

UACM

Universidad Autónoma
de la Ciudad de México

Nada humano me es ajeno

UNIVERSIDAD AUTÓNOMA DE LA CIUDAD DE MÉXICO

COLEGIO DE CIENCIA Y TECNOLOGÍA

POSGRADO DE CIENCIAS GENÓMICAS

**EFFECTO DE LA RECOMBINACIÓN GENÉTICA INTRASUBTIPO EN LA
EVOLUCIÓN DE GEN *ENV* EN EL VIRUS DE INMUNODEFICIENCIA
HUMANA -1**

TESIS

QUE PARA OPTAR POR EL GRADO DE:

MAESTRA EN CIENCIAS GENÓMICAS

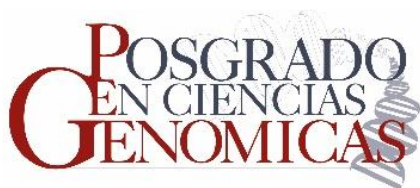
PRESENTA:

Biol. Anahí Tania Sánchez León

DIRECTORA

Dra. Claudia Selene Zárate Guerra

Ciudad de México, Enero, 2023



SISTEMA BIBLIOTECARIO DE INFORMACIÓN Y DOCUMENTACIÓN



UNIVERSIDAD AUTÓNOMA DE LA CIUDAD DE MÉXICO COORDINACIÓN ACADÉMICA

RESTRICCIONES DE USO PARA LAS TESIS DIGITALES

DERECHOS RESERVADOS[©]

La presente obra y cada uno de sus elementos está protegido por la Ley Federal del Derecho de Autor; por la Ley de la Universidad Autónoma de la Ciudad de México, así como lo dispuesto por el Estatuto General Orgánico de la Universidad Autónoma de la Ciudad de México; del mismo modo por lo establecido en el Acuerdo por el cual se aprueba la Norma mediante la que se Modifican, Adicionan y Derogan Diversas Disposiciones del Estatuto Orgánico de la Universidad de la Ciudad de México, aprobado por el Consejo de Gobierno el 29 de enero de 2002, con el objeto de definir las atribuciones de las diferentes unidades que forman la estructura de la Universidad Autónoma de la Ciudad de México como organismo público autónomo y lo establecido en el Reglamento de Titulación de la Universidad Autónoma de la Ciudad de México.

Por lo que el uso de su contenido, así como cada una de las partes que lo integran y que están bajo la tutela de la Ley Federal de Derecho de Autor, obliga a quien haga uso de la presente obra a considerar que solo lo realizará si es para fines educativos, académicos, de investigación o informativos y se compromete a citar esta fuente, así como a su autor ó autores. Por lo tanto, queda prohibida su reproducción total o parcial y cualquier uso diferente a los ya mencionados, los cuales serán reclamados por el titular de los derechos y sancionados conforme a la legislación aplicable.

INTEGRACIÓN DEL JURADO.

Presidente: Dra. Rosa Martha Eugenia Yocupicio Monroy. UACM

Secretario: Dra. Claudia Selene Zárate Guerra. UACM

Vocal: Dr. José Alberto Campillo Balderas. UNAM

Plantel de adscripción:

PLANTEL DEL VALLE, UACM.

DIRECTOR

Dra. Claudia Selene Zárate Guerra. UACM

Universidad Autónoma de la Ciudad de México.

Agradecimientos.

Agradezco al Posgrado en Ciencias Genómicas de la UACM

A la beca número 739180 Otorgada por CONACYT, Convocatoria de Becas CONACYT Nacionales 2021, para la realización de mi proyecto de maestría con duración de 24 meses.

A los miembros de mi comité tutorial: la Dra Martha Yocupicio, la Dra Selene Zárate y el Dr Campillo, por su tiempo.

Dedicatoria.

A mis bebés TOGO+, BALTO y ATENA que me han hecho ser una mejor persona y que cada día me han levantado para permanecer en este mundo y continuar con ellos. Los incondicionales <3.

Si pudiéramos ver a través del cráneo la fragilidad de cerebro podríamos notar el diseño de la imaginación y el potencial, que es tan fácil de tirar y derrumbar para nunca volver al mismo lugar.

ÍNDICE

- ii. ÍNDICE DE TABLAS
- iii. INIDCE DE FIGURAS
- iv. ABREVIATURAS
- v. RESUMEN EN ESPAÑOL
- vi. ABSTRACT

1. INTRODUCCIÓN.

- 1.1.1. DESCUBRIMIENTO DEL VIRUS DE INMUNODEFICIENCIA HUMANA
- 1.1.2. EPIDEMIOLOGÍA EN EL MUNDO
- 1.1.3. GENERALIDADES DEL VIH
- 1.1.4. ORIGEN Y CLASIFICACIÓN
- 1.1.5. ESTRUCTURA DEL VIH
- 1.1.6. GENOMA Y PROTEÍNAS
- 1.1.7. PROTEÍNAS ESTRUCTURALES
- 1.1.8. CICLO DE REPLICACIÓN
- 1.1.9. RECONOCIMIENTO Y ADHESIÓN:
- 1.1.10. ENTRADA:
- 1.1.11. PROTEÍNAS REGULADORAS
- 1.1.12. PROTEÍNAS ACCESORIAS
- 1.1.13. HISTORIA NATURAL DE LA ENFERMEDAD
- 1.1.14. EL VIH Y LA RESPUESTA DE LINFOCITOS T CITOTÓXICOS

2. ANTECEDENTES

2.1.1. MODELOS DE RECOMBINACIÓN.

2.1.2. RECOMBINACIÓN Y HOTSPOTS

3. HIPÓTESIS.

4. JUSTIFICACIÓN.

5. OBJETIVOS

5.1.1. OBJETIVO GENERAL

5.1.2. OBJETIVOS PARTICULARES

6. ESTRATEGIA EXPERIMENTAL.

7. MATERIALES Y MÉTODOS.

7.1.1. SECUENCIAS

7.1.2. DETECCIÓN DE CLÚSTERS DE TRANSMISIÓN

7.1.3. PRUEBA DE ESTADÍSTICO F

7.1.4. DETECCIÓN DE RECOMBINACIÓN

7.1.5. PREDICCIÓN DE ESTRUCTURA SECUNDARIA

7.1.6. ANÁLISIS DE POSICIONES BAJO SELECCIÓN NATURAL

7.1.7. ANALISIS DE VARIABILIDAD GENÉTICA

7.1.8. BÚSQUEDA DE EPÍTOPOS CTL/CD8

7.1.9. ANÁLISIS DE SITIOS DE GLICOSILACIÓN

7.1.10. ANÁLISIS DE USO DE CORRECEPTOR

8. RESULTADOS

- 8.1.1. DESCRIPCIÓN DE SECUENCIAS DEL GEN ENV DEL VIH-1 SUBTIPO B
- 8.1.2. DETECCIÓN DE CLÚSTERS DE TRANSMISIÓN
- 8.1.3. ESTADÍSTICA DE LOS CLÚSTERS
- 8.1.4. EXTRACCIÓN DE SECUENCIAS POR CLÚSTERS PARA SU ALINEAMIENTO
- 8.1.5. PRUEBA DE ESTADÍSTICO F
- 8.1.6. ANÁLISIS DE RECOMBINACIÓN
- 8.1.7. DISTINTOS PATRONES DE POSICIONES BAJO PRESIÓN SELECTIVA POSITIVA EN LAS SECUENCIAS
- 8.1.8. DIFERENTES PATRONES DE VARIABILIDAD ENTRE SECUENCIAS RECOMBINANTES Y NO RECOMBINANTES
- 8.1.9. BÚSQUEDA DE EPÍTOPOS DE CTL/CD8
- 8.1.10. EPÍTOPOS Y LOCALIZACIÓN EN LAS SECUENCIAS.
- 8.1.11. ANÁLISIS DE SITIOS DE GLICOSILACIÓN
- 8.1.12. ANÁLISIS DE ESCUDO DE GLICANOS
- 8.1.13. ANÁLISIS DE USO DEL USO DE CORRECEPTOR

9. DISCUSIÓN.

10. CONCLUSIONES

- vii. BIBLIOGRAFÍA O FUENTES DE CONSULTA.
- viii. ANEXOS

Índice de tablas.

- Tabla 1. Estadística regional de Europa y Norteamérica, México y el mundo 2021.
- Tabla 2. Características de la búsqueda.
- Tabla 3. Parámetros utilizados para la inferencia de Red
- Tabla 4. Sitios de glicosilación por número de secuencias.
- Tabla 5. Número de secuencias por año y región
- Tabla 6. Secuencias totales por región y número de cluster por año.
- Tabla 7 Descripción de los grupos para Norte América, Estados Unidos
- Tabla 8. Descripción de los grupos para Europa/Alemania
- Tabla 9. Estadística descriptiva.
- Tabla 10. Secuencias totales por país para el análisis de Recombinación.
- Tabla 11. Clusters dónde no se detectó recombinación.
- Tabla 12. Número de secuencias recombinantes por cluster.
- Tabla 13. Número de breakpoints por año y frecuencias de recombinación.
- Tabla 14. Breakpoints por estructura.
- Tabla 15. Posiciones bajo selección positiva y negativa por grupo entre las secuencias recombinantes y no recombinantes de ambos análisis.
- Tabla 16. Porcentaje de epítomos en las secuencias recombinantes y No recombinantes, mostrando en forma descendente de cada una para mejor comparación.
- Tabla 17. Mapa de epítomos reportados en pacientes de Estados Unidos y su posición en la secuencia de referencia
- Tabla 18 Epitopes identificados en los hotspot y su proporción.
- Tabla 19. Sitios de glicosilación identificados en los hotposts.
- Tabla 20. Distribución de agujeros de glicano en secuencias recombinantes y no recombinantes a lo largo de las estructuras gp120 y 41.
- Tabla 21. Valores de predicción de uso de correceptor

Índice de figuras.

- Figura 1. Descripción esquemática de la clasificación del VIH.
- Figura 2. Subtipos de VIH-1 en su distribución Global.
- Figura 3. Esquema de la organización del genoma del VIH I.
- Figura 4: Esquema de la estructura del VIH-1.

- Figura 5. Representación lineal de la glicoproteína de Env.
- Figura 6. Descripción esquemática de los eventos que ocurren durante el ciclo de replicación del VIH.
- Figura 7. Historia natural de la enfermedad causada por VIH.
- Figura 8. Naturaleza pseudodiploide del VIH.
- Figura 9. Proceso de transcripción reversa del VIH-1.
- Figura 10. Tres modelos de recombinación retroviral.
- Figura 11. Muestra un clúster de Norte América, año 2015 con VIH-TRACE.
- Figura 12. HeatMap de los puntos de recombinación.
- Figura 13. Sitios bajo selección positiva en el esquema de breakpoints.
- Figura 14. Mutaciones sinónimas y no sinónimas en la secuencia consenso del grupo de recombinantes por el método SLAC.
- Figura 15. Entropía de Shannon en las secuencias del gen *env*.
- Figura 16. Mapa de epítomos reportados en pacientes de Estados Unidos y su posición en la secuencia de referencia. Las líneas azul marino indican los sitios de hotspots.
- Figura 17. Estructura de gp120.
- Figura 18. Muestra el escudo de glicano predicho desde 3 puntos de vista.

Abreviaturas.

a.a: Aminoácido

CRF: Formas Recombinantes Circulantes, del inglés.

CTLs: Linfocitos T Citotóxicos, del inglés.

ADN: Ácido Desoxirribonucleico.

dS: Sustituciones Sinónimas, en inglés

OMS: Organización Mundial de la Salud.

ORF: Fragmentos de Lectura Abierta, en inglés

ARN: Ácido Ribonucleico.

RT -> Retro-transcriptasa.

SIDA -> Síndrome de Inmunodeficiencia Adquirida.

VIH -> Virus de Inmunodeficiencia Humana.

RESUMEN

El VIH-1 tiene la tasa de recombinación más alta reportada para cualquier entidad biológica, y este proceso contribuye en gran medida a la variabilidad del genoma viral. Aunque la recombinación puede ocurrir en todo el genoma, se ha reportado que existe un punto de alta frecuencia de recombinación ubicado en el gen *env*, que codifica las proteínas de superficie gp120 y gp41.

Dado que la recombinación ocurre en cada ciclo de transcripción inversa, su detección depende de la frecuencia de superinfecciones o coinfecciones con virus genéticamente distinguibles en una población.

El objetivo de este trabajo ha sido analizar el efecto de la recombinación genética intrasubtipo en la evolución del gen *env* en el virus de inmunodeficiencia humana-1

Las secuencias se obtuvieron de la base de datos de VIH de Los Alamos, usando como criterios que las secuencias Env estuvieran completas, y se tratara de virus que circularon en dos regiones en un período de 4 años consecutivos, para maximizar la probabilidad de detección de recombinación. Posteriormente, las secuencias se analizaron con VIH-TRACE para identificar grupos de transmisión. Posteriormente se usó RAPR para detectar eventos de recombinación dentro de cada grupo y finalmente se realizaron análisis de detección de selección natural, de diversidad, presencia de epítopes CTL, y de sitios de glicosilación y predicciones de uso de correceptor comparando las secuencias recombinantes y no recombinantes.

Se lograron identificar 70 clusters de transmisión pertenecientes a Estados Unidos y Alemania, y se analizaron aquellos con más de 20 secuencias. Con estos datos se calculó que la frecuencia de recombinación intrasubtipo era de 13.4 eventos de recombinación por genoma. Además, se encontró que las secuencias recombinantes tienden a ser más variables, tener más sitios de glicosilación, y lo que podría indicar que la recombinación podría contribuir al escape de sistema inmune del VIH.

ABSTRACT.

VIH-1 has the highest recombination rate reported, and this process greatly contributes to the variability of the viral genome. Although recombination can occur throughout the genome, the presences of a recombination hotspot located in the gene env, which encodes for the surface proteins gp120 and gp41, has been reported.

Since recombination events occur at each reverse transcription cycle its detection depends on the frequency of superinfections or coinfections with genetically distinct viruses in a population.

The aim of this study was to analyze the effect of intra-subtype genetic recombination on the evolution of the env gen in the human immunodeficiency virus-1. To this end sequence data was gathered from Los Alamos VIH Databases, with the following criteria: full Env sequences from two geographical regions collected in four consecutive years. Subsequently, the data sets were analyzed with VIH-TRACE to identify transmission groups. RAPR was then used to detect recombination events within each group, and analysis of natural selection detection, diversity, CTL epitope detection, glycosylation site detection, and co-receptor analysis were performed between recombinant and non-recombinant sequences.

In this analysis more than 70 transmission groups were identified, which corresponded to sequences from the United States and Germany. Data sets with more than 20 sequences were analyzed for recombination, estimating a recombination rate of 13.4 recombination events per genome for 417 recombinant sequences. Additionally, the diversity a number of glycosilation site is larger in recombinant sequences suggesting a role for recombination in immune escape.

INTRODUCCIÓN.

A casi 39 años de la identificación del virus de inmunodeficiencia humana (VIH), la evolución de este virus ha sido objeto de investigaciones por las consecuencias clínicas y epidemiológicas que ha tenido su diversificación conforme se ha esparcido a distintas regiones, países, grupos étnicos, hasta en estratos socioeconómicos. Además, dentro de cada paciente infectado el VIH evoluciona continuamente, infectando diversas poblaciones celulares, estableciendo nichos de latencia, y generando subpoblaciones capaces de escapar del control inmunológico. El VIH fue identificado en 1983, y se demostró que era la causa del síndrome de inmunodeficiencia adquirida (SIDA) en 1984, y los esfuerzos de investigación se han enfocado a la generación de fármacos antivirales y la creación de una vacuna, por lo que se identificaron las proteínas virales que pueden ser el blanco de fármacos y anticuerpos que permiten controlar y prevenir la infección [1].

DESCUBRIMIENTO DEL VIRUS DE INMUNODEFICIENCIA HUMANA

El VIH-1 se identificó a partir de la descripción de una nueva enfermedad en la década de los 80's en jóvenes y adultos previamente sanos que no debieran presentar inmunodeficiencias, se trataba principalmente de individuos del sexo masculino con afecciones como diarrea crónica, neumonías, linfadenopatías, y cánceres inusuales como el sarcoma de Kaposi o linfomas No-Hodgkins [2,3]. Los síntomas poco a poco se fueron asociando en sectores específicos de la población como usuarios de drogas inyectadas, homosexuales, trabajadoras sexuales y personas que recibieron transfusiones sanguíneas, además de una alta incidencia de transmisión materno-fetal [4]. Esta información permitió establecer que el riesgo de transmisión aumentaba con el contacto con sangre de personas infectadas y prácticas sexuales sin protección. Adicionalmente, los estudios de laboratorio indicaron una disminución significativa en el recuento de las células sanguíneas y linfocitos TCD4+ [5, 2, 6].

En 1986 el Doctor Luc Montagnier y colaboradores en el instituto Pasteur de Francia, reportaron el aislamiento de un nuevo agente a partir de nódulos linfáticos de individuos asintomáticos, que presentaban linfadenopatía generalizada de origen desconocido, posteriormente asignada como LAV o linfadenopatía asociada a virus por sus siglas en inglés, además se detectó una alta actividad de transcriptasa reversa magnesio dependiente (RT, reverse transcriptase) característica de retrovirus previamente

reportados, y un tropismo por células mononucleares de sangre periférica (PBMcs) específicamente linfocitos TCD4+ [14]. De manera paralela el grupo de trabajo del doctor Robert Gallo en el National Cancer Institute en Estados Unidos también aisló el agente y pudo demostrar que este virus era el causante de las afecciones reportadas [5, 7, 8].

EPIDEMIOLOGÍA EN EL MUNDO

Mundialmente, el VIH / SIDA continúa siendo una pandemia global que representa un gran problema de salud pública de primer orden, con grandes repercusiones demográficas, sociales y económicas. Esto se constata cada año con los datos de la ONUSIDA del 2022, se reportó que en el año 2021 existían 38.4 millones [33.9–43.8 millones] de personas viviendo con el VIH en el mundo. En ese mismo año, alrededor de 1.5 millones [1.0 –2.0 millones] de personas se infectaron por primera vez con el VIH y 860,000 personas fallecieron por enfermedades que se relacionan con el SIDA [9].

Desde la identificación del VIH, las infecciones han crecido exponencialmente y el desarrollo de medidas preventivas y terapéuticas se volvieron prioridad. En 2021, el 75% [56%–88%] de toda la población que vive con VIH obtuvo acceso al tratamiento. Desde el pico alcanzado en 1997, las nuevas infecciones por el VIH se han reducido en un 54%, aun así, los nuevos casos no cesan y tristemente el 49% de todas las nuevas infecciones es representado por mujeres y niñas en 2021. Otro dato actual y perturbador en el progreso de esta epidemia fue el hecho de tener confinamiento y diversas restricciones que se han dado por otra pandemia en desarrollo como lo es la de COVID-19 la cual llevó a interrumpir las pruebas y tratamientos de VIH, lo cual provocó fuertes caídas en los diagnósticos y por consiguiente retraso en el inicio del tratamiento contra el VIH.

Se sabe que el número de nuevas infecciones se ha incrementado en el este de Europa y en el sur y centro de Asia en un 37%. En 2019, se estimaron 34,800 nuevas infecciones por VIH en Estados Unidos [10]. Para el 2020 se estima que en Alemania hay 110,000 personas viviendo con VIH con el aproximadamente 83% con acceso a tratamientos retrovirales y para ese mismo año hubo 3,100 nuevas infecciones (Tabla 1) [9, 10].

Tabla 1. Estadística regional de Europa y Norteamérica, México y el mundo 2021.

Región	Personas que viven con VIH en el 2021	Número de infecciones por VIH, 2021 total	Número de infecciones por VIH con más de 15 años	Muertes relacionadas con el SIDA 2021	Personas con acceso al tratamiento 2021	Porcentaje de adultos mayores a 15 años que viven con VIH con acceso a terapia antirretroviral.
Europa Occidental, Central y América del Norte	2.2 millones [1.9 millones– 2.6 millones]	67 000 [53 000– 81 000]	66 000 [53 000– 80 000]	13 000 [9200– 17 000]	1.9 millones [1.8 millones– 1.9 millones]	83% [68–97%]
México	360 000 [320 000 - 400 000]	18 000 [14 000 - 23 000]	17 000 [13 000 - 22 000]	4600 [3600 - 5700]	220 000	61% [54 - 68]
Totales mundiales	37.7 millones [30.2 millones– 45.1 millones]	1.5 millones [1.0 millones– 2.0 millones]	1.3 millones [910 000– 1.8 millones]	680 000 [480 000– 1.0 millones]	27.5 millones [26.5 millones– 27.7 millones]	73% [56–88%]

Fuente: Estimaciones epidemiológicas de ONUSIDA 2021

GENERALIDADES DEL VIH

ORIGEN Y CLASIFICACIÓN

El origen del VIH ha sido el resultado de diferentes saltos zoonóticos de retrovirus entre humanos y monos. El SIDA es causado por dos tipos de VIH: VIH-1 y VIH-2, el primero se distribuye por todo el mundo y el segundo es endémico de África Oriental, en este último las afecciones inmunológicas son distintas y con una progresión lenta, periodos asintomáticos más largos y una tasa de mortalidad menor [11].

Se ha descrito que el VIH tipo 1 se origina de otro muy similar que es el virus de inmunodeficiencia simiana del chimpancé (SIVcpz, por sus siglas en inglés), [5, 7, 11,12]. Utilizando modelos evolutivos que utilizan reloj molecular, se pudo estimar cuando que el salto de especie tuvo lugar hacia principios del siglo XX con más de una introducción a la población humana, posiblemente ocasionados por la caza de chimpancés destinados a su consumo [7, 12]. Por otro lado, el VIH tipo 2 tiene su origen del SIV, pero del mono mangabey ahumado (sooty mangabey en inglés, SIVsmm) [13] teniendo una historia evolutiva distinta.

Cada uno de los eventos de transmisión zoonótica ha generado un linaje distinto (denominados grupos) del VIH: dentro del VIH-1 se encuentran los grupos M (“main”), O (“outlier”), N (“new”) y P; mientras que en el VIH-2 los grupos van de la A a la H. El grupo M perteneciente al VIH-1 es responsable de la pandemia y se ha diversificado en nueve subtipos (A, B, C, D, F, G, H, J, K) (Figura 1), los cuales tienen algunas diferencias en su biología y en su distribución geográfica [13].

En cuanto a la prevalencia de este virus, Sudáfrica, en particular, sufre la peor parte de la epidemia del VIH-1, donde aproximadamente el 20% de los adultos viven con VIH-1 y más del 98% de las infecciones son causadas por el VIH-1 subtipo C. A pesar de esto, la mayoría de las investigaciones sobre el VIH-1 se han centrado en el subtipo B, que constituye solo

el 12% de las infecciones y es prevalente en Europa occidental y central, América y Oceanía (Fig 2) [13].

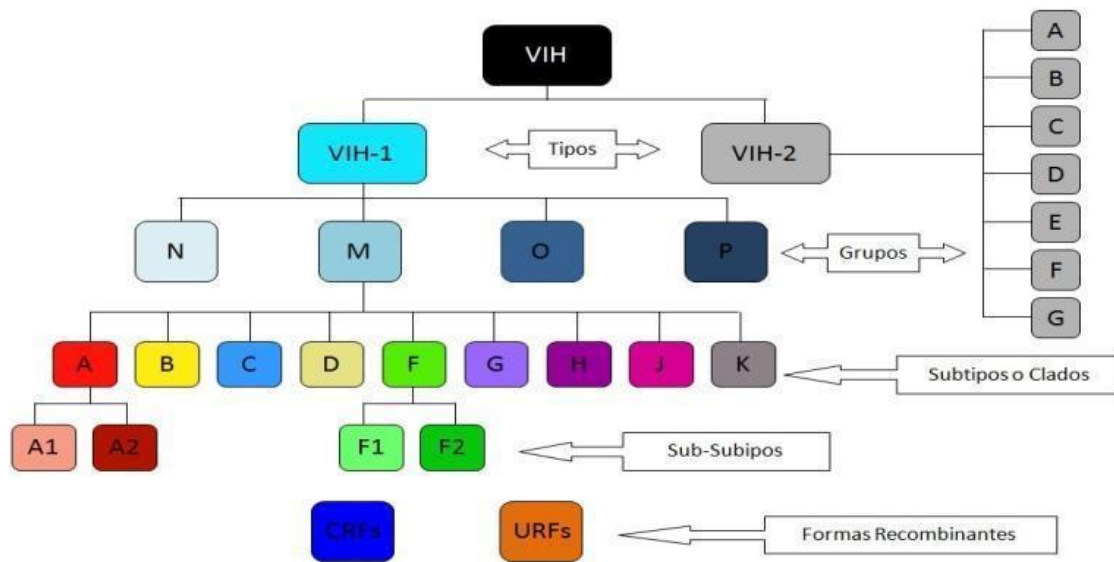


Figura 1. Descripción esquemática de la clasificación del VIH. Esta se basa en tipos, grupos, subtipos, sub-subtipos y formas recombinantes.

Se ha podido describir que las altas tasas de mutación en el virus es lo que proporciona una gran variación creando linajes que son distintos entre sí pero que coexisten en cualquier momento en un individuo infectado, identificando a estas variaciones como cuasiespecies, que contribuyen de manera importante al escape del control inmunológico y a la generación de virus resistentes a los fármacos antivirales [1].

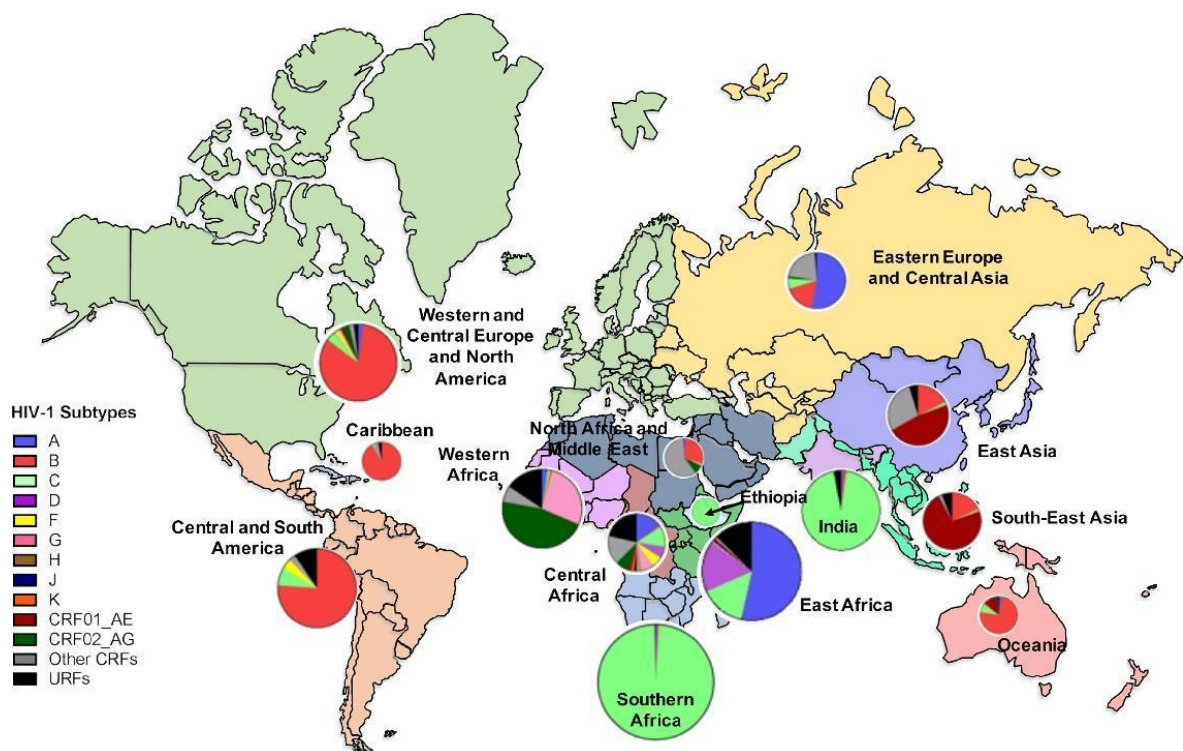


Figura 2. Subtipos de VIH-1 en su distribución Global. Se representa con círculos de colores los diferentes subtipos y el tamaño representa el número de personas que viven infectadas en cada región. En Europa y el continente americano el subtipo que predomina es el B [12].

ESTRUCTURA DEL VIH GENOMA Y PROTEÍNAS

El VIH taxonómicamente se ubica en el género *Lentivirus*, el cual pertenece a la familia *Retroviridae*. Estos virus se van a caracterizar por tener un genoma que presenta dos copias de una molécula de ARN+ de cadena sencilla, de ~10 kb en un virión, por lo que se denominan pseudodiploides. Los retrovirus particularmente se encuentran en vertebrados, y el ciclo biológico de estas entidades va a depender de una polimerasa, conocida como transcriptasa reversa (RT), que utiliza el genoma viral de ARN como templado y sintetiza una molécula de ADN de doble cadena, llamada provirus. El ADN proviral se integra al genoma de la célula hospedera y por ambos lados contiene secuencias largas repetidas (o LTR, por sus iniciales en inglés) que se generan durante la retro-transcripción [1,14].

El VIH en su genoma va a incluir tres genes, los cuales son primordiales de los retrovirus, los cuales van a codificar para proteínas estructurales y enzimas necesarias implicadas en la replicación y maduración del virus. Los 3 genes principales son: *gag* (gen de antígeno grupo-específico), *pol* (polimerasa) y *env* (envoltura). Este genoma también contiene

marcos abiertos de lectura (ORF) que están traslapados y que codifican para proteínas reguladoras o accesorias (Figura 3). Las proteínas que cumplen un papel de regulación, que son: Tat y Rev, se involucran en los mecanismos tempranos de transcripción y splicing; además, las tenemos proteínas accesorias como: Vpu, Vpr, Nef y Vif, que benefician los procesos para la integración del genoma lo que permite la replicación del virus *in vivo* y la evasión de la respuesta antiviral (Figura 4) [14,16].

Las LTRs van a servir como enlace al genoma celular lo que va a facilitar la unión de diversos factores en la transcripción de los genes virales [15].

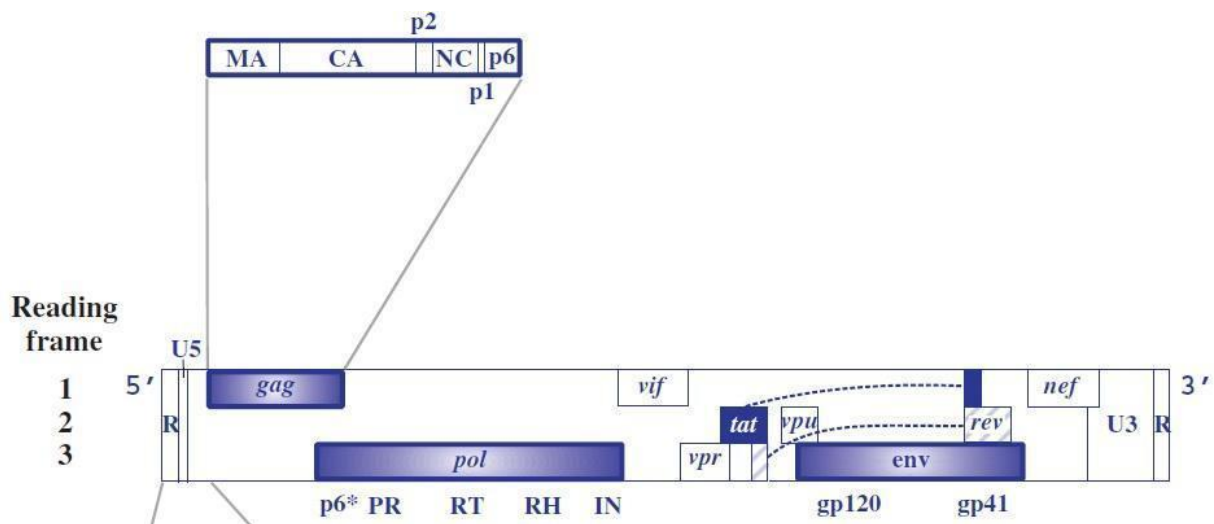


Figura 3. Esquema de la organización del genoma del VIH-1. Se muestran los tres marcos principales de lectura abiertos (ORF) los cuales son: *gag*, *pol* y *env* y seis ORFs que codifican para las proteínas accesorias. [16].

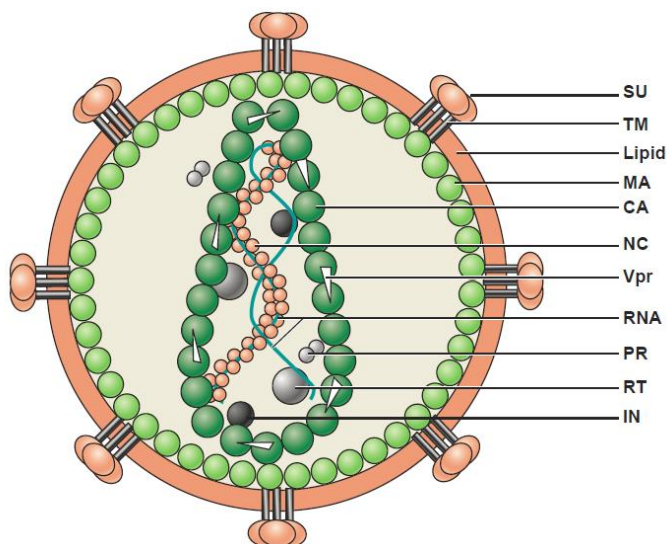


Figura 4: Esquema de la estructura del VIH-1. Se muestran las proteínas que componen la estructura del virión. SU: proteína de superficie, TM: transmembrana, MA: matriz, CA: cápside, NC: nucleocápside, PR: proteasa, RT: retrotranscriptasa, IN, integrasa [16].

PROTEÍNAS ESTRUCTURALES

El gen *gag* precursora de la poliproteína Gag (Pr55Gag) da lugar a seis proteínas maduras: p1, p2, p6, p7, p17 y p24 [17]. CA (p24) forma la cápside del virus, la cual va a favorecer las interacciones entre Gag. p7 forma la nucleocápside (NC) que facilita la transcripción reversa, protege al genoma de las nucleasas, facilita la incorporación de un tARN que funciona como iniciador en la retrotranscripción a la partícula viral, y que es fundamental para contener el genoma del virus.

El gen *pol* es una poliproteína que se conoce como Gag-Pol (Pr160/Gag-Pol) y da lugar a tres enzimas: proteasa (PR), transcriptasa reversa (RT) e integrasa (IN) [17,18]. La proteasa será la encargada de procesar los polipéptidos en el proceso de gemación del nuevo virión y esto permite que ocurran cambios morfológicos y estructurales durante la maduración viral [15]. La retrotranscriptasa es un ADN-polimerasa dependiente de ARN y ADN que lleva a cabo actividad de ARNasH que va a degradar la cadena original de ARN viral, sin actividad de exonucleasa 3'-5'. La RT tiene una baja fidelidad lo cual da espacio a una alta tasa de mutación $\sim 3.4 \times 10^{-5}$ errores por ciclo de replicación, y esto provoca recombinación durante el proceso de la retrotranscripción ya que se está cambiando de molde al usar ambas cadenas de ARN viral. Por último, la integrasa (IN) es la que permite que se integre el genoma del virus dentro de la célula hospedero [14].

El gen *env* codifica para un polipéptido que es una glicoproteína de 160kDa (gp160), este se fragmenta a través de una proteasa celular que da lugar a dos glicoproteínas, una de superficie (SU) o gp120 la cual se encuentra, como su nombre lo indica, en la superficie del virus y la otra glicoproteína de gp41 o transmembranal (TM) que se localiza justo en la región transmembranal del virión. SU o gp120 contiene cinco dominios variables (V1-V5) y cinco dominios conservados (C1-C5). gp41 consta de un dominio extracelular, que contiene el péptido de fusión (FP), regiones de héptadas repetidas (HR1 y HR2), la región externa proximal de la membrana (MPER), un dominio transmembranal (TMD) y una cola citoplásmica (CT). Se muestra una representación ampliada de la gp41 CT para resaltar varios motivos: la señal de internalización YSPL, la secuencia de Kennedy (ks), las hélices

α anfipáticas LLP-1, -2, -3 y un motivo de dileucina terminal C (LL) implicado en la endocitosis y la distribución intracelular de Env (Figura 5) [5, 19].

Estas proteínas al asociarse con gag van a formar trímeros los cuales van a producir espículas en la envoltura del virus. Gp120 presenta motivos de unión a los receptores celulares de CD4 y a los receptores de quimiocinas CXCR4 y CCR5.

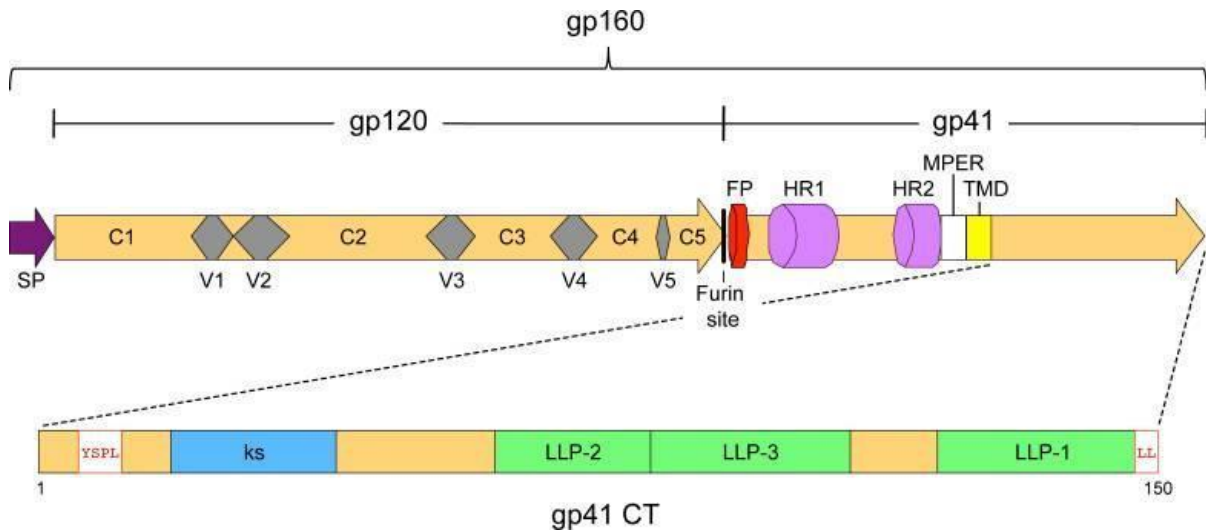


Figura 5. Representación lineal de la glicoproteína de Env. Se muestra el sitio de corte de gp160 dando a gp120 y gp41, SP indica péptido señal. En gp120 los dominios variables están indicados de V1 a V5 y los conservados de C1 a C5. En el ectodominio de gp41 se indica FP: péptido de fusión N-terminal, HR1 y HR2: dos regiones heptadas. La región de transmembrana, TMD esta indicada como el rectángulo a amarillo y a partir del aminoácido 150 se inicia la cola citoplasmática o CT. Los motivos implicados en el transporte e internalización son YSPL: Tyr-Ser-Pro-Leu (tirosina+Serina-Prolina-Leucina). Motivos helicoidales putativos o péptidos lentivirales líticos: LLP-1. LPP-2 y LPP-3[5].

CILCO DE REPLICACIÓN

Reconocimiento y adhesión.

Este virus va a infectar aquellas células que posean receptores CD4, identificado como el receptor principal, este reconocimiento da inicio a la replicación. El receptor CD4 se encuentra presente en linfocitos T cooperadores, macrófagos CD4⁺ y en algunas células dendríticas. El procesamiento proteolítico de gp160 es absolutamente necesario para la actividad fusogénica de Env y, por tanto, es esencial para la infectividad viral. Después de la escisión, gp120 y gp41 permanecen asociadas por interacciones no covalentes, conformando una espícula homotrimérica en la superficie del virión.

En el proceso de entrada del VIH a la célula es necesaria la unión del loop V3 de gp120 a un correceptor celular ya sea CXCR4, CCR5 o ambos. Se sabe que en su mayoría se usa al correceptor CCR5, pero conforme la infección progresa, los viriones generados pueden usar el correceptor CXCR4, o ambos, esta habilidad de usar uno u otro correceptor, está mediada por la variación genética en el loop 3 o región variable (V3) de gp120[21], el cual modifica el tropismo del virus.

Entrada

La unión de gp120 con el receptor y correceptor produce cambios conformacionales en la partícula viral exteriorizando el dominio de fusión de gp41, lo que permite la unión de la membrana viral y celular dando la entrada del virus a la célula hospedero [20]. El VIH-1 ingresa a las células al fusionar su membrana con la membrana plasmática (PM) de las células diana. La vía de internalización a nivel de superficie celular es la vía principal para la entrada del VIH en las células que expresan CD4 permisivas. Las GTPasas Rho: RhoA, Rac1 y Cdc42 participan en este proceso, al promover la estabilización del complejo de fusión y la formación y expansión de poros de fusión en la superficie de la célula hospedera a través de la remodelación del citoesqueleto. El poro de entrada se estabiliza mediante la actividad de GTPasa de la dinamina, lo que facilita la liberación de la cápside que contiene el genoma del virus en el citosol. También se ha descrito una ruta de entrada alternativa que no ha sido del todo caracterizada, alternativa, en la que las partículas de VIH-1 entran por endocitosis mediada por clatrina (CCV). En esta vía también está involucrada la actividad de GTPasa de la dinamina tanto en la formación de las vesículas, como en la fusión entre el virus y la membrana del endosoma. Esta vía también implica que Rab5 y Rab7 actúen en la maduración endosomal. Una vez en el citoplasma, inicia el proceso de retrotranscripción en el que el ARN viral se transcribe a ADN de doble cadena a través de la RT. Consecutivamente, se formará el complejo de preintegración o PIC el que va a transportar proteínas virales (IN, NC, Vpr y el ADN viral). El PIC es llevado al poro nuclear por CypA, TNPO3 y el transportador nuclear GTPasa Ran-GDP y una vez en el nucleoplasma, el provirus del VIH-1 se integra en la cromatina del hospedero [30], mediante la IN que realiza un corte en el genoma del hospedero en una región transcripcionalmente activa. Al ADN viral integrado en el genoma de la célula hospedero conocido como provirus el cuál adquiere una función de gen celular en el lapso de vida de la célula infectada. [16].

Posterior a la inserción se inicia con la retrotranscripción de los genes de virus, en este punto es importante la maquinaria celular como proteínas regulatorias y factores de transcripción celular. Los primeros ARNm virales llevan a cabo el splicing, dando lugar a la traducción de las proteínas como Tat, Rev, Nef, Vif, Vpr y Vpu. Conforme incrementa la acumulación de Rev en el núcleo, se ve beneficiada la salida de los RNAm virales, lo cual provoca que estos eviten el splicing. Lo que genera que los ARNm producidos al final del ciclo se caractericen, por solo pasar uno o ningún evento de splicing, lo que induce la formación de Gag, Pol, Env y ARNs genómicos que se empaquetan en nuevas partículas virales [21].

Ensamble y salida de los nuevos virones

Las proteínas Env, Pol, Gag y genoma recién formado de ARN es transportado hacia la membrana de la célula para poder ser ensamblados. El precursor de la glicoproteína Env sin procesar (gp160) contiene una secuencia señal del retículo endoplásmico (ER) en su extremo N-terminal, que dirige a Env a la membrana RER. Este péptido señal es eliminado co-traduccionalmente por peptidasas celulares dentro del RE. El dominio transmembranal (TMD) de gp41 contiene una señal de paro de transferencia hidrófoba (anclaje de membrana) que evita que gp160 se libere por completo en el lumen del RE. Por tanto, las porciones extracelulares de Env se extienden hacia la luz del RE, mientras que la cola citoplasmática (CT) de gp41 se extiende hacia el citoplasma y más tarde hacia el interior del virión [20].

Al salir de la célula los viriones recién formados toman parte de la membrana lipídica de la célula en donde se incrustan gp120 y gp41. En el proceso de maduración, la proteasa va a cortar cada una de las poliproteínas, lo que permite formar proteínas virales funcionales, las cuales se van a ensamblar para crear la matriz, cápside y nucleocápside del virus. Así el VIH que ha madurado es capaz de infectar a otras células para continuar con un nuevo el ciclo de replicación (Figura 6) [14].

Proteínas reguladoras

Tat es el transactivador transcripcional de los mensajeros virales ya que favorece la elongación de la transcripción de estos mensajeros al fosforilar la región CTD de ARN Pol II. Rev va a inducir la transición de la expresión de genes de la etapa temprana a la tardía al modular el proceso de splicing [22].

Proteínas accesorias

Nef, una proteína reguladora de aproximadamente 27 kDa, la cual provoca la disminución de la expresión de CD4 en la superficie de las células infectadas favoreciendo su endocitosis. La proteína viral r ó Vpr actúa como factor para el transporte del ADN viral hacia el núcleo de la célula, también arresta a la célula en la fase G2 del ciclo celular, deteniendo la apoptosis. Vpu es una proteína de 16 kDa, única del VIH-I, que disminuye la expresión de CD4 y ayuda a liberar a los viriones de la célula infectada, al contrarrestar la acción del factor de restricción viral teterina promoviendo su degradación. Vif de 23 kDa, es necesaria en la replicación y en el correcto empaquetamiento de las proteínas del virus, además, promueve la degradación de los factores de restricción APOBEC que producen hipermutación [22,23].

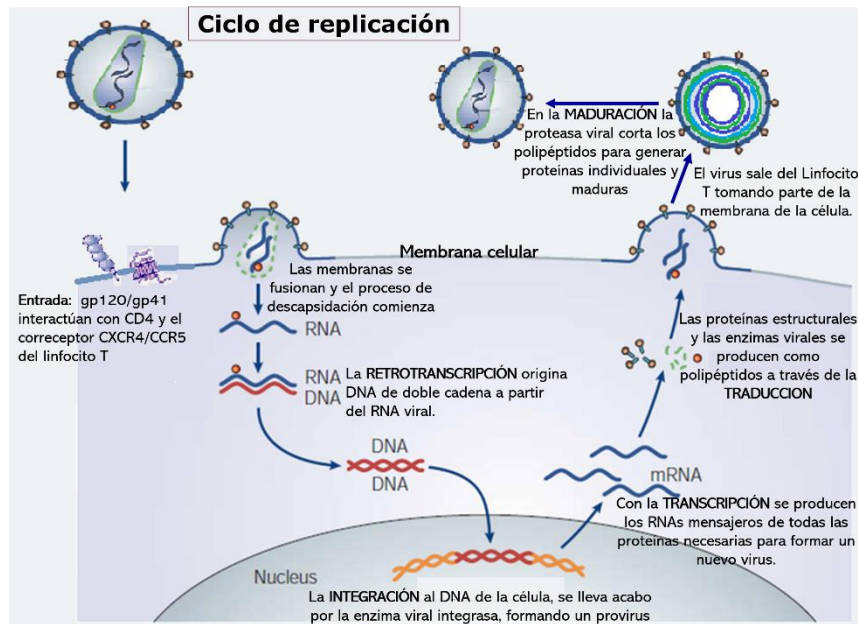


Figura 6. Descripción general del ciclo de replicación del VIH. Se muestra la interacción de gp120 y CD4, así como al de V3 y CCR5 o CXCR4. Seguido de la fusión de membranas gracias a gp41, la formación de PIC y su transporte al interior del núcleo para poder integrar el genoma viral. Finalizando con el transporte y ensamblaje de las proteínas virales, la liberación de los viriones y maduración. Tomado y modificado [10].

HISTORIA NATURAL DE LA ENFERMEDAD

Clínicamente la infección que produce el VIH se va a dividir en etapas que se van a diferenciar de acuerdo a la respuesta inmune que produce el hospedero.

- Etapa aguda o infección primaria: caracterizada por elevadas cargas virales (CV), que se asocian a una replicación exponencial del virus, seguida de una baja temporal en sangre de linfocitos T CD4⁺.
- Etapa de latencia clínica o crónica: es aquella en la que se va a estabilizar en sangre la CV y el número de linfocitos T CD4⁺, esto a concentraciones aproximadas de 200 a 500 células/μl.
- SIDA ó etapa tardía: en este punto las personas infectadas muestran concentraciones de 200 células T CD4⁺/μl o menores, dejando a un lado el poco control de la replicación viral lo que lleva a que la carga viral aumente en sangre provocando que se presenten enfermedades oportunistas. [24, 20].

En las primeras seis semanas después de la infección, se desarrolla un síndrome agudo, con valores muy elevados de carga viral en sangre periférica, resultando en signos y síntomas parecidos a una gripe común (Figura 7). La capacidad del sistema inmune para crear una respuesta específica o CTL's durante el inicio de la infección es muy importante y relevante como pronóstico y avance de la infección [25].

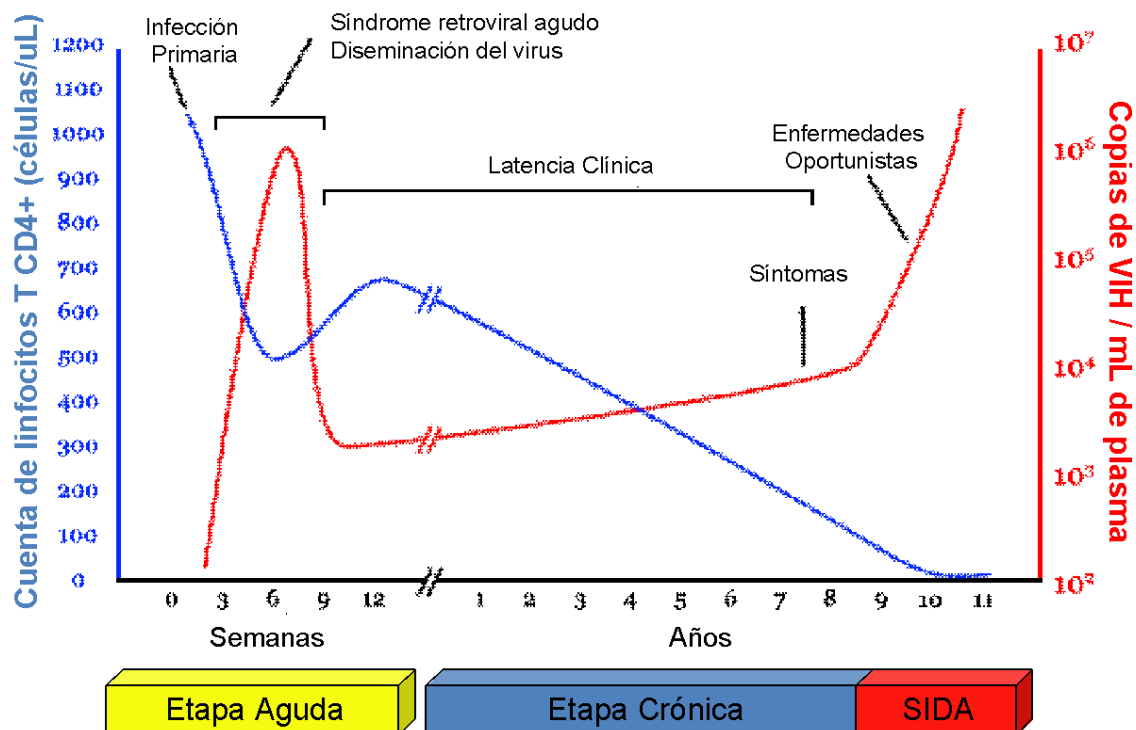


Figura 7. Historia natural de la infección ocasionada por VIH. Se muestra la variación a lo largo del tiempo de los linfocitos T CD4⁺ y del número de copias de ARN viral. Se marcan las etapas en cajas de colores bajo el gráfico. Tomado y modificado [108].

RESPUESTA DE LINFOCITOS T CITOTOXICOS EN LA INFECCIÓN POR VIH

El sistema inmune enfrenta a todos los agentes patógenos, durante su evolución ha desarrollado múltiples estrategias para luchar contra la invasión de agentes infecciosos. Una característica particular de la inmunidad adaptativa es que es capaz de identificar y deletar moléculas y organismos extraños de forma selectiva.

Una de las herramientas en este proceso son las células T que presentan en su membrana una molécula única de unión a antígeno llamada receptor de célula T o TCR (por sus siglas en inglés).

Mayoritariamente los receptores de células T solo identifican antígenos unidos a proteínas de membrana, MHC (por sus siglas en inglés). Las moléculas del MHC, también llamadas HLA (por sus siglas en inglés), son glicoproteínas polimórficas, y se encuentran en las membranas celulares. Existen dos tipos: moléculas MHC clase I, que expresan casi todas las células nucleadas, y las moléculas MHC clase II, que sólo se localizan en células presentadoras de antígeno o APC (por sus siglas en inglés). Las cuales van a ligar péptidos antigénicos que serán presentados a distintas subpoblaciones celulares del sistema inmune.

Los CTL cumplen una función indispensable en el control y vigilancia de las células, por ejemplo, en infecciones virales [27], y son estimulados al presentar antígenos a su TCR. Las células que son infectadas van a procesar las proteínas virales en forma de péptidos que se van a unir a moléculas HLA clase I.

Los CTLs regulan la apoptosis por medio de dos vías, una granulo-dependiente que libera moléculas citotóxicas como granzimas y perforinas que causa la eliminación de la célula que fue infectada y dos, esta la vía independiente de estas moléculas en donde se usa el ligando Fas o molécula FasL el cual conecta con el receptor Fas (CD95) en la célula infectada provocando la muerte celular [28].

La importancia de los CTLs es gracias a estudios previos realizados en simios que estaban infectados con SIV. Al eliminar, *in vivo*, de los macacos *Rhesus* los CTLs, se identificó que había una pérdida en el control de la viremia, mostrando que, si el proceso tardaba más de 28 días provoca que la viremia primaria no se controle nunca induciendo la muerte de los macacos por adquirir un síndrome parecido al SIDA. En estos organismos la disminución de CTLs suele ser transitoria ya que viven con la infección de forma crónica sin morir por ello de forma silvestre, lo que indica que la carga viral vuelve a niveles normales para los macacos una vez que la población de linfocitos TCD8⁺ se recupera. [25].

ANTECEDENTES

Una de las características más destacables del VIH-1 es la amplitud genética y la plasticidad de sus poblaciones [31]. La recombinación y las mutaciones puntuales han contribuido con esta variación. La recombinación en el VIH se ha documentado pero su importancia en la alta diversidad de estas poblaciones virales se comprende poco [32].

Un virus recombinante es una secuencia genética que lleva regiones de dos cepas parentales genéticamente distintas. Se ha demostrado que 90 recombinantes entre subtipos son recurrentes entre los virus VIH-1 circulantes [33].

La recombinación en el VIH-1 está bien documentada, pero se no se comprende en su totalidad la importancia en la diversidad generada dentro del hospedero como lo es la recombinación intrasubtipo [33].

Se sabe que los genomas recombinantes reemplazan rápidamente a los transmitidos/ linajes fundadores (T/F), en un plazo de 27 días, aumentando la complejidad genética de la población viral [33].

De acuerdo con estudios de Song y colaboradores la recombinación ayuda a propagar mutaciones que confieren resistencia a fármacos y de acuerdo a su investigación, los mecanismos moleculares de recombinación viral contribuyen a la persistencia e inmunopatogénesis del VIH-1 y tiene implicaciones para los estudios de transmisión y evolución del VIH *in vivo*.

Las mutaciones puntuales se acumulan de manera bastante constante a lo largo de generaciones y, por lo tanto, se pueden usar para registrar la divergencia de cepas virales [34]. Aunque la selección dicta que la variación no es constante en los genomas del VIH-1, la densidad de mutaciones acumuladas se puede utilizar para determinar si un aislado viral determinado ha pasado por rondas extensas de replicación o si se ha reactivado recientemente a partir de un provirus establecido hace mucho tiempo.

En contraste, la recombinación puede restablecer el reloj molecular modificando el contenido genético. Esto puede dar lugar a combinaciones beneficiosas de mutaciones, la pérdida de mutaciones perjudiciales o nuevos puntos de partida para la evolución viral. Siempre que se observan sustituciones agrupadas, la variación surgió, más probablemente a través de la recombinación que a través de mutaciones puntuales en serie [35]. Algunos casos de cambio fenotípico, incluido el cambio de correceptor y la readquisición de

resistencia a los medicamentos, se han relacionado con mutaciones agrupadas (embedded) en secuencias localizadas que difieren significativamente de las secuencias flanqueantes (flanking), lo que proporciona evidencia de recombinación intrasubtipo dentro de las poblaciones de virus de pacientes individuales [37].

Cuando en 1997 Kuwata y cols coinfectaron experimentalmente células en cultivo con dos cepas distintas con aptitudes (fitness) similares, se descubrió que más del 20 % de la población proviral era recombinante, lo que sugiere que la recombinación era excepcionalmente frecuente [38]. Además, la recombinación del virus de la inmunodeficiencia de simio (SIV) se detectó fácilmente en monos coinfectados experimentalmente, lo que demuestra que la recombinación de lentivirus similares al VIH también ocurre *in vivo* [39].

La recombinación del VIH-1 no implica la ruptura y la reincorporación de un fragmento genómico, sino que resulta del cambio de la plantilla de la transcriptasa inversa (RT) entre los ARN virales durante la síntesis del provirus.

De las propiedades fundamentales en los retrovirus que son críticas para su alta frecuencia de recombinación, hay dos. La primera es que los ARN genómicos retrovirales (ARNg) se encapsulan en pares. Tras la infección de una nueva célula, la proximidad de los dos ARNg facilita el cambio de plantilla que es mucho más frecuente que en otros virus. A pesar de albergar dos ARNg completos por partícula, los retrovirus no son verdaderamente diploides y se describen mejor como "pseudodiploides" (Fig. 8). Esto se debe a que solo se sintetiza uno o menos moléculas de ADN por virión y, por lo tanto, solo se transmite un alelo en cada locus en el ADN de la progenie.

Parte de la razón por la que no se produce más de un ADN por partícula viral es estocástica: probablemente menos del 1 % de todos los viriones generan provirus infecciosos y, por lo tanto, la probabilidad de generar dos es $<0,01$ %. Además, el cambio de plantilla durante la síntesis de la cadena negativa prácticamente impide la generación de más de un ADN por virión debido a la degradación de los segmentos de la plantilla por parte de la función de ARNasaH de la RT. De la misma manera, la alta frecuencia de cambio de plantilla del VIH-1 limita efectivamente el número de moléculas de ADN generado por cada dímero de ARNg a uno [40].

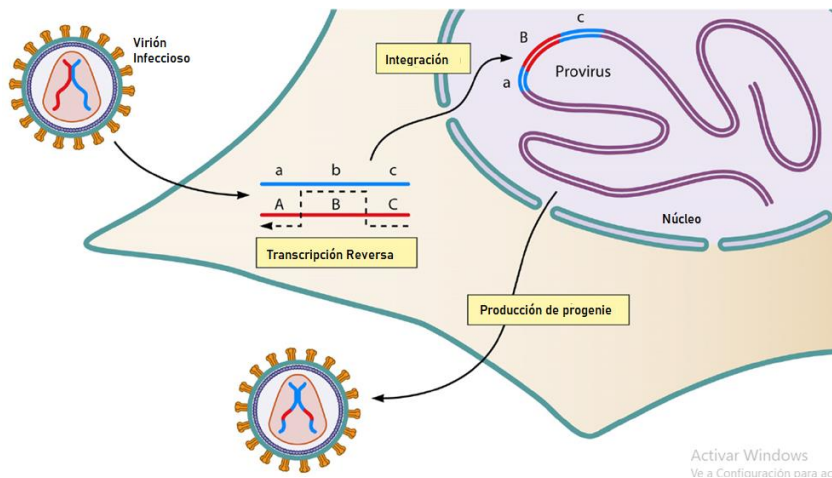


Figura 8: Naturaleza pseudodiploide del VIH Los viriones que empaquetan con dos copias genéticamente distintas de gRNA viral (líneas rojas y azules) tienen el potencial de generar provirus recombinantes si se producen uno o más cambios en las secuencias molde (indicados con una línea negra discontinua) durante la transcripción reversa. Tomado y modificado [42].

La segunda propiedad de los retrovirus que es fundamental para su alta frecuencia de recombinación es su maquinaria de replicación propensa a la recombinación. Se planteó la hipótesis de que los retrovirus son propensos al cambio de molde o plantilla debido a la necesidad de realizar dos cambios de plantilla replicativos mecánicamente similares durante cada ronda de síntesis de ADN viral [41] (Fig. 9).

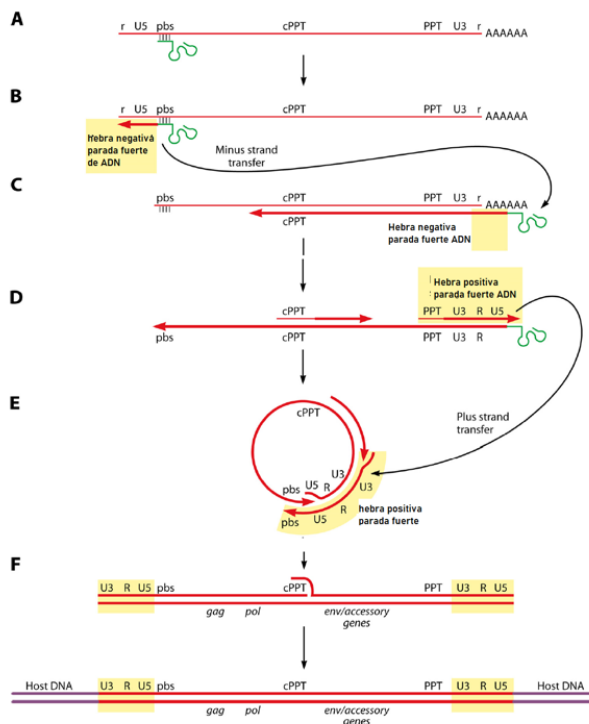


Figura 9: Proceso de transcripción reversa del VIH-1. (A) Un iniciador de ARNt (tRNA_{Lys}) se une a secuencias complementarias en el sitio de unión del iniciador (PBS) en el ARNg.(B) Inicia la síntesis de ADN de cadena negativa. (C) Cambio de templado del complejo replicativo conocido como transferencia de cadena negativa y continuación de la síntesis de ADN de cadena negativa. (D) Síntesis de la cadena positiva a partir de tracto de polipurinas (PPT) (E) Degradación del ARN restante y circularización de la molécula ADN de doble cadena parcial (F) Se completa la síntesis de ADN, generando ADN de doble cadena con repeticiones terminales largas. La integrasa cataliza la integración del provirus y las enzimas de reparación del hospedero eliminan los restantes y los espacios. Las líneas rojas finas representan ARN viral y las líneas más gruesas representan el ADN viral. El iniciador ARNt se representa en verde. Tomado y modificado [42].

Como se mencionó anteriormente, los genomas retrovirales se componen de ARN monocatenarios, denominados ARN de "cadena positiva" porque contienen marcos de lectura abiertos que son reconocibles por los ribosomas del hospedero. En el proceso de transcripción reversa los primeros intermediarios de ADN sintetizados son de cadena negativa o antisentido. La generación de un ADN retroviral no es tan simple como la copia de ARN de sentido positivo en ADN de cadena negativa, seguido de la síntesis de un complemento de cadena positiva.

La RT es una enzima, que se desliza rápidamente entre los extremos de la plantilla hasta que localiza el extremo 3' del iniciador. Su tiempo de permanencia allí está modulado por la firmeza con la que los dominios de los dedos y el pulgar se sujetan en su lugar, característica notable de la capacidad de la RT para cambiar a la orientación de polimerización una vez que alcanza el término del iniciador sin disociarse [54]. Un ARNt de lisina (tRNA_{Lys}) funciona como iniciador al unirse a una secuencia complementaria en el sitio de unión del iniciador (PBS) en los ARNg de VIH-1. La síntesis de ADN de cadena negativa se inicia a partir de este iniciador y se detiene cuando llega al extremo 5' del ARNg. La actividad de RNasaH de la polimerasa viral degrada el templado de ARN y desenmascara al ADN naciente. En este punto se produce el cambio de templado replicativo conocido como transferencia de cadena negativa, ya que la región R del ADN sintetizado es complementaria a la región R del extremo 3' del ARNg, por lo que continúa la síntesis de ADN de cadena negativa. A medida que avanza la síntesis de la cadena negativa, se degrada el ARNg. La síntesis de ADN de cadena positiva inicia a partir de

oligorribonucleótidos que persisten en una región rica en purina PPT y son más resistentes a la degradación.. Finalmente se completa la síntesis de ADN, generando ADN de doble cadena con repeticiones terminales largas que se han producido durante este proceso (Figura 9)

Durante el procesos de transcripción reversa ocurren dos cambios de plantilla replicativos obligatorios, también conocidos como transferencias de cadena de parada fuerte o "saltos" (el extremo 5' del ARN, se hibrida con el extremo 3' del ARN) y se duplican secuencias que se encuentran solo una vez en el ARNg para reconstituir las repeticiones terminales largas en los límites del ADN preintegrativo [43]. Además otros cambios de plantilla recombinogénico puede ocurrir desde cualquier posición en el genoma retroviral [A12]. En su forma más simple, la recombinación retroviral implica la copia de parte de un ARNg, seguido por el salto de la RT a una región homóloga en la otra molécula de ARNg para completar la síntesis de ADN viral [45]. Esto puede conducir a genomas recombinantes si los ARN empaquetados conjuntamente contienen diferencias alélicas.

Para el año 2009 se había identificado en estudios experimentales que el cambio de plantilla recombinogénico del VIH-1 se produce a una tasa de aproximadamente cada 2 kb o posiblemente incluso con mayor frecuencia en algunas condiciones fisiológicas [46]. Como lo reportado por Letvin y colaboradores en el 2006 que indican que en cada replicación viral ocurre de 7 a 30 eventos de recombinación por genoma y ciclo de replicación [47]. Mientras que algunas estimaciones de las tasas de recombinación *in vivo* varían entre los diferentes estudios (1.4×10^{-5} a 2×10^{-4} puntos de corte por sitio por generación) comparables a la tasa de mutación puntual de $2.2\text{-}5.4 \times 10^{-5}$ por base por generación [33]. Se ha observado que después de un cambio en la terapia con medicamentos antirretrovirales se observa un aumento en la recombinación, lo que sugiere que esta desempeña un papel importante en la evolución de la resistencia a los medicamentos. La aceleración de la evolución debida a la recombinación depende de los parámetros que rigen la dinámica de la población, muchos de los cuales no se conocen con suficiente precisión [48].

Los primeros estudios de recombinación del VIH-1 involucraron virus infecciosos en cultivos de tejidos o, en el caso de SIV, animales infectados [38]. Eventualmente para los análisis se inició con el estudio de herramientas bioinformáticas que pudieran predecir la

tasa de recombinación como lo fue RDP3 en un estudio sobre la recombinación intrasubtipo del subtipo C en Tanzania. Identificando que el número de sujetos con evidencia de recombinación aumentó de 69% a 82% durante un año de seguimiento. Se reportó que el 84% de sujetos mostraron evidencia de recombinación intrasubtipo en ENV, lo que sugiere que este evento es más común de lo que se pensaba anteriormente 5.6 eventos de recombinación por genoma [49]. En contraste a Zhuang y cols en el 2002 que estimaron que la recombinación ocurre aproximadamente 2.8 veces por genoma por ciclo de replicación viral (coincidiendo con el promedio del 2009)[90].

Experimentalmente, algunos recombinantes mostraron al menos 10 cruces, cuando el promedio fue inferior a 3 [50]. Quizás la mejor estimación actual de las tasas de recombinación intersubtipo del VIH-1 es de cuatro a cinco cruces por genoma, una tasa realmente sorprendente que sugiere que la recombinación es una parte normal de la generación de cada ADN viral. Una limitación obvia de estos ensayos es que se utilizan para describir y reportar la recombinación del VIH-1 es la capacidad de las herramientas bionfórmáticas.

La adición de la proteína chaperona de ácido nucleico viral, NC (nucleocápside), promueve el cambio de plantilla. Las frecuencias de recombinación informadas *in vitro* son más bajas que las observadas durante la replicación viral [51]. Es probable que esto se deba en parte a las propiedades físicas del complejo de transcripción inversa que resulta de la eliminación del revestimiento del virus. Algunos factores de restricción del hospedero afectan los primeros pasos de la replicación, como la retrotranscripción, pueden modular la recombinación. De hecho, se ha relacionado con APOBEC3G (A3G) [52] cierto deterioro de la elongación de RT y el cambio de plantilla replicativa. Sin embargo, aunque no se pueden descartar las contribuciones de los factores del hospedero y puede ser necesaria la activación del complejo de transcripción inversa en el hospedero, no se sabe que ningún otro factor, aparte de RT y NC, contribuya a la recombinación del VIH-1 [53].

Modelos de Recombinación.

Numerosas publicaciones sugieren que la recombinación genética del VIH-1 ocurre por un mecanismo de elección de copia, es decir, el uso alternativo de dos plantillas durante la síntesis de un solo ADN viral [44]. En concepto, la recombinación retroviral podría resultar

del cambio de plantilla durante la síntesis de ADN de cadena negativa o positiva, y se han propuesto modelos distintos para cada uno [45].

Durante el descubrimiento del modelo de "asimilación por desplazamiento de cadena" o recombinación durante la síntesis de ADN de cadena positiva [80] se basó en micrografías electrónicas de genomas de retrovirus aviares con transcripción inversa parcial visualizados después de la interrupción de la actividad del virión [55] ya que los viriones contienen una actividad enzimática que podría promover el desplazamiento de cadena implícito en tales modelos [55]. Las imágenes revelaron ADN unidos en estructuras ramificadas en "H" que se forman cuando colas libres monocatenarias (formadas por la hebra del extremo 5' que es desplazada por la síntesis continua) invaden el ADN que se sintetiza en el segundo genoma en el virión Fig. 10-A[109].

El modelo de "elección de copia forzada" explica cómo el cambio de plantilla podría generar un ADN proviral intacto a partir de ARNg fragmentados y se basó en observaciones de que los ARNg aislados de retrovirus animales a menudo aparecían cortados y, sin embargo, generaban eficientemente ADN viral infeccioso [41] (Fig. 10-B). En este modelo, los ADN nacientes se disocian de las plantillas de gRNA del donante cuando se alcanzan los extremos cortados de la plantilla, lo que permite que las hebras del cebador se asocien con regiones de homología del aceptor en el gRNA empaquetado. Una variación del modelo, basada en la alta frecuencia de eliminación repetida sugiere que no se requiere la ruptura del ARN [56].

Un segundo modelo de recombinación de cadena negativa, llamado "intercambio de cadena negativa" o "invasión de aceptor", como el modelo preferido para la mayoría de las recombinaciones [57] (Fig. 3C).

La degradación de la plantilla mediada por RNasa H expone el ADN naciente de cadena negativa, dando el apareamiento entre pares de bases del ADN sintetizado en un gARN y las regiones complementarias en el segundo gARN. Seguido de la realineación de la cadena del cebador por migración de rama, con la reanudación de la síntesis de ADN en la plantilla aceptora que produce la unión de recombinación [42].

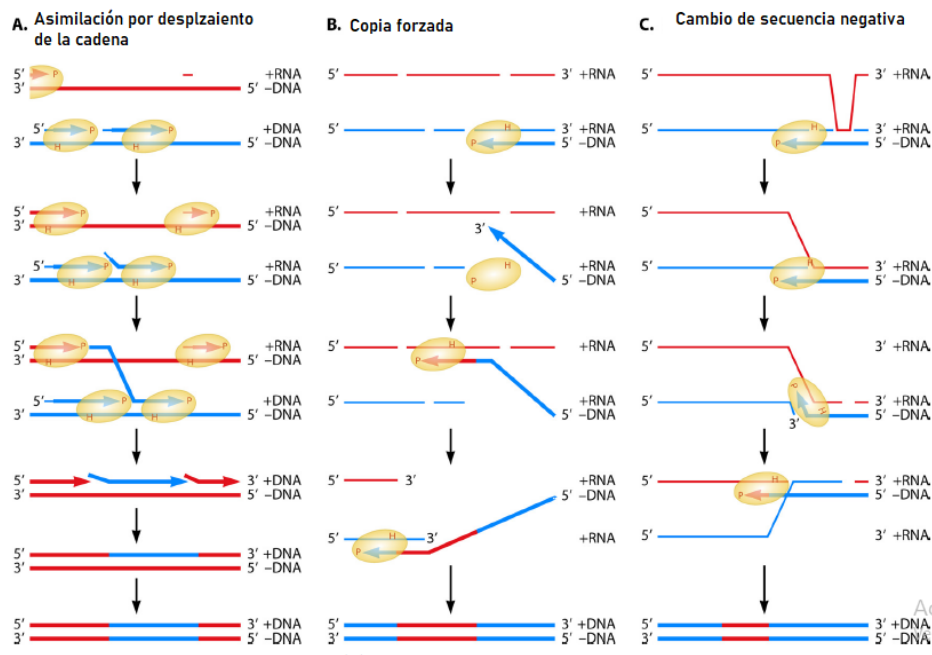


Figura 10: Tres modelos de recombinación retroviral. (A) Asimilación de desplazamiento de cadena que muestra recombinación durante la síntesis de cadena positiva. (B) Elección de copia forzada. (C) Cambio de secuencia negativa. Las líneas finas son ARN retrovirales con roturas en las hebras. Las líneas gruesas representan el ADN naciente; las flechas indican la dirección de la síntesis de ADN. Los óvalos dorados representan la RT del VIH-1. Tomado [42]

Los exámenes de las tasas de recombinación y las frecuencias de complementación de los RT mutantes de RNase H formaron la base del modelo de "elección de copia forzada dinámica". Este modelo sugiere que un equilibrio entre las actividades de la RNasa H dependientes e independientes de la polimerasa determina las tasas de recombinación [58]. Este modelo sugiere que la actividad de la RNasa H es necesaria para la disociación de la plantilla-cebador antes del cambio de plantilla y contrasta con el intercambio de cadena negativa, que prevé un complejo de precambio que contiene tanto al donante como al aceptor, en lugar de la liberación de la plantilla del donante antes que reconocimiento del aceptor [59]. Indicando que la recombinación no es necesariamente causada por daño en el ARN, sino que depende del proceso dinámico de síntesis de ADN. A medida que RT sintetiza el ADN, la ARNasa H degrada la plantilla de ARN, dejando el ADN naciente libre para aparearse con el ARN empaquetado conjuntamente. Si la síntesis de ADN se vuelve lenta o la ARNasa H degrada el molde de ARN más rápidamente, el complejo RT se vuelve menos estable, se disocia del molde y puede cambiar al uso del ARN empaquetado como

molde. Por lo tanto, la recombinación depende de la dinámica de la síntesis de ADN, no es necesariamente un evento obligatorio y puede ocurrir incluso en ausencia de daño de ARN. Además, este modelo se basa en el hallazgo de que la recombinación es promovida por la disminución de la polimerasa o el aumento de la actividad de la ARNasa H. Se estima que la RT del VIH-1 cambia de molde de 8 a 10 veces durante la síntesis de un solo genoma. La recombinación ha dado lugar a la aparición de más de 96 formas recombinantes circulantes (CRF) de diferentes subtipos del grupo M del VIH-1, y estos CRF causan una proporción sustancial de infecciones en todo el mundo. Además, se ha demostrado que la recombinación produce variantes que confieren resistencia a múltiples fármacos o evaden la respuesta inmune del huésped. Por tanto, la recombinación frecuente actúa como una barrera para el desarrollo de vacunas y terapias antivirales eficaces. En conjunto, estos hallazgos demuestran que la recombinación juega un papel importante en la generación de diversidad viral y en la configuración de la actual pandemia mundial del VIH-1 [60]. La tesis central del modelo dinámico de elección de copia, que propone que la recombinación depende de la estabilidad cinética del complejo RT y que los cruces pueden ser forzados o no forzados. La replicación eficaz del VIH-1 requiere tanto los ARN empaquetados como la recombinación para mantener la integridad del genoma. Estos hallazgos revelan la posibilidad de que se puedan desarrollar nuevas estrategias antivirales para controlar la replicación del VIH-1 dirigiéndose al proceso de recombinación esencial o al empaquetamiento de dos genomas de ARN [61].

Michael Worobey y Edward C. Holmes sugieren que la recombinación puede ser más común de lo que se supone a menudo, pero puede pasar desapercibida debido a la acción de una fuerte selección purificadora que eliminará nuevas combinaciones deletéreas de mutaciones [62]. A la luz de estos estudios, está claro que lo que se entiende por "frecuencia de recombinación" (un término que generalmente se usa sin unidades específicas) depende fundamentalmente de si estamos evaluando los eventos de recombinación antes o después de que la selección haya actuado.

RECOMBINACIÓN Y HOTSPOTS

Una debilidad de los ensayos que usan frecuencias cruzadas en un intervalo definido para deducir las tasas de recombinación de todo el genoma es que asumen que la recombinación ocurre en frecuencias uniformes en lugar de principalmente en sitios

específicos. Aunque la recombinación se observa en todos los genomas del VIH-1 y es más o menos proporcional a la longitud, las frecuencias no son uniformes y existen puntos "calientes" (o hotspots) y/o "fríos" (cold spots) para la recombinación [63].

La agrupación de puntos de ruptura de recombinación en cepas clínicas se observa con mayor frecuencia en los bordes de los genes virales [65]. Un estudio en el 2007 que examinó los puntos de corte en los recombinantes intersubtipo clínicos encontró un uso frecuente de intervalos genéticos que flanquean el gen *env* [66]. Estos no parecían estar dictados por la homología, la estructura del ARN o las regiones implicadas por estudios experimentales o hacia la adquisición de sitios bajo selección positiva. En cambio, apareció alguna selección para reemplazar toda la región de codificación *env*. Cruces similares en VIH propagados *in vitro* sugirieron que estos son hotspots de cruce reales. [67].

Solo una pequeña fracción de todos los recombinantes generados sobrevive dentro de las poblaciones de pacientes [68]. De manera similar, se ha observado una baja infectividad para muchos genes de envoltura quiméricos generados *in vitro* [69]. Por lo tanto, los cruces más comunes observados *in vivo* son más probablemente el resultado de la selección en lugar de hotspots. Presiones similares, superpuestas sobre diferentes presiones selectivas *in vivo*, se manifiestan en formas como la selección de cambios compensatorios agrupados que acompañan a las mutaciones requeridas para el escape inmunitario de respuestas específicas de células T citotóxicas [70].

La recombinación contribuye a la genética del VIH-1 en varios niveles: desde los orígenes del virus hasta su adaptabilidad en pacientes individuales, conocida como recombinación intrasubtipo [71]. Dentro de pacientes individuales, la recombinación (intrasubtipo) es una fuente extensa y continua de diversidad viral [72]. Uno de los *hot spots* identificados en el gen *env*, se localiza en una estructura secundaria de ARN que particularmente promueve la recombinación en esta región, y la eliminación de dicha estructura reduciría la tasa de recombinación. En los últimos años se ha propuesto que una región cercana al tracto central de poli-purina (cPPT) puede ser un *hot spot* de recombinación, por lo cual se hipotetiza que es una región rica en guaninas (Gs) que forma cuartetos G intermoleculares, facilitando la dimerización y la recombinación del ARN [73.]

Otra característica del VIH es que fija sucesivamente mutaciones que le permiten evadir las respuestas inmunitarias, especialmente en el gen *env*. Los estudios de la evolución del VIH intra-hospedero indican que el alcance y la estructura de la diversidad genética viral

es más un marcador de la carrera (teoría de la reina roja) entre el hospedero y el virus. Las filogenias del VIH intra-hospedero tienen una fuerte estructura temporal, lo que refleja la fijación sucesiva de mutaciones ventajosas y la extinción de linajes desfavorables. Tradicionalmente se ha minimizado el papel de la recombinación intrasubtipo. Sin embargo, estudios genéticos de poblaciones recientes indican que la recombinación también es una fuerza dominante dentro de los subtipos y un factor muy importante en la evolución del VIH-1 y en la adecuación de *env*. Sin mencionar que la recombinación frecuente y la selección natural elevan aún más su tasa de cambio evolutivo [62]

Para el 2018 Rawson y cols., muestran que el bloqueo de la recombinación en una región de 2.5 kb del genoma del VIH-1 resultó en la pérdida de la mitad de los provirus con genomas intactos. Estos hallazgos indican que se requiere la recombinación del VIH-1 para una retro transcripción eficaz y el mantenimiento de la estructura del genoma intacto, dando pauta a profundizar en este mecanismo [60].

Rousssue y cols. desde el 2007 indicaban que la recombinación entre dos cepas virales muy similares es difícil de detectar, por lo cual muchas secuencias de VIH-1 nunca se identificarán como recombinantes [74]. Esto ha resultado en que se haya estudiado con más profundidad la recombinación intersubtipo que intrasubtipo, sin embargo, dado que la recombinación ocurre en cada ciclo de transcripción reversa debería ser igualmente común y depender de la frecuencia de superinfecciones (personas que se infectan después de una primoinfección) o coinfecciones en una población. Taylora y cols., han propuesto que la tasa de superinfección intrasubtipo podría ser de hasta del 15% de la tasa de infección en esa población a partir de un modelo coalescente con recombinación [75]. En diferentes poblaciones se ha calculado que las tasas de incidencia de superinfección oscilan entre 0% to 7.7% por año [76] y esta fluctuación depende de la población de estudio, la frecuencia de uso de anti-retrovirales, el tiempo de seguimiento de los pacientes y los métodos usados para la detección, por ejemplo, un estudio reciente en Uganda que utiliza secuenciación de nueva generación reporto que las tasas de superinfección son tan altas como las de primera infección [77].

En un principio se creyó que la recombinación ocurría a azar, pero Smyth y cols., en el 2013 mostraron que no ocurre al azar y que depende de un conjunto complejo de factores, como la disponibilidad de sustratos de nucleótidos (nt), la secuencia de ARN en sí, la

similitud de secuencia general y el contexto de secuencia local de las secuencias de recombinación. Utilizando tanto ensayos *in vitro* como vectores de VIH-1 de ciclo único, identificaron hotspots de recombinación en las regiones no traducidas (UTR), en *gag* y en *env* [78].

En ese mismo año (2013) Kiwelu y cols., indicaron que si bien la recombinación intrasubtipo ocurre en el VIH-1, generalmente se subestima. Y para analizar la presencia de este tipo de recombinación utilizaron siete métodos implementados dentro del programa de detección de recombinación, RDP3 y encontraron que en 38 (84%) sujetos se mostró evidencia de recombinación intrasubtipo, incluidos 22 con el subtipo A1 del VIH-1, 13 con el subtipo C del VIH-1 y 3 con el subtipo D del VIH-1 esto en Tanzania, donde co-circulan múltiples subtipos de VIH-1. La recombinación intrasubtipo del VIH-1 aumenta la diversidad viral, lo cual podría ser un desafío en el diseño de la vacuna [49].

Usando el programa RARP Song y col., en el 2018, observaron que, al reensamblar genotipos que han aparecido a través de eventos independientes la recombinación se puede generar una gama más amplia de nuevos fenotipos que tomarían mucho más tiempo o serían menos propensos a aparecer solo a través de la mutación, además del hecho de que Env está bajo un grado particularmente alto de presión selectiva *in vivo* debido a la selección mediada por anticuerpos para el escape inmune, lo que podría explicar el enriquecimiento de la recombinación en regiones variables [33].

En este punto debemos tomar en cuenta las comparaciones con otras herramientas ya existentes como lo es:

- RIP que fue la primera herramienta bioinformática en detectar automáticamente recombinantes entre subtipos (no lo hace para intra subtipos).
- RECCO encuentra recombinantes nominalmente significativos comparando el costo de obtener cada secuencia por mutación o por recombinación. Sin embargo, no enumera las cepas parentales ni enumera distintos eventos de recombinación.
- RDP4 además de implementar el enfoque del programa de detección de recombinación (RDP), incluye varias otras herramientas existentes, las cuales varían en métodos y estrategias. RDP se basa en métodos filogenéticos y, como demostraron Posada y Crandall, en escenarios de baja divergencia como las

infecciones recientes presentadas en este estudio, estos métodos tienen menos poder para detectar recombinantes.

RARP se basa en el hecho de que los eventos de recombinación pueden identificarse mediante la herencia de bloques de ADN que portan bases distintivas. Por lo tanto, encuentra recombinantes comparando cada conjunto de tres secuencias en la alineación y aplicando la prueba de rachas de Wald-Wolfowitz para verificar la aleatoriedad en la distribución de posiciones variables, identificando regiones en una secuencia que son significativamente más similares a un padre candidato que a otro. Luego clasifica la población viral en grupos de virus transmitidos/founder (T / F) y sus descendientes que presentan mutaciones explicables por base distribuida aleatoriamente.

Debido a su sensibilidad para detectar la recombinación incluso en escenarios de baja diversidad, RAPR también resulta útil cuando se estudia la evolución viral por recombinación en especies virales menos diversas como los virus de la hepatitis B, enterovirus y norovirus.

Corroborando el papel de *env*, en un trabajo realizado en el grupo por Caballero en 2020 se identificó que gp120 predominó en la selección positiva y que en relaciones co-evolutivas, lo que indica que se producen cambios recíprocos en las secuencias de este gen que preservan la interdependencia funcional y estructural [110], y se presentó mayoritariamente fue la de gp120-gp120 para el VIH -1 del subtipo B, adicionando una pauta más para mirar estos sitios en recombinación el gen *env* [79].

Comprender la recombinación en *Env* es una pieza fundamental en el rompecabezas de la historia evolutiva del VIH-1 puede ayudar con el desarrollo de tratamientos futuros o con el diseño de vacunas ya que al estar expuesta Env al ambiente celular es un inmunógeno principal al que se dirigen los anticuerpos neutralizantes y un objetivo al que los inhibidores pueden acceder con relativa facilidad.

HIPÓTESIS.

La frecuencia de recombinación será más alta en sitios que codifican para regiones no estructuradas contribuyendo a la variabilidad de estas regiones.

JUSTIFICACIÓN.

La recombinación es fundamental en la replicación del VIH-1, y debido a los patrones de circulación de este virus la frecuencia de eventos de recombinación intrasubtipo deberá ser mucho mayor que la intersubtipo. Por lo anteriormente descrito, se evaluará, analizará y estudiará el efecto de la recombinación genética intrasubtipo en la evolución del gen *env* en el virus de inmunodeficiencia humana. Este estudio nos permitirá conocer el impacto de la recombinación intrasubtipo en la evolución del VIH. Para alcanzar este objetivo se utilizarán secuencias del subtipo B que codifican para la región *env*, ya que por una parte están ampliamente disponibles en las bases de datos y por otro lado se ha reportado un hotspot de recombinación dentro de este gen lo que aumentará la probabilidad de detectar eventos de recombinación. Los resultados de este estudio contribuyen a entender la recombinación intrasubtipo en los patrones de diversidad observados en el VIH y su papel en las mutaciones de escape y cambios en el uso de correceptor que tiene impacto en el desarrollo clínico de la enfermedad.

OBJETIVO GENERAL

Analizar el efecto de la recombinación genética intrasubtipo en la evolución del gen *env* en el virus de inmunodeficiencia humana-1

Objetivos Particulares.

- 1.- Identificar clústers de transmisión de pacientes infectados con VIH-1 subtipo B
- 2.- Detectar sitios de recombinación y mapear los sitios sobre las proteínas gp120 y gp41
- 3.- Determinar los posibles efectos de recombinación sobre la variabilidad de secuencia, el uso de correceptor y el escape antigénico

ESTRATEGIA EXPERIMENTAL.



METODOLOGÍA

SECUENCIAS

Las secuencias que se utilizaron para este proyecto fueron seleccionadas de la base de datos de los Álamos VIH Database [83]. Los criterios de búsqueda fueron: secuencias codificantes completas (CDS, por sus siglas en inglés) del gen env del VIH tipo 1, subtipo B. Con el objetivo de obtener conjuntos de más de 100 secuencias del mismo año/región geográfica que pudieran analizarse para identificar clústers de transmisión, se hicieron búsquedas dirigidas a las regiones de Norte América y Europa. Además, se utilizó como filtro al año de colecta, seleccionando aquellos que tuvieran conjuntos de más de 100 secuencias. (Ver tabla M1).

Tabla 2. Características de la búsqueda realizadas.

Años seleccionados	2015-2016-2017-2018
VIH	1
Subtipo	B
Región genómica	CDS completo de env
Región geográfica	Norte América y Europa

Las secuencias obtenidas de la base de datos de los Alamos VIH Databases se descargaron por región y por año.

DETECCIÓN DE CLÚSTERS DE TRANSMISIÓN

Con el objetivo de identificar posibles clústers de transmisión para detectar recombinación en conjuntos de datos de la misma región y año de colecta, se utilizó una herramienta a gran escala de epidemiología molecular de VIH-1, que se llama VIH-TRACE (TRANsmission Clúster Engine). Esta herramienta es capaz de identificar grupos de transmisión del VIH-1 en grandes bases de datos de vigilancia que comprenden decenas o cientos de miles de secuencias casi en tiempo real. Esta herramienta permite seleccionar los conjuntos con mayor probabilidad de eventos de recombinación ya que pertenecerán al mismo cluster epidemiológico.

Tabla 3. Parámetros utilizados para la inferencia de Red

Parámetros clave usados en la inferencia de la red que se tomaron en cuenta			
Parámetro	Usado	Significado	Fase
* r - reference	HXB2_env	Secuencia de Referencia	Alineamiento
* t -threshold	0.015	Umbral de distancia para dibujar un enlace de red	Construcción de la Red
*a ambigüedades	avarage	Establece una política para el manejo de nucleótidos ambiguos	Estimación de distancia
*m: minioverlap	500	Las secuencias deben de tener al menos cierto número de caracteres alienados	Estimación de distancia
*g - fracción	0.05	Establece la fracción máxima de nucleótidos resueltos	Estimación de distancia
*u -curate	report	Establece la política para manejar posibles contaminantes	Construcción de la Red
*f -filtro	remove	Elimina conexiones transitivas falsas	Construcción de la Red
*s - strip, drams	lewis	En mascara los sitios asociados con la resistencia a medicamentos contra VIH-1	Estimación de distancia

VIH-TRACE, primero alinea cada una de las secuencias de entrada con una única secuencia de referencia utilizando una extensión del algoritmo de programación dinámica de Smith-Waterman basada en codones [106].

En este trabajo se usó la secuencia HXB2 como referencia por pertenecer al subtipo B.

Los alineamientos pareados se combinan en una única alineación de secuencia múltiple (MSA). Después se calculan todas las distancias genéticas por pares $N(N - 1) / 2$ bajo el modelo de sustitución de nucleótidos Tamura-Nei 93 (TN93) [111]. Este modelo considera tasas de sustitución diferentes entre transiciones y transversiones y diferencia los dos tipos de transición. La red de transmisión se infiere del arcVIHo de distancias pareadas. Los nodos dentro de la red están todos codificados con el nombre de la secuencia completa, con lo cual se establece un vínculo entre dos individuos, si y sólo si las distancias por pares entre cualquiera de las secuencias emparejadas de estos individuos están por debajo de un umbral especificado o threshold (definido en párrafos previos como 0.0015). Al final el análisis arroja arcVIHos tipo JSON que son explorados utilizando una aplicación JavaScript llamada VIHtrace-viz que se crea desde una línea de comandos para la visualización de la información, ó en la página VIH-TRACE (veg.github.io). Este se basa en la biblioteca de visualización de datos de código abierto d3.js.

PRUEBA DE ESTADÍSTICO F

La prueba del estadístico F o índice de fijación (F_{ST}) es una medida de variación de la población de acuerdo a la estructura genética y se define [81]:

$$F_{ST} = \frac{\sigma_S^2}{\sigma_T^2} = \frac{\sigma_S^2}{\bar{p}(1 - \bar{p})}$$

La fórmula propuesta por Wright muestra que F_{ST} mide la variación genética que se puede explicar por medio de la estructura de la población. Además, se puede considerar como una fracción de la diversidad total, que no es a consecuencia de la diversidad promedio en subpoblaciones [82].

$$F_{ST} = \frac{f_0 - \bar{f}}{1 - \bar{f}}$$

Entonces para la fórmula anterior f_0 es la probabilidad de identidad por descendencia de dos individuos pertenecientes a la misma subpoblación, y $\bar{f}$ será la probabilidad de que dos individuos sean idénticos por descendencia en la población. Con esto, F_{ST} se interpreta como medida de variación, de que tan cerca están dos individuos en la misma subpoblación en comparación de la población total. Si la tasa de mutación es baja se interpreta al vincular la probabilidad de identidad por descendencia con los tiempos coalescentes, T_0 y T , que serían el tiempo promedio de coalescencia para individuos de la misma subpoblación y la población total. Dando,

$$F_{ST} \approx 1 - \frac{T_0}{T}$$

La ventaja de la fórmula es que el tiempo esperado de coalescencia se puede estimar a partir de datos genéticos.

DETECCIÓN DE RECOMBINACIÓN

Se utilizó el programa RAPR (Recombination Analysis PRogram) el cual se basa en la prueba de rachas de Wald-Wolfowitz y permite la detección computacionalmente eficiente de recombinantes. al reconstruir una jerarquía de eventos de recombinación en serie a medida que surgen recombinantes. En este punto se utilizaron las secuencias obtenidas de los clusters seleccionados de VIHTRACE las cuales fueron previamente alineados y manejando un valor de FDR (false Discovery rate) de 0.1, la asignación de linajes transmisor fundador fue designada de acuerdo a los TF identificados como

consenso y al tener secuencias correspondientes al mismo año en la herramienta se toma en cuenta que son datos contemporáneos por lo cual no se le asigna valores para secuencias más viejas, y con esto cada triplete identificado en el alineamiento será pareado tres veces con cada una de las secuencias como posibles descendientes.

PREDICCIÓN DE ESTRUCTURA SECUNDARIA

Se realizó una predicción de la estructura secundaria de las secuencias recombinantes y no recombinantes esto con la herramienta Jpred 4 (<https://www.compbio.dundee.ac.uk/jpred/>) que es un servidor de Predicción de Estructura Secundaria de Proteínas. JPred incorpora el algoritmo Jnet para poder hacer predicciones más precisas. Además de la estructura secundaria de la proteína. Se realiza un consenso de las secuencias ingresadas y nos muestra solo un resultado.

ANÁLISIS DE POSICIONES BAJO SELECCIÓN NATURAL

Se identificaron los codones con evidencia de selección en los grupos de secuencias recombinantes y no recombinantes obtenidas del análisis de RARP. El modelo basado en codones en evolución molecular permite inferir sitios seleccionados en un alineamiento de secuencias homólogas con lo cual estima el índice relativo de sustituciones sinónimas por sitio sinónimo (dS) y no sinónimas por sitio no sinónimo (dN). Para este propósito se utilizó el servidor DataMonkey (www.datamonkey.org), que incluye varios métodos como SLAC (Single-Likelihood Ancestor Counting), el cual se basa en la reconstrucción de secuencias ancestrales, sitios por sitio, para obtener mayor congruencia al estimar los valores de cada posición. Fue así como se identificaron posibles codones bajo presión selectiva positiva y negativa [101]. Además, para corroborar la sensibilidad del método también se utilizó la herramienta FUBAR (Fast Unconstrained Bayesian AppRoximation) del mismo servidor anteriormente empleado. Este es un método bayesiano jerárquico aproximado que utiliza una rutina de método Monte Carlo con cadenas de Markov (MCMC) que garantiza la solidez contra la especificación incorrecta del modelo al promediar un gran número de clases de sitios predefinidos. Permite que los sitios que experimentan una selección positiva y purificadora se identifiquen en un orden de magnitud más rápido que

con los métodos existentes. Ya que presenta una mayor potencia y su velocidad permite analizar conjuntos de datos más grandes que otros métodos [100].

ANÁLISIS DE VARIACIÓN GENÉTICA

Se calculó la variabilidad de los aminoácidos en las secuencias recombinantes y no recombinantes, estimando la entropía de Shannon, en Entropy (<http://www.VIH.lanl.gov/content/sequence/ENTROPY/entropy/entropy.html>) la cual es una herramienta de la base de datos de los Alamos. La cual estima la entropía (cambio) de cada una de las posiciones de los aminoácidos en el alineamiento, indicando la variación de la población. La herramienta considera cada columna en un alineamiento de secuencias de forma independiente y usa la entropía de Shannon como una forma de asignar una puntuación a cada columna que refleje la variabilidad en esa columna. Esta aplicación es potencialmente útil para situaciones en las que simplemente intenta evaluar la diversidad de la población en un sentido transversal.

La entropía de Shannon es:

$$H_n(p_1, p_2, \dots, p_n) = - \sum_{i=1}^n p_i \log_2 p_i.$$

Los datos de entrada se aleatorizan para crear tantos conjuntos de datos aleatorios sean indicados, al seleccionar 1000 aleatorizaciones se permite detectar situaciones en las que la entropía en un conjunto de datos fuera mayor que en otro con una $p < 0,001$.

Se obtiene una lista completa de las diferencias de entropía en cada posición comparando el primer conjunto y el segundo conjunto.

Al realizar las comparaciones de las secuencias recombinantes y No recombinantes con la finalidad de observar el grado de variabilidad entre éstos, se indicó un límite sugerido por la herramienta de 5, este límite es para el número de datos aleatorios que da un grado mayor o igual a la diferencia observada en los datos proporcionados que se aleatorizan 1000 veces, teniendo entonces un valor de $p=0.005$.

BÚSQUEDA DE EPÍTOPOS CTL/CD8

Se obtuvo una lista de epítomos CTL previamente descritos en la base de datos de inmunología molecular de VIH (VIH molecular immunology database) dentro de la base

de datos de los Alamos con la herramienta: Search CTL/CD8+ T-Cell Epitope Database (Search CTL/CD8+ T-Cell Epitope Database (lanl.gov)). En el análisis solo se incluyeron aquellos que se hubieran reportado en Estados Unidos o Alemania, y se localizan en la proteína ENV, subtipo B, y tanto como la ubicación de la proteína y el epitopo tenían que coincidir exactamente con la ubicación de consulta en las secuencias analizadas.

También se utilizó la herramienta PepMap de la base de datos de los Alamos (PepMap (lanl.gov)) para obtener mapas de péptidos, mostrando una tabla con ubicaciones de péptidos tanto en formato HTML como de texto.

ANÁLISIS DE SITIOS DE GLICOSILACIÓN

Para este análisis se utilizó la herramienta N-GlycoSite d([N-Linked Glycosylation Analysis \(lanl.gov\)](http://www.lanl.gov/N-GlycoSite)). La herramienta N-GlycoSite marca y cuenta las ubicaciones donde se produce el patrón NX[ST]Z, donde el aminoácido en la posición X o Z pueden ser determinantes importantes de la eficiencia de la glicosilación. Por ejemplo, una prolina en la posición X o Z desfavorece fuertemente la glicosilación ligada a N. Las señales de glicosilación ligadas a O son más difíciles de predecir y no son parte del análisis de esta herramienta, pero se pueden estimar sus posiciones usando el programa NetPhos en el Centro de Análisis de Secuencias Biológicas (RMG).

Además, se llevó a cabo un análisis de escudo de glicanos con la herramienta Glycan Shield Mapping (Glycan Shield Mapping (lanl.gov)). Se tomaron solo 10 secuencias, 6 pertenecientes a las secuencias recombinantes y 4 a las no recombinantes que incluyeron los sitios de glicosilación con mayor porcentaje en ambos grupos ver tabla 4.

Tabla 4. Sitios de glicosilación por número de secuencias.

Recombinantes			No recombinantes		
Número de sitios de glicosilación	Secuencias que los presentan	Sitos de glicosilación	Número de sitios de glicosilación	Secuencias que los presentan	Sitos de glicosilación
11	96%	88, 156, 241, 262, 276, 301, 386, 448, 392, 611, 637	6	99%	156, 241, 262, 276, 301, 611
8	84%	130, 160, 197, 295, 332, 339, 616, 625.	6	98%	88, 197, 386, 448, 637, 625
7	63%	136, 289, 356, 362, 406, 463, 816	13	74%	130, 136, 160, 289, 295, 332, 339, 356, 392, 406, 616, 816
11	43%	49, 141, 186, 230, 397, 398, 411, 413, 460, 462, 465	9	41%	141, 186, 230, 362, 397, 411, 413, 460, 463

Se muestra en la primera columna el número total de residuos glicosilados a lo largo de una secuencia, en la segunda columna se muestra la proporción de secuencias que incluyen los sitios previamente mostrados y en la tercera columna los sitios de glicosilación, marcando en verde los que son diferentes de las no recombinantes para la primera fila y los que tienen una línea abajo son los que no están en ningún grupo de los no recombinantes.

ANÁLISIS DE USO DE CORRECEPTOR

Para este análisis se utilizó la herramienta VIHCor que se localiza en el servidor: <http://codes.bio/VIHcor/>. Con esta herramienta logramos definir el correceptor usado. Para este paso se necesitó seleccionar solo la zona correspondiente a V3, en este caso con la ayuda de MEGA (Molecular Evolutionary Genetics Analysis) se tomaron solo 36 aminoácidos de la región variable 3 de cada uno de los grupos de secuencias.

VIHCoR se construye utilizando un modelo de Random Forest (RF) que es un método de conjunto que integra muchos árboles de clasificación y regresión (CART) (105). El modelo RF aprovecha dos poderosas técnicas de aprendizaje automático: empaquetado y selecciones aleatorias de características.

En el rendimiento de la predicción, se usan cuatro parámetros estadísticos estándar, precisión (Ac), sensibilidad (Sn), especificidad (Sp) y coeficiente de correlación de Matthews (MCC). Se dan ecuaciones que describen los detalles de los cuatro parámetros estadísticos:

$$Ac = \frac{TP + TN}{(TP + TN + FP + FN)} \times 100$$

$$Sn = \frac{TP}{(TP + FN)} \times 100$$

$$Sp = \frac{TN}{(TN + FP)} \times 100$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

donde TP es el número de péptidos predichos correctamente en el conjunto de datos positivos y TN es el número de péptidos predichos correctamente en el conjunto de datos negativos (105). Mientras tanto, FP representa el número de péptidos predichos incorrectamente en el conjunto de datos positivos y FN representa el número de péptidos predichos incorrectamente en el conjunto de datos negativos. Además, se utilizaron las curvas ROC (características operativas del receptor) para demostrar la capacidad de predicción del modelo propuesto (104).

Los resultados de predicción son validados mediante la predicción estadística de validación cruzada de N veces que es uno de los tres métodos que se utilizan a menudo para identificar el modelo predictivo empírico por su robustez en aplicaciones prácticas (105).

RESULTADOS

DESCRIPCIÓN DE SECUENCIAS DEL GEN ENV DEL VIH-1 SUBTIPO B

Se seleccionaron secuencias completas de env que cumplieran con los criterios establecidos en la sección de Métodos. En resumen, se seleccionaron 6797 secuencias del VIH-1 subtipo B de Norte América y Europa correspondientes a 4 años consecutivos del 2015 al 2018.

-

En este proceso se agruparon las secuencias obtenidas por año. ya que no en todos los años se obtuvieron secuencias para ambas regiones (Norte América y Europa). Para el año 2015 se descargaron 2206 secuencias, para el 2016 fueron 2628 secuencias, para el 2017 solo se encontraban disponibles 207 secuencias, y para el 2018 en dónde s se tuvieron

1256 secuencias, obteniendo un total de 6797 secuencias con las cuales inició el proyecto. La distribución de secuencias por año y región se muestra en la 5.

Tabla 5. Número de secuencias por año y región

Año	Región	Número de secuencias
2015	Norte América	2206
2016	Norte América y Europa	2628
2017	Norte América	207
2018	Norte América y Europa	1256
	TOTAL	6797

DETECCIÓN DE CLÚSTERS DE TRANSMISIÓN

Con el objetivo de identificar posibles clústers de transmisión se utilizó una herramienta a gran escala de epidemiología molecular de VIH-1, que se llama VIH-TRACE (TRANsmision Clúster Engine).

Al utilizar esta herramienta se obtuvieron un gran número de clusters por cada uno de los años, en los cuales se observó una alta variación en el número de secuencias que los conforman (ver tabla 6). El año para el que se obtuvo un mayor número de clusters fue en 2015 con 133 clusters perteneciente a la región de Norte América. El menor número de clusters fue de 2 que corresponden a Europa para el año 2016. En la tabla 6 se indican el número de secuencias totales por año y por región, así como el número de clusters totales obtenidos por año. De las 6797 secuencias analizadas, 5866 fueron clasificadas en uno de los 309 clusters formados, y 931 secuencias quedaron sin agrupación.

Tabla 6. Secuencias totales por región, número de cluster y año.

Región	Clusters totales	Año	Secuencias Totales
Europa	30	2016	596
Europa	2	2018	181
Norte América	133	2015	2016
Norte América	92	2016	1859
Norte América	14	2017	196
Norte América	38	2018	1018
Totales	309	-	5866

Los sets de secuencias no alineadas en arcVIHo FASTA, fueron procesados por VIH-TRACE para generar un arcVIHo JSON que contiene la descripción de las redes de transmisión inferida como nodos (individuos) y enlaces (potenciales socios de transmisión).

La visualización general de los clústers del primer set analizado. correspondiente a la región de Norte América, país Estados Unidos del año 2015 se muestra en la Figura 11.

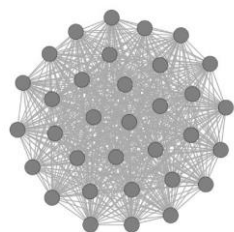


Figura 11. Muestra un clúster de Norte América, año 2015 con VIH-TRACE.

La información se organizó por años y región particularmente para los análisis se decidió que los clusters que continuarán siendo analizados serían aquellos con al menos 20 secuencias y limitando a 10 el número máximo de clusters por año y región geográfica. Al seleccionar los 10 clusters más grandes por año se identificó que estos corresponden a sólo dos países. En la región de Norte América los clusters identificados pertenecen a Estados Unidos y para la Región Europea pertenecen a Alemania.

En la tabla 7 se muestra el tamaño de los clústers de Estados Unidos que cumplen con los criterios para llevar a cabo el análisis de recombinación. De los grupos 1, 2 y 4 los 10 primeros clústers cumplieron el criterio de mínimo de secuencias. Con respecto al grupo 3, se tomaron en cuenta los primeros cuatro clusters, aunque el último de estos solo contaba con 19 secuencias con el fin de evaluar si el poder de nuestro programa de detección de recombinación permite el uso de conjuntos de datos pequeños.

Tabla 7. Descripción de los grupos para Norte América, Estados Unidos

2015 Número de Cluster	Número de secuencias	2016 Número de Cluster	Número de secuencias	2017 Número de Cluster	Número de secuencias	2018 Número de Cluster	Número de secuencias
1	102	1	122	1	57	1	149
2	99	2	102	2	41	2	144
3	83	3	96	3	24	3	113
4	78	4	91	4	19	4	91
5	73	5	82			5	87
6	64	6	78			6	69
7	58	7	76			7	65
8	57	8	75			8	51
9	52	9	69			9	40
10	51	10	66			10	23

Como se observa en la tabla 6, en el grupo correspondiente para el año 2016 de Europa/Alemania se obtuvieron 30 clusters, pero de estos sólo 7 tienen el número mínimo de secuencias, mientras que para el año 2018 solo se identificaron dos clusters y ambos fueron incluidos en el análisis (tabla 8).

Tabla 8. Descripción de los grupos para Europa/Alemania

2016 Número de Cluster	Número de secuencias	2018 Número de Cluster	Número de secuencias
1	92	1	106
2	91	2	75
3	74	-	-
4	60	-	-
5	53	-	-
6	46	-	-
7	44	-	-
8	32	-	-
9	27	-	-

ANÁLISIS ESTADÍSTICO DE LOS CLÚSTERS

Adicionalmente VIH-TRACE reporta la estadística descriptiva del total de nodos, los cuales corresponden a las secuencias que representan a los individuos dentro del clúster y número de enlaces considerados como potenciales socios de transmisión, así como la media, mediana, rango y rango intercuartil para el número de enlaces entre los nodos y el tamaño de los clústers (Tabla 9). esto referido a los datos globales de cada set, 132 clústers del primer grupo, 92 para el segundo, 13 para el tercero, 38 para el cuarto, 30 para el quinto y 2 para el sexto.

Tanto la mediana como la media miden la conectividad de la red y, podemos observar que para el grupo 4 en el link de nodos los datos son simétricos ya que la media y la mediana son similares, pasa lo mismo en el grupo número 6 para el tamaño de los clústers. En este último grupo, las secuencias utilizadas para hacer links fueron ocupadas casi en su totalidad usando 181 de 184.

El mayor número de enlaces fue para el grupo número 4 con 28026 potenciales socios de transmisión para el año 2018 con 1018 individuos. Lo que nos puede sugerir una relación más estrecha en estos clústers.

El rango y el rango intercuartil nos hablan de la dispersión de los datos y como esta se hace más grande en algunos grupos identificados, por ejemplo el rango con respecto al link de los nodos del segundo grupo que nos muestra un rango de 1 a 113 a diferencia del tercer grupo que va de 1 a 55 siendo este con una menor distancia entre los números indicados. Para el tamaño de los clústers el rango más amplio está en el grupo 4 que va de 2 a 149 y el más bajo corresponde al 6 con un rango de 75 a 106, con respecto al rango intercuartil en esta sección tenemos que el más amplio corresponde al set número 3 y que los dos más bajos corresponden al grupo 1 y 6 (ver tabla 10).

Tabla 9: Estadística descriptiva. Datos estadísticos obtenidos del análisis de las 6797 secuencias en HIV-TRACE

	Note América/Estados Unidos				Europa/Alemania	
	2015	2016	2017	2018	2016	2018
Arista	22420	35714	1515	28026	8223	6248
Nodos	2016	1859	196	1018	596	181
Secuencias usadas para hacer links	2016	1859	196	1018	596	181
clústers	133	93	14	39	31	3
Secuencias únicas (singletons)	192	124	13	56	82	4
Nodos						
media	22	38	15	55	28	69
mediana	15	32	14	33	15	100
rango	1 – 94	1-113	1.0 -55	1 - 143	1.0 - 90	2.0 - 104
Rango intercuartil	5 – 32	9 – 61.5	3.0 - 24	ago-35	6.0 - 35.25	28 - 104
Tamaño de los clústers						
media	15	20	15	27	20	91
mediana	6	8.5	10	7	4.5	91
rango	2 – 102	2 - 122	2.0 - 57	2 - 149	2.0 -92	75 - 106
Rango intercuartil	3 – 19	2 - 25.5	2.0 - 57	3.25 - 23	2.0 - 30.75	82.75 - 98.25

EXTRACCIÓN DE SECUENCIAS POR CLÚSTERS PARA SU ALINEAMIENTO

Para cada uno de los clústers seleccionados por su tamaño se obtuvieron las secuencias alineadas. Los lineamientos fueron examinados con el programa MEGA 5.05 (Molecular Evolutionary Genetics Analysis), para verificar su correcta alineación y posteriormente detectar señales de recombinación.

PRUEBA DE ESTADÍSTICO F

Para determinar las diferencias genéticas entre los clusters de Estados Unidos y Alemania. Se realizó una prueba del estadístico F el cual nos indicó un valor de p de cero. En este caso nos indica que los grupos de secuencias de las dos regiones efectivamente corresponden a poblaciones independientes las cuales pueden ser analizadas de forma separada.

Análisis de Recombinación

Se realizó la detección de recombinación de 43 clusters de transmisión en el programa RAPR, para un total de 3184 secuencias (Tabla 10). En 12 sets no se detectó recombinación (tabla 11). Los sets y los años, así como el número de clusters en estos y el número de secuencias en cada uno se representan en la tabla 11R.

Tabla 10. Secuencias totales por país para el análisis de Recombinación.

Año y País	Secuencias totales analizadas en RAPR
2015 EU	717
2016 A	519
2016 EU	857
2017 EU	141
2018 A	181
2018 EU	769
Total	3184

Tabla 11. Clusters dónde no se detectó recombinación.

Año y País	Cluster
2015 EU	4-8
2016 A	2
2016 EU	2-3-6-7
2017 EU	1-3
2018 A	1
2018 EU	2-8

Tabla 12.. Número de secuencias recombinantes por cluster.

Año, País y cluster	Secuencias recombinantes
C 1 - 2015 EU	7
C 2 - 2015 EU	4
C 3 - 2015 EU	49
C 5 - 2015 EU	4
C 6 - 2015 EU	5
C 7 - 2015 EU	25
C 9 - 2015 EU	2
C 10 - 2015 EU	18
C 1 - 2016 EU	19
C 4 - 2016 EU	17
C 5 - 2016 EU	5
C 8 - 2016 EU	18
C 9 - 2016 EU	4
C 10 - 2016 EU	24
C 2 - 2017 EU	8
C 4 - 2017 EU	1
C 1 - 2018 EU	6
C 3 - 2018 EU	14
C 4 - 2018 EU	16
C 5 - 2018 EU	10
C 6 - 2018 EU	27
C 7 - 2018 EU	16
C 1 - 2016 A	12
C 3 - 2016 A	34
C 4 - 2016 A	12
C 5 - 2016 A	8
C 6 - 2016 A	12
C 7 - 2016 A	15
C 8 - 2016 A	8
C 9 - 2016 A	11
C 1-2018 A	5
Total	416

En un conteo global, se detectaron 986 puntos de recombinación estadísticamente significativos (breakpoints) fueron 986, el promedio de la frecuencia de recombinación fue 0.20810594. El grupo que presentó más puntos de recombinación significativos fue el correspondiente al año 2015 de Estados Unidos (EU) seguido del de Alemania del año 2016, estos dos grupos también presentan mayor número de secuencias recombinantes, en contraparte los grupos con menor número de breakpoints fueron Estados Unidos 2017

y Alemania 2018 con 29 y 10, respectivamente, siendo los mismos con el menor número de secuencias recombinantes. Detalles en la tabla 13.

Tabla 13. Número de breakpoints por año y frecuencias de recombinación.

Año-País	Breakpoints totales	Breakpoints significativos	Porcentajes de breakpoints significativos	Secuencias Recombinantes	Frecuencia de recombinación
2015-EU	605	285	29%	114	0.2114
2016-EU	455	206	21%	87	0.1798
2017-EU	83	29	3%	9	0.1278
2018-EU	780	186	19%	89	0.1844
2016-A	784	270	27%	112	0.2814
2018-A	31	10	1%	5	0.0676
Global	2738	986	100%	416	0.20810594

De toda la información generada se construyó un mapa genético para identificar en qué estructuras se localizan estos puntos de recombinación. De forma global la mayoría se localizan en la glicoproteína 120, específicamente en las regiones variables 1 y 2 contando con 205 de los sitios. La región con menos puntos de recombinación fue el dominio conservado 5 (C5) y la región héptada 2 (HR2) con 6 y 7 puntos respectivamente. Ver tabla 14

Tabla 14. Breakpoints por estructura.

Estructuras de ENV	Breakpoints significativos	Estructuras de ENV	Breakpoints significativos
SP	44	FP*	22
C1*	62	HR1*	15
V1*V2	205	HR2*	7
C2*	104	MPER*	144
V3*	36	TMD*	79
C3*	79	CT*	25
V4*	42	-	-
C4*	106	-	-
V5*	101	-	-
C5*	6	-	-
Total gp120	785	Total gp41	153

*En gp120, dominios variables (V1 - V5), secuencias conservadas (C1 - C5). En gp41: el péptido de fusión (FP), dos regiones de héptadas repetidas (HR1 y HR2), el dominio transmembranal (TMD) y la cola citoplasmática (CT).

En la Figura 12 se muestra la descripción gráfica de los breakpoints detectados y agrupados por país. En la figura se muestra la frecuencia de breakpoints en ventanas de 50 aa., lo que quiere decir es que se dividió el número de breakpoints de cada una de las ventanas entre el número de secuencias analizadas para poder comparar los dos grupos de secuencias. Además, se muestra el número total de breakpoints en cada ventana para tener una mejor interpretación.

Cabe resaltar que los puntos de recombinación se agrupan en ciertas regiones. Se observan hotspots en las ventanas de las posiciones 151 a 200 y de 401 a 451 teniendo las zonas aledañas con una frecuencia media de puntos de recombinación. Notablemente vemos que la mayoría de los puntos se localizan en gp120 para todos los grupos, observando menos en la región de gp41 y variando de acuerdo al país esto de acuerdo al número de secuencias analizadas. Ver figura 12.

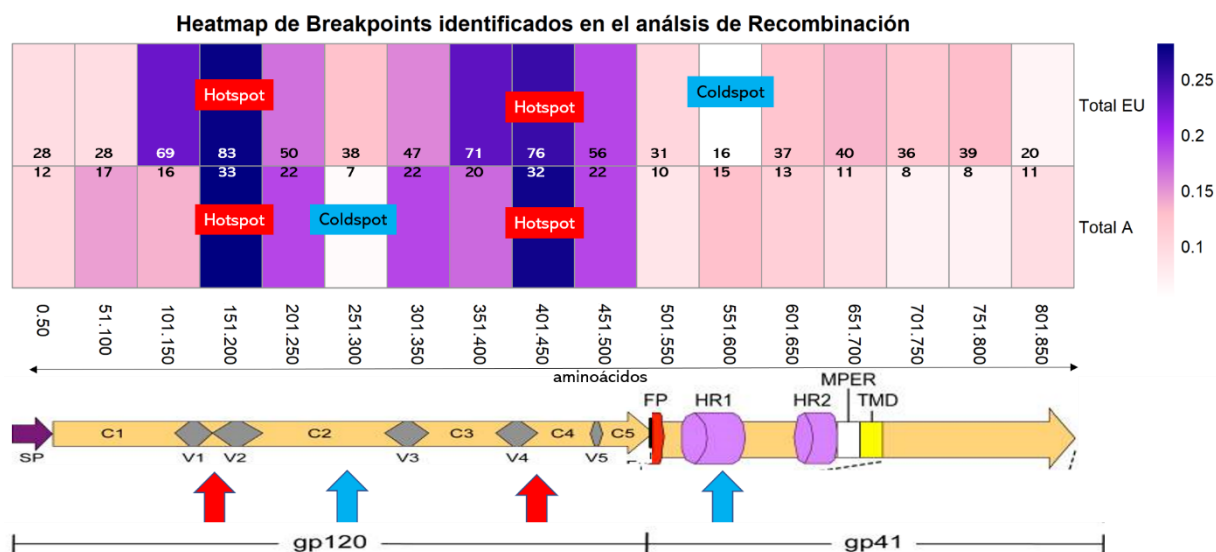


Figura 12. HeatMap de los puntos de recombinación usando una ventana deslizante de 50 aminoácidos. Los puntos de recombinación se agruparon por país. La secuencia de Env está dividida en 17 ventanas de 50 aminoácidos cada una. La acumulación de breakpoints se identifica en una escala de color para indicar las diferencias en la frecuencia de breakpoints. Se identifican en cuadros rojos los hotspots y en azul coldspots. Además de que se señalan estos sitios en la estructura de env que se muestra en la parte inferior.

PATRONES DISTINTOS EN LAS POSICIONES BAJO PRESIÓN SELECTIVA POSITIVA DE LAS SECUENCIAS.

Para este análisis se compararon conjuntos de secuencias recombinantes y no recombinantes. Para detectar sitios que evolucionaron bajo selección se utilizaron dos herramientas con distintos enfoques (verosimilitud en el caso de SLAC y estadística bayesiana del programa FUBAR). Se localizaron para cada año sitios específicos bajo presión selectiva positiva. En los seis grupos analizados se obtuvieron 137 sitios bajo presión selectiva positiva en las secuencias recombinantes y 219 en las no recombinantes, cuando se utilizó en SLAC ($p < 0.01$). Para los mismos datos en FUBAR detectó 152 sitios para las recombinantes y 266 sitios en las secuencias no recombinantes con una probabilidad posterior de 0.9 (Ver tabla 15).

Tabla 15. Posiciones bajo selección positiva y negativa por grupo entre las secuencias recombinantes y no recombinantes de ambos análisis.

Presión selectiva	Recombinantes	Año	No recombinantes	Presión selectiva	Recombinantes	Año	No recombinantes	Presión selectiva
+	18	A2016	43	+	10	A2016	31	+
-	137		180	-	223		275	-
+	0	A2018	0	+	8	A2018	14	+
-	0		6	-	7		11	-
+	22	2018US	72	+	45	2018US	91	+
-	78		126	-	187		240	-
+	0	2017US	2	+	6	2017US	32	+
-	12		32	-	28		51	-
+	13	2016US	39	+	26	2016US	46	+
-	50		109	-	160		220	-
+	84	2015US	63	+	57	2015US	52	+
-	278		166	-	359		249	-
+	137	Totales	219	+	152	Totales FUBAR	266	+
-	555	SLAC	619	-	964		1046	-

Al realizar el mismo análisis con ambas herramientas, pero con los datos agrupados por país, no por años, podemos notar que las proporciones cambian obteniendo con el análisis de SLAC se localizaron 84 sitios de recombinación positiva para todas las secuencias recombinantes en cambio en el análisis de FUBAR se identificaron 57.

Se muestra un esquema del número de sitios bajo presión selectiva positivas de las secuencias recombinantes totales obtenidas del análisis en FUBAR y SLAC en las ventanas de 50 a.a., para identificar si se localizan en hotspots de recombinación (Figura 13), y notamos que la distribución de estos sitios no coincide con los hotspots, pero si se encuentran cerca de las zonas de mayor número de breakpoints, como por ejemplo los sitios aledaños a la ventana correspondiente a 401-450 a.a. que corresponden a 351-450 a.a y 301-350 a.a , tiene 7 y 10 sitios bajo presión selectiva positiva, en SLAC, 8 y 4 sitios bajo presión selectiva en FUBAR respectivamente.

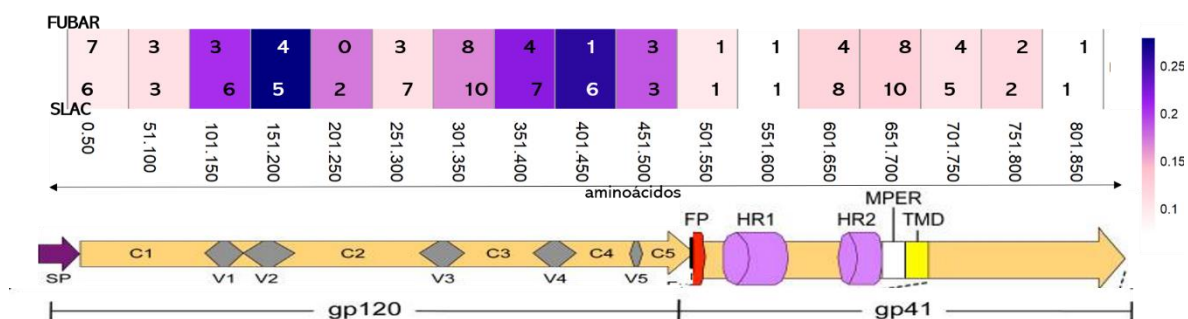


Figura 13. Sitios bajo selección positiva en el esquema de breakpoints. Se muestra el número de posiciones con selección positiva en cada una de las ventanas. Se agrega la estructura de Env con gp120 y gp41

Se observó que hay más posiciones bajo selección negativa que positiva. En general, la capacidad de identificación de sitios bajo presión selectiva fue proporcional al número de secuencias analizadas en cada uno de los sets. Se muestran las gráficas obtenidas del análisis en SLAC las mutaciones sinónimas y las no sinónimas comparando las secuencias recombinantes con las no recombinantes (Figura 14).

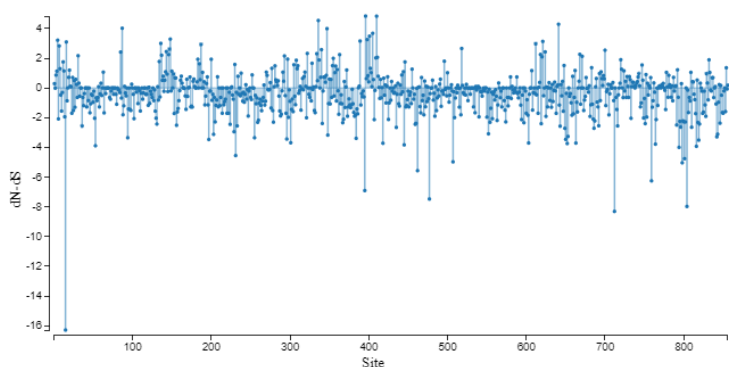


Figura 14. Gráfica de mutaciones no sinónimas y sinónimas en la secuencia consenso del grupo de recombinantes por el método SLAC. Se observan las mutaciones sinónimas a partir de cero dirección hacia arriba y las no sinónimas del cero hacia abajo en cada uno de los aminoácidos.

DIFERENTES PATRONES DE VARIABILIDAD ENTRE SECUENCIAS RECOMBINANTES Y NO RECOMBINANTES

Para explorar la diferencia en variabilidad de las secuencias recombinantes y no recombinantes se realizaron comparaciones de entropía de Shannon, que se usa para estimar la variabilidad en cada una de las posiciones del alineamiento teniendo en cuenta la frecuencia y el número de bases presentes. Esto con la finalidad de observar el grado de

variabilidad entre los dos grupos de secuencias (recombinantes y no recombinantes). Lo que podemos observar es que si hay sitios específicos en dónde se localizan diferencias entre los grupos, por ejemplo, en las no recombinantes se notan sitios de alta variabilidad en la parte inicial de la secuencia correspondiente a las primeras posiciones del péptido señal y en entre los aminoácidos 250 y 410 ambas secciones en gp120, en cambio en las Recombinantes se nota más a la diferencia en las posiciones correspondientes a C4 y V5 que va entre los a.a., 400 a 425.

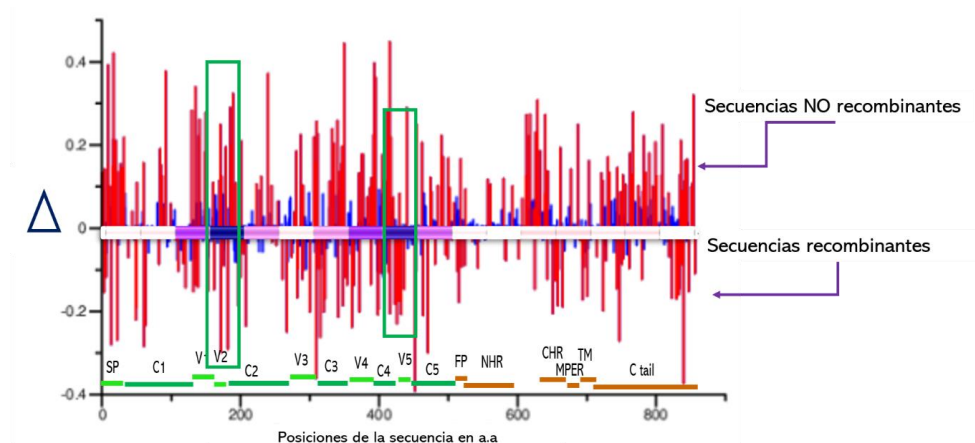


Figura 15. Entropía de Shannon en las secuencias del gen *env*. Se muestran dos conjuntos de secuencias alineadas, secuencias No recombinantes parte superior y Recombinantes parte inferior. Cada posición con una diferencia significativa en la variabilidad entre estos dos conjuntos se resalta en color rojo. Se muestran las estructuras de las glicoproteínas, así como su longitud.

BÚSQUEDA DE EPÍTOPOS DE CTL/CD8

La identificación de epítomos reportados se llevó a cabo en la base de datos de inmunología molecular de VIH dentro de la base de datos de los Álamos en la cual se identificaron 140 epítomos específicos en los pacientes de Estados Unidos y 8 de Alemania en los cuales sólo 80 de estas muestras de Estados Unidos fueron identificados en las secuencias analizadas, con diferentes proporciones entre las secuencias recombinantes y no Recombinantes. En las secuencias de Alemania sólo dos epítomos de ocho fueron identificados en las secuencias (Tabla 16). En la tabla se observa que no todos los epítomos están presentes en las mismas proporciones entre los dos grupos de secuencias, por ejemplo: TTLFCASDAK y KLTPLCVTL están presentes en un 80% en las secuencias recombinantes mientras que en las no recombinantes estos epítomos están presentes solo en un 31 y 27% respectivamente. De la misma forma podemos marcar los dos más abundantes en las no

recombinantes que corresponden a: WLWYIKIFI y RPVVSTQLLL con un 78% y 54% para cada uno, tomando en cuenta que solo WLWYIKIFI está en un porcentaje alto en las recombinantes y que RPVVSTQLLL no está en los primeros 25 epítomos más abundantes.

Tabla 16. Porcentaje de epítomos en las secuencias recombinantes y No recombinantes, mostrando en forma descendente de cada una para mejor comparación.

Epítopo	Recombinantes	No recombinantes	Epítopo	Recombinantes	No recombinantes
TTLFCASDAK	84.41	31.45	WLWYIKIFI	75.3	78.66
KLTPLCVTL	80.1	27.64	RPVVSTQLLL	44.12	54.06
NOGGEFFYCN	77.22	20.67	VPVWKEATTTTL	35.25	52.95
IVNRVRQGGVSP	76.74	19.57	KYVWVNLQV	59.47	53.73
WLWYIKIFI	75.3	78.66	ERYLKDQQL	59.47	53.62
LFCASDAKAY	72.66	23.38	QELKNSAVSL	58.75	52.57
HSPNCGGEF	71.7	16.58	VPVWKEATTTTL	32.37	51.8
KPCVKLTPLC	70.5	26.98	PCASDAK	61.63	51.02
SPNCGGEFFY	70.26	13.49	YLKDQQL	58.51	50.14
SPNCGGEFF	70.26	13.49	RYLKDQQL	58.51	50.05
VKVKIEPI	61.63	51.02	CTRPNNNTRK	42.45	49.7
KYVWVNLQV	59.47	53.73	GRGWWEALK	45.56	47.37
ERYLKDQQL	59.47	53.62	RRGWWEALKY	45.56	47.37
QELKNSAVSL	58.75	52.57	RIRQGLERA	57.55	47.32
YLKDQQL	58.51	50.14	RNNNTRKSI	48.92	45.19
RYLKDQQL	58.51	50.05	NVTENFMW	51.8	44.61
RIRQGLERA	57.55	47.32	VTFENFMWKN	49.4	44.17
CKNVSTVQC	52.04	39.8	IKCSSNITGL	35.25	42.07
NVTENFMW	51.8	44.61	IPCRHQH	30.7	41.02
VTFENFMWKN	49.4	44.17	VERYLKDQQL	44.6	40.96
RNNNTRKSI	48.92	45.19	IPRRIRQGL	47.96	40.08
IPRRIRQGL	47.96	40.08	CKNVSTVQC	52.04	39.8
GRGWWEALK	45.56	47.37	RIRQGLERALL	42.45	38.47
RRGWWEALKY	45.56	47.37	NMWQEVGKAM	29.26	36.48
VERYLKDQQL	44.6	40.96	RAIEAQQHL	26.62	35.66
RPVVSTQLLL	44.12	54.06	RLRDLLLV	31.41	32.23
RIRQGLERALL	42.45	38.47	RIKQHNMW	22.06	31.67
CTRPNNNTRK	42.45	49.7	TTLFCASDAK	84.41	31.45
RGPDRIPEGIEEE	35.97	19.07	RLRDLLLVTR	29.74	31.01
GGEDR					
VPVWKEATTTTL	35.25	53.95	SLNATAIAY	30.22	30.96
IKCSSNITGL	35.25	42.07	PVNTTQLP	34.53	30.51
PVNTTQLP	34.53	30.51	RLISCNSTV	28.78	27.81
NLLQYWSQELK	33.81	22.33	KLTPLCVTL	80.1	27.64
NSAVSL					
VPVWKEATTTTL	32.37	51.8	KPCVKLTPLC	70.5	26.98
PCASDAK					
RLRDLLLV	31.41	32.23	LFCASDAKAY	72.66	23.38
IPCRHQH	30.7	41.02	TAVPWNASW	24.22	22.39
SLNATAIAY	30.22	30.96	NLLQYWSQELK	33.81	22.33
NSAVSL					
RLRDLLLVTR	29.74	31.01	NOGGEFFYCN	77.22	20.67
NMWQEVGKAM	29.26	36.48	IVNRVRQGGVSP	76.74	19.57
RLISCNSTV	28.78	27.81	RGPDRIPEGIEEE	35.97	19.07
RAIEAQQHL	26.62	35.66	GGEDR		
VKIEPLGVAPTK	24.7	10.67	YSPLSLQTL	17.75	17.63
AKRRVVQR					
TAVPWNASW	24.22	22.39	HSPNCGGEF	71.7	16.58
RIKQHNMW	22.06	31.67	VGKAMYAPPPIR	16.31	15.98
IVTRIVELL	19.42	12.94	GGEDR		
IPSVIRIRDLIL	19.42	12.94	SPNCGGEFFY	70.26	13.49
IVTRIV			SPNCGGEFF	70.26	13.49
YSPLSLQTL	17.75	17.63	IVTRIVELL	19.42	12.94
VGKAMYAPPPIR	16.31	15.98	IPSVIRIRDLIL	19.42	12.94
GGEDR			IVTRIV		
HPRRIRQGLER	16.07	9.78	INRWQEVGKA	11.27	12.77
ALL	13.67	11.77	MYAPPPI		
VAPVNGGI	13.19	7.68	YRLISCNSTVIT	9.83	12.33
AASITLVQARQ	12.95	9.23	QACPIV	13.67	11.77
L.LNGIV	12.23	7.46	VAPVNGGI	13.67	11.77
INRWQEVGKA	11.27	12.77	VKIEPLGVAPTK	24.7	10.67
MYAPPPI			AKRRVVQR		
AENLWVTVY	11.03	6.74	LHIPTRIRQGLE	2.4	10.17
IVQQQSLLRAI	10.07	3.92	RAI	16.07	9.78
CAQGHM	9.83	12.33	HPRRIRQGLER	16.07	9.78
YRLISCNSTVIT	9.83	12.33	ALL	13.67	11.77
QACPIV	9.59	3.37	IVFNQSSGGDPE	8.39	9.67
L.NATAIAVARG	9.59	3.37	IVMHSE		
TDIRIK	8.87	2.93	AENLWVTVY	11.03	6.74
DIWDNMTWM	8.87	2.93	AKRRVVQREKR	6.24	4.81
IVFNQSSGGDPE	8.39	9.67	AVGICAM		
IVMHSE					
AKRRVVQREKR	6.24	4.81	AKRRVVQREKR	6.24	4.81
AVGICAM					
EKLWVTVVY	6	8.79	RACRAILH	6	4.75
RACRAILH	6	4.75	RAIHPRIRHQ	4.8	4.53
IEVVQRAICRAIL	5.76	1.66	GLERAIL	4.08	4.15
HIPRI	5.52	3.43	VQRTCRAIL	4.08	4.15
RRIRQGLEKILL	5.52	3.43	IVQQQSLLRAI	10.07	3.92
EAQGHM					
NEIWDNMTW	5.04	7.57	RRIRQGLEKILL	5.52	3.43
RAIHPRIRHQ	4.8	4.53	ELNATAIAVARG	9.59	3.37
GLERAIL	4.32	3.15	TDIRIK	8.87	2.93
VQRTCRAIL	4.08	4.15	DIWDNMTWM	8.87	2.93
LHIPTRIRQGLE	2.4	10.17	IEVVQRAICRAIL	5.76	1.66
RAI	1.92	1.22	HIPRI	1.92	1.22
IGRAIHPR	1.92	1.22	IGRAIHPR	1.92	1.22
DPNPQEMVL	0.72	0.44	AVLSVVRV	0.24	1
ELMISAT	0.48	0.39	DLRLGLESYH	0.24	0.88
IVQENENVINC	0.48	0.39	RLRDLL	0.48	0.61
THIRN			WEVLKCVWVNL	0.48	0.61
WEVLKCVWVNL	0.48	0.61	LYWSQEL		
LYWSQEL					
AVLSVVRV	0.24	1	DPNPQEMVL	0.72	0.44
IVQENENVINC	0.48	0.39			
THIRN					
ELMISAT	0.48	0.28			
IVQENENVINC	0.48	0.39			
THIRN					
WEVLKCVWVNL	0.48	0.61			
LYWSQEL					
PCASDAKSY	0.24	0.06			
RLRDLL	0.24	0.88			
PCASDAKSY	0.24	0.06			

EPÍTOPOS Y LOCALIZACIÓN EN LAS SECUENCIAS.

Para visualizar e integrar la información y realizar la comparación con los hotspots identificados se mapearon solo los epitopes que se encontraron en estas regiones, ver Figura 17. Esta nos muestra que los epitopes más frecuentes no se localizaron en los sitios identificados como hotspots de recombinación (marcados con una línea azul marino). Sólo un epitope abundante en las Recombinantes identificadas se encuentra en un hotspot que va de los residuos 401 a 450. Se muestra en la tabla 18 que indica cuales son los epitopes identificados en los hotspots y su porcentaje en las secuencias recombinantes y no recombinantes.

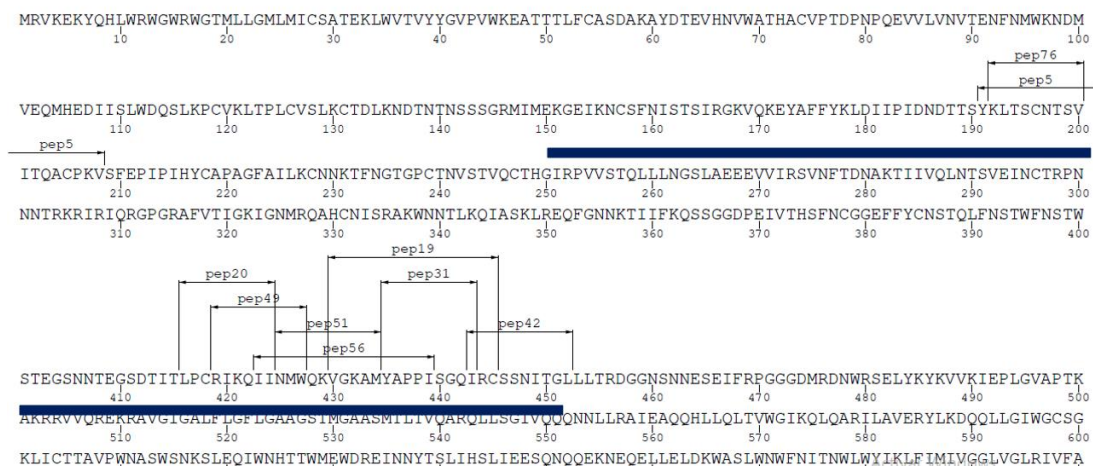


Figura 17. Mapa de epítomos reportados en pacientes de Estados Unidos y su posición en la secuencia de referencia. Las líneas azul marino indican los sitios de hotspots.

Tabla 18. Epitopes identificados en los hotspot y su proporción.

		Porcentaje en secuencias recombinantes	Porcentaje en secuencias No recombinantes
pep5	YRLISCVITQACPKV	9.83	12.33
pep76	RLISCVTSV	28.78	27.81
pep56	IINRWQEVGKAMYAPPI	11.27	12.77
pep51	NMWQEVGKAM	29.26	36.48
pep49	RIKQIINMW	22.06	31.67
pep42	IRCSSNITGL	35.25	42.07
pep31	YAPPISSGQI	13.67	11.77
pep19	VGKAMYAPPISSGQIRC	16.31	15.98
pep20	LPCRIKQII	30.70	41.02

ANÁLISIS DE SITIOS DE GLICOSILACIÓN

En este análisis se comparó el número de sitios de glicosilación entre las secuencias recombinantes y no recombinantes. En la tabla 19 se muestran las posiciones glicosiladas correspondientes a los hotspots identificados y se marca la proporción entre las recombinante y no recombinantes que no muestran un cambio en el porcentaje entre los

grupos, de hecho se pueden indicar que las recombinantes no presentan dos sitios glicosilados que las no recombinantes sí (sitios 150 y 197), al contrario estas no presentan el sitio 189 y las recombinantes sí, mostrando este como uno de los más abundantes para las recombinantes con el 92% y el 186 con el 43% contra el 29% de las no recombinantes, esto para el primer hotspot, en el segundo se muestran dos sitios que son más abundantes para las recombinantes correspondientes al 413 y 415 con un 37% y 15% respectivamente. De las 23 posiciones glicosiladas 7 fueron identificadas con Asparagina que es un aminoácido sin carga, todas las demás fueron muy variables sin poder indicar un consenso.

Tabla 19. Sitios de glicosilación identificados en los hotposts de gp120.

Sitios de Glicosilación	Porcentaje en secuencias Recombinantes	Porcentaje en secuencias No recombinantes
1er hotspot	150	0
	156	100%
	160	88%
	185	3%
	186	43%
	189	92%
	197	0
2do hotspot	400	7%
	401	19%
	402	1%
	403	17%
	405	2%
	406	59%
	407	9%
	408	2%
	409	4%
	410	2%
	411	29%
	413	37%
	442	15%
	444	7%
	446	3%
	448	97%

En este análisis se identificaron 1942 sitios de Glicosilación en las 417 Secuencias Recombinantes, contra 50558 en las 1810 secuencias de las no recombinantes. Además de que de acuerdo a la longitud de la secuencia en las recombinantes se localizaron 88 sitios que se glicosilan y en las secuencias no recombinantes 93 sitios. La mayoría de estos sitios de glicosilación identificados se localizan en gp120 y solo 5 sitios se localizaron en gp41.

ANÁLISIS DE ESCUDO DE GLICANOS

Para este análisis primero se tomaron las secuencias más representativas que presentaran el mayor número de sitios glicosilados identificados previamente.

El análisis de la evolución del escudo de glicanos de las secuencias recombinantes y no recombinantes mostró un poco de variación en el número de agujeros en el escudo de glicanos oscilando entre 113 a 94 en las recombinantes y 107 a 99 en las no recombinantes. Estos fueron identificados y colocado en cada una de las ventanas de 50 a.a., identificando 3 zonas en dónde se acumulan más, en el rango de 151 a 200, 301-350 y en 601-650, como se muestra en la tabla 20 y estos se localizan en V2, C2 en CHR (repetición heptada terminal). Casi todos los sitios para todas las secuencias son los mismos, de hecho, los 10 sitios analizados para gp120 en ventanas de 50 en 50 a.a presentes en las 10 secuencias corresponden a los mismos sitios en todas, y para gp41 tampoco hay variación en los 3 segmentos que casi no presenta variación. No se presentan diferencias en los agujeros de glicanos entre las secuencias recombinantes y no recombinantes.

Tabla 20. Distribución de agujeros de glicano en secuencias recombinantes y no recombinantes a lo largo de las estructuras gp120 y 41.

gp120										gp41			
	0-50	51-100	101-150	151-200	201-250	251-300	301-350	351-400	401-450	451-500	501-550	551-600	601-650
No R	0	6	4	19	7	10	20	7	7	8	1	0	10
	0	6	4	19	7	9	19	17	8	8	0	0	10
	0	6	4	26	7	10	19	7	9	7	0	0	10
	0	6	4	26	7	13	21	4	7	7	1	0	10
	0	6	4	26	7	13	21	4	7	7	1	0	10
	0	6	4	26	7	12	19	8	9	7	1	0	10
	0	6	4	26	7	9	17	4	10	7	1	0	10
	0	6	4	26	7	13	19	8	9	7	0	0	10
	0	6	4	26	7	8	19	17	8	8	0	0	10
	0	6	4	19	7	7	20	7	7	6	1	0	10
Recombinantes	0	6	4	19	7	10	20	7	7	8	1	0	10
	0	6	4	19	7	9	19	17	8	8	0	0	10
	0	6	4	26	7	10	19	7	9	7	0	0	10
	0	6	4	26	7	13	21	4	7	7	1	0	10
	0	6	4	26	7	13	21	4	7	7	1	0	10
	0	6	4	26	7	12	19	8	9	7	1	0	10
	0	6	4	26	7	9	17	4	10	7	1	0	10
	0	6	4	26	7	13	19	8	9	7	0	0	10
	0	6	4	26	7	8	19	17	8	8	0	0	10
	0	6	4	19	7	7	20	7	7	6	1	0	10

Se puede visualizar la distribución de estos agujeros de glicano en la siguiente figura la cual fue dividida con líneas rojas por casa 50 a.a., y se colocaron los sitios de los agujeros de glicano en cada sección, esta solo representa a gp120 iniciando a partir de C1.

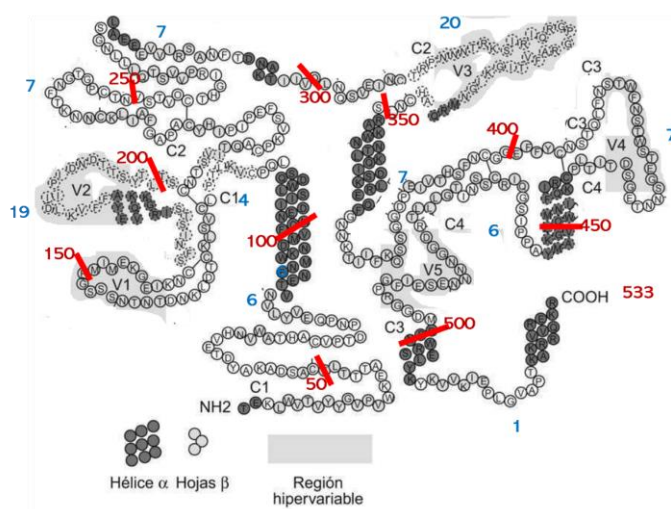


Figura 17. Estructura de proteína gp120. La proteína está dividida cada 50 a.a con una línea roja y números rojos indicando hasta que posición llega y en números azules se indican el número de sitios de agujeros de glicano en cada segmento. Se marca la estructura secundaria con círculos en gris oscuro y en círculos en gris claro las hojas beta, además de marcar en regiones sombreadas las regiones variables de la 1 a la 5.

En la figura 18 se muestra una representación en una estructura cristalina de Env trimérica de referencia, asumiendo que cada PNGS está ocupado por glicanos. Para aproximar la dinámica de los glicanos, asumimos que cada glicano protege la superficie de la proteína Env dentro de un radio de 10 Å del átomo de carbono alfa de la asparagina PNGS, un límite utilizado para caracterizar la superficie de la proteína Env en la vecindad de los glicanos. Mediante el mapeo de áreas protegidas alrededor de cada PNGS en la superficie de la proteína trímica, se puede estimar el grado de protección de glucano para cada trímero y se muestra una representación de una secuencia recombinante.

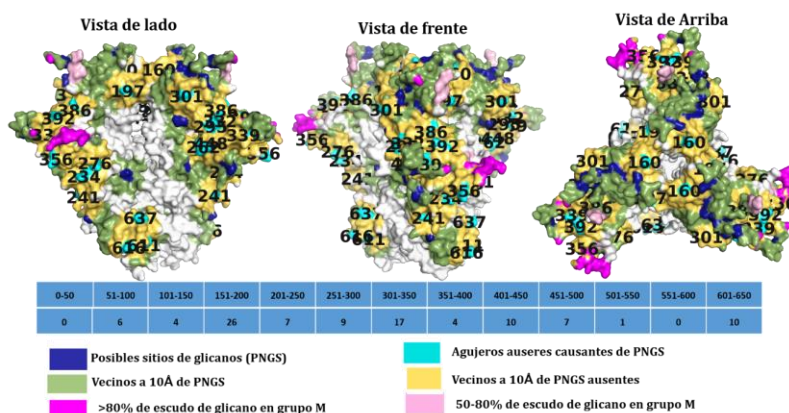


Figura 18. Agujeros de glicano en la estructura de Env vista de 3 ángulos. Se Muestra el escudo de glicano predicho desde 3 puntos de vista, de izquierda a derecha: 1) "lado" gira

la vista anterior $\sim 120^\circ$ grados en sentido contrario a las agujas del reloj alrededor del eje vertical; 2) "frente" es la vista con la membrana viral en la parte inferior y 3) "superior" muestra la vista a lo largo del eje del trímero desde el vértice del trímero. Así como el número de sitios por cada ventana de a.a.

ANÁLISIS DE USO DEL USO DE CORRECEPTOR

Este análisis se realizó en 416 secuencias recombinantes y 1809 No recombinantes, en las que la longitud de secuencia del loop V3 era la mínima para llevar a cabo el análisis (36 a.a.). En todos los casos obtuvimos una alta probabilidad de uso del mismo correceptor que fue CCR5 indicándonos que las muestras pudieron ser tomadas en estadios tempranos de la infección. Los valores de la predicción variaron en algunos decimales, pero en ambos grupos, recombinantes y no recombinantes se obtuvo el mismo resultado (Ver tabla 21).

Tabla 21. Valores de predicción de uso de correceptor.

No Recombinantes				Recombinantes		
Secuencia	Valor de CCR5	Valor de CXCR4	Predicción	Valor de CCR5	Valor de CXCR4	Secuencia
R.US.2015.A53	0.615	0.305	CCR5	0.598	0.402	R.US.2015.A53
40A01.w005.K						40A01.w007.K
X587000						X587024.1
R.US.2015.A53	0.615	0.305	CCR5	0.607	0.393	R.US.2015.A01
40A01.w005.K						CD4T.C1.15A4
X587011						MH26375.L1
R.US.2015.A53	0.615	0.305	CCR5	0.672	0.328	R.US.2015.A01
40A01.w005.K						CD4T.C1.15C5
X587012						MH263756.2
R.US.2015.A53	0.607	0.393	CCR5	0.598	0.402	R.US.2015.A01
40A01.w005.K						CD4T.C1.15A3
X587014						MH263756.1
R.US.2015.A53	0.607	0.393	CCR5	0.63	0.37	R.US.2015.A01
40A01.w005.K						CD4T.C1.16G12
X587015						MH263784.1
R.US.2015.A53	0.607	0.393	CCR5	0.67	0.33	R.US.2015.A01
40A01.w005.K						CD4T.C1.12C3M
X587016						H263800.1
R.US.2015.A53	0.615	0.385	CCR5	0.67	0.33	R.US.2015.A01
40A01.w005.K						CD4T.C1.15D7
X587017						MH263768.1
R.US.2015.A53	0.607	0.393	CCR5	0.81	0.19	R.US.2015.A02
40A01.w005.K						CD4T.C1.18D11
X587019						MH263897.1
R.US.2015.A53	0.607	0.393	CCR5	0.81	0.19	R.US.2015.A02
40A01.w005.K						CD4T.C1.18F5
X587020						MH263907.1
R.US.2015.A53	0.615	0.385	CCR5	0.596	0.404	R.US.2015.A02
40A01.w005.K						CD4T.C1.18H1
X587021						MH263913.1
R.US.2015.A53	0.615	0.385	CCR5	0.596	0.404	R.US.2015.A02
40A01.w007.K						CD4T.C1.15D11
X587022						MH263933.1
R.US.2015.A53	0.607	0.393	CCR5	0.675	0.325	R.US.2015.1HC
40A01.w007.K						2.D0.012016.8
X587023						12.1
R.US.2015.A53	0.607	0.393	CCR5	0.675	0.325	R.US.2015.1HC
40A01.w007.K						2.D0.012016.3
X587025						5.1
R.US.2015.A53	0.607	0.393	CCR5	0.675	0.325	R.US.2015.1HC
40A01.w007.K						2.D0.012016.0
X587026						4.1
R.US.2015.A53	0.615	0.385	CCR5	0.821	0.179	R.US.2015.1HC
40A01.w007.K						2.W1.071116.8
X587027						4.1
R.US.2015.A53	0.615	0.385	CCR5	0.675	0.325	R.US.2015.1HC
40A01.w007.K						2.D0.012016.A
X587028						5.1
R.US.2015.A53	0.615	0.385	CCR5	0.675	0.325	R.US.2015.1HC
40A01.w007.K						2.D0.012016.C7
X587029						1
R.US.2015.A53	0.607	0.393	CCR5	0.675	0.325	R.US.2015.1HC
40A01.w007.K						2.D0.012016.C8
X587030						1
R.US.2015.A53	0.615	0.385	CCR5	0.675	0.325	R.US.2015.1HC
40A01.w007.K						2.D0.012016.C9
X587031						1
R.US.2015.A53	0.615	0.385	CCR5	0.675	0.325	R.US.2015.1HC
40A01.w007.K						2.D0.012016.E1
X587032						1
R.US.2015.A53	0.598	0.402	CCR5	0.675	0.325	R.US.2015.1HC
40A01.w007.K						2.D0.012016.F5
X587033						1
R.US.2015.A53	0.615	0.385	CCR5	0.675	0.325	R.US.2015.1HC
40A01.w007.K						2.D0.012216.C6
X587034						1
R.US.2015.A01	0.672	0.328	CCR5	0.675	0.325	R.US.2015.1HC
CD4T.C1.16C5						2.D0.012216.C8
MH263773						1
R.US.2015.A01	0.598	0.402	CCR5	0.675	0.325	R.US.2015.1HC
Q1.13M28C2.M						2.D0.012216.E2
H263839						1
R.US.2015.A01	0.607	0.393	CCR5	0.675	0.325	R.US.2015.1HC
R2.20151014.1						2.D0.012216.H
59G7.MH26388						7.1
R.US.2015.A01	0.628	0.372	CCR5	0.675	0.325	R.US.2015.1HC
CD4T.C1.2F12						2.W1.070816.B
MH263806						12.1
R.US.2015.A53	0.647	0.353	CCR5	0.767	0.233	R.US.2015.1HC
40A05.w004.K						2.W1.070816.C
X587186						2.1
R.US.2015.A53	0.647	0.353	CCR5	0.675	0.325	R.US.2015.1HC
40A05.w004.K						2.W1.070816.E
X587187						3.1

DISCUSIÓN

En este proyecto se ha descrito por primera vez la recombinación intrasubtipo en el gen env de VIH-1 subtipo B con herramientas bioinformáticas, para dos países, Estados Unidos y Alemania con una N de más de 400 secuencias lo cual no se había realizado anteriormente.

Como ya lo había indicado Sergei et al., [84] VIHTRACE sirvió adecuadamente para detectar grupos de transmisión en las regiones de Norte América y Alemania lo cual nos delimitó los grupos en los cuales fue más probable detectar eventos de recombinación ya que pertenecen al mismo cluster epidemiológico. Dando pauta a estudiar la recombinación intrasubtipo en un cluster epidemiológico específico.

La estrategia usada para detectar recombinación se basó en utilizar conjuntos de secuencias con relación epidemiológica, lo que suponemos aumentará la probabilidad de que estas secuencias hayan sufrido eventos de recombinación entre sí. Usando RARP se detectó se identificaron 417 secuencias recombinantes con un total de 987 puntos de recombinación estadísticamente significativos, lo cual, corresponde a una frecuencia de recombinación de 0.20810594 por sitio por generación lo que corresponde a un rango de 1 a 49 eventos de recombinación por secuencia que se considera como alta. Smyth y cols., que en promedio obtuvieron entre 19-20 eventos de recombinación por genoma para gag y pol [85]. Estudios previos indican que la tasa de recombinación intrasubtipo en estudios in vivo depende de la población de estudio, la frecuencia de uso de anti-retrovirales, el tiempo de seguimiento de los pacientes y los métodos usados para la detección [75]. Otros reportes de recombinación intersubtipo habían indicado 2.4 cruces por genoma en un estudio con 16 secuencias en el 2002 [107] además Rhodes en el 2003 indicó que la recombinación se produce a una tasa notable de aproximadamente cada 2 kb o posiblemente incluso con mayor frecuencia en algunas condiciones fisiológicas [46]. Letvin y colaboradores indicaron que en cada replicación viral ocurren de 7 a 30 eventos de recombinación por genoma y ciclo de replicación (RT)[47]. Otras estimaciones de las tasas de recombinación *in vivo* varían entre los diferentes estudios (1.4×10^{-5} a 2×10^{-4} puntos de corte por sitio por generación) [32].

Estimaciones *in vitro* han demostrado una tasa de recombinación de 3×10^{-4} por sitio y generación [48]. Y específicamente para un estudio de recombinación intrasubtipo

identificaron 5.6 eventos de recombinación por genoma [49]. Zhuang y col estimaron que la recombinación intersubtipo ocurre aproximadamente 2.8 veces por genoma por ciclo de replicación viral (104). Tomando en cuenta que esto varía de acuerdo al estudio y región tenemos en cuenta también que otros genes estructurales presentan tasas variables como en el estudio de Smyth y cols., que en promedio obtuvieron una tasa de $2.0 \cdot 10^{-3}$ eventos de recombinación por nucleótido por ronda de infección o REPN (por sus siglas en inglés: recombination events per nucleotide per round of infection) lo cual corresponde aproximadamente entre 19-20 eventos de recombinación por genoma para gag y pol.[85].

Adicionalmente como lo reportado previamente en este análisis y confirmando la hipótesis, se localizaron dos hotspots entre los segmentos de 151-200 y 401- 450 nucleótidos de Env, justo el segundo segmento identificado como hotspot coincide con lo que Salemi y cols, [86] encontraron su estudio de recombinación intrasubtipo para env teniendo dos breakpoints en las posiciones 300 y 430, coincidiendo en el segundo sitio con uno de nuestros hotspots. Adicionalmente Wei y cols [87], identificaron 5 breakpoints por medio de patrones de recombinación y cuando corroboraron esto por medio de RARP identificaron 7 coincidiendo con nosotros con 2 de sus sitios en nuestros dos hotpost las otras 5 se distribuyen en otras regiones de env.

En el primer hot spot identificado correspondiente a la ventana de 151 a 200 nucleótidos se localizan las regiones hipervariables de VI y V2 que son conformaciones helicoidales las cuales cumplen un papel importante al exponer el sitio de unión de $\alpha 4\beta 7$ [88] la cual es una integrina que se sobre expresa en linfocitos T CD4+ activados localizados en mucosa, también se expresa en altos niveles de CCR5 por tanto su presencia representan una población ideal para una reproducción infectiva, durante la fase aguda de la infección facilita la migración de linfocitos al intestino. Gp120 se une a $\alpha 4\beta 7$ mediante un tripéptido conservado en el loop V2 (en las posiciones 179 y 181[87]) recientemente en las posiciones 170-173 [89], el cual es estructuralmente muy parecido a al ligando natural de esta integrina (MadCAM, VCAM) pero no es necesaria para la replicación [85, 90].

En el segundo hot spot correspondiente a 401-450 se identifica parte de V4 y C4. Se sabe que las regiones V4 y V5 tienen una función no muy específica en cuanto a neutralización,

pero estas afectan la conformación del empaquetamiento del escudo de glicanos, limitando estéricamente la disponibilidad de determinantes de neutralización [91], ya que Wei y cols. han indicado que las afectaciones a V4, pueden inhibir estrictamente el acceso principal de epítopes neutralizantes en la superficie del virus [91]. C4 se ha identificado como sitio implicado en el escudo de glicanos junto con V5 [92]. Y es un tramo anfipático de aminoácidos que, según los análisis mutacionales, constituye un componente principal de la región de unión a CD4 en gp120 [93].

Los cold-spots localizados fueron dos, de 551 a 600 y de 851 a 856 aminoácidos en la sección de gp41 mostrando claramente menor número de breakpoints en sus dominios bien definidos. En el primer cold spot tenemos parte de las regiones de héptadas repetidas (HR) terminales N y C que finalmente forman un haz de seis hélices que impulsa la aposición y fusión de la membrana, lo que finalmente conduce a la formación de poros de fusión, estas han sido objetivos para el desarrollo de numerosos inhibidores de fusión como T20 (Enfuvirtide), además en el segundo cold spot se localiza curiosamente, la cola citoplasmática de gp41 que no solo regula la incorporación de env sino que también regula la fusión. Se sabe que la región de la cola de gp41 contiene 3 regiones LLP (péptidos líticos lentivirales) que interactúan con la membrana viral y regulan la fusión, como es evidente en los mutantes de truncamiento [94].

Con esta información podemos inferir de forma más clara que la recombinación no se da de forma aleatoria sino en zonas específicas dentro de las estructuras de env.

Los análisis de variabilidad del gen Env entre las secuencias recombinantes y no recombinantes mostraron diferencias entre las posiciones a lo largo de la secuencias, aunque el número de secuencias a sido menor en el grupo de recombinantes la variación presente en las secuencias elevada, y esta se puede identificar mayoritariamente en las posiciones de 400 a 420 a.a que incluye estructuras como C4 y V5 esta zona coincide con la zona identificada como hotspot de 400 a 450 a.a., además para el mismo grupo se detectan posiciones de mayor variabilidad como en 253-258 que está al final de V3 y al inicio de C3, ambos grupos mostraron varias posiciones con una frecuencia de polimorfismos alta para los dos grupos estudiados.

Por el contrario, la zona correspondiente a C1, C5, FP y HR resultaron ser las más conservadas al presentar una menor frecuencia de polimorfismos. Nuestros resultados muestran concordancia en cuanto al grupo de secuencias NO recombinantes con lo detectado previamente por YAO y colaboradores con posiciones en las regiones V1 y V4 en estudios entre subtipos diferentes como el A, C B y D pero estas regiones presentaron mayor variabilidad [95].

Con respecto a las recombinantes que, si presentan posiciones polimórficas, pero no con el mismo grado que las no recombinantes de hecho para este punto se asocia a lo ya indicado por Schneider y colaboradores que mostraron que las diversidades virales siempre alcanzan un punto de equilibrio que es característico de la interacción virus-huésped. Si la frecuencia de mutación es un reflejo de la tasa de mutación, es posible que los virus con tasas de mutación más altas no toleren tasas de recombinación altas y que estén demasiado cerca de la extinción, y las frecuencias de mutación más bajas pueden reflejar tasas de recombinación más altas que resuelven las mutaciones [96].

Continuando con la comparación entre las secuencias recombinantes y no recombinantes, se observó que la variabilidad fue alta en secuencias recombinantes, considerando que el número de secuencias con las que se realizó el análisis fue mucho menor (416), en contraste con el grupo de secuencias NO recombinantes (1809).

La conservación de C1, C5, FP y HR se podría explicarse a partir de su importancia estructural y funcional, ya que las variaciones en estas estructuras pueden afectar la capacidad de replicación del virus. Aparentemente, las estructuras de gp120 que corresponden a V1-V2 y V4-C4 facilitan la flexibilidad en cuanto a la presencia de polimorfismos, permitiendo conservar sus funciones básicas.

Con respecto a la comparación entre los dos grupos de secuencias y determinar de mejor forma la variabilidad entre estos, se notó que hubo diferencia en los valores de entropía entre las posiciones de las secuencias recombinantes y las No recombinantes en V4-C4 correspondiente al segundo hotspot confirmando mayor variabilidad en las Recombinantes.

Aunado a esto, se identificó la presencia de un número considerable de posiciones bajo presión selectiva positiva a lo largo de las secuencias recombinantes de env las cuales se esperaba que las zonas de mayor número coincidieran con los hotspots encontrados

anteriormente ya que las mutaciones no sinónimas están asociadas al escape del sistema inmune como se ha indicado por Neher *et. al.* en el 2010 dónde la recombinación puede facilitar la rápida adquisición de múltiples mutaciones de resistencia debido a los cambios no sinónimos [48], lo cual no fue el caso ya que las herramientas utilizadas nos indicaron un mayor número de mutaciones no sinónimas en zonas aledañas a los hotspots que también presentaron un número considerable de posiciones bajo presión selectiva pero no el número más alto como se pudo haber imaginado.

Aunque se esperaba que los epítomos localizados en ambos grupos fueran identificados en o cerca de los hotspots, pero no fue así, de los 80 solo uno de cada uno de mayor frecuencia para ambos grupos llegan a coincidir con nuestros hotspots dejando a la mayoría en la zona de la cola citoplasmática de hecho interesantemente se encontraron en diferentes frecuencias entre grupos, los 10 más abundantes de las recombinantes son mucho más frecuentes que en las no recombinantes en contraste la frecuencia de los 10 más abundantes de las no recombinantes son similares entre las recombinantes. Las secuencias recombinantes tienen mayor abundancia en sitios de reconocimiento para epítomos CTL/CD8. El hecho de que no se hayan localizado coincidencias en estas zonas se puede deber a que al ser recombinantes se afirma que la respuesta inmunitaria impone una presión de selección diversificadora sobre la población de virus, lo que da como resultado un escape secuencial en cada epítomo. Una variante del virus no reconocida por el sistema inmunitario tiene una ventaja de aptitud y comienza a hacerse cargo de la población del virus, provocando finalmente una respuesta inmunitaria y, por lo tanto, la selección de nuevas variantes de escape [97].

Con respecto a los sitios de glicosilación se identificó mayor número de sitios a través de toda la longitud de env en las secuencias no recombinantes con 92 posiciones en contraste con los 88 sitios de las recombinantes. Y de acuerdo al porcentaje de sitios de glicosilados por posición por secuencias no hubo gran variación, muy pocos presentaron diferencia entre los grupos recombinantes y no recombinantes. Tomemos en cuenta que a mayor número de sitios de glicosilación se infiere que hay un menor reconocimiento por el sistema inmune, a menor número de sitios de glicosilación hay un mejor reconocimiento. El hecho de que no se hayan encontrado diferencias en los PNGS de los dos grupos de secuencias recombinantes y no recombinantes puede deberse a que ya en otros estudios como en el de Cao y cols [98] se ha visto que en la superficie del ectodominio Env hay

~95 sitios de glicosilación en promedio, y el análisis de espectrometría de masas (MS) ha revelado que existe heterogeneidad composicional entre sitios, así como en el mismo sitio entre diferentes copias de Env.

Adicionalmente el uso del correceptor para todas las secuencias fue CCR5 (R5) sin distinción para ninguno de los dos grupos, indicando que la recombinación puede no afectar la acumulación o pérdida de motivos considerables en estas secuencias analizadas.

La distribución de PNGS fue casi idéntica entre todas las secuencias analizadas de ambos grupos recombinantes y no recombinantes, es posible que estos patrones en la distribución de PNGS pueda constituir una ventaja durante la entrada del virus como protección como otros casos han notado en la falta de variación quizá por otros motivos aún no identificados [99].

Los resultados de todos los análisis realizados en este proyecto se agruparon en un mapa del gen env que resume las secciones recombinantes, así como la recombinación y su asociación con presión selectiva positiva y la distribución de epítopes y sitios de mayor glicosilación para ambos grupos recombinante y no recombinantes.

Este estudio muestra características únicas de la recombinación intrasubtipo en el gen env del VIH-1 subtipo B, de dos países con muestras de más de 400 secuencias, lo cual puede guiar estudios para comprender de una mejor manera la recombinación intrasubtipo del VIH -1 subtipo B a un nivel regional y poblacional.

CONCLUSIONES.

En este estudio se recabaron y utilizaron 6797 secuencias con las cuales se trabajó para realizar un análisis de recombinación de 3184 secuencias de *env* de dos países que formaron una cohorte representativa de los virus que circulan en Estados Unidos y Alemania.

Los breakpoints que se identificaron demostraron que no hay aleatoriedad a lo largo del gen env, y sobre todo esto se vio representado en las zonas de V1-V2 y V4-C4, coincidiendo con polimorfismos que marcan la variabilidad, que en su mayoría esto podría explicarse por la presión selectiva provocada por el sistema inmune. Se observó que la estructuras

como FP, HR, CT tiende a preservarse debido a la importancia en el correcto funcionamiento de la fusión de membranas e inserción.

Los resultados demuestran por vez primera un panorama de la influencia de la recombinación intrasubtipo en la evolución del VIH-1 subtipo B, enfocándose en la importancia de esta y las interacciones y variación a través del gen de *env*.

Con estos resultados se podría revelar la existencia de nuevas asociaciones en el control y regulación del sistema inmune con respecto al VIH-1 subtipo B, que puedan ser relevantes en la aportación de conocimiento para el acervo del diseño de vacunas que puedan ser basadas en CTLs.

El progresar con estos estudios e incluir a otros países y diferentes poblaciones permitirá tener una perspectiva más amplia de la recombinación intrasubtipo y la evolución del VIH-1 subtipo B lo cual ayuda a identificar blancos específicos no definidos para otras poblaciones que pudieran revelar información importante en el tratamiento y prevención de la infección así como servir como blanco para estrategias antirretrovirales.

En nuestros análisis, las secuencias recombinantes no contienen diferencias significativas en las características, como el tropismo del correceptor, la cantidad de sitios PNG u otros motivos que se han descrito previamente en la literatura como potencialmente ventajosos en la transmisión o infección temprana.

Quizá desde otra perspectiva el escape de epítomos inmunes de anticuerpos y linfocitos T citotóxicos puede impulsar la selección temprana de recombinantes como en el estudio de Keri y cols [99].

Como parte del seguimiento de esta investigación, el continuar con estudios sobre el impacto de recombinación intrasubtipo en poblaciones de diferentes tamaños y con muestras de los mismos individuos y en tiempos consecutivos para todos, permitiría poder realizar comparaciones más claras y realizar análisis de fitness viral tomando en cuenta la capacidad replicativa versus la replicación viral en estudios *in vitro*. Esto nos daría mucha más información, pero sería muy específica de los pacientes evaluados en una población y del subtipo de VIH específico en esa región, que nos llevaría a una medicina especializada. Este tipo de proyectos también nos permiten revelar el valor que tiene el impacto de la adaptación viral en los individuos que se asocia con la progresión de la enfermedad, continuar los estudios en ello, nos permite ahondar en aspectos de la

evolución del virus por individuos a nivel regional y poblacional, contribuyendo con información, para poder generar estrategias para el control y prevención en un futuro cercano de la infección por VIH.

Bibliografía

1. Cordoba Villalobos, J. A.; De León Rosales, S.P., Valdespino, J. L.(Eds.) (2009). 25 años de SIDA en México. Logros, desaciertos y retos. (2ª Ed.). México: Instituto Nacional de Salud Pública. 135-191. ISBN
2. Hoffman–Rockstroch-Kamps. 2006. VIH Medicine by Flying Publisher – Paris, Cagliari, Wuppertal
3. Hull, Mark Wa; Lima, Viviane Da,b; Hogg, Robert Sa,c; Harrigan, P Richarda; Montaner, Julio SGa,b. Epidemiology of treatment failure: a focus on recent trends. Current Opinion in VIH and AIDS: November 2009 - Volume 4 - Issue 6 - p 467-473 doi: 10.1097/COH.0b013e328331d353
4. UNAIDS, 2014
5. Fields, Bernard N. et al. Fields Virology. Fifth edition. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2007. Print.
6. Carlson J.M.; Le A.Q.; Shahid A.; Brumme Z.L. (2014).VIH adptation to HLA: a window into virus-host immune interactions. Trends in Micribiology XX(2015) 1-13
7. Van Heuverswyn F, Peeters M. The origins of VIH and implications for the global epidemic. Curr Infect Dis Rep. 2007 Jul;9(4):338-46. doi: 10.1007/s11908-007-0052-x. PMID: 17618555.
8. Rowland-Jones SL. Timeline: AIDS pathogenesis: what have two decades of VIH research taught us? Nat Rev Immunol. 2003 Apr;3(4):343-8. doi: 10.1038/nri1058. PMID: 12669024.
9. ONUSIDA 2021
10. VIH.gov June 02, 2022
11. Rambaut, A., Robertson, D., Pybus, O. et al. Phylogeny and the origin of VIH-1. Nature 410, 1047–1048 (2001). <https://doi.org/10.1038/35074179>

12. Sharp PM, Hahn BH. Origins of VIH and the AIDS pandemic. *Cold Spring Harb Perspect Med.* 2011 Sep;1(1):a006841. doi: 10.1101/cshperspect.a006841. PMID: 22229120; PMCID: PMC3234451.
13. Hemelaar, J. 2011. The origin and diversity of the VIH-1 pandemic. *Trends in Molecular Medicine*, March 2012, Vol. 18, No. 3
14. Freed, E.O.; Martin, M.A. (2007). VIHs and their replication. Philadelphia: Knipe DM, Vol. II. (5a Ed.) 2108-2183. (Era el 21)
15. Freed, E.O. (1998). VIH-1 Gag proteins: Diverse functions in the virus life cycle. *Virology*, 251: 1-15
16. Nicholas H. Acheson. 2011. *Fundamentals of molecular virology*. 2nd ed. United States. ISBN 978-0-470-90059-8
17. Barrie KA, Perez EE, Lamers SL, Farmerie WG, Dunn BM, Sleasman JW, Goodenow MM. Natural variation in VIH-1 protease, Gag p7 and p6, and protease cleavage sites within gag/pol polyproteins: amino acid substitutions in the absence of protease inhibitors in mothers and children infected by human immunodeficiency virus type 1. *Virology*. 1996 May 15;219(2):407-16. doi: 10.1006/viro.1996.0266. PMID: 8638406.
18. Tebit, D. M., & Arts, E. J. 2011. Tracking a century of global expansion and evolution of VIH to drive understanding and to combat disease. *The Lancet Infectious Diseases*, 11(1): 45-56. doi:10.1016/s1473-3099(10)70186-9.
19. Checkley et al. 2012. VIH-1 Envelope Glycoprotein Biosynthesis, Trafficking, and Incorporation. *J Mol Biol.* 2011 Jul 22; 410(4): 582-608. doi: 10.1016/j.jmb.2011.04.042
20. Heinrich G. Gottlinger. 2011. VIH-1 Gag: a Molecular Machine Driving Viral Particle Assembly and Release.
21. Carlson, J. M., Listgarten, J., Pfeifer, N., Tan, V., Kadie, C., Walker, B. D., ... Heckerman, D. (2012). Widespread Impact of HLA Restriction on Immune Control and Escape Pathways of VIH-1. *Journal of Virology*, 86(9), 5230-5243. <http://doi.org/10.1128/JVI.06728-11>
22. Brumme ZL, Brumme CJ, Heckerman D, Korber BT, Daniels M, et al. (2007) Evidence of Differential HLA Class I-Mediated Viral Evolution in Functional and Accessory/Regulatory Genes of VIH-1. *PLoS Pathog* 3(7): e94. doi: 10.1371/journal.ppat.0030094

23. Alter, G., & Altfeld, M. (2009). NK cells in VIH-1 infection: Evidence for their role in the control of VIH-1 infection. *Journal of Internal Medicine*, 265(1), 29–42. <http://doi.org/10.1111/j.1365-2796.2008.02045.x>
24. IDS Res Hum Retroviruses. 1995 Nov;11(11):1413-6. Automated subtyping of VIH-1 genetic sequences for clinical and surveillance purposes: Performance evaluation of the new REGA version 3 and seven other tools. Peña ACP, Faria NR, Imbrechts S, Libin P, Abecasis AB, Deforche K, Gomez A, Camacho RJ, de Oliveira T, Vandamme A-M. *Infectious Genetics and Evolution* 2013; 19:337-48. doi: 10.1016/j.meegid.2013.04.03
25. Kindt, T.J. Goldsby, R. A. Osborne B. A. 2007 INMUNOLOGÍA de Kuby. 6ta ed. McGRAW-HILL INTERAMERICANA EDITORES, S.A. de C.V.
26. Goulder, P. J. R., & Watkins, D. I. (2008). Impact of MHC class I diversity on immune control of immunodeficiency virus replication. *Nature Reviews. Immunology*, 8(8), 619–630. <http://doi.org/10.1038/nri2357>
27. McLaren PJ, Coulonges C, Ripke S, van den Berg L, Buchbinder S, et al. (2013) Association Study of Common Genetic Variants and VIH-1 Acquisition in 6,300 Infected Cases and 7,200 Controls. *PLoS Pathog* 9(7): e1003515. doi: 10.1371/journal.ppat.1003515
28. Goulder, P. J. R., & Walker, B. D. (2012). VIH and HLA Class I: an evolving relationship. *Immunity*, 37(3), 426–440. <http://doi.org/10.1016/j.immuni.2012.09.005>
29. Freed O. 2002. Viral Late Domains. *JOURNAL OF VIROLOGY*, May 2002, p. 4679–4687, 0022-538X/02/\$04.000 DOI: 10.1128/JVI.76.10.4679–4687.2002
30. Tavares Lucas A., Januário Yunan C., daSilva Luis L. P. 2021. VIH-1 Hijacking of Host ATPases and GTPases That Control Protein Trafficking. *Frontiers in Cell and Developmental Biology*. 10.3389/fcell.2021.622610
31. Coffin, J. M. 1986. Genetic variation in AIDS viruses. *Cell* 46:1–4.
32. Vincent Racaniello. 2009. Increased fidelity reduces viral fitness. *Virology blog*, about viruses and viral diseases. <https://bit.ly/3bilHam>
33. Song et. al. 2018. Tracking VIH-1 recombination to resolve its contribution to VIH-1 evolution in natural infection. *Nature Communications*. DOI: 10.1038/s41467-018-04217-5

34. Leitner, T., and J. Albert. 1999. The molecular clock of VIH-1 unveiled through analysis of a known transmission history. *Proc. Natl. Acad. Sci. USA* 96:10752–10757
35. Malim, M. H., and M. Emerman. 2001. VIH-1 sequence variation: drift, shift, and attenuation. *Cell* 104:469–472., 338
36. Gunthard, H. F., A. J. Leigh-Brown, R. T. D'Aquila, V. A. Johnson, D. R. Kuritzkes, D. D. Richman, and J. K. Wong. 1999. Higher selection pressure from antiretroviral drugs in vivo results in increased evolutionary distance in VIH-1. *J. Virol.* 259:154–165.,
37. Mild, M., J. Esbjornsson, E. M. Fenyo, and P. Medstrand. 2007. Frequent intrapatient recombination between human immunodeficiency virus type 1 R5 and X4 envelopes: implications for coreceptor switch. *J. Virol.* 81:3369–3376
38. Kuwata, T., Y. Miyazaki, T. Igarashi, J. Takehisa, and M. Hayami. 1997. The rapid spread of recombinants during a natural in vitro infection with two human immunodeficiency virus type 1 strains. *J. Virol.* 71:7088–7091
39. Fultz, P., L. You, Q. Wei, and M. Girard. 1997. Human immunodeficiency virus type 1 intersubtype (B/E) recombination in a superinfected chimpanzee. *J. Virol.* 71:7990–7995., 347
- Wooley, D. P., R. A. Smith, S. Czajak, and R. C. Desrosiers. 1997. Direct demonstration of retroviral recombination in a rhesus monkey. *J. Virol.* 71:9650–9653.
40. Zhuang, J., S. Mukherjee, Y. Ron, and J. P. Dougherty. 2006. High rate of genetic recombination in murine leukemia virus: implications for influencing proviral ploidy. *J. Virol.* 80:6706–6711.
41. Coffin, J. M. 1979. Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. *J. Gen. Virol.* 42:1–26., 320
- Temin, H. M. 1993. Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. *Proc. Natl. Acad. Sci. USA* 90:6900–6903.
42. Adewunmi Onafuwa-Nuga and Alice Telesnitsky. 2009. The Remarkable Frequency of Human Immunodeficiency Virus Type 1 Genetic Recombination. *MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS*, Sept. 2009, p. 451–480 Vol. 73, No. 3. doi:10.1128/MMBR.00012-09
43. Gilboa, E., S. W. Mitra, S. Goff, and D. Baltimore. 1979. A detailed model of reverse transcription and tests of crucial aspects. *Cell* 18:93–100., A11
- Telesnitsky, A., and S. P. Goff. 1993. Strong-stop strand transfer during reverse transcription, p. 49–83.

- In A. M. Skalka and S. P. Goff (ed.), Reverse transcriptase. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 318 A11
44. An, W., and A. Telesnitsky. 2002. VIH-1 genetic recombination: experimental approaches and observations. *AIDS Rev.* 4:195–212.
 45. Coffin, J. 1996. Retroviridae: the viruses and their replication. In B. N. Fields, D. M. Knipe, P. M. Howley, et al. (ed.), *Fundamental virology*, 3rd ed. Lippincott, Philadelphia, PA
 46. Rhodes, T., H. Wargo, and W. S. Hu. 2003. High rates of human immunodeficiency virus type 1 recombination: near-random segregation of markers one kilobase apart in one round of viral replication. *J. Virol.* 77:11193– 11200
 47. Letvin et al 2006
 48. Neher RA, Leitner T (2010) Recombination Rate and Selection Strength in VIH Intra-patient Evolution. *PLoS Comput Biol* 6(1): e1000660. doi:10.1371/journal.pcbi.1000660
 49. Kiwelu IE, Novitsky V, Margolin L, Baca J, Manongi R, et al. (2013) Frequent Intra-Subtype Recombination among VIH-1 Circulating in Tanzania. *PLoS ONE* 8(8): e71131. doi:10.1371/journal.pone.0071131
 50. Jetzt, A. E., H. Yu, G. J. Klarmann, Y. Ron, B. D. Preston, and J. P. Dougherty. 2000. High rate of recombination throughout the human immunodeficiency virus type 1 genome. *J. Virol.* 74:1234–1240.
 51. Negroni, M., and H. Buc. 2001. Mechanisms of retroviral recombination. *Annu. Rev. Genet.* 35:275–302., 228 Negroni, M., M. Ricchetti, P. Nouvel, and H. Buc. 1995. Homologous recombination promoted by reverse transcriptase during copying of two distinct RNA templates. *Proc. Natl. Acad. Sci. USA* 92:6971–6975.
 52. Iwatani, Y., D. S. Chan, F. Wang, K. S. Maynard, W. Sugiura, A. M. Gronenborn, I. Rouzina, M. C. Williams, K. Musier-Forsyth, and J. G. Levin. 2007. Deaminase-independent inhibition of VIH-1 reverse transcription by APOBEC3G. *Nucleic Acids Res.* 35:7096–7108.
 53. Goff, S. P. 2007. Host factors exploited by retroviruses. *Nat. Rev. Microbiol.* 5:253–263., 153 Jones, K. L., S. Sonza, and J. Mak. 2008. Primary T-lymphocytes rescue the replication of VIH-1 DIS RNA mutants in part by facilitating reverse transcription. *Nucleic Acids Res.* 36:1578–1588.

54. Liu, S., E. A. Abbondanzieri, J. W. Rausch, S. F. Le Grice, and X. Zhuang. 2008. Slide into action: dynamic shuttling of VIH reverse transcriptase on nucleic acid substrates. *Science* 322:1092–1097.
193. Liu, Y., M. E. Curlin, K. Diem, H. Zhao, A.
55. Junghans, R. P., L. R. Boone, and A. M. Skalka. 1982. Products of reverse transcription in avian retrovirus analyzed by electron microscopy. *J. Virol.* 43:544–554.
158. Junghans, R. P., L. R. Boone, and A. M. Skalka. 1982. Retroviral DNA H structures: displacement-assimilation model of recombination. *Cell* 30:53–62.
56. Xu, H., and J. D. Boeke. 1987. High-frequency deletion between homologous sequences during retrotransposition of Ty elements in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 84:8553–8557
57. Roda, R. H., M. Balakrishnan, J. K. Kim, B. P. Roques, P. J. Fay, and R. A. Bambara. 2002. Strand transfer occurs in retroviruses by a pause-initiated two-step mechanism. *J. Biol. Chem.* 277:46900–46911
58. Hwang, C. K., E. S. Svarovskaia, and V. K. Pathak. 2001. Dynamic copy choice: steady state between murine leukemia virus polymerase and polymerase- dependent RNase H activity determines frequency of in vivo template switching. *Proc. Natl. Acad. Sci. USA* 98:12209–12214
59. Basu, V. P., M. Song, L. Gao, S. T. Rigby, M. N. Hanson, and R. A. Bambara. 2008. Strand transfer events during VIH-1 reverse transcription. *Virus Res.* 134:19–38., 251
60. Rawson et. al. 2018. NAR Breakthrough Article Recombination is required for efficient VIH-1 replication and the maintenance of viral genome integrity. *Nucleic Acids Research*. Vol. 46, No. 20 10535–10545. doi: 10.1093/nar/gky910
61. Rambaut 2004
62. Michael Worobey and Edward C. Holmes. 1999. Evolutionary aspects of recombination in RNA viruses. *Journal of General Virology* Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK]
63. Dykes, C., M. Balakrishnan, V. Planelles, Y. Zhu, R. A. Bambara, and L. M. Demeter. 2004. Identification of a preferred region for recombination and mutation in VIH-1 gag. *Virology* 326:262–279., 148, A29
- Levy, D. N., G. M. Aldrovandi, O. Kutsch, and G. M. Shaw. 2004. Dynamics of VIH-1 recombination in its natural target cells. *Proc. Natl. Acad. Sci. USA* 101:4204–4209.,
64. 189

65. Takehisa, J., L. Zekeng, E. Ido, Y. Yamaguchi-Kabata, I. Mboudjeka, Y. Harada, T. Miura, L. Kaptue, and M. Hayami. 1999. Human immunodeficiency virus type 1 intergroup (M/O) recombination in Cameroon. *J. Virol.* 73:6810–6820.
66. Fan, J., M. Negroni, and D. L. Robertson. 2007. The distribution of VIH-1 recombination breakpoints. *Infect. Genet. Evol.* 7:717–723.
67. Quinones-Mateu, M. E., Y. Gao, S. C. Ball, A. J. Marozsan, A. Abraha, and E. J. Arts. 2002. In vitro intersubtype recombinants of human immunodeficiency virus type 1: comparison to recent and circulating in vivo recombinant forms. *J. Virol.* 76:9600–9613.
68. Archer, J., J. W. Pinney, J. Fan, E. Simon-Loriere, E. J. Arts, M. Negroni, and D. L. Robertson. 2008. Identifying the important VIH-1 recombination breakpoints. *PLoS Comput. Biol.* 4:e1000178.
69. Laird, M. E., T. Igarashi, M. A. Martin, and R. C. Desrosiers. 2008. Importance of the V1/V2 loop region of simian-human immunodeficiency virus envelope glycoprotein gp120 in determining the strain specificity of the neutralizing antibody response. *J. Virol.* 82:11054–11065.
70. Allen, T. M., M. Altfeld, S. C. Geer, E. T. Kalife, C. Moore, K. M. O'Sullivan, I. Desouza, M. E. Feeney, R. L. Eldridge, E. L. Maier, D. E. Kaufmann, M. P. Lahaie, L. Reyor, G. Tanzi, M. N. Johnston, C. Brander, R. Draenert, J. K. Rockstroh, H. Jessen, E. S. Rosenberg, S. A. Mallal, and B. D. Walker. 2005. Selective escape from CD8₊ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (VIH-1) sequence diversity and reveals constraints on VIH-1 evolution. *J. Virol.* 79:13239–13249.
71. Ramirez, B. C., E. Simon-Loriere, R. Galetto, and M. Negroni. 2008. Implications of recombination for VIH diversity. *Virus Res.* 134:64–73., 323Thomson, M. M., and R. Najera
72. Charpentier, C., T. Nora, O. Tenaillon, F. Clavel, and A. J. Hance. 2006. Extensive recombination among human immunodeficiency virus type 1 quasispecies makes an important contribution to viral diversity in individual patients. *J. Virol.* 80:2472–2482.
73. Nikolaitchik, O., Keele, B., Gorelick, R., Alvord, W. G., Mazurov, D., Pathak, V. K., & Hu, W.- S. (2015). High recombination potential of subtype A VIH-1. *Virology*, 484(1): 334-340.

74. Roussie CM. et al. 2007. Extensive Intrasubtype Recombination in South African Human Immunodeficiency Virus Type 1 Subtype C Infection. *JOURNAL OF VIROLOGY*, May 2007, p. 4492–4500 Vol. 81, No. 9 0022-538X/07/\$08.000 doi:10.1128/JVI.02050-06
75. Taylor J. E., and Korber B. T. 2004. HIV-1 intra-subtype superinfection rates: estimates using a structured coalescent with recombination. Published by Elsevier B.V. doi:10.1016/j.meegid.2004.07.001
76. Redd A. D., et al. 2013. Frequency and implications of HIV superinfection. *Lancet Infect Dis* 2013; 13: 622–28. Doi.org/10.1016/S1473-3099(13)70066-5
77. AD Redd, CE Mullis, D Serwadda, et al. The rates of HIV superinfection and primary HIV incidence in a general population in Rakai, Uganda *J Infect Dis*, 206 (2012), pp. 267-274
78. Smyth R. P. 2013. Identifying Recombination Hot Spots in the HIV-1 Genome. *Journal of Virology* p. 2891–2902.
79. Datos sin publicar
80. Hu, Wei-Shau. Temin, Howard M. 1990. Retroviral Recombination and Reverse Transcription. *Science*. P 1227-1233. V 250. 4985. doi:10.1126/science.1700865
81. Holsinger, Kent E.; Bruce S. Weir (2009). «Genetics in geographically structured populations: defining, estimating and interpreting F_{ST} ». *Nat Rev Genet* 10 (9): 639-650. ISSN 1471-0056. PMC 4687486. PMID 19687804. doi:10.1038/nrg2611
82. Richard Durrett (2008). *Probability Models for DNA Sequence Evolution*. Springer. ISBN 978-0-387-78168-6. Consultado el 25 de octubre de 2012.
83. lanl.gov
84. Sergei L Kosakovsky Pond, Steven Weaver, Andrew J Leigh Brown, Joel O Wertheim, HIV-TRACE (TRANsmision Cluster Engine): a Tool for Large Scale Molecular Epidemiology of HIV-1 and Other Rapidly Evolving Pathogens, *Molecular Biology and Evolution*, Volume 35, Issue 7, July 2018, Pages 1812–1819, <https://doi.org/10.1093/molbev/msy016>
85. Redmond P. Smyth, Timothy E. Schlub, Andrew J. Grimm, Caryll Waugh, Paula Ellenberg, Abha Chopra, Simon Mallal, Deborah Cromer, Johnson Mak and Miles P. Davenport J. Virol. 2014. Identifying Recombination Hot Spots in the HIV-1 Genome. *Journal of Virology*. 88(5):2891. DOI: 10.1128/JVI.03014-13.

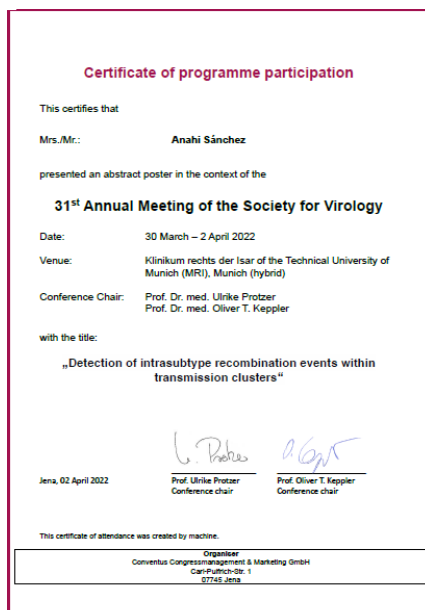
86. Marco Salemi, Rebecca R. Gray, Maureen M. Goodenow. 2008. An exploratory algorithm to identify intra-host recombinant viral sequences. *Molecular Phylogenetics and Evolution* 49 (2008) 618–628. Elsevier Inc. doi:10.1016/j.ympev.2008.08.017
87. Huamian Wei, Danwei Yu, Xiuzhu Geng and Yuxian He. 2020. Defective VIH-1 envelope gene promotes the evolution of the infectious strain through recombination in vitro. *BMC Infectious Diseases*. <https://doi.org/10.1186/s12879-020-05288-w>
88. Wibmer, C.K., Richardson, S.I., Yoltz, J. et al. Common helical V1V2 conformations of VIH-1 Envelope expose the $\alpha 4\beta 7$ binding site on intact virions. *Nat Commun* 9, 4489 (2018). <https://doi.org/10.1038/s41467-018-06794-x>
89. Tassaneetrithep, B. et al. Cryptic determinant of $\alpha 4\beta 7$ binding in the V2 loop of VIH-1 gp120. *PLoS ONE* 9, e108446 (2014)
90. Cicala, C., Arthos, J. & Fauci, A.S. VIH-1 envelope, integrins and co-receptor use in mucosal transmission of VIH. *J Transl Med* 9 (Suppl 1), S2 (2011). <https://doi.org/10.1186/1479-5876-9-S1-S2>
91. Timofeeva, Anna, Sergey Sedykh, and Georgy Nevinsky. 2022. "Post-Immune Antibodies in VIH-1 Infection in the Context of Vaccine Development: A Variety of Biological Functions and Catalytic Activities" *Vaccines* 10, no. 3: 384. <https://doi.org/10.3390/vaccines10030384>
92. Wei, X.; Decker, J.M.; Wang, S.; Hui, H.; Kappes, J.C.; Wu, X.; Salazar-Gonzalez, J.F.; Salazar, M.G.; Kilby, J.M.; Saag, M.S.; et al. Antibody neutralization and escape by VIH-1. *Nature* 2003, 422, 307–312.
93. Teeraputon, S.; Louisirojchanakul, S.; Auewarakul, P. N-Linked Glycosylation in C2 Region of VIH-1 Envelope Reduces Sensitivity to Neutralizing Antibodies. *Viral Immunol.* 2005, 18, 343–353
94. Garg H, Blumenthal R. Role of VIH Gp41 mediated fusion/hemifusion in bystander apoptosis. *Cell Mol Life Sci.* 2008 Oct;65(20):3134-44. doi: 10.1007/s00018-008-8147-6. PMID: 18500445; PMCID: PMC2574860.
95. Neng YAO1, Chunxia ZHANG1, Qi LIU, Jun LIU, Chiyu ZHANG. 2015. Polymorphism characteristics of VIH-1 gp120 and 5 hypervariable regions. *Turk J Med Sci.* 45: 47-54. doi:10.3906/sag-1307-132

96. Schneider WL, Roossinck MJ. 2000. Evolutionarily related Sindbis-like plant viruses maintain different levels of population diversity in a common host. *J Virol* 74:3130–3134. doi: 10.1128/JVI.74.7.3130-3134.2000.
97. Immonen TT, Conway JM, Romero-Severson EO, Perelson AS, Leitner T (2015) Recombination Enhances VIH-1 Envelope Diversity by Facilitating the Survival of Latent Genomic Fragments in the Plasma Virus Population. *PLOS Computational Biology* 11(12): e1004625. <https://doi.org/10.1371/journal.pcbi.1004625>
98. Cao L, Pauthner M, Andrabi R, Rantalainen K, Berndsen Z, Diedrich JK, Menis S, Sok D, Bastidas R, Park SR, Delahunty CM, He L, Guenaga J, Wyatt RT, Schief WR, Ward AB, Yates JR 3rd, Burton DR, Paulson JC. Differential processing of VIH envelope glycans on the virus and soluble recombinant trimer. *Nat Commun*. 2018 Sep 12;9(1):3693. doi: 10.1038/s41467-018-06121-4. PMID: 30209313; PMCID: PMC6135743.
99. Keri B. Sanborn¹, Mohan Somasundaran¹, Katherine Luzuriaga¹ and Thomas Leitner. 2015. Recombination elevates the effective evolutionary rate and facilitates the establishment of VIH-1 infection in infants after mother-to-child transmission. *Sanborn et al. Retrovirology* (2015) 12:96 DOI 10.1186/s12977-015-0222-0
100. Murrell, B et al. "FUBAR: A Fast, Unconstrained Bayesian Approximation for inferring selection." *Mol. Biol. Evol.* 30, 1196–1205 (2013).
101. Reza, FM. *An Introduction to Information Theory*. Dover Publications, Inc. NY (1994).
102. Alexey Drozdetskiy, Christian Cole, James Procter, Geoffrey J. Barton, JPred4: a protein secondary structure prediction server, *Nucleic Acids Research*, Volume 43, Issue W1, 1 July 2015, Pages W389–W394, <https://doi.org/10.1093/nar/gkv332>
103. Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.
104. Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves *BMC. Bioinform* 12 (1), 77.
105. Hongjaisee S, Nantasenamat C, Carraway TS, Shoombuatong W. VIHCoR: A sequence-based tool for predicting VIH-1 CRF01_AE coreceptor usage. *Comput Biol*

- Chem. 2019 Jun;80:419-432. doi: 10.1016/j.compbiolchem.2019.05.006. Epub 2019 May 20. PMID: 31146118.
106. Smith TF, Waterman MS. Identification of common molecular subsequences. J Mol Biol. 1981 Mar 25;147(1):195-7. doi: 10.1016/0022-2836(81)90087-5.
 107. Zhuang Jianling, Jetzt Amanda E., Sun Guoli, Yu Hong, Klarmann George, Ron Yacov, Preston Bradley D., Dougherty Joseph P. 2002. Human Immunodeficiency Virus Type 1 Recombination: Rate, Fidelity, and Putative Hot Spots. Journal of Virology. 11273-1128. American Society for Microbiology. doi: 10.1128/JVI.76.22.11273-11282.20022
 108. Langford E.S., Ananworanich J., Cooper D.A. 2007. Predictors of disease progression in HIV infection: a review. *AIDS Research and Therapy* 2007, 4:1
 109. Katz RA, Skalka AM. Generation of diversity in retroviruses. Annu Rev Genet. 1990;24:409-45. doi: 10.1146/annurev.ge.24.120190.002205. PMID: 1708222
 110. Juan Rodríguez Rivas. 2019. La coevolución en regiones de interacción entre proteínas: estudio y desarrollo de métodos computacionales. Universidad Autónoma de Madrid Programa de Doctorado en Biociencias Moleculares. Centro Nacional de Supercomputación Barcelona Supercomputing Center
 111. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 1993 May;10(3):512-26. doi: 10.1093/oxfordjournals.molbev.a040023. PMID: 8336541.

Anexos







The 3WBDS LA Organising Committee certifies that

Anahí Tania Sánchez León

has presented the poster entitled

"Effect of intra-subtype recombination on the evolution of env in HIV-1 "

at the 3rd WBDS LA Conference, virtually held from 20th to 23rd September 2022



Acti