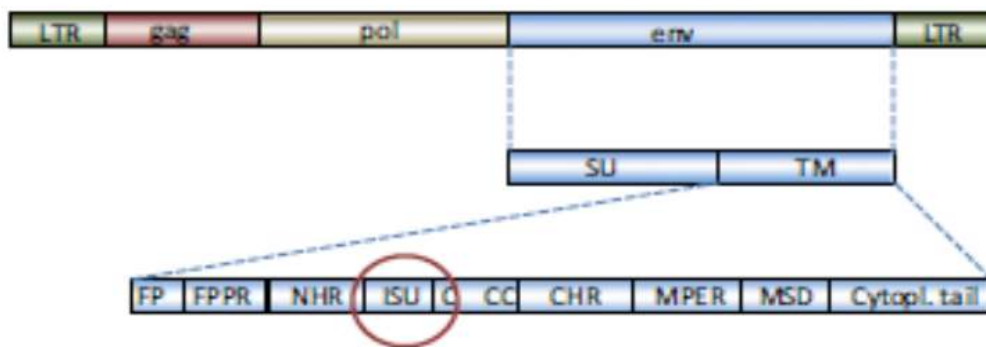


Figura 3. Esquema de la secuencia del retrovirus integrado (provirus). LTR: long terminal repeat; gag: antígeno específico de grupo, codificación de las proteínas del grupo; pol: gen de la polimerasa, codificación para la transcriptasa inversa, proteasa e integrasa; env: proteínas de la envuelta; SU: proteínas de superficie; TM: proteínas transmembrana; FP: péptido fusión; FPPR: región proximal del péptido fusión; NHR: N región helicoidal; ISU: dominio inmunosupresivo; C-CC: lazo cisteína-cisteína; CHR: región C-terminal helicoidal; MPER: región externa proximal de la membrana; MSD: dominio de membrana; Cytopl. tail: cola citoplasmática.

La placenta humana es un tejido que presenta alta expresión de HERVs, ya que hay estudios comparativos de la expresión de las proteínas de la envuelta en la placenta y en otros 18 tejidos, y se ha demostrado una mayor expresión en la placenta, por lo que se afirma esta hipótesis. Los retrovirus endógenos están regulados hormonalmente (Chen et al., 2006). Para que funcione un sincitio, los HERVs deben siempre ser fusogénicos (ser capaces de inducir la fusión de células), expresarse específicamente en la placenta y estar altamente conservados (Cornelis et al., 2015).

La sincitina cumple estos tres criterios para que el sincitio sea funcional: es fusogénica, tiene expresión específica en la placenta y está altamente conservada. En primates, existen dos clases de sincitina, la sincitina 1 y la sincitina 2 (Denner, 2016), que aparecen incluidas en la clasificación de las proteínas de la envuelta retroviral realizadas por este mismo autor (Tabla 2). La

sincitina 1 aparece integrada en los antepasados de los monos hominoides y en los catarrinos (Monos del Viejo Mundo).

Table 2. Overview about the syncytin-like retroviral envelope proteins involved in placentation of eutherian mammals

Name, retroviral origin	Species	Order/family/subfamily	Type of placenta	Time of integration Mya	Fusion capacity	Immuno-suppressive properties ¹
Syncytin-1	Human	Primates	Haemochorial	25	Yes	No
Syncytin-2	Human	Primates	Haemochorial	>40	Yes	Yes
ERV3	Human	Primates	Haemochorial		No	Yes
EnvV1 syncytin	Old world monkeys	Primates	Haemochorial	40	Yes/no ³	
EnvV2 syncytin	Old world monkeys	Primates	Haemochorial	45	Yes/no ³	
Env-Cav1	Guinea pig	Caviomorpha	Endotheliochorial	30	Yes	
Syncytin-Rum1	Cattle	Ruminantia	Synepitheliochorial	30	Yes	Yes
BERV-K1 (Fermatrin-1)	Cattle	Ruminantia	Synepitheliochorial	20	Yes	
bERVE	Cattle	Ruminantia	Synepitheliochorial	na		
BERV-P	Cattle	Ruminantia	Synepitheliochorial	17		
enJSRV	Sheep	Caprinae	Synepitheliochorial	7	Yes	
Syncytin-A	Mouse	Rodentia	Haemochorial	40	Yes	No
Syncytin-B	Mouse	Rodentia	Haemochorial	40	Yes	Yes
Syncytin-Carl	Cat, dog	Carnivora	Endotheliochorial	60-85		
Mar-1	Squirrel	Rodentia (Sciurida)		25	Yes	
Syncytin-Opo1 ²	Opossum	Marsupialis			Yes	Yes
X ²	Opossum	Marsupialis		80	No	Yes
Syncytin-Ten1	Lesser Hedgehog tenrec	Afrosoricida Theracidae	Haemochorial	40-60	Yes	
Syncytin-Ory1	Rabbit, Hares	Lagomorpha (Leporidae)		12	Yes	
?? ⁵	Pig	Artiodactyla	Epitheliochorial		?	?

Tabla 2. Clasificación de las proteínas de la envoltura retroviral de tipo sincitina implicadas en la placentación de mamíferos euterios (Denner, 2016).

El gen de la supresina se une a los receptores de la proteína sincitina (syn1) e inhibe la fusión celular o también llamada sincitización en mamíferos. La sincitina 1 y la sincitina 2 fueron observadas de forma muy conservada en catarrinos (monos del viejo mundo) y en plattirinos (monos del nuevo mundo) respectivamente (Gibb et al., 2015; Greenwood et al., 2005). El fenómeno evolutivo de la placentación aparece hace 150 Mya. Los sincitios originales han sido reemplazados por otros nuevos. La expresión de al menos dos sincitinas diferentes en una especie está asociada a que han evolucionado a lo largo del tiempo, es decir, una especie adquirió primero una sincitina y más tarde la otra gracias a su actividad fusogénica. Los HERVs expresados en la placenta están regulados hormonalmente por la sincitina 1, que tiene como función fusionar las vellosidades del citotrofoblasto en la placenta humana, para formar el sincitiotrofoblasto multinucleado (Denner, 2016). La sincitina 1 no tiene función inmunosupresora, es decir, no va a inhibir componentes del sistema inmunitario. La sincitina 2 sí que presenta función inmunosupresora (Denner, 2016). Estas proteínas de la envuelta retroviral (sincitina 1 y sincitina 2) van a tener un papel importante en el desarrollo, ya que son esenciales en la formación de sincitios o unión de células en la placenta (Greenwood et al., 2005).

Existe un inhibidor que es codificado por las células huésped, llamado ASCT-2 que participa en la transmisión viral. En los estudios realizados sobre este tema, se detectó supresina en la hormona gonadotropina coriónica humana (hCG) en células de la placenta humana en el primer y tercer trimestre del embarazo. Existen evidencias de una posible doble función de Fb1 en su retrovirus exógeno ancestral, o lo que es lo mismo, un mecanismo de interferencia viral o superinfección. La proteína Fb1 pudo haber sido parte de un mecanismo de superinfección en el que participaban codificado HERV-Fb y HERV-W, que compartían la preferencia por la entrada mediada por ASCT-2 en las células huésped (Sugimoto et al., 2013).

2. OBJETIVOS

Con este trabajo se va a investigar sobre Retrovirus Endógenos Humanos (HERVs), se va a mejorar el manejo de las bases de datos, también va a permitir ampliar conocimientos científicos sobre la supresina y los primates. Por supuesto, también se va a mejorar mucho la técnica de saber elegir y seleccionar artículos científicos, así como el manejo de gestores bibliográficos y otras muchas herramientas científicas.

3. MATERIAL Y MÉTODOS

3. 1. Búsqueda de secuencias provirales de interés

Lo primero que se llevó a cabo fue la búsqueda de secuencias en el “*National Center for Biotechnology Information*” (NCBI, <https://www.ncbi.nlm.nih.gov/>), empleando como palabra clave el nombre de la secuencia proviral completa (AB610407) que fue proporcionada por mi tutor, la cual corresponde a la secuencia completa de ERVH48-1 en *Homo sapiens*. La base de datos “Gene” proporcionó la secuencia completa, en la que está incluida la secuencia de la supresina, cuya secuencia completa se descargó en formato fasta para su posterior análisis.

Otro paso esencial para llevar a cabo el análisis fue tratar de encontrar posibles inserciones de HERV-Fb1 en otros genomas, por lo que empleamos la herramienta “*Basic Local Alignment Search Tool*” (BLAST, <https://blast.ncbi.nlm.nih.gov/Blast.cgi>). BLAST permite comparar nuestra “query” o secuencia de interés y numerosas secuencias del “GenBank”, realizando un alineamiento. La “query” empleada fue la secuencia proviral completa.

Lo siguiente que se llevó a cabo fue utilizar la herramienta “*Blast Like Alignment Tools*” (BLAT, <https://genome.ucsc.edu/cgi-bin/hgBlat>). Esta herramienta permite realizar una búsqueda directa sobre un genoma completo, permitiendo encontrar secuencias con un 95% de similitud. La base de datos

empleada del genoma fue (“assembly”) utilizada fue la (GRCh38/hg38). Al igual que antes, la secuencia empleada fue la proviral completa.

Por último, se hizo uso del navegador genómico “*Ensembl Genome Browser 90*” (<http://www.ensembl.org/index.html>). Este navegador genético anota genes, calcula múltiples alineamientos, predice la función reguladora y recopila datos sobre enfermedades. Se seleccionó la opción “all primates”, para comparar el grado de conservación tanto en *Homo sapiens* como en otras especies de primates.

3.2. Identificación de las TSD (*Target Site Duplications*)

Las *Target Site Duplications* (TSD) son repeticiones directas que se producen en sitios de inserción de elementos transponibles que se cree que ocurren debido al relleno de los extremos adhesivos causados por el corte escalonado por transposasas. Ellos flanquean elementos transponibles y pueden ser utilizados para encontrar sus loci en el genoma (Linheiro & Bergman, 2012).

Gracias a la identificación de las TSD se pudo determinar si hubo más de una inserción a lo largo de toda la historia evolutiva. Es decir, nos ha permitido saber si consisten en inserciones retrovirales originales. La búsqueda de esas posibles inserciones fue realmente sencilla, ya que esa información fue proporcionada por el programa LTR_STRUC (versión 1.1) . Este programa es necesario para el estudio de la divergencia de las LTR, por lo que se profundizará más sobre el mismo en el siguiente apartado (“Estudio de las LTR”).

3.3. Estudio de las LTR (*Long Terminal Repeat*)

Las LTR mantienen la estructura de las regiones *gag*, *pro*, *pol* y *env* del retrovirus (Cordeiro & Taroco, 1980). Además de cumplir esta función estructural, gracias a ellas se pudo determinar el tiempo de inserción del retrovirus en el genoma. Si las LTR son iguales nos indica que el retrovirus es reciente. A medida que transcurre el tiempo se van acumulando diferencias entre ellas, debidas a sustituciones nucleotídicas e indels (Coffin, Hughes, & Varmus, 1997).

Desde la página del NCBI (<https://www.ncbi.nlm.nih.gov/>) con el acceso de referencia (AB610407), en la base de datos “Gene” se amplió la región analizada hasta 1000 pb “aguas arriba” y otros tantos “aguas abajo” para poder encontrar las LTR.

Para comparar la secuencia de *Homo sapiens* con otras posibles inserciones en otras especies, se fue al apartado de “orthologs” en la página del NCBI para buscar posibles coincidencias en otras especies. Las coincidencias encontradas fueron 15 en total pero finalmente se seleccionaron estas: *Gorilla gorilla*, *Pan paniscus*, *Pongo abelii*, *Nomascus leucogeny* y *Pan troglodytes*. Al igual que antes, se añadieron a la secuencia de cada especie 1000 pares de bases (pb) también a cada extremo, para la observación de las LTR. Las inserciones de estas especies estaban todas en el cromosoma 21 al igual que la del *Homo sapiens*, excepto la del gibbon (*Nomascus leucogenys*) que estaba en el cromosoma 25. Una vez añadidos los pares de bases a cada lado, se procedió a realizar un alineamiento múltiple para el estudio de la divergencia de las LTR. La aplicación empleada fue “*CrustalW Multiple Alignment*” (<http://www.genome.jp/tools-bin/clustalw>), que aparece incluida en el programa de análisis de secuencias BioEdit v. 7.2.5 (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>).

Una vez abierta la aplicación de escritorio de BioEdit, en la opción “File”----“New Alignment” se abrió una pantalla nueva dentro del programa y después a la opción “File”----“Import”----“Sequence alignment file” y se añadieron las secuencias descargadas anteriormente de las seis especies en formato fasta. Por último, con la opción “Accessory Application”----“CrustalW Multiple Alignment” y tras varios minutos de espera se realizó hizo el alineamiento completo. Gracias a este alineamiento fue posible también sacar la secuencia consenso (secuencia ideal que representa los nucleótidos o aminoácidos que se encuentran con mayor frecuencia en cada posición de un fragmento de DNA o de una proteína) y las regiones con similitud, que aparecían marcadas con un *.

El programa LTR_STRUC (versión 1.1) (<https://omictools.com/ltr-retrotransposon-structure-program-tool>) identifica y analiza automáticamente los retrotransposones LTR en las bases de datos del genoma mediante la búsqueda de las características estructurales características de dichos elementos. Este programa se encarga de encontrar LTR-retrotransposones utilizando un algoritmo que abarca una serie de tareas que de otro modo tendrían que ser iniciadas individualmente por el usuario. Para cada LTR-retrotransposón encontrado, LTR_STRUC generaba automáticamente un análisis de una variedad de características estructurales de interés biológico. Este programa es una aplicación de consola compatible con el PC. Una vez creado el directorio de trabajo en el ordenador y la carpeta correspondiente el programa empezó a trabajar. LTR_STRUC identifica elementos en “hits” con la categoría de LTR (*Long Terminal Repeat*), retrotranscriptasa (RT) y PBS (*Primer Binding Site*) y colocaba los pares de bases (pb) de cada nuevo elemento en una lista a parte dentro de una carpeta de LTR_STRUC. Este programa es para identificar las diferentes secuencias dentro del retrovirus

(McCarthy & McDonald, 2003). En el archivo “flist.txt” e escribían los nombres de los ortólogos que aparecían en el NCBI . Una vez hecho esto, el programa proporcionaba los distintos archivos donde nos da toda la información sobre las LTR de cada especie estudiada. Los archivos se corresponden con los nombres: “_fsta.txt”; “_orfs.txt”; “_rppt.txt” y “_trns.txt”.

Lo siguiente fue hallar las LTR solitarias y el tiempo de inserción de las mismas. El archivo “_rppt.txt” del programa LTR_STRUC fue necesario para llevar a cabo esto ya que me proporcionó información como: la longitud del “contig”, la longitud de las LTR, los dinucleótidos de las LTR, las TSD y la secuencia de cada LTR.

Para la comprobación de si eran LTR solitarias o por el contrario, eran copias incompletas o truncadas, se compararon los resultados del “Human BLAT” (GRCh38/hg38), con la secuencia completa del trasposón y con las de las dos LTR (5’ Y 3’). Para ello, lo primero fue copiar la secuencia completa del trasposón (proporcionada por LTR_STRUC) y se pegó en el BLAT de la Universidad de California (UCSC, <https://genome.ucsc.edu/cgi-bin/hgBlat>). Luego se repitió el mismo proceso pero con la secuencia de la LTR 5’ y luego con la LTR 3’ y finalmente se hizo una comparación de los resultados de las tres coincidencias para determinar qué copias eran incompletas o truncadas y cuales correspondían a LTR solitarias.

Para el cálculo del tiempo de inserción de las LTR, es decir, para saber cuantos millones de años hace (Mya) que se insertaron en el genoma, se utilizó como reloj molecular la tasa de evolución entre *Homo sapiens* y *Pan troglodytes* (divergencia de 1% en posiciones nucleotídicas silenciosas durante los 6 millones de años transcurridos desde la separación de sus linajes; Siepel, 2009). Hallando el cociente de tasa de sustitución por lugar/Mya de años de divergencia se determina la tasa de evolución entre esas especies. El programa LTR_STRUC proporcionó las identidades en porcentajes de las especies *Homo sapiens*, *Gorilla gorilla*, *Pan paniscus* y *Pan troglodytes*. Con todo esto fue realmente sencillo determinar la divergencia entre estas especies así como su tiempo de inserción en millón de años (Mya).

3.4. Identificación de las regiones codificantes retrovirales

La herramienta CD-Search del NCBI, (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) es la interfaz de NCBI para buscar en la base de datos de dominio conservado con secuencias de búsqueda de proteínas o nucleótidos. Los resultados de CD-Search se presentan como una anotación de dominios de proteínas en la secuencia de consulta del usuario (ejemplo ilustrado), y se pueden visualizar como alineaciones de secuencias múltiples de dominio con consultas de usuario

integradas. Las asociaciones de alta confianza entre una secuencia de consulta y dominios conservados se muestran como aciertos específicos. La ayuda de búsqueda de CD proporciona detalles adicionales. (Marchler-Bauer et al., 2015).

Para esto, se copió y pegó la secuencia del elemento completo (ERVH48-1) y se seleccionó la opción “Submit” y aparecieron los dominios que están conservados.

Lo siguiente va a ser traducir la secuencia en todas las posibles pautas de lectura. Para ello, utilizamos sólo la región de la envuelta, ya que es la que nos interesa. Vamos al enlace <https://www.ncbi.nlm.nih.gov/orffinder/> para realizarlo. ORF Finder busca marcos de lectura abiertos (ORF) en la secuencia de ADN que ingresas. El programa devuelve el rango de cada ORF, junto con su traducción de proteínas. Se utiliza para buscar ADN recientemente secuenciado para posibles segmentos de codificación de proteínas.

Para esto, se pegó la secuencia completa del retroelemento en formato fasta (al igual que en los apartados anteriores) y se seleccionó la zona en la que se encuentra la supresina (From: 6593, To: 7075). Estas posiciones las indica “GenBank”, (<https://www.ncbi.nlm.nih.gov/nuccore>).

3.5. Reconstrucción de la secuencia activa

Llamamos secuencia activa a la secuencia retroviral original. La reconstrucción de esta secuencia activa u original fue llevada a cabo para ver los cambios evolutivos y las mutaciones que fueron pasadas a las siguientes generaciones tanto por transmisión vertical como horizontal (infección retroviral de otras células).

Para realizar esta reconstrucción, se llevó cabo una búsqueda desde el “National Centre for Biotechnology Information” (NCBI; <https://www.ncbi.nlm.nih.gov/>). Como palabra clave para la búsqueda se puso AB610407 y en el apartado de “orthologs” se descargaron todas las secuencias ortólogas en formato fasta: *Gorilla gorilla*, *Pan paniscus*, *Pan troglodytes*, *Pongo abelli*, *Nomascus leucogenys*, *Pliocolobus tephrosceles*, *Cercocebus atys*, *Mandrillus leucophaeus*, *Macaca nemestrina*, *Papio anubis*, *Aotus nancymaea*, *Manacus vitellinus*, *Callithrix jacchus* y *Cebus capucinus imitator*. Las últimas cuatro especies se descartaron ya que, *Manacus vitellinus* se trataba de un ave passeriforme y en el caso de *Aotus nancymaea*, *Callithrix jacchus* y *Cebus capucinus imitator* la secuencia de ERVH48-1 estaba muy deteriorada. En este caso no se tuvieron que añadir 1000 pares de bases a cada lado como en el apartado de “Estudio de las LTR”, ya que no queremos encontrar las *Long Terminal Repeat* (LTR). Para hacer la reconstrucción de la

secuencia retroviral original, se empleó la herramienta “CrustalW” (<http://www.genome.jp/tools-bin/clustalw>) encontrada dentro de la aplicación de escritorio del BioEdit.

En la aplicación de escritorio de BioEdit y pinchando sobre “Open” se cargó el archivo nuevo con los 11 ortólogos y se le dio a la opción “Accessory Application”----“CrustalW multiple alignment”.

Para reconstruir la secuencia original completa del retrovirus se hizo uso de un RetroTector online (<http://www.neuro.w.se/fysiologi/jbgs>) que es una versión sencilla que no requiere procedimientos de instalación específicos. Gracias a este programa fue una tarea fácil analizar los elementos retrovirales. Algunas de las funciones de RetroTector son: anotar y presentar las secuencias retrovirales conocidas, detectar cadenas províricas en secuencias genómicas desconocidas y traducir proteínas gracias a varios algoritmos. (Sperber et al., 2009).

Desde el navegador se fue a RetroTector online, (<http://retrotector.neuro.uu.se/pub/queue.php?show=submit>), se seleccionó el archivo de interés en formato fasta y se marcó la opción “all primates”. Una vez que el programa ejecutó el análisis, con la opción “RetroTector Viewer.jar” que se localiza en el mismo programa, se pudieron ver los resultados, ya que el ordenador generó nueva carpeta llamada “results” en la que me aparecieron automáticamente todos los resultados y visualizaciones realizadas por este programa en formato bloc de notas.

DnaSP, DNA Sequence Polymorphism (v 6.11.01), (<http://www.ub.edu/dnasp/>) es un paquete de software para el análisis de polimorfismos de ADN. DnaSP puede estimar varias medidas de variación de secuencia de ADN dentro y entre poblaciones en sitios no codificantes, sinónimos o no sinónimos, o en varios tipos de posiciones de codones), así como en desequilibrio de ligamiento, recombinación, flujo de genes y parámetros de conversión de genes. Los resultados de los análisis se muestran en forma tabular y gráfica. (Rozas et al., 2017).

Se descargó en el ordenador esta versión de DNAsp (v 6.11.01) para poder calcular el grado de polimorfismo de la secuencia proviral completa. Una vez abierto el programa en “File”----“Open Data File” se cargó el archivo. Este programa necesita trabajar con un formato especial, llamado formato “.phy”. Luego se fue a “Data”----“Define Domain Sets” y una vez aquí se seleccionaron solo las especies *Homo sapiens*, *Pan paniscus*, *Pan troglodytes*, *Gorilla gorilla*, *Nomascus leucogenys* y *Pongo abelii*. Por último, al darle a la opción “Analysis”----“DNA polymorphism”, el programa ya mostró los resultados en una tabla y en una gráfica.

Con el alineamiento de los ortólogos (las 11 especies) y con ayuda del BioEdit se pudo extraer la secuencia de la supresina. Para ello, se fue a “Edit”-----“Select positions”, se marcó “From: 6593” y “To: 7075” (que es la región en donde se localiza la supresina), tal y como indicaba “GenBank”. Una vez que la supresina aparece extraída de la secuencia proviral, se pudo realizar el nuevo alineamiento, al igual que antes, pero en este caso con la secuencia de la supresina, que después será necesario para la filogenia.

3.6. Reconstrucción filogenética

Para llevar a cabo el último apartado de este trabajo (la parte filogenética) se fue a <http://phylogeny.lirmm.fr/phylo.cgi/phylogeny.cgi>. Una vez en la página, se eligió la opción “a la Carte”, que ofrece más flexibilidad al ir seleccionando y configurando los pasos necesarios. (Dereeper et al., 2008). Se descartó la opción de alineamiento, ya que estaba hecho anteriormente del apartado de “Reconstrucción de la secuencia activa” y se seleccionó la opción “step by step”, con lo cual iba indicando paso a paso todo lo que el programa iba haciendo. Se añadió el archivo con el alineamiento en formato fasta. Se hizo la filogenia con el archivo de los 11 ortólogos del ERVH48-1 y después con el alineamiento de la secuencia extraída de la supresina también de las 11 especies de ortólogos.

Este análisis filogenético se divide en 3 partes: la primera es Curation (Gblocks), la segunda es Phylogeny (PhyML) y la tercera es Tree Rendering (TreeDyn) (Dereeper et al., 2008).

En la primera parte se utilizó el método Gblocks, que selecciona bloques de alineaciones para su uso en análisis filogenéticos. El grado de conservación de cada posición de la alineación múltiple se evalúa y clasifica como no conservado, conservado o altamente conservado. Se rechazan todos los tramos de posiciones contiguas no conservadas; en los bloques restantes, se examinan los flancos; se eliminan todas las posiciones con espacios; se eliminan los pequeños bloques que quedan después de la limpieza del espacio (Castresana, 2000).

La segunda parte es la propiamente filogenética “*Phylogeny: PhyML*” y este paso es imprescindible para la construcción de árboles (Dereeper et al., 2008).

La tercera y última parte es la visualización de árboles. El programa incluye herramientas para la manipulación y anotación de árboles (Chevenet et al., 2006).

3.7. Gestor bibliográfico

Se descargó la aplicación para escritorio de “Mendeley Desktop”, (<https://www.mendeley.com/>) como gestor bibliográfico para almacenar todas las referencias empleadas para el trabajo y también para facilitar y/o evitar dificultades a la hora de citar y/o insertar la bibliografía. Para las citas y la lista de referencias en Word, fue necesario descargar un plugin de Mendeley. El icono de Mendeley aparecía en todo momento en el margen superior derecho del navegador “Google Chrome”, que fue el navegador empleado en todo momento para buscar las referencias bibliográficas. Cada vez que se encontraba alguna referencia nueva para el trabajo, se pinchaba sobre el icono de Mendeley, y el programa añadía directamente la referencia a la librería del programa Mendeley.

4. RESULTADOS

4.1 Búsqueda de secuencias provirales de interés

Al realizar la búsqueda en “Nucleotide” en el NCBI con AB610407 se encontró una secuencia completa proviral que pertenece a ERVH48-1 en *Homo sapiens*. Al ir a “GenBank” dentro del NCBI, indica que este producto es un precursor de la supresina (HERV-Fb1) y que la inserción se localiza en el cromosoma 21q22.3 en *Homo sapiens*.

NUCLEOTIDE SEQUENCE

Homo sapiens HERV-Fb1 mRNA for suppressyn, complete cds

human

2,295 bp mRNA sequence

AB610407.1

[FASTA](#) [Gene](#)

Figura 4. Resultados búsqueda con clave de acceso AB610407.

También desde el NCBI también, utilizando la herramienta BLAST con la secuencia proviral completa, se consiguieron distintos alineamientos con otras secuencias del “GenBank”. Se seleccionaron los 5 primeros alineamientos, ya que presentaban una identidad del 100%.

Gracias al Genome Browser UCSC (University of California Santa Cruz) utilizando la herramienta BLAT del Genome, encontró secuencias similares

utilizando la secuencia de la inserción proviral completa. Con un 100% de identidad aparece la secuencia del elemento completo (8348pb), localizado en el cromosoma 21. Indica también las posiciones que ocupa la inserción en el cromosoma (42917251-42925598) y que es de cadena negativa.

Después, empleando como “query” la inserción de la supresina, con el navegador genómico “*Ensembl Genome Browser 90*”, aparecieron varios alineamientos de los cuales se seleccionó el primero, ya que presentaba un 100% de identidad y mostraba como localización genómica el cromosoma 21. También muestra en detalle la región del cromosoma en la que se localiza la inserción.

4.2. Identificación de las TSD (*Target Site Duplications*)

El programa LTR_STRUC nos indica que la secuencia debe empezar por TG y terminar por CA. Las target site duplications (TSD) son unas repeticiones idénticas de 5 pares de bases que se localizan justo antes y después de los dinucleótidos TG y CA. Hay presencia de TSD en el provirus (CTGAT), por lo tanto, indica que consiste en una inserción retroviral. Mediante estas indicaciones proporcionadas por LTR_STRUC y gracias a los resultados del navegador “*Ensembl Genome Browser 90*”, fue muy sencillo encontrar las target site duplications.

```
>chromosome:GRCh38:21:42916442:42926176:-1
42926176 CTCTGCCTCCCAGGTTTCAGGCCATTCTCCTGGCTCAGCCTCCCAAGTAGCTGGGACTACA 42926117
42926116 GGCGCCTGCAACCATGCCCGGCTAATTTTTTGATTTTTAGTAGAGACGGGGTTTCACCG 42926057
42926056 TGTTAGCCAGGACGGTCTCGATCTCCTGACCTCGTGATCTGCCCGCCTGGGCCTCCCAA 42925997
42925996 GTGCTGGGATTACAGGTGTGAGCCACCGTGCCCGGCTGAGACAGGTCTCTTGTTAATTTA 42925937
42925936 GAAAGTTTATTTTGCCAAGGTTGGGCAGGGACCTGTGACACCGCCTCAGGAAGTCTGAT 42925877
42925876 TGAAAGATTCTCCAGGGGCTGAAAGCTTGAGGGAATGAGTAACTCCTCCTCCTCAGGC 42925817
42925816 CCAGTCCCAACGCTCAAACCTGCACCAGCAGCAAGATAGCAGAGGCAGGAAAAGAGCTGG 42925757
42925756 CTGGAAGACACATACTCTCTGAAGATCAAGGGGGAGGCCGTCCGCGTACTACATAGCAGT 42925697

.....

.....

42916876 CACTTGGGGCTGATGCCATTTTAGGCCTCAGCCTGTCTGCATGCAGGCGCTCATTAAAAC 42916817
42916816 AGCATGTTGCTCCACACCGCCTCATGTTGTCTGTTGGCGCGCTCTCGGGGTTCCAACCAA 42916757
42916756 TGCAAGAGCCTTGCACTGATGACATGTCCCCAGGGTGGTTGGGCCACAGCTTGGTTTTAT 42916697
42916696 ACATTTTAGGGAGATATGAGACACCAATCAATATATGTAAGAAGCACCTTGGTTCTTCTG 42916637
42916636 GAAAGGTGGGACAACCTTGAAGCAAAGGCAGGAACACTCCAAGCGGGGAGGGAGCTTGCTG 42916577
42916576 GGCACAGAAAGGTGAGACACTGGAGGCTGCATTCTTTTGAGTTTCTGATCAGCCTTTCCA 42916517
42916516 AAGGAGGCAATGAGATATGCATCTATCTCAGTGAGCAGAGGAGTGACTTTGAATAGAATG 42916457
42916456 GGAGGCAGGTTTCCC 42916442
```

Figura 5. Representación en amarillo de las *Target Site Duplications* (CTGAT)

4.3. Estudio de las LTR (*Long Terminal Repeat*)

Además de cumplir una función estructural (Cordeiro & Taroco, 1980), las *Long Terminal Repeat* (LTR) son útiles para calcular el tiempo de inserción del retrovirus dentro del genoma (Coffin, Hughes, & Varmus, 1997).

LTR solitarias	21
LTR truncadas en 5´	20
LTR truncadas en 3´	3
Proviral completa	1

Tabla 3. Inserciones de cada categoría.

Tras el procedimiento explicado en “Material y Métodos” para poder distinguir entre LTR solitarias y copias truncadas o fragmentadas, los resultados obtenidos fueron: una inserción proviral completa (la que aparece en el cromosoma 21q22.3), LTR solitarias (21), LTR truncadas en 5´ (20) y LTR truncadas en 3´ (3). Por lo tanto, hay bastantes LTR solitarias.

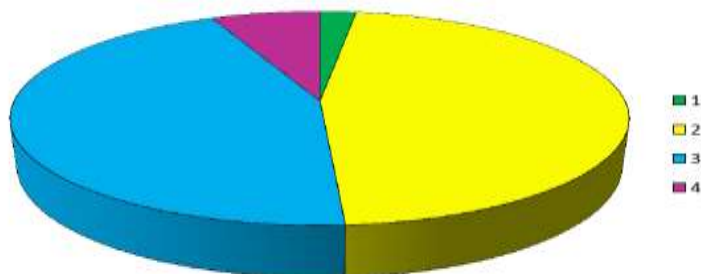


Gráfico 1. Representación LTR (proviral completa en verde, solitarias en amarillo, truncadas en 5´ en azul, truncadas en 3´ en violeta).

En la siguiente figura aparecen representadas con un círculo azul las LTR solitarias y con un cuadrado rojo aparece la localización de la inserción en el cromosoma 21.

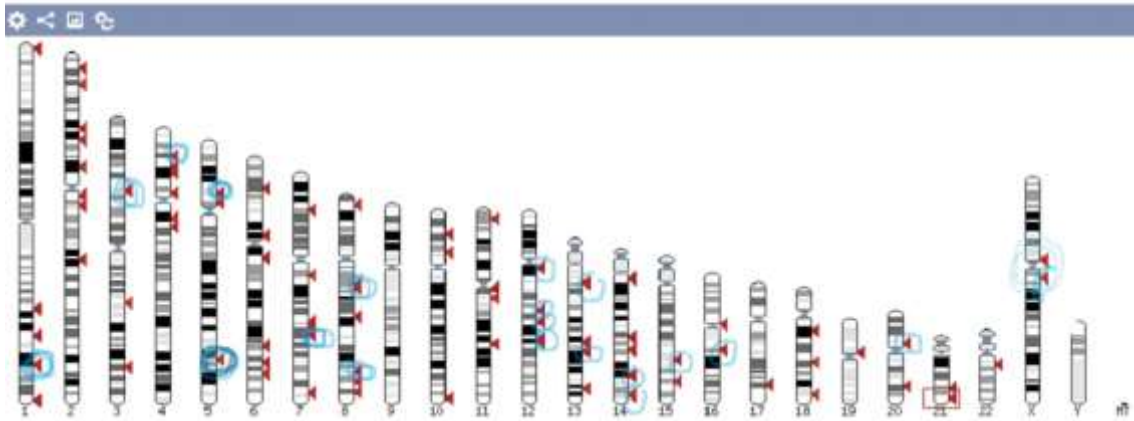


Figura 6. Esquema cromosomas con la localización de las LTR solitarias.

El tiempo de inserción de las LTR es muy antiguo (42,15 Mya), por lo tanto, la inserción de la supresina en el genoma va a aparecer hace mucho tiempo evolutivamente hablando. Y si, comparamos el tiempo de inserción por separado, vemos que la inserción también es muy antigua. En el caso de *Homo sapiens* es de 44,4 Mya, en *Gorilla gorilla* es de 39,6 Mya, en *Pan paniscus* es de 40,2 Mya y en *Pan troglodytes* es de 44,4 Mya.

Reloj molecular			
	Tiempo div (MA)	Sust por lugar	Tasa evo
Homo vs Pan	6	0,01	0,00166667

		Divergencia	Tiempo
Identidad Humano (%)	92,6	7,4	44,4
Identidad Gorila (%)	93,4	6,6	39,6
Identidad Bonobó (%)	93,3	6,7	40,2
Identidad Chimpancé (%)	92,6	7,4	44,4
PROMEDIO	92,975	7,025	42,15

Figura 7. Resultados identidad, tiempo de divergencia y tiempo de inserción en *Homo sapiens*, *Gorilla gorilla*, *Pan paniscus* y *Pan troglodytes*.

4.4. Identificación de las regiones codificantes retrovirales

La búsqueda de las regiones codificantes retrovirales a través de la herramienta CD-Search determinó qué mutaciones se han originado a lo largo del tiempo. Se observó también que la secuencia retroviral aparece completa.

Por otra parte, la herramienta ORF-Finder busca pautas de lectura abiertas.

En esta sección no obtuvimos gran información, pero en el siguiente apartado (“Reconstrucción de la secuencia activa”) aparecen profundizadas a detalle toda la información sobre las regiones retrovirales.

4.5. Reconstrucción de la secuencia activa

Gracias al RetroTector, y ayudándose del alineamiento, fue fácil distinguir las diferentes regiones del retrovirus: región *gag*, región *pol*, RNAsa, integrasa, transcriptasa inversa y región *env*.

	<i>Homo sapiens</i>	<i>Pan paniscus</i>	<i>Pan troglodytes</i>	<i>Gorilla gorilla</i>
<i>gag</i>	1942-2442	1659-2159	1646-2146	1665-2165
<i>pol</i>	4269-4887	3610-4218	3596-4205	3617-4238
<i>rna</i>	5255-5985	4586-5264	4574-5251	4606-5284
<i>int</i>	5544-6262	4875-5541	4862-5528	4895-5561
<i>rt</i>	6292-6852	5571-6131	5558-6118	5591-6151
<i>env</i>	6593-7351	6151-6624	6138-6611	6170-6650

Tabla 4. Representación de las distintas regiones del retrovirus.

Se incluye un anexo (Anexo I) con un esquema sobre las distintas regiones del retrovirus realizado por el programa RetroTector, en el que se aprecian las posiciones.

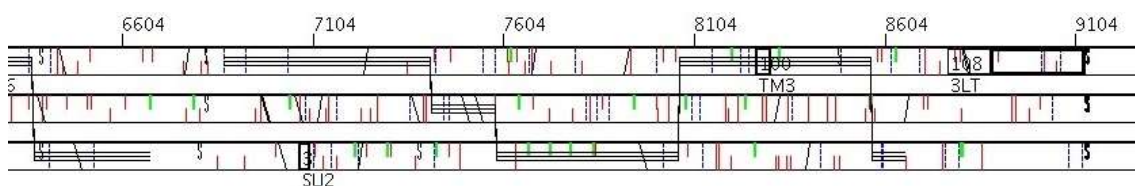


Figura 8. Esquema de la región de la envuelta realizado por RetroTector.

Por otra parte, se utilizó el programa DNAsp, que proporciona datos numéricos de las distintas partes de la secuencia completa y un gráfico de forma detallada donde muestra las zonas con polimorfismos elevados y también con polimorfismos bajos. La región de la supresina (región *env*) aparece muy conservada pero con un polimorfismo bajo (6593-7075). Por el contrario, la región de la proteasa (región *pro*) presenta un polimorfismo elevado. Gracias a los resultados del programa RetroTector, se sabe que la proteasa que aparece representada en la gráfica de los resultados es PR3 (proteína 3).

Nos va a decir si en la supresina hay más conservación que en el resto de la secuencia. Menos polimorfismo en la supresina.

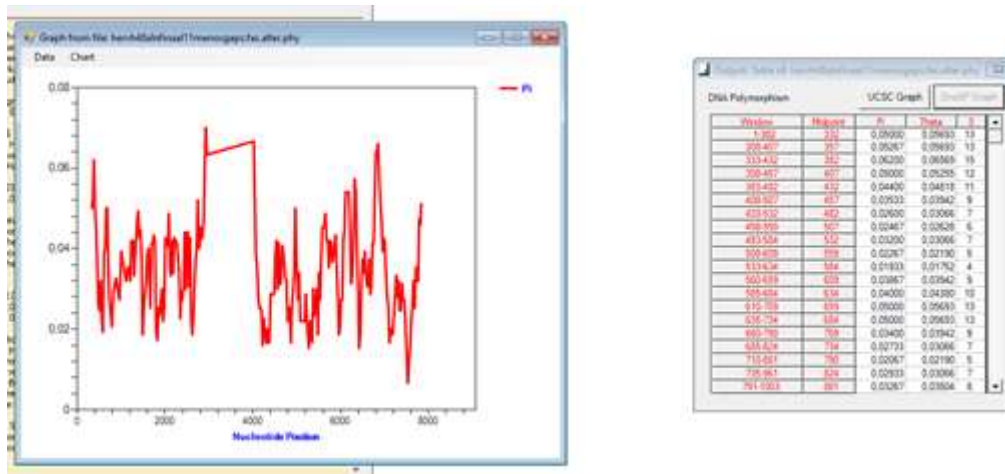


Figura 9. Representación en gráfica y tabla de los resultados de DNase

4.6. Reconstrucción filogenética

Los resultados de la reconstrucción filogenética muestran claramente un mismo ancestro en común, que luego se va a diversificar en dos ramas, catarrinos y platirrininos. La inserción va a ser muy antigua (42,15 Mya). Tal y como fueron dan los resultados anteriores, en platirrininos la inserción de la supresina aparecía muy deteriorada. Sin embargo, en catarrinos la inserción aparecía muy conservada a lo largo del tiempo. Esto era un claro ejemplo de lo que se apreciaba en los homínidos (*Homo sapiens*, *Pan paniscus*, *Pan troglodytes* y *Gorilla gorilla*), en los que la inserción estaba muy conservada a lo largo del tiempo, sin apreciar casi mutaciones.

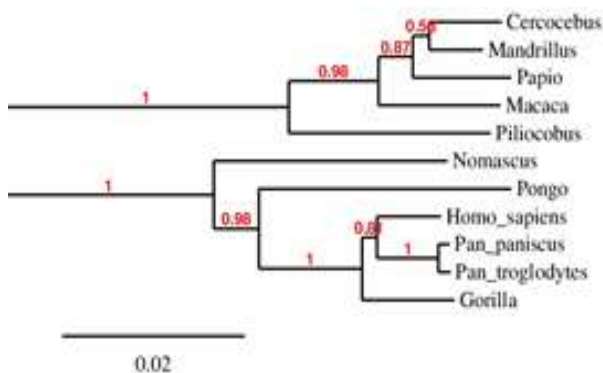


Figura 10. Reconstrucción filogenética primates con “PhyML”.

La estrella roja indica el momento en el que se produjo la inserción en el genoma, que coincide justamente con la separación de catarrinos y platirrininos. Eso explica porque en los catarrinos la inserción se encuentra mucho más conservada que en platirrininos.

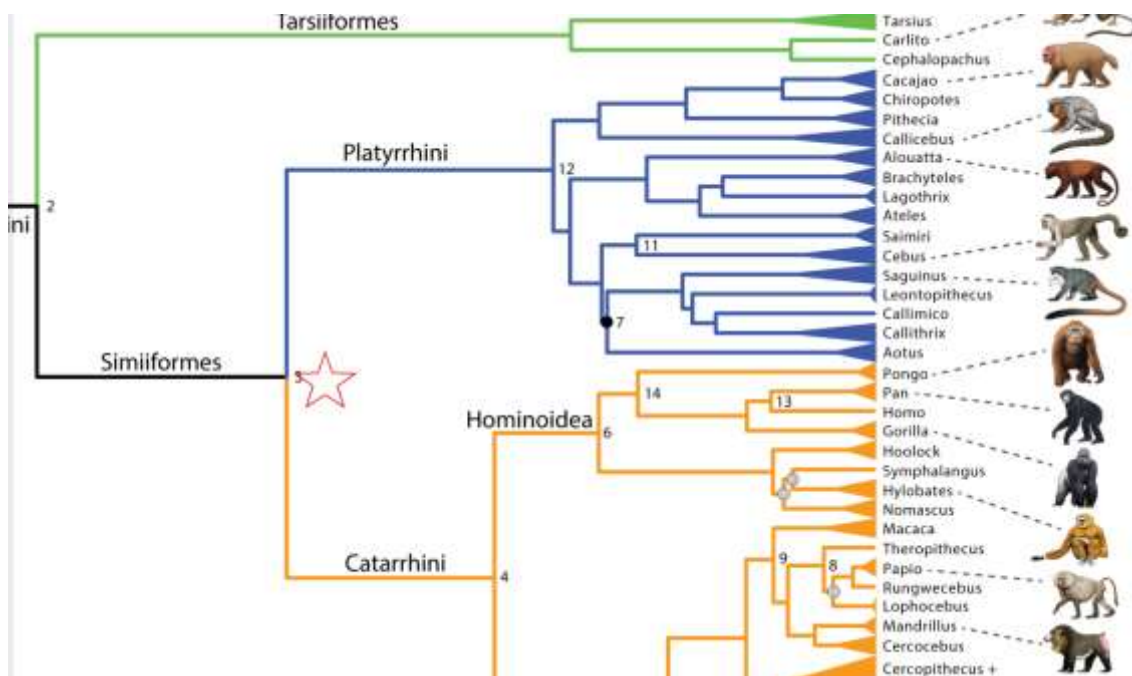


Figura11. Filogenia de primates.

5. DISCUSIÓN

5.1. Alto grado de conservación del grupo HERV-Fb1

El grupo HERV-Fb1 presenta una peculiar inserción en su genoma que lo diferencia de otros grupos de retrovirus, y es que presenta una inserción de un gen llamado supresina en el cromosoma 21q22.3. Esta inserción aparece en catarrinos, y sobre todo aparece altamente conservado en cuatro especies de homínidos: *Homo sapiens*, *Pan paniscus*, *Pan troglodytes* y *Gorilla gorilla*.

Se trata de una secuencia proviral completa con idénticas TSD o *Target Site Duplications*, por lo que va a derivar de una secuencia retroviral original o activa. Se observaron LTR solitarias, lo que indica que el retrovirus original no ha sufrido mutaciones a lo largo del tiempo (Coffin, Hughes, & Varmus, 1997), y que por lo tanto, esta inserción presenta un alto grado de conservación.

El tiempo promedio de la inserción en estas especie fue de 42,15 Mya (millones de años), por lo tanto, todo esto nos indica que se trata de una inserción muy antigua filogenéticamente hablando y a su vez muy conservada a lo largo del tiempo.

5.2. Funciones supresina

La supresina es un gen derivado del HERV-Fb1 localizado en la envuelta retroviral, para ser más exactos, en la región transmembrana. Este gen es parte de un provirus de retrovirus endógeno que tiene expresión específica en la placenta. La función de la supresina es la de inhibir la fusión celular o sincitización al competir con la proteína sincitina por la unión a un receptor celular (Wolf et al., 1979). La proteína sincitina permite la fusión celular del citotrofoblasto y es producto del gen *env* de un miembro de la familia HERV-W (Mi et al., 2000). HERV-W es una familia de HERVs propia de los primates catarrinos. (Berencsi, 2012).

5.3. Funciones de los HERVs en el proceso de la placentación y patologías relacionadas

En la placenta se detecta una gran actividad de la transcriptasa inversa (Larsson, Andersson, & Nilsson, 1994). Los retrovirus son inmunosupresores y fusogénicos gracias a las características de las proteínas codificadas por el gen *env*, lo que los convierte en candidatos idóneos, ya que sus productos mantienen las mismas propiedades (Denner, 1998). Además de supresores y fusogénicos, deben tener expresión específica en la placenta para poder ser funcionales (Cornelis et al., 2015). La placenta fue importante en la evolución de los retrovirus endógenos, que han sido fundamentales en la evolución de animales placentarios y el desarrollo de viviparidad, gracias a sus propiedades de fusogenicidad e inmunosupresión (Denner, 2016). Una LTR retroviral proporciona el enhancer específico de placenta al gen de la leptina (Bi et al., 1997). Este enhancer promueve una alta tasa de expresión de este gen en la placenta, sobre todo en el sincitiotrofoblasto y contribuye al incremento de las concentraciones circulantes leptina durante el embarazo (Lin, Xu, & Rote, 1999). La leptina en la placenta presenta también un papel angiogénico e inmunomodulador mediante mecanismos autocrinos y paracrinos y puede afectar al crecimiento fetal y al desarrollo mediante su unión a los receptores de leptina específicos presentes en los órganos fetales (Henson & Castracane, 2000). Por lo tanto, la desregulación de la leptina placentaria y/o de sus receptores, se ha asociado con problemas en el embarazo, como preeclampsia y diabetes materna, retraso en el crecimiento fetal, insuficiencia materna, etc. (Sentís, 2016). Así mismo, los HERVs actuarían en la placenta mediante una infección local sobre las células circundantes, de forma que provocarían la inmunosupresión de los macrófagos materno próximos que son los encargados del rechazo fetal, mientras que el resto del sistema inmune materno seguiría defendiendo a la madre y al feto de los agentes patógenos (Mi et al., 2000).

6. CONCLUSIONES

El grupo HERV-Fb1 de retrovirus endógenos humanos presenta una característica que los separa de otros grupos de retrovirus. Esta familia se caracteriza por ser el precursor del gen de la supresina, que tiene alta expresión en la placenta de los primates. Después de toda la parte experimental, se llegó a la conclusión final de que la inserción de la supresina en el cromosoma 21q22.3 es muy antigua (aproximadamente 42 Mya) y de que está muy conservada en los homínidos *Homo sapiens*, *Pan paniscus*, *Pan troglodytes* y *Gorilla gorilla*.

O grupo de HERV-Fb1 de retrovirus endógenos humanos presenta una característica que os separa de outros grupos de retrovirus. Esta familia caracterízase por ser precursora do xen da supresina, que ten alta expresión na placenta dos primates. Despois de toda a parte experimental, chegouse á conclusión final de que a inserción da supresina no cromosoma 21q22.3 é moi antiga (aproximadamente 42 Mya) e de que está moi conservada nos homínidos *Homo sapiens*, *Pan paniscus*, *Pan troglodytes* e *Gorilla gorilla*.

The HERV-Fb1 group of human endogenous retroviruses has a characteristic that separates them from other groups of retroviruses. This family is characterized for being the precursor of the gene of the supressyn, that has high expression in the placenta of the primates. After the whole experimental part, the final conclusion was reached that the insertion of supressyn on chromosome 21q22.3 is very old (approximately 42 Mya) and that it is highly conserved in hominids *Homo sapiens*, *Pan paniscus*, *Pan troglodytes* and *Gorilla gorilla*.

7. BIBLIOGRAFÍA

- Berencsi, G. (2012). *Maternal fetal transmission of human viruses and their influence on tumorigenesis. Maternal Fetal Transmission of Human Viruses and their Influence on Tumorigenesis* (Vol. 9789400742). Springer Netherlands. <https://doi.org/10.1007/978-94-007-4216-1>
- Bi, S., Gavrilova, O., Gong, D. W., Mason, M. M., & Reitman, M. (1997). Identification of a placental enhancer for the human leptin gene. *The Journal of Biological Chemistry*, 272(48), 30583–30588. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9374555>
- Castresana, J. (2000). Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution*,

17(4), 540–552. <https://doi.org/10.1093/oxfordjournals.molbev.a026334>

- Chen, C.-P., Wang, K.-G., Chen, C.-Y., Yu, C., Chuang, H.-C., & Chen, H. (2006). Altered placental syncytin and its receptor ASCT2 expression in placental development and pre-eclampsia. *BJOG: An International Journal of Obstetrics and Gynaecology*, 113(2), 152–158. <https://doi.org/10.1111/j.1471-0528.2005.00843.x>
- Chevenet, F., Brun, C., Bañuls, A.-L., Jacq, B., & Christen, R. (2006). TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics*, 7(1), 439. <https://doi.org/10.1186/1471-2105-7-439>
- Coffin, J. M., Hughes, S. H., & Varmus, H. E. (1997). *Retroviruses*. *Retroviruses*. Cold Spring Harbor Laboratory Press. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/21433340>
- Cordeiro, N., & Taroco, R. (1980). Retrovirus y VIH. *Science*, 449–476.
- Cornelis, G., Vernochet, C., Carradec, Q., Souquere, S., Mulot, B., Catzeflis, F., ... Heidmann, T. (2015). Retroviral envelope gene captures and syncytin exaptation for placentation in marsupials. *Proceedings of the National Academy of Sciences of the United States of America*, 112(5), E487-96. <https://doi.org/10.1073/pnas.1417000112>
- Denner, J. (1998). Immunosuppression by retroviruses: implications for xenotransplantation. *Annals of the New York Academy of Sciences*, 862, 75–86. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9928208>
- Denner, J. (2016). Expression and function of endogenous retroviruses in the placenta. *APMIS*, 124(1–2), 31–43. <https://doi.org/10.1111/apm.12474>
- Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., ... Gascuel, O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*, 36(Web Server), W465–W469. <https://doi.org/10.1093/nar/gkn180>
- 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. <https://doi.org/10.1016/j.virol.2005.01.045>
- Henson, M. C., & Castracane, V. D. (2000). Leptin in Pregnancy. *Biology of Reproduction*, 63(5), 1219–1228. <https://doi.org/10.1095/biolreprod63.5.1219>
- Larsson, E., Andersson, A. C., & Nilsson, B. O. (1994). Expression of an endogenous retrovirus (ERV3 HERV-R) in human reproductive and embryonic tissues--evidence for a function for envelope gene products. *Uppsala Journal of Medical Sciences*, 99(2), 113–120. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7716822>

- Lin, L., Xu, B., & Rote, N. S. (1999). Expression of Endogenous Retrovirus ERV-3 Induces Differentiation in BeWo, a Choriocarcinoma Model of Human Placental Trophoblast. *Placenta*, *20*(1), 109–118. <https://doi.org/10.1053/PLAC.1998.0337>
- Linheiro, R. S., & Bergman, C. M. (2012). Whole Genome Resequencing Reveals Natural Target Site Preferences of Transposable Elements in *Drosophila melanogaster*. *PLoS ONE*, *7*(2), e30008. <https://doi.org/10.1371/journal.pone.0030008>
- Mi, S., Lee, X., Li, X., Veldman, G. M., Finnerty, H., Racie, L., ... McCoy, J. M. (2000). Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature*, *403*(6771), 785–789. <https://doi.org/10.1038/35001608>
- Nelson, P. N., Carnegie, P. R., Martin, J., Davari Ejtehad, H., Hooley, P., Roden, D., ... Murray, P. G. (2003). Demystified. Human endogenous retroviruses. *Molecular Pathology: MP*, *56*(1), 11–18. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12560456>
- Sentís, C. (2016). Retrovirus endógenos humanos: Significado biológico e implicaciones evolutivas. *Arbor*, *677*, 135–166.
- 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. <https://doi.org/10.1038/srep01462>
- Sverdlov, E. D. (2005). Retroviruses and Primate Retroviruses and Primate Genome Evolution. *Eurekah/Landes Bioscience*, 161–171.
- Vargiu, L., Tomé, P. R., Sperber, G. O., Cadeddu, M., Grandi, N., Blikstad, V., ... Blomberg, J. (2016). Classification and characterization of human endogenous retroviruses ; mosaic forms are common. *Retrovirology*, 1–29. <https://doi.org/10.1186/s12977-015-0232-y>
- Wolf, G., Kiorpes, T. C., Masushige, S., Schreiber, J. B., Smith, M. J., & Anderson, R. S. (1979). Recent evidence for the participation of vitamin A in glycoprotein synthesis. *Federation Proceedings*, *38*(11), 2540–2543. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/90625>

Páginas web:

<https://wildlifesnps.files.wordpress.com/2016/02/journal-pone-0049521-g001.png>

ANEXO I

Detalle de gráfica resultados RetroTector.

