UNIVERSIDADE DA CORUÑA

FACULTADE DE CIENCIAS

ANÁLISIS DE LA VARIACIÓN GENÓMICA Y LA EVOLUCIÓN DEL GRUPO FC1 DE RETROVIRUS ENDÓGENOS HUMANOS

ANÁLISE DA VARIACIÓN XENÓMICA E A EVOLUCIÓN DO GRUPO FC1 DE RETROVIRUS ENDÓXENOS HUMANOS
ANALYSIS OF GENOMIC VARIATION OF THE ERV-FC1 GROUP OF HUMAN ENDOGENOUS RRETROVIRUSES

GRADO EN BIOLOGÍA

Estudiante: André Flores Bello

Tutor: Horacio Naveira Fachal

22 de julio de 2014

Resumen

El grupo de retrovirus endógenos humanos ERV-Fc1 parece poseer ciertas peculiaridades frente a la mayoría de retrovirus endógenos. Destaca su alto grado de conservación y su origen a partir de una única y reciente integración en un ancestro común a gorila, chimpancé, bonobo y humano. En este estudio se llevó a cabo un análisis del grupo mediante la búsqueda de posibles inserciones en los genomas, así como la determinación de la integridad de sus regiones estructurales más importantes. Por otro lado, se trató de reconstruir la posible secuencia activa original y el tiempo evolutivo en el que la inserción podría haber tenido lugar. Además, se intentaron detectar posibles indicios de presiones selectivas purificadoras sobre el grupo, lo que podría demostrar que desempeñan alguna importante función en el genoma. Por último, se reconstruyó la filogenia de las secuencias completas y de las regiones individualmente, con la finalidad de determinar si el grupo sigue un patrón evolutivo que coincide con el conocido y detectar posibles casos de recombinación, transferencia horizontal o evolución en mosaico de sus regiones.

Índice

1. Introducción

Gran parte de los genomas se encuentran constituidos por un elevado porcentaje de elementos transponibles, frente a un bajo porcentaje de DNA propiamente codificante. En los genomas eucariotas (de animales principalmente), parte de esos elementos transponibles la constituyen los retrovirus endógenos (ERVs). Así, en el genoma humano alrededor del 45 % está conformado por dichos elementos, entre los que aproximadamente un 8-10% corresponde a retrovirus endógenos humanos (HERVs), mientras que únicamente el 3% de genoma está ocupado por DNA codificante [1–4].

Los retrovirus exógenos tienen capacidad infectiva y se transmiten horizontalmente entre los individuos de las poblaciones, afectando generalmente a las células somáticas. Sin embargo, hay ocasiones en las que estas infecciones son capaces de traspasar la línea germinal, originando lo que se conoce como retrovirus endógeno [3,4]. Éstos, consisten en secuencias de DNA presentes en el genoma, procedentes de integraciones retrovirales que tuvieron lugar en células de la línea germinal, de manera que quedaron fijadas en el genoma del organismo infectado, comenzándose a transmitir verticalmente a la descendencia [3,5]. Van a presentar una estructura genómica homóloga a la secuencia del retrovirus exógeno original, sin embargo la inmensa mayoría van a ser defectivos por acumulación de múltiples mutaciones que impiden su completa expresión, quedando únicamente sujetos a procesos de transposición en el genoma y transmisión vertical [1–6].

Los retrovirus van a contener originalmente RNA en sus viriones, de manera que para su integración en un genoma en forma de provirus, va a ser necesaria la retrotranscripción del RNA a un DNA linear de cadena doble intermediario, capaz de acoplarse al genoma hospedador. Ésta es la característica básica de los retrovirus, y es posible gracias a la enzima retrotranscriptasa o transcriptasa inversa, que cataliza la reacción [3,5–7].

La organización genómica general de un provirus (Fig. 1) va a consistir en una LTR (*Long Terminal Repeat*) en cada extremo, tres grandes dominios codificantes (*gag*, *pol* y *env*) y un dominio codificante menor (*pr*) situado entre *gag* y *pol*. Por orden de 5' a 3' podemos encontrar, en primer lugar, *gag,* que va a codificar las proteínas estructurales internas de los viriones, como son las de la matriz, cápside y nucleocápside; *pr*, que codifica la proteasa; *pol* que va a codificar las enzimas esenciales para la actividad retroviral, es decir, retrotranscriptasa (*rt*), ribonucleasa H *(RNAsaH, rh*) e integrasa (*in*); la actividad *RNAsaH* está codificada en un dominio discreto que se encuentra ligado covalentemente al dominio polimerasa de la retrotranscriptasa [7]; y finalmente *env*, que va a codificar las proteínas de la envuelta, tanto las de superficie como las transmembrana.

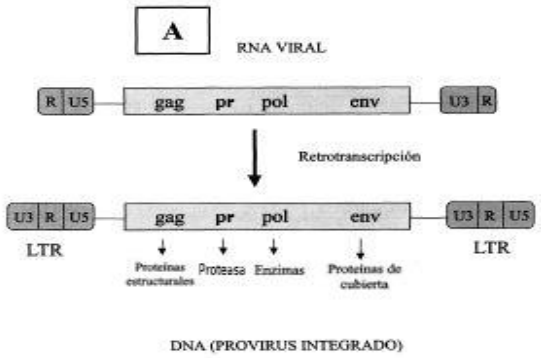


Fig. 1. Organización genómica general de un retrovirus exógeno libre (RNA) e integrado como un provirus (DNA y con LTRs).

El mecanismo de síntesis proteica va a ser similar en todos los grupos de retrovirus, presentando únicamente pequeñas variaciones entre algunos. *Gag*, *pr* y *pol* se encuentran naturalmente incluidos en un gran ORF (*Open Reading Frame* o marco abierto de lectura) iniciado por un codón de inicio ATG, originándose así un gran mRNA que codifica para una proteína de fusión precursora *gag-pr-pol* [7], que tiene que ser posteriormente procesada para generar cada una de las proteínas finales. Entre *gag* y *pr*-pol, a pesar de formar parte del mismo ORF, va a existir un codón *stop* natural que forma parte de un mecanismo destinado a incrementar la expresión de *gag* con respecto a *pr* y *pol*, ya que es necesaria a altos niveles para la replicación del retrovirus. Esto hace posible la síntesis de *gag* independientemente al delimitar dicho codón stop un ORF único para *gag*. Pero además se sigue sintetizando la proteína de fusión precursora *gag-pr-pol*, lo que es posible porque se produce un salto programado del codón *stop*, que puede tener lugar por medio de dos mecanismos [7]: en algunos grupos de retrovirus se lleva a cabo una mala lectura intencionada, de manera que se detecta como codificante y se introduce un aminoácido para continuar la traducción de *pr* y *pol*; sin embargo, el mecanismo más común es un cambio programado de la pauta de lectura (-1) por parte del ribosoma, llevado a cabo ocasionalmente durante la traducción de *gag*, de modo que se produce un salto al marco de lectura en el que se encuentran codificadas *pro-pol*. Por otro lado, las proteínas *env* se encuentran codificadas en un ORF independiente a *gag-pr-pol*, iniciado también por un codón ATG [7].

En cuanto a las LTR, consisten en repeticiones terminales que flanquean las regiones codificantes de la secuencia y no se encuentran presentes en el RNA viral del retrovirus exógeno, originándose como tales durante el proceso de retrotranscripción (Fig.1), antes de su inserción en un genoma. Estas repeticiones terminales van a contener la mayor parte de los elementos reguladores de la transcripción del provirus, incluyendo promotores y *enhancers* (potenciadores o activadores de la expresión génica), y son necesarias para la integración en el DNA hospedador junto con la enzima integrasa [3,7,8].

El proceso de integración del retrovirus, parece que, a pesar de no tener lugar en una secuencia específica, no sucede aleatoriamente en el genoma como se creía [9], sino que existen algunos lugares susceptibles como el DNA nucleosomal [10], genes activos o regiones reguladoras [9,11] en función del grupo retroviral. Durante la inserción, una corta secuencia (4-6 pb) del *target site* (sitio diana de inserción) se ve duplicada como resultado de la reparación de los *"gaps"* originados a ambos extremos de la secuencia retroviral [7] (Fig. 2).

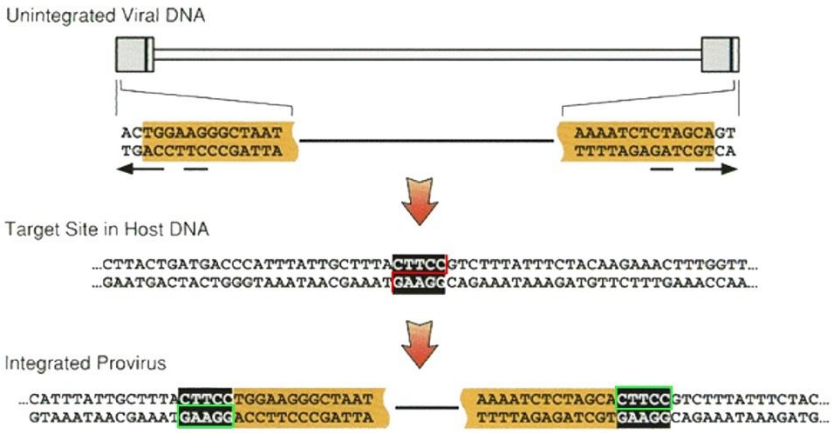


Fig. 2. Integración del retrovirus exógeno en forma de provirus. En rojo, corte escalonado del target site. En verde, target site duplications (TSDs).

La unión al DNA huésped se lleva a cabo por el complejo integrasa-DNA viral, que es seguida de una reacción catalizada por la integrasa, en la que los extremos 3'-OH libres del DNA viral, originados por la escisión de 2 nucleótidos por parte de la integrasa, son utilizados para atacar los enlaces fosfodiéster de las cadenas opuestas en la doble hélice de DNA diana. Esto tiene lugar en posiciones escalonadas a 4-6 bases en dirección 5' (Fig. 3). La energía liberada en la ruptura de los enlaces fosfodiéster del DNA huésped es aprovechado para la formación de nuevos enlaces con los extremos 3' del DNA viral, de manera que sus extremos 5' quedan libres, separados del DNA huésped por esas 4-6 bases antes mencionadas. A continuación, tiene lugar la síntesis de DNA, guiada por las proteínas virales o dirigida por la retrotranscriptasa, desde los extremos 3' libres del DNA huésped hacía los extremos 5' libres del DNA viral, llenando los *gaps* que se habían originado y dando lugar en ambos extremos de la secuencia a la duplicación del *target site*, en el que tuvo lugar el corte escalonado y la inserción del retrovirus [7]. Cada grupo retroviral va a originar unas *target site duplications* (TSD) específicas, que se van a caracterizar por su longitud, mientras que su secuencia no va a ser la misma para todo el grupo, pues la inserción, tal y como ya se dijo, no tiene lugar en una secuencia específica.

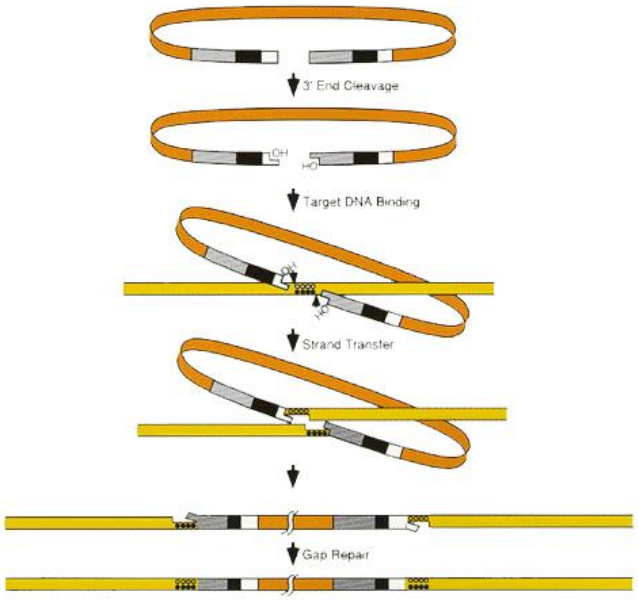


Fig. 3. Esquematización del proceso de integración del DNA retroviral.

Cuando un retrovirus consigue insertarse en la línea germinal, dando lugar a un retrovirus endógeno, queda a partir de ese momento integrado de forma irreversible en el genoma y expuesto a los diferentes mecanismos de variabilidad genética que tienen lugar en el genoma hospedador [3]. De este modo, su secuencia va a ir divergiendo de la del provirus original a lo largo de la historia evolutiva por acumulación de diversas mutaciones y alteraciones, lo que las lleva a perder su capacidad individual de replicación. Así pues, algunas secuencias van a mantener su estructura proviral completa, aunque defectiva; mientras que otras van a experimentar diferentes eventos de retrotransposición en el genoma gracias a la presencia de las LTR y la retrotranscriptasa, generándose multitud de copias, parciales o completas; va a ser común la recombinación homóloga entre las LTR del inserto, dando lugar a la formación de LTRs solitarias en lugar del elemento proviral [12].

Los retrovirus endógenos, a pesar de encontrase en la gran mayoría de los casos como secuencias defectivas en cuanto a su actividad viral, van a presentar diversos efectos en los

organismos hospedadores. En primer lugar, sus eventos de retrotransposición y recombinación otorgan una mayor plasticidad al genoma, permitiendo cambios más rápidos que los que tendrían lugar mediante mutaciones normales, pudiendo derivar en el truncamiento de algún gen de importante función o su traslado a otra región del genoma. Además, las LTR, como reguladores transcripcionales que son, pueden influenciar sobre la transcripción de genes cercanos, alterando su expresión como consecuencia de una activación o un silenciamiento, lo que puede derivar en serios problemas dependiendo de la función de dicho gen [1–3,5]. Numerosos grupos de retrovirus endógenos han sido relacionados con diversas patologías, fundamentalmente procesos cancerosos y autoinmunes, como la esclerosis múltiple en el humano, en la que además de otros grupos de ERVs, se ha detectado gran implicación del grupo ERV-Fc1 [13–18], que es el de interés en este estudio. Sin embargo, resulta difícil concluir si la expresión de los retrovirus endógenos relacionada con una patología, se trata de la causa o la consecuencia de la de la misma, no pudiendo determinar si el ERV contribuye a la enfermedad o es la propia enfermedad la que desemboca en la activación del ERV [3,15]. Por otro lado, se han descubierto también funciones beneficiosas en los genomas e implicaciones en procesos biológicos esenciales. Van a existir multitud de genes que tienen un origen viral (Tabla 1) [19]. Es posible que un nuevo ERV, contenga uno o más genes que resulten beneficiosos para el hospedador, de manera que tras su inserción tiene lugar un proceso de reclutamiento de genes con nuevas funciones positivas, actuando sobre ellos una presión selectiva que los mantiene funcionales a lo largo de la evolución a la vez que el resto del retrovirus endógeno degenera hasta hacerse defectivo [620]. Son abundantes los estudios que hacen referencia al papel que desempeñan ciertos retrovirus endógenos en procesos como la placentación y desarrollo [3,21,22], resistencia antiviral [5,23], incremento de la diversidad genética, estabilización de los extremos cromosómicos junto con la telomerasa [23]…

*Tabla 1. Ejemplos de genes celulares de origen viral. *Hace referencia a la sección del gen que codifica la RNAsaH y la integrasa. CarERV: Carnivore endogenous retrovirus; EBLN1: Endogenous Bornavirus-like nucleoprotein 1; ERVV1: Endogenous retrovirus group V, number 1; Fv1: Friend virus susceptibility 1; HERV: human endogenous retrovirus; MuERV: murine endogenous retrovirus.*

Gene	Virus progenitor (viral gene or domain)	Species distribution (age)	Function and activities
Syncytin 1 (also known as *ERVW1*)	HERV-W (*env*)	Catarrhine primates: humans, apes, Old World monkeys (25–40 million years)	Placenta-specific expression, fusogenic activities
Syncytin 2 (also known as *ERVFRD1*)	HERV-FRD (*env*)	Anthropoid primates: catarrhines and New World monkeys (40–65 million years)	Placenta-specific expression, fusogenic and immunosuppressive activities
Syncytin A (*Syna*)	HERV-F or HERV-H (*env*)	Murid rodents (20–30 million years)	Placenta formation (layer I of syncytiotrophoblast); placenta-specific expression, fusogenic activities *ex vivo*
Syncytin B (*Synb*)	HERV-F or HERV-H (*env*)	Murid rodents (20–30 million years)	Placenta formation (layer II of syncytiotrophoblast); placenta-specific expression, fusogenic and immunosuppressive activities
Syncytin-Ory1	Type D retroviruses (*env*)	Leporids: rabbits and hares (12–30 million years)	Placenta-specific expression, fusogenic activities
Syncytin-Car1	CarERV3 (class I)	Carnivores (65–80 million years)	Placenta-specific expression, fusogenic activities
ERVV1 and *ERVV2*	HERV-V (*env*)	Anthropoid primates (40–65 million years)	Placenta-specific expression, unknown function
Fv1	MuERV-L (*gag*) (class III)	*Mus* subgenera: mice (5–10 million years)	Confer resistance to murine leukaemia virus (MLV), binds MLV capsid
CGIN1 (also known as *NYNRIN*)	Retrovirus (*pol* (RNaseH, integrase)*)	Therians: placental and marsupial mammals (125–180 million years)	Unknown
EBLN1, *EBLN2*, *EBLN3* and *EBLN4*	Bornavirus (nucleoprotein)	Anthropoid primates (40–65 million years)	Unknown, but EBLN2 appears to interact with several cellular proteins
Iris	*Kanga* errantivirus (F-type *env*)	*Drosophila* melanogaster and obscura subgroups (25–35 million years)	Third instar larva- and adult-specific expression, localized to mitochondria

La especie humana es en la que mejor estudiados están los retrovirus endógenos, diferenciándose hasta 31 familias de HERVs [24]. A pesar de que algunas de ellas pueden ser encontradas en especies muy alejadas filogenéticamente al haberse producido una inserción

en algún ancestro común, la mayor parte son características de primates, y algunas incluso específicas de humanos. Así pues, queda patente que las diferencias entre especies próximas, no se debe únicamente a la existencia de muchos genes estructurales distintos, sino a la expresión diferencial entre los genes compartidos, ya que en cada una de ellas éstos van a presentar un grado de conservación y funcionalidad determinado [3]. Gracias a ello, es posible llevar a cabo reconstrucciones filogenéticas de diferentes especies que compartan un determinado tipo de ERV basándose en la comparación de sus secuencias, ya que en principio deberían reflejar los cambios producidos a los largo de la historia evolutiva en el genoma de cada especie.

Este trabajo se va a centrar en el estudio del grupo de retrovirus endógenos ERV-Fc1, perteneciente a la familia HERV-F, que se incluye en el Grupo I de la Clase I en la clasificación de los HERVs realizada por Nelson *et al.* en 2003 [5], en la que se agrupan por homología de secuencia con los retrovirus animales (Tabla 2). Esta familia se encuentra fuertemente relacionada con la familia HERV-H, con la que aparece formando un cluster al llevar a cabo la reconstrucción filogenética de las distintas familias de retrovirus endógenos [4,24] (Anexo I).

Tabla 2. Clasificación de las diferentes familias de HERVs

HERV family	Representative accession number
Class I HERVs (type C related HERVs)	
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-ERI	
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
Class II HERVs (type A, B, and D related HERVs)	
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-NMWV2	AF015995
Group 6, HERV-K (HML-6)	
HERV K (HML-6p)	U86698
Group 7, HERV-K (HML-7)	
HERV-K-NMWV7	AF016000
Group 8, HERV-K (HML-8)	
HERV-K-NMWV3	AF015996
Group 9, HERV-K (HML-9)	
HERV-K-NMWV9	AF016001
Group 10, HERV-K (HML-10)	
HERV-KC4	U07856
Class III Foamy virus related HERVs	
HERV-L	X89211

2. Material y métodos

2.1. Búsqueda de secuencias provirales de interés

En primer lugar, se llevó a cabo una búsqueda directa desde la base de datos "*Nucleotide"* del *National Center for Biotechnology Information* (NCBI, http://www.ncbi.nlm.nih.gov/),

empleando como palabra clave el nombre del retrovirus de interés (ERV-Fc1). De los cuatro resultados encontrados, dos correspondieron a *Pan troglodytes* (Chimpancé) y dos a *Gorilla gorilla* (Gorila), pero únicamente uno de cada especie aparecía como una secuencia proviral completa (*ERV-Fc1 Chimpanzee*: AJ507127.1; *ERV-Fc1 Gorilla*: AJ507128.1).

A continuación, tratando de encontrar posibles inserciones de ERV-Fc1 en otros genomas, se llevó a cabo una búsqueda masiva empleando la herramienta "*Basic Local Alignment Search Tool"* (BLAST, http://blast.ncbi.nlm.nih.gov/Blast.cgi) del NCBI, que encuentra las similitudes existentes entre nuestra secuencia de interés (*query*) y numerosas secuencias recogidas en diferentes bases de datos del *GenBank* (http://www.ncbi.nlm.nih.gov/genbank), que permite elegir, basándose en la realización de un alineamiento. Se origina así, una salida con las secuencias encontradas que presenten la mayor similitud con nuestro *query,* además de un alineamiento significativo. En esta salida se indica: su puntuación, el p-valor de la relación, la identidad entre las secuencias, el porcentaje que cubre del *query* y su número de acceso. El tipo de BLAST realizado fue un *nblast* (*nucleotide blast*), que permite buscar en bases de datos con secuencias nucleotídicas empleando como *query* otra secuencia nucleotídica de interés. Se realizó usando como *query* la secuencia de chimpancé (AJ507127.1) sobre la *Nucleotide collection (nr/nt) database*, y optimizando el proceso llevándolo a cabo como un *megablast* (selecciona solo aquellas secuencias con una similitud muy elevada). Los resultados obtenidos correspondieron únicamente a primates, de manera que se afinó la búsqueda empleando la base de datos *"Refseq_genomics"* y limitándola a primates.

Para el rastreo de posibles inserciones en el genoma humano se empleó la herramienta *Blast Like Alignment Tool* (BLAT, https://genome.ucsc.edu/cgi-bin/hgBlat?command=start) desarrollado por la *University of California, Santa Cruz* (UCSC), un algoritmo similar al BLAST, pero con una estructura diferente, ya que en lugar de buscar secuencias en el *GenBank*, lleva a cabo una búsqueda directa sobre un genoma completo, encontrando rápidamente aquellas que presentan una similitud del 95% o mayor, con una longitud mínima de 40 pares de bases (pb). De nuevo, la secuencia de chimpancé fue utilizada como *query*. Por último, como ensamblaje se seleccionó la base de datos del genoma humano correspondiente a "Feb. 2009 (GRCh37/hg19)".

Finalmente, para buscar las secuencias se empleó el navegador genómico correspondiente: *"Ensembl Genome Browser"* (http://www.ensembl.org/index.html) o *GenBank* en el NCBI (http://www.ncbi.nlm.nih.gov/nuccore/); y se descargaron como un archivo en formato FASTA.

Por otro lado, se realizó una búsqueda de posibles copias de las secuencias provirales y LTRs solitarias que pudiese haber en el genoma de las especies en las que se encontró alguna secuencia de ERV-Fc1. Para ello, se llevó a cabo un *nblast* en el BLAST del NCBI sobre la *Nucleotide collection (nr/nt) database* y, para buscar copias se emplearon como *query* cada una de las secuencias encontradas, mientras que para la búsqueda de LTRs solitarias se utilizó la LTR 5' de chimpancé, cuyas coordenadas venían indicadas en el GenBank.

2.2. Identificación de las *target site duplications*

En base a las TSD, se podrá determinar si ha habido más de una inserción a lo largo de la historia evolutiva, ya que si una secuencia proviral se encontrase en más de un genoma como un ERV y procediese de un mismo evento de integración, coincidirían tanto la secuencia como la longitud de las TSD en cada genoma en el que se hubiese hallado. Así, si se encontrasen más

copias del mismo ERV, pero con TSD diferentes, se podría concluir que tuvieron lugar diferentes eventos de integración.

Para tratar de confirmar que consisten en inserciones retrovirales originales y no son el resultado de algún proceso de transposición secundaria o recombinación, y poder determinar si hubo más de un evento de inserción. Se extendieron las secuencias detectadas anteriormente por BLAT/BLAST a 20 pares de bases extra a ambos extremos de la secuencia, lo que permitiría inferir si existen o no tales TSD flanqueando la secuencia proviral. De nuevo, se usaron los navegadores genómicos correspondientes para buscar las secuencias correspondientes con esos 20 nucleótidos extra a cada extremo y se descargaron en formato FASTA. La búsqueda y reconocimiento de esas posibles TSD fue llevaba a cabo manualmente en un editor de texto.

2.3. Divergencia de las LTR y tiempo de inserción

Partiendo de que en el momento de la inserción, el provirus va a presentar una secuencia idéntica a la del retrovirus [7] y que, tal y como se acaba de explicar, las LTR son originalmente idénticas entre sí, se puede calcular el tiempo de inserción del retrovirus en el genoma observando las diferencias existentes entre ambos LTRs, ya que reflejarán las mutaciones que se hayan producido desde su integración [7].

Basándose en las coordenadas de las LTR indicadas en la secuencia de chimpancé (AJ507127.1) y gorila (AJ507128.1) en el Genbank, se pudieron identificar en el resto de secuencias llevando a cabo un alineamiento mediante la aplicación *ClustalW Multiple Alignment* incluido en el programa de alineamiento y análisis de secuencias *Bioedit* v. 7.2.5 (disponible en http://www.mbio.ncsu.edu/bioedit/bioedit.html). Posteriormente se procedió a la eliminación de las posiciones CpG para evitar una sobreestimación de la divergencia como consecuencia de la elevada tasa de mutación que estas presentan [25,26] por la desaminación de la citosina a uracilo. El uracilo no forma parte del DNA, por lo que es transformado a timina, que se empareja con adenina dando lugar a transiciones:

5'CG3' → 5'TG3'
3'GC5' 3'AC5'

De esta forma, se eliminaron las posiciones ocupadas a la vez por una posición CG y alguna de las mutaciones derivadas de la desaminación de la citosina: CA y TG. Para ello se siguió un criterio que considera significativas las posiciones de un conjunto de secuencias cuando >70% de ellas presenta el mismo nucleótido en esa posición [26].

Posteriormente, se fueron aislando las LTR 5' y 3' de la secuencia proviral de cada especie en un archivo FASTA y alineándolas nuevamente con el *ClustalW Multiple Alignment* en el *Bioedit v. 7.2.5*.

A continuación, se llevó a cabo el cálculo de la edad aproximada de la secuencia del provirus. Para ello se crearon intervalos de tiempo para las LTR de cada secuencia, basándonos en dos tasas de mutación (μ) [4], $2.1x10^{9}$ y $1.3x10^{9}$ sustituciones por lugar nucleotídico y por año, lo que quiere decir que cada año se produciría una sustitución por cada $1/(2.1x10^{9})$ y $1/(1.3x10^{9})$ nucleótidos (476190476.2 y 769230769.2 respectivamente). En primer lugar, se calculó, empleando el programa MEGA v6.06 [27] (disponible en http://www.megasoftware.net/), el número de sustituciones nucleotídicas por posición (S) entre las LTR 5' y 3' de cada secuencia usando el modelo Kimura 2-parámetros, considerando tanto las transiciones como las

transversiones. A continuación, se procedió al cálculo de las edades (T) en millones de años (Mya) a partir de las S obtenidas [26,28]:

$$S = t[años] \cdot 2 \cdot \mu$$

$$t = \frac{S}{2 \cdot \mu} \rightarrow T[Mya] = 10^{-6} \cdot \frac{S}{2 \cdot \mu}$$

Se construyó así, un intervalo de edad (llevando a cabo el cálculo con ambas tasas de mutación) para cada secuencia que se pudo comparar, permitiendo observar si éstos son muy similares y se solapaban, lo que puede significar que se trata de la misma inserción, y localizar en qué momento aproximado de la historia evolutiva pudo haber tenido lugar la integración del retrovirus.

2.4. Identificación de las regiones retrovirales codificantes y su estudio

Se identificaron las regiones retrovirales codificantes más importantes. En primer lugar, la poliproteína gag-pr-pol, que incluye: proteínas estructurales (*gag*), proteasa (*pr*), retrotranscriptasa (*rt*), *RNAsaH (rh)*, integrasa (*in*); y por otro lado la poliproteína de los elementos de la envoltura (*env*). Además, se determinó el grado de funcionalidad aparente de cada una mediante la búsqueda de ORFs, permitiendo detectar la presencia de codones de paro prematuros y modificaciones en la pauta de lectura.

Para la búsqueda de posibles ORFs que contuviesen las diferentes regiones o parte de ellas se empleó el programa *Bioedit* v. 7.2.5. Se tradujeron las secuencias mediante la función *"Unsorted Six-Frame Translation"* limitando la búsqueda de ORFs con un tamaño mínimo de 20 pares de bases y cualquier codón de inicio (Como *gag* y *env* son las únicas regiones retrovirales codificantes que han de comenzar por ATG, una vez encontrados los ORFs que contengan dichas regiones, se repite la búsqueda de ORFs indicando como codón de inicio ATG para afinar las coordenadas; solamente variará la coordenada de inicio, por que resultará sencillo identificarlos). En principio, se fueron anotando las coordenadas de los ORFs de tamaño relativamente alto (>120).

Por otro lado, para buscar las regiones se empleó la herramienta *CD-Search* (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) del *NCBI*, que permite detectar dominios conservados comparando nuestro *query* con numerosas bases de datos. El programa devuelve una salida en la que se representan los dominios encontrados, con su marco de lectura o *reading frame* y sus coordenadas en el *query* contadas en aminoácidos, además de dar información acerca de la proteína. En primer lugar, se empleó como *query* cada una de las secuencias completas para poder tener una visión global de la propia secuencia y sus regiones retrovirales, pudiendo así tomar las coordenadas en aminoácidos de cada una de ellas y detectar posibles cambios en la pauta de lectura que las hayan truncado. A continuación, para poder determinar las coordenadas de cada región en pares de bases, se utilizó el *Bioedit* v.7.2.5 para llevar a cabo la traducción de la secuencia a estudiar en el marco de lectura que el *CD-Search* indicaba que se situaba la región que queremos identificar.

Posteriormente, para estudiar la integridad de las regiones identificadas se llevó a cabo una búsqueda en el CD-Search con cada uno de los posibles ORFs encontrados con *Bioedit*, con lo que se pudo ir observando qué regiones se encontraban en cada uno de ellos y si estaban o no truncadas, pudiendo inferir en tal caso, observando en *Bioedit* las zonas sospechosas, si se debía a una mutación sin sentido o a un cambio en la pauta de lectura.

Finalmente, se elaboró un archivo con todas las secuencias encontradas y se llevó a cabo un alineamiento, empleando nuevamente la herramienta *ClustalW Multiple Alignment* incluida en el *Bioedit*. En este alineamiento se creó una nueva secuencia sobre el resto en la que se fueron delimitando cada una de las regiones y sus codones para poder usarlo posteriormente como guía, permitiendo una rápida búsqueda de cada región.

2.5. Reconstrucción de la secuencia activa

Entendemos como secuencia activa a aquella secuencia retroviral original y funcional que llevó a cabo su inserción en el genoma como consecuencia de una infección que consiguió traspasar la línea germinal en algún momento de la historia evolutiva, y ha ido acumulando cambios y divergiendo en cada una de las líneas evolutivas en las que se encuentra en la actualidad. Ésta ha de aproximarse lo máximo posible a una secuencia que, en principio, no presente truncamientos en sus regiones codificantes gag-pr-pol y env, además de presentar unas LTR idénticas y unos sitios CpGs no mutados. Todo ello partiendo de que la secuencia retroviral se encontrase intacta en el momento de la inserción. Este estudio se centra en la reconstrucción de las regiones codificantes del ERV-Fc1 y se llevó a cabo un proceso similar al realizado por Lee y Bieniasz en 2007 [29].

En primer lugar, a partir del alineamiento con todas las secuencias y sus regiones identificadas, se creó su secuencia consenso con el *Bioedit* empleando la función ubicada en la pestaña *"Alignment"*. Esta secuencia representa los nucleótidos o aminoácidos que se encuentran con mayor frecuencia en cada posición de un conjunto de secuencias y por tanto será la que más se aproxime a la original. A continuación, se llevó a cabo la corrección manual de los gaps que no se generaron correctamente (serán aquellos que se encuentren en la mayoría de las secuencias del alineamiento pero no en la secuencia consenso creada) y de las sustituciones en las posiciones CpGs en las que aparecía CA o TG en lugar de CG tal y como se explicó en el punto 2.3..

El siguiente paso fue tratar de conseguir que las regiones codificantes gag-pr-pol y env se tradujesen al completo como si de la secuencia retroviral original se tratase, para lo que se observó en qué posiciones se truncaba la traducción y poder así realizar las modificaciones pertinentes, si fuese posible, para tratar de corregir los cambios de pauta de lectura y los codones stop prematuros que pudiésemos encontrar. Con el fin de obtener una secuencia externa con la que poder comparar la secuencia consenso en posibles casos de incertidumbre, se buscó la secuencia más similar llevando a cabo un BLAST desde el NCBI, usando como *query* la secuencia consenso. Se realizó sobre la base de datos *Nucleotide collection (nr/nt)* empleando el algoritmo *discontinuous megablast*, que se trata de un algoritmo más sensible. Una vez obtenida, se prosiguió llevando a cabo el alineamiento de todas las secuencias, para lo que se empleó el algoritmo *MUSCLE (Multiple Sequence Comparison by Log- Expectation)* del *EMBL (European Molecular Biology Laboratory),* disponible en http://www.ebi.ac.uk/Tools/msa/muscle/. Para determinar qué truncamientos se podían encontrar en la secuencia se siguió el mismo proceso que para la identificación de las regiones en las secuencias provirales, es decir, se empleó el *CD-Search* para detectar posibles cambios en la pauta de lectura y se llevó a cabo la traducción de la secuencia a través de *Bioedit* para detectar los ORFs y sus coordenadas, pudiendo identificar así en el alineamiento si éstos coincidían con los encontrados en las otras secuencias y si presentaban algún codón stop.

2.6. Puesta a prueba de hipótesis de la selección en la evolución del grupo

Tal y como ya se ha indicado, numerosos estudios hacen referencia a la implicación de ciertos retrovirus endógenos en múltiples procesos complejos de los organismos como: la placentación y desarrollo embrionario [3,21,22], resistencia antiviral [5,23], incremento de la diversidad genética, estabilización de los extremos cromosómicos junto con la telomerasa [23]… Suponiendo así, una ventaja para el hospedador, lo que explicaría cómo muchos retrovirus endógenos se han mantenido durante millones de años con múltiples ORFs intactos [23]. Si éste fuera el caso de ERV-Fc1, se tendrían que detectar indicios de selección en la evolución del mismo, observándose una baja divergencia en aquellas regiones que estuviesen desempeñando alguna función en beneficio del hospedador, como consecuencia de una menor influencia de la selección en ellas [6,20,22]. En este trabajo, se trató de buscar dichos indicios en las secuencias provirales encontradas, para realizar posteriormente un test estadístico con la finalidad de determinar si la selección en la evolución del grupo retroviral ERV-Fc1 es significativa o no.

Se partió del archivo FASTA creado con anterioridad que incluía el alineamiento de todas las secuencias provirales encontradas mediante el BLAT/BLAST, y en el que se había llevado a cabo la corrección manual de los gaps y la eliminación de las posiciones variables CpGs. En primer lugar, empleando el programa DNAsp v5.10.01 (disponible en http://www.ub.edu/dnasp/) desarrollado por la Universidad de Barcelona, se llevó a cabo un análisis del polimorfismo presente en las secuencias. Se obtuvo la diversidad nucleotídica (π) mediante el modelo *Jukes & Cantor* (predeterminado por DNAsp) y se representó en una gráfica de ventanas deslizantes con *Window lenght*=300 posiciones y *Step size*= 10 posiciones. En la gráfica se va representando la variación de π a lo largo de las posiciones de la secuencia, de manera que podemos identificar aquellas regiones en las que π se encuentra próximo a 0, lo que va a significar que casi no va a presentar variación entre las secuencias y por lo tanto que se ha mantenido muy estable desde su inserción en el genoma. Estas regiones con una baja divergencia entre las secuencias, tal y como se explicó, son candidatas a estar desempeñando alguna función positiva en el genoma del hospedador, manteniéndose a lo largo de la evolución por la actuación de alguna presión selectiva [6,20,22].

Una vez identificadas las posibles regiones con función positiva, se llevó a cabo un análisis descriptivo de las sustituciones sinónimas y no sinónimas de cada una de las mismas para rastrear evidencias de selección. Para llevar a cabo este análisis es necesario que en el alineamiento estén presentes las posiciones CpGs, ya que se busca analizar los sitios sinónimos y no sinónimos en los codones de las secuencias, de modo que si se llevase a cabo sobre un alineamiento en el que previamente se hubiesen eliminado tales posiciones, los codones que tomaría el programa no se corresponderían con los de las secuencias originales, viéndose alterados los resultados. Así pues, se aislaron en archivos independientes las regiones sospechosas a partir de un alineamiento en el que las secuencias presentasen las CpGs, y se fueron eliminando manualmente a través del *Bioedit* las posiciones en las que se hubiesen podido originar *gaps*, guardándose posteriormente en formato FASTA. A continuación, por medio del DNAsp, se llevó a cabo el propio análisis descriptivo de las sustituciones sinónimas y no sinónimas en cada uno de los archivos, para lo que fue necesario, en primer lugar, asignar las regiones codificantes desde el menú *Data* (se guardó cada archivo con formato NEXUS para mantener la configuración y acelerar el proceso en posibles análisis posteriores); en el análisis fueron consideradas las sustituciones sinónimas en las regiones codificantes y no codificantes. El programa devuelve una salida de texto en la que se explican las características de la

secuencia analizada, indicando el número total de sitios y codones analizados, así como las regiones codificantes y no codificantes. Por otro lado, genera una tabla en la que aparecen comparadas dos a dos las secuencias del alineamiento y en cada caso se da el valor de *SilentDif* (número total de diferencias sinónimas o silenciosas), *SilentPos* (número total de sitios silenciosos o sinónimos) , K_s (número de sustituciones sinónimas o silenciosas por sitio sinónimo o silencioso), *NSynDif* (número total de diferencias no sinónimas), *NSynPos* (número total de sitios no sinónimos) y K_a (número de sustituciones no sinónimas por sitio no sinónimo).

A partir del cociente K_a/K_s, podemos determinar si está actuando algún tipo de selección o no sobre una secuencia y de qué tipo se trata [30]. De este modo, podremos distinguir tres casos:

- **$K_a/K_s=1$**: el número de sustituciones sinónimas y no sinónimas por sitio son iguales, de modo que todas las mutaciones sinónimas y no sinónimas son neutras y no se está produciendo ningún tipo de selección. Estaríamos ante un **pseudogen**, es decir, una secuencia de DNA que carece de función por acumulación mutaciones a lo largo la evolución [31]. De este modo, una desviación de este valor, estaría dando evidencias sobre la actuación de alguna fuerza selectiva sobre la secuencia [30].
- **$K_a/K_s<1$**: el número de mutaciones no sinónimas es menor que el número de mutaciones sinónimas, es decir, existe una fuerza selectiva que tiende a mantener funcional la secuencia, impidiendo la fijación de mutaciones deletéreas. Estamos ante un caso de **selección purificadora**. Es el caso más común, ya que las mutaciones no sinónimas que modifican una proteína son mucho menos probables que las mutaciones sinónimas [30] en la tercera posición de los codones.
- **$K_a/K_s>1$**: el número de mutaciones no sinónimas es mayor que el de mutaciones sinónimas, lo que supone una evidencia de la presencia de una fuerza selectiva que actúa favoreciendo el cambio de la secuencia, es decir, se trata de una **selección positiva** [30].

Una secuencia funcional tenderá a presentar $k_a<k_s$, ya que se encuentran respaldadas por mecanismos de reparación del DNA, y por lo tanto a estar sometida a una selección purificadora que haga posible su persistencia a lo largo del tiempo. Así, este análisis permite predecir, antes de realizar un test estadístico, si se tratan o no de pseudogenes las regiones identificadas como sospechosas de ser beneficiosas para el hospedador, pudiendo detectar si hay indicios de que se esté produciendo una selección sobre ellas.

Finalmente, se llevó a cabo un test estadístico para determinar si existe un apoyo significativo que dé una mayor seguridad a la hora de concluir si existe una selección o no. El test realizado fue un test Z de selección desde el programa MEGA v6.06 en cada una de las regiones sospechosas de realizar alguna función importante en el genoma, excluyendo del análisis la secuencia consenso en cada caso. Este test se basa en el estadístico k_s-k_a y toma como hipótesis nula la presencia de neutralidad, H_0: $k_a=k_s$ (ausencia de selección), frente a la hipótesis alternativa H_1: $k_a<k_s$ ó $k_a>k_s$ en función de si queremos evaluar si existe selección purificadora o positiva, respectivamente. En este estudio, se quiere conocer si las regiones encontradas por medio del análisis de polimorfismos a través del *DNAsp,* desempeñan realmente alguna función positiva para el hospedador, teniendo que estar, en tal caso, sujetas a una selección purificadora. De este modo, se llevó a cabo dicho test Z con una H_0: $k_a=k_s$ y H_1: $k_a<k_s$ a un nivel de significación de $\alpha=0.05$, comparando las secuencias por pares y eliminando las posiciones con *gaps*. Además, la varianza del estadístico fue computada siguiendo el método *Bootstrap* con 10000 replicas y los análisis fueron realizados siguiendo el modelo de

sustitución Nei-Gojobori [32]. Rechazaríamos, por lo tanto, H_0 en aquellas comparaciones dos a dos de las secuencias en las que el test fuese significativo al 5% (p-valor<0.05), lo que permitiría aceptar H_1 al no tener pruebas de lo contrario.

A priori, cabría pensar que se trata de pseudogenes al consistir en un DNA exógeno integrado en un genoma, que en un principio no tendría por qué tener una función positiva en el huésped, de manera que iría acumulando indistintamente mutaciones sinónimas y no sinónimas sin existir sobre la secuencia proviral una fuerza selectiva.

2.7. Reconstrucción filogenética de la secuencia completa y de cada región

Con el fin de conocer cómo tuvo lugar la evolución de las secuencias provirales encontradas, para poder así observar si coincide con la historia evolutiva conocida [33,34], e identificar posibles casos de recombinación o transferencia horizontal, se llevó a cabo la reconstrucción filogenética del conjunto de secuencias provirales encontradas. A mayores, se obtuvo la filogenia para cada región retroviral individualmente, con el propósito de determinar si existe una evolución en mosaico de las regiones, es decir, que no se ajusten a la del resto como consecuencia de diferencias en sus tasas evolutivas, y si es el caso, tratar de determinar las posibles causas de ello.

El primer paso, fue buscar la raíz u *outgroup* que mejor se ajustase a nuestros datos. Ésta tendría que ser una secuencia externa a nuestro grupo, pero con alta similaridad, de modo que se utilizó la secuencia que había sido encontrada por medio del BLAST empleando como *query* la secuencia consenso para reconstruir la secuencia activa. A continuación, se alinearon las cuatro secuencias de ERV-Fc1 con el *outgroup* y se descartaron las LTRs. Además, fueron eliminadas las posiciones CpGs a lo largo del alineamiento por lo referido anteriormente. Posteriormente, se creó un archivo FASTA de cada una de las regiones alineadas. Finalmente, se llevó a cabo la reconstrucción filogenética partir del alineamiento de la secuencia completa y del de cada región.

La reconstrucción filogenética fue realizada mediante dos métodos, empleando como plataforma el programa MEGA v6.06.:

- **Método de Máxima Parsimonia (*Maximum Parsimony method*):** se utilizó el método *Bootstrap* con 100 réplicas para el test filogenético y el algoritmo "*Subtree-Pruning-Regrafting"* (SPR) [35] para la obtención el árbol más parsimonioso. Además, los *gaps* fueron excluidos del análisis.
- **Método de Máxima Verosimilitud (*Maximum Likelihoodmethod*):** se realizó considerando las sustituciones nucleotídicas en base al modelo "*General Time Reversible model"*. El cálculo del árbol inicial se utilizó el método automático "*Default – NJ/BioNJ*". Para determinar las diferencias de tipo evolutivo entre los sitios se utilizó una distribución discreta de tipo Gamma. Finalmente, los *gaps* fueron excluidos del análisis.

Los resultados fueron representados sobre los árboles de Máxima Parsimonia obtenidos, indicándose por encima de las ramas el apoyo *Bootstrap* obtenido con este método, y por debajo el obtenido con el método de Máxima Verosimilitud. Como valor de *cutoff* tomamos 70%, de manera que las ramas con un valor *Bootstrap* por debajo de 70 no se consideran informativamente significativas.

3. Resultados

3.1. Búsqueda de secuencias provirales de interés

Mediante la búsqueda directa de ERV-Fc1 en *Nucleotide*, se encontraron dos secuencias provirales completas (Fig. 4) pertenecientes al genoma de *Pan troglodytes* (AJ507127.1) y *Gorilla gorilla* (AJ507128.1).

Fig. 4. Resultados obtenidos en la búsqueda introduciendo "ERV-Fc1" Nucleotide.

Por medio de la herramienta BLAT de la UCSC para la búsqueda de inserciones de ERV-Fc1 en el genoma humano, empleando como query la inserción en chimpancé, AJ507127.1, se obtuvo un resultado que parecía corresponderse a una inserción completa al presentar un 98.8% de identidad con el *query*, además de cubrirlo por completo. Esta inserción presentó una longitud de 7943 pb, con coordenadas X:97096480-97104422 (Fig. 5). La misma secuencia fue encontrada por medio de un BLAST en el NCBI, con coordenadas 40963-48905 en el número de acceso AL354685.17.

Human BLAT Results

BLAT Search Results

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
browser details	YourSeq	7753	1	7997	7997	98.8%	X	+	97096480	97104422	7943
browser details	YourSeq	580	1626	7920	7997	77.7%	11	+	5951397	5959516	8120
browser details	YourSeq	329	7629	7997	7997	95.1%	X	+	97096480	97096855	376
browser details	YourSeq	267	239	2440	7997	74.9%	7	-	153109124	153111021	1898
browser details	YourSeq	262	77	380	7997	92.3%	X	+	97104124	97104422	299
browser details	YourSeq	219	239	7920	7997	77.4%	2	-	84101125	84106271	5147
browser details	YourSeq	194	2119	7920	7997	72.1%	Y	+	19723314	19726804	3491
browser details	YourSeq	186	241	7932	7997	76.4%	7	-	64295312	64300317	5006
browser details	YourSeq	166	2119	7920	7997	71.4%	Y	-	20452970	20456461	3492
browser details	YourSeq	130	2747	3322	7997	61.3%	4	-	8486767	8487342	576
browser details	YourSeq	116	2747	3331	7997	60.9%	1	+	222124164	222124748	585
browser details	YourSeq	111	2747	3235	7997	61.4%	2	+	36829074	36829562	489
browser details	YourSeq	110	2747	3328	7997	59.9%	1	-	23327696	23328276	581
browser details	YourSeq	108	2747	3238	7997	61.0%	3	-	86589880	86590371	492
browser details	YourSeq	107	2747	3235	7997	61.0%	3	+	80436925	80437413	489
browser details	YourSeq	98	1470	1931	7997	64.7%	2	-	79324539	79325005	467
browser details	YourSeq	95	241	7920	7997	69.6%	17	+	26593460	26595613	2154
browser details	YourSeq	94	2747	3182	7997	60.8%	2	-	45263977	45264412	436
browser details	YourSeq	84	2747	3178	7997	59.8%	2	-	209083366	209083797	432
browser details	YourSeq	84	3027	3238	7997	69.9%	2	+	207885146	207885357	212

Fig. 5. Resultado de la búsqueda de secuencias provirales empleando la herramienta BLAT de la UCSC. Query: "ERV-Fc1 Chimpanzee". En rojo, el resultado significado a la búsqueda.

Llevando a cabo el BLAST del NCBI limitando la búsqueda a primates y empleando de nuevo como *query* la inserción proviral de chimpancé, AJ507127.1, además de encontrar las secuencias ya obtenidas, se obtuvo un nuevo resultado, correspondiente a una secuencia proviral de *Pan paniscus* (Bonobo), con coordenadas 8858721-8866716 en el número de acceso NW_003870294.1. Esta secuencia parecía tratarse del provirus completo, ya que presentaba una identidad del 99% y una cobertura del *query* de 100%.

Es decir, finalmente se encontraron 4 secuencias provirales aparentemente completas: Cimpanzee ERV-Fc1 (AJ507127.1), Gorilla ERV-Fc1 (AJ507128.1), Human ERV-Fc1 (X: 97096480-97104422) y Bonobo ERV-Fc1 (8858721-8866716 en NW_003870294.1). Además, ninguna copia de las secuencias ni ninguna LTR solitaria fueron detectadas en el genoma [36], lo que podría significar que se trata de una inserción relativamente reciente y estable.

3.2. Identificación de las *target site duplications*

Se pudo observar la presencia de TSD en los cuatro provirus, de una longitud de 5 pares de bases. Además, todas ellas coincidieron, presentando la secuencia AAAAT (Fig. 6), lo que indica que las secuencias provirales proceden de una misma inserción retroviral en un antecesor común a chimpancé, gorila, bonobo y humano, que fue divergiendo a lo largo de la evolución.

```
>X dna:chromosome chromosome:GRCh37:X:97096460:97104442:1
TGGTTTTATTAACATAAAATTGTTAGGCAGGTCACCCAAGATGGCCGTTCCTCCAGGACC
CAAGATGGCAGCACCAACCCCTTCTCCCCCCACCCCCGCCCCCCGCCCGCTTGGAATCTC
CCACCAGATTTTCCTGCTGGACGGGCACTTTCAGATGACTGCAGCCCCGAGAAGTCGAAA
CCTATCCCAGAAAACCGAAACTTACTAAGCCCCTCCCCGCGTGCTCTATAAAAACCCTCT
ACTGCCCCAGTCGGGCGCGACTTCCCTGGCCCTCCTTGTTAGGACCAGTGAACCTCGCCC
GAGAGCTCCATTAATAAAGCAGGTCGCCTCTGACCATTAGTCACCTAAATTCTGTGCGGT
AGTTCTCATTGGATACCTGTCTTCCCAAGCCGGACATTGGTGCCAAAACCCGGGAGGAGA
CCCCTCTCTGACCCAGGGTCGGGGAGCATCTCCTCTCCCTACCTGCCAGGAACCAGACTC
                              ...
TCAGCGAAGTCTCCCGGGTGACGGTCAACCAAATGTTACTACACCCTTACTCCCGTCTTC
CGACCTCCGAAGACCACTATGACGACGCCCTCACTCAGCAGGAAGCAGCCAGATGATTAC
GTCGCCCCTTTTTCTTACAGTATGAGGTCGGAATGTTAGGCAGGTCACCCAAGATGGCTG
TTCCCCCAGGACCCAAGATGGCGGCACGAACCCCTTCTTCCCCGCCCCCCCCACCGCTTG
GAGTCTCCCACCAGATTTTCCCGCCCGACGGGCACTTTCCGATGACAGCAGCCCCGAGAA
GTCGAAACCTATCCCAGAAAACCGAAACTTACTAAGCCCCTCCCCACACGCTCTATAAAA
ACCCTCTACTGCCCCAGTCGGGTGCGACTTCCCTGGCCCTCCTTGTTAGGACCAGTGAAC
CTCGCCCGAGAGCTCCATTAATAAAGCAGGTCGCCTCTGACCATTAGTTACCTAAATTCT
GTGCGGCACTTCTCATTGGATACCTGTCTTCCCAAGCCGGACAAAAATGAATAAAACAAA
ATT
```

Fig. 6. Representación de los extremos 5' y 3' de la ERV-Fc1 en humano con sus TSD (en rojo)

3.3. Divergencia de las LTR y tiempo de inserción

Las localización de las LTR en el provirus de cada especie permitió obtener las coordenadas en la secuencia retroviral (Tabla 3).

Tabla 3. Coordenadas de LTR 5' y LTR 3' en las secuencias de los ERVs de Pan troglodytes, Pan paniscus, Homo sapiens y Gorilla gorilla.

	LTR 5'	*LTR 3'*
P. troglodytes	1-380	7629-7997
P. paniscus	1-377	7628-7996
H. sapiens	1-376	7574-7943
G. gorilla	1-374	7630-7998

Tras el cálculo del número de sustituciones nucleotídicas por sitio entre las LTR de cada secuencia proviral, se llevó a cabo el cálculo de la edad de dichas inserciones en base a las dos tasas de mutación indicadas ya en materiales y métodos, obteniendo los resultados obtenidos en la tabla 4.

Tabla 4. Datos obtenidos para el número de sustituciones nucleotídicas existente entre las LTRs 5' y 3' de cada secuencia proviral y el cálculo de sus edades en millones de años. T_1: edad basada en una $\mu=2.1x10^{-9}$; T_2: edad basada en una $\mu=1.3x10^{-9}$.

	Sust. nucleotídicas	*T_1*	*T_2*	*Intervalo de edad*
P. troglodytes	0.05	11.9047619	19.23076923	11.90-19.23
P. paniscus	0.053	12.61904762	20.38461538	12.62-20.38
H. sapiens	0.053	12.61904762	20.38461538	12.62-20.38
G. gorilla	0.059	14.04761905	22.69230769	14.05-22.69

Se puede observar que la inserción es relativamente reciente [12] en comparación con la mayoría de retrovirus endógenos, cuya antigüedad puede incluso superar los 100 Mya [37–39] . Además, a pesar del solapamiento entre los intervalos de edades estimados, el de gorila resulta ligeramente superior al de bonobo, humano y chimpancé, que en el caso de los dos primeros son iguales y muy próximos al del último. Esto podría deberse a que las diferencias entre chimpancé, bonobo y humano son menores que entre gorila. Todo ello lleva a pensar que tuvo lugar una única inserción de ERV-Fc1 en la historia evolutiva, y que se produjo en algún ancestro común a todos ellos, originándose posteriormente dos vías evolutivas [12]: la del gorila por un lado, y la de chimpancé, bonobo y humano por otra; lo que coincidiría con la realidad de la historia evolutiva conocida [34] y con el resultado de diversos estudios filogenéticos llevados a cabo a partir de ERVs [33,39].

3.4. Identificación de las regiones retrovirales codificantes y su estudio

La búsqueda de las regiones retrovirales y el estudio de su grado de funcionalidad a partir de la búsqueda de ORFs en las secuencias provirales y su posterior rastreo a través del CD-Search fue completada, permitiendo determinar qué mutaciones se han originado a lo largo de su evolución, dando lugar al actual provirus defectivo [3,5].

Se pudo observar que todas las secuencias provirales obtenidas se encuentran completas (5'LTR-gag-pr-rt-rh-in-env-LTR3'), y que se trata de un retrovirus endógeno con la estructura proviral general [1,3,5,7,14]. Por otro lado, también fue notable el alto grado de conservación que éstas presentaban, tanto al observarlas independientemente como al compararlas entre sí. En cuanto a indels en las secuencias, únicamente se podría destacar uno de corta longitud (55 pb) en el ERV del genoma humano, pero que no se encuentra afectando a ninguna región codificante (Fig. 7). Además, la secuencia proviral de gorila parecía nuevamente presentar una trayectoria ligeramente diferente a las de los otros tres primates. De nuevo, se puede sugerir que se trata de una inserción relativamente reciente que tuvo lugar en un momento próximo a la divergencia de las líneas evolutivas del gorila y del chimpancé, bonobo y humano. Además, el alto grado de conservación de las secuencias lleva a pensar que podría estar desempeñando alguna función que resultase ventajosa para el hospedador como ocurre en otros ERVs [3,5,24].

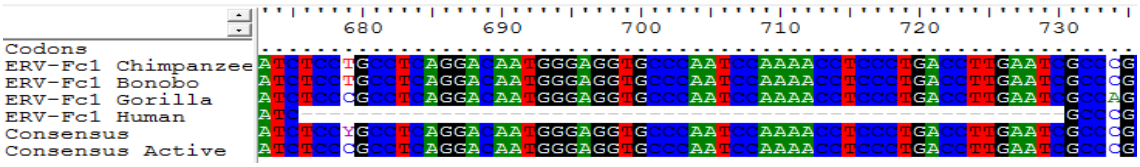


Fig. 7. Posible gap en el ERV-Fc1 de humano observado en el alineamiento de las secuencias de ERV-Fc1 de chimpancé, bonobo, gorila y humano. Se incluyen también la secuencia consenso y la secuencia activa reconstruida a partir de la secuencia consenso.

Las coordenadas encontradas de las regiones en cada secuencia proviral coincidieron en el alineamiento, lo que indica que la búsqueda se realizó exitosamente. En los archivos anexos del CD se recogen las tablas de coordenadas de los ORFs encontrados en cada una de las secuencias provirales junto con el de las regiones codificadoras.

Las regiones *gag-pro-pol* y *env* se solapan, pero esto no sería un problema a la hora de traducirse cuando se trataba de un provirus activo, ya que tal y como vimos en la introducción, cada región se traduciría independientemente, dando lugar a dos mRNA que finalmente se traducirían originando dos poliproteínas que serían finalmente procesadas [7].

Los ORFs fueron los mismos para "*ERV-Fc1 Chimpanzee"* y "*ERV-Fc1 Human"*, pero localizados en distintos marcos de lectura o *Reading frames* (RF) pareciendo haber un corrimiento +1 en humano (Anexo 2 y 4). En ellos, únicamente se observó truncamiento por mutación de cambio del marco de lectura o *frameshift mutation* en la *rh*, de RF+3→RF+1 en "*ERV-Fc1 Chimpanzee"* y RF+2→RF+3 en "*ERV-Fc1 Human"*. En cuanto al provirus en Bonobo "*ERV-Fc1 Bonobo*", además de la *frameshift mutation* de la *rh* (RF+1→RF+2), se observa una segunda en *pr* (RF+2→RF+1) (Anexo 5). Por otro lado, el provirus de gorila "*ERV-Fc1 Gorilla*" parece encontrase más degenerado, pudiéndose observar tres *frameshift mutations*: en *gag* (RF+2→RF+3), también en la *rh* (RF+3→RF+1) como el resto y en *env* (RF+2→RF+3) (Anexo 3).

Además de las *frameshift mutations*, se detectaron diversos codones stop prematuros como consecuencia de alguna mutación sin sentido o *nonsense mutation*. En todas las secuencias provirales se detectó el codón stop (TGA) característico entre *gag* y *pr*, además de encontrarse ambos en la misma RF, lo que significa que el retrovirus original pertenecería a esa minoría de retrovirus en los que el salto del codón stop para la síntesis de *gag-pr-pol* se llevaría a cabo mediante una mala lectura del mismo y la colocación de un aminoácido en su lugar. A mayores, se encontró, también en todas ellas, una *nonsense mutation* (TGA) entre los ORF que contienen la *rh* e *in* (Fig. 8).

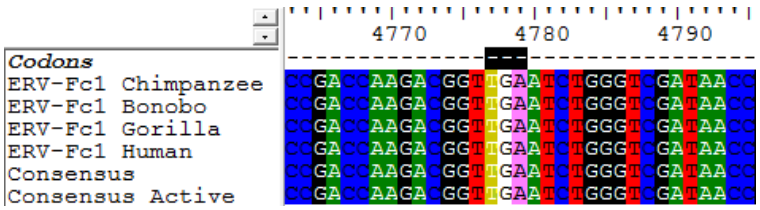


Fig. 8. Representación gráfica del codón stop prematuro entre los ORF de la rh e in sobre el alineamiento de las secuencias.

Finalmente, el provirus de gorila vuelve a ser el que presenta un mayor grado de degeneración con respecto a los otros, lo que advierte que pudo tomar una evolución independiente al resto. En él, se pudieron observar a mayores diversas *nonsense mutations* a lo largo de la secuencia: en *gag* se observó un codón stop prematuro anterior al natural (TAA, posiciones 1995-1997; Fig.9A), entre los ORF que incluían *pr* y *rt* (TGA, posiciones 2778-2780; Fig. 9B), los ORF de *rt* y *rh1* (TGA, posiciones 3492-3494; Fig. 9C); y en *env* (TGA, posiciones 6374-6376; Fig. 9D).

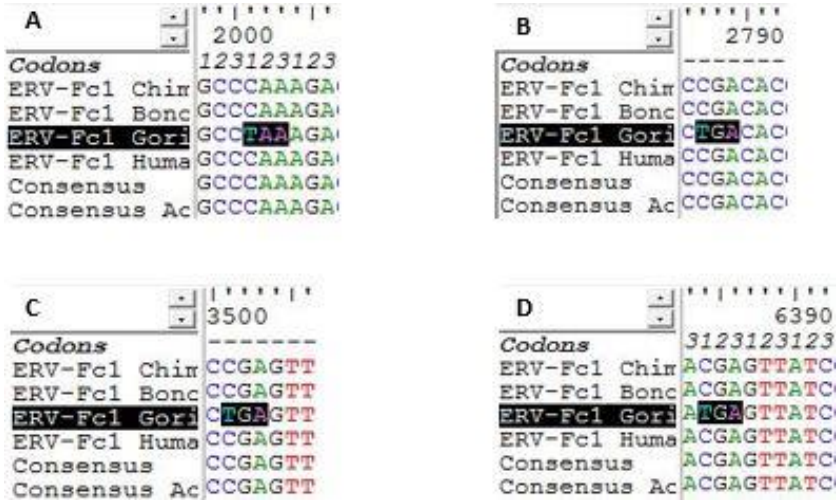


Fig. 9. Representación en el alineamiento de las secuencias de ERV-Fc1 de las mutaciones nonsense encontradas en la secuencia del gorila "ERV-Fc1 Gorilla". A: nonsense mutation en gag; B: nonsense mutation entre los ORF de pr y rt; C: nonsense mutation entre los ORFs de rt y rh; D: nonsense mutatin en env.

En resumen, el ERV-Fc1 de chimpancé, humano y bonobo parece interrumpido en *pol* por una *frameshift mutation* en la *rh* y una *nonsense mutation* entre los ORFs de *rh* e *in* [14] (Fig. 17), además de una segunda *frameshift mutation* en la región *pr* del "*ERV-Fc1 Bonobo"* (Fig. 17). Por otro lado, *env* parece estar intacto con un ORF [17,36], y con *gag*¸ a diferencia de lo leído en otros estudios, parece ocurrir lo mismo [17] salvo por el TGA natural que delimita con *pr*. En cuanto al ERV-Fc1 del gorila, *gag* va a estar interrumpido por una *frameshift mutation* y una *nonsense mutation*, sin contar con el TGA natural del provirus; la región *pr-pol*, va a presentar *nonsense mutation* entre los ORF de *pr* y *rt*, entre los de *rt* y *rh* y los de *rh* e *in,* además de una *frameshift mutation* en la *rh*; por último, en la región *env* también se detectaron truncamientos, distinguiéndose una *nonsense mutation* y una *frameshift mutation* (Fig. 17).

Es importante destacar que las coordenadas de *env* encontradas se corresponden con aquellas indicadas en las secuencias AJ507127.1 (*ERV-Fc1 chimpanzee*) y AJ507128.1 (*ERV-Fc1 gorilla*) en el GenBank, mientras que las de *gag* no, empezando en el siguiente ATG, lo que excluiría un codón stop TGA ubicado entre el ATG dado por el GenBank y el ATG encontrado en este estudio. Esto explicaría por qué, a diferencia de lo visto en otros estudios que sitúan dos codones stop en *gag* [13,14], en éste únicamente se encontró el codón stop que delimita con *pr*.

3.5. Reconstrucción de la secuencia activa

Como resultado del BLAST realizado para obtener la secuencia externa con la que tratar de solucionar las ambigüedades que pudieran surgir, se obtuvo una secuencia del genoma de macaco, con número de acceso AC210233.5 y coordenadas 132983-139912, que cubría el 93% del *query* (secuencia consenso) con una identidad del 69%. Durante la reconstrucción de las regiones codificantes de ERV-Fc1, se detectaron ciertas posiciones de la secuencia consenso obtenida en las que existía ambigüedad nucleotídica, consecuencia de que dos de las secuencias provirales presentaban un mismo nucleótido y las otras dos secuencias otro (Fig. 10), por lo que fueron corregidas, tomando como referencia la secuencia de macaco tras haberlas alineado. Al mismo tiempo se pudieron ir delimitando las coordenadas precisas que las regiones ocuparían en la secuencia proviral activa; fueron dadas teniendo en cuenta los LTR (Tabla 5). De este modo, teniendo en cuenta que *pol* no llega solo hasta donde el CD-Search indica que acaba la integrasa (última región de *pol*), sino hasta el codón stop del ORF, podemos

definir las coordenadas de los dos únicos ORFs que la secuencia activa original debería presentar, *gag-pr-pol*: 743-5903 y *env*: 5836-7590 (Anexo 6).

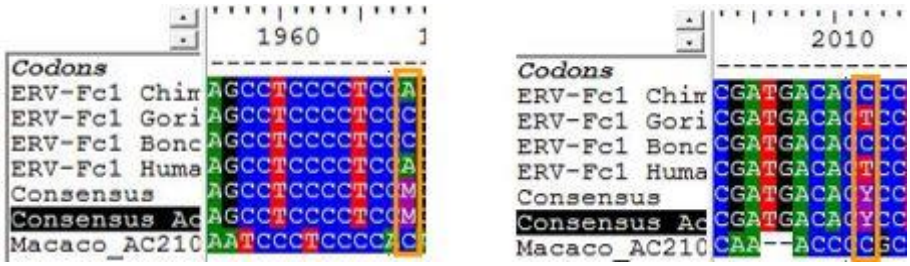


Fig. 10. Ejemplos de posiciones con ambigüedad nucleotídica.

Una vez reconstruida lo máximo posible, como era de suponer, el codón stop natural TGA entre *gag* y *pr* seguía presente en la secuencia. Lo sorprendente fue seguir encontrando el cambio de la pauta de lectura en la *rh* y el codón stop TGA entre la *rh* y la *in*, ya que en principio, en la secuencia retroviral original esto resultaría en un truncamiento de la poliproteína *pol* impidiendo su correcta traducción. Por otro lado, *env* resultó intacto con un solo ORF.

Tabla 5. Coordenadas de las regiones codificantes enteras en la secuencia ERV-Fc1 de cada genoma y en el provirus activo reconstruido (en rojo).

	ERV-Fc1 Chimpanzee	*ERV-Fc1 Bonobo*	*ERV-Fc1 Human*	*ERV-Fc1 Gorilla*	*ERV-Fc1 Active*
gag	744-2156	743-2155	686-2098	740-2156	743-2155
pr	2370-2618	2369-2616	2312-2563	2370-2618	2369-2617
rt	2826-3455	2824-3453	2771-3400	2826-3455	2825-3454
rh	4218-4644	4216-4642	4163-4589	4218-4644	4217-4643
in	5002-5349	5000-5348	4946-5294	5002-5349	5001-5348
env	5837-7591	5836-7590	5782-7536	5837-7592	5836-7590

3.6. Puesta a prueba de hipótesis de la selección en la evolución del grupo

Tras llevar a cabo el análisis del polimorfismo en el conjunto de secuencias províricas, solo se pudo detectar una zona con un aparente alto grado de conservación (presentando $\pi<0.005$) y candidato a estar llevando a cabo alguna función positiva importante para el hospedador (Fig. 11). Prestando atención a las coordenadas, se pudo comprobar que se trata de alguna pequeña zona ubicada entre las posiciones 6500 y 7000, lo que va a corresponder a la región *env*. Como ya se vio, esta región va a estar truncada en el gorila, por lo que se sabe de antemano que, en el posible caso de que desempeñase algún papel importante en las otras especies, en el gorila no sería viable.

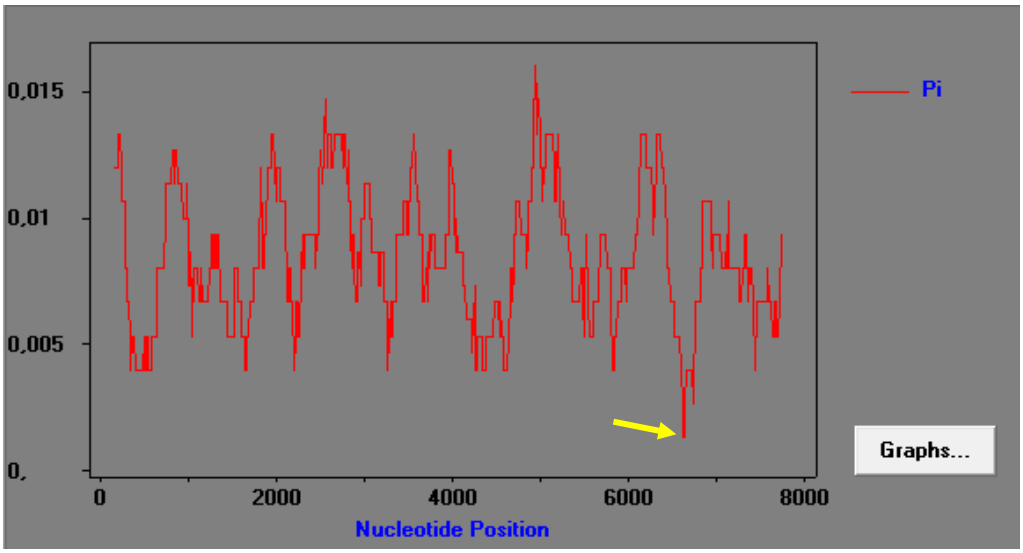


Fig. 11. Gráfica de la diversidad nucleotídica (π) a lo largo de las posiciones del alineamiento de las secuencias ERV-Fc1 de chimpancé, bonobo, humano y gorila. La flecha amarilla indica una zona con una π próxima a 0, y por lo tanto con un alto grado de conservación, que se corresponde con una zona dentro de la región env.

Una vez aislada la región *env* del alineamiento, se procedió a realizar el análisis inicial de las sustituciones sinónimas y no sinónimas. Los resultados generales del análisis vienen recogidos en la salida de texto del DNAsp (Fig. 12).

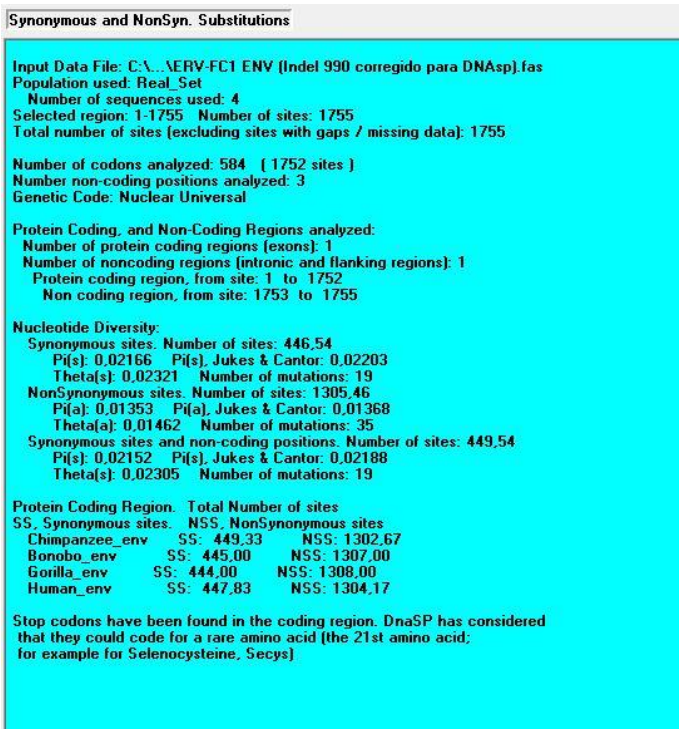


Fig. 12. Salida de texto con los resultados generales del análisis de las sustituciones sinónimas y no sinónimas realizado con el DNAsp.

Pero lo realmente importante es la comparación dos a dos de las secuencias que se origina en forma de tabla (Tabla 6). Se puede observar que, aunque algunas más y otras menos, todas las relaciones presentan una relación $K_a/K_s<1$, lo que quiere decir que el número de sustituciones no sinónimas es menor que el de sustituciones sinónimas y podría estar sometida a una selección purificadora que evita la acumulación de mutaciones deletéreas, haciendo posible

que la secuencia haya mantenido un grado de conservación tan alto a lo largo de la historia evolutiva. Además, parece que esta relación es menor siempre que una de las secuencias comparadas es el *env* del ERV de chimpancé, lo que podría estar indicando que esa posible selección purificadora actúa con mayor fuerza en esta secuencia. Por otro lado, en gorila, al no ser ya funcional la región *env*, ésta debería tender a equilibrarse en cuanto a sustituciones sinónimas y no sinónimas ($K_a/K_s \approx 1$), ya que estaría actuando como un pseudogen en el que no existiría ninguna fuerza selectiva. Esta observación se ve apoyada por los resultados obtenidos, ya que en todas las comparaciones con el *env* del ERV de gorila, se observa un valor elevado de la relación K_a/K_s, incluso cuando se compara con el del chimpancé. Además, la comparación entre las regiones *env* de los ERVs de humano y bonobo, se obtiene también una elevada relación K_a/K_s, lo que lleva a sospechar que en ellas o no existe selección purificadora o es muy baja, siendo la función beneficiosa para el hospedador escasa o ausente.

Tabla 6. Resultados de la comparación dos a dos de las secuencias ERV-Fc1 durante el análisis de las sustituciones sinónimas y no sinónimas. Seq 1 y 2: secuencias que se comparan; K_s: número de sustituciones sinónimas por sitio sinónimo; K_a: número de sustituciones no sinónimas por sitio no sinónimo; K_a/K_s: cociente entre K_a y K_s.

Seq 1	Seq 2	K_s	K_a	K_a/K_s
Chimpanzee_env	***Bonobo_env***	0.0157	0.0077	0.4904
Chimpanzee_env	***Gorilla_env***	0.0318	0.0210	0.6604
Chimpanzee_env	***Human_env***	0.0157	0.0054	0.3439
Bonobo_env	***Gorilla_env***	0.0296	0.0209	0.7061
Bonobo_env	***Human_env***	0.0090	0.0069	0.7666
Gorilla_env	***Human_env***	0.0295	0.0202	0.6847

Finalmente, tras realizar el test Z de selección, se observa que solo resulta significativo al 5% para la relación entre "*Chimpanzee_env*" y "*Human_env*" (Tabla 7), rechazándose en este caso la $\mathbf{H_0: k_a=k_s}$ (selección neutra) en favor de $\mathbf{H_1: k_a<k_s}$ (selección purificadora). Es decir, que en base al test, es muy posible que en alguna de las dos especies, la región *env* del ERV esté desempeñando alguna importante función para el hospedador. Examinando los p-valores obtenidos para el resto de relaciones se puede observar que, siguiendo con lo visto en el análisis de las sustituciones sinónimas y no sinónimas, siempre que está presente la secuencia de chimpancé el p-valor es de los más bajos obtenidos.

Tabla 7. Resultado del test Z realizado en la región env con el programa MEGA v6.06, para una H0: $k_a<k_s$, siguiendo el método Bootstrap con 10000 replicacs para el cálculo de la varianza del estadístico y el modelo de sustitución Nei-Gojobori para el análisis. Por debajo de la diagonal se indica el valor de los P-valores para cada comparación dos a dos de las secuencias ERV-Fc1. Por encima se indica el valor del estadístico ks-ka.

	1	2	3	4
1. Chimpanzee env		1.292	1.265	1.798
2. Bonobo env	0.099		1.030	0.433
3. Gorilla env	0.104	0.152		1.183
4. Human env	0.037	0.333	0.120	

Por lo tanto, en base a los resultados, la región *env* parece estar sometida a una selección purificadora en la secuencia de chimpancé, lo que podría estar indicando que tiene un papel importante en el organismo; por otro lado, en humano y bonobo esto no parece estar ocurriendo, y en gorila, como ya se indicó, la región se encuentra truncada por lo que se sabe de antemano que no puede ser funcional.

Por otro lado, aunque *gag* no apareciese como posible candidato en el análisis del polimorfismo, esta región se encuentra intacta en las secuencias ERV-Fc1 de chimpancé, bonobo y humano como sucedía con *env*, de modo que se llevó a cabo un test Z de selección

sobre ella. Sin embargo, no se obtuvieron resultados significativos al 5% de selección en ninguna de las secuencias (Tabla 8).

Tabla 8. Resultado del test Z realizado en la región gag con el programa MEGA v6.06, para una H0: $k_a<k_s$, siguiendo el método Bootstrap con 10000 replicacs para el cálculo de la varianza del estadístico y el modelo de sustitución Nei-Gojobori para el análisis. Por debajo de la diagonal se indica el valor de los P-valores para cada comparación dos a dos de las secuencias ERV-Fc1. Por encima se indica el valor del estadístico ks-ka.

	1	2	3	4
1. Chimpanzee Gag ERV-Fc1		-0.662	-1.157	0.492
2. Pan paniscus Gag ERV-Fc1	1.000		-1.048	0.589
3. Gorilla Gag ERV-Fc1	1.000	1.000		-1.706
4. Human Gag ERV-Fc1	0.312	0.279	1.000	

3.7. Reconstrucción filogenética de la secuencia completa y de cada región

Las filogenias obtenidas son bastante congruentes entre ellas y con la historia evolutiva conocida de los primates (Fig. 16). Todas coinciden en que a partir de un ancestro común surgieron dos líneas evolutivas independientes, la del gorila por un lado, y la del chimpancé, bonobo y humano por otro.

En la reconstrucción filogenética de la secuencia completa (Fig. 13), se puede discriminar con un elevado apoyo *bootstrap* la divergencia del ERV-Fc1 de gorila en una línea independiente a la de los ERV-Fc1 de chimpancé, bonobo y humano. Sin embargo, en el clado formando por estas tres últimas, el apoyo resulta demasiado débil en ambos métodos como para que lo reflejado en el árbol sobre las secuencias del ERV-Fc1 de chimpancé, bonobo y humano sea una información significativamente fiable. Así pues, en base a esta secuencia, no va a ser posible discriminar cómo se separaron en la historia evolutiva estos tres primates una vez tomaron una vía evolutiva diferente a la del gorila, posiblemente porque los tiempos de divergencia de las tres especies están demasiado próximos evolutivamente hablando como para que en la secuencia de ERV-Fc1 se hayan acumulado las mutaciones suficientes que permitan discriminar su divergencia.

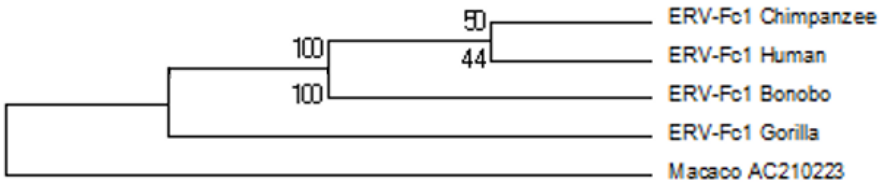


Fig. 13. Reconstrucción filogenética de las secuencia completas del grupo ERV-Fc1 de retrovirus endógenos. Por encima de las ramas, el valor en porcentaje del bootstrap obtenido con 100 replicaciones en la reconstrucción filogenética por el método de Máxima Parsimonia. Por debajo de las ramas, el valor en porcentaje del boostrap obtenido con 100 replicaciones en la reconstrucción filogenética por el método de Máxima Verosimilitud.

En cuanto a la reconstrucción filogenética de las regiones, estos resultados se repitieron para la *rt, rh, in* y *env,* de modo que solo se representa el árbol obtenido a partir de la región *env* como ejemplo (Fig. 14). Por otro lado, a partir de *gag* se obtuvo un árbol (Fig. 15) en el que, además de conseguir discriminar significativamente la diversificación independiente del ERV-Fc1 de gorila de los de chimpancé, bonobo y humano, es la única con la que se obtiene un valor de apoyo significativo mediante el método de Máxima Parsimonia (aunque bajo teniendo en cuenta un *cutoff* de 70%) para la divergencia de éstas tres últimas, ajustándose a la filogenia conocida de primates al agrupar al bonobo y chimpancé en un clado independiente a humano [33,34,39]. Para *pr* no se pudieron obtener árboles significativos.

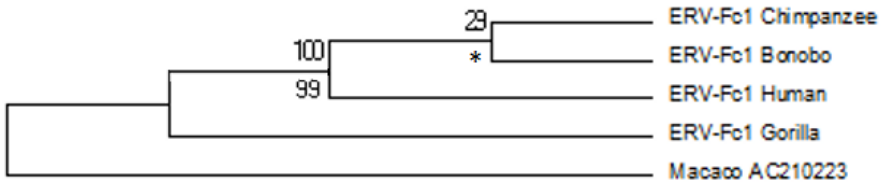


*Fig. 14. Reconstrucción filogenética de la región env del grupo ERV-Fc1 de retrovirus endógenos. Por encima de las ramas, el valor en porcentaje del bootstrap obtenido con 100 replicaciones en la reconstrucción filogenética por el método de Máxima Parsimonia. Por debajo de las ramas, el valor en porcentaje del boostrap obtenido con 100 replicaciones en la reconstrucción filogenética por el método de Máxima Verosimilitud. *,se creaba una politomía sin especificar el valor bootstrap.*

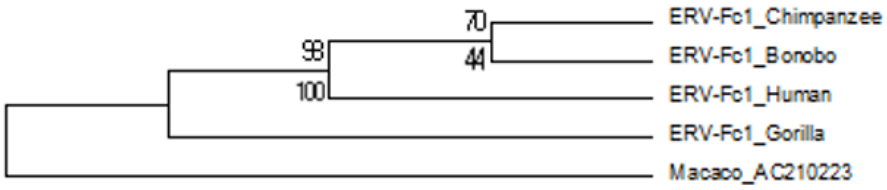


Fig. 15. Reconstrucción filogenética de la región gag del grupo ERV-Fc1 de retrovirus endógenos. Por encima de las ramas, el valor en porcentaje del bootstrap obtenido con 100 replicaciones en la reconstrucción filogenética por el método de Máxima Parsimonia. Por debajo de las ramas, el valor en porcentaje del boostrap obtenido con 100 replicaciones en la reconstrucción filogenética por el método de Máxima Verosimilitud.

4. Discusión

4.1. Un único y reciente evento de inserción en el grupo ERV-Fc1

A diferencia de lo que suele ocurrir con los diferentes grupos de retrovirus endógenos, el grupo ERV-Fc1 parece haber tenido un único evento de inserción en el genoma. Este grupo parece estar formado a partir de una sola copia original del retrovirus, que quedó reflejada en las cuatro secuencias encontradas: chimpancé, bonobo, humano y gorila. Además, se trata de secuencias completas y con idénticas TSD, lo que demuestra que proceden de una única inserción del retrovirus original que, a la vista de las secuencias encontradas, se tendría que haber producido en algún ancestro común a los cuatro primates hominoideos en los que se encontró el ERV-Fc1 (Fig. 16), a partir del cual comenzaría a acumular mutaciones e iría divergiendo en cada una de las líneas evolutivas en las que se encuentra actualmente. Es decir, el grupo ERV-Fc1 se encuentra constituido por cuatro secuencias provirales ortólogas que proceden de una única inserción proviral. que tuvo que tener lugar en el cromosoma X [14,15,36], ya que es donde se encontraron las secuencias en chimpancé, humano y gorila. En bonobo, aunque el *GenBank* no especificase el cromosoma en el que se encontró la secuencia, al tratarse de una única inserción proviral y encontrarse en el las otras tres especies en el mismo cromosoma, se puede concluir en ésta también.

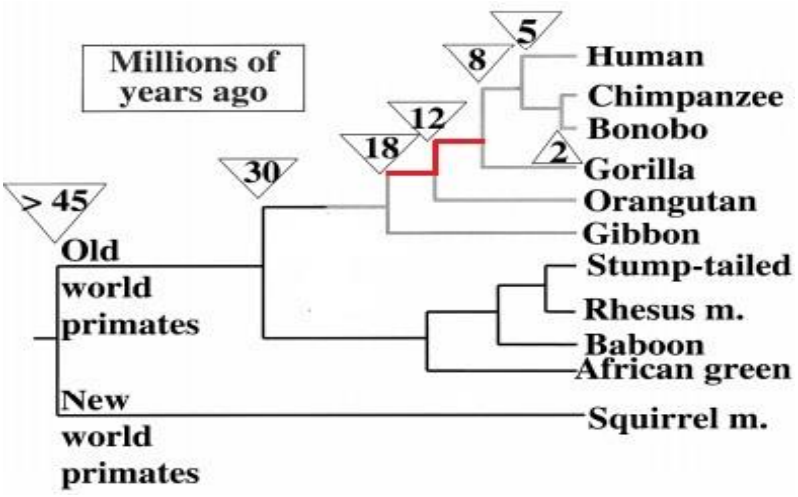


Fig. 16. Reconstrucción filogenética simplificada de los primates del Viejo Mundo en el que se detallan los tiempos de divergencia de cada línea en millones de años. En rojo se indica el intervalo de tiempo en el que parece haberse producido la inserción del ERV-Fc1 en el genoma de un ancestro común al chimpancé, bonobo, humano y gorila.

Por otro lado, se encontraron diversas evidencias que sostienen que se trata de una inserción relativamente reciente en comparación con muchos otros grupos de ERVs [37–39]. En primer lugar, se observó un alto grado de conservación de las secuencias y la ausencia de posibles copias y LTRs solitarias [3,12,20], resultado de procesos de recombinación y transposición secundaria en los genomas como suele ocurrir entre los ERVs [1,3–5]. Sin embargo, la evidencia más directa es el intervalo de edad calculado para cada secuencia proviral (Tabla 4), de los que podemos extraer un intervalo aproximado para la inserción de la secuencia original calculando el promedio, 12.79-20.67 Mya, que coincide con el intervalo de tiempo en el que tuvo lugar la diversificación de las líneas evolutivas del chimpancé, bonobo, humano y gorila (Fig. 16). Además, el solapamiento de los intervalos de edad calculados para cada secuencia, y que estos sean más similares entre chimpancé, bonobo y humano, siendo el de gorila algo superior, refleja que, efectivamente, se habría producido una única inserción en algún ancestro común a ellos, y que ésta evolucionaría posteriormente divergiendo, por un lado hacia el gorila, y por otro, hacía la vía evolutiva de chimpancé, bonobo y humano [12], tal y como se refleja en los resultados de la reconstrucción filogenética (Fig. 15). Sin embargo, en base a las filogenias obtenidas en este estudio a partir de las secuencias de ERV-Fc1, solo podemos concluir con base sólida que la primera secuencia en tomar una vía evolutiva distinta fue la del gorila, de ahí que sea el que posee mayor número de diferencias respecto a los otros tres. Esto indica que la separación del humano de la línea evolutiva de chimpancé y bonobo es muy reciente en términos evolutivos (Fig. 16), por lo que no se han acumulado suficientes mutaciones como para conseguir una correcta discriminación de su historia evolutiva.

4.2. Alto grado de conservación e indicios de selección en la evolución del grupo ERV-Fc1

Tal y como ya se vio, todas las secuencias provirales ERV-Fc1 encontradas consisten en secuencias completas con la estructura general [1,3,5,7,14]. Esto, junto con del estudio individual y en conjunto de las secuencias, así como la altísima similitud con la secuencia activa reconstruida, refleja gran conservación como consecuencia de una reciente inserción. Además, las reconstrucciones filogenéticas fueron sencillas, no encontrándose evidencias de transmisión horizontal entre las especies ni evolución en mosaico entre las regiones de las secuencias de ERV-Fc1. En cuanto a evidencias de procesos de recombinación no es posible

decir nada al respecto, ya que al tratarse de una única inserción en el genoma no sería posible que tuviesen lugar.

Aún con una elevada conservación, diversas *nonsense* y *frameshift mutations* evidencian cierta divergencia entre las secuencias. Aunque todas presentaron truncadas las regiones *pr-pol*, únicamente en gorila aparecían también interrumpidas *gag* y *env* (Fig. 17), lo que vuelve a suponer que el provirus original tomó dos líneas evolutivas diferentes tras su inserción, conservándose mejor en la que dio lugar finalmente a chimpancé, bonobo y humano. Así, es de suponer que *Gag* y *Env* podrían seguir expresándose en chimpancé, bonobo y humano, ya que presentan un ORF completo desde el codón de inicio (ATG) de la secuencia original hasta un codón stop, mientras que en gorila ambos van a estar truncados por mutaciones *nonsense* y *frameshift*. Esto coincide con lo referido en algunos estudios, que indican que se han encontrado evidencias de su expresión en tejidos sanos, acompañadas de un incremento de sus niveles en el caso de pacientes enfermos con esclerosis múltiple activa [13–16,22,36].

Por otro lado, que el cambio de pauta de lectura en *rh* y el codón stop entre ésta e *in* (Fig. 17) estén presentes en las cuatro secuencias provirales, procedentes de un único evento de inserción, y además no sea posible corregirlas en la secuencia consenso, podría suponer que se encontraban de forma natural en el retrovirus activo antes de producirse la inserción, o que se trate de mutaciones que se originaron posteriormente a la integración en un ancestro común a las cuatro especies, contribuyendo así a las formación del provirus defectivo. No se conoce ningún retrovirus que tenga de forma natural estas alteraciones y, como en el caso del codón stop entre *gag* y *pr*, esté programado para que el ribosoma los ignore, por lo que lo más seguro es que la segunda teoría planteada sea la correcta.

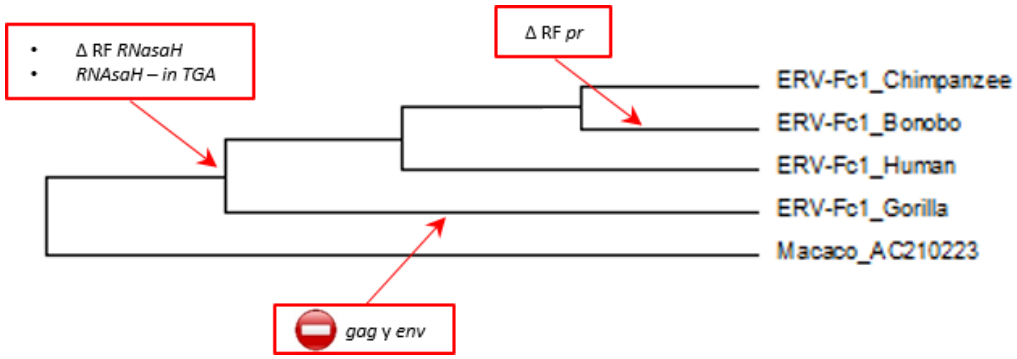


Fig. 17.Representación de las mutaciones más características del grupo ERV-Fc1 sobre su reconstrucción filogenética. Δ: cambio; RF: Reading Frame.

Además de por la relativa juventud del provirus en el genoma, en vista a los resultados obtenidos en el estudio de la selección, la elevada conservación podría deberse también a la posible función positiva que éste estaría desempeñando en el hospedador [20]. Tal y como se explicó en la introducción, son múltiples las funciones beneficiosas para el hospedador las que se le han atribuido a los retrovirus endógenos [3,6,7,20,22]. En el conjunto de secuencias del grupo ERV-Fc1, la región *env* presenta el mayor grado de conservación de todas las regiones retrovirales (Fig. 11) además de constituir un ORF intacto exceptuando solo al del genoma de gorila, en el que por lo tanto no sería posible que mantuviese alguna función. En base al test estadístico de selección se puede concluir que, de alguna forma, desempeña un papel importante en el chimpancé y está sometida a una selección purificadora, no pudiendo concluirlo con apoyo estadístico en humano y bonobo, pero sí observarse una tendencia a ello, posiblemente porque hasta no hace mucho seguía desempeñando tal/tales funciones

beneficiosas y de momento apenas ha degenerado, o porque todavía las sigue desempeñando pero esto no queda bien reflejado en el test, como consecuencia de una baja robustez estadística debido al bajo número de posiciones nucleotídicas implicadas. Las evidencias obtenidas en el estudio sobre alguna función de la región *env* del grupo ERV-Fc1 de retrovirus endógenos, se ven apoyadas por numerosas publicaciones, referidas a importantes implicaciones de la región *env* de otros grupos de ERVs en los procesos del organismo hospedador y su alto grado de conservación como consecuencia de ello (Tabla 1). Uno de los papeles más importantes de *env* a los que se hace referencia en múltiples estudios, es su intervención en el desarrollo embrionario gracias a sus capacidades fusogénicas e inmunomoduladoras [3]. La región *env* de muchos ERVs va a mantener un gen que codifica la sincitina, una proteína que hace posible la fusión de las membranas entre la envuelta viral y la membrana de las células del hospedador durante el proceso de infección, de modo que las células de un organismo que presente un ERV con esta región intacta, podrían ser capaces de fusionarse con otras gracias a la expresión de esta proteína [6]. Esto permitiría a las células del trofoblasto, en las que se ha detectado expresión de *Env* de otros grupos de retrovirus endógenos [40,41], fusionarse contribuyendo al desarrollo de la placenta [3,5,6,22]. Por otro lado, se cree que podría estar involucrada en la tolerancia del feto por parte del sistema inmunitario materno, impidiendo que sea detectado como un agente extraño. La poliproteína *Env* va a presentar en su dominio transmembrana una región inmunosupresora, de manera que su expresión en los tejidos placentarios llevaría a cabo un proceso de inmunosupresión que protege al feto [5,6,22,42]. Además de estar involucrada en el proceso embrionario, otra importante función beneficiosa es la resistencia frente a nuevas infecciones por retrovirus exógenos, debido a la interferencia de la *Env* procedente de los retrovirus endógenos y la de los retrovirus exógeno infectivo dada la competencia que se crea entre ambas por el receptor [5,6].

En cuanto a la función de *env* en el grupo de retrovirus endógenos ERV-Fc1, no se han encontrado referencias, pero su aparente función podría estar relacionada con alguno de los ejemplos aquí citados, lo que deja abierta una posible vía de investigación que permita completar el conocimiento sobre este grupo de retrovirus endógenos.

Por otro lado, a pesar de que algunos estudios hablan sobre funciones positivas vinculadas a la región *gag* [19], como es el caso del gen Fv1 del MuERV-L en ratón (Tabla 1), en el grupo de retrovirus endógenos ERV-Fc1 no parece desempeñar ninguna función.

5. Conclusiones

El grupo ERV-Fc1 de retrovirus endógenos humanos presenta unas características peculiares que se separan de lo observado en la mayor parte de los grupos de ERVs. Parece haberse originado a partir de una única copia del retrovirus original, por medio de un solo evento de inserción en algún ancestro común a gorila, chimpancé, bonobo y humano, que son los únicos genomas en los que se encontró la secuencia del retrovirus endógeno.

Todas estas secuencias van a estar completas y muy bien conservadas, a pesar de que en el gorila parece seguir un patrón evolutivo diferente a las de los otros tres primates al presentar truncadas la mayoría de las regiones, lo que indica que se estaría comportando como un pseudogen. En todas las secuencias de ERV-Fc1 se encontró una mutación de cambio en la pauta de lectura en la *rh* y un codón stop entre la *rh* e *in*. Además, se encontró un codón stop TGA entre *gag* y *pr*, encontrándose ambas regiones en el mismo RF, de modo que el retrovirus original pertenecería a esa minoría de retrovirus en los que el salto del codón stop para la

síntesis de *gag-pr-pol* se llevaría a cabo mediante una mala lectura del mismo y la colocación de un aminoácido en su lugar. Por otro lado, *gag* y *env* parecen estar intactos en chimpancé, bonobo y humano, por lo que se deduce que todavía se estarían expresando estos organismos.

En cuanto a la región *env* intacta de chimpancé, bonobo y humano, parece estar sometida a una presión purificadora (no pudiéndolo asegurar en base a los resultados del test para humano y bonobo) por estar desempeñando alguna función importante y beneficiosa en el organismo, lo que la mantiene con un alto grado de conservación.

Finalmente, el grupo ERv-Fc1 de retrovirus endógenos humanos no parece haber experimentado ningún proceso de transmisión horizontal entre especies ni evolución en mosaico entre sus regiones.

6. Bibliografía

1. Griffiths, D. J. Endogenous retroviruses in the human genome sequence. *Genome Biol.* **2,** 1–5 (2001).

2. Nelson, P. N. *et al.* Human endogenous retroviruses: transposable elements with potential? *Clin. Exp. Immunol.* **138,** 1–9 (2004).

3. Sentís, C. Retrovirus endógenos humanos: Significado biológico e implicaciones evolutivas. *Arbor* **172,** 135–166 (2010).

4. Tristem, M. Identification and Characterization of Novel Human Endogenous Retrovirus Families by Phylogenetic Screening of the Human Genome Mapping Project Database. *J. Virol.* **74,** 3715–3730 (2000).

5. Nelson, P. N. *et al.* Demystified... Human endogenous retroviruses. *J. Clin. Pathol. Mol. Pathol.* **56,** 11–18 (2003).

6. Liu, Y. & Soper, C. The Natural History of Retroviruses: Exogenization vs Endogenization. *Answers Res. J.* **2,** 97–106 (2009).

7. Coffin, J. M., Hughes, S. H. & Varmus, H. E. *Retroviruses*. 843 (Cold Sprin, 1997).

8. Lewin, B. *Genes IX*. 892 (McGraw-Hill, 2008).

9. Fischer, A. & Cavazzana-Calvo, M. Integration of retroviruses: a fine balance between efficiency and danger. *PLoS Med.* **2,** e10 (2005).

10. Müller, H. P. & Varmus, H. E. DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. *EMBO J.* **13,** 4704–4714 (1994).

11. Ambrosi, A., Cattoglio, C. & Di Serio, C. Retroviral integration process in the human genome: is it really non-random? A new statistical approach. *PLoS Comput. Biol.* **4,** e1000144 (2008).

12. Bénit, L., Calteau, A. & Heidmann, T. Characterization of the low-copy HERV-Fc family: evidence for recent integrations in primates of elements with coding envelope genes. *Virology* **312,** 159–168 (2003).

13. Laska, M. J. *et al.* Expression of HERV-Fc1, a human endogenous retrovirus, is increased in patients with active multiple sclerosis. *J. Virol.* **86,** 3713–22 (2012).

14. Nissen, K. K., Pedersen, F. S. & Nexø, B. A. Expression of Gag and Pol from reconstructed HERV-Fc1, associated with multiple sclerosis. *Retrovirology* **8,** P56 (2011).

15. Nissen, K. K. *et al.* Endogenous retroviruses and multiple sclerosis-new pieces to the puzzle. *BMC Neurol.* **13,** 111 (2013).

16. Laska, M. *et al.* Pathogenesis of multiple sclerosis: expression of HERV-Fc1: a human endogenous retrovirus. *Retrovirology* **8,** O23 (2011).

17. Nexø, B. A. *et al.* The etiology of multiple sclerosis: genetic evidence for the involvement of the human endogenous retrovirus HERV-Fc1. *PLoS One* **6,** e16652 (2011).

18. Nexø, B. A. *et al.* Involvement of the endogenous retroviral locus HERV-Fc1 on the human X-chromosome in multiple sclerosis. *Retrovirology* **8,** P54 (2011).

19. Feschotte, C. & Gilbert, C. Endogenous viruses: insights into viral evolution and impact on host biology. *Nat. Rev. Genet.* **13,** 283–96 (2012).

20. Magiorkinis, G., Belshaw, R. & Katzourakis, A. "There and back again": revisiting the pathophysiological roles of human endogenous retroviruses in the post-genomic era. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368,** 20120504 (2013).

21. Boyd, M. T., Bax, C. M., Bax, B. E., Bloxam, D. L. & Weiss, R. A. The human endogenous retrovirus ERV-3 is upregulated in differentiating placental trophoblast cells. *Virology* **196,** 905–9 (1993).

22. Kjeldbjerg, A. L., Villesen, P., Aagaard, L. & Pedersen, F. S. Gene conversion and purifying selection of a placenta-specific ERV-V envelope gene during simian evolution. *BMC Evol. Biol.* **8,** 11 (2008).

23. Kurth, R. & Bannert, N. Beneficial and detrimental effects of human endogenous retroviruses. *Int. J. Cancer* **126,** 306–14 (2010).

24. Sverdlov, E. *Retroviruses and primate genome evolution*. 250 (Landes Bio, 2005).

25. Bird, A. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* **8,** 1499–1504 (1980).

26. López-Sánchez, P., Costas, J. C. & Naveira, H. F. Paleogenomic record of the extinction of human endogenous retrovirus ERV9. *J. Virol.* **79,** 6997–7004 (2005).

27. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30,** 2725–9 (2013).

28. Costas, J. & Naveira, H. Evolutionary history of the human endogenous retrovirus family ERV9. *Mol. Biol. Evol.* **17,** 320–30 (2000).

29. Lee, Y. N. & Bieniasz, P. D. Reconstitution of an infectious human endogenous retrovirus. *PLoS Pathog.* **3,** e10 (2007).

30. Hurst, L. D. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet.* **18,** 486–487 (2002).

31. Pagon, R. A. *et al.* in *GeneReviews* (University of Washington). at <http://www.ncbi.nlm.nih.gov/books/NBK5191>

32. Nei, M. & Gojobori, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3,** 418–26 (1986).

33. Purvis, A. A composite estimate of primate phylogeny. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **348,** 405–21 (1995).

34. Perelman, P. *et al.* A molecular phylogeny of living primates. *PLoS Genet.* **7,** e1001342 (2011).

35. Nei, M. & Kumar, S. Molecular Evolution and Phylogenetics. *Heredity (Edinb).* **86,** 385–386 (2001).

36. Nissen, K., Laska, M., Hansen, B., Pedersen, F. & Nexø, B. No additional copies of HERV-Fc1 in the germ line of multiple sclerosis patients. *Virol. J.* **9,** 188 (2012).

37. Lee, A., Nolan, A., Watson, J. & Tristem, M. Identification of an ancient endogenous retrovirus, predating the divergence of the placental mammals. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368,** 20120503 (2013).

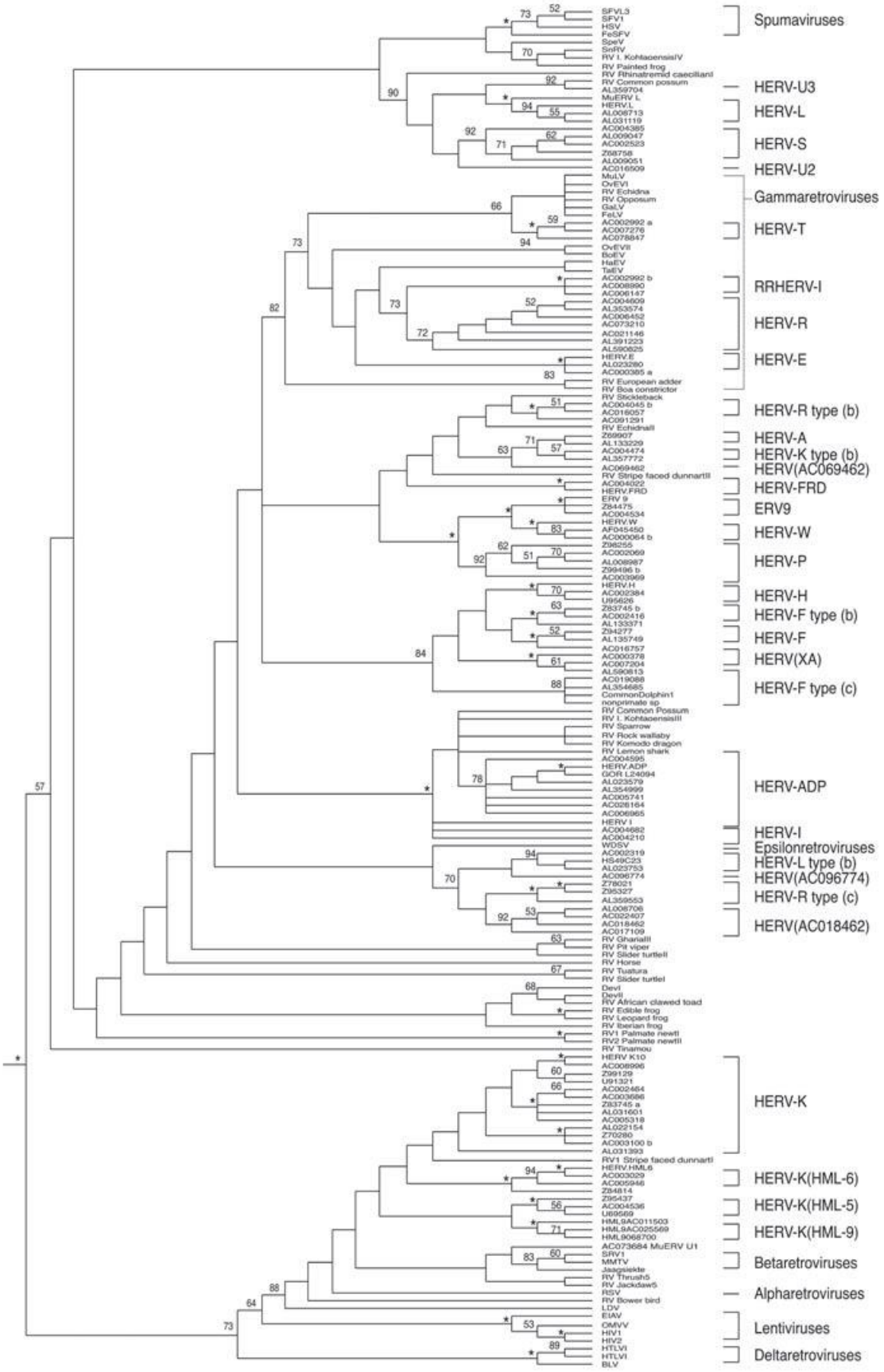
38. Kjellman, C., Sjögren, H. O. & Widegren, B. HERV-F, a new group of human endogenous retrovirus sequences. *J. Gen. Virol.* **80 (Pt 9),** 2383–92 (1999).

39. Johnson, W. E. & Coffin, J. M. Constructing primate phylogenies from ancient retrovirus sequences. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 10254–60 (1999).

40. Frendo, J.-L. *et al.* Direct involvement of HERV-W Env glycoprotein in human trophoblast cell fusion and differentiation. *Mol. Cell. Biol.* **23,** 3566–74 (2003).

41. Malassiné, A. *et al.* Human endogenous retrovirus-FRD envelope protein (syncytin 2) expression in normal and trisomy 21-affected placenta. *Retrovirology* **5,** 1–10 (2008).

42. Sverdlov, E. D. Retroviruses and primate evolution. *Bioessays* **22,** 161–71 (2000).



Anexo 1. Árbol filogenético de las familias de HERVs.

ERV-Fc1 Chimpanzee

RF +1	ORF1 (4285-4765)	*RNAsa2* (...4428-4644)
	ORF2 (4771-5904)	*in* (5002-5349)
RF+2	ORF* (5837-7591)	*env* (5837-7591)
RF +3	ORF1* (744-2156)	*gag* (744-2156)
	ORF3 (2157-4517)	*pr* (2370-2618)
		rt (2826-3455)
		RNAsa1 (4218-4493...)

Anexo 2. Coordenadas de los ORFs y las regiones encontrados en la secuencia del ERV-Fc1 de chimpancé.

ERV-Fc1 Gorilla

RF +1	ORF1 (4285-4767)	*RNAsa2* (...4429-4644)
	ORF2 (4771-5904)	*in* (5002-5349)
RF +2	ORF1* (740-1711)	*gag1* (740-1711...)
	ORF2* (5837-6373)	*env1* (5837-6373...)
	ORF3 (6377-6823)	*env2* (...6377-6823...)
RF +3	ORF1 (1599-1994)	*gag2* (...1599-1994...)
	ORF2 (1998-2153)	*gag3* (...1998-2156)
	ORF3 (2325-2777)	*pr* (2370-2618)
	ORF4 (2781-3491)	*rt* (2826-3455)
	ORF5 (3495-4517)	*RNAsa1* (4218-4430...)
	ORF6 (6627-7589)	*env3* (...6627-7592)

Anexo 1. Coordenadas de los ORFs y las regiones encontradas en la secuencia del ERV-Fc1 de gorila.

ERV-FC1 Human

RF +1	ORF1* (5782-7533)	*env* (5782-7533)
RF +2	ORF1* (686-2098)	*gag* (686-2098)
	ORF2 (2267-4462)	*pr* (2312-2563)
		rt (2771-3400)
		RNAsa1 (4163-4375...)
RF +3	ORF1 (4230-4712)	*RNAsa2* (...4374-4589)
	ORF2 (4716-5849)	*in* (4946-5294)

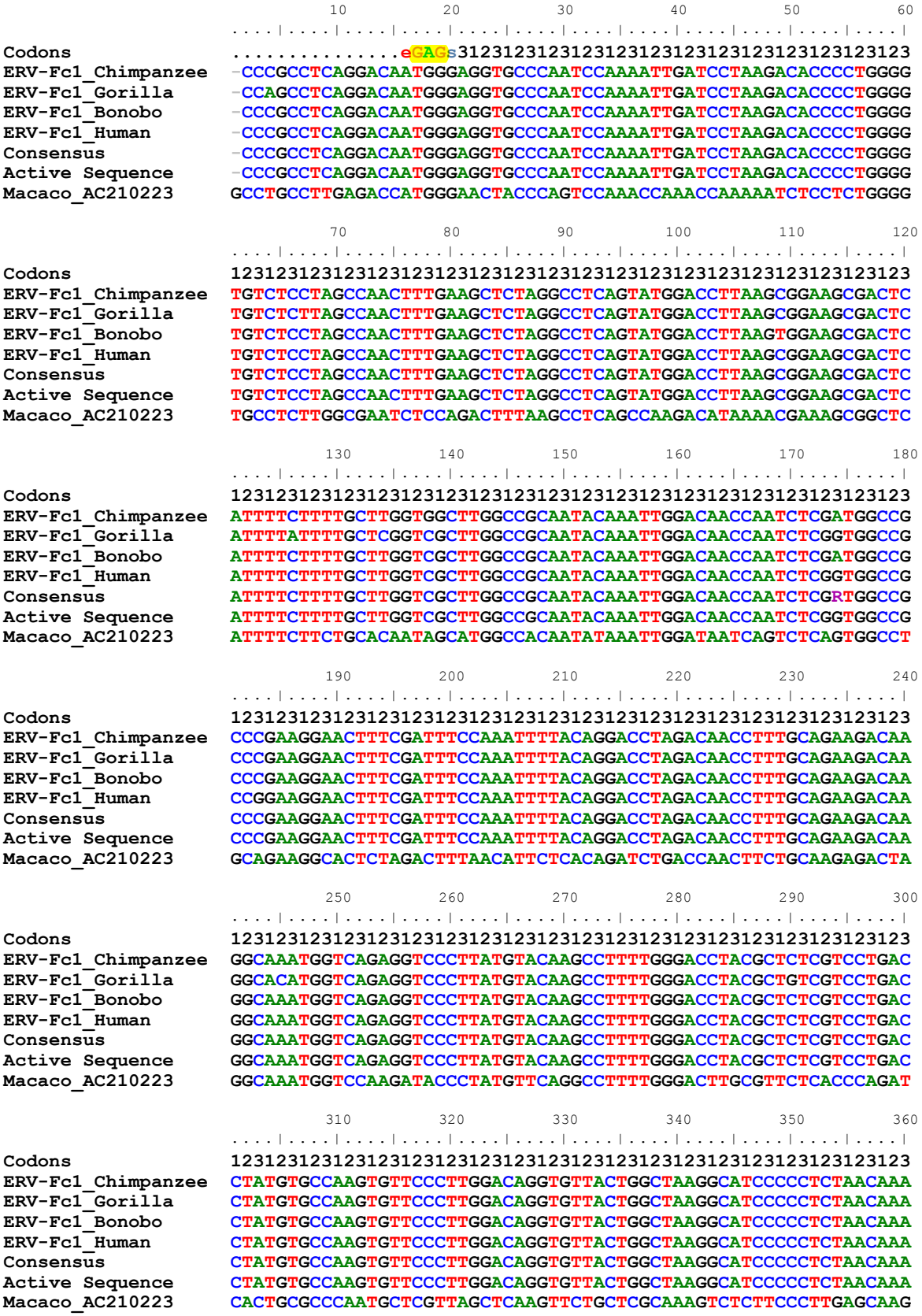
Anexo 4. Coordenadas de los ORFs y las regiones encontradas en la secuencia del ERV-Fc1 de humano.

ERV-Fc1 Bonobo

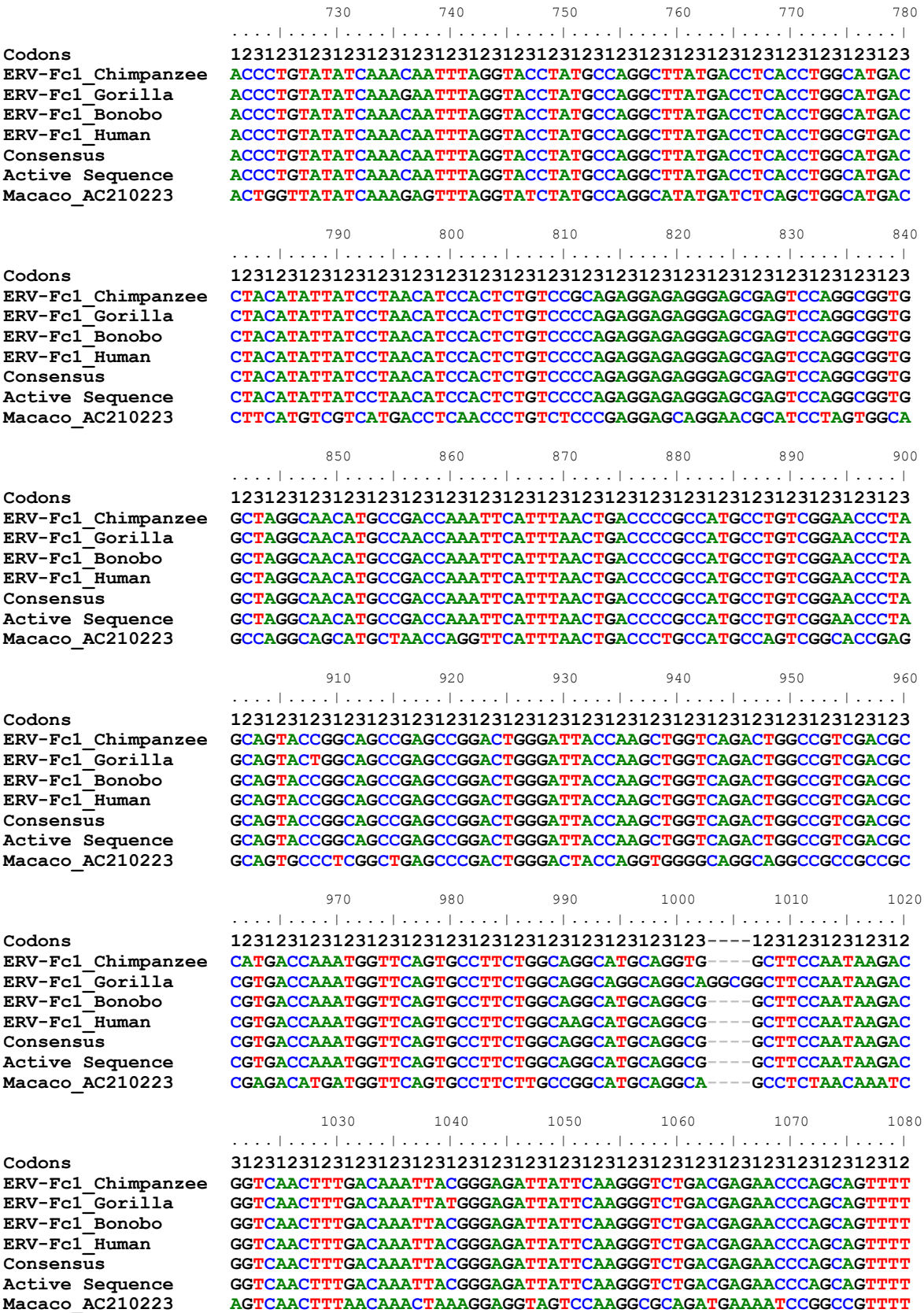
RF +1	ORF1 (2491-4515)	*pr2* (...2497-2616)
		rt (2824-3453)
		RNAsa1 (4216-4428...)
	ORF2* (5773-7587)	*env* (5836-7590)
RF +2	ORF1* (743-2152)	*gag* (743-2155)
	ORF2 (2156-2620)	*pr1* (2369-2491...)
	ORF3 (4283-4765)	*RNAsa2* (...4427-4642)
	ORF4 (4769-5903)	*In* (5000-5348)

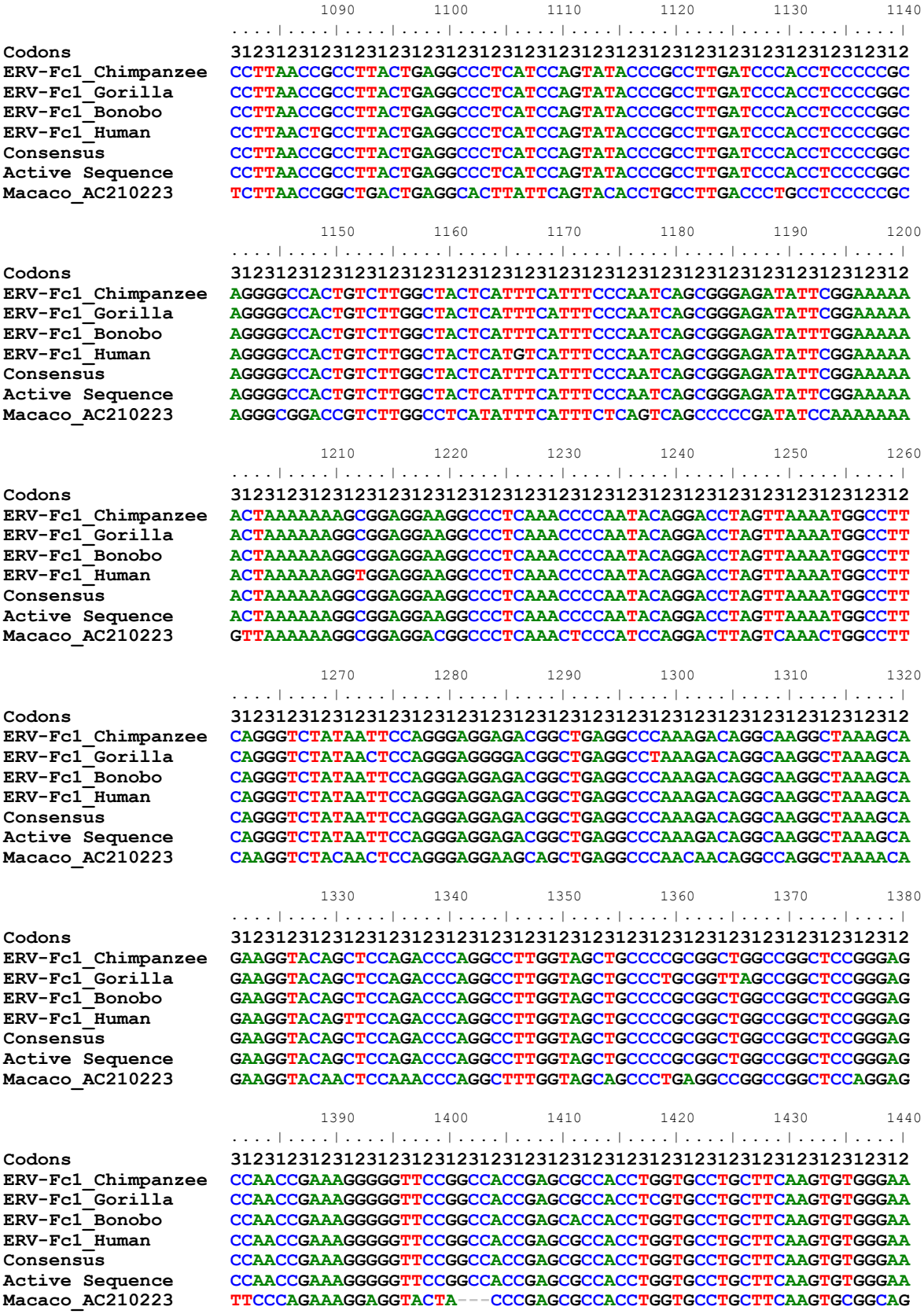
Anexo 3. Coordenadas de los ORFs y las regiones encontradas en la secuencia del ERV-Fc1 de bonobo.

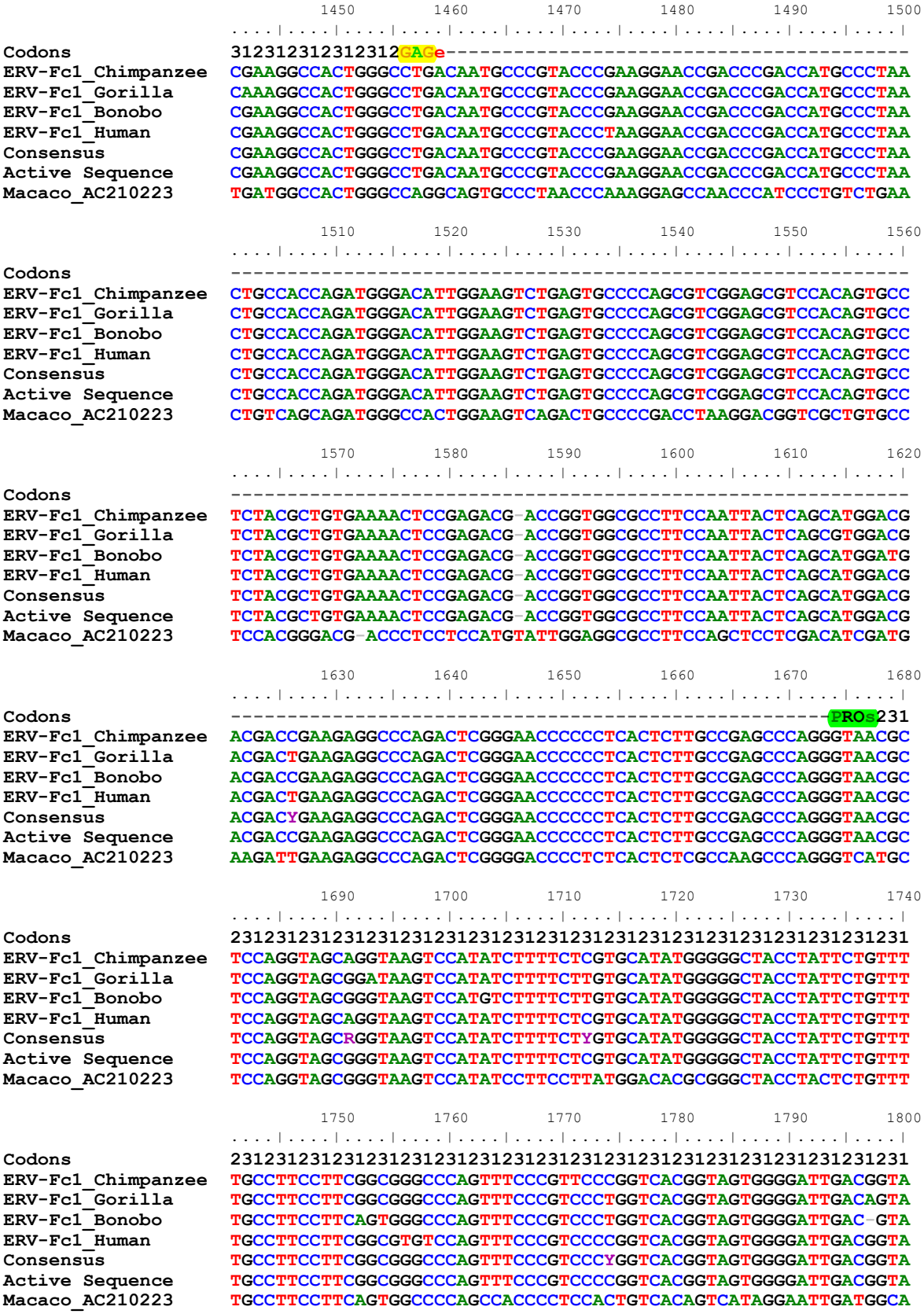
Anexo 6. Alineamiento de las secuencias encontradas del grupo ERV-Fc1, delimitando sus regiones e indicando los codones. Podemos encontrar también la secuencia consenso (Consensus) obtenida a partir de las cuatro secuencias de ERV-Fc1, la secuencia activa reconstruida (Active sequence) y la secuencia de macaco usada como outgroup (Macaco_AC210223).

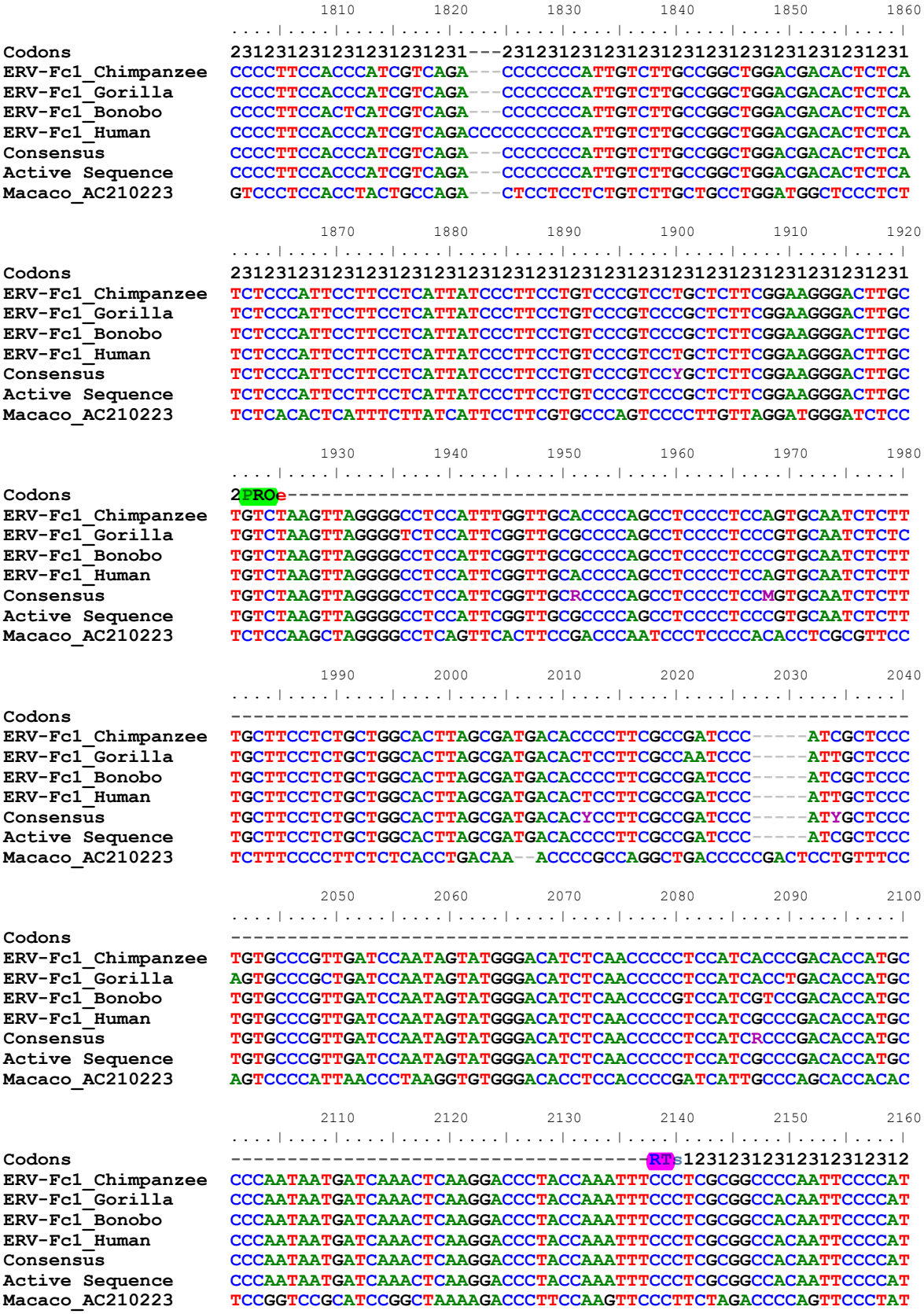


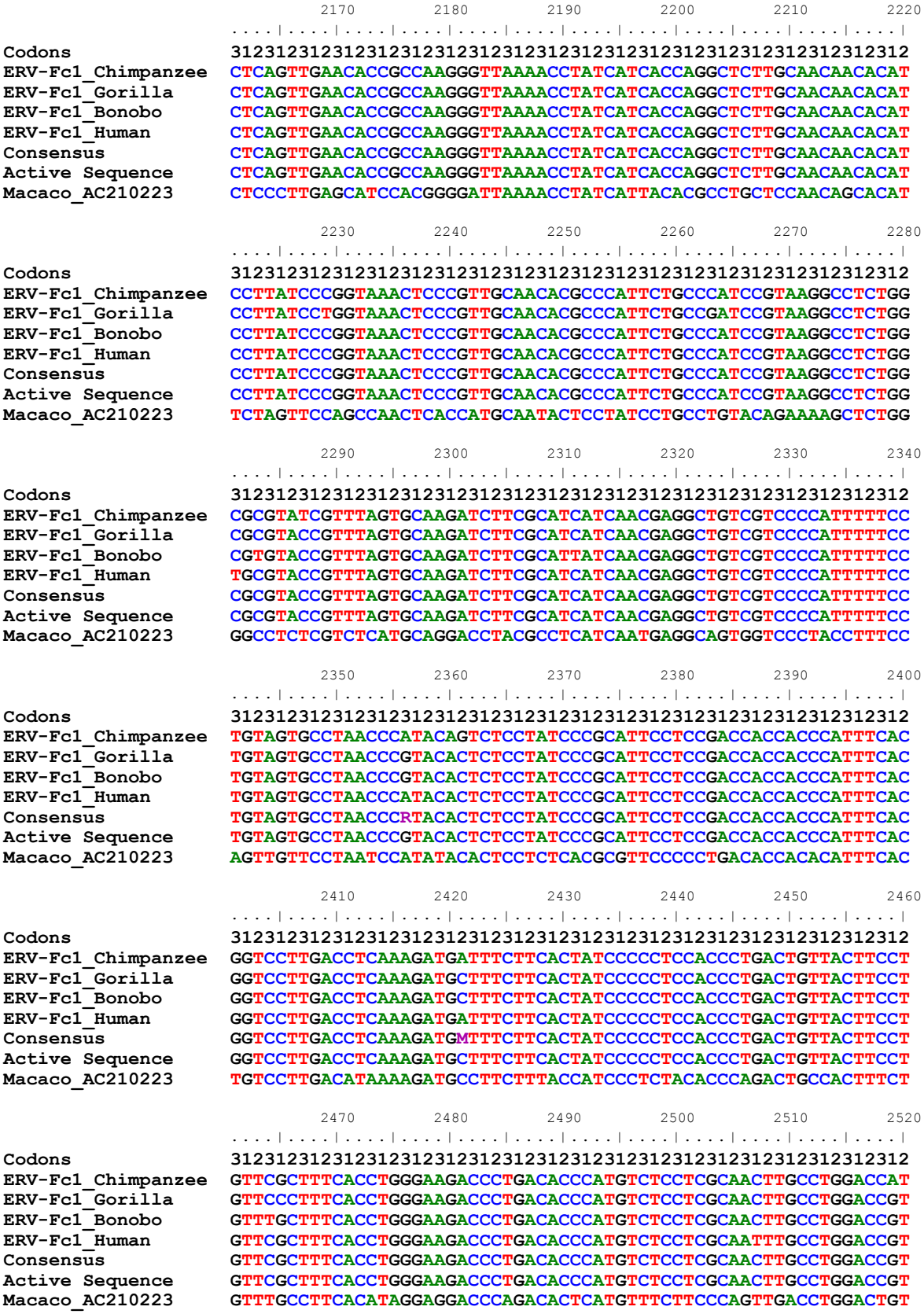
```
                                 370       380       390       400       410       420
                         ....|....|....|....|....|....|....|....|....|....|....|....|
Codons                   1231---2312312312312312312312312312312312312312312312312312-
ERV-Fc1_Chimpanzee       GAAC---CTGATTCCTCCCCTCTCTCCGAGCCTCCTGAAGCCCTTGCTTTACCACCATT-
ERV-Fc1_Gorilla          GAAC---CTGATTCCTCCCCTCTCTCCGAGCCTCCTGAAGCCCTCGCTTTACCACCATT-
ERV-Fc1_Bonobo           GAAC---CTGATTCCTCCCCTCTCTCCGAGCCTCCTGAAGCCCTTGCTTTACCACCATT-
ERV-Fc1_Human            GAAC---CTGATTCCTCCCCTCTCTCCGAGCCTCCTGAAGCCCTCGCTTTACCACCATT-
Consensus                GAAC---CTGATTCCTCCCCTCTCTCCGAGCCTCCTGAAGCCCTYGCTTTACCACCATT-
Active Sequence          GAAC---CTGATTCCTCCCCTCTCTCCGAGCCTCCTGAAGCCCTCGCTTTACCACCATT-
Macaco_AC210223          GAGAAGGATGATTCCTCCCCTTTTTCTGAGCCTCCTGACACCCTATCCCGGCCACCCCTC

                                 430       440       450       460       470       480
                         ....|....|....|....|....|....|....|....|....|....|....|....|
Codons                   --31231231231231231231231231231231--231231231231231----23123
ERV-Fc1_Chimpanzee       --GCCAGCGGCGCTCCCTCCTCCCTATCCAGGAT--CCTCTGGCCCCACCC----CAACG
ERV-Fc1_Gorilla          --GCCAGCGGCGCTCCCTCCTCCCTATCCAGGAT--CCTCTGGCCCCTCCC----CAACG
ERV-Fc1_Bonobo           --GCCAGCCGCGCTCCCTCCTCCCTATCCAGGAT--CCTCTGGCCCCACCC----CAACG
ERV-Fc1_Human            --GCCAGCGGCGCTCCCTCCTCCCTATCCAGGAT--CCTCTGGCCCCACCC----CAACG
Consensus                --GCCAGCGGCGCTCCCTCCTCCCTATCCAGGAT--CCTCTGGCCCCACCC----CAACG
Active Sequence          --GCCAGCGGCGCTCCCTCCTCCCTATCCAGGAT--CCTCTGGCCCCACCC----CAACG
Macaco_AC210223          CGGTCTCCTGCCCAACATCCTCCTTACCCTGACCCACCGTTATCCCCATCCTCGTCAACG

                                 490       500       510       520       530       540
                         ....|....|....|....|....|....|....|....|....|....|....|....|
Codons                   1231231231231231231231231231231231231231231231231231231----2
ERV-Fc1_Chimpanzee       GCTCCTCCGCTACCTCCTACACCACCTTCCTCTCCCGCTAACCCTCCCGCTTCTG----C
ERV-Fc1_Gorilla          GCTCCTCTGCTACCTCCTACATCACTTTCCTCTCCCGCTAACCCTCCCGCTTCTG----C
ERV-Fc1_Bonobo           GCTCCTCCGCTACCTCCTACACCACCTTCCTCTCCCGCTAACCCTCCCGCTTCTG----C
ERV-Fc1_Human            GCTCCTCCGCTACCTCCTACACCACCTTCCTCTCCCGCTAACCCTCCCGCTTCTG----C
Consensus                GCTCCTCCGCTACCTCCTACACCACCTTCCTCTCCCGCTAACCCTCCCGCTTCTG----C
Active Sequence          GCTCCTCCGCTACCTCCTACACCACCTTCCTCTCCCGCTAACCCTCCCGCTTCTG----C
Macaco_AC210223          CCTCC-CCACCCTTCCCCCCATCCTGTGCCCTTACCA--AACCCAACAGATTCTGTCTCC

                                 550       560       570       580       590       600
                         ....|....|....|....|....|....|....|....|....|....|....|....|
Codons                   3123123123123-----------123123123123123123123123123123123123
ERV-Fc1_Chimpanzee       TCTGCCACCGCCT-----------TCCCCTGTATCTGCGCACACTCGGTCGAAGACGGAC
ERV-Fc1_Gorilla          TCTGCCACCGCCT-----------TCCCCTGTATCTGCGCACACTCGGTCGAAGATGGAC
ERV-Fc1_Bonobo           TCTGCCACCGCCT-----------TCCCCTGTATCTGCGCACACTCGGTCGAAGACGGAC
ERV-Fc1_Human            TCTGCCACCGCCT-----------TCCCCTGTATCTGCGCACACTCGGTCGAAGACGGAC
Consensus                TCTGCCACCGCCT-----------TCCCCTGTATCTGCGCACACTCGGTCGAAGACGGAC
Active Sequence          TCTGCCACCGCCT-----------TCCCCTGTATCTGCGCACACTCGGTCGAAGACGGAC
Macaco_AC210223          CCTACCTCCACCTCCTCCCCTTCCTCGCCTGTCTCAGCCCATACTCAGTCCAGGACCGAC

                                 610       620       630       640       650       660
                         ....|....|....|....|....|....|....|....|....|....|....|....|
Codons                   123123123123123123123123123123123123123123123123123123123123
ERV-Fc1_Chimpanzee       CTCTTGTGTCCGCTCCATGAAGTTGCCGGTGCGGAAGGCGTGGTCAGGGTCCATGTTCCC
ERV-Fc1_Gorilla          CTCTTGTGTCCGCTCCGTGAAGTTGCCGGTGCGGAAGGCGTGGTCAGGGTCCATGTTCCC
ERV-Fc1_Bonobo           CTCTTGTGTACGCTCCGTGAAGTTGCCGGTGCGGAAGGCGTGGTCAGGGTCCATGTTCCC
ERV-Fc1_Human            CTCTTGTGTCTGCTCCGTGAAGTTGCCGGTGCGGAAGGCGTGGTCAGGGTCCATGTTCCC
Consensus                CTCTTGTGTCCGCTCCGTGAAGTTGCCGGTGCGGAAGGCGTGGTCAGGGTCCATGTTCCC
Active Sequence          CTCTTGTGTCCGCTCCGTGAAGTTGCCGGTGCGGAAGGCGTGGTCAGGGTCCATGTTCCC
Macaco_AC210223          C---TGTGTCCCTTACAAGAGGTAGCAGGCGCCAAAGGAGTAGTCCGGGTCCATGTGCCG

                                 670       680       690       700       710       720
                         ....|....|....|....|....|....|....|....|....|....|....|....|
Codons                   123123123123123123123123123123123123123123123123123123123123
ERV-Fc1_Chimpanzee       TTTTCTCTTACTGACTTATCTAAAATAGAGAAGCGGCCTGGGTCCTTCTCTGCCAACCCA
ERV-Fc1_Gorilla          TTTTCTCTTACTGACTTATCTAAAATAGAGAAGTGGCTTGGGTCCTTCTCTGCCAACCCA
ERV-Fc1_Bonobo           TTTTCTCTTACTGACTTATCTAAAATAGAGAAGCGGCCTGGGTCCTTCTCTGCCAACCCA
ERV-Fc1_Human            TTTTCTCTTACTGACTTATCTAAAATAGAGAAGCGGCCTGGGTCCTTCTCTGCCAACCCA
Consensus                TTTTCTCTTACTGACTTATCTAAAATAGAGAAGCGGCCTGGGTCCTTCTCTGCCAACCCA
Active Sequence          TTTTCTCTTACTGACTTATCTAAAATAGAGAAGCGGCCTGGGTCCTTCTCTGCCAACCCA
Macaco_AC210223          TTTTCATTGGCAGACTTGTCTAAGGTTGATGAGCGTGCAGGCTCCTTTTCGGCCAATCCT
```


730 740 750 760 770 780
Codons 123123123123123123123123123123123123123123123123123123
ERV-Fc1_Chimpanzee ACCCTGTATATCAAACAATTTAGGTACCTATGCCAGGCTTATGACCTCACCTGGCATGAC
ERV-Fc1_Gorilla ACCCTGTATATCAAAGAATTTAGGTACCTATGCCAGGCTTATGACCTCACCTGGCATGAC
ERV-Fc1_Bonobo ACCCTGTATATCAAACAATTTAGGTACCTATGCCAGGCTTATGACCTCACCTGGCATGAC
ERV-Fc1_Human ACCCTGTATATCAAACAATTTAGGTACCTATGCCAGGCTTATGACCTCACCTGGCGTGAC
Consensus ACCCTGTATATCAAACAATTTAGGTACCTATGCCAGGCTTATGACCTCACCTGGCATGAC
Active Sequence ACCCTGTATATCAAACAATTTAGGTACCTATGCCAGGCTTATGACCTCACCTGGCATGAC
Macaco_AC210223 ACTGGTTATATCAAAGAGTTTAGGTATCTATGCCAGGCATATGATCTCAGCTGGCATGAC
790 800 810 820 830 840
Codons 123123123123123123123123123123123123123123123123123123
ERV-Fc1_Chimpanzee CTACATATTATCCTAACATCCACTCTGTCCGCAGAGGAGAGGGAGCGAGTCCAGGCGGTG
ERV-Fc1_Gorilla CTACATATTATCCTAACATCCACTCTGTCCCCAGAGGAGAGGGAGCGAGTCCAGGCGGTG
ERV-Fc1_Bonobo CTACATATTATCCTAACATCCACTCTGTCCCCAGAGGAGAGGGAGCGAGTCCAGGCGGTG
ERV-Fc1_Human CTACATATTATCCTAACATCCACTCTGTCCCCAGAGGAGAGGGAGCGAGTCCAGGCGGTG
Consensus CTACATATTATCCTAACATCCACTCTGTCCCCAGAGGAGAGGGAGCGAGTCCAGGCGGTG
Active Sequence CTACATATTATCCTAACATCCACTCTGTCCCCAGAGGAGAGGGAGCGAGTCCAGGCGGTG
Macaco_AC210223 CTTCATGTCGTCATGACCTCAACCCTGTCTCCCGAGGAGCAGGAACGCATCCTAGTGGCA
850 860 870 880 890 900
Codons 123123123123123123123123123123123123123123123123123123
ERV-Fc1_Chimpanzee GCTAGGCAACATGCCGACCAAATTCATTTAACTGACCCCGCCATGCCTGTCGGAACCCTA
ERV-Fc1_Gorilla GCTAGGCAACATGCCAACCAAATTCATTTAACTGACCCCGCCATGCCTGTCGGAACCCTA
ERV-Fc1_Bonobo GCTAGGCAACATGCCGACCAAATTCATTTAACTGACCCCGCCATGCCTGTCGGAACCCTA
ERV-Fc1_Human GCTAGGCAACATGCCGACCAAATTCATTTAACTGACCCCGCCATGCCTGTCGGAACCCTA
Consensus GCTAGGCAACATGCCGACCAAATTCATTTAACTGACCCCGCCATGCCTGTCGGAACCCTA
Active Sequence GCTAGGCAACATGCCGACCAAATTCATTTAACTGACCCCGCCATGCCTGTCGGAACCCTA
Macaco_AC210223 GCCAGGCAGCATGCTAACCAGGTTCATTTAACTGACCCTGCCATGCCAGTCGGCACCGAG
910 920 930 940 950 960
Codons 123123123123123123123123123123123123123123123123123123
ERV-Fc1_Chimpanzee GCAGTACCGGCAGCCGAGCCGGACTGGGATTACCAAGCTGGTCAGACTGGCCGTCGACGC
ERV-Fc1_Gorilla GCAGTACTGGCAGCCGAGCCGGACTGGGATTACCAAGCTGGTCAGACTGGCCGTCGACGC
ERV-Fc1_Bonobo GCAGTACCGGCAGCCGAGCCGGACTGGGATTACCAAGCTGGTCAGACTGGCCGTCGACGC
ERV-Fc1_Human GCAGTACCGGCAGCCGAGCCGGACTGGGATTACCAAGCTGGTCAGACTGGCCGTCGACGC
Consensus GCAGTACCGGCAGCCGAGCCGGACTGGGATTACCAAGCTGGTCAGACTGGCCGTCGACGC
Active Sequence GCAGTACCGGCAGCCGAGCCGGACTGGGATTACCAAGCTGGTCAGACTGGCCGTCGACGC
Macaco_AC210223 GCAGTGCCCTCGGCTGAGCCCGACTGGGACTACCAGGTGGGGCAGGCAGGCCGCCGCCGC
970 980 990 1000 1010 1020
Codons 123123123123123123123123123123123123----12312312312312
ERV-Fc1_Chimpanzee CATGACCAAATGGTTCAGTGCCTTCTGGCAGGCATGCAGGTG----GCTTCCAATAAGAC
ERV-Fc1_Gorilla CGTGACCAAATGGTTCAGTGCCTTCTGGCAGGCAGGCAGGCAGGCGGCTTCCAATAAGAC
ERV-Fc1_Bonobo CGTGACCAAATGGTTCAGTGCCTTCTGGCAGGCATGCAGGCG----GCTTCCAATAAGAC
ERV-Fc1_Human CGTGACCAAATGGTTCAGTGCCTTCTGGCAAGCATGCAGGCG----GCTTCCAATAAGAC
Consensus CGTGACCAAATGGTTCAGTGCCTTCTGGCAGGCATGCAGGCG----GCTTCCAATAAGAC
Active Sequence CGTGACCAAATGGTTCAGTGCCTTCTGGCAGGCATGCAGGCG----GCTTCCAATAAGAC
Macaco_AC210223 CGAGACATGATGGTTCAGTGCCTTCTTGCCGGCATGCAGGCA----GCCTCTAACAAATC
1030 1040 1050 1060 1070 1080
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee GGTCAACTTTGACAAATTACGGGAGATTATTCAAGGGTCTGACGAGAACCCAGCAGTTTT
ERV-Fc1_Gorilla GGTCAACTTTGACAAATTATGGGAGATTATTCAAGGGTCTGACGAGAACCCAGCAGTTTT
ERV-Fc1_Bonobo GGTCAACTTTGACAAATTACGGGAGATTATTCAAGGGTCTGACGAGAACCCAGCAGTTTT
ERV-Fc1_Human GGTCAACTTTGACAAATTACGGGAGATTATTCAAGGGTCTGACGAGAACCCAGCAGTTTT
Consensus GGTCAACTTTGACAAATTACGGGAGATTATTCAAGGGTCTGACGAGAACCCAGCAGTTTT
Active Sequence GGTCAACTTTGACAAATTACGGGAGATTATTCAAGGGTCTGACGAGAACCCAGCAGTTTT
Macaco_AC210223 AGTCAACTTTAACAAACTAAAGGAGGTAGTCCAAGGCGCAGATGAAAATCCGGCCGTTTT

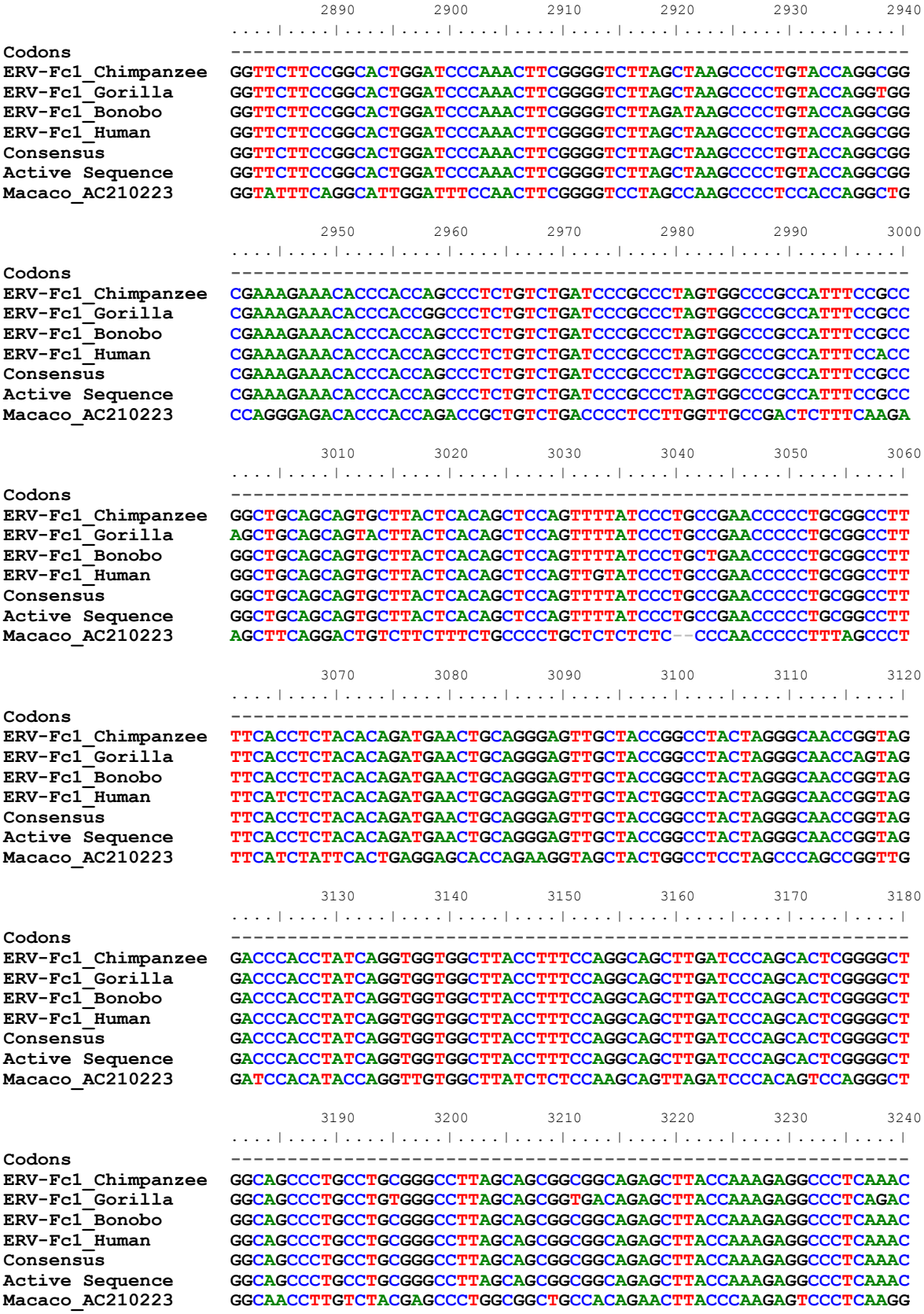

1090 1100 1110 1120 1130 1140
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee CCTTAACCGCCTTACTGAGGCCCTCATCCAGTATACCCGCCTTGATCCCACCTCCCCCGC
ERV-Fc1_Gorilla CCTTAACCGCCTTACTGAGGCCCTCATCCAGTATACCCGCCTTGATCCCACCTCCCCGGC
ERV-Fc1_Bonobo CCTTAACCGCCTTACTGAGGCCCTCATCCAGTATACCCGCCTTGATCCCACCTCCCCGGC
ERV-Fc1_Human CCTTAACTGCCTTACTGAGGCCCTCATCCAGTATACCCGCCTTGATCCCACCTCCCCGGC
Consensus CCTTAACCGCCTTACTGAGGCCCTCATCCAGTATACCCGCCTTGATCCCACCTCCCCGGC
Active Sequence CCTTAACCGCCTTACTGAGGCCCTCATCCAGTATACCCGCCTTGATCCCACCTCCCCGGC
Macaco_AC210223 TCTTAACCGGCTGACTGAGGCACTTATTCAGTACACCTGCCTTGACCCTGCCTCCCCCGC
1150 1160 1170 1180 1190 1200
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee AGGGGCCACTGTCTTGGCTACTCATTTCATTTCCCAATCAGCGGGAGATATTCGGAAAAA
ERV-Fc1_Gorilla AGGGGCCACTGTCTTGGCTACTCATTTCATTTCCCAATCAGCGGGAGATATTCGGAAAAA
ERV-Fc1_Bonobo AGGGGCCACTGTCTTGGCTACTCATTTCATTTCCCAATCAGCGGGAGATATTTGGAAAAA
ERV-Fc1_Human AGGGGCCACTGTCTTGGCTACTCATGTCATTTCCCAATCAGCGGGAGATATTCGGAAAAA
Consensus AGGGGCCACTGTCTTGGCTACTCATTTCATTTCCCAATCAGCGGGAGATATTCGGAAAAA
Active Sequence AGGGGCCACTGTCTTGGCTACTCATTTCATTTCCCAATCAGCGGGAGATATTCGGAAAAA
Macaco_AC210223 AGGGCGGACCGTCTTGGCCTCATATTTCATTTCTCAGTCAGCCCCCGATATCCAAAAAAA
1210 1220 1230 1240 1250 1260
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee ACTAAAAAAAGCGGAGGAAGGCCCTCAAACCCCAATACAGGACCTAGTTAAAATGGCCTT
ERV-Fc1_Gorilla ACTAAAAAAAGGCGGAGGAAGGCCCTCAAACCCCAATACAGGACCTAGTTAAAATGGCCTT
ERV-Fc1_Bonobo ACTAAAAAAGGCGGAGGAAGGCCCTCAAACCCCAATACAGGACCTAGTTAAAATGGCCTT
ERV-Fc1_Human ACTAAAAAAGGTGGAGGAAGGCCCTCAAACCCCAATACAGGACCTAGTTAAAATGGCCTT
Consensus ACTAAAAAAGGCGGAGGAAGGCCCTCAAACCCCAATACAGGACCTAGTTAAAATGGCCTT
Active Sequence ACTAAAAAAGGCGGAGGAAGGCCCTCAAACCCCAATACAGGACCTAGTTAAAATGGCCTT
Macaco_AC210223 GTTAAAAAAGGCGGAGGACGGCCCTCAAACTCCCATCCAGGACTTAGTCAAACTGGCCTT
1270 1280 1290 1300 1310 1320
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee CAGGGTCTATAATTCCAGGGAGGAGACGGCTGAGGCCCAAAGACAGGCAAGGCTAAAGCA
ERV-Fc1_Gorilla CAGGGTCTATAACTCCAGGGAGGGGACGGCTGAGGCCTAAAGACAGGCAAGGCTAAAGCA
ERV-Fc1_Bonobo CAGGGTCTATAATTCCAGGGAGGAGACGGCTGAGGCCCAAAGACAGGCAAGGCTAAAGCA
ERV-Fc1_Human CAGGGTCTATAATTCCAGGGAGGAGACGGCTGAGGCCCAAAGACAGGCAAGGCTAAAGCA
Consensus CAGGGTCTATAATTCCAGGGAGGAGACGGCTGAGGCCCAAAGACAGGCAAGGCTAAAGCA
Active Sequence CAGGGTCTATAATTCCAGGGAGGAGACGGCTGAGGCCCAAAGACAGGCAAGGCTAAAGCA
Macaco_AC210223 CAAGGTCTACAACTCCAGGGAGGAAGCAGCTGAGGCCCAACAACAGGCCAGGCTAAAACA
1330 1340 1350 1360 1370 1380
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee GAAGGTACAGCTCCAGACCCAGGCCTTGGTAGCTGCCCCGCGGCTGGCCGGCTCCGGGAG
ERV-Fc1_Gorilla GAAGGTACAGCTCCAGACCCAGGCCTTGGTAGCTGCCCTGCGGTTAGCCGGCTCCGGGAG
ERV-Fc1_Bonobo GAAGGTACAGCTCCAGACCCAGGCCTTGGTAGCTGCCCCGCGGCTGGCCGGCTCCGGGAG
ERV-Fc1_Human GAAGGTACAGTTCCAGACCCAGGCCTTGGTAGCTGCCCCGCGGCTGGCCGGCTCCGGGAG
Consensus GAAGGTACAGCTCCAGACCCAGGCCTTGGTAGCTGCCCCGCGGCTGGCCGGCTCCGGGAG
Active Sequence GAAGGTACAGCTCCAGACCCAGGCCTTGGTAGCTGCCCCGCGGCTGGCCGGCTCCGGGAG
Macaco_AC210223 GAAGGTACAACTCCAAACCCAGGCTTTGGTAGCAGCCCTGAGGCCGGCCGGCTCCAGGAG
1390 1400 1410 1420 1430 1440
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee CCAACCGAAAGGGGGTTCCGGCCACCGAGCGCCACCTGGTGCCTGCTTCAAGTGTGGGAA
ERV-Fc1_Gorilla CCAACCGAAAGGGGGTTCCGGCCACCGAGCGCCACCTCGTGCCTGCTTCAAGTGTGGGAA
ERV-Fc1_Bonobo CCAACCGAAAGGGGGTTCCGGCCACCGAGCACCACCTGGTGCCTGCTTCAAGTGTGGGAA
ERV-Fc1_Human CCAACCGAAAGGGGGTTCCGGCCACCGAGCGCCACCTGGTGCCTGCTTCAAGTGTGGGAA
Consensus CCAACCGAAAGGGGGTTCCGGCCACCGAGCGCCACCTGGTGCCTGCTTCAAGTGTGGGAA
Active Sequence CCAACCGAAAGGGGGTTCCGGCCACCGAGCGCCACCTGGTGCCTGCTTCAAGTGTGGGAA
Macaco_AC210223 TTCCCAGAAAGGAGGTACTA---CCCGAGCGCCACCTGGTGCCTGCTTCAAGTGCGGCAG

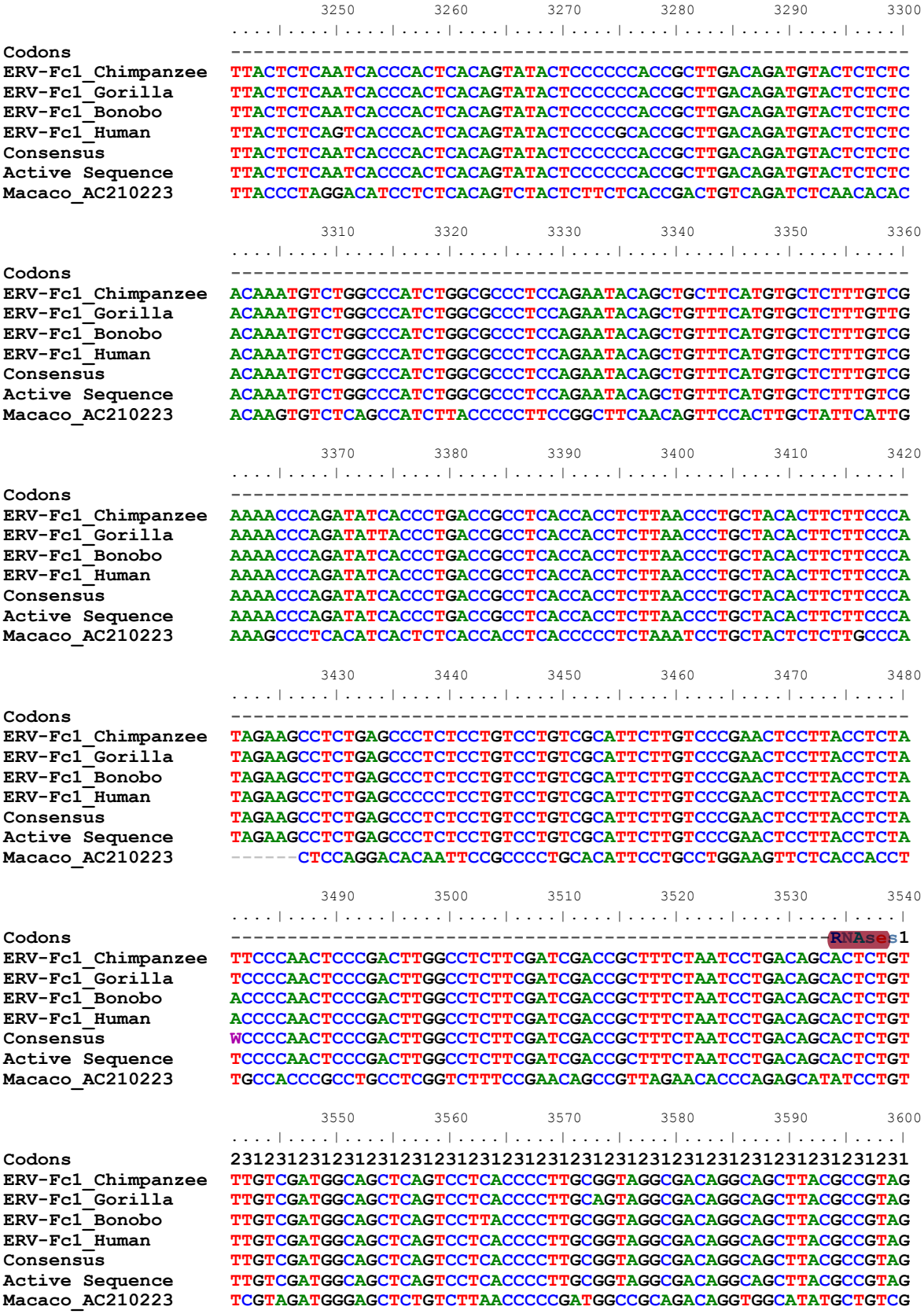

1450 1460 1470 1480 1490 1500
Codons 312312312312312GAGe----------
ERV-Fc1_Chimpanzee CGAAGGCCACTGGGCCTGACAATGCCCGTACCCGAAGGAACCGACCCGACCATGCCCTAA
ERV-Fc1_Gorilla CAAAGGCCACTGGGCCTGACAATGCCCGTACCCGAAGGAACCGACCCGACCATGCCCTAA
ERV-Fc1_Bonobo CGAAGGCCACTGGGCCTGACAATGCCCGTACCCGAAGGAACCGACCCGACCATGCCCTAA
ERV-Fc1_Human CGAAGGCCACTGGGCCTGACAATGCCCGTACCCTAAGGAACCGACCCGACCATGCCCTAA
Consensus CGAAGGCCACTGGGCCTGACAATGCCCGTACCCGAAGGAACCGACCCGACCATGCCCTAA
Active Sequence CGAAGGCCACTGGGCCTGACAATGCCCGTACCCGAAGGAACCGACCCGACCATGCCCTAA
Macaco_AC210223 TGATGGCCACTGGGCCAGGCAGTGCCCTAACCCAAAGGAGCCAACCCATCCCTGTCTGAA
1510 1520 1530 1540 1550 1560
Codons ----------
ERV-Fc1_Chimpanzee CTGCCACCAGATGGGACATTGGAAGTCTGAGTGCCCCAGCGTCGGAGCGTCCACAGTGCC
ERV-Fc1_Gorilla CTGCCACCAGATGGGACATTGGAAGTCTGAGTGCCCCAGCGTCGGAGCGTCCACAGTGCC
ERV-Fc1_Bonobo CTGCCACCAGATGGGACATTGGAAGTCTGAGTGCCCCAGCGTCGGAGCGTCCACAGTGCC
ERV-Fc1_Human CTGCCACCAGATGGGACATTGGAAGTCTGAGTGCCCCAGCGTCGGAGCGTCCACAGTGCC
Consensus CTGCCACCAGATGGGACATTGGAAGTCTGAGTGCCCCAGCGTCGGAGCGTCCACAGTGCC
Active Sequence CTGCCACCAGATGGGACATTGGAAGTCTGAGTGCCCCAGCGTCGGAGCGTCCACAGTGCC
Macaco_AC210223 CTGTCAGCAGATGGGCCACTGGAAGTCAGACTGCCCCGACCTAAGGACGGTCGCTGTGCC
1570 1580 1590 1600 1610 1620
Codons ----------
ERV-Fc1_Chimpanzee TCTACGCTGTGAAAACTCCGAGACG-ACCGGTGGCGCCTTCCAATTACTCAGCATGGACG
ERV-Fc1_Gorilla TCTACGCTGTGAAAACTCCGAGACG-ACCGGTGGCGCCTTCCAATTACTCAGCGTGGACG
ERV-Fc1_Bonobo TCTACGCTGTGAAAACTCCGAGACG-ACCGGTGGCGCCTTCCAATTACTCAGCATGGATG
ERV-Fc1_Human TCTACGCTGTGAAAACTCCGAGACG-ACCGGTGGCGCCTTCCAATTACTCAGCATGGACG
Consensus TCTACGCTGTGAAAACTCCGAGACG-ACCGGTGGCGCCTTCCAATTACTCAGCATGGACG
Active Sequence TCTACGCTGTGAAAACTCCGAGACG-ACCGGTGGCGCCTTCCAATTACTCAGCATGGACG
Macaco_AC210223 TCCACGGGACG-ACCCTCCTCCATGTATTGGAGGCGCCTTCCAGCTCCTCGACATCGATG
1630 1640 1650 1660 1670 1680
Codons ----------PROs231
ERV-Fc1_Chimpanzee ACGACCGAAGAGGCCCAGACTCGGGAACCCCCCTCACTCTTGCCGAGCCCAGGGTAACGC
ERV-Fc1_Gorilla ACGACTGAAGAGGCCCAGACTCGGGAACCCCCCTCACTCTTGCCGAGCCCAGGGTAACGC
ERV-Fc1_Bonobo ACGACCGAAGAGGCCCAGACTCGGGAACCCCCCTCACTCTTGCCGAGCCCAGGGTAACGC
ERV-Fc1_Human ACGACTGAAGAGGCCCAGACTCGGGAACCCCCCTCACTCTTGCCGAGCCCAGGGTAACGC
Consensus ACGACYGAAGAGGCCCAGACTCGGGAACCCCCCTCACTCTTGCCGAGCCCAGGGTAACGC
Active Sequence ACGACCGAAGAGGCCCAGACTCGGGAACCCCCCTCACTCTTGCCGAGCCCAGGGTAACGC
Macaco_AC210223 AAGATTGAAGAGGCCCAGACTCGGGGACCCCTCTCACTCTCGCCAAGCCCAGGGTCATGC
1690 1700 1710 1720 1730 1740
Codons 231231231231231231231231231231231231231231231231231231231231
ERV-Fc1_Chimpanzee TCCAGGTAGCAGGTAAGTCCATATCTTTTCTCGTGCATATGGGGGCTACCTATTCTGTTT
ERV-Fc1_Gorilla TCCAGGTAGCGGATAAGTCCATATCTTTTCTTGTGCATATGGGGGCTACCTATTCTGTTT
ERV-Fc1_Bonobo TCCAGGTAGCGGGTAAGTCCATGTCTTTTCTTGTGCATATGGGGGCTACCTATTCTGTTT
ERV-Fc1_Human TCCAGGTAGCAGGTAAGTCCATATCTTTTCTCGTGCATATGGGGGCTACCTATTCTGTTT
Consensus TCCAGGTAGCRGGTAAGTCCATATCTTTTCTYGTGCATATGGGGGCTACCTATTCTGTTT
Active Sequence TCCAGGTAGCGGGTAAGTCCATATCTTTTCTCGTGCATATGGGGGCTACCTATTCTGTTT
Macaco_AC210223 TCCAGGTAGCGGGTAAGTCCATATCCTTCCTTATGGACACGCGGGCTACCTACTCTGTTT
1750 1760 1770 1780 1790 1800
Codons 231231231231231231231231231231231231231231231231231231231231
ERV-Fc1_Chimpanzee TGCCTTCCTTCGGCGGGCCCAGTTTCCCGTTCCCGGTCACGGTAGTGGGGATTGACGGTA
ERV-Fc1_Gorilla TGCCTTCCTTCGGCGGGCCCAGTTTCCCGTCCCTGGTCACGGTAGTGGGGATTGACAGTA
ERV-Fc1_Bonobo TGCCTTCCTTCAGTGGGCCCAGTTTCCCGTCCCTGGTCACGGTAGTGGGGATTGAC-GTA
ERV-Fc1_Human TGCCTTCCTTCGGCGTGTCCAGTTTCCCGTCCCCGGTCACGGTAGTGGGGATTGACGGTA
Consensus TGCCTTCCTTCGGCGGGCCCAGTTTCCCGTCCCYGGTCACGGTAGTGGGGATTGACGGTA
Active Sequence TGCCTTCCTTCGGCGGGCCCAGTTTCCCGTCCCCGGTCACGGTAGTGGGGATTGACGGTA
Macaco_AC210223 TGCCTTCCTTCAGTGGCCCCAGCCACCCCTCCACTGTCACAGTCATAGGAATTGATGGCA

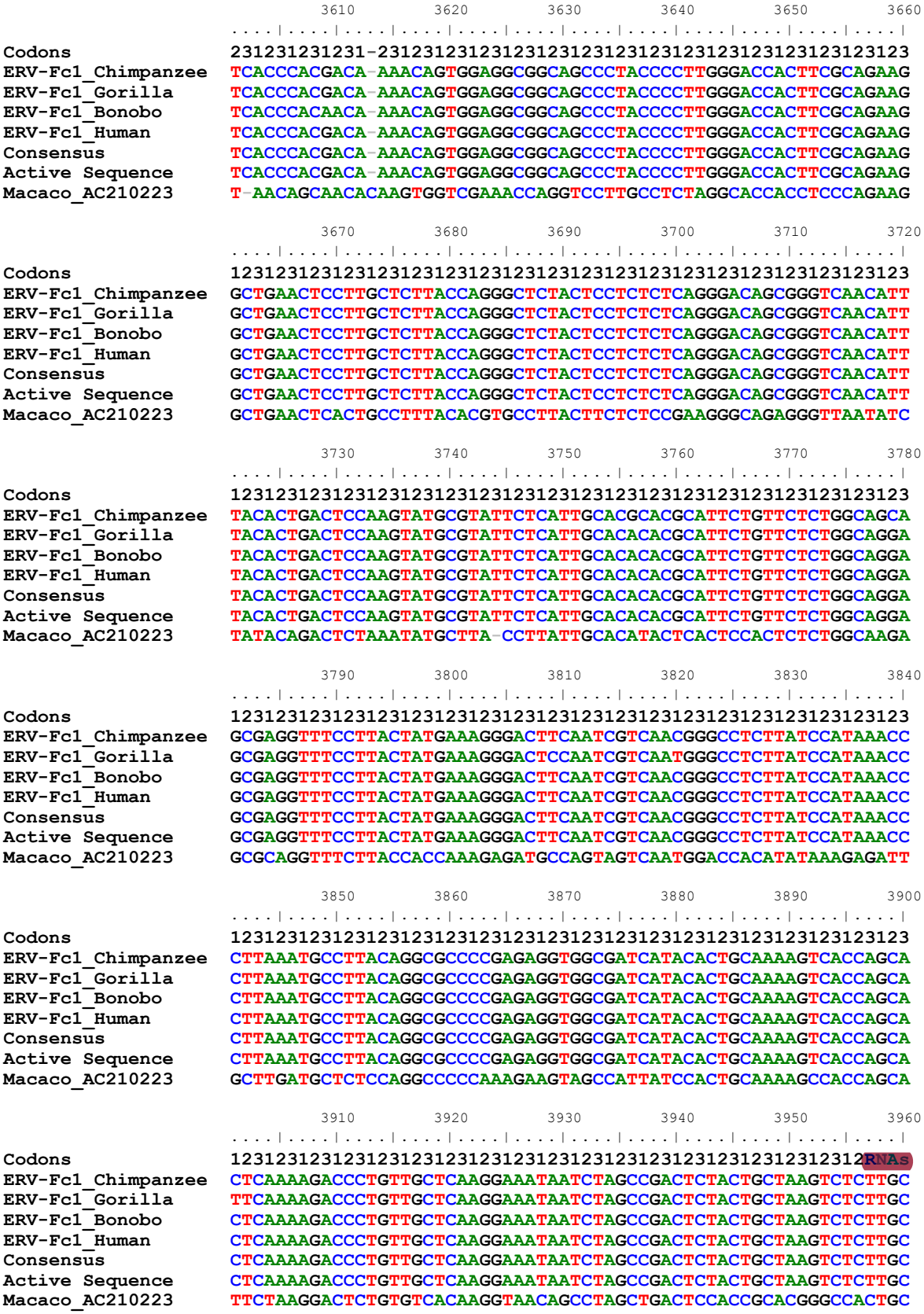

1810 1820 1830 1840 1850 1860
Codons 2312312312312312312---2312312312312312312312312312312312
ERV-Fc1_Chimpanzee CCCCTTCCACCCATCGTCAGA---CCCCCCCATTGTCTTGCCGGCTGGACGACACTCTCA
ERV-Fc1_Gorilla CCCCTTCCACCCATCGTCAGA---CCCCCCCATTGTCTTGCCGGCTGGACGACACTCTCA
ERV-Fc1_Bonobo CCCCTTCCACTCATCGTCAGA---CCCCCCCATTGTCTTGCCGGCTGGACGACACTCTCA
ERV-Fc1_Human CCCCTTCCACCCATCGTCAGACCCCCCCCCCCATTGTCTTGCCGGCTGGACGACACTCTCA
Consensus CCCCTTCCACCCATCGTCAGA---CCCCCCCATTGTCTTGCCGGCTGGACGACACTCTCA
Active Sequence CCCCTTCCACCCATCGTCAGA---CCCCCCCATTGTCTTGCCGGCTGGACGACACTCTCA
Macaco_AC210223 GTCCCTCCACCTACTGCCAGA---CTCCTCCTCTGTCTTGCTGCCTGGATGGCTCCCTCT
1870 1880 1890 1900 1910 1920
Codons 2312312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee TCTCCCATTCCTTCCTCATTATCCCTTCCTGTCCCGTCCTGCTCTTCGGAAGGGACTTGC
ERV-Fc1_Gorilla TCTCCCATTCCTTCCTCATTATCCCTTCCTGTCCCGTCCCGCTCTTCGGAAGGGACTTGC
ERV-Fc1_Bonobo TCTCCCATTCCTTCCTCATTATCCCTTCCTGTCCCGTCCCGCTCTTCGGAAGGGACTTGC
ERV-Fc1_Human TCTCCCATTCCTTCCTCATTATCCCTTCCTGTCCCGTCCTGCTCTTCGGAAGGGACTTGC
Consensus TCTCCCATTCCTTCCTCATTATCCCTTCCTGTCCCGTCCYGCTCTTCGGAAGGGACTTGC
Active Sequence TCTCCCATTCCTTCCTCATTATCCCTTCCTGTCCCGTCCCGCTCTTCGGAAGGGACTTGC
Macaco_AC210223 TCTCACACTCATTTCTTATCATTCCTTCGTGCCCAGTCCCCTTGTTAGGATGGGATCTCC
1930 1940 1950 1960 1970 1980
Codons 2PROe----------
ERV-Fc1_Chimpanzee TGTCTAAGTTAGGGGCCTCCATTTGGTTGCACCCCAGCCTCCCCTCCAGTGCAATCTCTT
ERV-Fc1_Gorilla TGTCTAAGTTAGGGGTCTCCATTCGGTTGCGCCCCAGCCTCCCCTCCCGTGCAATCTCTC
ERV-Fc1_Bonobo TGTCTAAGTTAGGGGCCTCCATTCGGTTGCGCCCCAGCCTCCCCTCCCGTGCAATCTCTT
ERV-Fc1_Human TGTCTAAGTTAGGGGCCTCCATTCGGTTGCACCCCAGCCTCCCCTCCAGTGCAATCTCTT
Consensus TGTCTAAGTTAGGGGCCTCCATTCGGTTGCRCCCCAGCCTCCCCTCCMGTGCAATCTCTT
Active Sequence TGTCTAAGTTAGGGGCCTCCATTCGGTTGCGCCCCAGCCTCCCCTCCCGTGCAATCTCTT
Macaco_AC210223 TCTCCAAGCTAGGGGCCTCAGTTCACTTCCGACCCAATCCCTCCCCACACCTCGCGTTCC
1990 2000 2010 2020 2030 2040
Codons ----------
ERV-Fc1_Chimpanzee TGCTTCCTCTGCTGGCACTTAGCGATGACACCCCTTCGCCGATCCC-----ATCGCTCCC
ERV-Fc1_Gorilla TGCTTCCTCTGCTGGCACTTAGCGATGACACTCCTTCGCCAATCCC-----ATTGCTCCC
ERV-Fc1_Bonobo TGCTTCCTCTGCTGGCACTTAGCGATGACACCCCTTCGCCGATCCC-----ATCGCTCCC
ERV-Fc1_Human TGCTTCCTCTGCTGGCACTTAGCGATGACACTCCTTCGCCGATCCC-----ATTGCTCCC
Consensus TGCTTCCTCTGCTGGCACTTAGCGATGACACYCCTTCGCCGATCCC-----ATYGCTCCC
Active Sequence TGCTTCCTCTGCTGGCACTTAGCGATGACACCCCTTCGCCGATCCC-----ATCGCTCCC
Macaco_AC210223 TCTTTCCCCTTCTCTCACCTGACAA--ACCCCGCCAGGCTGACCCCCGACTCCTGTTTCC
2050 2060 2070 2080 2090 2100
Codons ----------
ERV-Fc1_Chimpanzee TGTGCCCGTTGATCCAATAGTATGGGACATCTCAACCCCCTCCATCACCCGACACCATGC
ERV-Fc1_Gorilla AGTGCCCGCTGATCCAATAGTATGGGACATCTCAACCCCCTCCATCACCTGACACCATGC
ERV-Fc1_Bonobo TGTGCCCGTTGATCCAATAGTATGGGACATCTCAACCCCGTCCATCGTCCGACACCATGC
ERV-Fc1_Human TGTGCCCGTTGATCCAATAGTATGGGACATCTCAACCCCCTCCATCGCCCGACACCATGC
Consensus TGTGCCCGTTGATCCAATAGTATGGGACATCTCAACCCCCTCCATCRCCCGACACCATGC
Active Sequence TGTGCCCGTTGATCCAATAGTATGGGACATCTCAACCCCCTCCATCGCCCGACACCATGC
Macaco_AC210223 AGTCCCCATTAACCCTAAGGTGTGGGACACCTCCACCCCGATCATTGCCCAGCACCACAC
2110 2120 2130 2140 2150 2160
Codons ----------RTs12312312312312312
ERV-Fc1_Chimpanzee CCCAATAATGATCAAACTCAAGGACCCTACCAAATTTCCCTCGCGGCCCCAATTCCCCAT
ERV-Fc1_Gorilla CCCAATAATGATCAAACTCAAGGACCCTACCAAATTTCCCTCGCGGCCACAATTCCCCAT
ERV-Fc1_Bonobo CCCAATAATGATCAAACTCAAGGACCCTACCAAATTTCCCTCGCGGCCACAATTCCCCAT
ERV-Fc1_Human CCCAATAATGATCAAACTCAAGGACCCTACCAAATTTCCCTCGCGGCCACAATTCCCCAT
Consensus CCCAATAATGATCAAACTCAAGGACCCTACCAAATTTCCCTCGCGGCCACAATTCCCCAT
Active Sequence CCCAATAATGATCAAACTCAAGGACCCTACCAAATTTCCCTCGCGGCCACAATTCCCCAT
Macaco_AC210223 TCCGGTCCGCATCCGGCTAAAAGACCCTTCCAAGTTCCCTTCTAGACCCCAGTTCCCTAT

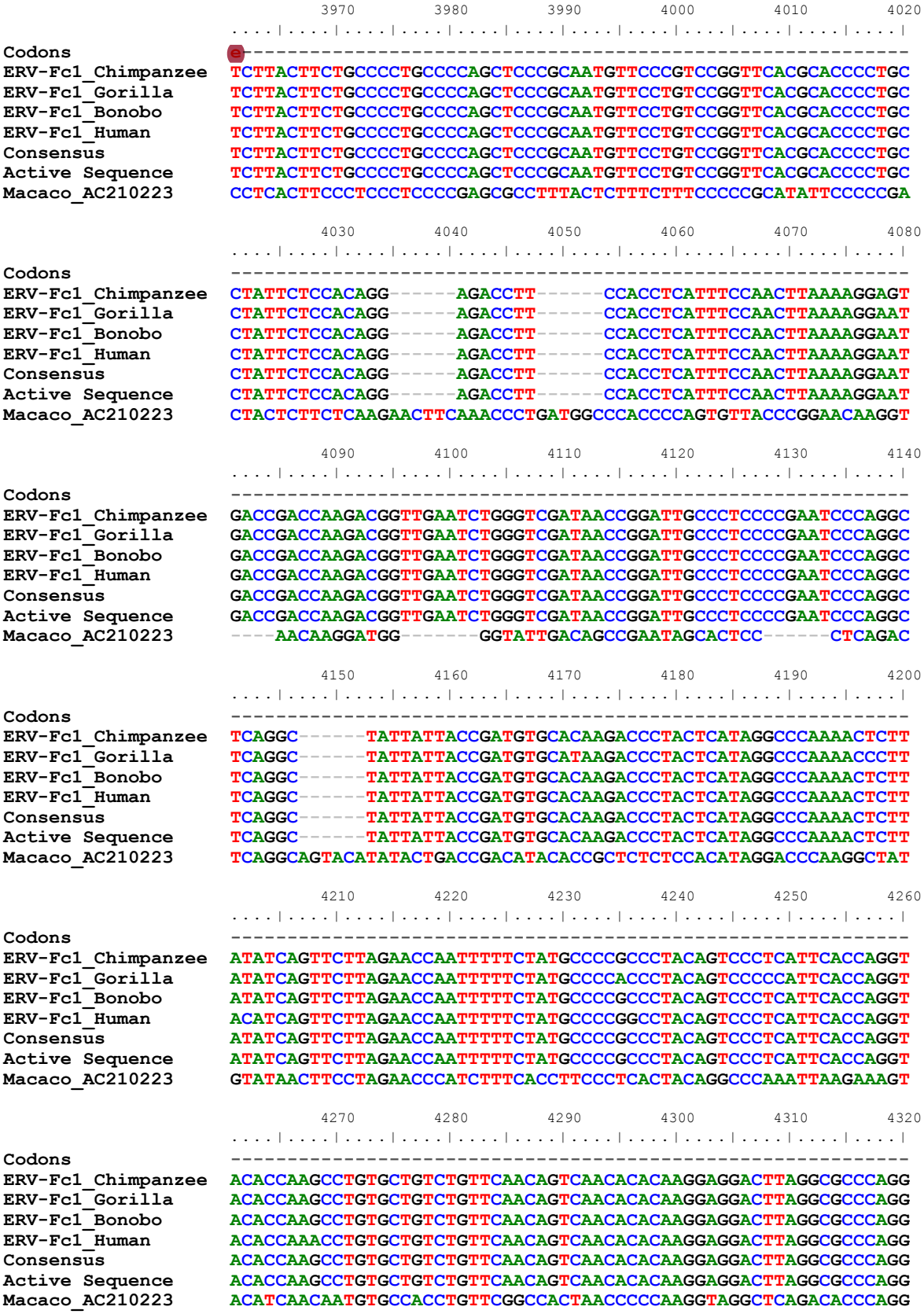

2170 2180 2190 2200 2210 2220
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee CTCAGTTGAACACCGCCAAGGGTTAAAACCTATCATCACCAGGCTCTTGCAACAACACAT
ERV-Fc1_Gorilla CTCAGTTGAACACCGCCAAGGGTTAAAACCTATCATCACCAGGCTCTTGCAACAACACAT
ERV-Fc1_Bonobo CTCAGTTGAACACCGCCAAGGGTTAAAACCTATCATCACCAGGCTCTTGCAACAACACAT
ERV-Fc1_Human CTCAGTTGAACACCGCCAAGGGTTAAAACCTATCATCACCAGGCTCTTGCAACAACACAT
Consensus CTCAGTTGAACACCGCCAAGGGTTAAAACCTATCATCACCAGGCTCTTGCAACAACACAT
Active Sequence CTCAGTTGAACACCGCCAAGGGTTAAAACCTATCATCACCAGGCTCTTGCAACAACACAT
Macaco_AC210223 CTCCCTTGAGCATCCACGGGGATTAAAACCTATCATTACACGCCTGCTCCAACAGCACAT
2230 2240 2250 2260 2270 2280
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee CCTTATCCCGGTAAACTCCCGTTGCAACACGCCCATTCTGCCCATCCGTAAGGCCTCTGG
ERV-Fc1_Gorilla CCTTATCCTGGTAAACTCCCGTTGCAACACGCCCATTCTGCCGATCCGTAAGGCCTCTGG
ERV-Fc1_Bonobo CCTTATCCCGGTAAACTCCCGTTGCAACACGCCCATTCTGCCCATCCGTAAGGCCTCTGG
ERV-Fc1_Human CCTTATCCCGGTAAACTCCCGTTGCAACACGCCCATTCTGCCCATCCGTAAGGCCTCTGG
Consensus CCTTATCCCGGTAAACTCCCGTTGCAACACGCCCATTCTGCCCATCCGTAAGGCCTCTGG
Active Sequence CCTTATCCCGGTAAACTCCCGTTGCAACACGCCCATTCTGCCCATCCGTAAGGCCTCTGG
Macaco_AC210223 TCTAGTTCCAGCCAACTCACCATGCAATACTCCTATCCTGCCTGTACAGAAAAGCTCTGG
2290 2300 2310 2320 2330 2340
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee CGCGTATCGTTTAGTGCAAGATCTTCGCATCATCAACGAGGCTGTCGTCCCCATTTTTCC
ERV-Fc1_Gorilla CGCGTACCGTTTAGTGCAAGATCTTCGCATCATCAACGAGGCTGTCGTCCCCATTTTCC
ERV-Fc1_Bonobo CGTGTACCGTTTAGTGCAAGATCTTCGCATTATCAACGAGGCTGTCGTCCCCATTTTTCC
ERV-Fc1_Human TGCGTACCGTTTAGTGCAAGATCTTCGCATCATCAACGAGGCTGTCGTCCCCATTTTTCC
Consensus CGCGTACCGTTTAGTGCAAGATCTTCGCATCATCAACGAGGCTGTCGTCCCCATTTTTCC
Active Sequence CGCGTACCGTTTAGTGCAAGATCTTCGCATCATCAACGAGGCTGTCGTCCCCATTTTTCC
Macaco_AC210223 GGCCTCTCGTCTCATGCAGGACCTACGCCTCATCAATGAGGCAGTGGTCCCTACCTTTCC
2350 2360 2370 2380 2390 2400
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee TGTAGTGCCTAACCCATACAGTCTCCTATCCCGCATTCCTCCGACCACCACCCATTTCAC
ERV-Fc1_Gorilla TGTAGTGCCTAACCCGTACACTCTCCTATCCCGCATTCCTCCGACCACCACCCATTTCAC
ERV-Fc1_Bonobo TGTAGTGCCTAACCCGTACACTCTCCTATCCCGCATTCCTCCGACCACCACCCATTTCAC
ERV-Fc1_Human TGTAGTGCCTAACCCATACACTCTCCTATCCCGCATTCCTCCGACCACCACCCATTTCAC
Consensus TGTAGTGCCTAACCCRTACACTCTCCTATCCCGCATTCCTCCGACCACCACCCATTTCAC
Active Sequence TGTAGTGCCTAACCCGTACACTCTCCTATCCCGCATTCCTCCGACCACCACCCATTTCAC
Macaco_AC210223 AGTTGTTCCTAATCCATATACACTCCTCTCACGCGTTCCCCCTGACACCACACATTTCAC
2410 2420 2430 2440 2450 2460
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee GGTCCTTGACCTCAAAGATGATTTCTTCACTATCCCCCTCCACCCTGACTGTTACTTCCT
ERV-Fc1_Gorilla GGTCCTTGACCTCAAAGATGCTTTCTTCACTATCCCCCTCCACCCTGACTGTTACTTCCT
ERV-Fc1_Bonobo GGTCCTTGACCTCAAAGATGCTTTCTTCACTATCCCCCTCCACCCTGACTGTTACTTCCT
ERV-Fc1_Human GGTCCTTGACCTCAAAGATGATTTCTTCACTATCCCCCTCCACCCTGACTGTTACTTCCT
Consensus GGTCCTTGACCTCAAAGATGMTTTCTTCACTATCCCCCTCCACCCTGACTGTTACTTCCT
Active Sequence GGTCCTTGACCTCAAAGATGCTTTCTTCACTATCCCCCTCCACCCTGACTGTTACTTCCT
Macaco_AC210223 TGTCCTTGACATAAAAGATGCCTTCTTTACCATCCCTCTACACCCAGACTGCCACTTTCT
2470 2480 2490 2500 2510 2520
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee GTTCGCTTTCACCTGGGAAGACCCTGACACCCATGTCTCCTCGCAACTTGCCTGGACCAT
ERV-Fc1_Gorilla GTTCCCTTTCACCTGGGAAGACCCTGACACCCATGTCTCCTCGCAACTTGCCTGGACCGT
ERV-Fc1_Bonobo GTTTGCTTTCACCTGGGAAGACCCTGACACCCATGTCTCCTCGCAACTTGCCTGGACCGT
ERV-Fc1_Human GTTCGCTTTCACCTGGGAAGACCCTGACACCCATGTCTCCTCGCAATTTGCCTGGACCGT
Consensus GTTCGCTTTCACCTGGGAAGACCCTGACACCCATGTCTCCTCGCAACTTGCCTGGACCGT
Active Sequence GTTCGCTTTCACCTGGGAAGACCCTGACACCCATGTCTCCTCGCAACTTGCCTGGACCGT
Macaco_AC210223 GTTTGCCTTCACATAGGAGGACCCAGACACTCATGTTTCTTCCCAGTTGACCTGGACTGT

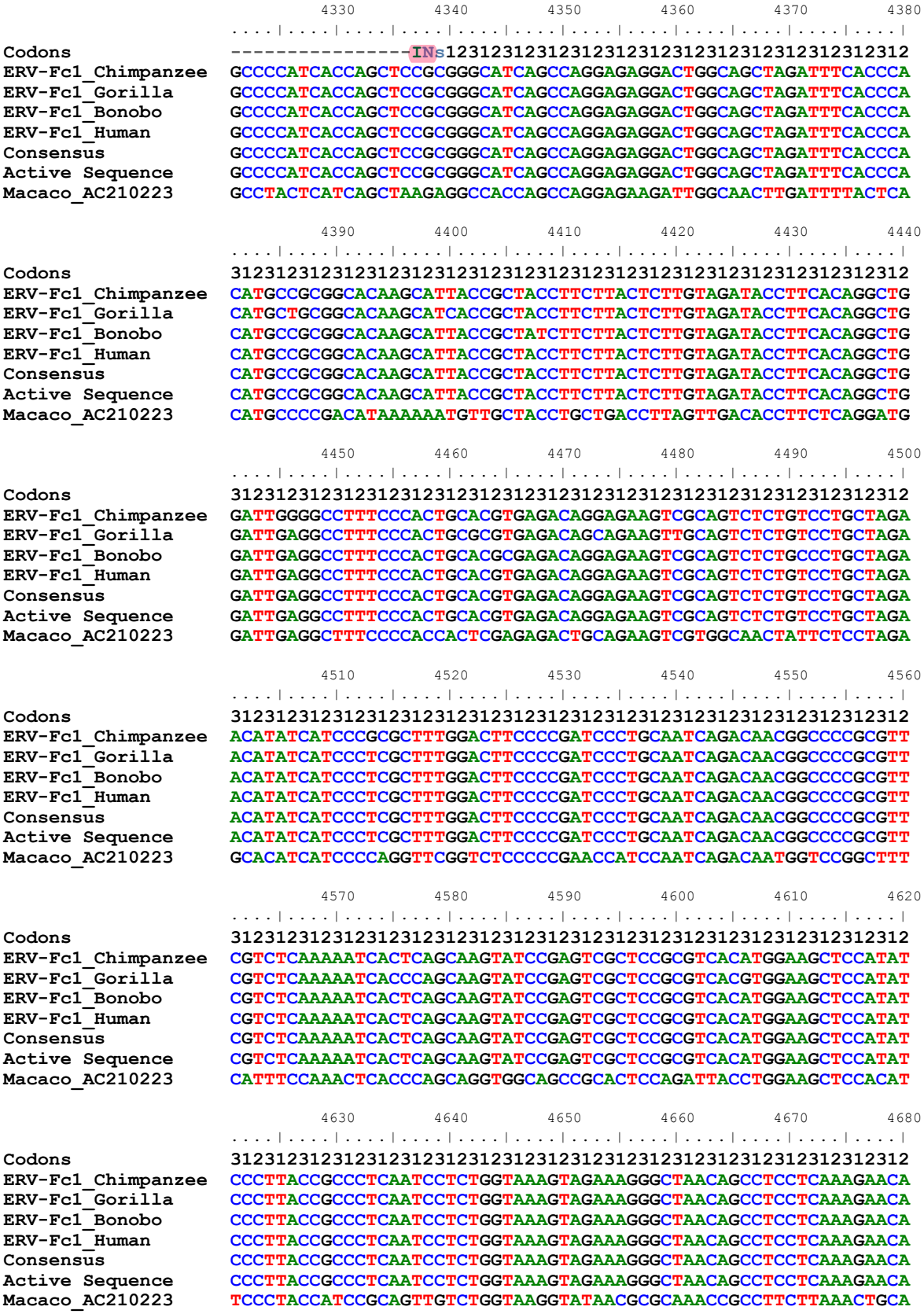
```
                          2530        2540        2550        2560        2570        2580
                          .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Codons                    312312312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee        TCTCCCGCAAGGCTTCCGAGACAGCCCTCACCTCTTTGGACAGGCTCTAGCTAAAGACCT
ERV-Fc1_Gorilla           TCTCCCGCAAGGCTTCCGAGACAGCCCTCACCTCTTTGGACAGGCTCTAGCTAAAGACCT
ERV-Fc1_Bonobo            TCTCCCGCAAGGCTTCCGAGACAGCCCTCACCTCTTTGGACAGGCTCTAGCTAAAGACCT
ERV-Fc1_Human             TCTCCCGCAAGGCTTCCGAGACAGCCCTCACCTCTTTGGACAGGCTCTAGCTAAAGACCT
Consensus                 TCTCCCGCAAGGCTTCCGAGACAGCCCTCACCTCTTTGGACAGGCTCTAGCTAAAGACCT
Active Sequence           TCTCCCGCAAGGCTTCCGAGACAGCCCTCACCTCTTTGGACAGGCTCTAGCTAAAGACCT
Macaco_AC210223           CTTGCCTCAAGGGTTCCGAGATAGCCCCCATTTTTTCGGCCAGGCACTGGCACAGGACAT

                          2590        2600        2610        2620        2630        2640
                          .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Codons                    312312312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee        CAGTACATGCACTTTGGCCGACAGCACCCTTCTCCTGTATGTTGATGACCTTCTCCTTTG
ERV-Fc1_Gorilla           CAGTACATGCACTTTGGCCGACAGCACCCTTCTCCAGTATGTTGATGACCTTCTCCTTTG
ERV-Fc1_Bonobo            CAGTACATGCACTTTGGCCGACAGCACCCTTCTCCTGTATGTTGATGACCTTCTCCTTTG
ERV-Fc1_Human             CAGTACATGCACTTTGGCCGACAGCACCCTTCTCCTGTATGTTGATGACCTTCTCCTTTG
Consensus                 CAGTACATGCACTTTGGCCGACAGCACCCTTCTCCTGTATGTTGATGACCTTCTCCTTTG
Active Sequence           CAGTACATGCACTTTGGCCGACAGCACCCTTCTCCTGTATGTTGATGACCTTCTCCTTTG
Macaco_AC210223           CCTCCTCTGCCCCCTAACTCATAGCACCCTTCTACAATACGTAGATGATCTATTACTATG

                          2650        2660        2670        2680        2690        2700
                          .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Codons                    3123123123123----123123123123123123123123123123123123123123
ERV-Fc1_Chimpanzee        CAGTCCTTCCCTG----TCTGTCTCGCAGCAAGATACAGCCACAATCCCTAATTTCTTAG
ERV-Fc1_Gorilla           CAGTCCTTCCCTG----TCTGTCTCGCAGCAAGATACAGCCACAATCCTTAATTTCTTAG
ERV-Fc1_Bonobo            CAGTCCTTCCCTG----TCTGTCTCGCAGCAAGATACAGCCACAATCCTTAATTTCTTAG
ERV-Fc1_Human             CAGTCCTTCCCTG----TCTGTCTCGCAGCAAGATACAGCCACAATCCTTAATTTCTTAG
Consensus                 CAGTCCTTCCCTG----TCTGTCTCGCAGCAAGATACAGCCACAATCCTTAATTTCTTAG
Active Sequence           CAGTCCTTCCCTG----TCTGTCTCGCAGCAAGATACAGCCACAATCCTTAATTTCTTAG
Macaco_AC210223           TAGTCCTTCCTGGGAGTGCTCCCTTGC----AGACACTGCTACACTTCTAAATTTTCTAG

                          2710        2720        2730        2740        2750        2760
                          .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Codons                    2312312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee        GAAAACAAGGGTATCGAGTTACCCCTCACAAAGTTCAGCTCTGCACCCCGACAGTCACAT
ERV-Fc1_Gorilla           GAAAACAAGGGTATCGAGTTACCCCTCACAAAGTTCAGCTCTGCACCCCAACAGTCACAT
ERV-Fc1_Bonobo            GAAAACAAGGGTATCGAGTTACCCCTCACAAAGTTCAGCTCTGCACCCCGACAGTCACAT
ERV-Fc1_Human             GAAAACAAGGGTATCGAGTTACCCCTCACAAAGTTCAGCTCTGCACCCCGACAGTCACAT
Consensus                 GAAAACAAGGGTATCGAGTTACCCCTCACAAAGTTCAGCTCTGCACCCCGACAGTCACAT
Active Sequence           GAAAACAAGGGTATCGAGTTACCCCTCACAAAGTTCAGCTCTGCACCCCGACAGTCACAT
Macaco_AC210223           GCAACGGAGGTTATCGGGTTACCCCGGCTAAGGCTCAACTTTGCACCCCTTCTGTCACCT

                          2770        2780        2790        2800        2810        2820
                          .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Codons                    23123123RTe-------------------------------------------------
ERV-Fc1_Chimpanzee        ACCTAGGCATTTCTCTCACCGCCACCACCAAAAGCCTCACCACAGACCGAGTTAGCCTCA
ERV-Fc1_Gorilla           ACCTAGGCATTTCTCTCACCCCCACCACCAAAAGCCTCACCACAGACTGAGTTAGCCTCA
ERV-Fc1_Bonobo            ACCTAGGCATTTCTCTCACCGCCACCACCAAAAGCCTCGCCACAGACCGAGTTAGCCTCA
ERV-Fc1_Human             ACCTAGGCATTTCTCTCACCGCCACCACCAAAAGCCTCACCACAGACCGAGTTAGCCTCA
Consensus                 ACCTAGGCATTTCTCTCACCGCCACCACCAAAAGCCTCACCACAGACCGAGTTAGCCTCA
Active Sequence           ACCTAGGCATTTCTCTCACCGCCACCACCAAAAGCCTCACCACAGACCGAGTTAGCCTCA
Macaco_AC210223           ACCTAGGCATACTACTCACACCCACTACAAAAAGCCTTACGGCAGATAGAATAAGCCTCA

                          2830        2840        2850        2860        2870        2880
                          .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Codons                    -------------------------------------------------------------
ERV-Fc1_Chimpanzee        TTAAAGACCTCCAACTTCCCCAGGACGCAGATAAGATCCTCTCCTTCGTAGGGCTAGTAG
ERV-Fc1_Gorilla           TTAAAGACCTCCAACCTCCCCAGGACGCAGATAAGATCCTCTCCTTCCTAGGGTTAGTAG
ERV-Fc1_Bonobo            TTAAAGACCTCCAACTTCCCCAGGACGCAGATAAGATCCTCTCCTTCGTAGGGCTAGTAG
ERV-Fc1_Human             TTAAAGACCTCCAACTTCCCCAGGACGCAGATAAGATCCTCTCCTTCGTAGGGCTAGTAG
Consensus                 TTAAAGACCTCCAACTTCCCCAGGACGCAGATAAGATCCTCTCCTTCGTAGGGCTAGTAG
Active Sequence           TTAAAGACCTCCAACTTCCCCAGGACGCAGATAAGATCCTCTCCTTCGTAGGGCTAGTAG
Macaco_AC210223           TTAAAACTCTCCAGCCTCCTCAGGATGCGGAAGAGATCTTGTCCTTCCTAGGACTGGTAC
```

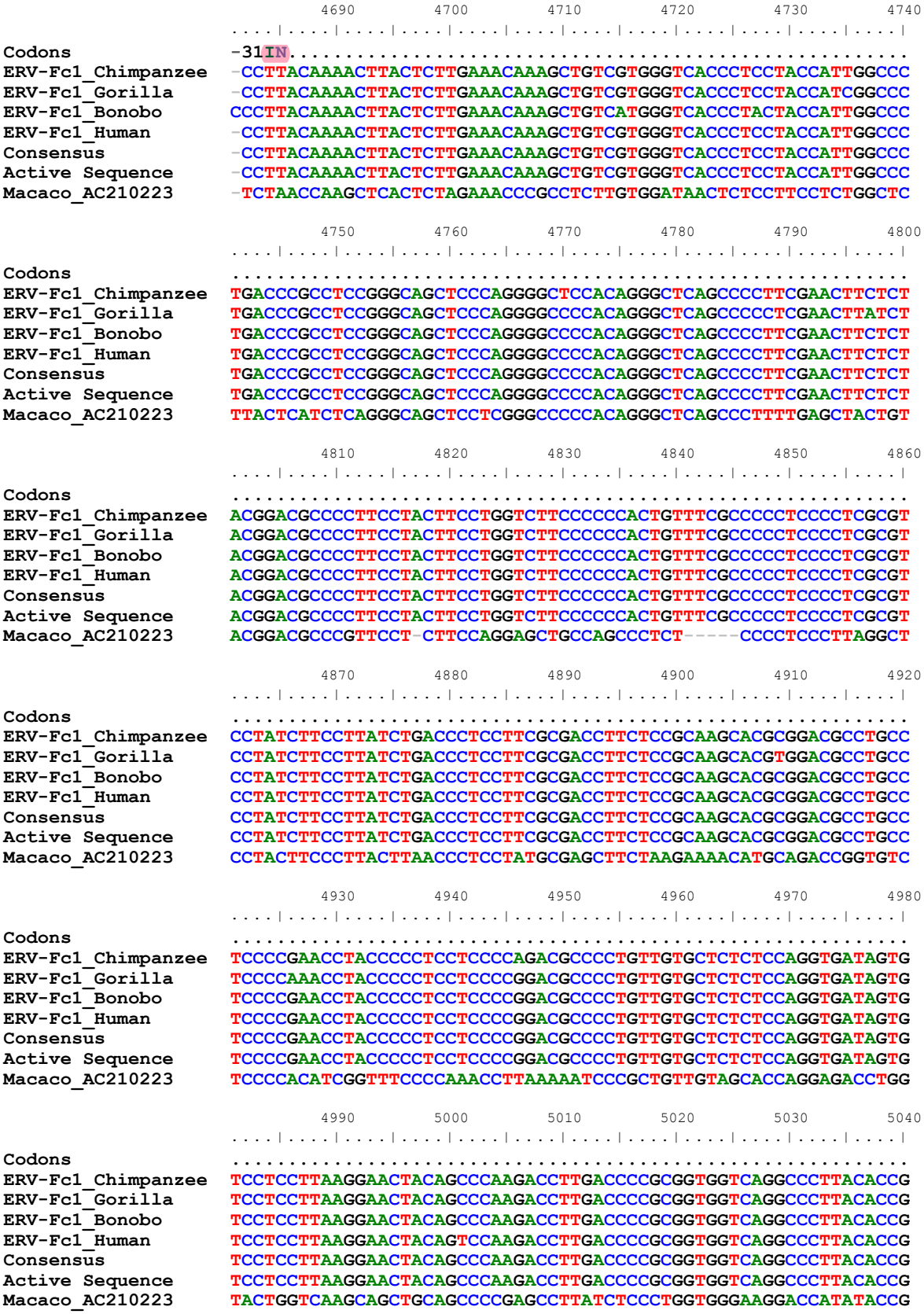

2890 2900 2910 2920 2930 2940
Codons
ERV-Fc1_Chimpanzee GGTTCTTCCGGCACTGGATCCCAAACTTCGGGGTCTTAGCTAAGCCCCTGTACCAGGCGG
ERV-Fc1_Gorilla GGTTCTTCCGGCACTGGATCCCAAACTTCGGGGTCTTAGCTAAGCCCCTGTACCAGGTGG
ERV-Fc1_Bonobo GGTTCTTCCGGCACTGGATCCCAAACTTCGGGGTCTTAGATAAGCCCCTGTACCAGGCGG
ERV-Fc1_Human GGTTCTTCCGGCACTGGATCCCAAACTTCGGGGTCTTAGCTAAGCCCCTGTACCAGGCGG
Consensus GGTTCTTCCGGCACTGGATCCCAAACTTCGGGGTCTTAGCTAAGCCCCTGTACCAGGCGG
Active Sequence GGTTCTTCCGGCACTGGATCCCAAACTTCGGGGTCTTAGCTAAGCCCCTGTACCAGGCGG
Macaco_AC210223 GGTATTTCAGGCATTGGATTTCCAACTTCGGGGTCCTAGCCAAGCCCCTCCACCAGGCTG
2950 2960 2970 2980 2990 3000
Codons
ERV-Fc1_Chimpanzee CGAAAGAAACACCCACCAGCCCTCTGTCTGATCCCGCCCTAGTGGCCCGCCATTTCCGCC
ERV-Fc1_Gorilla CGAAAGAAACACCCACCGGCCCTCTGTCTGATCCCGCCCTAGTGGCCCGCCATTTCCGCC
ERV-Fc1_Bonobo CGAAAGAAACACCCACCAGCCCTCTGTCTGATCCCGCCCTAGTGGCCCGCCATTTCCGCC
ERV-Fc1_Human CGAAAGAAACACCCACCAGCCCTCTGTCTGATCCCGCCCTAGTGGCCCGCCATTTCCACC
Consensus CGAAAGAAACACCCACCAGCCCTCTGTCTGATCCCGCCCTAGTGGCCCGCCATTTCCGCC
Active Sequence CGAAAGAAACACCCACCAGCCCTCTGTCTGATCCCGCCCTAGTGGCCCGCCATTTCCGCC
Macaco_AC210223 CCAGGGAGACACCCACCAGACCGCTGTCTGACCCCTCCTTGGTTGCCGACTCTTTCAAGA
3010 3020 3030 3040 3050 3060
Codons
ERV-Fc1_Chimpanzee GGCTGCAGCAGTGCTTACTCACAGCTCCAGTTTTATCCCTGCCGAACCCCCTGCGGCCTT
ERV-Fc1_Gorilla AGCTGCAGCAGTACTTACTCACAGCTCCAGTTTTATCCCTGCCGAACCCCCTGCGGCCTT
ERV-Fc1_Bonobo GGCTGCAGCAGTGCTTACTCACAGCTCCAGTTTTATCCCTGCTGAACCCCCTGCGGCCTT
ERV-Fc1_Human GGCTGCAGCAGTGCTTACTCACAGCTCCAGTTGTATCCCTGCCGAACCCCCTGCGGCCTT
Consensus GGCTGCAGCAGTGCTTACTCACAGCTCCAGTTTTATCCCTGCCGAACCCCCTGCGGCCTT
Active Sequence GGCTGCAGCAGTGCTTACTCACAGCTCCAGTTTTATCCCTGCCGAACCCCCTGCGGCCTT
Macaco_AC210223 AGCTTCAGGACTGTCTTCTTTCTGCCCCTGCTCTCTCTC--CCCAACCCCCTTTAGCCCT
3070 3080 3090 3100 3110 3120
Codons
ERV-Fc1_Chimpanzee TTCACCTCTACACAGATGAACTGCAGGGAGTTGCTACCGGCCTACTAGGGCAACCGGTAG
ERV-Fc1_Gorilla TTCACCTCTACACAGATGAACTGCAGGGAGTTGCTACCGGCCTACTAGGGCAACCAGTAG
ERV-Fc1_Bonobo TTCACCTCTACACAGATGAACTGCAGGGAGTTGCTACCGGCCTACTAGGGCAACCGGTAG
ERV-Fc1_Human TTCATCTCTACACAGATGAACTGCAGGGAGTTGCTACTGGCCTACTAGGGCAACCGGTAG
Consensus TTCACCTCTACACAGATGAACTGCAGGGAGTTGCTACCGGCCTACTAGGGCAACCGGTAG
Active Sequence TTCACCTCTACACAGATGAACTGCAGGGAGTTGCTACCGGCCTACTAGGGCAACCGGTAG
Macaco_AC210223 TTCATCTATTCACTGAGGAGCACCAGAAGGTAGCTACTGGCCTCCTAGCCCAGCCGGTTG
3130 3140 3150 3160 3170 3180
Codons
ERV-Fc1_Chimpanzee GACCCACCTATCAGGTGGTGGCTTACCTTTCCAGGCAGCTTGATCCCAGCACTCGGGGCT
ERV-Fc1_Gorilla GACCCACCTATCAGGTGGTGGCTTACCTTTCCAGGCAGCTTGATCCCAGCACTCGGGGCT
ERV-Fc1_Bonobo GACCCACCTATCAGGTGGTGGCTTACCTTTCCAGGCAGCTTGATCCCAGCACTCGGGGCT
ERV-Fc1_Human GACCCACCTATCAGGTGGTGGCTTACCTTTCCAGGCAGCTTGATCCCAGCACTCGGGGCT
Consensus GACCCACCTATCAGGTGGTGGCTTACCTTTCCAGGCAGCTTGATCCCAGCACTCGGGGCT
Active Sequence GACCCACCTATCAGGTGGTGGCTTACCTTTCCAGGCAGCTTGATCCCAGCACTCGGGGCT
Macaco_AC210223 GATCCACATACCAGGTTGTGGCTTATCTCTCCAAGCAGTTAGATCCCACAGTCCAGGGCT
3190 3200 3210 3220 3230 3240
Codons
ERV-Fc1_Chimpanzee GGCAGCCCTGCCTGCGGGCCTTAGCAGCGGCGGCAGAGCTTACCAAAGAGGCCCTCAAAC
ERV-Fc1_Gorilla GGCAGCCCTGCCTGTGGGCCTTAGCAGCGGTGACAGAGCTTACCAAAGAGGCCCTCAGAC
ERV-Fc1_Bonobo GGCAGCCCTGCCTGCGGGCCTTAGCAGCGGCGGCAGAGCTTACCAAAGAGGCCCTCAAAC
ERV-Fc1_Human GGCAGCCCTGCCTGCGGGCCTTAGCAGCGGCGGCAGAGCTTACCAAAGAGGCCCTCAAAC
Consensus GGCAGCCCTGCCTGCGGGCCTTAGCAGCGGCGGCAGAGCTTACCAAAGAGGCCCTCAAAC
Active Sequence GGCAGCCCTGCCTGCGGGCCTTAGCAGCGGCGGCAGAGCTTACCAAAGAGGCCCTCAAAC
Macaco_AC210223 GGCAACCTTGTCTACGAGCCCTGGCGGCTGCCACAGAACTTACCCAAGAGTCCCTCAAGG

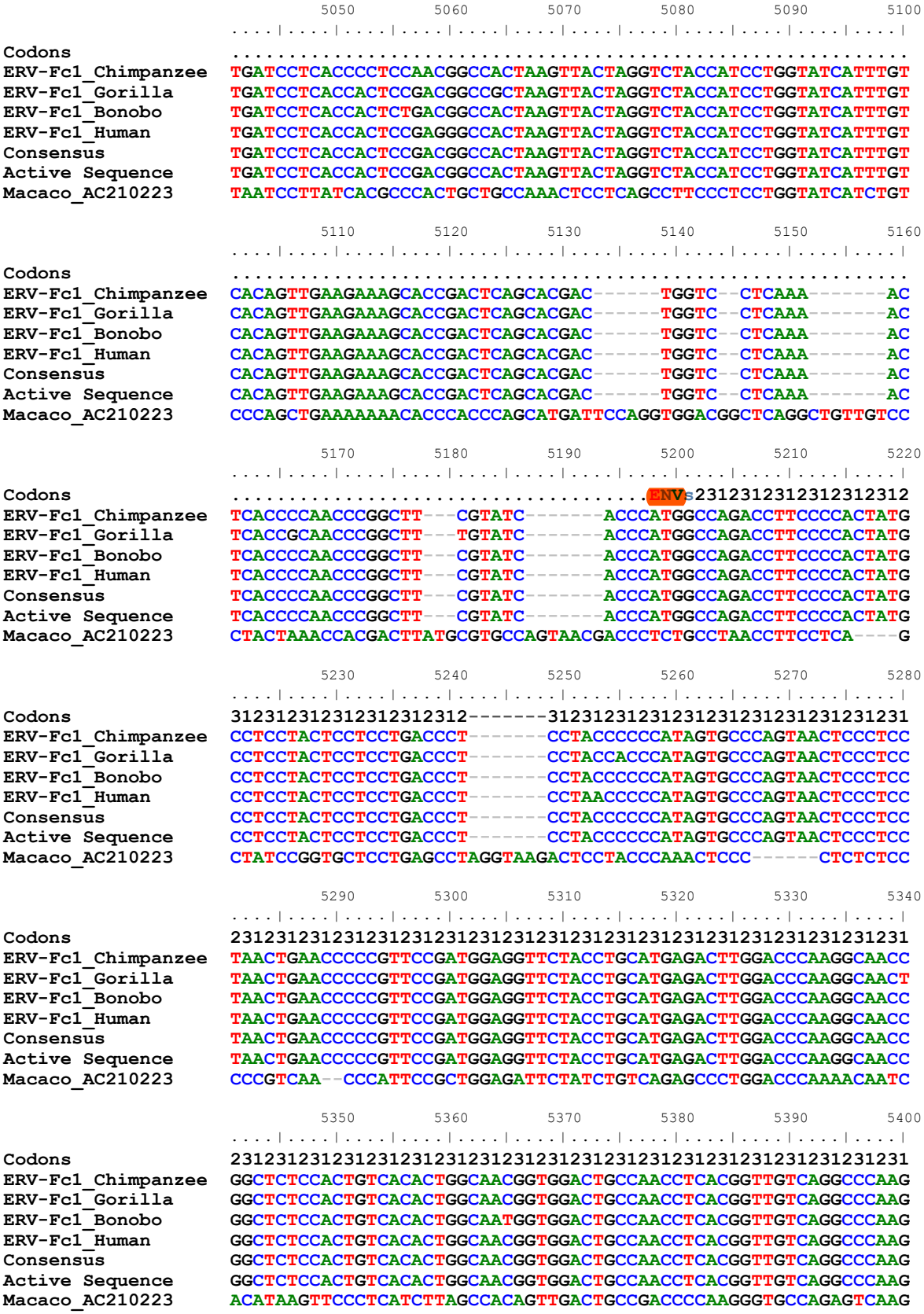

3250 3260 3270 3280 3290 3300
Codons
ERV-Fc1_Chimpanzee TTACTCTCAATCACCCACTCACAGTATACTCCCCCCACCGCTTGACAGATGTACTCTCTC
ERV-Fc1_Gorilla TTACTCTCAATCACCCACTCACAGTATACTCCCCCCACCGCTTGACAGATGTACTCTCTC
ERV-Fc1_Bonobo TTACTCTCAATCACCCACTCACAGTATACTCCCCCCACCGCTTGACAGATGTACTCTCTC
ERV-Fc1_Human TTACTCTCAGTCACCCACTCACAGTATACTCCCCGCACCGCTTGACAGATGTACTCTCTC
Consensus TTACTCTCAATCACCCACTCACAGTATACTCCCCCCACCGCTTGACAGATGTACTCTCTC
Active Sequence TTACTCTCAATCACCCACTCACAGTATACTCCCCCCACCGCTTGACAGATGTACTCTCTC
Macaco_AC210223 TTACCCTAGGACATCCTCTCACAGTCTACTCTTCTCACCGACTGTCAGATCTCAACACAC
3310 3320 3330 3340 3350 3360
Codons
ERV-Fc1_Chimpanzee ACAAATGTCTGGCCCATCTGGCGCCCTCCAGAATACAGCTGCTTCATGTGCTCTTTGTCG
ERV-Fc1_Gorilla ACAAATGTCTGGCCCATCTGGCGCCCTCCAGAATACAGCTGTTTCATGTGCTCTTTGTTG
ERV-Fc1_Bonobo ACAAATGTCTGGCCCATCTGGCGCCCTCCAGAATACAGCTGTTTCATGTGCTCTTTGTCG
ERV-Fc1_Human ACAAATGTCTGGCCCATCTGGCGCCCTCCAGAATACAGCTGTTTCATGTGCTCTTTGTCG
Consensus ACAAATGTCTGGCCCATCTGGCGCCCTCCAGAATACAGCTGTTTCATGTGCTCTTTGTCG
Active Sequence ACAAATGTCTGGCCCATCTGGCGCCCTCCAGAATACAGCTGTTTCATGTGCTCTTTGTCG
Macaco_AC210223 ACAAGTGTCTCAGCCATCTTACCCCCTTCCGGCTTCAACAGTTCCACTTGCTATTCATTG
3370 3380 3390 3400 3410 3420
Codons
ERV-Fc1_Chimpanzee AAAACCCAGATATCACCCTGACCGCCTCACCACCTCTTAACCCTGCTACACTTCTTCCCA
ERV-Fc1_Gorilla AAAACCCAGATATTACCCTGACCGCCTCACCACCTCTTAACCCTGCTACACTTCTTCCCA
ERV-Fc1_Bonobo AAAACCCAGATATCACCCTGACCGCCTCACCACCTCTTAACCCTGCTACACTTCTTCCCA
ERV-Fc1_Human AAAACCCAGATATCACCCTGACCGCCTCACCACCTCTTAACCCTGCTACACTTCTTCCCA
Consensus AAAACCCAGATATCACCCTGACCGCCTCACCACCTCTTAACCCTGCTACACTTCTTCCCA
Active Sequence AAAACCCAGATATCACCCTGACCGCCTCACCACCTCTTAACCCTGCTACACTTCTTCCCA
Macaco_AC210223 AAAGCCCTCACATCACTCTCACCACCTCACCCCCCTCTAAATCCTGCTACTCTCTTGCCCA
3430 3440 3450 3460 3470 3480
Codons
ERV-Fc1_Chimpanzee TAGAAGCCTCTGAGCCCTCTCCTGTCCTGTCGCATTCTTGTCCCGAACTCCTTACCTCTA
ERV-Fc1_Gorilla TAGAAGCCTCTGAGCCCTCTCCTGTCCTGTCGCATTCTTGTCCCGAACTCCTTACCTCTA
ERV-Fc1_Bonobo TAGAAGCCTCTGAGCCCTCTCCTGTCCTGTCGCATTCTTGTCCCGAACTCCTTACCTCTA
ERV-Fc1_Human TAGAAGCCTCTGAGCCCCCTCCTGTCCTGTCGCATTCTTGTCCCGAACTCCTTACCTCTA
Consensus TAGAAGCCTCTGAGCCCTCTCCTGTCCTGTCGCATTCTTGTCCCGAACTCCTTACCTCTA
Active Sequence TAGAAGCCTCTGAGCCCTCTCCTGTCCTGTCGCATTCTTGTCCCGAACTCCTTACCTCTA
Macaco_AC210223 ------CTCCAGGACACAATTCCGCCCCTGCACATTCCTGCCTGGAAGTTCTCACCACCT
3490 3500 3510 3520 3530 3540
Codons RNAse1
ERV-Fc1_Chimpanzee TTCCCAACTCCCGACTTGGCCTCTTCGATCGACCGCTTTCTAATCCTGACAGCACTCTGT
ERV-Fc1_Gorilla TCCCCAACTCCCGACTTGGCCTCTTCGATCGACCGCTTTCTAATCCTGACAGCACTCTGT
ERV-Fc1_Bonobo ACCCCAACTCCCGACTTGGCCTCTTCGATCGACCGCTTTCTAATCCTGACAGCACTCTGT
ERV-Fc1_Human ACCCCAACTCCCGACTTGGCCTCTTCGATCGACCGCTTTCTAATCCTGACAGCACTCTGT
Consensus WCCCCAACTCCCGACTTGGCCTCTTCGATCGACCGCTTTCTAATCCTGACAGCACTCTGT
Active Sequence TCCCCAACTCCCGACTTGGCCTCTTCGATCGACCGCTTTCTAATCCTGACAGCACTCTGT
Macaco_AC210223 TGCCACCCGCCTGCCTCGGTCTTTCCGAACAGCCGTTAGAACACCCAGAGCATATCCTGT
3550 3560 3570 3580 3590 3600
Codons 231231231231231231231231231231231231231231231231231231
ERV-Fc1_Chimpanzee TTGTCGATGGCAGCTCAGTCCTCACCCCTTGCGGTAGGCGACAGGCAGCTTACGCCGTAG
ERV-Fc1_Gorilla TTGTCGATGGCAGCTCAGTCCTCACCCCTTGCAGTAGGCGACAGGCAGCTTACGCCGTAG
ERV-Fc1_Bonobo TTGTCGATGGCAGCTCAGTCCTTACCCCTTGCGGTAGGCGACAGGCAGCTTACGCCGTAG
ERV-Fc1_Human TTGTCGATGGCAGCTCAGTCCTCACCCCTTGCGGTAGGCGACAGGCAGCTTACGCCGTAG
Consensus TTGTCGATGGCAGCTCAGTCCTCACCCCTTGCGGTAGGCGACAGGCAGCTTACGCCGTAG
Active Sequence TTGTCGATGGCAGCTCAGTCCTCACCCCTTGCGGTAGGCGACAGGCAGCTTACGCCGTAG
Macaco_AC210223 TCGTAGATGGGAGCTCTGTCTTAACCCCCGATGGCCGCAGACAGGTGGCATATGCTGTCG


3610 3620 3630 3640 3650 3660
Codons 231231231231-231231231231231231231231231231231231231231231231231
ERV-Fc1_Chimpanzee TCACCCACGACA-AAACAGTGGAGGCGGCAGCCCTACCCCTTGGGACCACTTCGCAGAAG
ERV-Fc1_Gorilla TCACCCACGACA-AAACAGTGGAGGCGGCAGCCCTACCCCTTGGGACCACTTCGCAGAAG
ERV-Fc1_Bonobo TCACCCACAACA-AAACAGTGGAGGCGGCAGCCCTACCCCTTGGGACCACTTCGCAGAAG
ERV-Fc1_Human TCACCCACGACA-AAACAGTGGAGGCGGCAGCCCTACCCCTTGGGACCACTTCGCAGAAG
Consensus TCACCCACGACA-AAACAGTGGAGGCGGCAGCCCTACCCCTTGGGACCACTTCGCAGAAG
Active Sequence TCACCCACGACA-AAACAGTGGAGGCGGCAGCCCTACCCCTTGGGACCACTTCGCAGAAG
Macaco_AC210223 T-AACAGCAACACAAGTGGTCGAAACCAGGTCCTTGCCTCTAGGCACCACCTCCCAGAAG
3670 3680 3690 3700 3710 3720
Codons 123123123123123123123123123123123123123123123123123123123123123
ERV-Fc1_Chimpanzee GCTGAACTCCTTGCTCTTACCAGGGCTCTACTCCTCTCTCAGGGACAGCGGGTCAACATT
ERV-Fc1_Gorilla GCTGAACTCCTTGCTCTTACCAGGGCTCTACTCCTCTCTCAGGGACAGCGGGTCAACATT
ERV-Fc1_Bonobo GCTGAACTCCTTGCTCTTACCAGGGCTCTACTCCTCTCTCAGGGACAGCGGGTCAACATT
ERV-Fc1_Human GCTGAACTCCTTGCTCTTACCAGGGCTCTACTCCTCTCTCAGGGACAGCGGGTCAACATT
Consensus GCTGAACTCCTTGCTCTTACCAGGGCTCTACTCCTCTCTCAGGGACAGCGGGTCAACATT
Active Sequence GCTGAACTCCTTGCTCTTACCAGGGCTCTACTCCTCTCTCAGGGACAGCGGGTCAACATT
Macaco_AC210223 GCTGAACTCACTGCCTTTACACGTGCCTTACTTCTCTCCGAAGGGCAGAGGGTTAATATC
3730 3740 3750 3760 3770 3780
Codons 123123123123123123123123123123123123123123123123123123123123123
ERV-Fc1_Chimpanzee TACACTGACTCCAAGTATGCGTATTCTCATTGCACGCACGCATTCTGTTCTCTGGCAGCA
ERV-Fc1_Gorilla TACACTGACTCCAAGTATGCGTATTCTCATTGCACACACGCATTCTGTTCTCTGGCAGGA
ERV-Fc1_Bonobo TACACTGACTCCAAGTATGCGTATTCTCATTGCACACACGCATTCTGTTCTCTGGCAGGA
ERV-Fc1_Human TACACTGACTCCAAGTATGCGTATTCTCATTGCACACACGCATTCTGTTCTCTGGCAGGA
Consensus TACACTGACTCCAAGTATGCGTATTCTCATTGCACACACGCATTCTGTTCTCTGGCAGGA
Active Sequence TACACTGACTCCAAGTATGCGTATTCTCATTGCACACACGCATTCTGTTCTCTGGCAGGA
Macaco_AC210223 TATACAGACTCTAAATATGCTTA-CCTTATTGCACATACTCACTCCACTCTCTGGCAAGA
3790 3800 3810 3820 3830 3840
Codons 123123123123123123123123123123123123123123123123123123123123123
ERV-Fc1_Chimpanzee GCGAGGTTTCCTTACTATGAAAGGGACTTCAATCGTCAACGGGCCTCTTATCCATAAACC
ERV-Fc1_Gorilla GCGAGGTTTCCTTACTATGAAAGGGACTCCAATCGTCAATGGGCCTCTTATCCATAAACC
ERV-Fc1_Bonobo GCGAGGTTTCCTTACTATGAAAGGGACTTCAATCGTCAACGGGCCTCTTATCCATAAACC
ERV-Fc1_Human GCGAGGTTTCCTTACTATGAAAGGGACTTCAATCGTCAACGGGCCTCTTATCCATAAACC
Consensus GCGAGGTTTCCTTACTATGAAAGGGACTTCAATCGTCAACGGGCCTCTTATCCATAAACC
Active Sequence GCGAGGTTTCCTTACTATGAAAGGGACTTCAATCGTCAACGGGCCTCTTATCCATAAACC
Macaco_AC210223 GCGCAGGTTTCTTACCACCAAAGAGATGCCAGTAGTCAATGGACCACATATAAAGAGATT
3850 3860 3870 3880 3890 3900
Codons 123123123123123123123123123123123123123123123123123123123123123
ERV-Fc1_Chimpanzee CTTAAATGCCTTACAGGCGCCCCGAGAGGTGGCGATCATACACTGCAAAAGTCACCAGCA
ERV-Fc1_Gorilla CTTAAATGCCTTACAGGCGCCCCGAGAGGTGGCGATCATACACTGCAAAAGTCACCAGCA
ERV-Fc1_Bonobo CTTAAATGCCTTACAGGCGCCCCGAGAGGTGGCGATCATACACTGCAAAAGTCACCAGCA
ERV-Fc1_Human CTTAAATGCCTTACAGGCGCCCCGAGAGGTGGCGATCATACACTGCAAAAGTCACCAGCA
Consensus CTTAAATGCCTTACAGGCGCCCCGAGAGGTGGCGATCATACACTGCAAAAGTCACCAGCA
Active Sequence CTTAAATGCCTTACAGGCGCCCCGAGAGGTGGCGATCATACACTGCAAAAGTCACCAGCA
Macaco_AC210223 GCTTGATGCTCTCCAGGCCCCCAAAGAAGTAGCCATTATCCACTGCAAAAGCCACCAGCA
3910 3920 3930 3940 3950 3960
Codons 123123123123123123123123123123123123123123123123123123RNAs
ERV-Fc1_Chimpanzee CTCAAAAGACCCTGTTGCTCAAGGAAATAATCTAGCCGACTCTACTGCTAAGTCTCTTGC
ERV-Fc1_Gorilla TTCAAAAGACCCTGTTGCTCAAGGAAATAATCTAGCCGACTCTACTGCTAAGTCTCTTGC
ERV-Fc1_Bonobo CTCAAAAGACCCTGTTGCTCAAGGAAATAATCTAGCCGACTCTACTGCTAAGTCTCTTGC
ERV-Fc1_Human CTCAAAAGACCCTGTTGCTCAAGGAAATAATCTAGCCGACTCTACTGCTAAGTCTCTTGC
Consensus CTCAAAAGACCCTGTTGCTCAAGGAAATAATCTAGCCGACTCTACTGCTAAGTCTCTTGC
Active Sequence CTCAAAAGACCCTGTTGCTCAAGGAAATAATCTAGCCGACTCTACTGCTAAGTCTCTTGC
Macaco_AC210223 TTCTAAGGACTCTGTGTCACAAGGTAACAGCCTAGCTGACTCCACCGCACGGGCCACTGC

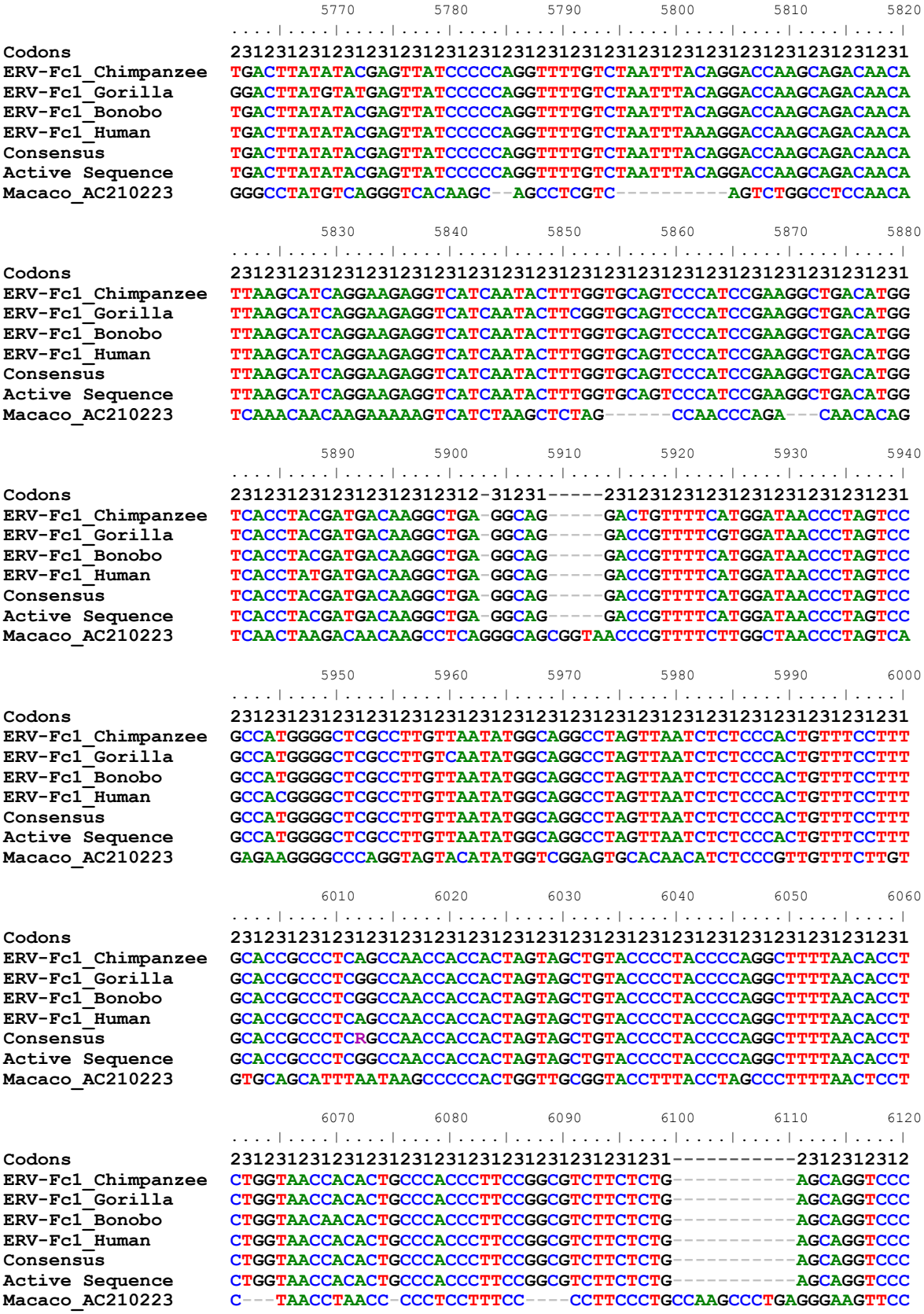

3970 3980 3990 4000 4010 4020
Codons
ERV-Fc1_Chimpanzee TCTTACTTCTGCCCCTGCCCCAGCTCCCGCAATGTTCCCGTCCGGTTCACGCACCCCTGC
ERV-Fc1_Gorilla TCTTACTTCTGCCCCTGCCCCAGCTCCCGCAATGTTCCTGTCCGGTTCACGCACCCCTGC
ERV-Fc1_Bonobo TCTTACTTCTGCCCCTGCCCCAGCTCCCGCAATGTTCCTGTCCGGTTCACGCACCCCTGC
ERV-Fc1_Human TCTTACTTCTGCCCCTGCCCCAGCTCCCGCAATGTTCCTGTCCGGTTCACGCACCCCTGC
Consensus TCTTACTTCTGCCCCTGCCCCAGCTCCCGCAATGTTCCTGTCCGGTTCACGCACCCCTGC
Active Sequence TCTTACTTCTGCCCCTGCCCCAGCTCCCGCAATGTTCCTGTCCGGTTCACGCACCCCTGC
Macaco_AC210223 CCTCACTTCCCTCCCTCCCCGAGCGCCTTTACTCTTTCTTTCCCCCGCATATTCCCCCGA
4030 4040 4050 4060 4070 4080
Codons
ERV-Fc1_Chimpanzee CTATTCTCCACAGG------AGACCTT------CCACCTCATTTCCAACTTAAAAGGAGT
ERV-Fc1_Gorilla CTATTCTCCACAGG------AGACCTT------CCACCTCATTTCCAACTTAAAAGGAAT
ERV-Fc1_Bonobo CTATTCTCCACAGG------AGACCTT------CCACCTCATTTCCAACTTAAAAGGAAT
ERV-Fc1_Human CTATTCTCCACAGG------AGACCTT------CCACCTCATTTCCAACTTAAAAGGAAT
Consensus CTATTCTCCACAGG------AGACCTT------CCACCTCATTTCCAACTTAAAAGGAAT
Active Sequence CTATTCTCCACAGG------AGACCTT------CCACCTCATTTCCAACTTAAAAGGAAT
Macaco_AC210223 CTACTCTTCTCAAGAACTTCAAACCCTGATGGCCCACCCCAGTGTTACCCGGAACAAGGT
4090 4100 4110 4120 4130 4140
Codons
ERV-Fc1_Chimpanzee GACCGACCAAGACGGTTGAATCTGGGTCGATAACCGGATTGCCCTCCCCGAATCCCAGGC
ERV-Fc1_Gorilla GACCGACCAAGACGGTTGAATCTGGGTCGATAACCGGATTGCCCTCCCCGAATCCCAGGC
ERV-Fc1_Bonobo GACCGACCAAGACGGTTGAATCTGGGTCGATAACCGGATTGCCCTCCCCGAATCCCAGGC
ERV-Fc1_Human GACCGACCAAGACGGTTGAATCTGGGTCGATAACCGGATTGCCCTCCCCGAATCCCAGGC
Consensus GACCGACCAAGACGGTTGAATCTGGGTCGATAACCGGATTGCCCTCCCCGAATCCCAGGC
Active Sequence GACCGACCAAGACGGTTGAATCTGGGTCGATAACCGGATTGCCCTCCCCGAATCCCAGGC
Macaco_AC210223 -----AACAAGGATGG-------GGTATTGACAGCCGAATAGCACTCC------CTCAGAC
4150 4160 4170 4180 4190 4200
Codons
ERV-Fc1_Chimpanzee TCAGGC------TATTATTACCGATGTGCACAAGACCCTACTCATAGGCCCAAAACTCTT
ERV-Fc1_Gorilla TCAGGC------TATTATTACCGATGTGCATAAGACCCTACTCATAGGCCCAAAACCCTT
ERV-Fc1_Bonobo TCAGGC------TATTATTACCGATGTGCACAAGACCCTACTCATAGGCCCAAAACTCTT
ERV-Fc1_Human TCAGGC------TATTATTACCGATGTGCACAAGACCCTACTCATAGGCCCAAAACTCTT
Consensus TCAGGC------TATTATTACCGATGTGCACAAGACCCTACTCATAGGCCCAAAACTCTT
Active Sequence TCAGGC------TATTATTACCGATGTGCACAAGACCCTACTCATAGGCCCAAAACTCTT
Macaco_AC210223 TCAGGCAGTACATATACTGACCGACATACACCGCTCTCTCCACATAGGACCCAAGGCTAT
4210 4220 4230 4240 4250 4260
Codons
ERV-Fc1_Chimpanzee ATATCAGTTCTTAGAACCAATTTTTCTATGCCCCGCCCTACAGTCCCTCATTCACCAGGT
ERV-Fc1_Gorilla ATATCAGTTCTTAGAACCAATTTTTCTATGCCCCACCCTACAGTCCCCCATTCACCAGGT
ERV-Fc1_Bonobo ATATCAGTTCTTAGAACCAATTTTTCTATGCCCCGCCCTACAGTCCCTCATTCACCAGGT
ERV-Fc1_Human ACATCAGTTCTTAGAACCAATTTTTCTATGCCCCGGCCTACAGTCCCTCATTCACCAGGT
Consensus ATATCAGTTCTTAGAACCAATTTTTCTATGCCCCGCCCTACAGTCCCTCATTCACCAGGT
Active Sequence ATATCAGTTCTTAGAACCAATTTTTCTATGCCCCGCCCTACAGTCCCTCATTCACCAGGT
Macaco_AC210223 GTATAACTTCCTAGAACCCATCTTTCACCTTCCCTCACTACAGGCCCAAATTAAGAAAGT
4270 4280 4290 4300 4310 4320
Codons
ERV-Fc1_Chimpanzee ACACCAAGCCTGTGCTGTCTGTTCAACAGTCAACACACAAGGAGGACTTAGGCGCCCAGG
ERV-Fc1_Gorilla ACACCAAGCCTGTGCTGTCTGTTCAACAGTCAACACACAAGGAGGACTTAGGCGCCCAGG
ERV-Fc1_Bonobo ACACCAAGCCTGTGCTGTCTGTTCAACAGTCAACACACAAGGAGGACTTAGGCGCCCAGG
ERV-Fc1_Human ACACCAAACCTGTGCTGTCTGTTCAACAGTCAACACACAAGGAGGACTTAGGCGCCCAGG
Consensus ACACCAAGCCTGTGCTGTCTGTTCAACAGTCAACACACAAGGAGGACTTAGGCGCCCAGG
Active Sequence ACACCAAGCCTGTGCTGTCTGTTCAACAGTCAACACACAAGGAGGACTTAGGCGCCCAGG
Macaco_AC210223 ACATCAACAATGTGCCACCTGTTCGGCCACTAACCCCCAAGGTAGGCTCAGACACCCAGG

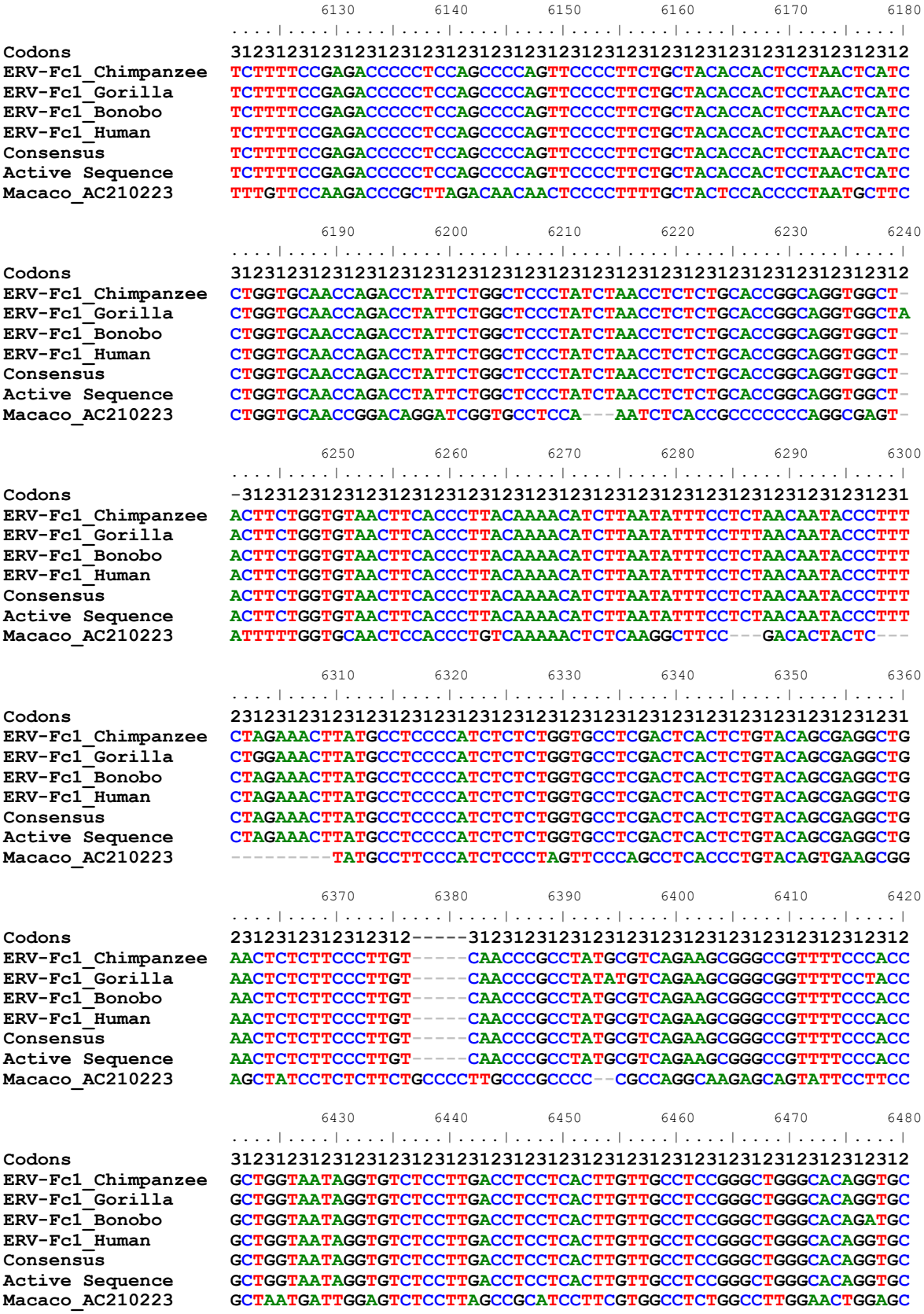

4330 4340 4350 4360 4370 4380
Codons ----------INs123123123123123123123123123123123123
ERV-Fc1_Chimpanzee GCCCCATCACCAGCTCCGCGGGCATCAGCCAGGAGAGGACTGGCAGCTAGATTTCACCCA
ERV-Fc1_Gorilla GCCCCATCACCAGCTCCGCGGGCATCAGCCAGGAGAGGACTGGCAGCTAGATTTCACCCA
ERV-Fc1_Bonobo GCCCCATCACCAGCTCCGCGGGCATCAGCCAGGAGAGGACTGGCAGCTAGATTTCACCCA
ERV-Fc1_Human GCCCCATCACCAGCTCCGCGGGCATCAGCCAGGAGAGGACTGGCAGCTAGATTTCACCCA
Consensus GCCCCATCACCAGCTCCGCGGGCATCAGCCAGGAGAGGACTGGCAGCTAGATTTCACCCA
Active Sequence GCCCCATCACCAGCTCCGCGGGCATCAGCCAGGAGAGGACTGGCAGCTAGATTTCACCCA
Macaco_AC210223 GCCTACTCATCAGCTAAGAGGCCACCAGCCAGGAGAAGATTGGCAACTTGATTTTACTCA
4390 4400 4410 4420 4430 4440
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee CATGCCGCGGCACAAGCATTACCGCTACCTTCTTACTCTTGTAGATACCTTCACAGGCTG
ERV-Fc1_Gorilla CATGCTGCGGCACAAGCATCACCGCTACCTTCTTACTCTTGTAGATACCTTCACAGGCTG
ERV-Fc1_Bonobo CATGCCGCGGCACAAGCATTACCGCTATCTTCTTACTCTTGTAGATACCTTCACAGGCTG
ERV-Fc1_Human CATGCCGCGGCACAAGCATTACCGCTACCTTCTTACTCTTGTAGATACCTTCACAGGCTG
Consensus CATGCCGCGGCACAAGCATTACCGCTACCTTCTTACTCTTGTAGATACCTTCACAGGCTG
Active Sequence CATGCCGCGGCACAAGCATTACCGCTACCTTCTTACTCTTGTAGATACCTTCACAGGCTG
Macaco_AC210223 CATGCCCCGACATAAAAAATGTTGCTACCTGCTGACCTTAGTTGACACCTTCTCAGGATG
4450 4460 4470 4480 4490 4500
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee GATTGGGGCCTTTCCCACTGCACGTGAGACAGGAGAAGTCGCAGTCTCTGTCCTGCTAGA
ERV-Fc1_Gorilla GATTGAGGCCTTTCCCACTGCGCGTGAGACAGCAGAAGTTGCAGTCTCTGTCCTGCTAGA
ERV-Fc1_Bonobo GATTGAGGCCTTTCCCACTGCACGCGAGACAGGAGAAGTCGCAGTCTCTGCCCTGCTAGA
ERV-Fc1_Human GATTGAGGCCTTTCCCACTGCACGTGAGACAGGAGAAGTCGCAGTCTCTGTCCTGCTAGA
Consensus GATTGAGGCCTTTCCCACTGCACGTGAGACAGGAGAAGTCGCAGTCTCTGTCCTGCTAGA
Active Sequence GATTGAGGCCTTTCCCACTGCACGTGAGACAGGAGAAGTCGCAGTCTCTGTCCTGCTAGA
Macaco_AC210223 GATTGAGGCTTTCCCCACCACTCGAGAGACTGCAGAAGTCGTGGCAACTATTCTCCTAGA
4510 4520 4530 4540 4550 4560
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee ACATATCATCCCGCGCTTTGGACTTCCCCGATCCCTGCAATCAGACAACGGCCCCGCGTT
ERV-Fc1_Gorilla ACATATCATCCCTCGCTTTGGACTTCCCCGATCCCTGCAATCAGACAACGGCCCCGCGTT
ERV-Fc1_Bonobo ACATATCATCCCTCGCTTTGGACTTCCCCGATCCCTGCAATCAGACAACGGCCCCGCGTT
ERV-Fc1_Human ACATATCATCCCTCGCTTTGGACTTCCCCGATCCCTGCAATCAGACAACGGCCCCGCGTT
Consensus ACATATCATCCCTCGCTTTGGACTTCCCCGATCCCTGCAATCAGACAACGGCCCCGCGTT
Active Sequence ACATATCATCCCTCGCTTTGGACTTCCCCGATCCCTGCAATCAGACAACGGCCCCGCGTT
Macaco_AC210223 GCACATCATCCCCAGGTTCGGTCTCCCCCGAACCATCCAATCAGACAATGGTCCGGCTTT
4570 4580 4590 4600 4610 4620
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee CGTCTCAAAAATCACTCAGCAAGTATCCGAGTCGCTCCGCGTCACATGGAAGCTCCATAT
ERV-Fc1_Gorilla CGTCTCAAAAATCACCCAGCAAGTATCCGAGTCGCTCCGCGTCACGTGGAAGCTCCATAT
ERV-Fc1_Bonobo CGTCTCAAAAATCACTCAGCAAGTATCCGAGTCGCTCCGCGTCACATGGAAGCTCCATAT
ERV-Fc1_Human CGTCTCAAAAATCACTCAGCAAGTATCCGAGTCGCTCCGCGTCACATGGAAGCTCCATAT
Consensus CGTCTCAAAAATCACTCAGCAAGTATCCGAGTCGCTCCGCGTCACATGGAAGCTCCATAT
Active Sequence CGTCTCAAAAATCACTCAGCAAGTATCCGAGTCGCTCCGCGTCACATGGAAGCTCCATAT
Macaco_AC210223 CATTTCCAAACTCACCCAGCAGGTGGCAGCCGCACTCCAGATTACCTGGAAGCTCCACAT
4630 4640 4650 4660 4670 4680
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee CCCTTACCGCCCTCAATCCTCTGGTAAAGTAGAAAGGGCTAACAGCCTCCTCAAAGAACA
ERV-Fc1_Gorilla CCCTTACCGCCCTCAATCCTCTGGTAAAGTAGAAAGGGCTAACAGCCTCCTCAAAGAACA
ERV-Fc1_Bonobo CCCTTACCGCCCTCAATCCTCTGGTAAAGTAGAAAGGGCTAACAGCCTCCTCAAAGAACA
ERV-Fc1_Human CCCTTACCGCCCTCAATCCTCTGGTAAAGTAGAAAGGGCTAACAGCCTCCTCAAAGAACA
Consensus CCCTTACCGCCCTCAATCCTCTGGTAAAGTAGAAAGGGCTAACAGCCTCCTCAAAGAACA
Active Sequence CCCTTACCGCCCTCAATCCTCTGGTAAAGTAGAAAGGGCTAACAGCCTCCTCAAAGAACA
Macaco_AC210223 TCCCTACCATCCGCAGTTGTCTGGTAAGGTATAACGCGCAAACCGCCTTCTTAAACTGCA

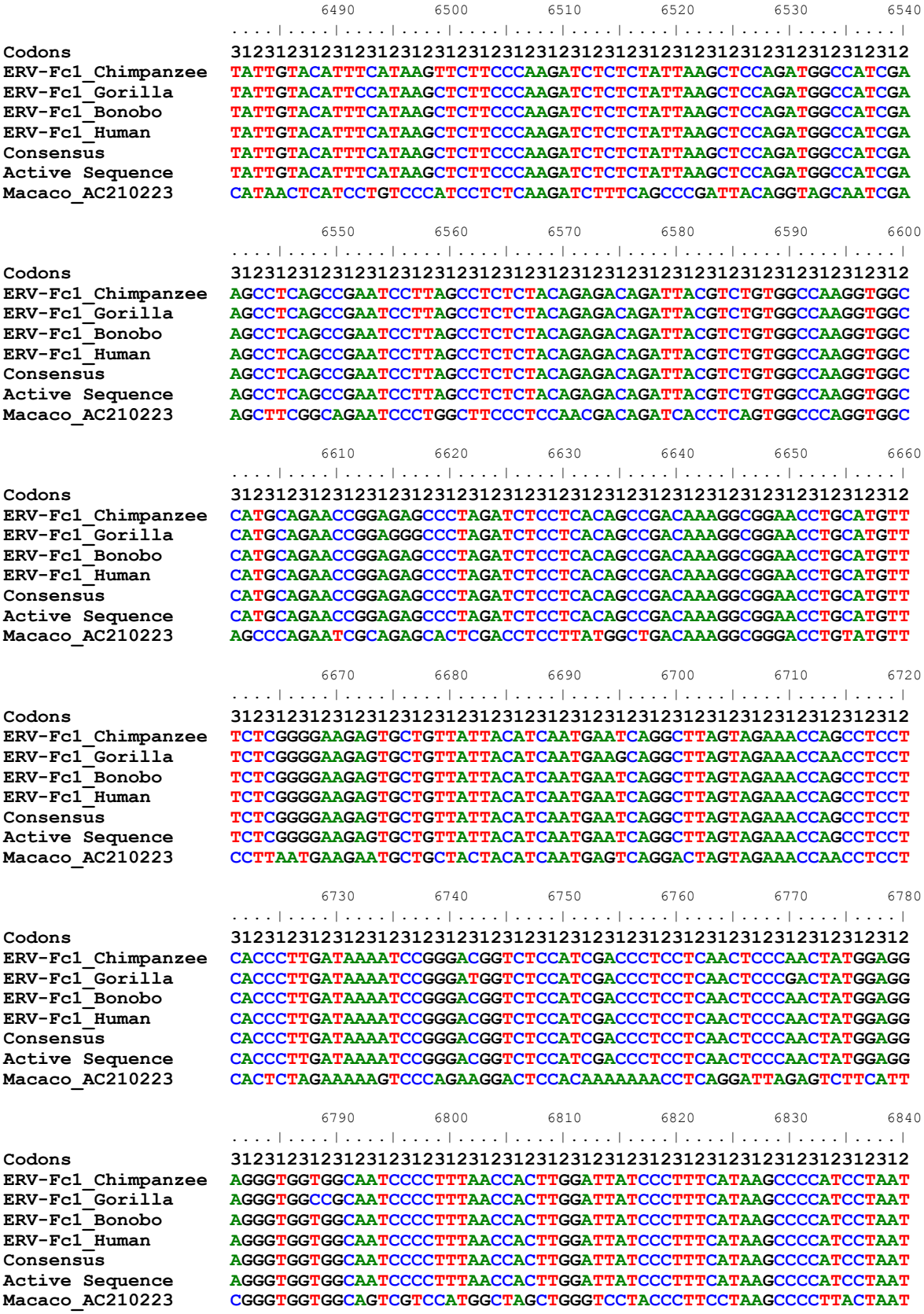

4690 4700 4710 4720 4730 4740
Codons -31IN
ERV-Fc1_Chimpanzee -CCTTACAAAACTTACTCTTGAAACAAAGCTGTCGTGGGTCACCCTCCTACCATTGGCCC
ERV-Fc1_Gorilla -CCTTACAAAACTTACTCTTGAAACAAAGCTGTCGTGGGTCACCCTCCTACCATCGGCCC
ERV-Fc1_Bonobo CCCTTACAAAACTTACTCTTGAAACAAAGCTGTCATGGGTCACCCTACTACCATTGGCCC
ERV-Fc1_Human -CCTTACAAAACTTACTCTTGAAACAAAGCTGTCGTGGGTCACCCTCCTACCATTGGCCC
Consensus -CCTTACAAAACTTACTCTTGAAACAAAGCTGTCGTGGGTCACCCTCCTACCATTGGCCC
Active Sequence -CCTTACAAAACTTACTCTTGAAACAAAGCTGTCGTGGGTCACCCTCCTACCATTGGCCC
Macaco_AC210223 -TCTAACCAAGCTCACTCTAGAAACCCGCCTCTTGTGGATAACTCTCCTTCCTCTGGCTC
4750 4760 4770 4780 4790 4800
Codons
ERV-Fc1_Chimpanzee TGACCCGCCTCCGGGCAGCTCCCAGGGGCTCCACAGGGCTCAGCCCCTTCGAACTTCTCT
ERV-Fc1_Gorilla TGACCCGCCTCCGGGCAGCTCCCAGGGGCCCCACAGGGCTCAGCCCCCTCGAACTTATCT
ERV-Fc1_Bonobo TGACCCGCCTCCGGGCAGCTCCCAGGGGCCCCACAGGGCTCAGCCCCTTCGAACTTCTCT
ERV-Fc1_Human TGACCCGCCTCCGGGCAGCTCCCAGGGGCCCCACAGGGCTCAGCCCCTTCGAACTTCTCT
Consensus TGACCCGCCTCCGGGCAGCTCCCAGGGGCCCCACAGGGCTCAGCCCCTTCGAACTTCTCT
Active Sequence TGACCCGCCTCCGGGCAGCTCCCAGGGGCCCCACAGGGCTCAGCCCCTTCGAACTTCTCT
Macaco_AC210223 TTACTCATCTCAGGGCAGCTCCTCGGGCCCCCACAGGGCTCAGCCCTTTTGAGCTACTGT
4810 4820 4830 4840 4850 4860
Codons
ERV-Fc1_Chimpanzee ACGGACGCCCCTTCCTACTTCCTGGTCTTCCCCCCACTGTTTCGCCCCCTCCCCTCGCGT
ERV-Fc1_Gorilla ACGGACGCCCCTTCCTACTTCCTGGTCTTCCCCCCACTGTTTCGCCCCCTCCCCTCGCGT
ERV-Fc1_Bonobo ACGGACGCCCCTTCCTACTTCCTGGTCTTCCCCCCACTGTTTCGCCCCCTCCCCTCGCGT
ERV-Fc1_Human ACGGACGCCCCTTCCTACTTCCTGGTCTTCCCCCCACTGTTTCGCCCCCTCCCCTCGCGT
Consensus ACGGACGCCCCTTCCTACTTCCTGGTCTTCCCCCCACTGTTTCGCCCCCTCCCCTCGCGT
Active Sequence ACGGACGCCCCTTCCTACTTCCTGGTCTTCCCCCCACTGTTTCGCCCCCTCCCCTCGCGT
Macaco_AC210223 ACGGACGCCCGTTCCT-CTTCCAGGAGCTGCCAGCCCTCT-----CCCCTCCCTTAGGCT
4870 4880 4890 4900 4910 4920
Codons
ERV-Fc1_Chimpanzee CCTATCTTCCTTATCTGACCCTCCTTCGCGACCTTCTCCGCAAGCACGCGGACGCCTGCC
ERV-Fc1_Gorilla CCTATCTTCCTTATCTGACCCTCCTTCGCGACCTTCTCCGCAAGCACGTGGACGCCTGCC
ERV-Fc1_Bonobo CCTATCTTCCTTATCTGACCCTCCTTCGCGACCTTCTCCGCAAGCACGCGGACGCCTGCC
ERV-Fc1_Human CCTATCTTCCTTATCTGACCCTCCTTCGCGACCTTCTCCGCAAGCACGCGGACGCCTGCC
Consensus CCTATCTTCCTTATCTGACCCTCCTTCGCGACCTTCTCCGCAAGCACGCGGACGCCTGCC
Active Sequence CCTATCTTCCTTATCTGACCCTCCTTCGCGACCTTCTCCGCAAGCACGCGGACGCCTGCC
Macaco_AC210223 CCTACTTCCCTTACTTAACCCTCCTATGCGAGCTTCTAAGAAAACATGCAGACCGGTGTC
4930 4940 4950 4960 4970 4980
Codons
ERV-Fc1_Chimpanzee TCCCCGAACCTACCCCCTCCTCCCCAGACGCCCCTGTTGTGCTCTCTCCAGGTGATAGTG
ERV-Fc1_Gorilla TCCCCAAACCTACCCCCTCCTCCCCGGACGCCCCTGTTGTGCTCTCTCCAGGTGATAGTG
ERV-Fc1_Bonobo TCCCCGAACCTACCCCCTCCTCCCCGGACGCCCCTGTTGTGCTCTCTCCAGGTGATAGTG
ERV-Fc1_Human TCCCCGAACCTACCCCCTCCTCCCCGGACGCCCCTGTTGTGCTCTCTCCAGGTGATAGTG
Consensus TCCCCGAACCTACCCCCTCCTCCCCGGACGCCCCTGTTGTGCTCTCTCCAGGTGATAGTG
Active Sequence TCCCCGAACCTACCCCCTCCTCCCCGGACGCCCCTGTTGTGCTCTCTCCAGGTGATAGTG
Macaco_AC210223 TCCCCACATCGGTTTCCCCAAACCTTAAAAATCCCGCTGTTGTAGCACCAGGAGACCTGG
4990 5000 5010 5020 5030 5040
Codons
ERV-Fc1_Chimpanzee TCCTCCTTAAGGAACTACAGCCCAAGACCTTGACCCCGCGGTGGTCAGGCCCTTACACCG
ERV-Fc1_Gorilla TCCTCCTTAAGGAACTACAGCCCAAGACCTTGACCCCGCGGTGGTCAGGCCCTTACACCG
ERV-Fc1_Bonobo TCCTCCTTAAGGAACTACAGCCCAAGACCTTGACCCCGCGGTGGTCAGGCCCTTACACCG
ERV-Fc1_Human TCCTCCTTAAGGAACTACAGTCCAAGACCTTGACCCCGCGGTGGTCAGGCCCTTACACCG
Consensus TCCTCCTTAAGGAACTACAGCCCAAGACCTTGACCCCGCGGTGGTCAGGCCCTTACACCG
Active Sequence TCCTCCTTAAGGAACTACAGCCCAAGACCTTGACCCCGCGGTGGTCAGGCCCTTACACCG
Macaco_AC210223 TACTGGTCAAGCAGCTGCAGCCCCGAGCCTTATCTCCCTGGTGGGAAGGACCATATACCG


5050 5060 5070 5080 5090 5100
Codons
ERV-Fc1_Chimpanzee TGATCCTCACCCCTCCAACGGCCACTAAGTTACTAGGTCTACCATCCTGGTATCATTTGT
ERV-Fc1_Gorilla TGATCCTCACCACTCCGACGGCCGCTAAGTTACTAGGTCTACCATCCTGGTATCATTTGT
ERV-Fc1_Bonobo TGATCCTCACCACTCTGACGGCCACTAAGTTACTAGGTCTACCATCCTGGTATCATTTGT
ERV-Fc1_Human TGATCCTCACCACTCCGAGGGCCACTAAGTTACTAGGTCTACCATCCTGGTATCATTTGT
Consensus TGATCCTCACCACTCCGACGGCCACTAAGTTACTAGGTCTACCATCCTGGTATCATTTGT
Active Sequence TGATCCTCACCACTCCGACGGCCACTAAGTTACTAGGTCTACCATCCTGGTATCATTTGT
Macaco_AC210223 TAATCCTTATCACGCCCACTGCTGCCAAACTCCTCAGCCTTCCCTCCTGGTATCATCTGT
5110 5120 5130 5140 5150 5160
Codons
ERV-Fc1_Chimpanzee CACAGTTGAAGAAAGCACCGACTCAGCACGAC------TGGTC--CTCAAA-------AC
ERV-Fc1_Gorilla CACAGTTGAAGAAAGCACCGACTCAGCACGAC------TGGTC--CTCAAA-------AC
ERV-Fc1_Bonobo CACAGTTGAAGAAAGCACCGACTCAGCACGAC------TGGTC--CTCAAA-------AC
ERV-Fc1_Human CACAGTTGAAGAAAGCACCGACTCAGCACGAC------TGGTC--CTCAAA-------AC
Consensus CACAGTTGAAGAAAGCACCGACTCAGCACGAC------TGGTC--CTCAAA-------AC
Active Sequence CACAGTTGAAGAAAGCACCGACTCAGCACGAC------TGGTC--CTCAAA-------AC
Macaco_AC210223 CCCAGCTGAAAAAAACACCCACCCAGCATGATTCCAGGTGGACGGCTCAGGCTGTTGTCC
5170 5180 5190 5200 5210 5220
Codons ENVs2312312312312312312
ERV-Fc1_Chimpanzee TCACCCCAACCCGGCTT---CGTATC-------ACCCATGGCCAGACCTTCCCCACTATG
ERV-Fc1_Gorilla TCACCGCAACCCGGCTT---TGTATC-------ACCCATGGCCAGACCTTCCCCACTATG
ERV-Fc1_Bonobo TCACCCCAACCCGGCTT---CGTATC-------ACCCATGGCCAGACCTTCCCCACTATG
ERV-Fc1_Human TCACCCCAACCCGGCTT---CGTATC-------ACCCATGGCCAGACCTTCCCCACTATG
Consensus TCACCCCAACCCGGCTT---CGTATC-------ACCCATGGCCAGACCTTCCCCACTATG
Active Sequence TCACCCCAACCCGGCTT---CGTATC-------ACCCATGGCCAGACCTTCCCCACTATG
Macaco_AC210223 CTACTAAACCACGACTTATGCGTGCCAGTAACGACCCTCTGCCTAACCTTCCTCA----G
5230 5240 5250 5260 5270 5280
Codons 312312312312312312-------312312312312312312312312312312
ERV-Fc1_Chimpanzee CCTCCTACTCCTCCTGACCCT-------CCTACCCCCCATAGTGCCCAGTAACTCCCTCC
ERV-Fc1_Gorilla CCTCCTACTCCTCCTGACCCT-------CCTACCACCCATAGTGCCCAGTAACTCCCTCC
ERV-Fc1_Bonobo CCTCCTACTCCTCCTGACCCT-------CCTACCCCCCATAGTGCCCAGTAACTCCCTCC
ERV-Fc1_Human CCTCCTACTCCTCCTGACCCT-------CCTAACCCCCATAGTGCCCAGTAACTCCCTCC
Consensus CCTCCTACTCCTCCTGACCCT-------CCTACCCCCCATAGTGCCCAGTAACTCCCTCC
Active Sequence CCTCCTACTCCTCCTGACCCT-------CCTACCCCCCATAGTGCCCAGTAACTCCCTCC
Macaco_AC210223 CTATCCGGTGCTCCTGAGCCTAGGTAAGACTCCTACCCAAACTCCC------CTCTCTCC
5290 5300 5310 5320 5330 5340
Codons 231231231231231231231231231231231231231231231231231231231231
ERV-Fc1_Chimpanzee TAACTGAACCCCCGTTCCGATGGAGGTTCTACCTGCATGAGACTTGGACCCAAGGCAACC
ERV-Fc1_Gorilla TAACTGAACCCCCGTTCCGATGGAGGTTCTACCTGCATGAGACTTGGACCCAAGGCAACT
ERV-Fc1_Bonobo TAACTGAACCCCCGTTCCGATGGAGGTTCTACCTGCATGAGACTTGGACCCAAGGCAACC
ERV-Fc1_Human TAACTGAACCCCCGTTCCGATGGAGGTTCTACCTGCATGAGACTTGGACCCAAGGCAACC
Consensus TAACTGAACCCCCGTTCCGATGGAGGTTCTACCTGCATGAGACTTGGACCCAAGGCAACC
Active Sequence TAACTGAACCCCCGTTCCGATGGAGGTTCTACCTGCATGAGACTTGGACCCAAGGCAACC
Macaco_AC210223 CCCGTCAA--CCCATTCCGCTGGAGATTCTATCTGTCAGAGCCCTGGACCCAAAACAATC
5350 5360 5370 5380 5390 5400
Codons 231231231231231231231231231231231231231231231231231231231231
ERV-Fc1_Chimpanzee GGCTCTCCACTGTCACACTGGCAACGGTGGACTGCCAACCTCACGGTTGTCAGGCCCAAG
ERV-Fc1_Gorilla GGCTCTCCACTGTCACACTGGCAACGGTGGACTGCCAACCTCACGGTTGTCAGGCCCAAG
ERV-Fc1_Bonobo GGCTCTCCACTGTCACACTGGCAATGGTGGACTGCCAACCTCACGGTTGTCAGGCCCAAG
ERV-Fc1_Human GGCTCTCCACTGTCACACTGGCAACGGTGGACTGCCAACCTCACGGTTGTCAGGCCCAAG
Consensus GGCTCTCCACTGTCACACTGGCAACGGTGGACTGCCAACCTCACGGTTGTCAGGCCCAAG
Active Sequence GGCTCTCCACTGTCACACTGGCAACGGTGGACTGCCAACCTCACGGTTGTCAGGCCCAAG
Macaco_AC210223 ACATAAGTTCCCTCATCTTAGCCACAGTTGACTGCCGACCCCAAGGGTGCCAGAGTCAAG

```
                                   5410      5420      5430      5440      5450      5460
                          ....|....|....|....|....|....|....|....|....|....|....|....|
Codons                    2312312312312312312312312312-3123123123123123123123123123123
ERV-Fc1_Chimpanzee        TAACTTTTAACTTCACTTCCTTTAAAAG-TGTTCTGCGGGGCTGGTCCAATCCCACCATC
ERV-Fc1_Gorilla           TAACTTTTAACTTCACTTCCTTTAAAAG-TGTTCTGCGGGGCTGGTCCAATCCCACCATC
ERV-Fc1_Bonobo            TAACTTTTAACTTCACTTCCTTTAAAAG-CGTTCTGTGGGGCTGGTCCAATCCCACCATC
ERV-Fc1_Human             TAACTTTTAACTTCACTTCCTTTAAAAG-TGTTCTGCGGGGCTGGTCCAATCCCACCATC
Consensus                 TAACTTTTAACTTCACTTCCTTTAAAAG-TGTTCTGCGGGGCTGGTCCAATCCCACCATC
Active Sequence           TAACTTTTAACTTCACTTCCTTTAAAAG-TGTTCTGCGGGGCTGGTCCAATCCCACCATC
Macaco_AC210223           TTACCTTTAACTTCTCCGCCTTTAACAGTTGCCCT----GACTGGTGGAACCCAGTCATA

                                   5470      5480      5490      5500      5510      5520
                          ....|....|....|....|....|....|....|....|....|....|....|....|
Codons                    123123123123123123123123123123123123123123123123123123123123
ERV-Fc1_Chimpanzee        GGCTTTGTCTATGATCAAACACACAGCAACTGCCGCGACTATTGGGTGGACACAAACGGA
ERV-Fc1_Gorilla           TGCTTTGTCTATGATCAAACACACAGCAACTGTCGCGACTATTGGGCGGACACAAACGGA
ERV-Fc1_Bonobo            TGCTTTGTCTATGATCAAACACACAGCAACTGCCGTGACTATTGGGTGGACACAAACGGA
ERV-Fc1_Human             TGCTTTGTCTATGATCAAACACACAGCAACTGCCGCGACTATTGGGTGGACACAAACGGA
Consensus                 TGCTTTGTCTATGATCAAACACACAGCAACTGCCGCGACTATTGGGTGGACACAAACGGA
Active Sequence           TGCTTTGTCTATGATCAAACACACAGCAACTGCCGCGACTATTGGGTGGACACAAACGGA
Macaco_AC210223           TGCTTTCTCTATGATCAAGTAGAATATAACTGTCTCAATTACTGGGTAGAAACCAATGGC

                                   5530      5540      5550      5560      5570      5580
                          ....|....|....|....|....|....|....|....|....|....|....|....|
Codons                    12312312312312312312312312312312312312312312312------3123123
ERV-Fc1_Chimpanzee        GGATGCCCCTATGCCTATTGTCGTATGCATGTGACCCAGCTCGATAC------CGCCAAG
ERV-Fc1_Gorilla           GGATGCCCCTATGCCTATTGTCATATGCATGTGACCCAGCTCGATAC------TGCCAAG
ERV-Fc1_Bonobo            GGATGCCCCTATGCCTATTGTCGTATGCATGTGACCCAGCTCGATAC------CGCCAAG
ERV-Fc1_Human             GGATGCCCCTATGCCTATTGTCGTATGCATGTGACCCAGCTCCATAC------CGCCAAG
Consensus                 GGATGCCCCTATGCCTATTGTCGTATGCATGTGACCCAGCTCGATAC------CGCCAAG
Active Sequence           GGATGCCCCTATGCCTATTGTCGTATGCATGTGACCCAGCTCGATAC------CGCCAAG
Macaco_AC210223           GGGTGTCCATATCATTATTGTAACATGCATTTTACTTACCTTGACATGTCATACACCAAG

                                   5590      5600      5610      5620      5630      5640
                          ....|....|....|....|....|....|....|....|....|....|....|....|
Codons                    1231231231--23123123123123-------123123123123--1231-------23
ERV-Fc1_Chimpanzee        AAAGTCCAAC--ACACCTATCGCCTG-------ACATCTGATGGA--AGGA-------CA
ERV-Fc1_Gorilla           AAACTCCAAC--ACACCTATCGCCTG-------ACATCTGATGGA--AGGA-------CA
ERV-Fc1_Bonobo            AAACTCCAAC--ACACCTATCGCCTG-------ACATCTGATGGA--AGGA-------CA
ERV-Fc1_Human             AAACTCCAAC--ACACCTATCGCCTG-------ACATCTGATGGA--AGGA-------CA
Consensus                 AAACTCCAAC--ACACCTATCGCCTG-------ACATCTGATGGA--AGGA-------CA
Active Sequence           AAACTCCAAC--ACACCTATCGCCTG-------ACATCTGATGGA--AGGA-------CA
Macaco_AC210223           TGGCAGCAACCGGCATCAACAGTTCGGTTAGTCAGATCATACGGACGAGGAGAAGTGCCT

                                   5650      5660      5670      5680      5690      5700
                          ....|....|....|....|....|....|....|....|....|....|....|....|
Codons                    1231231231231231231231231231231231231231231--2312---31231231
ERV-Fc1_Chimpanzee        ACTTACTTCCTGACCATCCCAGACCCATGGGATTCTCGGTGGG--TCAG---TGGAGTCA
ERV-Fc1_Gorilla           ACTTACTTCCTGACCATCCCAGACCCATGGGATTCTCGGTGGG--TCAG---TAGAGTCA
ERV-Fc1_Bonobo            ACTTACTTCCTGACCATCCCAGACCCATGGGATTCTCGGTGGG--TCAG---TGGAGTCA
ERV-Fc1_Human             ACTTACTTCCTGACCATCCCAGACCCATGGGATTCTCGGTGGG--TCAG---TGGAGTCA
Consensus                 ACTTACTTCCTGACCATCCCAGACCCATGGGATTCTCGGTGGG--TCAG---TGGAGTCA
Active Sequence           ACTTACTTCCTGACCATCCCAGACCCATGGGATTCTCGGTGGG--TCAG---TGGAGTCA
Macaco_AC210223           ACATTCTTCCTTACCATTCCTGACCCGTGGGACCCTCGGTGGGCATCAGGTATAGAGGC-

                                   5710      5720      5730      5740      5750      5760
                          ....|....|....|....|....|....|....|....|....|....|....|....|
Codons                    231231231231231231231231231231231231231231231231231231231231
ERV-Fc1_Chimpanzee        CTGGTCGACTGTACCGGTGGCCCACTGACTCCTACCCGGTCGGCAAACTCCGGATATTCC
ERV-Fc1_Gorilla           CTGGTCGACTGTACCAGTGGCCCACCGACTCCTACCCAGTCAGCAAACTCCGGATTTTCC
ERV-Fc1_Bonobo            CTGGTCGACTGTACCGGTGGCCCACCGACTCCTACCCAGTTGGCAAACTCCGGATATTCC
ERV-Fc1_Human             CTGGTCGACTGTACCGGTGGCCCACCGACTCCTACCCAGTTGGCAAACTCCGGATATTCC
Consensus                 CTGGTCGACTGTACCGGTGGCCCACCGACTCCTACCCAGTYGGCAAACTCCGGATATTCC
Active Sequence           CTGGTCGACTGTACCGGTGGCCCACCGACTCCTACCCAGTCGGCAAACTCCGGATATTCC
Macaco_AC210223           ----TCGCCTTTACCGGCACGGCTACGAATCTTATCCCGTAGCCCGACTCAAGATTTATA
```


5770 5780 5790 5800 5810 5820
Codons 231231231231231231231231231231231231231231231231231231231231
ERV-Fc1_Chimpanzee TGACTTATATACGAGTTATCCCCCAGGTTTTGTCTAATTTACAGGACCAAGCAGACAACA
ERV-Fc1_Gorilla GGACTTATGTATGAGTTATCCCCCAGGTTTTGTCTAATTTACAGGACCAAGCAGACAACA
ERV-Fc1_Bonobo TGACTTATATACGAGTTATCCCCCAGGTTTTGTCTAATTTACAGGACCAAGCAGACAACA
ERV-Fc1_Human TGACTTATATACGAGTTATCCCCCAGGTTTTGTCTAATTTAAAGGACCAAGCAGACAACA
Consensus TGACTTATATACGAGTTATCCCCCAGGTTTTGTCTAATTTACAGGACCAAGCAGACAACA
Active Sequence TGACTTATATACGAGTTATCCCCCAGGTTTTGTCTAATTTACAGGACCAAGCAGACAACA
Macaco_AC210223 GGGCCTATGTCAGGGTCACAAGC--AGCCTCGTC----------AGTCTGGCCTCCAACA
5830 5840 5850 5860 5870 5880
Codons 231231231231231231231231231231231231231231231231231231231231
ERV-Fc1_Chimpanzee TTAAGCATCAGGAAGAGGTCATCAATACTTTGGTGCAGTCCCATCCGAAGGCTGACATGG
ERV-Fc1_Gorilla TTAAGCATCAGGAAGAGGTCATCAATACTTCGGTGCAGTCCCATCCGAAGGCTGACATGG
ERV-Fc1_Bonobo TTAAGCATCAGGAAGAGGTCATCAATACTTTGGTGCAGTCCCATCCGAAGGCTGACATGG
ERV-Fc1_Human TTAAGCATCAGGAAGAGGTCATCAATACTTTGGTGCAGTCCCATCCGAAGGCTGACATGG
Consensus TTAAGCATCAGGAAGAGGTCATCAATACTTTGGTGCAGTCCCATCCGAAGGCTGACATGG
Active Sequence TTAAGCATCAGGAAGAGGTCATCAATACTTTGGTGCAGTCCCATCCGAAGGCTGACATGG
Macaco_AC210223 TCAAACAACAAGAAAAAGTCATCTAAGCTCTAG------CCAACCCAGA---CAACACAG
5890 5900 5910 5920 5930 5940
Codons 231231231231231231231-31231-----231231231231231231231231231
ERV-Fc1_Chimpanzee TCACCTACGATGACAAGGCTGA-GGCAG-----GACTGTTTTCATGGATAACCCTAGTCC
ERV-Fc1_Gorilla TCACCTACGATGACAAGGCTGA-GGCAG-----GACCGTTTTCGTGGATAACCCTAGTCC
ERV-Fc1_Bonobo TCACCTACGATGACAAGGCTGA-GGCAG-----GACCGTTTTCATGGATAACCCTAGTCC
ERV-Fc1_Human TCACCTATGATGACAAGGCTGA-GGCAG-----GACCGTTTTCATGGATAACCCTAGTCC
Consensus TCACCTACGATGACAAGGCTGA-GGCAG-----GACCGTTTTCATGGATAACCCTAGTCC
Active Sequence TCACCTACGATGACAAGGCTGA-GGCAG-----GACCGTTTTCATGGATAACCCTAGTCC
Macaco_AC210223 TCAACTAAGACAACAAGCCTCAGGGCAGCGGTAACCCGTTTTCTTGGCTAACCCTAGTCA
5950 5960 5970 5980 5990 6000
Codons 231231231231231231231231231231231231231231231231231231231231
ERV-Fc1_Chimpanzee GCCATGGGGCTCGCCTTGTTAATATGGCAGGCCTAGTTAATCTCTCCCACTGTTTCCTTT
ERV-Fc1_Gorilla GCCATGGGGCTCGCCTTGTCAATATGGCAGGCCTAGTTAATCTCTCCCACTGTTTCCTTT
ERV-Fc1_Bonobo GCCATGGGGCTCGCCTTGTTAATATGGCAGGCCTAGTTAATCTCTCCCACTGTTTCCTTT
ERV-Fc1_Human GCCACGGGGCTCGCCTTGTTAATATGGCAGGCCTAGTTAATCTCTCCCACTGTTTCCTTT
Consensus GCCATGGGGCTCGCCTTGTTAATATGGCAGGCCTAGTTAATCTCTCCCACTGTTTCCTTT
Active Sequence GCCATGGGGCTCGCCTTGTTAATATGGCAGGCCTAGTTAATCTCTCCCACTGTTTCCTTT
Macaco_AC210223 GAGAAGGGGCCCAGGTAGTACATATGGTCGGAGTGCACAACATCTCCCGTTGTTTCTTGT
6010 6020 6030 6040 6050 6060
Codons 231231231231231231231231231231231231231231231231231231231231
ERV-Fc1_Chimpanzee GCACCGCCCTCAGCCAACCACCACTAGTAGCTGTACCCCTACCCCAGGCTTTTAACACCT
ERV-Fc1_Gorilla GCACCGCCCTCGGCCAACCACCACTAGTAGCTGTACCCCTACCCCAGGCTTTTAACACCT
ERV-Fc1_Bonobo GCACCGCCCTCGGCCAACCACCACTAGTAGCTGTACCCCTACCCCAGGCTTTTAACACCT
ERV-Fc1_Human GCACCGCCCTCAGCCAACCACCACTAGTAGCTGTACCCCTACCCCAGGCTTTTAACACCT
Consensus GCACCGCCCTCRGCCAACCACCACTAGTAGCTGTACCCCTACCCCAGGCTTTTAACACCT
Active Sequence GCACCGCCCTCGGCCAACCACCACTAGTAGCTGTACCCCTACCCCAGGCTTTTAACACCT
Macaco_AC210223 GTGCAGCATTTAATAAGCCCCCACTGGTTGCGGTACCTTTACCTAGCCCTTTTAACTCCT
6070 6080 6090 6100 6110 6120
Codons 231231231231231231231231231231231231----------2312312312
ERV-Fc1_Chimpanzee CTGGTAACCACACTGCCCACCCTTCCGGCGTCTTCTCTG----------AGCAGGTCCC
ERV-Fc1_Gorilla CTGGTAACCACACTGCCCACCCTTCCGGCGTCTTCTCTG----------AGCAGGTCCC
ERV-Fc1_Bonobo CTGGTAACAACACTGCCCACCCTTCCGGCGTCTTCTCTG----------AGCAGGTCCC
ERV-Fc1_Human CTGGTAACCACACTGCCCACCCTTCCGGCGTCTTCTCTG----------AGCAGGTCCC
Consensus CTGGTAACCACACTGCCCACCCTTCCGGCGTCTTCTCTG----------AGCAGGTCCC
Active Sequence CTGGTAACCACACTGCCCACCCTTCCGGCGTCTTCTCTG----------AGCAGGTCCC
Macaco_AC210223 C---TAACCTAACC-CCCTCCTTTCC----CCTTCCCTGCCAAGCCCTGAGGGAAGTTCC


6130 6140 6150 6160 6170 6180
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee TCTTTTCCGAGACCCCCTCCAGCCCCAGTTCCCCTTCTGCTACACCACTCCTAACTCATC
ERV-Fc1_Gorilla TCTTTTCCGAGACCCCCTCCAGCCCCAGTTCCCCTTCTGCTACACCACTCCTAACTCATC
ERV-Fc1_Bonobo TCTTTTCCGAGACCCCCTCCAGCCCCAGTTCCCCTTCTGCTACACCACTCCTAACTCATC
ERV-Fc1_Human TCTTTTCCGAGACCCCCTCCAGCCCCAGTTCCCCTTCTGCTACACCACTCCTAACTCATC
Consensus TCTTTTCCGAGACCCCCTCCAGCCCCAGTTCCCCTTCTGCTACACCACTCCTAACTCATC
Active Sequence TCTTTTCCGAGACCCCCTCCAGCCCCAGTTCCCCTTCTGCTACACCACTCCTAACTCATC
Macaco_AC210223 TTTGTTCCAAGACCCGCTTAGACAACAACTCCCCTTTTGCTACTCCACCCCTAATGCTTC
6190 6200 6210 6220 6230 6240
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee CTGGTGCAACCAGACCTATTCTGGCTCCCTATCTAACCTCTCTGCACCGGCAGGTGGCT-
ERV-Fc1_Gorilla CTGGTGCAACCAGACCTATTCTGGCTCCCTATCTAACCTCTCTGCACCGGCAGGTGGCTA
ERV-Fc1_Bonobo CTGGTGCAACCAGACCTATTCTGGCTCCCTATCTAACCTCTCTGCACCGGCAGGTGGCT-
ERV-Fc1_Human CTGGTGCAACCAGACCTATTCTGGCTCCCTATCTAACCTCTCTGCACCGGCAGGTGGCT-
Consensus CTGGTGCAACCAGACCTATTCTGGCTCCCTATCTAACCTCTCTGCACCGGCAGGTGGCT-
Active Sequence CTGGTGCAACCAGACCTATTCTGGCTCCCTATCTAACCTCTCTGCACCGGCAGGTGGCT-
Macaco_AC210223 CTGGTGCAACCGGACAGGATCGGTGCCTCCA---AATCTCACCGCCCCCCCAGGCGAGT-
6250 6260 6270 6280 6290 6300
Codons -312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee ACTTCTGGTGTAACTTCACCCTTACAAAACATCTTAATATTTCCTCTAACAATACCCTTT
ERV-Fc1_Gorilla ACTTCTGGTGTAACTTCACCCTTACAAAACATCTTAATATTTCCTTTAACAATACCCTTT
ERV-Fc1_Bonobo ACTTCTGGTGTAACTTCACCCTTACAAAACATCTTAATATTTCCTCTAACAATACCCTTT
ERV-Fc1_Human ACTTCTGGTGTAACTTCACCCTTACAAAACATCTTAATATTTCCTCTAACAATACCCTTT
Consensus ACTTCTGGTGTAACTTCACCCTTACAAAACATCTTAATATTTCCTCTAACAATACCCTTT
Active Sequence ACTTCTGGTGTAACTTCACCCTTACAAAACATCTTAATATTTCCTCTAACAATACCCTTT
Macaco_AC210223 ATTTTTGGTGCAACTCCACCCTGTCAAAAACTCTCAAGGCTTCC---GACACTACTC---
6310 6320 6330 6340 6350 6360
Codons 231231231231231231231231231231231231231231231231231231
ERV-Fc1_Chimpanzee CTAGAAACTTATGCCTCCCCATCTCTCTGGTGCCTCGACTCACTCTGTACAGCGAGGCTG
ERV-Fc1_Gorilla CTGGAAACTTATGCCTCCCCATCTCTCTGGTGCCTCGACTCACTCTGTACAGCGAGGCTG
ERV-Fc1_Bonobo CTAGAAACTTATGCCTCCCCATCTCTCTGGTGCCTCGACTCACTCTGTACAGCGAGGCTG
ERV-Fc1_Human CTAGAAACTTATGCCTCCCCATCTCTCTGGTGCCTCGACTCACTCTGTACAGCGAGGCTG
Consensus CTAGAAACTTATGCCTCCCCATCTCTCTGGTGCCTCGACTCACTCTGTACAGCGAGGCTG
Active Sequence CTAGAAACTTATGCCTCCCCATCTCTCTGGTGCCTCGACTCACTCTGTACAGCGAGGCTG
Macaco_AC210223 ---------TATGCCTTCCCATCTCCCTAGTTCCCAGCCTCACCCTGTACAGTGAAGCGG
6370 6380 6390 6400 6410 6420
Codons 2312312312312312-----312312312312312312312312312312312312
ERV-Fc1_Chimpanzee AACTCTCTTCCCTTGT-----CAACCCGCCTATGCGTCAGAAGCGGGCCGTTTTCCCACC
ERV-Fc1_Gorilla AACTCTCTTCCCTTGT-----CAACCCGCCTATATGTCAGAAGCGGGCGGTTTTCCTACC
ERV-Fc1_Bonobo AACTCTCTTCCCTTGT-----CAACCCGCCTATGCGTCAGAAGCGGGCCGTTTTCCCACC
ERV-Fc1_Human AACTCTCTTCCCTTGT-----CAACCCGCCTATGCGTCAGAAGCGGGCCGTTTTCCCACC
Consensus AACTCTCTTCCCTTGT-----CAACCCGCCTATGCGTCAGAAGCGGGCCGTTTTCCCACC
Active Sequence AACTCTCTTCCCTTGT-----CAACCCGCCTATGCGTCAGAAGCGGGCCGTTTTCCCACC
Macaco_AC210223 AGCTATCCTCTCTTCTGCCCCTTGCCCGCCCC--CGCCAGGCAAGAGCAGTATTCCTTCC
6430 6440 6450 6460 6470 6480
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee GCTGGTAATAGGTGTCTCCTTGACCTCCTCACTTGTTGCCTCCGGGCTGGGCACAGGTGC
ERV-Fc1_Gorilla GCTGGTAATAGGTGTCTCCTTGACCTCCTCACTTGTTGCCTCCGGGCTGGGCACAGGTGC
ERV-Fc1_Bonobo GCTGGTAATAGGTGTCTCCTTGACCTCCTCACTTGTTGCCTCCGGGCTGGGCACAGATGC
ERV-Fc1_Human GCTGGTAATAGGTGTCTCCTTGACCTCCTCACTTGTTGCCTCCGGGCTGGGCACAGGTGC
Consensus GCTGGTAATAGGTGTCTCCTTGACCTCCTCACTTGTTGCCTCCGGGCTGGGCACAGGTGC
Active Sequence GCTGGTAATAGGTGTCTCCTTGACCTCCTCACTTGTTGCCTCCGGGCTGGGCACAGGTGC
Macaco_AC210223 GCTAATGATTGGAGTCTCCTTAGCCGCATCCTTCGTGGCCTCTGGCCTTGGAACTGGAGC


6490 6500 6510 6520 6530 6540
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee TATTGTACATTTCATAAGTTCTTCCCAAGATCTCTCTATTAAGCTCCAGATGGCCATCGA
ERV-Fc1_Gorilla TATTGTACATTCCATAAGCTCTTCCCAAGATCTCTCTATTAAGCTCCAGATGGCCATCGA
ERV-Fc1_Bonobo TATTGTACATTTCATAAGCTCTTCCCAAGATCTCTCTATTAAGCTCCAGATGGCCATCGA
ERV-Fc1_Human TATTGTACATTTCATAAGCTCTTCCCAAGATCTCTCTATTAAGCTCCAGATGGCCATCGA
Consensus TATTGTACATTTCATAAGCTCTTCCCAAGATCTCTCTATTAAGCTCCAGATGGCCATCGA
Active Sequence TATTGTACATTTCATAAGCTCTTCCCAAGATCTCTCTATTAAGCTCCAGATGGCCATCGA
Macaco_AC210223 CATAACTCATCCTGTCCCATCCTCTCAAGATCTTTCAGCCCGATTACAGGTAGCAATCGA
6550 6560 6570 6580 6590 6600
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee AGCCTCAGCCGAATCCTTAGCCTCTCTACAGAGACAGATTACGTCTGTGGCCAAGGTGGC
ERV-Fc1_Gorilla AGCCTCAGCCGAATCCTTAGCCTCTCTACAGAGACAGATTACGTCTGTGGCCAAGGTGGC
ERV-Fc1_Bonobo AGCCTCAGCCGAATCCTTAGCCTCTCTACAGAGACAGATTACGTCTGTGGCCAAGGTGGC
ERV-Fc1_Human AGCCTCAGCCGAATCCTTAGCCTCTCTACAGAGACAGATTACGTCTGTGGCCAAGGTGGC
Consensus AGCCTCAGCCGAATCCTTAGCCTCTCTACAGAGACAGATTACGTCTGTGGCCAAGGTGGC
Active Sequence AGCCTCAGCCGAATCCTTAGCCTCTCTACAGAGACAGATTACGTCTGTGGCCAAGGTGGC
Macaco_AC210223 AGCTTCGGCAGAATCCCTGGCTTCCCTCCAACGACAGATCACCTCAGTGGCCCAGGTGGC
6610 6620 6630 6640 6650 6660
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee CATGCAGAACCGGAGAGCCCTAGATCTCCTCACAGCCGACAAAGGCGGAACCTGCATGTT
ERV-Fc1_Gorilla CATGCAGAACCGGAGGGCCCTAGATCTCCTCACAGCCGACAAAGGCGGAACCTGCATGTT
ERV-Fc1_Bonobo CATGCAGAACCGGAGAGCCCTAGATCTCCTCACAGCCGACAAAGGCGGAACCTGCATGTT
ERV-Fc1_Human CATGCAGAACCGGAGAGCCCTAGATCTCCTCACAGCCGACAAAGGCGGAACCTGCATGTT
Consensus CATGCAGAACCGGAGAGCCCTAGATCTCCTCACAGCCGACAAAGGCGGAACCTGCATGTT
Active Sequence CATGCAGAACCGGAGAGCCCTAGATCTCCTCACAGCCGACAAAGGCGGAACCTGCATGTT
Macaco_AC210223 AGCCCAGAATCGCAGAGCACTCGACCTCCTTATGGCTGACAAAGGCGGGACCTGTATGTT
6670 6680 6690 6700 6710 6720
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee TCTCGGGGAAGAGTGCTGTTATTACATCAATGAATCAGGCTTAGTAGAAACCAGCCTCCT
ERV-Fc1_Gorilla TCTCGGGGAAGAGTGCTGTTATTACATCAATGAAGCAGGCTTAGTAGAAACCAACCTCCT
ERV-Fc1_Bonobo TCTCGGGGAAGAGTGCTGTTATTACATCAATGAATCAGGCTTAGTAGAAACCAGCCTCCT
ERV-Fc1_Human TCTCGGGGAAGAGTGCTGTTATTACATCAATGAATCAGGCTTAGTAGAAACCAGCCTCCT
Consensus TCTCGGGGAAGAGTGCTGTTATTACATCAATGAATCAGGCTTAGTAGAAACCAGCCTCCT
Active Sequence TCTCGGGGAAGAGTGCTGTTATTACATCAATGAATCAGGCTTAGTAGAAACCAGCCTCCT
Macaco_AC210223 CCTTAATGAAGAATGCTGCTACTACATCAATGAGTCAGGACTAGTAGAAACCAACCTCCT
6730 6740 6750 6760 6770 6780
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee CACCCTTGATAAAATCCGGGACGGTCTCCATCGACCCTCCTCAACTCCCAACTATGGAGG
ERV-Fc1_Gorilla CACCCTTGATAAAATCCGGGATGGTCTCCATCGACCCTCCTCAACTCCCGACTATGGAGG
ERV-Fc1_Bonobo CACCCTTGATAAAATCCGGGACGGTCTCCATCGACCCTCCTCAACTCCCAACTATGGAGG
ERV-Fc1_Human CACCCTTGATAAAATCCGGGACGGTCTCCATCGACCCTCCTCAACTCCCAACTATGGAGG
Consensus CACCCTTGATAAAATCCGGGACGGTCTCCATCGACCCTCCTCAACTCCCAACTATGGAGG
Active Sequence CACCCTTGATAAAATCCGGGACGGTCTCCATCGACCCTCCTCAACTCCCAACTATGGAGG
Macaco_AC210223 CACTCTAGAAAAAGTCCCAGAAGGACTCCACAAAAAAACCTCAGGATTAGAGTCTTCATT
6790 6800 6810 6820 6830 6840
Codons 312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee AGGGTGGTGGCAATCCCCTTTAACCACTTGGATTATCCCTTTCATAAGCCCCATCCTAAT
ERV-Fc1_Gorilla AGGGTGGCCGCAATCCCCTTTAACCACTTGGATTATCCCTTTCATAAGCCCCATCCTAAT
ERV-Fc1_Bonobo AGGGTGGTGGCAATCCCCTTTAACCACTTGGATTATCCCTTTCATAAGCCCCATCCTAAT
ERV-Fc1_Human AGGGTGGTGGCAATCCCCTTTAACCACTTGGATTATCCCTTTCATAAGCCCCATCCTAAT
Consensus AGGGTGGTGGCAATCCCCTTTAACCACTTGGATTATCCCTTTCATAAGCCCCATCCTAAT
Active Sequence AGGGTGGTGGCAATCCCCTTTAACCACTTGGATTATCCCTTTCATAAGCCCCATCCTAAT
Macaco_AC210223 CGGGTGGTGGCAGTCGTCCATGGCTAGCTGGGTCCTACCCTTCCTAAGCCCCTTACTAAT

```
                          6850        6860        6870        6880        6890        6900
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|
Codons                    312312312312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee        CATTTGCCTTTTACTTCTCATAGCCCCCTGTGTCCTCAAGTTCATCAAAAACCGCATCAG
ERV-Fc1_Gorilla           CATTTGCCTTTTACTTCTCATAGCCCCCTGTGTCCTCAAGTTCATCAAAAACCGCATCAG
ERV-Fc1_Bonobo            CATTTGCCTTTTACTTCTCATAGCCCCCTGTGTCCTCAAGTTCATCAAAAACTGCATCAG
ERV-Fc1_Human             CATTTGCCTTTTACTTCTCATAGCCCCCTGTGTCCTCAAGTTCATCAAAAACCGCATCAG
Consensus                 CATTTGCCTTTTACTTCTCATAGCCCCCTGTGTCCTCAAGTTCATCAAAAACCGCATCAG
Active Sequence           CATTTGCCTTTTACTTCTCATAGCCCCCTGTGTCCTCAAGTTCATCAAAAACCGCATCAG
Macaco_AC210223           CATTGGCTTCTTACTGCTCAGAGCTCCCTGTATCATCCGCTTCATCCAGAACCGCATAAA

                          6910        6920        6930        6940        6950        6960
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|
Codons                    312312312312312312312312312312312312312312312312312312312312312
ERV-Fc1_Chimpanzee        CGAAGTCTCCCGGGTGACGGTCAACCAAATGTTACTACACCCTTACTCTCGTCTTCCGAC
ERV-Fc1_Gorilla           CGAAGTCTCCTGGGTGACGGTCAACCAAATGTTACTACACCCTTACTCCCGTCTTCCGAC
ERV-Fc1_Bonobo            CGAAGTCTCCCGGGTGACGGTCAACCAAATGTTACTACACCCTTACTCCCGTCTTCCGAC
ERV-Fc1_Human             CGAAGTCTCCCGGGTGACGGTCAACCAAATGTTACTACACCCTTACTCCCGTCTTCCGAC
Consensus                 CGAAGTCTCCCGGGTGACGGTCAACCAAATGTTACTACACCCTTACTCCCGTCTTCCGAC
Active Sequence           CGAAGTCTCCCGGGTGACGGTCAACCAAATGTTACTACACCCTTACTCCCGTCTTCCGAC
Macaco_AC210223           AGAAGTCTCCCGGATCACTTTCAATCAGATGCTGCTCCACCCCTATACCCGAGTTCTGAC

                          6970        6980        6990        7000        7010        7020
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|
Codons                    312312312312312312312312312312312312312312312ENVe........
ERV-Fc1_Chimpanzee        CTCCGAAGACCACTATGACGTCGCCCTCACTCAGCAGGAAGCAGCCAGATGATTACGTCG
ERV-Fc1_Gorilla           CTCCGAAGACCACTATGACGATGCCCTCACTCAGCAGGAAGCAGCCAGATGATTACGTCG
ERV-Fc1_Bonobo            CTCCGAAGACCACTATGACGACGCCCTCACTCAGCAGGAAGCAGCCAGATGATTATGTCG
ERV-Fc1_Human             CTCCGAAGACCACTATGACGACGCCCTCACTCAGCAGGAAGCAGCCAGATGATTACGTCG
Consensus                 CTCCGAAGACCACTATGACGACGCCCTCACTCAGCAGGAAGCAGCCAGATGATTACGTCG
Active Sequence           CTCCGAAGACCACTATGACGACGCCCTCACTCAGCAGGAAGCAGCCAGATGATTACGTCG
Macaco_AC210223           CTCCGAAAAACCCCACTACGA---CCCAAAACAGCAGGAAGCAGCCAGATGAACAGGTCG

                          7030        7040        7050        7060
                          ....|....|....|....|....|....|....|....|..
Codons                    ..........................................
ERV-Fc1_Chimpanzee        CCCCTTTTTCTTA-CAGTATGAGGTCGGAATGTTAGGCAGG
ERV-Fc1_Gorilla           CCCCTTTTTCTTA-TAGTATGAGGTCGGAATGTTAGGCAGG
ERV-Fc1_Bonobo            CCCCTTTTTCTTA-CAGTATGAGGTCAGAATGTTAGGCAGG
ERV-Fc1_Human             CCCCTTTTTCTTA-CAGTATGAGGTCGGAATGTTAGGCAGG
Consensus                 CCCCTTTTTCTTA-CAGTATGAGGTCGGAATGTTAGGCAGG
Active Sequence           CCCCTTTTTCTTA-CAGTATGAGGTCGGAATGTTAGGCAGG
Macaco_AC210223           CCCGTTTTTCTTATTAGAAAGAGGTAGGAATGTTAGGCAGG
```