

**MASTER EN
MICROBIOLOGÍA Y PARASITOLOGÍA: I+D**

TRABAJO FIN DE MÁSTER



**UNIVERSIDAD COMPLUTENSE
MADRID**

ALUMNA: PALOMA TROYANO HERNÁEZ

TÍTULO DEL TRABAJO:

“MARCADORES GENÉTICOS Y DOMINIOS CONSERVADOS EN LA PROTEÍNA DE LA CÁPSIDE DE LAS DISTINTAS VARIANTES DEL VIH”.

“GENETIC MARKERS AND CONSERVED DOMAINS OF THE CAPSID PROTEIN ACROSS HIV VARIANTS”

CURSO ACADÉMICO: 2018-2019

CONVOCATORIA: JUNIO



Directora de trabajo: Dra. África Holguín

Tutor UCM: María Molina

Curso académico: 2018-2019

Periodo de realización: Enero – Junio 2019

Lugar de realización: Laboratorio de Epidemiología Molecular del VIH-1. Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS). Servicio de Microbiología y Parasitología. Hospital Universitario Ramón y Cajal. Madrid.

ÍNDICE

RESUMEN	2
INTRODUCCIÓN	3
1.1 Origen y epidemiología del VIH.....	3
1.2 Diversidad genética del VIH.....	3
1.3 Clasificación del VIH.....	3
1.4 Estructura del VIH y ciclo replicativo.....	4
1.5 Estructura del genoma de HIV-1 y gen <i>gag</i>	5
1.6 Proteína p24/p26 (CA) y cápside:.....	6
JUSTIFICACIÓN Y OBJETIVOS	7
MATERIALES Y MÉTODOS	7
3.1 Descarga, alineamiento y traducción de secuencias de la proteína CA del VIH.....	7
3.2 Evaluación de la función de conservación de secuencias de aminoácidos en la herramienta bioinformática EpiMolVIH.....	7
3.3 Identificación de los dominios conservados en la proteína <i>gag</i> p24/p26 (CA) en las distintas variantes del VIH-1 y VIH-2.....	8
3.4 Generación de secuencias consenso.....	8
3.5 Identificación de marcadores genéticos en CA específicos de cada variante del.....	9
3.6 Estructura secundaria de la cápside viral.....	10
3.7 Mutaciones que afectan el ensamblaje de la cápside inmadura.....	10
3.8 Naturaleza de los aa en marcadores de variante y en conservación de residuos clave ..	10
3.9 Clasificación de variantes del VIH empleando análisis filogenético con secuencias CA	10
RESULTADOS	11
4.1 Secuencias analizadas.....	11
4.2 Conservación de CA entre variantes del VIH.....	11
4.3 Conservación de la estructura secundaria de la proteína de la cápside viral.....	14
4.3.1 Región Mayor de Homología (MHR).....	14
4.3.2 <i>Loop</i> de unión a ciclofilina A.....	14
4.3.3 Poro del hexámero.....	17
4.4 Marcadores específicos de variante en la proteína de la cápside del VIH.....	17
4.4.1 Marcadores de VIH-2.....	17
4.4.2 Marcadores de Grupo de VIH-1.....	17
4.4.3 Marcadores de las distintas variantes del grupo M del VIH-1 y su presencia en estructuras secundarias.....	18
4.5. Mutaciones que afectan el ensamblaje de la cápside inmadura:.....	19
4.6 Clasificación de variantes del VIH empleando análisis filogenético con secuencias CA	20
DISCUSIÓN	21
CONCLUSIONES	23
BIBLIOGRAFÍA	24

RESUMEN

La proteína CA de gag que forma la cápside del VIH posee un papel estructural y funcional esencial para el ciclo viral, resultando un interesante objetivo para el diseño de test moleculares diagnósticos y nuevas drogas antirretrovirales. En este trabajo estudiamos en detalle la conservación de esta proteína y de los elementos de su estructura secundaria. Localizamos los marcadores genéticos asociados a cada variante del VIH-1 y VIH-2 y los polimorfismos en aminoácidos clave para su función descritos en la literatura. También evaluamos la utilidad de secuencias de la CA para clasificar correctamente las variantes del VIH por análisis filogenético. Para todo ello se descargaron 39.010 secuencias de la proteína CA de la base de datos del Laboratorio Nacional de los Álamos de EE.UU de todas las variantes del VIH (tipos, grupos, subtipos, sub-subtipos y CRFs) y se analizaron utilizando una nueva herramienta bioinformática desarrollada en el laboratorio y verificada durante este estudio. CA presentó una alta conservación en las 105 variantes analizadas, encontrando estructuras menos conservadas como la horquilla- β , la hélice- α 6 y el *loop* de unión a ciclofilina A. Se encontraron marcadores genéticos de variante en todos los tipos, subtipos, sub-subtipos y recombinantes del VIH. Se hallaron mutaciones en posiciones relevantes para la funcionalidad de estructuras destacadas de CA y en residuos altamente conservados cuyos polimorfismos se asocian a defectos del ensamblaje de la cápside inmadura según la literatura. Por último, se constató que CA es una buena región para caracterizar gran parte de las variantes del VIH.

ABSTRACT

The CA gag protein that forms the HIV capsid has an essential structural and functional role for the viral cycle, being an interesting objective for the design of diagnostic molecular tests and new antiretroviral drugs. In this work we studied in detail the conservation of this protein and of the elements in its secondary structure. We located the genetic markers associated with each HIV-1 and HIV-2 variant and the polymorphisms in key aminoacids for its functions described in the literature. We also evaluated its usefulness to correctly classify HIV variants by phylogenetic analysis. For this purpose, we downloaded 39.010 CA sequences from all HIV variants (types, groups, subtypes, sub-subtypes and CRFs) from the US Los Alamos HIV database and we analyzed them using a new bioinformatics tool developed in the laboratory and verified during this study. CA presented a high conservation in the 105 analyzed HIV variants. We identified structures with lower conservation such as the β hairpin, the α helix 6 and the cyclophilin A binding loop. HIV-1 variant associated genetic markers were found in all types, subtypes, sub-subtypes and HIV CRFs. We found mutations in CA residues relevant for the functionality of CA structures and in highly conserved residues whose polymorphisms have been associated in the literature to defective assembly of the immature capsid. Finally, we found that CA coding region is a good region to characterize a large number of HIV variants.

INTRODUCCIÓN

1.1 Origen y epidemiología del VIH

El virus de la inmunodeficiencia humana (VIH), descubierto en 1983, es el agente etiológico del Síndrome de Inmunodeficiencia Adquirida (SIDA) en humanos. Es un virus ARN del género *Lentivirus* y de la familia *Retroviridae*. El VIH presenta su origen en múltiples transmisiones zoonóticas de virus de la inmunodeficiencia de simios (VIS) desde primates a humanos en África central y occidental. Uno de estos eventos de transmisión a partir del VIS de chimpancés en el Sudeste de Camerún, dio lugar al VIH-1 del grupo M, principal causante de la pandemia de SIDA del siglo XX (American Society for Microbiology and Versalovic, 2011). Según el informe de 2018 del Programa Conjunto de las Naciones Unidas sobre el VIH/Sida (ONUSIDA), en 2017 casi 37 millones de personas vivían con el VIH, 1,8 millones contrajeron la infección ese año y 940.000 fallecieron a causa de enfermedades relacionadas con el SIDA (UNAIDS, 2018).

1.2 Diversidad genética del VIH

El VIH presenta una rápida evolución y alta heterogenicidad genética que le permite adaptarse a nuevos hospedadores y presiones selectivas. Su origen se debe a la elevada frecuencia de mutación y de recombinación durante la actividad de la retrotranscriptasa (RT) viral y a la alta tasa de replicación del virus. Así, la RT carece de actividad exonucleasa 3'-5' correctora y, por tanto, es incapaz de eliminar los nucleótidos incorrectamente incorporados durante la retrotranscripción, dando lugar a la generación continua de variantes. Esta diversidad genética presenta importantes implicaciones en la patogenia, transmisión, inmunogenicidad y tratamiento del VIH (Hemelaar, 2013) y le confiere una estructura poblacional de "cuasiespecie viral", entendida como un espectro de mutantes constituido por distintas variantes del virus con genomas relacionados, pero no idénticos, que infectan a un mismo individuo (Domingo *et al.*, 2012).

1.3 Clasificación del VIH

El VIH se clasifica en dos tipos: VIH-1 y VIH-2. Como resultado de la alta variabilidad genética, el VIH-1 se ha clasificado en base a la homología genética de secuencias génicas completas o parciales en cuatro grandes grupos: M (*main* o principal), O (*outlier*), N (no M, no O), y P, del que casi no se han descrito infecciones salvo casos aislados en Camerún. Los virus de los grupos O y N del VIH-1 presentan baja prevalencia, siendo mayoritariamente encontrados en África subsahariana occidental. Prácticamente la mayoría de infecciones por VIH-1 están causadas por variantes del grupo M, que ha sido subdividido en 9 subtipos (A-D, F-H, J, K), varios sub-subtipos (A1-A6 dentro del subtipo A; y F1 y F2 dentro del subtipo F), 98 formas recombinantes circulantes (CRFs) e innumerables formas recombinantes únicas (URFs) (Robertson *et al.*, 2000) (**Figura 1**).

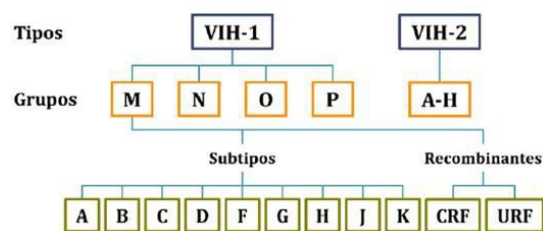


Figura 1. Clasificación de las variantes del VIH

Aunque el VIH-2 también es capaz de causar SIDA, presenta importantes diferencias epidemiológicas, geográficas y clínicas con el VIH-1: es menos patogénico, presentando períodos de latencia clínica más largos, una progresión más lenta y menor carga viral en individuos asintomáticos. A su vez, las tasas de transmisión horizontal y vertical son menores. El VIH-2 es resistente de manera natural a los inhibidores de la RT no análogos de nucleósidos y es más cercano filogenéticamente al VIS que al VIH-1. Actualmente causa de 1 a 2 millones de infecciones, con altas prevalencias en África Occidental. La organización genómica del VIH-2 es similar a la del VIH-1, pero su regulación genética es diferente. El VIH-2 se clasifica en 9 grupos (A-I), representando cada grupo un único evento de transmisión del virus entre especies, y 2 formas recombinantes: una forma recombinante circulante o CRF (CRF01_AB) y una forma recombinante única o URF (Visseaux *et al.*, 2016).

1.4 Estructura del VIH y ciclo replicativo

El VIH presenta forma esférica, con un diámetro de 100-120 nanómetros. El virión posee tres capas: una envoltura externa lipídica, procedente de la membrana de la célula humana infectada, con las glicoproteínas virales gp120 y gp41 en forma de trímeros, la matriz proteica (MA) y la cápside icosaédrica (CA), dentro de la cual se encuentran las enzimas virales: RT, proteasa (PR) e integrasa (IN), la nucleocápside (NC) y dos copias de ARN viral de polaridad positiva en su interior (**Figura 2A**).

El ciclo replicativo del virus puede dividirse en varias etapas (**Figura 2B**): (1) Interacción inicial entre el virión, mediante la glicoproteína viral gp120, y la célula diana (principalmente linfocitos T CD4) (2) Fusión de la envuelta del virión con la membrana celular y decapsidación del genoma vírico. (3) Transcripción inversa y síntesis de ADN proviral (4) Transporte al núcleo del ADN proviral acoplado a factores celulares y virales en el complejo de preintegración, e integración en el genoma hospedador gracias a la IN y a las secuencias largas repetidas (LTR) virales que hay en el extremo del genoma viral. (5) Transcripción del ADN proviral a ARNm viral por la ARN polimerasa celular y traducción a poliproteínas precursoras. (6) Procesamiento y ensamblaje de las proteínas virales y salida de los viriones inmaduros por gemación. (7) Maduración por acción de la PR viral de las poliproteínas *gag* y *gag-pol* originando viriones maduros infectivos (Alcamí and Coiras, 2011).

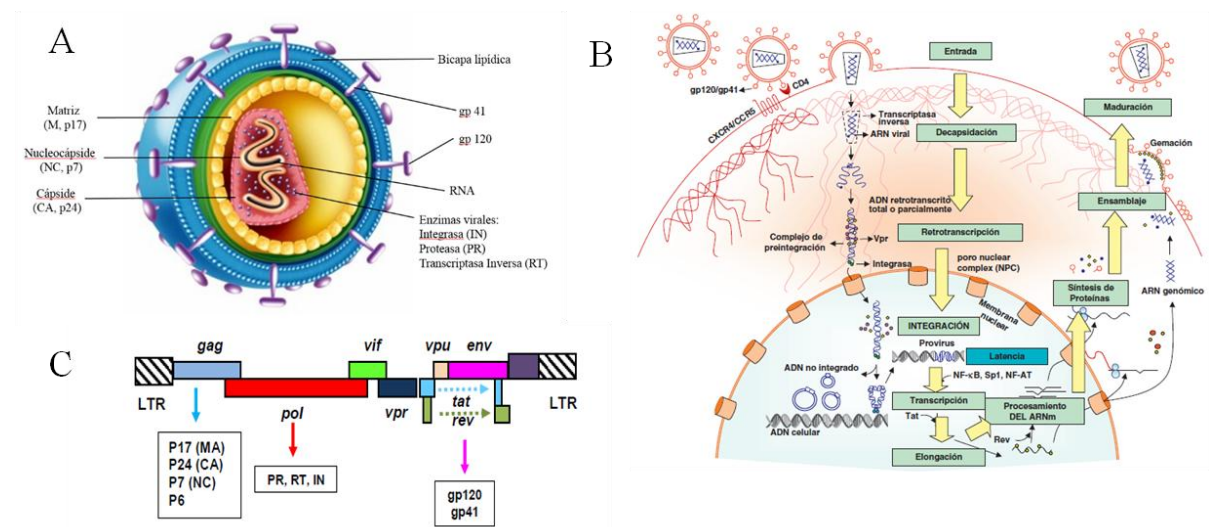


Figura 2.A. Estructura del VIH.

Adaptada de <https://www.encyclopediasalud.com/categorias/enfermedades/articulos/el-virus-del-sida>.
B. Ciclo biológico del VIH. **C.** Genoma del VIH. Tomada de Alcamí and Coiras, 2011.

1.5 Estructura del genoma de HIV-1 y gen gag

El genoma del VIH está constituido por tres genes principales comunes a todos los retrovirus: *gag*, *pol* y *env* (**Figura 2C**). El gen *gag* codifica principalmente proteínas estructurales. El gen *pol* codifica la PR, RT e IN virales y *env* codifica al precursor de las glicoproteínas gp120 y gp41. Además, existen seis genes reguladores virales: *tat*, *rev*, *vif*, *vpr*, *nef*, *vpu* (sólo en VIH-1) y *vpx* (sólo en VIH-2). En los extremos 5' y 3' se localizan las LTR no codificantes. El gen *gag* codifica las proteínas implicadas en estructura y tráfico viral, ensamblaje, control de las proteínas *pol*, interacción con proteínas celulares y salida viral. Por tanto, *gag* determina la estructura y funciones enzimáticas en el VIH. Las principales proteínas estructurales *gag* son p17 o MA, p24 o CA (denominada p26 en VIH-2) y p7 o NC (**Tabla 1**). Estas proteínas, P6 *gag* y dos espaciadores (P1 y P2) se sintetizan a partir de una serie de reacciones proteolíticas mediadas por la PR en unos sitios de corte específicos llamados sitios de procesamiento proteico o *cleavage sites* (CS) ubicados en las poliproteínas precursoras *gag* (Pr55*gag*) y *gagpol* (Pr160*gagpol*) (**Figura 3**). Ésta última poliproteína se genera gracias a un marco de lectura ribosomal adicional durante la traducción del precursor *gag* (Torrecilla *et al.*, 2014).

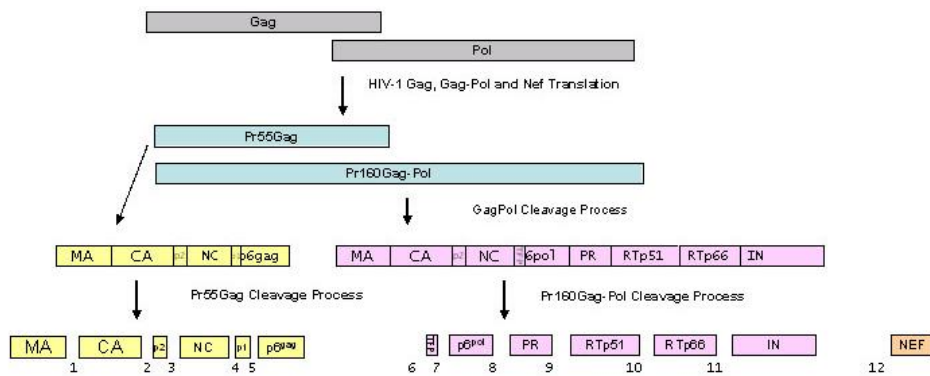


Figura 3. Procesamiento de *gagpol*. Adaptada de <http://www.bioafrica.net/proteomics/HIVcleavagesites.html>

Tabla 1. Longitud, localización y función de las proteínas *gag* del VIH-1 generadas durante la maduración viral

HIV-1	Localización HXB2	HIV-2	Localización BEN	Función	Detalles
p17 (MA) 131aa	2-132	p15 (MA) 134 aa	2-135	Estructural	Formación de la MA. Estabiliza la envoltura. Acompaña al ADN al interior del núcleo antes de su integración.
p24 (CA) 231 aa	133-363	p26 (CA) 230 aa	136-365	Estructural	Formación de la cápside viral.
p2 14aa	378-432	p2 17 aa	366-382	Reguladora	Ensamblaje y morfogénesis. Estabilidad del dímero ARN-VIH. Infectividad viral.
p7 (NC) 55aa	378-432	p8 (NC) 49 aa	383-431	Estructural y reguladora	Formación de la NC. Reconocimiento viral e incorporación del ARN al nuevo virión.
p1 16 aa	433-448	p1 14 aa	432-445	Reguladora	Regulación del número de proteínas <i>gag</i> y <i>pol</i> .
p6 52 aa	449-500	p6 76 aa	446-521	Reguladora	Incorporación de <i>vpr</i> al virión.

Adaptado de HIV *Sequence Compendium* 2018 Los Alamos y Uniprot: <https://www.uniprot.org.html>

1.6 Proteína p24/p26 (CA) y cápside:

En este PFM estudiaremos en detalle los marcadores genéticos asociados a cada variante del VIH y los dominios conservados en una de las proteínas *gag*: la p24 o CA, que forma la cápside viral. Se ubica entre los nucleótidos 1186 y 1879 del aislado subtipo B HXB2 del VIH-1. Esta proteína, con 231 aminoácidos (aa), tiene un dominio N-terminal (NTD) de 145 aa con una horquilla β y 7 hélices α (H), así como un dominio C-terminal (CTD) de 85 aa con 4 hélices α y una región sin estructura de 11 residuos. Ambas regiones están unidas por una Región de Unión Interdominio (RUI) (aa 146-150). También presenta una Región de Homología Mayor o MHR (aa 153-172) en el CTD con 20 aa altamente conservados y un *loop* (aa 85-93) en el NTD que se une a la ciclofilina A (**Figura 4A**). La cápside madura está formada por aproximadamente 1.100 monómeros de P24 ensamblados en una matriz de hexámeros con 12 variaciones pentaméricas, donde NTD es responsable de los contactos intra-hexaméricos y CTD forma dímeros de unión a los hexámeros adyacentes (**Figura 4B**) (Rihn *et al.*, 2013). En el centro de cada hexámero hay un poro rodeado por seis residuos de arginina cargados positivamente. El poro está cubierto por la horquilla β que puede cambiar de conformación para abrirlo o cerrarlo. Una molécula IP6 (inositol-hexakisfosfato) se une al centro del poro estabilizando el hexámero (Jaques *et al.*, 2016, Obr and Kräusslich, 2018) (**Figura 4C**).

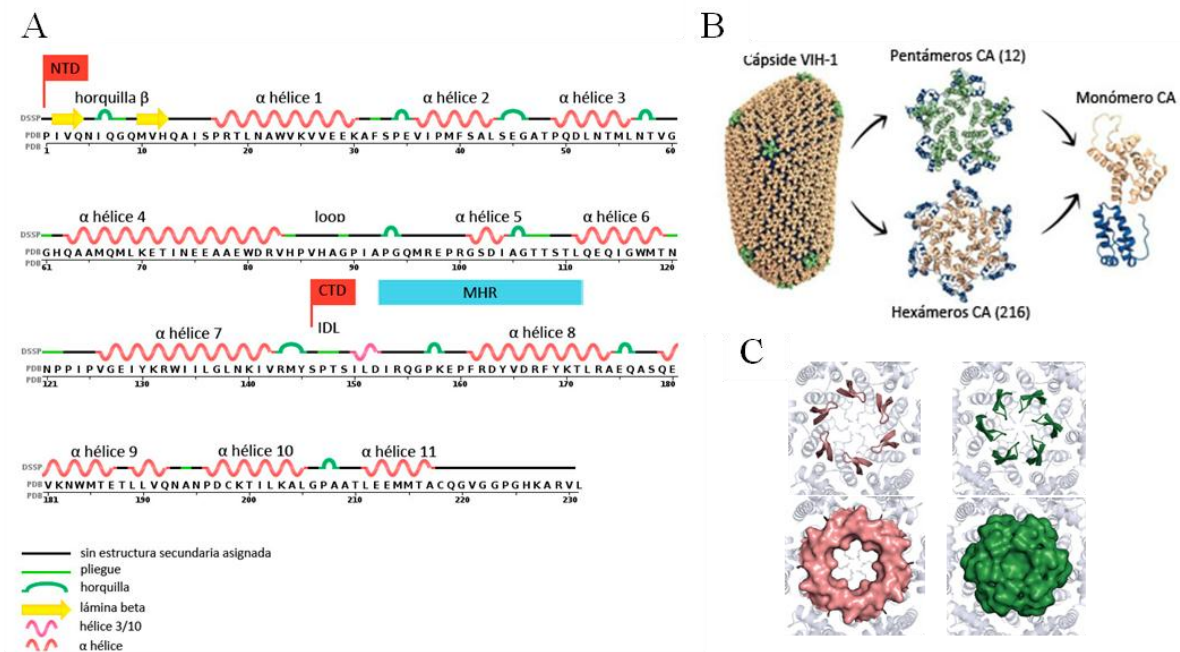


Figura 4.A. Estructura secundaria de la cápside viral. Adaptada de RCSB *Protein Data Bank*

(<http://www.rcsb.org.html>). **B.** Estructura terciaria de la cápside viral. Adaptada de Quinn *et al.*, 2018. **C.**

Horquilla β en la conformación de poro abierto (rosa) y poro cerrado (verde). Tomada de Jaques *et al.*, 2016

Estudios previos de p24 con mutantes generados por mutagénesis dirigida han revelado que la cápside es extremadamente intolerante a las sustituciones no sinónimas, produciendo virus defectivos o de infectividad reducida (Perilla and Gronenborn, 2016, Rihn *et al.*, 2013). Ello le hace una región muy interesante para el diseño de test moleculares diagnósticos que detecten precozmente al VIH en niños y adultos infectados. A su vez, dado que es tanto un componente estructural vital para el VIH como un importante participante de funciones básicas del virus, tales como la retrotranscripción, resulta un interesante objetivo para el diseño de nuevas drogas antirretrovirales (Jaques *et al.*, 2016, Rihn *et al.*, 2013).

JUSTIFICACIÓN Y OBJETIVOS

En este trabajo se plantearon los siguientes objetivos:

1. Comprobar el correcto funcionamiento de una nueva herramienta bioinformática desarrollada en el laboratorio de Epidemiología Molecular del VIH para identificar regiones conservadas del VIH a partir de secuencias del virus.
2. Identificar los dominios conservados en la proteína *gag* p24 en las distintas variantes del VIH (tipos, grupos, subtipos, sub-subtipos y recombinantes CRF) con secuencia disponible en LANL a fecha de estudio y en los elementos de la estructura secundaria de la cápside.
3. Identificar los polimorfismos naturales o marcadores genéticos asociados a cada una de las variantes del VIH en la proteína *gag*p24 (VIH-1) y p26 (VIH-2).
4. Analizar el grado de conservación, en las distintas variantes del VIH-1, de aminoácidos de p24 previamente identificados como esenciales para su estructura y ensamblaje viral.
5. Confirmar la utilidad de las secuencias de *gag* p24 para clasificar correctamente las variantes del VIH-1 por análisis filogenético e identificar las variantes que peor se definen en esta región.

MATERIALES Y MÉTODOS

3.1 Descarga, alineamiento y traducción de secuencias de la proteína CA del VIH

Entre enero y marzo de 2019 se descargaron todas las secuencias de la proteína *gag* CA en nt y formato *fasta* de las variantes del VIH (tipos, grupos, subtipos, sub-subtipos y CRFs) disponibles a fecha de estudio en la base de datos del Laboratorio Nacional de los Álamos de EE.UU (LANL: <http://www.hiv.lanl.gov/>), que recoge las secuencias depositadas en *GenBank*. Tras su descarga, las secuencias se alinearon y tradujeron a aa con el programa *MEGA v6.06* (*Molecular Evolutionary Genetics Analysis*: <https://www.megasoftware.net/>), utilizando el programa *MUSCLE* (*Multiple Sequence Comparison by Log Expectation*) para alinear los aa de CA en cada variante. Tras el alineamiento, se emplearon las secuencias de los aislados de referencia BEN de VIH-2 (subtipo A de VIH-2, nº acceso del GenBank U38293) y HXB2 de VIH-1 (subtipo B de VIH-1, nº acceso del GenBank K03455) para verificar la localización de los aa de CA en cada tipo del virus. Las variantes con menos de 3 secuencias se descartaron para el análisis posterior de conservación y marcadores.

3.2 Evaluación de la función de conservación de secuencias de aminoácidos en la herramienta bioinformática EpiMolVIH

Para realizar el análisis de la conservación de los aa de CA en las distintas variantes de VIH-1 y VIH-2 se utilizó la función “Conservación” de una nueva herramienta bioinformática (EpiMolVIH) desarrollada y programada en el laboratorio por el experto Roberto Reinoso Fernández. La finalidad de esta herramienta bioinformática es el análisis de secuencias del VIH-1 y VIH-2 en formato *fasta*, tanto para nt como para aa. El programa está desarrollado en el lenguaje de programación Java usando el *openjdk* en su versión 1.8.0_212 y está disponible tanto a 32 como a 64 bits. La razón principal del uso de Java es ser multiplataforma y funcionar más rápido que otras opciones con dichas características. Esta herramienta actualmente sirve para analizar mutaciones de resistencias a distintas familias de fármacos antirretrovirales, para identificar polimorfismos frente a una secuencia deseada, y para analizar conservación a nivel de aa y nts. El software genera tablas y listas donde se presentan los datos de forma clara y concisa. Actualmente el programa se encuentra en constante

desarrollo, y se van añadiendo nuevas funciones y formas de presentación de los datos según las necesidades del laboratorio EpiMolVIH.

La función de conservación, recientemente programada, reconoce los alineamientos de secuencias de aa o nt, introducidas en formato *fasta* previo alineamiento y traducción en *MEGA v6.06*. Calcula el grado de conservación para cada residuo según su posición en la proteína, teniendo en cuenta los gaps en la secuencia y genera un listado o tabla con la secuencia consenso compuesta por el residuo o nucleótido más frecuente en cada posición de las secuencias alineadas y el porcentaje de conservación de éste, siempre y cuando sea mayor del 50% (**Figura 5**).

VARIANTE	HIV2_A_p24.fas	HIV2_A_p24.fas	HIV2_A_p24.fas	HIV2_A_p24.fas	HIV2_A_p24.fas	HIV2_A_p24.fas
POSICIÓN	1	2	3	4	5	6
AA	P	V	Q	Q		G
CONSERVACIÓN	100,00%	98,80%	98,80%	57,40%		64,20%
TOTAL	162	162	162	162	162	162; 1 gap; 1?

Figura 5. Ejemplo de tabla de conservación generada por el programa EpiMolVIH.

Para verificar el correcto funcionamiento de dicha herramienta bioinformática, durante abril se estableció una estrategia de verificación en dos pasos. Primero se procedió al diseño de 202 secuencias de aa de diferentes longitudes empleando el programa *MEGA v6.06*, con conservación conocida para cada residuo del alineamiento de secuencias introducidas, incluyendo también gaps (“-“), aa no identificados tras la traducción (“?”) y “errores” generados como símbolos que el programa no debería considerar: “(“, “)”, “X”, “%”. Después se usaron para comprobar los resultados de la función “Conservación” del programa EpiMolVIH. En segundo lugar, se volvió a comprobar la función utilizando otros dos grupos de secuencias de la CA del VIH: 188 secuencias de HIV-2 y 31618 secuencias de HIV-1 descargadas de LANL y con conservación conocida en posiciones de aa concretas. Una vez comprobado el correcto funcionamiento del programa para analizar el grado de conservación, se usó para analizar las secuencias reales de CA de la base de datos de Los Álamos.

3.3 Identificación de los dominios conservados en la proteína gag p24/p26 (CA) en las distintas variantes del VIH-1 y VIH-2

Se buscaron los dominios conservados en la secuencia de CA de las distintas variantes del VIH-1 y VIH-2 empleando la función “Conservación” dentro del programa EpiMolVIH previamente indicada. Ésta generó una tabla con una secuencia consenso y el grado de conservación para cada residuo de p24 (VIH-1) o p26 (VIH-2) en cada posición y variante analizada. Dicha tabla se modificó utilizando la herramienta Microsoft Excel de Windows, generando un código de colores para cada intervalo de porcentajes de conservación: rojo ($\leq 50\%$), naranja: (50.1-74.9%), amarillo (75-89.9%), verde claro (90-99.9%) y verde oscuro (100%). Para identificar los aa más conservados en CA, se seleccionaron los residuos con conservación $\geq 90\%$, es decir, aquellos que aparecían en al menos el 90% de las secuencias analizadas en esa posición, para cada variante del VIH-1 y VIH-2 con al menos 3 secuencias disponibles en LANL.

3.4 Generación de secuencias consenso

Las secuencias consenso se definen como aquellas secuencias que tienen en cada posición aminoacídica, el aa más frecuente de la cuasiespecie viral. Con el procedimiento previamente

descrito, se generaron secuencias consenso de CA para los grupos B y A y el CRF H2_01_AB del VIH-2, para los grupos O, P, y N del VIH-1, y para cada subtipo, sub-subtipo y CRF del Grupo M del VIH-1. Por otro lado, generamos tres secuencias consenso de consensos a partir de los consensos previamente generados mediante el programa EpiMol (**Tabla 2**). Para ello, se crearon manualmente archivos *fasta* de los consensos generados y se introdujeron en el programa “Conservación” de EpiMol, generándose la secuencia consenso de consensos correspondiente con la conservación de cada aa. Las 3 secuencias consensos de consensos creadas fueron:

- (I) Consenso de consensos de la CA del VIH-1 empleando las secuencias p24 consenso generadas para los grupos M, N, O y P del VIH-1
- (II) Consenso de consensos de la CA del grupo M de VIH-1 a partir de las 99 secuencias consenso p24 que representaban a cada uno de los 8 subtipos, 7 sub subtipos y 84 CRFs con secuencia disponible en LANL a fecha de estudio.
- (III) Consenso de consensos de la CA del VIH-2 a partir de las secuencias p26 consensos de sus 3 variantes con secuencias disponibles en LANL.

Es importante señalar que las secuencias consenso de consenso se emplearon para el análisis de conservación, búsqueda de marcadores de variante y análisis de estructuras y aa de interés de CA, tomándose como referencia en los alineamientos, como se irá indicando en los apartados posteriores.

Tabla 2. Secuencias de la CA empleadas para establecer cada consenso.

Consenso	Secuencias empleadas	Método
Consenso de cada una de las 99 variantes del VIH-1: grupos O, P y N y subtipos, sub-subtipos y CRFs del grupo M	Secuencias de p24 disponibles en LANL. El nº total de secuencias por variante se describe en la Tabla 4 .	Generados directamente desde el programa EpiMol.
(I) Consenso de consensos del VIH-1	Generada a partir de las 4 secuencias consenso de cada grupo del VIH-1.	Generada por el programa tras introducir un nuevo archivo <i>fasta</i> de consensos creado manualmente.
(II) Consenso de consensos del grupo M del VIH-1	Generada a partir de las 99 secuencias consenso de cada variante del grupo M.	Generada por el programa tras introducir un nuevo archivo <i>fasta</i> de consensos creado manualmente.
Consenso de cada una de las 3 variantes del VIH-2	Secuencias de p26 disponibles en LANL. El nº total de secuencias por variante se describe en la Tabla 4 .	Generados automáticamente con programa EpiMol.
(III) Consenso de consensos del VIH-2	Generada a partir de las 3 secuencias consenso de cada variante del VIH-2.	Generada por el programa tras introducir un nuevo archivo <i>fasta</i> de consensos creado manualmente.

3.5 Identificación de marcadores genéticos en CA específicos de cada variante del VIH

Para buscar marcadores genéticos en la secuencia de aa de la CA específicos de cada variante del VIH, se empleó la tabla previamente diseñada para el estudio de la conservación. Así, seleccionamos los cambios de aa que presentasen una conservación $\geq 75\%$ para esa posición en cada variante con respecto a las secuencias consenso de consenso generadas. Así, la secuencia consenso de consensos del VIH-1 (I) se usó para ver marcadores en los 4 grupos O, P, M y N del VIH-1, la secuencia consenso de consensos del Grupo M (II) para ver marcadores de cada subtipo, sub-subtipo y CRF del Grupo M del VIH-1 y la secuencia consenso de consensos del VIH-2 (III) para analizar marcadores de grupos A, B y

CRFH2_01_AB del VIH-2. En las posiciones donde la secuencia consenso presentaba una conservación de aa $\leq 50\%$, se consideró marcador de variante cualquier residuo conservado en $\geq 75\%$ de secuencias del alineamiento, aunque éste no fuese diferente al consenso.

3.6 Estructura secundaria de la cápside viral

Los dominios conservados y los marcadores genéticos específicos identificados para cada variante se ubicaron dentro de la estructura secundaria de la proteína CA descrita en RCSB Protein Data Bank (DOI: 10.2210/pdb3GV2/pdb).

3.7 Mutaciones que afectan el ensamblaje de la cápside inmadura

En base a bibliografía actualizada de PubMed, se identificaron aquellas mutaciones de CA que afectan al ensamblaje de la cápside inmadura en las 39.010 secuencias del VIH-1 descargadas de LANL. Dado que la bibliografía suele hacer referencia al grupo mayoritario M, los grupos P, O y N, se analizaron separadamente.

3.8 Naturaleza de los aa en marcadores de variante y en conservación de residuos clave

Para caracterizar los cambios de aa presentes en algunos marcadores de variante y residuos clave en la proteína CA del VIH, clasificamos los aminoácidos en base a la polaridad de su cadena lateral en 4 tipos: 1) apolares; 2) neutros polares; 3) catiónicos o básicos; 4) aniónicos o ácidos. A su vez se tuvieron en cuenta aquellos cambios de aa aromáticos que se caracterizan porque en su estructura posee un anillo de benceno. Son relativamente apolares (hidrofóbicos): fenilalanina (F), tirosina (Y) y triptófano (W) (**Tabla 3**).

Tabla 3. Clasificación de aminoácidos según naturaleza de su cadena lateral.

SEGÚN POLARIDAD		
	TIPO	aa
1	APOLARES	alanina (A), valina (V), leucina (L), isoleucina (I), metionina (M), prolina (P), fenilalanina (F), triptófano (W)
2	POLARES SIN CARGA O NEUTROS POLARES	serina (S), treonina (T), glutamina (Q), asparagina (N), tirosina (Y), cisteína (C), glicina (G)
3	CATIÓNICOS o BÁSICOS	lisina (K), arginina (R), histidina (H)
4	ANIÓNICOS o ÁCIDOS	ácido aspártico (D), ácido glutámico (E)
SEGÚN NATURALEZA Y PROPIEDADES		
1	AROMÁTICOS	fenilalanina (F), tirosina (Y), triptófano (W).

3.9 Clasificación de variantes del VIH empleando análisis filogenético con secuencias CA

Se descargaron de LANL 2 secuencias de nucleótidos en formato fasta de cada una de las variantes del VIH en p24 (VIH-1) y p26 (VIH-2). Tras alinearse con MUSCLE, los árboles se realizaron con el programa MEGA v6.06 por Maximum Likelihood con el programa FastTee, usando como modelo evolutivo el GTR+G+I y 1000 réplicas (*bootstrap*). Para evaluar la fiabilidad de las ramas, consideramos estadísticamente fiable un valor de *bootstrap* mayor o igual a 70%.

RESULTADOS

4.1 Secuencias analizadas

El total de secuencias descargadas por variante empleadas en este trabajo se describe en la **Tabla 4**. Se descargaron un total de 39.021 secuencias de la proteína CA de VIH de la base de datos LANL, 188 pertenecientes a VIH-2 y 38.833 a VIH-1: 119 de los grupos no-M (O, N y P) y 38.714 del grupo mayoritario M. De éste último las variantes con mayor representación fueron los subtipos B, C, sub-subtipo A1 y el CRF01_AE. En LANL no se encontraron secuencias disponibles de p24/p26 del subtipo F, CRF30_0206, CRF76_01B, CRF80_0107 y CRF81_cpx del VIH-1 y de los grupos C, D y E del VIH-2, aunque sí se encontraban almacenadas las secuencias de otras regiones del genoma del VIH. En cambio, del sub-subtipo A5, CRF75_BF, CRF84_A1D, CRF89_BF, CRF91_01C, CRF94_02B, CRF95_02B y CRF97_01B del VIH-1 no existen secuencias de ninguna región del VIH en LANL, por lo que no figuran en la **Tabla 4**. Descartando las variantes con menos de 3 secuencias, finalmente se emplearon 38.827 secuencias de HIV-1 y 183 secuencias de HIV-2 para este estudio.

4.2 Conservación de CA entre variantes del VIH

La **Tabla 5** muestra el grado de conservación en el total de secuencias analizadas por variante en la proteína de la cápside viral de cada tipo, subtipo, sub-subtipo, grupo y CRF del VIH.

En general todos los tipos, subtipos, sub-subtipos, grupos y CRFs del VIH presentaron una alta conservación de aa en la proteína CA. Así, en el consenso de consensos de VIH-1 (I) generado con las 4 secuencias consenso de los grupos O, P, N y M, un total de 169 (73.2%) de los 231 aa de p24 presentaron una conservación del 100%, mientras que en el consenso de consensos de VIH-2 la conservación fue mayor (206 residuos, 89.6%). El grupo de VIH-1 con CA más conservada fue el grupo P (90.5% de aa con conservación 100%). Entre el resto de grupos, el grupo N presentó un mayor nº de aa conservados (75.8%) en el 100% de sus secuencias, seguido del grupo M (69.3%) y grupo O (48.1%). Si consideráramos sólo los conservados en $\geq 90\%$, los porcentajes fueron 90.5% (grupo P), 88.3% (grupo N), 84.8% (grupo M) y 84.4% (grupo N).

En los subtipos del grupo M de VIH-1, la conservación $\geq 90\%$ osciló entre el 77.9% (subtipo J, 7 secuencias) y 93.9% (sub-subtipo A6, 76 secuencias). Los subtipos con mayor nº de aa 100% conservados en p24 fueron los que tenían menos secuencias (A4, K). Considerando los subtipos con más de 10 secuencias, el A3 (16 secuencias) y el H (14 secuencias) presentaron el mayor porcentaje de aa (77.5% y 77.3% de aa respectivamente) conservados en el 100% de sus secuencias analizadas. Las variantes A1, B, C, D y CRF01_AE con mayor nº de secuencias analizadas (entre 1.320 y 14.577) no presentaron ningún residuo totalmente conservado. La conservación $\geq 90\%$ en los CRFs de VIH-1 con más de 10 secuencias disponibles oscilaba entre el 87% de aa en el CRF11_cpx (29 secuencias) y el 99.1% de aa en el CRF83_cpx (11 secuencias), que a su vez presentó el mayor nº de residuos totalmente conservados (219 aa, 94.8%).

La variante de VIH-2 con mayor conservación en su cápside fue el CRFH2_01_AB (13 secuencias), presentando el 96.5% de aa con una conservación $\geq 90\%$ y 90.4% de aa conservados al 100%. La de menor conservación fue el grupo B, presentando un 88,3% de aa conservados $\geq 90\%$ y 87.8% de aa totalmente conservados entre las 8 secuencias analizadas.

Tabla 4. Secuencias de CA del VIH-1 y VIH-2 descargadas por variante

VIH-1			Variantes		Secuencias
VIH-1	Grupos No M	N		11	
		O		104	
		P		4	
	Grupo M	Subtipos	A		19
			A1		4002
			A2		71
			A3		16
			A4		3
			A6		76
			B		14577
			C		12310
			D		1320
			F		0
			F1		312
			F2		31
			G		176
			H		14
			J		7
	K		4		
	VIH-1	Grupo M	CRF	01_AE	3764
				02_AG	537
				03_AB	6
				04_cpx	12
				05_DF	6
				06_cpx	42
				07_BC	575
				08_BC	294
				09_cpx	11
				10_CD	3
				11_cpx	29
				12_BF	11
				13_cpx	12
				14_BG	15
15_01B				7	
16_A2D				4	
17_BF				7	
18_cpx				7	
19_cpx				5	
20_BG				4	
21_A2D				3	
22_01A1				20	
23_BG				2	
24_BG				11	
25_cpx				5	
26_A5U				4	
27_cpx				3	
28_BF				5	
29_BF				8	
30_0206				0	
31_BC				3	
32_06A6				5	
33_01B				18	
34_01B	3				
35_AD	23				
36_cpx	4				
37_cpx	5				
38_BF	5				
39_BF	3				
40_BF	4				
41_CD	3				
VIH-2	Grupo M	CRF	42_BF	17	
			43_02G	8	
			44_BF	3	
			45_cpx	10	
			46_BF	7	
			47_BF	5	
			48_01B	6	
			49_cpx	8	
			50_A1D	5	
			51_01B	7	
			52_01B	3	
			53_01B	4	
			54_01B	3	
			55_01B	15	
			56_cpx	4	
			57_BC	5	
			58_01B	6	
			59_01B	9	
			60_BC	4	
			61_BC	4	
			62_BC	3	
			63_02A	11	
			64_BC	9	
			65_cpx	16	
			67_01B	3	
			68_01B	3	
			69_01B	7	
			70_BF1	5	
			71_BF1	14	
			72_BF1	6	
			73_BG	2	
			74_01B	6	
			76_01B	0	
77_cpx	4				
78_cpx	3				
79_0107	3				
80_0107	0				
81_cpx	0				
82_cpx	6				
83_cpx	11				
85_BC	9				
86_BC	3				
87_cpx	3				
88_BC	3				
90_BF1	11				
92_C2U	3				
93_cpx	3				
96_cpx	3				
98_06B	2				
99_BF	3				
VIH-2	Grupos	A	162		
		AB	1		
		B	8		
		C	0		
		D	0		
		E	0		
		F	1		
U	2				
CRF	H2_01_AB	13			

En rojo, variantes con <3 secuencias disponibles no utilizadas para los análisis de conservación y de marcadores de variante. En negro, variantes con secuencias utilizadas para la generación de consensos de consensos y los análisis de conservación y marcadores de variante. Las variantes sin secuencia de ninguna región del VIH en la base de datos LANL no figuran en la tabla.

Tabla 5. Número y porcentaje de aa conservados en el $\geq 90\%$ y 100% de las secuencias en cadatipo, grupo, subtipo, sub-subtipo y CRF del VIH.

		Conservación de secuencias de VIH					
		$\geq 90\%$		100%			
Variante	Secuencias	N° aa	Porcentaje	N° aa	Porcentaje		
I) Consenso de consensos VIH-1	4	169	73,2%	169	73,2%		
Grupo O	104	195	84,4%	111	48,1%		
Grupo P	4	209	90,5%	209	90,5%		
Grupo N	11	204	88,3%	175	75,8%		
II) Consenso de consensos Grupo M	99	196	84,8%	160	69,3%		
Subtipos	A	19	193	83,5%	151	65,4%	
	A1	4002	205	88,7%	0	0,0%	
	A2	71	213	92,2%	132	57,1%	
	A3	16	206	89,2%	179	77,5%	
	A4	3	213	92,2%	213	92,2%	
	A6	76	217	93,9%	130	56,3%	
	B	14577	211	91,3%	0	0,0%	
	C	12310	204	88,3%	0	0,0%	
	D	1320	212	91,8%	0	0,0%	
	F1	312	210	90,9%	26	11,3%	
	F2	31	207	89,6%	157	68,0%	
	G	176	204	88,3%	73	31,6%	
	H	14	199	86,1%	174	75,3%	
	J	7	180	77,9%	180	77,9%	
	K	4	196	84,8%	196	84,8%	
	CRFs	01_AE	3764	214	92,6%	0	0,0%
		02_AG	537	204	88,3%	23	10,0%
		03_AB	6	206	89,2%	206	89,2%
		04_cpx	12	211	91,3%	199	86,1%
		05_DF	6	210	90,9%	210	90,9%
06_cpx		42	210	90,9%	176	76,2%	
07_BC		575	222	96,1%	40	17,3%	
08_BC		294	214	92,6%	57	24,7%	
09_cpx		11	212	91,8%	186	80,5%	
10_CD		3	211	91,3%	211	91,3%	
11_cpx		29	201	87,0%	173	74,9%	
12_BF		11	217	93,9%	200	86,6%	
13_cpx		12	207	89,6%	195	84,4%	
14_BG		15	221	95,7%	207	89,6%	
15_01B		7	207	89,6%	207	89,6%	
16_A2D		4	211	91,3%	211	91,3%	
17_BF		7	202	87,4%	202	87,4%	
18_cpx		7	193	83,5%	193	83,5%	
19_cpx		5	216	93,5%	216	93,5%	
20_BG		4	219	94,8%	219	94,8%	
21_A2D		3	222	96,1%	222	96,1%	
22_01A1		20	214	92,6%	183	79,2%	
24_BG		11	219	94,8%	214	92,6%	
25_cpx		5	198	85,7%	198	85,7%	
26_A5U		4	204	88,3%	204	88,3%	
27_cpx		3	203	87,9%	203	87,9%	
28_BF		5	201	87,0%	201	87,0%	
29_BF		8	201	87,0%	201	87,0%	
31_BC		3	210	90,9%	210	90,9%	
32_06A6		5	212	91,8%	212	91,8%	
33_01B		18	211	91,3%	189	81,8%	
34_01B		3	225	97,4%	225	97,4%	
35_AD		23	209	90,5%	177	76,6%	
36_cpx		4	217	93,9%	217	93,9%	

		Conservación de secuencias de VIH			
		$\geq 90\%$		100%	
Variante	Secuencias	N° aa	Porcentaje	N° aa	Porcentaje
37_cpx	5	202	87,4%	202	87,4%
38_BF	5	204	88,3%	204	88,3%
39_BF	3	218	94,4%	218	94,4%
40_BF	4	221	95,7%	221	95,7%
41_CD	3	227	98,3%	227	98,3%
42_BF	17	222	96,1%	203	87,9%
43_02G	8	205	88,7%	205	88,7%
44_BF	3	211	91,3%	211	91,3%
45_cpx	10	214	92,6%	183	79,2%
46_BF	7	206	89,2%	206	89,2%
47_BF	5	214	92,6%	214	92,6%
48_01B	6	225	97,4%	225	97,4%
49_cpx	8	195	84,4%	195	84,4%
50_A1D	5	220	95,2%	220	95,2%
51_01B	7	216	93,5%	216	93,5%
52_01B	3	212	91,8%	212	91,8%
53_01B	4	207	89,6%	207	89,6%
54_01B	3	225	97,4%	225	97,4%
55_01B	15	222	96,1%	217	93,9%
56_cpx	4	224	97,0%	224	97,0%
57_BC	5	222	96,1%	222	96,1%
58_01B	6	199	86,1%	199	86,1%
59_01B	9	220	95,2%	220	95,2%
60_BC	4	216	93,5%	216	93,5%
61_BC	4	224	97,0%	224	97,0%
62_BC	3	223	96,5%	223	96,5%
63_02A	11	223	96,5%	205	88,7%
64_BC	9	219	94,8%	219	94,8%
65_cpx	16	227	98,3%	215	93,1%
67_01B	3	221	95,7%	221	95,7%
68_01B	3	225	97,4%	225	97,4%
69_01B	7	213	92,2%	213	92,2%
70_BF1	5	198	85,7%	198	85,7%
71_BF1	14	203	87,9%	176	76,2%
72_BF1	6	201	87,0%	201	87,0%
74_01B	6	216	93,5%	216	93,5%
77_cpx	4	228	98,7%	228	98,7%
78_cpx	3	215	93,1%	215	93,1%
79_0107	3	226	97,8%	226	97,8%
82_cpx	6	211	91,3%	211	91,3%
83_cpx	11	229	99,1%	219	94,8%
85_BC	9	219	94,8%	219	94,8%
86_BC	3	224	97,0%	224	97,0%
87_cpx	3	224	97,0%	224	97,0%
88_BC	3	227	98,3%	227	98,3%
90_BF1	11	211	91,3%	201	87,0%
92_C2U	3	203	87,9%	203	87,9%
93_cpx	3	207	89,6%	207	89,6%
96_cpx	3	225	97,4%	225	97,4%
99_BF	3	216	93,5%	216	93,5%
III) Consenso de consensos de VIH-2	3	206	89,6%	206	89,6%
Grupo A	162	215	93,5%	109	47,4%
Grupo B	8	203	88,3%	202	87,8%
CRF H2_01_AB	13	222	96,5%	208	90,4%

La descripción de cómo se generaron las secuencias consenso I, II, y III se muestra en el apartado 3.4.

4.3 Conservación de la estructura secundaria de la proteína de la cápside viral

Una vez determinados los residuos más conservados en el VIH, procedimos a localizarlos en cada elemento de la estructura secundaria de la cápside. Es importante recordar que ésta tiene un dominio N-terminal (NTD) con la horquilla β , 7 hélices α (H) y el *Loop* de unión a Cyp A, y un dominio C-terminal (CTD) con la Región de Unión Interdominios (RUI), 4 hélices α y la Región de Homología Mayor (MHR). La ubicación de cada región, su conservación y nº de secuencias analizadas de VIH-1 (I) y grupos O, P, N y M (II) se describe en la **Figura 6**.

Tras sumar el valor de conservación de cada residuo en la estructura y dividir por el nº total de residuos de la misma, observamos que la conservación de ciertas estructuras difería entre variantes y tipo de estructura (**Figura 7**). En el consenso de VIH-1 (I), la estructura de p24 con mayor conservación fue H8 (98.2%), seguida de MHR (97.5%) y RUI (95%). Las estructuras con menor conservación fueron H6 (69.4%), H3 (83.3%) y el *loop* de unión a CypA (86.1%) (**Figura 7A**). Al comparar la conservación de las estructuras secundarias de cada uno de los 4 grupos (**Figura 7A**), se observó que el consenso del grupo M (II) presentaba una mayor conservación de H6 (93.3%). A su vez, el grupo O presentó una menor conservación de H11 en comparación al resto de grupos (96.27% vs. 98.6-100%) y una mayor conservación de H9 (98.5 vs. 90-96%) y de H10 (99.6 vs. 90-94%). El grupo N presentó una mayor conservación de MHR (98.6% vs. 96%) y de H4 (98.2% vs. 95-96%). El grupo P fue el único que presentó un 100% de conservación en H5, H1, H2, H11, H3 y en la horquilla β , y la mayor conservación en H9 (96.1%), aunque presentó la menor conservación en RUI (80%). Se calculó la conservación media total de cada una de las estructuras de p24 entre los 4 grupos (**Figura 7B**) para conocer aquellas más robustas (más tolerantes al cambio o menos conservadas) y las más frágiles (menos tolerantes a cambios o más conservadas). Las estructuras con mayor conservación fueron H1 (99%), H2 (99%) y H5 (99%). Las hélices α H11 (98%), H7 (97%), H8 (97%) y MHR (97%) presentaban una conservación media alta (97-98%) y H4 (96%), H3 (95%), H9 (95%), H10 (94%) y RUI (94%) una conservación media baja (94-96%). Por último, las estructuras con menor conservación fueron la horquilla β (93%), H6 (91%) y el *loop* de unión a CypA (90%).

4.3.1 Región de Homología Mayor (MHR)

La **Figura 8** muestra el alineamiento de aa en MHR en los 4 grupos del VIH-1. Fue una región altamente conservada entre grupos del VIH-1 (97%) y en el consenso del VIH-1 (I) (97.5%), siendo ligeramente mayor en el Grupo N (98,6%). A su vez estuvo altamente conservada en VIH-2 (98%), presentando los cambios Q162, S163 y S171 con respecto al VIH-1 que comparte con el lentivirus de macacos SIVmac239 (Tanaka *et al.*, 2016).

4.3.2 *Loop* de unión a ciclofilina A

La **Figura 9** muestra el alineamiento del *loop* de unión a CypA en los 4 grupos del VIH-1. Identificamos 3 residuos (P85, G89 y P90) con una conservación del 100% en los 4 grupos. En las posiciones P92 y P93, predominó P en los consensos de grupo, aunque la conservación encontrada no fue del 100%. Las posiciones 86, 87, 88 y 91 fueron las más variables del *loop*. También encontramos variabilidad en el residuo A88, aa apolar relevante para la unión de p24 a la CypA según la bibliografía (Gamble *et al.*, 1996), que sólo presentaba una conservación del 100% en el consenso del grupo M (II), pero no en los grupos no-M del VIH-1. Así, los 3 grupos no-B presentaron V88 (aa apolar) con distinta frecuencia: 47.1%, 75% y 90.9% en grupo O, P y N, respectivamente.

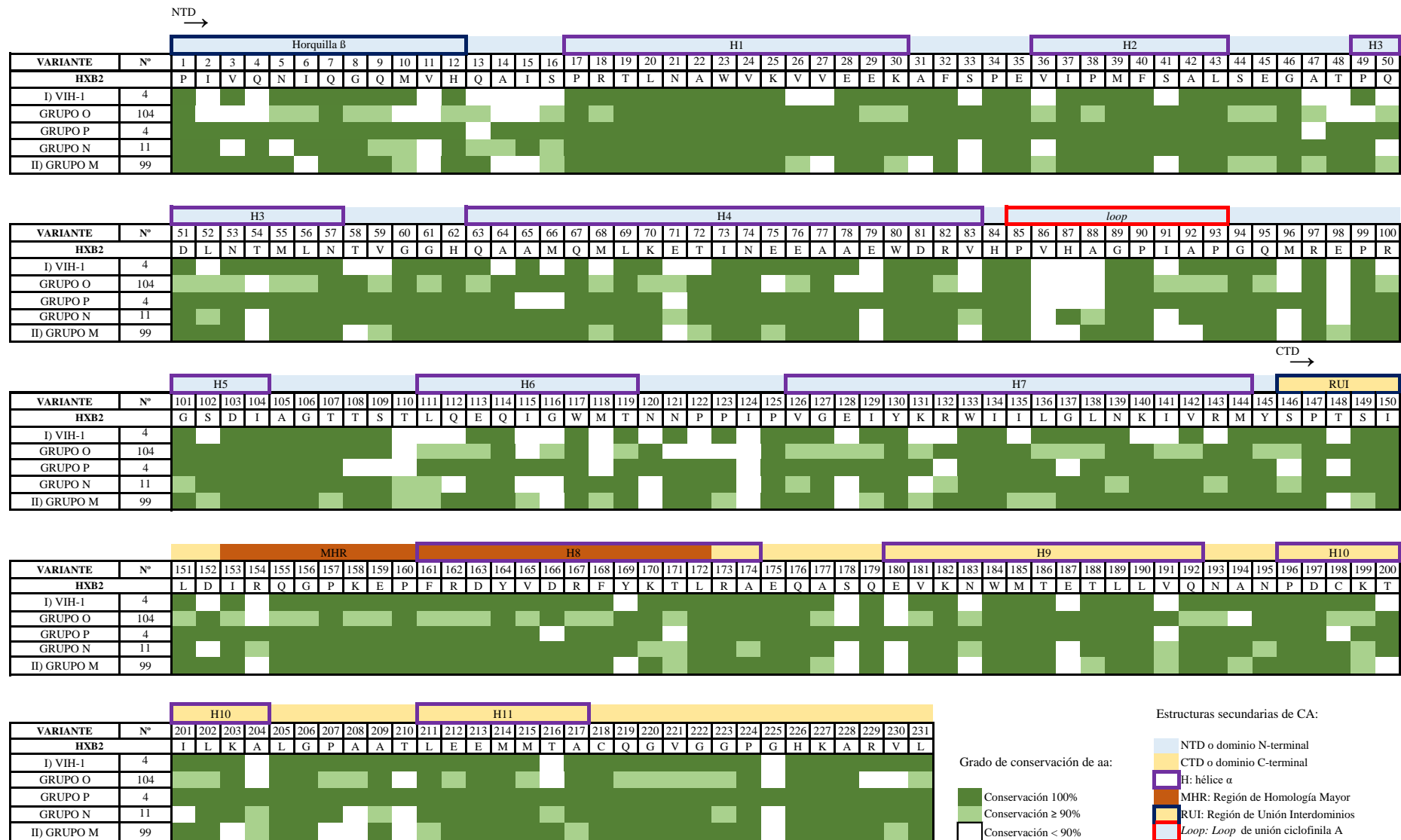


Figura 6. Conservación de los 4 grupos del VIH-1 en la estructura secundaria de la cápside viral y número de secuencias analizadas.

A)

Conservación	I) Consenso de consensos VIH-1	II) Consenso de consensos Grupo M	Grupo O	Grupo P	Grupo N
100%			H5	H5, H2, H1, H11, H3, Horquilla	H1
98-99.9%	H8	H5, H11	H2, H1, RUI, H10, H8, H7, H9		H11, MHR, H4, RUI, H8
96-97.9%	MHR	H8, H7, H2, H1, RUI, MHR	MHR, H11	H4, MHR, H9, H7	H5, H2, H7
94-95.9%	RUI, H1, H10	H4, H3, Horquilla	H4, H3	H8, H10, H6	
92-93.9%	H5, H11	H6			loop, H10, H9
90-91.9%	Horquilla, H2, H9	H9, H10		loop	H3, Horquilla
88-89.9%	H4				H6
86-87.9%	H7, loop	loop	loop, H6		
84-85.9%			Horquilla		
82-83.9%	H3				
80-81.9%				RUI	
<80%	H6				
Nº secuencias	4	99	104	4	11
Conservación media	90,19%	95,31%	95,73%	96,00%	95,58%

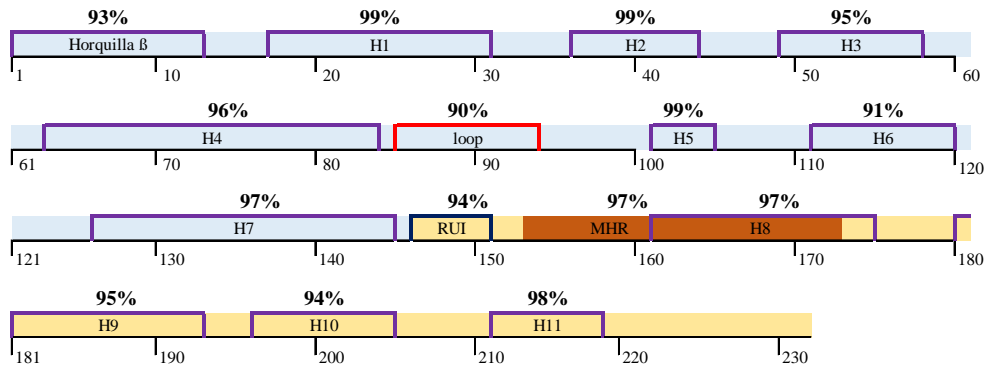


Figura 7. Conservación de cada elemento de la estructura secundaria de la cápside en el VIH-1 y en cada uno de sus grupos(A) y conservación media (B) H: Hélice α ; RUI; Región de Unión Interdominio; MHR: Región de Homología Mayor (naranja). Se muestra la región NTD en celeste y la región CTD en amarillo.

	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		
I) Consenso de consensos VIH-1	I	K	Q	G	P	K	E	P	F	R	D	Y	V	D	R	F	Y	K	T	L
GRUPO O	I	K	Q	G	P	K	E	P	F	R	D	Y	V	D	R	F	Y	K	T	L
GRUPO P	I	K	Q	G	P	K	E	P	F	R	D	Y	V	D	R	F	Y	K	T	L
GRUPO N	I	K	Q	G	P	K	E	P	F	R	D	Y	V	D	R	F	Y	K	T	L
II) Consenso de consensos GRUPO M	I	R	Q	G	P	K	E	P	F	R	D	Y	V	D	R	F	F	K	T	L
III) Consenso de consensos VIH-2	I	K	Q	G	P	K	E	P	F	Q	S	Y	V	D	R	F	Y	K	S	L
SIVmac239	V	K	Q	G	P	K	E	P	F	Q	S	Y	V	D	R	F	Y	K	S	L
	153																			172

Figura 8. Alineamiento de MHR.

Letra en **rojo**: aa invariables de MHR en retrovirus. Letra **negrita**: conservación del 100%. Asteriscos: aa idénticos en todas las secuencias. Secuencia SIVmac329 (LANL).

				*	*	*			
I) Consenso de consensos VIH-1	P	V	P	V	G	P	L	P	P
GRUPO O	P	P	P	V	G	P	L	P	P
GRUPO P	P	V	P	V	G	P	L	P	P
GRUPO N	P	V	P	V	G	P	L	P	P
II) Consenso de consensos GRUPO M	P	V	H	A	G	P	I	P	P
	85			88	89	90			93

Figura 9. Alineamiento de secuencias del *loop* de unión a ciclofilina A. En **negrita**: conservación del 100%. En *cursiva*, conservación $\geq 90\%$. Con asterisco: posiciones de aa importantes para la unión de Cyp A según bibliografía. En **rojo**: residuos que coinciden con los aa importantes para la unión de Cyp A según bibliografía.

4.3.3 Poro del hexámero

La cápside permite el acceso de los 4 nt necesarios para una transcripción reversa eficiente a través de un poro cargado positivamente en cada hexámero. Se han descrito 4 aa claves para el correcto funcionamiento del poro en la horquilla β de p24 (P1, H12, R18 y D51), ya que cambios en ellos afectan a la estructura del poro y, por tanto, a la eficiencia de la transcripción reversa (Jaques *et al.*, 2016). En nuestro estudio, P1 (aa apolar) estaba conservado al 100% en los grupos O, P y N. Sin embargo 64 (0.2%) de las 38714 secuencias del consenso del grupo M (II) presentaron el cambio A1 (aa apolar) en 3 variantes, predominando el subtipo C (31 secuencias, 0.3%). En el residuo 12, en lugar de H12 (aa básico), el grupo O presentó Y12 (aa aromático) en 4 (3.8%) secuencias. El consenso del grupo M presentó cambios en esta posición en 889 secuencias (2.3%), siendo también Y12 (aa aromático) el polimorfismo mayoritario, presente en 636 secuencias (1.6%) de 21 variantes diferentes. Y12 se encontró con mayor frecuencia en el sub-subtipo F1 (49 secuencias, 15.7%), seguido del CRF02_AG (30 secuencias, 5.6%) y del subtipo D (60 secuencias, 4.5%). En R18 (aa básico) no se encontraron polimorfismos en los grupos O, P y N, pero sí en 114 (0.3%) secuencias del consenso del grupo M, siendo K18 (aa básico) el cambio más prevalente (60 secuencias, 0.2%). K18 se encontró en 12 variantes, predominando el sub-subtipo F1 (4 secuencias, 1.3%), seguido de los subtipos B (23 secuencias) y C (19 secuencias), ambos con una prevalencia del 0.2%. En D51 (aa ácido) tampoco se encontraron cambios en los grupos O, P y N, pero sí en 68 (0.2%) secuencias del consenso del grupo M, siendo el más prevalente el aa neutro polar N51 (45 secuencias), principalmente en el subtipo D (14 secuencias, 1%).

4.4 Marcadores específicos de variante en la proteína de la cápside del VIH

4.4.1 Marcadores de VIH-2

Los marcadores genéticos específicos de cada variante de VIH-2 con secuencia (Grupos A, B y CRF H2_01_AB) se describen en la **Figura 10**. Es importante destacar que no se encontraron polimorfismos en MHR de ninguna de las variantes de VIH-2. La gran mayoría de cambios de residuo supuso cambios entre aa del mismo tipo o bien cambios entre aa apolares y neutros polares. El cambio más frecuente fue el de V por I o viceversa, ambos residuos apolares. Sin embargo el CRFH2_01_AB presentó un marcador con cambio de aa neutro polar por aa ácido (Q186E) y otro con cambio de aa neutro polar por aa básico (N207H). Además, presentó el mayor número de marcadores específicos de variante con máxima (100%) conservación.

VARIANTE	N° SECUENCIAS	Posición de aa																				Conservación:
		5	10	35	60	67	96	117	119	122	135	152	178	181	182	186	190	206	207	220		
BEN		A	V	V	D	I	D	Y	P	P	I	I	P	K	N	Q	I	M	N	V		
CONSENSO VIH-2		I	V	V	E	I	D	Y	A	P	I/L	I	P	K	N	Q	I/V	M	N	I		
A	162		T		D		E	F			I									V		
B	8																I					
H2_01_AB	13	V		I		V			G	S	L	V	T	R	A	E		P	H			

Figura 10. Polimorfismos marcadores de variante en VIH-2.

4.4.2 Marcadores de Grupo de VIH-1

Los polimorfismos de los grupos de VIH-1 se describen en la **Figura 11**. El grupo P fue el grupo que presentó mayor número de marcadores de variante conservados en un 100% (26

residuos). Los grupos N y M compartían 15 marcadores en distintas posiciones y, a su vez, los grupos O y P también compartían hasta 12 marcadores. No se encontraron marcadores de grupo en MHR, RUI, ni en las hélices α H8, H10 o H11 de p24. Los cambios de residuo más frecuentes fueron L-I, V-I, A-V (todos residuos apolares), V-T (cambio entre aa apolar y neutro polar) y K-R (aa básicos). Se observaron 2 posiciones que presentaron marcadores de grupo con una alta variabilidad de residuo en cuanto al tipo de aa y naturaleza de éste: posición 50 (H3) y 111 (H6). En la posición 50, el grupo P presentó el marcador D50 (aa ácido) con una conservación del 100%. El grupo O presentó Y50 (aa aromático) y el grupo M, Q50 (aa neutro polar), ambos con una conservación $\geq 90\%$. En posición 111, el grupo P presentó el marcador K111 (aa básico), conservación al 100%. El grupo O presentó el marcador Q111 (aa neutro polar) y los grupos N y M, el marcador L111 (aa apolar), los tres con una conservación $\geq 90\%$. Por último, cabe destacar los marcadores de la posición 100 en H5, una de las estructuras más conservadas en los grupos de VIH-1, donde los grupos O y P presentaban el marcador T100 (aa neutro polar) con una conservación $\geq 90\%$ y del 100% respectivamente. Los grupos N y M presentaron el marcador R100 (aa básico) con una conservación del 100%.

Variante	N° secuencias	HB		H1				H2		H3				H4				L	H5						
		2	11	13	14	16	26	27	33	36	47	48	50	52	58	59	64	66	68	72	86	88	96	100	102
HXB2		I	V	Q	A	S	V	V	S	V	A	T	Q	L	T	V	A	M	M	T	V	A	M	R	S
I) Consenso de consensos VIH-1		I	V	Q	P	S	A/V	V	N	I/V	A	L/T	D ^V /S ^O	I/L	A/T	V	A/G	L/M	M/V	V	V	V	L	T/R	S
Grupo O	104						A			I			Y	I	A	I	G	L	V		P		I	T	
Grupo P	4	I	T	L	P		A			I	S	I	D	I	A		G	L	V			V	L	T	G
Grupo N	11	L			P	T	V	I		V		T		L	T		A	M	M			V		R	
II) Consenso de consensos Grupo M		I					V		S	V		T	Q	L			A	M	M	T		A		R	

Variante	N° secuencias	H6				H7								H9				Conservación:				
		111	115	116	118	120	122	124	131	132	135	140	141	143	152	178	183		187	191	216	225
HXB2		L	I	G	M	N	P	I	K	R	I	K	I	R	D	S	N	E	V	T	G	
I) Consenso de consensos VIH-1		L	I/V	A ^G /H ^T	M	R	N/P	I	R	K/R	V	K	I	K/R	D	T	N	D/E	V	T	T	
Grupo O	104	Q		H		R		I		K			M	K				E		V	T	
Grupo P	4	K	I	T		R	N			K			V	K		S	S	D	I	N	T	
Grupo N	11	L			M		P			R			R	I	R	E					T	
II) Consenso de consensos Grupo M		L	I		M		P		K	R	I		I	R							T	

Figura 11. Marcadores específicos de cada grupo del VIH-1. Estructuras: HB: horquilla β , H: hélice α , L: loop, NTD: color celeste, CTD: color amarillo

4.4.3 Marcadores de las distintas variantes del grupo M del VIH-1 y su presencia en estructuras secundarias

Los marcadores genéticos específicos de cada subtipo, sub-subtipo y CRF del grupo M del VIH-1 se describen en el **Anexo I**. Debido al alto número de marcadores, se escogió analizar aquellos que presentaron cambios significativos en la naturaleza de su aa y los presentes en estructuras de p24 de especial interés (*loop* de unión a CypA y MHR) y en posiciones altamente variables donde el consenso del grupo M (II) presentaba una baja conservación.

Se encontraron marcadores genéticos específicos de variante en todos los subtipos, sub-subtipos y CRF del grupo M del VIH-1 analizados. No hubo ninguna estructura secundaria de p24 que no presentase algún marcador. Los cambios de residuo más frecuentes fueron entre aa apolares o entre aa apolares y neutros polares. En cuanto a los marcadores más relevantes, encontramos W10 en la horquilla β del CRF21_A2D con un 100% de conservación, con respecto a M10 (aa apolar) del consenso del grupo M.

La Región de Homología Mayor (MHR), como se comentó en el apartado 4.3.1, tuvo una alta conservación. El consenso del grupo M (I) presentó dos posiciones (154 y 169) donde R154 (aa básico) pasaba a ser K154 (aa básico) en 23 variantes y F169 (aa aromático) era Y169 (aa aromático) en 28 variantes, con conservación <75%. No se encontraron más marcadores de variante en MHR, excepto en la posición 171. Mientras el consenso del grupo M presentaba T171 (aa neutro polar), el CRF 47_BF tenía V171 (aa apolar) y el sub-subtipo A3 presentaba C171 (aa neutro polar) en $\geq 90\%$ de sus secuencias.

El *loop* de unión a ciclofilina A, fue altamente variable como se comentó en el apartado 4.3.2. De sus 9 aa, el consenso del grupo M (I) presentó 100% de conservación en las posiciones 85, 88, 89, 90 y 93 (**Figura 9**), por lo que no se hallaron marcadores de variante en éstas. Sin embargo, las posiciones 86, 87, 91 y 92 presentaron alta variabilidad en algunas variantes, siendo la posición 91 de las más variables de p24, con hasta 7 polimorfismos. Según la naturaleza de los aa destacaron los siguientes tres marcadores con respecto a la secuencia consenso del grupo M: 1) H87Q (aa básico por aa neutro polar), con 100% de conservación en 11 variantes, 2) I91H (aa apolar por aa básico) en CRF36_cpx (100% de conservación) y CRF83_cpx ($\geq 90\%$ de conservación) y 3) I91F (aa apolar por aa aromático), marcador en 3 variantes (sub-subtipo A6 y CRF03_AB y CRF32_06A6), con una conservación intermedia (75-90%).

Otras posiciones con alta variabilidad de polimorfismos fueron G116 (H6), A14, I15, L83 (H4) y S120, destacando los marcadores con cambio en la naturaleza del aa G116R (de aa neutro polar por aa básico) en CRF25_cpx (conservación del 100%) y S120H (cambio de aa neutro polar por aa básico) en los recombinantes CRF36_cpx y CRF90_BF1 con 75-90% de conservación y en CRF69_01B con un 100% de conservación.

4.5. Mutaciones que afectan el ensamblaje de la cápside inmadura:

Para un correcto ensamblaje de la cápside inmadura, la proteína funcional CA debe presentar V181/K182, W184/M185 y L189/L190 en la hélice α 9 (H9) que forma la interfaz de dímeros del extremo CTD entre hexámeros. Dobles mutantes en dichas posiciones alteran los enlaces hidrofóbicos que estabilizan la interacción de la interfaz de dímeros con la pareja homóloga. En la base del extremo CTD hay 3 aa en regiones no contiguas que están muy próximas entre sí en la estructura terciaria: K158 (MHR), D197 (H10) y P224 (extremo 3' CTD). Las mutaciones en estos aa afectan la multimerización de gag en la membrana. Por último, en las hélices α 4-6, las mutaciones dobles en las posiciones 75/76 de H4, 100/102 en H5 y 107/108 y 110/112 en H6, donde la proteína funcional debe presentar E/E, R/S, T/T y T/Q respectivamente, inhiben el paso final del ensamblaje, provocando el acúmulo de intermediarios de la ruta sin salida de viriones (Lingappa *et al.*, 2014). En este trabajo se buscaron estos 3 tipos de mutaciones que afectan el ensamblaje de la cápside inmadura en distintas localizaciones de CA en estas regiones, con los siguientes resultados:

Mutantes en la hélice α 9. En los grupos O, P y N no observamos cambios en dichas posiciones. En el grupo M se encontraron los dobles mutantes C184T185, R184T185, R184K185, L184I185 y R184L185 en 6 (0.01%) de las 38.708 secuencias, 5 (0.03%) de ellas del subtipo B. También aparecieron dobles mutantes V189V190, W189W190 y C189V190 en 3 (<0.01%) secuencias del grupo M (2 subtipos D, 1 subtipo C).

Mutantes en la base del extremo CTD. En los grupos O, P y N, los residuos K158, D197 y P224 estaban bien conservados y sólo una secuencia del grupo O presentó E158. En el grupo

M se encontraron mutantes en las 3 posiciones, principalmente R158 en 40 (0.1%) secuencias, mayoritariamente en el subtipo D (en 2% de sus 1320 secuencias), G197 en 26 (0.06%) secuencias y N197 en 22 (0.05%) secuencias, predominando en el sub-subtipo A1 (0.4% y 0.2% de sus 4002 secuencias respectivamente), y L224 en 16 (0.04%) secuencias.

Mutantes en las hélices α 4-6. En los grupos O, P y N, no se hallaron cambios en las posiciones 75/76 ni 107/108. Sin embargo, en las posiciones 100/102 del grupo O predominaba T100S102 (98% secuencias), mientras que el grupo P presentó T100G102 en todas sus secuencias. En las posiciones 110/112 se observaron diferencias en el grupo N, donde predominó T110A112 (91% secuencias) y en el grupo O, aunque predominaba T110Q112, se presentaron polimorfismos en el 5.8% de sus secuencias. En el grupo M se observaron los cuatro tipos de mutantes de hélices α 4-6, siendo la posición 110/112 la que presentó más polimorfismos (14), predominando el doble mutante N110A112 en 120 (0.3%) secuencias del grupo M, presentándose en 77 (0.6%) de las secuencias del subtipo C.

4.6 Clasificación de variantes del VIH empleando análisis filogenético con secuencias CA

Tras el análisis filogenético de secuencias de p24, se observó que la mayoría de variantes del VIH se clasificaban con un buen soporte estadístico (*bootstrap* ≥ 70) (**Anexo II**). Los grupos de VIH-1, O, P y N se clasificaron correctamente. En el grupo M, hubo una buena clasificación de 6 de los 8 subtipos, 3 de los 7 sub-subtipos y en 69 (82%) de los 84 CRF con secuencia disponible estudiados. En VIH-2 se clasificaron correctamente ambos grupos estudiados y el único CRF. Las variantes bien clasificadas se describen en la **Tabla 6**.

Tabla 6. Variantes bien clasificadas en el análisis filogenético (Bootstrap ≥ 70).

	Grupos	N, P, O
	Subtipos	A, B, G, H, J y K
	Sub-subtipos	A1, A2 y A3
VIH-1	CRFs	02_AG, 04_cpx, 05_DF, 06_cpx, 07_BC, 08_BC, 09_cpx, 11_cpx, 12_BF, 13_cpx, 14_BG, 18_cpx, 19_cpx, 20_BG, 21_A2D, 22_01A1, 24_BG, 25_cpx, 26_A5U, 29_BF, 31_BC, 32_06, A6, 33_01B, 34_01B, 35_AD, 36_cpx, 37_cpx, 38_BF, 39_BF, 40_BF, 41_CD, 42_BF, 43_02G, 44_BF, 47_BF, 48_01B, 49_cpx, 50_A1D, 51_01B, 54_01B, 55_01B, 56_cpx, 57_BC, 58_01B, 59_01B, 60_BC, 61_BC, 62_BC, 63_02A, 64_BC, 65_cpx, 69_01B, 70_BF1, 72_BF1, 74_01B, 77_cpx, 79_0107, 82_cpx, 83_cpx, 85_BC, 86_BC, 87_cpx, 88_BC, 90_BF1, 92_C2U, 93_cpx, 96_cpx, 98_cpx y 99_BF
VIH-2	Grupos	A y B
	CRF	H2_01_AB

Las variantes que no se clasificaron bien en el AF en p24: subtipos C y D, sub subtipos F1, F2, A4, y A6; CRF01_AE, CRF03_AB, CRF10_CD, CRF15_01B, CRF16_A2D, CRF17_BF, CRF23_BG, CRF27_cpx, CRF28_BF, CRF45_cpx, CRF46_BF, CRF52_01B, CRF53_01B, CRF67_01B, CRF68_01B, CRF71_BF1, CRF73_BG, y CRF78, aunque bastantes de ellas se clasificaban bien con un bootstrap muy cercano al 70%. Hay que recordar que las variantes no incluidas en el AF por falta de secuencia en LANL fueron el subtipo F, sub subtipo A5, CRF30, CRF75, CRF76, CRF80, CRF81, CRF84, CRF89, CRF91, CRF94, CRF95 y CRF97.

DISCUSIÓN

En este estudio se analizó la conservación de la proteína de la cápside viral (CA) a partir de un total de 39010 secuencias de 105 variantes (tipos, subtipos, sub-subtipos y CRFs) del VIH disponibles en el momento de estudio en la base de datos de LANL, así como de sus secuencias consenso generadas. Los resultados se analizaron en el contexto de la estructura secundaria de la proteína CA descrita en RCSB Protein Data Bank, profundizando en aquellas estructuras de mayor relevancia para su función. A su vez, hemos identificado por primera vez los polimorfismos naturales que se pueden considerar marcadores genéticos de cada variante del VIH-1 y VIH-2. Ambos procesos se llevaron a cabo utilizando una nueva herramienta bioinformática desarrollada en el laboratorio y testada durante el inicio del proyecto con dicha finalidad. Dada el elevado número de información generada en este trabajo, y la limitación en extensión de la versión escrita, sólo discutiremos los datos más relevantes encontrados.

Nuestro estudio se ha centrado en la proteína de la cápside viral porque es extremadamente importante para el VIH. La proteína CA juega un papel crítico en el ensamblaje de los viriones inmaduros. A través del ensamblaje de una matriz de hexámeros, se origina una cápside cónica en los virus maduros que alberga las enzimas y el ARN viral, permitiendo su transporte hasta el núcleo de la célula huésped. Pero además de esta función estructural, la CA interviene en diferentes procesos de la fase temprana de la infección viral, como el tráfico citoplasmático a través de los microtúbulos, la decapsidación y la importación nuclear del complejo de preintegración viral. A su vez, establece interacciones con proteínas celulares del huésped necesarias para el ciclo viral, como la ciclofilina A (Campbell and Hope, 2015). La CA permanece unida de alguna manera al complejo viral de transcripción reversa tras la decapsidación por un mecanismo aún desconocido, por el que la CA interacciona con proteínas del poro nuclear mediando en la importación nuclear del complejo de preintegración viral, e incluso permaneciendo unida a éste dentro del núcleo de la célula huésped. Estos hallazgos aumentan notablemente la importancia del papel funcional de esta proteína, aumentando su interés como diana terapéutica (Novikova *et al.*, 2019).

A diferencia de las enzimas virales del VIH, la CA es una proteína extremadamente frágil (Rihn *et al.*, 2013), es decir, que no acepta fácilmente cambios de aa resultado de mutaciones no sinónimas, ya que éstos reducen drásticamente el *fitness* o capacidad replicativa del virus. Esta fragilidad puede explicarse por la necesidad que tiene de mantener su estructura e interactuar con las diversas proteínas del huésped. Así, cada monómero de CA debe interactuar con, al menos, otros 3. Además, algunos deben ser capaces de adoptar estructuras y posiciones respecto a otros ligeramente distintas para formar las variantes pentaméricas que permiten cerrar la cápside (Perilla and Gronenborn, 2016). Esta alta conservación de CA la convierte en una región de especial interés para el desarrollo de test moleculares diagnósticos y como diana de nuevos antirretrovirales. Aun así, se sabe que la robustez genética de algunas estructuras puede variar. Así, se consideran regiones más robustas a las que admiten más cambios de aa y menos robustas o más frágiles a las que son menos tolerantes al cambio. Estudios previos han observado que la horquilla β , el *loop* de unión a Cyp A y RUI son regiones bastante robustas (menos conservadas) mientras que las hélices α son menos tolerantes a los cambios (más conservadas), siendo las más frágiles H2, H5, H6 y H7 (Rihn *et al.*, 2013). Nuestros resultados han revelado que, en general todos los tipos, subtipos, grupos y recombinantes del VIH1 presentan una alta conservación de aa en la proteína CA, pudiendo considerarse ésta una proteína altamente conservada (o frágil) a lo largo de su secuencia. Sin embargo, en cuanto al estudio de la conservación de estructuras individuales, hubo dos discrepancias: RUI, que presentó una conservación media baja (94%, la misma que H10) y

H6 que presentó una conservación incluso menor a la horquilla β (91%). Uno de los posibles motivos de esta discordancia es el mayor estudio en la bibliografía del subtipo mayoritario B del grupo M. En el análisis individualizado de las secuencias del subtipo B observamos que la conservación media de RUI descendía hasta el 91.8%. Sin embargo, la conservación de H6 era intermedia (95,1%).

A través de la modulación de la apertura de los poros de los hexámeros que conforman la cápside, ésta protege los componentes virales de los sensores citosólicos y nucleasas, mientras que permite el acceso de los nucleótidos para una transcripción reversa eficiente. Se han descrito 4 aa claves en la horquilla β (P1, H12, R18 y D51) para el correcto funcionamiento del poro, donde las mutaciones en P1 y D51 producen partículas virales no infecciosas (Jaques *et al.*, 2016). En nuestro estudio, se hallaron polimorfismos en el grupo M de ambos aa, aunque en muy baja frecuencia (0.2%). Destacamos el cambio H12Y encontrado en 3.8% de las secuencias del grupo O y en el 1.6% del grupo M, que favorece la conformación cerrada del poro reduciendo la cinética de la retrotranscripción (Jaques *et al.*, 2016). Nuestro trabajo también ha incluido el estudio de conservación a nivel de aa del *loop* de unión de a ciclofilina A (región de 9 aa en posiciones 85-93 de CA en la región NTD de p24). La CypA es una chaperona con actividad peptidil isomerasa que tiene un papel general en el plegamiento proteico. Es necesaria para el correcto plegamiento de *gag* del VIH, interviene en el proceso de entrada del virus mediante la interacción con CA y promueve la retrotranscripción del VIH-1 en las células humanas (Chuang *et al.*, 2016, Peng *et al.*, 2019). A pesar de que la unión del virus a CypA es indispensable para la infección del VIH, *el loop* de unión a CypA se ha considerado bastante robusto genéticamente (Rihn *et al.*, 2013). Esto también se ha confirmado en nuestro trabajo, ya que observamos que es una de las regiones de p24 con menor conservación en el consenso de consensos de VIH-1 (86,11%) y entre grupos del VIH-1 (90%). Hemos identificado 3 aa (P85, G89 y P90) en el *loop* totalmente conservados entre los 4 grupos del VIH-1 que, probablemente, cumplen una función esencial en el mantenimiento de su estructura o en la unión con CypA. Previamente ya se describió que 2 de ellos (G89 y P90), junto con A88, forman el sitio de enlace canónico a CypA (Gamble *et al.*, 1996). Sin embargo, en nuestro estudio, el residuo A88 sólo apareció conservado en el 100% de secuencias del grupo M, mientras que en los grupos no-M predominaba V88 con una conservación del 90.9% en grupo N, 75% en grupo P y 34.6% en grupo O. En un futuro se debería confirmar si el residuo A88 es también crítico para la unión de CypA en los grupos no-M o si estas variantes tienen aa alternativos de unión. De nuevo se refuerza la importancia de tener en cuenta a todas las variantes del VIH-1, y no sólo al subtipo B del grupo M, en este tipo de estudios. A su vez, confirmamos el alto nivel de conservación de MHR, motivo de 20 aminoácidos en la región CTD de p24 (aa 153-172), en todos los grupos del VIH-1 y del VIH-2, como se ha visto en todos los retrovirus. Ello se explicaría por ser indispensable para el correcto ensamblaje de los viriones, concretamente en el paso de estabilización del oligómero *gag* tras asociarse a la membrana, donde MHR forma parte de la interfaz intrahexamérica de la cápside inmadura. (Tanaka *et al.*, 2016). También nuestros consensos de VIH-1 y VIH-2 presentaban los dos aa Q155 y E159 totalmente conservados e invariables en todos los retrovirus. Aunque se encontró algún marcador de variante en esta región, no hubo cambios radicales de la naturaleza de los aa encontrados.

En nuestro trabajo encontramos marcadores genéticos en todas las variantes del VIH, y algunos eran compartidos por 2 o más variantes. Algunos grupos del VIH-1 compartían numerosos marcadores (grupos P y O y grupos M y N), sugiriendo mayor relación genética entre ellos. A su vez, algunos marcadores presentaron cambios radicales en la naturaleza de su aa con respecto al consenso. También se buscaron mutaciones en p24 relacionadas con una

alteración en el ensamblaje de la cápside inmadura del virus, cuyo hallazgo fue anecdótico, excepto para algunas variantes y posiciones concretas, pero siempre en una baja proporción (<0.6%). Se sabe que la cápside admite más cambios de aa en los residuos expuestos al exterior, mientras que las mutaciones en la parte interna de la cápside son raramente compatibles con la viabilidad (Perilla and Gronenborn, 2016). Dada la complejidad de la estructura de la cápside y el gran número de datos generados, no hemos podido profundizar en la significación de dichos hallazgos. Sin embargo, dado que las secuencias LANL proceden de plasma o células de pacientes infectados por virus infectivos y viables, se puede hipotetizar que dichos cambios de aa encontrados en residuos clave de p24 puedan estar compensados por otros cambios dentro o fuera de la CA que pudieran ejercer efectos a distancia y recuperar el *fitness* viral disminuido por efecto de dichas mutaciones encontradas. Otra posible explicación sería que los cambios se encontraran en regiones de menor relevancia para la conformación de una estructura tridimensional correcta de la cápside.

Finalmente se realizó el análisis filogenético de las distintas variantes del VIH utilizando las secuencias de la proteína gag p24 del VIH disponibles, observando que es una buena región para clasificar gran parte de las variantes. Este estudio es el primero que, empleando el mayor nº de variantes diferentes del VIH hasta la fecha, identifica a las variantes del VIH-1 que no se clasifican bien empleando AF de secuencias p24 y que, seguramente, caracterizaríamos mejor analizando secuencias gag más largas.

Nuestro estudio presentó limitaciones en cuanto al número de secuencias disponibles en LANL de algunas variantes, ausente para algunas de ellas y con bajo nº en otras, lo que dificultó la capacidad de extraer conclusiones para dichas variantes. A su vez, debido a que CA no ha sido tan estudiada como otras proteínas del VIH, aún existen lagunas y controversias con respecto a cuáles son todos los aa claves para el mantenimiento de su estructura y para la interacción con proteínas celulares, así como algunas funciones de las estructuras de la CA y su papel en la formación del virión maduro. En cuanto a perspectivas futuras, creemos que resultaría de gran interés el estudio de CA integrada dentro de gag empleando secuencias de gag completas y no sólo de p24/p26, lo que permitiría investigar la hélice 12 que conforman p24 y p2 y la interacción de CA con MA y NC a través de SP1 en la cápside inmadura. También sería interesante el estudio de los polimorfismos en el contexto de los epítomos antigénicos y la caracterización a mayor profundidad de las mutaciones halladas consideradas defectivas para el virus, estudios que serán continuados por el grupo.

CONCLUSIONES

1. La proteína CA que conforma la cápside del VIH presenta una alta conservación a lo largo de su secuencia en las 105 variantes del VIH-1 y VIH-2 analizadas.
2. Las estructuras de la cápside con mayor conservación son las hélices α 1, 2 y 5, mientras que la horquilla β , la hélice α 6 el *loop* de unión a CypA presentan la menor conservación.
3. A pesar de la elevada conservación de CA, se encontraron marcadores genéticos de variante en todos los tipos, subtipos, sub-subtipos y recombinantes del VIH-1 y VIH-2.
4. Se encontraron mutaciones en posiciones relevantes para la funcionalidad de estructuras destacadas de CA, como el *loop* de unión a ciclofilina A y el poro del hexámero.
5. Aunque de forma minoritaria, se encontraron mutaciones asociadas con un ensamblaje defectivo de la cápside inmadura que, supuestamente, darían lugar a virus defectivos.
6. CA es una buena región para caracterizar gran parte de las variantes del VIH-1.

7. Es necesario un mayor estudio de la estructura y funcionalidad de esta proteína que ofrece interesantes oportunidades para el desarrollo de nuevas técnicas diagnósticas y fármacos antirretrovirales dado su importante papel en el ciclo viral y su elevada conservación.
8. La herramienta bioinformática desarrollada en el laboratorio, EpiMolVIH ha resultado ser útil y práctica para el estudio de conservación de la proteína de la cápside viral.

BIBLIOGRAFÍA

- AMERICAN SOCIETY FOR MICROBIOLOGY & VERSALOVIC J. 2011. Manual of Clinical Microbiology. In: GRIFFITH, B., CAMPBELL, S. & CALIENDO A. M. (eds.) *Human Immunodeficiency Viruses*. Washington, DC: ASM Press. pp. 1302-1303.
- ALCAMÍ, J. & COIRAS, M. 2011. Inmunopatogenia de la infección por el virus de la inmunodeficiencia humana. *Enferm Infecc Microbiol Clin*, 29, 216–226.
- CAMPBELL, E. M. & HOPE, T. J. 2015. HIV-1 capsid: the multifaceted key player in HIV-1 infection. *Nat Rev Microbiol*, 13, 471–483.
- DOMINGO, E., SHELDON, J. & PERALES, C. 2012. Viral quasispecies evolution. *Microbiol Mol Biol Rev*, 76, 159-216
- GAMBLE, T. R., VAJDOS, F. F., YOO, S., WORTHYLAKE, D. K., HOUSEWEART, M., SUNDQUIST, W. I. & HILL, C. P. 1996. Crystal Structure of Human Cyclophilin A Bound to the Amino-Terminal Domain of HIV-1 Capsid. *Cell*, 87, 1285–1294.
- HEMELAAR, J. 2013. Implications of HIV diversity for the HIV-1 pandemic. *J Infec*, 66, 391-400.
- JACQUES, D. A., MCEWAN, W. A., HILDITCH, L., PRICE, A. J., TOWERS, G. J. & JAMES, L. C. 2016. HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis. *Nature*, 536, 349–353.
- LINGAPPA, J. R., REED, J. C., TANAKA, M., CHUTIRAKA, K. & ROBINSON, B.A. 2014. How HIV-1 Gag assembles in cells: Putting together pieces of the puzzle. *Virus Res*, 193, 89–107.
- LIU, C., PERILLA, J. R., NING, J., LU, M., HOU, G., RAMALHO, R., HIMES, B. A., ZHAO, G., BEDWELL, G. J., BYEON, I. J., AHN, J., GRONENBORN, A. M., PREVELIGE, P.E., ROUSSO, I., AIKEN, C., POLENOVA, T., SCHULTEN, K. & ZHANG, P. 2016. Cyclophilin A stabilizes the HIV-1 capsid through a novel non-canonical binding site. *Nat Commun*, 7, 10714.
- NOVIKOVA, M., ZHANG, Y., FREED, E. O. & PENG, K. 2019. Multiple Roles of HIV-1 Capsid during the Virus Replication Cycle. *Viol Sin*, 34, 119–134.
- OBR, M. & KRÄUSSLICH, H. G. 2018. The secrets of the stability of the HIV-1 capsid. *Elife*, 7, e38895.

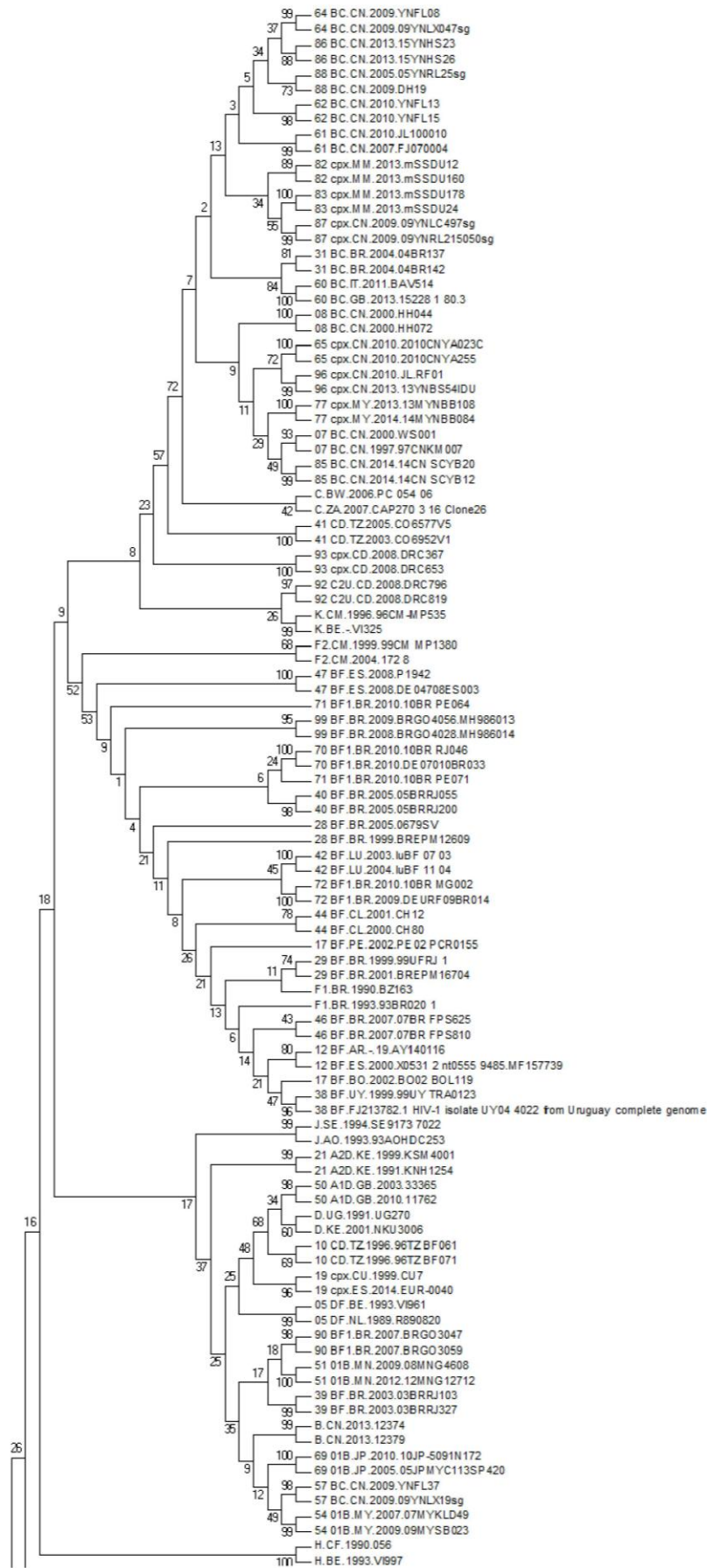
- PENG, W., SHI, J., MÁRQUEZ, C. L., LAU, D., WALSH, J., FAYSAL, K. M. R., BYEON, C. H., BYEON, I. L., AIKEN, C. & BÖCKING, T. 2019. Functional analysis of the secondary HIV-1 capsid binding site in the host protein cyclophilin A. *Retrovirology*, 16, 10.
- PERILLA, J. R. & GRONENBORN, A. M. 2016. Molecular Architecture of the Retroviral Capsid. *Trends Biochem Sci*, 41, 410–420.
- QUINN, C. M., WANG, M., FRITZ, M. P., RUNGE, B., AHN, J., XU, C., PERILLA, J. R., GRONENBORN, A. M. & POLENOVA, T. 2018. Dynamic regulation of HIV-1 capsid interaction with the restriction factor TRIM5 α identified by magic-angle spinning NMR and molecular dynamics simulations. *Proc Natl Acad Sci U S A*, 115, 11519-11524.
- RIHN, S. J., WILSON, S. J., LOMAN, N.J., ALIM, M., BAKKER, S. E., BHELLA, D., GIFFORD, R. J., RIXON, F.J. & BIENIASZ, P. D. 2013. Extreme genetic fragility of the HIV-1 capsid. *PLoS Pathog*, 9, e1003461
- ROBERTSON, D. L., ANDERSON, J. P., BRADAC, J. A., CARR, J. K., FOLEY, B., FUNKHOUSER, R. K., GAO, F., HAHN, B. H., KALISH, M. L., KUIKEN, C., LEARN, G. H., LEITNER, T., MCCUTCHAN, F., OSMANOV, S., PEETERS, M., PIENIAZEK, D., SALMINEN, M., SHARP, P. M., WOLINSKY, S. & KORBER, B. 2000. HIV-1 nomenclature proposal. *Science*, 288, 55-56.
- TANAKA, M., ROBINSON, B. A., CHUTIRAKA, K., GEARY, C. D., REED, J. C. & LINGAPPA, J. R. 2016. Mutations of Conserved Residues in the Major Homology Region Arrest Assembling HIV-1 Gag as a Membrane-Targeted Intermediate Containing Genomic RNA and Cellular Proteins. *J Virol*, 90, 1944–1963.
- TORRECILLA, E., LLÁCER DELICADO, T. & HOLGUIN, A. 2014. New findings in cleavage sites variability across groups, subtypes and recombinants of human immunodeficiency virus type 1. *PLoS One*, 9, e88099.
- UNAIDS. 2018. UNAIDS data 2018. URL <https://www.unaids.org/en/resources/documents/2018/unaid-data-2018.html>
- VISSEAU, B., DAMOND, F., MATHERON, S., DESCAMPS, D. & CHARPENTIER, C. 2016. Hiv-2 molecular epidemiology. *Infect Genet Evol*, 46, 233-240.

ABREVIATURAS

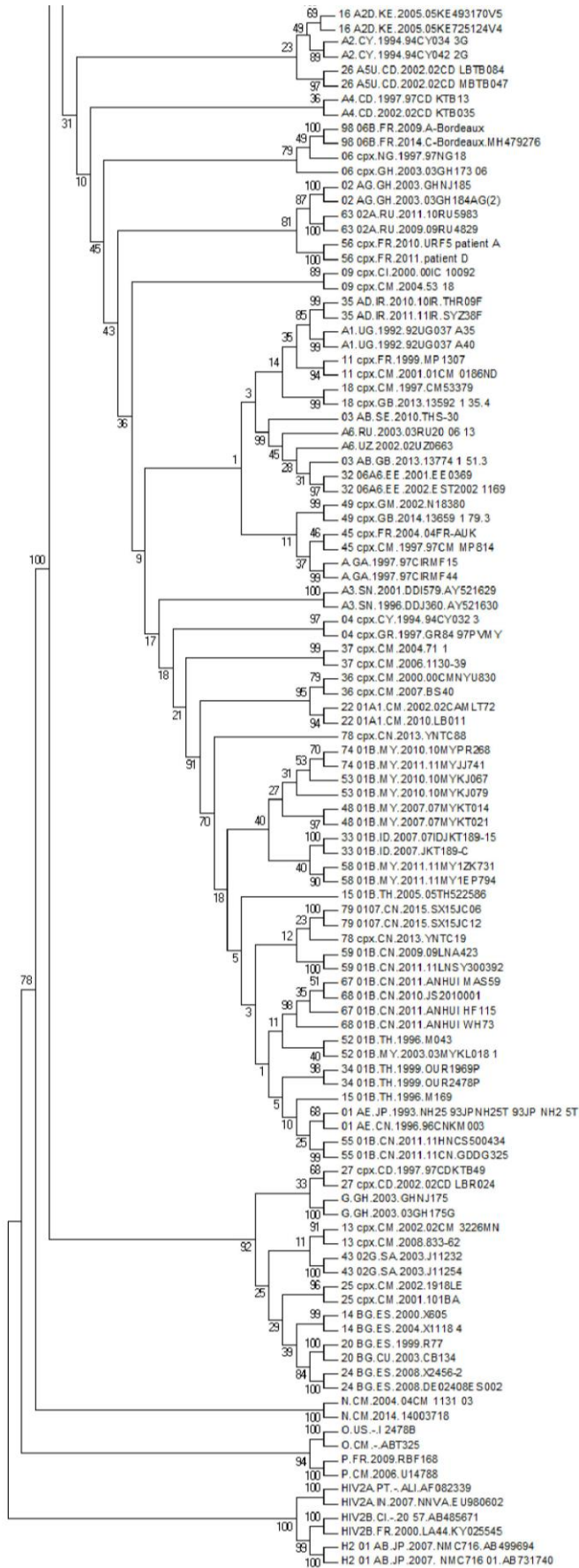
VIH	Virus de la inmunodeficiencia humana
SIDA	Síndrome de Inmunodeficiencia Adquirida
VIS	Virus de la Inmunodeficiencia de Simios
RT	Retrotranscriptasa viral
PR	Proteasa viral
IN	Integrasa viral
NC	Nucleocápside
MA	Matriz
CA	Cápside o proteína de la cápside
ARN	Ácido ribonucleico
ADN	Ácido desoxirribonucleico
ARNm	ARN mensajero
LTR	Secuencias largas repetidas virales
CS	Sitios de procesamiento proteico o <i>cleavage sites</i>
CRF	Forma recombinante circulante
URF	Forma recombinante única
NTD	Dominio N-terminal
CTD	Dominio C-terminal
H	Hélice α
RUI	Región de Unión Interdominio
MHR	Región de Homología Mayor
CypA	Ciclofilina A
AF	Análisis filogenético

Aminoácidos:

A	Alanina
C	Cisteína
D	Ácido aspártico
E	Ácido glutámico
F	Fenilalanina
G	Glicina
H	Histidina
I	Isoleucina
K	Lisina
L	Leucina
M	Metionina
N	Asparagina
P	Prolina
Q	Glutamina
R	Arginina
S	Serina
T	Treonina
V	Valina
W	Triptófano
Y	Tirosina



Anexo II. Análisis filogenético de secuencias de p24 del VIH (continúa en la página siguiente).



Anexo II. Análisis filogenético de secuencias de p24 del VIH (continuación de la página anterior).