



**UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO**

---

---

PROGRAMA DE DOCTORADO EN CIENCIAS BIOMÉDICAS  
FACULTAD DE MEDICINA

**“Evolución del virus de inmunodeficiencia humana:  
Polimorfismos en el gen *pol* y presión selectiva por  
antígeno leucocitario humano y terapia antirretroviral  
en población mexicana”**

**T E S I S**

QUE PARA OBTENER EL GRADO DE  
**DOCTOR EN CIENCIAS BIOMÉDICAS**

PRESENTA:

**SANTIAGO AVILA RÍOS**

TUTOR:

**DR. GUSTAVO REYES TERÁN**

COMITÉ TUTORAL:

**DR. EDUARDO GARCÍA ZEPEDA**

**DR. LUIS PADILLA NORIEGA**

MÉXICO D. F., 2009



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.

*A mamá y papá, ejemplo cotidiano y cariño incondicional.*

*A Diego, siempre presente.*

*A Lupe, por su enorme interés y cariño.*

*A Julián, mi apoyo constante.*

泣いて 笑って 悩んで 起きて  
感じる全てが自分になってく  
生きてるって 自分であるって  
当たり前の幸せが嬉しい  
信じるって 頑張るって  
その度何度もやり直して  
人間ってそんなものね  
許し逢えるって素晴らしい

助け合って 肩貸し合って  
少し進んでは立ち止まって  
息を吸って 深呼吸して  
たまにはため息ついたりしながら  
泣いて 笑って 悩んで おきて  
感じる全てが自分になってく  
人間ってそんなものね  
許し逢えるってすばらしい

De:「人間ってそんなものね」 por Kokia

Lloramos, reimos, nos preocupamos, despertamos  
Todos nuestros sentimientos nos hacen ser quienes somos  
Estar vivos, existir  
Esas cosas simples nos hacen ser felices  
Creemos, nos esforzamos  
Volvemos a empezar una y otra vez  
Los seres humanos somos así  
El podernos perdonar es maravilloso

Nos ayudarnos unos a otros, nos prestamos un hombro  
para recargarnos,  
Avanzamos un poco solo para luego de pronto detenernos  
Respiramos, tomamos un respiro profundo,  
Mientras ocasionalmente suspiramos  
Lloramos, reimos, nos preocupamos, despertamos  
Todos nuestros sentimientos nos hacen ser quienes somos  
Los seres humanos somos así  
El podernos perdonar es maravilloso

De: “Los seres humanos somos así” por Kokia

## **AGRADECIMIENTOS**

Agradezco al Dr. Gustavo Reyes Terán, director del Centro de Investigación en Enfermedades Infecciosas (CIENI) del Instituto Nacional de Enfermedades Respiratorias (INER) y mi tutor, por su enorme y constante apoyo, por su confianza, por su interés en mi formación y por abrirme las puertas de su grupo de investigación.

Agradezco a Humberto Valenzuela, Daniela Garrido, Claudia García y Juan Blanco, del CIENI del INER, por su valioso trabajo, constante apoyo y su amistad. Este trabajo no habría sido posible sin su ayuda.

Agradezco a Christopher E. Ormsby, del CIENI del INER, su enorme apoyo y esfuerzo en el análisis de datos y su valiosa contribución en la discusión del contenido de este estudio.

Agradezco a Ramón Hernández y Verónica Quiroz, del Laboratorio de Diagnóstico Viroológico del CIENI del INER, por su enorme contribución al presente estudio mediante la realización de pruebas de carga viral y genotipo de VIH.

Agradezco a Edna Rodríguez, del CIENI del INER, por su valiosa contribución con los ensayos de conteo de linfocitos T CD4 positivos y su siempre alegre disposición de apoyo.

Agradezco a Carolina Demeneghi, Mario Preciado, Silvia del Arenal, Raymundo González, Israel Molina por su valioso trabajo en la sección de toma de muestra del CIENI del INER, con una constante disposición para apoyar el presente estudio.

Agradezco a los médicos del CIENI del INER Akio Murakami, María Gómez-Palacio, José Luis Sandoval, Daniela de la Rosa, Ricardo Stanley Vega, Cristina Sánchez, Karla Romero, Jorge Ibarra por su enorme ayuda en el reclutamiento de pacientes para el presente estudio.

Agradezco a los médicos de los estados Indiana Torres (Puebla), Beatriz Ramírez (Estado de México), Adakatia Armenta (Guerrero), Jaime Andrade y Lucero González (Jalisco), por contribuir en el reclutamiento de pacientes de diversas regiones del país para integrar la cohorte de estudio del presente trabajo.

Agradezco a Zeidy Arenas, del CIENI del INER, por su intensa labor administrativa y su valiosísimo y constante apoyo hacia mi persona y hacia el presente proyecto.

Agradezco a Sandra Zamora, Rosalinda Hernández, Eduardo López, Carolina Pérez, Lourdes Medina, del CIENI del INER, por su apoyo administrativo y organizacional, fundamental para la realización del presente estudio.

Agradezco al Dr. Enrique Espinosa, del CIENI del INER, por su iniciativa y apoyo en el diseño del presente estudio, así como por sus aportaciones en la discusión de los resultados obtenidos.

Agradezco al Dr. Joel Vázquez Pérez, del CIENI del INER, por su constante disposición de ayuda en los aspectos técnicos asociados con las metodologías de laboratorio empleadas en el presente estudio.

Agradezco al Dr. Jonathan M. Carlson y al Dr. David Heckerman, del eScience Group, Microsoft Research, Redmond, Washington, USA, por su valiosa contribución en la aplicación del modelo evolutivo de redes de dependencia filogenética a nuestros datos, así como sus aportaciones en la discusión de resultados y la elaboración del manuscrito.

Agradezco a la Dra. Zabrina L Brumme, del Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology and Harvard, Boston, Massachusetts, USA y de la Faculty of Health Sciences, Simon Fraser University, Burnaby, British Columbia, Canada, por su enorme ayuda en la discusión de resultados, en la elaboración del manuscrito y por su genuina disposición para colaborar con nuestro grupo.

Agradezco al Dr. Simon Mallal y a la Dra. Mina John, del Center for Clinical Immunology and Biomedical Statistics, Royal Perth Hospital and Murdoch University, Perth, Australia, por su interés de colaboración, sus contribuciones en la elaboración del manuscrito y por proveer acceso a los datos de los pacientes australianos de la cohorte IHAC.

Agradezco al Dr. Bruce Walker, al Dr. Richard Harrigan y al British Columbia Centre for Excellence in HIV/AIDS por permitir el uso de los datos de las cohortes IHAC y HOMER.

Agradezco al Dr. Eduardo García Zepeda y al Dr. Luis Padilla Noriega por su asesoría académica como parte de mi comité tutorial.

Agradezco a la Universidad Nacional Autónoma de México y al Programa de Doctorado en Ciencias Biomédicas por su apoyo para la realización de este proyecto.

Agradezco al Instituto Nacional de Enfermedades Respiratorias, sede de realización del presente proyecto.

Agradezco a CONACyT por su apoyo económico (becario 161138).

Agradezco a todos los pacientes que voluntariamente decidieron participar en el presente estudio. Su contribución fue sumamente valiosa para la generación de información sobre la epidemia de VIH/SIDA en nuestro país con posibles aplicaciones al conocimiento mundial de esta enfermedad.

Este trabajo fue financiado por CONACyT U48159M, Fundación México Vivo y la Comisión de Equidad y Género de la H. Cámara de Diputados.

## ABREVIATURAS ÚTILES

Ac	Anticuerpo
ANOVA	Análisis de Varianza
AP2	Proteína Adaptadora 2
ARF6	Factor de ADP-ribosilación 6
ARV	Antirretroviral
CA	Cápside (p24)
CCD	Dominio del Núcleo Catalítico
CDK9	Cinasa dependiente de Ciclina 9
CIENI	Centro de Investigación en Enfermedades Infecciosas
CMV	Citomegalovirus
cPPT	Tracto Central de Polipurina
CRF	Forma Recombinante Circulante
CTD	Dominio Carboxilo Terminal
CTL	Linfocito T Citotóxico
CTS	Secuencia Central de Terminación
Cul5-EloBC	Culina 5 – Elonguina B – Elonguina C
CYCT1	Ciclina T1
dN	Tasa de Sustituciones No Sinónimas
DNA	Ácido Desoxirribonucleico
dS	Tasa de Sustituciones Sinónimas
ELB	Excoffier-Laval-Balding
EM	Maximización de Expectancia (Expectation-Maximization)
FDR	Razón de Descubrimientos Falsos
GALT	Tejido Linfoide Asociado al Intestino
HAART	Tratamiento Antirretroviral Altamente Activo
HLA	Antígeno Leucocitario Humano
HR	Heptad Repeats
HSH	Hombres que tienen sexo con hombres
IBD	Dominio de Unión a Integrasa
IHAC	International HIV Adaptive Collaboration
IN	Integrasa
INER	Instituto Nacional de Enfermedades Respiratorias
Inr	Iniciador
K2P	Kimura 2 Parámetros
LTR	Repeticiones Terminales Largas
MA	Matriz (p17)
MAC	Complejo <i>Mycobacterium avium</i>
MHC	Complejo Principal de Histocompatibilidad
ML	Máxima Verosimilitud (Maximum Likelihood)
mRNA	Acido Ribonucleico Mensajero
MVB	Cuerpos Multivesiculares
NC	Nucleocápside (p7)
Nef	Factor Negativo
NES	Secuencia de Exportación Nuclear
NJ	Neighbor-Joining
NLS	Señal de Localización Nuclear
NNRTI	Inhibidores de la Transcriptasa Inversa No Análogos de Nucleósidos
NRTI	Inhibidores de la Transcriptasa Inversa Análogos de Nucleósidos
NTD	Dominio Amino Terminal
ORF	Marco Abierto de Lectura
PAK2	Cinasa 2 Activada por p21
PBMC	Células Mononucleares de Sangre Periférica

PCR	Reacción en Cadena de la Polimerasa
PDN	Redes de Dependencia Filogenética
PI	Inhibidores de Proteasa
PIC	Complejo de Preintegración
PLSD	Diferencias Mínimas Significativas Protegidas
PR	Proteasa
R	Sitio de Inicio de la Transcripción
RE	Retículo Endoplásmico
Rev	Regulador de la Expresión Viral
RNA	Acido Ribonucleico
RNAPII	RNA Polimerasa II
RR	Riesgo Relativo
RRE	Elemento de Respuesta a Rev
RT	Transcriptasa Inversa
RTC	Complejo de Retrotranscripción
SIDA	Síndrome de Inmunodeficiencia Adquirida
SIV	Virus de Inmunodeficiencia del Simio
SLAC	Single-Likelihood Ancestor Counting
SSP	Iniciador Secuencia Específico
SU	Glicoproteína de superficie (gp120)
TAR	Elemento de Respuesta de Transactivación
Tat	Proteína Transactivadora
TCR	Receptor de Células T
TM	Glicoproteína Transmembranal (gp41)
UNG2	Uracil DNA Glicosilasa 2
URF	Formas Recombinantes Únicas
Vif	Factor de Infectividad Viral
VIH	Virus de Inmunodeficiencia Humana
WHO	Organización Mundial de la Salud



# ÍNDICE

<b>0. RESUMEN</b>	<b>10</b>
<b>1. INTRODUCCIÓN</b>	<b>11</b>
<b>1.1 PANORAMA GLOBAL DE LA EPIDEMIA DE VIH/SIDA</b>	<b>11</b>
<b>1.2 LA EPIDEMIA DE VIH/SIDA EN LATINOAMÉRICA Y EN MÉXICO</b>	<b>13</b>
<b>1.3 ORGANIZACIÓN DEL GENOMA Y ESTRUCTURA DEL VIH</b>	<b>15</b>
<b>1.4 ASPECTOS BÁSICOS DEL CICLO DE REPLICACIÓN DEL VIH</b>	<b>18</b>
1.4.1 Entrada del virus a las células	18
1.4.2 Eventos tempranos después de la entrada	21
1.4.3 Integración	23
1.4.4 Transcripción del genoma viral	24
1.4.5 Salida del virus de las células	25
1.4.6 Funciones de las proteínas accesorias/reguladoras	27
1.4.6.1 Vpr	27
1.4.6.2 Vpu	28
1.4.6.3 Vif	29
1.4.6.4 Nef	30
<b>1.5 HISTORIA NATURAL DE LA INFECCIÓN POR VIH</b>	<b>32</b>
1.5.1 Etapas clínicas de la infección por VIH	32
1.5.2 El tratamiento antirretroviral y sus implicaciones para el control viral	34
<b>1.6 EPIDEMIOLOGÍA MOLECULAR DEL VIH</b>	<b>35</b>
1.6.1 Diversidad Genética	36
1.6.2 Origen del VIH	37
1.6.3 Resistencia a fármacos antirretrovirales y variabilidad viral	39
<b>1.7 EVOLUCIÓN DEL VIH</b>	<b>39</b>
1.7.1 El VIH es un virus con gran capacidad de variación	39
1.7.2 Los genes HLA de clase I definen y restringen las respuestas de linfocitos T citotóxicos	41
1.7.3 La respuesta inmune como presión selectiva actuando sobre el VIH	43
1.7.4 La respuesta de CTLs restringida por los genes HLA de clase I es una fuerza selectiva importante dirigiendo la evolución intra-individuo del VIH	45
1.7.5 La selección por CTLs tiene un impacto en la evolución del VIH a nivel poblacional	46
1.7.6 Limitaciones y dificultades para calcular la magnitud de la influencia de los genes HLA en la evolución del VIH a nivel poblacional	48
1.7.7 Implicaciones del polimorfismo de los genes HLA de clase I para la evolución del VIH	51
<b>2. PLANTEAMIENTO DEL PROBLEMA</b>	<b>52</b>
<b>3. OBJETIVOS</b>	<b>54</b>
<b>3.1 OBJETIVO GENERAL</b>	<b>54</b>
<b>3.2 OBJETIVOS PARTICULARES</b>	<b>54</b>
<b>4. HIPÓTESIS</b>	<b>55</b>
<b>5. MÉTODOS</b>	<b>55</b>
<b>6. RESULTADOS</b>	<b>65</b>
<b>6.1 CARACTERÍSTICAS GENERALES DE LA COHORTE ESTUDIADA</b>	<b>65</b>
<b>6.2 CARACTERÍSTICAS INMUNOGENÉTICAS DE LA COHORTE DE INDIVIDUOS VIH POSITIVOS MEXICANOS</b>	<b>68</b>

6.2.1 Frecuencias alélicas y haplotípicas de los genes HLA-A, B y C en una cohorte de individuos VIH positivos mexicanos	68
6.2.2 Diferencias en frecuencias alélicas HLA entre la cohorte de individuos VIH positivos mexicanos y la población mexicana general	71
6.2.3 Contexto inmunogenético único e la cohorte de individuos VIH positivos mexicanos: comparación con otras cohortes de individuos VIH positivos y VIH negativos del mundo	73
6.2.4 Desequilibrio de ligamiento	76
6.2.5 Desviaciones del equilibrio de Hardy-Weinberg	78
<b>6.3 ASOCIACIONES ENTRE LA PRESENCIA DE ALELOS HLA Y LOS VALORES DE CARGA VIRAL EN LA COHORTE DE PACIENTES MEXICANOS</b>	<b>78</b>
<b>6.4 ANÁLISIS DE RESISTENCIA PRIMARIA DE VIH EN LA COHORTE DE INDIVIDUOS MEXICANOS</b>	<b>82</b>
<b>6.5 RELACIONES FILOGENÉTICAS ENTRE SECUENCIAS DE VIH CIRCULANTES EN EL CENTRO / SUR DE MÉXICO</b>	<b>81</b>
<b>6.6 EVOLUCIÓN DEL VIH MEDIADA POR LOS GENES HLA DE CLASE I EN UNA COHORTE DE INDIVIDUOS MEXICANOS</b>	<b>86</b>
<b>6.7 DIFERENCIAS EN LA EVOLUCIÓN DEL VIH ENTRE SECUENCIAS DE VIRUS LIBRES EN PLASMA Y PROVIRUS INTEGRADOS EN PBMCs</b>	<b>95</b>
<b>6.8 DETECCIÓN DE POSICIONES EN EL GENOMA VIRAL BAJO PRESIÓN SELECTIVA A TRAVÉS DE MODELOS MATEMÁTICOS DE CODONES</b>	<b>99</b>
<b>6.9 EFECTO DEL TRATAMIENTO ARV EN LA EVOLUCIÓN DEL VIH A NIVEL POBLACIONAL</b>	<b>103</b>
<b>7. DISCUSIÓN</b>	<b>107</b>
<b>7.1 CARACTERÍSTICAS ÚNICAS DE LA EVOLUCIÓN DEPENDIENTE DE HLA DEL VIH EN UNA COHORTE DE INDIVIDUOS MEXICANOS</b>	<b>107</b>
<b>7.2 DIFERENCIAS EN LA EVOLUCIÓN ASOCIADA A HLA DEL VIH EN EL COMPARTIMENTO DE PROVIRUS INTEGRADOS EN PBMCs Y DE VIRUS LIBRE EN PLASMA</b>	<b>114</b>
<b>7.3 IMPACTO DEL TRATAMIENTO ARV EN LA EVOLUCIÓN DEL VIH A NIVEL POBLACIONAL</b>	<b>118</b>
<b>7.4 POSIBLES FACTORES GENÉTICOS ASOCIADOS A PROGRESIÓN RÁPIDA O LENTA DE LA ENFERMEDAD POR VIH, O A UNA MAYOR SUSCEPTIBILIDAD DE INFECCIÓN POR EL VIRUS EN LA POBLACIÓN MEXICANA</b>	<b>119</b>
<b>8. CONCLUSIONES</b>	<b>121</b>
<b>9. PERSPECTIVAS</b>	<b>123</b>
<b>10. REFERENCIAS</b>	<b>125</b>
<b>11. APÉNDICES</b>	<b>139</b>

## 0. RESUMEN

**Antecedentes:** Evidencia abundante apunta a que la evolución mediada por los genes del antígeno leucocitario humano (HLA) de clase I del virus de inmunodeficiencia humana (VIH) sigue patrones altamente estereotípicos que resultan en huellas asociadas a HLA a nivel poblacional. Sin embargo, no se conoce si distribuciones de frecuencias HLA características en diferentes poblaciones han resultado en huellas únicas adicionales en la evolución del virus.

**Métodos:** Se aplicó el modelo de redes de dependencia filogenética, recientemente descrito, para evaluar la evolución mediada por HLA en secuencias del gen *pol* del VIH provenientes de virus libre en plasma y de provirus integrados en células mononucleares de sangre periférica (PBMCs) de una cohorte inmunogenéticamente única de individuos del centro/sur de México. Nuestros datos se compararon con datos obtenidos de la cohorte IHAC, una cohorte multicéntrica compuesta de individuos mayoritariamente caucásicos de Canadá, Australia y Estados Unidos.

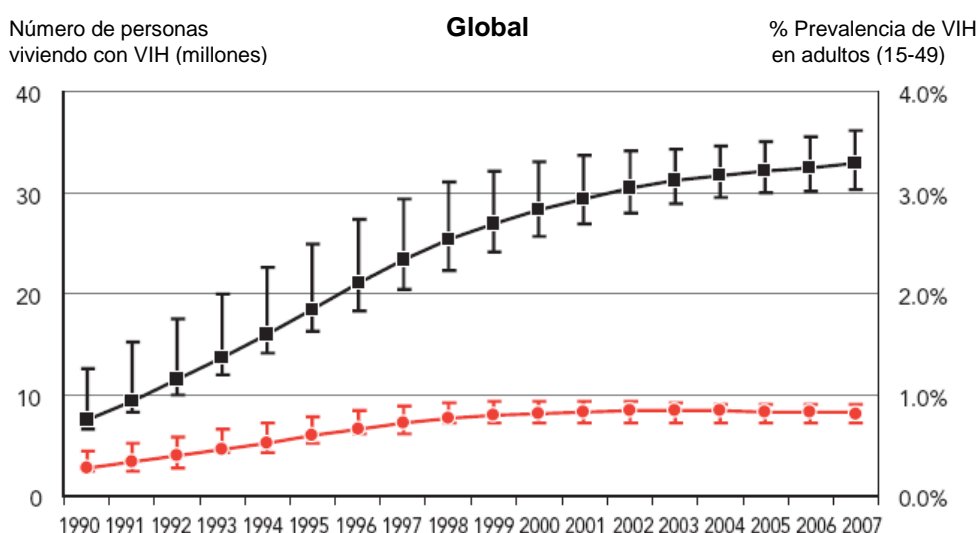
**Resultados:** Se observaron 43 asociaciones HLA-codón de VIH, representando 30 pares HLA-codón de VIH en la cohorte mexicana ( $q < 0.2$ ). Sorprendentemente, 23 (53%) de estas asociaciones fueron diferentes a las observadas en la cohorte IHAC, caracterizada por un alto poder estadístico para detectar asociaciones, lo cual sugiere la existencia de características únicas en la evolución del VIH mediada por HLA a nivel poblacional en la cohorte mexicana. Más aún, 17 de las 23 asociaciones nuevas involucraron alelos HLA cuyas frecuencias no fueron significativamente diferentes de las observadas en la cohorte IHAC, sugiriendo que su detección no fue debida a un mayor poder estadístico sino a posibles diferencias en los patrones de inmunodominancia de los epítomos blanco en la cohorte mexicana. Interesantemente, se observaron cuatro posiciones en el consenso poblacional que fueron diferentes entre las dos cohortes y tres de estas diferencias pudieron ser explicadas por selección asociada a HLA. Adicionalmente, se observaron asociaciones HLA-codón de VIH diferentes al comparar la selección mediada por HLA en virus de plasma y provirus archivados en PBMCs a nivel poblacional, con un número significativamente menor de asociaciones en las secuencias provirales.

**Conclusiones:** Nuestros datos apoyan una evolución universal del VIH mediada por HLA a nivel poblacional, resultando en huellas evidentes asociadas a HLA en los virus circulantes. Sin embargo, nuestros datos también sugieren fuertemente que contextos inmunogenéticos únicos en diferentes poblaciones infectadas por el virus pueden dirigir la evolución del VIH de una manera particular, al observarse asociaciones HLA-codón de VIH particulares que son determinadas por distribuciones de frecuencias HLA específicas. Nuestro análisis también sugiere una evolución asociada a HLA dinámica en el VIH con menos asociaciones HLA-codón de VIH en el compartimento proviral, el cual está presuntamente enriquecido en secuencias virales tempranas archivadas, en comparación con el compartimento de virus libre en plasma. Estos resultados resaltan la importancia de realizar estudios comparables de evolución del VIH en poblaciones inmunogenéticamente diferentes a nivel global.

# 1. INTRODUCCIÓN

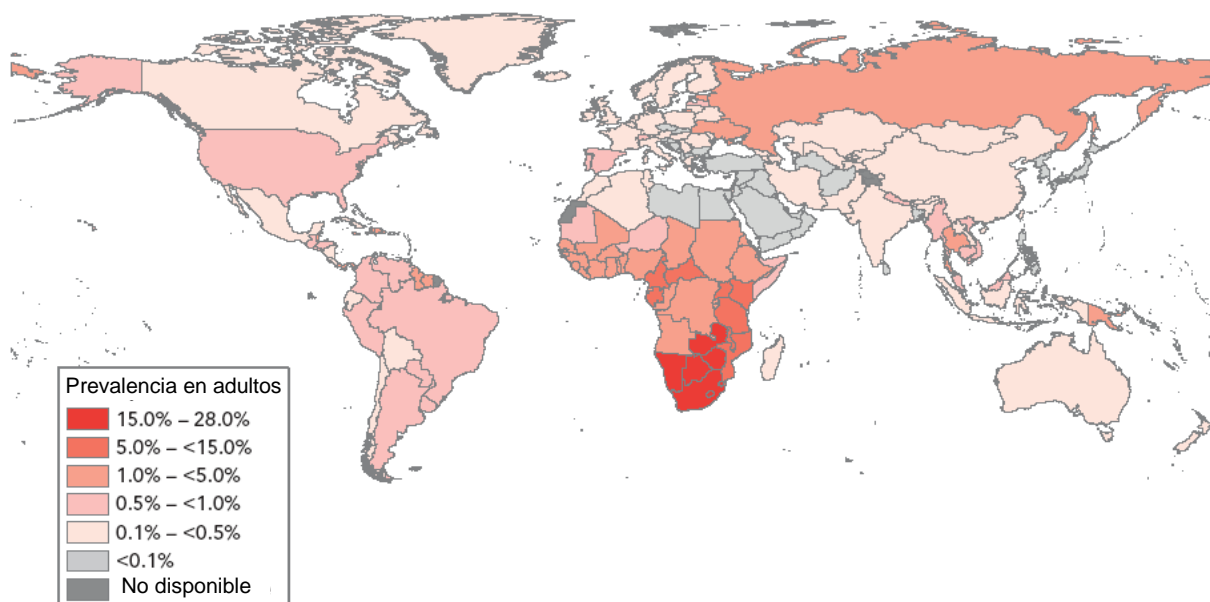
## 1.1 PANORAMA GLOBAL DE LA EPIDEMIA DE VIH/SIDA

A 25 años del descubrimiento del virus de la inmunodeficiencia humana (VIH) como agente causal del síndrome de inmunodeficiencia adquirida (SIDA), la epidemia continúa imponiendo una carga enorme no sólo en el aspecto de la salud, sino también en las esferas social, política y económica a nivel mundial. Un estimado de 33.2 millones de personas viven actualmente con el virus en el mundo (Figura 1.1) con una tasa de prevalencia altamente variable en diferentes regiones del mundo que permanece menor al 1% a nivel global (Figura 1.1 y 1.2) [1]. Los últimos estimados disponibles indican que durante el 2007 se registraron 2.7 millones de infecciones nuevas con un 45% de éstas presentándose en personas jóvenes de 15 a 24 años de edad [1]. Desde 1981, unos 58 millones de personas han sido infectadas por el virus, de las cuales 25 millones han fallecido [2]. A nivel global, la epidemia afecta por igual a hombres y mujeres, componiendo estas últimas el 50% de la población infectada (Figura 1.3). El problema por el VIH/SIDA es especialmente grave en África Subsahariana donde se concentran 22 millones de personas infectadas (67% del total de las personas infectadas en el mundo), con una prevalencia tan alta como el 28% en la población general en algunas regiones y siendo mujeres cerca del 60% de la población afectada [1, 2] (Figura 1.2). La región de Africa Subsahariana alberga también a cerca del 90% de los 2 millones de niños menores de 15 años que viven con VIH [1].



**Figura 1.1. Número total estimado y prevalencia de personas viviendo con VIH a nivel global de 1990 a 2007.** Se muestran datos de personas mayores de 15 años. En círculos se muestran los datos globales de prevalencia y en cuadrados el número total de individuos infectados en el mundo. Las barras verticales representan rangos alrededor del estimado. Tomado de [1].

Aunque la terapia antirretroviral (ARV) ha transformado el manejo y pronóstico de la infección trayendo una gran esperanza para las personas que viven con VIH, solo una tercera parte de las personas en países en desarrollo que los requieren tienen acceso a los medicamentos antirretrovirales. Dada la magnitud y el carácter global que ha tomado la epidemia del VIH/SIDA, el interés y el destino de recursos para el estudio del virus ha sido enorme. Sin embargo, aún con todo el esfuerzo realizado en el entendimiento de la virología e inmunopatogénesis de este notable patógeno, una estrategia factible que pueda solucionar el problema mundial del VIH/SIDA se encuentra todavía lejos. El fracaso del primer ensayo clínico avanzado de una prometedora vacuna candidata basada en respuestas de células T ha representado un fuerte golpe para el optimismo de la comunidad científica en el campo de la vacunología contra el VIH [3]. Más aún, la gran diversidad genética y capacidad adaptativa características del virus, aunadas a la notable variabilidad y complejidad de la respuesta inmune en las diferentes poblaciones del mundo, representan un reto enorme para el desarrollo de una vacuna efectiva contra el virus. Profundizar en el conocimiento de la replicación del VIH, en las múltiples interacciones de los componentes virales con la maquinaria molecular de la célula hospedera y en las características particulares de la compleja inmunopatogénesis viral, así como en la identificación de las respuestas inmunes efectivas para el control del virus podría llevar finalmente a la elucidación de estrategias útiles de control retroviral.



**Figura 1.2. Prevalencia de infección por VIH en diferentes regiones del mundo en 2007.** Se muestran prevalencias de infección por VIH en personas mayores de 15 años. Tomado de [1].

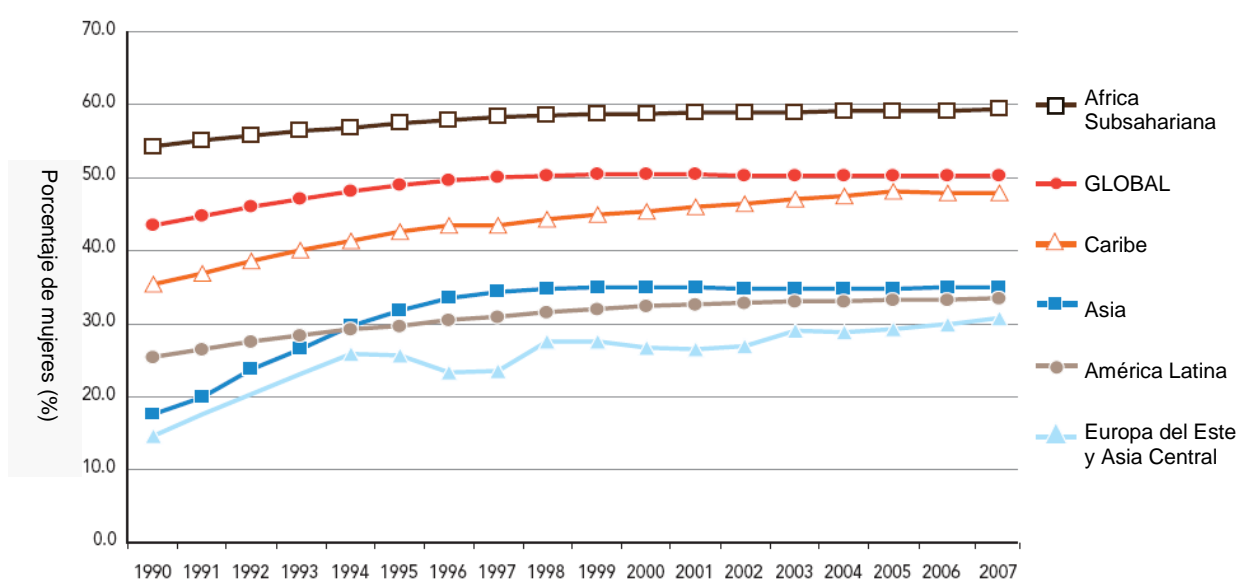
Adicionalmente, estudios comparativos enfocados a diferentes poblaciones infectadas serán necesarios para obtener un panorama más completo de la capacidad de adaptación del virus a diversas poblaciones humanas y sus posibilidades de evolución a nivel poblacional, llevando a una posibilidad más real de desarrollar una vacuna candidata efectiva para el control de la epidemia de VIH/SIDA.

## **1.2 LA EPIDEMIA DE VIH/SIDA EN LATINOAMÉRICA Y EN MÉXICO**

El VIH/SIDA en Latinoamérica tiene características únicas que plantean problemas específicos de la región para el establecimiento de programas de prevención, tratamiento y atención médica [4]. En particular, la discriminación y el estigma continúan siendo un obstáculo importante para el manejo de la epidemia. El VIH/SIDA en Latinoamérica se caracteriza por ser una epidemia centrada en grupos de mayor riesgo, incluyendo hombres que tienen sexo con hombres (HSH), trabajadores sexuales, usuarios de drogas intravenosas y migrantes (Figura 1.3) [1, 5]. Sin embargo, se ha observado un aumento en la transmisión heterosexual del virus en varios países de la región [1, 6]. Los esfuerzos de prevención, incluyendo campañas de educación, son en general desorganizados y cuentan con poco apoyo debido a que los presupuestos se dirigen principalmente al tratamiento [7]. Aunque los gobiernos de la mayoría de los países latinoamericanos han hecho esfuerzos para garantizar el acceso a fármacos ARVs a las personas infectadas con VIH que lo requieren, puntos como el acceso a y calidad de la atención médica, la detección oportuna, la prevención y el monitoreo clínico de los pacientes en tratamiento están aún muy lejos de ser adecuados [8, 9]. La falta de monitoreo de los pacientes lleva a la aparición de resistencia y a la disminución de la eficiencia del tratamiento ARV, que van de la mano con el requerimiento de esquemas de tratamiento secundarios con fármacos ARVs de mayor costo. Asimismo, el problema se complica por la falta de detección temprana, característica de la mayoría de los países latinoamericanos, con la mayoría de las personas buscando tratamiento hasta la etapa sintomática de la infección [7, 8]. Por otra parte, el fenómeno migratorio es de gran importancia para la epidemia de VIH/SIDA en Latinoamérica [9], observándose un aumento notorio en conductas de riesgo en la población de migrantes como son el cambio de prácticas sexuales y el uso de drogas intravenosas [10].

Considerando el número total de casos reportados, México ocupa el tercer lugar de prevalencia de personas infectadas con VIH en el continente americano, después de Estados Unidos y Brasil. Desde 1983, año en que se reportó el primer caso de SIDA en México, hasta 2009, se han contabilizado 124,505 casos de infección en el país [11]; sin embargo, tanto el subregistro como el retraso en la notificación son problemas presentes en

las estadísticas mexicanas [12]. De los casos de infección reportados, se estima que el 83% se presenta en hombres y que un 50% se concentra en solo cinco entidades federativas: Distrito Federal, Estado de México, Veracruz, Jalisco y Puebla. Sin embargo, al considerar tasas de infección, ajustando por número total de habitantes en cada entidad federativa, se ha observado que los estados más afectados por la epidemia son el Distrito Federal, seguido de Baja California, Yucatán, Morelos, Jalisco, Veracruz y Quintana Roo [12]. La vía de transmisión sexual es predominante (92.5%) en aquellos casos para los cuales se conoce la forma de transmisión, con 46% de los casos a nivel nacional en HSH, llegando éstos a constituir el 63% de los casos en algunas zonas urbanas [12].



**Figura 1.3. Porcentaje de adultos (mayores de 15 años) viviendo con VIH que son mujeres.** Se muestra la tendencia de cambio en el porcentaje de mujeres que viven con VIH a nivel global en comparación con regiones específicas del mundo, de 1990 a 2007. Tomado de [1].

Así, el VIH/SIDA en México puede considerarse como una epidemia concentrada, que se caracteriza por una prevalencia relativamente baja en la población general (0.3%) (Figura 1.2) y una prevalencia alta en ciertos grupos que presentan alta frecuencia de actividades de riesgo y que incluyen a HSH, hombres trabajadores sexuales, mujeres trabajadoras sexuales y usuarios de drogas intravenosas (con prevalencias aproximadas de 10-15%, 15%, 1% y 2-4% respectivamente) [12, 13]. Se cree que el problema es más notorio en ciudades fronterizas, donde la prostitución y el uso de drogas intravenosas son amplios [14, 15]. El problema en zonas rurales se ha hecho más importante con el tiempo y se cree que en gran parte está asociado al fenómeno migratorio [8]. La vía heterosexual de transmisión es más importante en zonas rurales que en zonas urbanas, con una razón de casos hombre/mujer de 4:1, en comparación con 6:1 en zonas urbanas, lo que hace

evidente la mayor vulnerabilidad de la mujer en este sector de la población [12]. Es notorio el aumento en conductas de riesgo en la población de migrantes como cambios en las prácticas sexuales y el uso de drogas intravenosas [10]; sin embargo, la insuficiencia de información y la calidad deficiente de la información disponible, aunadas a la vulnerabilidad asociada a los grupos con mayor riesgo como los migrantes, trabajadores sexuales y HSH, hacen difícil estimar el potencial de dispersión de la infección, así como su impacto en la población general, especialmente en el sector rural [10, 16]. Por eso, es imperativo mejorar los métodos de recolección de información y el introducir nuevas herramientas para el monitoreo de tendencias conductuales y del impacto de programas de intervención [6, 8].

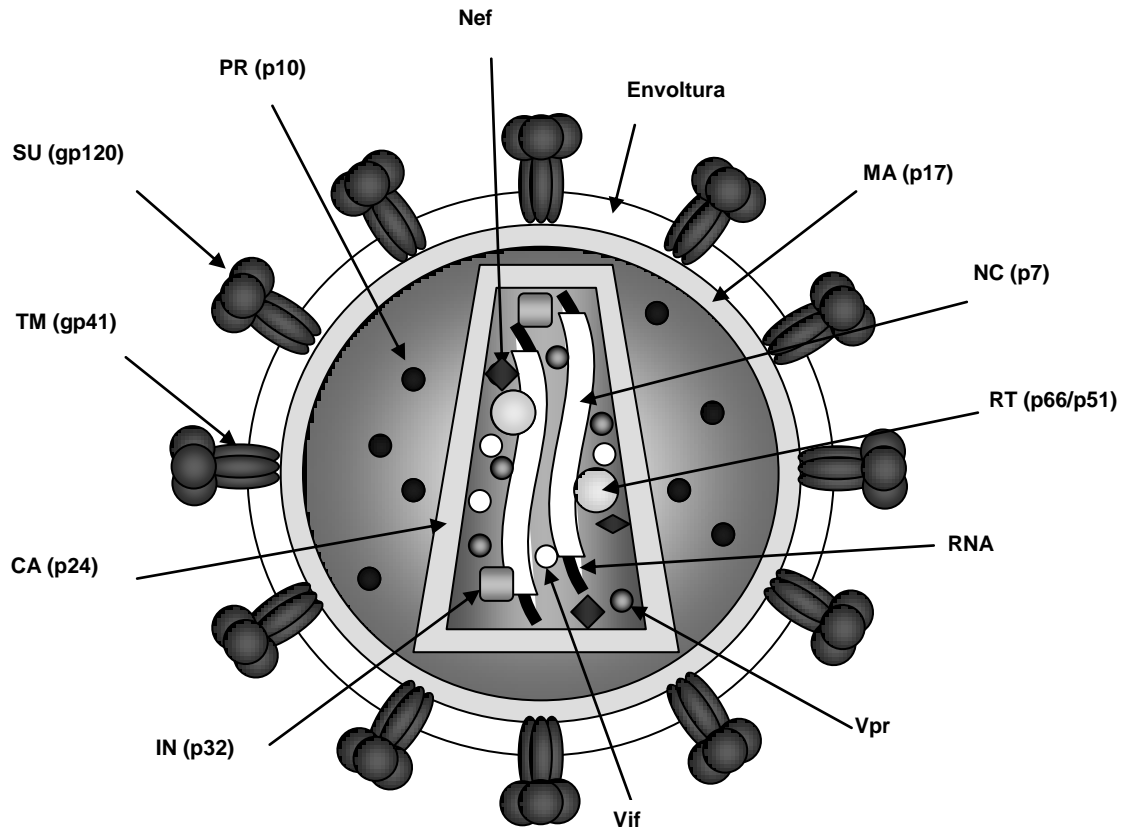
### 1.3 ORGANIZACIÓN DEL GENOMA Y ESTRUCTURA DEL VIH

El VIH pertenece al género de los lentivirus, dentro de la familia de los retrovirus. Como otros retrovirus típicos, el VIH es un virus envuelto con un genoma de RNA de cadena sencilla positiva que tiene la capacidad de transcribir su genoma a DNA en la célula hospedera mediante la enzima llamada transcriptasa inversa (Figura 1.4). El VIH comparte los tres genes característicos que codifican para las proteínas estructurales y enzimas requeridas para infecciones productivas en todos los retrovirus: *gag* (gen antígeno grupo-específico), que codifica para las proteínas de la cápside y nucleocápside virales, *pol* (polimerasa), que codifica para las enzimas virales y *env* (envoltura) que codifica para las glicoproteínas de la envoltura viral. Notablemente, el genoma del VIH contiene también varios marcos abiertos de lectura (ORFs) sobrelapados que codifican para las proteínas reguladoras / accesorias Vif, Vpr, Vpu, Tat, Rev y Nef (Figura 1.5). Las funciones de estas proteínas son variadas y muchas veces complejas. Ahora es claro que las proteínas reguladoras / accesorias son requeridas para una replicación eficiente del genoma viral y para varios aspectos de la patogénesis del virus *in vivo* (ver más adelante).

Los viriones de VIH al microscopio electrónico son partículas envueltas, de 100 a 120 nm de diámetro, con morfología heterogénea y una cápside de forma cónica. Dentro de la cápside viral se encuentran dos copias (generalmente pero no necesariamente idénticas) del genoma viral de aproximadamente 9.3 kb. La cápside y la matriz proteica del virus están conformadas por las proteínas codificadas por *gag* (las proteínas virales por convención se nombran con un número que corresponde a su tamaño): p24 (CA), que constituye la cápside viral; p17 (MA), que forma la matriz proteica asociada a la cara interna de la envoltura viral; p7 (NC), que constituye la nucleocápside asociada al RNA viral y p6, que es importante para la salida de los viriones nuevos. Las proteínas Gag son generadas como una poliproteína precursora (Pr55<sup>Gag</sup>) en la que la ubicación de las proteínas maduras sigue el orden



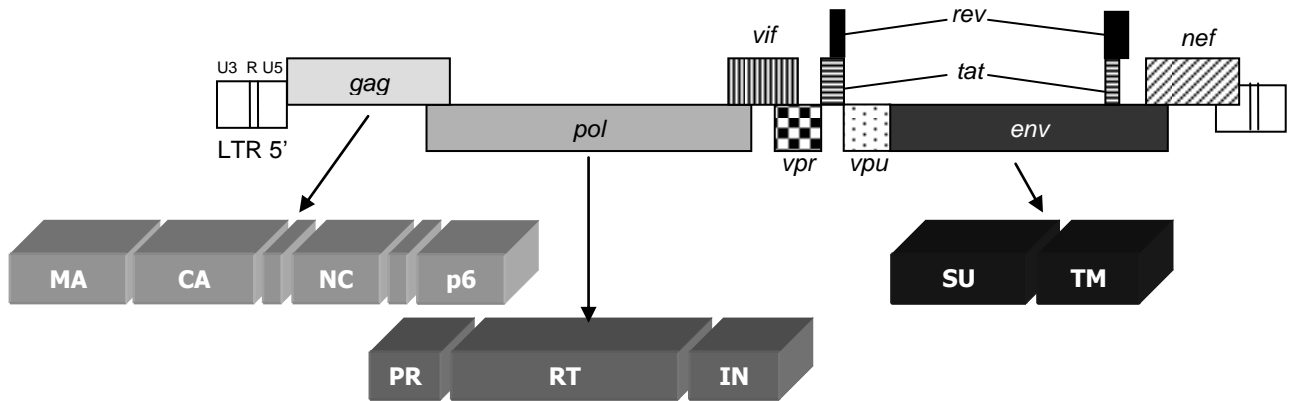
p17/p24/p2/p7/p1/p6, donde p1 y p2 son péptidos espaciadores. Las proteínas individuales Gag son generadas mediante proteólisis por la proteasa viral, en el momento de la salida de los viriones nuevos [17, 18] (Figuras 1.4 y 1.5).



**Figura 1.4. Estructura del VIH.** La membrana lipídica que recubre al virus es adquirida durante la salida de la célula hospedera y contiene las glicoproteínas virales de envoltura, gp120 y gp41, codificadas por el gen *env*. La proteína de matriz, p17, se encuentra ligada a la membrana por su miristilación. La proteína p24 constituye la cápside viral, dentro de la cual se encuentra el genoma viral, ligado a la nucleocápside, p7. Las proteínas p17, p24 y p7 son productos del gen *gag*. El gen *pol* codifica para las tres enzimas virales: proteasa (PR), transcriptasa inversa (RT) e integrasa (IN), todas contenidas dentro de los viriones. Las proteínas accesorias Vpr, Vif y Nef también forman parte del virión [17]. Adaptado de [19].

Las enzimas codificadas por *pol* se encuentran también formando parte de los viriones. Las enzimas virales se producen por autocatálisis del precursor Gag-Pol (Pr160<sup>Gag-Pol</sup>), resultando en la producción de la proteasa (PR, p10), la transcriptasa inversa (RT, p66 / p51), y la integrasa (IN, p32). La transcriptasa inversa es una DNA polimerasa dependiente de RNA heterodimérica que posee una actividad adicional de RNasa H. Esta enzima produce una copia de DNA de doble cadena a partir del RNA viral y se encuentra asociada al RNA viral en el virion. La integrasa se localiza también en la cápside y es fundamental para la integración del genoma viral en cromosomas de la célula hospedera. La proteasa viral se encuentra fuera de la cápside viral y se encarga de cortar las poliproteínas precursoras Pr55<sup>Gag</sup> y Pr160<sup>Gag-Pol</sup> en el momento de la gemación del nuevo virión,

produciendo rearrreglos estructurales y cambios morfológicos importantes conocidos como maduración. Se ha estimado que la razón de precursores Gag:Gag-Pol es de 20:1, generándose de esta manera las cantidades requeridas de las diferentes proteínas para la formación del virión completo [18] (Figura 1.4).



**Figura 1.5. El genoma del VIH.** El VIH tiene un genoma retroviral típico que incluye tres genes que codifican para las proteínas estructurales: *gag*, *pol* y *env*. Además existen seis marcos abiertos de lectura (ORFs) que originan las proteínas accesorias/reguladoras del virus: *vif*, *vpr*, *vpu*, *tat*, *rev* y *nef*. Muchos de los ORFs del genoma viral se encuentran sobrelapados. En los extremos del genoma se encuentran las repeticiones terminales largas (LTRs) que incluyen el sitio de inicio de la transcripción (R) contienen regiones promotoras (U3) y de unión de factores de transcripción celulares (U5), así como regiones que favorecen la dimerización y encapsidamiento del genoma viral dentro de los viriones. MA – matriz, CA – cápside, NC – nucleocápside, PR – proteasa, RT – transcriptasa inversa, IN – integrasa, SU – glicoproteína de superficie (gp120), TM – glicoproteína transmembranal (gp41)[17]. Adaptado de [19].

El virus adquiere una envoltura lipídica al salir de la célula infectada, la cual contiene a las proteínas virales codificadas por *env*. Las proteínas virales Env provienen de la proteólisis por enzimas celulares de la glicopoliproteína precursora gp160. La proteólisis de gp160 origina a gp120 (SU), una glicoproteína externa, y gp41 (TM), una glicoproteína transmembranal. En conjunto, gp120 y gp41 se ensamblan en trímeros para formar las espículas características de la envoltura del virus, cuya estructura tridimensional nativa ha sido recientemente resuelta [20]. Las espículas son fundamentales para la unión del virus a células blanco del hospedero y para la entrada del virus en ellas mediante la fusión de la membrana viral con la membrana de la célula hospedera. gp120, la proteína con más alta variabilidad del VIH, contiene los motivos de unión a CD4, receptor característico del virus en células blanco, y CCR5 y CXCR4, correceptores más importantes para la entrada del virus. El ensamblaje de los nuevos viriones ocurre en balsas lipídicas, lo cual se refleja en una alta razón molar colesterol/fosfolípidos de la envoltura viral. Las proteínas accesorias Nef, Vpr y Vif también se han encontrado dentro de las partículas virales (Figura 1.4), además de algunas proteínas celulares como componentes del citoesqueleto y ciclofilina A.

El genoma del VIH está flanqueado por las repeticiones terminales largas (LTR) (Figura 1.5). El LTR ha sido subdividido en tres dominios: U3, R y U5. En el contexto de RNA genómico viral, la región R media un paso crítico de la transcripción inversa al formar estructuras híbridas RNA-DNA que permiten a la RT continuar su actividad polimerizante y generar hebras de DNA de sentido negativo completas. Los LTRs contienen también una secuencia necesaria para el empaquetamiento del RNA genómico viral en los nuevos viriones. Además, los LTRs favorecen la integración del DNA viral al guiar su inserción en cromosomas de la célula hospedadora. Una vez integrado, el LTR funciona como regulador de la síntesis de mRNA. El promotor transcripcional del VIH y varios elementos reguladores están situados en la región U3 del LTR 5', siendo su principal función reclutar a la RNA polimerasa II (RNAPII) al sitio de inicio de replicación del genoma viral definido como el primer nucleótido de la región R. La región R también contiene señales de poliadenilación que son activas solamente en el LTR 3'. La región U5 contiene sitios de unión para factores de transcripción que regulan la síntesis de mRNA durante la infección por VIH [17].

## **1.4 ASPECTOS BÁSICOS DEL CICLO DE REPLICACIÓN DEL VIH**

### **1.4.1 Entrada del virus a las células**

El VIH se une a sus células blanco en el hospedero a través de la molécula CD4, la cual funciona como receptor principal del virus. Así, el VIH puede infectar a linfocitos T ayudadores CD4+, células CD4+ del linaje de los macrófagos y a algunas poblaciones de células dendríticas. CD4 se une a la glicoproteína de superficie viral gp120, lo que induce cambios conformacionales en la proteína viral que exponen los dominios de unión a un grupo de receptores de quimiocinas con siete dominios transmembranales, los cuales funcionan como correceptores del virus. Más de una docena de receptores de quimiocinas han sido implicados como correceptores del virus, sin embargo, dos son los más usados en el hospedero: CCR5 y CXCR4 [18] (Figura 1.6). La resolución de la estructura cristalina del núcleo de la glicoproteína gp120 ha mostrado que existen dos grandes dominios denominados externo e interno, unidos por la llamada hoja beta conectora [21]. El sitio de unión a CD4 en gp120 incluye una parte del dominio externo y la hoja beta conectora, y está rodeado de regiones altamente glicosiladas. El sitio de unión a los correceptores del VIH en gp120 involucra al asa V3, lo que es consistente con la influencia de esta parte de gp120 en el tropismo del virus. Además, se sabe que regiones conservadas de gp120 que son expuestas después de la unión de CD4 también tienen un papel en la interacción con el correceptor [21].

Los aislados de VIH se clasifican en la actualidad de acuerdo al uso de los correceptores virales en R5, X4 y R5/X4. Los aislados R5, también llamados macrófago-trópicos, utilizan el correceptor CCR5 para entrar en las células, pueden infectar monocitos / macrófagos y son incapaces de inducir sincicios en células en cultivo. Los virus R5 en general se replican lentamente y generan pequeñas cantidades de partículas virales progenie, aunque existen excepciones. Los aislados X4, también llamados linfotrópicos, utilizan como correceptor a CXCR4, infectan exclusivamente células T CD4+ e inducen sincicios en cultivos celulares. Los virus X4 muestran cinéticas de replicación más rápidas y liberan títulos altos de viriones. Los aislados R5/X4 tienen propiedades intermedias y se caracterizan por poder usar ambos correceptores para entrar a la célula, ser capaces de infectar monocitos, macrófagos y células T CD4+ y causar sincicios en células cultivadas, aunque en menor grado que los aislados CXCR4 [17, 18, 22]. Los virus R5 median tanto la transmisión por mucosas como la transmisión por vía intravenosa de la infección por VIH, causando una depleción marcada de linfocitos T CD4+ de memoria en el tracto gastrointestinal durante las etapas tempranas de la infección. Los virus X4 aparecen comúnmente en etapas avanzadas de la infección en un 50% de las personas infectadas, asociándose a una depleción acelerada de células T CD4+ y una progresión más rápida a SIDA. Es notorio que los ligandos naturales de los correceptores del virus pueden interferir con la infección por el VIH. Así, las  $\beta$ -quimiocinas MIP-1 $\alpha$ , MIP-1 $\beta$  y RANTES, ligandos de CCR5, interfieren con la entrada de virus R5, mientras que la  $\alpha$ -quimiocina SDF-1, evita la unión de virus X4 [23]. Estas observaciones han sido utilizadas para el desarrollo de fármacos ARVs análogos de quimiocinas para inhibir la entrada del virus a las células del hospedero al competir por la unión a los correceptores virales. Hoy se utilizan ya los antagonistas de CCR5 como una nueva familia de fármacos ARVs y algunos antagonistas de CXCR4 están en estudio como posibles candidatos para la terapia anti-VIH [24].

Además de CCR5 y CXCR4, varios otros receptores de quimiocinas se han propuesto como correceptores del virus. Estos incluyen a CCR2b, CCR3, CCR8, CCR9, APJ, STRL33/Bonzo, GPR15/BOB, CX3CR1, US28, D6 y RDC-1. La importancia biológica de la existencia de estos correceptores alternativos aún no es clara, pero su estudio podría esclarecer aspectos del tropismo selectivo de variantes virales, la transmisión de variantes macrófago-trópicas, la aparición de virus linfotrópicos durante la progresión a SIDA y diferencias en susceptibilidad hacia la infección y hacia el avance de la enfermedad en diferentes individuos [25].

La unión secuencial de CD4 y los receptores de quimiocinas a gp120 desata cambios conformacionales que exponen el dominio de fusión de gp41, la proteína transmembranal de la envoltura viral. La proteína gp41 contiene dos dominios HR (heptad repeats), uno en la región N-terminal y otro en la región C-terminal, denominados HR1 y HR2 respectivamente. Las regiones HR1 y HR2 de gp41 tiene una estructura helical de tipo “coiled coil”. Al igual que gp120, gp41 se asocia en trímeros para formar la región transmembranal de las espículas virales. En el extremo N-terminal de gp41 se localiza el péptido de fusión. Después de la unión de gp120 a CD4 y los correceptores de quimiocinas el dominio de fusión es expuesto y proyectado hacia la célula hospedera de manera parecida a un arpón. Esto trae como consecuencia el acercamiento e interacción de HR1 y HR2 de las tres moléculas gp41 asociadas, para formar una estructura en forma de racimo con seis hélices antiparalelas. Este movimiento análogo al de un pasador permite el acercamiento de la membrana viral a la membrana celular y provoca su fusión, liberando al virus en el interior de la célula [26].

Los inhibidores de fusión ya son parte del repertorio de fármacos ARVs usados en la actualidad. T20, péptido aprobado para uso clínico, es parte de la familia de los péptidos C, al ser análogo al dominio HR2 de la región C-terminal de gp41. Así, T20 compite con HR2 para interactuar con HR1 en el dominio N-terminal de gp41, impidiendo la formación del racimo de seis hélices necesario para la fusión de la membrana viral con la membrana celular. Los péptidos N, análogos de HR1, han mostrado tener menor actividad inhibitoria que los péptidos C [27].

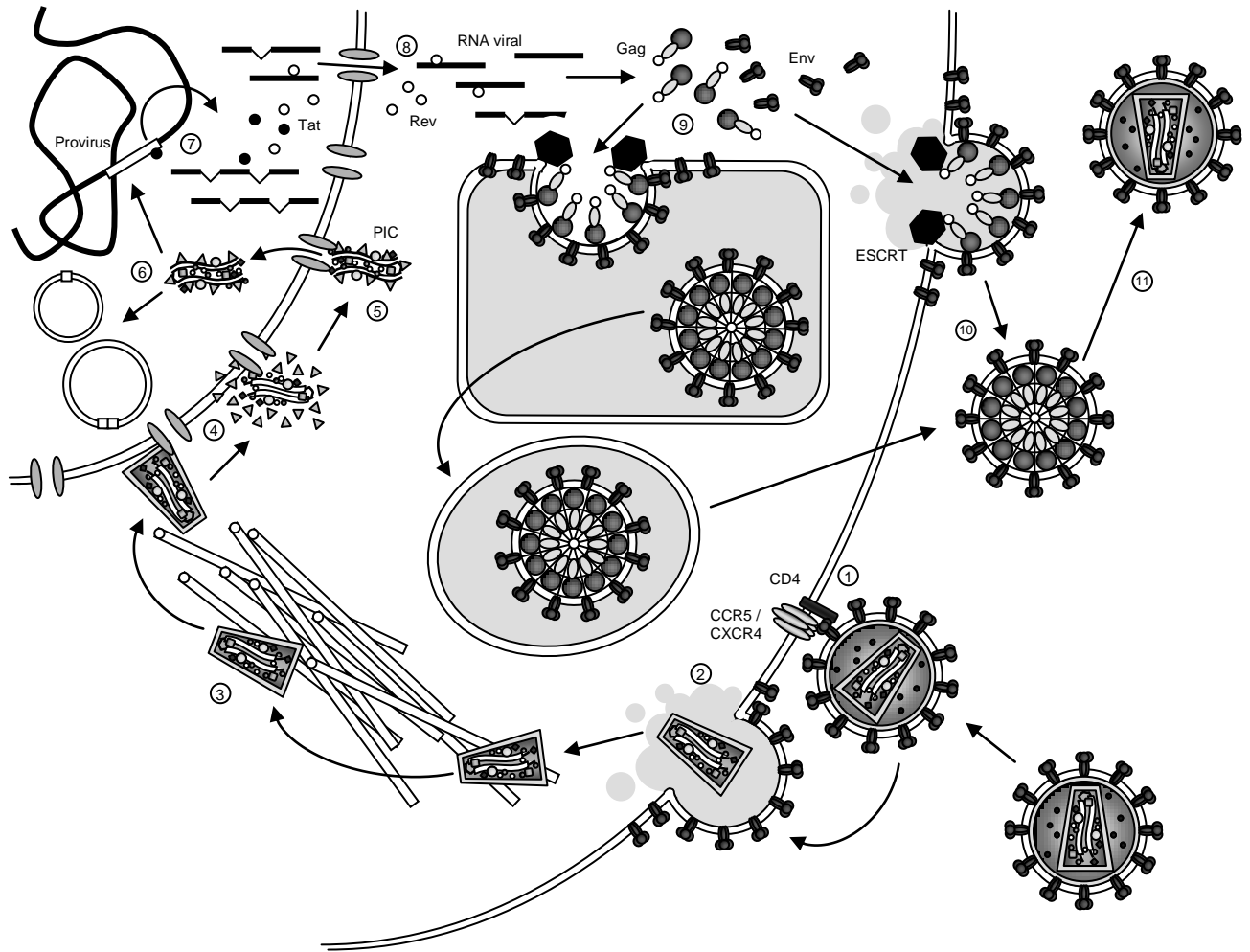
Las células dendríticas de la submucosa presentan una estructura alterna a la cual se une el VIH, llamada DC-SIGN. Esta lectina de tipo C se une con alta afinidad a gp120, pero no desencadena los cambios conformacionales necesarios para la exposición del dominio de fusión en gp41 y la consecuente entrada del virus a la célula. La unión a DC-SIGN promueve la internalización de las partículas virales a compartimentos ácidos en la célula dendrítica para ser posteriormente expuestas en la superficie celular, una vez que la célula dendrítica ha madurado y migrado a nodos linfáticos secundarios. Así, las células dendríticas promueven la infección en *trans* de células T CD4+ al llevar en su interior virus que son posteriormente expuestos en los órganos linfoides secundarios, de una manera reminiscente al histórico “Caballo de Troya” [28].

Recientemente se demostró que gp120 se une a la integrina alfa4beta7. Esta es una molécula de adhesión que media la migración de linfocitos al tejido linfoide asociado a la mucosa intestinal. La interacción con alfa4beta7 es mediada por un motivo tripeptídico del

asa V2 de gp120 que mimetiza la estructura presentada por los ligandos naturales de alfa4beta7. La unión de gp120 a alfa4beta7 induce rápidamente la expresión de LFA-1, una integrina central para el establecimiento de las sinapsis virológicas [29]. El término sinapsis virológica se ha utilizado en los últimos años para describir las zonas de contacto entre linfocitos T CD4+ y células dendríticas que facilitan la transmisión del VIH al concentrar localmente al virus y a los receptores virales [30].

#### **1.4.2 Eventos tempranos después de la entrada**

Aunque los eventos que ocurren inmediatamente después de la entrada del virus a la célula aún son controversiales, el modelo más reciente sugiere que la desencapsidación del virus no ocurre inmediatamente después de la entrada [31]. Se ha demostrado que las cápsides completas viajan hacia el núcleo dirigidas en un principio por microtúbulos y posteriormente por filamentos de actina en las proximidades del núcleo. Movimientos lentos dirigidos por actina llevan a las cápsides a encallar en los poros nucleares (Figura 1.6). La transcripción inversa del genoma viral ocurre en el llamado complejo de retrotranscripción (RTC), dentro de la cápside, en el poro nuclear. Una vez retrotranscrito el genoma viral y a nivel del poro nuclear ocurre la desencapsidación, la cual es mediada por una estructura característica de los genomas lentivirales llamada ala (flap) central de DNA. El ala central de DNA es una estructura de triple cadena de DNA que se forma por la interacción en *cis* de dos secuencias lentivirales: el tracto central de polipurina (cPPT) y la secuencia central de terminación (CTS). DNA viral carente del ala central de DNA se acumula en los poros nucleares, pero no es transportado al interior del núcleo [31]. La desencapsidación favorece la organización del complejo de preintegración (PIC), compuesto del cDNA de doble cadena, IN, RT, MA, NC y Vpr [22]. A través de un proceso poco caracterizado, el PIC es transportado al interior del núcleo, donde el cDNA de doble cadena puede integrarse en la cromatina de la célula hospedadora o circularizarse formando estructuras episomales no funcionales (Figura 1.6). Se cree que señales de localización nuclear en varias proteínas como IN [32], MA [33] y Vpr [34] son importantes para dirigir el PIC al interior del núcleo.



**Figura 1.6. Ciclo de replicación del VIH.** 1) Al unirse gp120 a CD4, receptor principal del VIH en células del hospedero, ocurren cambios conformacionales en gp120 que exponen los dominios de unión al correceptor en gp120. Los receptores de quimiocinas CCR5 y CXCR4 funcionan como los principales correceptores del virus. 2) La unión del correceptor causa otro cambio conformacional que expone el dominio de fusión en gp41, a través del cual se unen las membranas celular y viral, permitiendo la entrada del virus a la célula. 3) La cápside viaja a través de microtúbulos para anclarse en el poro nuclear. 4) Al ocurrir la transcripción inversa del genoma viral y formarse el ala central de DNA, ocurre la desencapsidación y se forma el complejo de preintegración (PIC), incluyendo a MA, RT, IN y Vpr. 5) El PIC es transportado al interior del núcleo gracias a señales de localización nuclear en MA, IN y Vpr. 6) Una vez en el núcleo, el DNA viral puede integrarse en la cromatina de la célula por la actividad de la integrasa. De manera alternativa, el DNA viral puede permanecer en forma episomal como círculos de dos LTRs, círculos de un LTR, y formas circulares reorganizadas. Estas formas episomales no generan infecciones productivas. 7) Una vez integrado, el genoma viral es transcrito, produciéndose RNAs con múltiples eventos de corte y empalme (splicing) en etapas tempranas. Estos RNAs tempranos originan a las proteínas reguladoras Tat, Rev y Nef. Tat funciona como transactivador de la transcripción de los genes virales, permitiendo la formación de transcritos largos. En etapas tardías de la transcripción se forman RNAs con uno o sin ningún evento de corte y empalme, los cuales originan a las proteínas estructurales Env, Pol y Env así como a las proteínas accesorias Vif, Vpr y Vpu. 8) Los RNAs tardíos son transportados al citoplasma por Rev, evitando así los eventos de corte y empalme. 9) Los precursores de las proteínas estructurales y los genomas virales se ensamblan en balsas lipídicas en membranas de endosomas tardíos o en la membrana plasmática. MA se dirige a membranas por su característica miristilación. 10) Para la salida de nuevos viriones de la célula, el virus usurpa a los sistemas ESCRT, ordenadores del tránsito intracelular de vesículas. Gracias a interacciones con proteínas celulares ESCRT y a la contribución de la proteína viral Vpu, los nuevos viriones pueden gemar hacia el exterior a partir de la membrana plasmática o gemar hacia el lumen de cuerpos multivesiculares (MVB), para después ser liberados por fusión de estas vesículas con la membrana plasmática. 11) Los nuevos viriones deben madurar para ser infecciosos. La proteasa viral lleva a cabo este proceso cortando las poliproteínas Gag y Gag-Pol, para generar proteínas maduras [17]. Adaptado de [19].

### 1.4.3 Integración

Para el establecimiento de una infección productiva, el DNA viral debe ser integrado en un cromosoma de la célula hospedera a través de la enzima integrasa. Esta enzima remueve inicialmente dos nucleótidos de los extremos del genoma viral (originalmente con extremos romos), para después catalizar la reacción de transferencia de hebra de DNA, generando un corte escalonado en el DNA del hospedero y uniendo el DNA viral al DNA del hospedero [17]. El DNA viral también puede tener otros tres destinos posibles que no generan infecciones productivas: Los dos extremos del DNA viral pueden unirse, formando círculos de 2 LTRs; el genoma viral puede sufrir recombinación homóloga, originando círculos de un LTR o el genoma viral puede autointegrarse en él mismo, produciendo una estructura circular rearmada [22]. Los sitios de integración en los cromosomas de la célula hospedera son importantes para el virus, ya que influyen en la expresión eficiente del genoma viral. Asimismo, los sitios de integración viral también tienen importancia para el hospedero, al poder activar o inactivar la transcripción de genes celulares, e. g. la activación insertional de oncogenes que trae como consecuencia la transformación celular hacia un fenotipo maligno. En general, se ha observado que la mayoría de los provirus se integran en sitios transcripcionalmente activos de la cromatina celular, lo que favorece una expresión eficiente de los genes virales después de la integración. Interesantemente, se ha sugerido cierta participación de factores celulares como la proteína LEDGF/p75 en dirigir la integración del VIH [35]. Fisiológicamente, LEDGF/p75 tiene un papel como proteína adaptadora, enlazando otras proteínas a la cromatina [36]. En la infección por VIH, LEDGF/p75 se une a IN y favorece la integración del DNA viral al funcionar como un enlace entre IN y la cromatina celular [37-39]. Se ha observado que LEDGF/p75 se une a la región del núcleo catalítico (CCD) de IN. Alteraciones sutiles en la interfase entre el CCD y LEDGF/p75 impiden la integración del genoma del VIH sin correlacionar necesariamente con una pérdida de la función catalítica de IN. Sin embargo, estas alteraciones afectan muy posiblemente la unión de LEDGF/p75 a IN [40]. Varios grupos han caracterizado mutaciones en IN que impiden la unión IN-LEDGF/p75, manteniendo la actividad catalítica de la enzima [40-43]. En general, estas mutaciones no hacen contacto directo con el dominio de unión a IN de LEDGF/p75 (IBD), pero muy probablemente modifican la conformación de la interfase CCD-IBD [41].

En la actualidad se utilizan ya los inhibidores de IN como parte de la terapia ARV. Los compuestos disponibles en el mercado inhiben la integración del VIH al intervenir con la reacción de transferencia de hebra de DNA. Como es esperado, mutaciones de resistencia que impiden la acción de esta familia de fármacos ya se han caracterizado en IN [44]. La



interacción LEDGF/p75-IN representa un nuevo punto de intervención para el desarrollo de posibles ARVs que funcionarían de manera diferente a los actuales inhibidores de integrasa y no compartirían los mismos patrones de mutaciones de resistencia.

#### 1.4.4 Transcripción del genoma viral

El LTR 5' del virus funciona de manera similar a como lo hacen unidades transcripcionales eucarióticas. El LTR contiene secuencias promotoras y enhancers con sitios de unión para varios factores de transcripción y una señal de poliadenilación [45]. La región U3 del LTR, río arriba del sitio de inicio de la transcripción (R), incluye al iniciador (Inr), la caja TATA y tres sitios Sp1 y su principal función es el reclutamiento de la RNA polimerasa II (RNAPII) para el inicio de la transcripción de los genes virales. Río arriba de la región promotora se encuentra el enhancer, que une factores de transcripción presentes en el núcleo de células activadas, como NF- $\kappa$ B, NFAT y proteínas de la familia ETS. El LTR también presenta una región reguladora conocida como elemento de respuesta de transactivación (TAR). Esta estructura de RNA presente en todos los mensajeros virales se une a la proteína reguladora viral Tat (proteína transactivadora). En la ausencia de Tat, la transcripción del VIH puede iniciar, pero no es eficiente para producir mensajeros largos. Tat y su cofactor celular P-TEFb, se unen cooperativamente a TAR con alta afinidad, permitiendo la producción de mensajeros largos por la RNAPII. P-TEFb está compuesto de la ciclina T1 (CYCT1) y la cinasa dependiente de ciclina 9 (CDK9) y su función consiste en fosforilar la región C-terminal de la RNAPII para la activación de la enzima [45].

Los eventos de corte y empalme son fundamentales para la producción de diversas especies de mRNA, responsables de la síntesis de las proteínas virales durante el ciclo de replicación del VIH. mRNAs con pocos eventos de corte y empalme originan las proteínas estructurales virales, mientras que mRNAs con múltiples eventos de corte y empalme originan típicamente a las proteínas accesorias / reguladoras. La cantidad relativa de mRNAs con múltiples eventos de corte y empalme contra los no cortados o con un solo evento de corte y empalme es controlada por la proteína Rev (regulador de la expresión viral). Rev es una pequeña proteína que interacciona con un asa compleja de RNA presente en el gen *env* llamada RRE (elemento de respuesta a Rev), permitiendo la exportación de mRNA con pocos eventos de corte y empalme fuera del núcleo para la formación de las proteínas estructurales necesarias para el ensamblaje de nuevos viriones [17] (Figura 1.6). Rev posee una señal de localización nuclear (NLS) y una secuencia de exportación nuclear (NES) rica en leucina. El transporte a través del núcleo se da por la interacción de la NES de

Rev con CRM1/exportina-1, una proteína de la familia de las beta-importinas, y es mediado por RanGTP [22].

#### 1.4.5 Salida del virus de las células

Estudios con microscopía electrónica han mostrado que el ensamblaje y salida del VIH ocurre predominantemente en la membrana plasmática. Sin embargo, también es claro que en algunos tipos de células (principalmente monocitos / macrófagos), el ensamblaje y gemación de los viriones ocurre en compartimentos intracelulares definidos como endosomas tardíos o cuerpos multivesiculares (MVB) [46, 47] (Figura 1.6). El proceso de salida del VIH está dirigido fundamentalmente por las poliproteínas Pr55<sup>Gag</sup> y Pr160<sup>Gag-Pol</sup>, pero también está influenciado por Env e involucra el reclutamiento de dos copias del RNA genómico viral, así como de Vpr [22]. El precursor Pr55<sup>Gag</sup>, producido a partir del mRNA sin eventos de corte y empalme, es dirigido a membranas inmediatamente después de su síntesis a través del dominio MA. La afinidad de MA por membranas está dada fundamentalmente por su miristilación en un residuo de glicina del extremo amino terminal. Debido a la miristilación de MA, los precursores Pr55<sup>Gag</sup> se ensamblan preferentemente en microdominios membranales ricos en colesterol y glicolípidos, llamados balsas lipídicas, que parecen favorecer la liberación y la estabilidad de los viriones nuevos, así como la fusión de sus membranas a nuevas células blanco. La salida del virus ocurre a través de estas regiones especializadas de la membrana, produciéndose viriones con envolturas ricas en colesterol [48].

El procesamiento del precursor de la poliproteína Gag ocurre en una serie ordenada de cortes para generar a las proteínas estructurales MA, CA y NC. La liberación final de CA madura es fundamental para el ensamblaje de partículas infectivas. El último paso para la generación de CA madura es el corte del precursor p25, que contiene a CA y p2. Es precisamente en este paso de la maduración del VIH donde se ha encontrado actividad inhibitoria de compuestos derivados del ácido betulínico [49, 50]. Estos compuestos podrían servir como base para la generación de una nueva familia de compuestos con actividad antiviral: los inhibidores de la maduración [51].

De 2000 a 4000 copias de la poliproteína Gag son necesarias para el ensamblaje de la cápside viral. La proteína madura CA forma la cápside viral después de que el precursor Gag es procesado por la proteasa viral. CA contiene dos dominios estructurales y funcionales llamados dominio N-terminal (NTD) y dominio C-terminal (CTD), conectados por una asa flexible. Alteraciones en el NTD inhiben la maduración y liberación de viriones,

mientras que alteraciones en el CTD causan problemas en el ensamblaje de la cápside. CA se asocia en hexámeros en la cápside cónica con la inserción de algunos pentámeros, cuyo espaciamiento determina la forma cónica característica de la cápside del VIH. Cada hexámero de CA se compone de un anillo interno de NTDs y un anillo externo de CTDs que forman uniones diméricas para conectar hexámeros vecinos [52]. Estudios *in vitro* de inhibidores del ensamblaje de la cápside han mostrado que la alteración de las interfases de dimerización de CA evitan la formación de las redes de proteína que constituyen la cápside madura. Estos estudios demuestran que la cápside es un blanco factible para la inhibición de la replicación del virus [53, 54].

La función principal de la proteína NC es la encapsidación del genoma completo viral de RNA en los viriones. NC contiene dos motivos C-(X)<sub>2</sub>-C-(X)<sub>4</sub>-H-(X)<sub>4</sub>-C conocidos como dedos de zinc, que son fundamentales para su interacción con el RNA viral, junto con algunos residuos básicos que flanquean a los dedos de zinc [55]. La actividad de NC como chaperona de ácidos nucleicos favorece la transcripción inversa mediante la desestabilización de estructuras secundarias que interferirían con el procesamiento de la RT, mediante la protección del RNA viral hacia degradación por nucleasas y mediante la colocación del iniciador de tRNA y la promoción de reacciones de alineamiento durante las reacciones de transferencia de hebra. NC también promueve la integración del genoma viral, actividad para la cual requiere de ambos dedos de zinc y de los residuos básicos adyacentes. Adicionalmente, NC une de manera selectiva el RNA viral para su dimerización y empaquetamiento en partículas virales nuevas a través de sus dedos de Zn. La especificidad para encapsidar genomas de VIH está dada por la interacción de NC y una secuencia de 120 nucleótidos localizada entre el LTR 5' y el codón de inicio de Gag, conocida como señal de empaquetamiento o sitio psi que característicamente forma cuatro asas tallo [17].

La proteína p6 es fundamental para la salida de los viriones nuevos de la célula infectada. La salida de los viriones depende de manera crítica de las secuencias “dominio terminal” presentes en p6. Estas secuencias unen proteínas de enlace con los complejos ESCRT de la vía de ordenamiento vacuolar de proteínas. Los complejos ESCRT seleccionan las proteínas que serán incorporadas a vesículas en los MVBs para su posterior degradación lisosomal. Los MVBs se forman cuando fragmentos de endosomas tardíos geman hacia el lumen del endosoma. Una vez formados los MVBs, éstos se fusionan con lisosomas para la degradación de su contenido. Ahora es claro que el VIH ha usurpado esta vía de tránsito de vesículas para lograr su propio escape de las células infectadas. La importancia de Gag en la liberación de partículas virales se demostró al observar que la

deleción de p6 causaba una acumulaci3n de partculas virales que permanecían unidas a la membrana celular [56]. p6 contiene al menos dos sitios de uni3n para proteínas de enlace que reclutan proteínas de los complejos ESCRT [57]. Estos sitios de Gag, fundamentales para la salida del virus, se conocen como dominios tardíos. El motivo PTAP en la regi3n amino terminal de p6 se une a la proteína Tsg101, componente de ESCRT-I [58]. De manera similar, el motivo LYPLTSLRSL se une al factor AIP1, también conocido como Alix, que contiene un sitio de uni3n para la proteína CHMP4 de ESCRT-III [59].

#### **1.4.6 Funciones de las proteínas accesorias/reguladoras**

##### **1.4.6.1 Vpr**

A pesar de su pequeño tamaño, Vpr presenta varias funciones importantes en el ciclo de replicaci3n viral [60]. Vpr se incorpora en grandes cantidades en los viriones al interactuar con la regi3n carboxilo terminal de p6. Sin embargo, su alta afinidad por la proteína NC madura, podría determinar su ubicaci3n dentro de la cápside del virus. Se ha calculado que podrían existir hasta 275 moléculas de Vpr por viri3n. Esta propiedad ha permitido usar a Vpr como una herramienta para introducir proteínas o fármacos al interior de los viriones para estudiar varios aspectos del ciclo de replicaci3n viral. Vpr forma parte del complejo de transcripci3n inversa y se ha demostrado que esta proteína tiene la capacidad de modular la tasa de mutaci3n del VIH al influenciar la fidelidad de la transcripci3n inversa. Esta actividad correlaciona con la interacci3n de Vpr con la uracil DNA glicosilasa 2 (UNG2), una enzima que remueve residuos de uracilo del DNA que pueden incorporarse equivocadamente o generarse por desaminaci3n de citosina. La ausencia de Vpr causa tasas de mutaci3n hasta 4 veces más altas que en sistemas que contienen a la proteína viral [60]. Vpr tiene también un papel en el transporte del DNA viral al núcleo después de la transcripci3n inversa, al promover interacciones directas o indirectas con la maquinaria celular que regula el transporte de proteínas entre el núcleo y el citoplasma. Asimismo, Vpr forma parte del PIC muy probablemente debido a su capacidad de interactuar con nucleoporinas y es altamente probable que Vpr funcione como un ancla del PIC en la envoltura nuclear, mientras que IN y MA actúan cooperativamente con el ala central de DNA para dirigir el DNA viral al interior del núcleo. En el núcleo, Vpr puede funcionar como un transactivador débil de la expresi3n de genes de VIH por LTR y de genes celulares por promotores celulares. Vpr puede interactuar directamente con factores de transcripci3n celulares como TFIIB a través de un motivo WXXF [17, 60].

Otra función bien caracterizada de Vpr es su capacidad de inducir el arresto de las células en la fase G<sub>2</sub> del ciclo celular, durante las primeras horas de la infección, por inactivación del complejo cinasa dependiente de ciclina (CDK) p34/cdc2. La CDK p34/cdc2 se asocia a ciclina B1 en la fase G<sub>2</sub> para regular la transición de G<sub>2</sub> a M. Esta función de Vpr aún no ha sido totalmente entendida, pero se cree que el LTR es más activo durante la fase G<sub>2</sub>, implicando que un arresto en la fase G<sub>2</sub> conferiría un ambiente celular favorable para la transcripción eficiente de genes del VIH. Vpr causa además hernias nucleares transitorias, las cuales pueden contribuir al arresto del ciclo celular por la modificación de la estructura de la membrana nuclear y pérdida en la compartimentalización de proteínas reguladoras clave del ciclo celular [17, 60]. Más aún, Vpr se ha asociado a la producción de apoptosis tanto por vías dependientes, como por vías independientes de caspasas. Un hallazgo interesante es que pacientes no progresores tienen mutaciones más frecuentes en el residuo R77 de Vpr que los progresores normales. Este residuo parece ser crucial para que Vpr cause apoptosis por permeabilización de la membrana mitocondrial [60].

#### **1.4.6.2 Vpu**

Vpu es una proteína característica del VIH-1, encontrándose ausente en el VIH-2 y los SIVs conocidos (ver abajo). Esta proteína integral de membrana se traduce a partir de mRNAs bicistrónicos *vpu-env* en cantidades intracelulares similares a Gag. Sin embargo, no se ha detectado dentro de los viriones. Vpu es una proteína de membrana con dos dominios con diferentes actividades: el dominio transmembranal (TM), involucrado en la salida de nuevos viriones de las células infectadas y el dominio citoplásmico, implicado en la degradación de CD4 [61]. Vpu causa la degradación de CD4 al servir de enlace entre esta molécula y proteínas de la vía de degradación del proteosoma. Se cree que esta función es importante para liberar a la proteína Env de su interacción con CD4, permitiéndole continuar su transporte hacia la membrana. Vpu se une directamente a la cola citoplásmica de CD4 en el retículo endoplásmico y a la proteína h-βTrCP, marcador para proteólisis mediada por ubiquitinación [17].

Hace algunos meses se identificó una actividad en células humanas que inhibía la liberación de partículas retrovirales, la cual era antagonizada por Vpu. La proteína responsable de esta actividad ARV se ha denominado enlacina (tetherin, en inglés) y corresponde a la proteína de membrana CD317. La enlacina puede ser inducida por interferón-alfa y causa retención de viriones completamente formados en la superficie celular. Vpu se une a CD137, permitiendo la liberación de los viriones [62].

### 1.4.6.3 Vif

El factor de infectividad viral (Vif) deriva su nombre de su capacidad de promover la infectividad del virus. Ahora se sabe que la actividad de Vif consiste en contrarrestar la actividad de factores celulares de restricción para la infección retroviral de la familia de las proteínas APOBEC. Vif es incorporado en los viriones en cantidades que se han calculado de 7 a 20 moléculas por partícula viral. Su ausencia permite que las proteínas APOBEC restrinjan la replicación del VIH a través de su actividad de citidín desaminasas, provocando hipermutación en el virus. Algunos de los miembros de la familia de la superfamilia APOBEC de citidín desaminasas, principalmente APOBEC3G y APOBEC3F han sido identificados como factores de restricción celulares hacia retrovirus. Su actividad consiste en la desaminación de dC para producir dU en el DNA viral de cadena sencilla negativa que se origina durante la transcripción inversa. Esta actividad tiene como resultado la hipermutación del genoma viral al producirse sustituciones G→A en la hebra positiva del DNA proviral. La hipermutación introduce frecuentemente cambios deletéreos, así como codones de paro que dañan irreversiblemente al virus, impidiendo su replicación [63-65]. Los pacientes con virus hipermutados han demostrado tener viremias pre-tratamiento significativamente más bajas [66].

La proteína viral Vif se contrapone a la actividad de las proteínas APOBEC al promover la degradación dependiente de ubiquitinación de APOBEC en el proteosoma [67-69]. Vif causa la ubiquitinación de APOBEC vía una ligasa E3 cuyo reclutamiento depende de un complejo proteico de Culina 5 – Elonguina B – Elonguina C (Cul5-EloBC) [70, 71]. El complejo Cul5-EloBC se une a APOBEC a través de Vif, que funciona como una proteína de enlace. Se sabe que Vif se une a EloBC y a Cul5 a través de un motivo caja BC (aminoácidos 144 a 153) y un sitio H-(X)<sub>5</sub>-C-(X)<sub>17-18</sub>-C-(X)<sub>3-5</sub>-H (HCCH) de coordinación de zinc (aminoácidos 108 a 139) respectivamente [72, 73]. De la misma manera, recientemente se describieron las regiones Y<sup>40</sup>RHHY<sup>44</sup> y D<sup>14</sup>RMR<sup>17</sup> de Vif que interaccionan directamente con APOBEC3G y con APOBEC3F respectivamente [74]. Estos sitios son de gran interés para el desarrollo de fármacos ARVs que promuevan la actividad de APOBEC, impidiendo su degradación por Vif. En un estudio poblacional reciente, se demostró la existencia de polimorfismos en Vif que aparecían exclusivamente en secuencias hipermutadas y que no podían ser atribuidos a sustituciones G→A mediadas por APOBEC [66]. Esto sugiere que algunas mutaciones podrían intervenir con la función de Vif, produciéndose hipermutación en el virus al poderse expresar la actividad de APOBEC. Debido al impacto que Vif puede tener en la replicación del virus, se ha sugerido que esta proteína viral puede ser un buen candidato en el desarrollo de posibles vacunas contra el VIH [75]. También recientemente se

describieron los motivos de unión de APOBEC a Vif, identificándose el requerimiento de los residuos 128 a 130 de esta proteína celular para la interacción. Además, se reportó que los residuos 124 a 127 son importantes para el empaquetamiento de APOBEC3G en las partículas virales [76]. Esta información podría ser útil para el diseño de fármacos que inhiban la interacción Vif-APOBEC, sin intervenir con la actividad de APOBEC [24].

Interesantemente, APOBEC3G y APOBEC3F parecen haber tenido un papel evolutivo como presión de selección moldeando el genoma del VIH. Es posible que su actividad haya resultado en un sesgo en el contenido de nucleótidos y el uso de codones en el genoma viral [77]. Esta hipótesis surgió de la observación de que el genoma del VIH, a diferencia de otros retrovirus, es rico en A. Adicionalmente, de los seis codones que codifican para arginina, CGG, posible blanco de APOBEC3G, es muy poco usado para codificar este aminoácido en el VIH. Consistentemente, existe una tendencia marcada a usar AGA, el cual no es blanco de APOBEC3G [78].

#### **1.4.6.4 Nef**

Nef (factor negativo) es una proteína presente únicamente en lentivirus de primates que posee un papel fundamental en la patogénesis del VIH. A pesar de su nombre, ahora se sabe que Nef aumenta la replicación e infectividad virales. En general, Nef modifica el ambiente interno de la célula infectada para optimizar la replicación viral. Es notorio que macacos infectados con SIVs carentes del gen *nef* o individuos infectados por VIHs con proteínas Nef codificadas por genes que contienen deleciones importantes en su secuencia y que son seleccionadas *in vivo*, típicamente progresan lentamente hacia SIDA [79, 80]. La potenciación de la virulencia del VIH causada por Nef está asociada a su capacidad de funcionar como una proteína adaptadora, modulando funciones celulares como el tráfico de proteínas, transducción de cascadas de señalización y vías apoptóticas a través de interacciones proteína-proteína [81]. Nef es una de las primeras proteínas virales que se producen, además de ser una de las más abundantes, y típicamente lleva a cabo sus funciones en membranas, a donde es reclutada por miristilación. Además, Nef se encuentra también dentro del virión, asociado a la cápside.

Entre las funciones específicas atribuidas a la proteína Nef se encuentra la regulación negativa de varias proteínas celulares de superficie, con importantes consecuencias para el reconocimiento inmune de las células infectadas. Es bien sabido que Nef tiene la capacidad de regular negativamente la expresión de CD4 al enlazar el dominio citoplásmico de este receptor con pozos recubiertos de clatrina a través de la interacción

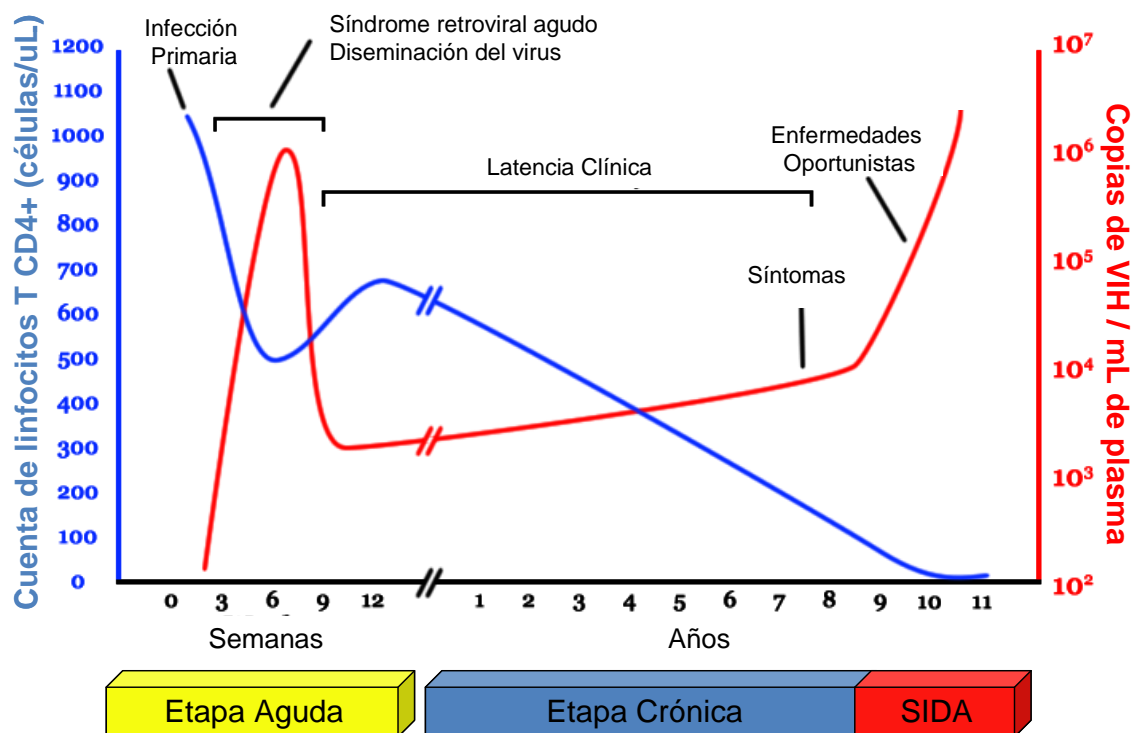
con la proteína adaptadora 2 (AP2). Esta interacción promueve la endocitosis de CD4. Una vez en el endosoma, Nef conecta a CD4 con coatómeros COP-I, dirigiendo a CD4 hacia degradación lisosomal [82, 83]. Esta actividad de Nef limita la superinfección y mitiga los efectos inhibitorios de CD4 en la liberación de nuevos viriones y en la incorporación de Env [17, 45, 84]. Nef también regula negativamente la expresión de moléculas del complejo mayor de histocompatibilidad (MHC) de clase I, pero a través de un mecanismo distinto al que utiliza para disminuir la expresión de CD4 [85]. Este mecanismo involucra la formación de complejos Nef-MHC-I y su redireccionamiento dependiente de AP1 de la vía trans-Golgi a lisosomas [86] o de manera alternativa, la internalización directamente de la membrana plasmática a través de una vía dependiente del factor de ADP-ribosilación 6 (ARF6) [87]. Esta función de Nef limita la actividad de los linfocitos T citotóxicos (CTLs) para reconocer y eliminar células infectadas, proporcionando al VIH un mecanismo de evasión de esta respuesta inmune. La presencia de mutaciones que evitan la regulación negativa de MHC-I por Nef aumenta significativamente la susceptibilidad de las células infectadas a ser eliminadas por CTLs *in vitro* [88-90], induce respuestas de CTLs más altas en macacos infectados por SIV [91], y permite el establecimiento de respuestas de CTLs más efectivas [92]. Además, aunque la disminución de expresión de MHC-I hace a las células infectadas susceptibles de ser destruidas por células NK, Nef puede evitar esta respuesta al no disminuir la expresión de HLA-C y HLA-E, moléculas MHC-I asociadas con protección a destrucción por células NK [93]. Esta especificidad radica en diferencias en aminoácidos entre las colas citoplásmicas de las moléculas sensibles y resistentes a la acción de Nef. Adicionalmente, Nef regula negativamente la expresión de CD28 [94], del complejo CD3-receptor de células T [95] y de moléculas MHC-II [96], funcionando típicamente como una proteína adaptadora entre estas proteínas y complejos celulares AP.

Varios estudios sugieren que Nef es capaz de potenciar la infectividad del VIH. Esta función de Nef parece manifestarse en etapas tempranas post-entrada del ciclo de replicación del VIH, posiblemente ayudando a la desencapsidación del virus o tal vez regulando el transporte de la cápside viral. Nef también podría intervenir en el transporte de proteínas virales o factores celulares involucrados en la entrada o salida del virus. Se ha reportado que Nef afecta el tráfico y procesamiento proteolítico de Env. Nef aumenta la localización de Env en membranas endosomales en células productivas e incrementa su incorporación en las partículas virales. También se sabe que Nef se une a la poliproteína Gag-Pol. Sin embargo, los mecanismos a través de los cuales Nef aumenta la infectividad viral necesitan todavía ser elucidados [17].



Nef promueve la infección de linfocitos T CD4+ al provocar su activación y hacerlos más susceptibles a ser infectados. Para lograr esto, Nef altera la cascada de señalización intracelular desencadenada por el receptor de células T (TCR). Específicamente, se sabe que Nef interacciona con proteínas de la cascada del TCR como Vav, cinasa 2 activada por p21 (PAK2), Rac, CDC42 y el complejo DOCK2/ELMO1 [81]. Interacciones con la vía de señalización del TCR también llevan a la regulación positiva del ligando de Fas en la superficie celular, lo cual podría proteger a la célula infectada promoviendo apoptosis en linfocitos T citotóxicos vecinos. Para proteger a la célula infectada de sufrir una muerte prematura por la acción de FasL, Nef podría también suprimir la actividad de ASK1, una cinasa responsable de transducir señales apoptóticas de FasL y del receptor de TNF. Adicionalmente, Nef puede inhibir la apoptosis mediada por p53 [81].

## 1.5 HISTORIA NATURAL DE LA INFECCIÓN POR VIH



**Figura 1.7. Historia natural de la infección por VIH-1.** Se muestra la variación típica de dos parámetros clínicos clásicos para el seguimiento de la infección: la cuenta de linfocitos T CD4+ en sangre y la carga viral de VIH en plasma. Se resaltan la etapa aguda, la etapa crónica y la etapa de SIDA. Modificado de <http://en.wikipedia.org/wiki/File:Hiv-timecourse.png>.

### 1.5.1 Etapas clínicas de la infección por VIH.

La infección típica por VIH se asocia a un decremento progresivo en la cuenta de linfocitos T CD4+ en sangre y a un aumento progresivo de la carga viral plasmática. Estos

dos marcadores clínicos se han utilizado tradicionalmente para el seguimiento de la infección por el virus. La infección por VIH puede dividirse básicamente en tres etapas: la etapa aguda, la etapa crónica asintomática y la fase de SIDA (Figura 1.7), aunque el curso de la enfermedad es variable de persona a persona. El tiempo de la infección aguda al desarrollo de SIDA, definido por cuentas de linfocitos T CD4+ menores a 200 células/uL o por la aparición de infecciones oportunistas definitorias de SIDA puede ser tan rápido como 6 meses [97], mientras que algunas personas llevan más de 25 años infectadas manteniendo niveles normales de linfocitos T CD4+ sin mostrar inmunodeficiencia y sin necesidad de tratamiento ARV [98].

La etapa aguda inicia con la transmisión del virus a través de fluidos corporales de una persona infectada a otra. Esta etapa se caracteriza por una replicación viral rápida inmediatamente después de la exposición del individuo al VIH, alcanzándose niveles de  $10^6$  a  $10^7$  copias de RNA viral por mL de sangre en el pico de viremia y estabilizándose éstos en un punto de equilibrio promedio de alrededor de 30,000 copias/mL dentro de los primeros 6 a 12 meses de la infección (Figura 1.7). Sin embargo, este punto de equilibrio es variable y su valor se relaciona con el riesgo de progresión de la enfermedad (valores altos correlacionan con progresión rápida, mientras que valores bajos correlacionan con progresión lenta) [99]. Durante la infección aguda ocurre una infección del 30 al 60% de los linfocitos T CD4+ del tejido linfoide asociado al intestino (GALT), resultando en una pérdida masiva e irrecuperable de los linfocitos T CD4+ de memoria efectora locales en un periodo de tan solo 4 días [100, 101]. Durante esta etapa temprana de la enfermedad por VIH más de la mitad de los linfocitos T CD4+ de memoria se pierden, dejando al cuerpo con un repertorio severamente restringido de células T para enfrentarse a la fase crónica de la infección [98]. Importantemente, el daño al componente inmunológico de la mucosa intestinal por el VIH, aunado al daño en el microambiente epitelial del intestino con sus funciones antimicrobianas características, resultan en la translocación de productos bacterianos del lumen intestinal, los cuales correlacionan con el desarrollo de activación inmune generalizada, característica de la fase crónica de la enfermedad [102].

La etapa aguda de la enfermedad por VIH va acompañada en un 50 al 70% de los casos por síntomas inespecíficos que inician dos semanas después de la exposición y que incluyen frecuentemente fiebre, faringitis, dolor de cabeza, artralgias, mialgias, malestar, pérdida de peso, exantemas maculopapulares no pruríticos y linfadenopatía generalizada [98]. Estos síntomas definitorios del llamado síndrome retroviral agudo se resuelven en 3 a 4 semanas. Una vez resuelta la fase primaria de la infección se establece un estado virológico cuasi-estable que define al periodo crónico de la enfermedad por VIH. Aunque la mayoría de

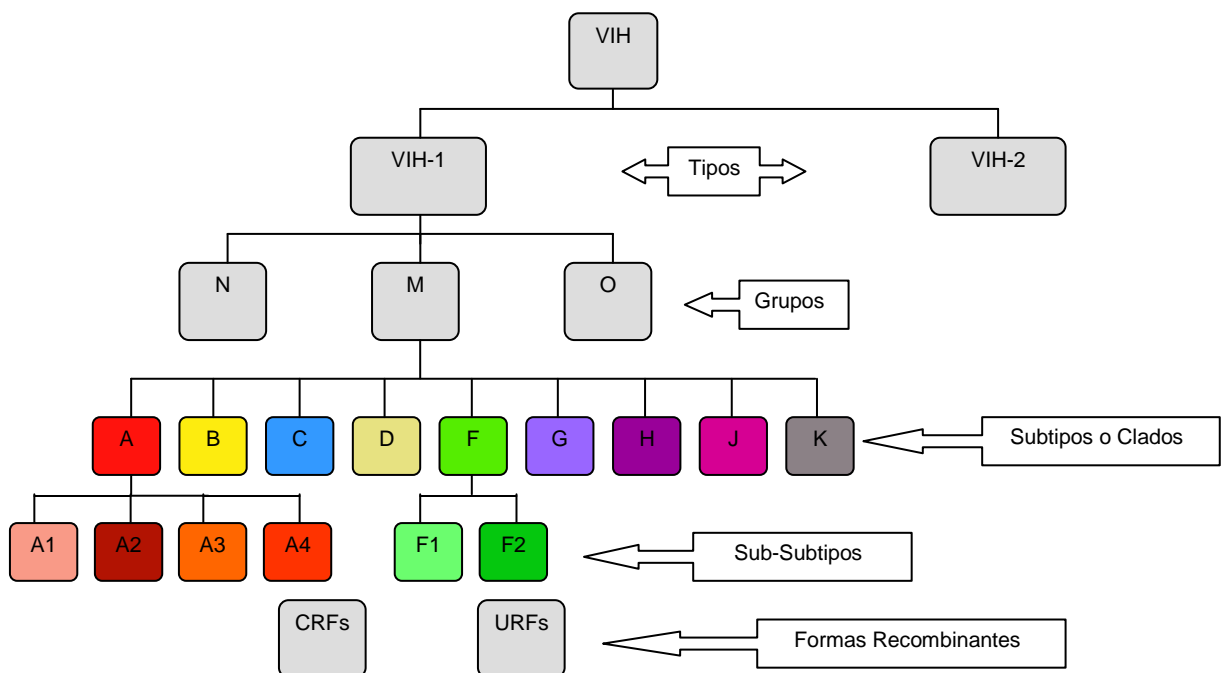
los pacientes permanecen asintomáticos durante este periodo, existe una replicación viral constante que causa una depleción lenta pero continua de linfocitos T CD4+ (Figura 1.7). Las infecciones oportunistas y cánceres son raros en personas infectadas con cuentas de linfocitos T CD4+ superiores a 500 células/uL, pero su frecuencia aumenta a medida que este valor disminuye. La candidiasis oral, infecciones neumococales, tuberculosis, reactivación de herpes simplex y varicella zoster son comunes. Con cuentas de células T CD4+ menores a 200 células/uL se incrementa altamente el riesgo de infecciones oportunistas que ponen en riesgo la vida del paciente definiendo la etapa avanzada de la enfermedad o SIDA. Infecciones oportunistas comunes incluyen neumonía por *Pneumocystis jiroveci*, esofagitis por *Candida*, histoplasmosis diseminada, infecciones fúngicas sistémicas, encefalitis por *Toxoplasma* y meningitis criptococal. La presencia de infección por el complejo *Mycobacterium avium* (MAC), reactivación de infección por citomegalovirus (CMV), criptosporidiosis, microsporidiosis y leucoencefalopatía multifocal por virus JC son indicadores de inmunodeficiencia profunda y se presentan con cuentas de linfocitos T CD4+ menores a 50 células/uL. Los cánceres asociados al SIDA se asocian comúnmente a etiologías virales incluyendo al sarcoma de Kaposi, causado por el herpesvirus humano 8; linfomas asociados con el virus Epstein-Barr y carcinomas anales y cervicales asociados al virus del papiloma humano [98]. El tiempo promedio desde la seroconversión hasta el desarrollo de SIDA en ausencia de tratamiento es de entre 9 y 10 años (aunque este tiempo es ampliamente variable de persona a persona debido a múltiples factores inmunológicos, genéticos y virales), mientras que la expectativa de vida a partir del desarrollo de SIDA es de un poco más de 9 meses [103]. Sin embargo, la terapia ARV ha cambiado radicalmente la historia natural de la infección por VIH.

### **1.5.2 El tratamiento antirretroviral y sus implicaciones para el control viral.**

El tratamiento ARV empleado actualmente para tratar la infección por VIH consiste en la combinación de 3 o 4 fármacos de al menos dos familias y se conoce como terapia ARV altamente activa (HAART, por sus siglas en inglés). Los regímenes típicos de HAART consisten de dos inhibidores de la transcriptasa inversa análogos de nucleósidos (NRTIs) y un inhibidor de la transcriptasa inversa no análogo de nucleósido (NNRTI) o un inhibidor de la proteasa (PI) potenciado. Las clases relativamente más nuevas de fármacos ARVs que incluyen a los inhibidores de fusión, antagonistas de CCR5 e inhibidores de integrasa, proporcionan opciones para individuos infectados por virus resistentes a los fármacos de las familias más comunes. Sin embargo, muchas veces estos fármacos nuevos no están ampliamente disponibles y no son accesibles en países con recursos limitados. El HAART exitoso ha transformado la historia natural típica de la infección por VIH al reducir la viremia

a niveles indetectables y permitiendo una reconstitución parcial gradual del número de linfocitos T CD4+. Siendo el único modo de control viral disponible en la actualidad, la terapia ARV ha cambiando completamente el pronóstico de la enfermedad por VIH, trayendo gran esperanza para las personas que viven con el virus. El espectro de complicaciones que emergen en pacientes tratados exitosamente, sin embargo, se ha modificado drásticamente desde el advenimiento del HAART. Las enfermedades definitorias de SIDA típicas han sido sustituidas por nuevas condiciones de co-morbilidad que pueden amenazar incluso a los pacientes que mantienen supresión virológica. En particular, se debe poner una atención especial en el manejo rutinario adecuado del riesgo cardiovascular, el diagnóstico temprano de cánceres no asociados a SIDA (incluyendo neoplasmas relacionados con el papilomavirus humano), desórdenes neurocognitivos, coinfecciones por los virus de la hepatitis B y C, enfermedad de hígado graso, leucoencefalopatías multifocales progresivas y tuberculosis, además de posibles lipoatrofias y toxicidades en hueso y riñón a largo plazo [104]. La identificación y manejo oportuno del síndrome inflamatorio por reconstitución inmune es de gran interés, especialmente en pacientes que inician HAART con una severa depleción de linfocitos T CD4+, como es muchas veces el caso en los pacientes latinoamericanos [7, 104].

## 1.6 EPIDEMIOLOGÍA MOLECULAR DEL VIH



**Figura 1.8. Subtipificación del VIH-1.** Se muestra la clasificación del virus en tipos, grupos, subtipos, sub-subtipos y formas recombinantes. CRFs – Formas recombinantes circulantes; URFs – Formas recombinantes únicas.

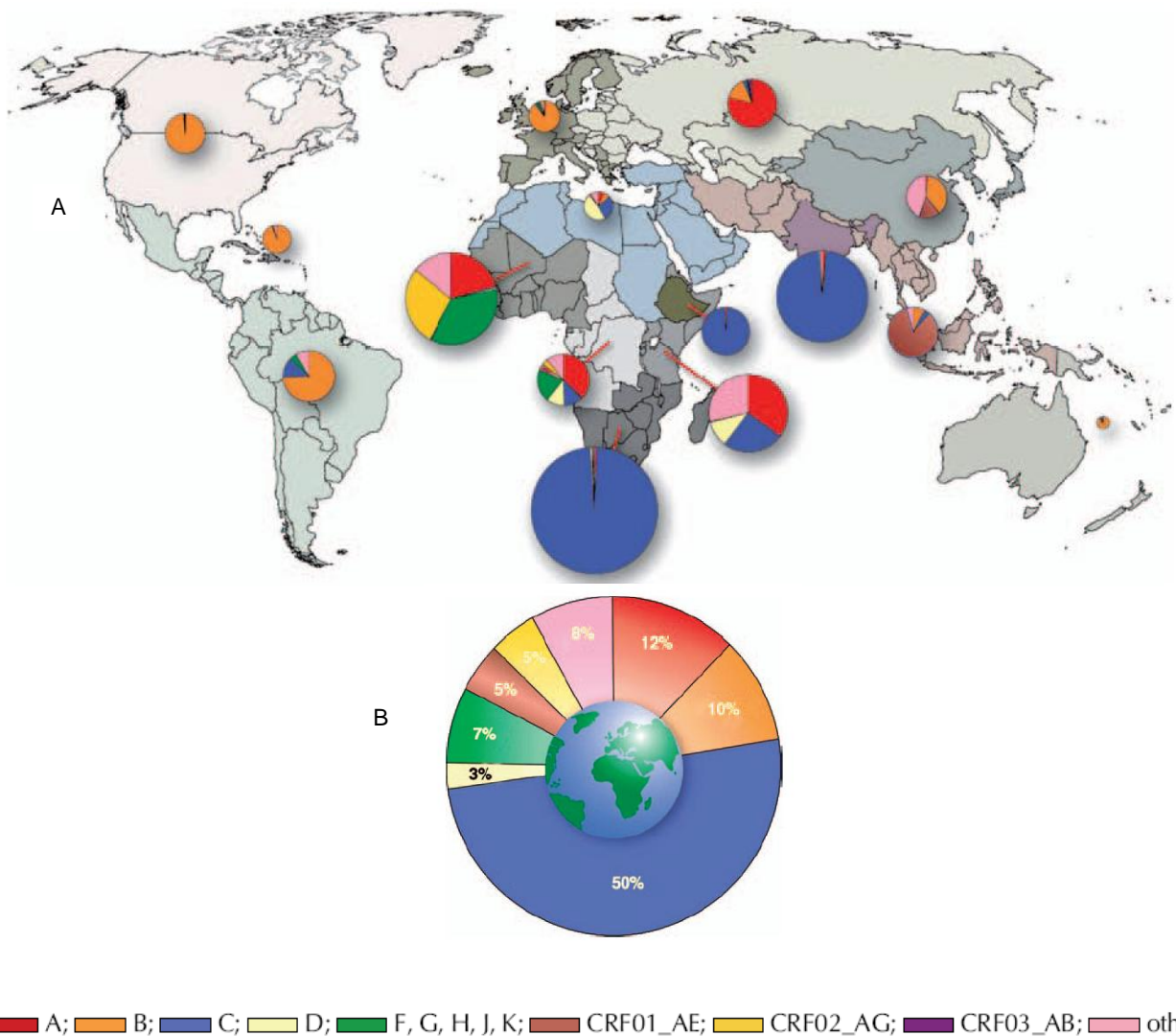
### 1.6.1 Diversidad Genética

El VIH se caracteriza por una notable diversidad genética que ha derivado en su clasificación en tipos, grupos, subtipos o clados, sub-subtipos, formas recombinantes circulantes (CRFs) y formas recombinantes únicas (URFs) (Figura 1.8) [105, 106]. El VIH-1, causante de la mayoría de los casos mundiales de infección, puede clasificarse en 3 grupos filogenéticos: M (major), O (outlier) y N (non-M, non-O) (Figura 1.8), cada uno correspondiente a un salto interespecie independiente de los virus de inmunodeficiencia del simio (SIV) de chimpancés (para los grupos M y N) y posiblemente gorilas (para el grupo O) en África centro-occidental (ver abajo). El grupo M es el causante del 90% de las infecciones a nivel mundial y es el que más se ha diversificado, pudiéndose distinguir diferentes subtipos (A, B, C, D, F, G, H, J y K) y sub-subtipos (A1-A4, F1-F2) dentro de él (Figura 1.8) [107]. África central es la única región en donde se han identificado todos los grupos del VIH-1 y todos los subtipos del grupo M (Figura 1.9) [1, 106].

Las formas recombinantes de VIH han surgido por los frecuentes eventos de recombinación de virus de diferentes clados, que han generado virus con regiones genéticas derivadas de dos o más cepas virales. Las llamadas CRFs (CRF01-CRF37) han adquirido la capacidad de propagarse en diversas poblaciones (<http://www.hiv.lanl.gov/>), las URFs han permanecido como casos aislados. Hasta la fecha se han identificado formas recombinantes en prácticamente todo el mundo donde circulan dos o más subtipos [105]. La distribución geográfica de los varios subtipos M del VIH-1 es muy heterogénea en el mundo, con diferentes subtipos presentándose a frecuencias variables en distintos continentes o incluso de país a país (Figura 1.9) [106, 108]. En Norteamérica predomina el subtipo B, aunque el subtipo C es el más abundante a nivel mundial (Figura 1.9). Recientemente se ha reportado un incremento constante en la prevalencia de subtipos no-B (2.93% entre 2004 y 2008), especialmente A y C, en Estados Unidos, con una prevalencia relativamente estable de CRFs [109]. La información para México, sin embargo, es sumamente escasa y poco representativa existiendo una gran necesidad de evaluar la Epidemiología Molecular del VIH a nivel nacional [110, 111].

La diversidad del VIH tiene implicaciones importantes para múltiples aspectos de la pandemia como son el diagnóstico, pruebas de seguimiento, diseño de vacunas, susceptibilidad a fármacos ARVs, capacidad de transmisión y virulencia de los virus

circulantes y progresión de la enfermedad [112]. De ahí que sea imperativo monitorear las distribuciones globales y regionales de subtipos y formas recombinantes del virus.

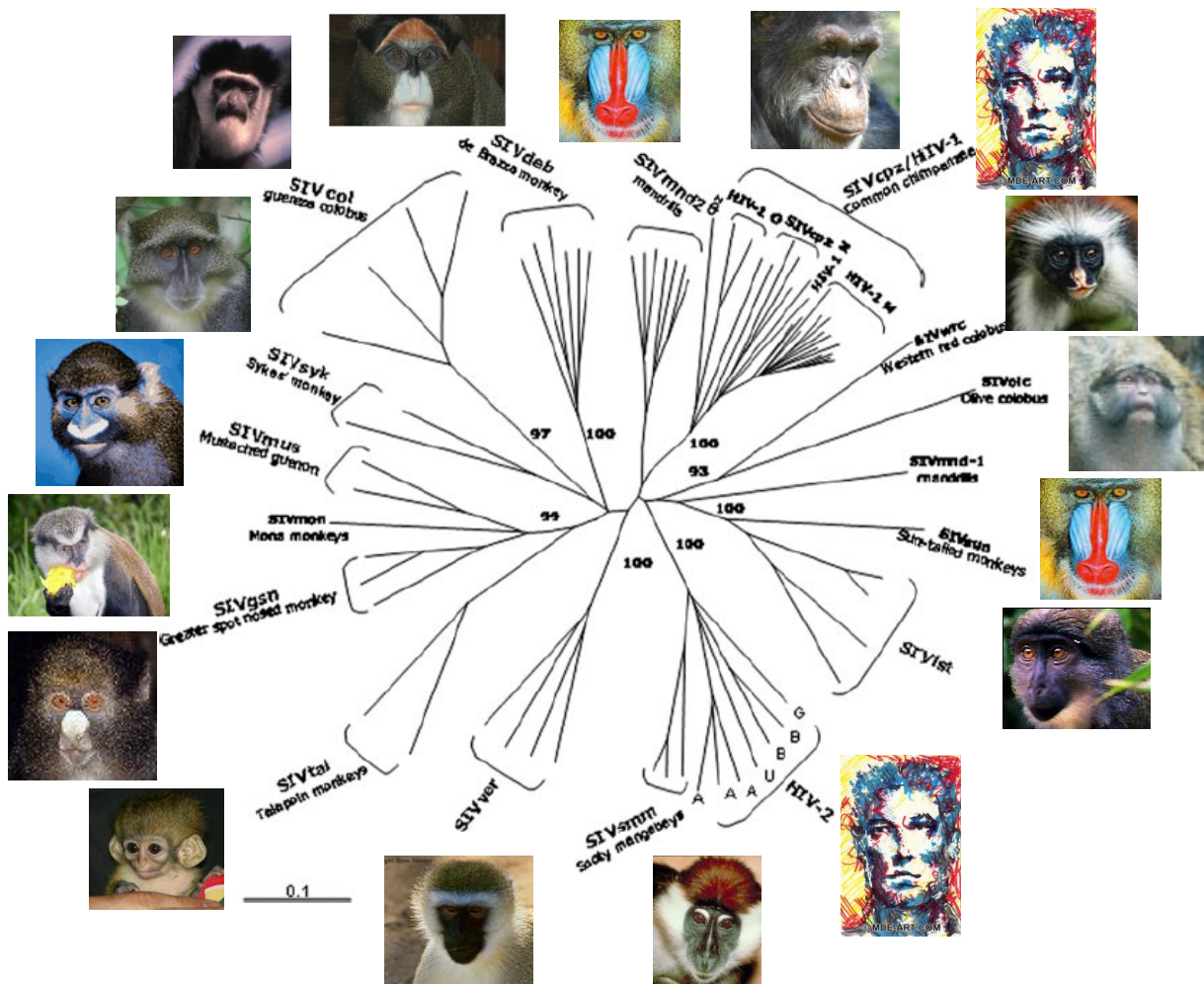


**Figura 1.9. Diversidad genética del VIH en el mundo.** A) Se muestran las distribuciones de subtipos de VIH-1 y formas recombinantes por región, de acuerdo a datos de 23,874 muestras de 70 países reunidos por la Organización Mundial de la Salud (WHO/UNAIDS) en 2004. Las regiones consideradas están sombreadas en diferentes colores. Se muestran gráficas de pastel con las distribuciones características de subtipos de VIH-1 por región. El tamaño de las gráficas de pastel es proporcional al número relativo de personas viviendo con VIH en cada región. B) Proporción de infecciones causadas por los diferentes subtipos de VIH-1 con respecto al número total de infecciones a nivel global. Tomado y adaptado de [106].

### 1.6.2 Origen del VIH

Está bien establecido que el VIH-1 y el VIH-2 surgieron como resultado de saltos inter-especie de virus de primates a humanos [113]. Los antecesores más cercanos de los VIHs se han determinado con base en estudios filogenéticos detallados de los diversos SIVs de primates (Figura 1.10). El mono mangabey ahumado (sooty mangabey, *Cercocebus atys*) se ha identificado como el reservorio del VIH-2. Se estima que al menos 8 saltos inter-

especie dieron origen al VIH-2 a partir del SIV del mangabey ahumado (SIVsm) en Africa Occidental [113] (Figura 1.10). Similarmente, el reservorio del VIH-1 ha sido claramente identificado como una de las subespecies de chimpancés que hospedan SIVs: *Pan troglodytes troglodytes*, característico de África Centro-Occidental. Todos los grupos conocidos del VIH-1, incluyendo el M, N y O están fuertemente relacionados con cepas de SIV de este linaje de chimpancés (SIVcpz), más no con el SIVcpz de la subespecie de chimpancé *Pan troglodytes schweinfurthii*, característica de África Oriental [114]. Más aún, el hábitat natural de *P. t. troglodytes* coincide de manera única con áreas de endemicidad de VIH-1 de grupo M, N y O, sugiriendo fuertemente que este simio es el reservorio principal del VIH-1 con al menos tres introducciones independientes del SIVcpz en la población humana [114, 115]. Consiguientemente, estos virus simianos se adaptaron gradualmente a su nuevo hospedero humano culminando en el surgimiento del VIH-1, identificándose la proteína de matriz (p17) como un modulador de la aptitud replicativa viral después de la transmisión a su nuevo hospedero [116].



**Figura 1.10. Relaciones filogenéticas entre VIH-1, VIH-2 y SIVs de diferentes primates.** Se observa un agrupamiento característico entre los VIH-1 de los grupos M, N y O y los SIVcpz característicos de los chimpancés, mientras que los VIH-2 se agrupan claramente con los SIVsm de los monos mangabey ahumados.



Interesantemente, se ha encontrado una fuerte relación filogenética entre los SIVs de gorilas (SIVgor) de África Centro-Occidental y el VIH-1 del grupo O, sugiriendo que este grupo de VIH-1 posiblemente se originó a partir del SIVgor. Sin embargo, estudios recientes sugieren que los chimpancés *P. t. troglodytes* son reservorio del SIVgor y que estos simios son responsables de la epidemia de VIH-1 de grupo O [117]. De cualquier manera, la posibilidad de que los gorilas hayan servido como hospederos intermedios aún no puede ser descartada. Aunque el origen de los VIHs se ha establecido cada vez más claramente, aún quedan preguntas sobre patogenicidad y extensión epidémica de algunas variantes de VIH/SIV que requieren ser exploradas. Más aún, debido a que los humanos estamos expuestos a una gran cantidad de lentivirus de primates a través de la caza y el manejo de carne de estos animales, la posibilidad de que ocurran nuevos saltos zoonóticos de lentivirus desde otros primates debe ser considerada [113].

### **1.6.3 Resistencia a fármacos antirretrovirales y variabilidad viral.**

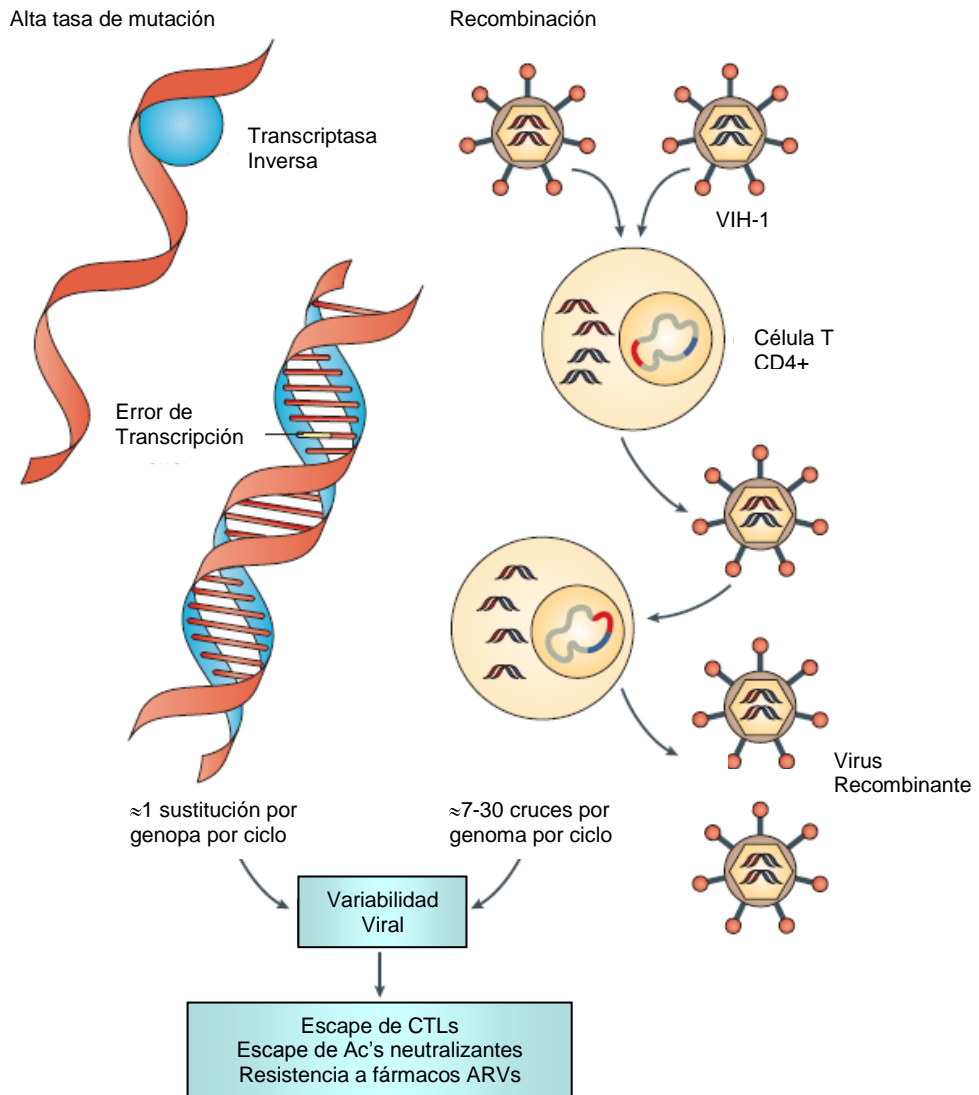
La resistencia hacia fármacos ARV es un aspecto clave en el manejo clínico de la infección por VIH. Es fundamental determinar si existen diferencias en susceptibilidad hacia los fármacos disponibles en los virus circulantes en diferentes regiones así como diferencias en su capacidad de desarrollar resistencia. Hasta el momento, la evidencia disponible parece indicar que los diferentes subtipos del grupo M presentan susceptibilidades semejantes a los fármacos ARV usados actualmente [106, 112, 118-120]. Sin embargo, existen diferencias sutiles entre subtipos en barreras genéticas para desarrollar mutaciones de resistencia mayores y en la presencia de mutaciones menores, principalmente en la proteasa viral, que pueden aparecer como polimorfismos naturales acelerando el desarrollo de resistencia [119, 121]. Asimismo, la presencia de mutaciones de resistencia en virus de individuos no tratados previamente con fármacos ARV y su impacto en la efectividad del tratamiento y la progresión de la enfermedad permanece como una preocupación importante tanto a nivel individual como poblacional. Por esta razón, el conocimiento de la prevalencia de estas mutaciones de resistencia primaria y sus tendencias de cambio en diferentes poblaciones es fundamental, siendo un punto importante en la investigación epidemiológica actual [122-128].

## **1.7 EVOLUCIÓN DEL VIH**

### **1.7.1 El VIH es un virus con gran capacidad de variación.**



El VIH ha desarrollado una capacidad extraordinaria de variación que representa un reto enorme para el desarrollo de una posible vacuna y que tiene implicaciones en la evolución del virus y su capacidad de adaptación a diferentes fuerzas selectivas que actúan sobre él [129]. Esta extraordinaria capacidad de variación del VIH se debe principalmente a fenómenos intrínsecos del ciclo de replicación viral que resultan en una alta tasa de mutación y a una alta frecuencia de eventos de recombinación (Figura 1.11).



**Figura 1.11. Fuentes e variabilidad del VIH.** La transcriptasa inversa es responsable de gran parte de la variabilidad del VIH al carecer de un mecanismo de corrección de errores durante la transcripción del genoma viral, introduciendo al menos un error por genoma por ciclo de replicación. Este mecanismo de variación está aunado a la posibilidad de recombinación entre las dos hebras de RNA presentes en el virión para formar virus mosaico con una frecuencia de 7 a 30 eventos de recombinación por genoma por ciclo de replicación. La variabilidad viral generada de esta manera resulta en la adaptación del virus a presiones selectivas naturales y artificiales como las respuestas inmunes celulares, humorales y la terapia ARV. CTLs – linfocitos T citotóxicos, Ac's – anticuerpos, ARV – antirretroviral. Adaptado de [130].

La transcriptasa inversa, enzima propensa a cometer errores durante la transcripción del genoma viral, introduce en promedio una mutación y origina de 7 a 30 eventos de

recombinación por cada ciclo de replicación [130]. Considerando además el hecho de que existe una producción abundante y constante de nuevos viriones durante la infección (cerca de 1,000 millones por día), las posibilidades de variación del VIH resultan en verdad notables. Como resultado de estos mecanismos de variación, dentro de un individuo infectado ocurre una expansión progresiva en la diversidad viral a lo largo del curso de la infección, con múltiples variantes co-existiendo como un enjambre de “cuasiespecies” en cualquier momento dado [131]. Esta amplia variabilidad viral permite la adaptación del VIH a presiones selectivas impuestas por el sistema inmune mediante la aparición e mutaciones de escape (ver abajo) y a la terapia antirretroviral mediante la aparición de mutaciones de resistencia.

### **1.7.2 Los genes HLA de clase I definen y restringen las respuestas de linfocitos T citotóxicos.**

Los genes del antígeno leucocitario humano (HLA) de clase I, se agrupan en la región más distal del complejo principal de histocompatibilidad (MHC) en el brazo corto del cromosoma 6, abarcando unas 2 Mb [132]. Algunos genes HLA de clase I codifican para proteínas que se expresan en la superficie de la mayoría de las células somáticas de los vertebrados con la función de presentar péptidos provenientes del procesamiento intracelular de proteínas de patógenos intracitoplásmicos a células T CD8+. Algunos genes HLA de clase I parecen haber perdido esta función y se denominan genes atípicos, mientras que otros como el HLA-A, B y C sí llevan a cabo la función de presentación de péptidos y son llamados clásicos [132, 133].

Las moléculas de clase I constan de dos cadenas polipeptídicas unidas de forma no covalente: una cadena alfa (cadena pesada) codificada por genes HLA de 6 o 7 exones y una subunidad no perteneciente al MHC denominada  $\beta_2$ microglobulina, codificada en el cromosoma 15 (Figura 1.12). El extremo amino terminal de la cadena alfa está expuesto al medio extracelular y se caracteriza por presentar una plataforma de 8 hebras beta antiparalelas que soporta dos hélices alfa paralelas. Esta estructura forma una hendidura cerrada que permite la unión y presentación de péptidos de 8 a 11 aminoácidos [133] (Figura 1.12). La unión de péptidos a las moléculas HLA es una interacción no covalente mediada por residuos tanto en los péptidos como en las hendiduras de las moléculas HLA. Los residuos del péptido cuyas cadenas laterales interaccionan directamente con residuos de la hendidura de la molécula HLA se denominan residuos de anclaje. Cada péptido que se une a la molécula HLA suele contener uno o dos residuos de anclaje que pueden localizarse en el centro o en los extremos del péptido. Sin embargo, las interacciones específicas de los

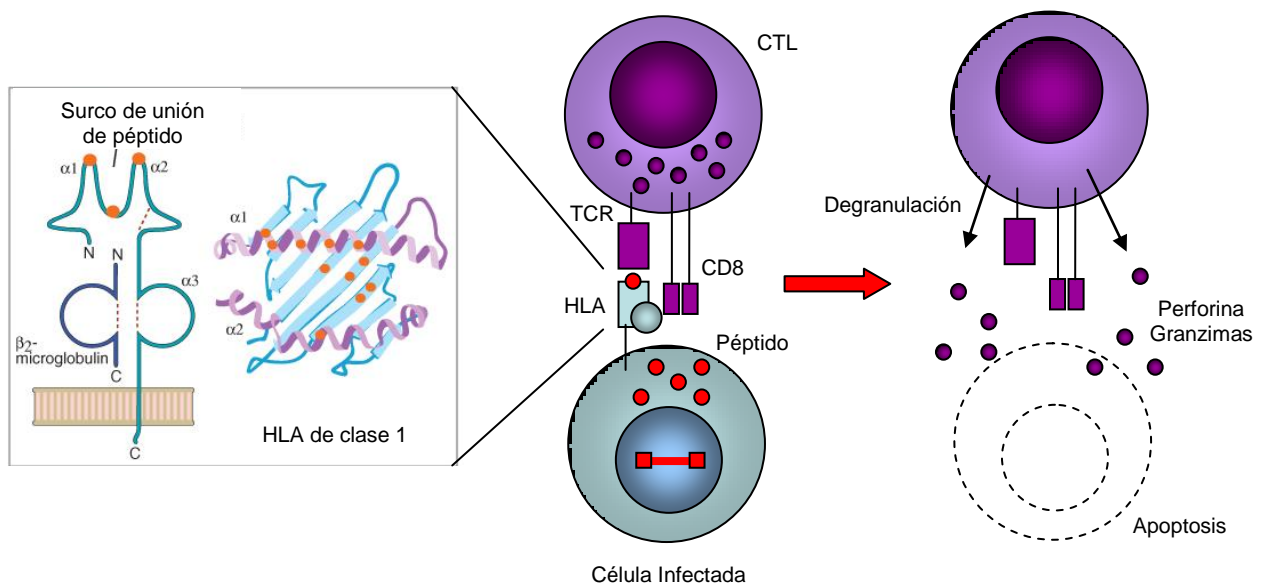
péptidos con las hélices alfa laterales de la hendidura del HLA contribuyen también a la fijación formando puentes de hidrógeno o interacciones polares [132, 133]. Los péptidos que se unen a clase I contienen normalmente aminoácidos básicos o hidrofóbicos en sus extremos carboxílicos que contribuyen también a la interacción [134]. Los péptidos asociados al MHC clase I se producen por la degradación proteolítica de proteínas citosólicas, a través de un complejo proteico codificado por el MHC, conocido como proteosoma. Los péptidos generados por el proteosoma en el citosol se transportan mediante un transportador especializado (TAP) hacia el retículo endoplásmico (RE), donde existen moléculas del MHC de clase I recién sintetizadas que unen a estos péptidos para después ser transportadas a la superficie celular a través de vesículas del complejo de Golgi. Tanto aminoácidos del péptido antigénico como aminoácidos de la molécula HLA contribuyen al reconocimiento antigénico por los linfocitos T a través del receptor de células T (TCR) [132]. Los linfocitos T CD8+ que reconocen al complejo péptido-HLA son activados con la ayuda de señales adicionales como citocinas, convirtiéndose en linfocitos T citotóxicos (CTLs). Los CTLs que reconocen a la célula infectada mediante la unión de su TCR al complejo péptido-HLA liberan el contenido de gránulos citotóxicos resultando en la destrucción de la célula blanco. Estos gránulos citotóxicos contienen perforina, una proteína que forma poros en la membrana celular de la célula blanco y granzimas que activan caspasas produciendo apoptosis en la célula blanco [132, 133] (Figura 1.12). Los CTLs también pueden destruir la célula blanco mediante la expresión de ligando de Fas, provocando apoptosis en la célula infectada.

Los genes HLA se heredan en bloque en forma de haplotipos, por lo que cada individuo tiene una combinación de genes que se segregaron del padre y otra combinación de la madre. La recombinación puede existir a muy bajos niveles (<1%). Los genes HLA se expresan de forma codominante en cada individuo, es decir, cada persona expresa todos los alelos HLA heredados de ambos padres. Notoriamente, determinados alelos HLA presentes en diferentes loci se heredan conjuntamente con más frecuencia de lo que se espera por una asignación al azar. Este es fenómeno se denomina desequilibrio de ligamiento [133].

Los genes MHC son los genes más polimórficos en el genoma humano [133]. La variabilidad de los genes HLA de clase I radica en los exones 2 y 3 que conforman los dominios de unión al péptido en el extremo amino terminal de la cadena alfa (Figura 1.12). Diferentes alelos HLA favorecen la unión de péptidos con diferentes características. Hacia mediados del 2009 se conocen cerca de 1050 proteínas HLA-B, 650 proteínas HLA-A y 360 proteínas HLA-C (IMGT/HLA Database, <http://www.ebi.ac.uk/imgt/hla>), existiendo una gran gama de posibles péptidos que pueden ser presentados por distintos alelos. Adicionalmente,

debido a la poligenia del MHC, cada persona expresará un máximo de seis moléculas clásicas HLA de clase I diferentes (dos HLA-A, dos HLA-B y dos HLA-C), pudiendo presentar un repertorio variado de péptidos a células T CD8+ [133, 135].

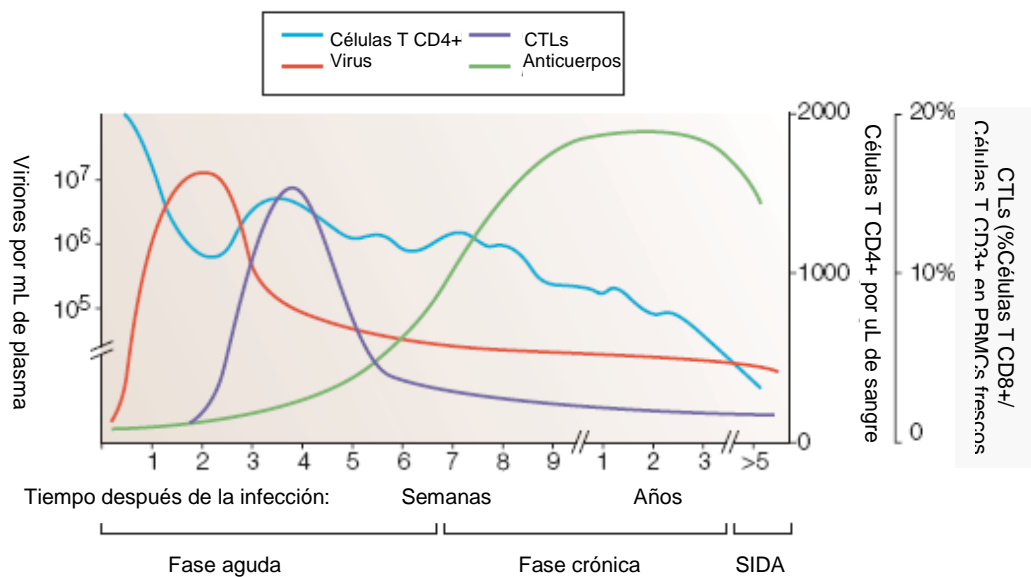
Los alelos HLA se denominan con la letra del locus correspondiente seguida de un número (que se asigna de manera consecutiva según el orden de descubrimiento del alelo), separados mediante un asterisco cuando el alelo está definido a nivel de secuencia de DNA (y no de manera serológica). Los primeros dos dígitos describen el tipo o grupo de alelo. El tercero y cuarto dígitos designan los subtipos o alelos específicos. Los alelos que difieren en los primeros cuatro dígitos son diferentes en una o más posiciones en la secuencia proteica. Los alelos que difieren por una sustitución sinónima dentro de la secuencia codificante son distinguidas con el uso del quinto y sexto dígito, mientras que los alelos que sólo difieren por polimorfismos en las secuencias intrónicas son distinguidos con el uso del séptimo y octavo dígito (Anthony Nolan Research Institute, [www.anthonynolan.org.mx](http://www.anthonynolan.org.mx)).



**Figura 1.12. Las respuestas de CTLs están definidas y restringidas por moléculas HLA de clase I.** Células infectadas por patógenos intracitoplásmicos procesan péptidos antigénicos que son expresados en su superficie en el contexto de moléculas HLA de clase I. Los CTLs reconocen células infectadas por la interacción altamente específica de su TCR con los complejos péptido-HLA expresados en la superficie de la célula blanco. Este reconocimiento, junto con otras señales provoca la degranulación de los CTLs con la consiguiente liberación de las moléculas citotóxicas perforina y granzima, llevando a la muerte de la célula blanco. El recuadro muestra un detalle de la molécula HLA de clase I, compuesta por una cadena alfa pesada y un polipéptido llamado beta2-microglobulina no codificado en el MHC. El sitio de unión del péptido se encuentra en el extremo amino terminal de la cadena alfa y se caracteriza por una hoja beta con 8 hebras antiparalelas que soporta dos hélices alfa paralelas. Los residuos polimórficos que interactúan directamente con el péptido se muestran como puntos rojos. CTLs – Linfocitos T citotóxicos, HLA – Antígeno leucocitario humano, TCR – Receptor de células T, MHC – Complejo principal de histocompatibilidad. Modificado y adaptado de [133].

### 1.7.3 La respuesta inmune como presión selectiva actuando sobre el VIH

El sistema inmune del hospedero actúa como una fuerza de selección importante en el curso natural de la infección por VIH, dirigiendo la evolución del virus por un proceso continuo y dinámico conocido como “escape inmune” [131]. Los anticuerpos contra VIH reconocen epítodos tridimensionales principalmente en las proteínas de la envoltura del virus, neutralizando viriones libres y marcándolos para su destrucción por células efectoras o por el sistema del complemento. El escape de la respuesta inmune de anticuerpos se da por mutaciones puntuales, inserciones y deleciones en los genes que codifican para las proteínas de envoltura. Este tipo de escape inmune es responsable en gran parte de la diversidad viral en el gen *env* tanto a nivel intra-individuo como a nivel global [136]. Sin embargo, aunque la presión selectiva ejercida por los anticuerpos sobre las proteínas de envoltura del VIH es evidente, el papel de la respuesta de anticuerpos en el control de la replicación viral y su impacto en el avance de la enfermedad por VIH podría ser poco significativo en comparación con el papel de la respuesta de linfocitos T CD8+ citotóxicos (CTLs) [137].



**Figura 1.12. Respuestas inmunes celulares y humorales en la infección por VIH.** La aparición de los linfocitos T citotóxicos (CTLs) coincide con la disminución inicial de la viremia. La aparición de la respuesta de anticuerpos es tardía y no correlaciona con cambios en el control viral. Adaptado de [138].

Los CTLs eliminan células infectadas por el VIH mediante el reconocimiento de péptidos lineales cortos derivados de las proteínas virales (epítodos), procesados intracelularmente y presentados en la superficie de las células infectadas a través de moléculas HLA de clase I (ver arriba). La presentación de epítodos virales por moléculas HLA de clase I permite el reconocimiento de las células infectadas por CTLs, los cuales las

destruyen por liberación de citotoxinas como perforina que forma poros en las membranas celulares y serín proteasas llamadas granzimas que inducen apoptosis. Consecuentemente, las respuestas de CTLs están definidas y restringidas por los genes HLA de clase I. Varias líneas de investigación sugieren un papel importante de la respuesta de CTLs en el control del VIH (ver más adelante). Este hecho aunado a la observación de que la aparición de la respuesta de CTLs coincide con la selección de mutantes de escape inmune, resalta el potencial de esta respuesta inmune de funcionar como una fuerte presión selectiva *in vivo* sobre el virus.

#### **1.7.4 La respuesta de CTLs restringida por los genes HLA de clase I es una fuerza selectiva importante dirigiendo la evolución intra-individuo del VIH.**

Dentro de las líneas de evidencia que sugieren la importancia de la respuesta de CTLs en la replicación viral destaca la observación de que la aparición de CTLs específicos contra el VIH correlaciona temporalmente con la disminución de la viremia en etapas tempranas de la infección [139] (Figura 1.12). Asimismo, estudios en macacos infectados por SIV en los cuales se administraron anticuerpos anti-CD8 con una consecuente abrogación de la reducción inicial de la viremia, confirman la importancia de los CTLs en el control de la replicación viral [137, 140-142]. Más aún, existe evidencia epidemiológica entre la expresión de moléculas HLA de clase I específicas y velocidades diferentes de progresión de la enfermedad por VIH, lo cual sugiere que la especificidad de las respuestas de CTLs, así como las características de los epítomos blanco pueden influenciar fuertemente la efectividad del control antiviral [143-145] (revisado en [134]; ver más adelante). Además, es claro que individuos homocigotos para alguno de los genes HLA de clase I, cuyas respuestas de CTLs tienen consecuentemente una variedad menor de blancos posibles en el genoma viral, progresan más rápido hacia SIDA [146, 147]. Sin embargo, la evidencia más fuerte que apoya el hecho de que los CTLs son una respuesta determinante del control del VIH y que demuestra que esta respuesta puede ejercer una fuerte presión selectiva dirigiendo la evolución del virus, es tal vez el surgimiento del escape inmune [131].

La presión selectiva por CTLs selecciona virus con mutaciones de escape inmune que evitan la unión de los péptidos virales a las moléculas HLA, que impiden el procesamiento intracelular de los péptidos virales antes de ser presentados en la superficie de la célula o que disminuyen la afinidad de la unión del receptor de células T (TCR) de los CTLs a los complejos péptido-HLA [138, 148-150]. Debido a que péptidos generados a partir de cualquiera de las proteínas virales tienen la capacidad de unirse a y ser presentados por

diferentes moléculas HLA de clase I, la respuesta de CTLs restringida por los genes HLA de clase I, puede seleccionar mutaciones de escape inmune en todo el proteoma viral [131, 151], a diferencia de la respuesta humoral, dirigida preferentemente hacia las glicoproteínas de envoltura del virus. Sin embargo, recientemente se ha reportado que existe una diferencia en la capacidad de respuestas de CTLs dirigidas hacia diferentes proteínas virales para controlar la replicación del VIH, siendo las respuestas dirigidas hacia Gag, las únicas efectivas en el control viral [152]. La selección de mutaciones de escape inmune ocurre tanto en la etapa aguda [153] como en la crónica [138, 154] de la infección, pudiéndose considerar la evolución dirigida por la respuesta de CTLs como un proceso continuo a lo largo de la infección [131]. Las mutaciones de escape se seleccionan frecuentemente dentro de epítomos reconocidos por CTLs; sin embargo, el escape puede asociarse también a mutaciones fuera de epítomos de CTLs relevantes. Algunas mutaciones de escape aparecen en regiones flanqueantes de epítomos, interfiriendo con el procesamiento intracelular y presentación de los péptidos antigénicos [155]. Asimismo, se ha reportado la aparición de mutaciones compensatorias, que estabilizan mutaciones primarias con altos costos en la aptitud replicativa del virus y que surgen como consecuencia de la presión selectiva inmune [156].

Recientemente, se han identificado mutaciones de escape a CTLs seleccionadas de manera reproducible en el contexto de su restricción específica HLA [131]. Asimismo, se ha observado que la aparición de las mutaciones de escape seleccionadas por alelos HLA específicos sigue también un orden reproducible [138, 157]. Esto trajo como consecuencia la importante observación de que la evolución del VIH sigue patrones altamente predecibles en respuesta a la presión selectiva inmune. El hecho de que los sitios y los patrones de escape del VIH sean predecibles, implica que aún con la gran diversidad en secuencias del VIH en el mundo, existen restricciones importantes para la evolución del virus [158], cuyo conocimiento podría incorporarse al diseño de inmunógenos. Sin embargo, la descripción de un mapa comprensivo de las mutaciones de escape a CTLs posibles para el VIH ha estado limitada por estudios observacionales pequeños y por sesgos hacia el estudio de alelos HLA considerados como protectores hacia el avance de la enfermedad por VIH o que se presentan frecuentemente en algunas poblaciones altamente estudiadas [131] (revisado en [134]).

### **1.7.5 La selección por CTLs tiene un impacto en la evolución del VIH a nivel poblacional.**

El hecho de que la respuesta de CTLs, definida por los genes HLA de clase I, dirige y restringe de manera predecible la evolución del VIH en un individuo infectado, trae la consecuencia inmediata de que la evolución poblacional del virus y su adaptación a diferentes poblaciones humanas podría también ser un fenómeno predecible [131]. Las mutaciones de escape frecuentemente se asocian a un costo sustancial en la aptitud replicativa (fitness) del virus y revierten al ser transmitidas a un individuo carente del alelo HLA selector [159-163]. Sin embargo, se han identificado también mutaciones neutrales para la aptitud replicativa del virus o mutaciones cuyo costo en aptitud replicativa es compensado por otras mutaciones asociadas, las cuales pueden transmitirse y mantenerse en un nuevo hospedero aún sin la presencia del alelo HLA selector [149, 164-166]. La hipótesis de la “huella” de los genes HLA de clase I en la evolución del virus a nivel poblacional fue propuesta por primera vez por Moore y colaboradores [167]. Esta hipótesis sugiere la adaptación del virus a las respuestas de CTLs restringidas por los alelos HLA más frecuentes en una población, la cual se refleja en la secuencia consenso de los virus circulantes. Esta huella de la respuesta inmune sobre el VIH a nivel poblacional se ha demostrado a partir de entonces en varias cohortes mediante asociaciones entre la presencia de polimorfismo en posiciones específicas de la secuencia viral y la expresión de alelos HLA determinados [167-173]. Esta evidencia de la “huella” poblacional de los genes HLA en la evolución del virus ha llevado a sugerir que la selección repetida de mutaciones de escape neutrales en sus costos en aptitud replicativa para el virus en el contexto de alelos HLA expresados en altas frecuencias en la población, pueden llegar a la fijación de epítomos de CTLs “escapados” en la población, con la consecuente adaptación del virus a las respuestas inmunes más frecuentes a medida que la epidemia progresa (revisado en [131]). La factibilidad de esta hipótesis se ha reforzado con la descripción de las llamadas “asociaciones negativas” en las que la presencia de un alelo HLA determinado correlaciona con la falta de polimorfismo en posiciones específicas del genoma viral [160, 167]. Esta asociación entre respuestas de CTLs específicas y la conservación de la secuencia viral se ha explicado como sitios donde mutaciones de escape se han fijado en la secuencia consenso de los virus circulantes, haciendo al VIH menos inmunogénico hacia respuestas inmunes frecuentes en la población [160].

El papel de los CTLs en la evolución del virus a nivel poblacional está influenciado por una compleja interacción entre múltiples fuerzas selectivas conflictivas que complican la estimación de la magnitud del impacto de esta respuesta inmune en la evolución del virus a nivel poblacional. Ejemplos notables de estas fuerzas selectivas conflictivas son el balance entre los beneficios del escape inmune contra los costos asociados de aptitud replicativa [156, 160, 161, 163], la complejidad de la respuesta de CTLs, muchas veces con varias



respuestas con diferentes restricciones HLA actuando sobre una misma parte de la secuencia viral [174] o con el posible reconocimiento de mutantes de escape por respuestas con restricciones HLA distintas [175-177], la presencia de eventos fundadores que promuevan la acumulación de ciertas mutaciones [134], las jerarquías de inmunodominancia de las respuestas de CTLs [169, 178, 179] y la distribución de frecuencias de los alelos HLA en diversas poblaciones [134, 180]. Sin embargo, a pesar de esta complejidad, la vasta evidencia reunida hasta el momento sugiere fuertemente que la existencia de una huella poblacional de los genes HLA en la evolución del VIH es un fenómeno evidente.

### **1.7.6 Limitaciones y dificultades para calcular la magnitud de la influencia de los genes HLA en la evolución del VIH a nivel poblacional**

El efecto de la “huella” de los genes HLA de clase I en la secuencia de los virus circulantes en diferentes cohortes se ha demostrado a través de asociaciones correlativas entre la presencia de polimorfismo en posiciones específicas de la secuencia viral y la expresión de alelos HLA determinados que demuestran la selección de mutaciones de escape inmune por la respuesta de CTLs a nivel poblacional [167-173]. Sin embargo, se ha criticado el uso de pruebas estadísticas paramétricas (prueba exacta de Fisher, regresión logística) para obtener este tipo de asociaciones correlativas ya que asumen independencia entre las secuencias virales analizadas, cuando éstas podrían estar relacionadas filogenéticamente [131]. Este problema es especialmente aparente en cohortes heterogéneas con la presencia de varios subtipos virales, donde las pruebas estadísticas podrían identificar correlaciones entre polimorfismos específicos de algún subtipo viral y alelos HLA enriquecidos en individuos infectados por ese subtipo. En este caso, la asociación no estaría mostrando necesariamente que el polimorfismo se está seleccionando bajo presión inmune contemporánea, sino que estaría reflejando un evento fundador. Incluso en cohortes relativamente homogéneas, el no considerar las relaciones evolutivas entre los virus circulantes puede incrementar la varianza de las pruebas estadísticas tradicionales y reducir la certeza en la interpretación de los resultados [131]. Para abordar este problema, en un trabajo reciente por Bhattacharya y colaboradores [181] se propusieron métodos de corrección filogenética para identificar sitios de selección positiva por CTLs. Estos métodos permiten identificar asociaciones HLA – polimorfismo de VIH, que pueden ser explicadas mejor por eventos fundadores, además de identificar asociaciones nuevas que podrían pasarse por alto al no considerar la estructura del árbol filogenético de las secuencias de los virus circulantes [181].

Aunque se ha debatido la importancia de la evolución poblacional dirigida por los genes HLA en el VIH, la evidencia experimental y el desarrollo de análisis estadísticos sofisticados que incorporan correcciones filogenéticas, demuestran un papel sustancial de la respuesta de CTLs en la dirección de la evolución del virus a nivel poblacional. Un trabajo reciente por Brumme y colaboradores [170] demostró que al menos un 40% de la variabilidad en los codones del subtipo B está dirigida por la presión inmune asociada a los genes HLA de clase I. Este estudio identificó diferencias importantes en el número de eventos de selección observados en diferentes proteínas del VIH, siendo Nef la proteína con mayor evidencia de adaptación inmune con 40% de sus codones identificados como bajo presión inmune. Esto contrasta con proteínas como la proteasa, RT o Vpr, donde 10 – 15% de los codones mostraron evidencia de selección por respuestas asociadas a los genes HLA de clase I [170].

Una limitación importante de la mayoría de los estudios que han tratado de caracterizar el impacto de la presión inmune sobre la evolución poblacional del VIH, es la potencia estadística. La incorporación de correcciones filogenéticas en los análisis de correlaciones HLA – polimorfismo de VIH, reduce considerablemente el número de asociaciones relevantes [181] pudiéndose menospreciar equivocadamente el papel de los CTLs en moldear la evolución poblacional del virus. Sin embargo, ahora se sabe que el tamaño de muestra es un factor determinante en la potencia del análisis para detectar asociaciones. Se ha calculado de manera preliminar que una cohorte de aproximadamente 700 individuos, como la estudiada en el trabajo de Brumme y colaboradores [170], tendría poder estadístico para detectar solamente un 50% de las asociaciones HLA – polimorfismo de VIH existentes [131]. Este hecho demuestra la importancia del estudio de cohortes grandes para aumentar la potencia para identificar sitios bajo presión selectiva asociada a los genes HLA de clase I.

Otros factores que confunden la identificación de posiciones bajo presión selectiva por CTLs en la secuencia viral y que limitan las conclusiones de los análisis del impacto de las respuestas asociadas a HLA en la evolución del VIH son el efecto del desequilibrio de ligamiento entre los genes HLA y la co-evolución de aminoácidos en las secuencias virales [131]. Estos factores han sido poco abordados en los estudios existentes y requerirán mayor atención en estudios posteriores. El fenómeno de desequilibrio de ligamiento es común entre los genes del complejo mayor de histocompatibilidad, cuya proximidad hace que con frecuencia no se hereden de manera independiente. Así, es común que alelos HLA determinados tiendan a heredarse de manera conjunta, incrementando la frecuencia de ciertos haplotipos HLA en la población. El fenómeno de desequilibrio de ligamiento puede

provocar que se encuentren asociaciones HLA – polimorfismo de VIH no con el alelo HLA selector, sino con alelos HLA que se heredan de manera conjunta con él, confundiendo las conclusiones de los análisis evolutivos. Algunas mutaciones de escape ocurren en el contexto de mutaciones secundarias o compensatorias. Debido a la existencia de esta co-evolución entre codones de la secuencia viral, es posible que se encuentren asociaciones HLA – polimorfismo de VIH tanto en posiciones de mutaciones primarias de escape como en posiciones de mutaciones compensatorias. El análisis del papel de los genes HLA en la aparición de mutaciones compensatorias ha resultado complejo ya que los efectos filogenéticos son un factor de confusión más grande cuando varios codones comparten la misma historia evolutiva [182]. Recientemente, Carlson y colaboradores propusieron un modelo evolutivo de redes genéticas de dependencia en el que se toman en cuenta los principales factores que causan confusión en la caracterización de la evolución del VIH dependiente de HLA: co-variación de codones virales, desequilibrio de ligamiento de alelos HLA y la filogenia de los virus circulantes en la población [183]. Este modelo de redes filogenéticas de dependencia (PDN) confirmó que los sitios de escape, orden del escape y la aparición de mutaciones compensatorias son grandemente consistentes, incluso entre virus de diferentes clados. Sin embargo, también confirmó que existen diferencias entre la evolución del virus en poblaciones infectadas por virus de diferentes clados. El modelo PDN ha mostrado ser capaz de reconstruir vías de mutaciones de escape y compensatorias previamente definidas y ser concordante con información reciente de patrones de reconocimiento de epítomos [131]. La existencia de este tipo de modelos comprensivos representa una oportunidad para estudiar la evolución del VIH de manera sistemática en poblaciones inmunogenéticamente diferentes para establecer la importancia de la existencia de diferentes contextos HLA en la evolución del virus a nivel poblacional.

El comprender la complejidad y el grado de consistencia del escape inmune del VIH a nivel poblacional tiene implicaciones importantes para el diseño de vacunas basadas en respuestas de CTLs. Sin embargo, a pesar de que se han realizado grandes avances en el conocimiento de los mecanismos poblacionales de evolución del virus, aún falta mucho por comprender para desentrañar la complejidad de este fenómeno. Aunque se ha estudiado exhaustivamente el impacto de los genes HLA en la evolución del virus en cohortes seleccionadas con un gran número de individuos, no se han realizado comparaciones entre poblaciones inmunogenéticamente diferentes en el contexto de sus distribuciones variables de frecuencias de alelos HLA. Estas diferencias en frecuencias podrían poner de manifiesto características únicas de evolución del virus en poblaciones distintas, enriqueciendo el conocimiento universal de la evolución del virus y contribuyendo a la formación de un mapa universal de posiciones en el genoma viral sometidas a presión selectiva por respuestas de

CTLs. La identificación de sitios de evolución preferentemente observados en poblaciones específicas es de gran relevancia para la efectividad regional de posibles vacunas basadas en respuestas de CTLs. Estudios de evolución del VIH en poblaciones inmunogenéticamente diferentes, aún contando con cohortes con un número relativamente limitado de individuos, podrían tener potencias estadísticas variables para identificar diferentes asociaciones HLA – polimorfismo de VIH. Asociaciones que difícilmente podrían identificarse en una población, podrían observarse con alta potencia en poblaciones con distribuciones de frecuencias HLA diferentes. De ahí que sea importante contar con estudios evolutivos comparables en poblaciones diversas.

### 1.7.7 Implicaciones del polimorfismo de los genes HLA de clase I para la evolución del VIH

Tabla 1.1 Asociaciones entre alelos HLA de clase I y resultados en la infección por VIH.

Alelo HLA de clase I	Avance de la enfermedad	Setpoint de carga viral	Referencias	Epítomos implicados	Proteína del VIH	Referencias
HLA B*1302	Lento	Bajo	[184, 185]	RQANFLGKI	Gag 429-437	[134]
HLA B*1801	Rápido	Alto	[144, 186]	-	-	-
HLA B*2705	Lento	Bajo	[187, 188]	KRWIILGLNK	Gag 263-272	[149, 154, 189, 190]
HLA B*3502	Rápido	Alto	[146, 191]	-	-	-
HLA B*3503						
HLA B*5101	Lento	Bajo	[187, 188]	TAFTIPSI	RT 128-135	[134, 180]
HLA B*5701	Lento	Bajo	[186-188, 192-194]	TSTLQEQIAW	Gag 240-249	[160, 192]
HLA B*5702				KAFSPEVIPMF	Gag 162-172	[152, 195]
HLA B*5703				ISPRTLNAW	Gag 147-155	[148, 192]
				IVLPEKDSW	RT 235-242	[196]
HLA B*5801	Lento	Bajo	[144, 193]	TSTLQEQIAW	Gag 240-249	[160, 192]
				ISPRTLNAW	Gag 147-155	[148, 192]
HLA B*5802	Rápido	Alto	[144, 152]	QTRVLAIERYL	gp41 577-587	[197]
HLA B*8101	Lento	Bajo	[144]	-	-	-

\*Modificado y adaptado de [134].

Los genes del complejo mayor de histocompatibilidad (MHC), localizados en el brazo corto del cromosoma 6, son centrales en la respuesta inmune al restringir la especificidad de la respuesta de CTLs. De los tres loci MHC de clase I clásicos, HLA-A, HLA-B y HLA-C, el locus B es el más polimórfico no sólo dentro del cromosoma 6, sino de todo el genoma humano [198]. Este polimorfismo notable se restringe a residuos que conforman el sitio de unión al péptido, los cuales definen el tipo de péptidos que pueden unirse a cada molécula HLA. Este elevado polimorfismo, se aúna a la poligenia de los genes HLA de clase I,

pudiendo cada individuo expresar hasta 6 moléculas clásicas diferentes (2 moléculas A, 2 moléculas B y 2 moléculas C) con la posibilidad de reconocimiento de varios epítomos simultáneos en un patógeno determinado. La importancia del polimorfismo de los genes HLA de clase I radica en que las pequeñas diferencias existentes entre las múltiples moléculas HLA y entre los péptidos que cada una de ellas es capaz de unir y presentar a CTLs, pueden representar diferencias funcionales cruciales en la respuesta hacia un patógeno determinado que incluso pueden estar sujetas a selección natural Darwiniana [134]. En el caso del VIH, existen alelos HLA asociados a un buen control del virus, con un avance consecuentemente más lento hacia SIDA y por el contrario, se han descrito también alelos HLA de riesgo, asociados con un avance rápido de la enfermedad por VIH (revisado en [134]; Tabla 1).

## 2. PLANTEAMIENTO DEL PROBLEMA

La respuesta de linfocitos T citotóxicos (CTLs), definida y restringida por las moléculas del antígeno leucocitario humano (HLA) de clase I, es una presión selectiva importante en la evolución del virus de la inmunodeficiencia humana (VIH) [138, 153, 175, 199, 200]. Los CTLs reconocen células infectadas a través de la unión de sus receptores de células T (TCR) con pequeños fragmentos de 8 a 12 aminoácidos (epítomos) de proteínas virales presentados en la superficie de células infectadas por las altamente polimórficas moléculas HLA de clase I. Las células infectadas reconocidas son entonces destruidas a través de la liberación de citotoxinas como perforina y granulisina que forman poros en las membranas celulares y serín proteasas llamadas granzimas que inducen apoptosis [201].

Como consecuencia de la presión selectiva por los CTLs, se seleccionan virus con mutaciones de escape que evitan la unión de los péptidos virales a las moléculas HLA, que intervienen con el procesamiento de los epítomos antes de ser presentados o que disminuyen la afinidad de la unión del TCR a los complejos péptido-HLA [159, 160, 163]. Sin embargo, muchas mutaciones de escape frecuentemente se asocian a un costo sustancial en la aptitud replicativa (fitness) del virus y revierten al ser transmitidas a un individuo carente del alelo HLA selector [159-161, 163]. A pesar de esto, existe evidencia de que algunas mutaciones de escape pueden ser transmitidas y mantenidas en un hospedero aún sin la presencia del alelo HLA selector, favoreciéndose de esta manera su acumulación y fijación en la población, muchas veces gracias a la aparición de mutaciones compensatorias [149, 164-166]. El impacto relativo de diversos factores que podrían influenciar la persistencia de mutaciones de escape en un número grande de individuos dentro de una

población aún no se conoce [172, 180]. La gran variedad de estos factores – tales como los costos en aptitud replicativa que implica para el virus, la inmunogenicidad del epítipo que la contiene, su tiempo de aparición en la infección, la frecuencia poblacional del alelo HLA que la selecciona y la existencia de fuerzas selectivas opuestas actuando en la misma sección del genoma viral – resaltan la complejidad de la adaptación viral a la respuesta inmune a nivel poblacional [159-161, 163, 174, 180].

A pesar de esta complejidad, existe evidencia abundante de que una gran cantidad de mutaciones de escape a CTLs se seleccionan de manera reproducible en el contexto de restricciones HLA específicas, lo cual ha llevado a la observación fundamental de que la evolución del VIH sigue patrones mutacionales generalmente predecibles en respuesta a respuestas inmunes restringidas por HLA específicas [167] (revisado en [131]). Esta huella de la respuesta de CTLs en el VIH se ha demostrado mediante asociaciones correlativas entre la presencia de polimorfismo en posiciones determinadas de la secuencia viral y la expresión de alelos HLA específicos [167, 168, 170, 173]. Otro hallazgo notable que apoya la evidencia de que la presión selectiva por HLA está dejando huella en la evolución del VIH a nivel poblacional, es la observación de que algunos alelos HLA se asocian a la conservación (falta de polimorfismo) de ciertas posiciones en la secuencia consenso del virus [202]. Estas asociaciones “negativas” han sido explicadas como sitios donde existían epítopos previos que se extinguieron debido a la selección de mutaciones de escape por alelos HLA comunes al madurar la epidemia en una población determinada [203, 204].

Debido al alto polimorfismo de los genes HLA, los alelos frecuentes en cada población son diferentes, pudiendo impactar de manera distinta la evolución del VIH a nivel poblacional. Dada la consistencia del efecto de la selección restringida por HLA en la evolución del VIH y las diferentes frecuencias alélicas HLA existentes en diferentes poblaciones, es probable que ciertas asociaciones HLA – polimorfismo de VIH se observen preferentemente en diferentes poblaciones, posiblemente determinando características únicas de la evolución del VIH en diferentes grupos humanos [134, 205]. Para explorar esta posibilidad, se estudió la evolución del VIH restringida por HLA en una cohorte de individuos infectados por virus del clado B, provenientes del centro y sur de México. Los resultados fueron comparados con los previamente observados en una cohorte grande multicéntrica de individuos infectados por virus de clado B provenientes de British Columbia, Canadá; Estados Unidos y la parte occidental de Australia (cohorte internacional colaborativa de adaptación del VIH, IHAC) [170-172, 183, 206]. Para determinar hasta qué grado la huella poblacional del VIH es un fenómeno generalizado, estudios en poblaciones inmunogenéticamente únicas que podrían reflejar diferencias en presión selectiva en

comparación con la observada en otras poblaciones estudiadas serían informativos. La población mexicana se caracteriza por características inmunogenéticas únicas que reflejan principalmente la mezcla de haplotipos HLA caucásicos y amerindios [207, 208]. Hasta el momento no conocemos otros estudios en los cuales se haya estudiado la evolución del VIH en cohortes latinoamericanas como blanco principal. Nuestros datos sugieren que una frecuencia de alelos HLA única en una población previamente no caracterizada está dejando huellas únicas en la evolución del VIH a nivel poblacional. Este hecho subraya la importancia de expandir sistemáticamente el conocimiento del escape de CTLs y la evolución del VIH en poblaciones inmunogenéticamente distintas. Este conocimiento tiene implicaciones importantes para el diseño de vacunas y otras estrategias de tratamiento basadas en respuestas de CTLs.



### **3. OBJETIVOS**

#### **3.1 OBJETIVO GENERAL**

Caracterizar la evolución dependiente de HLA a nivel poblacional en el VIH circulante en la población mexicana, mediante el uso de modelos evolutivos de vanguardia, comparando los resultados obtenidos con los datos conocidos para poblaciones inmunogenéticamente distintas.

#### **3.2 OBJETIVOS PARTICULARES**

- Describir las frecuencias alélicas de los genes HLA de clase I en una población de individuos VIH positivos mexicanos.
- Buscar posibles alelos HLA relacionados con avance rápido o lento de la infección por VIH en el contexto inmunogenético de la población mexicana.
- Buscar asociaciones entre la presencia de polimorfismo en los codones de la proteasa y transcriptasa inversa virales y la expresión de alelos HLA determinados en una cohorte de pacientes mexicanos.
- Buscar evidencias de evolución característica del virus comparando las asociaciones HLA-codón de VIH encontradas en una cohorte mexicana con las previamente observadas en otras cohortes.

- Comparar la evolución mediada por HLA del VIH a nivel poblacional en el compartimento de virus libre en plasma y el compartimento de provirus integrados en células mononucleares de sangre periférica.
- Evaluar el impacto de la terapia antirretroviral en la presión selectiva ejercida por HLA sobre el VIH en el contexto inmunogenético de la población mexicana.

#### **4. HIPÓTESIS**

El contexto inmunogenético único de la población mexicana originará características únicas en la evolución dependiente de HLA del VIH a nivel poblacional en comparación con otras poblaciones inmunogenéticamente diferentes. Estas características únicas de la evolución del VIH a nivel poblacional en la población mexicana se observarán como huellas asociadas a la presión selectiva por HLA ausentes o distintas a las observadas en otras poblaciones con características inmunológicas diferentes.

## 5. MÉTODOS

**PACIENTES.** Se obtuvieron muestras de sangre periférica de 303 individuos infectados por VIH, sin tratamiento ARV previo, en etapa crónica de la enfermedad, del Centro y Sur de la República Mexicana. Los individuos participantes fueron reclutados en diversos centros de salud en el D. F., Estado de México, Puebla, Jalisco, Oaxaca, Guerrero y Chiapas. La participación de todos los individuos en este estudio fue voluntaria, previa firma de consentimiento informado. Las muestras sanguíneas de los individuos participantes fueron enviadas a y procesadas en el Centro de Investigación en Enfermedades Infecciosas (CIENI) del Instituto Nacional de Enfermedades Respiratorias (INER) en la Ciudad de México. Todas las consideraciones éticas asociadas a este proyecto fueron evaluadas y aprobadas por el Comité de Ciencia y Bioética del INER. Para cada paciente se separaron alícuotas de plasma y células mononucleares de sangre periférica (PBMC), las cuales se criopreservaron a -80 y -140°C respectivamente.

**TIPIFICACIÓN DE HLA.** Se purificó DNA genómico a partir de al menos 6 millones de PBMCs utilizando el estuche de reactivos QIAamp DNA Blood Mini Kit (QIAGEN, Valencia, CA) de acuerdo a las especificaciones del fabricante. Los genes A, B y C de la región HLA de clase I se tipificaron a baja/mediana resolución para cada paciente participante en el estudio por SSP-PCR (reacción en cadena de la polimerasa con iniciadores secuencia-específicos). Se utilizó el estuche de reactivos ABC SSP UniTray Kit (Invitrogen, Brown Deer, WI). Brevemente, DNA genómico de cada paciente a 75-125 ng/ $\mu$ l

se utilizó como templado para 95 PCRs con pares de iniciadores específicos diseñados para detectar polimorfismos útiles para la asignación de alelos. Los productos de reacción se corrieron en gel de agarosa (Promega, Madison, WI) al 2.0%. El patrón de amplificación se analizó con el software provisto por el fabricante (UniMatch v3.2), utilizando bases de datos actualizadas para determinar el haplotipo HLA. Todas las reacciones incluyeron un control interno de amplificación para validarse y todas las tipificaciones incluyeron un control de reactivos para detectar posibles contaminaciones.

**ANÁLISIS DE FRECUENCIAS HLA, DETERMINACIÓN DE HAPLOTIPOS HLA Y ANÁLISIS HLA RELACIONADOS.** Las frecuencias alélicas y poblacionales para la cohorte mexicana fueron obtenidas con la herramienta Frequency Analysis de la Base de Datos de VIH de Los Alamos ([www.hiv.lanl.gov/content/immunology/tools-links.html](http://www.hiv.lanl.gov/content/immunology/tools-links.html)). Las frecuencias de haplotipos HLA fueron obtenidas con el software Arlequín v3.11. Debido al hecho de que la cohorte estaba compuesta por individuos no relacionados con antecedentes familiares desconocidos, se realizó una estimación de la fase gamética más probable para cada individuo usando un algoritmo pseudo-Bayesiano para reconstruir la fase gamética de genotipos de múltiples loci, incluido en el software Arlequin v3.11 (Excoffier-Laval-Balding, ELB) [209, 210]. Para buscar los alelos que se encontraban en desequilibrio de ligamiento en la cohorte mexicana se utilizaron las herramientas de la Base de Datos de Inmunología de VIH de Los Alamos ([www.hiv.lanl.gov/content/immunology/hla/index.html](http://www.hiv.lanl.gov/content/immunology/hla/index.html)). Los análisis de desviaciones del equilibrio de Hardy-Weinberg se realizaron con el paquete Arlequín v3.11.

**COMPARACIONES DE FRECUENCIAS HLA ENTRE DIFERENTES COHORTES Y OTROS ANÁLISIS ESTADÍSTICOS.** Las comparaciones de frecuencias de alelos HLA entre la cohorte del presente estudio y las cohortes HOMER, IHAC y la muestra de población abierta de Barquera, et al [207], se llevaron a cabo mediante pruebas de chi cuadrada con pruebas de significancia post hoc de 2X2 determinadas por pruebas exactas de Fisher, corrigiendo para múltiples comparaciones mediante valores q [211]. Los valores significativos se consideraron con un umbral de  $q < 0.05$ . Estos análisis se llevaron a cabo con el ambiente estadístico "R" v2.8.1, usando el paquete qvalue v1.1.

**GENOTIPIFICACIÓN DE VIH A PARTIR DE VIRUS LIBRE EN PLASMA.** Se purificó RNA viral a partir de 1 ml de plasma para cada paciente participante, utilizando el estuche de reactivos QIAamp Viral RNA Mini Kit (QIAGEN, Valencia, CA) de acuerdo a las recomendaciones del fabricante. Se secuenció un fragmento del gen *pol* del VIH, incluyendo

toda la proteasa (99 aminoácidos) y 334 posiciones de la transcriptasa inversa a partir de RNA viral de cada paciente. Las secuencias se obtuvieron en el equipo 3100-Avant Genetic Analyzer (Applied Biosystems, Foster City, CA), utilizando el estuche de reactivos ViroSeq HIV-1 Genotyping System (Celera Diagnostics, Alameda, CA) de acuerdo a las especificaciones del fabricante. Brevemente, fragmentos de 1.3 Kbp del gen *pol* se amplificaron por RT-PCR a partir de RNA viral proveniente del suero de cada paciente. Los productos de PCR fueron purificados con columnas de ultrafiltración y cuantificados en gel de agarosa (Promega, Madison, WI) al 1.5%. Siete PCRs de secuenciación fueron realizadas con siete iniciadores diferentes, asegurando que toda la región en estudio estuviera cubierta con por lo menos dos secuencias. Las secuencias obtenidas fueron ensambladas y alineadas al consenso HXB2, utilizando el software provisto por el fabricante.

**GENOTIPIFICACIÓN DE VIH PROVIRAL EN PBMCs.** Se obtuvieron secuencias de DNA proviral de toda la proteasa y de las primeras 334 posiciones de la transcriptasa inversa del gen *pol* viral, a partir de DNA genómico de PBMCs. El DNA genómico fue purificado para cada paciente participante en el estudio con el estuche de reactivos QIAamp DNA Blood Mini Kit (QIAGEN, Valencia, CA). Un fragmento de aproximadamente 1.5 Kbp fue amplificado mediante PCR anidada utilizando Platinum Taq DNA Polymerase (Invitrogen, Carlsbad, CA), de acuerdo a las recomendaciones del fabricante. Los iniciadores utilizados fueron PR 5' OUTER 5'-CCCTAGGAAAAAGGGCTGTTG-3' y RT 3' OUTER 5'-GTTTTTCAGATTTTTAAATGGCTCTTG- 3', para la primera reacción, y PR 5' INNER 5'-TGAAAGATTGTA CTGAGAGACAGG -3' y RT 3' INNER 5'-GGCTCTTGATAAATTTGATATGTCC -3' para la segunda reacción. Las condiciones de amplificación de las dos reacciones fueron un ciclo de 94°C por 3 min, seguido por 35 ciclos de 94°C por 30 s, 60°C por 30s y 72°C por 2 min y un ciclo de 72°C por 5 min, con 2 mM final de Mg<sup>++</sup>, 0.2 mM final de dNTPs, 0.4 mM final de cada iniciador y 20 ng/μl final de DNA genómico (transfiriendo 10% del producto de la primera reacción como templado de la segunda). En todos los casos se incluyeron controles de contaminación. Los productos de PCR fueron purificados con columnas de sílica con el estuche de reactivos QIAquick PCR Purification Kit (QIAGEN, Valencia, CA) y cuantificados en gel de agarosa (Promega, Madison, WI) al 2.0%. Se realizaron siete PCRs de secuenciación con las mezclas de reacción provistas en el estuche de reactivos ViroSeq HIV-1 Genotyping System (Celera Diagnostics, Alameda, CA), incluyendo para cada muestra siete iniciadores que aseguraran tener al menos dos secuencias en toda la región genética analizada. Las secuencias obtenidas fueron ensambladas y alineadas al consenso HXB2 de VIH, utilizando el software provisto por el fabricante.

GENOTIPIFICACIÓN ULTRASENSIBLE DE VIH A PARTIR DE VIRUS RESIDUAL EN PLASMA DE INDIVIDUOS EN TRATAMIENTO ANTIRRETROVIRAL Y SUPRESIÓN VIRAL EXITOSA. Se obtuvieron secuencias del gen *pol* de virus residual en plasma de individuos con tres meses de tratamiento ARV y supresión viral exitosa (Carga Viral <50 copias/mL sangre), mediante un método modificado de extracción de RNA viral y amplificación se la secuencia genética de interés. Se purificó RNA viral a partir de 10 a 20 mL de plasma para cada individuo, utilizando el estuche de reactivos QIAamp Viral RNA Mini Kit (QIAGEN, Valencia, CA) de acuerdo a las recomendaciones del fabricante, previa ultracentrifugación de las partículas virales a 17,000 rpm por 2 h a 4°C. Un fragmento de aproximadamente 1.5 Kpb fue amplificado mediante RT-PCR anidada utilizando el estuche SuperScript One-Step RT-PCR (Invitrogen, Carlsbad, CA) para la primera amplificación y el el estuche Platinum Taq DNA Polymerase (Invitrogen, Carlsbad, CA) para la segunda amplificación anidada. Los iniciadores utilizados fueron PR 5' OUTER 5'-CCCTAGGAAAAAGGGCTGTTG-3' y RT 3' OUTER 5'-GTTTTTCAGATTTTTAAATGGCTCTTG- 3', para la primera reacción, y PR 5' INNER 5'-TGAAAGATTGTA CTGAGAGACAGG -3' y RT 3' INNER 5'-GGCTCTTGATAAATTTGATATGTCC -3' para la segunda reacción. Las condiciones de amplificación para la primera amplificación fueron un ciclo de 50° por 30 min y un ciclo de 94°C por 3 min, seguidos de 35 ciclos de 94°C por 30 s, 60°C por 30s y 72°C por 2 min. Para la segunda reacción se utilizó un ciclo de 94°C por 3 min, seguido por 35 ciclos de 94°C por 30 s, 60°C por 30s y 72°C por 2 min y un ciclo de 72°C por 5 min y un ciclo de 72°C por 5 min para la segunda amplificación. Para ambas amplificaciones se utilizó 2 mM final de Mg<sup>++</sup>, 0.2 mM final de dNTPs y 0.4 mM final de cada iniciador (transfiriendo 10% del producto de la primera reacción como templado de la segunda). Para aumentar las posibilidades de amplificación, se utilizó todo el RNA viral extraído, realizando 5 reacciones de amplificación anidada para cada individuo. En todos los casos se incluyeron controles de contaminación. Los productos de PCR fueron purificados con columnas de sílica con el estuche de reactivos QIAquick PCR Purification Kit (QIAGEN, Valencia, CA) y cuantificados en gel de agarosa (Promega, Madison, WI) al 2.0%. Se realizaron siete PCRs de secuenciación con las mezclas de reacción provistas en el estuche de reactivos ViroSeq HIV-1 Genotyping System (Celera Diagnostics, Alameda, CA), incluyendo para cada muestra siete iniciadores que aseguraran tener al menos dos secuencias en toda la región genética analizada. Las secuencias obtenidas fueron ensambladas y alineadas al consenso HXB2 de VIH, utilizando el software provisto por el fabricante.

ANÁLISIS EVOLUTIVOS. Se aplicó el modelo de redes de dependencia filogenética (PDN) por Carlson y colaboradores [183], para inferir patrones de escape de CTLs y de covariación de codones en secuencias del gen *pol* de VIH de virus de plasma y de provirus de PBMCs. El modelo PDN se diseñó para considerar simultáneamente la covariación de codones de VIH, el desequilibrio de ligamiento de los genes HLA y los efectos confusores de la filogenia del VIH para evaluar la evolución viral mediada por HLA [183]. Brevemente, el modelo PDN se basa en un modelo multivariado que representa las dependencias probabilísticas entre un atributo blanco (en este caso la presencia o ausencia de aminoácidos en cada uno de los codones de una proteína del VIH) y una serie de atributos predictores (en este caso la presencia o ausencia de aminoácidos específicos en todos los codones de la proteína viral adicionales al codón blanco y/o la presencia o ausencia de todos los posibles alelos HLA), corrigiendo al mismo tiempo para la estructura filogenética de las secuencias virales (Figura 5.1). Una red de dependencia muestra gráficamente qué atributos HLA o codones predicen cada uno de los atributos codón blanco, asociando una distribución de probabilidad para cada atributo codón blanco condicionada por varios atributos codón y/o HLA. Cada distribución de probabilidad local es corregida para la estructura filogenético de las secuencias de VIH.

Para determinar si existe una asociación significativa HLA – codón de VIH o codón de VIH – codón de VIH se compara la verosimilitud de un modelo nulo que supone que la variable blanco no se encuentra bajo presión selectiva del atributo predictor, contra la de un modelo alternativo que supone que la variable blanco está bajo presión selectiva del atributo predictor. Se añaden predictores múltiples al modelo de manera iterativa, usando selección hacia adelante (forward) en la cual el atributo asociado más significativamente se incorpora iterativamente al modelo hasta que ningún atributo alcance un valor de  $p < 0.05$ . El uso de un modelo multivariado evita asociaciones espurias explicadas por la presencia de desequilibrio de ligamiento entre los alelos HLA y por covariación de codones de VIH. Para cada atributo predictor añadido, se registra el tipo de distribución de hoja más significativa a saber: escape, reversión, atracción o repulsión (ver abajo).

La significancia de las asociaciones obtenidas por el modelo PDN se identifica usando valores  $q$  [211], los cuales estiman de manera conservadora la razón de descubrimientos falsos (false discovery rate, FDR) para cada valor  $p$ . La FDR se define como la proporción esperada de falsos positivos entre los resultados obtenidos como significativos, a un valor umbral dado  $t$ , e.g. un valor de FDR de 5% indica que entre todas

las asociaciones significativas obtenidas, en promedio 5% son falsos positivos. En el caso del modelo PDN se usó la aproximación:

$$\text{FDR}(t) \approx \{E[F(t)] / E[S(t)]\}$$

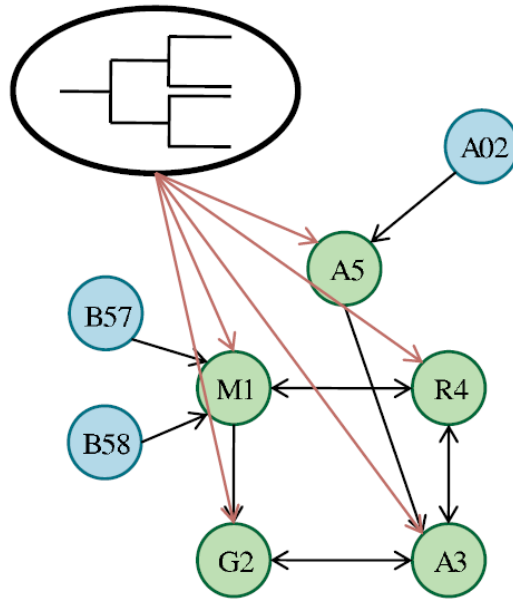
siguiendo el razonamiento de Storey y Tibshirani [211], donde  $S(t)$  es el número de asociaciones identificadas como significativas al valor umbral  $t$  y  $F(t)$  es el número de verdaderos falsos positivos al valor umbral  $t$ . Para estimar el valor de  $E[F(t)]$ , Storey y Tibshirani señalan que la presencia de uniformidad en cada uno de los valores  $p$  (denominados  $p_i$ ) permite hacer la aproximación:

$$E[F(p_i)] \approx \pi_0 p_i m$$

donde  $m$  es el número de valores  $p$  determinados y  $\pi_0$  es una estimación conservadora de la proporción de todas las hipótesis que son verdaderamente nulas. En este caso, se asume *a priori* que la gran mayoría de las hipótesis probadas serán nulas ya que se espera que la mayoría de los codones de VIH y de los alelos HLA no tengan efecto en un atributo blanco determinado. Así, el modelo PDN define  $\pi_0 = 1$ .

Tomando en cuenta estas consideraciones, la significancia del atributo predictor con respecto al atributo blanco en el modelo PDN se computa utilizando FDRs basadas en una prueba de verosimilitud (likelihood ratio test) en la cual el modelo nulo y el alternativo se condicionan a todos los predictores significativos que fueron identificados en iteraciones previas de selección hacia adelante. Para cada valor  $p$ , se reporta un valor  $q$  correspondiente que representa la FDR mínima entre regiones de rechazo que incluyen el valor  $p$ , utilizando el método de Storey y Tibshirani con el parámetro  $\pi_0$  definido conservadoramente como uno [211]. Se excluyeron atributos como posibles predictores cuando el par correspondiente predictor-blanco tenía una tabla de contingencia de 2X2 con un valor de tres o menos en cualquier celda. Se consideraron significativas las asociaciones con valores de  $q < 0.2$ , lo cual significa que se esperaría que un 20% de las asociaciones identificadas como significativas con el modelo PDN, son en realidad asociaciones falsas positivas. Se calcularon las FDR de manera separada para las asociaciones HLA - codón de VIH y codón de VIH – codón de VIH.

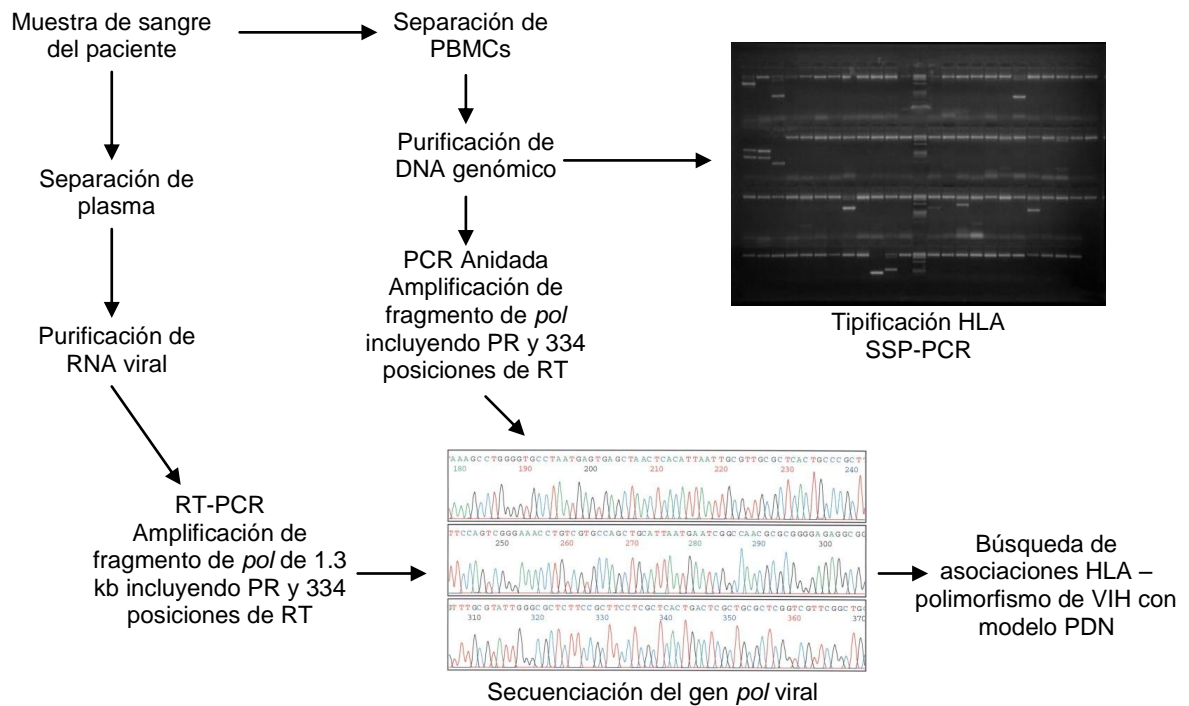




**Figura 5.1. Red de dependencia filogenética.** Una red de dependencia filogenética es un modelo gráfico compuesto de atributos blanco, cuyo resultado es una función probabilística de atributos predictores. En este ejemplo se observan los atributos blanco (círculos verdes) M1, G2, A3, R4 y A5 que representan la presencia o ausencia de aminoácidos en codones de la secuencia viral. Estos atributos blanco dependen de atributos predictores (círculos azules) que pueden ser otros codones (co-variación de codones) o alelos HLA (escape mediado por HLA). Las flechas negras muestran las dependencias encontradas entre los atributos blanco y los atributos predictores. Además, se asume que todos los atributos blanco están influenciados por la filogenia de las secuencias (flechas rojas). La red de dependencia filogenética se compone de probabilidades condicionales locales, cada una de las cuales relaciona a un atributo blanco con la filogenia y con varios atributos predictores y se predice de manera independiente. En este caso, B\*57 y B\*58 predicen a M1 y A\*02 predice a A5. A5 a su vez predice a A3 y M1, G2, A3 y R4 tienen una dependencia cíclica bidireccional. Tomado de [183].

Las reglas precisas que gobiernan las transiciones del atributo blanco (i.e. los cambios en aminoácidos en posiciones específicas del proteoma viral), condicionado por los atributos predictores (i.e. la presencia de aminoácidos en otras posiciones del virus o la expresión de alelos HLA específicos) y la filogenia de las secuencias, están dadas por una distribución hoja univariada que se asume es la misma para cada individuo. Se definen cuatro distribuciones hoja posibles: *atracción*, tener el predictor hace más probable tener el blanco; *repulsión*, no tener el predictor hace menos probable tener el blanco; *escape*, tener el predictor hace menos probable tener el blanco; *reversión*, no tener el predictor hace más probable tener el blanco. El par *atracción / repulsión* corresponde a correlaciones positivas, mientras que el par *escape / reversión* corresponde a correlaciones negativas.

La figura 5.2 muestra un esquema de la estrategia metodológica completa seguida en el presente estudio desde la obtención de las muestras sanguíneas de los individuos participantes hasta la realización de estudios evolutivos de correlación entre la expresión de alelos HLA y la presencia de polimorfismo en la secuencia del VIH.



**Figura 5.2. Esquema de la metodología general seguida para la realización del presente estudio evolutivo.** Se muestra la foto de un gel de tipificación HLA de un paciente, con 96 reacciones para definir la presencia de polimorfismos importantes para tipificar los genes HLA-A, B y C a baja/mediana resolución. Cada línea presenta un control de amplificación, con bandas de amplificación adicionales en los casos de reacción positiva para el polimorfismo analizado. PR – proteasa, RT – transcriptasa inversa, SSP-PCR – reacción en cadena de la polimerasa con iniciadores secuencia específicos, PDN – modelo de redes de dependencia filogenética.

VALIDACIÓN DE ASOCIACIONES HLA-CODÓN DE VIH MEDIANTE MAPEO DE EPÍTOPOS. Las asociaciones HLA-codón de VIH que resultaron estadísticamente significativas al aplicar el modelo PDN fueron validadas usando mapas de epítomos conocidos y predichos. Se buscaron epítomos reportados para cada asociación utilizando la Base de Datos de Inmunología de VIH de Los Alamos ([www.hiv.lanl.gov/content/immunology/index.html](http://www.hiv.lanl.gov/content/immunology/index.html)), dibujando los epítomos relevantes en un mapa genético del virus. Adicionalmente, se realizó una predicción de epítomos a través de motivos de anclaje para cada una de las asociaciones restantes, utilizando las herramientas de la Base de Datos de Inmunología de VIH de Los Alamos ([www.hiv.lanl.gov/content/immunology/tools-links.html](http://www.hiv.lanl.gov/content/immunology/tools-links.html)). Los epítomos relevantes fueron dibujados en el mapa genético del virus.

ANÁLISIS DE SELECCIÓN POSITIVA CON MODELOS DE CODONES. Se buscaron posiciones en el genoma viral con evidencia de estar sujetas a selección positiva, utilizando un modelo de evolución molecular de codones basado en razones de las tasas de sustituciones sinónimas (dS) y no sinónimas (dN), disponible en el servidor Datamonkey ([www.datamonkey.org](http://www.datamonkey.org)). Se utilizó el análisis SLAC (Single-Likelihood Ancestor Counting)

para estimar la presencia de selección positiva [212]. Brevemente, SLAC es un método de conteo para estimar selección positiva, que se basa en calcular el número de sustituciones sinónimas y no sinónimas que han ocurrido en cada codón de la secuencia en estudio a través de su historia evolutiva. SLAC ajusta un modelo de nucleótidos al alineamiento de secuencias y construye el árbol filogenético correspondiente con el método Neighbor-Joining (NJ), utilizando métodos de máxima verosimilitud (ML, maximum likelihood) para obtener la longitud de las ramas y las tasas de sustitución. Posteriormente, SLAC incorpora un modelo de codones al modelo de nucleótidos de la fase anterior, manteniendo los parámetros de longitud de las ramas y tasas de sustitución previamente calculados constantes, con el objetivo de obtener un valor global de  $\omega$ , donde  $\omega = dN/dS$ . A continuación, se reconstruyen secuencias de codones ancestrales sitio por sitio, utilizando métodos de ML e incorporando los parámetros determinados en los ajustes de modelos de nucleótidos y de codones anteriores. Finalmente, se infiere si existe selección calculando los valores normalizados esperados (ES y EN) y observados (NS y NN) de sustituciones sinónimas y no sinónimas para cada sitio. SLAC calcula las razones  $dN = NN/EN$  y  $dS = NS/ES$ , y determina si  $dN < (\text{ó } >) dS$  para cada codón, estimando la significancia mediante valores  $p$  derivados de una distribución binomial de dos colas. Si  $dN > dS$ , entonces se determina que la posición se encuentra bajo presión selectiva positiva; si  $dN < dS$ , entonces se determina que la posición se encuentra bajo presión negativa. Se consideraron como significativas posiciones que presentaron  $dN \neq dS$  con  $p < 0.05$ .

**ANÁLISIS DE CORRELACIÓN ENTRE VALORES SIGNIFICATIVAMENTE MÁS ALTOS O BAJOS DE CARGA VIRAL Y LA EXPRESIÓN DE ALELOS O HAPLOTIPOS HLA ESPECÍFICOS.** Se estimaron las distribuciones de valores de carga viral contra la expresión de alelos HLA y contra la expresión de haplotipos HLA de dos y tres genes en la cohorte mexicana, mediante el programa StatView. Se consideraron en el análisis exclusivamente alelos y haplotipos con frecuencias mayores a 0.5%. Las distribuciones de valores de carga viral para cada alelo o haplotipo HLA se compararon con análisis de varianza (ANOVA), con prueba posthoc de Diferencias Mínimas Significativas Protegidas (PLSD) de Fisher, con un nivel de significancia de  $p < 0.05$ .

**ANÁLISIS DE RESISTENCIA A FÁRMACOS ANTIRRETROVIRALES.** Todas las secuencias fueron analizadas utilizando el programa HIVdb de la base de Datos de Resistencia de la Universidad de Stanford (<http://hivdb.stanford.edu/pages/algs/HIVdb.html>). Este programa calcula un puntaje de riesgo relativo (RR) de resistencia hacia cada uno de los 19 fármacos ARVs disponibles que tienen como blanco a la proteasa viral y a la RT, de

acuerdo a las mutaciones asociadas con resistencia encontradas en cada secuencia. El puntaje de resistencia asociado a cada mutación para cada fármaco antirretroviral se estima mediante algoritmos basados en datos clínicos y experimentos *in vitro*.

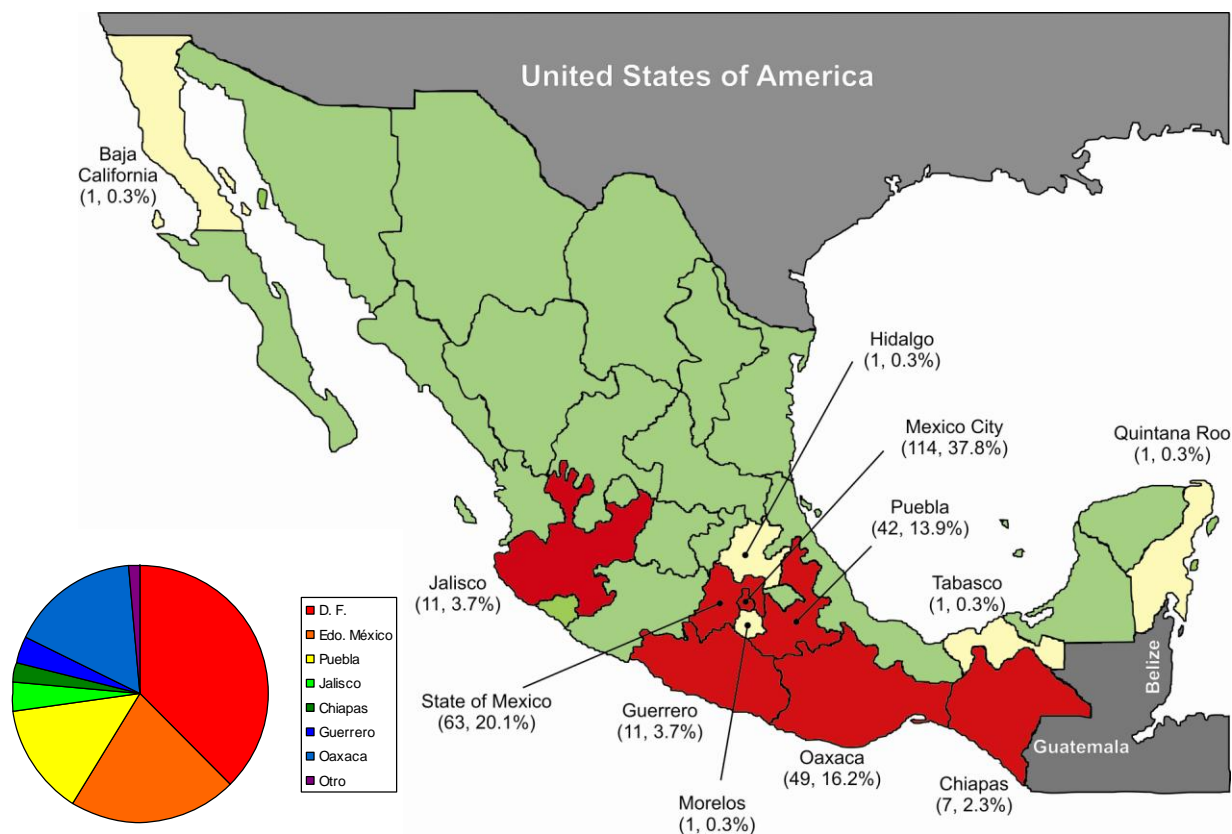
Los individuos con posibilidad de presentar resistencia transmitida se contaron y se estratificaron en las categorías de riesgo relativo de resistencia propuestas en la Base de Datos de la Universidad de Stanford. La interpretación de las categorías del puntaje de riesgo relativo son las siguientes: 1) Pacientes susceptibles (RR 0-9), cuyos virus no muestran sensibilidad reducida a fármacos ARVs; 2) Pacientes potencialmente con nivel de resistencia bajo (RR 10-14), cuyos virus presentan mutaciones que por sí mismas no causan resistencia pero que indican una posibilidad de selección por fármacos previa; 3) Pacientes con nivel de resistencia bajo (RR 15-29), cuyos virus presentan susceptibilidad reducida *in vitro* a fármacos, pudiendo causar respuestas virológicas subóptimas al tratamiento ARV; 4) Pacientes con nivel de resistencia intermedio (RR 30-59), cuyos virus presentan resistencia en un nivel intermedio entre el bajo y alto; y 5) Pacientes con nivel de resistencia alto (RR  $\geq 60$ ), cuyos virus presentan niveles de resistencia altos *in vitro*, trayendo como consecuencia respuestas virológicas deficientes o nulas al tratamiento ARV.

**ANÁLISIS FILOGENÉTICOS.** Se construyó un árbol filogenético basado en matrices de distancias para las secuencias mexicanas a partir de un fragmento de de 1305 pb del gen *pol* del VIH, abarcando todo el gen de la proteasa y dos terceras partes de la RT, en el programa MEGA 4. Se utilizó el método de Neighbor-Joining (NJ), calculando las distancias evolutivas con el modelo de dos parámetros de Kimura (K2P). NJ es un método algorítmico que proporciona una buena aproximación heurística para encontrar el árbol de evolución mínima más corto en distancia genética. Este método secuencialmente encuentra vecinos que minimizan la longitud total del árbol. El modelo K2P es un modelo paramétrico de evolución de sustitución de nucleótidos simple que considera una misma frecuencia de bases ( $\pi_A = \pi_T = \pi_C = \pi_G$ ), pero diferentes tasas de sustitución transicionales/transversionales ( $\alpha \neq \beta$ ).

La resolución del árbol filogenético construido se probó mediante el método de bootstrap, realizando 1000 pseudorréplicas de muestreo aleatorio con reemplazo para estimar un árbol consenso. Se incluyeron 93 secuencias de referencia abarcando los principales subtipos y formas recombinantes del VIH-1, obtenidas de la Base de Datos de Secuencias de Los Alamos (<http://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html>).

## 6. RESULTADOS

### 6.1 CARACTERÍSTICAS GENERALES DE LA COHORTE ESTUDIADA.



**Figura 6.1. Distribución geográfica de los individuos participantes en el estudio.** El mapa muestra datos de 302 individuos VIH positivos, sin tratamiento ARV previo. Los estados en rojo conjuntaron el 98.3% de los individuos participantes, los estados en blanco representaron el 1.7% de los individuos de la cohorte. La gráfica e pastel muestra las proporciones de individuos participantes por entidad federativa.

Para el presente estudio se reunió de manera prospectiva una cohorte de 302 individuos VIH positivos, en etapa crónica de la infección y sin tratamiento ARV previo. La residencia geográfica de los individuos participantes se muestra en la Figura 6.1. Como es característico de las cohortes latinoamericanas [7, 213], cerca de la mitad de los individuos de la cohorte se encontraba en etapa tardía de la infección (conteo de linfocitos T CD4+  $\leq 200$  células/ $\mu$ l) al momento del reclutamiento, con aproximadamente la mitad de estos pacientes tardíos presentándose con menos de 50 células T CD4+ por  $\mu$ l de sangre (Tabla 6.1). Solo un 10% de los individuos reclutados presentaron infecciones tempranas (conteo de linfocitos T CD4+  $> 500$  células/ $\mu$ l). Considerando a toda la cohorte, la media del conteo de linfocitos T CD4+ fue menor a 200 células/ $\mu$ l, reflejando el problema de la detección tardía de la infección. Asimismo, la relación hombres:mujeres en la cohorte fue de 3:1 (Tabla 6.1), con una prevalencia en mujeres ligeramente mayor a la previamente reportada para la

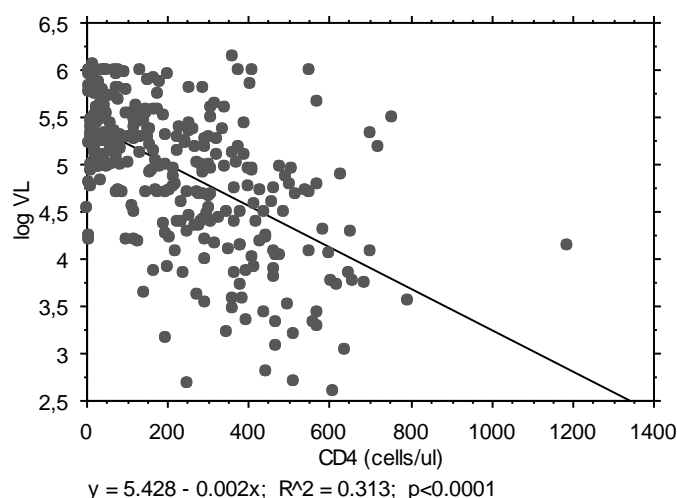
población mexicana [12]. Esto podría reflejar la tendencia de aumento en la proporción de mujeres infectadas en Latinoamérica en los últimos años [213] (<http://www.unaids.org>) o de manera alternativa un sesgo en el reclutamiento de la cohorte incluida en el presente estudio.

Tabla 6.1. Parámetros clínicos relevantes para la cohorte de 303 individuos VIH positivos del estudio.

Parámetros Clínicos	Media	Mediana	Error Estándar	Desviación Estándar
Conteo linfocitos T CD4 <sup>+</sup> (células/μl de sangre)*	238.4	196,5	11.6	201.1
Porcentaje de células T CD4+ (%)*	14.0	12.0	0.6	10.4
Carga Viral (copias RNA de VIH/ml de plasma)	233,528	105,000	16,927	292,691
Log Carga Viral	4.917	5.021	0.044	0.767
<b>Estratificación por conteo de linfocitos T CD4+* [n (%)]</b>				
>500 células/μl	35 (11.6%)	}	148 (49.3%)	
201 – 500 células/μl	113 (37.7%)			
50 – 200 células/ μl	82 (27.3%)			
<50 células/μl	70 (23.3%)	}	152 (50.7%)	
<b>Género [n (%)]</b>				
Masculino			229 (75.6%)	
Femenino			74 (24.4%)	

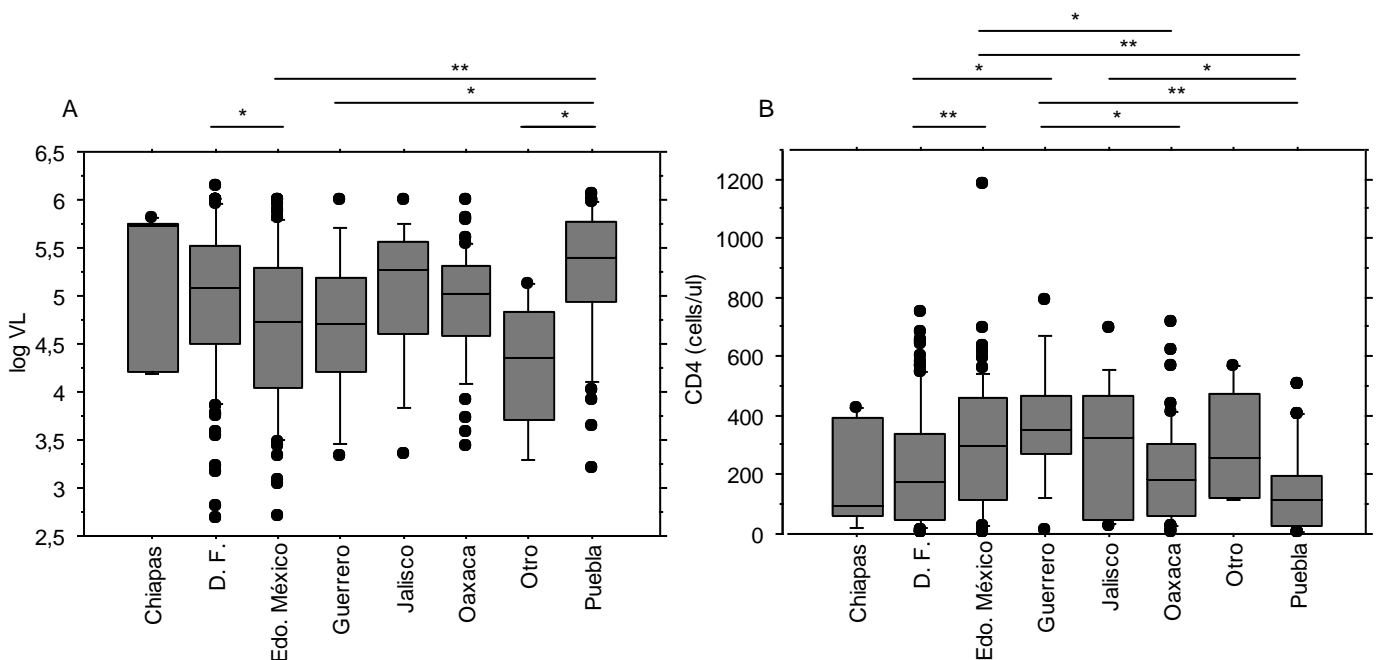
\*Datos para 3 pacientes no disponibles

Como era esperado, se observó una correlación negativa típica entre los valores de carga viral y las cuentas de células T CD4+ para la cohorte mexicana ( $p < 0.0001$ ), con la mayoría de los pacientes presentándose con valores de carga viral altos y cuentas de linfocitos T CD4+ bajas (Figura 6.2). Se observó un aumento promedio de 0.1 logaritmos en la carga viral por cada 50 células T CD4+ por μl de sangre perdidas.



**Figura 6.2. Correlación negativa entre las cuentas de linfocitos T CD4+ y los valores de carga viral.** Se muestran datos de regresión lineal para 300 individuos VIH+ de la cohorte del estudio.

Adicionalmente, se observaron diferencias en las distribuciones de cuentas de linfocitos T CD4+ (Figura 6.3A) y de cargas virales (Figura 6.3B) en los pacientes reclutados en diferentes estados de la República, lo cual sugiere características distintas en las cohortes de pacientes atendidos en diferentes regiones de México. Característicamente, los pacientes reclutados en Puebla presentaron valores de carga viral más altos y cuentas de linfocitos T CD4+ más bajas que varias otras regiones del país. Las cohortes del D. F. y Jalisco (esta última compuesta mayoritariamente por pacientes provenientes de Guadalajara) presentaron distribuciones similares de valores de carga viral, conformándose típicamente por pacientes con cargas virales altas, mostrando cierta similitud en las condiciones de los pacientes atendidos en las grandes urbes. Esto contrasta con lo observado en el Estado de México, que aún con su cercanía al D. F. muestra pacientes en ligeramente mejores condiciones (cargas virales más bajas y cuentas de linfocitos T CD4+ más altas) que los atendidos en la capital (Figura 6.3). En conjunto, estas observaciones son representativas de una cohorte mexicana típica, compuesta principalmente de individuos en etapas tardías de la infección, frecuentemente diagnosticados como VIH positivos al momento de su presentación en los centros de salud debido a síntomas de infecciones oportunistas asociadas al SIDA.

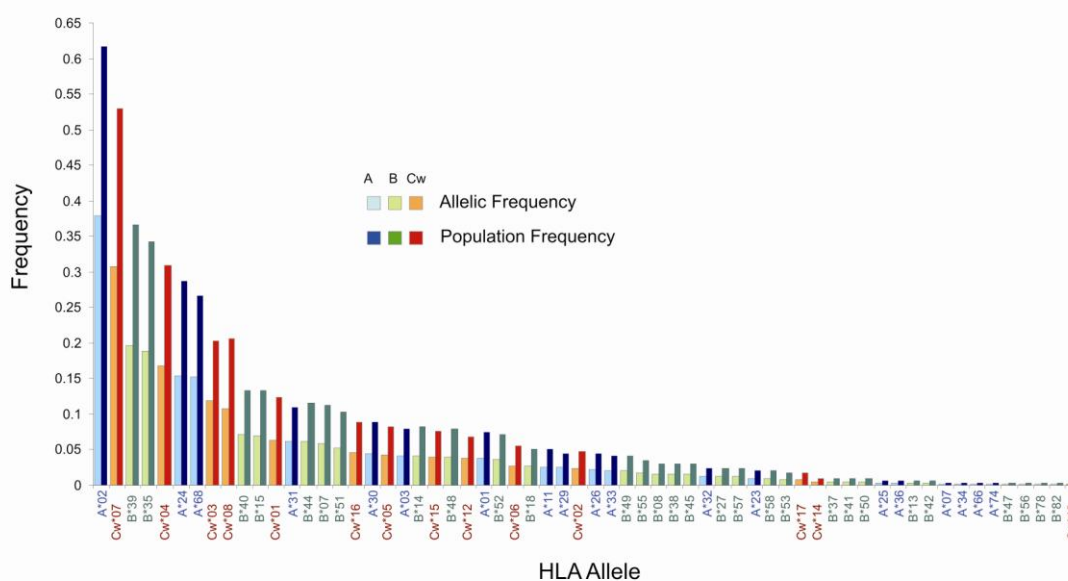


**Figura 6.3. Distribución de valores de carga viral (n=284) (A) y cuentas de linfocitos T CD4+ (n=286) (B) por entidad federativa.** Las cajas representan los percentiles 25 a 75 (50% de los datos). Las cajas más las barras lineales representan los percentiles 10 a 90 (80% de los datos). Los puntos representan datos fuera del percentil 10-90 (20% de los datos). La línea central de las cajas corresponde a la mediana de cada categoría. \*p<0.002, \*\*p<0.0001, ANOVA con prueba posthoc de Bonferroni/Dunn.

## 6.2 CARACTERÍSTICAS INMUNOGENÉTICAS DE LA COHORTE DE INDIVIDUOS VIH POSITIVOS MEXICANOS

### 6.2.1 Frecuencias alélicas y haplotípicas de los genes HLA-A, B y C en una cohorte de individuos VIH positivos mexicanos.

Para describir inmunogenéticamente a la cohorte de pacientes VIH positivos de este estudio, se determinaron las frecuencias alélicas y poblacionales de los loci HLA de clase I A, B y C en 292 individuos con tipificación HLA disponible (Figura 6.4). Los grupos alélicos A más frecuentes fueron A\*02, A\*24, A\*68 y A\*31; los grupos alélicos B más frecuentes fueron B\*39, B\*35, B\*40 y B\*15 y los grupos alélicos C predominantes fueron Cw\*07, Cw\*04, Cw\*03 y Cw\*08 (Figura 6.4). Característicamente, más del 60% de la cohorte presentó algún alelo A\*02, más del 50% presentó algún alelo Cw\*07 y más de una tercera parte presentó alelos B\*39 y/o B\*35.



**Figura 6.4. Frecuencias alélicas y poblacionales para los genes HLA-A, B y C en la cohorte de individuos VIH positivos del centro y sur de la República Mexicana.** El cálculo de frecuencias se realizó con la herramienta HLA Frequency Analysis de la Base de Datos de VIH de Los Alamos ([www.hiv.lanl.gov/content/immunology/tools-links.html](http://www.hiv.lanl.gov/content/immunology/tools-links.html)). La tipificación HLA se realizó a baja/mediana resolución por el método de SSP-PCR como se describe en la sección de materiales y métodos. Se incluyeron un total de 292 individuos y 584 alelos. Se observaron 19, 29 y 14 grupos alélicos diferentes para los genes HLA-A, B y C respectivamente.

Para describir de manera más precisa las características inmunogenéticas de la cohorte del estudio, así como para enriquecer los análisis de la influencia de los genes HLA de clase I en el control de la carga viral, se estimaron las frecuencias de los haplotipos más comunes de dos y tres genes para los loci HLA-A, B y C en la cohorte de pacientes VIH positivos mexicanos. Debido a que la cohorte del estudio estaba compuesta por individuos

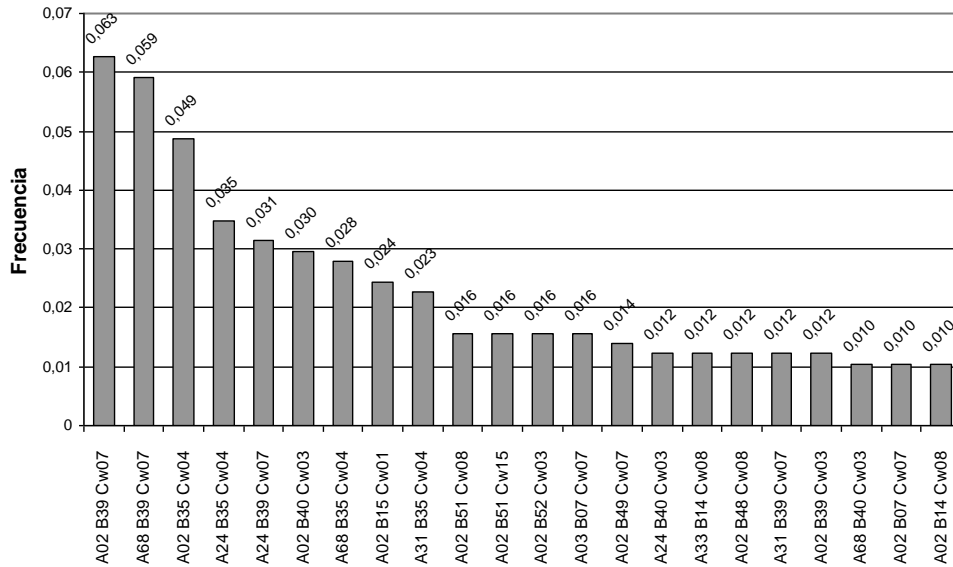


no relacionados, cuyos antecedentes genéticos familiares eran desconocidos, se realizó una estimación de la fase gamética más probable para cada individuo. Es decir, se estimaron las combinaciones alélicas originales de los loci HLA-A, B y C que cada individuo heredó de sus padres. Esta estimación es necesaria para describir de manera precisa la distribución haplotípica de cualquier población con antecedentes familiares desconocidos.

Se utilizó un algoritmo pseudo-Bayesiano (Excoffier-Laval-Balding, ELB) [210], diseñado para reconstruir la fase gamética de genotipos con múltiples loci, incorporado al programa de análisis de genética de poblaciones Arlequín v3.11. Este algoritmo realiza actualizaciones de fase con base en una ventana de loci vecinos cuyo tamaño varía de acuerdo al nivel de desequilibrio de ligamiento local en los datos analizados (para una descripción del fenómeno de desequilibrio de ligamiento en la cohorte mexicana ver más adelante). La aplicación del algoritmo ELB permitió obtener una lista de los individuos participantes con sus alelos HLA-A, B y C ordenados de acuerdo a su fase gamética más probable. A partir de esta lista ordenada se estimaron entonces las frecuencias genéticas de los haplotipos HLA-A/B/C (Figura 6.5A), y A/B, A/C, B/C (Figura 6.5B) presentes en la cohorte de pacientes VIH positivos del presente estudio.

Por otro lado, se han utilizado algoritmos de máxima verosimilitud (ML, maximum likelihood) para el análisis de datos genotípicos con fase gamética desconocida [209]. El método de estimación de frecuencias haplotípicas por ML se basa en un algoritmo iterativo de Maximización de Expectancia (EM, Expectation-Maximization) en el cual se estiman las desviaciones estándar de las frecuencias haplotípicas mediante un análisis de bootstrap paramétrico a partir de muestras aleatorias de la población. En cada iteración del análisis de bootstrap se asume que estas muestras tienen frecuencias haplotípicas iguales a sus valores de ML y se aplica el algoritmo EM para obtener frecuencias haplotípicas de ML nuevas. Para establecer la conveniencia de utilizar el algoritmo ELB con respecto al algoritmo EM, se realizó una comparación de las frecuencias haplotípicas HLA obtenidas por cada uno de los dos métodos en nuestra cohorte de pacientes VIH positivos mexicanos, utilizando el programa Arlequín v3.11 (Figura 6.6). Aunque en general las frecuencias haplotípicas obtenidas por ambos métodos fueron similares, se observaron frecuentemente casos donde la frecuencia fue subestimada con el algoritmo EM, respecto al algoritmo ELB. Asimismo, se observaron discordancias en la definición de haplotipos, especialmente en los de menor frecuencia (Figura 6.6). Dados estos resultados, se decidió usar el algoritmo ELB para estimar la fase gamética de los loci HLA estudiados, previamente al cálculo de las frecuencias haplotípicas, dándonos mayor confianza en los análisis haplotípicos posteriores.

A

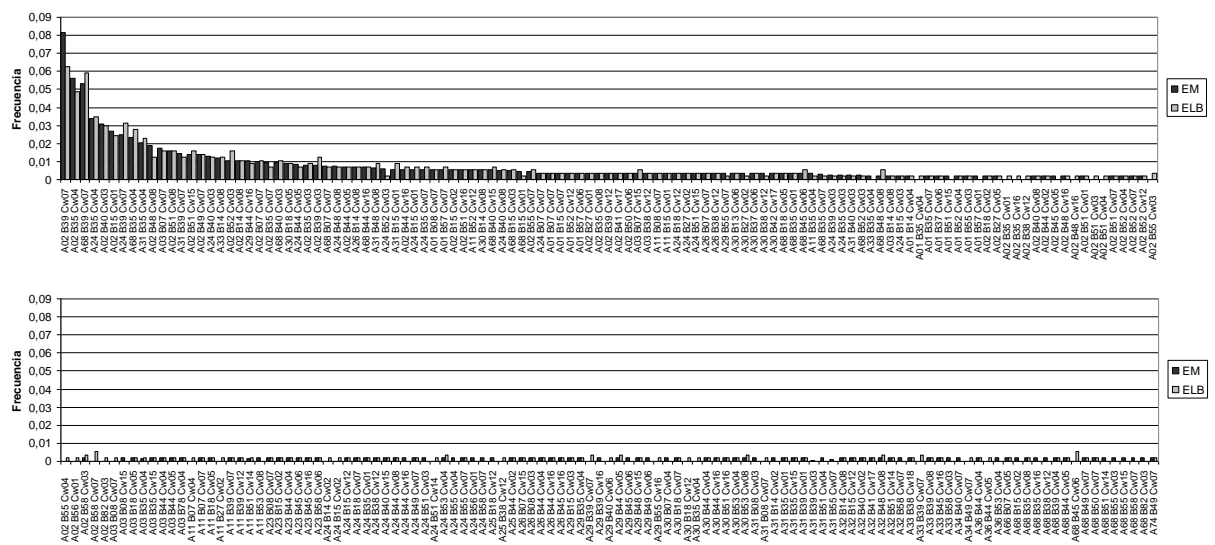


B

HLA A-B	Frecuencia	HLA B-C	Frecuencia	HLA A-C	Frecuencia
A*02 B*39	0.0938	B*39 Cw*07	0.17422	A*02 Cw*07	0.10627
A*02 B*35	0.0486	B*35 Cw*04	0.14286	A*68 Cw*07	0.07840
A*24 B*35	0.0486	B*40 Cw*03	0.05401	A*02 Cw*03	0.07143
A*68 B*39	0.0486	B*07 Cw*07	0.04530	A*02 Cw*08	0.05226
A*02 B*51	0.0451	B*15 Cw*01	0.04530	A*02 Cw*04	0.04878
A*68 B*35	0.0451	B*14 Cw*08	0.03833	A*24 Cw*07	0.04355
A*02 B*40	0.0399	B*48 Cw*08	0.03659	A*02 Cw*01	0.03833
A*02 B*15	0.0330	B*44 Cw*16	0.02613	A*24 Cw*04	0.03833
A*31 B*35	0.0278	B*44 Cw*05	0.02265	A*31 Cw*04	0.02613
A*24 B*39	0.0243	B*49 Cw*07	0.02091	A*02 Cw*15	0.02265
A*03 B*07	0.0208	B*52 Cw*03	0.01916	A*68 Cw*04	0.02265
A*24 B*15	0.0208	B*51 Cw*15	0.01916	A*24 Cw*03	0.02265
A*02 B*52	0.0208	B*35 Cw*07	0.01742	A*01 Cw*07	0.02091
A*31 B*39	0.0191	B*08 Cw*07	0.01568	A*31 Cw*07	0.01916
A*02 B*48	0.0156	B*18 Cw*05	0.01568	A*68 Cw*03	0.01916
A*02 B*49	0.0139	B*51 Cw*08	0.01568	A*02 Cw*16	0.01568
A*24 B*40	0.0139	B*38 Cw*12	0.01394	A*03 Cw*07	0.01568
A*30 B*18	0.0122	B*39 Cw*03	0.01394	A*24 Cw*08	0.01394
A*68 B*40	0.0122	B*35 Cw*03	0.01045	A*33 Cw*08	0.01394
A*24 B*44	0.0122	B*52 Cw*12	0.01045	A*29 Cw*16	0.01220
A*29 B*44	0.0122	B*55 Cw*07	0.01045	A*30 Cw*05	0.01045
A*33 B*14	0.0122			A*31 Cw*01	0.01045
A*01 B*57	0.0104			A*24 Cw*05	0.01045
A*02 B*07	0.0104				
A*02 B*44	0.0104				
A*02 B*55	0.0104				

**Figura 6.5. Haplotipos de tres (A) y dos (B) genes HLA de clase I más frecuentes en una cohorte de 292 pacientes VIH positivos del centro y sur de la República Mexicana.** La fase gamética de cada individuo se estimó con el algoritmo pseudo-Bayesiano ELB, del software Arlequín v3.11. Se muestran los haplotipos con frecuencias mayores a 1%.

Aplicando el algoritmo ELB, se identificaron un total de 192 haplotipos HLA A-B-C posibles, de los cuales sólo 22 presentaron una frecuencia mayor al 1% en la cohorte. Los haplotipos más frecuentes fueron A\*02/B\*39/Cw\*07, A\*68/B\*39/Cw\*07 y A02/B\*35/Cw\*04, con frecuencias mayores al 4.5% en la cohorte (Figura 6.5A). Considerando dos loci, se encontraron un total de 121 haplotipos posibles para HLA-A/B, 82 para HLA-B/C y 92 para HLA-A/C. Los haplotipos más frecuentes de dos loci resultaron A\*02/B\*39, A\*02/B\*35, A\*24/B\*35 y A\*68/B\*39, para HLA-A/B; B\*39/Cw\*07, B\*35/Cw\*04 y B\*40/Cw\*03, para HLA-B/C; y A\*02/Cw\*07, A\*68/Cw\*07 y A\*02/Cw03, para HLA-A/C (Figura 6.5B). En general, hubo menos variabilidad en los haplotipos HLA-B/C que en los haplotipos HLA-A/C y HLA-A/B, muy probablemente debido al fenómeno de desequilibrio de ligamiento observado frecuentemente entre genes B y C (Tabla 6.2, ver abajo).



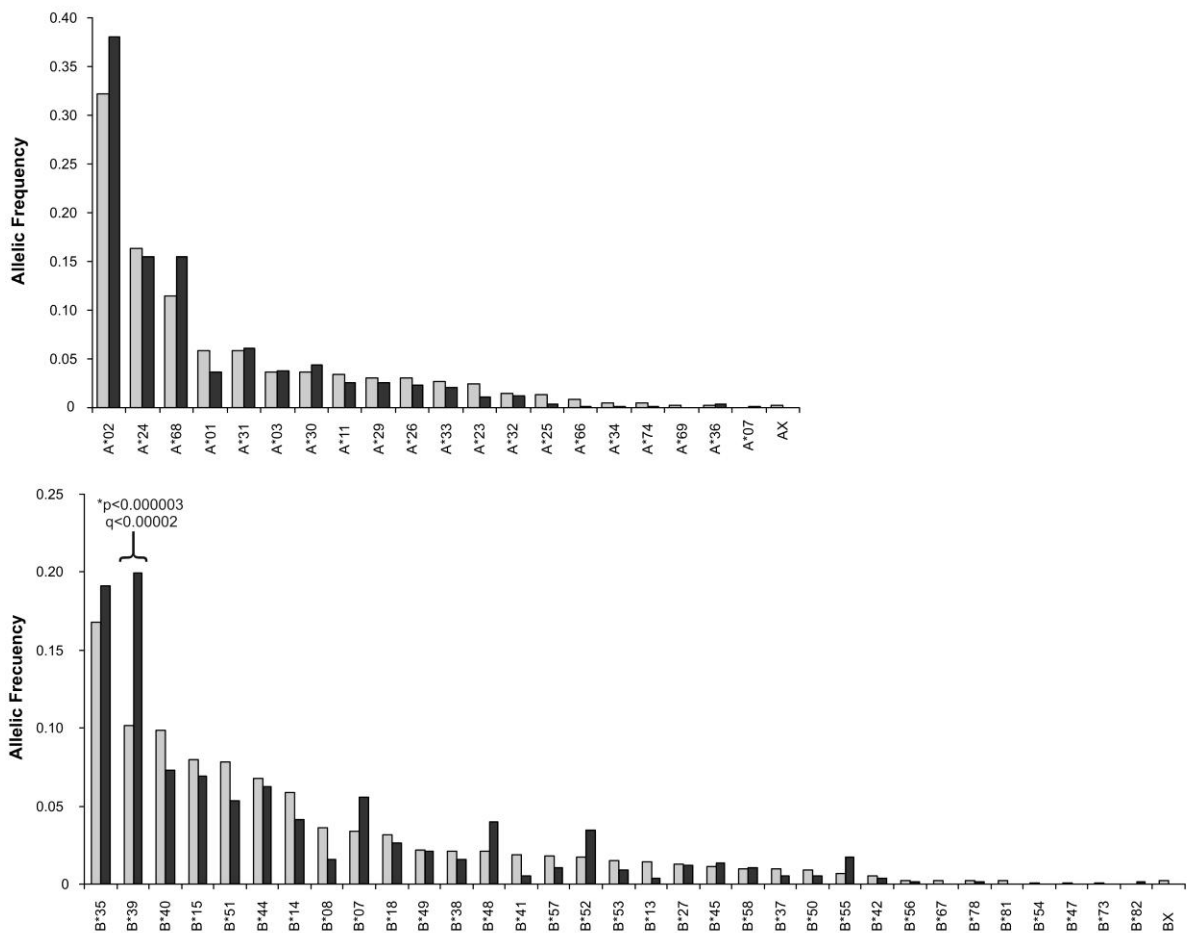
**Figura 6.6. Comparación de los algoritmos EM y ELB para estimar frecuencias haplotípicas en una población de 292 pacientes VIH positivos del centro y sur de la República Mexicana con antecedentes familiares desconocidos.** Se comparó un método tradicional de estimación de frecuencias haplotípicas por máxima verosimilitud, mediante un algoritmo de Maximización de Expectancia (EM, Expectation-Maximization) con un método pseudo-bayesiano de estimación de fase gamética en genotipos con loci múltiples (ELB) mediante el software Arlequín v3.11.

Dentro de nuestro conocimiento, este es el primer reporte formal de frecuencias de alelos HLA de clase I en una cohorte típica de individuos VIH positivos mexicanos.

### 6.2.2 Diferencias en frecuencias alélicas HLA entre la cohorte de individuos VIH positivos mexicanos y la población mexicana general.

Con el objetivo de buscar posibles diferencias en las frecuencias HLA de individuos VIH+ con respecto a la población mexicana en general, las frecuencias alélicas HLA de la

cohorte de individuos VIH+ del presente estudio se compararon con las reportadas en un estudio reciente por Barquera y colaboradores en una cohorte de 381 individuos de 191 familias mexicanas del centro y norte del país [207] (Figura 6.7). Aunque el origen geográfico de la cohorte de individuos del estudio de Barquera varía con respecto a la de este estudio, una alta representación de individuos del Centro del país y el hecho de que la metodología de tipificación HLA para ambos estudios fue la misma, hacen a los datos del estudio por Barquera una buena referencia de una cohorte VIH negativa mexicana para la comparar nuestros datos de individuos VIH positivos.



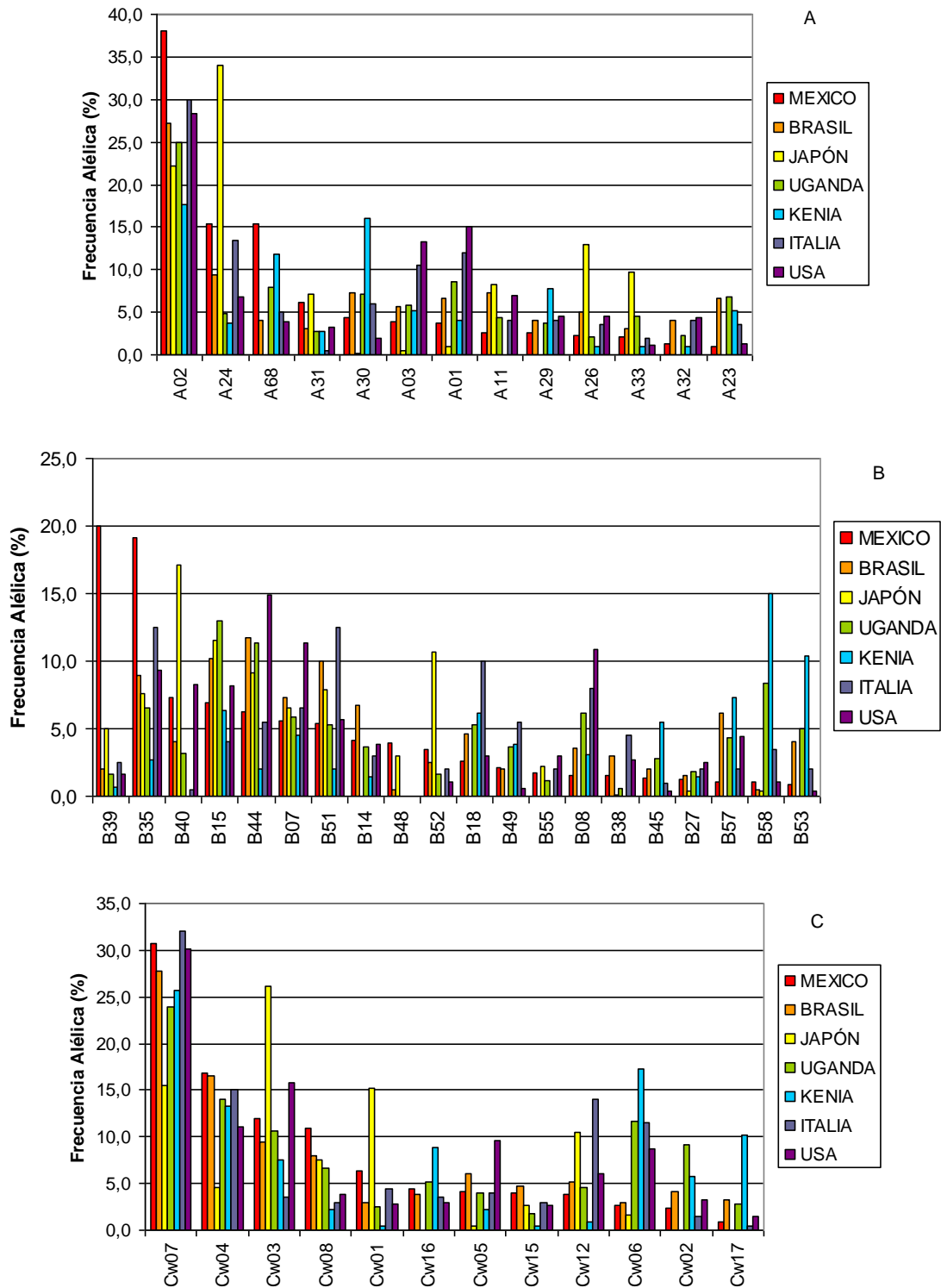
**Figura 6.7. Diferencias en frecuencias HLA entre individuos VIH+ y VIH- mexicanos.** Las frecuencias alélicas de los genes HLA-A y B de la cohorte de 292 individuos VIH+ del presente estudio (gris oscuro) se compararon con las reportadas en un estudio reciente para una cohorte de 381 individuos de 191 familias mexicanas [207] (gris claro). La tipificación HLA en ambos estudios se realizó por SSP-PCR como se indica en la sección de materiales y métodos. Para hacer la comparación, se sustituyó la nomenclatura de uso en histocompatibilidad del estudio de Barquera *et al*, por sus equivalentes genéticos: B65 y B64 se incluyeron dentro de los alelos B\*14; B62, B63, B70, B71, B72 y B75 fueron considerados alelos B\*15; y B61 y B60 se contaron como alelos B\*40, de acuerdo a la nomenclatura aceptada por el Comité de Nomenclatura de Factores del Sistema HLA de la Organización Mundial de la Salud (OMS) (<http://www.ebi.ac.uk/imgt/hla/dictionary.html>).

La distribución de frecuencias alélicas de los loci HLA-A y B fue significativamente diferente entre las dos cohortes ( $\chi^2=99.39$ ,  $p=0.00008$ ), con diferencias evidentes en B\*39

( $p=2.25E-06$ ,  $q=1.19E-04$ ), un alelo HLA típicamente amerindio que mostró una frecuencia cerca de dos veces más alta en el grupo de individuos VIH positivos que en el de VIH negativos (Figura 6.7). Aún queda por confirmar si el presentar alelos B\*39 representa un factor de riesgo para la infección por VIH en México, debido a que la alta frecuencia de este grupo alélico podría estar reflejando también un fenómeno epidemiológico causante del enriquecimiento del alelo en los sectores poblacionales más afectados por la epidemia, o simplemente un sesgo de muestreo en alguno de los dos estudios en comparación.

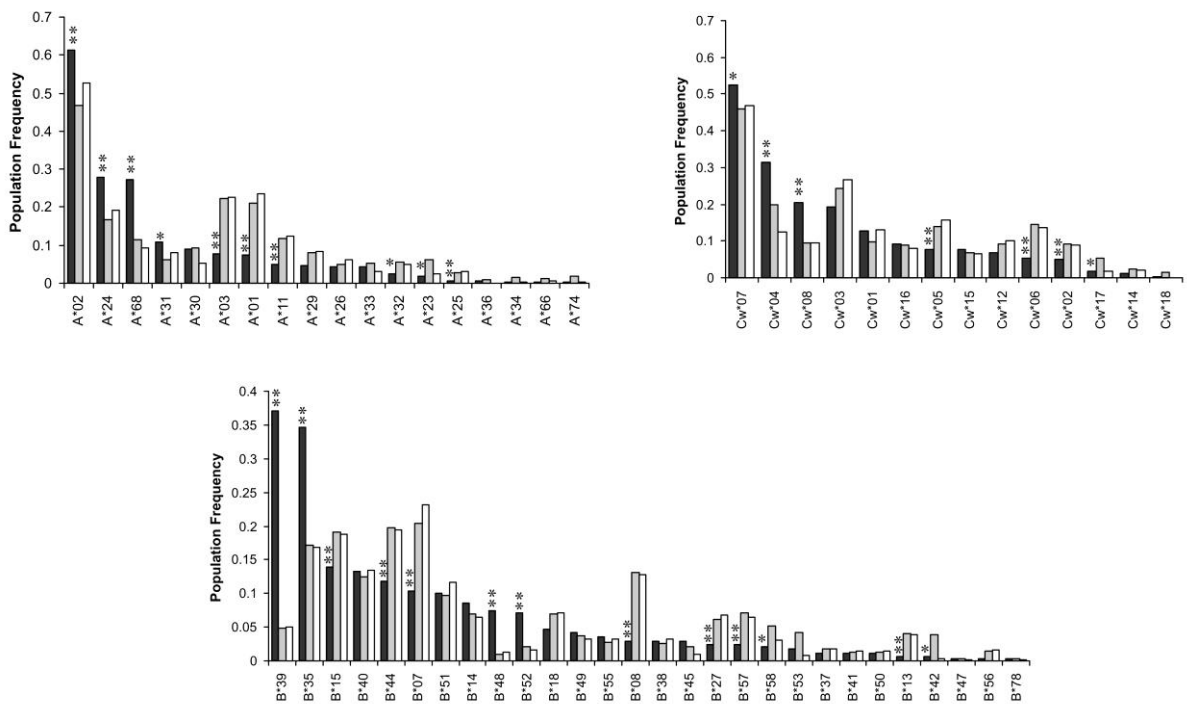
### **6.2.3 Contexto inmunogenético único de la cohorte de individuos VIH positivos mexicanos: comparación con otras cohortes de individuos VIH positivos y VIH negativos del mundo.**

Es bien sabido que existe una diversidad extraordinaria en la expresión de alelos HLA en diferentes poblaciones humanas [134]. Para demostrar la distribución alélica única de la población mexicana con respecto a otras poblaciones, se realizó una comparación entre frecuencias alélicas de alelos selectos de la cohorte de pacientes VIH positivos de este estudio con las reportadas para otras cohortes representativas de población abierta de diferentes partes del mundo de la base de datos Allele Frequencies (<http://www.allelefrequencias.net/test/default1.asp>) (Figura 6.8). Aunque las metodologías de tipificación HLA pueden variar, se seleccionaron poblaciones con métodos de tipificación genética y no serológica para hacer más válida la comparación. Como era esperado, se observó una gran diversidad en frecuencias alélicas entre las diferentes poblaciones comparadas. Importantemente, la cohorte de pacientes VIH positivos mexicanos del presente estudio mostró diferencias notables con respecto a otras poblaciones, las cuales podrían determinar características únicas en la adaptación del VIH a las respuestas inmunes más frecuentes a nivel poblacional en México. Notoriamente, la frecuencia de los alelos B\*39 en la cohorte de pacientes VIH+ mexicanos fue más de tres veces mayor a la observada en poblaciones de otras partes del mundo. De igual manera, los alelos B\*35 presentaron una frecuencia más de dos veces más alta en la cohorte del presente estudio que la observada en otras poblaciones. En contraste, la frecuencia de alelos B\*57 en mexicanos VIH+ fue menor a la observada en las otras poblaciones comparadas. Alelos como A\*03, A\*01, A\*11 tuvieron frecuencias menores en la cohorte VIH+ mexicana que en cohortes de caucásicos norteamericanos; alelos como B\*53, B\*58, Cw\*06, Cw\*02 tuvieron frecuencias menores en la cohorte mexicana que en cohortes africanas de población negra y alelos como A\*24, A\*26, A\*33 B\*40, B\*52, Cw\*03, Cw\*01 presentaron frecuencias menores que en cohortes asiáticas (Figura 6.8).



**Figura 6.8. Características inmunogenéticas únicas en una cohorte de individuos VIH positivos mexicanos.** Las frecuencias alélicas para los genes HLA de clase I A (A) B (B) y C (C) fueron calculadas para una cohorte de 288 pacientes VIH positivos del Centro / Sur de México, utilizando la herramientas de análisis de frecuencias HLA de la Base de Datos de VIH de Los Alamos ([www.hiv.lanl.gov/content/immunology/tools-links.html](http://www.hiv.lanl.gov/content/immunology/tools-links.html)). Los genes HLA fueron tipificados a baja / mediana resolución por SSP-PCR como se indica en la sección de Materiales y Métodos. Las frecuencias alélicas de algunos alelos HLA fueron comparadas con las de cohortes seleccionadas de distintas partes del mundo (<http://www.allelefreqs.net/test/default1.asp>). Los alelos están ordenados de acuerdo a su frecuencia en la población mexicana VIH positiva del presente estudio. Mexico: Cohorte de pacientes VIH+ del presente estudio (n=288); Brasil: Sao Paulo (n=239); Japón: Región Central (n=371); Uganda: Kampala (n=161); Kenya: General (n=144); Italia: Roma (n=100); USA: Caucásicos (n=265).

Con el objetivo de comparar la evolución del VIH mediada por HLA a nivel poblacional en cohortes inmunogenéticamente diferentes se seleccionaron dos cohortes de individuos VIH positivos para las cuales existe información extensa sobre la evolución poblacional del virus y cuya comparación con la cohorte mexicana podría ayudar a identificar características únicas de la evolución del VIH asociadas a distribuciones alélicas HLA diferentes. Así, las frecuencias de alelos HLA de clase I en la cohorte HOMER de 1,045 individuos VIH positivos de British Columbia, Canadá y en la cohorte combinada IHAC de 1,845 individuos VIH positivos de British Columbia, Canadá; Australia Occidental y Estados Unidos, se compararon con las obtenidas en la cohorte de 292 individuos VIH positivos del centro/sur de México (Figura 6.9). Aunque tanto la cohorte HOMER como la IHAC incluían una minoría de individuos autoidentificados como hispanos, se observaron diferencias importantes en las distribuciones alélicas de las tres cohortes que reflejan el mestizaje genético característico de la población mexicana [207, 208].



**Figura 6.9. Diferencias marcadas en las distribuciones de frecuencias alélicas en tres cohortes de individuos VIH positivos.** Las frecuencias poblacionales para los genes HLA de clase I A, B y C fueron comparadas entre la cohorte mexicana del centro/sur de México descrita en este estudio (n=292) (gris oscuro), la cohorte combinada IHAC de individuos VIH positivos de British Columbia, Canadá; Australia Occidental y Estados Unidos (n=1,845) (gris claro) [171, 206] y la cohorte HOMER de individuos VIH positivos de British Columbia, Canadá (n=1,045) (blanco) [170]. \*\*Diferencias significativas ( $q < 0.05$ ) entre ambas cohortes, HOMER y IHAC, y la cohorte mexicana. \*Diferencias significativas ( $q < 0.05$ ) entre la cohorte IHAC exclusivamente y la cohorte mexicana.

Como era esperado, se observaron diferencias significativas entre las frecuencias alélicas de la cohorte mexicana y las cohortes HOMER y IHAC ( $\chi^2=597.41$  y  $782.13$ ,  $p<10^{-88}$  y  $10^{-125}$ , respectivamente). Se observaron frecuencias significativamente más altas en la cohorte mexicana que en las cohortes HOMER y IHAC para los alelos A\*68, B\*35, B\*39, B\*48, B\*52, Cw\*04 y Cw\*08 ( $p<0.005$ ,  $q<0.01$ ), los cuales representan alelos amerindios típicos [207, 208, 214]. Por otro lado se observaron frecuencias significativamente menores en la cohorte mexicana que en las cohortes HOMER y IHAC para los alelos A\*01, A\*03, A\*11, B\*07, B\*08, B\*13, B\*27, B\*44, B\*57, Cw\*05 y Cw\*06 ( $p<0.005$ ,  $q<0.01$ ), los cuales están enriquecidos en poblaciones caucásicas [207, 208, 214] (Figura 6.9). Adicionalmente, los alelos A\*02 y A\*24 tuvieron frecuencias significativamente mayores y los alelos A\*25, B\*15 y Cw\*02 tuvieron frecuencias significativamente menores en la cohorte mexicana sin encontrarse específicamente enriquecidos en poblaciones Amerindias o Caucásicas. Notablemente, la frecuencia del grupo alélico HLA B\*39 resultó ser más de 7 veces más alta en la cohorte mexicana que para las cohortes HOMER y IHAC (Figura 6.9). En conjunto, estos resultados confirman la mezcla característica de genes amerindios y caucásicos en la población mestiza mexicana en una cohorte típica de individuos infectados por VIH de la región Centro/Sur del país y revelan un contexto previamente no caracterizado para el estudio de la evolución asociada a HLA en el VIH a nivel poblacional.

#### **6.2.4 Desequilibrio de ligamiento**

El fenómeno de desequilibrio de ligamiento es altamente frecuente para los loci HLA [133]. La consideración de este fenómeno es importante para el análisis de asociaciones entre la expresión de alelos HLA específicos y la presencia de polimorfismos asociados a escape inmune en la secuencia del VIH, ya que algunas asociaciones pueden observarse con alelos en desequilibrio de ligamiento con el alelo selector (ver más adelante). El fenómeno de desequilibrio de ligamiento también es importante para la definición y distribución de haplotipos HLA en una población determinada. Se encontraron 21 pares de alelos (Tabla 6.2) así como una terna de alelos (Tabla 6.3) en desequilibrio de ligamiento en la cohorte de pacientes VIH positivos mexicanos. Estos resultados demuestran la importancia de considerar el fenómeno de desequilibrio de ligamiento en análisis del impacto de los genes HLA en la evolución del VIH a nivel poblacional en México. Adicionalmente, estos datos muestran la importancia de considerar el fenómeno de desequilibrio de ligamiento en la determinación de frecuencias haplotípicas HLA de cohortes mexicanas sin antecedentes familiares conocidos para una adecuada descripción inmunogenética de la población.



Tabla 6.2. Análisis de desequilibrio de ligamiento para los genes HLA-A, B y C en una población de 292 pacientes VIH positivos mexicanos.

HLA1	HLA2	Valor p	Ambos	Sólo HLA1	Sólo HLA2	Ninguno	P(HLA1)	P(HLA2)	P(HLA1 & HLA2)	E(P(HLA1 & HLA2))
B*35	Cw*04	1,30E-33	78	22	11	177	0.3472	0.3090	0.2708	0.1073
B*39	Cw*07	4,40E-21	97	10	55	126	0.3715	0.5278	0.3368	0.1961
B*15	Cw*01	8,99E-15	25	13	10	240	0.1319	0.1215	0.0868	0.0160
B*40	Cw*03	1,08E-11	29	10	29	220	0.1354	0.2014	0.1007	0.0273
B*14	Cw*08	5,83E-11	22	2	38	226	0.0833	0.2083	0.0764	0.0174
B*48	Cw*08	3,63E-10	21	2	39	226	0.0799	0.2083	0.0729	0.0166
B*44	Cw*16	3,58E-07	16	18	9	245	0.1181	0.0868	0.0556	0.0102
B*40	Cw*07	2,84E-06	3	36	149	100	0.1354	0.5278	0.0104	0.0715
B*38	Cw*12	1,02E-05	8	1	12	267	0.0312	0.0694	0.0278	0.0022
B*44	Cw*05	2,66E-05	14	20	9	245	0.1181	0.0799	0.0486	0.0094
B*18	Cw*05	3,92E-04	9	5	14	260	0.0486	0.0799	0.0312	0.0039
A*03	B*07	4,23E-04	12	9	19	248	0.0729	0.1076	0.0417	0.0078
A*01	B*57	7,22E-04	6	15	0	267	0.0729	0.0208	0.0208	0.0015
B*41	Cw*17	2,54E-02	3	0	2	283	0.0104	0.0174	0.0104	0.0002
B*07	Cw*07	4,64E-02	28	3	124	133	0.1076	0.5278	0.0972	0.0568
A*30	B*39	5,35E-02	0	25	107	156	0.0868	0.3715	0.0000	0.0323
B*52	Cw*03	6,68E-02	13	7	45	223	0.0694	0.2014	0.0451	0.0140
A*33	B*14	6,88E-02	7	5	17	259	0.0417	0.0833	0.0243	0.0035
B*51	Cw*15	9,22E-02	10	20	12	246	0.1042	0.0764	0.0347	0.0080
A*29	Cw*16	1,92E-01	7	6	18	257	0.0451	0.0868	0.0243	0.0039
A*30	B*18	3,63E-01	7	18	7	256	0.0868	0.0486	0.0243	0.0042

Ambos, número de individuos que expresan tanto HLA1 como HLA2. Sólo HLA1, número de individuos que expresan HLA1, pero no HLA2. Sólo HLA2, número de individuos que expresan HLA2, pero no HLA1. Ninguno, Número de individuos que no expresan ni HLA1 ni HLA2. P(HLA1), frecuencia observada de individuos con HLA1 en la población. P(HLA2), frecuencia observada de individuos con HLA2 en la población. P(HLA1 & HLA2), frecuencia observada de individuos tanto con HLA1 como con HLA2 en la población. E(P(HLA1 & HLA2)), frecuencia esperada de P(HLA1 & HLA2) si P(HLA1) y P(HLA2) son independientes. Se muestran valores p de prueba exacta de Fisher de dos colas, estimados de la probabilidad de que P(HLA1) y P(HLA2) sean independientes dada la información observada. Un valor pequeño de p indica que la ocurrencia de HLA1 y HLA2 está ligada en la población. Número de pruebas de dos vías realizado =  $A*B+A*C+B*C = 1190$ . Valor p para comparaciones de dos vías =  $0.05/(A*B+A*C+B*C) = 4.202e-05$  para nivel de confianza de 95%.

Tabla 6.3. Análisis de desequilibrio de ligamiento para los genes HLA-A, B y C en una población de 292 pacientes VIH positivos mexicanos.

HLA1	HLA2	HLA3	Valor p	Todos	Sólo HLA1 & HLA2	Sólo HLA3	Ninguno	P(HLA1 & HLA2)	P(HLA3)	P(HLA1 & HLA2 & HLA3)	E(P(HLA1 & HLA2 & HLA3))
B*07	Cw*07	A*03	4,23E-04	12	19	9	248	0.1076	0.0729	0.0417	0.0078

Todos, número de individuos que expresan HLA1, HLA2 y HLA3. Sólo HLA1 & HLA2, número de individuos que expresan HLA1 y HLA2, pero no HLA3. Sólo HLA3, número de individuos que expresan HLA3, pero no HLA1 ni HLA2. Ninguno, Número de individuos que no expresan HLA1, HLA2 ni HLA3. P(HLA1 & HLA2), frecuencia observada de individuos con HLA1 y HLA2 en la población. P(HLA3), frecuencia observada de individuos con HLA3 en la población. P(HLA1 & HLA2 & HLA3), frecuencia observada de individuos con HLA1, HLA2 y HLA3 en la población. E(P(HLA1 & HLA2 & HLA3)), frecuencia esperada de P(HLA1 & HLA2 & HLA3) si P(HLA1 & HLA2) y P(HLA3) son independientes. Se muestran valores p de prueba exacta de Fisher de dos colas, estimados de la probabilidad de que P(HLA1 & HLA2) y P(HLA3) sean independientes dada la información observada. Un valor pequeño de p indica que la ocurrencia de HLA1, HLA2 y HLA3 está ligada en la población. Número de pruebas de dos vías realizado =  $3*A*B*C = 22344$ . Valor p para comparaciones de dos vías =  $0.05/(3*A*B*C) = 2.238e-06$  para nivel de confianza de 95%.

### 6.2.5 Desviaciones del equilibrio de Hardy-Weinberg

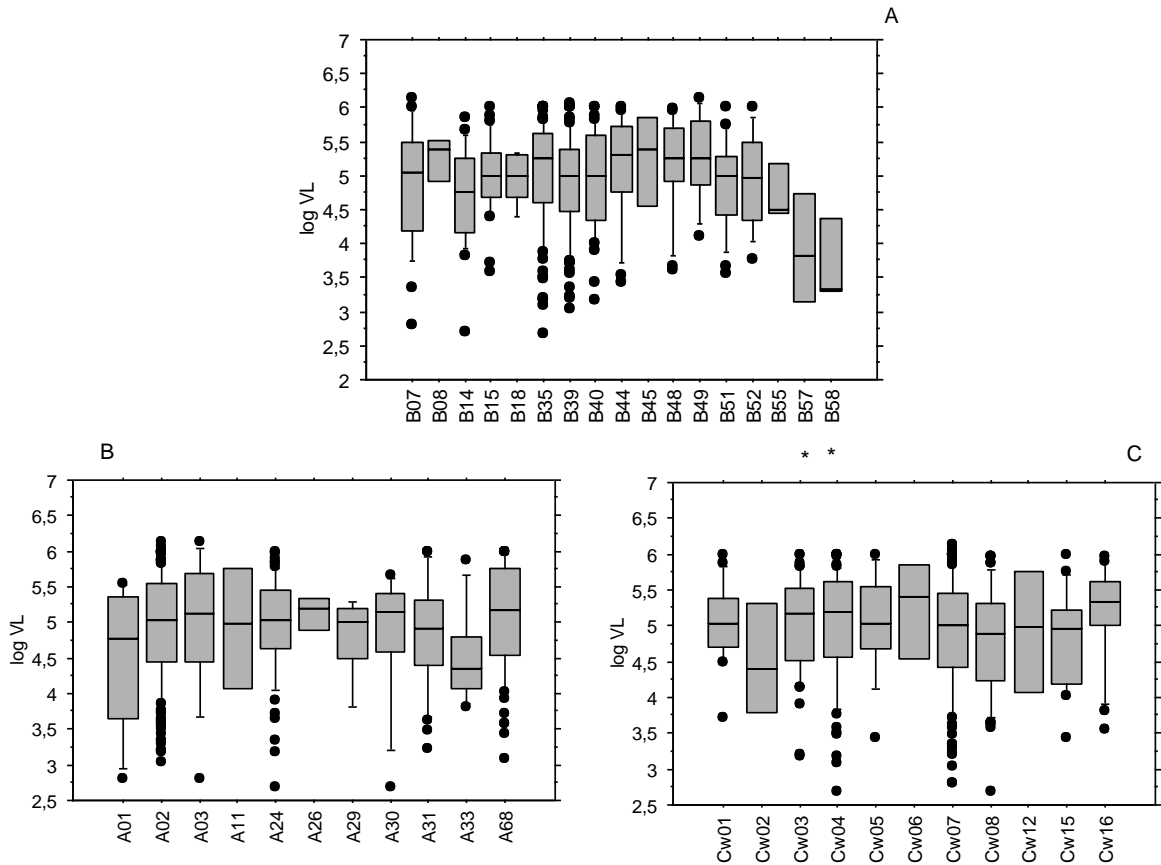
El principio de Hardy-Weinberg establece que, tras una generación de apareamiento al azar, las frecuencias de los genotipos de un locus individual se fijarán en un valor de equilibrio particular, siendo estas frecuencias al equilibrio una función sencilla de las frecuencias alélicas en ese locus. Para analizar si las frecuencias alélicas y genotípicas de los genes HLA-A, B y C en la cohorte de pacientes VIH positivos mexicanos se encontraban en equilibrio al transmitirse los diferentes alelos de generación en generación, se realizó un análisis de desviación del equilibrio de Hardy-Weinberg. Es sabido que eventos como apareamiento no aleatorio (endogamia, emparejamiento selectivo), poblaciones de tamaño pequeño, selección natural, mutaciones y migración, pueden causar cambios en las frecuencias genéticas que se observan a nivel poblacional como desviaciones al equilibrio de Hardy-Weinberg. Se buscó si la cohorte analizada presentaba este tipo de desviaciones comparando las frecuencias genotípicas reales de la población y las frecuencias genotípicas esperadas en caso de existir equilibrio de Hardy-Weinberg, utilizando pruebas exactas de Fisher. No se observaron desviaciones significativas del equilibrio de Hardy-Weinberg para ninguno de los loci HLA (A, B ó C). Esto puede entenderse al no observarse diferencias ( $p > 0.05$ ) entre las frecuencias genotípicas reales y las esperadas según el principio de Hardy-Weinberg para cada locus (Tabla 6.4).

Tabla 6.4. Equilibrio de Hardy-Weinberg\* en los genes HLA A, B y C de la cohorte de pacientes VIH positivos del centro y sur de México.

Locus	Número de Genotipos	Frecuencia Heterocigotos Observada	Frecuencia Heterocigotos Esperada	Valor p
A	287	0.77352	0.79804	0.11002
B	287	0.90244	0.89866	0.05907
C	287	0.82230	0.84003	0.45641

\*Se analizó cada locus (HLA A, B y C) por separado en Arlequín v3.11 mediante pruebas exactas de Fisher. Valores de  $p > 0.05$  indican que no se encontraron diferencias significativas entre las frecuencias observadas y esperadas y que por tanto no se detectaron desviaciones del equilibrio de Hardy-Weinberg.

### 6.3 ASOCIACIONES ENTRE LA PRESENCIA DE ALELOS HLA Y LOS VALORES DE CARGA VIRAL EN LA COHORTE DE PACIENTES MEXICANOS

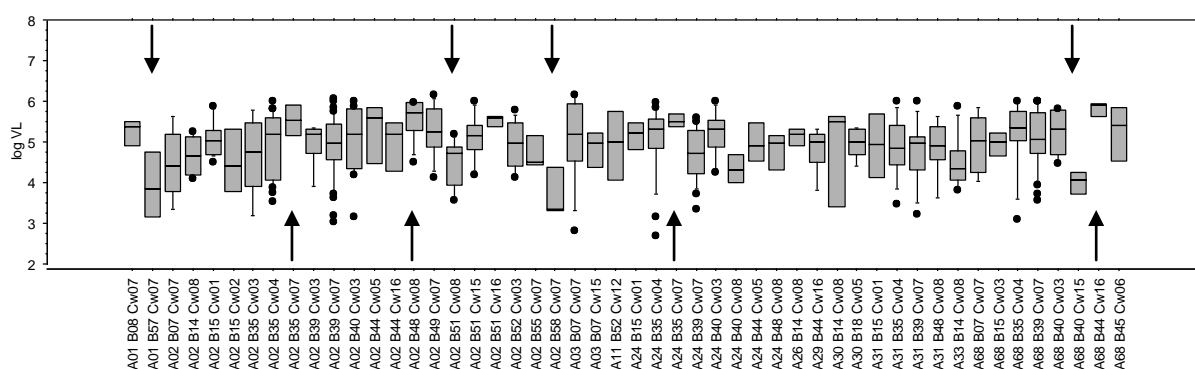


**Figura 6.10. Asociaciones entre la presencia de alelos HLA A (A), B\* (B\*) y Cw\* (C) específicos y los valores de carga viral en plasma en una cohorte de pacientes mexicanos.** Se incluyeron en el análisis exclusivamente los alelos con frecuencias mayores al 0.5% en la cohorte de estudio (n final = 400). Las cajas representan los percentiles 25 a 75 (50% de los datos). Las cajas más las barras lineales representan los percentiles 10 a 90 (80% de los datos). Los puntos representan datos fuera del percentil 10-90 (20% de los datos). La línea central de las cajas corresponde a la mediana de cada categoría. \*Significancia ( $p < 0.05$ ) en prueba ANOVA con prueba posthoc de Diferencias Mínimas Significativas Protegidas (PLSD) de Fisher.

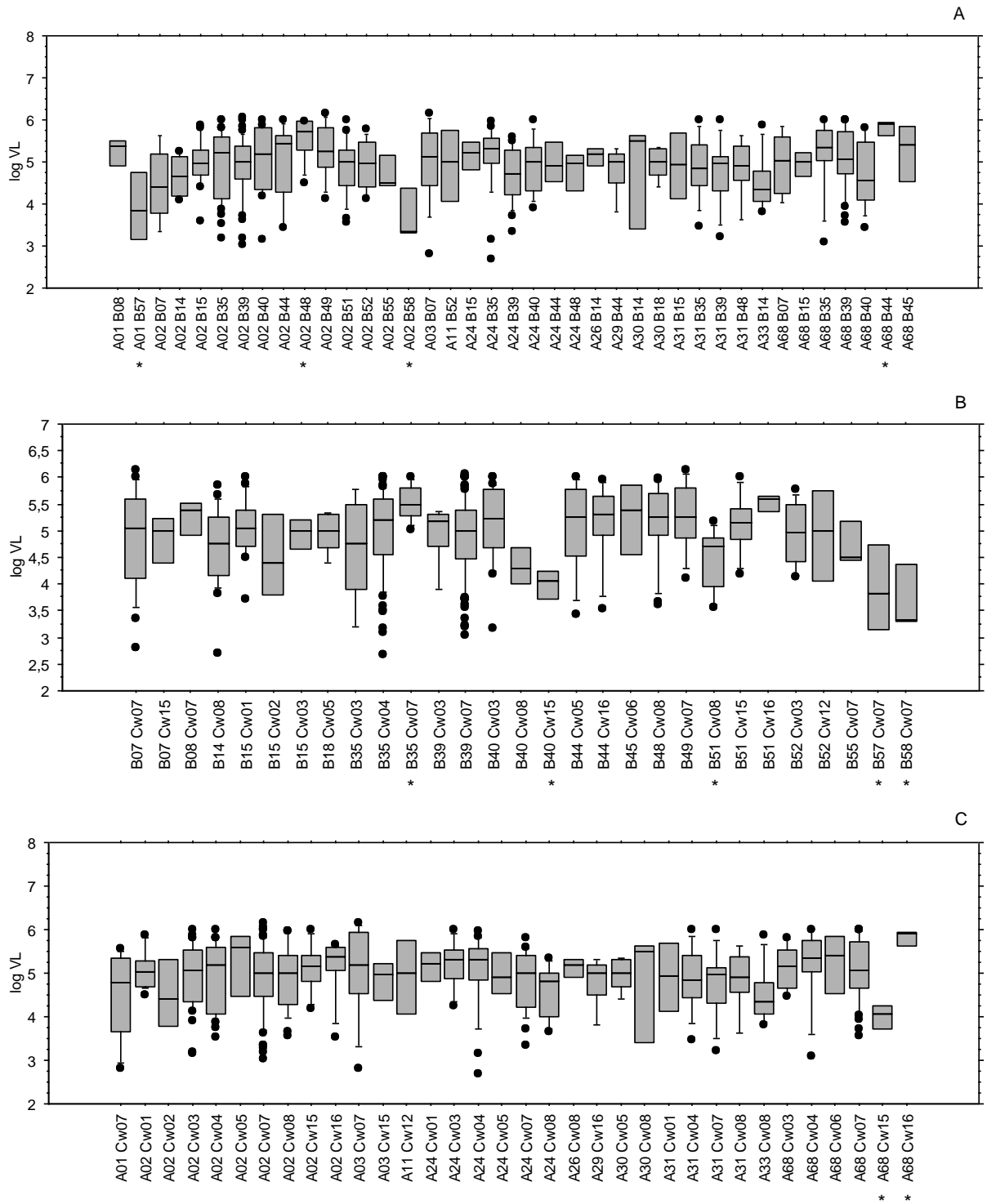
Se buscaron posibles asociaciones entre la expresión de alelos HLA determinados y la presencia de valores altos o bajos de carga viral en la cohorte de pacientes VIH+ mexicanos del presente estudio, con el objetivo de buscar factores de riesgo o protección hacia el avance de la enfermedad por VIH en la población mexicana. Aún cuando la mayoría de los pacientes se encontraba en etapas intermedias o tardías de la infección, se pudieron observar algunas asociaciones importantes que podrían representar características particulares de susceptibilidad en la población mexicana. Al considerar alelos HLA por separado, sólo se observaron asociaciones significativas de la presencia de alelos B\*57 y B\*58 con cargas virales bajas (Figura 6.10). Estas asociaciones son consistentes en prácticamente todas las cohortes estudiadas (revisado en [134]).

Al buscar asociaciones entre valores de carga viral y haplotipos HLA particulares, se observaron haplotipos como A\*01/B\*57, A\*02/B\*58, B\*57/Cw\*07 y B\*58/Cw\*07, que se asociaron a valores bajos de carga viral, reflejando claramente el efecto protector de los

alelos B\*57 y B\*58. Sin embargo, los haplotipos B\*40/Cw\*15, B\*51/Cw\*08 y A\*68/Cw\*15 también se asociaron a valores bajos de carga viral, independientemente del efecto de B\*57 y B\*58. Por el contrario, los haplotipos A\*02/B\*48, A\*68/B\*44, B\*35/Cw\*07 y A\*68/Cw\*16 se asociaron a valores altos de carga viral y podrían representar factores de riesgo para el avance rápido hacia SIDA en la infección por VIH en mexicanos (Figura 6.11). Algunas de estas asociaciones fueron confirmadas al observar efectos protectores o de riesgo para haplotipos de tres genes (Figura 6.12). En particular, A\*68/B\*40/Cw\*15 y A\*02/B\*51/Cw\*08 se asociaron con cargas virales bajas, mientras que A\*68/B\*44/Cw\*16, A\*02/B\*35/Cw\*07, A\*24/B\*35/Cw\*07 y A\*02/B\*48/Cw\*08 se asociaron con cargas virales altas. Asimismo, los haplotipos A\*02/B\*58/Cw\*07 y A\*01/B\*57/Cw\*07 confirmaron el efecto protector de B\*57 y B\*58 al asociarse con cargas virales bajas. Estos resultados en conjunto sugieren que ciertos haplotipos HLA podrían representar un factor de riesgo o de protección hacia el avance a SIDA en la infección por VIH, considerando el contexto de frecuencias HLA característicos de la población mexicana y la posible adaptación del virus a las respuestas citotóxicas frecuentes en nuestra población.



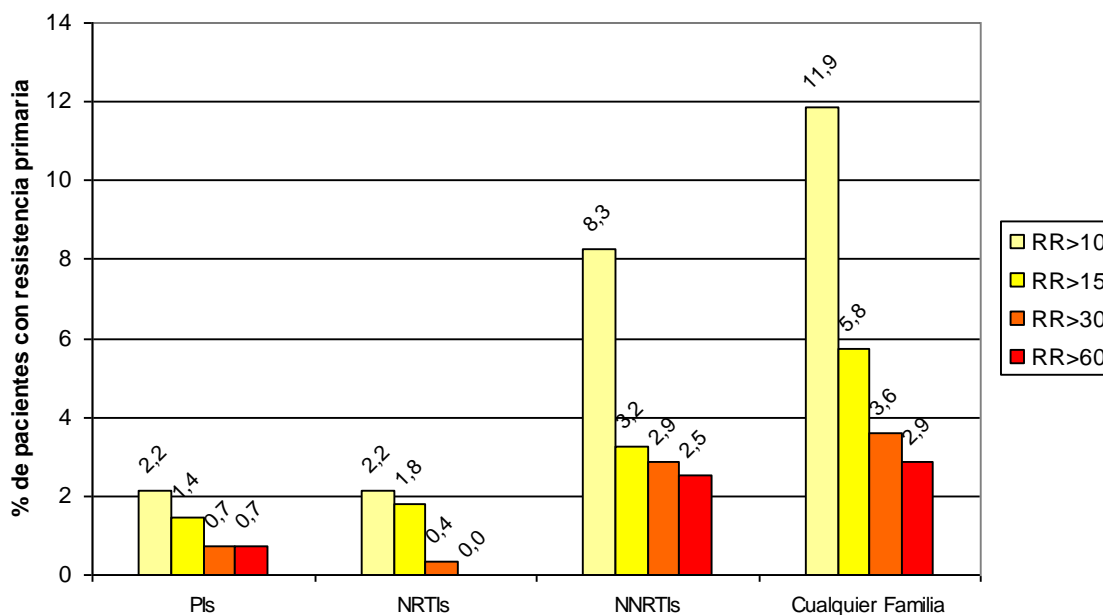
**Figura 6.12. Asociaciones entre la presencia de haplotipos triples HLA A-B-C específicos y los valores de carga viral en plasma en una cohorte de pacientes mexicanos.** Se incluyeron en el análisis exclusivamente los haplotipos con frecuencias mayores al 0.5% en la cohorte de estudio (n final = 400). Las cajas representan los percentiles 25 a 75 (50% de los datos). Las cajas más las barras lineales representan los percentiles 10 a 90 (80% de los datos). Los puntos representan datos fuera del percentil 10-90 (20% de los datos). La línea central de las cajas corresponde a la mediana de cada categoría. Las flechas ascendentes señalan haplotipos asociados a valores de carga viral altos y las flechas descendentes haplotipos asociados a valores de carga viral bajos.



**Figura 6.11. Asociaciones entre la presencia de haplotipos dobles HLA A-B (A), B-C (B) y A-C (C) específicos y los valores de carga viral en plasma en una cohorte de pacientes mexicanos.** Se incluyeron en el análisis exclusivamente los haplotipos con frecuencias mayores al 0.5% en la cohorte de estudio (n final = 400). Las cajas representan los percentiles 25 a 75 (50% de los datos). Las cajas más las barras lineales representan los percentiles 10 a 90 (80% de los datos). Los puntos representan datos fuera del percentil 10-90 (20% de los datos). La línea central de las cajas corresponde a la mediana de cada categoría. \*Significancia ( $p < 0.05$ ) en prueba ANOVA con prueba posthoc de Diferencias Mínimas Significativas Protegidas (PLSD) de Fisher.

## 6.4 ANÁLISIS DE RESISTENCIA PRIMARIA DE VIH EN LA COHORTE DE INDIVIDUOS MEXICANOS

Se realizó un análisis de riesgo relativo de resistencia primaria en 278 personas VIH positivas, vírgenes a tratamiento, del Centro / Sur de México, con secuencias disponibles del gen *pol*/ viral (incluyendo toda la proteasa y dos terceras partes de la transcriptasa inversa), de acuerdo a los algoritmos de la base de Datos de Resistencia de la Universidad de Stanford (<http://hivdb.stanford.edu/pages/algs/HIVdb.html>), como se explica en la sección de Métodos. Se observó que en la cohorte del presente estudio, un 5.8% de los pacientes presentaron riesgo bajo, intermedio o alto de resistencia primaria ( $RR \geq 15$ ) hacia al menos una familia de fármacos ARVs (Figura 6.13). La frecuencia de pacientes con riesgo de resistencia primaria intermedio-alto ( $RR \geq 30$ ) fue de 3.6% y la frecuencia de pacientes con riesgo de resistencia primaria alto ( $RR \geq 60$ ) fue de 2.9% (Figura 6.13).

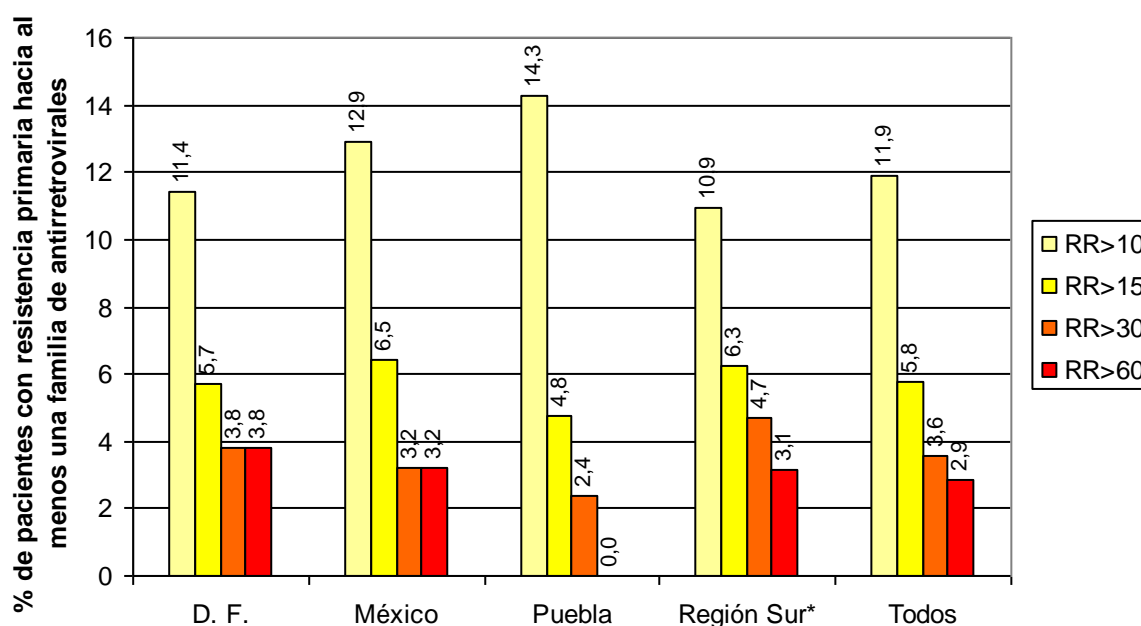


**Figura 6.13. Resistencia primaria en una cohorte de pacientes del Centro y Sur de México por familias de fármacos antirretrovirales.** Se muestran frecuencias acumuladas de pacientes con diversos grados de riesgo relativo (RR) de presentar resistencia, de acuerdo a las categorías definidas en la Base de Datos de Resistencia de la Universidad de Stanford (ver interpretación en el texto). Los puntajes de RR se obtuvieron con el programa HIVdb de la Base de Datos de Resistencia de la Universidad de Stanford a partir de 278 secuencias abarcando todo el gen de la proteasa y dos terceras partes del gen de la RT. PIs – Inhibidores de proteasa, NRTIs – Inhibidores de RT análogos de nucleósidos, NNRTIs – Inhibidores de RT no análogos de nucleósidos.

La familia de fármacos ARVs con mayor riesgo de resistencia fue la de inhibidores de la RT no análogos de nucleósidos (NNRTIs) con 3.2% de los pacientes con riesgo (bajo, intermedio o alto) de resistencia y 2.5% de los pacientes con alto riesgo de resistencia (Figura 6.13). Un 1.8% de los pacientes presentaron riesgo de resistencia a inhibidores de la RT análogos de nucleósidos/nucleótidos (NRTIs) y un 1.4% hacia inhibidores de la proteasa

(PIs). Sin embargo, no se observaron virus con niveles altos de resistencia hacia NRTIs, contrastando con 0.7% de los aislados presentando alto riesgo de resistencia hacia PIs (Figura 6.13).

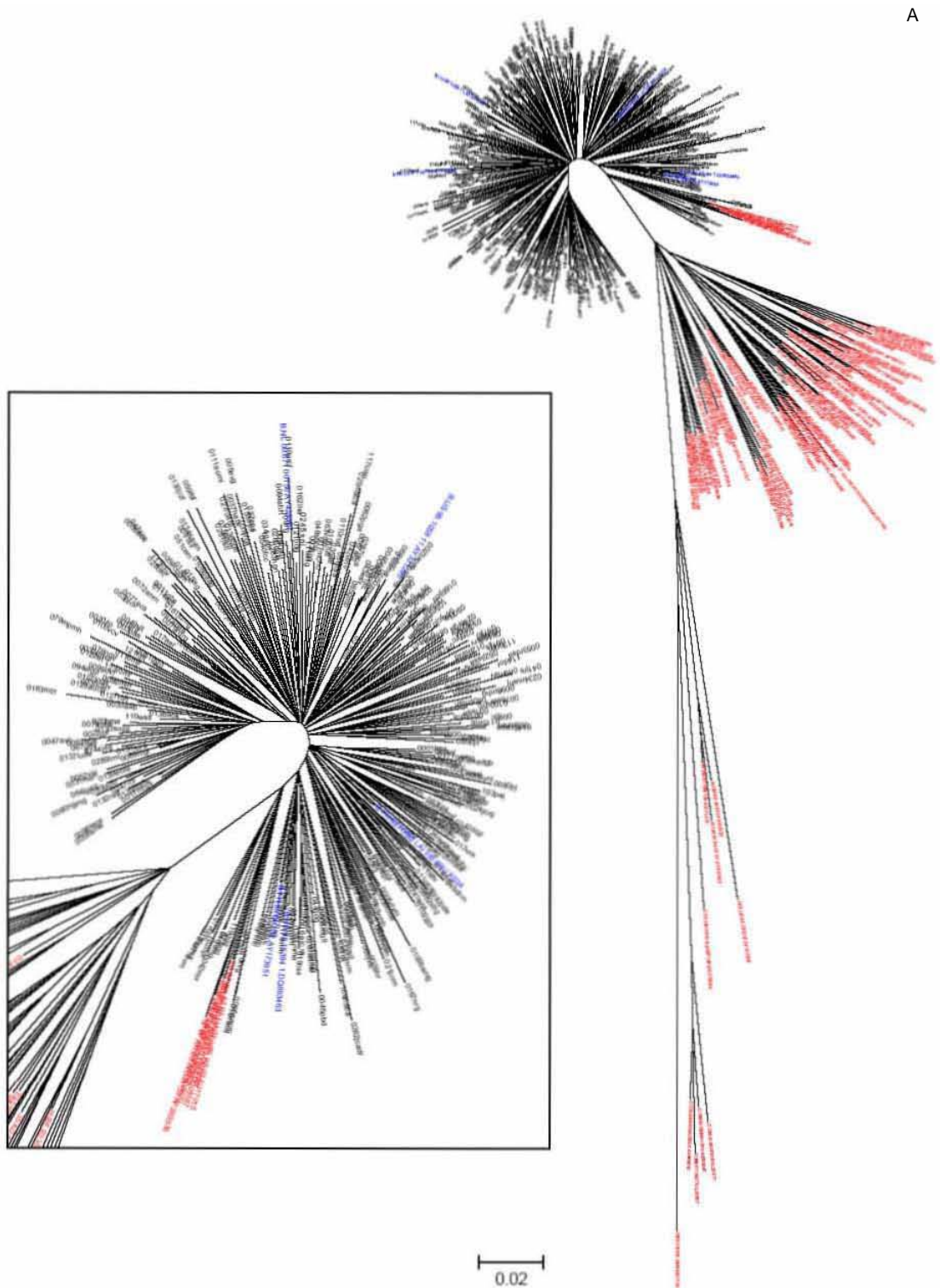
En general, no se observaron diferencias importantes entre los niveles de resistencia primaria en las diferentes regiones geográficas estudiadas. Las frecuencias de resistencia primaria en el D. F., Estado de México y Puebla fueron semejantes entre sí y semejantes al promedio de toda la cohorte (Figura 6.14). Las prevalencias de resistencia primaria en el Centro del país fueron similares también a las observadas en la Región Sur (Chiapas, Guerrero y Oaxaca) (Figura 6.14). Sin embargo, considerando por separado los estados de Chiapas y Guerrero, los niveles de resistencia tendieron a ser más altos (20% de pacientes con riesgo de resistencia primaria en ambos casos) (datos no mostrados), aunque estas observaciones podrían ser producto de un número pequeño de pacientes representando estos estados en la cohorte.



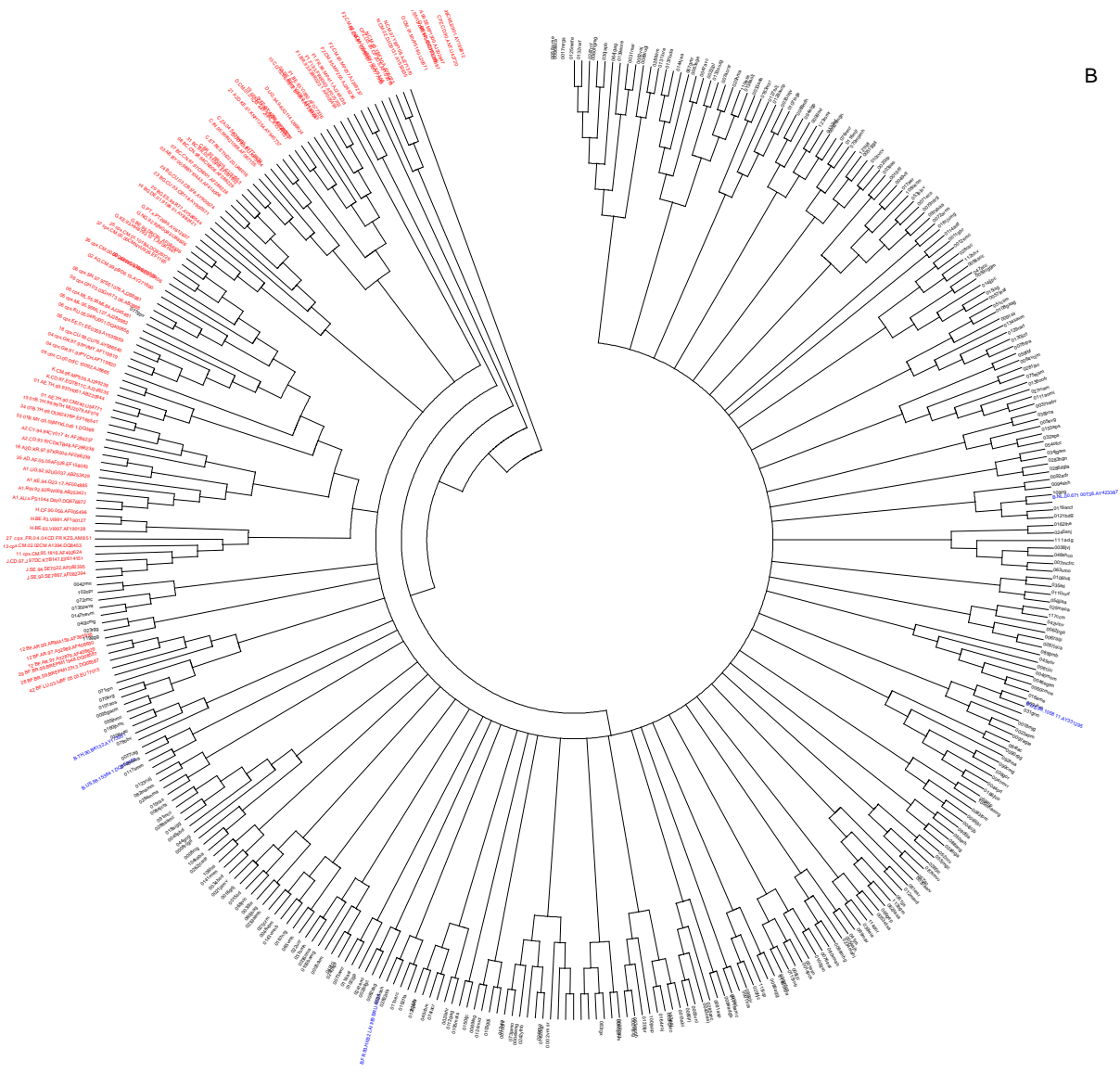
**Figura 6.14. Resistencia primaria en una cohorte de pacientes del Centro y Sur de México por zona geográfica.** Se muestran frecuencias acumuladas de pacientes con diversos grados de riesgo relativo (RR) de presentar resistencia, de acuerdo a las categorías definidas en la Base de Datos de Resistencia de la Universidad de Stanford (ver interpretación en el texto). Los puntajes de RR se obtuvieron con el programa HIVdb de la Base de Datos de Resistencia de la Universidad de Stanford a partir de 278 secuencias abarcando todo el gen de la proteasa y dos terceras partes del gen de la RT (D. F., n=105; México, n=62; Puebla, n=42; Región Sur, n=64; otros, n=5). \*Región Sur incluye secuencias de Chiapas, Guerrero y Oaxaca.

## 6.5 RELACIONES FILOGENÉTICAS ENTRE SECUENCIAS DE VIH CIRCULANTES EN EL CENTRO / SUR DE MÉXICO

Para estudiar el grado de homogeneidad y las relaciones filogenéticas existentes entre las secuencias virales de los pacientes de la cohorte mexicana, se construyó un árbol filogenético utilizando el método de neighbor-joining (NJ) y el modelo de dos parámetros de Kimura (Figura 6.15). Se incluyeron secuencias consenso de los principales subtipos y formas recombinantes de VIH en la construcción del árbol.







**Figura 6.15. Relaciones filogenéticas entre VIH circulante en el centro y sur de México.** Se muestran árboles filogenéticos de distancias construidos con 278 secuencias de 1305 pb del gen *pol* viral abarcando todo el gen de la proteasa y dos terceras partes de la RT. Se empleó el programa MEGA 4, utilizando el método de Neighbor-Joinig. Se muestra el árbol consenso inferido a partir de 1000 réplicas de bootstrap en formato radial (A) y circular (B). Se incluyeron 93 secuencias de referencia de la Base de Datos de Secuencias de Los Alamos, abarcando los principales subtipos y formas recombinantes del VIH-1. Las secuencias mexicanas se muestran en negro, las secuencias de referencia del subtipo B se muestran en color azul y las secuencias de referencia de otros subtipos y formas recombinantes se muestran en color rojo. El inserto en A muestra un detalle del subgrupo de secuencias mexicanas. Las ramas correspondientes a particiones reproducidas en menos del 50% de las réplicas de bootstrap están colapsadas. Las distancias evolutivas fueron calculadas con el modelo de dos parámetros de Kimura y se encuentran en unidades de número esperado de sustituciones por sitio. Todas las posiciones con información faltante fueron eliminadas por comparación pareada de secuencias (pairwise comparison).

Se observó una agrupación clara de las secuencias mexicanas en un grupo bien diferenciado, junto con las secuencias de referencia del consenso B. Aunque las secuencias consenso de las formas recombinantes BF se insertaron dentro del cluster de

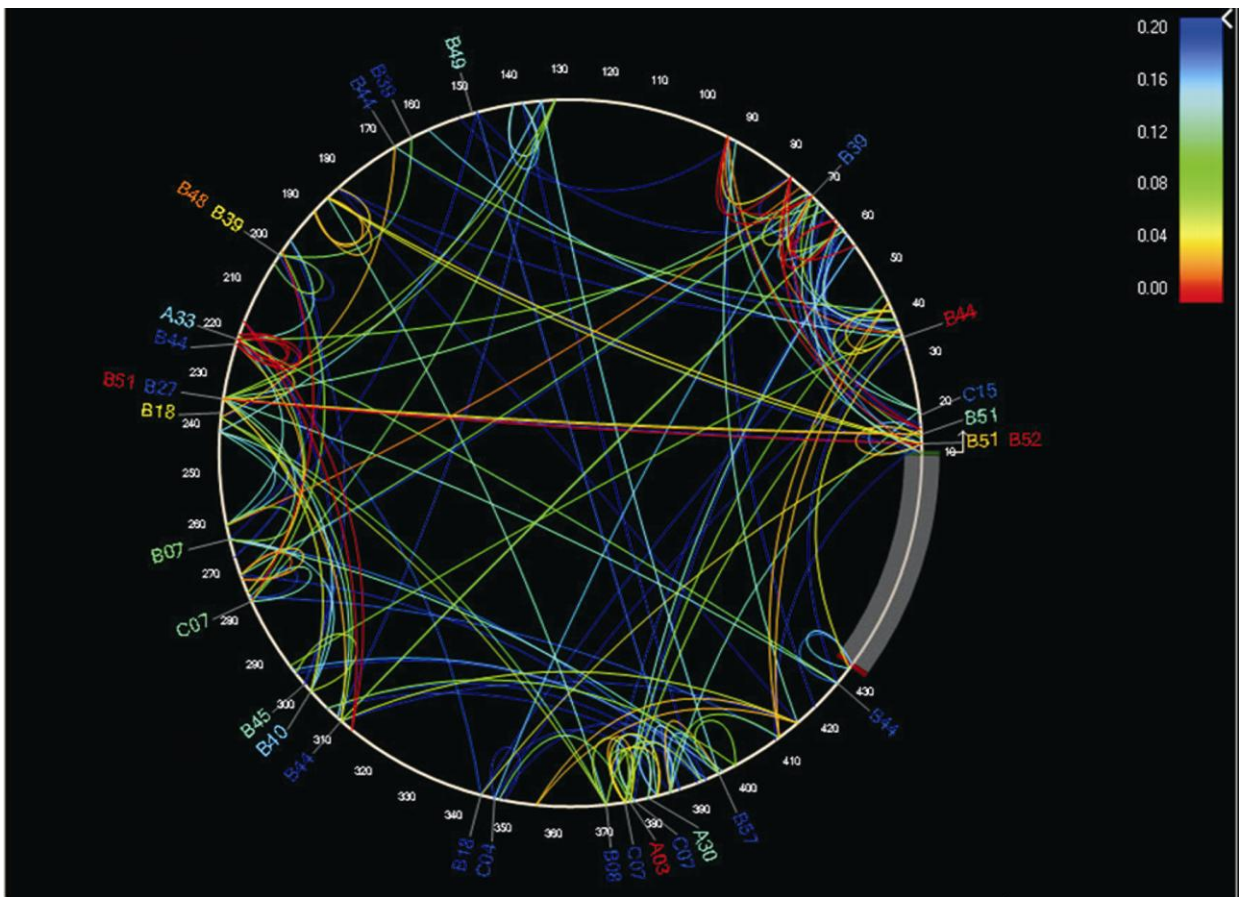
secuencias mexicanas, un análisis de subtipificación posterior en el programa REGA (datos no mostrados) de todas las secuencias utilizadas en el estudio permitió confirmar que un 99.6% de las secuencias en la corte eran de subtipo B. Un solo virus resultó ser una forma recombinante compleja (CRF06-cpx), agrupándose con el consenso de estas formas recombinantes en el árbol filogenético (Figura 6.15).

El hecho de que las secuencias mexicanas no se agruparon en clusters definidos por zonas de procedencia o por líneas de diversificación, sugiere una alta homogeneidad en los virus circulantes en la región centro y sur del país. Esta alta homogeneidad se apoya con un inesperadamente alto predominio de virus del subtipo B en comparación con otros países de norteamérica, aún con los altos flujos migratorios característicos de la población mexicana.

## **6.6 EVOLUCIÓN DEL VIH MEDIADA POR LOS GENES HLA DE CLASE I EN UNA COHORTE DE INDIVIDUOS MEXICANOS**

La evolución del VIH mediada por selección por HLA a nivel poblacional se estudió en un fragmento de 433 codones cubriendo toda la proteasa (99 aminoácidos) y 334 aminoácidos de la transcriptasa inversa en 280 individuos con infección crónica de la cohorte mexicana. El modelo de redes de dependencia filogenética (PDN) por Carlson y colaboradores [183], actualmente uno de los modelos más completos para evaluar la evolución de VIH mediada por HLA, se aplicó para inferir patrones de escape de CTLs y covariación de codones en el VIH en la cohorte mexicana. Nuestros resultados se compararon con los obtenidos previamente al aplicar el modelo PDN a la bien caracterizada cohorte multicéntrica IHAC, compuesta de individuos provenientes de British Columbia, Canadá; Australia Occidental y Estados Unidos. La corte IHAC y la cohorte mexicana presentaron contextos inmunogenéticos claramente diferentes al comparar sus distribuciones de frecuencias HLA (Figura 6.9). Ambas cohortes estaban infectadas mayoritariamente por el clado B viral [171, 183, 206] (Figura 6.15). El modelo PDN se utilizó para identificar asociaciones HLA – polimorfismo de VIH, así como asociaciones polimorfismo de VIH – polimorfismo de VIH significativas, usando un umbral de valor  $q$  de 0.2. Debido a que el modelo PDN se basa en un modelo multivariado en el que varios atributos predictores (i. e. la presencia o ausencia de un alelo HLA particular o un aminoácido en un codón de VIH específico) pueden asociarse con la presencia o ausencia de un aminoácido específico en un codón de VIH blanco, asociaciones indirectas explicadas por la existencia del fenómeno de desequilibrio de ligamiento entre genes HLA y por covariación de codones de VIH se minimizaron.

Se identificaron un total de 43 asociaciones HLA – codón de VIH y 251 asociaciones codón de VIH – codón de VIH, representando 30 diferentes pares HLA – codón de VIH y 135 pares codón de VIH – codón de VIH (Tabla 6.5). Esta red de asociaciones se representó de manera gráfica con el visualizador del modelo PDN, PhyloDv (<http://www.codeplex.com/MSCompBio>) [183] (Figura 6.16), mostrando la secuencia de aminoácidos del VIH como un círculo, con líneas uniendo los HLA asociados a codones de VIH fuera del círculo y arcos uniendo codones de VIH covariantes en el interior del círculo. Incluso con un número relativamente pequeño de individuos en la cohorte, se observó una red de asociaciones densa a un valor de  $q < 0.2$  que revela patrones característicos de covariación de codones de VIH y sustituciones mediadas por HLA en la cohorte mexicana.



**Figura 6.16. Red de dependencia filogenética para la proteasa y la transcriptasa inversa de 280 individuos del centro/sur de México.** Se obtuvo un mapa de la red filogenética de dependencia con el software PhyloDv (<http://www.codeplex.com/MSCompBio>). Las posiciones de Pol están dibujadas sobre el círculo en sentido contrario a las manecillas del reloj, con el extremo N-terminal de la proteasa a las 3 y el primer codón de la transcriptasa inversa correspondiendo a la posición 100. Las líneas indican asociaciones entre codones (dentro del círculo) o entre alelos HLA y codones de VIH (fuera del círculo). Los colores indican los valores  $q$  de las asociaciones con más alta significancia entre los dos atributos. Se incluyeron asociaciones con  $q < 0.02$ .

Se encontraron asociaciones con alelos HLA en 6.1% de los codones de la proteasa y 7.1% de los codones de la transcriptasa inversa. Como se ha descrito previamente para Gag [183], los codones covariantes fueron más comunes dentro de una subproteína (75.6% en total, 20% en la proteasa y 55.6% en la transcriptasa inversa) que entre subproteínas (24.4%) ( $p < 0.001$ ). 28 de los 135 (20.7%) pares de codones de VIH covariantes se presentaron dentro de 10 posiciones uno del otro, sugiriendo una alta proximidad en una proporción importante de mutaciones compensatorias, o bien la presencia de HLAs con varios epítomos blanco en una región de la proteína viral. Notablemente, 46.7% de las asociaciones HLA – codón de VIH predijeron sustituciones en otros codones, sugiriendo vías de escape mediadas por HLA complejas.

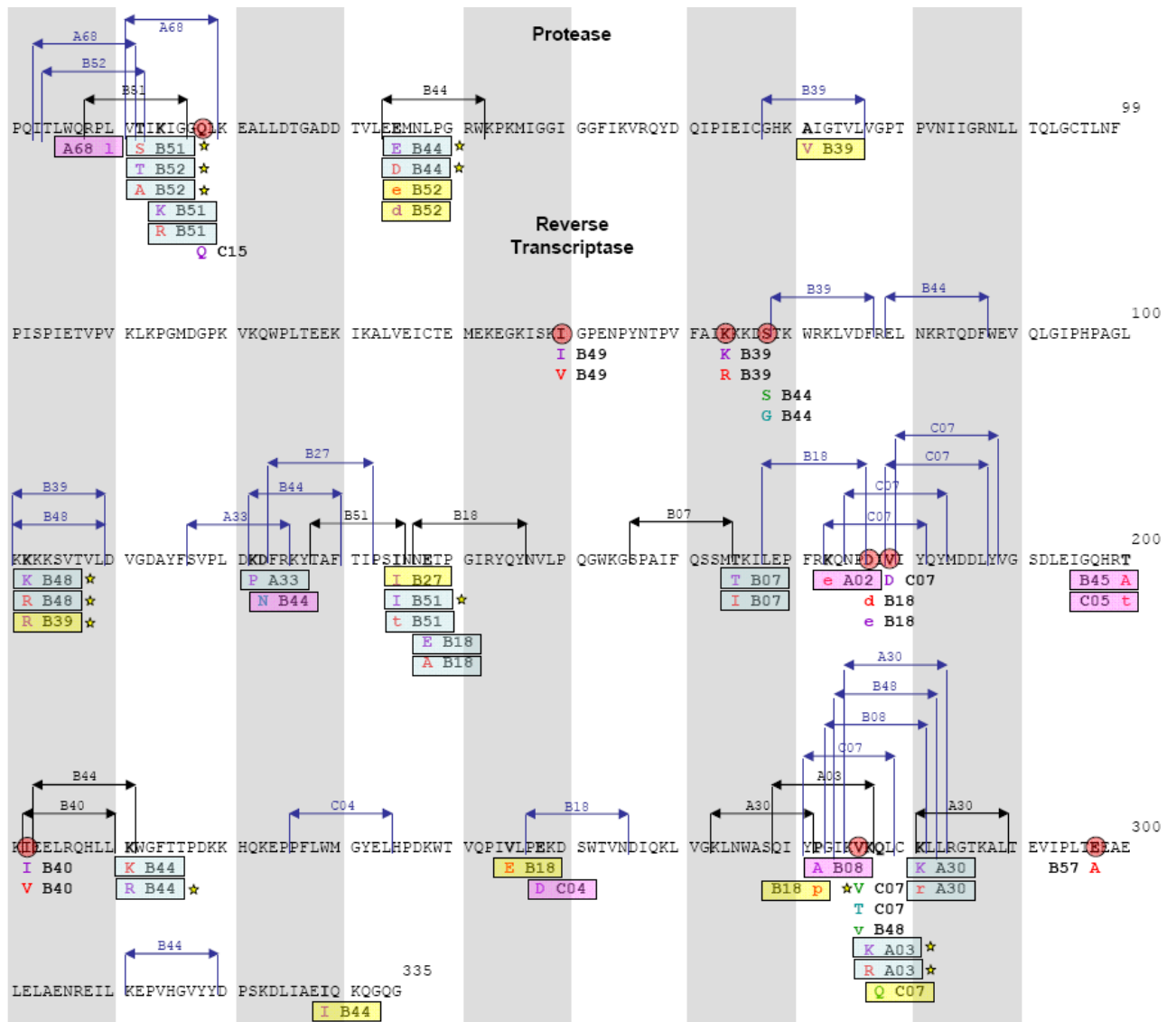
Interesantemente, sólo se observaron dos sitios de pol previamente asociados a resistencia a fármacos ARVs que también se encontraban asociados a presión selectiva por HLA. B\*18 se asoció con un cambio de E a A en la posición 138 de la transcriptasa inversa. El polimorfismo 138A se asocia con una respuesta disminuida a los NNRTIs, incluyendo etravirina (Base de Datos de Resistencia de VIH de la Universidad de Stanford, <http://hivdb.stanford.edu/>). De manera similar, Cw\*07 se asoció con menor probabilidad de presentar D y una tendencia a la conservación de V en la posición 179 de la RT. El polimorfismo 179D se asocia a bajo nivel de resistencia a NNRTIs (Base de Datos de Resistencia de VIH de la Universidad de Stanford, <http://hivdb.stanford.edu/>). Estas observaciones muestran que la evolución mediada por HLA en el VIH puede influenciar la resistencia a fármacos ARVs tanto promoviendo como previniendo la presencia de polimorfismos asociados a resistencia. Este fenómeno de presión selectiva dual ha sido descrito previamente [173, 176, 215], sin embargo, su frecuencia e impacto a nivel poblacional en la cohorte mexicana tendrán que evaluarse en estudios futuro

Tabla 6.5. Asociaciones HLA – codón de VIH en los genes de la proteasa y la transcriptasa inversa en una cohorte de 280 individuos del centro/sur de México§.

Variable Predictora (HLA)	Variable Blanco (Codón VIH)	VV	VF	FV	FF	Valor p	Valor q	Variables Condicionantes (Codones en coevolución. HLA)	Referencia	Observaciones
A*03♦	RT 277R	19	2	56	195	1.20E-10	1.82E-06		[170]. [169]. [171]	
A*03♣	RT 277K	2	19	195	56	1.20E-10	1.82E-06		[170]. [169]. [171]	
A*30♣	RT 281K	18	5	251	3	0.00073383	0.13153195		[171]	
A*33♣	RT 122P	1	11	67	191	0.0009406	0.14940166	RT 98A♥	[171]*	Consistente (Escape de P en Mex, atracción a E en IHAC)
B*07♦	RT 165I	7	18	10	241	0.0005105	0.1175217		[170]. [171]	
B*07♣	RT 165T	18	7	240	10	0.00052215	0.1175217		[170]. [171]	
B*08♣	RT 272A	1	7	122	143	0.00174628	0.19533002		[169]*. [171]*	Diferente (Escape de P por B*42 en IHAC. B*08 frecuencia mucho menor en Mex)
B*18♦	RT 138A	3	10	1	266	0.00013697	0.05164684		[169]. [171]	
B*18♣	RT 138E	10	3	265	1	0.00013862	0.05164684		[169]. [171]	
B*18♦	RT 245E	4	8	22	240	0.00144187	0.18595855	PR 45K♥. B58♦	[170]*. [171]*	Diferente (Escaé de V a L por B*57/58. Escape de E por B*44 en IHAC; no hay asociación con B*18 con p<0.05 en IHAC; no hay diferencia en frecuencia HLA)
B*27♣	RT 135I	1	6	153	94	0.00140863	0.1832374	B*51♣. RT 202I♣. PR 12N♣	[170]*. [169]*. [171]*	Diferente (Escape de I a T/M por B*51/52 en IHAC; B*27 menor frecuente en Mex)
B*39♣	PR 71V	6	85	23	140	0.00121556	0.17636733	PR 93L♣. PR 57K♦	[169]*. [171]*	Diferente (Escape de T por B*53 en IHAC; B*39 mucho más frecuente en Mex)
B*39♣	RT 102R	0	96	17	153	0.00011626	0.04678005	B*48♦. RT 64K♣	[170]*. [171]*	Diferente (K a R por B*48; B*39 mucho más frecuente en Mex)
B*39♦	RT 64R	5	91	4	166	0.00166664	0.1905207	RT 102R♣	NR	No hay asociaciones parecidas en IHAC
B*39♣	RT 64K	91	5	166	4	0.00166664	0.1905207	RT 102R♥	NR	No hay asociaciones parecidas en IHAC
B*40♦	RT 202V	5	27	10	186	0.00100069	0.15487079	RT 135I♣. RT 197E♦. PR 35E♣. RT 334L♣. RT 165I♣. RT 122P♦	NR	No hay asociaciones parecidas en IHAC; no hay diferencia en frecuencia HLA entre cohortes.
B*40♣	RT 202I	27	5	186	10	0.00100069	0.15487079	RT 135I♦. RT 197E♣. PR 35E♥. RT 334L♦. RT 165I♥. RT 122P♣	NR	No hay asociaciones parecidas en IHAC; no hay diferencia en frecuencia HLA entre cohortes.
B*44♦	PR 35D	21	10	33	210	2.78E-09	1.40E-05		[170]. [169]. [171]	
B*44♣	PR 35E	9	22	210	33	6.65E-10	6.69E-06		[170]. [169]. [171]	
B*44♦	RT 211K	22	6	106	115	0.00185231	0.19533002	PR 43K♥. RT 196E♣	[170]*. [171]*	Consistente (R a K en Mex, sólo escape de R en IHAC)
B*44♣	RT 211R	2	28	94	128	4.19E-05	0.02767678		[170]. [171]	
B*44♣	RT 329I	14	15	203	30	0.00130602	0.18323349	RT 334L♦	[171]*	Diferente (Escape de V por B*53 en IHAC; B*44 menor frecuente en Mex)

B*44♠	RT 123N	7	25	3	236	0.00152115	0.18684048	RT 122E♠	[170]*. [169]*	Diferente (D a E por B*35 D enIHAC)
B*44♥	RT 68S	26	7	232	7	0.00185709	0.19533002	PR 41R♥. RT 139T♠	NR	No hay asociaciones parecidas en IHAC
B*44♠	RT 68G	7	26	7	232	0.00185709	0.19533002	PR 41R♠. RT 139T♥	NR	No hay asociaciones parecidas en IHAC
B*45♦	RT 200A	5	3	37	220	0.00072663	0.13153195		[170]*. [171]*	Diferente (Atracción a I por B*40/41, Escape de A por B*08, Escape de T por B*37 en IHAC; no hay diferencia en frecuencia de B*45)
B*48♦	RT 102R	10	8	7	249	2.54E-09	1.40E-05		[170]. [171]	
B*48♠	RT 102K	8	10	227	28	2.83E-05	0.02083046		[170]. [171]	
B*49♦	RT 50V	3	9	3	265	0.00078914	0.13651316		NR	No hay asociaciones parecidas en IHAC; no hay diferencia en frecuencia HLA entre cohortes.
B*49♠	RT 50I	9	3	265	3	0.00078914	0.13651316		NR	No hay asociaciones parecidas en IHAC; no hay diferencia en frecuencia HLA entre cohortes
B*51♦	PR 12S	5	23	2	249	7.70E-05	0.03639581		[170]*. [171]*	Consistente (escape de T, atracción débil a S (p=0.02) en IHAC)
B*51♦	PR 14R	10	13	20	212	0.00064885	0.13057972	RT 135I♠	[170]. [171]*	
B*51♠	PR 14K	12	10	212	20	0.00052685	0.1175217	RT 135I♦	[170]. [171]*	
B*51♠	RT 135I	4	20	155	84	6.58E-08	0.00016543		[170]. [169]*. [171]	
B*52♦	PR 12A	7	8	9	247	2.44E-06	0.00319527		[170]. [171]*	
B*52♠	PR 12T	4	13	215	37	9.21E-08	0.00019546		[170]. [171]*	
B*57♦	RT 297A	4	3	33	239	0.0014485	0.18601806		NR	No hay asociaciones parecidas en IHAC (B*58 con atracción a R con p=0.03, q=0.88)
C*04♠	RT 248D	0	88	11	180	0.00160017	0.18790507		[169]*. [171]*	Diferente (Escape de E por Cw*02, escape de D por B*15 en IHAC; B*15 parece estar en ligero desequilibrio de ligamiento negative con Cw*04)
C*07♥	RT 276V	111	32	116	8	7.21E-05	0.03639581		NR	Atracción a V casi vista como asociación con Cw*08 en IHAC (q=0.22); no hay diferencia en frecuencia HLA entre cohortes
C*07♠	RT 276T	11	131	1	130	0.00152896	0.18684048	RT 272P♠	NR	Atracción a V casi vista como asociación con Cw*08 en IHAC (q=0.22); no hay diferencia en frecuencia HLA entre cohortes
C*07♠	RT 179D	0	147	8	125	0.00061063	0.12536221		NR	No hay asociaciones parecidas en IHAC; Cw*08 débilmente asociado con escape de V (p=0.006; q=0.42)
C*07♥	RT 278Q	137	8	129	1	0.00130929	0.18323349	RT 281R♠	[169]*. [171]*	Diferente (Atracción a V por Cw*12 cerca, q=0.22)
C*15♠	PR 18Q	18	4	253	4	0.00118543	0.17492044		NR	No hay asociaciones parecidas con p<0.05 en IHAC; no hay diferencia en frecuencia HLA entre cohortes.

§Asociaciones con q<0.2. NR – No reportado; \*el codón de VIH se ha reportado previamente como bajo presión por HLA, pero la especificidad HLA y/o el aminoácido blanco son diferentes a los observados en el presente estudio; VV. VF. FV. FF: tablas de contingencia de 2x2, el HLA es el primer valor, el aminoácido es el segundo valor; ♣ escape (tener el predictor hace menos probable tener el blanco), ♦ atracción (tener el predictor hace más probable tener el blanco), ♥ reversión (no tener el predictor hace más probable tener el blanco), ♠repulsión (no tener el predictor hace menos probable tener el blanco).



**Figura 6.17. Mapa de epítomos para comprobar asociaciones HLA – polimorfismo de VIH obtenidas por el modelo de redes de dependencia filogenético en la cohorte mexicana.** Se buscaron asociaciones HLA – polimorfismo de VIH en secuencias de *pol* derivadas de 280 pacientes del centro/sur de México utilizando el modelo de PDN previamente descrito [183], que corrige para los efectos de confusión de la filogenia del VIH, covariación de codones de VIH y desequilibrio de ligamiento de los alelos HLA. Las asociaciones significativas ( $q < 0.2$ ) fueron localizadas en un mapa de epítomos. Las asociaciones fuertemente significativas ( $q < 0.05$ ) están marcadas con estrellas amarillas. Los epítomos de CTLs definidos experimentalmente se muestran en negro ([www.hiv.lanl.gov/content/immunology/index.html](http://www.hiv.lanl.gov/content/immunology/index.html)), los epítomos predichos por motivos de unión de péptido a las moléculas HLA se muestran en azul (Herramienta Motif Scan de la Base de Datos de VIH de Los Alamos, [www.hiv.lanl.gov/content/immunology/tools-links.html](http://www.hiv.lanl.gov/content/immunology/tools-links.html)). Las posiciones blanco se muestran con negritas, incluyendo los alelos HLA predictores y los aminoácidos asociados en diferentes colores dependiendo de las distribuciones hoja de las asociaciones determinadas con el modelo PDN, i. e. escape (morado, tener el HLA predictor hace menos probable tener el aminoácido blanco), atracción (rojo, tener el HLA predictor hace más probable tener el aminoácido blanco), reversión (verde, no tener el HLA predictor hace más probable tener el aminoácido blanco), repulsión (azul claro, no tener el HLA predictor hace menos probable tener el aminoácido blanco). Los aminoácidos en minúscula indican asociaciones obtenidas sin tomar en cuenta la covariación de codones. Las asociaciones HLA – polimorfismo de VIH en la cohorte mexicana fueron comparadas con las observadas en la cohorte combinada IHAC. Las asociaciones compartidas en ambas cohortes se muestran en rectángulos azules. Las asociaciones con posiciones de VIH y aminoácidos compartidos en las dos cohortes pero con diferentes HLA predictores se muestran en rectángulos amarillos. Las asociaciones con posiciones de VIH compartidas en las dos cohortes pero con diferentes aminoácidos blanco y diferentes HLA predictores se muestran en rectángulos rosas. Las posiciones de VIH asociadas a HLA exclusivamente en la cohorte mexicana se marcaron con círculos rojos.

Las asociaciones significativas HLA – codón de VIH ( $q < 0.02$ ) fueron presentadas en un mapa de epítomos para confirmar la validez de las asociaciones (Figura 6.17). 10 pares HLA – codón de VIH pudieron ser explicados por epítomos confirmados experimentalmente, de los cuales 5 han sido definidos de manera óptima (Base de Datos de VIH de Los Alamos, [www.hiv.lanl.gov/content/immunology/index.html](http://www.hiv.lanl.gov/content/immunology/index.html)). 12 pares HLA – codón de VIH adicionales pudieron ser confirmados por predicción de epítomos por motivos de unión de péptidos (Herramienta “Motif Scan” de la Base de Datos de VIH de Los Alamos, [www.hiv.lanl.gov/content/immunology/tools-links.html](http://www.hiv.lanl.gov/content/immunology/tools-links.html)). 8 pares HLA – codón de VIH no pudieron ser explicados por mapeo de epítomos, posiblemente debido a la falta de información sobre motivos de unión de péptido para los HLAs asociados (e.g. B\*39 y B\*49), o debido a la presencia de asociaciones falsas positivas (con  $q < 0.2$ , se espera que un 20% de las asociaciones sean falsas). También existe la posibilidad de que estas asociaciones representen mutaciones de escape dentro de epítomos inusuales o nuevos, así como mutaciones de procesamiento que pueden ocurrir lejos del epítopo asociado. Asociaciones indirectas o “de salto” del tipo  $a \rightarrow b \rightarrow c$ , donde el HLA “a” se muestra como predictor del polimorfismo “c”, serían improbables, debido a que el modelo multivariado del modelo PDN las minimiza. Lo mismo es cierto para asociaciones con alelos HLA en desequilibrio de ligamiento con el alelo selector ya que el fenómeno de desequilibrio de ligamiento es tomado en cuenta por el modelo PDN. Algunas asociaciones adicionales observadas sin tomar en cuenta la covariación de codones también se muestran (Figura 6.17). El no considerar la covariación de codones incrementa el poder para detectar asociaciones, pero permite la presencia de asociaciones indirectas.

Las asociaciones HLA – codón de VIH encontradas para la cohorte mexicana fueron comparadas con las previamente descritas para la cohorte combinada IHAC [171, 206] (Figura 6.17, Tabla 6.5). El hecho de encontrar asociaciones HLA – codón de VIH en la cohorte mexicana relativamente pequeña, que no se observen en la cohorte IHAC cuyo tamaño proporciona alta potencia para encontrar asociaciones, podría ser indicativo de la presencia de evolución única dirigida por HLA en el VIH en cohortes inmunogenéticamente diferentes. Como era esperado, muchas de las asociaciones HLA – codón de VIH observadas en la cohorte mexicana también fueron predichas en la cohorte IHAC, apoyando la observación de patrones mutacionales altamente conservados en la evolución de VIH mediada por HLA (Figura 6.17, Tabla 6.5). Sin embargo, se encontraron también diferencias importantes entre las dos cohortes. De las 43 asociaciones HLA – codón de VIH de la cohorte mexicana, 23 fueron identificadas como asociaciones nuevas, no observadas previamente en la cohorte IHAC (ni en otros estudios previos similares [170, 172, 181],



representando un total de 18 pares HLA – codón de VIH diferentes. Aunque es probable que algunas de estas asociaciones sean falsos positivos debido al umbral de FDR de 20% utilizado, el hecho de que 53% de las asociaciones no se hayan encontrado en la cohorte IHAC, incluso con su alta potencia para detectar asociaciones, es sorprendente. Más aún, de estos 18 nuevos pares HLA – codón de VIH, 8 involucraron asociaciones con codones que no estaban asociados a presión selectiva por HLA en la cohorte IHAC y 4 involucraron asociaciones con B\*39, grupo alélico sustancialmente más frecuente en la cohorte mexicana que en la cohorte IHAC. Notoriamente, 17 de las 23 nuevas asociaciones en la cohorte mexicana involucraron HLAs cuyas frecuencias fueron estadísticamente indistinguibles de las de la cohorte IHAC, sugiriendo que su presencia no se debía a una potencia estadística mayor sino posiblemente a patrones de reconocimiento de epítomos. Además, 2 de las nuevas asociaciones involucraron a B\*27 o B\*08, dos alelos significativamente menos frecuentes en la cohorte del presente estudio que en IHAC, lo cual puede reflejar diferente reconocimiento de epítomos entre las cohortes o la fijación (y la resultante disminución de la potencia estadística para detectar asociaciones) de mutaciones de escape en la cohorte IHAC [134].

Interesantemente, aunque previamente se identificaron como asociados a HLA en la cohorte IHAC, algunos codones de VIH mostraron diferentes especificidades HLA o diferentes aminoácidos blanco en sus asociaciones en la cohorte mexicana. Este fue el caso de 13 de las 23 asociaciones HLA – codón de VIH nuevas (10 de 18 pares HLA – codón de VIH), e.g. la posición 71V en la proteasa se asoció a B\*39 en la cohorte mexicana, mientras que la asociación fue con B\*15 en IHAC; de la misma manera, la posición 245E en RT se asoció a B\*18 en la cohorte mexicana, mientras que la asociación fue con B\*57 en IHAC. Como era esperado, B\*39 es mucho más frecuente en la cohorte mexicana ( $p=1.80E-44$ ,  $q=1.21E-42$ ), mientras que B\*15 fue más frecuente en IHAC ( $p=0.00964$ ,  $q=0.0231$ ) (Figura 6.9). Por otro lado, B\*57 fue menos frecuente en la cohorte mexicana ( $p=0.000430$ ,  $q=0.00180$ ) y no se encontraron diferencias en las frecuencias de B\*18. Estas asociaciones pueden ser explicadas por epítomos confirmados o predichos tanto en la cohorte mexicana como en IHAC (Figura 6.17, Base de Datos de VIH de Los Alamos, [www.hiv.lanl.gov/content/immunology/index.html](http://www.hiv.lanl.gov/content/immunology/index.html)). Tanto B\*15 como B\*39 tienen un epítopo posible en la posición PR 68-76. Estas observaciones muestran la presencia de sitios de alta variabilidad en el genoma del VIH en los cuales HLAs de diferentes especificidades pueden originar huellas similares a nivel poblacional y apoyan observaciones previas mostrando que alelos HLA diferentes pueden seleccionar mutaciones de escape idénticas (además de opuestas) [216].

Interesantemente, se encontraron diferentes aminoácidos en cuatro posiciones de Pol en los consensos de la población mexicana y de la cohorte HOMER (Tabla 6.6). Dos de estos sitios en el VIH estuvieron asociados a HLA en ambas cohortes (RT 272 y RT 277), uno estuvo asociado a HLA exclusivamente en la cohorte HOMER (PR 93) y para otro no se encontró asociación con HLA (RT 293) (aunque la posibilidad de una asociación no detectada con HLA no puede ser descartada). Esta observación apoya la posibilidad de observar diferentes huellas asociadas a HLA en diferentes poblaciones, incluso entre cohortes predominantemente infectadas por virus del mismo subtipo. El papel de la frecuencia de los alelos HLA en la fijación de mutaciones de escape fue evidente en la posición RT 277, asociada a A\*03, donde la forma de escape R se ha fijado en el consenso Pol de HOMER, permaneciendo como la forma susceptible K en el consenso Pol de la cohorte mexicana. La frecuencia de A\*03 en HOMER fue tres veces más alta que en la cohorte mexicana ( $p=7.08E-10$ ,  $q=1.00E-08$ ) (Tabla 6.6, Figura 6.9).

Tabla 6.6. Diferencias en las secuencias consenso de la cohorte Mexicana y la cohorte HOMER.

Posición del VIH	Sus.	Esc.	Consenso mexicano	Consenso HOMER	Consenso B	HLA asociado cohorte mexicana	HLA asociado cohorte HOMER	Frec. HLA cohorte mexicana	Frec. HLA cohorte HOMER
PR 93	I	L	L	I	I	-	B*15	0.0702	0.1131
RT 211	R	G, K	K	K	R	B*44	B*44, B*15	0.0616§	0.1059§
RT 272	P, A	S	P	A	A	B*08	B*42	0.0154§	0.0693§
RT 277	K	R	K	R	K	A*03	A*03	0.0411	0.1235
RT 293	-	-	I	V	I	-	-	-	-

§Se muestra la frecuencia del alelo HLA asociado en la cohorte mexicana. Sus. Forma susceptible a la respuesta específica de CTLs. Esc. Forma de escape a la respuesta específica de CTLs.

Como se ha observado previamente [144, 180], los alelos HLA B estuvieron involucrados en la mayoría de las asociaciones (73% de los pares HLA – codón de VIH), en comparación con los alelos HLA A y C en la cohorte mexicana ( $p<0.001$  en ambos casos). Interesantemente, 16.6% de los pares HLA – codón de VIH se debieron a alelos HLA C. En modelos previos de evolución mediada por HLA, muchas de las asociaciones aparentemente definidas por alelos HLA C representaban asociaciones indirectas con alelos HLA B o A debido al fenómeno de desequilibrio de ligamiento de los genes HLA. Sin embargo, el modelo PDN usado en el presente estudio considera el desequilibrio de ligamiento HLA, minimizando el riesgo de encontrar este y otros tipos de asociaciones indirectas. Es importante notar que la habilidad del modelo PDN para corregir por desequilibrio de ligamiento HLA correlaciona positivamente con el tamaño de muestra y negativamente con la fuerza del desequilibrio de ligamiento. Por esta razón, todavía se podrían encontrar asociaciones falsas positivas cuando existen patrones fuertes de desequilibrio de ligamiento y el ruido aleatorio hace difícil distinguir las asociaciones

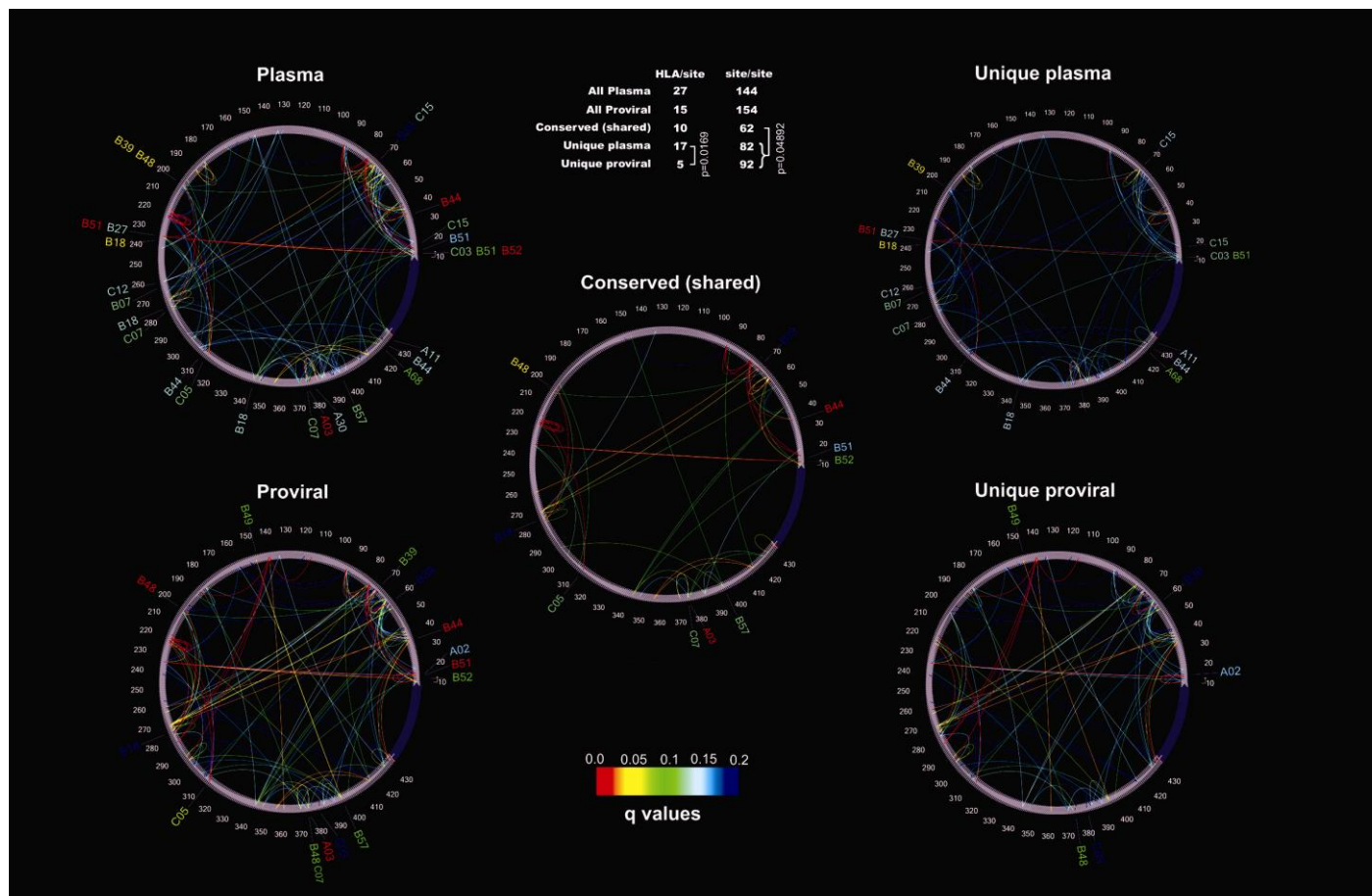
verdaderas. Este podría ser el caso de la posición PR 18, asociada a Cw\*15 y que no pudo ser explicada por predicción de epítomos. Cw\*15 y B\*51 se encuentran en fuerte desequilibrio de ligamiento en la cohorte mexicana (Tabla 6.2) y existe un epítomo experimentalmente confirmado de B\*51 que podría explicar la asociación. Aunque es cierto que muchas asociaciones con alelos HLA C presentaron valores  $q$  altos (Tabla 6.5) y podrían representar asociaciones falsas positivas, es también cierto que algunas de ellas estuvieron fuertemente asociadas y/o estuvieron en consonancia con epítomos predichos (Figura 6.17). Estas observaciones sugieren un papel importante de los alelos HLA C en la evolución poblacional del VIH en la cohorte mexicana.

Por otro lado, el grupo alélico B\*44 por sí solo fue responsable del 16.6% de los pares HLA – codón de VIH, seguido por B\*51 y B\*39, responsables del 10% de los pares HLA – codón de VIH cada uno (Figura 6.17, Tabla 6.5). Esta observación podría estar mostrando diferentes patrones de inmunodominancia y escape inmune entre las dos cohortes e ilustra peculiaridades en la evolución del VIH en cohortes específicas debidas a selección por HLA.

En conjunto, estos resultados apoyan observaciones previas que sugieren patrones mutacionales altamente conservados en la evolución del VIH mediada por HLA, que determinan una huella universal en el virus a nivel poblacional. Sin embargo, los resultados también sugieren la existencia de características únicas de la evolución poblacional del VIH mediada por HLA en poblaciones inmunogenéticamente diferentes, que pueden ser detectadas incluso con cohortes con un número relativamente pequeño de individuos.

## **6.7 DIFERENCIAS EN LA EVOLUCIÓN DEL VIH ENTRE SECUENCIAS DE VIRUS LIBRES EN PLASMA Y PROVIRUS INTEGRADOS EN PBMCs**

Una muestra de 250 individuos VIH positivos para los cuales se tenían secuencias del gen *pol* de provirus integrados en PBMCs además de secuencias de virus libre en plasma y tipificación HLA, fue utilizada para comparar el papel de los genes HLA en la evolución del VIH en dos compartimentos virales diferentes a nivel poblacional. El modelo PDN se aplicó a ambos grupos de secuencias y los resultados fueron mostrados gráficamente con el visualizador de PDN PhyloDv [183] (Figura 6.18).



**Figura 6.18. Asociaciones HLA – codón de VIH para virus libre en plasma y provirus integrados en PBMCs en secuencias de Pol.** Mapas de redes filogenéticas de dependencia fueron generadas con el software PhyloDv (<http://www.codeplex.com/MSCCompBio>). Cada mapa muestra asociaciones con HLA como líneas apuntando a sus sitios correspondientes en los círculos y asociaciones codón de VIH – codón de VIH como arcos dentro de los círculos conectando los sitios de VIH asociados. Plasma: todas las asociaciones de virus libres en plasma en 250 individuos que tenían sus secuencias provirales de PBMCs correspondientes. Proviral: todas las asociaciones de provirus en PBMCs para 250 individuos que tenían sus secuencias de virus en plasma correspondientes. Conserved (shared): asociaciones HLA – codón de VIH o codón de VIH – codón de VIH compartidas entre las secuencias provirales y de virus en plasma. Si un aminoácido en algún sitio de VIH estaba asociado con dos aminoácidos en otro sitio (e.g. uno con escape y otro con reversión), solamente se contó una asociación. Uique plasma: asociaciones encontradas como significativas exclusivamente en virus libres en plasma. Unique proviral: asociaciones encontradas como significativas exclusivamente en provirus de PBMCs. La significancia fue calculada con la prueba binomial. Los valores q se representan como un mapa de calor mostrado en la parte inferior. Solamente se muestran asociaciones con  $q < 0.2$  (todas las asociaciones tuvieron  $p < 0.002$ ). El inserto de la parte superior muestra el número de asociaciones significativas en cada mapa.

Tabla 6.7. Asociaciones HLA – codón del VIH en proteasa y RT de virus libre en plasma y provirus integrado en PBMCs para una cohorte de 250 individuos mexicanos.

Secuencias de <i>pol</i> de virus libre en plasma					Secuencias de <i>pol</i> de provirus integrados en PBMCs				
HLA	Codón de VIH asociado	Variables condicionantes	Valor p	Valor q	HLA	Codón de VIH asociado	Variables condicionantes	Valor p	Valor q
A*03♦	RT 277R		1,14E-08	1,10E-05	A*02♥	PR 15I	PR 77V♥,PR 36M♠, PR 13I♣	0,0010442	0,1522376
A*03♣	RT 277K		1,14E-08	1,10E-05	A*03♦	RT 277R		1,74E-09	1,48E-05
A*11♦	RT 333E	PR 35E♥,RT 293I♣, RT 207Q♠,PR 62I♠	0,00113806	0,1336789	A*03♣	RT 277K		1,74E-09	1,48E-05
A*30♥	RT 281K	RT 278Q♠	0,00109204	0,1336789					
A*68♣	RT 326V		0,00035689	0,1048983					
B*07♦	RT 165I		0,00067347	0,1191747					
B*07♣	RT 165T		0,00067347	0,1191747					
B*18♦	RT 138A		0,00012948	0,0494756					
B*18♣	RT 138E		0,00012948	0,0494756					
B*18♦	RT 245E	PR 45K♥	0,00068267	0,1191747					
B*18♦	RT 177D		0,00094587	0,1312293	B*18♦	RT 177D	RT 123D♠,RT 121D♠	0,0017257	0,19555
B*18♣	RT 177E		0,00094513	0,1312293	B*18♣	RT 177E		0,0006388	0,1183628
B*18♣	RT 245V		0,00115452	0,1336789					
B*27♣	RT 135I	B*51♣,PR 12T♠, RT 202I♠	0,00096164	0,1312293					
B*39♣	RT 102R	B*48♦,RT 64K♣	0,00016533	0,0539628					
B*39♥	PR 71V	PR 93L♠,PR 39P♣, PR 57K♠,C*15♦	0,00193116	0,1994313	B*39♥	PR 71V	PR 57R♣,PR 93I♥,C*15♠, RT 178M♦,PR 39P♣	0,0003592	0,0888581
					B*39♣	PR 60E	PR 39Q♦,PR 37D♦, PR 37E♦	0,0019258	0,199068
B*44♦	PR 35D		1,16E-08	1,10E-05	B*44♦	PR 35D		5,77E-08	0,0002091
B*44♣	PR 35E		1,16E-08	1,10E-05	B*44♣	PR 35E		5,77E-08	0,0002091
B*44♣	RT 211R		0,00112279	0,1336789					
B*44♣	RT 329I	RT 334L♦	0,00120957	0,1359348					
B*48♦	RT 102R		2,47E-08	1,89E-05	B*48♦	RT 102R		1,53E-11	3,88E-07
B*48♣	RT 102K		0,00016947	0,0539628	B*48♣	RT 102K		1,90E-05	0,0120849
					B*48♣	RT 276V	RT 272P♥,C*07♥	0,0005419	0,1049938

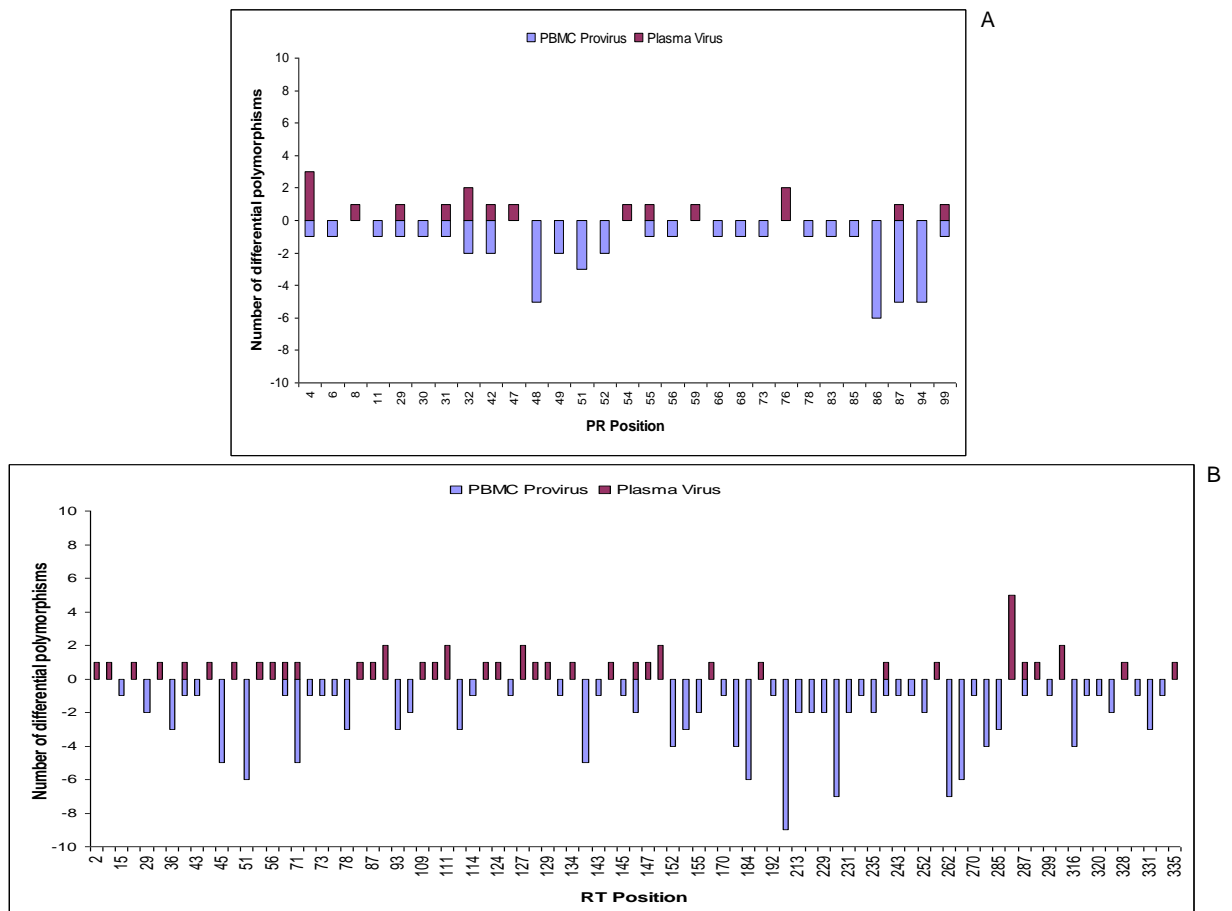
					B*49♣	RT 48S		0,0004979	0,1035941
					B*49♦	RT 48T		0,0004979	0,1035941
B*51♣	RT 135I		4,03E-07	0,0002198					
B*51♦	PR 12S		0,00038855	0,1060476					
B*51♦	PR 14R	RT 135I♣,RT 200A*♣	0,00144043	0,1528856	B*51♦	PR 14R		7,41E-06	0,005226
B*51♣	PR 14K	RT 135I♦,RT 200A*♦	0,00138006	0,1506635	B*51♣	PR 14K		7,41E-06	0,005226
B*52♦	PR 12A		5,59E-07	0,0002671	B*52♦	PR 12A	RT 135T♠	0,0005416	0,1049938
B*52♣	PR 12T		3,73E-08	2,38E-05	B*52♣	PR 12T		1,60E-06	0,0016939
B*57♦	RT 297A		0,00060553	0,1191747	B*57♦	RT 297A		0,0004121	0,0941468
					C*03♦	RT 283I		0,0017747	0,1964929
					C*03♣	RT 283L		0,0017747	0,1964929
C*03♦	PR 12P	PR 14R♠,PR 67C♥	0,00070538	0,1191747					
C*05♦	RT 214L	RT 118I♦	0,00065426	0,1191747	C*05♦	RT 214L	RT 118V♥	0,0002149	0,0665192
C*05♣	RT 214F	RT 118I♣	0,00065426	0,1191747	C*05♣	RT 214F	RT 118V♠	0,0002149	0,0665192
C*07♥	RT 276V	RT 272P♥	0,00054742	0,1191747	C*07♥	RT 276V	RT 272P♥	0,0004154	0,0941468
C*07♣	RT 179D		0,00072749	0,1191747					
C*12♦	RT 162S	PR 64L♦,RT 165T♥, PR 91T♥,RT 135V♣	0,00106536	0,1336789					
C*15♣	PR 18Q		0,00074855	0,1191747					
C*15♦	PR 71V	PR 93L♠,PR 39P♣, PR 57K♠	0,0009574	0,1312293					

♣ escape (tener el predictor hace menos probable tener el blanco), ♦ atracción (tener el predictor hace más probable tener el blanco), ♥ reversión (no tener el predictor hace más probable tener el blanco), ♠ repulsión (no tener el predictor hace menos probable tener el blanco).

En total se encontraron 36 asociaciones HLA – codón de VIH en virus libres en plasma y sólo 24 para provirus integrados en PBMCs, representando 27 y 15 pares HLA – codón de VIH respectivamente (Figura 6.18, Tabla 6.7). Interesantemente, el número de pares HLA – codón de VIH únicos en virus libre en plasma fue significativamente mayor que el número de pares únicos en provirus de PBMCs ( $p=0.0169$ ). Sólo 10 de los pares HLA – codón de VIH fueron observados en ambos compartimentos. Estos resultados son consistentes con un estudio reciente que reporta la presencia de mutaciones de escape asociadas a HLA en secuencias de virus de plasma que son raras en el compartimento proviral en algunos individuos infectados [217]. La observación de una evolución general diferente en secuencias de virus libre en plasma y secuencias provirales, con un número significativamente menor de posiciones de VIH asociadas a HLA en el compartimento proviral es consistente con el modelo que sugiere que las secuencias provirales representan secuencias de VIH tempranas, archivadas en el reservorio latente y que las secuencias de virus plasmático representan una población viral que ha evolucionado en respuesta a la presión selectiva inmune del hospedero. Más aún, estas observaciones sugieren un desarrollo dinámico de respuestas de CTLs a lo largo de la infección, de manera que respuestas tempranas de CTLs quedan archivadas en el compartimento proviral mientras que el compartimento plasmático refleja la presión de respuestas de CTLs más recientes [218, 219]. Sin embargo, las asociaciones HLA – codón de VIH provirales no correlacionaron con epítomos previamente definidos como de escape rápido al estar bajo presión por HLA [169]. Este hecho puede reflejar diferencias en velocidades de escape en cohortes inmunogenéticamente divergentes o diferencias en presión selectiva por CTLs en diferentes compartimentos que simplemente reflejan la naturaleza de archivo de las secuencias provirales.

Interesantemente, el número de codones de VIH en covariación que fueron comunes para ambos compartimento fue menor que el número de sitios covariantes observados exclusivamente en secuencias de virus de plasma o provirales ( $p=0.0489$ ). Más aún, se observó un número elevado de sitios covariantes en secuencias provirales, reflejando probablemente remanentes de la adaptación viral a hospederos pasados.

En conjunto, estos resultados muestran diferencias en la selección mediada por HLA en secuencias de virus de plasma y provirus de PBMCs, sugiriendo una evolución asociada a HLA altamente dinámica en el VIH, ya que muchas de las asociaciones HLA – codón de VIH en virus libre en plasma no fueron evidentes en el compartimento proviral enriquecido en secuencias tempranas archivadas que probablemente todavía no reflejan adaptación a las respuestas restringidas por HLA del hospedero.



**Figura 6.19. Mayor polimorfismo general en secuencias provirales de PBMCs que en secuencias de virus libre en plasma.** Se muestra el número total de polimorfismos para cada posición de 250 secuencias del gen de la proteasa (A) y de la RT (B) obtenidas de manera pareada a partir de virus libre en plasma y a partir de provirus integrado en PBMCs, en relación a la secuencia de referencia HXB2. Los polimorfismos diferenciales en los dos compartimentos fueron comparados para cada posición con la prueba de McNemar corregida para comparaciones múltiples por métodos de razón de descubrimientos falsos (FDR).

Adicionalmente, se comparó el grado general de polimorfismo en las secuencias provirales de PBMCs contra las secuencias correspondientes de virus libre en plasma tomando como referencia la secuencia de referencia HXB2 (Figura 6.19). En conjunto, las secuencias provirales mostraron un grado de polimorfismo general mayor que las secuencias de virus libre en plasma ( $\chi^2$  de McNemar=14.9031,  $p=0.0001132$ ). Este mayor grado de polimorfismo en el compartimento proviral podría ser representativo de la adaptación del virus a respuestas tempranas de CTLs, las cuales se reflejarían en secuencias virales tempranas archivadas en este compartimento. Más aún, se buscaron diferencias significativas entre las secuencias provirales y de virus libre en plasma en el número de polimorfismos para cada codón mediante pruebas de McNemar, corrigiendo para comparaciones múltiples con métodos de razón de descubrimientos falsos (FDR) [220]. La



posición RT 272 fue la única con polimorfismo diferencial significativo entre los dos compartimentos virales ( $\chi^2$  de McNemar=7.1111,  $p=0.0076$ ,  $q=0.1877$ ). Interesantemente, la posición RT 272 se asoció a presión selectiva por HLA tanto en el compartimento proviral como en el compartimento de virus plasmático en la cohorte mexicana, así como en la cohorte IHAC (Figura 6.17, Tabla 6.8) y se identificó como una posición bajo selección positiva por modelos matemáticos (Tabla 6.8, ver abajo). Estos datos sugieren que esta posición de la RT podría representar un sitio de escape a respuestas de CTLs tempranas, con mutaciones que tienden a revertir parcialmente en la infección crónica al modificarse la respuesta de CTLs durante el avance de la infección.

## 6.8 DETECCIÓN DE POSICIONES EN EL GENOMA VIRAL BAJO PRESIÓN SELECTIVA A TRAVÉS DE MODELOS MATEMÁTICOS DE CODONES

Tabla 6.8. Detección de selección positiva por modelos de codones en secuencias de VIH de la cohorte mexicana.

Codón bajo Presión selectiva*	Virus Plasma		Provirus PBMCs		Codones en co-evolución	Asociación con HLA
	dN-dS Normalizado	Valor p	dN-dS Normalizado	Valor p		
PR 12	293.477	0.00132042	232.404	0.00253196	PR 14, RT 311	IHAC, Mex
PR 19	225.935	0.0221282		NS	RT 297, RT 329	IHAC
PR 37	566.508	0.286704	53.956	0.432598	-	IHAC
PR 60	160.967	0.0380266	187.507	0.0166046	PR 39, PR 62	No
PR 71	21.978	0.0156142	19.315	0.0258959	PR 35, PR 61, RT 173	IHAC, Mex
PR 72	370.026	0.0184568		NS	-	No
PR 77	360.819	0.00136812	332.379	0.00287733	-	No
PR 93	280.756	0.0027707	357.102	0.000396848	PR 70	IHAC
RT 35	291.189	0.00524808	277.644	0.00890384	-	No
RT 102	270.613	0.000432898	309.018	0.000129144	RT 179	IHAC, Mex
RT 135	400.189	0.00988806	360.805	0.0116994	RT 178	IHAC, Mex
RT 165	1.031	0.0397925		NS	RT 239	IHAC, Mex
RT 200	311.007	0.00153852	379.241	0.000275201	-	IHAC, Mex
RT 211	378.873	0.00375352	325.552	0.0124905	-	IHAC, Mex
RT 245	368.705	0.00246414	376.575	0.00212562	-	IHAC, Mex
RT 272	464.309	0.256069	443.778	0.402341	-	IHAC, Mex
RT 288	168.356	0.0233416		NS	RT 103	No
RT 297	500.481	0.00047012	487.857	0.00082719	PR 19	IHAC, Mex
RT 322	106.442	0.0382113		NS	PR 70	No
RT 334	432.638	0.00333132	411.403	0.00724173	-	IHAC

\*Aplicando un modelo de evolución molecular de codones a un alineamiento de 150 secuencias del gen *pol* obtenidas de forma pareada a partir de virus libre en plasma y de provirus de PBMCs de individuos de la cohorte mexicana. Las huellas de selección se infirieron por estimación de las tasas de sustituciones sinónimas (dS) y no sinónimas (dN) con el análisis SLAC del servidor Datamonkey ([www.datamonkey.org](http://www.datamonkey.org)) [212]. Se muestran posiciones con  $dN>dS$  ( $p<0.05$ ). Se muestra además si las posiciones identificadas como bajo presión selectiva se encontraron asociadas a HLA en la cohorte IHAC o la cohorte mexicana. NS – No significativo.

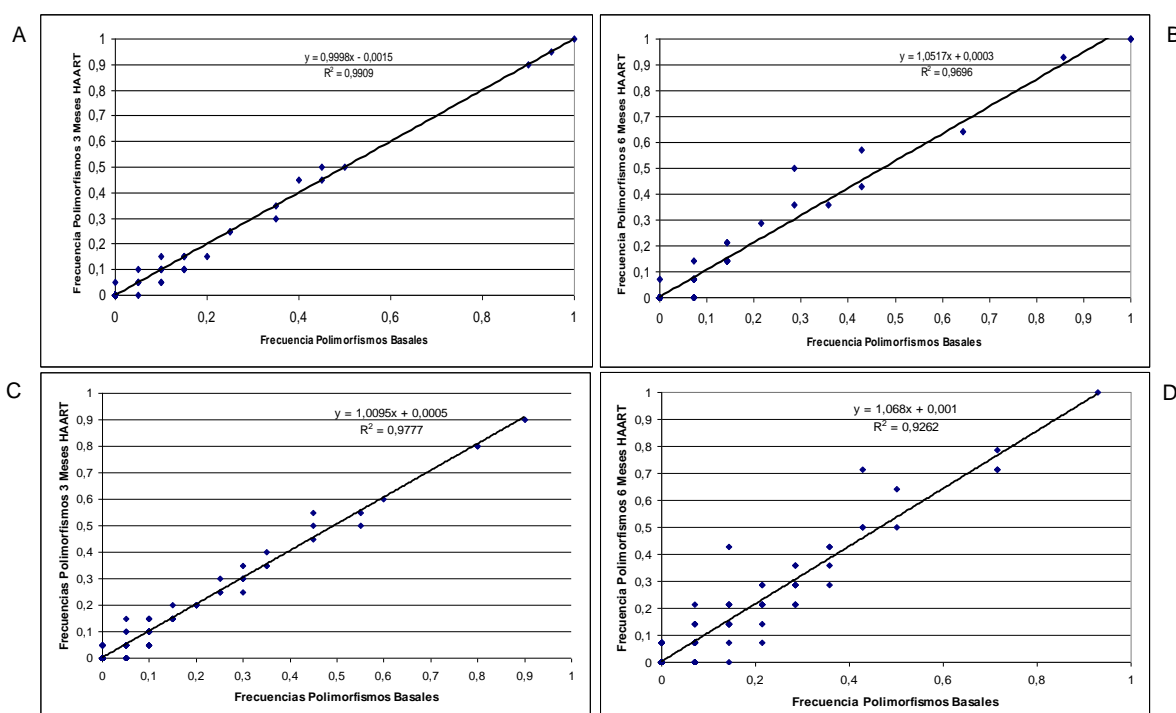
Se aplicó un modelo de codones para buscar evidencias de selección positiva en las secuencias virales del compartimento de virus plasmático y del compartimento de provirus integrados en PBMCs, con el objetivo de determinar la magnitud del impacto de la presión selectiva por HLA dentro del contexto de las varias presiones selectivas posibles que pueden afectar la evolución del VIH a nivel poblacional. Los sitios del genoma viral con evidencia de selección positiva se determinaron mediante el programa SLAC del servidor Datamonkey, como se describe en la sección de Métodos. Estos sitios fueron definidos como aquéllos en los que la tasa de sustituciones no sinónimas era significativamente mayor que la tasa de sustituciones sinónimas ( $dN > dS$ ,  $p < 0.05$ ).

En total se encontraron 20 posiciones en la región de la proteasa y la RT analizada con evidencia de selección positiva para las secuencias plasmáticas mexicanas (Tabla 6.8). De éstas, 5 resultaron no significativas para el compartimento proviral, sugiriendo un aumento en la gama de presiones selectivas que actúan sobre el virus a medida que avanza la infección, las cuales se reflejan en el virus replicativamente activo, en comparación con la gama de presiones selectivas existentes en la etapa inicial de la infección, reflejadas en secuencias tempranas archivadas en el compartimento proviral. Interesantemente, 14 (70%) de las posiciones con evidencia de selección positiva fueron identificadas como sitios de presión selectiva asociada a HLA en la cohorte mexicana y/o en la cohorte IHAC, sugiriendo una fuerte influencia y dominio de esta presión selectiva en la evolución del VIH a nivel poblacional, dentro del contexto de las varias presiones selectivas que pueden actuar sobre el virus (Tabla 6.8). Notablemente, 3 de las 5 posiciones con evidencia de selección positiva en el compartimento de virus plasmático pero no en el compartimento proviral, no se identificaron como posiciones asociadas a HLA en la cohorte mexicana ni en la cohorte IHAC. Esto sugiere la presencia de presiones asociadas a HLA que van surgiendo a lo largo de la infección y que se reflejan en los virus replicativamente activos, pero que no se evidencian en el compartimento proviral, rico en secuencias tempranas archivadas. Esta observación está en concordancia con las evidencias que muestran una evolución de las respuestas inmunes restringidas por HLA a lo largo de la infección, con un aumento característico en la amplitud de estas respuestas que se asocia con escape viral [218, 219].

Interesantemente, cerca de la mitad de las posiciones bajo presión positiva se asociaron a otras posiciones del proteoma viral en coevolución, incluso en proteínas distintas, sugiriendo alta complejidad en los patrones mutacionales del virus (Tabla 6.8). Esta observación es consistente con la complejidad de los patrones mutacionales asociados a la evolución mediada por HLA en el virus (Figuras 6.16 y 6.18).

La mitad de las posiciones con evidencia de selección positiva se asociaron a presión selectiva por HLA en la cohorte mexicana, mientras que un 70% de estas posiciones se asociaron a presión selectiva por HLA en la cohorte IHAC (Tabla 6.8). Esto sugiere la existencia de asociaciones HLA-codón de VIH que no fueron identificadas para la cohorte mexicana, dado el poder estadístico del presente análisis debido muy probablemente al tamaño de muestra utilizado.

## 6.9 EFECTO DEL TRATAMIENTO ARV EN LA EVOLUCIÓN DEL VIH A NIVEL POBLACIONAL



**Figura 6.20. Correlaciones entre la frecuencia de polimorfismos en secuencias del gen *pol* de virus libre en plasma antes y después del inicio del tratamiento ARV.** La frecuencia de polimorfismos se calculó para secuencias pareadas antes y después del inicio del tratamiento ARV en relación a la secuencia de referencia HXB2. Se muestran regresiones lineales entre la frecuencia de polimorfismos en 20 secuencias del gen de la proteasa antes del inicio del tratamiento ARV y después de tres (A) y seis (B) meses de su inicio; así como regresiones lineales entre la frecuencia de polimorfismos en 20 secuencias del gen de la RT antes del inicio del tratamiento ARV y después de tres (C) y seis (D) meses de su inicio. En todos los casos  $p < 0.001$ .

El tratamiento ARV altamente activo (HAART) disminuye los niveles de virus en plasma hasta niveles indetectables por métodos estándar de laboratorio. Esta notable limitación de la replicación viral tiene importantes consecuencias para la capacidad de variación del VIH, pudiendo limitar su capacidad de adaptación a la presión selectiva impuesta por la respuesta inmune. Para evaluar el efecto de la terapia ARV sobre la evolución del virus a nivel poblacional de manera general se comparó el grado de polimorfismo en relación a la secuencia de referencia HXB2 en un grupo de secuencias del gen *pol* del VIH antes y después del inicio del tratamiento ARV, de manera que cada

secuencia de virus bajo tratamiento ARV tuvo como control a su secuencia correspondiente antes del inicio del tratamiento (Figura 6.20).

Se observó una alta similitud en el polimorfismo general de las secuencias antes y después del inicio del tratamiento ARV, tanto en la proteasa como en la RT. Esta similitud fue especialmente notoria para las secuencias obtenidas a los 3 meses del inicio del tratamiento ARV (Figura 6.20 A y C) y se confirmó al construir un árbol filogenético con secuencias pareadas antes y después de tres meses de tratamiento ARV, donde los pares de secuencias antes de y bajo tratamiento ARV se agruparon fuertemente en casi todos los casos, mostrando en general grados sutiles de divergencia genética (Figura 6.21). Sin embargo, se observó una ligera tendencia de aumento en la dispersión de los puntos (evidente en los coeficientes de correlación) en las comparaciones de polimorfismos de las secuencias obtenidas a los 6 meses el inicio del tratamiento ARV (Figura 6.20 B y D). Además, se observó un ligero aumento en las pendientes de las correlaciones entre los polimorfismos de virus antes del tratamiento ARV y después de 6 meses de su inicio, indicando posiblemente un aumento en el polimorfismo general de las secuencias de virus bajo tratamiento ARV (Figura 6.20 B y D). Estas observaciones son sugerentes de evolución continua del VIH aún bajo la presión selectiva del tratamiento ARV o de manera alternativa, de un reemplazo de las variantes virales dominantes en la viremia residual provenientes de reservorios viral diferentes [221, 222]. Este aumento aparente en la variación del virus durante el tratamiento ARV fue más evidente en la RT que en la proteasa (Figura 6.20 B y D). Estudios prospectivos a largo plazo en individuos en tratamiento ARV serán necesarios para apoyar estas observaciones.

Para buscar de manera más detallada diferencias en la evolución viral antes y durante el tratamiento ARV, se realizó una búsqueda comparativa de posiciones con evidencia de selección positiva en secuencias de virus plasmático en individuos sin tratamiento ARV previo y sus correspondientes secuencias amplificadas de la viremia residual en plasma de individuos con tres meses de tratamiento ARV y supresión de viremia exitosa (Tabla 6.9). Se aplicó el análisis SLAC a través del servidor Datamonkey, como se describe en la sección de Métodos. Estos sitios fueron definidos como aquéllos en los que la tasa de sustituciones no sinónimas era significativamente mayor que la tasa de sustituciones sinónimas ( $dN > dS$ ,  $p < 0.1$ ). Se incluyeron 40 secuencias de individuos con carga viral indetectable y sus correspondientes secuencias de virus libre en plasma antes del inicio del tratamiento ARV en el análisis.

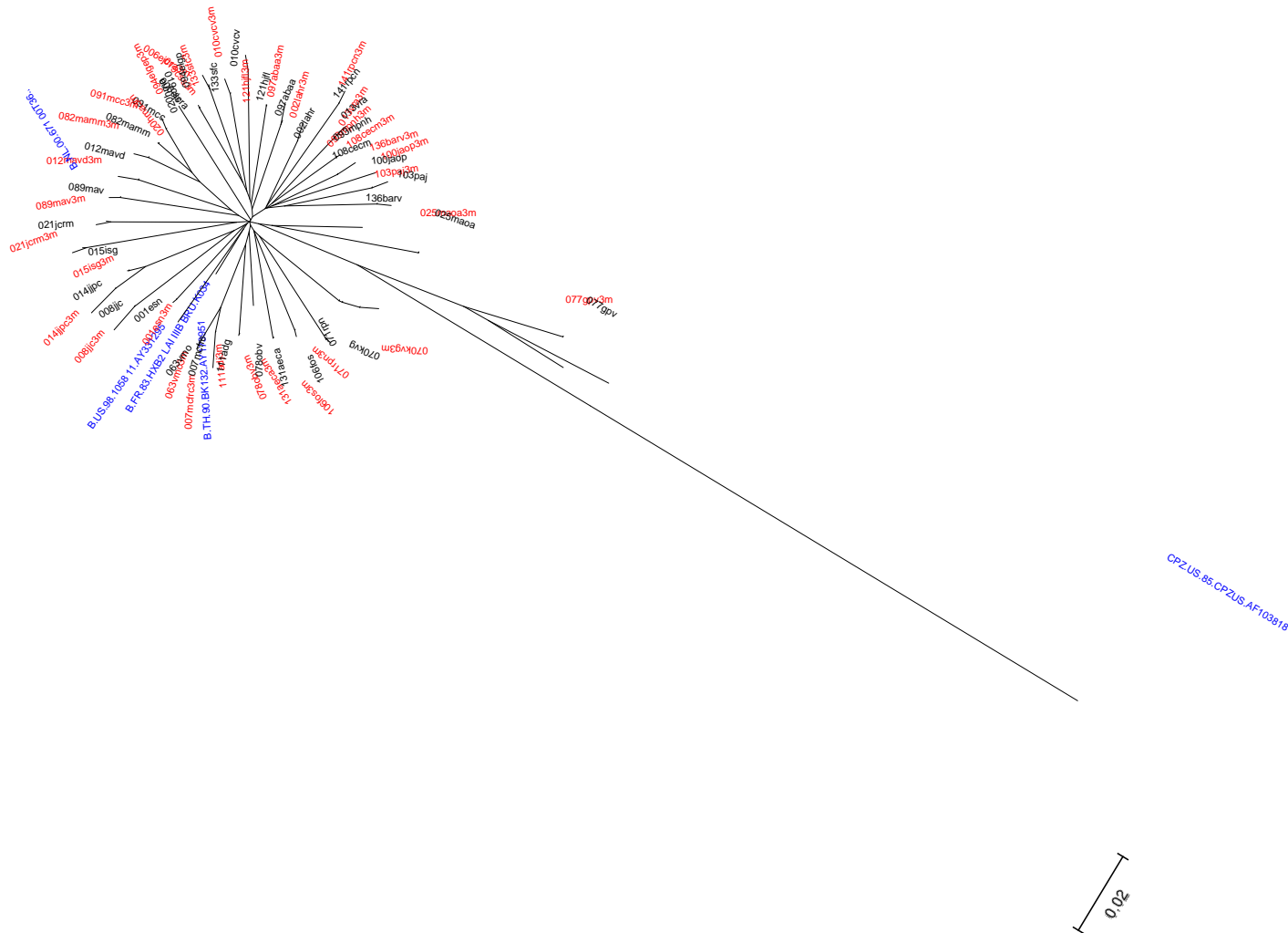
Tabla 6.9. Detección de selección positiva en secuencias de VIH de virus plasmático antes y después del inicio del tratamiento ARV.

Codones bajo selección positiva*	Virus plasmático sin tratamiento		Virus plasmático 3 meses HAART		Asociación con HLA
	dN-dS Normalizado	Valor p	dN-dS Normalizado	Valor p	
PR 35	1.17666	0.0643593	0.961361	0.0874984	IHAC, Mex
PR 37	2.12604	0.00769769	2.07797	0.00438316	IHAC
PR 41	1.00495	0.0664684	1.35141	0.020513	No
PR 93	1.36623	0.0153369	1.68988	0.00505695	IHAC
RT 35	1.34145	0.0436785	1.18398	0.0536475	No
RT 200	1.09174	0.0339921	1.32100	0.0152359	IHAC, Mex
RT 207	NS		1.91131	0.0376481	IHAC
RT 211	1.48918	0.0689142	1.43035	0.0587883	IHAC, Mex
RT 272	1.17302	0.0259531	1.40706	0.0114976	IHAC, Mex
RT 276	0.947417	0.0555735	1.01850	0.0400374	Mex
RT 297	1.95391	0.018626	1.71144	0.0273014	IHAC, Mex

\*Aplicando un modelo de evolución molecular de codones a un alineamiento de 40 secuencias del gen *pol* obtenidas de forma pareada a partir de virus libre en plasma antes del inicio del tratamiento ARV y después de 3 meses de inicio del tratamiento ARV en individuos con supresión viral exitosa de la cohorte mexicana. Las huellas de selección se infirieron por estimación de las tasas de sustituciones sinónimas (dS) y no sinónimas (dN) con el análisis SLAC del servidor Datamonkey ([www.datamonkey.org](http://www.datamonkey.org)) [212]. Se muestran posiciones con  $dN > dS$  ( $p < 0.1$ ). Se muestra además si las posiciones identificadas como bajo selección positiva se encontraron asociadas a HLA en la cohorte IHAC o la cohorte mexicana. NS – No significativo, HAART – Terapia antirretroviral altamente activa.

Prácticamente las mismas posiciones identificadas como bajo selección positiva en la población de virus plasmáticos antes del inicio del tratamiento ARV se identificaron también en virus de plasma después de 3 meses del inicio del tratamiento, sugiriendo que las mutaciones de adaptación a las diferentes presiones selectivas actuando sobre el virus antes del tratamiento ARV, principalmente a la presión selectiva por HLA, tienden a conservarse a nivel poblacional al menos durante los primeros tres meses del inicio del tratamiento ARV. Este tiempo es suficiente para reducir la carga viral plasmática a niveles indetectables en la mayoría de los individuos en HAART, limitando considerablemente las posibilidades de variación del virus a nivel intra-individuo. Como se describió arriba, una alta proporción de las posiciones bajo selección positiva (más del 80%) se encontraron asociadas a HLA en la cohorte IHAC y/o en la cohorte mexicana. En este análisis se estableció un umbral de significancia más laxo ( $p < 0.1$ ) debido al menor número de secuencias disponibles. Este número menor de secuencias también podría explicar diferencias respecto al análisis mostrado en la Tabla 6.8, con más poder estadístico para detectar posiciones bajo selección positiva. Una sola posición, RT 207, se encontró asociada a selección positiva exclusivamente en virus plasmático bajo tratamiento ARV, pero no antes del inicio del tratamiento (Tabla 6.9). La mutación RT 207R/A/E no se ha asociado a resistencia a fármacos ARVs (Base de Datos de Resistencia de VIH de Stanford, <http://hivdb.stanford.edu/index.html>) y por tanto es improbable que haya sido seleccionada

directamente por la presión selectiva del tratamiento ARV. Sin embargo, esta posición se ha asociado a adaptación hacia la presión selectiva por HLA-B\*15 en la cohorte IHAC [206]. Esta asociación con HLA no fue detectada en la cohorte mexicana.



**Figura 6.21. Relaciones filogenéticas entre VIH antes y después del inicio de la terapia ARV.** Se muestra un árbol filogenético de distancias construido con 35 secuencias pareadas de fragmentos de 1305 bp del gen *pol* viral abarcando todo el gen de la proteasa y dos terceras partes de la RT, de virus plasmáticos antes y después de tres meses del inicio de la terapia ARV. El árbol se realizó empleando el programa MEGA 4, utilizando el método de Neighbor-Joinig. Se muestra el árbol consenso inferido a partir de 1000 réplicas de bootstrap en formato radial. Se incluyeron 4 secuencias de referencia del subtipo B, 2 secuencias de referencia de la CRF 06 cpx y una secuencia de referencia de SIV de chimpancé, obtenidas de la Base de Datos de Secuencias de Los Alamos y mostradas en azul. Las secuencias mexicanas antes del tratamiento ARV se muestran en negro, las secuencias mexicanas después de tres meses de tratamiento ARV se muestran en rojo. Las ramas correspondientes a particiones reproducidas en menos del 50% de las réplicas de bootstrap están colapsadas. Las distancias evolutivas fueron calculadas con el modelo de dos parámetros de Kimura y se encuentran en unidades de número esperado de sustituciones por sitio. Todas las posiciones con información faltante fueron eliminadas por comparación pareada de secuencias (pairwise comparison).

Variando el umbral de significancia de  $p < 0.1$  a  $p < 0.05$ , se observarían algunos cambios entre los virus antes y después del inicio del tratamiento ARV que podrían ser importantes y requerirán análisis con un mayor número de secuencias para confirmarse. La presencia de selección positiva en posiciones como PR 41 y RT 276 parecería volverse más

significativa una vez iniciado el tratamiento ARV y la presencia de selección positiva en posiciones como RT 35 parecería perder significancia al iniciarse el tratamiento (Tabla 6.9). Estas diferencias, de establecerse, podrían apoyar observaciones previas de cambios radicales en el reservorio viral a partir del cual se genera la viremia después del inicio del tratamiento ARV [221, 222], con la posibilidad de que estos reservorios culpables de la viremia residual estuvieran sujetos a presiones selectivas distintas a los virus plasmáticos sin la presencia de la presión selectiva del tratamiento.

## **7. DISCUSIÓN**

### **7.1 CARACTERÍSTICAS ÚNICAS DE LA EVOLUCIÓN DEPENDIENTE DE HLA DEL VIH EN UNA COHORTE DE INDIVIDUOS MEXICANOS**

En este estudio se ha presentado evidencia que sugiere que una distribución única de frecuencias de alelos HLA en una cohorte de individuos VIH positivos mexicanos ha dejado huellas únicas en la evolución del virus a nivel poblacional. Se estudió la evolución del VIH mediada por HLA en una cohorte de individuos mexicanos infectada mayoritariamente por virus de subtipo B, comparando nuestros datos con los obtenidos para la cohorte IHAC, la más grande cohorte de individuos infectados principalmente por virus de subtipo B reunida hasta ahora para estudiar evolución mediada por HLA del VIH a nivel poblacional y compuesta por individuos provenientes de British Columbia, Canadá; Australia Occidental y Estados Unidos [171, 206]. Las dos cohortes presentaron contextos inmunogenéticos claramente diferentes, con una notable mezcla de genes amerindios en la cohorte del centro/sur de México (Figura 6.9). Estos diferentes contextos inmunogenéticos permitieron evaluar el papel de dos distribuciones alélicas HLA diferentes en la evolución del VIH en dos cohortes infectadas por virus del mismo subtipo. La cohorte mexicana del presente estudio reflejó de manera representativa las características típicas de las cohortes VIH positivas mexicanas, enriquecidas en individuos en etapas avanzadas de la enfermedad por VIH y con una distribución alélica HLA similar a la población general (Figura 6.7) con algunas excepciones específicas notorias, como la alta frecuencia del grupo alélico B\*39 en población VIH positiva, que tendrán que ser evaluadas en estudios posteriores como posibles factores de riesgo para la infección por el virus en la población mexicana.

Aunque la cohorte en estudio reflejó las características típicas de las cohortes latinoamericanas infectadas por VIH, no podemos asegurar que sea representativa de la población mexicana infectada en general. Más aun, reportes recientes han sugerido



diferencias en distribuciones de frecuencias HLA incluso dentro de la población mestiza mexicana, existiendo la posibilidad de observar diferencias en adaptación viral en poblaciones mestizas de distintas áreas geográficas dentro del territorio nacional [207]. Sin embargo, las comparaciones realizadas en el presente trabajo con cohortes de individuos predominantemente caucásicos muestran diferencias importantes en distribuciones de frecuencias HLA que definen un contexto inmunogenético único en la cohorte mexicana, previamente no estudiado, de gran interés para evaluar evolución mediada por HLA del VIH, incluso con la existencia de posibles sesgos de muestreo. De cualquier manera, comparaciones de frecuencias HLA entre la cohorte de individuos mexicanos infectados por VIH y una cohorte de población abierta VIH negativa con diferencias en procedencia geográfica, mostraron pocas diferencias significativas, apoyando la representatividad de nuestros análisis. La representatividad limitada de la cohorte del presente estudio limita también las conclusiones de homogeneidad genética y de prevalencia de resistencia transmitida en virus circulantes en México del presente estudio. Sin embargo, un estudio comprensivo de Epidemiología Molecular del VIH en México en la cohorte más grande de individuos mexicanos infectados por VIH reunida hasta el momento en el país, incluyendo a más de 1500 personas sin tratamiento ARV previo, provenientes de los estados con mayor contribución al número total de infecciones en el país, se está llevando a cabo en nuestros laboratorios del CIENI del INER. Este estudio a gran escala en la población mexicana infectada ha confirmando en parte los hallazgos observados en la cohorte del presente estudio (datos no publicados). Eventualmente, la conclusión de este estudio comprensivo con alta representatividad a nivel nacional proporcionará información molecular importante, hasta ahora no existente, acerca de la epidemia de VIH en México.

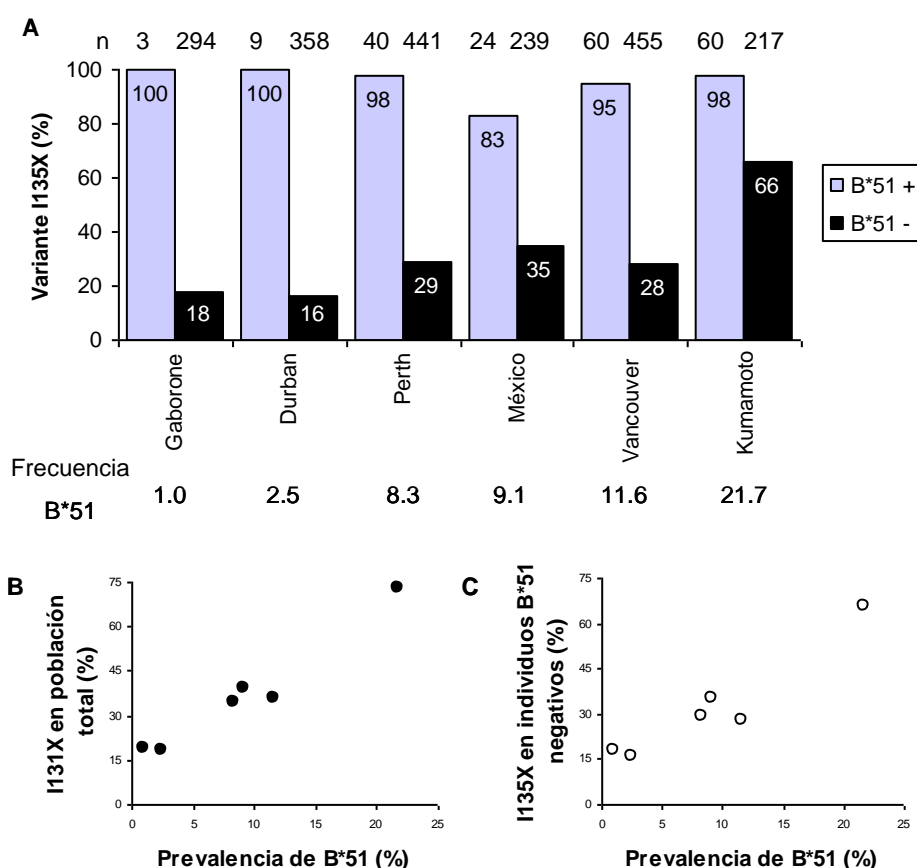
Nuestro análisis por modelos matemáticos de evolución basados en razones de las tasas de sustituciones sinónimas y sustituciones no sinónimas confirmó el importante papel de la presión de selección por HLA en la evolución poblacional del VIH en la cohorte mexicana, al encontrarse asociadas a HLA al menos 70% de las posiciones con evidencia de selección positiva identificadas (Tabla 6.8). Además, el hecho de que varias posiciones asociadas a HLA fueron también identificadas como posiciones con evidencia de selección positiva por modelos matemáticos de codones aumenta la validez y confiabilidad de los resultados obtenidos en el análisis evolutivo (Tabla 6.8),

Estudios previos han sugerido que la evolución poblacional del VIH sigue patrones mutacionales asociados a presión selectiva por CTLs ampliamente predecibles y altamente conservados [160, 167-171, 173, 202]. Las conclusiones obtenidas de la comparación directa entre diferentes estudios en diferentes poblaciones han estado limitadas por el uso

de modelos distintos para evaluar la evolución viral mediada por HLA, en los cuales frecuentemente no se han considerado fuentes de confusión importantes. En este estudio se aplicó el modelo PDN, recientemente descrito [183], a una cohorte de individuos mexicanos infectados por virus de subtipo B. El modelo PDN toma en cuenta simultáneamente los efectos de las principales fuentes de confusión presentes al evaluar la evolución mediada por HLA a nivel poblacional en el VIH: el desequilibrio de ligamiento entre los genes HLA, la covariación de codones del VIH y los efectos de la filogenia del VIH. Nuestros datos fueron comparados con los obtenidos al aplicar el modelo PDN a la cohorte multicéntrica IHAC, permitiéndonos de esta manera evaluar el impacto de los genes HLA en la evolución del virus en una población inmunogenéticamente única, previamente no caracterizada.

Nuestros datos apoyaron las observaciones de huellas poblacionales asociadas a HLA altamente conservadas en el proteoma viral ya que varias de las asociaciones HLA – codón de VIH encontradas para la cohorte mexicana han sido consistentemente observadas en la cohorte IHAC así como en otros estudios previos con cohortes diversas [167, 168, 170, 172, 173]. Interesantemente, sin embargo, nuestros datos predijeron la existencia de huellas asociadas a HLA únicas en la evolución del VIH en la cohorte mexicana que podrían estar influenciadas por una distribución de frecuencias alélicas HLA únicas en diferentes poblaciones infectadas. Las características únicas de la evolución mediada por HLA del virus en la cohorte mexicana se evidenciaron no solamente por la presencia de pares HLA – codón de VIH únicos no detectados en la cohorte IHAC, sino también por la presencia de posiciones en el proteoma viral previamente identificadas como bajo presión por HLA, pero con diferentes especificidades HLA y/o aminoácidos blanco en las dos cohortes. El grado en el que estas asociaciones representan un fenómeno biológico verdadero y no un fenómeno estadístico tendrá que ser confirmado con datos experimentales; sin embargo, la evidencia presentada en este estudio sugiere fuertemente la existencia de diferencias reales entre las dos cohortes. Aunque la cohorte mexicana era bastante más pequeña que la cohorte IHAC (la potencia para detectar asociaciones se incrementa dramáticamente con el tamaño de la muestra [183]), resultando en una capacidad de detección de sólo un 20% de las asociaciones previamente definidas en la cohorte del presente estudio, el hecho de que 53% de las asociaciones HLA – codón de VIH fueron novedosas en la cohorte mexicana sugiere fuertemente diferencias en la evolución mediada por HLA en las dos cohortes. Aunque estas asociaciones nuevas podrían representar falsos negativos de la cohorte IHAC, esta cohorte multicéntrica es lo suficientemente grande para que la razón de falsos negativos esperada sea muy pequeña y cualquier falso negativo sería un evento raro en ella [183]. Otra posibilidad es que las asociaciones nuevas representen falsos positivos en la cohorte del presente estudio; sin embargo, aún con una razón de falsos positivos esperada de hasta

20% debido al umbral de  $q < 0.2$  definido, el número de asociaciones novedosas encontradas en la cohorte mexicana es sorprendente. Además, es importante mencionar que de los 18 pares HLA – codón de VIH novedosos de la cohorte mexicana, al menos 11 (61%) pueden ser explicados por epítomos confirmados o potenciales de CTLs (Figura 6.17, Tabla 6.5), lo cual apoya la validez de estas asociaciones y sugiere la existencia de diferencias biológicas reales en la selección dependiente de HLA entre las dos cohortes.



**Figura 7.1. Correlación entre la frecuencia de la mutación de escape I135X y la frecuencia del alelo selector HLA-B\*51 en diferentes poblaciones.** (A) Frecuencia de la mutación I135X, dentro del epítomo TAFTIPSI (RT 128-135), en individuos HLA-B\*51 positivos y negativos en 6 cohortes de estudio. (B) Correlación entre la frecuencia de I135X en la población total y la prevalencia de HLA-B\*51 en 6 cohortes de estudio. (C) Correlación entre la frecuencia de I135X en individuos carentes del alelo selector HLA-B\*51 y la prevalencia de HLA-B\*51 en 6 cohortes de estudio. Modificado y adaptado de [180] incluyendo datos del presente estudio para México.

Otra evidencia que apoya el impacto diferencial de la selección por HLA en la evolución del VIH a nivel poblacional en diferentes poblaciones es la observación de diferencias puntuales en las secuencias consenso poblacionales de la cohorte mexicana y la cohorte IHAC que pueden mapearse a sitios asociados a HLA (Tabla 6.6). Este fue el caso de la posición RT 277, asociada con A\*03 tanto en la cohorte mexicana como en IHAC, en la cual la forma adaptada 277R se ha fijado al consenso de IHAC, permaneciendo como la forma susceptible 277K en el consenso mexicano. No es sorprendente que la frecuencia de

A\*03 en IHAC haya sido tres veces mayor que en la cohorte mexicana ( $p=7.08E-10$ ,  $q=1.00E-08$ ), mostrando un papel importante de la frecuencia alélica HLA en la fijación de mutaciones de escape a nivel poblacional (Figura 6.9, Tabla 6.6). De manera similar, la posición PR 93 se asoció con B\*15 en IHAC con la forma susceptible 93I en el consenso de esta cohorte, mientras que la forma adaptada 93L se ha fijado en el consenso de la cohorte mexicana. Aunque no se encontró una asociación directa con HLA en la cohorte mexicana para este sitio (probablemente debido a falta de potencia estadística), la posición PR 93 se asoció a otros sitios de VIH tales como PR 71 en donde sí se observaron asociaciones con HLA (Figura 6.16). Así, cambios en la secuencia consenso poblacional pueden ligarse a selección mediada por HLA.

En un trabajo reciente [180], Kawashima y colaboradores demostraron que la frecuencia de algunas mutaciones de escape seleccionadas por alelos HLA capaces de producir respuestas protectoras hacia el avance de la infección por VIH correlacionaba fuertemente con la prevalencia del alelo HLA selector en diferentes poblaciones del mundo, lo cual es sugerente de una posible adaptación del VIH a respuestas restringidas por HLA a nivel poblacional. Esta observación queda claramente ilustrada con el epítipo TAFTIPSI de la transcriptasa inversa (RT 128-135), donde la frecuencia de la mutación de escape I135X correlacionó fuertemente con la prevalencia del alelo selector B\*51 en 9 cohortes de diferentes partes del mundo [180]. En Japón, donde B\*51 es altamente prevalente (más del 20% de la población), la frecuencia de I135X fue cercana al 75% con la consiguiente fijación de la mutación en el consenso poblacional y la progresiva extinción del epítipo TAFTIPSI en la población japonesa infectada (Figura 7.1A). Interesantemente, existe evidencia de que la efectividad del alelo HLA-B\*51 como protector hacia el avance de la infección por VIH ha ido disminuyendo con el tiempo hasta llegar al grado de ya no ser considerado como protector en la población japonesa [180]. Notablemente, los resultados obtenidos en el presente estudio para la cohorte de individuos VIH positivos mexicanos concuerdan con lo observado por Kawashima y colaboradores. Con una frecuencia de B\*51 cercana al 10% en la cohorte mexicana, la mutación I135X se presentó en cerca del 40% de la población, mostrándose aún una asociación correlativa entre la presencia de valores de carga viral bajos y la expresión de algunos haplotipos HLA que incluyen alelos B\*51 (Figuras 6.11 y 7.1A). Notoriamente, el grado de acumulación de la mutación I135X en la población B\*51 negativa (y por tanto el grado de fijación de la mutación de escape a nivel poblacional) en diversas poblaciones del mundo, incluyendo el estimado para la población mexicana, fue proporcional a la frecuencia de este alelo en cada población (Figura 7.1B y C).

También de interés es el hecho de que B\*44 estuvo asociado con 5 de los 30 pares HLA – codón de VIH identificados para la cohorte mexicana. Dos de estas asociaciones han sido descritas en la cohorte IHAC y dos han sido previamente identificadas como posiciones bajo presión selectiva por HLA, pero con distintas especificidades en la cohorte mexicana (Figura 6.17). La fuerte influencia de B\*44 en la evolución del VIH mediada por HLA en la cohorte mexicana podría estar reflejando diferencias en las jerarquías de inmunodominancia de las respuestas de CTLs en el contexto de diferentes distribuciones de frecuencias de alelos HLA. Mientras que respuestas de CTLs fuertemente inmunodominantes podrían enmascarar el efecto de otras respuestas menos inmunodominantes en una cohorte, estas respuestas podrían tener un mayor impacto en la evolución viral en otra cohorte en la cual las respuestas más inmunodominantes sean poco frecuentes. Es notable que la frecuencia de muchos alelos HLA inmunodominantes como B\*57, B\*27, B\*08, B\*07, A\*03 y A\*11 [169], fue menor en la cohorte mexicana que en la cohorte IHAC ( $q < 0.05$ ) (Figura 6.9). Es posible que en la cohorte IHAC, las respuestas de CTL restringidas por estos alelos enmascaren el efecto de otros alelos menos inmunogénicos que se observan frecuentemente en la población mexicana. En efecto, este podría ser el caso de Cw\*07, el grupo alélico HLA C más frecuente en la población mexicana, que explica 10% de los pares HLA – codón de VIH observados en nuestro análisis. Estas asociaciones son interesantemente únicas para la cohorte mexicana y pueden ser confirmadas con epítomos predichos, fuertes asociaciones estadísticas ( $q < 0.05$ ) (RT 276) y/o evidencia de selección positiva mediante el uso de modelos matemáticos de codones (RT 276) (Figura 6.17, Tabla 6.9).

El caso de B\*39 es también notable, al ser el grupo alélico HLA-B más frecuente en la cohorte mexicana con una frecuencia 7 veces mayor que la observada en la cohorte IHAC ( $p = 1.80E-44$ ,  $q = 1.21E-42$ ) (Figura 6.9). B\*39 explicó otro 10% de los pares HLA – codón de VIH identificados en la cohorte mexicana, sugiriendo ya sea una fuerte influencia de este grupo alélico en la evolución del VIH en el contexto inmunogenético de la cohorte mexicana o una más alta potencia estadística para detectar asociaciones con este grupo alélico HLA. Interesantemente, las asociaciones restringidas por B\*39 tuvieron principalmente distribuciones de hoja de escape (tener el HLA asociado hace menos probable tener el aminoácido blanco) en el modelo PDN, con aminoácidos blanco diferentes al aminoácido consenso, lo cual sugiere que la secuencia consenso podría ser la forma de escape del virus (Figura 6.9). Esto podría sugerir un papel frecuente de B\*39 en la conservación de la secuencia viral en la cohorte mexicana a nivel poblacional. La conservación de la secuencia viral asociada a HLA ha sido previamente descrita con HLAs frecuentes que promueven la acumulación de variantes adaptadas a las respuestas de CTLs en diferentes poblaciones [180, 202].

En general, dos aspectos clave podrían explicar la observación de asociaciones diferentes en cohortes con frecuencias HLA distintas, pero infectadas mayoritariamente por virus del mismo subtipo: 1) Diferentes patrones de inmunodominancia, que apoyan la existencia de diferencias reales en los epítomos blancos de las respuestas de CTLs más importantes en la población; y 2) Diferentes potencias estadísticas para detectar asociaciones, que sugieren un efecto estadístico más que una diferencia biológica real en la evolución del VIH mediada por HLA en cohortes inmunogenéticamente diferentes. Como se ha descrito arriba, diferentes patrones de inmunodominancia determinados por diferentes distribuciones de frecuencias de alelos HLA en una población, podrían evidenciar el efecto de alelos HLA menos frecuentes en la ausencia de las respuestas fuertemente inmunodominantes en algunas poblaciones. Para ser capaces de confirmar esta posibilidad a nivel poblacional se requiere ser capaces de reducir las proporciones de falsos positivos y falsos negativos esperadas. Aunque la proporción de falsos negativos en la cohorte IHAC es baja, se requiere información experimental para confirmar este punto. Por otro lado, diferencias en las frecuencias HLA pueden simplemente cambiar la potencia estadística para detectar asociaciones y apoyan la importancia de evaluar la selección mediada por HLA en diversas cohortes. La posibilidad también existe de que un simple problema de potencia estadística podría resolverse al combinar diferentes cohortes infectadas por el mismo subtipo viral para obtener cohortes de referencia más grandes, lo cual apoyaría la creación de un banco universal de asociaciones que podrían irse actualizando a medida que se añadan nuevas secuencias. Este tipo de base de asociaciones podría permitir la extrapolación a partir de un grupo grande de referencia a nuevos grupos demográficos para los cuales la formación de cohortes y la colección de información fuera difícil. El hecho de que 17 de las 23 asociaciones HLA – codón de VIH nuevas en la cohorte mexicana involucraran HLAs cuyas frecuencias no fueron significativamente diferentes de las observadas en la cohorte IHAC, sugiere fuertemente que su presencia no se debe a un aumento en potencia estadística en su detección, sino más bien a diferencias en los epítomos blanco de las respuestas de CTLs más frecuentes. De todas maneras, es muy probable que tanto efectos de inmunodominancia como de potencia estadística se pudieran presentar al mismo tiempo en el análisis de una cohorte determinada. Ejemplos de ambos fenómenos han sido descritos para la cohorte mexicana (ver arriba), sugiriendo que un grupo de cohortes diversas inmunogenéticamente hablando podrían enriquecer grandemente los estudios de evolución del VIH sin la necesidad de grandes cohortes. También es importante mencionar el hecho de que la resolución hasta sólo dos dígitos de los alelos HLA podría no revelar posibles divergencias adicionales en la presión selectiva por HLA, ya que se pueden definir subtipos HLA con motivos de unión de péptido diferentes

para ciertos grupos alélicos como B\*35, B\*40, B\*51, B\*58, A\*02, todos con distribuciones características en diferentes poblaciones [134, 144]. Así, divergencias significativas de selección en algunos casos, podrían ser explicadas por diferencias en prevalencia de subtipos HLA de cuatro dígitos predominantes dentro del grupo alélico grueso en las poblaciones comparadas. Este hecho podría tener un impacto en la potencia estadística para detectar asociaciones debidas a los diferentes subtipos HLA en diferentes poblaciones y apoya aún más la existencia de huellas poblacionales únicas en diversas poblaciones.

En resumen, aunque existen limitaciones importantes en el análisis de la evolución del VIH mediada por HLA en la cohorte mexicana, particularmente la presencia de asociaciones falsas positivas y la baja potencia estadística para detectar asociaciones, nuestros análisis mostraron evidencia sólida que sugiere la existencia de características únicas reales en la evolución del VIH en la cohorte mexicana, incluyendo la proporción sorprendente de asociaciones HLA – codón de VIH únicas para esta cohorte, muchas de las cuales pueden ser confirmadas a través de epítomos confirmados o predichos; la presencia de diferencias puntuales asociadas a HLA en las secuencias consenso de la cohorte mexicana y la cohorte HOMER que reflejan la fijación de mutaciones de escape a CTLs a nivel poblacional con alta dependencia de la frecuencia HLA; así como la existencia de una alta proporción de asociaciones novedosas que involucran HLAs cuyas frecuencias fueron similares en la cohorte mexicana y la cohorte IHAC, lo cual sugiere que al menos algunas asociaciones no se observan por un fenómeno estadístico relacionado con la potencia de detección de asociaciones sino que representan diferencias biológicas reales tal vez asociadas a fenómenos de inmunodominancia.

## **7.2 DIFERENCIAS EN LA EVOLUCIÓN ASOCIADA A HLA DEL VIH EN EL COMPARTIMENTO DE PROVIRUS INTEGRADOS EN PBMCs Y DE VIRUS LIBRE EN PLASMA**

Para obtener mayor información sobre la evolución del VIH mediada por HLA, se compararon las asociaciones HLA – codón de VIH en virus libres en plasma y provirus integrados en PBMCs en un grupo de 250 individuos de la cohorte mexicana, para los cuales estaban disponibles las secuencias de los virus de ambos compartimentos. Como se observó al dibujar las PDNs de ambos compartimentos virales, se presentaron diferentes patrones mutacionales y diferentes asociaciones HLA – codón de VIH en virus replicativos en plasma y provirus archivados en PBMCs a nivel poblacional. Un número significativamente menor de asociaciones HLA – codón de VIH se observó en las

secuencias provirales y un mayor número de asociaciones codón de VIH – codón de VIH únicas se observó en los virus de plasma o provirus que asociaciones codón de VIH – codón de VIH compartidas en ambos compartimentos (Figura 6.18, Tabla 6.7). Este resultado es congruente con el obtenido en el análisis de selección positiva por modelos matemáticos de codones basados en razón de sustituciones sinónimas y no sinónimas, el cual mostró también un mayor número de posiciones bajo selección positiva en el compartimento de virus libre en plasma con respecto al compartimento proviral (Tabla 6.8). Estos resultados podrían ser explicados por la observación de que las secuencias provirales frecuentemente representan un reservorio estable rico en secuencias de VIH archivadas de manera temprana en la infección [223], mientras que los virus en plasma representan secuencias posteriores en el curso de la infección. Así, las secuencias provirales podrían haber sido archivadas antes de que se produjeran respuestas de CTLs hacia algunos epítomos, o antes de que se seleccionaran mutaciones de escape en epítomos que ya eran blanco de la respuesta de CTLs, reflejando de manera alternativa mutaciones de escape asociadas a respuestas tempranas de CTLs, cuya jerarquía de inmunodominancia muchas veces disminuye a lo largo de la infección crónica para ser sustituidas por otras respuestas de CTLs. Este fenómeno cronológico podría consecuentemente causar una pérdida de potencia para detectar asociaciones HLA – codón de VIH potenciales en las secuencias provirales a nivel poblacional, por lo que no es sorprendente que se hayan observado menos asociaciones HLA – codón de VIH en este compartimento en comparación con el de virus libres en plasma. En efecto, estudios previos han mostrado la presencia de mutaciones de escape asociadas a HLA en virus de plasma que son raras en provirus dentro de individuos infectados [224]. Sin embargo, los pares HLA – codón de VIH provirales no pudieron ser correlacionados con epítomos de escape temprano conocidos [225] en el presente análisis, aunque la posibilidad existe de que una cohorte más grande o el estudio de otros genes virales evidencie esta correlación. Asimismo, debido a que la cohorte del estudio descrito en [225] mostraba una distribución HLA similar a la cohorte IHAC, se podría esperar que existieran diferentes patrones de escape que en la inmunogenéticamente única cohorte mexicana. Las asociaciones provirales en la cohorte mexicana podrían por tanto representar eventos de escape temprano en el contexto de una cohorte latinoamericana.

Sorprendentemente, algunas asociaciones con HLA detectadas en secuencias provirales no se observaron en los virus de plasma. Algunos de los pares HLA – codón de VIH observados exclusivamente en el compartimento proviral presentaron valores de  $q$  altos, sugiriendo este hecho una posible presencia de asociaciones falsas positivas. Por otro lado, las asociaciones provirales únicas podrían también sugerir un rearrreglo cronológico de la evolución del VIH mediada por HLA, reflejando mutaciones de escape de rápida reversión



que son perdidas tan pronto como el virus es transmitido a individuos sin el alelo HLA selector. Esta posibilidad estaría apoyada por la observación de un mayor grado de polimorfismo general en las secuencias provirales que en las secuencias de virus libre en plasma que podrían reflejar la reversión de mutaciones de adaptación a respuestas inmunes del hospedero pasado y el surgimiento de mutaciones de adaptación a respuestas inmunodominantes tempranas en el nuevo hospedero (Figura 6.19). Adicionalmente, la posición RT 272, única con polimorfismo diferencial significativo entre virus libre en plasma y provirus integrado en PBMCs a nivel poblacional, además de estar asociada a presión selectiva por HLA en ambos compartimentos virales, ilustraría la posible reversión gradual de mutaciones asociadas a HLA seleccionadas por respuestas de CTLs en etapas tempranas de la infección, las cuales muchas veces dejan de ser una presión de selección importante con el avance de la infección y los consiguientes cambios en las respuestas inmunodominantes [169]. Una explicación adicional de las diferencias en asociaciones HLA – codón de VIH observadas en los compartimentos proviral y plasmático sería la existencia de compartimentalización de las variantes de VIH en un individuo infectado. La compartimentalización por compartimento anatómicos, órganos específicos e incluso diferentes subpoblaciones celulares y su relación con diferencias en selección positiva dentro de un individuo infectado han sido descritas previamente [226]. Este fenómeno podría explicar diferencias poblacionales entre los virus activamente replicativos provenientes de un compartimento con presiones selectivas características y los provirus archivados que permanecen como un reservorio viral originado en diferentes compartimentos anatómicos y/o celulares.

Una posibilidad factible que podría explicar las diferencias en polimorfismo entre la secuencia plasmática y proviral de PBMCs observada en algunos de los individuos del estudio, sería un posible fenómeno de redistribución celular provocada por la infección por VIH. Células infectadas productoras de virus replicativamente activo contemporáneo, podrían ser secuestradas en órganos linfoides y mucosas, haciendo que células de memoria, previamente infectadas, con virus latente predominen en la circulación. El secuestro de células en ganglios linfáticos durante la infección y diferencias en subpoblaciones de células T CD4+ en cuanto a susceptibilidad de infección, activación y tendencias de migración hacia mucosas y órganos linfoides en la infección por VIH se han descrito ampliamente [227-229]. De esta manera, la secuenciación gruesa de provirus a partir de PBMCs usada en el presente estudio mostraría diferencias entre el virus replicativamente activo observado en plasma y los provirus integrados en células circulantes, posiblemente infectadas por variantes más tempranas archivadas, mientras que las secuencias provirales de células

productivamente infectadas se reflejarían minoritariamente al no encontrarse frecuentemente en circulación.

Las asociaciones compartidas entre los virus de plasma y los provirus de PBMCs podrían reflejar sitios del proteoma viral con una presión continua de respuestas de CTLs a lo largo de la infección crónica, una característica que podría ser de interés en la selección de blancos candidatos para vacunas. Por otro lado, estas asociaciones aparentemente más estables podrían también reflejar epítomos que son blanco de respuestas de CTLs tempranas, las cuales cambian con el tiempo y dejan de ejercer presión selectiva pero sin que ocurra reversión en las mutaciones de escape, implicando costos bajos en aptitud replicativa impuestos por estas mutaciones. Si este último caso fuera cierto, algunas asociaciones compartidas en los dos compartimentos podrían ser más propensas a fijarse en la población en el futuro. Esta posibilidad tendría implicaciones para nuestro entendimiento y capacidades predictivas de la adaptación del VIH a diferentes poblaciones humanas.

De manera análoga a lo encontrado en las asociaciones HLA – codón de VIH, se observaron pares de codones de VIH en coevolución únicos en las secuencias provirales y en las secuencias de virus libre en plasma. Los pares codón de VIH – codón de VIH únicos en el compartimento proviral podrían explicarse como una reorganización de los patrones mutacionales en la evolución del VIH con vías de reversión remanentes asociadas a la adaptación al hospedero previo y nuevas vías asociadas a la adaptación al hospedero actual. Esta reorganización de patrones mutacionales se estaría reflejando adicionalmente en un mayor grado de polimorfismo general de las secuencias provirales con respecto a las secuencias de virus libre en plasma (Figura 6.19). Los pares codón de VIH – codón de VIH únicos en virus de plasma podrían estar reflejando huellas secuenciales dejadas por la adaptación viral a las respuestas restringidas por HLA en la infección crónica. Estas observaciones traen consecuencias importantes para nuestro entendimiento de la evolución del VIH mediada por HLA, sugiriendo que la apariencia y densidad de una PDN para una población específica son altamente dinámicas y podrían variar con el tiempo. El desarrollo dinámico de las respuestas de CTLs a lo largo de la infección dentro de un individuo ha sido previamente reportado [218, 219]. Estudios posteriores en cohortes de seguimiento o en cohortes transversales cuidadosamente estratificadas, podrían ayudar a apoyar o refutar estas observaciones iniciales.

En resumen, el presente estudio mostró diferencias notables en la evolución poblacional de virus libres en plasma y provirus integrados en PBMCs que pueden

explicarse por la observación de que el compartimento proviral es rico en secuencias virales tempranas archivadas que no reflejan el impacto de presiones selectivas desarrolladas durante el transcurso de la etapa crónica de la infección. Estas diferencias reflejan la dinámica de cambio de las respuestas de CTLs a lo largo de la infección y sugieren patrones de inmunodominancia cambiantes a nivel poblacional, teniendo importantes aplicaciones en el diseño de vacunas candidatas.

### **7.3 IMPACTO DEL TRATAMIENTO ARV EN LA EVOLUCIÓN DEL VIH A NIVEL POBLACIONAL**

Debido a la importante limitación en la capacidad de variación del VIH impuesta por el tratamiento ARV efectivo, y la implicación de esta limitación de la variabilidad viral en las posibilidades de adaptación del VIH a las diversas presiones selectivas que actúan sobre él, se evaluó la evolución del virus antes y después del inicio del tratamiento ARV a nivel poblacional. Estudios previos han mostrado que las poblaciones virales en la viremia residual de individuos bajo tratamiento ARV son altamente homogéneas, no muestran evidencia de evolución importante y podrían provenir de reservorios previamente no identificados con presiones selectivas desconocidas y posiblemente diferentes a las que actúan sobre los virus replicativamente activos previamente a la terapia ARV [221, 222]. Nuestros datos mostraron poca evidencia de cambio entre el polimorfismo general de los virus antes y después del inicio del tratamiento ARV (Figuras 6.20 y 6.21). Sin embargo, el análisis de polimorfismo general entre las secuencias antes y después del inicio del tratamiento ARV a nivel poblacional sugirió una ligera tendencia de variación en los virus bajo tratamiento, la cual fue más notoria a seis meses que a tres meses del inicio de la terapia ARV (Figura 6.20). Estos resultados van aparentemente en contra de la relativa ausencia de evolución de los virus de la viremia residual en plasma de individuos bajo tratamiento reportada previamente [221] y requieren confirmarse por estudios a largo plazo en un mayor número de individuos bajo tratamiento ARV efectivo. Sin embargo, debido a que el presente estudio se realizó en individuos con inicio de tratamiento ARV reciente, el efecto de aumento en la variación viral observado podría explicarse también como el lapso de surgimiento de las variantes dominantes de la viremia residual a partir de reservorios virales nuevos, las cuales son generalmente diferentes a los virus circulantes antes del tratamiento [221, 222]

Aunque análisis por modelos matemáticos de codones, basados en razones de las tasas de sustituciones sinónimas y no sinónimas para detectar posiciones con evidencia de selección positiva no mostraron diferencias aparentes entre la población de virus antes y

después del inicio del tratamiento ARV, estudios en cohortes más grandes, con mayor poder estadístico y confiabilidad para inferir la presencia de selección positiva serán necesarios para confirmar estas observaciones. Interesantemente, variando el umbral de significancia ( $p < 0.1$  a  $p < 0.05$ ) del análisis de selección positiva se observarían algunos cambios entre los virus antes y después del inicio del tratamiento ARV que podrían ser indicativos de las diferencias descritas entre los viriones de la viremia residual y los viriones de la viremia sin la presión selectiva del tratamiento ARV, algunas de las cuales estarían asociadas con presión selectiva por HLA (Tabla 6.9) [221, 222]. Este podría ser el caso de la posición RT 207, identificada como posición bajo selección positiva exclusivamente en virus bajo tratamiento ARV, o de la posición RT 276, identificada como posición bajo selección positiva con mayor significancia previamente al inicio del tratamiento ARV. El cambio en las variantes dominantes en la viremia residual en presencia de tratamiento ARV se asocia muy probablemente con la generación de virus a partir de un reservorio hasta ahora desconocido [222], sujeto a presiones selectivas diferentes a las que actúan en los virus sin presión selectiva impuesta por el tratamiento ARV. La posibilidad de que este reservorio viral alternativo esté influenciado por presiones selectivas restringidas por HLA diferentes tendrá que ser evaluada más a fondo en estudios posteriores y será importante para la generación de estrategias de control de la viremia residual.

#### **7.4 POSIBLES FACTORES GENÉTICOS ASOCIADOS A PROGRESIÓN RÁPIDA O LENTA DE LA ENFERMEDAD POR VIH, O A UNA MAYOR SUSCEPTIBILIDAD DE INFECCIÓN POR EL VIRUS EN LA POBLACIÓN MEXICANA**

El presente estudio mostró evidencias importantes que pueden guiar nuevas líneas de investigación para determinar factores de protección para la adquisición de la infección por VIH o que determinan una progresión más rápida o lenta que el promedio de la enfermedad por el virus en la población mexicana. La alta frecuencia del grupo alélico HLA-B\*39 en la población infectada con respecto a la población general mexicana fue notoria (Figura 6.7). La influencia de este grupo alélico en el riesgo de adquisición de la infección tendrá que comprobarse mediante estudios en individuos expuestos al virus en múltiples ocasiones pero que permanecen sin infección aparente y mediante experimentos funcionales. La frecuencia de distintos alelos B\*39 en poblaciones nativas latinoamericanas es alta, pudiendo implicar un mayor reclutamiento de personas con ascendencia indígena, muchas veces dentro de los grupos sociales con alto grado de marginación y fuertemente afectadas por el VIH/SIDA, entre los individuos participantes en el presente estudio. Sin embargo, aunque es posible que la alta frecuencia de B\*39 en nuestra cohorte de individuos

VIH positivos mexicanos se deba a un sesgo en el reclutamiento, las semejanzas en las frecuencias de otros alelos HLA con la población general, pero no en este grupo alélico en específico, sugieren un papel de B\*39 como factor de riesgo para adquirir la infección. Estudios de caracterización de los alelos B\*39 más frecuentes en la población mestiza mexicana y su papel en el desarrollo de la infección por VIH están ya siendo realizados por nuestro grupo como seguimiento de los resultados del presente estudio.

Se realizaron análisis de correlación entre valores altos y bajos de carga viral y la expresión de alelos HLA determinados en la presente cohorte de individuos VIH positivos mexicanos. Estos análisis tienen la importante limitación de que una gran proporción de los individuos que componen la cohorte del presente estudio se encontraban en etapas tardías de la infección, donde los valores de carga viral tienden a elevarse y separarse significativamente del valor de equilibrio (setpoint viral) establecido al principio de la etapa crónica de la infección (Figura 1.7). Aún con este inconveniente, el efecto de alelos ampliamente identificados como protectores, como B\*57 y B\*58 (Tabla 1.1), se conservó en la cohorte mexicana al observarse asociaciones correlativas entre valores significativamente más bajos de carga viral y la expresión de estos alelos HLA o haplotipos que los contienen (Figuras 6.10, 6.11 y 6.12). Interesantemente, haplotipos no relacionados con B\*57 y B\*58, como B\*40/Cw\*15, B\*51/Cw\*08, A\*68/Cw\*15, A\*68/B\*40/Cw\*15 y A\*02/B\*51/Cw\*08 también se asociaron a valores bajos de carga viral. B\*51 se ha asociado previamente a progresión lenta a SIDA (Figura 1.1) e interesantemente, el epítipo RT 128-135, implicado en la respuesta protectora asociada a este alelo, puede explicar asociaciones HLA – codón de VIH observadas en la cohorte mexicana (Tablas 1.1 y 6.5, Figura 6.17). Más aún, los alelos B\*40 y Cw\*15 definieron asociaciones HLA – codón de VIH únicas en la población mexicana (Tabla 6.5), pudiendo las respuestas definidas por estos alelos determinar características de protección hacia el avance de la infección por VIH en nuestra población. Estos resultados deberán confirmarse mediante experimentos funcionales y estudios correlativos en cohortes de individuos en etapas tempranas de la infección.

Por otro lado, los haplotipos A\*02/B\*48, A\*68/B\*44, B\*35/Cw\*07, A\*68/Cw\*16, A\*68/B\*44/Cw\*16, A\*02/B\*35/Cw\*07, A\*24/B\*35/Cw\*07 y A\*02/B\*48/Cw\*08 se asociaron a valores altos de carga viral, pudiendo representar factores de riesgo para el avance rápido hacia SIDA en la infección por VIH en mexicanos (Figuras 6.11 y 6.12). Algunos alelos B\*35 se han identificado extensivamente como alelos de riesgo para un avance rápido de la infección por VIH (Tabla 1.1). Dada la alta frecuencia de B\*35 en la población mexicana (Figuras 6.4 y 6.7) y su posible papel en la rápida progresión a SIDA de los individuos mexicanos, sospechada por observaciones clínicas de los pacientes atendidos en nuestro

centro, estudios dedicados a caracterizar las respuestas restringidas por este grupo alélico en nuestra población, comparando con otras poblaciones del mundo, serán de gran importancia para el entendimiento de las respuestas protectoras y de riesgo en el contexto inmunogenético de la población mexicana. Adicionalmente, existen asociaciones HLA – codón de VIH determinadas por B\*44 y B\*48 en la población mexicana y que podrían reflejar respuestas de riesgo para un avance rápido a SIDA en esta población. En especial, B\*44 definió asociaciones HLA – codón de VIH únicas para la población mexicana, por lo que su papel como factor de riesgo específico de nuestra población deberá evaluarse especialmente en estudios posteriores.

Es importante mencionar que los estudios de asociación entre la expresión de alelos HLA específicos y valores altos y bajos de carga viral, serían más sensibles al utilizar tipificación HLA de alta resolución, la cual permite definir los alelos HLA al menos hasta 4 dígitos, y que el uso de baja resolución en la tipificación HLA podría enmascarar el efecto de alelos específicos. Diferencias importantes e incluso efectos opuestos se han observado en las respuestas definidas por alelos del mismo grupo. Un ejemplo importante es la asociación de B\*5801 con valores bajos de carga viral y avance lento a SIDA, en oposición a B\*5802, asociado con cargas virales altas y riesgo de avance rápido a SIDA (Tabla 1.1).

## 8. CONCLUSIONES

Nuestros datos derivados del análisis de la evolución del VIH mediada por HLA en una cohorte previamente no caracterizada e inmunogenéticamente única del centro/sur de México, apoya observaciones previas que evidencian un componente altamente conservado y predecible de la evolución mediada por HLA a nivel poblacional, que está dejando una huella real en los virus circulantes a nivel mundial. Este efecto es evidente incluso considerando objeciones importantes a la hipótesis de la huella poblacional en el VIH asociada a HLA tales como la rápida reversión de una parte considerable del total de las mutaciones de escape a CTLs en ausencia del alelo HLA selector [230], la complejidad de las respuestas de CTLs que frecuentemente imponen fuerzas selectivas conflictivas en el mismo sitio de la secuencia viral [174], la posibilidad de que una variante de escape seleccionada por un alelo HLA específico pueda ser blanco de otras respuestas de CTLs restringidas por otros alelos HLA [161, 176, 181] y las dinámicas jerárquicas de inmunodominancia observadas en las respuestas restringidas por HLA [169, 218, 219].

Notablemente, el análisis de la evolución del VIH mediada por HLA en la cohorte mexicana mostró asociaciones HLA – codón de VIH adicionales, que no se han descrito previamente en otras cohortes estudiadas, incluyendo la gran cohorte combinada IHAC, con un contexto inmunogenético claramente diferente a la cohorte mexicana. Este hecho apoya la posibilidad de que estas asociaciones específicas no estén impactando la evolución poblacional del VIH en otras cohortes, pero que sean significativas en el contexto inmunogenético de la población mexicana.

La realización de estudios evolutivos con modelos comparables que tomen en consideración efectos confusorios importantes tales como la covariación de codones de VIH, efectos de linaje del VIH y desequilibrio de ligamiento de los genes HLA, incluso realizados en cohortes relativamente pequeñas, pero inmunogenéticamente divergentes, podrían ser útiles para detectar las huellas asociadas a HLA de manera específica en diferentes poblaciones en el mundo y permitirían la construcción de un panorama más completo del impacto de la selección por HLA en la evolución del VIH a nivel mundial. Este conocimiento sería útil para el desarrollo de vacunas candidatas y para el desarrollo de estrategias terapéuticas dirigidas a poblaciones específicas.

La creación de una base de datos universal de sitios de VIH asociados a HLA empleando modelos comprensivos y comparables para evaluar la evolución del VIH mediada por HLA en poblaciones inmunogenéticamente divergentes, infectadas por virus de diferentes subtipos, mejoraría nuestro entendimiento de la evolución del virus a nivel global. Asimismo, evidencia experimental futura ayudará a entender las limitaciones impuestas por modelos estadísticos para detectar las huellas asociadas a HLA en la evolución del VIH en diferentes poblaciones.

Una comparación entre la evolución mediada por HLA en secuencias de virus libres en plasma y provirus integrados en PBMCs sugirió una evolución mediada por HLA altamente dinámica en el VIH, ya que muchas de las asociaciones HLA – codón de VIH observadas en el compartimento de virus libre en plasma no fueron evidentes en el compartimento proviral, presuntamente enriquecido en secuencias tempranas de VIH que no reflejan de manera completa la evolución intra-hospedero dirigida por HLA del virus. Además, las asociaciones HLA – codón de VIH compartidas en ambos compartimentos podrían ser de interés al reflejar epítomos que son blanco persistente de respuestas de CTLs a lo largo de la infección o, de manera alternativa, mutaciones de escape con bajo costo en aptitud replicativa que podrían llegar a fijarse en la población en el futuro. Estudios futuros con cohortes más grandes y utilizando varios genes virales podrían enriquecer estas



observaciones primarias e incrementar nuestro entendimiento de la adaptación del VIH a diferentes poblaciones en el mundo.

Un análisis preliminar del impacto del inicio de la terapia ARV sobre la evolución del virus a nivel poblacional mostró una ligera tendencia de variación en los virus componentes de la viremia residual bajo tratamiento ARV exitoso con respecto a virus antes del inicio del tratamiento, sugiriendo posibles diferencias en la selección por HLA entre virus antes y después del inicio del tratamiento. Estas posibles diferencias podrían apoyar observaciones previas de un cambio en el reservorio viral que origina la viremia una vez iniciado el tratamiento ARV, el cual podría estar sometido a diferentes presiones HLA con respecto a las que actúan sobre el virus replicativamente activo sin la presencia del tratamiento ARV. Sin embargo, las conclusiones de estos análisis están limitadas por el pequeño número de secuencias de virus bajo tratamiento disponibles y deberán confirmarse con estudios a largo plazo en cohortes grandes de individuos en tratamiento ARV.

El presente estudio confirmó observaciones previas de alelos HLA asociados de manera consistente a avance rápido o lento de la infección por VIH hacia SIDA en el contexto de la población mexicana. Este es el caso del efecto protector de los alelos B\*57, B\*58 y B\*51 así como el efecto de riesgo de B\*35. Adicionalmente, nuestro estudio sugirió la existencia de factores genéticos específicos de la población mexicana que podrían estar asociados a un avance lento o rápido hacia SIDA, cuyo papel en la infección tendrá que evaluarse más a fondo con estudios en cohortes de individuos en etapas tempranas de la infección con tipificación HLA a alta resolución disponible. Este es el caso de B\*40 y Cw\*15, asociados presuntamente con cargas virales bajas y avance lento hacia SIDA y B\*44 y B\*48, posiblemente asociados a cargas virales altas y a una rápida progresión hacia SIDA. El papel del altamente frecuente alelo B\*39 como un factor de riesgo para la adquisición de la infección por VIH en el contexto de la población mexicana, también requerirá estudios confirmatorios posteriores.

## **9. PERSPECTIVAS**

Los resultados del presente estudio derivan una serie de nuevas preguntas que originan oportunidades de investigación importantes para el entendimiento del papel de la respuesta inmune definida por HLA en la evolución poblacional del VIH en el contexto de la población mexicana y a nivel universal. Entre los puntos que merecen especial atención se encuentran los siguientes:

- Comprobar experimentalmente la existencia de asociaciones HLA – codón de VIH exclusivas en la población mexicana mediante estudios funcionales de respuestas de CTLs restringidas por los alelos de interés.
- Estudiar la existencia de posibles diferencias en la evolución poblacional del virus dentro de la gran diversidad genética de la población mexicana, inicialmente comparando la evolución dirigida por HLA en virus circulantes en el norte, centro y sur del país, cuyas poblaciones mayoritarias se caracterizan por tener rasgos inmunogenéticos distintos.
- Caracterizar y estudiar el papel de las respuestas restringidas por B\*39 en la evolución del VIH en el contexto de la población mexicana, así como su posible contribución como respuestas de riesgo para el avance de la enfermedad por el virus asociándose a una posible adaptación poblacional del VIH mediada por HLA a respuestas restringidas por este alelo altamente frecuente en los mexicanos.
- Caracterizar las respuestas restringidas por B\*35 en la población mexicana, analizando la diversidad alélica de este grupo HLA en los mexicanos, determinando posibles asociaciones con valores altos o bajos de carga viral en nuestra población y comparando estas asociaciones con las observadas en otras poblaciones, tratando de explicar las observaciones mediante estudios funcionales.
- Asociar correlativamente la expresión de alelos HLA específicos definidos a alta resolución con valores significativamente más altos o más bajos de carga viral en individuos en etapas tempranas de la infección, con el fin de determinar respuestas de protección y de riesgo hacia el avance de la infección por VIH.
- Evaluar detalladamente la influencia de la terapia ARV en la evolución mediada por HLA del VIH a nivel poblacional mediante estudios longitudinales en cohortes grandes de individuos bajo tratamiento ARV.
- Explorar la posibilidad de interacción de presiones selectivas que actúan sobre sitios compartidos del virus con el fin de buscar sitios de posible acorralamiento viral, que permitan limitar la variabilidad del virus mediante presiones selectivas opuestas.
- Estudiar el papel de otras ramas de la respuesta inmune en la evolución del virus a nivel poblacional, enfocándose en especial en la respuesta de células asesinas naturales (NK) y sus receptores KIR que tienen sus ligandos naturales en los alelos HLA. Las respuestas de células NK son importantes en el control inicial de la viremia, y el estudiar las características funcionales y genéticas específicas de esta respuesta en la población mexicana enriquecerá grandemente nuestro entendimiento del impacto e la respuesta inmune innata en la población mexicana.

Algunos de estos puntos ya están siendo abordados por nuestro grupo de investigación, y la exploración de otros ya se ha iniciado en colaboración con grupos importantes de investigación del mundo, dedicados al estudio del impacto de las respuestas restringidas por HLA sobre la evolución del VIH a nivel poblacional.

## 10. REFERENCIAS

1. **AIDS Epidemic Update.** [monografía en internet] [Disponible en: <http://www.unaids.org/en/KnowledgeCentre/HIVData/EpiUpdate/EpiUpdArchive/2007/> ]
2. **AVERT** [<http://www.avert.org/worldstats.html>]
3. **HIV vaccine failure prompts Merck to halt trial.** *Nature* 2007, **449**:390.
4. Cohen J: **HIV/AIDS: Latin America & Caribbean. Mexico & Central America.** *Science* 2006, **313**:477.
5. Soto RJ, Ghee AE, Nunez CA, Mayorga R, Tapia KA, Astete SG, Hughes JP, Buffardi AL, Holte SE, Holmes KK: **Sentinel surveillance of sexually transmitted infections/HIV and risk behaviors in vulnerable populations in 5 Central American countries.** *J Acquir Immune Defic Syndr* 2007, **46**:101-111.
6. Calleja JM, Walker N, Cuchi P, Lazzari S, Ghys PD, Zacarias F: **Status of the HIV/AIDS epidemic and methods to monitor it in the Latin America and Caribbean region.** *Aids* 2002, **16 Suppl 3**:S3-12.
7. Soto-Ramirez LE: **HIV/AIDS in Latin America.** *Science* 2008, **321**:465.
8. Cohen J: **HIV/AIDS: Latin America & Caribbean. Mexico: prevention programs target migrants.** *Science* 2006, **313**:478-479.
9. Cohen J: **HIV/AIDS: Latin America & Caribbean. Mexico: land of extremes: prevention and care range from bold to bleak.** *Science* 2006, **313**:477-479.
10. Magis-Rodriguez C, Gayet C, Negroni M, Leyva R, Bravo-Garcia E, Uribe P, Bronfman M: **Migration and AIDS in Mexico: an overview based on recent evidence.** *J Acquir Immune Defic Syndr* 2004, **37 Suppl 4**:S215-226.
11. **CONASIDA, Panorama Epidemiológico del VIH/SIDA en México** [<http://www.salud.gob.mx/conasida/>]
12. Magis-Rodríguez C, Hernández-Avila M: **Epidemiología del SIDA en México.** In *25 Años de SIDA en México Logros, desaciertos y retos.* First edition. Edited by Córdova-Villalobos J, Ponce de León-Rosales S, Valdespino J. Cuernavaca, Morelos, México: Instituto Nacional de Salud Pública; 2008: 85-103
13. Magis-Rodríguez C, Gayet C, Manrique B: *Hombres que tienen sexo con hombres, mujeres trabajadoras sexuales, hombres trabajadores sexuales y usuarios de drogas inyectables. Informe técnico.* México: CENSIDA; 2005.
14. Bucardo J, Brouwer KC, Magis-Rodriguez C, Ramos R, Fraga M, Perez SG, Patterson TL, Strathdee SA: **Historical trends in the production and consumption of illicit drugs in Mexico: implications for the prevention of blood borne infections.** *Drug Alcohol Depend* 2005, **79**:281-293.
15. Patterson TL, Semple SJ, Fraga M, Bucardo J, de la Torre A, Salazar J, Orozovich P, Staines H, Amaro H, Magis-Rodriguez C, Strathdee SA: **Comparison of sexual and drug use behaviors between female sex workers in Tijuana and Ciudad Juarez, Mexico.** *Subst Use Misuse* 2006, **41**:1535-1549.
16. Sanchez MA, Lemp GF, Magis-Rodriguez C, Bravo-Garcia E, Carter S, Ruiz JD: **The epidemiology of HIV among Mexican migrants and recent immigrants in California and Mexico.** *J Acquir Immune Defic Syndr* 2004, **37 Suppl 4**:S204-214.

17. Freed EO, Martin MA: **HIVs and their replication.** In *Fields Virology. Volume Two.* Fifth edition. Edited by Knipe DM, Howley PM. Philadelphia: Lippincott Williams & Wilkins; 2007: 2107-2185
18. Levy J: *HIV and the Pathogenesis of AIDS.* Third edn. Washington, DC: ASM Press; 2007.
19. Avila-Rios S, Reyes-Teran G: **Virología del VIH: Buscando nuevas estrategias antirretrovirales.** In *25 Años de SIDA en México.* First edition. Cuernavaca, Morelos, México: Instituto Nacional de Salud Pública; 2008: 135-160.[Córdova-Villalobos J, Ponce de León-Rosales S, Valdespino J (Series Editor)]
20. Liu J, Bartesaghi A, Borgnia MJ, Sapiro G, Subramaniam S: **Molecular architecture of native HIV-1 gp120 trimers.** *Nature* 2008, **455**:109-113.
21. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA: **Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody.** *Nature* 1998, **393**:648-659.
22. Greene WC, Peterlin BM: **Charting HIV's remarkable voyage through the cell: Basic science as a passport to future therapy.** *Nat Med* 2002, **8**:673-680.
23. Scarlatti G, Tresoldi E, Bjorndal A, Fredriksson R, Colognesi C, Deng HK, Malnati MS, Plebani A, Siccardi AG, Littman DR, et al: **In vivo evolution of HIV-1 co-receptor usage and sensitivity to chemokine-mediated suppression.** *Nat Med* 1997, **3**:1259-1265.
24. Flexner C: **HIV drug development: the next 25 years.** *Nat Rev Drug Discov* 2007, **6**:959-966.
25. Berger EA, Murphy PM, Farber JM: **Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease.** *Annu Rev Immunol* 1999, **17**:657-700.
26. Chan DC, Kim PS: **HIV entry and its inhibition.** *Cell* 1998, **93**:681-684.
27. Barbaro G, Scozzafava A, Mastrolorenzo A, Supuran CT: **Highly active antiretroviral therapy: current state of the art, new agents and their pharmacological interactions useful for improving therapeutic outcome.** *Curr Pharm Des* 2005, **11**:1805-1843.
28. Geijtenbeek TB, Kwon DS, Torensma R, van Vliet SJ, van Duijnhoven GC, Middel J, Cornelissen IL, Nottet HS, KewalRamani VN, Littman DR, et al: **DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells.** *Cell* 2000, **100**:587-597.
29. Arthos J, Cicala C, Martinelli E, Macleod K, Van Ryk D, Wei D, Xiao Z, Veenstra TD, Conrad TP, Lempicki RA, et al: **HIV-1 envelope protein binds to and signals through integrin alpha4beta7, the gut mucosal homing receptor for peripheral T cells.** *Nat Immunol* 2008, **9**:301-309.
30. Fackler OT, Alcover A, Schwartz O: **Modulation of the immunological synapse: a key to HIV-1 pathogenesis?** *Nat Rev Immunol* 2007, **7**:310-317.
31. Arhel NJ, Souquere-Besse S, Munier S, Souque P, Guadagnini S, Rutherford S, Prevost MC, Allen TD, Charneau P: **HIV-1 DNA Flap formation promotes uncoating of the pre-integration complex at the nuclear pore.** *Embo J* 2007, **26**:3025-3037.
32. Gallay P, Hope T, Chin D, Trono D: **HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway.** *Proc Natl Acad Sci U S A* 1997, **94**:9825-9830.
33. Bukrinsky MI, Haggerty S, Dempsey MP, Sharova N, Adzhubel A, Spitz L, Lewis P, Goldfarb D, Emerman M, Stevenson M: **A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells.** *Nature* 1993, **365**:666-669.
34. Heinzinger NK, Bukrinsky MI, Haggerty SA, Ragland AM, Kewalramani V, Lee MA, Gendelman HE, Ratner L, Stevenson M, Emerman M: **The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells.** *Proc Natl Acad Sci U S A* 1994, **91**:7311-7315.

35. Marshall HM, Ronen K, Berry C, Llano M, Sutherland H, Saenz D, Bickmore W, Poeschla E, Bushman FD: **Role of PSIP1/LEDGF/p75 in Lentiviral Infectivity and Integration Targeting.** *PLoS ONE* 2007, **2**:e1340.
36. Maertens GN, Cherepanov P, Engelman A: **Transcriptional co-activator p75 binds and tethers the Myc-interacting protein JPO2 to chromatin.** *J Cell Sci* 2006, **119**:2563-2571.
37. Bradley CM, Craigie R: **Seeing is believing: structure of the catalytic domain of HIV-1 integrase in complex with human LEDGF/p75.** *Proc Natl Acad Sci U S A* 2005, **102**:17543-17544.
38. Llano M, Saenz DT, Meehan A, Wongthida P, Peretz M, Walker WH, Teo W, Poeschla EM: **An essential role for LEDGF/p75 in HIV integration.** *Science* 2006, **314**:461-464.
39. Turlure F, Maertens G, Rahman S, Cherepanov P, Engelman A: **A tripartite DNA-binding element, comprised of the nuclear localization signal and two AT-hook motifs, mediates the association of LEDGF/p75 with chromatin in vivo.** *Nucleic Acids Res* 2006, **34**:1653-1675.
40. Rahman S, Lu R, Vandegraaff N, Cherepanov P, Engelman A: **Structure-based mutagenesis of the integrase-LEDGF/p75 interface uncouples a strict correlation between in vitro protein binding and HIV-1 fitness.** *Virology* 2006.
41. Cherepanov P, Ambrosio AL, Rahman S, Ellenberger T, Engelman A: **Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75.** *Proc Natl Acad Sci U S A* 2005, **102**:17308-17313.
42. Cherepanov P, Sun ZY, Rahman S, Maertens G, Wagner G, Engelman A: **Solution structure of the HIV-1 integrase-binding domain in LEDGF/p75.** *Nat Struct Mol Biol* 2005, **12**:526-532.
43. Emiliani S, Mousnier A, Busschots K, Maroun M, Van Maele B, Tempe D, Vandekerckhove L, Moisant F, Ben-Slama L, Witvrouw M, et al: **Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication.** *J Biol Chem* 2005, **280**:25517-25523.
44. Lataillade M, Chiarella J, Kozal MJ: **Natural polymorphism of the HIV-1 integrase gene and mutations associated with integrase inhibitor resistance.** *Antivir Ther* 2007, **12**:563-570.
45. Peterlin BM, Trono D: **Hide, shield and strike back: how HIV-infected cells avoid immune eradication.** *Nat Rev Immunol* 2003, **3**:97-107.
46. Orenstein JM, Meltzer MS, Phipps T, Gendelman HE: **Cytoplasmic assembly and accumulation of human immunodeficiency virus types 1 and 2 in recombinant human colony-stimulating factor-1-treated human monocytes: an ultrastructural study.** *J Virol* 1988, **62**:2578-2586.
47. Pelchen-Matthews A, Raposo G, Marsh M: **Endosomes, exosomes and Trojan viruses.** *Trends Microbiol* 2004, **12**:310-316.
48. Campbell SM, Crowe SM, Mak J: **Lipid rafts and HIV-1: from viral entry to assembly of progeny virions.** *J Clin Virol* 2001, **22**:217-227.
49. Li F, Goila-Gaur R, Salzwedel K, Kilgore NR, Reddick M, Matallana C, Castillo A, Zoumplis D, Martin DE, Orenstein JM, et al: **PA-457: a potent HIV inhibitor that disrupts core condensation by targeting a late step in Gag processing.** *Proc Natl Acad Sci U S A* 2003, **100**:13555-13560.
50. Sakalian M, McMurtrey CP, Deeg FJ, Maloy CW, Li F, Wild CT, Salzwedel K: **3-O-(3',3'-dimethylsuccinyl) betulinic acid inhibits maturation of the human immunodeficiency virus type 1 Gag precursor assembled in vitro.** *J Virol* 2006, **80**:5716-5722.
51. Salzwedel K, Martin DE, Sakalian M: **Maturation inhibitors: a new therapeutic class targets the virus structure.** *AIDS Rev* 2007, **9**:162-172.
52. Ganser-Pornillos BK, Cheng A, Yeager M: **Structure of full-length HIV-1 CA: a model for the mature capsid lattice.** *Cell* 2007, **131**:70-79.

53. Ternois F, Sticht J, Duquerroy S, Krausslich HG, Rey FA: **The HIV-1 capsid protein C-terminal domain in complex with a virus assembly inhibitor.** *Nat Struct Mol Biol* 2005, **12**:678-682.
54. Tang C, Loeliger E, Kinde I, Kyere S, Mayo K, Barklis E, Sun Y, Huang M, Summers MF: **Antiviral inhibition of the HIV-1 capsid protein.** *J Mol Biol* 2003, **327**:1013-1020.
55. Levin JG, Guo J, Rouzina I, Musier-Forsyth K: **Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism.** *Prog Nucleic Acid Res Mol Biol* 2005, **80**:217-286.
56. Gottlinger HG, Dorfman T, Sodroski JG, Haseltine WA: **Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release.** *Proc Natl Acad Sci U S A* 1991, **88**:3195-3199.
57. von Schwedler UK, Stuchell M, Muller B, Ward DM, Chung HY, Morita E, Wang HE, Davis T, He GP, Cimbora DM, et al: **The protein network of HIV budding.** *Cell* 2003, **114**:701-713.
58. Garrus JE, von Schwedler UK, Pornillos OW, Morham SG, Zavitz KH, Wang HE, Wettstein DA, Stray KM, Cote M, Rich RL, et al: **Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding.** *Cell* 2001, **107**:55-65.
59. Strack B, Calistri A, Craig S, Popova E, Gottlinger HG: **AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding.** *Cell* 2003, **114**:689-699.
60. Le Rouzic E, Benichou S: **The Vpr protein from HIV-1: distinct roles along the viral life cycle.** *Retrovirology* 2005, **2**:11.
61. Montal M: **Structure-function correlates of Vpu, a membrane protein of HIV-1.** *FEBS Lett* 2003, **552**:47-53.
62. Neil SJ, Zang T, Bieniasz PD: **Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu.** *Nature* 2008, **451**:425-430.
63. Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, Neuberger MS, Malim MH: **DNA deamination mediates innate immunity to retroviral infection.** *Cell* 2003, **113**:803-809.
64. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D: **Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts.** *Nature* 2003, **424**:99-103.
65. Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L: **The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA.** *Nature* 2003, **424**:94-98.
66. Pace C, Keller J, Nolan D, James I, Gaudieri S, Moore C, Mallal S: **Population level analysis of human immunodeficiency virus type 1 hypermutation and its relationship with APOBEC3G and vif genetic variation.** *J Virol* 2006, **80**:9259-9269.
67. Conticello SG, Harris RS, Neuberger MS: **The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G.** *Curr Biol* 2003, **13**:2009-2013.
68. Marin M, Rose KM, Kozak SL, Kabat D: **HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation.** *Nat Med* 2003, **9**:1398-1403.
69. Yu X, Yu Y, Liu B, Luo K, Kong W, Mao P, Yu XF: **Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex.** *Science* 2003, **302**:1056-1060.
70. Mehle A, Thomas ER, Rajendran KS, Gabuzda D: **A zinc-binding region in Vif binds Cul5 and determines cullin selection.** *J Biol Chem* 2006, **281**:17259-17265.
71. Xiao Z, Ehrlich E, Yu Y, Luo K, Wang T, Tian C, Yu XF: **Assembly of HIV-1 Vif-Cul5 E3 ubiquitin ligase through a novel zinc-binding domain-stabilized hydrophobic interface in Vif.** *Virology* 2006, **349**:290-299.

72. Mehle A, Goncalves J, Santa-Marta M, McPike M, Gabuzda D: **Phosphorylation of a novel SOCS-box regulates assembly of the HIV-1 Vif-Cul5 complex that promotes APOBEC3G degradation.** *Genes Dev* 2004, **18**:2861-2866.
73. Tian C, Yu X, Zhang W, Wang T, Xu R, Yu XF: **Differential requirement for conserved tryptophans in human immunodeficiency virus type 1 Vif for the selective suppression of APOBEC3G and APOBEC3F.** *J Virol* 2006, **80**:3112-3115.
74. Russell RA, Pathak VK: **Identification of two distinct human immunodeficiency virus type 1 Vif determinants critical for interactions with human APOBEC3G and APOBEC3F.** *J Virol* 2007, **81**:8201-8210.
75. Jeffrey Fessel W: **A new approach to an AIDS vaccine: creating antibodies to HIV vif will enable apobec3G to turn HIV-infection into a benign problem.** *Med Hypotheses* 2005, **64**:261-263.
76. Huthoff H, Malim MH: **Identification of amino acid residues in APOBEC3G required for regulation by human immunodeficiency virus type 1 Vif and Virion encapsidation.** *J Virol* 2007, **81**:3807-3815.
77. Berkhout B, Grigoriev A, Bakker M, Lukashov VV: **Codon and amino acid usage in retroviral genomes is consistent with virus-specific nucleotide pressure.** *AIDS Res Hum Retroviruses* 2002, **18**:133-141.
78. Yu Q, Konig R, Pillai S, Chiles K, Kearney M, Palmer S, Richman D, Coffin JM, Landau NR: **Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome.** *Nat Struct Mol Biol* 2004, **11**:435-442.
79. Churchill MJ, Rhodes DI, Learmont JC, Sullivan JS, Wesselingh SL, Cooke IR, Deacon NJ, Gorry PR: **Longitudinal analysis of human immunodeficiency virus type 1 nef/long terminal repeat sequences in a cohort of long-term survivors infected from a single source.** *J Virol* 2006, **80**:1047-1052.
80. Deacon NJ, Tsykin A, Solomon A, Smith K, Ludford-Menting M, Hooker DJ, McPhee DA, Greenway AL, Ellett A, Chatfield C, et al: **Genomic structure of an attenuated quasi species of HIV-1 from a blood transfusion donor and recipients.** *Science* 1995, **270**:988-991.
81. Roeth JF, Collins KL: **Human immunodeficiency virus type 1 Nef: adapting to intracellular trafficking pathways.** *Microbiol Mol Biol Rev* 2006, **70**:548-563.
82. Piguet V, Chen YL, Mangasarian A, Foti M, Carpentier JL, Trono D: **Mechanism of Nef-induced CD4 endocytosis: Nef connects CD4 with the mu chain of adaptor complexes.** *Embo J* 1998, **17**:2472-2481.
83. Piguet V, Gu F, Foti M, Demarex N, Gruenberg J, Carpentier JL, Trono D: **Nef-induced CD4 degradation: a diacidic-based motif in Nef functions as a lysosomal targeting signal through the binding of beta-COP in endosomes.** *Cell* 1999, **97**:63-73.
84. Lama J, Mangasarian A, Trono D: **Cell-surface expression of CD4 reduces HIV-1 infectivity by blocking Env incorporation in a Nef- and Vpu-inhibitable manner.** *Curr Biol* 1999, **9**:622-631.
85. Schwartz O, Marechal V, Le Gall S, Lemonnier F, Heard JM: **Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein.** *Nat Med* 1996, **2**:338-342.
86. Roeth JF, Williams M, Kasper MR, Filzen TM, Collins KL: **HIV-1 Nef disrupts MHC-I trafficking by recruiting AP-1 to the MHC-I cytoplasmic tail.** *J Cell Biol* 2004, **167**:903-913.
87. Blagoveshchenskaya AD, Thomas L, Feliciangeli SF, Hung CH, Thomas G: **HIV-1 Nef downregulates MHC-I by a PACS-1- and PI3K-regulated ARF6 endocytic pathway.** *Cell* 2002, **111**:853-866.
88. Ali A, Ng HL, Dagarag MD, Yang OO: **Evasion of cytotoxic T lymphocytes is a functional constraint maintaining HIV-1 Nef expression.** *Eur J Immunol* 2005, **35**:3221-3228.



89. Ali A, Pillai S, Ng H, Lubong R, Richman DD, Jamieson BD, Ding Y, McElrath MJ, Guatelli JC, Yang OO: **Broadly increased sensitivity to cytotoxic T lymphocytes resulting from Nef epitope escape mutations.** *J Immunol* 2003, **171**:3999-4005.
90. Yang OO, Nguyen PT, Kalams SA, Dorfman T, Gottlinger HG, Stewart S, Chen IS, Threlkeld S, Walker BD: **Nef-mediated resistance of human immunodeficiency virus type 1 to antiviral cytotoxic T lymphocytes.** *J Virol* 2002, **76**:1626-1631.
91. Swigut T, Alexander L, Morgan J, Lifson J, Mansfield KG, Lang S, Johnson RP, Skowronski J, Desrosiers R: **Impact of Nef-mediated downregulation of major histocompatibility complex class I on immune response to simian immunodeficiency virus.** *J Virol* 2004, **78**:13335-13344.
92. Calugi G, Montella F, Favalli C, Benedetto A: **The Entire Genome of a nef-Deleted Strain of Human Immunodeficiency Virus Type 1 recovered 20 years after Primary Infection: large pool of env-deleted proviruses.** *J Virol* 2006.
93. Cohen GB, Gandhi RT, Davis DM, Mandelboim O, Chen BK, Strominger JL, Baltimore D: **The selective downregulation of class I major histocompatibility complex proteins by HIV-1 protects HIV-infected cells from NK cells.** *Immunity* 1999, **10**:661-671.
94. Swigut T, Shohdy N, Skowronski J: **Mechanism for down-regulation of CD28 by Nef.** *Embo J* 2001, **20**:1593-1604.
95. Bell I, Ashman C, Maughan J, Hooker E, Cook F, Reinhart TA: **Association of simian immunodeficiency virus Nef with the T-cell receptor (TCR) zeta chain leads to TCR down-modulation.** *J Gen Virol* 1998, **79 ( Pt 11)**:2717-2727.
96. Stumptner-Cuvelette P, Morchoisne S, Dugast M, Le Gall S, Raposo G, Schwartz O, Benaroch P: **HIV-1 Nef impairs MHC class II antigen presentation and surface expression.** *Proc Natl Acad Sci U S A* 2001, **98**:12144-12149.
97. Markowitz M, Mohri H, Mehandru S, Shet A, Berry L, Kalyanaraman R, Kim A, Chung C, Jean-Pierre P, Horowitz A, et al: **Infection with multidrug resistant, dual-tropic HIV-1 and rapid progression to AIDS: a case report.** *Lancet* 2005, **365**:1031-1038.
98. Kuritzkes DR, Walker BD: **HIV-1: Pathogenesis, Clinical Manifestations and Treatment.** In *Fields Virology. Volume Two.* Fifth edition. Edited by Knipe DM, Howley PM. Philadelphia: Lippincott Williams & Wilkins; 2007: 2187-2214
99. Lu W, Arraes LC, Ferreira WT, Andrieu JM: **Therapeutic dendritic-cell vaccine for chronic HIV-1 infection.** *Nat Med* 2004, **10**:1359-1365.
100. Grossman Z, Meier-Schellersheim M, Paul WE, Picker LJ: **Pathogenesis of HIV infection: what the virus spares is as important as what it destroys.** *Nat Med* 2006, **12**:289-295.
101. Mattapallil JJ, Douek DC, Hill B, Nishimura Y, Martin M, Roederer M: **Massive infection and loss of memory CD4+ T cells in multiple tissues during acute SIV infection.** *Nature* 2005, **434**:1093-1097.
102. Brenchley JM, Price DA, Schacker TW, Asher TE, Silvestri G, Rao S, Kazzaz Z, Bornstein E, Lambotte O, Altmann D, et al: **Microbial translocation is a cause of systemic immune activation in chronic HIV infection.** *Nat Med* 2006, **12**:1365-1371.
103. Morgan D, Mahe C, Mayanja B, Okongo JM, Lubega R, Whitworth JA: **HIV-1 infection in rural Africa: is there a difference in median time to AIDS and survival compared with that in industrialized countries?** *Aids* 2002, **16**:597-603.
104. Llibre JM, Falco V, Tural C, Negredo E, Pineda JA, Munoz J, Ortega E, Videla S, Sirera G, Martinez E, et al: **The changing face of HIV/AIDS in treated patients.** *Curr HIV Res* 2009, **7**:365-377.
105. Lal RB, Chakrabarti S, Yang C: **Impact of genetic diversity of HIV-1 on diagnosis, antiretroviral therapy & vaccine development.** *Indian J Med Res* 2005, **121**:287-314.
106. Hemelaar J, Gouws E, Ghys PD, Osmanov S: **Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004.** *Aids* 2006, **20**:W13-23.

107. Frange P, Galimand J, Vidal N, Goujard C, Deveau C, Souala F, Peeters M, Meyer L, Rouzioux C, Chaix ML: **New and old complex recombinant HIV-1 strains among patients with primary infection in 1996-2006 in France: the French ANRS CO6 primo cohort study.** *Retrovirology* 2008, **5**:69.
108. Buonaguro L, Tornesello ML, Buonaguro FM: **Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications.** *J Virol* 2007, **81**:10209-10219.
109. Pyne MT, Holzmayer V, Hackett J, Hillyard DR: **Analysis of HIV subtype prevalence and geographic distribution in a large U. S. population.** In *16th Conference on Retroviruses and Opportunistic Infections, Abstract B-106, Poster 292, Session 59; Montreal, Canada.* 2009
110. Eyzaguirre L, Brouwer KC, Nadai Y, Patterson TL, Ramos R, Firestone Cruz M, Orozovich P, Strathdee SA, Carr JK: **First molecular surveillance report of HIV type 1 in injecting drug users and female sex workers along the U.S.-Mexico border.** *AIDS Res Hum Retroviruses* 2007, **23**:331-334.
111. Rivera-Morales LG, Novitsky VA, Trujillo JR, Lavallo-Montalvo C, Cano-Dominguez C, Ramos-Jimenez J, Jimenez-Rios E, Flores-Flores L, Lopez-Guillen P, Gilbert P, et al: **The molecular epidemiology of HIV type 1 of men in Mexico.** *AIDS Res Hum Retroviruses* 2001, **17**:87-92.
112. Geretti AM: **HIV-1 subtypes: epidemiology and significance for HIV management.** *Curr Opin Infect Dis* 2006, **19**:1-7.
113. Van Heuverswyn F, Peeters M: **The Origins of HIV and Implications for the Global Epidemic.** *Curr Infect Dis Rep* 2007, **9**:338-346.
114. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, Cummins LB, Arthur LO, Peeters M, Shaw GM, et al: **Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*.** *Nature* 1999, **397**:436-441.
115. Keele BF, Van Heuverswyn F, Li Y, Bailes E, Takehisa J, Santiago ML, Bibollet-Ruche F, Chen Y, Wain LV, Liegeois F, et al: **Chimpanzee reservoirs of pandemic and nonpandemic HIV-1.** *Science* 2006, **313**:523-526.
116. Wain LV, Bailes E, Bibollet-Ruche F, Decker JM, Keele BF, Van Heuverswyn F, Li Y, Takehisa J, Ngole EM, Shaw GM, et al: **Adaptation of HIV-1 to its human host.** *Mol Biol Evol* 2007, **24**:1853-1860.
117. Takehisa J, Kraus MH, Ayouba A, Bailes E, Van Heuverswyn F, Decker JM, Li Y, Rudicell RS, Learn GH, Neel C, et al: **Origin and biology of simian immunodeficiency virus in wild-living western gorillas.** *J Virol* 2009, **83**:1635-1648.
118. Kantor R, Katzenstein DA, Efron B, Carvalho AP, Wynhoven B, Cane P, Clarke J, Sirivichayakul S, Soares MA, Snoeck J, et al: **Impact of HIV-1 subtype and antiretroviral therapy on protease and reverse transcriptase genotype: results of a global collaboration.** *PLoS Med* 2005, **2**:e112.
119. van de Vijver DA, Wensing AM, Angarano G, Asjo B, Balotta C, Boeri E, Camacho R, Chaix ML, Costagliola D, De Luca A, et al: **The calculated genetic barrier for antiretroviral drug resistance substitutions is largely similar for different HIV-1 subtypes.** *J Acquir Immune Defic Syndr* 2006, **41**:352-360.
120. Wainberg MA: **HIV-1 subtype distribution and the problem of drug resistance.** *Aids* 2004, **18 Suppl 3**:S63-68.
121. Montes B, Vergne L, Peeters M, Reynes J, Delaporte E, Segondy M: **Comparison of drug resistance mutations and their interpretation in patients infected with non-B HIV-1 variants and matched patients infected with HIV-1 subtype B.** *J Acquir Immune Defic Syndr* 2004, **35**:329-336.
122. **Evidence of a decline in transmitted HIV-1 drug resistance in the United Kingdom.** *Aids* 2007, **21**:1035-1039.
123. Gale CV, Yirrell DL, Campbell E, Van der Paal L, Grosskurth H, Kaleebu P: **Genotypic variation in the pol gene of HIV type 1 in an antiretroviral treatment-**

- naive population in rural southwestern Uganda. *AIDS Res Hum Retroviruses* 2006, **22**:985-992.
124. Han X, Zhang M, Dai D, Wang Y, Zhang Z, Liu J, Geng W, Jiang Y, Takebe Y, Shang H: **Genotypic resistance mutations to antiretroviral drugs in treatment-naive HIV/AIDS patients living in Liaoning Province, China: baseline prevalence and subtype-specific difference.** *AIDS Res Hum Retroviruses* 2007, **23**:357-364.
  125. Palma AC, Araujo F, Duque V, Borges F, Paixao MT, Camacho R: **Molecular epidemiology and prevalence of drug resistance-associated mutations in newly diagnosed HIV-1 patients in Portugal.** *Infect Genet Evol* 2007, **7**:391-398.
  126. Perez L, Alvarez LP, Carmona R, Aragonés C, Delgado E, Thomson MM, Gonzalez Z, Contreras G, Perez J, Najera R: **Genotypic resistance to antiretroviral drugs in patients infected with several HIV type 1 genetic forms in Cuba.** *AIDS Res Hum Retroviruses* 2007, **23**:407-414.
  127. Varella RB, Ferreira SB, de Castro MB, Zalis MG, Tavares MD: **Human immunodeficiency virus type 1 protease and reverse transcriptase mutation patterns among treatment-naive patients in different stages of infection in Rio de Janeiro, Brazil.** *J Med Virol* 2007, **79**:1033-1039.
  128. Vergne L, Stuyver L, Van Houtte M, Butel C, Delaporte E, Peeters M: **Natural polymorphism in protease and reverse transcriptase genes and in vitro antiretroviral drug susceptibilities of non-B HIV-1 strains from treatment-naive patients.** *J Clin Virol* 2006, **36**:43-49.
  129. Malim MH, Emerman M: **HIV-1 sequence variation: drift, shift, and attenuation.** *Cell* 2001, **104**:469-472.
  130. Letvin NL: **Progress and obstacles in the development of an AIDS vaccine.** *Nat Rev Immunol* 2006, **6**:930-939.
  131. Carlson JM, Brumme ZL: **HIV evolution in response to HLA-restricted CTL selection pressures: a population-based perspective.** *Microbes Infect* 2008, **10**:455-461.
  132. Janeway C, Travers P, Walport M: *Immunobiology*. 7a. edn. New York: Garland Science; 2008.
  133. Abbas A, Litchman A: *Inmunología Celular y Molecular*. Madrid: Elsevier Imprint; 2004.
  134. Goulder PJ, Watkins DI: **Impact of MHC class I diversity on immune control of immunodeficiency virus replication.** *Nat Rev Immunol* 2008, **8**:619-630.
  135. Srivasta R, Ram B, Tyle P: *Immunogenetics of the Major Histocompatibility Complex*. New York, USA: VCH Publishers; 2007.
  136. Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X, Salazar-Gonzalez JF, Salazar MG, Kilby JM, Saag MS, et al: **Antibody neutralization and escape by HIV-1.** *Nature* 2003, **422**:307-312.
  137. Letvin NL, Walker BD: **Immunopathogenesis and immunotherapy in AIDS virus infections.** *Nat Med* 2003, **9**:861-866.
  138. Goulder PJ, Watkins DI: **HIV and SIV CTL escape: implications for vaccine design.** *Nat Rev Immunol* 2004, **4**:630-640.
  139. Koup RA, Safrit JT, Cao Y, Andrews CA, McLeod G, Borkowsky W, Farthing C, Ho DD: **Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome.** *J Virol* 1994, **68**:4650-4655.
  140. Jin X, Bauer DE, Tuttleton SE, Lewin S, Gettie A, Blanchard J, Irwin CE, Safrit JT, Mittler J, Weinberger L, et al: **Dramatic rise in plasma viremia after CD8(+) T cell depletion in simian immunodeficiency virus-infected macaques.** *J Exp Med* 1999, **189**:991-998.
  141. Matano T, Shibata R, Siemon C, Connors M, Lane HC, Martin MA: **Administration of an anti-CD8 monoclonal antibody interferes with the clearance of chimeric simian/human immunodeficiency virus during primary infections of rhesus macaques.** *J Virol* 1998, **72**:164-169.

142. Schmitz JE, Kuroda MJ, Santra S, Sasseville VG, Simon MA, Lifton MA, Racz P, Tenner-Racz K, Dalesandro M, Scallon BJ, et al: **Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes.** *Science* 1999, **283**:857-860.
143. Carrington M, O'Brien SJ: **The influence of HLA genotype on AIDS.** *Annu Rev Med* 2003, **54**:535-551.
144. Kiepiela P, Leslie AJ, Honeyborne I, Ramduth D, Thobakgale C, Chetty S, Rathnavalu P, Moore C, Pfafferott KJ, Hilton L, et al: **Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA.** *Nature* 2004, **432**:769-775.
145. Scherer A, Frater J, Oxenius A, Agudelo J, Price DA, Gunthard HF, Barnardo M, Perrin L, Hirschel B, Phillips RE, McLean AR: **Quantifiable cytotoxic T lymphocyte responses and HLA-related risk of progression to AIDS.** *Proc Natl Acad Sci U S A* 2004, **101**:12266-12270.
146. Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, Kaslow R, Buchbinder S, Hoots K, O'Brien SJ: **HLA and HIV-1: heterozygote advantage and B\*35-Cw\*04 disadvantage.** *Science* 1999, **283**:1748-1752.
147. Tang J, Costello C, Keet IP, Rivers C, Leblanc S, Karita E, Allen S, Kaslow RA: **HLA class I homozygosity accelerates disease progression in human immunodeficiency virus type 1 infection.** *AIDS Res Hum Retroviruses* 1999, **15**:317-324.
148. Draenert R, Le Gall S, Pfafferott KJ, Leslie AJ, Chetty P, Brander C, Holmes EC, Chang SC, Feeney ME, Addo MM, et al: **Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection.** *J Exp Med* 2004, **199**:905-915.
149. Goulder PJ, Brander C, Tang Y, Tremblay C, Colbert RA, Addo MM, Rosenberg ES, Nguyen T, Allen R, Trocha A, et al: **Evolution and transmission of stable CTL escape mutations in HIV infection.** *Nature* 2001, **412**:334-338.
150. Phillips RE, Rowland-Jones S, Nixon DF, Gotch FM, Edwards JP, Ogunlesi AO, Elvin JG, Rothbard JA, Bangham CR, Rizza CR, et al.: **Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition.** *Nature* 1991, **354**:453-459.
151. Frahm N, Korber BT, Adams CM, Szinger JJ, Draenert R, Addo MM, Feeney ME, Yusim K, Sango K, Brown NV, et al: **Consistent cytotoxic-T-lymphocyte targeting of immunodominant regions in human immunodeficiency virus across multiple ethnicities.** *J Virol* 2004, **78**:2187-2200.
152. Kiepiela P, Ngumbela K, Thobakgale C, Ramduth D, Honeyborne I, Moodley E, Reddy S, de Pierres C, Mncube Z, Mkhwanazi N, et al: **CD8+ T-cell responses to different HIV proteins have discordant associations with viral load.** *Nat Med* 2007, **13**:46-53.
153. Borrow P, Lewicki H, Wei X, Horwitz MS, Peffer N, Meyers H, Nelson JA, Gairin JE, Hahn BH, Oldstone MB, Shaw GM: **Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus.** *Nat Med* 1997, **3**:205-211.
154. Goulder PJ, Phillips RE, Colbert RA, McAdam S, Ogg G, Nowak MA, Giangrande P, Luzzi G, Morgan B, Edwards A, et al: **Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS.** *Nat Med* 1997, **3**:212-217.
155. Allen TM, Altfeld M, Yu XG, O'Sullivan KM, Lichtenfeld M, Le Gall S, John M, Mothe BR, Lee PK, Kalife ET, et al: **Selection, transmission, and reversion of an antigen-processing cytotoxic T-lymphocyte escape mutation in human immunodeficiency virus type 1 infection.** *J Virol* 2004, **78**:7069-7078.
156. Schneidewind A, Brockman MA, Yang R, Adam RI, Li B, Le Gall S, Rinaldo CR, Craggs SL, Allgaier RL, Power KA, et al: **Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in Gag is associated with a**

- dramatic reduction in human immunodeficiency virus type 1 replication. *J Virol* 2007, 81:12382-12393.**
157. Draenert R, Allen TM, Liu Y, Wrin T, Chappey C, Verrill CL, Sirera G, Eldridge RL, Lahaie MP, Ruiz L, et al: **Constraints on HIV-1 evolution and immunodominance revealed in monozygotic adult twins infected with the same virus. *J Exp Med* 2006, 203:529-539.**
  158. Allen TM, Yu XG, Kalife ET, Reyor LL, Lichtenfeld M, John M, Cheng M, Allgaier RL, Mui S, Frahm N, et al: **De novo generation of escape variant-specific CD8+ T-cell responses following cytotoxic T-lymphocyte escape in chronic human immunodeficiency virus type 1 infection. *J Virol* 2005, 79:12952-12960.**
  159. Friedrich TC, Dodds EJ, Yant LJ, Vojnov L, Rudersdorf R, Cullen C, Evans DT, Desrosiers RC, Mothe BR, Sidney J, et al: **Reversion of CTL escape-variant immunodeficiency viruses in vivo. *Nat Med* 2004, 10:275-281.**
  160. Leslie AJ, Pfafferott KJ, Chetty P, Draenert R, Addo MM, Feeney M, Tang Y, Holmes EC, Allen T, Prado JG, et al: **HIV evolution: CTL escape mutation and reversion after transmission. *Nat Med* 2004, 10:282-289.**
  161. Liu Y, McNevin J, Zhao H, Tebit DM, Troyer RM, McSweyn M, Ghosh AK, Shriner D, Arts EJ, McElrath MJ, Mullins JI: **Evolution of human immunodeficiency virus type 1 cytotoxic T-lymphocyte epitopes: fitness-balanced escape. *J Virol* 2007, 81:12179-12188.**
  162. Peyerl FW, Barouch DH, Letvin NL: **Structural constraints on viral escape from HIV- and SIV-specific cytotoxic T-lymphocytes. *Viral Immunol* 2004, 17:144-151.**
  163. Peyerl FW, Bazick HS, Newberg MH, Barouch DH, Sodroski J, Letvin NL: **Fitness costs limit viral escape from cytotoxic T lymphocytes at a structurally constrained epitope. *J Virol* 2004, 78:13901-13910.**
  164. Edwards CT, Pfafferott KJ, Goulder PJ, Phillips RE, Holmes EC: **Inpatient escape in the A\*0201-restricted epitope SLYNTVATL drives evolution of human immunodeficiency virus type 1 at the population level. *J Virol* 2005, 79:9363-9366.**
  165. Furutsuki T, Hosoya N, Kawana-Tachikawa A, Tomizawa M, Odawara T, Goto M, Kitamura Y, Nakamura T, Kelleher AD, Cooper DA, Iwamoto A: **Frequent transmission of cytotoxic-T-lymphocyte escape mutants of human immunodeficiency virus type 1 in the highly HLA-A24-positive Japanese population. *J Virol* 2004, 78:8437-8445.**
  166. Piontkivska H, Hughes AL: **Patterns of sequence evolution at epitopes for host antibodies and cytotoxic T-lymphocytes in human immunodeficiency virus type 1. *Virus Res* 2006, 116:98-105.**
  167. Moore CB, John M, James IR, Christiansen FT, Witt CS, Mallal SA: **Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science* 2002, 296:1439-1443.**
  168. Ahlenstiel G, Roomp K, Daumer M, Nattermann J, Vogel M, Rockstroh JK, Beerenwinkel N, Kaiser R, Nischalke HD, Sauerbruch T, et al: **Selective pressures of HLA genotypes and antiviral therapy on human immunodeficiency virus type 1 sequence mutation at a population level. *Clin Vaccine Immunol* 2007, 14:1266-1273.**
  169. Brumme ZL, Brumme CJ, Carlson J, Streeck H, John M, Eichbaum Q, Block BL, Baker B, Kadie C, Markowitz M, et al: **Marked epitope and allele-specific differences in rates of mutation in HIV-1 Gag, Pol and Nef CTL epitopes in acute/early HIV-1 infection. *J Virol* 2008.**
  170. Brumme ZL, Brumme CJ, Heckerman D, Korber BT, Daniels M, Carlson J, Kadie C, Bhattacharya T, Chui C, Szinger J, et al: **Evidence of Differential HLA Class I-Mediated Viral Evolution in Functional and Accessory/Regulatory Genes of HIV-1. *PLoS Pathog* 2007, 3:e94.**
  171. Brumme ZL, John M, Brumme CJ, Carlson J, Haubrich R, Riddler S, Swenson L, Tao I, Szeto S, Chan D, et al: **Relationship between HLA class I-driven evolution in**

- Gag, Pol and Nef and clinical markers of HIV disease: a multi-center collaborative study.** In *AIDS Vaccine, Abstract P09-01; Cape Town, South Africa.* 2008
172. Brumme ZL, Tao I, Szeto S, Brumme CJ, Carlson JM, Chan D, Kadie C, Frahm N, Brander C, Walker B, et al: **Human leukocyte antigen-specific polymorphisms in HIV-1 Gag and their association with viral load in chronic untreated infection.** *Aids* 2008, **22**:1277-1286.
  173. Mueller SM, Schaetz B, Eismann K, Bergmann S, Bauerle M, Schmitt-Haendle M, Walter H, Schmidt B, Korn K, Sticht H, et al: **Dual selection pressure by drugs and HLA class I-restricted immune responses on human immunodeficiency virus type 1 protease.** *J Virol* 2007, **81**:2887-2898.
  174. Iversen AK, Stewart-Jones G, Learn GH, Christie N, Sylvester-Hviid C, Armitage AE, Kaul R, Beattie T, Lee JK, Li Y, et al: **Conflicting selective forces affect T cell receptor contacts in an immunodominant human immunodeficiency virus epitope.** *Nat Immunol* 2006, **7**:179-189.
  175. Allen TM, Altfeld M, Geer SC, Kalife ET, Moore C, O'Sullivan K M, Desouza I, Feeney ME, Eldridge RL, Maier EL, et al: **Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution.** *J Virol* 2005, **79**:13239-13249.
  176. Avila-Rios S, Reyes-Teran G, Espinosa E: **Cornering HIV: taking advantage of interactions between selective pressures.** *Med Hypotheses* 2007, **69**:422-431.
  177. Mason RD, Grant MD: **A therapy-related point mutation changes the HLA restriction of an HIV-1 Pol epitope from A2 to B57 and enhances its recognition.** *Aids* 2005, **19**:981-984.
  178. Altfeld M, Kalife ET, Qi Y, Streeck H, Lichterfeld M, Johnston MN, Burgett N, Swartz ME, Yang A, Alter G, et al: **HLA Alleles Associated with Delayed Progression to AIDS Contribute Strongly to the Initial CD8(+) T Cell Response against HIV-1.** *PLoS Med* 2006, **3**:e403.
  179. Streeck H, Lichterfeld M, Alter G, Meier A, Teigen N, Yassine-Diab B, Sidhu HK, Little S, Kelleher A, Routy JP, et al: **Recognition of a defined region within p24 gag by CD8+ T cells during primary human immunodeficiency virus type 1 infection in individuals expressing protective HLA class I alleles.** *J Virol* 2007, **81**:7725-7731.
  180. Kawashima Y, Pfafferott K, Frater J, Matthews P, Payne R, Addo M, Gatanaga H, Fujiwara M, Hachiya A, Koizumi H, et al: **Adaptation of HIV-1 to human leukocyte antigen class I.** *Nature* 2009.
  181. Bhattacharya T, Daniels M, Heckerman D, Foley B, Frahm N, Kadie C, Carlson J, Yusim K, McMahon B, Gaschen B, et al: **Founder effects in the assessment of HIV polymorphisms and HLA allele associations.** *Science* 2007, **315**:1583-1586.
  182. Carlson J, Kadie C, Mallal S, Heckerman D: **Leveraging hierarchical population structure in discrete association studies.** *PLoS ONE* 2007, **2**:e591.
  183. Carlson JM, Brumme ZL, Rousseau CM, Brumme CJ, Matthews P, Kadie C, Mullins JI, Walker BD, Harrigan PR, Goulder PJ, Heckerman D: **Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag.** *PLoS Comput Biol* 2008, **4**:e1000225.
  184. Harrer EG, Bergmann S, Eismann K, Rittmaier M, Goldwisch A, Muller SM, Spriewald BM, Harrer T: **A conserved HLA B13-restricted cytotoxic T lymphocyte epitope in Nef is a dominant epitope in HLA B13-positive HIV-1-infected patients.** *Aids* 2005, **19**:734-735.
  185. Honeyborne I, Prendergast A, Pereyra F, Leslie A, Crawford H, Payne R, Reddy S, Bishop K, Moodley E, Nair K, et al: **Control of human immunodeficiency virus type 1 is associated with HLA-B\*13 and targeting of multiple gag-specific CD8+ T-cell epitopes.** *J Virol* 2007, **81**:3667-3672.
  186. Tang J, Tang S, Lobashevsky E, Myracle AD, Fideli U, Aldrovandi G, Allen S, Musonda R, Kaslow RA: **Favorable and unfavorable HLA class I alleles and**

- haplotypes in Zambians predominantly infected with clade C human immunodeficiency virus type 1.** *J Virol* 2002, **76**:8276-8284.
187. Kaslow RA, Carrington M, Apple R, Park L, Munoz A, Saah AJ, Goedert JJ, Winkler C, O'Brien SJ, Rinaldo C, et al: **Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection.** *Nat Med* 1996, **2**:405-411.
  188. O'Brien SJ, Gao X, Carrington M: **HLA and AIDS: a cautionary tale.** *Trends Mol Med* 2001, **7**:379-381.
  189. Ammaranond P, Zaunders J, Satchell C, van Bockel D, Cooper DA, Kelleher AD: **A new variant cytotoxic T lymphocyte escape mutation in HLA-B27-positive individuals infected with HIV type 1.** *AIDS Res Hum Retroviruses* 2005, **21**:395-397.
  190. Feeney ME, Tang Y, Roosevelt KA, Leslie AJ, McIntosh K, Karthas N, Walker BD, Goulder PJ: **Immune escape precedes breakthrough human immunodeficiency virus type 1 viremia and broadening of the cytotoxic T-lymphocyte response in an HLA-B27-positive long-term-nonprogressing child.** *J Virol* 2004, **78**:8927-8930.
  191. Gao X, Nelson GW, Karacki P, Martin MP, Phair J, Kaslow R, Goedert JJ, Buchbinder S, Hoots K, Vlahov D, et al: **Effect of a single amino acid change in MHC class I molecules on the rate of progression to AIDS.** *N Engl J Med* 2001, **344**:1668-1675.
  192. Goulder PJ, Bunce M, Krausa P, McIntyre K, Crowley S, Morgan B, Edwards A, Giangrande P, Phillips RE, McMichael AJ: **Novel, cross-restricted, conserved, and immunodominant cytotoxic T lymphocyte epitopes in slow progressors in HIV type 1 infection.** *AIDS Res Hum Retroviruses* 1996, **12**:1691-1698.
  193. Lazaryan A, Lobashevsky E, Mulenga J, Karita E, Allen S, Tang J, Kaslow RA: **Human leukocyte antigen B58 supertype and human immunodeficiency virus type 1 infection in native Africans.** *J Virol* 2006, **80**:6056-6060.
  194. Migueles SA, Sabbaghian MS, Shupert WL, Bettinotti MP, Marincola FM, Martino L, Hallahan CW, Selig SM, Schwartz D, Sullivan J, Connors M: **HLA B\*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors.** *Proc Natl Acad Sci U S A* 2000, **97**:2709-2714.
  195. Crawford H, Prado JG, Leslie A, Hue S, Honeyborne I, Reddy S, van der Stok M, Mncube Z, Brander C, Rousseau C, et al: **Compensatory mutation partially restores fitness and delays reversion of escape mutation within the immunodominant HLA-B\*5703-restricted Gag epitope in chronic human immunodeficiency virus type 1 infection.** *J Virol* 2007, **81**:8346-8351.
  196. Klein MR, van der Burg SH, Hovenkamp E, Holwerda AM, Drijfhout JW, Melief CJ, Miedema F: **Characterization of HLA-B57-restricted human immunodeficiency virus type 1 Gag- and RT-specific cytotoxic T lymphocyte responses.** *J Gen Virol* 1998, **79** ( Pt 9):2191-2201.
  197. Ngumbela KC, Day CL, Mncube Z, Nair K, Ramduth D, Thobakgale C, Moodley E, Reddy S, de Pierres C, Mkhwanazi N, et al: **Targeting of a CD8 T cell env epitope presented by HLA-B\*5802 is associated with markers of HIV disease progression and lack of selection pressure.** *AIDS Res Hum Retroviruses* 2008, **24**:72-82.
  198. Mungall AJ, Palmer SA, Sims SK, Edwards CA, Ashurst JL, Wilming L, Jones MC, Horton R, Hunt SE, Scott CE, et al: **The DNA sequence and analysis of human chromosome 6.** *Nature* 2003, **425**:805-811.
  199. Blankson JN, Bailey JR, Siliciano RF: **Crosscurrents in HIV-1 evolution.** *Nat Immunol* 2006, **7**:121-122.
  200. Price DA, Goulder PJ, Klenerman P, Sewell AK, Easterbrook PJ, Troop M, Bangham CR, Phillips RE: **Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection.** *Proc Natl Acad Sci U S A* 1997, **94**:1890-1895.
  201. Altfeld M, Allen TM: **Hitting HIV where it hurts: an alternative approach to HIV vaccine design.** *Trends Immunol* 2006, **27**:504-510.

202. Leslie A, Kavanagh D, Honeyborne I, Pfafferott K, Edwards C, Pillay T, Hilton L, Thobakgale C, Ramduth D, Draenert R, et al: **Transmission and accumulation of CTL escape variants drive negative associations between HIV polymorphisms and HLA.** *J Exp Med* 2005, **201**:891-902.
203. Altfeld M, Allen TM, Kalife ET, Frahm N, Addo MM, Mothe BR, Rathod A, Reyor LL, Harlow J, Yu XG, et al: **The majority of currently circulating human immunodeficiency virus type 1 clade B viruses fail to prime cytotoxic T-lymphocyte responses against an otherwise immunodominant HLA-A2-restricted epitope: implications for vaccine design.** *J Virol* 2005, **79**:5000-5005.
204. Frahm N, Kiepiela P, Adams S, Linde CH, Hewitt HS, Sango K, Feeney ME, Addo MM, Lichtenfeld M, Lahaie MP, et al: **Control of human immunodeficiency virus replication by cytotoxic T lymphocytes targeting subdominant epitopes.** *Nat Immunol* 2006, **7**:173-178.
205. Pond SL, Frost SD, Grossman Z, Gravenor MB, Richman DD, Brown AJ: **Adaptation to different human populations by HIV-1 revealed by codon-based analyses.** *PLoS Comput Biol* 2006, **2**:e62.
206. Brumme ZL, John M, Carlson JM, Brumme CJ, Chan D, Brockman MA, Swenson LC, Tao I, Szeto S, Rosato P, et al: **HLA-associated immune escape pathways in HIV-1 subtype B Gag, Pol and Nef proteins.** *PLoS One* 2009, **4**:e6687.
207. Barquera R, Zuniga J, Hernandez-Diaz R, Acuna-Alonzo V, Montoya-Gama K, Moscoso J, Torres-Garcia D, Garcia-Salas C, Silva B, Cruz-Robles D, et al: **HLA class I and class II haplotypes in admixed families from several regions of Mexico.** *Mol Immunol* 2008, **45**:1171-1178.
208. Gorodezky C, Alaez C, Vazquez-Garcia MN, de la Rosa G, Infante E, Balladares S, Toribio R, Perez-Luque E, Munoz L: **The genetic structure of Mexican Mestizos of different locations: tracking back their origins through MHC genes, blood group systems, and microsatellites.** *Hum Immunol* 2001, **62**:979-991.
209. Excoffier L, Slatkin M: **Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population.** *Mol Biol Evol* 1995, **12**:921-927.
210. Excoffier L, Laval G, Balding D: **Gametic phase estimation over large genomic regions using an adaptive window approach.** *Hum Genomics* 2003, **1**:7-19.
211. Storey JD, Tibshirani R: **Statistical significance for genomewide studies.** *Proc Natl Acad Sci U S A* 2003, **100**:9440-9445.
212. Kosakovsky Pond SL, Frost SD: **Not so different after all: a comparison of methods for detecting amino acid sites under selection.** *Mol Biol Evol* 2005, **22**:1208-1222.
213. Bastos FI, Caceres C, Galvao J, Veras MA, Castilho EA: **AIDS in Latin America: assessing the current status of the epidemic and the ongoing response.** *Int J Epidemiol* 2008, **37**:729-737.
214. Llop E: **El sistema mayor de histocompatibilidad humano, HLA.** In *Poblaciones chilenas: cuatro décadas de investigaciones bioantropológicas*  
 Edited by Rothhammer F, Llop E. Chile: Editorial Universitaria; 2004: 165-188
215. Karlsson AC, Deeks SG, Barbour JD, Heiken BD, Younger SR, Hoh R, Lane M, Sallberg M, Ortiz GM, Demarest JF, et al: **Dual pressure from antiretroviral therapy and cell-mediated immune response on the human immunodeficiency virus type 1 protease gene.** *J Virol* 2003, **77**:6743-6752.
216. Matthews PC, Prendergast A, Leslie A, Crawford H, Payne R, Rousseau C, Rolland M, Honeyborne I, Carlson J, Kadie C, et al: **Central role of reverting mutations in HLA associations with human immunodeficiency virus set point.** *J Virol* 2008, **82**:8548-8559.
217. Bailey JR, Brennan TP, O'Connell KA, Siliciano RF, Blankson JN: **Evidence of CD8+ T-cell-mediated selective pressure on human immunodeficiency virus type 1 nef in HLA-B\*57+ elite suppressors.** *J Virol* 2009, **83**:88-97.
218. Goulder PJ, Altfeld MA, Rosenberg ES, Nguyen T, Tang Y, Eldridge RL, Addo MM, He S, Mukherjee JS, Phillips MN, et al: **Substantial differences in specificity of**



- HIV-specific cytotoxic T cells in acute and chronic HIV infection.** *J Exp Med* 2001, **193**:181-194.
219. Karlsson AC, Iversen AK, Chapman JM, de Oliveira T, Spotts G, McMichael AJ, Davenport MP, Hecht FM, Nixon DF: **Sequential broadening of CTL responses in early HIV-1 infection is associated with viral escape.** *PLoS ONE* 2007, **2**:e225.
220. Storey JD, Tibshirani R: **Statistical methods for identifying differentially expressed genes in DNA microarrays.** *Methods Mol Biol* 2003, **224**:149-157.
221. Bailey JR, Sedaghat AR, Kieffer T, Brennan T, Lee PK, Wind-Rotolo M, Haggerty CM, Kamireddi AR, Liu Y, Lee J, et al: **Residual human immunodeficiency virus type 1 viremia in some patients on antiretroviral therapy is dominated by a small number of invariant clones rarely found in circulating CD4+ T cells.** *J Virol* 2006, **80**:6441-6457.
222. Brennan TP, Woods JO, Sedaghat AR, Siliciano JD, Siliciano RF, Wilke CO: **Analysis of HIV-1 Viremia and Provirus in Resting CD4+ T Cells Reveals a Novel Source of Residual Viremia in Patients on Antiretroviral Therapy.** *J Virol* 2009.
223. Rozera G, Abbate I, Bruselles A, Vlasi C, D'Offizi G, Narciso P, Chillemi G, Prosperi M, Ippolito G, Capobianchi MR: **Massively parallel pyrosequencing highlights minority variants in the HIV-1 env quasispecies deriving from lymphomonocyte sub-populations.** *Retrovirology* 2009, **6**:15.
224. Bailey JR, Zhang H, Wegweiser BW, Yang HC, Herrera L, Ahonkhai A, Williams TM, Siliciano RF, Blankson JN: **Evolution of HIV-1 in an HLA-B\*57-positive patient during virologic escape.** *J Infect Dis* 2007, **196**:50-55.
225. Browne EP, Littman DR: **Species-specific restriction of apobec3-mediated hypermutation.** *J Virol* 2008, **82**:1305-1313.
226. Borderia AV, Codoner FM, Sanjuan R: **Selection promotes organ compartmentalization in HIV-1: evidence from gag and pol genes.** *Evolution* 2007, **61**:272-279.
227. Biancotto A, Grivel JC, Iglehart SJ, Vanpouille C, Lisco A, Sieg SF, Debernardo R, Garate K, Rodriguez B, Margolis LB, Lederman MM: **Abnormal activation and cytokine spectra in lymph nodes of people chronically infected with HIV-1.** *Blood* 2007, **109**:4272-4279.
228. Brenchley JM, Hill BJ, Ambrozak DR, Price DA, Guenaga FJ, Casazza JP, Kuruppu J, Yazdani J, Migueles SA, Connors M, et al: **T-cell subsets that harbor human immunodeficiency virus (HIV) in vivo: implications for HIV pathogenesis.** *J Virol* 2004, **78**:1160-1168.
229. Brenchley JM, Schacker TW, Ruff LE, Price DA, Taylor JH, Beilman GJ, Nguyen PL, Khoruts A, Larson M, Haase AT, Douek DC: **CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract.** *J Exp Med* 2004, **200**:749-759.
230. Li B, Gladden AD, Altfeld M, Kaldor JM, Cooper DA, Kelleher AD, Allen TM: **Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution.** *J Virol* 2007, **81**:193-201.

Research

Open Access

## Unique features of HLA-mediated HIV evolution in a Mexican cohort: a comparative study

Santiago Avila-Rios\*<sup>1,2</sup>, Christopher E Ormsby<sup>1</sup>, Jonathan M Carlson<sup>3</sup>, Humberto Valenzuela-Ponce<sup>1</sup>, Juan Blanco-Heredia<sup>1</sup>, Daniela Garrido-Rodriguez<sup>1</sup>, Claudia Garcia-Morales<sup>1</sup>, David Heckerman<sup>3</sup>, Zabrina L Brumme<sup>4,5</sup>, Simon Mallal<sup>6</sup>, Mina John<sup>6</sup>, Enrique Espinosa<sup>1</sup> and Gustavo Reyes-Teran\*<sup>1</sup>

Address: <sup>1</sup>Center for Research in Infectious Diseases, National Institute of Respiratory Diseases, Mexico City, Mexico, <sup>2</sup>Faculty of Medicine, National Autonomous University of Mexico, Mexico City, Mexico, <sup>3</sup>Science Group, Microsoft Research, Redmond, Washington, USA, <sup>4</sup>Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology and Harvard, Boston, Massachusetts, USA, <sup>5</sup>Faculty of Health Sciences, Simon Fraser University, Burnaby, British Columbia, Canada and <sup>6</sup>Center for Clinical Immunology and Biomedical Statistics, Royal Perth Hospital and Murdoch University, Perth, Australia

Email: Santiago Avila-Rios\* - [santiago.avila@cieni.org.mx](mailto:santiago.avila@cieni.org.mx); Christopher E Ormsby - [christopher.ormsby@cieni.org.mx](mailto:christopher.ormsby@cieni.org.mx); Jonathan M Carlson - [carlson@microsoft.com](mailto:carlson@microsoft.com); Humberto Valenzuela-Ponce - [humberto.valenzuela@cieni.org.mx](mailto:humberto.valenzuela@cieni.org.mx); Juan Blanco-Heredia - [juan.blanco@cieni.org.mx](mailto:juan.blanco@cieni.org.mx); Daniela Garrido-Rodriguez - [daniela.garrido@cieni.org.mx](mailto:daniela.garrido@cieni.org.mx); Claudia Garcia-Morales - [claudia.garcia@cieni.org.mx](mailto:claudia.garcia@cieni.org.mx); David Heckerman - [heckerma@microsoft.com](mailto:heckerma@microsoft.com); Zabrina L Brumme - [zbrumme@partners.org](mailto:zbrumme@partners.org); Simon Mallal - [S.Mallal@murdoch.edu.au](mailto:S.Mallal@murdoch.edu.au); Mina John - [M.John@iidd.com.au](mailto:M.John@iidd.com.au); Enrique Espinosa - [enrique.espinosa@cieni.org.mx](mailto:enrique.espinosa@cieni.org.mx); Gustavo Reyes-Teran\* - [reyesteran@cieni.org.mx](mailto:reyesteran@cieni.org.mx)

\* Corresponding authors

Published: 10 August 2009

Received: 25 April 2009

*Retrovirology* 2009, **6**:72 doi:10.1186/1742-4690-6-72

Accepted: 10 August 2009

This article is available from: <http://www.retrovirology.com/content/6/1/72>

© 2009 Avila-Rios et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Mounting evidence indicates that HLA-mediated HIV evolution follows highly stereotypic pathways that result in HLA-associated footprints in HIV at the population level. However, it is not known whether characteristic HLA frequency distributions in different populations have resulted in additional unique footprints.

**Methods:** The phylogenetic dependency network model was applied to assess HLA-mediated evolution in datasets of HIV *pol* sequences from free plasma viruses and peripheral blood mononuclear cell (PBMC)-integrated proviruses in an immunogenetically unique cohort of Mexican individuals. Our data were compared with data from the IHAC cohort, a large multi-center cohort of individuals from Canada, Australia and the USA.

**Results:** Forty three different HLA-HIV codon associations representing 30 HLA-HIV codon pairs were observed in the Mexican cohort ( $q < 0.2$ ). Strikingly, 23 (53%) of these associations differed from those observed in the well-powered IHAC cohort, strongly suggesting the existence of unique characteristics in HLA-mediated HIV evolution in the Mexican cohort. Furthermore, 17 of the 23 novel associations involved HLA alleles whose frequencies were not significantly different from those in IHAC, suggesting that their detection was not due to increased statistical power but to differences in patterns of epitope targeting. Interestingly, the consensus differed in four positions between the two cohorts and three of these positions could be explained by HLA-associated

selection. Additionally, different HLA-HIV codon associations were seen when comparing HLA-mediated selection in plasma viruses and PBMC archived proviruses at the population level, with a significantly lower number of associations in the proviral dataset.

**Conclusion:** Our data support universal HLA-mediated HIV evolution at the population level, resulting in detectable HLA-associated footprints in the circulating virus. However, it also strongly suggests that unique genetic backgrounds in different HIV-infected populations may influence HIV evolution in a particular direction as particular HLA-HIV codon associations are determined by specific HLA frequency distributions. Our analysis also suggests a dynamic HLA-associated evolution in HIV with fewer HLA-HIV codon associations observed in the proviral compartment, which is likely enriched in early archived HIV sequences, compared to the plasma virus compartment. These results highlight the importance of comparative HIV evolutionary studies in immunologically different populations worldwide.

---

## Background

The cytotoxic CD8+ T lymphocyte (CTL) response has been identified as an important selective pressure driving HIV evolution within an infected host [1-5]. Strong lines of evidence support the importance of the CTL response in HIV control, including the temporal correlation between the appearance of HIV-specific CTLs *in vivo* and the decline of viremia in the early stages of HIV infection [6], as well as the lack of control of virus levels after experimental depletion of CD8+ cells in rhesus macaques prior to simian immunodeficiency virus (SIV) infection [7]. CTLs recognize and destroy infected cells through the binding of their T cell receptor (TCR) to viral peptides (epitopes) presented on the surface of infected cells by highly polymorphic molecules encoded by class I human leukocyte antigen (HLA) genes. Each HLA allele encodes a unique HLA molecule capable of presenting a broad range of possible epitopes derived from various areas of the HIV proteome. CTL recognition of these peptide-HLA complexes may be associated with different functional outcomes in the infection [8-10]. Importantly, as a result of CTL-mediated selective pressure, immune escape mutations are selected that hinder viral peptide binding to HLA molecules, prevent peptide processing before their presentation or lower TCR affinity of specific CTL clones to peptide-HLA complexes [4,11-13]. Therefore, both the processes of antigen presentation to CTLs and CTL escape are HLA-restricted [14].

Depending on their costs to viral fitness, some CTL escape mutations can be transmitted and maintained in a new host [15-18], even without the presence of the originally selective HLA allele [11,19-21]. Additionally, there is evidence supporting the notion that some immune escape mutations can accumulate in a large number of individuals and become fixed in the circulating virus consensus sequence, driving HIV evolutionary changes at the population level [9,19,22-24]. As a result, specific HLA epitopes could become extinct in the viral population, allowing HIV adaptation to HLA-associated immune control in a

certain region [22,23]. The relative impact of different factors that could influence the persistence of escape mutations in a large number of individuals remains incompletely understood [22,25]. The variety of these factors—such as the extent of reversion of immune escape mutations in the absence of the selecting HLA allele [15,16,18], selection of compensatory mutations that restore viral fitness [26], founder effects [9], conflicting evolutionary forces on clustered epitopes [27], development of novel CTL responses to escape variants [28,29], inter-clade differences in the circulating viruses [14,30], immunodominance hierarchies of CTL responses [25,31,32], and HLA allele frequency distributions in different populations [22]—highlight the complexity of viral adaptation to the immune response at the population level [9,25].

In spite of this complexity, mounting evidence indicating that a large number of CTL escape mutations are reproducibly selected in the context of specific HLA restrictions has led to the hallmark observation that HIV evolution follows generally predictable mutational patterns in response to specific HLA-restricted immune responses (reviewed in [10]). This "HLA footprint effect" on HIV has been shown at the population level through correlative associations between the presence (or absence) of polymorphisms at specific positions of the viral sequence and the expression of specific HLA alleles [24,25,33-35]. Detection of HLA-HIV polymorphism associations is potentially limited by important confounding effects, namely HIV phylogeny, HIV codon covariation, and linkage disequilibrium of HLA alleles [10,14]. Several studies have accounted for some of these confounding effects explicitly [24,30,34]; more recently, a comprehensive evolutionary model considering all these confounding sources was proposed [14]. This phylogenetic dependency network model was shown to be able to reconstruct previously defined escape and compensatory mutation pathways and agrees with emerging data on patterns of epitope targeting. The existence of this kind of comprehensive

models represents an opportunity to systematically study HIV evolution in immunogenetically different populations and assess the importance of different HLA backgrounds in HIV evolution at the population level. Due to the extensive polymorphism of HLA genes, allelic and haplotypic frequency distributions in distinct infected populations vary widely [9]. Given the highly consistent effect of HLA-restricted selection on HIV evolution and the distinct HLA allele distributions in differing populations, it is likely that specific HLA-HIV polymorphism associations will be preferentially observed in different populations, determining unique characteristics of HIV evolution in different human groups [9,36]. To explore this possibility, HLA-mediated HIV evolution at the population level was studied in a cohort of clade B-infected individuals from Central/Southern Mexico, and compared to previously reported studies in a large multicenter cohort of predominantly clade B-infected individuals from British Columbia, Canada; Western Australia; and the USA (the International HIV Adaptation Collaborative [IHAC] cohort) (Brumme ZL, John M, et al, PLoS ONE 2009, in press) [14,25,34,37]. In order to determine to what extent HLA imprinting on HIV is a general phenomenon, it is informative to study HIV evolution in an immunogenetically unique population that possibly reflects a different selective pressure to that observed in other studied populations. The Mexican population is known to have a unique immunogenetic background characterized by the admixture of mainly Amerindian and Caucasian HLA haplotypes [38,39]. To our knowledge, Latin American cohorts have not been the primary subject of HIV evolutionary studies. Our data suggest that the unique HLA frequency distribution in a previously uncharacterized, immunogenetically unique HIV-infected population is imprinting HIV evolution in a unique way. This fact underlines the importance of systematically expanding our understanding of CTL escape and HIV evolution in immunogenetically distinct populations. This knowledge has important implications for the design of CTL-based vaccines and treatment strategies.

## Methods

### Study population

Peripheral blood samples were prospectively obtained from 303 chronically-infected, HIV positive, antiretroviral treatment-naïve individuals from Central/Southern Mexico. Participating individuals were recruited with written informed consent at different health centers in Mexico City and from the states of Puebla, Jalisco, Oaxaca, Guerrero, the State of Mexico and Chiapas. Blood samples were shipped to and processed at the Center for Research in Infectious Diseases of the National Institute of Respiratory Diseases in Mexico City. All ethical issues related to this project were evaluated and approved by the Institutional Bioethics and Science Committee. For each patient,

plasma aliquots and peripheral blood mononuclear cells (PBMCs) were obtained and cryopreserved.

HLA frequency and HLA-mediated HIV evolution data obtained from the Mexican cohort were compared with that obtained from a previously described cohort of 1,045 HIV-positive, predominantly Caucasian individuals from British Columbia, Canada (HOMER cohort) [34] and the large multicenter International HIV Adaptation Combined (IHAC) cohort, including 1,845 predominantly Caucasian individuals from British Columbia, Canada; Western Australia and the USA [37] (Brumme ZL, John M, et al, PLoS ONE 2009, in press).

### HLA typing

Genomic DNA was extracted from at least 6 million PBMCs using QIAmp DNA Blood Mini Kit (QIAGEN, Valencia CA), according to the manufacturer's specifications. Class I HLA A, B and C genes were typed at low/medium resolution for each participating individual by sequence-specific primer polymerase chain reaction (SSP-PCR) using ABC SSP UniTray Kit (Invitrogen, Brown Deer, WI) according to the manufacturer's specifications. Briefly, genomic DNA from each participating individual at 75–125 ng/μL was used as template for 95 PCRs with different sequence-specific primers designed to detect relevant polymorphisms for typing. Reaction products were run on a 2.0% agarose gel (Promega, Madison, WI). Amplification patterns were analyzed with UniMatch v3.2 software using up-to-date data bases to determine HLA groups. All the reactions included an internal amplification control to be validated and each test included a reagent control to detect contamination.

### HLA frequency analyses and comparisons

HLA allelic and population frequencies for the Mexican cohort were obtained with the HLA Frequency Analysis tool of the Los Alamos HIV Database <http://www.hiv.lanl.gov/content/immunology/tools-links.html>. HLA haplotype frequencies were obtained with the Arlequin v3.11 software. Due to the fact that the cohort was composed of non-related individuals with unknown family genetic backgrounds, a gametic phase estimation was carried out for each individual using a pseudo-Bayesian algorithm designed to reconstruct the gametic phase of multi-loci genotypes, included in the Arlequin v3.11 software (Excoffier-Laval-Balding, ELB) [40]. Frequency analysis between the cohort reported here and the Canadian HOMER cohort, the multicenter IHAC cohort and a cohort of HIV-negative individuals from Central/Northern Mexico [38], was carried out by chi squared test, with post hoc two by two significance determined by Fisher's exact test, corrected for multiple comparisons by q values [41]. Significant values were considered to be  $q < 0.05$ . These analyses were carried out

with R statistical environment v2.8.1, using the package qvalue v1.1.

#### **HIV pol genotyping from free plasma virus**

Viral RNA from free plasma virus was purified from 1 mL of plasma using QIAmp Viral RNA Mini Kit (QIAGEN, Valencia, CA) according to the manufacturer's specifications. A fragment of the viral *pol* gene including the whole protease (PR) and 335 codons of the reverse transcriptase (RT) was bulk sequenced from plasma viral RNA for each participating individual. Sequences were obtained with a 3100-Avant Genetic Analyzer (Applied Biosystems, Foster City, CA), using ViroSeq HIV-1 Genotyping System (Celera Diagnostics, Alameda, CA) according to the manufacturer's specifications. Briefly, 1.3 Kbp fragments of the *pol* gene were amplified by RT-PCR from plasma viral RNA. PCR products were purified with ultra filtration columns and quantified in 1.5% agarose gels (Promega, Madison, WI). For each patient, sequencing PCRs were carried out with 7 different primers to assure that the whole genomic region was covered with at least two sequences. Sequences were assembled, aligned to the HXB2 consensus, and manually edited using the ViroSeq v2.7 software provided by the manufacturer.

#### **HIV pol genotyping from PBMC proviral DNA**

Genomic DNA was purified as described above. A fragment of approximately 1.5 Kbp covering the whole PR and the first 335 codons of RT was amplified by nested PCR with Platinum Taq DNA Polymerase (Invitrogen, Carlsbad, CA), and primers PR 5' OUTER 5'-CCCTAG-GAAAAGGGCTGTTG-3'/RT 3' OUTER 5'-GTTTTCA-GATTTTTAAATGGCTCTTG-3', for the first round of amplification, and PR 5' INNER 5'-TGAAAGATTGACT-GAGAGACAGG-3'/RT 3' INNER 5'-GGCTCTTGATAAATTTGATATGTCC-3' for the second round of amplification. PCR conditions were 1 cycle of 94°C, for 3 min, followed by 35 cycles of 94°C for 30 s, 60°C for 30 s and 72°C for 2 min and a cycle of 72°C for 5 min, with final concentrations of 2 mM Mg<sup>++</sup>, 0.2 mM dNTPs, 0.4 mM of each primer and 20 ng/μL genomic DNA for both amplification rounds (transferring 10% of the volume of first round PCR product to the second round). In all cases, contamination controls were included. PCR products were purified by QIAquick PCR Purification Kit (QIAGEN, Valencia, CA) according to the manufacturer's specifications, and quantified in 2.0% agarose gels (Promega, Madison, WI). Seven sequencing PCRs were carried out for each patient using seven primer mixes included in the ViroSeq HIV-1 Genotyping System Kit (Celera Diagnostics, Alameda, CA), in order to cover the whole analyzed region with at least two sequences. Bulk proviral *pol* sequences were obtained with a 3100-Avant Genetic Analyzer (Applied Biosystems, Foster City, CA). Sequences

were assembled, aligned to the HXB2 consensus, and edited manually using the ViroSeq v2.7 software.

#### **Evolutionary analyses**

The phylogenetic dependency network (PDN) model by Carlson, et al [14], was applied to infer patterns of CTL escape and codon covariation in the plasma and proviral sequence datasets, using the PhyloDv program <http://www.codeplex.com/MSCompBio>. The PDN model was designed to simultaneously account for HIV codon covariation, linkage disequilibrium among HLA alleles and the confounding effects of HIV phylogeny when attempting to identify HLA-associated polymorphisms in HIV [14]. Briefly, the PDN model is a multivariate model that represents the probabilistic dependencies among a set of target attributes (in this case the presence or absence of amino acids at all codons in an HIV protein) and a set of predictor attributes (in this case the presence or absence of amino acids at all codons other than that for the target attribute in the HIV sequence, as well as the presence or absence of all possible HLA alleles) while correcting for the phylogenetic structure of the viral sequences. A dependency network graphically depicts which HLA and codon attributes predict each target codon attribute, associating a probability distribution for each target codon attribute, conditioned on various HLA and codon attributes. Importantly, each local probability distribution is corrected for the phylogenetic structure of the HIV sequences. To determine the significance of a particular predictor-target pair, the likelihood of a null model that reflects the assertion that the target variable is not under selection pressure from the predictor attribute is compared to an alternative model that reflects the assertion that the target variable is under selection pressure from that predictor attribute. Multiple predictors are added to the model in an iterative fashion using forward selection, in which the most significantly associated attribute is iteratively added to the model until no attribute achieves  $p < 0.05$ . The use of a multivariate model minimizes spurious associations caused by the presence of linkage disequilibrium among HLA alleles and HIV codon covariation. For each added predictor attribute, the most significant leaf distribution is recorded (escape, reversion, attraction, or repulsion, see below). The statistical significance of a predictor with respect to a target attribute is computed using false discovery rates (FDRs), which are computed using a likelihood-ratio test in which both the null and the alternative models are conditioned on all significant predictors that were identified in previous iterations of forward selection. For each p-value, we report the corresponding q-value, which is the minimum FDR among rejection regions that include that p-value, as computed using the method of Storey and Tibshirani with the  $\pi_0$  parameter conservatively set to one [41]. Attributes were excluded as possible predictors when the corresponding predictor-tar-

get pair had a 2 × 2 contingency table in which any cell of the table had an observed or expected value of three or less.

The precise rules governing the transitions of the target attribute, conditioned on the predictor attributes and the sequence phylogeny, are given by a univariate leaf distribution, which is assumed to be the same for each individual. Four possible leaf distributions are defined: *Attraction*, having the predictor makes it more likely to have the target; *Repulsion*, not having the predictor makes it less likely to have the target; *Escape*, having the predictor makes it less likely to have the target; and *Reversion*, not having the predictor makes it more likely to have the target. The pair Attraction/Repulsion corresponds to a positive correlation between predictor and target, while the pair Escape/Reversion corresponds to a negative correlation between predictor and target.

**Results**

**General clinical and geographical characteristics of the Mexican cohort**

Figure 1 shows the geographical residence of the individuals included in the study. As is typical in Latin American cohorts [42,43], half of the individuals were found to be in relatively advanced stages of HIV infection (CD4+ T cell counts <200 cells/μL) at enrolment, with approximately half of these patients having less than 50 CD4+ T cells/μL. Only one of every 10 participating individuals was found to be at relatively early stages of the infection (CD4+ T cell count >500 cells/μL) (Table 1). Taking the cohort as a whole, the median CD4+ T cell count was lower than 200 cells/μL. The male-to-female ratio of infected individuals was 3 to 1 (Table 1), representing a slightly higher HIV prevalence among women than previously reported for the Mexican infected population [44], possibly suggesting a tendency towards increased HIV infection in females in the Latin American region [43]<http://www.unaids.org>. A

typical negative correlation was observed between CD4+ T cell counts and plasma viral loads (p < 0.0001), with a mean increase in viral load of 0.1 logarithms per 50 CD4+ T cell decrease. Taken together, these observations are representative of a typical Mexican cohort, comprised mainly of individuals in relatively advanced stages of HIV infection, often diagnosed at the moment of presentation at the health care centers due to AIDS-related opportunistic disease symptoms.

**HLA allelic and haplotypic frequencies in a cohort of HIV-positive Mexican individuals**

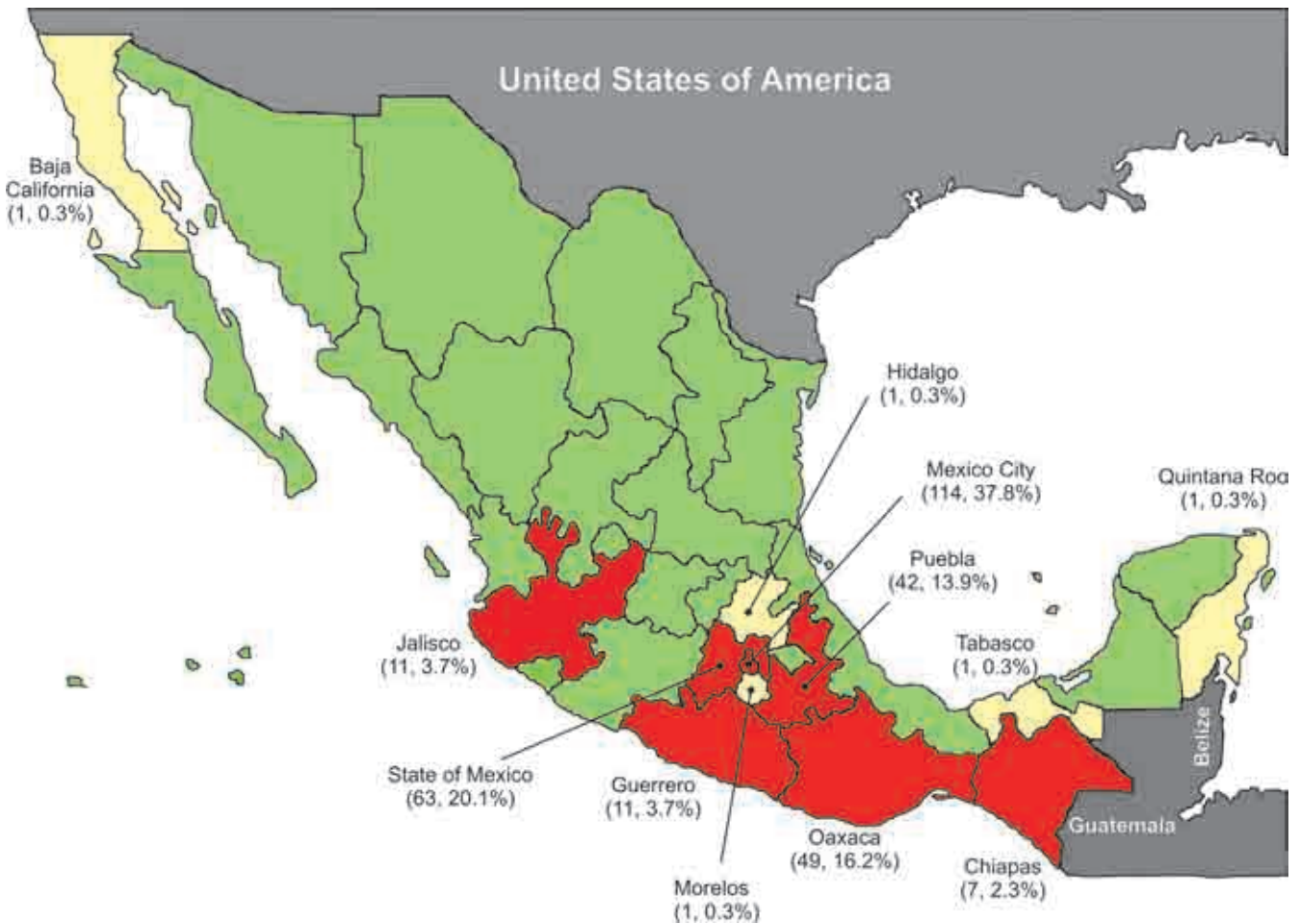
292 HIV-positive individuals from Central/Southern Mexico for whom class I HLA-A, B and C typing was available were used to characterize the immunogenetic background of this cohort. HLA allelic frequencies for the Mexican cohort are shown in the Additional file 1: Figure S1, Table S1. The most frequent alleles at the HLA-A locus were A\*02, A\*24, A\*68 and A\*31; the most frequent alleles at the HLA-B locus were B\*39, B\*35, B\*40 and B\*15; and the most frequent alleles at the HLA-C locus were Cw\*07, Cw\*04, Cw\*03 and Cw\*08 (Additional file 1: Figure S1). Characteristically, more than 60% of the participating individuals expressed A\*02, more than 50% expressed Cw\*07 and more than a third expressed B\*39 and/or B\*35 (Additional file 1: Figure S1, Table S1).

In order to more precisely describe the immunogenetic background of the HIV-positive cohort of Mexican individuals, the frequencies of two and three-gene class I HLA haplotypes were estimated. Due to the fact that the cohort was composed of non-related individuals with unknown family genetic backgrounds, a gametic phase estimation for each individual was carried out prior to the calculation of HLA haplotype frequencies as described in the Methods. A total of 192 different three-gene HLA haplotypes were identified, of which 22 occurred at a frequency higher than 1% (Figure 2). The most frequent three-gene

**Table 1: Relevant clinical parameters for a cohort of 303 Mexican individuals.**

Clinical Parameters	Mean	Median	Standard Error	Standard Deviation
CD4+ T cell count (cells/μL)*	238.4	196.5	11.6	201.3
CD4+ T cell %*	14.0	12.0	0.6	10.4
Viral Load (RNA copies/mL)	233,528	105,000	16,927	292,691
Log Viral Load	4.917	5.021	0.044	0.767
<b>CD4+ T cell count stratification* [n (%)]</b>				
>500 cells/μL	35 (11.6%)		148 (49.3%)	
201 – 500 cells/μL	113 (37.7%)			
50 – 200 cells/μL	82 (27.3%)		152 (50.7%)	
<50 cells/μL	70 (23.3%)			
<b>Gender [n (%)]</b>				
Male			229 (75.6%)	
Female			74 (24.4%)	

\*Data for 3 patients are missing



**Figure 1**  
**Geographical residence of the individuals included in the study.** The map shows data for 302 antiretroviral treatment-naïve HIV-infected individuals. States in red account for 98.3% of the individuals included in the study. States in white account for 1.7% of the individuals of the cohort.

haplotypes were A\*02/B\*39/Cw\*07, A\*68/B\*39/Cw\*07 and A\*02/B\*35/Cw\*04, all occurring at frequencies higher than 4.5%. Considering two loci, a total of 121 possible haplotypes were found for HLA-A/B, 82 for HLA-B/C and 92 for HLA-A/C. The most frequent two-gene haplotypes were A\*02/B\*39, A\*02/B\*35, A\*24/B\*35 and A\*68/B\*39 for HLA-A/B; B\*39/Cw\*07, B\*35/Cw\*04 and B\*40/Cw\*03 for HLA-B/C; and A\*02/Cw\*07, A\*68/Cw\*07 and A\*02/Cw03 for HLA-A/C (Table 2). In general, there was lower variability among the HLA-B/C haplotypes compared to the HLA-A/C and the HLA-A/B haplotypes, possibly due to the frequent linkage disequilibrium observed between HLA-B and C genes (Additional file 1: Table S2).

HLA-A and B allelic frequencies in this study were compared to those previously reported in an open population-based study of 381 individuals from 191 Mexican families

from Central/Northern Mexico [38] (Figure 3). Although the geographical origin of the individuals in the latter study differs somewhat from that of the individuals in the present study, the large number of individuals from the Central part of Mexico and the fact that the HLA typing method used was the same as ours, renders this study an adequate reference for a typical HIV-negative population in Mexico for comparison with our study. The HLA frequency distribution of loci A and B was significantly different between the two studies ( $\chi^2 = 99.39$ ,  $p = 0.00008$ ), with differences in residuals seen only in B\*39 ( $p = 2.25E-06$ ,  $q = 1.19E-04$ ), a typical Amerindian allele group, which showed a frequency nearly two-fold higher in HIV-positive individuals compared to HIV-negative individuals (Figure 3). Whether having B\*39 represents a risk factor for HIV infection in Mexico remains to be confirmed, as the high frequency of this allele could also reflect an epidemiological phenomenon such as B\*39

**Table 2: Most frequent two-gene class I HLA haplotypes in the cohort of HIV-positive Mexican individuals.†**

HLA A-B	Frequency	HLA B-Cw	Frequency	HLA A-Cw	Frequency
A*02 B*39	0.0938	B*39 Cw*07	0.17422	A*02 Cw*07	0.10627
A*02 B*35	0.0486	B*35 Cw*04	0.14286	A*68 Cw*07	0.07840
A*24 B*35	0.0486	B*40 Cw*03	0.05401	A*02 Cw*03	0.07143
A*68 B*39	0.0486	B*07 Cw*07	0.04530	A*02 Cw*08	0.05226
A*02 B*51	0.0451	B*15 Cw*01	0.04530	A*02 Cw*04	0.04878
A*68 B*35	0.0451	B*14 Cw*08	0.03833	A*24 Cw*07	0.04355
A*02 B*40	0.0399	B*48 Cw*08	0.03659	A*02 Cw*01	0.03833
A*02 B*15	0.0330	B*44 Cw*16	0.02613	A*24 Cw*04	0.03833
A*31 B*35	0.0278	B*44 Cw*05	0.02265	A*31 Cw*04	0.02613
A*24 B*39	0.0243	B*49 Cw*07	0.02091	A*02 Cw*15	0.02265
A*03 B*07	0.0208	B*52 Cw*03	0.01916	A*68 Cw*04	0.02265
A*24 B*15	0.0208	B*51 Cw*15	0.01916	A*24 Cw*03	0.02265
A*02 B*52	0.0208	B*35 Cw*07	0.01742	A*01 Cw*07	0.02091
A*31 B*39	0.0191	B*08 Cw*07	0.01568	A*31 Cw*07	0.01916
A*02 B*48	0.0156	B*18 Cw*05	0.01568	A*68 Cw*03	0.01916
A*02 B*49	0.0139	B*51 Cw*08	0.01568	A*02 Cw*16	0.01568
A*24 B*40	0.0139	B*38 Cw*12	0.01394	A*03 Cw*07	0.01568
A*30 B*18	0.0122	B*39 Cw*03	0.01394	A*24 Cw*08	0.01394
A*68 B*40	0.0122	B*35 Cw*03	0.01045	A*33 Cw*08	0.01394
A*24 B*44	0.0122	B*52 Cw*12	0.01045	A*29 Cw*16	0.01220
A*29 B*44	0.0122	B*55 Cw*07	0.01045	A*30 Cw*05	0.01045
A*33 B*14	0.0122			A*31 Cw*01	0.01045
A*01 B*57	0.0104			A*24 Cw*05	0.01045
A*02 B*07	0.0104				
A*02 B*44	0.0104				
A*02 B*55	0.0104				

†292 HIV positive individuals from Central/Southern Mexico were included. Gametic phase for each individual was estimated with the Arlequin v3.11 software as described in the Methods. Haplotypes with frequencies over 1% are shown.

being enriched in the most affected sectors of the population by HIV infection or simply be a sample bias of the individuals included in either of the two studies.

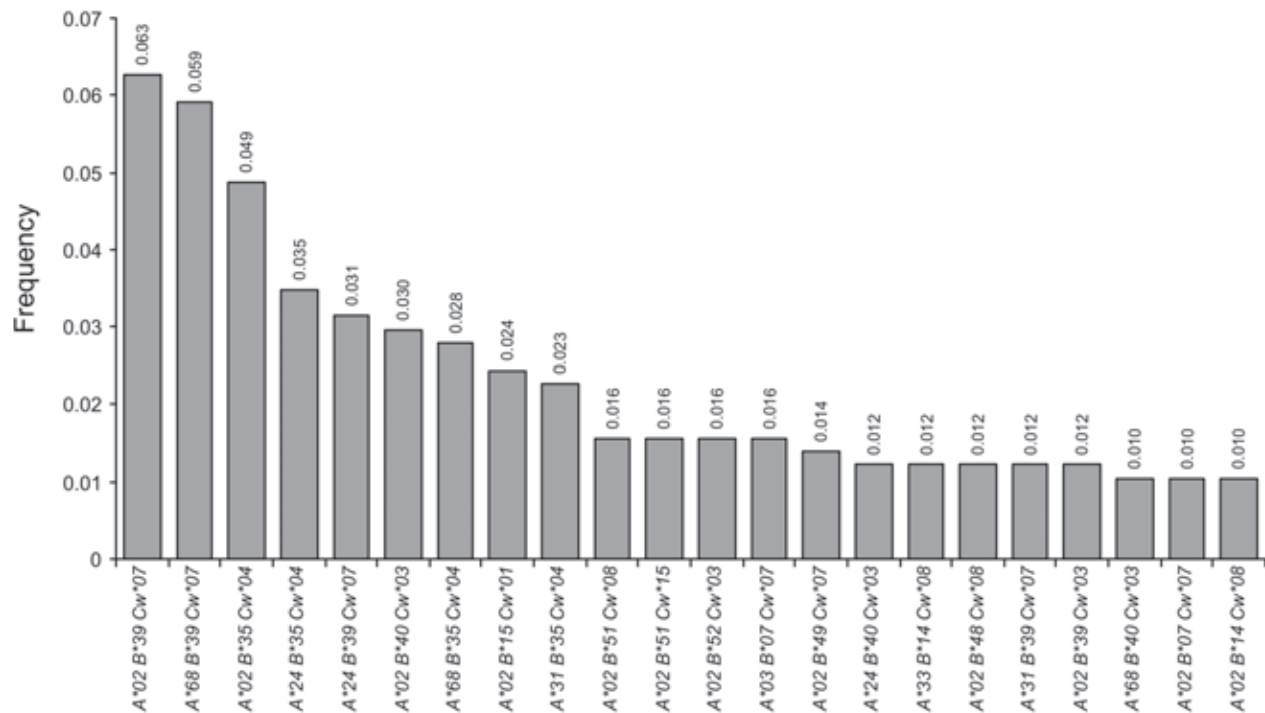
To our knowledge, this study is the first formal report of class I HLA frequencies in a typical HIV-infected Mexican cohort.

#### **Unique immunogenetic Background in a cohort of HIV-infected, antiretroviral treatment naïve individuals from Central/Southern Mexico**

In order to highlight the unique immunogenetic background of the Mexican population with respect to other populations in which HLA-associated HIV evolution has been studied, an HLA frequency comparison was carried out between our cohort of 292 HIV-positive individuals from Central/Southern Mexico, a previously described cohort of 1,045 HIV-positive individuals from British Columbia, Canada (HOMER cohort) [34] and the large International HIV Adaptation Combined (IHAC) cohort, including 1,845 individuals from British Columbia, Canada; Western Australia and the USA (Figure 4). Although both the HOMER cohort and the USA subset of the IHAC cohort include a minority of individuals self-identified as Hispanic, important differences were seen in HLA allele

distribution in the three cohorts that account for the typical genetic admixture of the Mexican population [38,39]. As expected, there were significant differences between the allele frequencies of the cohort reported here and the HOMER and IHAC cohorts ( $\chi^2 = 597.41$  and  $782.13$ ,  $p < 10^{-88}$  and  $10^{-125}$ , respectively). HLA-A\*68, B\*35, B\*39, B\*48, B\*52, Cw\*04 and Cw\*08 alleles were observed at significantly higher frequencies in the Mexican cohort compared to HOMER and IHAC cohorts ( $p < 0.005$ ,  $q < 0.01$ ), consistent with typical Amerindian alleles [38,39,45]. Similarly, HLA-A\*01, A\*03, A\*11, B\*07, B\*08, B\*13, B\*27, B\*44, B\*57, Cw\*05, and Cw\*06 alleles were observed at significantly lower frequencies in the Mexican cohort compared to HOMER and IHAC cohorts ( $p < 0.005$ ,  $q < 0.01$ ), consistent with the higher frequency of these alleles among Caucasians [38,39,45] (Figure 4). Additionally, HLA-A\*02 and A\*24 alleles had significantly higher frequencies, and HLA-A\*25, B\*15 and Cw\*02 alleles had significantly lower frequencies in the Mexican cohort than in HOMER and IHAC cohorts, not specifically reported to be enriched in Amerindian, or Caucasian groups. Notably, the frequency of HLA-B\*39 alleles was more than 7 times higher in the Mexican cohort than in HOMER and IHAC cohorts (Figure 4). Taken together, these results confirm the characteristic





**Figure 2**

**Most frequent three-gene class I HLA haplotypes in a Mexican cohort of HIV-positive individuals.** Genetic frequencies were calculated for 292 HIV-positive individuals from Central/Southern Mexico. Gametic phase for each individual was estimated using the pseudo-Bayesian algorithm ELB, using the program Arlequin v3.11. HLA-A, B and C genes were typed at low/medium resolution by SSP-PCR as described in Methods. Haplotypes with frequencies over 1% in the cohort are shown.

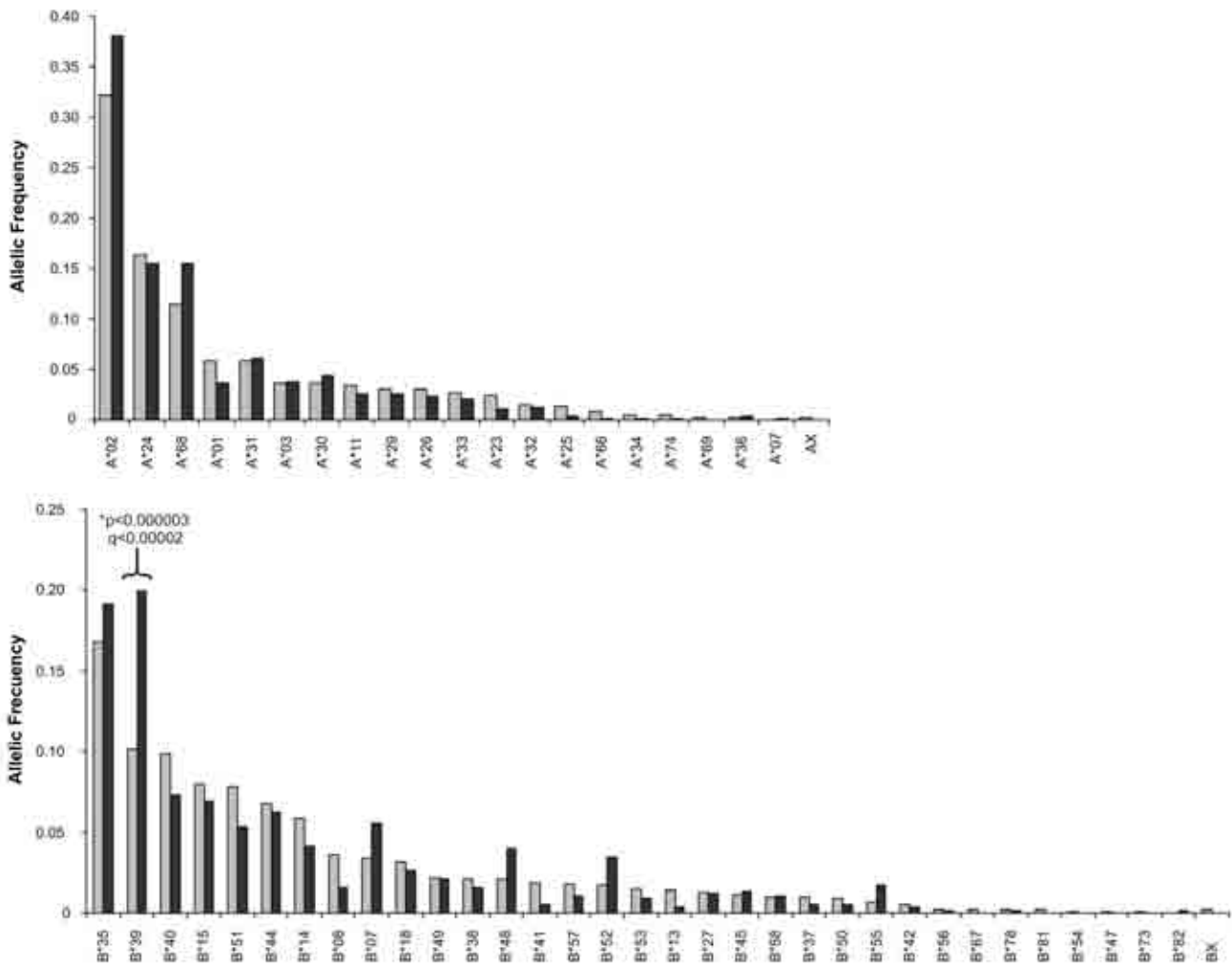
admixture of the mainly Amerindian and Caucasian genes of the Mexican mestizo population in a typical cohort of HIV-infected individuals from the Central/Southern region of the country, and reveal a previously uncharacterized, unique immunogenetic background for the study of HLA-associated HIV evolution at the population level.

#### **HLA-mediated HIV evolution in a Mexican cohort**

HIV evolution mediated by HLA selection at the population level was studied using a 434 amino acid fragment spanning the whole HIV protease and 335 codons of the reverse transcriptase in 280 chronically-infected individuals from this cohort. The phylogenetic dependency network (PDN) model by Carlson et al [14], currently one of the most comprehensive models to assess HLA-mediated HIV evolution, was applied to infer patterns of CTL escape and codon co-variation in the Mexican cohort. Our results were compared with those previously derived from applying the PDN model to a thoroughly characterized, multi-center, combined cohort of predominantly clade B-infected, antiretroviral treatment-naïve individuals from British Columbia, Canada; Western Australia; and the

USA (the IHAC cohort), with a clearly different immunogenetic background compared to the Mexican cohort [14,34,37,46] (Figure 4). The Mexican cohort was also found to be predominantly clade B-infected (99.64%) with only one subtype other than B/recombinant form (CRF\_06\_cpx) identified (REGA HIV-1 Subtyping Tool 2.0, <http://dbpartners.stanford.edu/RegaSubtyping/>). A phylogenetic tree for the Mexican *pol* sequences included in this study is shown in the Additional file 1: Figure S2.

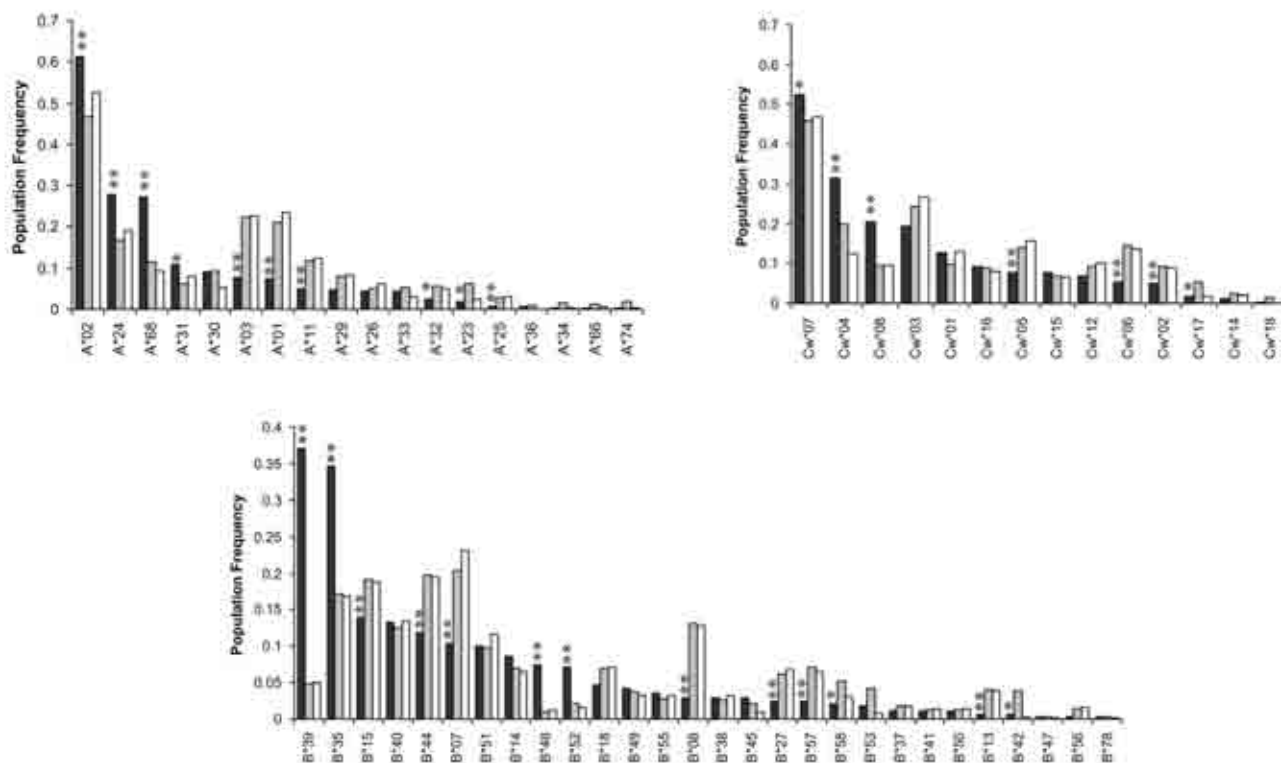
The PDN model was used to identify significant HLA-HIV codon as well as HIV codon-HIV codon associations, using a q-value threshold of 0.2. Due to the fact that the PDN model uses a multivariate model in which several predictor attributes (i.e. the presence or absence of a specific HLA or amino acid at an HIV codon) can be associated with the presence or absence of a specific amino acid at an HIV target codon, spurious associations explained by the presence of linkage disequilibrium among HLA alleles and HIV codon covariation were minimized. A total of 43 HLA-HIV codon and 251 HIV codon-HIV codon associations were identified, representing 30 differ-



**Figure 3**  
**Differences in HLA frequencies between HIV-positive and HIV-negative Mexicans.** Allelic frequencies were calculated for HLA-A and B genes, in the cohort of 292 HIV-positive individuals from this study (dark grey) and compared to those previously reported for a cohort of 381 individuals of 191 Mexican families by Barquera et al [38] (light grey). HLA typing in both cases was carried out by SSP-PCR as described in the Methods. For comparability, HLA nomenclature for histocompatibility used by Barquera et al. was substituted with its genetic equivalent, i.e. B65 and B64 were included as B\*14 alleles; B62, B63, B70, B71, B72, and B75 were included as B\*15 alleles; and B61, and B60 were included as B\*40 alleles, according to the equivalents accepted by the WHO Nomenclature Committee for Factors of the HLA System <http://www.ebi.ac.uk/imgt/hla/dictionary.html>.

ent HLA-HIV codon and 135 HIV codon-HIV codon pairs (Additional file 1: Table S3). This association network was depicted graphically with the PDN viewer PhyloDv <http://www.codeplex.com/MSCCompBio> [14] (Figure 5), showing the HIV amino acid sequence as a circle with lines joining HLA alleles and associated HIV codons outside the circle and arcs joining covarying HIV codons within the circle. Even with a relatively small number of individuals in the cohort, a dense association network was observed at  $q < 0.2$  that reveals characteristic patterns of HIV codon covariation and HLA-mediated substitutions

in the studied cohort. HLA associations were found at 6.1% of protease codons, and at 7.1% of RT codons. As previously described for HIV Gag [14], covarying codons were more frequent within a sub-protein (75.6% total: 20% within the protease and 55.6% within the reverse transcriptase) than between sub-proteins (24.4%;  $p < 0.001$ ). 28/135 (20.7%) of HIV codon pairs were within 10 positions of each other, suggesting a close proximity in an important proportion of compensatory mutations, or the targeting of multiple epitopes by the same HLA allele. Notably, 46.7% of HLA-HIV codon associations predicted



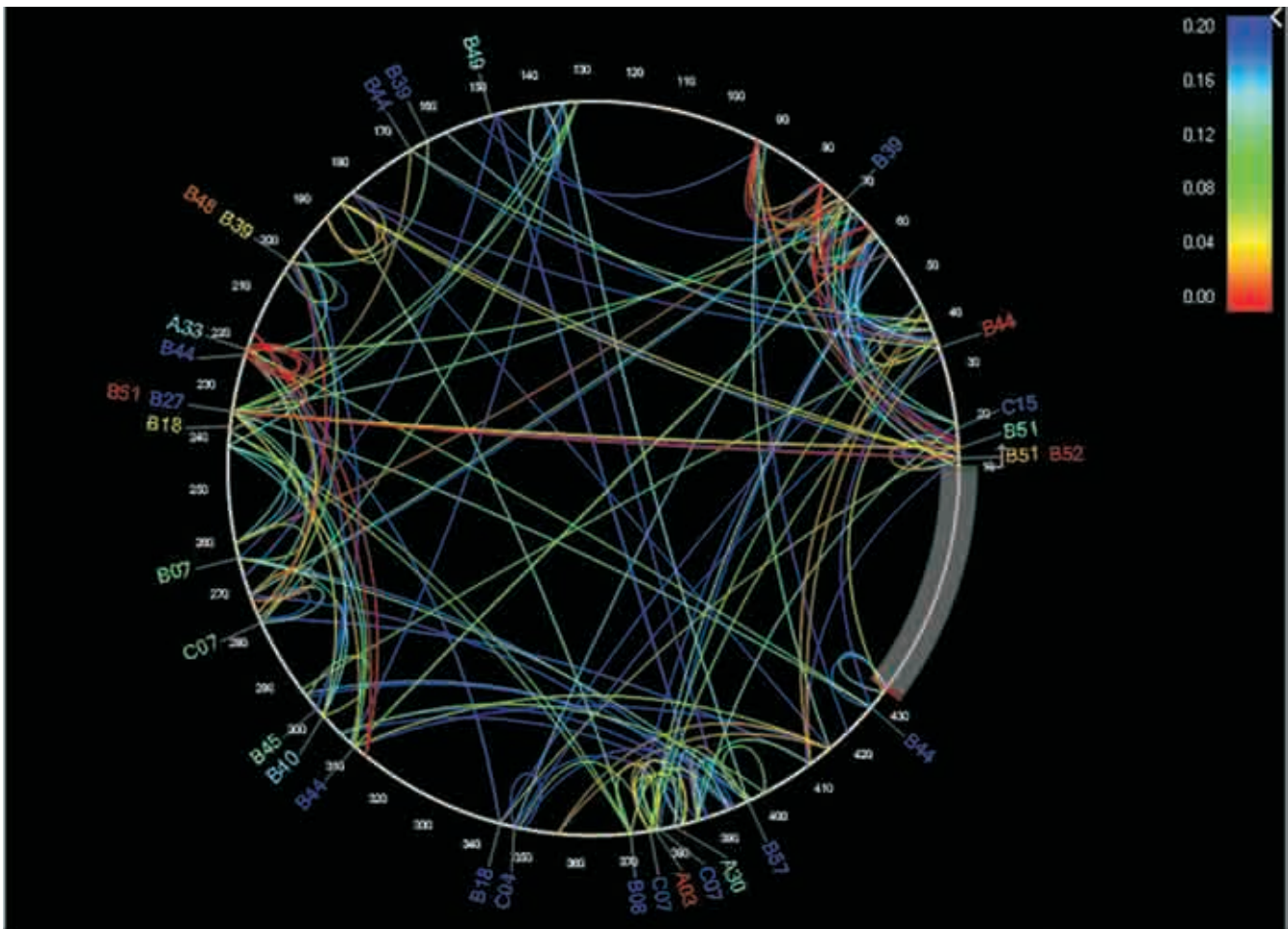
**Figure 4**  
**Marked differences in HLA allele frequencies in three clade B-infected cohorts.** Population frequencies for class I HLA genes A, B and C were compared between the Mexican cohort described in this study (n = 292) (dark grey). The combined IHAC cohort including individuals from British Columbia, Canada; Western Australia and the USA (n = 1845) [37] (Brumme ZL, John M, et al, PLoS ONE 2009, in press), and the British Columbia HOMER cohort (n = 1045) (white) described in detail previously [34]. \*\*Significant differences (q < 0.05) between the Mexican cohorts and both the IHAC and the HOMER cohorts, \*significant differences (q < 0.05) between the Mexican cohort and the IHAC cohort only.

substitutions at other codons, suggesting complex HLA-mediated escape pathways.

Interestingly, there were only two HIV *pol* sites previously associated with resistance to antiretroviral drugs that were also predicted to be associated with HLA selective pressure. B\*18 was associated with an E to A change in RT position 138. The polymorphism 138A is associated with decreased response to non-nucleoside RT inhibitors (NNRTIs), including etravirine (Stanford University HIV Drug Resistance Database, <http://hivdb.stanford.edu/>). Similarly, Cw\*07 was associated with a lower probability of having a D residue and a tendency for conservation of a V residue in RT position 179. The polymorphism 179D is associated with low level resistance to NNRTIs (Stanford University HIV Drug Resistance Database, <http://hivdb.stanford.edu/>). These observations show that HLA-mediated evolution can influence antiretroviral drug resistance, both promoting and preventing the presence of

drug-resistance-related polymorphisms. This dual pressure phenomenon has been described previously [35,47]; however, its frequency and population impact in the Mexican cohort will have to be assessed further.

HLA-HIV codon associations found for the Mexican cohort at q < 0.2 are presented in an epitope map in order to confirm the validity of the associations (Figure 6). 10 HLA-HIV codon pairs can be explained by experimentally confirmed epitopes, of which 5 have been optimally defined (Los Alamos HIV Database, <http://www.hiv.lanl.gov/content/immunology/index.html>). Twelve additional HLA-HIV codon pairs can be confirmed by epitope prediction with HLA peptide binding motifs (Motif Scan Tool, Los Alamos HIV Database, <http://www.hiv.lanl.gov/content/immunology/tools-links.html>). Eight HLA-epitope pairs could not be explained by epitope mapping, possibly because of lack of data on peptide binding motifs of associated HLA alleles

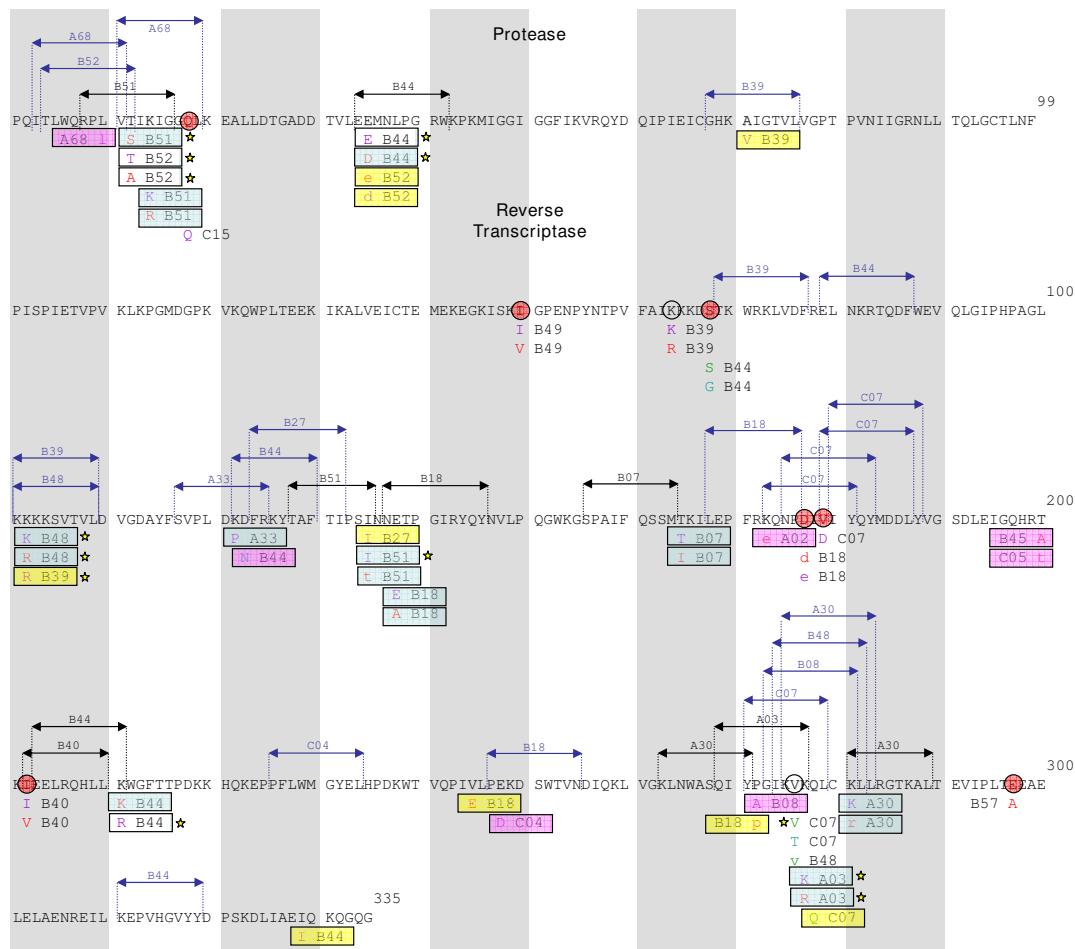


**Figure 5**  
**Protease and RT phylogenetic dependency network for the Mexican cohort.** A phylogenetic dependency network map was generated with the PhyloDv software <http://www.codeplex.com/MSCompBio>. Pol positions are drawn counter clockwise, with the N-terminus of the protease at the 3 o'clock position, and the first RT codon corresponding to position 100. Lines indicate associations between codons (inside the circle) or between HLA alleles and codons (outside the circle). Colors indicate q-values of the most significant associations between two attributes. Associations with  $q < 0.2$  for 280 individuals from Central/Southern Mexico were included.

(e.g. B\*39 and B\*49), or because of the presence of false positive associations (at  $q < 0.2$ , we expect 20% of the associations to be false positives). The possibility also exists that these associations represent escape mutations within unusual (novel) epitopes, or escape mutations that influence epitope processing that may occur far away from the actual epitope. Indirect or "one-hop" associations of the type  $a \rightarrow b \rightarrow c$ , where the HLA allele "a" is shown to predict the polymorphism "c", would be improbable as the multivariate model of the PDN model minimizes them. The same is true for associations with alleles in linkage disequilibrium with the selective allele, as linkage disequilibrium is accounted for by the PDN model. Some additional associations observed without taking codon covariation into account are also shown (Figure 6). Not

considering codon co-variation increases the power to detect associations, but allows the presence of indirect associations.

HLA-HIV codon associations found in the Mexican cohort were compared to the ones previously found in a dataset 1845 HIV *pol* sequences from the combined IHAC cohort [37] (Brumme ZL, John M, et al, PLoS ONE 2009, in press) (Figure 6, Additional file 1: Table S3). Finding HLA-HIV codon associations in the smaller Mexican cohort that were not observed in the well-powered IHAC cohort could be indicative of unique HLA-driven HIV evolution in immunogenetically distinct cohorts. Not surprisingly, many of the observed HLA-HIV codon associations in the Mexican cohort were also predicted in the IHAC cohort,



**Figure 6**  
**Epitope map supporting HLA-HIV polymorphism associations obtained by the Phylogenetic Dependency Network model in the Mexican cohort.** HLA-HIV polymorphism associations were searched in *pol* sequences derived from 280 individuals from Central/Southern Mexico using the previously described PDN model [14] that corrects for the confounding effects of HIV phylogeny, HIV codon covariation, and linkage disequilibrium of HLA alleles. Significant HLA-HIV polymorphism associations ( $q < 0.2$ ) were located in an epitope map. Strongly significant associations ( $q < 0.05$ ) are marked with yellow stars. Experimentally defined CTL epitopes are shown in black <http://www.hiv.lanl.gov/content/immunology/index.html>, epitopes predicted by HLA peptide binding motifs are shown in blue (Motif Scan, Los Alamos HIV Database, <http://www.hiv.lanl.gov/content/immunology/tools-links.html>). Target HIV positions are shown in bold, including the predictor HLA allele and the associated amino acid residue in different colors depending on the type of leaf distribution of the association obtained with the PDN model, i.e. escape (purple; having the predictor HLA makes it less likely to have the target amino acid), attraction (red; having the predictor HLA makes it more likely to have the target amino acid), reversion (green; not having the predictor HLA makes it more likely to have the target amino acid), repulsion (light blue; not having the predictor HLA makes it less likely to have the target amino acid). Amino acids in lower case indicate associations obtained without considering codon covariation. HLA-HIV polymorphism associations in the Mexican cohort were compared to the ones observed in the multi-center IHAC cohort including 1845 individuals from British Columbia, Canada, the USA and Western Australia [37] (Brumme ZL, John M, et al, PLoS ONE 2009, in press). Shared associations between the Mexican and IHAC cohorts are shown in blue rectangles; shared target HIV positions with shared target amino acids that were associated with different predictor HLA alleles in each cohort are shown in yellow rectangles; shared target positions with different HLA predictors and different associated target amino acids are shown in pink rectangles. HLA-associated HIV positions observed exclusively in the Mexican cohort are marked with red circles.

supporting the observation of highly conserved mutational patterns in HLA-driven HIV evolution (Figure 6, Additional file 1: Table S3). Nevertheless, important differences were also noted between the two cohorts. From the 43 HLA-HIV codon associations observed in the Mexican cohort, 23 were identified as novel associations, not previously observed in the larger IHAC cohort (nor in previous similar studies [25,30,34]), representing 18 different HLA-HIV codon pairs. Although several of these are likely to be false positives due to the 20% FDR, the fact that 53% of the present associations were not found in the well-powered IHAC cohort is striking. Furthermore, of these 18 new HLA-HIV codon pairs, 8 involved associations with codons that were not associated with any HLA allele in the IHAC cohort, and 4 involved associations with B\*39, which is substantially more frequent in the Mexican cohort than in the IHAC cohort. Remarkably, 17 of the 23 novel associations involved HLA alleles whose frequencies were statistically indistinguishable from those in IHAC, suggesting that their presence is not due to increased statistical power, but rather may be due to differences in patterns of epitope targeting. In addition, 2 of the novel associations involved B\*27 or B\*08, two alleles that were significantly less frequent in the present cohort than in IHAC, which may reflect differences in epitope targeting between the cohorts or the fixation (and resulting drop in statistical power) of escape mutations in the IHAC cohort [9]. Interestingly, although previously identified as HLA-associated in the IHAC cohort, some HLA-associated HIV codons in the Mexican cohort showed different HLA specificities and/or target amino acids. This was the case in 13 of the 23 novel HLA-HIV codon associations (10 of 18 HLA-HIV codon pairs). For example, Protease 71 V was associated with B\*39 in the Mexican cohort, while it was associated with B\*15 in the IHAC cohort; RT 245E was associated with B\*18 in the Mexican cohort, but with B\*57 in the IHAC cohort. Not surprisingly, B\*39 was much more frequent in the Mexican cohort ( $p = 1.80E-44$ ,  $q = 1.21E-42$ ), while B\*15 was more frequent in the IHAC cohort ( $p = 0.00964$ ,  $q = 0.0231$ ) (Figure 4). On the other hand B\*57 was less frequent in the Mexican cohort ( $p = 0.000430$ ,  $q = 0.00180$ ) and no significant difference was

found for B\*18. These associations can be explained by experimentally confirmed or predicted epitopes both in the Mexican cohort and in the IHAC cohort (Figure 6; Los Alamos HIV Immunology Database, <http://www.hiv.lanl.gov/content/immunology/index.html>).

Both B\*15 and B\*39 are predicted to have an epitope in position PR 68–76. These observations support the existence of sites in the HIV genome whose sequence variability at the population level reflects active selection pressure by different HLA alleles, and support previous observations that different HLA alleles may drive identical (as well as conflicting) escape mutations [48].

Interestingly, different consensus amino acids between the Mexican and the HOMER cohorts were detected at four Pol codons (Table 3). Two of these sites were HLA-associated in both cohorts, (RT 272 and RT 277), one was HLA-associated in only the HOMER cohort (PR 93) and one was not found to be associated to HLA in either cohort (RT 293) (although the possibility of an undetected association with HLA cannot be discarded). This observation supports the possibility of finding different HLA footprints in different populations, even between cohorts predominantly infected with viruses of the same clade. The finding that three of the four observed changes in consensus amino acids between the Mexican and the HOMER cohorts could have originated from HLA-driven pressure is also noteworthy. The role of HLA frequency in the fixation of escape mutations was evident in the A\*03-associated position RT 277, where the escaped form R has become fixed in the HOMER Pol consensus, while mainly remained as the susceptible form K in the Mexican Pol consensus. The A\*03 allelic frequency in the HOMER cohort was three times higher than in the Mexican cohort ( $p = 7.08E-10$ ,  $q = 1.00E-08$ ) (Table 3, Figure 4).

As previously observed [22,49], HLA-B alleles were involved in the majority of the associations (73% of HLA-HIV codon pairs), compared to HLA-A and C alleles in the Mexican cohort ( $p < 0.001$  for both cases). Interestingly, 16.6% of HLA-HIV codon pairs were due to HLA-C alleles. Although in previous studies many of the associations

**Table 3: Point differences in consensus sequences between the Mexican and the HOMER cohorts.**

HIV position	Susceptible	Escaped	Mexican Consensus	HOMER Consensus	B Consensus	Associated HLA Mexican cohort	Associated HLA HOMER cohort	HLA frequency Mexican cohort	HLA frequency HOMER cohort
PR 93	I	L	L	I	I	-	B*15	0.0702	0.1131
RT 211	R	G, K	K	K	R	B*44	B*44, B*15	0.0616†	0.1059†
RT 272	P, A	S	P	A	A	B*08	B*42	0.0154†	0.0693†
RT 277	K	R	K	R	K	A*03	A*03	0.0411	0.1235
RT 293	-	-	I	V	I	-	-	-	-

†Frequency of the associated HLA allele in the Mexican cohort is shown.

apparently defined by HLA-C alleles represented indirect associations with HLA-B or A alleles due to the HLA linkage disequilibrium phenomenon, the PDN model used in the present study accounts for HLA linkage disequilibrium, minimizing the risk of finding these and other kinds of indirect associations. Nevertheless, it is important to note, that the ability of the PDN model to correct for HLA linkage disequilibrium is positively correlated with sample size and negatively correlated with the strength of linkage disequilibrium. Thus, false positive associations could still be found when strong linkage disequilibrium patterns exist, and random noise makes it difficult to distinguish the true associations. This could be the case for Cw\*15-associated position PR 18, which could not be explained by epitope prediction. Cw\*15 and B\*51 are in strong linkage disequilibrium in the Mexican cohort (Additional file 1: Table S2), and an experimentally confirmed B\*51 epitope exists that could explain the association. Although many of the HLA-C associations had high *q*-values, and could represent false positive associations, some of them were strongly associated and were in consonance with predicted and verified epitope mapping (Additional file 1: Table S3, Figure 6). These observations suggest an important role of HLA-C alleles in shaping HIV evolution at the population level in the Mexican cohort.

On the other hand, B\*44 alone was responsible for 16.6% of HLA-HIV codon pairs, followed by B\*51 and B\*39, each responsible for 10% of the observed HLA-HIV codon pairs (Additional file 1: Table S3, Figure 6). This predominance is not observed in the IHAC cohort, possibly suggesting different patterns of immunodominance and HIV immune escape resulting from different epitope targeting between the two cohorts. These observations further support the existence of differential patterns of HIV selection by HLA alleles in populations worldwide.

Taken together, these results support the existence of highly conserved, universal HLA-mediated mutational patterns or "footprints" on HIV sequences at the population level. However, they also suggest that unique characteristics could exist in HLA-mediated HIV evolution in immunogenetically distinct populations, which can be detected even with cohorts of relatively small number of individuals.

#### **Differences in HIV evolution between free plasma virus and PBMC proviral sequences**

A subset of 250 HIV-infected individuals for whom HLA typing and both *pol* PBMC proviral sequences and free plasma virus *pol* RNA sequences were available was used to compare HLA associated polymorphisms in two different viral compartments at the population level. The PDN model was applied to both sequence datasets and results were graphically depicted with the PDN viewer PhyloDv

[14] (Figure 7). In all, 36 HLA-HIV codon associations were found for the free plasma virus sequences and only 24 for the PBMC proviral sequences, representing 27 and 15 HLA-HIV codon pairs respectively (Figure 7, Table 4). Interestingly, the number of unique HLA-HIV codon pairs observed in free plasma virus sequences was significantly higher than the number of unique pairs in proviral sequences ( $p = 0.0169$ ), and only 10 of the HLA-HIV codon pairs were observed in the two viral compartments. These results are consistent with a recent study that reported the presence of HLA-associated escape mutations in plasma sequences that were rarely seen in the proviral population within some infected individuals [50]. The observation that an overall different evolution was seen in plasma viral sequences compared to proviral sequences, with a significantly lower number of HLA-associated sites in proviral sequences, is consistent with the model that suggests that proviral sequences represent early archived HIV in the latent reservoir and that plasma sequences represent a population that has evolved further in response to immune selective pressure. Furthermore, these observations are suggestive of a dynamic development of CTL responses throughout the infection, such that early CTL responses are reflected in the archival proviral compartment while the plasma compartment reflects more recent CTL responses [51,52]. We note, however, that proviral HLA-HIV site associations did not correlate with previously defined rapidly escaping sites under HLA pressure in clade B-infected Caucasian individuals [25]. This fact might reflect different rates of escape between demographically divergent cohorts, or it might reflect differing compartment-based CTL selective pressures that are simply reflective of the archival nature of the proviral sequences.

Interestingly, the number of covarying HIV sites common to both compartments was lower than the number of covarying sites observed exclusively in plasma or in proviral sequences ( $p = 0.0489$ ). Moreover, a large number of covarying sites was seen in proviral HIV sequences possibly reflecting remnants of viral adaptation to previous hosts.

Overall, differences in HLA-mediated selection were observed in the plasma virus and the PBMC provirus compartments, suggesting a highly dynamic HLA-associated evolution in HIV, as many of the HLA-HIV codon associations in the free plasma virus compartment were not evident in the proviral dataset, which likely contains early archived HIV sequences that appear to reflect less adaptation to within host HLA-mediated immune responses.

#### **Discussion**

In this study we have presented evidence suggesting that a unique HLA allele frequency distribution in a cohort of clade B-infected Mexican individuals has left unique foot-

**Table 4: Protease and Reverse Transcriptase HLA-HIV codon associations in free plasma virus RNA and PBMC proviral DNA for the Mexican cohort.**

Free plasma virus <i>pol</i> sequences†					PBMC proviral <i>pol</i> sequences†				
HLA	Associated HIV Codon	Conditioning variables	p-value	q-value	HLA	Associated HIV codon	Conditioning variables	p-value	q-value
					A*02 <sup>c</sup>	PR 15I	PR 77V <sup>c</sup> , PR 36M <sup>d</sup> , PR 13I <sup>a</sup>	0,0010442	0,1522376
A*03 <sup>b</sup>	RT 277R		1,14E-08	1,10E-05	A*03 <sup>b</sup>	RT 277R		1,74E-09	1,48E-05
A*03 <sup>a</sup>	RT 277K		1,14E-08	1,10E-05	A*03 <sup>a</sup>	RT 277K		1,74E-09	1,48E-05
A*11 <sup>b</sup>	RT 333E	PR 35E <sup>c</sup> , RT 293I <sup>a</sup> , RT 207Q <sup>d</sup> , PR 62I <sup>d</sup>	0,00113806	0,1336789					
A*30 <sup>c</sup>	RT 281K	RT 278Q <sup>d</sup>	0,00109204	0,1336789					
A*68 <sup>a</sup>	RT 326V		0,00035689	0,1048983					
B*07 <sup>b</sup>	RT 165I		0,00067347	0,1191747					
B*07 <sup>a</sup>	RT 165T		0,00067347	0,1191747					
B*18 <sup>b</sup>	RT 138A		0,00012948	0,0494756					
B*18 <sup>a</sup>	RT 138E		0,00012948	0,0494756					
B*18 <sup>b</sup>	RT 245E	PR 45K <sup>c</sup>	0,00068267	0,1191747					
B*18 <sup>b</sup>	RT 177D		0,00094587	0,1312293	B*18 <sup>b</sup>	RT 177D	RT 123D <sup>d</sup> , RT 121D <sup>d</sup>	0,0017257	0,19555
B*18 <sup>a</sup>	RT 177E		0,00094513	0,1312293	B*18 <sup>a</sup>	RT 177E		0,0006388	0,1183628
B*18 <sup>a</sup>	RT 245V		0,00115452	0,1336789					
B*27 <sup>a</sup>	RT 135I	B*51 <sup>a</sup> , PR 12T <sup>d</sup> , RT 202I <sup>d</sup>	0,00096164	0,1312293					
B*39 <sup>a</sup>	RT 102R	B*48 <sup>b</sup> , RT 64K <sup>a</sup>	0,00016533	0,0539628					
B*39 <sup>c</sup>	PR 71V	PR 93L <sup>d</sup> , PR 39P <sup>a</sup> , PR 57K <sup>d</sup> , Cw*15 <sup>b</sup>	0,00193116	0,1994313	B*39 <sup>c</sup>	PR 71V	PR 57R <sup>a</sup> , PR 93I <sup>c</sup> , Cw*15 <sup>d</sup> , RT 178M <sup>b</sup> , PR 39P <sup>a</sup>	0,0003592	0,0888581
					B*39 <sup>a</sup>	PR 60E	PR 39Q <sup>b</sup> , PR 37D <sup>b</sup> , PR 37E <sup>b</sup>	0,0019258	0,199068
B*44 <sup>b</sup>	PR 35D		1,16E-08	1,10E-05	B*44 <sup>b</sup>	PR 35D		5,77E-08	0,0002091
B*44 <sup>a</sup>	PR 35E		1,16E-08	1,10E-05	B*44 <sup>a</sup>	PR 35E		5,77E-08	0,0002091
B*44 <sup>a</sup>	RT 211R		0,00112279	0,1336789					
B*44 <sup>a</sup>	RT 329I	RT 334L <sup>b</sup>	0,00120957	0,1359348					
B*48 <sup>b</sup>	RT 102R		2,47E-08	1,89E-05	B*48 <sup>b</sup>	RT 102R		1,53E-11	3,88E-07



**Table 4: Protease and Reverse Transcriptase HLA-HIV codon associations in free plasma virus RNA and PBMC proviral DNA for the Mexican cohort. (Continued)**

B*48 <sup>a</sup>	RT	102K		0,00016947	0,0539628	B*48 <sup>a</sup>	RT	102K		1,90E-05	0,0120849
						B*48 <sup>a</sup>	RT	276V	RT 272P <sup>c</sup> , Cw*07 <sup>c</sup>	0,0005419	0,1049938
						B*49 <sup>a</sup>	RT	48S		0,0004979	0,1035941
						B*49 <sup>b</sup>	RT	48T		0,0004979	0,1035941
B*51 <sup>a</sup>	RT	135I		4,03E-07	0,0002198						
B*51 <sup>b</sup>	PR	12S		0,00038855	0,1060476						
B*51 <sup>b</sup>	PR	14R	RT 135I <sup>a</sup> , RT 200A <sup>a</sup>	0,00144043	0,1528856	B*51 <sup>b</sup>	PR	14R		7,41E-06	0,005226
B*51 <sup>a</sup>	PR	14K	RT 135I <sup>b</sup> , RT 200A <sup>b</sup>	0,00138006	0,1506635	B*51 <sup>a</sup>	PR	14K		7,41E-06	0,005226
B*52 <sup>b</sup>	PR	12A		5,59E-07	0,0002671	B*52 <sup>b</sup>	PR	12A	RT 135T <sup>d</sup>	0,0005416	0,1049938
B*52 <sup>a</sup>	PR	12T		3,73E-08	2,38E-05	B*52 <sup>a</sup>	PR	12T		1,60E-06	0,0016939
B*57 <sup>b</sup>	RT	297A		0,00060553	0,1191747	B*57 <sup>b</sup>	RT	297A		0,0004121	0,0941468
						Cw*03 <sup>b</sup>	RT	283I		0,0017747	0,1964929
						Cw*03 <sup>a</sup>	RT	283L		0,0017747	0,1964929
Cw*03 <sup>b</sup>	PR	12P	PR 14R <sup>d</sup> , PR 67C <sup>c</sup>	0,00070538	0,1191747						
Cw*05 <sup>b</sup>	RT	214L	RT 118I <sup>b</sup>	0,00065426	0,1191747	Cw*05 <sup>b</sup>	RT	214L	RT 118V <sup>c</sup>	0,0002149	0,0665192
Cw*05 <sup>a</sup>	RT	214F	RT 118I <sup>a</sup>	0,00065426	0,1191747	Cw*05 <sup>a</sup>	RT	214F	RT 118V <sup>d</sup>	0,0002149	0,0665192
Cw*07 <sup>c</sup>	RT	276V	RT 272P <sup>c</sup>	0,00054742	0,1191747	Cw*07 <sup>c</sup>	RT	276V	RT 272P <sup>c</sup>	0,0004154	0,0941468
Cw*07 <sup>a</sup>	RT	179D		0,00072749	0,1191747						
Cw*12 <sup>b</sup>	RT	162S	PR 64L <sup>b</sup> , RT 165T <sup>c</sup> , PR 91T <sup>c</sup> , RT 135V <sup>a</sup>	0,00106536	0,1336789						
Cw*15 <sup>a</sup>	PR	18Q		0,00074855	0,1191747						
Cw*15 <sup>b</sup>	PR	71V	PR 93L <sup>d</sup> , PR 39P <sup>a</sup> , PR 57K <sup>d</sup>	0,0009574	0,1312293						

†Associations for 250 plasma virus sequences and the corresponding PBMC proviral sequences with  $q < 0.2$  are shown. <sup>a</sup> escape (having the predictor makes it less likely to have the target), <sup>b</sup> attraction (having the predictor makes it more likely to have the target), <sup>c</sup> reversion (not having the predictor makes it more likely to have the target), <sup>d</sup> repulsion (not having the predictor makes it less likely to have the target).

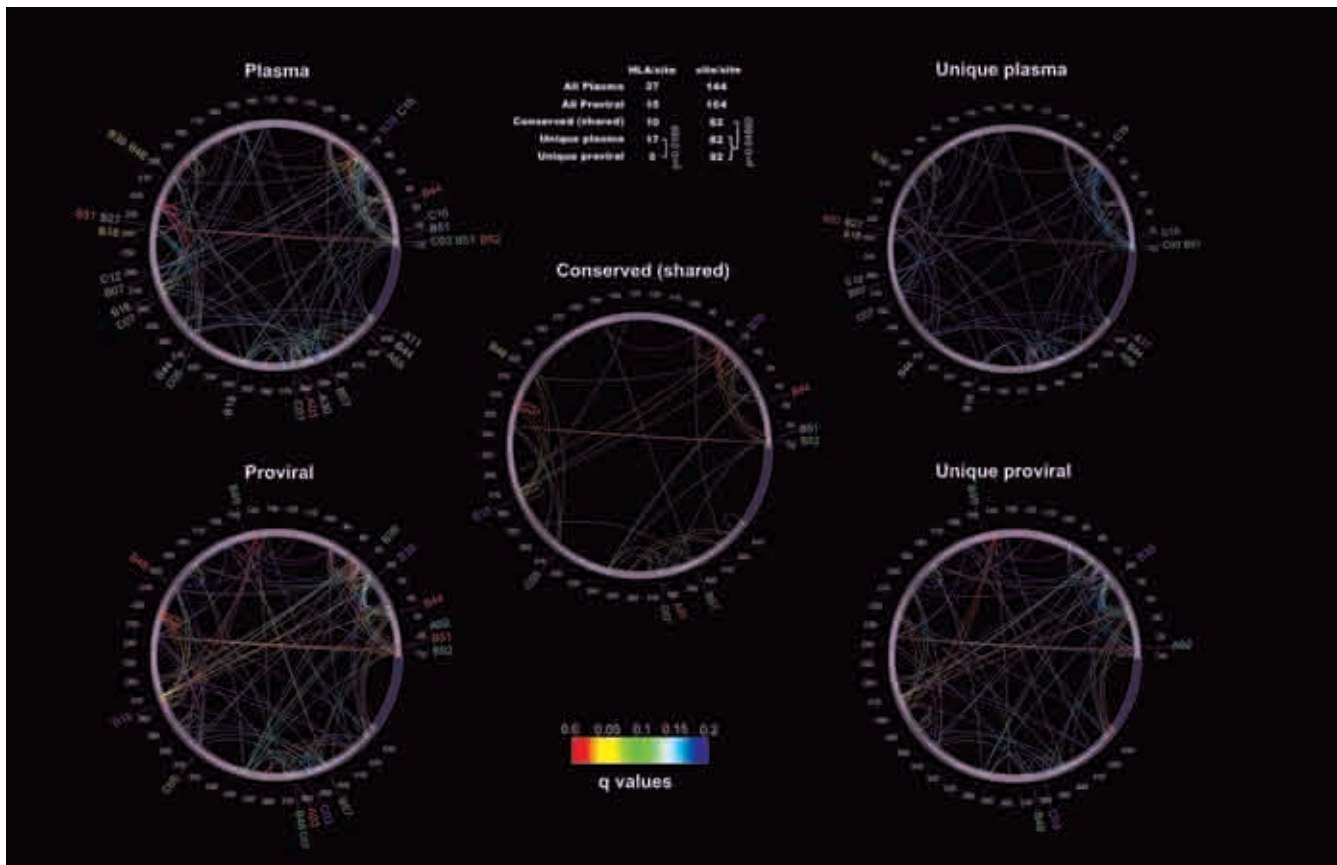
prints on HIV sequences at the population level. We studied HLA-mediated HIV evolution in a clade B-infected Mexican cohort, comparing our data with data from the IHAC cohort, the largest clade B-infected cohort used to assess HLA-mediated evolution so far, which is composed of individuals from Canada, Australia, and the USA [37] (Brumme ZL, John M, et al, PLoS ONE 2009, in press). The two cohorts were shown to present notably different immunogenetic backgrounds, with an important admixture of Amerindian genes in the population of the Central/Southern part of Mexico (Figure 4). These different immunogenetic backgrounds provided a chance to assess the role of different HLA allele distributions in HLA-mediated selection in two cohorts infected by viruses of the same clade. The present cohort was shown to reflect the typical characteristics of an HIV-infected Mexican cohort, enriched in individuals in relatively advanced stages of HIV disease and presenting a similar HLA allele frequency distribution to the general population (Figure 3), with some specific exceptions (e. g. B\*39) that will have to be assessed further in future studies.

Previous studies have suggested that HIV evolution at the population level follows broadly predictable, highly conserved mutational patterns associated with host CTL selective pressure [16,23,24,33-35]. Conclusions obtained from a direct comparison between different studies in different populations have been limited mainly due to the use of different methods and models for assessing HLA-mediated viral evolution in which important sources of confounding are frequently not accounted. We applied the recently described PDN model [14], which simultaneously accounts for HLA linkage disequilibrium, HIV codon co-variation and viral lineage effects, to clade B *pol* sequences (Additional file 1: Figure S2) from the present cohort and compared the results to the immunogenetically distinct population of the IHAC cohort. Our data support the observations of highly conserved, universal, HLA-associated footprints in the HIV proteome at the population level, as many of the HLA - HIV codon associations found in the Mexican cohort have consistently been observed in the IHAC cohort, as well as in previous studies with diverse cohorts [24,33-35,37].

Interestingly, however, our data also suggest the existence of unique HLA-associated footprints in HIV, which could be influenced by specific HLA frequency distributions in different HIV-infected populations. The unique characteristics of HLA-mediated selection in the Mexican cohort was revealed not only by the presence of unique HLA-HIV codon pairs not detected in the IHAC cohort, but also by the presence of HIV positions previously identified as HLA associated, and with different HLA specificities and/or target amino acids in the two cohorts. The extent to which these unique HLA-associated footprints represent a

real biological phenomenon and not a statistical effect will have to be further assessed with experimental data; nevertheless, evidence presented in this study strongly suggests the existence of real differences between the two cohorts. Although the Mexican cohort was much smaller than the IHAC cohort (the power to detect associations increases dramatically with sample size [14]), resulting in only 20% of the expected associations being confirmed in the present cohort, the fact that 53% of the HLA-HIV codon associations were novel in the Mexican cohort strongly suggests differences in HLA-mediated evolution between the two clade B-infected cohorts. Although these novel associations may represent false negatives from the IHAC cohort, that cohort is large enough that the false negative rate is expected to be quite small and any false negatives are likely to be rare events in it [14]. It is also possible that the novel associations represent false positives in the present cohort; however, with an expected 20% false-positive rate due to the  $q < 0.2$  threshold, the number of novel associations found in the Mexican cohort is striking. In addition, of the 18 novel HLA-HIV codon pairs in the Mexican cohort, at least 11 (61%) can be explained by confirmed or potential CTL epitopes (Figure 6), strongly arguing for the validity of these associations and for the existence of real biological differences in HLA-mediated selection between the two cohorts

The observation of point differences in the population consensus sequences of the two cohorts which were mapped to HLA-associated sites is a piece of evidence that further supports the differential impact of HLA selection in HIV evolution at the population level (Table 3). This was the case of position RT 277, associated with A\*03 both in the Mexican cohort and in the IHAC cohort, in which the adapted form 277R has become fixed in the IHAC consensus while the susceptible form 277K has remained in the Mexican consensus. Not surprisingly, the frequency of A\*03 was three times higher in the HOMER cohort than in the Mexican cohort ( $p = 7.08E-10$ ,  $q = 1.00E-08$ ), supporting an important role of HLA allele frequency in the fixation of HLA escape mutations at the population level (Figure 4, Table 3). Similarly, PR 93 was associated with B\*15 in the IHAC cohort with the susceptible form 93I observed in this cohort's consensus, but the adapted form 93L observed in the Mexican consensus. Although no direct HLA association was detected in the Mexican cohort at this site (probably due to statistical power issues), position PR 93 was associated with other HIV sites, such as PR 71, which is HLA associated (Figure 6). Thus, changes in population consensus sequences may be linked with HLA-mediated selection. Also of interest is the observation that B\*44 was associated with 5 of the 30 HLA-HIV codon pairs identified in the Mexican cohort. Two of these associations have been described in the IHAC cohort and two have been previously identified as



**Figure 7**  
**HLA-HIV codon and HIV codon-HIV codon associations for the free plasma virus and PBMC proviral Pol sequences.** Phylogenetic dependency network maps were generated with the PhyloDv software <http://www.codeplex.com/MSCompBio>. Each map shows HLA associations as tags pointing to their corresponding sites, and HIV codon-HIV codon associations as inner arcs connecting the associated sites. Plasma: all associations in free plasma virus from 250 patients that had a corresponding PBMC proviral sequence. Proviral: all associations for 250 PBMC proviral sequences. Conserved (shared): shared HLA-site or site-site associations between proviral and plasma sequences. If one amino acid at one site was associated with two different amino acids at another site (e.g. one for escape, one for reversion), only one association was counted. Unique plasma: associations only found to be significant in free plasma virus samples. Unique proviral: Associations only found to be significant in PBMC proviral samples. Significance was calculated with the binomial test; q-values are represented as a heat map shown in the lower inset. Associations with  $q < 0.2$  are shown (all associations had  $p < 0.002$ ). The upper inset table shows the number of significant associations in each map.

HLA-associated positions with different HLA specificities (Figure 6). The strong influence of B\*44 on HLA-mediated HIV evolution in the Mexican cohort could reflect differences in immunodominance hierarchies of CTL responses in the context of different HLA frequency distributions. Whereas strongly immunodominant CTL responses could be masking the effect of other less immunodominant responses in one cohort, these responses could have a greater impact on HLA-mediated HIV evolution in another cohort in which the immunodominant responses are infrequent. It is notable that the frequencies of many strongly immunodominant HLA alleles, such as B\*57, B\*27, B\*08, B\*07, A\*03, and A\*11 [25], are lower in the Mexican cohort compared to the IHAC cohort ( $q <$

0.05) (Figure 4). It is possible that in the latter cohort, CTL responses restricted by these alleles could be masking the effect of other less immunogenic alleles that are frequently seen in the Mexican population. Indeed, this could be the case for Cw\*07, the most frequent HLA-C allele group in the Mexican cohort, which explains 10% of HLA-HIV codon pairs observed in our analysis. These associations are unique to the Mexican cohort, and are supported by predicted epitopes and/or a strong statistical association ( $q < 0.05$ ) (Figure 6).

The case of B\*39 is also noteworthy, being the most frequent HLA-B allele group in the Mexican cohort with a frequency 7 times higher than that observed in the IHAC

cohort ( $p = 1.80E-44$ ,  $q = 1.21E-42$ ) (Figure 4). B\*39 explained another 10% of the HLA-HIV codon pairs identified in the Mexican cohort, suggesting either a strong influence of this allele in HIV evolution in the immunogenetic context of the Mexican cohort or a higher statistical power to detect associations. Interestingly, B\*39-restricted associations were primarily escape associations (where possession of B\*39 made it less likely to have the target amino acid in question), in which the target amino acid was a residue other than the consensus, suggesting that the consensus residue represents a possible escaped form for B\*39 at this position (Additional file 1: Table S3, Figure 6). This could be suggestive of a frequent role of the B\*39 allelic group in HIV codon conservation in the Mexican cohort. This HLA-associated conservation of sites has been previously described with highly frequent HLA alleles that promote the accumulation of CTL adapted variants in different populations [22,23].

Overall, two general key aspects could explain the observation of different associations in cohorts that are infected by viruses of the same clade but which have different HLA frequencies: 1) Different patterns of immunodominance, which argue for real differences in CTL epitope targeting; and 2) Different statistical power to detect associations, which argues for a statistical effect rather than a biological difference. For example, the absence of strong immunodominance patterns in certain populations could potentially facilitate the detection of HIV polymorphisms associated with less immunodominant alleles. Being able to confirm this possibility at the population level strongly relies on low false positive/negative rates. Although the false negative rate on the IHAC data is low, further experimental data is necessary to confirm this point. On the other hand, different HLA frequencies can simply change the statistical power to detect associations, thus supporting the importance of assessing HLA-mediated selection in a diverse set of cohorts. The possibility also exists that a simple statistical power issue could be resolved by combining different cohorts infected by the same viral clade to make a larger reference set, supporting the creation of a universal set of associations that could get updated periodically as new sequences are added. Such a sequence and association database would allow extrapolation from a large reference set to new demographic groups for which collection of cohorts would be difficult. The fact that 17 of the 23 novel HLA-HIV codon associations in the Mexican cohort involved HLA alleles whose frequencies were not significantly different from those in the IHAC cohort strongly suggests that their presence is not due to increased statistical power but rather may be due to differences in patterns of epitope targeting. Furthermore, immunodominance effects as well as statistical power issues depending on HLA frequencies could both exist in the same dataset. Examples of both phenomena have

been described above for the Mexican cohort, suggesting that a set of immunogenetically diverse cohorts could greatly enrich HIV evolutionary studies without the need of very large cohorts. It should also be noted that a broad two-digit HLA allele grouping does not reveal all possible divergence in HLA pressure, as a number of HLA subtypes with different peptide-binding motifs can be defined at four-digit level within some allelic groups such as B\*35, B\*40, B\*51, B\*58, A\*02, all with highly characteristic distributions in different populations [9,49]. Thus, significant divergence in selection in some cases could be explained by different dominant four-digit subtypes of the broad allele group in the compared cohorts. This fact could have an impact on statistical power to detect associations defined by different subtypes within a broad allele group in different populations and further argues for the unique HLA-associated imprinting of HIV in different populations.

In summary, although important limitations exist for the analysis of HLA-mediated HIV evolution in the Mexican population, including the presence of false positive associations and the low power to detect associations, our analysis yielded strong evidence suggesting that unique characteristics in HLA-mediated HIV evolution in the Mexican cohort indeed exist. These include the striking proportion of unique HLA-HIV codon associations in the Mexican cohort (many of which can be supported by predicted or confirmed CTL epitopes), the presence of HLA-associated differences in the consensus sequence with respect to the HOMER consensus (which reflects differential fixation of CTL escape mutations at the population level with a high dependency on HLA frequency), and the existence of a high proportion of novel associations that involve HLA alleles whose frequencies were similar in the Mexican and the IHAC cohorts (which argues against a statistical power issue in detecting at least some of the significant associations).

To further characterize HLA-mediated HIV evolution, HLA-HIV codon and HIV codon-HIV codon associations were compared in free plasma virus and PBMC proviral DNA in the cohort of Mexican individuals. As shown by graphically depicting the PDNs for the two viral compartments, different mutational patterns and different HLA-HIV codon associations were seen in actively replicating plasma viruses and PBMC-archived proviruses at the population level. A significantly lower number of HLA-HIV codon associations was observed in proviral sequences and there were more distinct than shared HIV codon-HIV codon associations in the two compartments (Figure 7). This could be explained by the observation that proviral sequences frequently represent a stable reservoir of HIV sequences archived early in the course of the infection [53], whereas plasma viruses represent sequences from

later in the course of infection. Thus, the proviral sequences may have been archived before some epitopes were targeted by host CTL responses, or before escape mutations had a chance of being selected at epitopes already being targeted by CTLs, resulting in fewer associations in proviral sequences than in the extant plasma sequences. Indeed, previous studies have shown the presence of HLA-associated escape mutations in plasma viruses that are rare in proviruses within infected individuals [54]. Nevertheless, proviral HLA-HIV codon pairs could not be mapped to known epitopes of early escape [25] in the present data, although the possibility exists that a larger cohort and analyses in other viral genes could further support this correlation. However, given the differences in escape association that we have observed between the Mexican and IHAC cohorts and the observation that the cohort described [25] is immunogenetically similar to the IHAC cohort, it may be that the discordance between proviral escape associations reported here and previously reported early-escape epitopes reflects different patterns of CTL epitope targeting and kinetics between the two populations. The proviral associations in the Mexican cohort could thus represent early escape events in a Latin American cohort setting.

Surprisingly, some HLA associations detected for proviral sequences were not seen in the plasma virus dataset. Some of these HLA-HIV codon pairs observed exclusively in proviral sequences have fairly high *q*-values, possibly suggesting the presence of false positive associations. However, unique proviral associations could also suggest a chronological reshaping of HLA-mediated HIV evolution, reflecting rapidly reverting mutations which are lost soon after transmission to HLA-mismatched individuals. Alternatively, the existence of organ compartmentalization of HIV variants within an infected host and its relation to positive selection has been described [55]. This phenomenon could explain population differences between actively replicating viruses coming from a specific compartment with characteristic selective pressures and archived proviruses, remaining as reservoir(s) originating from different anatomical and/or cellular compartments.

Shared associations between the plasma virus and the PBMC provirus compartments may reflect sites in the viral proteome with continuous CTL targeting throughout the chronic infection, a characteristic that might be of interest in the selection of candidate vaccine targets. On the other hand, these apparently more stable associations could also reflect epitopes with early CTL targeting that has stopped, but for which no reversion has occurred, suggesting low fitness costs for escape. If the latter case were true, some shared associations might be more likely to reach fixation at the population level in the future. This would

have implications for our understanding and predictive capabilities of HIV adaptation in human populations.

Similarly, unique coevolving HIV codon pairs were detected in proviral sequences and in plasma virus sequences, perhaps reflecting different patterns of compensatory mutations to the different HLA escape mutations observed in the two compartments. Alternatively, unique proviral HIV codon-HIV codon pairs could be explained as a reorganization of mutational patterns in HIV evolution that reflect escape mutations selected in previous hosts as well as new mutations selected in the current host, while unique plasma virus HIV codon-HIV codon pairs could reflect sequential footprints left by viral adaptation to HLA-restricted responses in chronic infection in the current host. These observations bring up interesting consequences for our understanding of HLA-mediated HIV evolution, suggesting that the appearance and density of the PDNs for a specific population are highly dynamic and could vary in time. The dynamic development of CTL responses over the course of infection within an individual has been previously reported [51,52]. Further studies in follow-up cohorts or in carefully stratified cross-sectional cohorts might be able to support or refute these observations.

## Conclusion

In conclusion, our data derived from analysis of HLA-mediated HIV evolution in a previously uncharacterized, immunogenetically unique cohort from Central/Southern Mexico, support a highly conserved and strongly predictable component of HLA-mediated HIV evolution at the population level, resulting in HLA-associated footprints in the circulating virus worldwide. This effect is evident even after considering important objections to the HLA-HIV population imprinting hypothesis, such as the rapid reversion of a considerable part of the total CTL escape mutations in the absence of the selective HLA allele [56], the complexity of the CTL response which frequently imposes conflicting selective forces in the same site of the viral sequence [27], the possibility that an escape variant selected by a specific HLA allele can be targeted by CTL responses restricted by different HLA alleles [18,30,57], and the dynamic immunodominance hierarchies observed in HLA-restricted responses [51,52]. Interestingly, the HLA-mediated evolution analysis in our cohort of Mexican individuals showed additional HLA-HIV codon associations that have not been described in previously studied cohorts, including the large multi-center IHAC cohort with a clearly different immunogenetic background to the Mexican cohort. This fact supports the possibility that these specific associations are not significantly impacting HIV evolution at the population level in other cohorts, but that they are significant in the immunogenetic context of the Mexican population. Comparative

HLA-mediated HIV evolution studies, with comparable methods that take into account important confounding factors such as HIV codon covariation, HIV lineage effects and HLA linkage disequilibrium, can thus be useful in identifying these distinct HLA-associated footprints in different populations worldwide. Extending such comparative studies to other immunogenetically distinct cohorts, would allow the reconstruction of a more complete panorama of the impact of HLA selection in HIV evolution worldwide. This knowledge may prove useful for the development of vaccine candidates and the development of therapeutic strategies directed to specific populations. The creation of a universal database of HLA-associated HIV sites applying comprehensive and comparable models to assess HLA-mediated evolution in immunogenetically divergent cohorts from different parts of the world, including cohorts predominantly infected by different viral clades, would greatly improve our understanding of HIV evolution worldwide. Importantly, further experimental evidence will help to understand the limitations imposed by statistical models to detect footprints of HLA-associated evolution in HIV in different populations.

Additionally, a comparison between HLA-mediated evolution in free plasma virus and PBMC proviral sequences suggested a highly dynamic HLA-associated evolution in HIV, as many of the HLA-HIV codon associations observed in the free plasma virus compartment are not evident in the proviral dataset, which is presumably enriched in early HIV sequences and does not reflect the full extent of within-host HLA-driven viral evolution. Moreover, shared HLA-HIV codon associations in both viral compartments could be of interest, reflecting epitopes with continuous CTL targeting throughout the chronic infection or, alternatively, escape mutations with low fitness costs that could reach fixation at the population level in the future. Further studies with larger cohorts and various viral genes could enrich these primary observations and increase our understanding of HIV adaptation to different populations worldwide.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

SAR, CEO, EE and GRT conceived and directed the project. SAR wrote the manuscript. CEO carried out and revised statistical analyses. JMC and DH carried out HLA-mediated HIV evolution analyses applying the PDN model. SAR, HVP, and JBH carried out HLA typing. SAR and JBH carried out proviral HIV *pol* sequencing. DGR and CGM carried out HIV *pol* sequencing and coordinated shipping and processing of blood samples. ZLB allowed the use of the HOMER cohort and provided data for the comparisons with the Mexican cohort. SM and MJ allowed the use

of the Perth and the USA cohort data. SAR, EE, GRT, CEO, JMC, DH, ZLB, MJ and SM were involved in critically revising the manuscript.

### Additional material

#### Additional file 1

**Two additional figures and three additional tables.** Figure S1 – Allelic and population frequencies for class I HLA-A, B and C genes in a cohort of HIV-infected individuals from Central/Southern Mexico. Frequencies were obtained with the HLA Frequency Analysis Tool of the Los Alamos HIV Immunology Database <http://www.hiv.lanl.gov/content/immunology/tools-links.html>. HLA typing was carried out by SSP-PCR as described in the Methods. A total of 292 individuals (584 HLA alleles) were included; 19, 29 and 14 distinct allelic groups were observed for the HLA-A, B and C genes respectively. Figure S2 – Phylogeny of 280 HIV *pol* sequences from a cohort of antiretroviral treatment-naïve individuals from Central/Southern Mexico. A Neighbor-Joining tree was inferred through the analysis of 1305 bp *pol* sequences including the whole protease and 335 codons of the RT in MEGA 4.0. The consensus tree from 1000 bootstrap replicas is shown. Evolutionary distances were calculated with Kimura's two-parameter model and are shown in substitutions per site. Positions with missing information were eliminated by pairwise comparison. 93 reference sequences of the main HIV-1 groups, subtypes and recombinant forms, obtained from the Los Alamos HIV Sequence Database <http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html> were included. Reference subtype B sequences are shown in blue, reference sequences of all other subtypes and recombinant forms are shown in red and Mexican sequences are shown in black. The inset shows a detail of the Mexican sequence cluster. Table S1 – HLA-A, B and C population and allelic frequencies in the Mexican cohort. Table S2 – Linkage disequilibrium for class I HLA-A, B and C genes in the Mexican cohort. Table S3 – Protease and RT HLA-HIV codon associations observed in the Mexican cohort.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1742-4690-6-72-S1.pdf>]

### Acknowledgements

We thank all patients of the Mexican cohort for their participation in this study; the physicians Akio Murakami, María Gomez-Palacio, José L. Sandoval, Daniela de la Rosa, Jorge Ibarra, Ricardo S. Vega and Cristina Sánchez of the Center for Research in Infectious Diseases of the National Institute of Respiratory Diseases in Mexico City, for their help in recruiting patients; Carolina Demeneghi, Mario Preciado, and Silvia del Arenal, for collection of blood samples; Ramón Hernández, and Verónica Quiroz, for viral load, and HIV genotyping assays; Edna Rodríguez for CD4+ T cell count assays; Zeidy Arenas, Sandra Zamora, Rosalinda Hernández, Eduardo López, for their administrative support; Dr. Joel Vázquez, for his technical guidance; Dr. Luis Padilla-Noriega and Dr. Eduardo García-Zepeda for their academic counselling. We thank Dr. Indiana Torres, Dr. Beatriz Ramírez, Dr. Adakatia Armenta, Dr. Jaime Andrade and Dr. Lucero González for providing blood samples of individuals from different states in Mexico. We thank the Doctorate in Biomedical Sciences Program and the National Autonomous University of Mexico, for their support. We thank Dr. Bruce Walker for his mentorship and support, and the BC Centre for Excellence in HIV/AIDS for providing access to data. We thank Dr. Richard Harrigan and the BC Centre for Excellence in HIV/AIDS for providing access to data on the HOMER cohort.

Sponsorship: This work was supported by grants from CONACYT U48159M, Fundación México Vivo, and Comisión de Equidad y Género de la H. Cámara de Diputados. SAR was supported by a scholarship from CONACYT. ZLB is supported by a post-doctoral fellowship from the Canadian Institutes for Health Research (CIHR).

## References

- Allen TM, Altfeld M, Geer SC, Kalife ET, Moore C, O'Sullivan KM, Desouza I, Feeney ME, Eldridge RL, Maier EL, et al.: **Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type I (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution.** *J Virol* 2005, **79**:13239-13249.
- Blankson JN, Bailey JR, Siliciano RF: **Crosscurrents in HIV-1 evolution.** *Nat Immunol* 2006, **7**:121-122.
- Borrow P, Lewicki H, Wei X, Horwitz MS, Peffer N, Meyers H, Nelson JA, Gairin JE, Hahn BH, Oldstone MB, Shaw GM: **Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus.** *Nat Med* 1997, **3**:205-211.
- Goulder PJ, Watkins DI: **HIV and SIV CTL escape: implications for vaccine design.** *Nat Rev Immunol* 2004, **4**:630-640.
- Price DA, Goulder PJ, Klenerman P, Sewell AK, Easterbrook PJ, Troop M, Bangham CR, Phillips RE: **Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection.** *Proc Natl Acad Sci USA* 1997, **94**:1890-1895.
- Koup RA, Safrit JT, Cao Y, Andrews CA, McLeod G, Borkowsky W, Farthing C, Ho DD: **Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type I syndrome.** *J Virol* 1994, **68**:4650-4655.
- Schmitz JE, Kuroda MJ, Santra S, Sasseville VG, Simon MA, Lifton MA, Racz P, Tenner-Racz K, Dalesandro M, Scallan BJ, et al.: **Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes.** *Science* 1999, **283**:857-860.
- Kiepiela P, Ngumbela K, Thobakgale C, Ramduth D, Honeyborne I, Moodley E, Reddy S, de Pierres C, Mncube Z, Mkhwanazi N, et al.: **CD8+ T-cell responses to different HIV proteins have discordant associations with viral load.** *Nat Med* 2007, **13**:46-53.
- Goulder PJ, Watkins DI: **Impact of MHC class I diversity on immune control of immunodeficiency virus replication.** *Nat Rev Immunol* 2008, **8**:619-630.
- Carlson JM, Brumme ZL: **HIV evolution in response to HLA-restricted CTL selection pressures: a population-based perspective.** *Microbes Infect* 2008, **10**:455-461.
- Goulder PJ, Brander C, Tang Y, Tremblay C, Colbert RA, Addo MM, Rosenberg ES, Nguyen T, Allen R, Trocha A, et al.: **Evolution and transmission of stable CTL escape mutations in HIV infection.** *Nature* 2001, **412**:334-338.
- Phillips RE, Rowland-Jones S, Nixon DF, Gotch FM, Edwards JP, Ogunlesi AO, Elvin JG, Rothbard JA, Bangham CR, Rizza CR, et al.: **Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition.** *Nature* 1991, **354**:453-459.
- Draenert R, Le Gall S, Pfafferoth KJ, Leslie AJ, Chetty P, Brander C, Holmes EC, Chang SC, Feeney ME, Addo MM, et al.: **Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection.** *J Exp Med* 2004, **199**:905-915.
- Carlson JM, Brumme ZL, Rousseau CM, Brumme CJ, Matthews P, Kadie C, Mullins JL, Walker BD, Harrigan PR, Goulder PJ, Heckerman D: **Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag.** *PLoS Comput Biol* 2008, **4**:e1000225.
- Friedrich TC, Dodds EJ, Yant LJ, Vojnov L, Rudersdorf R, Cullen C, Evans DT, Desrosiers RC, Mothe BR, Sidney J, et al.: **Reversion of CTL escape-variant immunodeficiency viruses in vivo.** *Nat Med* 2004, **10**:275-281.
- Leslie AJ, Pfafferoth KJ, Chetty P, Draenert R, Addo MM, Feeney M, Tang Y, Holmes EC, Allen T, Prado JG, et al.: **HIV evolution: CTL escape mutation and reversion after transmission.** *Nat Med* 2004, **10**:282-289.
- Peyerl FW, Bazick HS, Newberg MH, Barouch DH, Sodroski J, Letvin NL: **Fitness costs limit viral escape from cytotoxic T lymphocytes at a structurally constrained epitope.** *J Virol* 2004, **78**:13901-13910.
- Liu Y, McNevin J, Zhao H, Tebit DM, Troyer RM, McSweyn M, Ghosh AK, Shriner D, Arts EJ, McElrath MJ, Mullins JL: **Evolution of human immunodeficiency virus type I cytotoxic T-lymphocyte epitopes: fitness-balanced escape.** *J Virol* 2007, **81**:12179-12188.
- Edwards CT, Pfafferoth KJ, Goulder PJ, Phillips RE, Holmes EC: **Intra-patient escape in the A\*0201-restricted epitope SLYNT-VATL drives evolution of human immunodeficiency virus type I at the population level.** *J Virol* 2005, **79**:9363-9366.
- Furutsuki T, Hosoya N, Kawana-Tachikawa A, Tomizawa M, Odawara T, Goto M, Kitamura Y, Nakamura T, Kelleher AD, Cooper DA, Iwamoto A: **Frequent transmission of cytotoxic-T-lymphocyte escape mutants of human immunodeficiency virus type I in the highly HLA-A24-positive Japanese population.** *J Virol* 2004, **78**:8437-8445.
- Piontkivska H, Hughes AL: **Patterns of sequence evolution at epitopes for host antibodies and cytotoxic T-lymphocytes in human immunodeficiency virus type I.** *Virus Res* 2006, **116**:98-105.
- Kawashima Y, Pfafferoth K, Frater J, Matthews P, Payne R, Addo M, Gatanaga H, Fujiwara M, Hachiya A, Koizumi H, et al.: **Adaptation of HIV-1 to human leukocyte antigen class I.** *Nature* 2009, **458**:641-5.
- Leslie A, Kavanagh D, Honeyborne I, Pfafferoth K, Edwards C, Pillay T, Hilton L, Thobakgale C, Ramduth D, Draenert R, et al.: **Transmission and accumulation of CTL escape variants drive negative associations between HIV polymorphisms and HLA.** *J Exp Med* 2005, **201**:891-902.
- Moore CB, John M, James IR, Christiansen FT, Witt CS, Mallal SA: **Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level.** *Science* 2002, **296**:1439-1443.
- Brumme ZL, Brumme CJ, Carlson J, Streeck H, John M, Eichbaum Q, Block BL, Baker B, Kadie C, Markowitz M, et al.: **Marked epitope and allele-specific differences in rates of mutation in HIV-1 Gag, Pol and Nef CTL epitopes in acute/early HIV-1 infection.** *J Virol* 2008, **82**:9216-27.
- Schneidewind A, Brockman MA, Yang R, Adam RI, Li B, Le Gall S, Rinaldo CR, Craggs SL, Allgaier RL, Power KA, et al.: **Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in Gag is associated with a dramatic reduction in human immunodeficiency virus type I replication.** *J Virol* 2007, **81**:12382-12393.
- Iversen AK, Stewart-Jones G, Learn GH, Christie N, Sylvester-Hviid C, Armitage AE, Kaul R, Beattie T, Lee JK, Li Y, et al.: **Conflicting selective forces affect T cell receptor contacts in an immunodominant human immunodeficiency virus epitope.** *Nat Immunol* 2006, **7**:179-189.
- Mason RD, Grant MD: **A therapy-related point mutation changes the HLA restriction of an HIV-1 Pol epitope from A2 to B57 and enhances its recognition.** *Aids* 2005, **19**:981-984.
- Allen TM, Yu XG, Kalife ET, Reyrol LL, Lichtenfeld M, John M, Cheng M, Allgaier RL, Mui S, Frahm N, et al.: **De novo generation of escape variant-specific CD8+ T-cell responses following cytotoxic T-lymphocyte escape in chronic human immunodeficiency virus type I infection.** *J Virol* 2005, **79**:12952-12960.
- Bhattacharya T, Daniels M, Heckerman D, Foley B, Frahm N, Kadie C, Carlson J, Yusim K, McMahon B, Gaschen B, et al.: **Founder effects in the assessment of HIV polymorphisms and HLA allele associations.** *Science* 2007, **315**:1583-1586.
- Altfeld M, Kalife ET, Qi Y, Streeck H, Lichtenfeld M, Johnston MN, Burgett N, Swartz ME, Yang A, Alter G, et al.: **HLA Alleles Associated with Delayed Progression to AIDS Contribute Strongly to the Initial CD8(+) T Cell Response against HIV-1.** *PLoS Med* 2006, **3**:e403.
- Streeck H, Lichtenfeld M, Alter G, Meier A, Teigen N, Yassine-Diab B, Sidhu HK, Little S, Kelleher A, Routy JP, et al.: **Recognition of a defined region within p24 gag by CD8+ T cells during primary human immunodeficiency virus type I infection in individuals expressing protective HLA class I alleles.** *J Virol* 2007, **81**:7725-7731.
- Ahlenstiel G, Roomp K, Daumer M, Nattermann J, Vogel M, Rockstroh JK, Beerenwinkel N, Kaiser R, Nischalke HD, Sauerbruch T, et al.: **Selective pressures of HLA genotypes and antiviral therapy on human immunodeficiency virus type I sequence mutation at a population level.** *Clin Vaccine Immunol* 2007, **14**:1266-1273.



34. Brumme ZL, Brumme CJ, Heckerman D, Korber BT, Daniels M, Carlson J, Kadie C, Bhattacharya T, Chui C, Szinger J, et al.: **Evidence of Differential HLA Class I-Mediated Viral Evolution in Functional and Accessory/Regulatory Genes of HIV-1.** *PLoS Pathog* 2007, **3**:e94.
35. Mueller SM, Schaetz B, Eismann K, Bergmann S, Bauerle M, Schmitt-Haendle M, Walter H, Schmidt B, Korn K, Sticht H, et al.: **Dual selection pressure by drugs and HLA class I-restricted immune responses on human immunodeficiency virus type I protease.** *J Virol* 2007, **81**:2887-2898.
36. Pond SL, Frost SD, Grossman Z, Gravenor MB, Richman DD, Brown AJ: **Adaptation to different human populations by HIV-1 revealed by codon-based analyses.** *PLoS Comput Biol* 2006, **2**:e62.
37. Brumme ZL, John M, Brumme CJ, Carlson J, Haubrich R, Riddler S, Swenson L, Tao I, Szeto S, Chan D, et al.: **Relationship between HLA class I-driven evolution in Gag, Pol and Nef and clinical markers of HIV disease: a multi-center collaborative study.** *AIDS Vaccine, Abstract P09-01; Cape Town, South Africa* 2008.
38. Barquera R, Zuniga J, Hernandez-Diaz R, Acuna-Alonzo V, Montoya-Gama K, Moscoso J, Torres-Garcia D, Garcia-Salas C, Silva B, Cruz-Robles D, et al.: **HLA class I and class II haplotypes in admixed families from several regions of Mexico.** *Mol Immunol* 2008, **45**:1171-1178.
39. Gorodezky C, Alaez C, Vazquez-Garcia MN, de la Rosa G, Infante E, Balladares S, Toribio R, Perez-Luque E, Munoz L: **The genetic structure of Mexican Mestizos of different locations: tracking back their origins through MHC genes, blood group systems, and microsatellites.** *Hum Immunol* 2001, **62**:979-991.
40. Excoffier L, Laval G, Balding D: **Gametic phase estimation over large genomic regions using an adaptive window approach.** *Hum Genomics* 2003, **1**:7-19.
41. Storey JD, Tibshirani R: **Statistical significance for genomewide studies.** *Proc Natl Acad Sci USA* 2003, **100**:9440-9445.
42. Soto-Ramirez LE: **HIV/AIDS in Latin America.** *Science* 2008, **321**:465.
43. Bastos FI, Caceres C, Galvao J, Veras MA, Castilho EA: **AIDS in Latin America: assessing the current status of the epidemic and the ongoing response.** *Int J Epidemiol* 2008, **37**:729-737.
44. Magis-Rodríguez C, Hernández-Avila M: **Epidemiología del SIDA en México.** In *25 Años de SIDA en México Logros, desaciertos y retos* First edition. Edited by: Córdova-Villalobos J, Ponce de León-Rosales S, Valdespino J. Cuernavaca, Morelos, México: Instituto Nacional de Salud Pública; 2008:85-103.
45. Llop E: **El sistema mayor de histocompatibilidad humano, HLA.** In *Poblaciones chilenas: cuatro décadas de investigaciones bio-antropológicas* Edited by: Rothhammer F, Llop E. Chile: Editorial Universitaria; 2004:165-188.
46. Brumme ZL, Tao I, Szeto S, Brumme CJ, Carlson JM, Chan D, Kadie C, Frahm N, Brander C, Walker B, et al.: **Human leukocyte antigen-specific polymorphisms in HIV-1 Gag and their association with viral load in chronic untreated infection.** *Aids* 2008, **22**:1277-1286.
47. Karlsson AC, Deeks SG, Barbour JD, Heiken BD, Younger SR, Hoh R, Lane M, Sallberg M, Ortiz GM, Demarest JF, et al.: **Dual pressure from antiretroviral therapy and cell-mediated immune response on the human immunodeficiency virus type I protease gene.** *J Virol* 2003, **77**:6743-6752.
48. Matthews PC, Prendergast A, Leslie A, Crawford H, Payne R, Rousseau C, Rolland M, Honeyborne I, Carlson J, Kadie C, et al.: **Central role of reverting mutations in HLA associations with human immunodeficiency virus set point.** *J Virol* 2008, **82**:8548-8559.
49. Kiepiela P, Leslie AJ, Honeyborne I, Ramduth D, Thobakgale C, Chetty S, Rathnavalu P, Moore C, Pfafferoth KJ, Hilton L, et al.: **Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA.** *Nature* 2004, **432**:769-775.
50. Bailey JR, Brennan TP, O'Connell KA, Siliciano RF, Blankson JN: **Evidence of CD8+ T-cell-mediated selective pressure on human immunodeficiency virus type I nef in HLA-B\*57+ elite suppressors.** *J Virol* 2009, **83**:88-97.
51. Goulder PJ, Altfeld MA, Rosenberg ES, Nguyen T, Tang Y, Eldridge RL, Addo MM, He S, Mukherjee JS, Phillips MN, et al.: **Substantial differences in specificity of HIV-specific cytotoxic T cells in acute and chronic HIV infection.** *J Exp Med* 2001, **193**:181-194.
52. Karlsson AC, Iversen AK, Chapman JM, de Oliveira T, Spotts G, McMichael AJ, Davenport MP, Hecht FM, Nixon DF: **Sequential broadening of CTL responses in early HIV-1 infection is associated with viral escape.** *PLoS ONE* 2007, **2**:e225.
53. Rozera G, Abbate I, Bruxelles A, Vlassi C, D'Offizi G, Narciso P, Chillemi G, Prosperi M, Ippolito G, Capobianchi MR: **Massively parallel pyrosequencing highlights minority variants in the HIV-1 env quasispecies deriving from lymphomonocyte sub-populations.** *Retrovirology* 2009, **6**:15.
54. Bailey JR, Zhang H, Wegweiser BW, Yang HC, Herrera L, Ahonkhai A, Williams TM, Siliciano RF, Blankson JN: **Evolution of HIV-1 in an HLA-B\*57-positive patient during virologic escape.** *J Infect Dis* 2007, **196**:50-55.
55. Borderia AV, Codoner FM, Sanjuan R: **Selection promotes organ compartmentalization in HIV-1: evidence from gag and pol genes.** *Evolution* 2007, **61**:272-279.
56. Li B, Gladden AD, Altfeld M, Kaldor JM, Cooper DA, Kelleher AD, Allen TM: **Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type I evolution.** *J Virol* 2007, **81**:193-201.
57. Avila-Rios S, Reyes-Teran G, Espinosa E: **Cornering HIV: taking advantage of interactions between selective pressures.** *Med Hypotheses* 2007, **69**:422-431.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

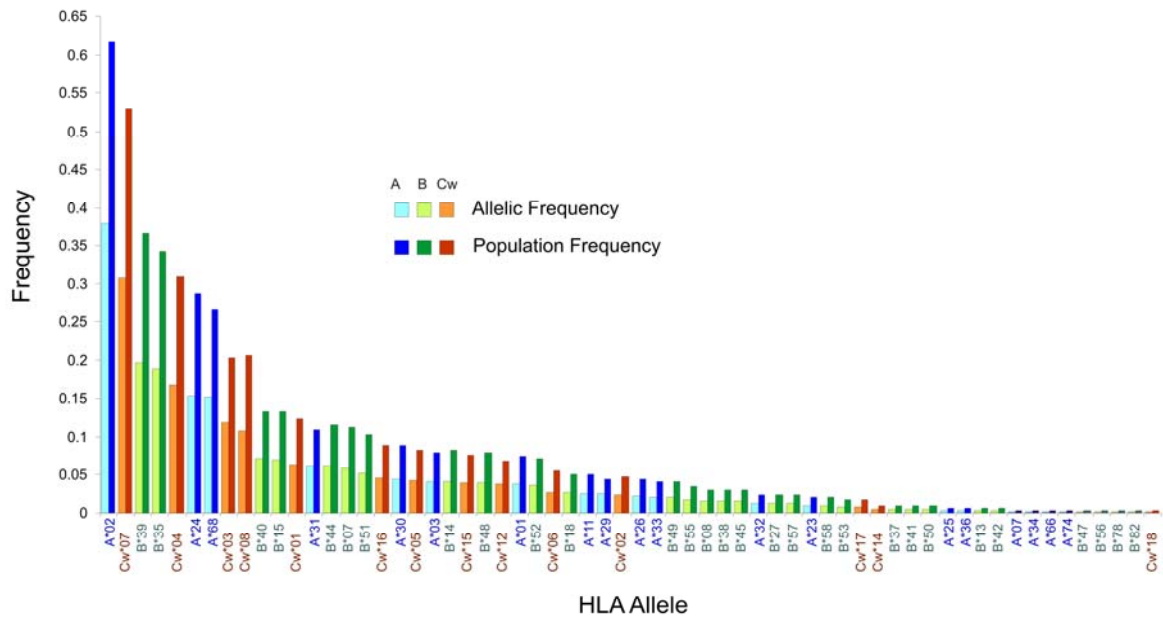
Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

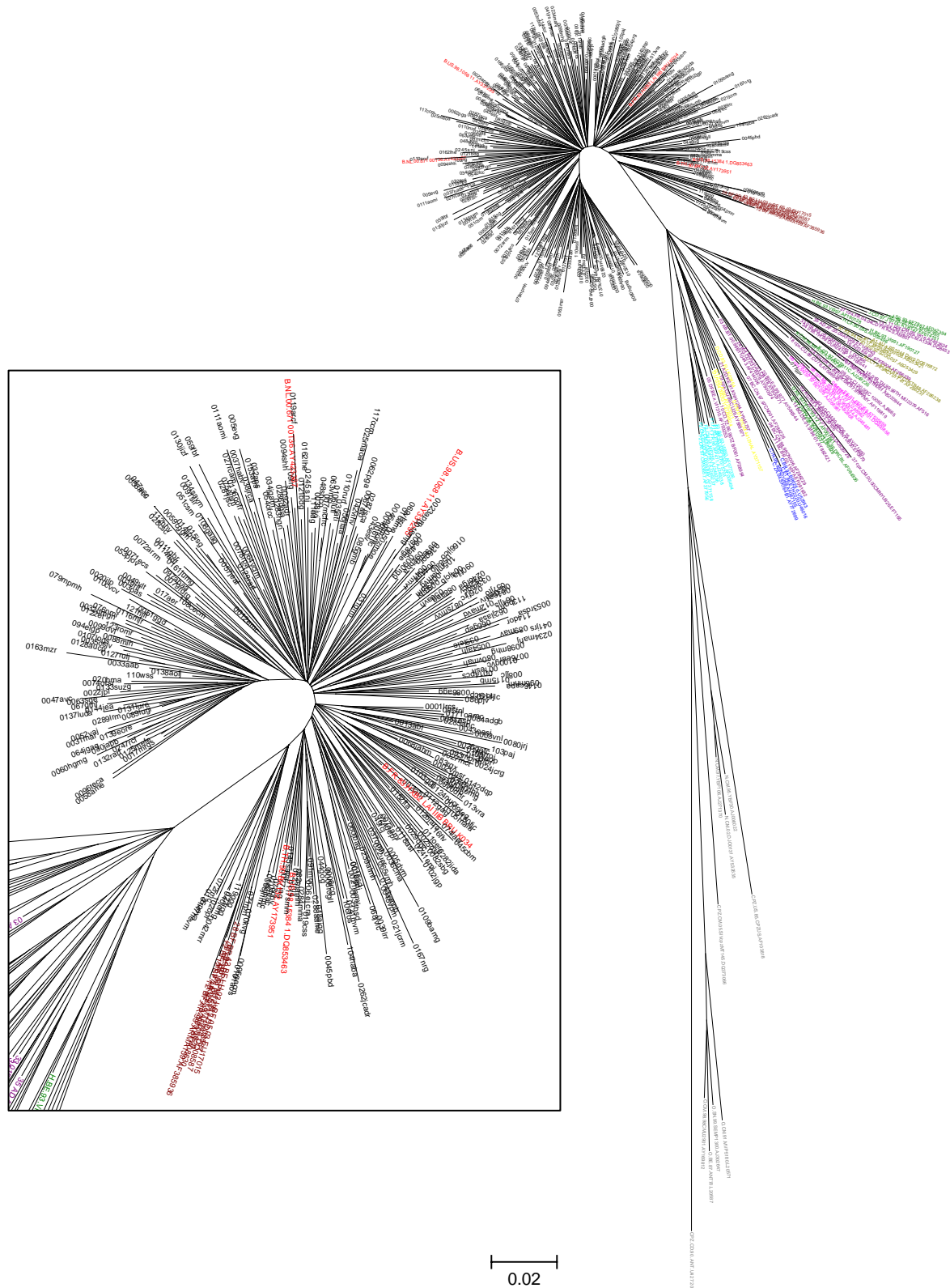






**Figure S1 - Allelic and population frequencies of class I HLA-A, B and C genes in a cohort of HIV-infected individuals from Central/Southern Mexico.**

Frequencies were obtained with the HLA Frequency Analysis Tool of the Los Alamos HIV Immunology Database ([www.hiv.lanl.gov/content/immunology/tools-links.html](http://www.hiv.lanl.gov/content/immunology/tools-links.html)). HLA typing was carried out by SSP-PCR as described in the Methods. A total of 292 individuals (584 HLA alleles) were included; 19, 29 and 14 distinct allelic groups were observed for the HLA-A, B and C genes respectively.



**Figure S2 - Phylogeny of 280 HIV *pol* sequences from a cohort of antiretroviral treatment-naïve individuals from Central/Southern Mexico.**

A Neighbor-Joining tree was inferred through the analysis of 1305 bp *pol* sequences including the whole protease and 335 codons of the RT in MEGA 4.0. The consensus tree from 1000 bootstrap replicas is shown. Evolutionary distances were calculated with Kimura's two-parameter model and are shown in substitutions per site. Positions with missing information were eliminated by pairwise comparison. 93 reference sequences of the main HIV-1 groups, subtypes and recombinant forms were obtained from the Los Alamos HIV Sequence Database (<http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html>) and included in the tree. Black: Mexican sequences, red: subtype B, dark red: BF recombinant forms, olive: subtype A, dark blue: subtype C, yellow: subtype D, light blue: subtype F, green: subtypes G, H, J, K, fucsia: 06\_cpx recombinant forms, purple: other recombinant forms, grey: O-N groups and SIVs. The inset shows a detail of the Mexican sequence cluster.

**Table S1 – HLA-A, B and C population and allelic frequencies in the Mexican cohort. §**

HLA	Allelic Frequency	Population Frequency
A*02	0.3784	0.6164
Cw*07	0.3076	0.5292
B*39	0.1969	0.3664
B*35	0.1884	0.3425
Cw*04	0.1684	0.3093
A*24	0.1541	0.2877
A*68	0.1524	0.2671
Cw*03	0.1186	0.2027
Cw*08	0.1082	0.2062
B*40	0.0719	0.1336
B*15	0.0702	0.1336
Cw*01	0.0636	0.1237
A*31	0.0616	0.1096
B*44	0.0616	0.1164
B*07	0.0582	0.113
B*51	0.0531	0.1027
Cw*16	0.0464	0.0893
A*30	0.0445	0.089
Cw*05	0.043	0.0825
A*03	0.0411	0.0788
B*14	0.0411	0.0822
Cw*15	0.0395	0.0756
B*48	0.0394	0.0788
Cw*12	0.0378	0.0687
A*01	0.0377	0.0753
B*52	0.036	0.0719
Cw*06	0.0275	0.055
B*18	0.0274	0.0514
A*11	0.0257	0.0514
A*29	0.0257	0.0445
Cw*02	0.0241	0.0481
A*26	0.0223	0.0445
A*33	0.0205	0.0411
B*49	0.0205	0.0411
B*55	0.0171	0.0342
B*08	0.0154	0.0308
B*38	0.0154	0.0308
B*45	0.0154	0.0308
A*32	0.012	0.024
B*27	0.012	0.024
B*57	0.012	0.024
A*23	0.0103	0.0205

B*58	0.0103	0.0205
B*53	0.0086	0.0171
Cw*17	0.0086	0.0172
Cw*14	0.0052	0.0103
B*37	0.0051	0.0103
B*41	0.0051	0.0103
B*50	0.0051	0.0103
A*25	0.0034	0.0068
A*36	0.0034	0.0068
B*13	0.0034	0.0068
B*42	0.0034	0.0068
A*07	0.0017	0.0034
A*34	0.0017	0.0034
A*66	0.0017	0.0034
A*74	0.0017	0.0034
B*47	0.0017	0.0034
B*56	0.0017	0.0034
B*78	0.0017	0.0034
B*82	0.0017	0.0034
Cw*18	0.0017	0.0034

---

§ HLA typing was carried out by SSP-PCR as described in Methods. A total of 292 individuals and 584 alleles were included.

**Table S2 - Linkage disequilibrium for class I HLA-A, B and C genes in the Mexican cohort. §**

HLA1	HLA2	p-Value	Both	Only HLA1	Only HLA2	None	P(HLA1)	P(HLA2)	P(HLA1 & HLA2)	E(P(HLA1 & HLA2))
B*35	Cw*04	1.30E-33	78	22	11	177	0.3472	0.3090	0.2708	0.1073
B*39	Cw*07	4.40E-21	97	10	55	126	0.3715	0.5278	0.3368	0.1961
B*15	Cw*01	8.99E-15	25	13	10	240	0.1319	0.1215	0.0868	0.0160
B*40	Cw*03	1.08E-11	29	10	29	220	0.1354	0.2014	0.1007	0.0273
B*14	Cw*08	5.83E-11	22	2	38	226	0.0833	0.2083	0.0764	0.0174
B*48	Cw*08	3.63E-10	21	2	39	226	0.0799	0.2083	0.0729	0.0166
B*44	Cw*16	3.58E-07	16	18	9	245	0.1181	0.0868	0.0556	0.0102
B*40	Cw*07	2.84E-06	3	36	149	100	0.1354	0.5278	0.0104	0.0715
B*38	Cw*12	1.02E-05	8	1	12	267	0.0312	0.0694	0.0278	0.0022
B*44	Cw*05	2.66E-05	14	20	9	245	0.1181	0.0799	0.0486	0.0094
B*18	Cw*05	3.92E-04	9	5	14	260	0.0486	0.0799	0.0312	0.0039
A*03	B*07	4.23E-04	12	9	19	248	0.0729	0.1076	0.0417	0.0078
A*01	B*57	7.22E-04	6	15	0	267	0.0729	0.0208	0.0208	0.0015
B*41	Cw*17	2.54E-02	3	0	2	283	0.0104	0.0174	0.0104	0.0002
B*07	Cw*07	4.64E-02	28	3	124	133	0.1076	0.5278	0.0972	0.0568
A*30	B*39	5.35E-02	0	25	107	156	0.0868	0.3715	0.0000	0.0323
B*52	Cw*03	6.68E-02	13	7	45	223	0.0694	0.2014	0.0451	0.0140
A*33	B*14	6.88E-02	7	5	17	259	0.0417	0.0833	0.0243	0.0035
B*51	Cw*15	9.22E-02	10	20	12	246	0.1042	0.0764	0.0347	0.0080
A*29	Cw*16	1.92E-01	7	6	18	257	0.0451	0.0868	0.0243	0.0039
A*30	B*18	3.63E-01	7	18	7	256	0.0868	0.0486	0.0243	0.0042

§ 288 individuals from Central/Southern Mexico were included. Both: number of individuals expressing HLA1 and HLA2. Only HLA1: number of individuals expressing HLA1, but not HLA2. Only HLA2: number of individuals expressing HLA2, but not HLA1. None: number of individuals expressing neither HLA1 nor HLA2. P(HLA1): observed frequency of individuals expressing HLA1 in the cohort. P(HLA2): observed frequency of individuals expressing HLA2 in the cohort. P(HLA1 & HLA2): observed frequency of individuals expressing both HLA1 and HLA2 in the cohort. E(P(HLA1 & HLA2)): expected frequency of P(HLA1 & HLA2) if P(HLA1) and P(HLA2) were independent. The p-value for two-tail Fisher exact tests is shown. Estimation of the probability that P(HLA1) and P(HLA2) are independent given the observed information. A low p-value indicates that the occurrence of HLA1 and HLA2 is linked in the cohort. Number of two-way tests performed =  $A*B+A*C+B*C = 1190$ ; p-value for two-way comparisons =  $0.05/(A*B+A*C+B*C) = 4.202e-05$  for a confidence level of 95%.

**Table S3 - Protease and RT HLA-HIV codon associations observed in the Mexican cohort. §**

Predictor Variable (HLA)	Target Variable (HIV codon)	TT	TF	FT	FF	p-value	q-value	Conditioning Variables (Coevolving codons, conditioning HLA)	Reference	Observations
A*03 <sup>b</sup>	RT 277R	19	2	56	195	1.20E-10	1.82E-06		[34], [25], [37]	
A*03 <sup>a</sup>	RT 277K	2	19	195	56	1.20E-10	1.82E-06		[34], [25], [37]	
A*30 <sup>a</sup>	RT 281K	18	5	251	3	0.00073383	0.13153195		[37]	
A*33 <sup>a</sup>	RT 122P	1	11	67	191	0.0009406	0.14940166	RT 98A <sup>c</sup>	[37]†	Consistent (Escape from P in Mex, attraction to E in IHAC).
B*07 <sup>b</sup>	RT 165I	7	18	10	241	0.0005105	0.1175217		[34], [37]	
B*07 <sup>a</sup>	RT 165T	18	7	240	10	0.00052215	0.1175217		[34], [37]	
B*08 <sup>a</sup>	RT 272A	1	7	122	143	0.00174628	0.19533002		[25]†, [37]†	Different (B*42 escape from P in IHAC, B*08 has a much lower frequency in Mex).
B*18 <sup>b</sup>	RT 138A	3	10	1	266	0.00013697	0.05164684		[25], [37]	
B*18 <sup>a</sup>	RT 138E	10	3	265	1	0.00013862	0.05164684		[25], [37]	
B*18 <sup>b</sup>	RT 245E	4	8	22	240	0.00144187	0.18595855	PR 45K <sup>c</sup> , B58 <sup>b</sup>	[34]†, [37]†	Different (B*57/58 escape from V to L, B*44 escape from E in IHAC; no B*18 association with p<0.05 in IHAC, no difference in HLA frequency).
B*27 <sup>a</sup>	RT 135I	1	6	153	94	0.00140863	0.1832374	B*51 <sup>a</sup> , RT 202I <sup>d</sup> , PR 12N <sup>a</sup>	[34]†, [25]†, [37]†	Different (B*51/52 escape from I to T/M in IHAC; B*27 less frequent in Mex).
B*39 <sup>a</sup>	PR 71V	6	85	23	140	0.00121556	0.17636733	PR 93L <sup>d</sup> , PR 57K <sup>b</sup>	[25]†, [37]†	Different (B*53 escape from T in IHAC; B*39 much more frequent in Mex).
B*39 <sup>a</sup>	RT 102R	0	96	17	153	0.00011626	0.04678005	B*48 <sup>b</sup> , RT 64K <sup>a</sup>	[34]†, [37]†	Different (B*48 K to R; B*39 much more frequent in Mex).
B*39 <sup>b</sup>	RT 64R	5	91	4	166	0.00166664	0.1905207	RT 102R <sup>d</sup>	NR	No close association in IHAC.
B*39 <sup>a</sup>	RT 64K	91	5	166	4	0.00166664	0.1905207	RT 102R <sup>c</sup>	NR	No close association in IHAC.
B*40 <sup>b</sup>	RT 202V	5	27	10	186	0.00100069	0.15487079	RT 135I <sup>a</sup> , RT 197E <sup>b</sup> , PR 35E <sup>d</sup> , RT 334L <sup>a</sup> , RT 165I <sup>d</sup> , RT 122P <sup>b</sup>	NR	No close association in IHAC; no difference in HLA frequency among cohorts.
B*40 <sup>a</sup>	RT 202I	27	5	186	10	0.00100069	0.15487079	RT 135I <sup>b</sup> , RT 197E <sup>a</sup> , PR 35E <sup>c</sup> , RT 334L <sup>b</sup> , RT 165I <sup>c</sup> , RT 122P <sup>a</sup>	NR	No close association in IHAC; no difference in HLA frequency among cohorts.
B*44 <sup>b</sup>	PR 35D	21	10	33	210	2.78E-09	1.40E-05		[34], [25], [37]	
B*44 <sup>a</sup>	PR 35E	9	22	210	33	6.65E-10	6.69E-06		[34], [25], [37]	
B*44 <sup>b</sup>	RT 211K	22	6	106	115	0.00185231	0.19533002	PR 43K <sup>c</sup> , RT 196E <sup>a</sup>	[34]†, [37]†	Consistent (R to K in Mex, just escape from R in IHAC).
B*44 <sup>a</sup>	RT 211R	2	28	94	128	4.19E-05	0.02767678		[34], [37]	

B*44 <sup>a</sup>	RT 329I	14	15	203	30	0.00130602	0.18323349	RT 334L <sup>b</sup>	[37]†	Different (B*53 escape from V in IHAC; B*44 less frequent in Mex).
B*44 <sup>d</sup>	RT 123N	7	25	3	236	0.00152115	0.18684048	RT 122E <sup>d</sup>	[34]†, [25]†	Different (B*35 D to E in IHAC).
B*44 <sup>c</sup>	RT 68S	26	7	232	7	0.00185709	0.19533002	PR 41R <sup>c</sup> , RT 139T <sup>d</sup>	NR	No close association in IHAC.
B*44 <sup>d</sup>	RT 68G	7	26	7	232	0.00185709	0.19533002	PR 41R <sup>d</sup> , RT 139T <sup>c</sup>	NR	No close association in IHAC.
B*45 <sup>b</sup>	RT 200A	5	3	37	220	0.00072663	0.13153195		[34]†, [37]†	Different (B*40/41 attraction to I, B*08 escape from A, B*37 escape from T in IHAC; no difference in B*45 frequency).
B*48 <sup>b</sup>	RT 102R	10	8	7	249	2.54E-09	1.40E-05		[34], [37]	
B*48 <sup>a</sup>	RT 102K	8	10	227	28	2.83E-05	0.02083046		[34], [37]	
B*49 <sup>b</sup>	RT 50V	3	9	3	265	0.00078914	0.13651316		NR	No close association in IHAC; no difference in HLA frequency among cohorts.
B*49 <sup>a</sup>	RT 50I	9	3	265	3	0.00078914	0.13651316		NR	No close association in IHAC; no difference in HLA frequency among cohorts.
B*51 <sup>b</sup>	PR 12S	5	23	2	249	7.70E-05	0.03639581		[34]†, [37]†	Consistent (escape from T, weak attraction to S [p=0.02] in IHAC).
B*51 <sup>b</sup>	PR 14R	10	13	20	212	0.00064885	0.13057972	RT 135I <sup>a</sup>	[34], [37]†	
B*51 <sup>a</sup>	PR 14K	12	10	212	20	0.00052685	0.1175217	RT 135I <sup>b</sup>	[34], [37]†	
B*51 <sup>a</sup>	RT 135I	4	20	155	84	6.58E-08	0.00016543		[34], [25]†, [37]	
B*52 <sup>b</sup>	PR 12A	7	8	9	247	2.44E-06	0.00319527		[34], [37]†	
B*52 <sup>a</sup>	PR 12T	4	13	215	37	9.21E-08	0.00019546		[34], [37]†	
B*57 <sup>b</sup>	RT 297A	4	3	33	239	0.0014485	0.18601806		NR	No close association in IHAC (B*58 has attraction to R with p=0.03, q=0.88).
Cw*04 <sup>a</sup>	RT 248D	0	88	11	180	0.00160017	0.18790507		[25]†, [37]†	Different (Cw*02 escape from E, B*15 escape from D in IHAC; B*15 appears to be in slight negative LD with Cw*04).
Cw*07 <sup>c</sup>	RT 276V	111	32	116	8	7.21E-05	0.03639581		NR	The attraction to V is nearly seen as a Cw*08 association in IHAC (q=0.22); no difference in HLA frequency among cohorts.
Cw*07 <sup>d</sup>	RT 276T	11	131	1	130	0.00152896	0.18684048	RT 272P <sup>d</sup>	NR	The attraction to V is nearly seen as a C*08 association in IHAC (q=0.22); no difference in HLA frequency among cohorts.
Cw*07 <sup>a</sup>	RT 179D	0	147	8	125	0.00061063	0.12536221		NR	No close association in IHAC; Cw*08 is weakly assoc with escape from V (p=0.006, q=0.42).
Cw*07 <sup>c</sup>	RT 278Q	137	8	129	1	0.00130929	0.18323349	RT 281R <sup>a</sup>	[25]†, [37]†	Different (Cw*12 attraction to V is somewhat close, q=0.22).
Cw*15 <sup>a</sup>	PR 18Q	18	4	253	4	0.00118543	0.17492044		NR	No close associations (p<0.05) in IHAC; no difference in HLA frequency among cohorts.

§ 280 individuals from Central/Southern Mexico were included. Associations with q<0.2 are shown. NR – Not previously reported associations. †HIV codon has been previously reported to be under HLA pressure, but HLA specificity and/or target amino acid are different to that observed in the present study. TT, TF, FT, FF: 2x2 contingency tables; HLA is the first value, amino acid is the second value. <sup>a</sup> escape (having the predictor makes it less likely to have the target), <sup>b</sup> attraction (having the predictor makes it more likely to have the target), <sup>c</sup> reversion (not having the predictor makes it more likely to have the target), <sup>d</sup> repulsion (not having the predictor makes it less likely to have the target).