



**UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO**

---

---

**FACULTAD DE CIENCIAS**

**CARACTERIZACIÓN DEL GEN *GAG* DE VIH-I CIRCULANTE  
EN LAS POBLACIONES MEXICANA Y GUATEMALTECA**

**T E S I S**

**QUE PARA OBTENER EL TÍTULO DE:**

**BIÓLOGA**

**P R E S E N T A:**

**FRIDA SOMMER OBANDO**



**DIRECTOR DE TESIS:  
DR. SANTIAGO ÁVILA RÍOS  
2012**



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



**CARACTERIZACIÓN DEL GEN *GAG* DE VIH-I  
CIRCULANTE EN LAS POBLACIONES  
MEXICANA Y GUATEMALTECA**

TESIS  
QUE PARA OBTENER EL TÍTULO DE:  
BIÓLOGA  
PRESENTA:

**Frida Sommer Obando**

Director de Tesis:  
**Dr. Santiago Ávila Ríos**

Centro de Investigación en Enfermedades Infecciosas (CIENI)  
Instituto Nacional de Enfermedades Infecciosas (INER)

Facultad de Ciencias  
Universidad Nacional Autónoma de México  
2012



<b>INTRODUCCIÓN .....</b>	<b>V</b>
<b>EPIDEMIA DEL VIH.....</b>	<b>7</b>
Epidemia del VIH a nivel global .....	7
Epidemia de VIH en Latinoamérica.....	10
Epidemia del VIH México.....	14
Epidemia del VIH en Guatemala .....	17
<b>CARACTERÍSTICAS DEL VIRUS DE LA INMUNODEFICIENCIA HUMANA (VIH).....</b>	<b>19</b>
Origen y filogenia del VIH .....	19
Estructura y organización del genoma viral.....	20
Ciclo de replicación del VIH .....	23
Patogénesis del VIH.....	26
<b>EVOLUCIÓN DEL VIH.....</b>	<b>28</b>
Evolución del VIH dentro del hospedero.....	28
Las moléculas HLA.....	30
Impacto de las moléculas HLA y el gen <i>gag</i> en la evolución del VIH a nivel poblacional .....	32
Evolución del VIH en las poblaciones mexicana y guatemalteca .....	35
<b>PLANTEAMIENTO DEL PROBLEMA .....</b>	<b>39</b>
<b>OBJETIVOS .....</b>	<b>41</b>
Objetivo general.....	41
Objetivos particulares.....	41
<b>HIPÓTESIS.....</b>	<b>43</b>
<b>METODOLOGÍA.....</b>	<b>45</b>
Amplificación del gen <i>gag</i> de VIH por los métodos de RT- PCR anidada y PCR anidada.....	45
Electroforesis en gel de agarosa .....	47
Secuenciación y análisis de genes virales.....	47
<b>RESULTADOS.....</b>	<b>51</b>
Descripción de las cohortes de pacientes de México y Guatemala.....	51
Filogenia del gen <i>gag</i> de VIH circulante en las poblaciones mexicana y guatemalteca.....	52
Variabilidad del gen <i>gag</i> de VIH que circula en México y Guatemala .....	55
Asociación entre moléculas HLA-I y posiciones del gen <i>gag</i> de VIH dentro de las poblaciones mexicana y guatemalteca .....	63
<b>DISCUSIÓN .....</b>	<b>77</b>
<b>CONCLUSIONES.....</b>	<b>85</b>
<b>REFERENCIAS.....</b>	<b>87</b>



# Capítulo 1

## Introducción

### 1.1. Epidemia del VIH

#### Epidemia del VIH a nivel global

Después de más de 30 años del descubrimiento del virus de la inmunodeficiencia humana (VIH), la epidemia global parece haberse estabilizado gracias a los múltiples programas internacionales enfocados en el monitoreo de la epidemia, la protección de grupos vulnerables, la prevención y la distribución de tratamiento antirretroviral (TAR) alrededor del mundo [1].

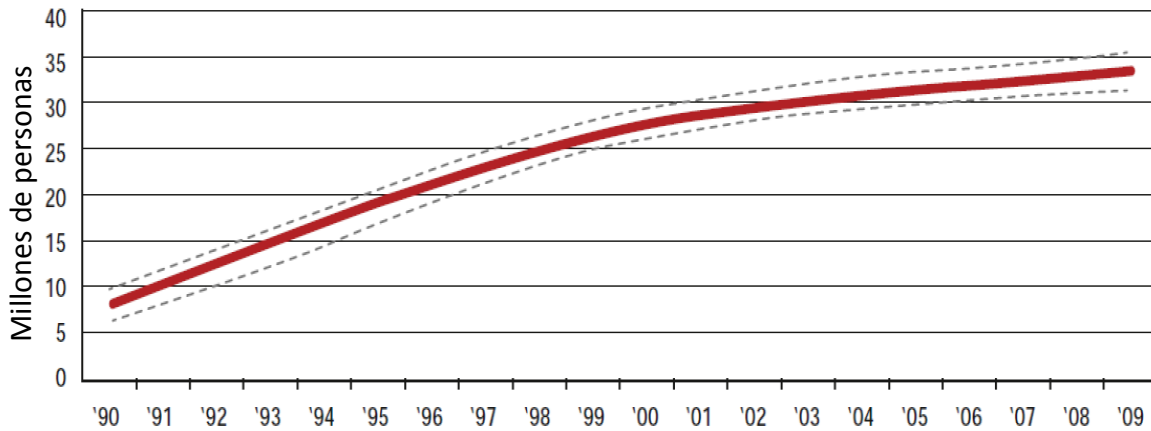
Desde finales de los años noventa se ha observado una disminución paulatina en el número anual de nuevas infecciones y en los últimos años, el número de muertes a causa de la infección por VIH también ha disminuido significativamente gracias al desarrollo, la optimización, distribución y facilitación del acceso al tratamiento antirretroviral a nivel global. Actualmente, 5 millones de personas en todo el mundo reciben terapia antirretroviral (aproximadamente el 35% del total de personas que necesitan este tratamiento [1]).

Con relación al 2001, el número de personas que viven con VIH ha aumentado de 28.6 millones a 34 millones en 2011 (Fig. 1.1.1) [1].

Este incremento se debe principalmente, al aumento en el acceso al tratamiento antirretroviral y la consecuente extensión de la expectativa de vida de las personas que se apegan a éste régimen terapéutico. El número de nuevas infecciones ha descendido un 25% en diez años a nivel global, de 3.1 millones en 1999 a 2.7 millones en 2011 (Fig. 1.1.2) [1].

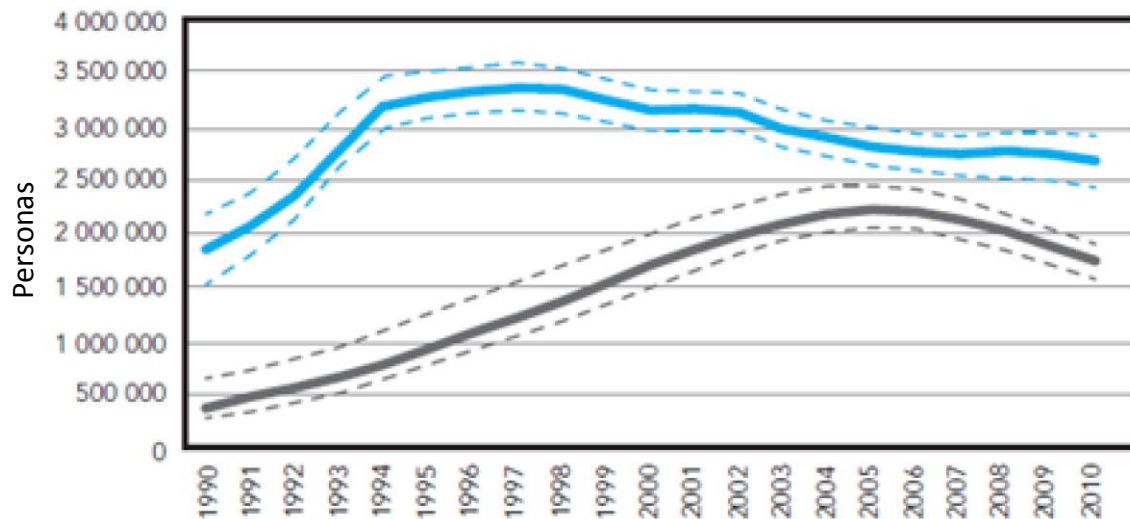


### Número total de personas infectadas con VIH.



**Figura 1.1.1. Número total de personas que viven con VIH en el mundo de 1990 a 2010.** La gráfica muestra el aumento en el número de personas que viven con VIH de 1990 a 2010. Se puede apreciar cómo se ha establecido la epidemia en los últimos años ya que el aumento en el número de personas que viven con la infección ha aumentado muy poco y este aumento se debe principalmente a la mayor expectativa de vida que tienen las personas con TAR [modificado de1].

### Nuevas infecciones y muertes relacionadas con la infección por VIH.



**Figura 1.1.2. Número de nuevas infecciones por VIH y muertes relacionadas a nivel global desde 1990.** La gráfica muestra cómo han disminuido tanto el número de nuevas infecciones (azul) por año como el número de muertes relacionadas con la infección por VIH (negro) de 1990 a 2010. Se puede reconocer que el número de nuevas infecciones por año ha disminuido paulatinamente desde 1998 gracias a la aplicación de medidas preventivas eficaces a nivel global. También se puede observar que desde el año 2005 el número de muertes relacionadas con la infección por VIH ha disminuido año con año gracias a la distribución de tratamiento antirretroviral en todo el mundo [modificado de1].

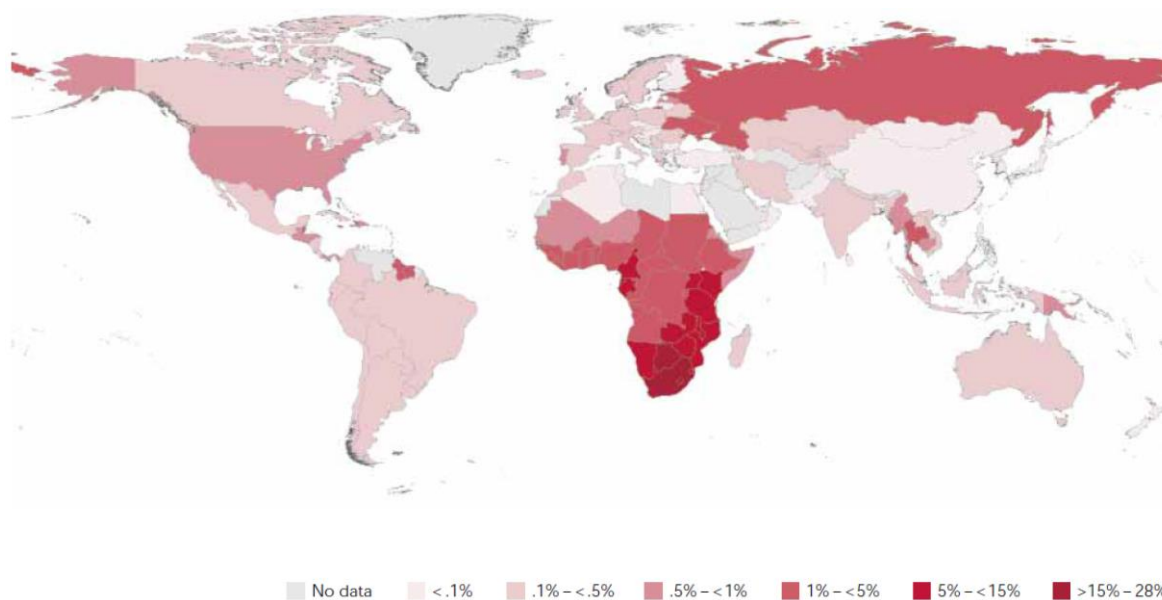
En África Subsahariana se concentran aproximadamente dos tercios del total de personas que viven con VIH (22 millones). Además, se registra el mayor número de nuevas infecciones en esta región (un promedio de 1.9 millones en 2011). La prevalencia promedio de la infección en adultos pasó de 5.9% en 2001 a 5.0% en 2011 y el número de muertes relacionadas con la infección pasó de 1.4 millones en 2001 a 1.2 millones en 2011<sup>[1]</sup>.

La región del Caribe es la más afectada después de África Subsahariana con una prevalencia en adultos del 1.1% en 2001 y 0.9% en 2011, seguida de Europa del Este y Asia central, región en la cual la prevalencia se ha duplicado en menos de 10 años, con una prevalencia de 0.85% en 2011 cuando era de 0.4% en 2001. Otras regiones que presentan un aumento en la prevalencia del VIH en su población adulta son el norte y el este de África (0.1% en 2001 y 0.6% en 2010), Oceanía (0.2% en 2001 y 0.3% en 2011) y América del Norte (0.4% en 2001 y 0.6% en 2011) <sup>[1]</sup>.

En Centro y Sudamérica la prevalencia del VIH en adultos es de 0.5% y ha permanecido inalterada desde 2001, al igual que la prevalencia de 0.2% que se ha mantenido en Europa central y occidental (figura 1.1.3) <sup>[1]</sup>.

El este de Asia registra la prevalencia más baja en el mundo con menos del 0.1% en 2010 (figura 1.1.3) <sup>[1]</sup>.

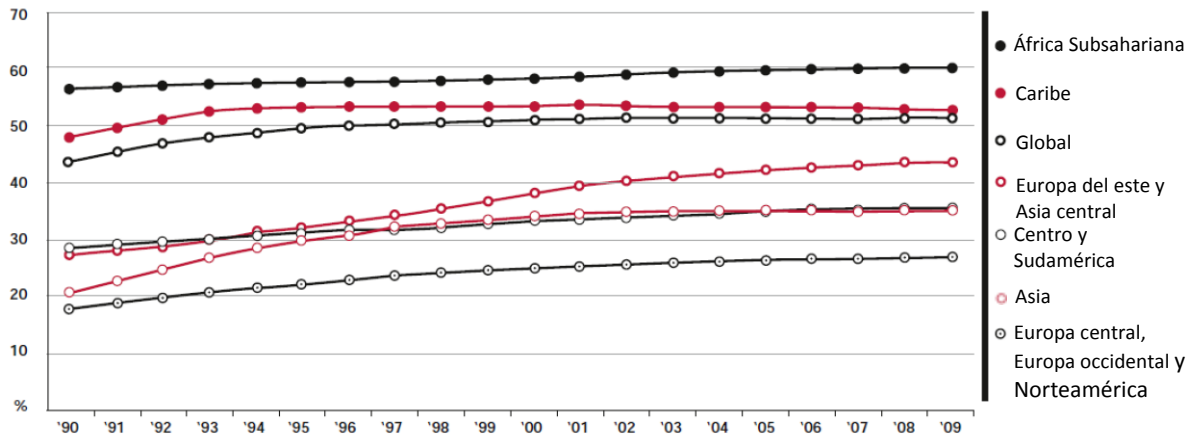
Prevalencia global de VIH, 2009



**Figura 1.1.3. Mapa de prevalencia mundial del VIH.** El mapa muestra la prevalencia de la infección por VIH en la población general de las diferentes regiones geográficas del mundo. África Subsahariana alberga cerca de dos tercios del número total de personas infectadas en el mundo y la prevalencia promedio en esta región es de 5.0% <sup>[tomado de1]</sup>.

A nivel global, el porcentaje de mujeres infectadas con VIH es ligeramente mayor que el de hombres infectados, con aproximadamente 51.6% (Fig. 1.1.4). Esto se debe principalmente a que en África Subsahariana el 60% de las personas infectadas

son mujeres y como se mencionó anteriormente, esta región alberga aproximadamente dos terceras partes del total de personas infectadas con VIH. En las regiones restantes el porcentaje de mujeres con VIH es inferior al de hombres con excepción del Caribe, dónde las mujeres infectadas representan aproximadamente un 53% del total de los casos (Fig. 1.1.4) [1].



**Figura 1.1.4. Proporción de mujeres infectadas con VIH-1 mayores de 14 años en el mundo.** El número de mujeres infectadas con VIH en el mundo supera el número de hombres infectados debido a que el 60% de las personas infectadas en África Subsahariana y el 52% en el Caribe, son mujeres y es en estas regiones dónde se concentra cerca del 70% de las personas infectadas en el mundo. En todas las regiones restantes, el número de hombres infectados es mayor que el de mujeres con la infección [modificado de1].

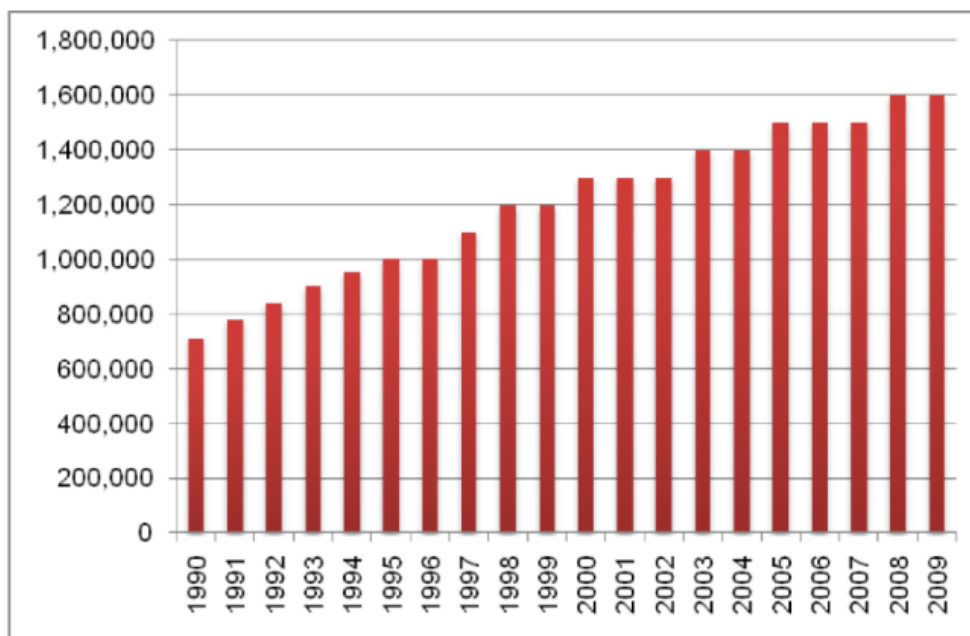
El 41% de las infecciones en 2010 ocurrieron en jóvenes de 15 a 24 años de edad y es la categoría de edad con el mayor riesgo de contraer la infección. Aunque la vía de infección en las diferentes regiones geográficas varía, a nivel global la vía sexual/heterosexual es la que predomina [1].

## Epidemia de VIH en Latinoamérica

La epidemia de VIH en Latinoamérica se ha caracterizado por ser particularmente estable, sobre todo, en los últimos años [2].

El número total de personas que viven con VIH en esta región continúa aumentando acorde con la tendencia global. En 2001 el número total de personas infectadas con VIH era de 1.1 millones y ascendió a 1.6 millones en 2009 principalmente, debido a la disponibilidad del tratamiento antirretroviral (Fig. 1.1.5) [2].

Más del 70% de las personas que viven con VIH residen en los cuatro países más grandes de la región, Argentina, Brasil, Colombia y México. Aproximadamente una tercera parte de las personas infectadas en Latinoamérica se concentra en Brasil [2].

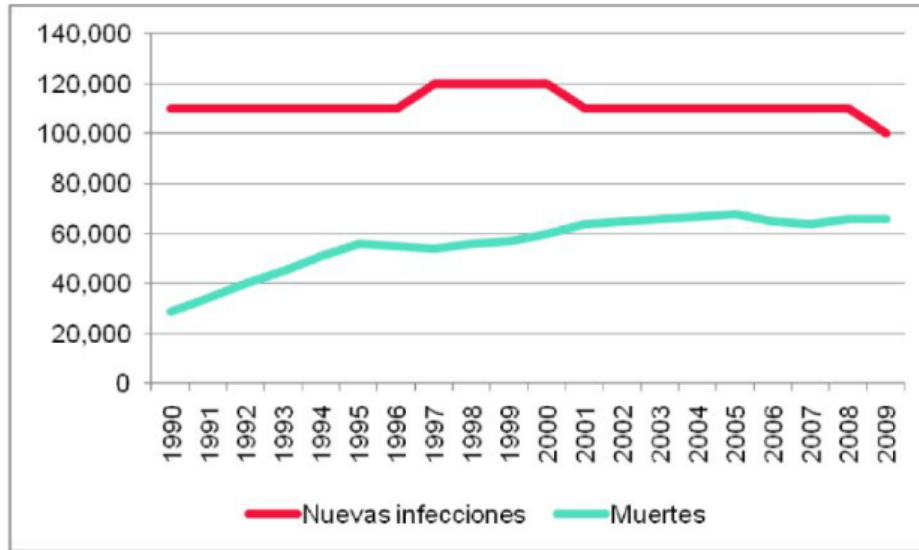


**Figura 1.1.5. Número aproximado de personas con VIH en 17 países de Latinoamérica.** La gráfica muestra el número de casos de VIH en diferentes países de Latinoamérica de 1990 a 2009. Se puede apreciar que el número de personas ha aumentado casi de manera constante cada año. En los últimos años el número de personas con VIH se ha mantenido estable en esta región [tomado de 2].

Se estima que en 2009 ocurrieron 92,000 nuevas infecciones con VIH en la región (Fig. 1.1.6) y se determinó que la vía de transmisión más común es la vía sexual-anal, que se da principalmente en redes de hombres que tienen sexo con hombres (HSH). También, se reconocen como significativos los números correspondientes a las infecciones entre los usuarios de drogas inyectables (UDI) y entre trabajadoras y trabajadores sexuales (MTS y HTS respectivamente) y sus clientes [2].

El número de defunciones relacionadas con la infección por VIH en 2011 fue de 58,000 cuando en 2001 era de 53,000 (Fig. 1.1.6) [2].

La prevalencia del VIH en la población adulta de América Latina está estimada en 0.4% y la epidemia del VIH/SIDA en esta región se clasifica como concentrada, dado que no se ha establecido en la población en general, sino que se concentra en sectores que se consideran clave. Como se indicó anteriormente, los sectores más destacados son los transexuales, hombres que tienen sexo con hombres (HSH), mujeres trabajadoras sexuales (MTS), hombres trabajadores sexuales (HTS) y usuario de drogas inyectables (UDI). La prevalencia del VIH en estos grupos está muy por encima de la prevalencia total en Latinoamérica.



**Figura 1.1.6. Número de nuevas infecciones con VIH y muertes relacionadas con la infección desde 1990 a 2009 en 17 países de Latinoamérica.** La gráfica muestra el número de nuevas infecciones (rojo) y el número de muertes relacionadas con la infección por VIH en Latinoamérica (azul). El número de nuevas infecciones se ha mantenido estable a lo largo del tiempo con excepción del período comprendido entre 1996 y 2001 que muestra un aumento y entre 2008 y 2009 se puede reconocer una disminución en el número de nuevas infecciones. Por esta razón se considera que la epidemia de VIH en Latinoamérica es estable. En cuanto al número de muertes relacionadas con la infección, se puede reconocer un aumento ligero cada año de manera sostenida [tomado de 2].

En el sector comprendido por transexuales, se ha reportado una prevalencia de hasta 36.5% (en Uruguay), en hombres que tienen sexo con hombres, la prevalencia del VIH puede llegar a 20.3% (en Chile) y es superior a 5% en todos los países de la región. Entre las trabajadoras sexuales, la prevalencia del VIH alcanza valores de hasta el 4.9% (Brasil), mientras que en hombres trabajadores sexuales de los países que reportan datos, puede alcanzar hasta el 22.8% (Argentina). Finalmente, en personas que usan drogas inyectables, la prevalencia de la infección supera el 5% en los países que reportan datos al respecto (Tabla 1.1.1) [2].

Tabla 1.1.1 Prevalencia del VIH en población adulta y joven, en grupos más expuestos a VIH y otras poblaciones vulnerables en algunos países de Latinoamérica.

	Población adulta	Mujeres jóvenes (15 a 24)	Hombres jóvenes (15 a 24)	Personas usuarias de droga		Trabajadores sexuales		Hombres que tienen sexo con hombres	Población trans	Población garífuna
				Intravenosa	Otras	Hombres	Mujeres			
<b>América Central y México</b>										
Costa Rica	0.3 [0.2-0.4]	0.1 [0.1-0.2]	0.2 [0.1-0.3]	...	...	...	...	12.7 <sup>e</sup>	...	...
El Salvador	0.8 [0.6-1.1]	0.3 [0.1-0.5]	0.4 [0.2-1.3]	...	...	...	4.1 <sup>a</sup>	9.8 <sup>a</sup>	...	...
Guatemala	0.8 [0.6-1.0]	0.3 [0.2-0.6]	0.5 [0.2-1.4]	...	...	...	1.0 <sup>b</sup>	18.3 <sup>b</sup>	...	...
Honduras	0.8 [0.5-1.0]	0.2 [0.1-0.4]	0.3 [0.1-1.1]	...	...	...	2.3 <sup>b</sup>	6.6 <sup>b</sup>	...	4.6 <sup>b</sup>
México	0.3 [0.3-0.4]	0.1 [0.1-0.2]	0.2 [0.1-0.2]	5.0	...	...	0.9 <sup>a</sup>	10.2 <sup>a</sup>	...	...
Nicaragua	0.2 [0.1-0.3]	0.1 [0.1-0.1]	0.1 [0.1-0.2]	...	...	...	...	4.2 <sup>a</sup>	...	...
Panamá	0.9 [0.6-1.5]	0.3 [0.1-0.5]	0.4 [0.2-1.3]	...	...	...	2.0 <sup>a</sup>	10.0 <sup>a</sup>	...	...
<b>Países Andinos</b>										
Bolivia	0.2 [0.1-0.3]	0.1 [<0.1-0.1]	0.1 [<0.1-0.3]	...	...	...	0.4 <sup>f</sup>	11.6 <sup>d</sup>	...	...
Colombia	0.5 [0.4-0.7]	0.1 [0.1-0.3]	0.2 [0.1-0.7]	...	...	...	1.6 <sup>d</sup>	...	...	...
Ecuador	0.4 [0.3-0.6]	0.2 [0.1-0.3]	0.2 [0.1-0.8]	...	...	...	3.8 <sup>b</sup>	19.2 <sup>c</sup>	...	...
Perú	0.4 [0.3-0.5]	0.1 [0.1-0.2]	0.2 [0.1-0.3]	...	...	...	0.5 <sup>d</sup>	10.1 <sup>a</sup>	29.6 <sup>f</sup>	...
Venezuela	...	...	...	...	...	...	...	...	...	...
<b>Cono Sur y Brasil</b>										
Argentina	0.5 [0.3-0.6]	0.2 [0.1-0.3]	0.3 [0.1-0.8]	...	4.8	22.8	1.9 <sup>d</sup>	11.8 <sup>d</sup>	34.0	...
Brasil	... [0.3-0.6]	... [0.1-0.4]	... [0.1-0.3]	5.9 <sup>a</sup>	...	...	4.9	12.6 <sup>a</sup>	...	...
Chile	0.4 [0.3-0.5]	0.1 [0.1-0.3]	0.2 [0.1-0.7]	...	...	...	0.7	20.3 <sup>a</sup>	...	...
Paraguay	0.3 [0.2-0.4]	0.1 [0.1-0.2]	0.2 [0.1-0.6]	9.0 <sup>b</sup>	...	10.7 <sup>b</sup>	1.8 <sup>b</sup>	8.7	...	...
Uruguay	0.6 [0.5-0.7]	0.2 [0.1-0.3]	0.3 [0.1-1.0]	18.5 <sup>a</sup>	9.5 <sup>a</sup>	19.3 <sup>b</sup>	...	9.1	36.5 <sup>b</sup>	...
<b>Total</b>	<b>0.4 [0.4-0.5]</b>	<b>0.2 [0.1-0.2]</b>	<b>0.2 [0.2-0.4]</b>	...	...	...	...	...	...	...

\*Nota. La población garífuna es un grupo étnico de origen africano que reside en varias regiones de Centroamérica y el Caribe. a=2004, b=2006, c=2007, d=2008, e=2009 y f=2010 [tomado de 2].

Doce de los dieciséis países de América Latina incluidos en el monitoreo anual de UNAIDS ofrecen 100% de tamizaje sistemático de las donaciones de sangre con criterios de calidad asegurada. Por esta razón, las infecciones por transfusión sanguínea representan un porcentaje mínimo (<1%) de los casos totales de VIH/SIDA. De igual manera, la vía de transmisión materno- fetal (perinatal) representa valores marginales con relación al total [2].

Un problema en el monitoreo y el manejo de la epidemia en Latinoamérica es la desaprobación social de los grupos en los que el riesgo de contraer la infección es más alto, razón por la cual no se reconocen, se ocultan o simplemente se ignoran eventos de transmisión en determinados grupos de la población [2].

## Epidemia del VIH en México

Los primeros casos de VIH en México se detectaron en 1983 y desde entonces la epidemia ha avanzado lentamente. En 1990 la prevalencia del VIH en México era de 0.13%. Actualmente, la prevalencia estimada del VIH es de 0.3% en la población adulta y comprende aproximadamente 220,000 personas, de las cuales cerca de dos terceras partes corresponden a HSH, el 22.3% a mujeres heterosexuales y 6.6% a clientes de trabajadoras sexuales, principalmente heterosexuales. El 3.1% corresponde a trabajadoras sexuales y el 2.0% a varones heterosexuales. Mientras que los UDI y las personas privadas de su libertad contribuyen cada uno con el 1.5% del total de personas que viven con VIH en México (Fig. 1.1.7) [3].

De acuerdo con la prevalencia de VIH en población adulta, que es el indicador que utiliza UNAIDS para comparaciones internacionales, México se ubicó en el décimo séptimo lugar en la región de América Latina y el Caribe (Fig. 1.1.8) [3].

Dentro de los sectores de la población que se consideran de alto riesgo (HSH, MTS, HTS UDI), los HSH presentan una prevalencia de la infección de aproximadamente 10.2%, entre los HTS de 8.9%, entre los UDI 5.0%, y las MTS presentan una prevalencia aproximada de 0.9% (Tabla 1.1.2).

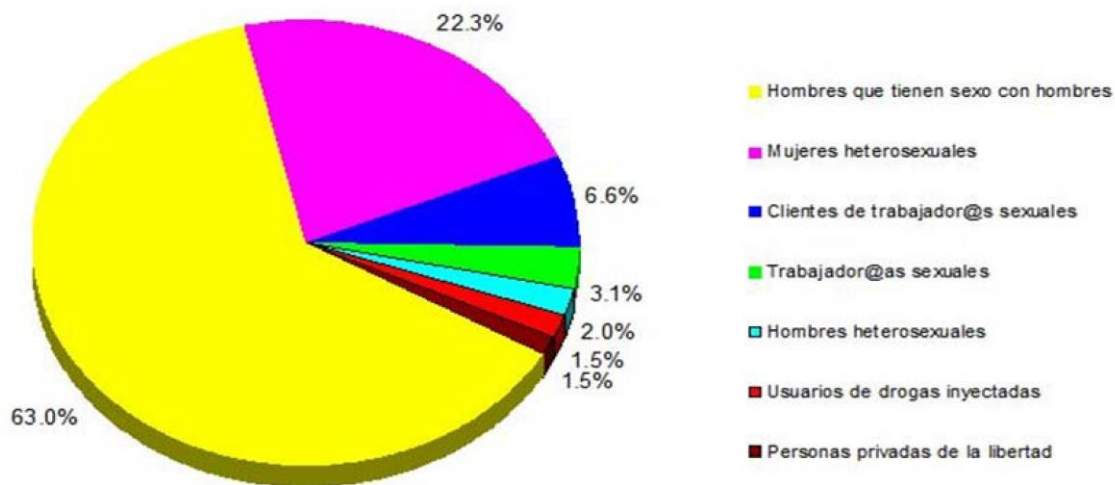
*Tabla 1.1.2 Prevalencia del VIH/SIDA en las poblaciones expuestas en México.*

Prevalencias: poblaciones más expuestas infectadas por el VIH. México 2009					
Población	Todos	Hombre	Mujer	<25	25+
Trabajadoras/es sexuales	1,7%	8,9%	0,9%	1,7%	1,7%
HSH	10,2%	-	-	7,9%	11,5%
UDI	5,0%	4,9%	4,6%	2,3%	5,7%

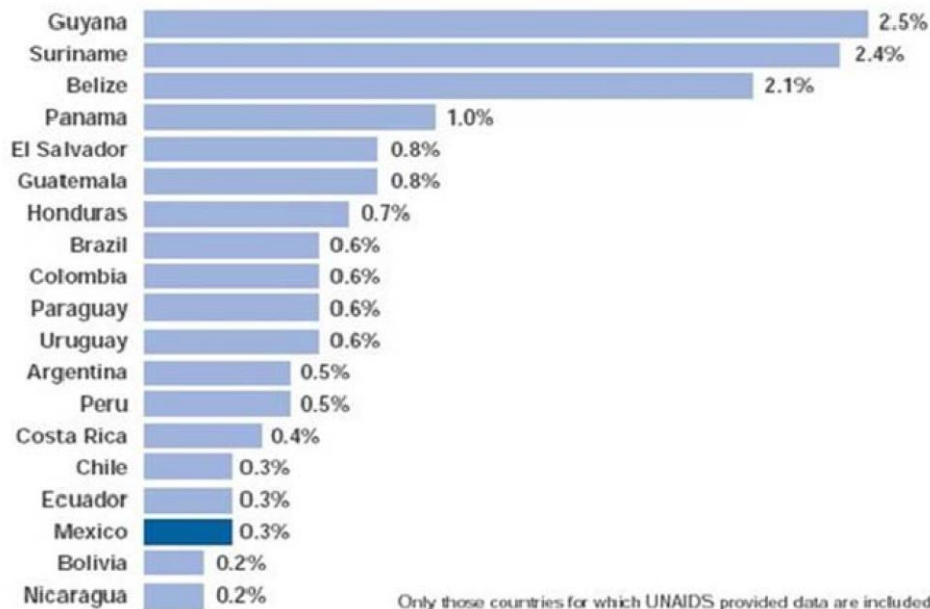
Fuente: Informe UNGASS /SIDA México -2010

\*HTS: hombres que tienen sexo con hombres, MTS: mujeres trabajadoras sexuales, UDI: usuarios de drogas inyectables [tomado de 3].





**Figura 1.1.7. Distribución porcentual de la infección por VIH/SIDA en los diferentes sectores de riesgo de la población mexicana.** La gráfica muestra el porcentaje correspondiente a los diferentes grupos de riesgo del total de los casos de VIH/SIDA en México. El 63% de los casos de VIH en México corresponden a HSH, el 22.3% a mujeres heterosexuales, el 6.6% clientes de trabajadoras y trabajadores sexuales, el 3.1% a MTS y HTS y tanto UDI como personas privadas de la libertad contribuyen con 1.5% de los casos cada uno [tomado de 3].



**Figura 1.1.8. Prevalencia del VIH/SIDA en 19 países de Latinoamérica y el Caribe.** En un estudio comparativo que contempló 19 países de Latinoamérica, México se ubicó en el lugar 17 con una prevalencia de 0.3% sólo por encima de Bolivia y Nicaragua, ambas con una prevalencia de 0.2 [tomado de 4].



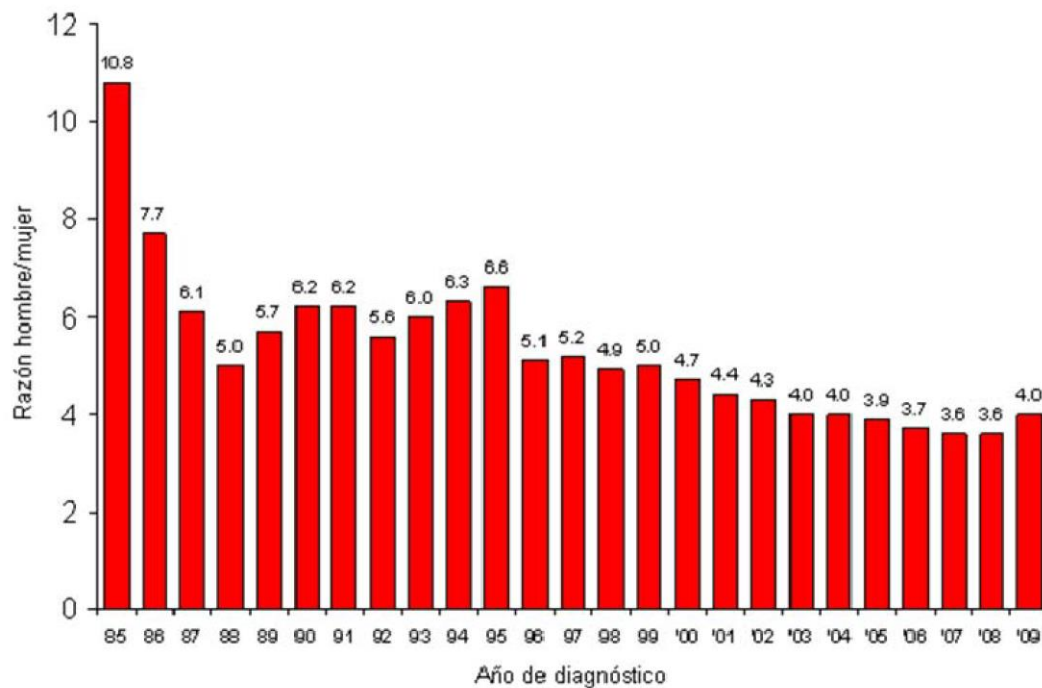
A pesar de que a finales de 2003, México alcanzó el acceso universal a TAR, en el año 2011 se registraron aproximadamente 9,200 muertes relacionadas con la infección [3, 6].

La vía de transmisión principal en México es la sexual, y representa más del 90% de los casos acumulados de SIDA en el país. Una característica que determinó en gran medida el curso de la epidemia en México fue que a finales de los años ochenta y principios de los noventa, los casos en HSH y HSH que también tenían sexo con mujeres llegaron a representar más del 90% del total de casos mientras que a finales de la década actual, representan alrededor del 50% [4].

Los casos de VIH/ SIDA ligados a la transfusión sanguínea, que llegaron a representar más del 80% de los casos al inicio de la epidemia, prácticamente han desaparecido como resultado del tamizaje obligatorio de la sangre a transfundir, así como por la prohibición de la comercialización de la misma, aplicada en nuestro país exitosamente en los primeros años de la epidemia [3].

La transmisión materno-fetal arroja números poco significativos con respecto al total de personas que viven con VIH [3].

La evolución de la razón hombre/mujer se ha reducido a lo largo de la epidemia del SIDA en México. Dicha razón disminuyó de 10.8 en 1985, hasta 3.6 en 2008 y aumentó a 4.0 para el año 2010. La vía de transmisión predominante en mujeres adultas (de 15 y más años de edad), es la vía sexual, en donde la transmisión de tipo heterosexual significa casi el 95% del total de casos registrados (Fig. 1.1.9) [3, 6].



**Figura 1.1.9 Proporción hombres/mujeres infectados con VIH/SIDA en México.** La razón hombre/mujer en México ha disminuido con el paso del tiempo. En 1985 había 10.8 hombres infectados con VIH por cada mujer infectada. En 2007 se redujo a 3.6 hombres infectados por cada mujer y en 2009 esta razón se estableció en cuatro hombres infectados por cada mujer infectada [tomado de3].

Con respecto a la distribución de la infección en los diferentes grupos de edad, el grupo de 25-34 años de edad es el grupo con mayor porcentaje de casos de SIDA, seguido del grupo de 35-44 años y en tercer lugar en importancia, aparecen los jóvenes de 15-24 años de edad. La tendencia histórica de la epidemia muestra una disminución en la proporción de casos de SIDA en los grupos de edad más afectados (25-44 años), y un aumento sostenido del porcentaje de casos de SIDA en jóvenes de entre 15 y 24 años [3].

La distribución geográfica de la infección en México reconoce dos aspectos como fundamentales en la actualidad; el aumento de la prevalencia en las fronteras del país, que parece estar relacionado con el hecho de que los países con los que México comparte fronteras, exhiben prevalencias de VIH más elevadas (Estados Unidos de América con 0.6%, Belice con 1% y Guatemala con 0.8%); y el reconocimiento de cárceles como focos concentrados de infección [3].

## **Epidemia del VIH en Guatemala**

Guatemala es el país más poblado de Centroamérica y el tercero en extensión territorial. El promedio de edad de la población guatemalteca es de 19 años [5].

El primer caso de SIDA en Guatemala se detectó en junio de 1984. Desde dicho año hasta noviembre 2009, el Centro Nacional de Epidemiología reporta 20,488 casos de SIDA y VIH, dos tercios de los cuales se presentan en hombres, aunque el nivel de sub-registro podría superar el 50% de los casos. En 2009, ocurrieron en promedio 18 nuevas infecciones por VIH al día y se estima que dos de cada tres personas con VIH desconocen su condición [5].

La prevalencia del VIH en este país es de 0.8% en la población adulta (15 años-49 años). La vía de transmisión sexual es la principal (94%), seguida por la vía materno-infantil (4%). El grupo de edad que presenta la prevalencia más elevada es la categoría de 15 años a 24 (0.34% del total) [5].

Al igual que en México y el resto de Latinoamérica, la epidemia se considera concentrada ya que afecta principalmente a sectores de la población tales como HSH (18.3%) y trabajadoras(es) sexuales (1.09%), individuos privados de libertad (3.24%), jóvenes en riesgo social (3.3%), a los pacientes con tuberculosis (12.9%) y a los uniformados (0.7%). No se han reportado a la fecha casos de transmisión en usuarios de drogas inyectables (UDI) [5].

La administración de antirretrovirales (ARV) en el país ha aumentado significativamente la esperanza de vida de las personas que viven con la infección y que están en tratamiento (82.61% sigue con vida después de 12 meses de TAR) [5].

La transmisión materno-infantil se redujo en un 30% en este país gracias a la administración de TAR a las madres durante el embarazo y a la implementación de asistencia médica adecuada al momento del parto. El 22.95% de mujeres embarazadas con VIH recibió TAR para prevenir la transmisión materno-infantil y 30% de los niños(as) nacido(as) de madres con VIH, tienen VIH [5].

Geográficamente, el 78% de los casos reportados se concentran en 7 de 22 departamentos de Guatemala. Los departamentos más afectados son, en primer lugar,

Guatemala, seguida por Retalhuleu, Izabal, Suchitepéquez, Quetzaltenango, Sacatepéquez y Escuintla (Fig. 1.1.10) [5].



**Figura 1.1.10. Departamentos de Guatemala en los que se concentran los casos de VIH/SIDA y que en conjunto, albergan aproximadamente el 80% de los casos reportados. Los departamentos más afectados son por la infección por VIH son Guatemala, Retalhuleu, Izabal, Suchitepéquez, Quetzaltenango, Sacatepéquez y Escuintla.**

Se estima que actualmente, el 69.24%, de las personas con VIH/SIDA, reciben tratamiento antirretroviral. Sin embargo, el acceso universal a servicios de prevención, atención, cuidados y apoyo para personas que viven con la infección, sigue siendo un reto para este país [5].

El 23.27% de las unidades de sangre donadas en el país no cumplen con criterios de control de calidad asegurada, únicamente el 3.71% de la población adulta se ha realizado la prueba de VIH durante los últimos 12 meses y conoce el resultado. El conocimiento sobre prevención del VIH y rechazo de ideas erróneas es limitado: 22.54% en jóvenes, 32.94% en HSH y 3.27% en MTS [5].

La relación Hombre: Mujer (H: M) de los casos de VIH y SIDA reportados ha disminuido de 7:1 en 1987 a 2.5:1 en el año 2009. Estas tasas demuestran que un mayor número de mujeres se está infectando con VIH y que gracias al tamizaje obligatorio en mujeres embarazadas ha aumentado el número de casos detectados [5].

## 1.2. Características del virus de la inmunodeficiencia humana (VIH)

### Origen y filogenia del VIH

El origen del VIH ha sido objeto de estudio y durante mucho tiempo, fue incierto. El hecho de que virus con características muy similares al VIH infecten a un gran número de vertebrados, condujo a investigadores a buscar en ellos el origen del VIH.

Actualmente, se sabe con certeza que la epidemia del VIH/SIDA se originó a partir de una serie de eventos zoonóticos de simios a humanos. Los virus de la inmunodeficiencia en simios (VIS) se consideran la evidencia más contundente de la hipótesis que postula que la cercana convivencia entre hombres y simios, favoreció eventos zoonóticos cuando los humanos entraron en contacto con sangre o tejidos infectados de simios. Algunas evidencias que sustentan la transmisión zoonótica del VIS de primates a humanos son:

- A. La similitud de los genomas de VIH y VIS.
- B. La prevalencia del VIS en sus hospederos naturales (chimpancés y otros simios).
- C. La distribución biogeográfica del VIH y VIS con secuencias muy similares abarca prácticamente la misma área [6].

El VIH pertenece a la familia Retroviridae y al género Lentivirus. Se ubica en el grupo VI (ss-RNA (+) –RT) de la clasificación de Baltimore [7], lo que implica que es un virus con un genoma constituido por dos hebras de RNA de cadena sencilla positiva y que requiere de la actividad de la reverso transcriptasa (RT) para sintetizar DNA de doble cadena a partir de una cadena sencilla de RNA [6].

Se dice que el VIH es un retrovirus exógeno ya que infecta células somáticas y por lo tanto, no se transmite verticalmente entre los hospederos, a diferencia de los retrovirus endógenos que infectan células germinales y por consiguiente, se transmiten de generación en generación en el organismo infectado [6].

La comparación de secuencias de proteínas y nucleótidos de una gran cantidad de muestras de VIH han demostrado la existencia de dos tipos de virus, el VIH-1 y el VIH-2 [6].

Múltiples análisis evolutivos de estos dos tipos virales establecen que se originaron independientemente, ya que mientras el VIH-1 está estrechamente relacionado con el SIV de chimpancés (VIS Scpz), el VIH-2 se ubica lejos del VIH-1 y es muy cercano al VIS de los monos verdes (o VISsm por *sooty mangabey* en inglés) [6].

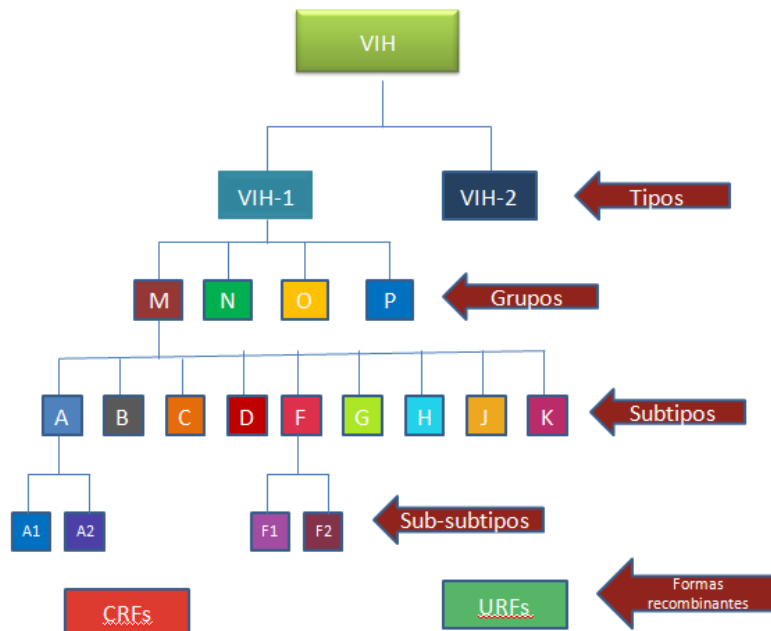
El VIH-1 es el tipo que predomina en el mundo y está presente en todos los continentes, mientras que la distribución del VIH-2 está restringida al continente Africano [6].

El VIH-1 se divide en cuatro grandes grupos que parecen haber surgido de cuatro eventos zoonóticos independientes (con VIS Scpz) y se denominan M, N, O y recientemente se identificó el grupo P, que se relaciona estrechamente con el VIS de go-

riila. Se reconocen nueve clados del virus dentro del grupo M y son nombrados de la A-D, de la F-H, J y K. Se han identificado algunos subtipos del grupo O y muy pocos del grupo N (Fig. 1.2.1) [6].

Los clados en el grupo M, se definieron porque difieren por lo menos en un 20% el uno del otro en la región de la envoltura y 15% en la región de *gag*. De manera general, la clasificación del VIH se hace con base en porcentajes de divergencia en los diferentes genes virales [6].

El subtipo B es el que predomina en el continente americano.



**Figura 1.2.1. Clasificación del VIH.** EL VIH se divide en dos grandes grupos, el VIH-1 y el VIH-2. El VIH-1 se divide a su vez en cuatro grupos, M, N, O y P. El grupo M es el más común y se divide en subtipos que van de la A a la D, F a H y J y K. Los subtipos A y F se dividen en dos sub-subtipos cada uno, A1 Y A2 Y F1 Y F2 respectivamente. También se reconocen formas recombinantes circulantes (CRFs) y formas recombinantes únicas (URFs).

## Estructura y organización del genoma viral

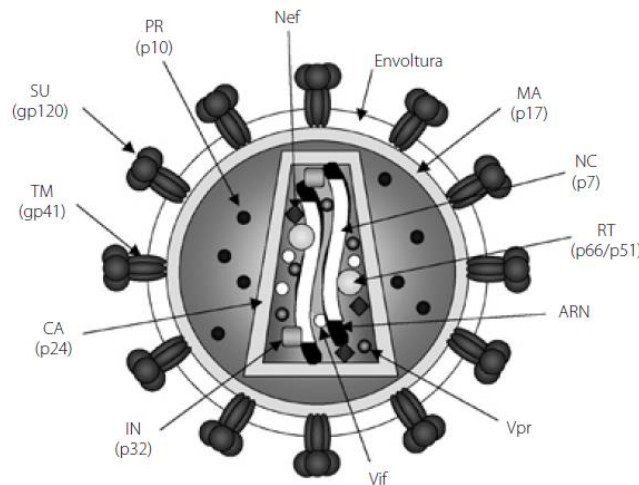
El VIH es una partícula de 100-120 nm de diámetro, de morfología heterogénea y con una cápside cónica recubierta por una membrana de origen celular (envoltura viral) que el virus adquiere al gemar de la célula hospedera [6].

En la envoltura viral se expresan cerca de 72 espículas formadas por trímeros asociados de las glicoproteínas virales gp 120 y gp 41. La molécula gp 120 se ubica en la superficie de la membrana (SU) y funge como receptor de las moléculas CD4 presentes en la superficie de las células que el VIH puede infectar (linfocitos T ayudado-

res CD4+ y células CD4+ del linaje de los macrófagos además de algunas células dendríticas). gp 41 es una proteína transmembranal (TM) indispensable para la fusión entre la partícula viral y la célula hospedera. Además de las glicoproteínas virales, en la envoltura se expresan proteínas y elementos celulares entre los que destacan las moléculas HLA-I y HLA II [6, 8,9].

La matriz es una capa proteica asociada a la cara interna de la envoltura que es fundamental para proteger la integridad de la partícula viral [6, 8, 9].

La cápside contiene el genoma viral, que consta de dos copias de (+)ss RNA, con una longitud de aproximadamente 9.3 Kb y que está recubierto por una capa de proteínas llamada nucleocápside. Además, la cápside viral contiene múltiples copias de las enzimas reverso transcriptasa (RT) e integrasa (IN) y dos RNAs de transferencia (tRNAs) celulares que actúan como primers en la reversotranscripción. La proteasa se ubica fuera de la cápside (Fig. 1.2.2) [6, 8].



**Figura 1.2.2. Estructura de una partícula de VIH.** La envoltura viral contiene trímeros asociados de las glicoproteínas gp120 y gp41 que forman estructuras de espículas, la matriz reviste la cápside viral t se encuentra asociada a la envoltura viral. La cápside viral contiene el genoma viral recubierto por la nucleocápside y múltiples copias de enzimas como la RT, IN, Vif, Nef y Vpr. La IN se ubica fuera de la cápside [tomado de 6].

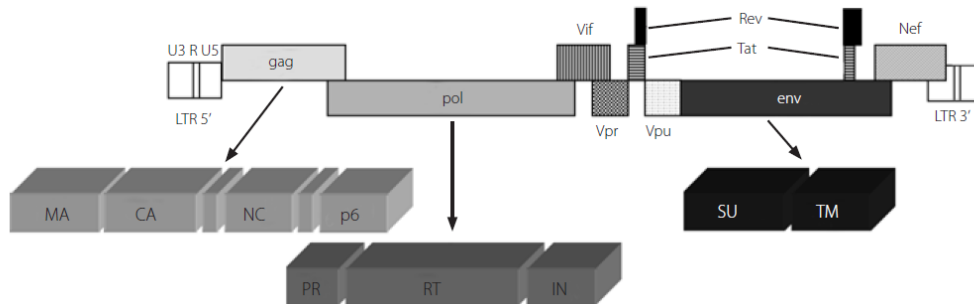
El genoma del VIH codifica 16 proteínas diferentes y consta de los tres genes característicos de los retrovirus, *gag* (gen antígeno grupo-específico), *pol* (gen que codifica la maquinaria enzimática del virus, principalmente la RT) y *env* (gen que codifica las glicoproteínas virales que se expresan en la envoltura) que codifican para las proteínas estructurales y las enzimas del virus. En sus extremos, el genoma del VIH presenta secuencias largas de repetición terminal (LTRs) que se subdividen en tres dominios, U3, R y U5. Estas secuencias cumplen diversas funciones entre la que destaca el reclutamiento de la RNA polimerasa II (RNAPII) y otros elementos reguladores necesarios para la transcripción del genoma viral integrado (Fig. 1.2.3) [6, 9].

El gen *gag* codifica las proteínas de matriz (MA o p17), la proteína de cápside (CA o p24), la proteína de nucleocápside (NC o p7), una proteína indispensable para la liberación de los viriones de la célula infectada, encargada de reclutar proteínas en la

membrana para facilitar el proceso de gemación llamada p6 y dos proteínas espaciadoras llamadas p1 y p2 [6].

La transcripción y traducción de *gag* da como resultado una poliproteína precursora llamada Pr.55gag, que es cortada por una proteasa viral para generar las proteínas individuales al momento de la salida de los nuevos viriones [6].

El gen *pol* codifica las enzimas indispensables para la retrotranscripción, integración del genoma viral y la maduración del virus, por un mecanismo de autocatálisis de un precursor llamado Gag-Pol o Pr. 160gag-pol [6].



**Figura 1.2.3 Genoma del VIH.** El genoma del VIH presenta dos secuencias largas de repetición terminal en sus extremos y consta de tres genes característicos de los retrovirus: *gag*, *pol* y *env*, además, los múltiples marcos de lectura abiertos codifican diferentes proteínas reguladoras y accesorias indispensables para el virus [tomado de 6].

La RT (p66/51, que se refiere a las dos subunidades que conforman la proteína) es una enzima similar a la DNA-polimerasa, que tiene como función sintetizar DNA de doble cadena a partir de una cadena sencilla de RNA y que además, tiene un dominio de RNasa H que le confiere la función de una endonucleasa inespecífica para hidrolizar RNA. La cadena doble de DNA que resulta de la actividad de RT se denomina provirus [6].

La actividad de la RT es responsable en gran medida de la alta variabilidad genética del VIH, ya que carece de la actividad de exonucleasa 3'-5' y por lo tanto, es incapaz de corregir los errores que ocurren durante la replicación. Este hecho conlleva a una tasa de mutación altísima en el genoma del virus que se estima en  $3.4 \times 10^{-5}$  mutaciones por par de bases por ciclo de replicación in vivo [6].

Se calcula que diariamente se producen  $10^9$  (mil millones) de viriones en una persona infectada con VIH/SIDA [6].

La IN (p32), como su nombre lo indica, es la proteína encargada de integrar el DNA viral resultante de la retrotranscripción al DNA cromosómico de la célula hospedera. La PR (p10) tiene como función principal generar proteínas individuales a partir de las poliproteínas precursoras Pr.55gag y Pr.160gag-pol al momento de la gemación. Antes de que los nuevos viriones puedan gemar de la célula hospedera se generan múltiples rearrreglos estructurales y morfológicos en el virión que se denominan *maduración* [6,8].

El gen *env* codifica las proteínas de envoltura gp120 y gp41 que, como se mencionó anteriormente, son indispensables para el reconocimiento de moléculas CD4 en

la superficie de la célula y la entrada del virus a la misma. Estas glicoproteínas son el resultado de la proteólisis del precursor gp 160 por parte de una proteasa celular [6]. gp 120 posee los motivos de unión a CD4 y receptores de quimiocinas como CCR5 y CXCR4, que funcionan como correceptores al momento de la unión entre el virus y la célula [6, 8, 9].

La presencia de múltiples marcos abiertos de lectura (ORFs) sobrelapados resulta en la síntesis de proteínas accesorias y reguladoras que son necesarias para que la replicación viral sea eficiente y que contribuyen en varios aspectos de la patogénesis del virus *in vivo* [6, 8, 9].

Dentro de las proteínas reguladoras se encuentran Tat (transactivador transcripcional) y Rev (regulador de la expresión de genes virales) [6, 8].

Como proteínas accesorias, el VIH-1 presenta Nef (efector negativo), Vif (factor de infectividad viral) y las proteínas virales r (Vpr) y u (Vpu). El VIH- 2 presenta Vpx en vez de Vpu [6].

## Ciclo de replicación del VIH

El VIH entra a la célula por la interacción de gp 120 y CD4, el receptor principal del virus que permite a éste infectar linfocitos T CD4+, células CD4+ del linaje de los macrófagos y ciertas poblaciones de células dendríticas. La unión de CD4 y gp 120 resulta en la expresión de receptores de quimiocinas en la molécula CD4 entre las cuales destacan CCR5 y CXCR4, que funcionan como correceptores virales. El correceptor que un virus usa, determina el tropismo del mismo, por lo tanto, los aislados de VIH se clasifican en R5 o macrofagotrópicos si el correceptor que usan es CCR5, X4 o linfotrópicos si usan CXCR4 y R5/X4 si pueden usar ambos [6].

La interacción de CD4 y los correceptores con gp120 desencadena una serie de cambios conformacionales que derivan en la exposición del dominio de fusión de la proteína gp41. Este dominio, se proyecta hacia la membrana de la célula hospedera de manera similar a la de un arpón, provocando que la membrana celular y la envoltura viral se fusionen y que la partícula viral ingrese a la célula hospedera [6, 8].

Algunas células dendríticas (DC), particularmente las que se ubican en las submucosas, presentan una lectina de tipo C con alta afinidad por gp120 (denominada DC-SIGN) que permite la unión entre la célula y la partícula viral, pero que no desata la serie de cambios conformacionales necesarios para que el virión pueda infectar exitosamente la célula. La unión de gp 120 con DC-SIGN promueve la endocitosis de la partícula viral y el almacenamiento de ésta en compartimientos ácidos celulares que, cuando las células maduran y migran a órganos linfáticos secundarios se exponen en la membrana celular y diseminan la infección por VIH, ya que en estos órganos hay una gran cantidad de células CD4+. Por esta razón, se dice que algunas células dendríticas actúan de manera similar a la de un caballo de Troya [6, 8].

Aunque no se conocen con certeza los eventos subsecuentes a la entrada del virus en las células que puede infectar exitosamente, se sugiere que la cápside viral viaja a las proximidades del núcleo celular por redes de microtúbulos y por filamentos de actina en la última etapa, que por pequeños movimientos, encallan la cápside en los



poros nucleares. En los poros nucleares y dentro de la cápside, se lleva a cabo la retrotranscripción y una vez concluida, ocurre la desencapsidación [6].

La desencapsidación está mediada por una estructura de triple cadena de DNA característica de los genomas lentivirales llamada DNA *flap*, un tracto central de polipurina (cPPT) y una secuencia central de terminación (CTS). La desencapsidación favorece la formación del complejo de preintegración (PIC) que está constituido por el DNA viral de doble cadena, la IN, RT, MA, NC y Vpr. El mecanismo de entrada del PIC al núcleo celular no se conoce con detalle, pero es posible que proteínas que poseen secuencias canónicas de reconocimiento nuclear como IN, MA y Vpr, jueguen un papel importante [6].

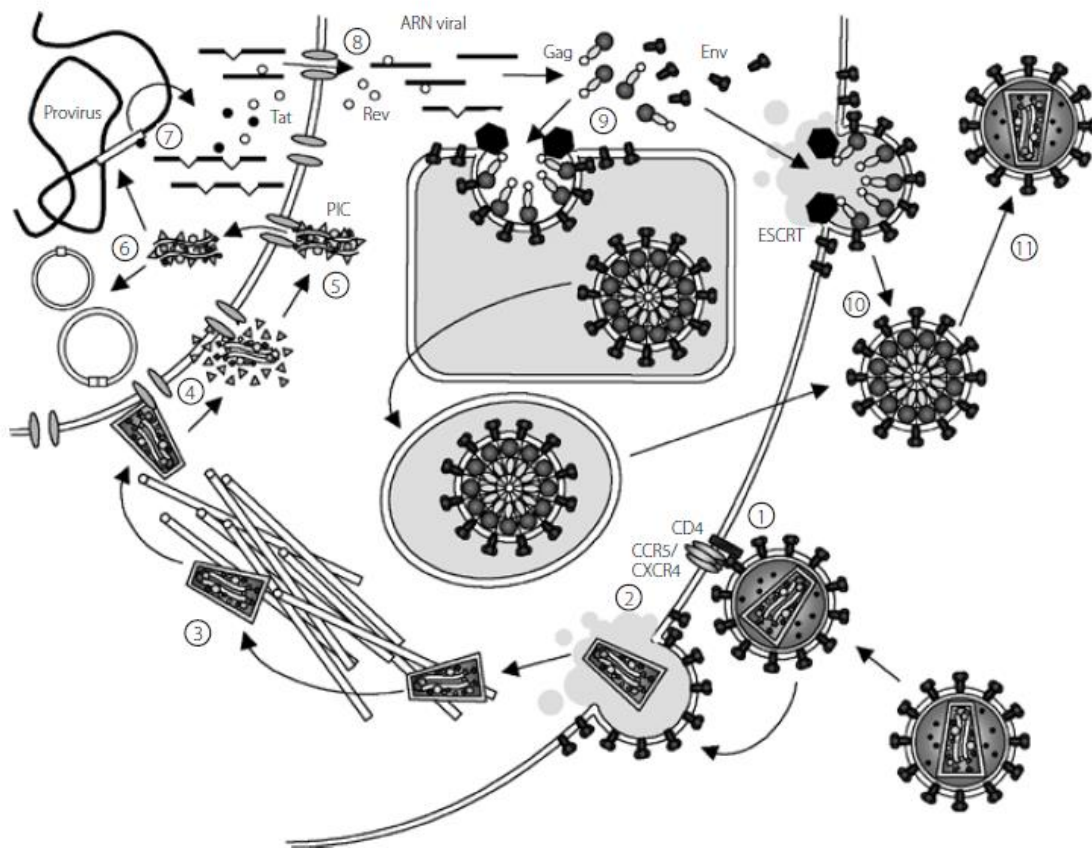
Una vez en el núcleo, el DNA viral puede integrarse al genoma de la célula hospedera o circularizarse y formar estructuras episomales no funcionales. Por lo tanto, la actividad de la IN es indispensable para que la infección sea productiva. Dicha actividad consiste en retirar dos nucleótidos en los extremos del DNA viral, realizar un corte escalonado en el DNA celular y finalmente, unir ambas cadenas de DNA. El genoma viral se integra comúnmente en regiones transcripcionalmente activas [6, 8, 9].

Es posible transcribir el genoma viral gracias al *primer* (cebador) y otros elementos reguladores de la transcripción ubicados en la región R de los LTRs y la actividad de Tat, que permite formar transcritos virales largos. Diferentes eventos de corte y empalme (*splicing*) dan como resultados las diferentes proteínas del VIH. Los transcritos tempranos con muchos eventos de corte y empalme, originan las proteínas accesorias y reguladoras mientras que los transcritos tardíos sin alteraciones o con un solo evento de corte y empalme, originan las proteínas estructurales y la maquinaria enzimática del virus. La proteína reguladora Rev se encarga de exportar los transcritos tardíos a las inmediaciones de la membrana celular (o en endosomas tardíos en el caso de monocitos/macrófagos) ya que es ahí donde el ensamblaje y la salida de nuevos viriones toman lugar [6].

La salida de los nuevos viriones está dirigida principalmente por las poliproteínas Pr.55 Gag y Pr.160Gag Pol, Env y Vpu. Las poliproteínas se asocian a regiones de la membrana celular ricas en colesterol y glicolípidos, llamadas balsas lipídicas, ya que favorecen la liberación, la estabilidad de la partícula viral y la fusión de la envoltura de los nuevos viriones con la membrana de nuevas células blanco [6, 8, 9].

Para que los viriones nuevos puedan salir, tiene que completarse el proceso de maduración a cargo de la PR. La proteína p6 es fundamental para la salida de los viriones maduros ya que contiene secuencias “dominio terminal” que unen proteínas de enlace, que facilitan la salida de los viriones maduros, con elementos del sistema ESCRT. El sistema ESCRT está involucrado con el tránsito de vesículas y cuerpos multivesiculares al lisosoma para su degradación, sin embargo, el VIH secuestra este sistema para promover la salida de nuevos viriones de la célula hospedera (figura 1.2.4) [6,

9].



**Figura 1.2.4. Ciclo de replicación del VIH-1.** La glicoproteína gp120 de envoltura se une al receptor CD4+ de la célula hospedera y desencadena múltiples cambios conformacionales en gp41 y la exposición de receptores de quimiocinas (CCR5 y CXCR4) en la molécula CD4+ necesarios para completar la fusión entre la célula hospedera y la envoltura viral. Una vez que la partícula viral entra a la célula hospedera, viaja a los poros nucleares por redes de microtúbulos y actina. En el poro nuclear y dentro de la cápside se lleva a cabo la retrotranscripción y posteriormente, ocurre la desencapsidación. Se forma el complejo de preintegración (PIC) y por mecanismos poco caracterizados aún, el genoma viral y múltiples enzimas virales ingresan al núcleo de la célula hospedera. La IN integra el genoma viral a una región transcripcionalmente activa del genoma de la célula hospedera y con ayuda de Tat, inicia la retrotranscripción del genoma viral. Los transcritos tempranos sufren muchos eventos de corte y empalme y dan origen a las proteínas accesorias y reguladoras. Los transcritos tardíos inalterados o con un solo evento de corte y empalme se exportan al citoplasma por Rev y originan las proteínas estructurales. Antes de la salida de los nuevos viriones, p6 recluta a los precursores (Pr.55 Gag y Pr.160. Gag pol) y a las proteínas virales y el RNA viral que van a conformar los nuevos viriones para que finalmente, la PR lleve a cabo el proceso de maduración y los nuevos viriones puedan salir de la célula hospedera para infectar otras células [tomado de 6].

## Patogénesis del VIH

La infección por VIH puede transmitirse principalmente por tres vías:

- **Vía sexual.** Sostener relaciones sexuales sin protección con personas infectadas es la vía de transmisión más común a nivel global.
- **Vía intravenosa o por contaminación de instrumentos.** La transfusión de sangre contaminada, el trasplante de órganos contaminados o incluso procedimientos de inseminación artificial con semen de personas infectadas pueden derivar en la transmisión del VIH. El uso compartido de jeringas contaminadas (común entre los UDI) o de instrumentos quirúrgicos contaminados también son una vía de transmisión del virus.
- **Vía materno- infantil.** La transmisión del VIH de madre a hijo se puede dar durante el embarazo, el parto o la lactancia sin embargo, se puede prevenir administrando TAR oportunamente a la madre [6, 10, 11].

Dado que las células blanco del VIH son primordialmente los linfocitos TCD4+, que juegan un papel central en la respuesta inmune adquirida, su destrucción conlleva a una inmunodeficiencia grave en las personas infectadas. Tradicionalmente, el desarrollo de la infección por VIH se divide en tres etapas: la infección aguda o primaria, la etapa crónica o asintomática y la etapa de SIDA (Fig. 1.2.5) [6, 10, 11].

La etapa aguda comprende las primeras cuatro semanas tras la infección y se caracteriza por un aumento acelerado de la carga viral (RNA viral) en el plasma sanguíneo y un desplome en la cantidad de linfocitos TCD4+ en sangre periférica. A partir de las dos semanas post-infección, se presentan diferentes síntomas no específicos como fiebre, faringitis, cefaleas y muchas veces malestares gastrointestinales (diarreas), que en conjunto se conocen como síndrome retroviral agudo. Existen también linfadenopatías generalizadas [9, 10, 11].

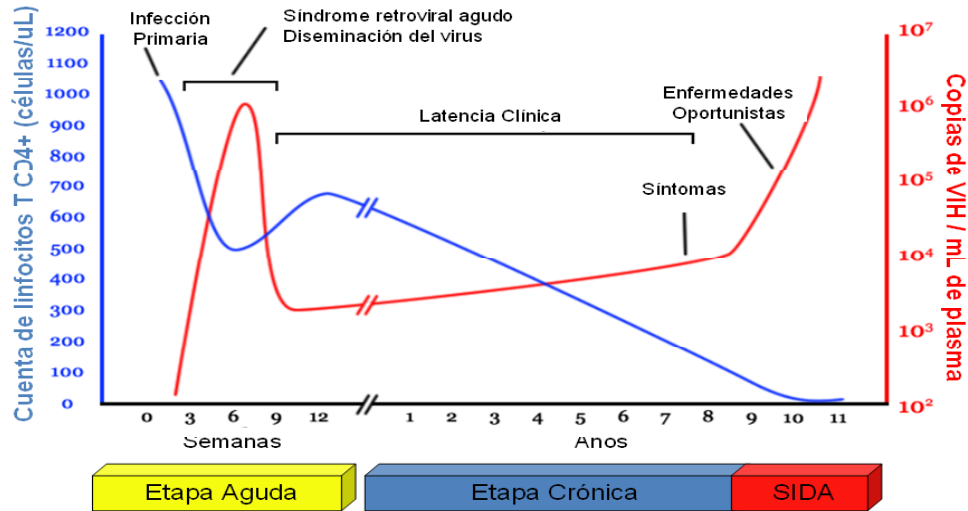
Se sabe con certeza que durante la etapa aguda, la replicación viral toma lugar en los nodos linfáticos y que a partir de ahí se disemina rápidamente a otros órganos como el bazo, la médula ósea y tejidos linfoides asociados a mucosas (particularmente el tejido linfoide asociado a la mucosa intestinal). Como efecto de la amplificación de la infección, en la etapa aguda se reconoce un aumento drástico en la carga viral hasta alcanzar un pico de viremia y una marcada disminución de linfocitos TCD4+ como resultado de dicho aumento [9]. La variabilidad viral en esta etapa es reducida y es posible reconocer una cuasiespecie dominante además de un aumento considerable en la cantidad de linfocitos TCD8+ activados [11, 12]. Los linfocitos TCD8+ o linfocitos citotóxicos (CTL por sus siglas en inglés) inducen la muerte celular programada (apoptosis) por medio de enzimas (perforina y granzima), al ser estimulados por los antígenos intracelulares presentados en el contexto del complejo principal de histocompatibilidad (MHC) de clase I y son responsables de un control parcial de la infección por VIH, que marca el fin de la etapa aguda y el inicio de la etapa crónica [10, 11, 12, 13].

En la etapa crónica, la carga viral es mucho menor que en la aguda y el número de linfocitos T CD4+ aumenta ligeramente como consecuencia de la aparición de la respuesta citotóxica de CTLs específicos [10, 11]. Se establece un punto de equilibrio de-

nominado “set point” viral que se refiere a la carga viral que presenta el individuo tras la etapa aguda [10, 11, 12, 14, 15, 1, 17]. El set- point viral es un reflejo fiel de la efectividad de la respuesta de los CTLs para controlar la viremia y por lo tanto, determina en gran medida el curso de la infección. Como consecuencia de la respuesta de CTLs, se pueden reconocer nuevas variantes virales como consecuencia de la presión selectiva que estas células ejercen sobre el virus [14, 15, 16, 18, 19 20].

En la etapa crónica, la replicación viral continúa pero a una tasa menor que en la etapa aguda y la disminución de linfocitos T CD4+ continúa también de manera paulatina. En la mayoría de los casos, la etapa crónica tiene una duración de entre diez y doce años, sin embargo, hay casos en los que la progresión a la etapa de SIDA se da de manera acelerada en uno o dos años (progresores rápidos) y otros casos en los que la progresión es muy lenta o nunca ocurre (controladores o controladores elite, respectivamente) [9, 10, 11, 12]. La progresión de la infección a la etapa de SIDA depende tanto de características inmunológicas propias de la persona infectada, como de factores genéticos, como el repertorio de moléculas HLA que presente el individuo [13, 14, 15, 16, 17, 18, 19, 20]. También influyen factores virológicos como la presencia de ciertas mutaciones que favorezcan o afecten de manera negativa la replicación viral o que favorezcan o dificulten el reconocimiento del virus por parte del sistema inmune [14, 15].

La etapa de SIDA empieza cuando la carga viral aumenta nuevamente y los linfocitos T CD4+ disminuyen (conteo de células T CD4+ < 200 células/ $\mu$ l) como resultado de un sistema inmune exhausto que aún con TAR, pierde el control parcial sobre la infección que había sostenido durante la etapa crónica. La etapa de SIDA se caracteriza por la aparición de infecciones oportunistas como candidiasis oral, tuberculosis, infecciones por *pneumococco*, *citomegalovirus* y *herpesvirus* entre otros. También pueden desarrollarse enfermedades malignas tales como el sarcoma de Kaposi, linfomas y diferentes tipos de carcinomas. La incapacidad del sistema inmune de responder adecuadamente ante las diferentes infecciones o enfermedades en la etapa de SIDA, deriva irremediablemente en la muerte de los individuos [10, 11, 17]. La figura 1.2.5 resume el desarrollo típico de la infección por VIH.



**Figura 1.2.5. Desarrollo de la infección por VIH-1.** En la etapa aguda (desde el momento de la infección hasta las primeras cuatro semanas) se distingue una pérdida acelerada de linfocitos TCD4+ y un aumento marcado de la viremia en el plasma sanguíneo hasta alcanzar un pico de viremia. En la etapa crónica la carga viral disminuye y el número de linfocitos TCD4+ aumenta hasta alcanzar un equilibrio. Esto se debe principalmente a la actividad de CTLs específicas y conlleva a su vez a una diversificación viral como respuesta ante la presión selectiva que ejercen dichas células. La etapa crónica suele ser asintomática y la presencia de síntomas o la presencia de infecciones oportunistas, suele indicar el principio de la etapa de SIDA que conlleva irremediablemente a la muerte a las personas [modificada de 11].

## 1.3. Evolución del VIH

### Evolución del VIH dentro del hospedero

La dinámica evolutiva del VIH dentro del hospedero, se fundamenta en poseer una alta variabilidad genética que le confiere plasticidad al virus para adaptarse y evadir las diferentes respuestas inmunes que el hospedero monte en su contra para controlar o erradicar la infección. La alta variabilidad genética del VIH proviene principalmente de la actividad de la RT, que como se mencionó con anterioridad, es una enzima similar a la DNA-polimerasa, que carece de la actividad de exonucleasa 3'-5' y por lo tanto, es muy propensa a cometer errores durante la replicación [6,14 15, 16, 18,19, 20].

Además de la mutación, la recombinación entre cadenas de DNA de un mismo virus o de virus diferentes en el núcleo de la célula hospedera, contribuye de manera importante con la enorme variabilidad genética del VIH [6, 13].

La alta tasa de error de la RT y la recombinación de genomas virales son responsables de que en cada evento de replicación se produzca un genoma viral único, diferente de todos los demás genomas virales circulantes. Por lo tanto, en un individuo infectado con VIH, coexisten poblaciones de variantes virales distintas pero relacionadas que se denominan "cuasiespecies" [12, 16, 17].

La existencia de múltiples cuasiespecies garantiza la permanencia del virus en el hospedero, ya que si éste monta una respuesta inmune específica, sólo neutralizará a una proporción de las variantes virales circulantes y la infección persistirá. Normalmente, el sistema inmune neutraliza la cuasiespecie dominante, por lo que una cuasiespecie diferente a ésta, la sustituirá cuando se haya erradicado a la primera y orillará al sistema inmune a montar otra respuesta específica hacia la nueva variante predominante [10, 11, 12]. Esta interacción ocurre durante todo el curso de la infección por VIH y este fenómeno evolutivo se describe en la “hipótesis de la reina roja” o “carrera armamentista” propuesta por Van Valen en 1973. La hipótesis de la reina roja describe una relación de co-evolución basada en adaptaciones recíprocas entre dos organismos. De manera general, la adaptación de un organismo a una condición impuesta por otro organismo, deriva en una “contra-adaptación” del segundo organismo ante la adaptación del primero en un ciclo permanente y que termina sólo cuando uno de los dos organismos muere o es erradicado [21].

En la infección por VIH, esta dinámica ocurre a diferentes niveles, pero el caso de la respuesta de CTLs contra el virus resulta de mucha importancia, ya que la respuesta citotóxica ha demostrado ser la más efectiva para controlar la replicación viral. Los CTLs reconocen células infectadas por patógenos intracitoplásmicos, destruyéndolas mediante mecanismos altamente específicos. El argumento principal a favor de las CTLs como agentes primordiales para el control de la replicación viral, es la asociación temporal entre la clara disminución de la carga viral al final de la etapa aguda y la aparición de la respuesta de CTLs específicas. Además, se llevaron a cabo experimentos en simios en los que se bloqueó la actividad de las células CD8+ y se observó una replicación descontrolada del virus y una progresión acelerada a la etapa de SIDA. La efectividad de la respuesta de CTLs está estrechamente ligada a moléculas muy variables del sistema inmune llamadas HLA [14, 15].

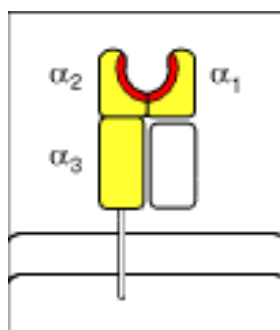
El término HLA (antígeno leucocitario humano) o MHC (complejo principal de histocompatibilidad) se refiere a una serie de genes altamente polimórficos, ubicada en el brazo corto (p) del cromosoma 6 de los seres humanos. Los productos más representativos de estos genes son proteínas de membrana que en su porción extracelular, presentan un surco para unir antígenos cuya estructura y composición es altamente variable y le permiten al organismo discriminar lo propio de lo no-propio [13].

Los antígenos capaces de unirse a moléculas HLA son péptidos derivados de estructuras proteicas de patógenos que se degradan intracelularmente. Los péptidos antigénicos son entonces presentados por moléculas HLA en la superficie de la célula infectada para que ésta pueda ser reconocida y eliminada por CTLs. La unión entre el receptor de células T (TCR) de las CTLs y el complejo HLA-péptido activa a estas células para que lleven a cabo la lisis de las células infectadas. Así, las moléculas HLA encargadas de presentar antígenos, juegan un papel central en la respuesta inmune adquirida [13].

## Las moléculas HLA

Existen tres tipos de HLAs que se diferencian en estructura, función y en el tipo de células en que se presentan.

El HLA de clase I (HLA-I) es una proteína heterodimérica que consta de una cadena  $\alpha$  con tres dominios llamados  $\alpha_1$ ,  $\alpha_2$  y  $\alpha_3$  y una molécula de  $\beta_2$ -microglobulina. La asociación de  $\alpha_1$  y  $\alpha_2$  forma el surco de unión a antígenos en la porción extramembranal de la proteína mientras que  $\alpha_3$  ancla la proteína a la membrana. La molécula de  $\beta_2$ -microglobulina está unida a la cadena  $\alpha$  de manera no covalente (Fig. 1.3.1). Las moléculas HLA-I pueden presentar únicamente antígenos de entre ocho y diez aminoácidos derivados de patógenos intracelulares (virus, por ejemplo) que se encuentren en el citosol o antígenos autólogos [13].

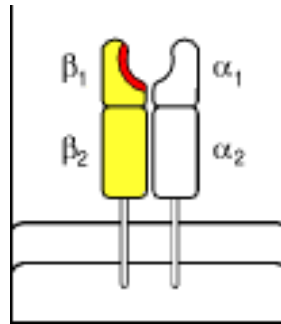


**Figura 1.3.1. Esquema de la estructura de la molécula HLA-I.** La molécula HLA-I consta de una cadena heterodimérica constituida por los dominios  $\alpha_1$ ,  $\alpha_2$  y  $\alpha_3$  y una molécula de  $\beta_2$ -microglobulina unida a la cadena  $\alpha$  de manera no covalente. Los dominios  $\alpha_1$  y  $\alpha_2$  forman el surco de unión a péptidos (rojo) en la porción extramembranal de la molécula. El surco de unión a péptidos es sumamente variable en su composición de aminoácidos y en su estructura [tomado de 13].

Las moléculas HLA-I se almacenan en el retículo endoplásmico (RE) y se exponen en la membrana hasta que se unen covalentemente a un antígeno. Los antígenos que presentan estas moléculas derivan de la ubiquitinación de proteínas de un patógeno intracelular (o propias) que conduce a las mismas a una estructura proteolítica llamada proteasoma. El proteasoma es un complejo multienzimático que genera pequeños péptidos que posteriormente, se transportan al RE y se unen al HLA-I para finalmente, exponerse en la membrana celular [13].

Los HLA-I se expresan en prácticamente todas las células nucleadas e interactúan con linfocitos TCD8+, permitiendo que éstos lleven a cabo su actividad citotóxica sobre células infectadas con patógenos intracelulares o que presentan proteínas autólogas disfuncionales [13]. Es importante enfatizar que las células son incapaces de reconocer antígenos libres o solubles ya que la molécula HLA forma parte del ligando de los receptores específicos de estas células. Hay células que sólo reconocen antígenos ligados a un alelo HLA. A este fenómeno se le denomina “restricción por HLA” [14, 15].

El HLA de clase II (HLA-II) está constituido por dos glicoproteínas transmembranales, una  $\alpha$  y otra  $\beta$ . Cada glicoproteína posee dos dominios,  $\alpha_1$  y  $\alpha_2$  y  $\beta_1$  y  $\beta_2$  respectivamente. La unión de  $\alpha_2$  y  $\beta_2$  forma el surco de unión a antígenos y se ubica en la porción extramembranal mientras que  $\alpha_1$  y  $\beta_1$  anclan al complejo proteico en la membrana (Fig. 1.3.2) [13].



**Figura 1.3.2. Esquema de la estructura de la molécula HLA-II.** La molécula HLA-II está formada por glicoproteínas transmembranales:  $\alpha$  y  $\beta$ . Cada glicoproteína consta de dos dominios  $\alpha_1$  y  $\alpha_2$  y  $\beta_1$  y  $\beta_2$  respectivamente. Los dominios  $\alpha_1$  y  $\beta_1$  forman el surco de unión a péptidos (rojo) mientras que  $\alpha_2$  y  $\beta_2$  anclan la molécula a la membrana celular [tomado de 13].

Las moléculas HLA-II presentan péptidos de mínimo trece aminoácidos y se procesan a partir de la endocitosis de patógenos extracelulares que se almacenan en vesículas una vez internalizados, se degradan dentro de éstas y se generan péptidos que se importan al RE, donde se unen al HLA-II para expresarse en la membrana celular. Sólo las células con capacidad endocítica pueden sintetizar antígenos de esta manera, por esta razón, sólo las células presentadoras de antígeno profesionales (APCs) como los macrófagos y las células dendríticas, además de algunos linfocitos B con capacidad endocítica, expresan moléculas HLA-II en su membrana [13].

Los genes HLA-III no codifican estructuras proteicas que se expresan en la membrana celular sino proteínas inmunitarias secretadas para desempeñar diversas funciones inmunológicas como lo son, elementos del sistema del complemento y citoquinas proinflamatorias [13].

A la combinación particular de alelos HLA ubicados en un cromosoma se le denomina haplotipo [13].

Se dice que la región HLA es poligénica y altamente polimórfica. Como consecuencia de estas características, los HLAs pueden presentar una enorme variedad de péptidos. Su variabilidad deriva de eventos de recombinación (durante la meiosis), acumulación de mutaciones puntuales y de procesos de conversión génica, en los cuales se substituye un fragmento de un gen por un fragmento de otro gen no homólogo [13].

Existen tres genes que codifican para moléculas HLA-I clásicas, capaces de presentar antígenos en la superficie de células infectadas: HLA-A, HLA-B y HLA-C, donde A, B y C representan 3 genes que codifican la cadena  $\alpha$ . Los alelos HLA-II se subdividen en HLA-DP, HLA-DQ y HLA-DR según los genes que codifican las cadenas  $\alpha$  y  $\beta$  [13].

Los genes A, B y C del HLA-I, son altamente polimórficos y se han identificado más de 1000 proteínas HLA distintas en los seres humanos, con propiedades altamente variables para unir péptidos antigénicos. Además, su expresión es codominante, por lo que cada ser humano puede expresar seis tipos de moléculas HLA-I en la membrana de sus células nucleadas (una molécula A, una B y una C de cada padre). Dado el alto polimorfismo de los genes HLA, lo más común es que los individuos sean heterocigos



con relación a sus alelos HLA, lo que resulta en una mayor capacidad de presentar péptidos y de manera general, en una mejor respuesta inmune [13].

Dentro de los loci del HLA-I, el alelo B es el más polimórfico de los tres con 2490 alelos identificados, mientras que los loci A y C presentan 1884 y 1384 respectivamente [23]. Estas cifras pueden generar más de 6.4 millones de combinaciones posibles, aunque se sabe que hay alelos que tienden a aparearse con mayor frecuencia entre sí, por lo que no es un proceso meramente aleatorio. A este fenómeno de apareamiento no aleatorio se le denomina “desequilibrio de enlace” [13].

La distribución geográfica de los alelos HLA-I en las poblaciones humanas, es muy heterogénea y varios estudios coinciden en que esto se explica principalmente por efectos fundador [13, 22]. También, se ha determinado que los alelos HLA-I están sujetos a las leyes de la evolución Darwiniana debido a que se han podido establecer asociaciones entre la susceptibilidad o el control de ciertas enfermedades o infecciones y la presencia o ausencia de alelos particulares HLA [13, 15, 15, 16, 18, 19, 22, 26]. Pequeñas variaciones en la composición de aminoácidos de una molécula HLA puede tener efectos funcionales significativos sobre la aptitud del sistema inmunológico para controlar o sucumbir ante la infección de un patógeno [14, 15, 18, 26]. De manera general, la alta variabilidad de las moléculas HLA se interpreta como efecto de una selección de tipo balanceadora, es decir, que favorece la heterocigocidad (y la variabilidad en general) ante la homocigocidad ya que un individuo heterocigo puede presentar un mayor número de antígenos que un individuo homocigo y por lo tanto, tiene más probabilidad de controlar o erradicar diferentes enfermedades e infecciones [13, 22].

De estos hechos deriva que haya poblaciones en las que ciertos alelos sean más frecuentes que otros y que por lo tanto, los individuos que presentan determinados alelos, exhiban una mayor o menor susceptibilidad a diferentes enfermedades o un mayor o menor control de ciertas infecciones [22, 26].

## **Impacto de las moléculas HLA y el gen *gag* en la evolución del VIH a nivel poblacional**

En el contexto del VIH, se ha podido constatar que la respuesta de CTLs es necesaria para controlar la replicación viral y que existe una relación clara entre determinadas moléculas HLA, principalmente HLA-B, y un mejor o menor control de la replicación viral [14, 15, 16, 18, 19, 26].

Dos moléculas similares HLA-B pueden presentar epítomos distintos y generar respuestas inmunes contrarias, ya que pequeñas diferencias en la estructura de estas moléculas tienen repercusiones importantes en los péptidos capaces de unirse a ellas y en la efectividad de la respuesta inmune que se monte. Por ejemplo, el alelo HLA-B\*5801 se relaciona con cargas virales bajas, producto de un buen control de la replicación viral, mientras que el alelo HLA-B\*5802 se relaciona con cargas virales altas, reflejo de un control deficiente de la replicación del virus. Estos alelos, con respuestas contrarias difieren únicamente en tres aminoácidos en el surco de unión a péptidos [14]. La tabla 1.3.1 muestra alelos que se relacionan de manera importante con la progresión de la infección por VIH.

Tabla 1.3.1 Alelos MHC-I que se relacionan de manera importante con la progresión de la infección por VIH. [Modificado de 14]

Alelos de MHC-I	Progresión	Set-point viral	Epítipo implicado	Proteína viral del VIH
<b>HLA-B*1302</b>	Lenta	Bajo	RQANFLGKI	Gag 429-437
<b>HLA-B*1801</b>	Rápida	Alta	-	-
<b>HLA-B*2705</b>	Lenta	Bajo	KRWIILGLNK	Gag 263-272
<b>HLA-B*3502 y HLA-B*3503</b>	Rápida	Alto	-	-
<b>HLA-B*5101</b>	Lenta	Bajo	TAFTIPSI	RT 128-135
<b>HLA-B*5701, HLA B*5702 y HLA-B*5703</b>	Lenta	Bajo	TSTLQEQIAW KAFSPEVIPMF ISPRTLNAW IVLPEKDSW	Gag 240-249 Gag 162-172 Gag 147-155 RT 235- 242
<b>HLA- B*5801</b>	Lenta	Bajo	TSTLQEQIAW ISPRTLNAW	Gag 240-249 Gag 147-155
<b>HLA- B*5802</b>	Rápida	Alto	QTRVLAIERYL	gp41 577-587
<b>HLA- B*8101</b>	Lenta	Bajo	-	-

Numerosos estudios [14, 15, 16, 18, 19] sugieren que los alelos HLA-B que presentan epítomos de las proteínas virales codificadas por el gen *gag* del VIH, son los más efectivos en el control de la replicación viral. Dentro de estos alelos destacan HLA-B\*27 Y HLA-B\*57 (alelos protectores) [14, 15, 16, 18, 19]. Algunas de las razones por las cuales las respuestas generadas por epítomos de Gag son sumamente eficientes son, que Gag es altamente inmunogénico, la secuencia del gen *gag* es muy conservada dentro del genoma viral y que la proteína de cápside p24 que codifica *gag*, se encuentra en grandes cantidades dentro de la célula y no requiere ser sintetizada *de novo* como muchas otras proteínas virales pudiendo presentarse antígenos derivados de esta proteína poco tiempo después de la infección. Este último hecho resulta de fundamental importancia ya que una de las funciones de la proteína viral accesoria Nef, es disminuir la expresión de moléculas HLA-I en la superficie celular para rezagar la respuesta de las CTLs contra las células infectadas [14, 15, 16, 18, 19, 20, 26].

Los epítomos de Gag mejor caracterizados en individuos que poseen los alelos protectores HLA-B\*57 y HLA-B\*5801 son TW10 (TSTLQEIQGW) que corresponde a las posiciones 240-249 del gen *gag* y se presenta principalmente en la fase aguda, KF11 (KAFSPEVIPMF, *gag* 162-172) y KK10 (KRWIILGLN, *gag* 263-272), ambos se presentan en la etapa crónica. Dentro de estos epítomos se han registrado mutaciones virales que le permiten al virus evadir la respuesta inmune [15]. A este tipo de mutaciones se les denomina mutaciones de escape y se dice que los HLAs que presentan epítomos que generan una respuesta inmune efectiva representan una presión selectiva para el virus (alelos selectores) [14].

Las mutaciones de escape del virus ante la presión selectiva de las CTLs, comúnmente evitan la unión del péptido al surco del HLA, evitan el procesamiento intracelular del péptido, de modo que éste no se pueda presentar por el HLA en la mem-

brana celular o provocan una disminución en la afinidad entre el receptor de las células T (TCR) y el complejo HLA-péptido para evitar la activación de las CTLs [14, 15, 20].

En la mayoría de los casos, las mutaciones de escape relacionadas con los epítopos de Gag derivan en una pérdida de *fitness* viral (capacidad replicativa relativa de un virus bajo condiciones particulares). Ante la pérdida de *fitness* viral y durante el resto de la infección, hay una acumulación paulatina de mutaciones compensatorias destinadas a restaurar poco a poco la capacidad replicativa del virus [14, 15, 16, 18].

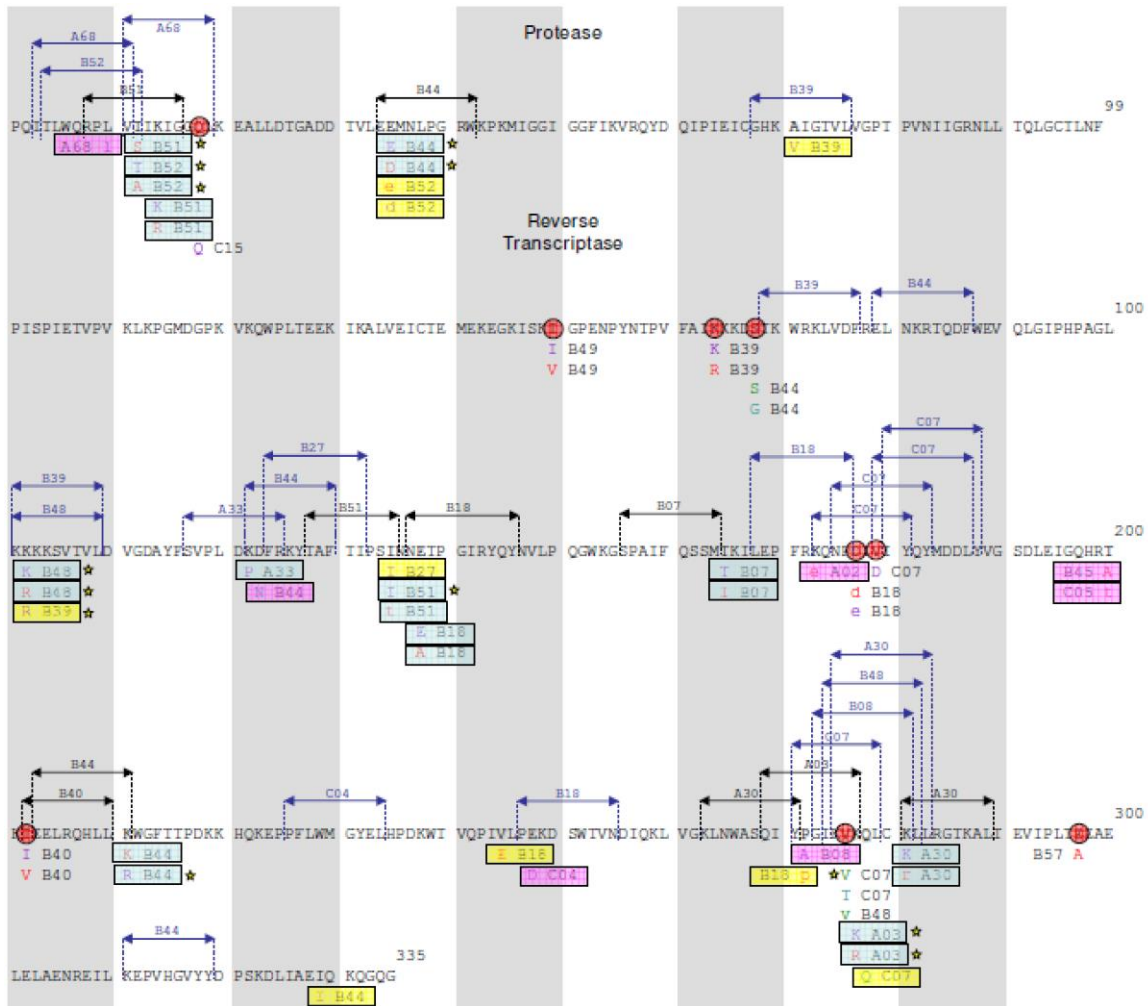
Con relación al efecto de la transmisión de variantes virales con mutaciones de escape y su efecto a nivel poblacional, se ha observado que en un número importante de casos, la transmisión de virus con mutaciones de escape inmune a individuos carentes del alelo HLA selector, conlleva a la reversión de estas mutaciones adaptativas y a la aparición de mutaciones de escape específicas de los alelos HLA del nuevo hospedero. Sin embargo, en algunas ocasiones, mutaciones de escape con bajos costos en el *fitness* viral pueden transmitirse y mantenerse en un nuevo hospedero aún sin la presión selectiva de las moléculas HLA. Dependiendo de las frecuencias de las moléculas HLA de la población, el virus circulante en una región puede acumular mutaciones adaptativas. Estas mutaciones incluso pueden llegar a fijarse en la mayoría de los virus circulantes y convertirse en el consenso poblacional. Esto traería como consecuencia la extinción de un epítipo, como resultado de la adaptación viral a la presión selectiva ejercida por los CTLs, lo cual tiene implicaciones importantes en el diseño de posibles vacunas contra el VIH [15, 16].

La evolución del VIH y la influencia del perfil inmunogenético de una población en dicho proceso, son fenómenos complejos puesto que actúan diferentes fuerzas evolutivas como la mutación (mutaciones de escape y compensatorias resultantes de la presión selectiva que ejercen diferentes respuestas inmunes), la deriva génica, la aparición de múltiples cuellos de botella y los efectos fundador (al momento de la transmisión por ejemplo). Sin embargo, varios estudios han determinado que la evolución dependiente de moléculas HLA del VIH a nivel poblacional, es un fenómeno altamente específico y predecible. Ante la presencia de determinadas moléculas HLA-I, el genoma viral normalmente sigue patrones mutacionales predecibles, ya que el efecto de las moléculas HLA-I presentes, se ve reflejado en la aparición de ciertos polimorfismos en regiones del genoma viral involucradas con la generación de antígenos durante la infección. Esta correlación entre moléculas HLA-I y polimorfismos determinados en el genoma viral permiten evaluar y predecir, en cierta medida, el desarrollo de la epidemia en una población dada y de la cual se conoce la composición inmunogenética. Más aún, la distribución heterogénea de las moléculas HLA en el mundo, le confieren características particulares a las diferentes poblaciones humanas en cuanto a las respuestas que los individuos de dicha población puedan generar ante la infección por VIH, dirigiendo consecuentemente la evolución poblacional del virus de manera específica en diferentes contextos inmunogenéticos [14, 15, 16, 18, 19, 22, 26].

## **Evolución del VIH en las poblaciones mexicana y guatemalteca**

La gran mayoría de los estudios evolutivos del VIH y del impacto de las moléculas HLA a nivel poblacional se han realizado en poblaciones caucásicas y en distintas poblaciones de África. Las poblaciones latinoamericanas que conservan características genéticas amerindias se han estudiado poco hasta ahora.

La población mexicana se caracteriza por tener un perfil inmunogenético mixto que conjuga principalmente elementos amerindios y caucásicos. Estudios antropológicos han reportado repetidas migraciones provenientes de África en las regiones de Oaxaca, Guerrero y Veracruz además de asentamientos recientes de individuos de origen asiático sobre todo, en Sinaloa y en menor proporción Sonora y Nayarit [24, 25]. La población mexicana es por lo tanto, sumamente heterogénea en cuanto a su perfil inmunogenético. Recientemente, nuestro grupo de investigación describió características únicas de la evolución del VIH en el contexto inmunogenético único de la población mexicana [26]. Este estudio acerca del impacto de las moléculas HLA presentes en la población mexicana reveló la existencia de 43 nuevas asociaciones entre moléculas HLA y mutaciones en el genoma viral. Veintitrés de las 43 asociaciones diferían consistentemente de las asociaciones reportadas en otras poblaciones, lo que sugiere patrones evolutivos únicos en la población mexicana. En este estudio se tipificaron los alelos HLA (haplotipos) que presentaban 280 individuos del centro y sur de México y se buscaron correlaciones entre la presencia de éstos alelos y polimorfismos en el gen *pol* del VIH (Fig. 1.3.3) [26].



**Figura 1.3.3. Mapa de asociaciones entre moléculas HLA y polimorfismos en las secuencias del gen *pol* de VIH en una cohorte mexicana.** En el mapa se indican las secuencias de aminoácidos del gen *pol* que codifican para la PR (1-99) y la RT (1-335). Las flechas bidireccionales encima de las secuencias, enmarcan el epítoto que reconocen moléculas HLA específicas. Los epítotos se obtuvieron de manera experimental, por medio de pruebas serológicas (enmarcados por flechas bidireccionales negras) y otros epítotos y su unión a moléculas específicas HLA se predijo a partir de los motivos de unión que presentarían ambas estructuras (flechas bidireccionales azules). Las flechas bidireccionales por encima de las secuencias, enmarcan el epítoto que moléculas HLA específicas reconocen. En las cajas de colores debajo de las secuencias se indica el aminoácido blanco, el que determina el resultado de la asociación entre una molécula HLA y el epítoto en la secuencia. Si el aminoácido está escrito en morado, indica escape (tener un alelo determinado HLA hace menos probable que se presente el aminoácido blanco), rojo implica atracción (tener un alelo determinado HLA hace más probable que se presente el aminoácido blanco), verde reversión (no tener un alelo determinado HLA hace más probable que se presente el aminoácido blanco) y azul repulsión (no tener un alelo determinado HLA hace más probable que se presente el aminoácido blanco). Los alelos en rectángulos de colores y los epítotos que reconocen representan asociaciones que comparten la población mexicana y otras poblaciones caucásicas sobre todo. Los aminoácidos rojos en círculos son posiciones que se relacionan con HLA's y exclusivas de la cohorte mexicana [Tomado de 25].

Al igual que la población mexicana, la población guatemalteca es sumamente heterogénea. La población mestiza representa el 51% y la indígena el 41% y está distribuida en 23 grupos étnicos, predominantemente ligados a la cultura Maya. El 9% restante está constituido por individuos de origen caucásico y asiático principalmente [5, 25]. Documentos históricos revelan que poblaciones del sur del continente (particularmente poblaciones asentadas en el delta del Orinoco) y esclavos provenientes de África asentados en Las Antillas invadieron territorios guatemaltecos, por lo que el grado de mestizaje en este país es sumamente elevado [24, 25].

Actualmente, el Centro de Investigación en Enfermedades Infecciosas (CIENI) lleva a cabo un proyecto que incluye a todos los países de América Central y tiene como objetivo caracterizar la epidemia del VIH en dichos países, considerando el perfil inmunogenético predominante en las poblaciones de origen amerindio.



## Capítulo 2

# Planteamiento del problema

Considerando la distribución heterogénea de las moléculas HLA en el mundo y el impacto de éstas en la dinámica evolutiva del VIH dentro de una población, resulta muy interesante investigar cómo se relacionan las moléculas HLA-I presentes en las poblaciones de origen amerindio y los virus circulantes en estas poblaciones poco caracterizadas hasta el día de hoy. Si bien no se conoce con detalle el perfil inmunogenético de las poblaciones latinoamericanas, en general, se sabe que conjugan elementos tanto de origen caucásico como de origen amerindio, africano y asiático en muchas regiones. Por esta razón, es probable que las poblaciones de origen amerindio, al haberse expuesto a múltiples eventos de mestizaje a lo largo de su historia, exhiban relaciones únicas, diferentes de las que se han observado en poblaciones ampliamente estudiadas (poblaciones de África y de origen caucásico) con respecto a la presión selectiva que ejercen las moléculas HLA-I sobre el VIH durante la infección a nivel poblacional.

Debido a que diversos estudios han demostrado que los alelos HLA-I están estrechamente relacionados con una respuesta de CTLs efectiva y una consecuente progresión lenta a la etapa de SIDA, sobre todo aquellos con la capacidad de presentar epítomos de las proteínas codificadas por el gen *gag*; este estudio se enfoca en las posibles relaciones entre los alelos HLA-I que presentan las poblaciones de México y Guatemala y los polimorfismos detectables en la secuencia de aminoácidos en el gen *gag* de individuos infectados con VIH de ambos países.

De manera general, la intención de este estudio es determinar cuál es la relación filogenética entre las secuencias del gen *gag* de VIH que circula en las poblaciones mexicana y guatemalteca, en qué población es más variable el gen *gag* y si esta variabilidad es consecuencia directa de una presión selectiva ejercida por moléculas HLA-I y finalmente, buscar correlaciones entre las posiciones polimórficas (bajo presión selectiva) y alelos HLA-I característicos de cada población, que pueden estar ligados a un mejor control de la replicación viral, a una progresión acelerada a la etapa de SIDA o simplemente representar asociaciones novedosas, no reportadas hasta la fecha.





# Capítulo 3

## Objetivos

### 3.1. Objetivo general

Evaluar y comparar la evolución mediada por HLA en el gen *gag* del VIH, en una cohorte de pacientes mexicanos y una cohorte de pacientes guatemaltecos.

### 3.2. Objetivos particulares

- Purificar RNA viral a partir del plasma de individuos VIH positivos de México y Guatemala.
- Retrotranscribir el RNA viral a DNA y amplificar el gen *gag* viral por reacciones en cadena de la polimerasa (PCRs).
- Purificar los productos de PCR y secuenciar el DNA.
- Comparar las secuencias del gen del VIH de los individuos de México y Guatemala para realizar análisis filogenéticos y de variabilidad para posteriormente, buscar posibles correlaciones entre polimorfismos en las secuencias virales de aminoácidos y la expresión de moléculas HLA-I.



# Capítulo 4

## Hipótesis

Dadas las diferencias en los componentes genéticos y los patrones de mestizaje de las poblaciones mexicana y guatemalteca, es factible encontrar perfiles inmunogenéticos distintos en cada una de las poblaciones, los cuales impactarán de manera distinta la evolución mediada por HLA a nivel poblacional en el VIH circulante en cada región. Esta adaptación diferencial del virus a la respuesta inmune en dos poblaciones inmunogenéticamente distintas, tendrá implicaciones importantes en la investigación de vacunas, el diseño de posibles inmunógenos y podría ayudar a identificar factores genéticos de riesgo o protección que influyen en la progresión de la enfermedad por VIH, que no se han identificado en otras poblaciones.



# Capítulo 5

## Metodología

### 5.1. Amplificación del gen *gag* de VIH por los métodos de RT- PCR anidada y PCR anidada

Se tomaron muestras de individuos de México y Guatemala sin ningún método particular de muestreo y se aisló el RNA viral del plasma de la sangre de los individuos involucrados en este trabajo de investigación con el kit “RNeasy Mini Kit” (Invitrogen-Life Technologies Carlsbad, CA).

El RNA correspondiente al gen *gag* de VIH-1, se retrotranscribió a DNA y se amplificó por una reacción de RT-PCR (reacción en cadena de la polimerasa con transcripción inversa) anidada, con primers diseñados en el laboratorio, según el protocolo de la tabla 5.1.

*Tabla 5.1 Protocolo para RT-PCR para el gen gag de VIH. Fi 623 es un primer en dirección 5' a 3' y 2cRx es un primer en dirección 3' a 5', ambos son específicos para la amplificación del gen gag. SuperScript™ III One-Step RT-PCR System with Platinum III Taq Pol (Life Technologies Carlsbad, CA) es el kit que se utilizó para la amplificación del gen gag de VIH.*

Reactivos	1x	Concentración de cada reactivo en la mezcla final
Agua	8.5µl	-
Buffer (2x)	12.5µl	2.5 X
Fi 623 (20 pmol/µl (primer fw)	0.5µl	0.4 µMol
2cRx (20 pmol/µl (primer rev)	0.5µl	0.4 µMol
Superscript III RT/Platinum Taq Pol	0.5µl	2.5 X
RNA viral (10 nmol-40 nmol)	2.5µl	-
<b>Total</b>	<b>25 µl</b>	-

\*Las “X” representan unidades preestablecidas en el kit SuperScript™ III One-Step RT-PCR System with Platinum III Taq Pol (Life Technologies Carlsbad, CA). La unidad del Buffer (2X) equivale a 0.5µL y la unidad de la enzima Superscript III RT/Platinum Taq Pol equivale a un volumen de 0.2 µl.

Protocolo de amplificación:

- 1) 1 ciclo:  
55°C x 30 minutos
- 2) 40 ciclos:  
94°C x 15 segundos  
55°C x 30 segundos  
68°C x 30 segundos
- 3) 1 ciclo:  
68°C x 5 minutos y 4° C ∞.

Posteriormente, se tomó el DNA resultante de la RT- PCR y se amplificó por una reacción de PCR anidada siguiendo el protocolo de la tabla 5.2:

*Tabla 5.2 Protocolo para PCR para el gen gag de VIH. PFG-1 es el primer específico para gag en sentido 5' -3' y PFG-10 en sentido 3' - 5'. Platinum Taq Polymerase (invitrogen- Life Technologies Carlsbad, CA) es la polimerasa termoestable que se usó en este ensayo.*

<b>Reactivos</b>	<b>1x</b>	<b>Concentración de cada reactivo en la mezcla final</b>
<b>Agua</b>	13.75µL	-
<b>Buffer (10x)</b>	2.5µL	5X
<b>MgCl (50 µMol)</b>	1µL	0.2 µM
<b>dNTPs (1.25 µMol)</b>	4µL	0.2 µM
<b>PFG1 (20 µMol)</b>	1µL	0.8 µM
<b>PRG10 (2 µMol)</b>	1µL	0.8 µM
<b>TaqPol (invitrogen)</b>	0.25µL	1.25 X
<b>DNA viral (10 nmol-40 nmol)</b>	2.5µL	-
<b>Total</b>	26 µL	-

\* La unidad (X) del Buffer (10X) equivale a 0.5µL y la unidad de la enzima Platinum Taq Polymerase (invitrogen- Life Technologies Carlsbad, CA) equivale a un volumen de 0.2 µL.

Protocolo de amplificación:

- 1) 1 ciclo:  
94°C x 15 segundos.
- 2) 40 ciclos:  
94°C x 2 minutos  
56°C x 30 segundos.  
72°C x 2 minutos
- 3) 1 ciclo:  
72°C x 5 minutos y 4° C ∞.

## 5.2. Electroforesis en gel de agarosa

Se preparó un gel de agarosa al 1% con buffer TBE (100mL) y 5 $\mu$ L de bromuro de etidio (esta sustancia es un agente intercalante en DNA que permite visualizar los productos de PCR cuando se exponen a luz UV).

Se llenaron los pozos con las muestras y con el marcador de peso molecular (1Kb DNA Ladder, Invitrogen) con 3 $\mu$ L de buffer de carga y 3 $\mu$ L de producto de PCR (total de 6 $\mu$ L). Se conectó la cámara de electroforesis a un generador de corriente eléctrica siguiendo las siguientes condiciones: 100V, 400 A, 30 min.

Una vez transcurrido el tiempo, se colocó el gel en un transiluminador, dónde se pudo observar los resultados de la amplificación.

Los productos de PCR se cuantificarán espectrofotométricamente utilizando un instrumento Nanodrop (Thermo Scientific, Wilmington, DE).

## 5.3. Secuenciación y análisis de genes virales

Se secuenció el gen *gag* de cada paciente utilizando 8 primers que permiten obtener al menos dos secuencias para cada parte de la región de interés, utilizando la química BigDye Terminator v3.1 (Life Technologies, Carlsbad, CA), en un secuenciador capilar ABI Prism 33730xl Genetic Analyzer (Life Technologies, Carlsbad, CA). Los primers que se utilizaron en este proceso son:

G2F	5' - GCGGCGACTGGTGAGTA - 3'	<i>gag</i> : 734- 750
GS1R	5' - TTATCTAAAGCTTCCTTGGTGTCT - 3'	<i>gag</i> : 2313- 2294
GAS3F	5' - CATCAATGAGGAAGCTGCAG - 3'	<i>gag</i> : 1401-1420
GAS4R	5' - GGTTCCTCATCTGGCCTGG - 3'	<i>gag</i> : 1481- 1461
GAS5F	5' - CTCTAAGAGCCGAGCAAGCT -3'	<i>gag</i> : 1696-1716
GAS6R	5' - AAAATAGTCTTACAATCTGG 3'	<i>gag</i> : 1790- 1771
HPR1977F	5' - GTTAAGTGTTC AATTGTGG - 3'	<i>gag</i> : 1957- 1977
GA22774R	5' - TCTTTATTGTGACGAGGGGTCG - 3'	<i>gag</i> : 2296- 2274

Las letras R y F al final del nombre de cada primer indica la dirección del mismo (forward o reverse).

Las secuencias exitosas se ensamblarán con el software "Geneious".

Las secuencias se alinearon utilizando el software MEGA 5.05 [27, 28] por el método de Clustal w [27, 28]. A partir de este alineamiento se construyeron árboles filogenéticos. Además, se realizaron estudios de variabilidad cuantificando manualmente el número de sustituciones de aminoácido en cada posición de las secuencias problema con respecto a una secuencia de referencia. Se obtuvieron secuencias consenso del gen *gag* de los virus circulantes en México y en Guatemala, registrando los aminoácidos mayoritarios en cada posición del gen, con la herramienta "Consensus Maker" disponible en la barra de herramientas de la Base de datos de Los Alamos [29] y se comparó la variabilidad de las secuencias en ambas poblaciones. Además, para estudiar más a fondo la variabilidad de las secuencias en ambas poblaciones se realizaron



análisis de entropía de Shannon [30] utilizando las herramientas “Entropy I” y “Entropy II” de la Base de Datos de datos de Los Alamos [29].

Adicionalmente, se realizaron estudios de presión selectiva por tasas de sustituciones sinónimas y no sinónimas, utilizando la herramienta en línea Data Monkey [31, 32]. Se buscaron correlaciones entre la presencia de alelos HLA y polimorfismos en la secuencia viral en ambas poblaciones mediante el uso de mapas de epítomos comprobados de linfocitos T citotóxicos disponibles en la Base de Datos de Inmunología de Los Alamos [29]. La tipificación de alelos HLA se realiza rutinariamente en los laboratorios del CIENI del INER mediante secuenciación y la información incluida en este estudio se obtuvo a partir de bases de datos disponibles en el laboratorio.

### Construcción de árboles filogenéticos

Se construyeron árboles filogenéticos para la población mexicana y guatemalteca por separado y un árbol que incluía a ambas con el software MEGA 5.05 [27, 28] por el método de máxima similitud, con un bootstrap=1000 y según el modelo general tiempo reversible (GTR) de sustitución de nucleótidos tomando como base diferentes estudios acerca de la filogenia de secuencias de VIH-I [33, 34, 33] y el tutorial que ofrece La Base de Datos de Los Alamos [29] para la construcción de árboles filogenéticos con secuencias de VIH-I.

### Construcción de secuencias consenso de las poblaciones mexicana y guatemalteca

Se construyeron secuencias consenso para las poblaciones mexicana y guatemalteca con la herramienta “Consensus Maker” disponible en la Base de datos de Los Alamos [29]. Este análisis requiere que se importen todas las secuencias para obtener una secuencias consenso representativa para el set de datos que se ingresaron en el servidor. El programa tomó como aminoácido consenso aquel que se presentaba en más del 50% de las secuencias en una posición determinada [29].

### Análisis de entropía como indicador de variabilidad en el gen *gag* del VIH

Estos análisis aplican la entropía de Shannon [30] como una medida de variación e incertidumbre en un conjunto de secuencias de aminoácidos (también se puede aplicar a secuencias de nucleótidos). Este análisis considera la entropía como la cantidad de información promedio contenida en un sistema en el cual los símbolos o las combinaciones de éstos con menor probabilidad son los que aportan mayor información [29, 30].

#### **-Entropy I-**

Nos permite evaluar el grado de incertidumbre y aleatoriedad en cada posición de un set de datos, e inferir el grado de variación en diferentes posiciones o regiones de un gen o proteína [29, 30].

### -Entropy II-

Este análisis compara la aleatoriedad en un set de datos contra la aleatoriedad de otro set de datos (en este caso México y Guatemala). Las posiciones en las cuales hay una diferencia significativa en la variabilidad de un set con relación al otro, se destacan en rojo [29, 30].

### Inferencia de posiciones bajo presión selectiva positiva con base en la tasa de sustituciones sinónimas y no sinónimas de aminoácidos por medio del servidor Data Monkey

Data Monkey [31, 32] es un servidor gratuito que identifica posiciones en un alineamiento que probablemente están bajo presión selectiva positiva y negativa. Lo hace según el modelo de SLAC que consta básicamente de cuatro pasos:

- 1) Ajusta un modelo de máximo similitud a los datos para determinar la longitud de las ramas y la tasa de sustituciones en las secuencias de nucleótidos para hacer una pequeña reconstrucción filogenética de los datos [31, 32].
- 2) Una vez que establece la relación filogenética de las secuencias, mantiene la longitud de las ramas y la tasa de sustituciones en las secuencias para después ajustar un modelo que usa codones como unidad. Con este modelo se estima la tasa de sustituciones no sinónimas/ sinónimas ( $\omega = dN/dS$ , donde  $dN$ = sustituciones no sinónimas y  $dS$ =sustituciones sinónimas) [31, 32].
- 3) Con los parámetros que se obtienen de los pasos uno y dos, el servidor construye secuencias ancestrales de codones en cada posición del alineamiento por medio de análisis de máxima verosimilitud, de manera que se maximicen las similitudes entre las secuencias y el que se supone, es su estado ancestral [31, 32].
- 4) Finalmente, con la información de los pasos anteriores, el programa infiere sitios que están bajo presión selectiva positiva y negativa considerando el número de posiciones con sustituciones no sinónimas y sinónimas esperadas y observadas [31, 32].

En condiciones de neutralidad, este análisis asume que la probabilidad de que una sustitución sea sinónima ( $P$ ) es igual al número de sustituciones sinónimas esperadas ( $ES$ ) dividido entre la suma de sustituciones sinónimas esperadas y no sinónimas esperadas ( $EN$ ) [31, 32].

$$P = \frac{ES}{ES + EN}$$

A partir de  $P$ , el programa estima qué tan probable es que una sustitución sea sinónima o no sinónima. De modo tal que,  $dN$  va a ser igual al número de sustituciones no sinónimas observadas ( $NN$ ) entre el número de sustituciones no sinónimas esperadas ( $dN = NN/EN$ ) y  $dS$  al número de sustituciones sinónimas esperadas ( $NS$ ) entre el número de sustituciones sinónimas observadas ( $dS = NS/ES$ ) [31, 32].

Asociación entre posiciones bajo presión selectiva determinadas por tasas de sustituciones sinónimas y no sinónimas y alelos HLA-I establecidas por el modelo de red de dependencia filogenética (PDN)<sup>[36]</sup>

El análisis de red de dependencia filogenética (PDN)<sup>[36]</sup> es un modelo probabilístico que infiere relaciones entre alelos HLA-I y polimorfismos en el genoma viral<sup>[36]</sup>. El modelo considera tanto el fenómeno del desequilibrio de enlace de moléculas HLA como la filogenia del VIH y la co-variación de codones, que se refiere a que la presencia de un determinado codón condiciona la presencia de otro codón; por ejemplo, una mutación de escape que trae una mutación compensatoria como consecuencia. De este modo, este modelo permite establecer las relaciones HLA- polimorfismo en la secuencia de VIH considerando aspectos que pueden resultar confusos y es por lo tanto, es un análisis más preciso que el análisis con el servidor Data Monkey<sup>[36]</sup>.

# Capítulo 6

## Resultados

### 6.1. Descripción de las cohortes de pacientes de México y Guatemala

Tabla 6.1. Descripción de la cohorte de pacientes de México y Guatemala.

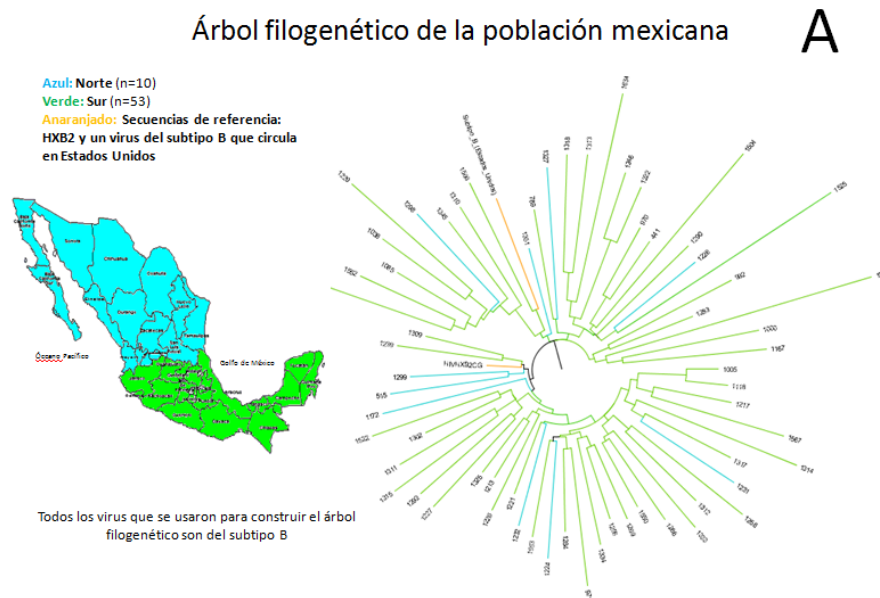
Población	Mexicana	Guatemalteca
Número total de pacientes	63	55
Género	16 mujeres 47 hombres	23 mujeres 32 hombres
Edad	<b>Mediana:</b> 27 años <b>Rango:</b> 16- 75 años	<b>Mediana:</b> 33 <b>Rango:</b> 19- 63 años
Conteo de células CD4+ (CD4+/ $\mu$ l de sangre)	<b>Mediana:</b> 485 <b>Rango:</b> 29-1, 374	<b>Mediana:</b> 289 <b>Rango:</b> 18- 1, 053
Carga viral (CV) (Número de copias de RNA viral/ mL de plasma)	<b>Mediana:</b> 30, 656 <b>Rango:</b> 348- 1, 134, 798	<b>Mediana:</b> 36, 429 <b>Rango:</b> 499- 4, 784, 974

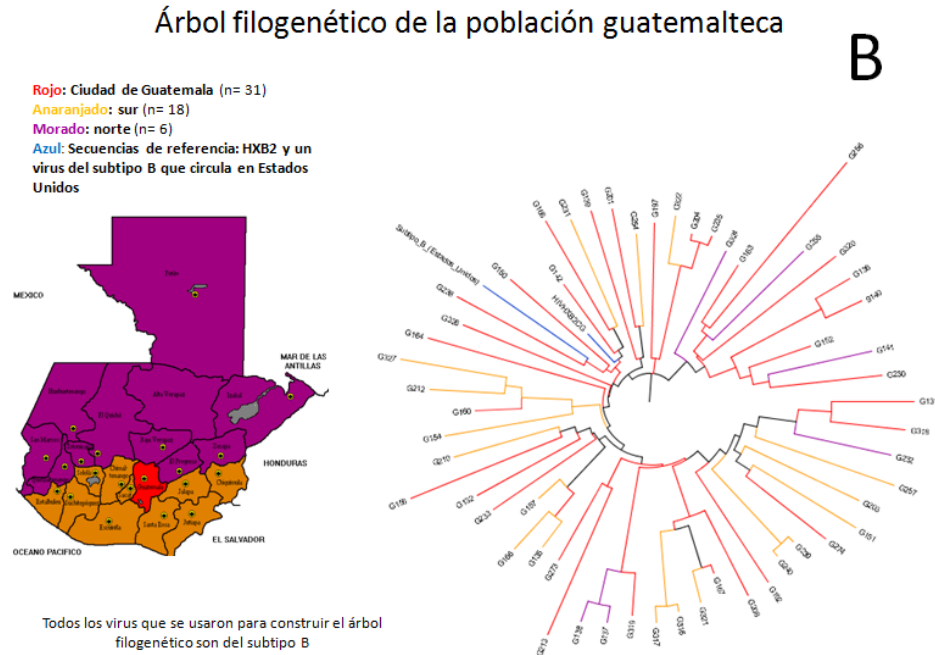
En el presente estudio se analizó un total de 118 secuencias del gen *gag* de VIH de las cuales 63 correspondían a la población mexicana y 55 a la población guatemalteca. De los 63 pacientes mexicanos, 16 eran mujeres y 47 hombres mientras que de los 55 pacientes guatemaltecos 23 eran mujeres y 32 hombres. La mediana de edad en la cohorte mexicana fue de 27 años en un rango que va desde los 16 años hasta los 75 años de edad. En la cohorte guatemalteca, la mediana de edad fue 33 años y el rango de edades va desde los 19 años hasta los 63 años de edad (Tabla 6.1).

Con relación a los parámetros clínicos de las cohortes analizadas en este estudio, la cohorte mexicana tuvo una mediana de células CD4+ de 485 células/ $\mu$ L de sangre (rango: 29-1374 células/ $\mu$ L de sangre) y una mediana de carga viral (CV) de 30 656 copias de RNA viral/ mL de plasma (rango: 348-1 134 798 copias de RNA viral/ mL de plasma). La mediana de células CD4+ en la cohorte guatemalteca fue de 289 células/ $\mu$ L de sangre (rango: 18- 1053/ $\mu$ L de sangre) y una mediana de CV de 36 429 copias de RNA viral/ mL de plasma (rango: 499-4 784 974 copias de RNA viral/ mL de plasma) (tabla 6.1).

## 6.2. Filogenia del gen *gag* de VIH circulante en las poblaciones mexicana y guatemalteca

Para explorar la diversidad de las secuencias de México y Guatemala incluidas en este estudio y para buscar evidencias de posibles eventos fundador regionales se realizaron análisis filogenéticos del gen *gag* de virus circulantes en cada región (Fig. 6.1). El objetivo principal de la construcción de árboles filogenéticos es determinar de qué manera se relacionan las secuencias de ambas poblaciones, qué tan estrecha o lejana es la distancia genética entre éstas y evaluar la probabilidad de que grupos de secuencias se hayan diversificado a través de ancestros comunes.





**Figura 6.1. Árboles filogenéticos de las poblaciones mexicana (arriba) y guatemalteca (abajo).** Los árboles se construyeron con el programa MEGA 5.05 con el modelo de máxima verosimilitud (con 1000 repeticiones de bootstrap). Se usaron las secuencias del gen *gag* del virus HXB2 (virus de referencia) y un virus del subtipo B circulante en Estados Unidos de América como referencias. Se analizaron las poblaciones mexicana y guatemalteca por separado para determinar las relaciones filogenéticas de las secuencias que se incluyeron en este estudio a nivel local. Se dividió el territorio mexicano en norte (todos los estados que se localizan al norte de los estados de Veracruz y Jalisco) y sur (todos los estados al sur de Veracruz y Jalisco) de acuerdo a estudios anteriores que identifican una mayor proporción de genes caucásicos en el norte y mayor proporción de genes amerindios en el sur [37]. En el caso de México, no se identificaron patrones de asociaciones de secuencias del norte o del sur, lo que indica que todas las secuencias incluidas en este análisis, independientemente de la región del país del que provengan, son similares y filogenéticamente cercanas entre sí. En el caso de la población guatemalteca, el territorio del país se dividió en Ciudad de Guatemala (porque 31 de las 55 muestras provenían de esta ciudad), provincia-norte (todos los departamentos que se ubican al norte de la Ciudad de Guatemala) y provincia- sur (todos los departamentos ubicados al sur de la Ciudad de Guatemala). Al igual que en la población mexicana, en las secuencias de Guatemala no se reconocen asociaciones de secuencias de una misma región, por lo que todas las secuencias usadas en el análisis son filogenéticamente cercanas entre sí.

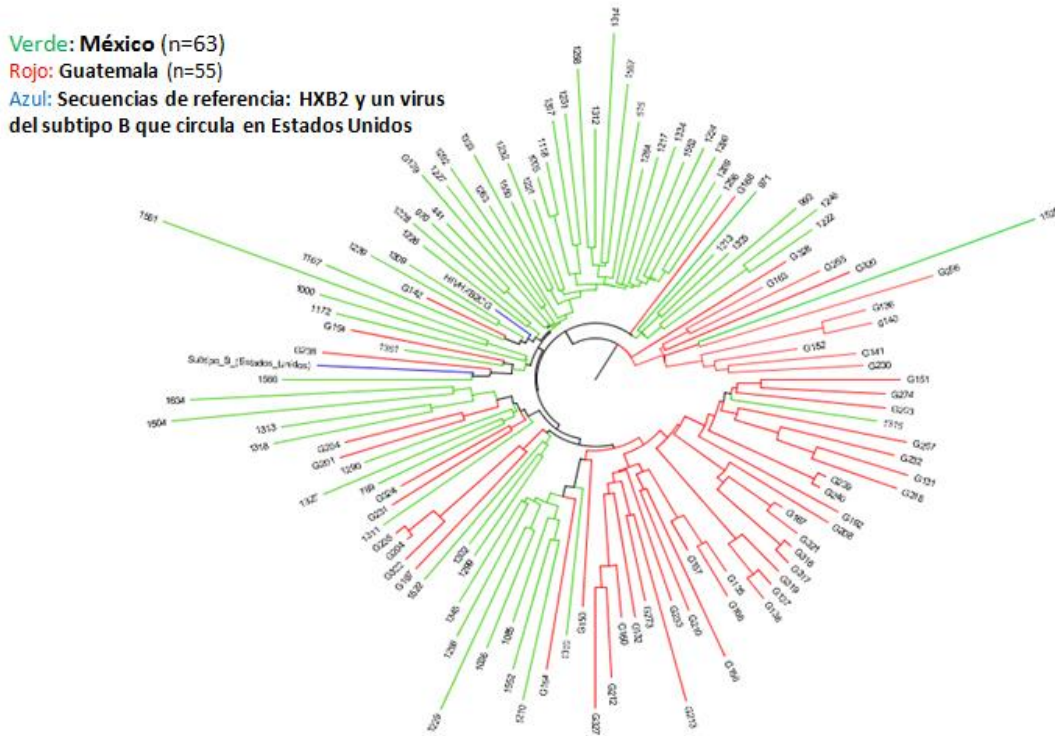
Para el caso de VIH circulante en la población mexicana, aunque en general las secuencias del norte y sur del país tendieron a entremezclarse en el árbol filogenético, se observaron agrupamientos de secuencias del norte y otros de secuencias del sur del país, lo cual habla de posibles eventos fundadores en la dispersión del virus en México, pero también de una alta homogeneidad de los virus circulantes en el país (Fig. 6.1A).

En el caso de Guatemala, no se identificaron agrupamientos geográficos claros entre los virus provenientes de departamentos del norte, departamentos del sur o de la Ciudad de Guatemala y las secuencias tendieron a entremezclarse (Fig. 6.1B). Esto sugiere que las secuencias del gen *gag* que circula en cada una de las poblaciones analizadas en este estudio son sumamente homogéneas a nivel local.

A continuación se construyó un árbol filogenético que incluía las secuencias de México y Guatemala para ver si el gen *gag* de VIH que circula en estas poblaciones y

que es sumamente homogéneo a nivel local, era diferente en cada población si se analizaban todas las muestras de manera conjunta (Fig. 6.2).

### Árbol filogenético construido con las secuencias de las poblaciones mexicana y guatemalteca



**Figura 6.2. Árbol filogenético construido con las secuencias de las poblaciones mexicana y guatemalteca.** El árbol se construyó con el programa MEGA 5.05 con el modelo de máxima similitud y un bootstrap=1000. En total se analizaron 120 secuencias con una longitud total de 1767 posiciones, de las cuales 63 eran de México, 55 de Guatemala y se usaron como referencias la secuencia del virus HXB2 (virus de referencia) y un virus del subtipo B circulante en los Estados Unidos de América. Se puede ver claramente la formación de agrupamientos bien definidos (clusters) de secuencias de Guatemala (rojo) y de México (verde). Esto indica que las secuencias de México son más parecidas entre sí que a las secuencias que circulan en Guatemala. De igual modo, las secuencias de Guatemala, son más parecidas entre sí que a las secuencias que circulan en México. La gran similitud local tanto en México como en Guatemala sugiere que la dinámica de la epidemia en ambos países es de carácter local y puede contribuir de manera importante con que los genes *gag* de VIH circulantes en México y Guatemala difieran entre sí.

Se pudo observar claramente que muchas de las secuencias de México y de Guatemala se asocian con secuencias del mismo país formando agrupamientos bien definidos (clusters). Estos clusters formados por secuencias de una misma población, indican que el gen *gag* de VIH que circula en ambas poblaciones (la mexicana y guatemalteca) exhibe un carácter local distintivo, es decir que las secuencias de México

son más parecidas entre sí que a las de Guatemala y éstas últimas, de igual forma, son más parecidas entre sí que a las de México.

### **6.3. Variabilidad del gen *gag* de VIH que circula en México y Guatemala**

Considerando que el árbol filogenético que incluía a ambas poblaciones mostró que el gen *gag* de VIH que circula en las poblaciones mexicana y guatemalteca es diferente, se construyó una secuencia consenso para cada población (Fig. 6.3 y tabla 6.2) con el objetivo de determinar si existían diferencias en las secuencias propias de cada país que se fijaron a nivel poblacional.

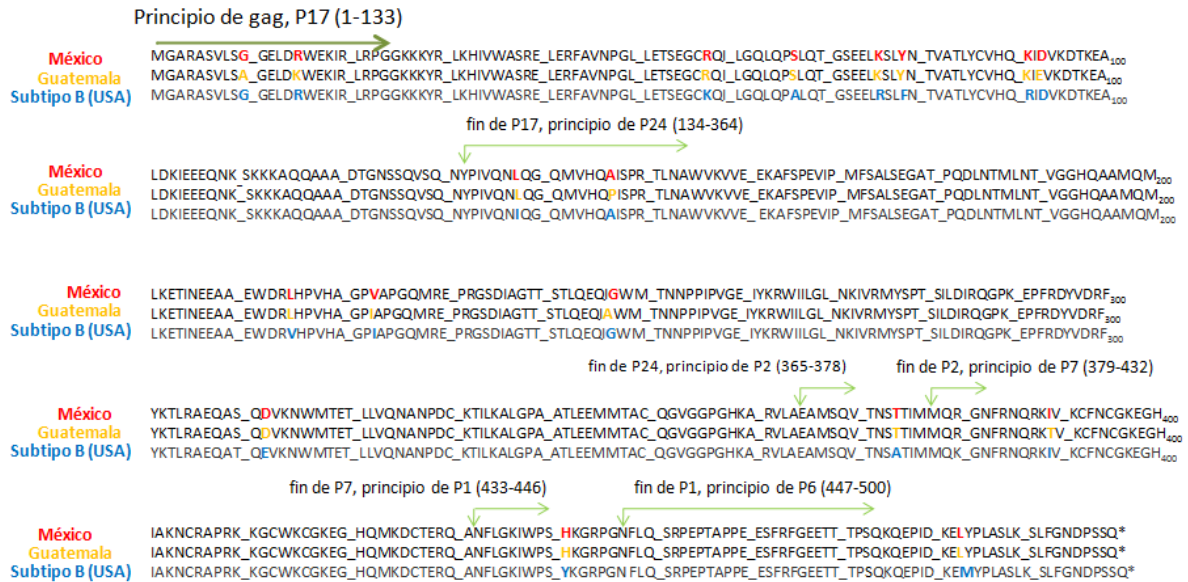
Las secuencias consenso de ambas poblaciones se compararon entre sí y con una secuencia consenso del gen *gag* del subtipo B circulante en Estados Unidos de América, con la intención de determinar qué tan similares o diferentes son las secuencias de dicho gen en las dos poblaciones mestizas y en una población predominantemente caucásica y geográficamente cercana (Fig. 6.3).

Al analizar las secuencias se determinó que las tres secuencias consenso eran distintas en 18 posiciones en total (tabla 6.2). Se observaron diferencias en las secuencias consenso de las poblaciones de México y Guatemala en siete posiciones: 3 posiciones en la porción de *gag* correspondiente a p17 (p17-*gag* 10, p17-*gag* 15 y p17-*gag* 93), tres en la porción de p24 (p24-*gag* 146, p24-*gag* 223, p24-*gag* 248) y una en el segmento que codifica a p7 (p7-*gag* 389).

La secuencia consenso de México fue diferente en 12 posiciones con respecto a la secuencia consenso de subtipo B circulante en Estados Unidos, mientras que la secuencia consenso de Guatemala presentó diferencias en 17 posiciones con respecto al consenso estadounidense. De las 7 posiciones en las cuales las secuencias consenso de México y de Guatemala difieren, 6 posiciones del consenso mexicano presentaron el mismo aminoácido que la secuencia consenso que circula en Estados Unidos. Por lo tanto, este análisis sugiere fuertemente que el gen *gag* del VIH que circula en la población Mexicana es más parecido al del virus circulante en Guatemala que al del virus circulante en Estados Unidos de América y que si comparamos las secuencias consenso de México y Guatemala con la secuencia consenso de Estados Unidos, es claro que las secuencias de México y Estados Unidos son más parecidas entre sí (12 diferencias) que las secuencias de Guatemala y Estados Unidos (17 diferencias).



Secuencias consenso del gen *gag* circulante en México y Guatemala comparadas entre sí y con una secuencia circulante en Estados Unidos



**Figura 6.3. Secuencias consenso de residuos de aminoácidos derivada de la secuencia de nucleótidos del gen *gag* circulante en México y Guatemala comparadas entre sí y con una secuencia circulante en Estados Unidos de América.** Las secuencias consenso de México (rojo), Guatemala (anaranjado) y la de un virus del subtipo B circulante en los Estados Unidos de América (azul) se construyeron con la herramienta “Consensus Maker” disponible en la base de datos de Los Alamos y se presentan alineadas. Se destaca con colores la posición en la que los aminoácidos son diferentes en alguna de las tres secuencias. Arriba de las secuencias se indica la proteína que corresponde a cada fragmento y los guiones bajos separan diez aminoácidos. Las secuencias consenso de México y Guatemala se compararon con una secuencia de referencia circulante en Estados Unidos, para ver qué tan similares o diferentes eran del VIH que circula en el Norte del Continente Americano. Se puede observar que las secuencias consenso de México y Guatemala son más similares entre sí (difieren en 7 posiciones) que a la secuencia consenso del subtipo B circulante en Estados Unidos de América (el consenso mexicano difiere del consenso circulante en Estados Unidos de América en 12 posiciones y el consenso guatemalteco difiere del consenso de Estados Unidos de América en 17 posiciones), y que el consenso de la población mexicana es más parecido a la secuencia consenso de Estados Unidos, que la secuencia consenso de Guatemala a este mismo consenso.

Tabla 6.2. Posiciones que difieren en las secuencias consenso de México, Guatemala y un virus del subtipo B, circulante en Estados Unidos de América.

Posición en el alineamiento	México	Guatemala	Subtipo B (USA)
p17- <i>gag</i> 10	G	<b>A</b>	G
p17- <i>gag</i> 15	R	<b>K</b>	R
p17- <i>gag</i> 58	R	R	<b>K</b>
p17- <i>gag</i> 67	P	P	<b>A</b>
p17- <i>gag</i> 76	K	K	<b>R</b>
p17- <i>gag</i> 79	Y	Y	<b>F</b>
p17- <i>gag</i> 91	K	K	<b>R</b>
p17- <i>gag</i> 93	D	<b>E</b>	D
p24- <i>gag</i> 138	N	N	<b>I</b>
p24- <i>gag</i> 146	A	<b>P</b>	A
p24- <i>gag</i> 215	L	L	<b>V</b>
p24- <i>gag</i> 223	<b>*V (46%)</b>	I	I
p24- <i>gag</i> 248	G	<b>A</b>	G
p24- <i>gag</i> 312	D	D	<b>E</b>
p2- <i>gag</i> 374	T	T	<b>A</b>
p7- <i>gag</i> 389	I	<b>T</b>	I
p1- <i>gag</i> 441	H	H	<b>Y</b>
p6- <i>gag</i> 483	L	L	<b>M</b>

\* Los aminoácidos en rojo señalan el aminoácido que difiere en alguna de las secuencias consenso con relación a las otras dos secuencias. \*En la posición p24-*gag* 223 en la población mexicana no presentó un aminoácido que apareciera en el 50% de las secuencias totales, el más representativo fue una valina (V) con un 46%.

Con el objetivo de analizar más detalladamente las diferencias entre las secuencias del gen *gag* de VIH circulante en las poblaciones mexicana y guatemalteca, se llevaron a cabo análisis de variabilidad que contemplaban la diversidad entre todas las secuencias de las dos poblaciones que se trataron en este trabajo y no sólo de los consensos de cada una de ellas.

En el caso del VIH, la variabilidad de un gen refleja, en cierta medida, la presión selectiva que se ejerce sobre éste. De manera general, un gen que es objeto de presión selectiva tiende a ser muy variable (de manera contraria al efecto homogenizador que tiene normalmente la selección natural), ya que la presión selectiva impulsa cambios en el mismo, que se ven reflejados en las variantes virales circulantes en un individuo infectado debido a la gran capacidad de mutar que posee el VIH y al corto período intergeneracional en la replicación del virus. Por el contrario, un gen conservado, puede considerarse libre de presión selectiva, bajo una presión selectiva baja o bajo una presión selectiva positiva que tiende a la conservación. Es importante señalar que la plasticidad o permisividad a mutaciones en los genes del VIH o en fragmentos de éstos no es igual ya que, mientras hay genes con una alta plasticidad como el gen *gp -120*, hay otros que, dada la importancia estructural de las proteínas para las cuales codifican, no pueden ser muy variables, como el caso de *gag*. Por esta razón, se analizó la distribución de la variabilidad en el gen *gag* de VIH que circula en las poblaciones de México y Guatemala para determinar si existían diferencias con respecto a las porciones del gen que están bajo presión selectiva en cada población.

Primero, se calculó de manera manual el porcentaje de secuencias de México y Guatemala que presentaban polimorfismos en cada posición del gen *gag* de VIH, con la intención de determinar si alguna de las proteínas codificadas por este gen, permitía una mayor variabilidad y para evaluar qué posiciones (y fragmentos) del alineamiento del gen eran las más variables (Fig. 6.4).

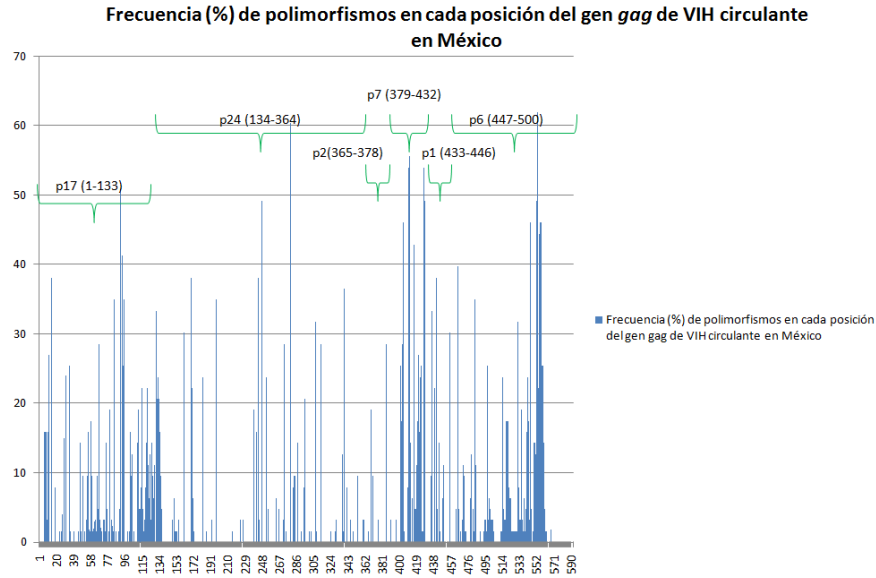


Figura 6.4a

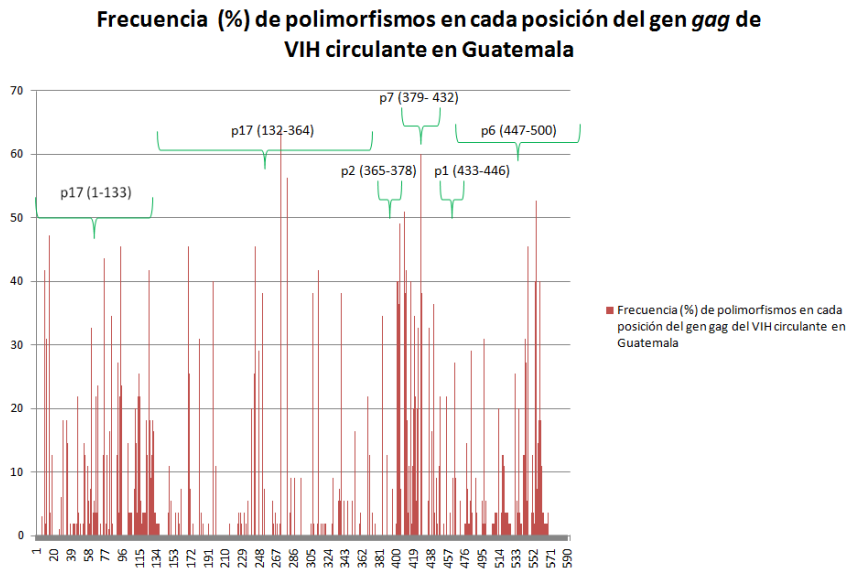
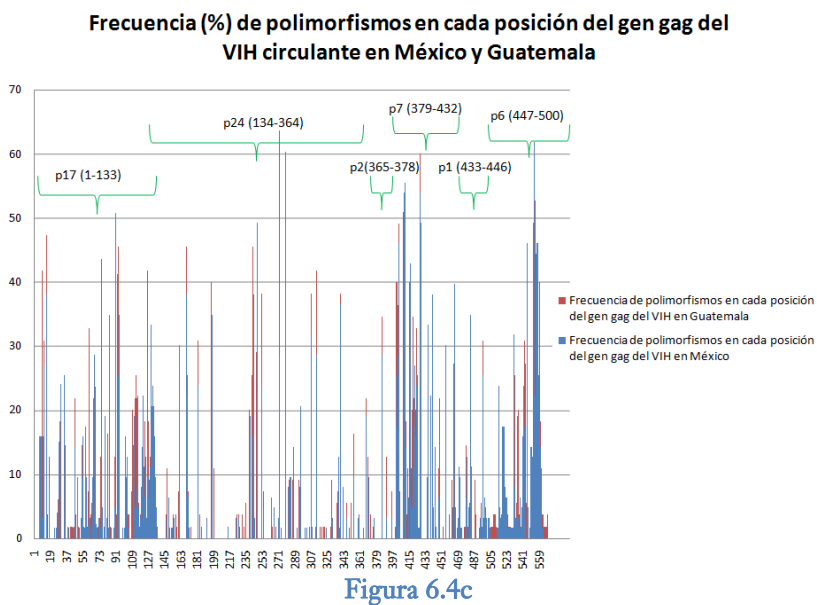


Figura 6.4b

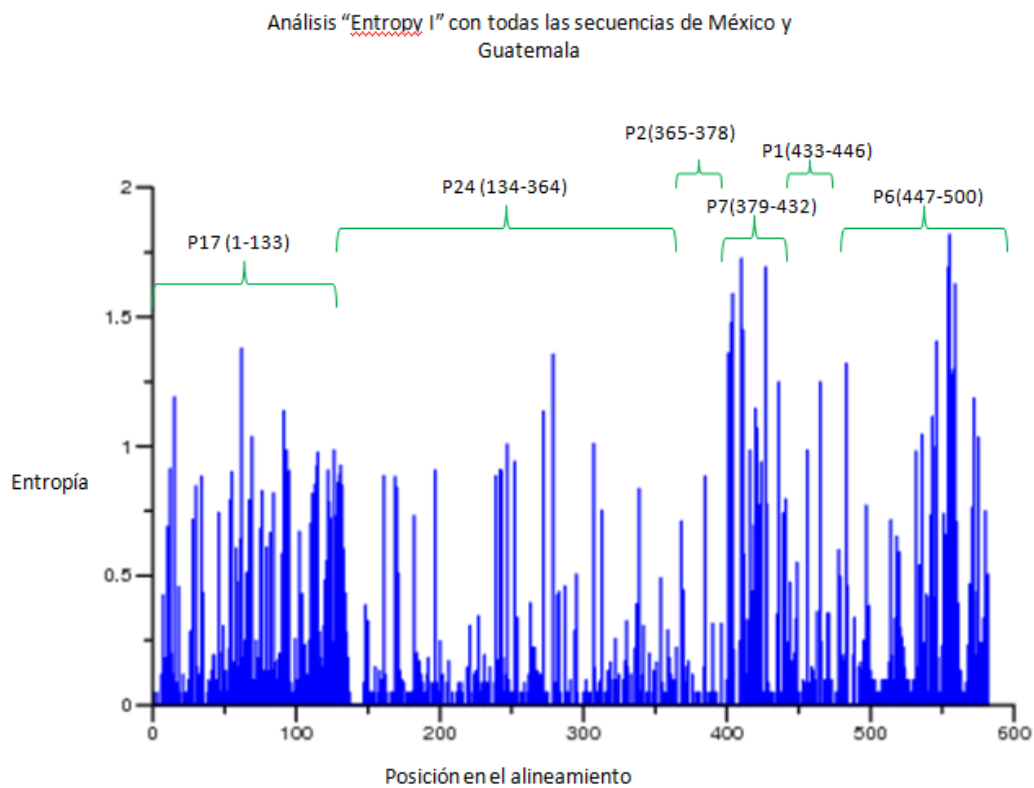


**Figura 6.4. Frecuencia de polimorfismos en cada posición del gen *gag* en México y Guatemala.** Las gráficas muestran el porcentaje de secuencias que presentaron polimorfismos en cada posición del gen *gag* en México (a) y Guatemala (b). En ambos casos se puede observar que la región que codifica la proteína de cápside, p24 es la menos variable de las proteínas del gen *gag*, mientras que la variabilidad se concentra en los extremos del gen, que corresponden a las proteínas p17, p7 y 6. La última gráfica (c) es la sobreposición de las gráficas de México y Guatemala y deja claro que el patrón de variabilidad es muy similar en los virus circulantes en ambas poblaciones.

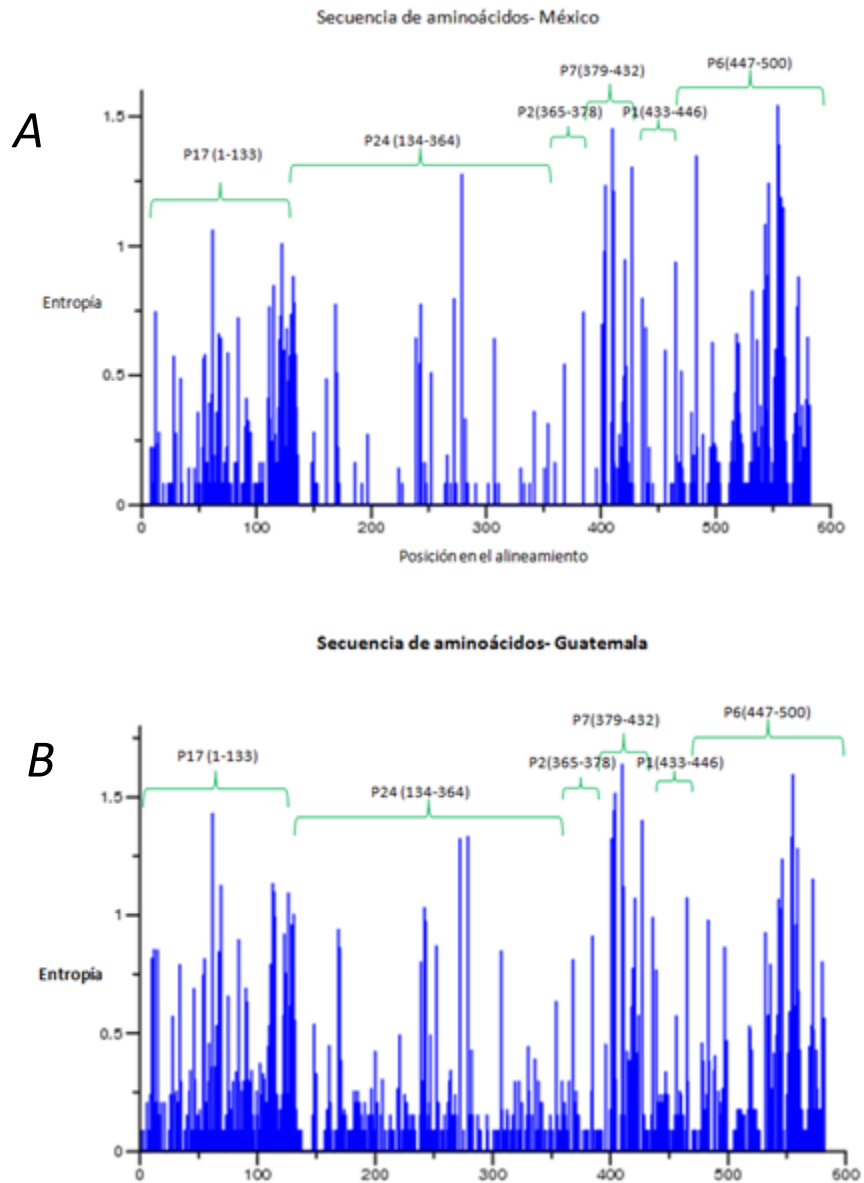
Tanto en México como en Guatemala la variabilidad se concentró en las proteínas p17, p7 y p6 mientras que la proteína de cápside p24 resultó ser la más conservada de las seis proteínas en este gen. Interesantemente, p24 mostró ser ligeramente más polimórfica en la población guatemalteca (Fig. 6.4).

Además del cálculo manual del porcentaje de secuencias que presentaban polimorfismos en cada posición del gen *gag*, se realizaron análisis de entropía (como medida de variabilidad) con las secuencias de México y Guatemala (Figs. 6.5 y 6.6), ya que los análisis de entropía son más precisos que el conteo manual y cuentan con análisis estadísticos que le dan validez a los resultados obtenidos.

Los análisis de entropía son una medida de la variabilidad de una posición en una secuencia de aminoácidos. En el contexto genético, la entropía se refiere al grado de desorden o aleatoriedad en la composición de aminoácidos (o nucleótidos) en una posición dentro de una secuencia, de tal modo que la entropía es mayor en posiciones en las que hay un número mayor de aminoácidos diferentes y/o en las cuales las proporciones de aminoácidos diferentes en una posición son más heterogéneas.



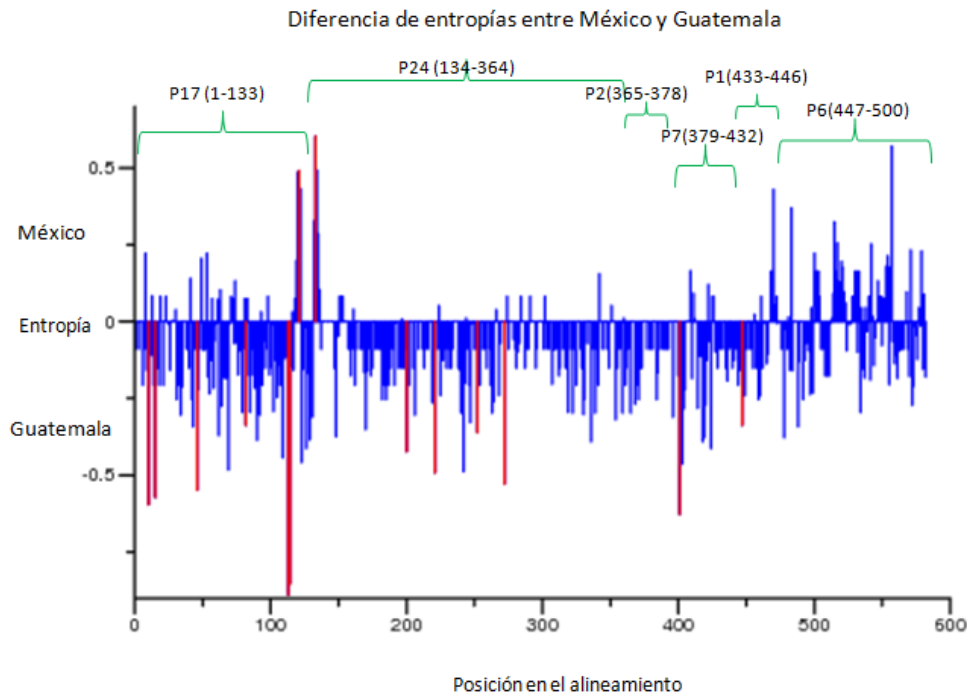
**Figura 6.5.** Análisis "Entropy I" con las secuencias de aminoácidos del gen *gag* circulante tanto en México como en Guatemala. La gráfica muestra cuáles son las posiciones más entrópicas (variables) en un set de datos comprendido por las secuencias de México y Guatemala. Resulta evidente que la región menos entrópica del gen *gag* en este set de datos es la que corresponde a la proteína p24 y las más entrópicas son las posiciones que comprenden las regiones que codifican p 17, p7 y p6.



**Figura 6.6. “Análisis Entropy II” con las secuencias de aminoácidos del gen *gag* circulante en México y Guatemala.** Las secuencias se analizaron a partir de dos sets de datos, un set correspondiente a México (A) y otro a Guatemala (B). Las gráficas muestran el grado de entropía en las secuencias de aminoácidos del gen *gag* del VIH circulante en México (A) y Guatemala (B) por separado. En ambos casos, la región menos entrópica (más conservada) es la que corresponde a p24. Las regiones que codifican las proteínas p17, p7 y p6 son las más entrópicas dentro del gen *gag*.

Se observó un patrón similar al de las gráficas de porcentaje de polimorfismos de ambas poblaciones (Fig. 6.4) si se consideran las dos poblaciones (mexicana y guatemalteca) como un set de datos (Fig. 6.5) y también si se les considera como sets separados (Fig. 6.6). El análisis de entropía coincide en que la región más conservada del gen *gag* circulante en México y Guatemala es p24 y las regiones más variables son las correspondientes a las proteínas p17, p7 y p6.

Si bien la distribución de la variabilidad en las secuencias del gen *gag* de VIH circulante en las poblaciones mexicana y guatemalteca resultó ser muy similar, se calculó la diferencia de variabilidad (entropía) entre las secuencias de México y Guatemala para determinar diferencias puntuales en la variabilidad de los diferentes codones de ambos grupos de secuencias (Fig. 6.7).



**Figura 6.7. Diferencia de entropías entre México y Guatemala.** El análisis “Entropy II” calcula la diferencia de entropía en cada posición de las secuencias del set de datos de México y el de Guatemala (entropía en México- entropía en Guatemala). Las columnas en la gráfica que se ubican por encima del cero corresponden a la población mexicana y son posiciones en las que las secuencias de México son más entrópicas que las secuencias de Guatemala. Las barras por debajo del cero son las posiciones en las que las secuencias de Guatemala son más entrópicas que las secuencias de México. Las barras rojas son posiciones en las cuales las diferencias de entropía entre los dos sets de datos es significativa ( $P < 0.05$ ). En total, el análisis detectó 14 posiciones en el alineamiento en las que la diferencia de entropía entre México y Guatemala es significativa. Doce de estas posiciones indican que el gen *gag* del VIH circulante en Guatemala es más entrópico que el que circula en México. Siete de las posiciones que difieren significativamente entre México y Guatemala se ubican en la región que codifica p17, seis de estas posiciones indican que Guatemala es más entrópico que México y sólo una posición es más entrópica en México que en Guatemala. En la porción que corresponde a p24, Guatemala presenta cuatro posiciones con una entropía significativamente mayor que en México, mientras que México presenta sólo una en la porción inicial de esta región. Guatemala también presenta dos posiciones significativamente más entrópicas que México en p2 y p7, una en cada proteína.

Al calcular la diferencia de entropía entre las secuencias de México y Guatemala (Fig. 6.7), se confirmó que las secuencias de Guatemala son significativamente ( $P < 0.05$ ) más polimórficas en la porción de p24 que las de México. El programa determinó que las secuencias de Guatemala son más variables en seis posiciones de p17, cuatro en p24 y dos en p7 y p1, una en cada proteína (Fig. 6.7). Las secuencias de México resultaron ser más entrópicas que las de Guatemala en 2 posiciones, una en p17 y otra en p24. Estos resultados concuerdan con lo observado en los estudios de frecuencias de polimorfismos y confirman una mayor variabilidad en los virus guatemaltecos. Además, estos datos localizan los puntos de mayor variabilidad de los virus guatemaltecos con respecto a los mexicanos principalmente en p24 y p17, sugiriendo un posible impacto diferencial de presiones selectivas en los virus de ambas poblaciones

#### **6.4. Asociación entre moléculas HLA-I y posiciones del gen *gag* de VIH dentro de las poblaciones mexicana y guatemalteca**

Para comprobar si la variabilidad observada en las secuencias del gen *gag* de VIH-I circulante en México y Guatemala por medio de los análisis de frecuencia de polimorfismos y entropía era consecuencia de una presión selectiva incidente, se identificaron sitios bajo presión selectiva positiva (se refiere al cambio de un nucleótido dentro de un codón que deriva en el cambio del aminoácido que codificaba el codón originalmente) por tasas de sustituciones sinónimas y no sinónimas, utilizando el servidor en línea Data Monkey. Además, dado el fuerte impacto de la presión selectiva inmune sobre la evolución del VIH, se evaluó la posibilidad de que las posiciones bajo presión selectiva positiva observadas, pudieran ser explicadas por la presión que ejercen los HLA-I, mediante la búsqueda de posibles epítomos de linfocitos T citotóxicos (disponibles en la Base de Datos de VIH de Los Alamos) que estuvieran asociados a cada posición bajo presión selectiva positiva que se identificara en el análisis.

En la población mexicana se detectaron dieciocho sitios bajo presión selectiva positiva mientras que en la población guatemalteca se detectaron catorce sitios ( $P < 0.05$ ) (tabla 6.3). Interesantemente, solo seis de las posiciones bajo presión selectiva positiva coincidieron entre las dos poblaciones.

En los segmentos del gen *gag* circulante en la población mexicana que codifican la proteína p17, se identificaron 4 posiciones bajo presión selectiva positiva al igual que en la población guatemalteca. Una de las 4 posiciones bajo presión selectiva en esta porción de *gag* coincidió en México y Guatemala (p17-*gag* 62). En el segmento que codifica la proteína p24, las secuencias de *gag* de la población guatemalteca presentaron cinco posiciones bajo presión selectiva, mientras que en la población mexicana se identificaron sólo cuatro. Dos posiciones bajo presión selectiva en este segmento coincidieron en las poblaciones mexicana y guatemalteca (p24-*gag* 173 y p24-*gag* 248). En el segmento que codifica la proteína espaciadora p2, se identificaron dos posiciones bajo presión selectiva positiva en la población guatemalteca y una posición en la población mexicana, todas las posiciones fueron distintas.



Tabla 6.3 Asociación entre posiciones bajo presión selectiva positiva determinadas con el servidor Data Monkey en la secuencia de gag y alelos HLA-I.

Posición del gen <i>gag</i> bajo presión selectiva positiva	Población en la que se presenta	P	Alelos HLA que presentan epítomos que incluyen los aminoácidos bajo presión selectiva positiva
<i>p17-gag</i> 7 ( <i>gag</i> 7 HXB2)	México	0.0204249	-
<i>p17-gag</i> 54 ( <i>gag</i> 54 HXB2)	Guatemala	0.00675898	-
<b><i>p17-gag</i> 62 (<i>gag</i> 62 HXB2)</b>	México y Guatemala	<b>México: 0.0363194 Guatemala: 0.0217921</b>	-
<i>p17-gag</i> 69 ( <i>gag</i> 69 HXB2)	Guatemala	0.017039	-
<i>p17-gag</i> 76 ( <i>gag</i> 76 HXB2)	Guatemala	0.00768387	-
<i>p17-gag</i> 84 ( <i>gag</i> 84 HXB2)	México	0.0124996	A*3002
<i>p17-gag</i> 111 ( <i>gag</i> 111 HXB2)	México	0.0418444	-
<i>p17-gag</i> 122 ( <i>gag</i> 117 HXB2)	México	0.03897	-
<i>p24-gag</i> 182 ( <i>gag</i> 159 HXB2)	Guatemala	0.0486926	A*0201, A*0202, B*1503, B*4501
<b><i>p24-gag</i> 197 (<i>gag</i> 173 HXB2)</b>	México y Guatemala	<b>México: 0.000028749 Guatemala: 0.00266355</b>	A*2601, A*2602, A*2603, C*0102
<i>p24-gag</i> 243 ( <i>gag</i> 219 HXB2)	México	0.0174885	B*5701, B*5703, B*5801, B*3910.
<b><i>p24-gag</i> 272 (<i>gag</i> 248 HXB2)</b>	México y Guatemala	<b>México: 0.0450768 Guatemala: 0.0419703</b>	A*0201, B*5701, B*5703, B*5801.
<i>p24-gag</i> 307 ( <i>gag</i> 280 HXB2)	Guatemala	0.00555509	C*1801, B*5201, C*102
<i>p24-gag</i> 313 ( <i>gag</i> 286 HXB2)	Guatemala	0.0314916	-

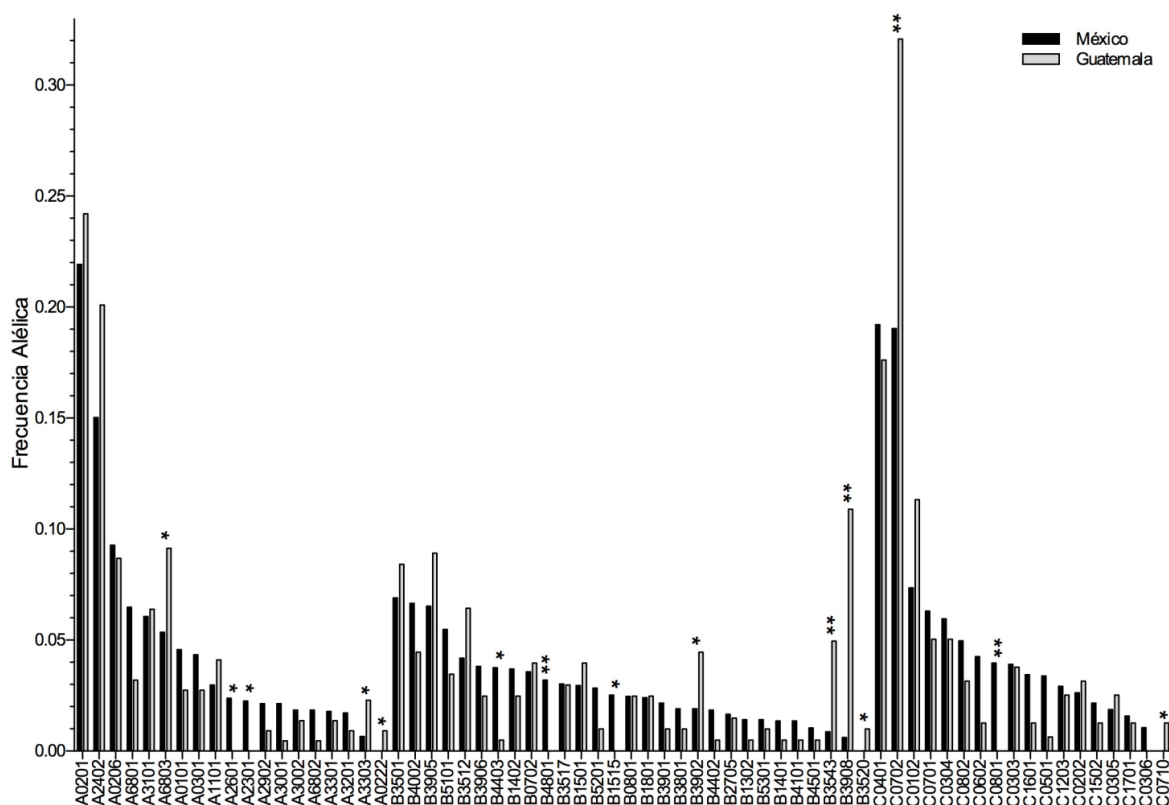
Posición del gen <i>gag</i> bajo presión selectiva positiva	Población en la que se presenta	P	Alelos HLA que presentan epítomos que incluyen los aminoácidos bajo presión selectiva positiva
<i>p24-gag 368 (gag 340 HXB2)</i>	México	0.0273037	-
<i>p2-gag 401 (gag 370 HXB2)</i>	Guatemala	0.02136	B*4002, B*4005,
<i>p2-gag 404 (gag 373 HXB2)</i>	México	0.00112308	B*4002, B*4501,
<i>p2-gag 410 (gag 374 HXB2)</i>	Guatemala	0.00033394	-
<i>p7-gag 427 (gag 389 HXB2)</i>	México y Guatemala	<b>México:</b> 0.000101149 <b>Guatemala:</b> 0.00270702	-
<i>p7-gag 428 (gag 390 HXB2)</i>	México	0.00877036	-
<i>p7-gag 465 (427 HXB2)</i>	México	0.0332177	-
<i>p1-gag 483 (gag 441 HXB2)</i>	México y Guatemala	<b>México:</b> 0.00290393 <b>Guatemala:</b> 0.0397946	A*0201, A*0205, A*3101.
<i>p6-gag 497 (gag 451 HXB2)</i>	México	0.0310415	-
<i>p6-gag 557 (gag 481 HXB2)</i>	México	0.0258507	B*0801, B*4001.
<i>p6-gag 559 (gag 483 HXB2)</i>	México y Guatemala	<b>México:</b> 0.000682691 <b>Guatemala:</b> 0.000638351	B*0801, B*4001, B*3501, B*0702.
<i>p6-gag 576 (gag 488 HXB2)</i>	México	0.0306621	B*4001, B*3501, B*0702.

\*La presión selectiva positiva se refiere al cambio de un nucleótido dentro de un codón que deriva en el cambio del aminoácido que codificaba el codón originalmente [27]. \*Se destacan en rojo las posiciones bajo presión selectiva positiva determinadas con el servidor Data Monkey que coinciden entre las poblaciones mexicana y guatemalteca.

En la porción que codifica la proteína p7 se identificaron tres sitios en la población mexicana y uno en la población guatemalteca, este último (p7-*gag* 389) coincidió entre las dos poblaciones. En la porción que codifica a la proteína espaciadora p1 se identificó un sitio en cada población y estos sitios son diferentes uno de otro. Finalmente, en el segmento que codifica a la proteína p6, se identificaron cuatro sitios bajo presión selectiva positiva en la población mexicana y uno en la guatemalteca, éste último (p6-*gag* 559) coincide con uno de los sitios identificado en la población mexicana (tabla 6.3).

Considerando estos datos en conjunto, se puede decir que hay un número mayor de posiciones bajo presión selectiva positiva en México (dieciocho) que en Guatemala (catorce). De las posiciones bajo presión selectiva positiva que se identificaron, 6 posiciones, que se encuentran repartidas en todas las proteínas que codifica el gen *gag*, coinciden en las poblaciones de México y Guatemala (tabla 6.3), con excepción de p2, en donde ninguno de los sitios identificados coincide. Estas observaciones sugieren rasgos comunes en la evolución del virus en México y Guatemala, pero también características únicas de evolución del virus en cada una de las dos poblaciones.

Se buscaron correlaciones entre posiciones del gen *gag* de VIH y alelos HLA-I disponibles en la base de datos de Los Alamos (se consideraron sólo los alelos HLA-I disponibles en la base de datos que se hubieran determinado por métodos de secuenciación). Al buscar asociaciones de posibles epítomos de linfocitos T citotóxicos con las posiciones bajo presión selectiva encontradas se observó que solo el 44% (8/18) de las posiciones del virus que circula en México y el 43% (6/14) de las posiciones del que circula en Guatemala se encontraron en regiones previamente identificadas como inmunogénicas, con epítomos comprobados experimentalmente para alelos HLA específicos (tabla 6.3). Sin embargo, se debe considerar que la Base de Datos de VIH de Los Alamos está constituida principalmente por datos obtenidos de cohortes predominantemente caucásicas y de origen africano, por lo que posiblemente, las correlaciones entre posiciones del gen *gag* de VIH y alelos HLA-I no sean los más adecuados para las poblaciones con componentes genéticos de origen amerindio como la mexicana y la guatemalteca. Por esta razón, se analizaron las frecuencias alélicas de genes HLA-I de las poblaciones mexicana y guatemalteca que se han tipificado hasta ahora en las cohortes de pacientes del laboratorio de epidemiología y evolución molecular del Centro de Investigación en Enfermedades Infecciosas (CIENI) del Instituto Nacional de Enfermedades Respiratorias (INER). Interesantemente, se observaron 13 alelos HLA con frecuencias distintas en ambas cohortes, lo cual podría estar determinando características únicas en la evolución del VIH en cada una de las dos poblaciones (tabla 6.4 y Fig. 6.8).



**Figura 6.8. Comparación de frecuencias alélicas HLA-I en México y Guatemala.** La figura muestra la frecuencia de los alelos HLA-I en México (negro) y Guatemala (blanco). Los alelos HLA-I se dividieron en HLA-A (izq.), HLA-B (centro) y HLA-C (der.). Los asteriscos indican diferencias significativas. (\*\*)  $p > 0.05$  y un asterisco (\*)  $P < 0.05$  y  $q < 0.2$ . Valenzuela-Ponce, et al, datos no publicados

*Tabla 6.4. Comparación de frecuencias alélicas HLA-I en México y Guatemala (Valenzuela-Ponce, et al, datos no publicados).*

HLA	FA México	FA Guatemala	A	b	c	d	P	q
A0201	0.219251337	0.242009132	369	53	1314	166	0.437609162	0.75120055
A2402	0.150326797	0.200913242	253	44	1430	175	0.059733587	0.265908628
A0206	0.092691622	0.086757991	156	19	1527	200	0.90105124	0.922932521
A6801	0.0647653	0.03196347	109	7	1574	212	0.069729892	0.292813517
A3101	0.060606061	0.063926941	102	14	1581	205	0.88039529	0.92903481
A6803	0.053475936	0.091324201	90	20	1593	199	0.030548256	0.16177458
A0101	0.045751634	0.02739726	77	6	1606	213	0.289324909	0.650431742
A0301	0.043374926	0.02739726	73	6	1610	213	0.366124904	0.706923238
A1101	0.029708853	0.04109589	50	9	1633	210	0.403594859	0.754759426
A2601	0.023767083	0	40	0	1643	219	0.01116979	0.107092375
A2301	0.022578728	0	38	0	1645	219	0.017651146	0.124626278
A2902	0.021390374	0.00913242	36	2	1647	217	0.306737596	0.665545354
A3001	0.021390374	0.00456621	36	1	1647	218	0.115666551	0.426277776
A3002	0.018419489	0.01369863	31	3	1652	216	0.790388049	0.89172158

HLA	FA México	FA Guatemala	A	b	c	d	P	q
A6802	0.018419489	0.00456621	31	1	1652	218	0.167438616	0.480843954
A3301	0.017825312	0.01369863	30	3	1653	216	1	1
A3201	0.017231135	0.00913242	29	2	1654	217	0.570152179	0.800768141
A3303	0.006535948	0.02283105	11	5	1672	214	0.029099978	0.169694603
A0222	0	0.00913242	0	2	1683	217	0.013204075	0.10931161
<b>B3501</b>	<b>0.068965517</b>	<b>0.084158416</b>	<b>112</b>	<b>17</b>	<b>1512</b>	<b>185</b>	<b>0.465309176</b>	<b>0.770890663</b>
<b>B4002</b>	<b>0.066502463</b>	<b>0.044554455</b>	<b>108</b>	<b>9</b>	<b>1516</b>	<b>193</b>	<b>0.285823989</b>	<b>0.660383594</b>
B3905	0.065270936	0.089108911	106	18	1518	184	0.233466555	0.609018091
B5101	0.054802956	0.034653465	89	7	1535	195	0.313715017	0.664230596
B3512	0.041871921	0.064356436	68	13	1556	189	0.146901335	0.475908784
B3906	0.03817734	0.024752475	62	5	1562	197	0.429427182	0.757401631
B4403	0.037561576	0.004950495	61	1	1563	201	0.011538712	0.095919174
B1402	0.036945813	0.024752475	60	5	1564	197	0.543830329	0.780194483
<b>B0702</b>	<b>0.035714286</b>	<b>0.03960396</b>	<b>58</b>	<b>8</b>	<b>1566</b>	<b>194</b>	<b>0.692468633</b>	<b>0.852899133</b>
B4801	0.032019704	0	52	0	1572	202	0.003062746	0.048879357
B3517	0.030172414	0.02970297	49	6	1575	196	1	1
B1501	0.02955665	0.03960396	48	8	1576	194	0.390182834	0.765036276
<b>B5201</b>	<b>0.028325123</b>	<b>0.00990099</b>	<b>46</b>	<b>2</b>	<b>1578</b>	<b>200</b>	<b>0.160270865</b>	<b>0.485989438</b>
B1515	0.025246305	0	41	0	1583	202	0.018995059	0.122853934
<b>B0801</b>	<b>0.024630542</b>	<b>0.024752475</b>	<b>40</b>	<b>5</b>	<b>1584</b>	<b>197</b>	<b>1</b>	<b>0.934201475</b>
B1801	0.024014778	0.024752475	39	5	1585	197	0.811599105	0.905350324
B3901	0.021551724	0.00990099	35	2	1589	200	0.423853923	0.767499308
B3801	0.01908867	0.00990099	31	2	1593	200	0.572787016	0.795064635
B3902	0.01908867	0.044554455	31	9	1593	193	0.035351511	0.193343093
B4402	0.018472906	0.004950495	30	1	1594	201	0.245246091	0.638279022
<b>B2705</b>	<b>0.016625616</b>	<b>0.014851485</b>	<b>27</b>	<b>3</b>	<b>1597</b>	<b>199</b>	<b>1</b>	<b>1</b>
B1302	0.014162562	0.004950495	23	1	1601	201	0.507545414	0.771565315
B5301	0.014162562	0.00990099	23	2	1601	200	1	1
B1401	0.013546798	0.004950495	22	1	1602	201	0.503433179	0.778303378
B4101	0.013546798	0.004950495	22	1	1602	201	0.503433179	0.778303378
<b>B4501</b>	<b>0.01046798</b>	<b>0.004950495</b>	<b>17</b>	<b>1</b>	<b>1607</b>	<b>201</b>	<b>0.711653792</b>	<b>0.851369019</b>
B3543	0.00862069	0.04950495	14	10	1610	192	0.000107382	0.002840167
B3908	0.006157635	0.108910891	10	22	1614	180	7.55E-15	2.04E-13
B3520	0	0.00990099	0	2	1624	200	0.012183829	0.108373865
C0401	0.192060712	0.176100629	329	28	1384	131	0.674023629	0.853700869
C0702	0.190309399	0.320754717	326	51	1387	108	0.000181984	0.003446145
<b>C0102</b>	<b>0.073555166</b>	<b>0.113207547</b>	<b>126</b>	<b>18</b>	<b>1587</b>	<b>141</b>	<b>0.08534066</b>	<b>0.340892379</b>
C0701	0.063047285	0.050314465	108	8	1605	151	0.609269954	0.809575931
C0304	0.059544658	0.050314465	102	8	1611	151	0.859475042	0.927742148
C0802	0.049620549	0.031446541	85	5	1628	154	0.435897651	0.757904142
C0602	0.042615295	0.012578616	73	2	1640	157	0.086169773	0.33381053
C0801	0.039696439	0	68	0	1645	159	0.003348768	0.041284087
C0303	0.039112668	0.037735849	67	6	1646	153	1	1

HLA	FA México	FA Guatemala	A	b	c	d	P	q
C1601	0.034442499	0.012578616	59	2	1654	157	0.164307079	0.4887051
C0501	0.033858727	0.006289308	58	1	1655	158	0.056484152	0.267509468
C1203	0.029188558	0.025157233	50	4	1663	155	1	0.920121032
C0202	0.026269702	0.031446541	45	5	1668	154	0.609338988	0.833062667
C1502	0.021599533	0.012578616	37	2	1676	157	0.76855344	0.884713168
C0305	0.018680677	0.025157233	32	4	1681	155	0.541075225	0.785935507
C1701	0.015761821	0.012578616	27	2	1686	157	1	0.942544834
C0306	0.010507881	0	18	0	1695	159	0.392919114	0.765283555
C0710	0	0.012578616	0	2	1713	157	0.007172566	0.083113694

\*Para este análisis se tomaron 870 pacientes mexicanos y 184 guatemaltecos. Los datos en verde indican que  $P < 0.05$   $q < 0.05$  y los datos en naranja indican que  $P < 0.05$   $q < 0.2$ . a= número de individuos positivos para un alelo HLA en la cohorte mexicana, b= número de individuos positivos para un alelo HLA en la cohorte guatemalteca, c= número de individuos negativos para un alelo HLA en la cohorte mexicana, d= número de individuos negativos para un alelo HLA en la cohorte guatemalteca. En negritas se muestran los alelos posiblemente asociados a posiciones bajo presión selectiva positiva, identificados a través de la base de datos de epítomos de linfocitos T citotóxicos de Los Alamos.

Dentro de los alelos que se asocian con posiciones bajo presión selectiva determinadas por el servidor Data Monkey, los que están presentes en las cohortes mexicana y guatemalteca según los datos disponibles en el CIENI son A\*3002, A\*0201, A\*2601, A\*3101, B\*5201, B\*4002, B\*0801, B\*3501, B\*0702 y C\*0102 (tabla 6.4).

Los alelos HLA-I que están presentes en ambas poblaciones según la base de datos del CIENI se muestran en negritas. Es importante señalar que cuatro sitios que mostraron relación con HLA-I coinciden en México y Guatemala (tabla 6.3). Estos sitios fueron:

- p24- *gag* 173 con **A\*2601**, A\*2602, A\*2603 y **C\*0102**
- p24- *gag* 248 con **A\*0201**, B\*5701, B\*5703 y B\*5801
- p1- *gag* 441 con **A\*0201**, A\*0205 y **A\*3101**
- p6- *gag* 483 con **B\*0801**, B\*4001, **B\*3501** y **B\*0702**
- 

Los sitios que mostraron relación con moléculas HLA-I en la población de México que no coincidían con los sitios identificados en la cohorte de Guatemala (tabla 6.3) fueron:

- p17- *gag* 84 con **A\*3002**
- p24- *gag* 219 con B\*5701, B\*5703, B\*5801 y B\*3910
- p2- *gag* 373 con **B\*4002** y **B\*4501**
- p6- *gag* 488 con B\*4001, **B\*3501** y **B\*0702**

Los sitios de la población guatemalteca que se asociaron con HLA-I que no coincidían con los sitios identificados en México (tabla 6.3) fueron:

- p24- *gag* 159 con **A\*0201**, A\*0202, B\*1503 Y B\*4501
- p2- *gag* 370 con **B\*4002** y B\*4005

Se puede observar que en todas las correlaciones entre alelos HLA-I y posiciones del gen *gag* hay por lo menos un alelo HLA-I que poseen las poblaciones mexicana y/o guatemalteca, con excepción de la posición p24- *gag* 219 que se correlaciona con alelos que no están presentes (o están en frecuencias muy bajas) en las poblaciones estudiadas. Al analizar la tabla 6.4, se puede apreciar que no existen diferencias significativas en la frecuencia de los alelos que se correlacionan con HLA en México y Guatemala con excepción de los alelos **C\*07** y **A\*2601**, aunque no se puede descartar que las frecuencias de estos alelos sean similares, ya que en el análisis comparativo de frecuencias alélicas HLA, se incluyó un número casi cinco veces mayor de pacientes de México.

Estos resultados sugieren que los perfiles inmunogenéticos de las poblaciones mexicana y guatemalteca presentan diferencias importantes que podrían determinar características únicas de la evolución del VIH en cada una de las dos poblaciones. Sin embargo, estos datos deben tomarse con reserva puesto que el tamaño de muestra de pacientes de los cuales se secuenció el gen *gag* de VIH que se empleó en este estudio es pequeño y las frecuencias alélicas de los genes HLA-I en la población guatemalteca no se han caracterizado completamente, por lo que los datos disponibles deberán confirmarse ampliando la cohorte.

Con el objetivo de emplear un modelo más adecuado a las poblaciones con componentes genéticos de origen amerindio para identificar posiciones en el genoma viral asociadas a selección por HLA, se realizó una comparación de los sitios bajo presión selectiva positiva identificados por razones de sustituciones sinónimas y no sinónimas, con las identificadas a través de un modelo evolutivo previamente desarrollado, denominado red de dependencia filogenética (PDN) [28]. El modelo PDN permite la identificación de sitios en el genoma viral asociados a selección por HLA, corrigiendo para los principales factores que confunden dichas asociaciones: efectos filogenéticos en el virus circulante, desequilibrio de enlace entre alelos HLA y coevolución de codones en el VIH. Este modelo fue aplicado a una cohorte de 530 pacientes mexicanos del CIENI, INER, con tipificación HLA y secuencia viral de *gag* disponibles (tabla 6.5).

Tabla 6.5. Asociación entre posiciones del gen gag de VIH bajo presión selectiva positiva y alelos HLA según el modelo de red de dependencia filogenética (PDN).

Posición del gen gag bajo presión selectiva positiva	Población en la que se presenta	Alelos HLA asociados a posiciones bajo presión selectiva positiva según el modelo de PDN <sup>a</sup>
<i>p17 gag 7 (gag 7 HXB2)</i>	México	-
<i>*p17-gag 54 (gag 54 HXB2)</i>	Guatemala	B*15 (atracción <sup>b</sup> ) p=0.000112
<i><b>p17- gag 62 (gag 62 HXB2)</b></i>	México y Guatemala	-
<i>*p17- gag 69 (gag 69 HXB2)</i>	Guatemala	C*07 (atracción) p= 0.000080
<i>p17- gag 76 (gag 76 HXB2)</i>	Guatemala	-
<i>p17-gag 84 (gag 84 HXB2)</i>	México	-
<i>p17- gag 111 (gag 111 HXB2)</i>	México	-
<i>p17- gag 122 (gag 117 HXB2)</i>	México	-
<i>p24- gag 182 (gag 159 HXB2)</i>	Guatemala	-
<i><b>p24- gag 197 (gag 173 HXB2)</b></i>	México y Guatemala	-
<i>p24- gag 243 (gag 219 HXB2)</i>	México	-
<i><b>*p24- gag 272 (248 en HXB2)</b></i>	México y Guatemala	C*0404 (atracción) p=0.000021
<i>*p24- gag 307 (gag 280 HXB2)</i>	Guatemala	B*15 (atracción) p=0.0000975
<i>*p24- gag 313 (gag 286 HXB2)</i>	Guatemala	C*01 (escape <sup>c</sup> ) p=0.000072
<i>p2-gag 368 (gag 340 HXB2)</i>	México	-
<i>p7- gag 401 (gag 370 HXB2)</i>	Guatemala	-
<i>p7- gag 404 (gag 373 HXB2)</i>	México	-
<i>p7- gag 410 (gag 374 HXB2)</i>	Guatemala	-
<i><b>p7-gag 427 (gag 389 HXB2)</b></i>	México y Guatemala	-
<i>*p7-gag 428 (gag 390 HXB2)</i>	México	B*2705 (escape) p= 0.000142
<i>p6-gag 465 (gag 427 HXB2)</i>	México	-
<i><b>p6- gag 483 (gag 441 HXB2)</b></i>	México y Guatemala	-
<i>p6- gag 497 (gag 451 HXB2)</i>	México	-
<i>p6- gag 557 (gag 481 HXB2)</i>	México	-
<i><b>p6- gag 559 (gag 483 HXB2)</b></i>	México y Guatemala	-
<i>p6- gag 576 (gag 488 HXB2)</i>	México	-

\* a= Modelo de red de dependencia filogenética para la identificación de posiciones del genoma viral asociadas a HLA [28]. b=Atracción: la presencia del alelo HLA hace más probable que se presente un aminoácido en una posición dada. c= Escape: la presencia del alelo HLA hace menos probable que se presente un aminoácido en una posición dada. \* p>0.05 indica qué tan fuerte es la correlación entre el alelo HLA y la posición en el alineamiento. \*Se muestran las posiciones bajo presión selectiva positiva determinadas por tasas de sustituciones sinónimas y no sinónimas (Data Monkey, tabla 6.3) en las poblaciones mexicana y guatemalteca. \*En rojo se señalan las posiciones bajo presión selectiva positiva que se identificaron en las dos poblaciones. (\*)= posiciones que mostraron relación con alelos HLA-I en el análisis de PDN. (\*\*)= posición bajo presión selectiva positiva y se puede correlacionar con alelos HLA-I tanto en la base de datos de Los Alamos como en el análisis de PDN.



De los dieciocho sitios bajo presión selectiva positiva que se detectaron con el servidor Data Monkey en la población mexicana, sólo dos se correlacionaron con moléculas HLA-I en el análisis de PDN. De los catorce sitios bajo presión selectiva positiva en Guatemala, cinco se correlacionaron con moléculas HLA-I en el análisis de PDN.

Las posiciones del gen *gag* de VIH que se asociaron con moléculas HLA-I según el análisis de PDN en la cohorte guatemalteca fueron p17-*gag* 54 que en el análisis PDN se relaciona con el alelo B\*15, de manera tal que la presencia del alelo hace más probable que se presente el aminoácido (S) en la posición 54 del gen. Las posiciones p17-*gag* 69 (Q) que se relaciona con el alelo C\*07 y la posición p24-*gag* 280(T) que se relaciona con el alelo B\*15 se relacionan de la misma manera que el caso anterior. Por el contrario, la posición p24-*gag* 286 se relaciona con el alelo C\*01 de tal forma, que la presencia del alelo HLA hace menos probable que se presente un aminoácido determinado (R) en la posición del péptido (tabla 6.5). La posición p24-*gag*- 248 (G en la secuencia consenso de México y A en la de Guatemala) se relaciona con el alelo C\*w07 y es una asociación que se encontró en ambas poblaciones. En la población mexicana además, se identificó la asociación de la posición p7-*gag* 390 con el alelo B\*2705 de tal manera que la presencia del alelo, hace menos probable la aparición del aminoácido (V) en esa posición.

De las seis posiciones bajo presión selectiva positiva que coincidieron en la población mexicana y guatemalteca en el análisis con Data Monkey, la posición p24-*gag* 248, es la única que se correlaciona a moléculas HLA-I tanto en la base de datos de Los Alamos como en el análisis de PDN y coincide en ambas poblaciones. Además, las frecuencias alélicas de determinados alelos HLA C\*07 difieren en las poblaciones mexicana y guatemalteca (tabla 6.4), lo que sugiere una presión selectiva por moléculas HLA diferentes en las poblaciones de México y Guatemala.

Finalmente, se buscaron correlaciones entre las posiciones del gen *gag* que difieren en las secuencias consenso de la población de México y la de Guatemala, con moléculas HLA-I tanto en la base de datos de Los Alamos como en el análisis de PDN que se practicó en una cohorte mexicana (tabla 6.5). Se realizó este análisis para determinar si las posiciones que difieren en las secuencias consenso son reflejo de adaptaciones virales a nivel poblacional que se fijaron ante la presión selectiva que ejerce el sistema inmune característico de cada población.

Tabla 6.6 Asociación entre posiciones del gen gag que difieren en las secuencias consenso de las poblaciones mexicana y guatemalteca y alelos HLA-I.

Codón del gen gag bajo presión selectiva positiva que difiere en las secuencias consenso de le las poblaciones mexicana y guatemalteca	Alelos HLA que presentan epítomos que incluyen los aminoácidos que difieren entre las secuencias consenso de las poblaciones mexicana y Guatemala (La Base de dato de Los Alamos)	Alelos HLA que presentan epítomos que incluyen los aminoácidos que difieren entre las secuencias consenso de las poblaciones mexicana y Guatemala (PDN <sup>a</sup> )
<i>p17- gag 10 en HXB2 y consensos</i>	B*4002	-
<i>p17- gag 15 en HXB2 y consensos</i>	B*4002	-
<i>p17- gag 93 en HXB2 y consensos</i>	B*4001, B*0801	-
<i>p24- gag 146 en HXB2 y consensos</i>	A*2501, C*07, B*1510, B*5802, B*4801.	C*07 (escape <sup>b</sup> ) P= 0.000101
<b>***p24- gag 248 en HXB2 y consensos</b>	A*0201, B*5701, B*5703, B*5801.	C*0404 (atracción <sup>c</sup> ) p=0.000021
<i>p7- gag 389 HXB2 y consensos</i>	-	-

\* a= Modelo de red de dependencia filogenética para la identificación de posiciones del genoma viral asociadas a HLA [28]. b= escape: la presencia del alelo HLA hace menos probable que se presente un aminoácido en una posición dada. c=atracción la presencia del alelo HLA hace más probable que se presente un aminoácido en una posición dada. \*Se destacan con rojo las posiciones bajo presión selectiva positiva determinadas por tasa de sustituciones sinónimas y no sinónimas con Data Monkey que coinciden entre las poblaciones mexicana y guatemalteca.\*En negritas se destaca el alelo HLA-I que coincide en la base de datos de Los Alamos y en el análisis de PDN. (\*\*\*)= posición bajo presión selectiva positiva determinada con Data Monkey que coincide en las poblaciones mexicana y guatemalteca, que se asocia con alelos HLA-I tanto en la base de datos de Los Alamos como con el análisis de PDN y difiere en las secuencias consenso de México y Guatemala.

De las siete posiciones en las que las secuencias consenso de México y Guatemala difieren, una posición (p7-*gag* 389) no mostró relación con alelos HLA-I en ninguno de los dos análisis, tres posiciones (p17-*gag* 10, p17-*gag* 15 y p17-*gag* 93) se correlacionaron con HLA-I según la base de datos de Los Alamos y dos (p24-*gag* 145 y p24-*gag* 248) posiciones se asociaron con alelos HLA-I tanto en el análisis de PDN como con la base de datos de Los Alamos. Estos datos sugieren que las diferencias en las secuencias consenso de México y Guatemala, son resultado de la fijación de mutaciones virales de escape ante una presión selectiva distinta en cada una de las poblaciones estudiadas. La posición p24-*gag* 248 resulta de especial interés ya que es una posición bajo presión selectiva positiva en ambas poblaciones según el análisis con Data Monkey, se correlaciona con alelos HLA-I tanto en la base de datos de Los Alamos como en el análisis de PDN y difiere en la secuencia consenso de las poblaciones mexicana (G) y guatemalteca (A). Esta posición en particular sugiere fuertemente la fijación de una mutación de escape viral a nivel poblacional diferente en México y Guatemala.

A manera de resumen, en lo que a la asociación entre moléculas HLA-I y posiciones del gen *gag* circulante en las poblaciones mexicana y guatemalteca respecta, se puede decir que al analizar las secuencias con el servidor Data Monkey se identificaron veintiséis posiciones bajo presión selectiva positiva, doce exclusivamente en la población mexicana, ocho exclusivamente en la guatemalteca y seis posiciones se identificaron en ambas poblaciones. De las veintiséis posiciones bajo presión selectiva positiva que se identificaron, once se asociaron con moléculas HLA-I de la base de datos de Los Alamos, cinco posiciones del gen *gag* circulante en la población mexicana, tres en la población guatemalteca y cuatro posiciones que se identificaron tanto en México como Guatemala (tabla 6.3).

Al buscar correlaciones entre las posiciones bajo presión selectiva según el análisis con Data Monkey y un análisis de red de dependencia filogenética (PDN) practicado en una cohorte mexicana se determinó que cuatro posiciones del gen *gag* de VIH circulante en la población guatemalteca tenían relación con alelos HLA-I tanto en el análisis de PDN como con la base de datos de Los Alamos. Sólo una posición del gen *gag* circulante en México correlacionó con alelos HLA-I de la base de datos de Los Alamos y el análisis de PDN y una posición bajo presión selectiva en común (de la secuencia de México y de Guatemala) correlacionó con alelos HLA en ambos análisis (tabla 6.5).

Al determinar correlaciones entre alelos HLA-I y posiciones en las cuales las secuencias consenso de las poblaciones mexicana y guatemalteca diferían, se determinó que todas las posiciones se correlacionan con moléculas HLA-I según la base de datos de Los Alamos con excepción de la posición p7-*gag* 389 HXB2, que no se asoció con ningún alelo. Sólo dos posiciones se correlacionan con moléculas HLA-I según el análisis de PDN (tabla 6.6 y Fig. 6.9). La posición p24-*gag* 146 se correlaciona con el alelo HLA-C\*07 en ambos análisis por lo que la asociación se puede considerar contundente aunque no está bajo presión selectiva (sólo difiere en las secuencias consenso de las poblaciones mexicana y guatemalteca) según el análisis con Data Monkey. La posición p24-*gag* 248 fue la única posición que se correlacionó con moléculas HLA-I tanto en el análisis de PDN como en la base de datos de Los Alamos, que está bajo presión selectiva según el análisis con Data Monkey y además es una posición que difiere

en las secuencias consenso de México y Guatemala (tabla 6.6 y Fig. 6.9), lo que sugiere la fijación de una adaptación viral diferente en las poblaciones mexicana y guatemalteca. La posición *p7-gag* 389 HXB2 que no se asoció con ningún alelo HLA, puede representar una asociación no caracterizada en la actualidad o una asociación espuria.



# Capítulo 7

## Discusión

El presente estudio sugiere fuertemente un papel fundamental de la respuesta de linfocitos T citotóxicos, definida y restringida por los alelos HLA de clase I, en la adaptación diferencial del VIH circulante en dos poblaciones aparentemente similares. Este estudio es evidencia de que la evolución del VIH presenta características únicas en distintas poblaciones, independientemente de la cercanía geográfica y de si existe una ancestría étnica común entre éstas. Nuestros datos resaltan la importancia de conocer el perfil inmunogenético de cada población para estudiar y predecir patrones de evolución del virus, lo cual es relevante en la investigación en vacunas. Nuestro estudio muestra que además de existir características comunes en la evolución del virus en México y Guatemala, pudiéndose aprovechar éstas en el diseño de una vacuna universal al contarse con información de otras poblaciones del mundo, existen también diferencias importantes que apoyan la idea del desarrollo de vacunas a nivel local [38].

En el caso concreto de este estudio, los resultados de los análisis de la filogenia del gen *gag* de VIH circulante en México y Guatemala indican que existen diferencias en las secuencias de dicho gen en cada población. Esto sugiere una dinámica de transmisión de la infección de carácter local, que puede deberse a que barreras geográficas naturales como lo son el río Usumacinta, que recorre cerca de 200 km de la frontera México- Guatemala (con una extensión total de 956 km), el intrincado terreno montañoso que caracteriza la frontera entre México y Guatemala y los múltiples cuerpos de agua ubicados en la zona, limitan en cierta medida, el flujo génico entre ambas poblaciones [39]. Aunado a estas barreras naturales, la frontera política, a pesar de ser una división artificial que no obedece patrones biogeográficos de distribución, en el caso de los seres humanos, representa una barrera significativa que limita el flujo génico entre poblaciones, favoreciendo la endogamia y contribuyendo a la conservación de caracteres particulares de cada población. Adicionalmente, las barreras culturales son un factor social que podría influir también en las redes de transmisión del VIH en cada país. Las diferencias que se observaron en las secuencias de *gag* en las poblaciones objeto de este estudio, también pueden reflejar diferencias resultantes de un mestizaje distinto en México y Guatemala a lo largo de la historia, que le otorgan un perfil inmunogenético diferente a cada población. El hecho de que los árboles filogenéticos de las muestras de México y Guatemala por separado (análisis local) fueran considerablemente homogéneas a nivel local e incluyeran pacientes de diferentes regiones de cada país, le da validez a estas observaciones ya que excluye la posibilidad de que las diferencias observadas entre ambas poblaciones en el árbol elaborado con las secuencias de México y Guatemala, se deba a una fuerte divergencia local (Figs. 6.1 y 6.2).

Al comparar los consensos del gen *gag* de VIH que circula en México y Guatemala, se pudieron identificar diferencias que sugieren adaptaciones virales a una presión selectiva inmunológica diferente y característica en cada población.

En la comparación de las secuencias consenso de las poblaciones mexicana y guatemalteca con una secuencia consenso del gen *gag* circulante en Estados Unidos, se pueden reconocer dos cosas: una es que las secuencias de México y Guatemala son más parecidas entre sí que a la secuencia circulante en Estados Unidos, lo que sugiere una cercanía mayor entre dichas poblaciones que puede deberse a una similitud en sus perfiles inmunogenéticos debido a que las moléculas HLA-I características de una población son producto de un efecto fundador y a que la presencia de genes amerindios es considerablemente más fuerte en México y Guatemala que en Estados Unidos, donde estos genes están presentes en grupos minoritarios [24]. Además, se puede observar que las secuencias consenso de México y Estados Unidos son más parecidas entre sí que el consenso guatemalteco al consenso estadounidense; hecho que puede explicarse por la cercanía geográfica entre México y Estados Unidos y un flujo génico mayor entre estas poblaciones que entre Estados Unidos y Guatemala. Las diferencias en las secuencias consenso de una población sugieren fuertemente la fijación de mutaciones virales como respuesta a una presión selectiva inmunológica específica. El análisis de las secuencias consenso es congruente con los árboles filogenéticos ya que las secuencias consenso de las poblaciones mexicana y guatemalteca presentan diferencias que se ven reflejadas a nivel poblacional y refuerza la noción de una composición de moléculas HLA-I estrechamente relacionada, aunque con diferencias puntuales notorias en las poblaciones mexicana y guatemalteca.

Debido a que el tamaño de muestra es reducido (n=118, México: 63 y Guatemala: 55) y no se puede considerar representativo de la población de un país, no se puede excluir la posibilidad de que los patrones observados en los análisis filogenéticos realizados reflejen sólo asociaciones geográficas, producto de un patrón de transmisión local. Sin embargo, hemos mostrado evidencia sólida de la influencia de la presión selectiva inmune en la determinación de cambios adaptativos diferenciales del VIH a nivel poblacional. De cualquier manera, nuestros datos pueden servir como base para estudios más amplios y detallados que permitan predecir patrones evolutivos característicos del virus en cada una de las dos poblaciones, así como buscar blancos para el posible diseño de inmunógenos dirigidos.

Los análisis de variabilidad del gen *gag* circulante en las poblaciones mexicana y guatemalteca confirman que los genes *gag* circulantes en México y en Guatemala difieren entre sí. Es importante señalar que una de las limitaciones de estos análisis es que no consideran la recombinación de genomas virales y asume que todo cambio es producto de una mutación.

Aunque de manera general, la variabilidad del gen muestra patrones similares en ambas poblaciones, la población guatemalteca resultó significativamente más polimórfica en diferentes proteínas codificadas por *gag*, entre las que destaca la proteína de cápside, p24. Por lo que si se considera la variabilidad como un indicador indirecto de presión selectiva, la presión selectiva es mayor en la porción p24 en la población guatemalteca que en la mexicana, en la cual la presión selectiva es más fuerte en las porciones que codifican las proteínas p17, p6 y p7. Cabe recalcar que la variabilidad no es forzosamente resultado de una presión selectiva incidente y puede ser conse-

cuencia de procesos de carácter azaroso como la deriva génica, por ejemplo. Sin embargo, la distribución diferente de la variabilidad en el gen *gag* debe considerarse un punto focal para el desarrollo de vacunas y tratamiento ya que puede sentar bases para desarrollar medicamentos que consideren los blancos principales del sistema inmunológico de las poblaciones de origen amerindio, por ejemplo, para diseñar esquemas de tratamiento complementarios y potenciadores de las respuestas naturales predominantes, que resulten más efectivos que esquemas utilizados en la actualidad. Cabe destacar la importancia de profundizar investigaciones acerca del gen *gag* (sobre todo de p24) debido a su importante papel en el control de la infección por VIH y complementar estas investigaciones con caracterizaciones de otros genes virales y de posibles relaciones sinérgicas entre éstos que amplíen las posibilidades de una terapia antirretroviral efectiva y adecuada.

Con el objetivo de determinar si la alta variabilidad de una posición en la secuencia del gen *gag* era consecuencia de una presión selectiva, se llevó a cabo el análisis de tasa de sustituciones sinónimas y no sinónimas con el servidor Data Monkey y se pudo observar que la presión selectiva positiva (se refiere a que hay un cambio de nucleótido para cambiar el aminoácido que codificaba el triplete originalmente) es mayor en la población mexicana (dieciocho posiciones bajo presión selectiva positiva) que en la guatemalteca (catorce posiciones bajo presión selectiva positiva). Aunque cinco posiciones se identificaron tanto en México como en Guatemala, de manera general, se puede decir que el análisis por tasas de sustituciones sinónimas y no sinónimas confirma lo observado en los análisis de variabilidad (frecuencia de polimorfismos y análisis de entropía); aunque la presión selectiva sigue un patrón similar en las poblaciones mexicana y guatemalteca, existen diferencias en cuanto a los sitios sobre los cuales recae la misma, lo cual nos lleva de nuevo a pensar en una posible asociación de la variabilidad observada con presión selectiva inmune determinada por perfiles inmunogenéticos distintos. Los análisis de red de dependencia filogenética (PDN) a su vez, confirman las observaciones tanto de los análisis filogenéticos como de variabilidad del gen *gag* circulante en México y Guatemala. El hecho de que un número mayor de posiciones del gen circulante en la población guatemalteca coincidieran con los resultados del análisis de PDN practicado en una cohorte mexicana, sugiere fuertemente que lo que parecen ser asociaciones entre moléculas HLA-I y codones del gen *gag* de VIH frecuentes en la población mexicana, son asociaciones poco frecuentes en la población guatemalteca. Este hecho se confirmó al comparar sitios bajo presión selectiva negativa (se genera un cambio de nucleótido para conservar el aminoácido que codifica el triplete original) por tasas de sustituciones sinónimas y no sinónimas utilizando el servidor Data Monkey en la población mexicana. Se identificaron 237 sitios bajo presión selectiva negativa y se buscaron 64 posiciones en el análisis de PDN, de las cuales 31 posiciones aparecieron asociadas a moléculas HLA-I (no se muestran los resultados). Esto se debe a que existe una zona híbrida o (rango “umbral”) entre una posición bajo presión selectiva positiva y presión selectiva negativa que depende de la frecuencia de una asociación entre moléculas HLA-I y una posición del genoma de VIH en una población. Por esta razón, una posición bajo presión selectiva positiva en una población, puede reportarse como una posición bajo presión selectiva negativa en una población diferente, siendo esta asociación más frecuente en la población en la cual la presión selectiva es negativa (población mexicana) y menos



frecuente en la cual la posición está bajo presión selectiva positiva (guatemalteca). Considerando este fenómeno de zona híbrida (o rango “umbral”), es muy probable que muchas diferencias o similitudes con relación a las posiciones bajo presión selectiva en las poblaciones mexicana y guatemalteca caigan en este rango y sólo puedan detectarse ampliando el tamaño de muestra o al refinar las herramientas de análisis.

Considerando las diferencias en las secuencias consenso de México y Guatemala, se trató de determinar si estas diferencias reflejaban la fijación de una mutación de escape viral a nivel poblacional y por esta razón, se buscaron correlaciones con las posiciones diferentes en los consensos y moléculas HLA-I. Todas las posiciones mostraron relación con moléculas HLA-I ya fuera por análisis de PDN o en la base de datos de Los Alamos, con la excepción de la posición p7-*gag* 389, que no mostró relación con alelos HLA-I en ninguno de los dos análisis. La posición p24-*gag* 146 se correlaciona con el alelo HLA-C\*07 en ambos análisis, por lo que la asociación se puede considerar contundente. Además, es importante señalar que las frecuencias alélicas de las moléculas HLA-C\*07 difieren significativamente en las poblaciones mexicana y guatemalteca, por lo que esta posición puede reflejar el efecto de un perfil inmunogenético diferente entre estas poblaciones. La posición p24-*gag* 248 fue la única posición que se correlacionó con moléculas HLA tanto en el análisis de PDN como en la base de datos de Los Alamos, que está bajo presión selectiva según el análisis con Data Monkey y además, es una posición que difiere en las secuencias consenso de México y Guatemala; lo que sugiere la fijación de una adaptación viral diferente en las poblaciones mexicana y guatemalteca.

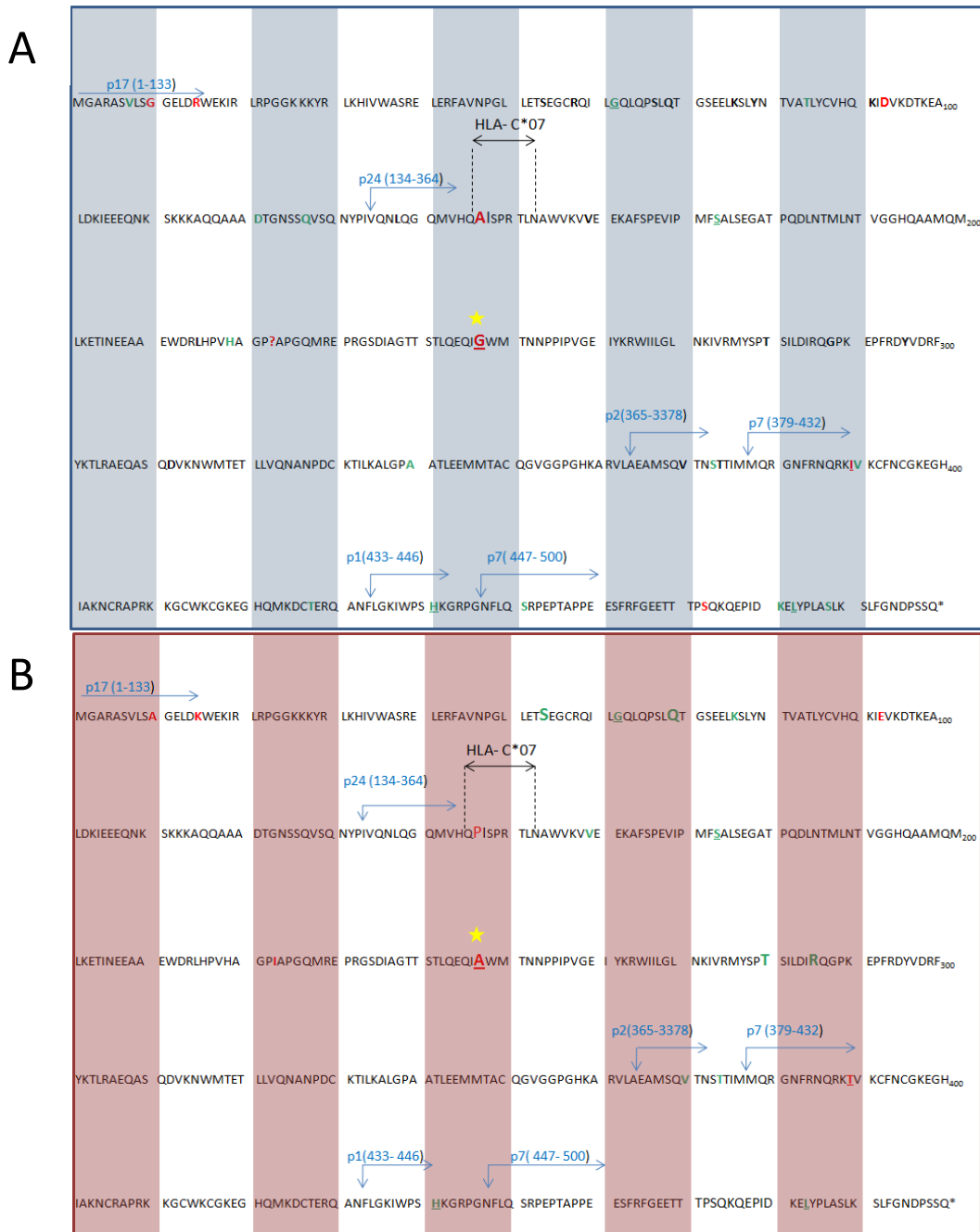
Establecer correlaciones entre moléculas HLA-I y posiciones del gen *gag* en las poblaciones estudiadas, resultó difícil ya que en lo que a la base de datos de Los Alamos respecta, los datos puestos a disposición tienen un sesgo hacia poblaciones caucásicas y poblaciones de África. Además, la falta de información del perfil inmunogenético de la población guatemalteca dificulta anticipar correlaciones claras. Se podrán determinar correlaciones más evidentes una vez que la base de datos de la cohorte guatemalteca del CIENI esté completa, ya que con la información que está disponible actualmente no se pueden reconocer diferencias significativas en las frecuencias alélicas de los genes HLA-I entre las poblaciones mexicana y guatemalteca que pueden estar influyendo en lo observado en este trabajo. Sin embargo, al analizar la Fig. 6.8 y la tabla 6.4, se puede observar que si hay varias diferencias significativas en las frecuencias de alelos HLA-I entre las poblaciones, lo que muy probablemente apoyará las observaciones realizadas en el presente estudio de adaptación viral diferencial en las dos poblaciones, una vez concluidos y ampliados los análisis.

No fue posible identificar nuevas asociaciones ni identificar epítomos presentados por los alelos HLA-I que arrojaron los análisis de PDN y la base de datos de Los Alamos; sin embargo, se pudieron identificar aminoácidos que juegan un papel importante en el desarrollo de la infección por VIH en las poblaciones mexicana y guatemalteca. Las posiciones destacadas por su relación con moléculas HLA-I y por ser diferentes en las poblaciones mexicana y guatemalteca son p24-*gag* 146, por su relación fuerte con los alelos HLA-C\*07 (las frecuencias alélicas de estos genes parece ser significativamente diferente en las poblaciones mexicana y guatemalteca) y p24-*gag* 248, por ser una posición que se relaciona con alelos HLA-I tanto en el análisis de PDN como en la base de datos de Los Alamos, también fue una posición bajo presión selectiva posi-

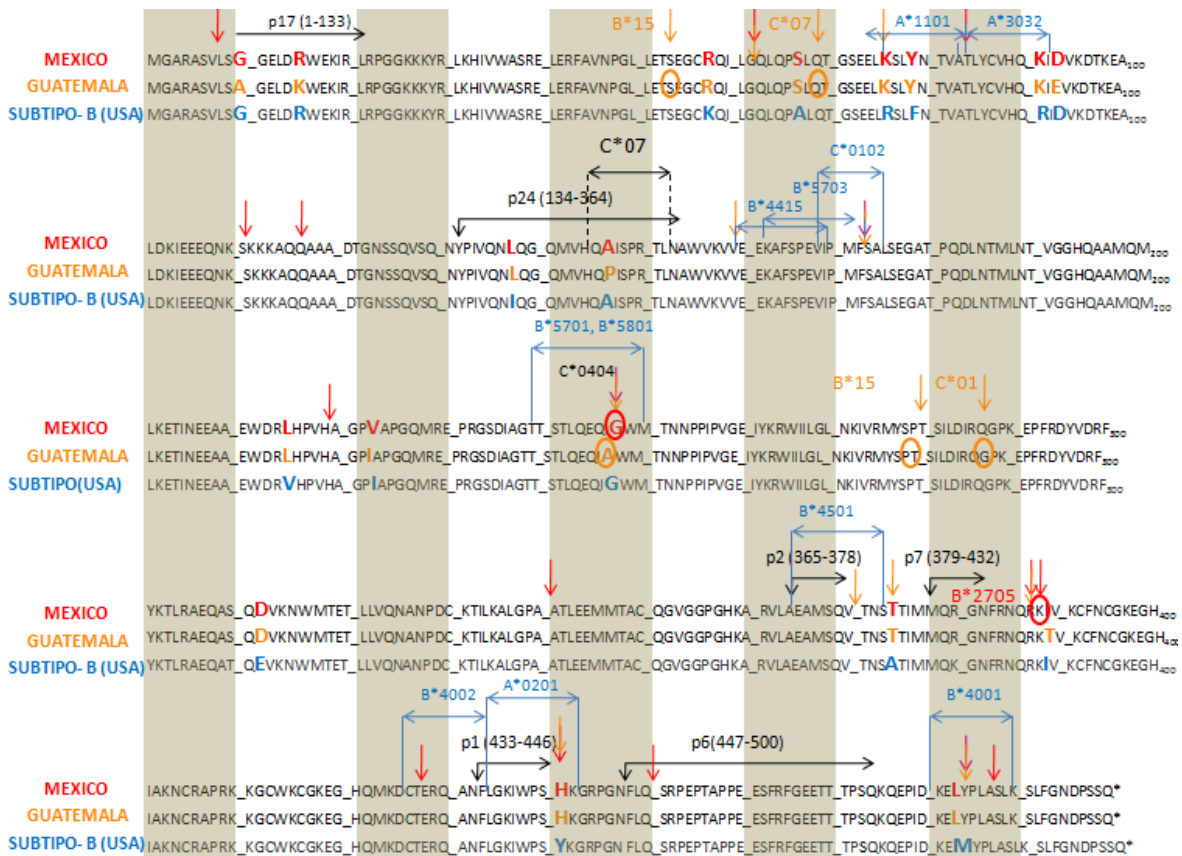
tiva según el análisis con Data Monkey tanto en la población guatemalteca como en la mexicana y además difiere en las secuencias consenso de las poblaciones. Estos resultados pueden dar pie a investigaciones más amplias y profundas sobre este tema. Estas posiciones además, resultan muy interesantes para el desarrollo de tratamiento antirretroviral ya que se ubican en la porción de *gag* que codifica la proteína p24, que es un blanco terapéutico ideal por ser una región sumamente conservada dentro del genoma del VIH.

La Fig. 6.9 resume los resultados obtenidos en los análisis con el servidor Data Monkey y las posibles correlaciones de las posiciones bajo presión selectiva positiva y alelos HLA-I que se llevaron a cabo en este estudio para caracterizar el gen *gag* que circula en las poblaciones mexicana y guatemalteca.

Aunque una de las mayores limitaciones de este estudio fue el reducido tamaño de muestra que se utilizó, tanto los análisis filogenéticos como los de variabilidad y los que buscan correlaciones entre moléculas HLA-I y posiciones del gen *gag* de VIH arrojan resultados interesantes que deben considerarse al llevar a cabo estudios más profundos y detallados sobre este tema.



**Figura 6.9. Caracterización del gen *gag* de VIH circulante en la población mexicana y guatemalteca.** La figura esquematiza el gen *gag* de VIH que circula en la población mexicana (A) y en la guatemalteca (B) caracterizado en este estudio. En rojo, se indican las posiciones en las que las secuencias consenso de México y Guatemala difieren. En verde se señalan los sitios bajo presión selectiva positiva según el análisis con el servidor Data Monkey. Las posiciones subrayadas son posiciones bajo presión selectiva positiva que coinciden en las poblaciones mexicana y guatemalteca. Los aminoácidos escritos con un tamaño de letra mayor, son aquellos que se correlacionan con moléculas HLA-I tanto en la base de datos de Los Alamos como en el análisis de red de dependencia filogenética (PDN). Las flechas bidireccionales indican una relación contundente (la posición se relaciona con el alelo HLA-I tanto en la base de datos de Los Alamos como en el análisis de PDN) entre una posición del gen *gag* de VIH y un alelo HLA-I. La estrella amarilla indica que una posición difiere en las secuencias consenso de México y Guatemala, está bajo presión selectiva positiva en ambas poblaciones y se correlaciona con alelos HLA-I en la base de datos de Los Alamos y en el análisis de PDN.



**Figura 6.10. Comparación de la proteína Gag de VIH-I de México, Guatemala y un subtipo- b circulante en Estados Unidos.** La figura muestra las secuencias de cada una de las poblaciones analizadas (Méx- rojo, Gua- anaranjado y USA- azul) destacando con colores las posiciones que difieren entre las tres secuencias. Las flechas indican las posiciones bajo presión selectiva positiva y el color de la flecha refiera a la población a la que pertenecen. A la izquierda de las flechas se indica la molécula HLA-I con la que se asocian ciertas posiciones bajo presión selectiva en el análisis de PDN. Las flechas bidireccionales azules enmarcan epítotos de alelos HLA-I frecuentes en las poblaciones caucásicas y el epítoto en negro (C\*07) representa la asociación sólida del alelo con la posición (p24- gag 146) bajo presión selectiva positiva en ambas poblaciones y tanto en La Base de Datos de Los Alamos como por el análisis de PDN. El alelo C\*0404 que asocia a la posición p24- gag 248 se indica en negro ya que es una asociación común en las poblaciones mexicana y guatemalteca pero no fue posible determinar el epítoto que la molécula HLA reconoce.



# Capítulo 8

## Conclusiones

A partir de lo observado en este estudio, se puede concluir que existen diferencias puntuales en la secuencia del gen *gag* de VIH circulante en las poblaciones mexicana y guatemalteca a pesar de ser poblaciones geográficamente cercanas y con una ancestría amerindia común. Por lo tanto, se confirma que tanto el perfil inmunogenético de una población como la ancestría, geografía y la dinámica de flujo génico entre diferentes poblaciones determinan en gran medida, las características del virus que circula en una población. Por esta razón, este análisis sienta criterios a considerar al diseñar regímenes terapéuticos y vacunas contra la infección por VIH.

La presión selectiva ejercida sobre el gen *gag* es diferente en México y Guatemala, reflejo de esto son las diferencias en la variabilidad de dichas poblaciones que si bien sigue un patrón similar en ambas poblaciones, presenta diferencias significativas en la porción de *gag* que codifica la proteína de cápside p24 ya que es mayor en la población guatemalteca.

El gen *gag* de VIH circulante en la población mexicana posee un número mayor de sitios bajo presión selectiva que la guatemalteca a pesar de ser menos variable que esta población. Esto puede explicarse por la fijación de mutaciones seleccionadas por HLA (identificadas como posiciones bajo presión selectiva negativa) en mayor proporción en la población mexicana que en la población guatemalteca, donde se observan más posiciones bajo presión selectiva positiva asociadas a HLA, en las cuales las mutaciones seleccionadas no han alcanzado niveles altos de fijación en los virus circulantes.

No fue posible trazar relaciones claras entre posiciones del gen *gag* de VIH y moléculas HLA-I pero se identificaron dos posiciones del gen que resultan de gran interés (*p24-gag* 146 y *p24-gag* 248), que deben considerarse en futuros estudios sobre este tema y que sugieren un impacto diferencial de las moléculas HLA-I en la evolución del VIH a nivel poblacional incluso en poblaciones geográficamente cercanas y con orígenes étnicos similares.



# Referencias

1. UNAIDS. (2011). Global Report- World AIDS Day Report 2011. Ingreso: 2.07.2011, 2011, de <http://www.unaids.org/en/>
2. UNAIDS. (2011). Global Report- World AIDS Day Report 2011., de <http://www.unaids.org/en/regionscountries/regions/latinamerica/>.
3. UNAIDS. (2011). Global Report- World AIDS Day Report 2011. de <http://www.unaids.org/en/regionscountries/countries/mexico>.
4. KAISER-FAMILY-FOUNDATION, H. J. (2011). HIV/AIDS Epidemic in Mexico (HIV/AIDS Policy Fact Sheet, July 2011). Ingreso: 2.07.11, 2011, de <http://www.kff.org/>
5. UNAIDS. (2011). Global Report- World AIDS Day Report 2011. de <http://www.unaids.org/en/regionscountries/countries/guatemala>
6. ÁVILA-RÍOS, S., & REYES-TERÁN, G. (2009). Virología del VIH: buscando nuevas estrategias antirretrovirales 25 años de SIDA en México. *Logros, desaciertos y retos*. (2da ed.). México: FCE/INER.
7. BALTIMORE, D. (1971). Expression of animal virus genomes. *Microbiology and Molecular Biology Reviews*, 35(3), 235.
8. GREENE, W. C., & MATIJA, B. M. (2002). Charting HIV's remarkable voyage through the cell: Basic science as a passport to future therapy. *Nature Medicine*, 8(7), 673-680.
9. PETERLIN, B. M., & TRONO, D. (2003). Hide, shield and strike back: how HIV-infected cells avoid immune eradication. *Nature Reviews Immunology*, 3, 97-107.
10. LEVY, J. A. (2009). HIV Pathogenesis: 25 Years of Progress and Persistent Challenges. *Official Journal of the International AIDS Society*, 23, 147-160.
11. LEVY, J. A. (2008). *El VIH y la patogénesis del SIDA*. México: FCE/INER.
12. SHRINER, D., LIU, Y., NICKLE, D., & MULLINS, D. C. (2006). Evolution of Intrahost HIV-1 Genetic Diversity during Chronic Infection. *Journal of Organic Evolution*, 60(6), 1165-1176.
13. JANEWAY, C. A., TRAVERS, P., WALPORT, M., & SHLOMCHIK, M. J. (2007). *Immunobiology- The Immune System in Health and Disease*. Nueva York: Garland Science.
14. GOULDER, P., & WATKINS, D. I. (2008). Impact of MHC class I diversity on immune control of immunodeficiency virus replication. *Nature Reviews Immunology*, 8(8), 619-630.



15. BROCKMAN, M. A., BRUMME, Z. L., BRUMME, C. J., MIURA, T., SELA, J., ROSATO, P. C., . . . JESSEN, H. (2010). Early Selection in Gag by Protective HLA Alleles Contributes to Reduced HIV-1 Replication Capacity That May Be Largely Compensated for in Chronic Infection. *Journal of Virology*, *84*(22), 11937–11949.
16. BRUMME, Z. L., POON, A. F. Y., CARLSON, J. M., & WALKER, B. D. (2010). Identifying HLA-Associated Polymorphisms in HIV-1. *HIV Molecular Immunology- Reviews*.
17. RUIZ-ARGÜELLES, A. (2009). Aspectos inmunológicos de la infección por VIH/SIDA 25 años de SIDA en México. *Logros, desaciertos y retos* (2da ed.). México: FCE/INER.
18. MARTINEZ-PICADO, J., PRADO, J. G., FRY, E. E., PFAFFEROTT, K., LESLIE, A., CHETTY, S., . . . GOULDER, P. (2006). Fitness Cost of Escape Mutations in p24 Gag in Association with Control of Human Immunodeficiency Virus Type 1. *Journal of Virology*, *80*(7), 3617–3623.
19. WRIGHT, J. K., BRUMME, Z. L., CARLSON, J. M., HECKERMAN, D., KADIE, C. M., BRUMME, C. J., . . . NDUN'U, T. (2010). Gag-protease-mediated replication capacity in HIV-1 subtype C chronic infection: associations with HLA type and clinical parameters. *Journal of Virology*, *84*(20), 10820-10831.
20. WAHEED, A. A., & FREED, E. O. (2011). HIV type 1 Gag as a target for antiviral therapy. *AIDS research and human retroviruses*, *28*(1), 54-75.
21. VAN-VALEN, L. (1973). A New Evolutionary Law. *MIT Press*, *1*(1), 1-30.
22. PRUGNOLLE, F., MANICA, A., CHARPENTIER, M., & GUÉGAN, J. F. (2005). Pathogen-Driven Selection and Worldwide HLA Class I Diversity. *Current Biology- Elsevier*, *15*, 1022-1027.
23. MIDDELTON, D., GONZÁLEZ-GALARZA, F. F., CHRISTMAS, S., & JONES, A. (2011). Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. Ingreso: 21.07.2011 <http://allelefrequencys.net/>
24. MÖRNER, M. (1962). *El mestizaje en la historia de la Iberoamérica*. México: Instituto Panamericano de Historia, Comisión de Historia.
25. BASTOS, S., & CAMUS, M. (2004). Multiculturalismo y pueblos indígenas: reflexiones a partir del caso de Guatemala. *Revista Centroamericana de Ciencias Sociales*, *1*(1).
26. AVILA-RÍOS, S., ORMSBY, C. E., CARLSON, J. M., VALENZUELA-PONCE, H., BLANCO-HEREDIA, J., GARRIDO-RODRÍGUEZ, D., . . . REYES-TERAN, G. (2009). Unique features of HLA-mediated HIV evolution in a Mexican cohort: a comparative study. *Journal of Retrovirology*, *6*(72), agosto 10.
27. TAMURA, K., PETERSON, D., PETERSON, N., STECHER, G., NEI, M., & KUMAR, S. (2011). Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*.

28. NEI, M., & KUMAR, S. (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press.
29. LANL. Los Alamos National Laboratory- HIV Database. Ingreso: 5.07.2011 <http://www.hiv.lanl.gov/content/index>
30. SHANNON, C., & (1948). A mathematical theory of communication. *Bell System Technology Journal*, 27(379), 423.
31. UCSD. (2012). Rapid detection of positive selection. Ingreso: 05.07.2011, 2011, de <http://datamonkey.org/>
32. KOSAKOVSKY-PONG, S. L., & FROST, S. D. W. (2005). UCSD. (2012). Rapid detection of positive selection. Ingreso: 05.07.2011, 2011, de <http://datamonkey.org/>. *Molecular Biology and Evolution*, 22(5), 1208-1222.
33. ÁVILA-RÍOS, S. (2009). *Evolución del virus de inmunodeficiencia humana: Polimorfismos en el gen pol y presión selectiva por antígeno leucocitario humano y terapia antirretroviral en población mexicana*. (PhD), Universidad Nacional Autónoma de México, México.
34. ARCHER, J. P. (2008). *The Diversity of HIV-1*. (PhD), The University of Manchester, Reino Unido
35. PILCHER, C. D., WONG, J. K., & PILLAI, S. K. (2008). Inferring HIV Transmission Dynamics from Phylogenetic Sequence Relationships. *PLOS- Medicine*, 5(3).
36. CARLSON, J. M., BRUMME, Z. L., ROUSSEAU, C. M., BRUMME, C. J., MATTHEWS, P., JAMES, C. K., . . . HECKERMANN, D. (2008). Phylogenetic Dependency Networks: Inferring Patterns of CTL Escape and Codon Covariation in HIV-1 Gag. *PLOS- Computational Biology*.
37. BARQUERA, R., ZÚÑIGA, J., HERNÁNDEZ-DÍAZ, R., ACUÑA-ALONZO, V., MONTOYA-GAMA, K., MOSCOSO, J., . . . GRANADOS, J. (2008). HLA class I and class II haplotypes in admixed families from several regions of Mexico. *Molecular Immunology*, 45(4), 1171-1178.
38. SHAO, Y., & WILLIAMSON, C. (2012). The HIV-1 Epidemic: Low- to Middle-Income Countries. *Cold Spring Harb Perspective Med.*, 2(3).
39. AGHG. Academia de Geografía e Historia de Guatemala. Ingreso: 30.08.2011, 2011, de <http://www.academiageohist.org.gt>