

UNIVERSIDAD AUTÓNOMA DE MADRID

Facultad de Ciencias

Departamento de Biología Molecular

Estudio sobre los cambios moleculares  
asociados con las variaciones de la  
eficacia biológica en el VIH-1 *in vitro*

**Antonio Vicente Borderia Giner**

Madrid, 2007

UNIVERSIDAD AUTÓNOMA DE MADRID

Facultad de Ciencias

Departamento de Biología Molecular

Estudio sobre los cambios moleculares  
asociados con las variaciones de la eficacia  
biológica en el VIH-1 *in vitro*

Memoria presentada para optar al grado de Doctor en Ciencias por:

**Antonio Vicente Borderia Giner**

Director de la Tesis:

Tutora de la Tesis:

**Dr. Cecilio López Galíndez**

**Dra. Amelia Nieto**

Centro Nacional de Microbiología

Centro Nacional de Biotecnología

Instituto de Salud Carlos III

Consejo Superior de Investigaciones Científicas

Madrid, Marzo de 2007



Cuando se empieza una Tesis Doctoral, muchas veces no se sabe como va a ir la cosa, es algo que se va viendo conforme se va recorriendo el largo trayecto que lleva por fin al momento de la escritura. Al final de este trayecto es cuando de verdad nos damos cuenta de todo el esfuerzo y trabajo invertido en la Tesis. Es un proceso constante de aprendizaje, tanto en lo profesional como en lo personal. Uno aprende de sus errores, y a como mejorar en cada experimento propuesto, además de a trabajar en grupo y en apoyarse en sus compañeros de laboratorio. Por ello, me gustaría dar mis más sinceras gracias a las siguientes personas:

Al Instituto Carlos III, por haberme concedido la beca intramural que me permitió realizar la presente Tesis Doctoral.

Al Dr. Cecilio López-Galíndez, por haberme dado su apoyo y ayuda en todo este trayecto hasta el final de la Tesis, e intentar en infinitas ocasiones que perdiera mi alergia al “aminoácido”. Por fin aquí estamos!!. Muchas gracias.

A la Dra. Amelia Nieto, tutora de esta Tesis, por su interés y por la evaluación previa de este trabajo, así como por su ayuda durante los trámites de presentación de la Tesis.

Al Dr. Esteban Domingo, por ofrecer su sincera ayuda y buenos consejos durante el desarrollo de la Tesis, y por inspirar buena parte del trabajo.

A la Dra. Eloisa Yuste, por ser quien empezó a abrir este camino antes que yo llegara al laboratorio, y ayudarme en mis primeros pasos en los experimentos.

Al Servicio de Secuenciación, Pablo, Aurora, Ana y Alicia. Sois los mejores!! sin vosotros no hubiera podido secuenciar los cientos y cientos de secuencias necesarios para tener todos los resultados de la Tesis.

A los Drs. Santiago Elena y Rosario Miralles, por ser quienes me ayudaron en mis primeros pasos en la ciencia, e inspiraron en mí el interés por la evolución en virus de ARN. A ellos debo estar donde estoy ahora.

A mis compañeras en el Servicio de Virología Molecular que estuvieron durante toda mi tesis ofreciéndome su apoyo y ayuda, y manteniendo el tan buen ambiente que reina en el laboratorio. Gracias por todo, sin vosotras esto no hubiera sido lo mismo: Concha, Isabel, Maria, Tamara, Virginia, y Rosa. Aunque también a todas las becarias FINNOVA y de verano que pasaron por el labo: Maricarmen, Ángeles, y las Beas. Entre ellas también Vir que junto a Tamara me apoyaron en tanto los buenos como los malos momentos. Por ultimo, también quiero agradecer a Dora las múltiples y variadas charlas

que compartimos a últimas horas en el labo cuando estábamos acabando experimentos, y animarme cuando mas “lamentoso” estaba.

A los que ya se fueron del labo, Gonzalo y Mariajose porque tanto dentro del labo como fuera, siempre estuvieron allí, aun están... aunque en la distancia!! Por muchos años!!

A mis compañeros mas allá de mi laboratorio: Eliana y Pachi, por los buenos tiempos tomando hierba mate, y las charlas “filosóficas” de los viernes para descansar cuerpo y mente, y como no por su amistad. También a Javi, Nuria, Esther, Cristina, también a Pilar por las horas irrepetibles de las comidas... lo que nos reíamos!!!! A Andrea, por esos pedazos de partidos de fútbol que nos pegábamos, y por mostrarnos como realmente baila Bisbal. A Enrique y Ana, por las charlas frente a la maquina de café y hacer las tardes mucho mas llevaderas. A Carolina, Fran y Yoli, por las risas que nos dábamos cuando nos veíamos por los pasillos. Chicos, se os echa un mundo de menos!!!

A la gente que compartió mi vida en Madrid fuera del labo, Marco, por su apoyo siempre que lo necesite y esas pedazo fiestas que montábamos en el piso, espero verte pronto. A Gustavo y Luisa, por hacer del piso un lugar en donde siempre pasaba algo interesante!! Se os echa mucho de menos!!. A Pepe, por muchas y variadas cosas, tantas que mejor lo dejo para decirlo en persona.

A mis amigos de la aventura Neoyorkina, por aguantarme durante la escritura de la tesis y sacarme de casa cuando mas lo necesitaba: Rafa, Miquel, Mariana, Fayna, Maria y Simón.

A mis amigos de Valencia, porque se que siempre estarán ahí para lo que necesite. Rafa, porque es un chico excelente y por ofrecerme su amistad desde siempre, espero que por muchos años!!. A Vicente, porque siempre tiene un abrazo cuando me ve, y siempre saca tiempo para una charla.

A Myriam por ser simplemente maravillosa. Por compartir todo conmigo e iluminarme la vida.

A mis padres y hermana, porque si, porque son mi familia y los quiero mogollón, aunque parezca que cada vez me intente alejar más de ellos, siempre los llevo en el corazón.

## **ABREVIATURAS**

A	Adenina
ADN	Acido desoxiribonucleico
ARN	Acido ribonucleico
BrEt	Bromuro de Etidio
C	Citidina
CTL	Linfocitos T citotóxicos
dN	Proporción de mutaciones no sinónimas
DO	Densidad óptica
dS	Proporción de mutaciones sinónimas
ECLIA	Inmunoensayo de electroquimioluminiscencia
ECP	Efecto citopático
EDTA	Acido etilendiaminotetraacético
FBS	Suero bovino fetal
FMDV	Virus de la fiebre aftosa
G	Guanidina
IN	Integrasa
HAART	Terapia antiretroviral altamente activa
Kb	Kilobases
LTR	Repeticiones largas
m.d.i.	Multiplicidad de infección
MA	Matriz
NC	Nucleocápside
Nd	Número de diferencias no sinónimas promedio
N <sub>e</sub>	Tamaño poblacional efectivo
nm	Nanometros
ONUSIDA	Organización de la Naciones Unidas sobre el VIH/SIDA
ORF	Pautas de lectura abiertas
PBS	Tampón fosfato salino
PCR	Reacción en cadena de la polimerasa
RT-PCR	Retrotranscripción y reacción en cadena de la polimerasa
Sd	Número de diferencias sinónimas promedio
SDS	Dodecil sulfato sódico
SIDA	Síndrome de la inmunodeficiencia adquirida

SIV	Virus de la inmunodeficiencia del simio
SIV <sub>cpz</sub>	Virus de la inmunodeficiencia del simio aislado de chimpancé
SIV <sub>smm</sub>	Virus de la inmunodeficiencia del simio aislado del simio mangabeye fuliginoso
SU	Proteína de la superficie
T	Timidina
TBE	Tris-borato-EDTA
TCID <sub>50</sub>	Número de dosis infectivas 50 por ml
TM	Transmembrana
VIH-1	Virus de la inmunodeficiencia humana tipo 1
VIH-2	Virus de la inmunodeficiencia humana tipo 2
VSV	Virus de la estomatitis vesicular



# **INDICE**

---

<b>Resumen</b>	1
<b>1.- Introducción</b>	4
1.1.- El virus de la inmunodeficiencia humana tipo 1	4
1.1.1.- Antecedentes y situación actual	4
1.1.2.- Clasificación del VIH-1	6
1.1.3.- Estructura del genoma y descripción de las proteínas virales	7
1.1.4.- Ciclo biológico del VIH-1	10
1.2.- Los virus de ARN y el VIH-1: Conceptos evolutivos	12
1.2.1.- Cuasiespecies, variabilidad y eficacia biológica	12
1.2.2.- Deriva genética	15
1.2.2.1.- Tamaño poblacional efectivo y deriva genética	15
1.2.2.2.- Trinquete de Muller	16
1.2.3.- Selección y evolución adaptativa	17
1.2.3.1.- Paisajes adaptativos	18
1.2.3.2.- Recuperación de eficacia biológica por medio de la evolución adaptativa	18
1.2.4.- Deriva genética y selección adaptativa en el contexto de las cuasiespecies y los paisajes adaptativos	19
1.3.- Antecedentes	20
<b>Objetivos</b>	23
<b>2.- Materiales y métodos</b>	25
2.1.- Cultivos celulares	26
2.1.1.- Células	26
2.1.2.- Medios y condiciones del cultivo celular	26
2.1.3.- Virus e infecciones	26
2.1.3.1.- Calculo de la viabilidad celular	26
2.1.3.2.- Virus s61	26
2.1.3.3.- Clones virales D15, E11, G7, H12, I15, K15	27
2.1.3.4.- Titulación de los virus	27
2.1.3.5.- Experimento para la recuperación de la eficacia biológica	28
2.1.3.6.- Competiciones virales para el calculo de la eficacia	28

biológica	
2.2.- Métodos moleculares	30
2.2.1.- Extracción del ácido desoxiribonucleico (ADN) de los cultivos infectados	30
2.2.2.- Extracción del ácido ribonucleico (ARN) de los cultivos infectados	30
2.2.3.- Amplificación de los genomas virales consenso	32
2.2.3.1.- A partir de ADN	32
2.2.3.2.- A partir de ARN	33
2.2.4.- Obtención de las cuasiespecies virales de cada clon biológico en los pases inicial y final	35
2.2.5.- Ensayo para la determinación de la eficacia biológica de los clones biológicos	35
2.2.6.- Secuenciación	36
2.2.7.- Inmunoensayo de electroquimiluminiscencia (ECLIA) para la determinación de los niveles de la proteína p24 viral	37
2.3.- Cálculo de la eficacia biológica y análisis de secuencia	38
2.3.1.- Cuantificación de los variantes virales en geles de acrilamida y cálculo de la eficacia biológica	38
2.3.2.- Obtención de los genomas consenso completos	38
2.3.3.- Alineamientos de secuencias	38
2.3.4.- Heterogeneidad, sustituciones sinónimas, sustituciones no sinónimas y generación de árboles filogenéticos	39
2.3.5.- Obtención de las dS y dN de cada clon viral y pase por medio de una ventana deslizante a lo largo del genoma	39
2.3.6.- Análisis estadístico	40
<b>3.- Resultados</b>	<b>42</b>
3.1.- Recuperación de la eficacia biológica de los clones virales	43
3.1.1.- Análisis del título y el valor por ECLIA de la proteína viral p24 de los clones virales en los pases inicial y final de la recuperación de la eficacia biológica	43
3.1.2.- Diferenciación por movilidad electroforética en geles de poli(acrilamida) de los clones virales iniciales y finales de los pases de recuperación de eficacia biológica	45
3.1.3.- Cálculo de la eficacia biológica en los clones virales iniciales y	46

finales	
3.1.3.1.- Clones virales del linaje D	46
3.1.3.2.- Clones virales del linaje G	48
3.1.3.3.- Clones virales del linaje I	49
3.1.3.4.- Clones virales del linaje K	51
3.1.3.5.- Clones virales del linaje H	53
3.1.3.6.- Clones virales del linaje E	54
3.1.3.7.- Análisis y cuantificación de las eficacias biológicas iniciales y finales	55
3.1.3.8.- Eficacias biológicas de los clones biológicos relativas al clon J1	57
3.2.- Comparación de los genomas consenso de los clones	59
3.2.1.- Árbol filogenético de los clones virales	59
3.2.2.- Análisis de las sustituciones fijadas en el genoma viral entre el pase inicial y final de recuperación de eficacia biológica	61
3.3.- Comportamiento de las cuasiespecies virales de los clones biológicos antes y después de la recuperación de eficacia biológica	64
3.3.1.- Análisis de los cambios de heterogeneidad entre el pase inicial y final de recuperación de eficacia biológica	65
3.3.2.- Análisis de los cambios en las sustituciones sinónimas (Sd) entre el pase inicial y final de recuperación de eficacia biológica	67
3.3.3.- Análisis de los cambios en las sustituciones no sinónimas (Nd) entre el pase inicial y final de recuperación de eficacia biológica	69
3.3.4.- Estudio del cociente dN/dS para las regiones estudiadas	70
3.3.5.- Análisis conjunto del aumento de heterogeneidad, Sd, Nd y eficacia	71
3.3.5.1.- Correlación entre heterogeneidad y Sd o Nd	72
3.3.5.2.- Correlación entre heterogeneidad y eficacia biológica	73
3.3.6.- Análisis de la acumulación de mutaciones en los clones moleculares estudiados	74
3.3.7.- Análisis de clones moleculares con codón de terminación	76
3.3.8.- Análisis de la distribución de las sustituciones sinónimas y no sinónimas en las distintas regiones estudiadas	77
3.3.8.1.- Región gen <i>gag</i> (p17/p24)	79
3.3.8.2.- Región gen <i>vpu</i>	81
3.3.8.3.- Región gen <i>env</i> (V1-V2)	83

---

3.3.8.4.- Región gen <i>env</i> (V3-V4)	85
3.3.9.- Aparición de deleciones en la región V1-V2 del gen de la envuelta después de los pases de recuperación	87
<b>4.- Discusión</b>	89
4.1.- Aumento de eficacia biológica, título viral y los niveles de la proteína p24	91
4.2.-Variación en los genomas consenso virales tras los pases seriados	92
4.3.- Posible efecto de la heterogeneidad de la población viral en la eficacia biológica	96
4.4.- Relación de la heterogeneidad, Sd, Nd, eficacia biológica y la proporción de dN/dS	98
4.5.- Acumulación de mutaciones en la población viral tras los pases seriados	99
4.6.- Deriva genética y selección en la cuasiespecie viral	100
4.7.- Análisis de los sitios variables y conservados durante los pases seriados	101
<b>5.- Conclusiones</b>	105
<b>Bibliografía</b>	107

## **RESUMEN**

The human immunodeficiency virus type 1 (HIV-1) has been greatly studied for its importance as human pathogen. Nowadays is widely distributed along the world, in the form of 9 different subtypes and at least 30 different recombinant forms. HIV-1 as other RNA viruses is known to be very variable due to the high error prone polymerase that possesses, coupled with the high number of viral particles and generations that are produced in a very limited amount of time. These characteristics produces in RNA viruses a specific population structure known as cuasiespecies, that enhances viruses adaptability to fluctuating environments.

Some studies have tried to shed light on how HIV-1 viral cuasiespecie is being affected by transmission events, antiretroviral treatments and effect of the compartmentalization *in vivo*. Although these studies are important, there is still a lack of basic knowledge in the way HIV-1 evolves after population fluctuations. Fluctuations in population sizes are known to be important in shaping other RNA viruses, as foot-and-mouth disease virus (FMDV) or vesicular stomatitis virus (VSV), evolution and viral fitness. This same fluctuations are not well understood in the case of HIV-1.

In order to understand populations fluctuations in the HIV-1, ten previously bottleneck passaged viruses that have seen its fitness reduced, were serially passaged with the aim of recovering their viral fitness. This situation was necessary to simulate a fluctuation condition. Our results indicated that after the serial pasajes there was a general increase in viral fitness in the clones studied. To characterize viral recovery of fitness, we sequenced the complete viral consensus genome and found that a low number of mutations were fixed (average of 1.6), although some viruses did not fix any mutation, indicating other parameters might be affecting viral fitness recovery. One of that parameters could be the heterogeneity of the viral cuasiespecies. We analyzed three different genomic regions and the viral heterogeneity in each region was studied. We found that there was a correlation between the increase of viral fitness and heterogeneity, this implied that the viruses needed the generation of variability to recover fitness, thus pointing out the importance of cuasiespecies structure in the viral evolution.

# **1.- INTRODUCCIÓN**



## **1. 1.- EL VIRUS DE LA INMUNODEFICIENCIA HUMANA TIPO 1.**

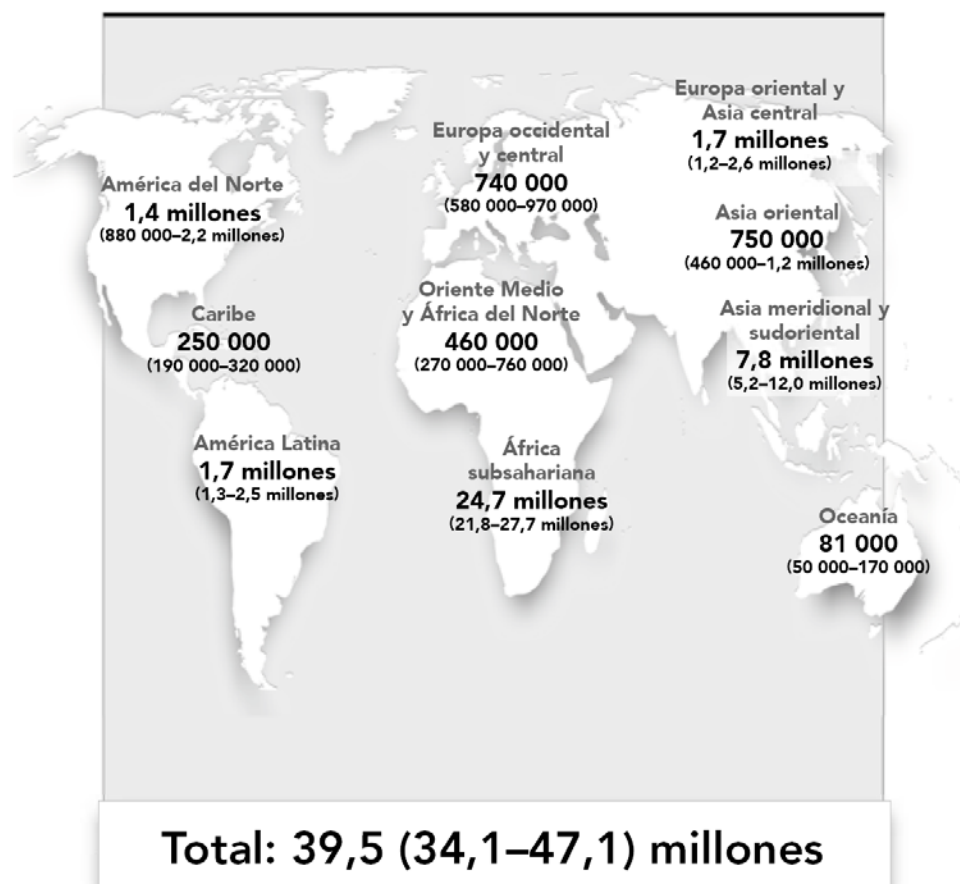
### 1. 1. 1.- Antecedentes y situación actual.

En el año 1981, el síndrome de inmunodeficiencia adquirida (SIDA) fue reconocido como una nueva entidad clínica (Gottlieb y cols 1981; Masur y cols 1981). Se caracterizaba por la presencia de anomalías inmunológicas generalmente acompañadas por infecciones oportunistas, desórdenes neurológicos y formas inusuales de cáncer. Dos años más tarde, en 1983, el virus de la inmunodeficiencia adquirida tipo 1 (VIH-1) fue aislado por vez primera a partir de nódulos linfáticos de un paciente con linfadenopatía (Barre-Sinoussi y cols 1983). En 1984 se demostró la presencia del virus en las muestras de suero de pacientes con SIDA (Gallo y cols 1984). Los primeros estudios epidemiológicos demostraron que este nuevo agente infeccioso era transmitido a través de la sangre o sus derivados, a través de los fluidos genitales y verticalmente de madre a hijo. La secuencia viral más antigua de VIH-1 procede de una muestra de plasma del Zaire del año 1959, y varios estudios sobre el origen de la epidemia postulan que la actual epidemia comenzó en 1930 (Zhu y cols 1998). Se han descrito dos tipos de virus que causan el SIDA, el VIH-1 y el VIH-2, cuyo origen se supone debido a eventos de salto de especie independientes desde el virus de la inmunodeficiencia del simio (SIV) al hombre. Se postula que el tipo 1 proviene del SIV de chimpancé (SIV<sub>cpz</sub>), mientras que el VIH tipo 2 proviene del simio mangabeye fuliginoso (sooty mangabey monkey), SIV<sub>smm</sub> (Heeney y cols 2006).

Globalmente, la pandemia de VIH/SIDA continúa dispersándose por los cinco continentes, aunque fundamentalmente por África y más recientemente por Asia y los países de Europa del este. Según las estimaciones para el año 2006 del Programa Conjunto de las Naciones Unidas sobre el VIH/SIDA (ONUSIDA) y la Organización Mundial de la Salud (OMS) ([http://www.unaids.org/en/HIV\\_data/2006GlobalReport/default.asp](http://www.unaids.org/en/HIV_data/2006GlobalReport/default.asp)), la epidemia de SIDA produjo 2.8 millones de muertes, y se estima que 4.3 millones de personas contrajeron el VIH a lo largo del año, lo que eleva a 39.5 millones el número de personas que viven con el virus en todo el mundo. Desde el comienzo de la epidemia, el África subsahariana es con mucho la región más afectada del mundo. En ésta región vive el 10% de la población mundial pero acumula el 62.5% del total de personas infectadas con el VIH (24.7 millones), siendo además la que acumula el mayor número

de nuevos casos de VIH (2.8 millones, 65% del total). En la Figura 1.1 se puede observar cómo están repartidos los casos de VIH en el mundo.

El número de casos por VIH ha ido en aumento desde el comienzo de la epidemia, aunque en los últimos años, gracias a la generalización del uso de las terapias antiretrovirales se ha frenado sustancialmente su expansión en los países desarrollados. En los últimos años, ONUSIDA y otras organizaciones han tratado de impulsar la implantación del tratamiento antiretroviral en algunos países de África. Por ello, entre los años 2002 y el 2005, el número de personas que recibían tratamiento se incrementó 5 veces, siendo este incremento mayor en el África subsahariana. Sin embargo, no podemos olvidar que por el momento el tratamiento antiretroviral sólo alcanza al 5-10% de la población subsahariana que lo necesita. Por ello, el uso del tratamiento tiene que generalizarse, para evitar nuevas infecciones y paliar los efectos de la enfermedad en los países afectados, sobre todo en África. Además estos datos ponen de relieve la importancia de desarrollar una vacuna para controlar la epidemia en el mundo.

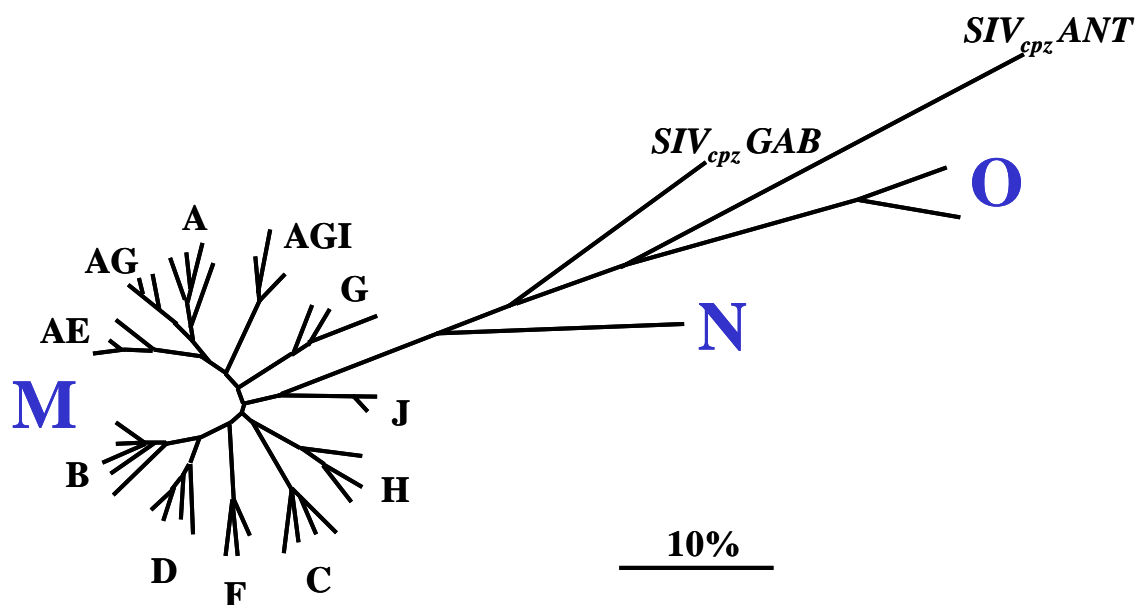


**Figura 1.1.** Número estimado de niños y adultos que vivían con el VIH a finales del 2006. Tomado de ONUSIDA/OMS ([http://www.unaids.org/en/HIV\\_data/epi2006/default.asp](http://www.unaids.org/en/HIV_data/epi2006/default.asp)).

### 1. 1. 2.- Clasificación del VIH-1

Sobre la base de sus características genéticas, morfológicas y biológicas, el VIH-1 se clasifica dentro del género *Lentivirus* (Ratner y cols 1985; Wain-Hobson y cols 1985) perteneciente a la familia *Retroviridae*. El VIH-1 se divide en tres grupos: M (responsable de la pandemia mundial), O (endémico de Camerún y de otros países del África central occidental) y N (que fue descrito a partir de 7 pacientes de Camerún (Simon y cols 1998)). Se postula que la existencia de estos 3 grupos es debida a distintas introducciones en humanos del SIV de primates (Nerrienet y cols 2005).

Dentro del grupo M se distinguen nueve subtipos A, B, C, D, F, G, H, J y K, con una distribución geográfica característica, aunque todos se encuentran representados en África subsahariana. Los subtipos se definieron en base a la distancia genética de las secuencias del gen de la envuelta en árboles filogenéticos (ver Figura 1.2), de manera que se consideró un subtipo al conjunto de secuencias en una rama del árbol y cuya divergencia en el árbol era mayor al 25% respecto a otras ramas, y se consideraron distintos grupos cuando ésta era del 35%. Asimismo se han descrito múltiples formas recombinantes (en la actualidad más de 30 (Louwagie y cols 1993; 1994; Myers 1994; Thomson y cols 2002), ver <http://www.hiv.lanl.gov/content/index>).



**Figura 1.2.** Análisis filogenético de los diferentes grupos y subtipos genéticos del VIH-1. Con letras azules se indican los tres grupos del VIH-1. Las letras A hasta la J indican los subtipos genéticos dentro del grupo M (el K no se encuentra presente en esta figura). En itálicas se señalan los lentivirus de chimpancés más cercanos al VIH-1.

1. 1. 3.- Estructura del genoma y descripción de las proteínas virales.

El examen por microscopía electrónica revela que el virión del VIH-1 tiene una estructura esférica de 80 a 130 nm de diámetro constituida por tres capas concéntricas. La capa interna (nucleoide o core), constituida por la proteína de la cápside (CA o p24), tiene forma de cono truncado y contiene dos copias de ARN viral de cadena sencilla y polaridad positiva (Hu y cols 1990) de aproximadamente 9.2 Kb, formando un complejo ribonucleoproteico constituido fundamentalmente por las proteínas transcriptasa reversa (p51/66), integrasa (IN ó p32) y la proteína de la nucleocápside (NC ó p7), así como también por la proteína vpr. La capa intermedia de forma icosaédrica está constituida por la proteína matriz (MA ó p17). La capa externa o envuelta está constituida por una bicapa lipídica derivada de la célula infectada que contiene proteínas celulares y por espículas formadas por oligómeros de las glicoproteínas de la envuelta vírica: la glicoproteína externa gp120 y la proteína de transmembrana gp41. Ambas proteínas se encuentran asociadas no covalentemente en la superficie del virión. En la Figura 1.3 se muestra una representación de la estructura del virión del VIH-1.

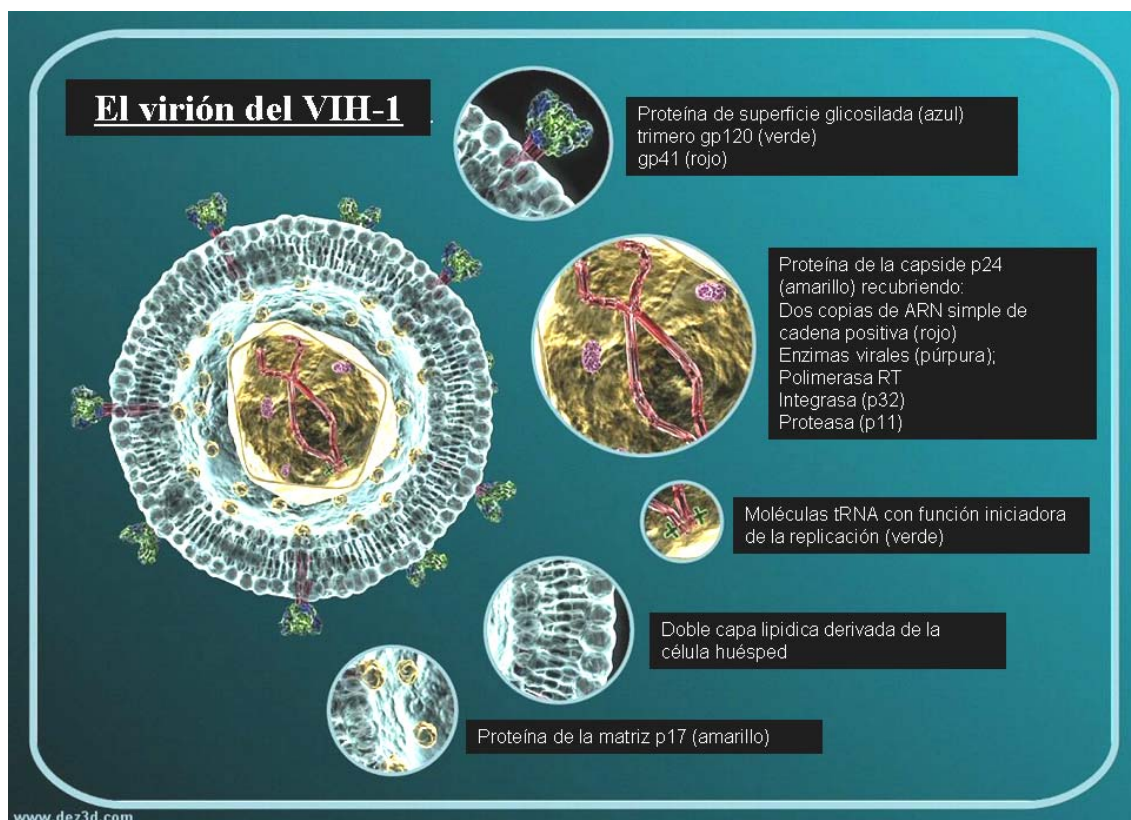
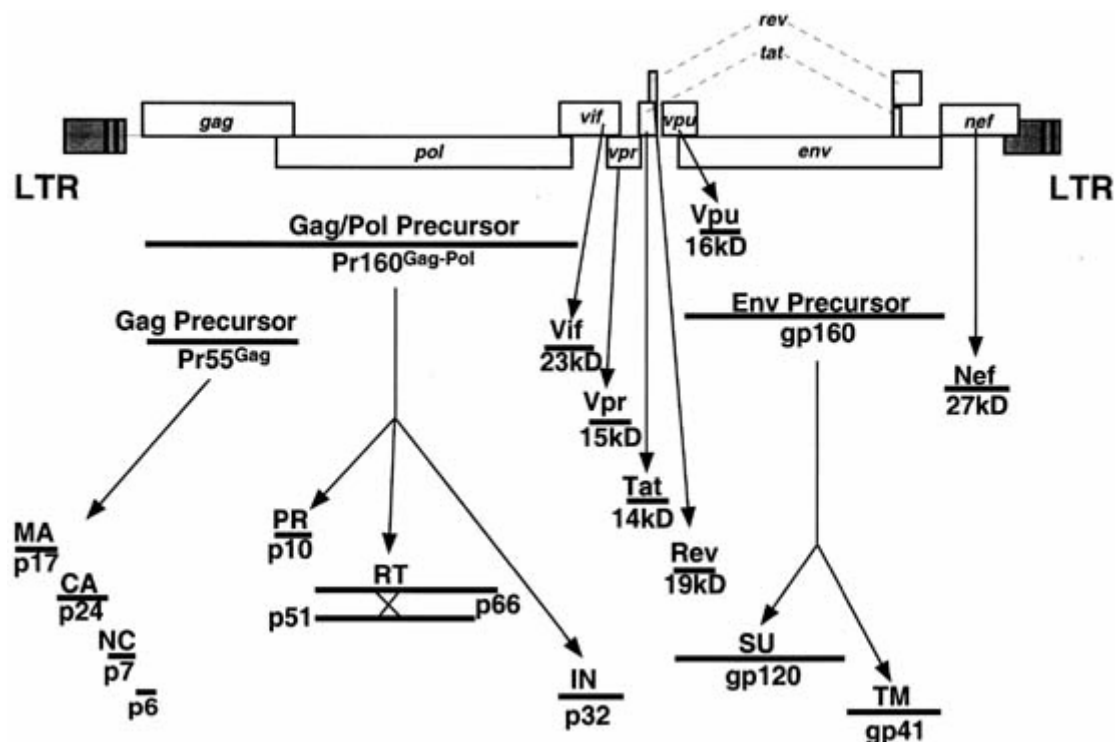


Figura 1.3. Representación de la estructura del virión del VIH-1. Tomado de [www.dez3d.com](http://www.dez3d.com).

El genoma viral del VIH-1 tiene 9 pautas de lectura abierta (open reading frames, ORF). Estas ORFs están flanqueadas por repeticiones terminales de varios cientos de nucleótidos denominadas repeticiones terminales largas (long terminal repeat, LTR). Tres de las ORF corresponden a los genes estructurales virales, *gag*, *pol* y *env*, y las seis ORF restantes corresponden a los genes reguladores, *tat* y *rev*, y los genes accesorios, *vif*, *nef*, *vpr* y *vpu*. Cada una de estas ORF da lugar directamente a proteínas virales, a excepción de los 3 genes estructurales, que codifican los precursores poliproteicos que son posteriormente procesados por proteasas celulares o virales para formar las proteínas virales estructurales maduras. El precursor de *gag* (pr55) da lugar a la proteína de matriz (MA, p17), la proteína de la cápside (CA, p24), la proteína de la nucleocápside (NC, p7), la proteína p6, y las proteínas p1 y p2. El precursor de *gag/pol* (pr160) da lugar a la proteasa (PR, p10), retrotranscriptasa (RT, formada por un heterodímero de las proteínas p51 y p66) y la integrasa (IN, p32). Finalmente el precursor de la envuelta (gp160) da lugar a la proteína de superficie (SU, gp120) y la proteína transmembrana (TM, gp41). En la Figura 1.4 podemos observar una representación de lo anteriormente descrito.



**Figura 1.4.** Representación del genoma del VIH-1. Transcritos primarios que se generan, y las proteínas resultantes de estos. Tomado de (Knipe y cols 2001).

Cada una de las distintas proteínas virales que se generan tienen importantes funciones en la infección viral (ver Tabla 1.1).

<b>Gen</b>	<b>proteína</b>	<b>Descripción</b>
<b>Genes Estructurales</b>		
<i>gag</i>	MA/p17	Forma la proteína de matriz que rodea a la cápside viral, media en el ensamblaje del virión (Yuan y cols 1993; Zhou y cols 1994; Freed y cols 1995) y en la liberación de éste de la membrana plasmática, y dirige al complejo de pre-integración hacia el núcleo celular (Bukrinsky y cols 1993; von Schwedler y cols 1994).
	CA/p24	Forma la cápside viral que protege al genoma viral.
	NC/p7	Está acomplejada con el ARN viral formando el denominado complejo ribonucleoproteico (Berkowitz y cols 1996).
	p6	Esta proteína favorece el ensamblaje del virión y su liberación de la membrana plasmática (Gottlinger y cols 1991; Huang y cols 1995; Zhu y cols 2004).
<i>pol</i>	PR/p10	Es la proteasa viral que proteoliza los precursores poliproteicos virales en sus componentes proteicos simples (Crawford y cols 1985; Katoh y cols 1985).
	RT/p51-p66	Es la polimerasa viral que convierte el ARN de cadena sencilla en ADN de doble cadena (Baltimore 1970; Temin y cols 1970). Está formada por un heterodímero entre la proteína p51 y la p66, después de procesar el fragmento que codifica para la ARNasa H.
	IN/p32	Es el enzima que media la integración del ADN viral en el genoma de la célula hospedador (Fujiwara y cols 1988; Brown y cols 1989).
<i>env</i>	SU/gp120	Es la proteína de superficie, implicada en la interacción y fusión con la membrana celular de la célula hospedador (White 1992). Interacciona con el receptor celular CD4 (Sattentau y cols 1993; Thali y cols 1993).
	TM/gp41	Es la proteína transmembrana, implicada en la interacción y fusión con la membrana celular de la célula hospedador (White 1992).
<b>Genes reguladores</b>		
<i>tat</i>	p14	Incrementa los niveles de transcripción viral por interacción con un motivo en el ARN viral en células infectadas (Dayton y cols 1986; Fisher y cols 1986).
<i>rev</i>	p19	Es la proteína viral que media el transporte de los mensajeros de ARN virales al citoplasma (Meyer y cols 1994).

Genes accesorios		
<i>vif</i>	p23	Esta proteína viral, también denominada factor viral de infectividad, contrarresta el efecto de APOBEC3G, el cual es un inhibidor de la infección viral (Schrofelbauer y cols 2004; Kremer y cols 2005), y por tanto la proteína <i>vif</i> es necesaria para la generación de viriones infectivos (Strebel y cols 1987) en células no permisivas.
<i>vpr</i>	p15	Estimula la transcripción de los genes virales, facilita el transporte al núcleo del complejo de preintegración (Lu y cols 1993; Heinzinger y cols 1994) y mantiene a la célula infectada en la fase G <sub>2</sub> del ciclo celular (Poon y cols 1998).
<i>vpu</i>	p16	Incrementa la liberación de virus (Schubert y cols 1996) y la degradación del CD4 (Chen y cols 1993; Bour y cols 1995).
<i>nef</i>	p27	Reduce la expresión en la superficie de la célula infectada del CD4 (García y cols 1991) y del complejo mayor de histocompatibilidad I (MHC I, (Schwartz y cols 1996)). También incrementa la infectividad viral (Chowers y cols 1994; Miller y cols 1994; Spina y cols 1994), y puede incrementar también la eficiencia de la retrotranscriptasa viral (Schwartz y cols 1995).

**Tabla 1.1.** Descripción de las proteínas virales del VIH-1.

#### 1. 1. 4.- Ciclo biológico del VIH-1.

La infección del VIH-1 comienza con la interacción de la partícula viral con el receptor CD4 y con los co-receptores de quimioquinas de la familia CC ó CXC (Choe y cols 1996; Deng y cols 1996; Doranz y cols 1996; Feng y cols 1996) de la célula huésped, fundamentalmente los linfocitos T CD4<sup>+</sup> cooperadores. Entre ellos los más importantes para la infección por el VIH-1 son: el CXCR4 y el CCR5 (Figura 1.5, paso 1). Esta interacción produce la fusión de la partícula viral con la membrana celular, permitiendo la entrada en la célula de la cápside viral (pasos 2 y 3). En este momento, la cápside se abre parcialmente y se produce la retrotranscripción del genoma viral por parte de la retrotranscriptasa viral (paso 4). Posteriormente, la cápside y el genoma viral retrotranscrito formando un complejo de pre-integración son transportados al núcleo (paso 5), donde se finaliza la retrotranscripción (paso 6). Seguidamente, se integra el genoma viral en el genoma de la célula hospedador con la ayuda de la integrasa viral, produciéndose la transcripción de los genes virales por la ARN polimerasa II de la

célula cuando ésta recibe señales de activación. En este paso, la proteína viral tat actúa favoreciendo la alta producción de los transcritos virales (paso 7). Estos transcritos virales generados son transportados al citoplasma (paso 8) gracias a la proteína viral rev. La traducción del precursor del gen de la envuelta se produce en el retículo endoplasmático (paso 8), mientras que las poliproteínas gag y gag-pol son traducidas en los polirribosomas (paso 9). Tanto las proteínas de la envuelta como las de los genes virales *gag* y *pol*, son transportadas a la membrana celular, donde comienzan a formar complejos virales en la membrana citoplasmática, dando lugar a partículas esféricas virales inmaduras que contienen las glicoproteínas de la envuelta TM y SU. La liberación de estas partículas de la membrana plasmática está mediada por la proteína p6. El procesamiento proteolítico de las poliproteínas gag y pol por la proteasa viral inmediatamente después de la liberación de la partícula viral de la célula hospedadora produce la maduración de la partícula viral, dando lugar a la forma característica en cono truncado de la cápside madura (pasos 10 y 11). En el paso 12 se muestra cómo algunas de las proteínas gp120 generadas que finalmente no acaban asociadas a viriones son liberadas al medio.

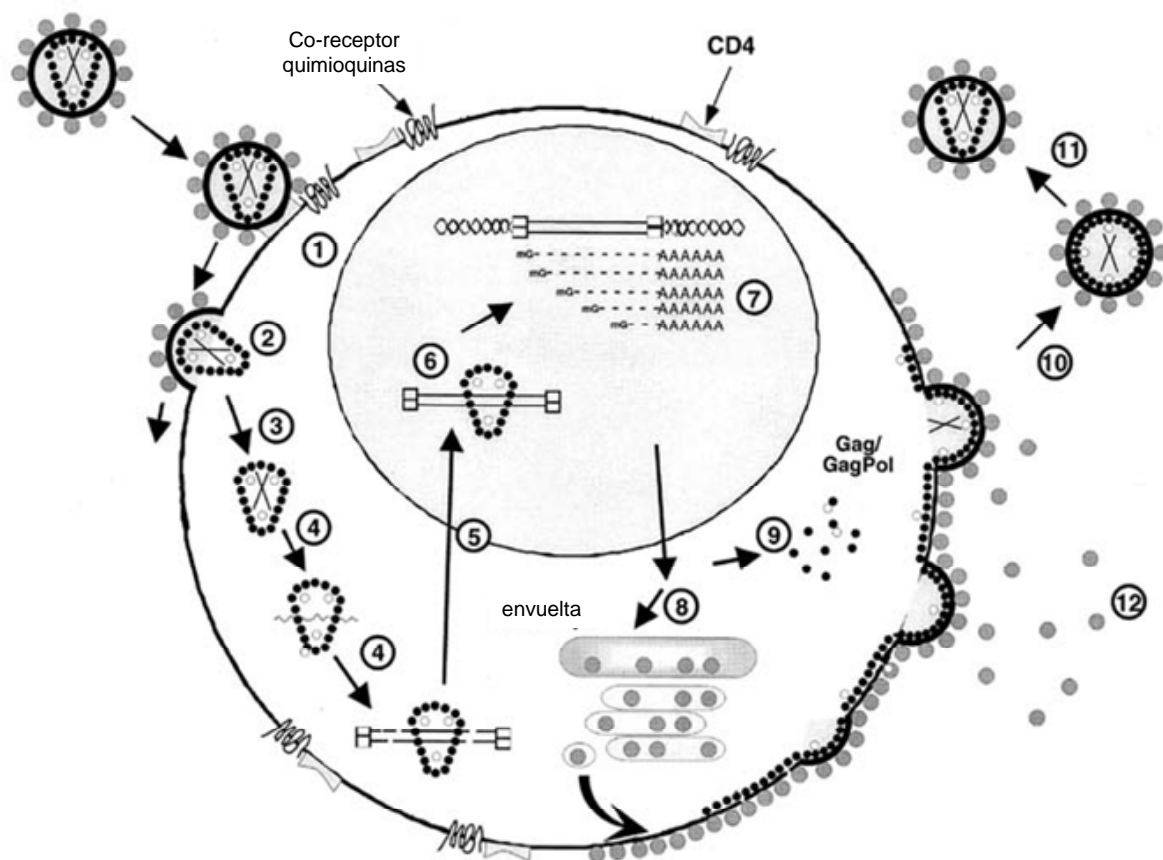


Figura 1.5. El ciclo biológico del VIH-1.



## **1. 2.- LOS VIRUS DE ARN Y EL VIH-1: CONCEPTOS EVOLUTIVOS.**

Por un lado, los virus de ARN son estudiados por su importancia como agentes etiológicos de enfermedades infecciosas y por su capacidad de generar epidemias tanto en el hombre como en animales. En segundo lugar debido a su gran capacidad evolutiva son excelentes modelos evolutivos (Domingo y cols 2001). El VIH-1 constituye una de las epidemias más importantes para el hombre. Por ello, los estudios sobre su evolución tanto en pacientes como en cultivos son importantes a la hora de entender mejor la patogenia de la enfermedad y la capacidad del virus para eludir el sistema inmune del hospedador y las terapias antiretrovirales.

A continuación presentaremos los conceptos evolutivos más importantes a la hora de entender el proceso evolutivo al que se ve sometido el VIH-1 durante la infección.

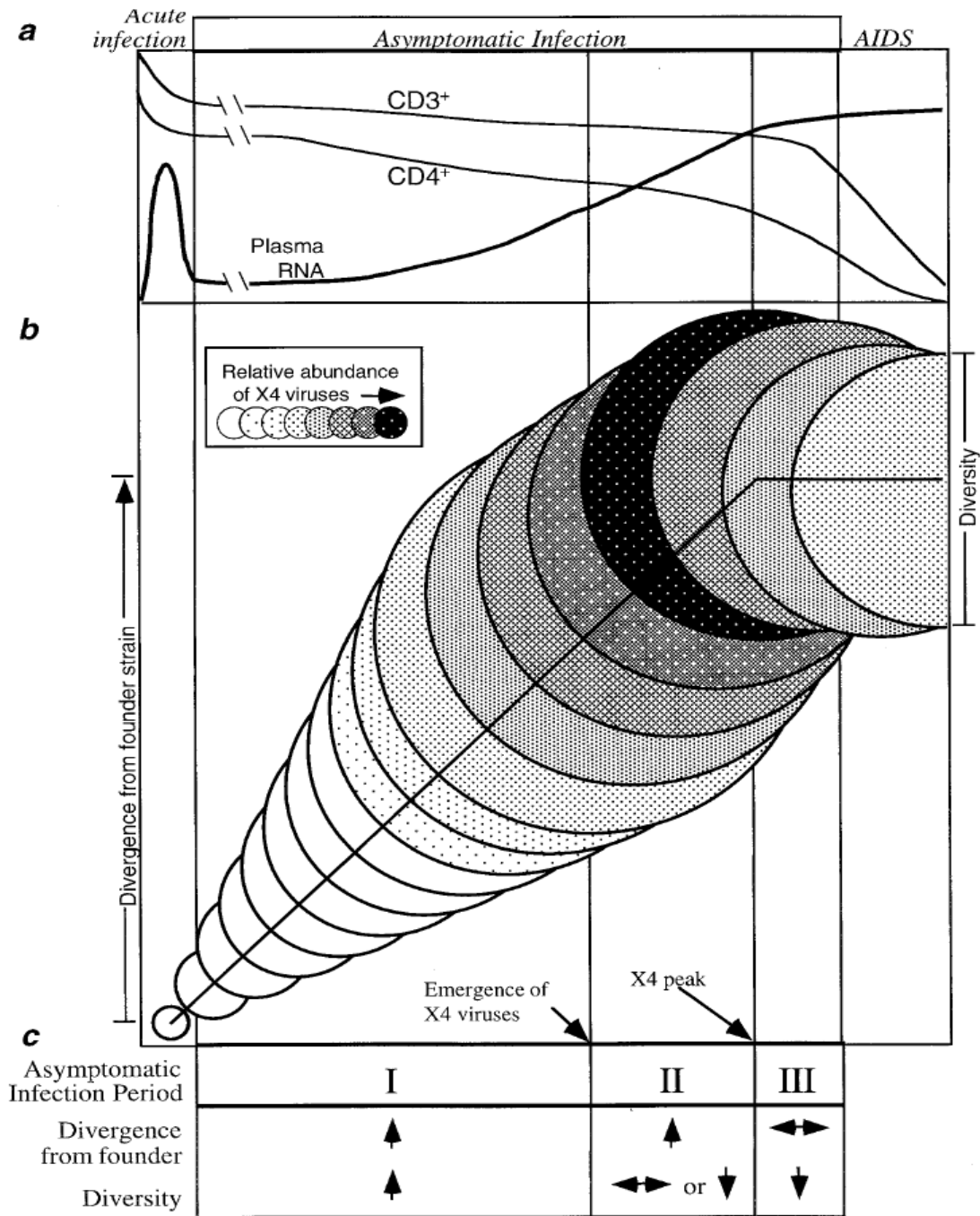
### 1. 2. 1.- Cuasiespecies, variabilidad y eficacia biológica.

Los virus de ARN tienen unas altas tasas de mutación (Drake y cols 1998), debido principalmente a que sus polimerasas virales no tienen mecanismos de corrección de errores, como por ejemplo la actividad 3' exonucleasa (Steinhauer y cols 1992). Los virus producen gran número de partículas virales en cada ciclo viral (Holland y cols 1982; Neumann y cols 1998), y tienen ciclos replicativos cortos. Todo esto hace que las poblaciones de virus de ARN estén formadas por un complejo espectro de mutantes que ha sido denominado cuasiespecie (Domingo y cols 1985; Eigen y cols 1988; Steinhauer y cols 1989). Todas estas características también las encontramos en el VIH-1. Tiene una alta tasa de mutación, medida en  $3 \times 10^{-5}$  sustituciones/sitio/generación (Mansky y cols 1995). Produce un gran número de partículas virales en cada ciclo viral de  $10^8$  a  $10^9$  viriones por día (Ho y cols 1995; Wei y cols 1995), además del gran número de células que se pueden llegar a infectar por día del orden de  $10^7$  a  $10^8$  (Chun y cols 1997). Además, la existencia de recombinación (Jung y cols 2002; Iglesias-Sanchez y cols 2005) aumenta la capacidad de variación y posibilita la regeneración de genomas libres de mutaciones deletéreas, o de genomas con combinaciones de mutaciones. La recombinación entre virus de distintos grupos y subtipos del VIH-1 ha sido ampliamente descrita, y ha dado lugar a las numerosas formas recombinantes (<http://www.hiv.lanl.gov/content/hiv-db/mainpage.html>). Aunque

más difícil de detectar, la recombinación entre virus pertenecientes al mismo subtipo genético también se ha descrito (Diaz *y cols* 1995; Sala *y cols* 1995; Zhu *y cols* 1995; Butto *y cols* 1997).

El VIH-1 se ve sometido a una gran y variada presión selectiva durante la infección, pero según lo explicado anteriormente es capaz de generar una gran variabilidad genética, lo que le permite adaptarse a los cambios ambientales de una manera muy rápida. Varios ejemplos de esta dinámica incluyen el escape por parte del virus del sistema inmunológico del paciente, tanto de la respuesta humoral (van 't Wout *y cols* 1998; Richman *y cols* 2003), como de los linfocitos-T citotóxicos (CTL, (Borrow *y cols* 1997; Price *y cols* 1997; Brander *y cols* 2003; Cao *y cols* 2003; Draenert *y cols* 2004), así como la generación de resistencias frente a la terapia antiretroviral suministrada a los pacientes para combatir al virus (Johnson *y cols* 2003). El VIH-1 también es capaz de adaptarse a los distintos ambientes que encontramos en los hospedadores, produciéndose aquí también una selección y compartimentalización de la población viral (Sanjuan *y cols* 2004; Borderia *y cols* 2007). Esta situación contribuye a la generación de variabilidad que se acumula a la anteriormente descrita.

La variabilidad no es simplemente una medida de la complejidad de la cuasiespecie, sino que puede tener importantes implicaciones para el progreso de la enfermedad en el VIH-1. Se han realizado numerosos estudios en los que se ha relacionado el incremento de variabilidad con la progresión clínica de la enfermedad (Wolinsky *y cols* 1996; Ganeshan *y cols* 1997; Essajee *y cols* 2000; Anastassopoulou *y cols* 2005). En concreto se ha descrito el aumento de la variabilidad en el VIH-1 durante una infección en pacientes con progresión típica (de 8 a 10 años, ver Figura 1.6) por Shankarappa *y colaboradores* en 1999. En este estudio se comprobó cómo a partir de una cuasiespecie que era homogénea en el momento de la transmisión, ésta iba aumentando su heterogeneidad a lo largo de la infección viral para estabilizarse justo antes de entrar en fase de SIDA. Esto mismo se observó para la divergencia (el número de sustituciones acumuladas durante el proceso evolutivo). Todos estos cambios se asociaron con la presión ejercida por el sistema inmunológico hasta que éste no podía soportar más la carga ejercida por la infección viral y se producía la disminución definitiva de células T CD4<sup>+</sup> (ver Figura 1.6).



**Figura 1.6.** Descripción esquemática de los cambios en la divergencia y diversidad a lo largo de la infección por VIH-1. A) Estadios de la infección por VIH-1 desde el inicio hasta la aparición del SIDA. En esta figura se muestran los niveles de ARN viral en plasma, de linfocitos T CD4+ y en general de las células CD3+. B) Generación de divergencia y diversidad. La diversidad media de la población viral viene representada por el diámetro de los círculos. C) Tabla resumen de divergencia y variabilidad según progresa la enfermedad. Los virus X4 son linfocito-tróficos, responsables de una progresión más rápida a SIDA. Figura tomada de (Shankarappa y cols 1999).

Un concepto importante a la hora de estudiar los virus es el de la eficacia biológica, que puede definirse como el parámetro que mide la adaptación replicativa de un organismo a su medio (Domingo y cols 1997; Domingo y cols 1999). Es una variable relativa ya que se mide con respecto a un virus de referencia. Generalmente se usa un marcador genotípico o genético que pueda diferenciar a las dos poblaciones virales (Holland y cols 1991). En un ambiente constante, la eficacia biológica está directamente influenciada por la aparición y fijación de mutaciones y por las fluctuaciones poblacionales a las que los virus se ven sometidas.

En una cuasiespecie, la distribución de secuencias nucleotídicas no es arbitraria sino que se centra en torno a una secuencia maestra, que es la de mayor eficacia biológica en esas condiciones ambientales, y que puede coincidir o no con la secuencia consenso de la población. La proporción de cada variante en la población global viene determinada por la eficacia biológica de cada una de ellas. En condiciones ambientales constantes, una cuasiespecie replicante se halla en equilibrio poblacional cuando mantiene su secuencia consenso invariable, aunque el espectro de mutantes sufra alteraciones por la generación estocástica de mutaciones. Este equilibrio puede alterarse por la generación de mutantes con mayor ventaja selectiva o por un cambio ambiental que confiera una ventaja selectiva a alguna variante hasta entonces minoritaria. Es decir, a menos que las condiciones ambientales permanezcan controladas y constantes, la superioridad o inferioridad de una variante en particular es fugaz. Durante las infecciones naturales las perturbaciones del equilibrio poblacional ocurren frecuentemente, por lo que la estructura en forma de cuasiespecie proporciona a los virus ARN la capacidad de responder y adaptarse rápidamente a los cambios ambientales, asegurando así su supervivencia (Domingo y cols 1997).

### 1. 2. 2.- Deriva genética.

#### 1. 2. 2. 1.- Tamaño poblacional efectivo y deriva genética.

El tamaño poblacional y la deriva genética están íntimamente ligados, pues la magnitud de deriva genética que observamos depende del tamaño poblacional. La deriva genética se puede definir como el muestreo aleatorio de los genotipos entre generación y generación. Cuando una población está afectada por deriva genética, ésta evoluciona de manera aleatoria por la acumulación de mutaciones (Kimura 1985). Por

probabilidad, la mayoría de estas mutaciones serán neutrales o con valores deletéreos, siendo relativamente poco probable la aparición de mutaciones con alto valor de selección. El parámetro que permite estimar la deriva que sufre una población es el tamaño poblacional efectivo ( $N_e$ ), que se define como el tamaño poblacional bajo el que, en ausencia de cualquier otro factor, existiría el mismo valor de deriva que la que hay en la población real (Caballero 1994). Cuanto más bajos sean los valores de  $N_e$ , mayor será el efecto de la deriva genética y por tanto menor el de la selección natural, y por el contrario a valores altos de  $N_e$ , mayor será el efecto de la selección natural y menor el de la deriva genética. El efecto de la deriva genética en el VIH-1 se ha comprobado *in vivo* en secuencias obtenidas del gen de la envuelta que abarcan desde la región C2 a la V5 (Shriner y cols 2004). Estos autores estimaron que el tamaño efectivo del VIH-1 *in vivo* sería de  $10^3$ , que es un valor consistente con estimaciones realizadas previamente por otros métodos (Brown 1997; Seo y cols 2002).

#### 1. 2. 2. 2.- Trinquete de Muller.

Por trinquete de Muller se conoce al proceso por el cual se produce una acumulación y fijación irreversible de mutaciones deletéreas en poblaciones asexuales de organismos debido a un bajo  $N_e$  (Muller 1964). Para simular este proceso se recurre a la realización de pases de virus en cuellos de botella. En poblaciones virales el proceso de acumulación de mutaciones por el trinquete Muller es irreversible por la falta de mecanismos de recombinación en general. Entonces, es posible que por deriva se produzca una disminución de la eficacia biológica. Si el proceso del trinquete de Muller se mantiene durante un largo tiempo se puede llegar a la extinción de la población (Lynch y cols 1993; Domingo y cols 2005; Domingo y cols 2005) y en el VIH-1 (Yuste y cols 1999). Aún así, al tratarse de un fenómeno estocástico, existe gran variabilidad entre réplicas, e incluso se han observado casos aislados de ganancia de eficacia (Shaw y cols 2000), también observado en el VIH-1 (Yuste y cols 1999). Los cuellos de botella son un fenómeno importante en evolución debido a que la mayoría de las transmisiones virales implican un paso por cuello de botella que condiciona inicialmente la infección viral (Elena y cols 2001). En el caso del VIH-1 el trinquete es operativo y se ha demostrado en estudios *in vitro* (Yuste y cols 1999). En otros sistemas biológicos el efecto del trinquete de Muller está ampliamente aceptado al haber sometido a repetidos cuellos de botella a diversas poblaciones experimentales, como por ejemplo *Drosophila*

melanogaster (Mukai y cols 1972), *Caenorhabditis elegans* (Keightley y cols 1997), *E. coli* (Kibota y cols 1996), o *Aspergillus nidulans* (Bruggeman y cols 2003). Asimismo, la acción del trinquete también ha sido estudiada en otros virus de ARN, como el bacteriófago  $\phi 6$  (Chao 1990), los virus de la estomatitis vesicular (VSV; (Duarte y cols 1992; Clarke y cols 1993)) y de la fiebre aftosa (FMDV; (Escarmis y cols 1996)).

### 1. 2. 3.- Selección y evolución adaptativa.

La selección es la fuerza modeladora de la evolución en poblaciones con valores de  $N_e$  elevados. Según el modelo seleccionista de evolución, la selección actúa directamente sobre las variantes seleccionando aquellos que son beneficiosas para la población, lo que se conoce como selección positiva. Por el contrario, según el modelo neutral de evolución la mayoría de los cambios a nivel molecular son debidos a mutaciones neutrales, que no confieren ninguna ventaja selectiva, y por tanto la evolución sería estocástica. Ambos modelos reconocen el efecto de la selección purificadora, que actúa sobre las mutaciones deletéreas eliminándolas de la población. Para diferenciar entre ambos modelos se recurre a los cálculos de las tasas de cambio no sinónimas (dN) y las tasas sinónimas (dS). Las sustituciones nucleotídicas no sinónimas son todas aquellas que producen un cambio en el aminoácido que codifica el codón, mientras que las sustituciones nucleotídicas sinónimas no promueven un cambio en el aminoácido codificado. Según esto, la teoría neutralista de la evolución predice que la mayoría de los cambios serán neutrales, lo que implica que el cociente entre dN/dS debería ser menor o igual a 1. Para los seleccionistas sin embargo, las sustituciones son selectivas y por tanto el cociente podría tomar tanto valores mayores, iguales o menores a 1. Ninguno de las dos teorías niega la existencia de mutaciones deletéreas. En general valores del cociente mayores a 1 indican selección positiva, mientras que valores menores a 1 indican selección purificadora. Para el VIH-1 se ha demostrado que para los genes de *gag*, *pol*, y *env* a lo largo de la epidemia existe un cociente  $dN/dS < 1$  (Brown y cols 1988; Li y cols 1988; Gojobori y cols 1990; Seibert y cols 1995; Smith y cols 2004; Edwards y cols 2006). Por ello parece que la evolución del VIH-1 está modelada mayoritariamente por una selección purificadora en pacientes infectados, aunque podrían existir momentos puntuales de selección positiva (Shriner y cols 2004; Bazykin y cols 2006).

1. 2. 3. 1.- Paisajes adaptativos.

El concepto de los paisajes adaptativos fue introducido por Sewall Wright en 1932 (Wright 1932). En biología evolutiva los paisajes adaptativos son usados para visualizar la relación entre el genotipo y la eficacia biológica. El paisaje estaría formado por montañas y valles, que representarían la eficacia biológica, siendo las montañas o picos adaptativos valores elevados de eficacia y los valles, valores bajos de eficacia. Según esto, los distintos genotipos se distribuirían a lo largo de este paisaje, tendiendo a agruparse los genotipos similares y a separarse los genotipos no relacionados. La rugosidad o complejidad de este paisaje vendrá determinada por la magnitud de epistasia del sistema (Wolf y *cols* 2000). La epistasia se produce cuando la acción de un gen está afectada por uno u otros genes independientes de aquel y juega un papel importante en los paisajes adaptativos. Una población en evolución tenderá a ascender los picos adaptativos por medio de cambios genéticos, hasta que consiga alcanzar la cima del pico adaptativo. En un paisaje adaptativo determinado tendremos un pico máximo y varios locales. Las poblaciones nunca se podrán mover en contra de la selección (descender un pico), dependen de la deriva genética (por medio del trinquete de Muller) para moverse de un pico adaptativo a otro.

1. 2. 3. 2.- Recuperación de eficacia biológica por medio de la evolución adaptativa.

Cuando se ha producido un evento de deriva genética tras un cuello de botella, y ha actuado el trinquete de Muller reduciendo la eficacia biológica de la población, este proceso habrá permitido que la población se desplace del pico adaptativo en el que se encontraba. En este momento, esta población podría intentar alcanzar un nuevo pico adaptativo con mayor eficacia biológica del que provenía anteriormente. Las poblaciones pueden recuperar eficacia por medio del aumento del  $N_e$ , lo cual permite que la selección natural actúe, en contraposición al efecto de la deriva genética, seleccionando así de la variabilidad disponible aquellas mutaciones beneficiosas para la población, ó eliminando aquellas deletéreas. Esto se ha demostrado experimentalmente realizando pases seriados en gran tamaño de población de poblaciones virales previamente debilitadas de diversos virus ARN, como por ejemplo, en el VSV (Elena y *cols* 1998), FMDV (Escarmis y *cols* 1999) y el VIH-1 (Yuste y *cols* 2005). La

recuperación de eficacia se puede producir por reversión de mutaciones deletéreas acumuladas, pero también pueden fijarse mutaciones compensatorias que compensen el efecto negativo producido por las mutaciones acumuladas. Los casos de reversión son poco probables pero pueden producirse, como se observó en el caso del VIH-1 (Yuste y cols 2005). Para que se produzca evolución compensatoria se tiene que asumir que existe epistasia positiva (Sanjuan y cols 2005). Tanto las reversiones como las compensaciones no son autoexcluyentes de manera que las dos pueden estar contribuyendo activamente a que se produzca el aumento de la eficacia biológica (Escarmis y cols 1999; Escarmis y cols 2002; Novella y cols 2004).

#### 1. 2. 4.- Deriva genética y selección adaptativa en el contexto de las cuasiespecies y los paisajes adaptativos.

Existe una estrecha relación entre deriva genética y selección adaptativa. Como ya hemos descrito, la primera permite cruzar valles adaptativos, fenómeno no permitido por la selección adaptativa. Sin embargo, la selección permite que una vez producido el salto se produzca una adaptación al nuevo pico adaptativo y por tanto una recuperación de la eficacia biológica perdida. Por ello, ambos son necesarios para la evolución de los virus, y se dan de manera constante a lo largo de la evolución de éstos. En términos de variabilidad genética ambos procesos podrían ser detectables, dado que cuando se produce una reducción del tamaño poblacional se da una homogenización de la población, es decir, se produce una reducción de la variabilidad genética de la cuasiespecie. Sin embargo en un proceso de selección es necesaria la variabilidad, por tanto la complejidad de la cuasiespecie debería aumentar, dado que ésta es la materia prima sobre la que actúa la selección para adaptarse a los ambientes cambiantes. Por todo ello, la acción conjunta de la deriva genética y la selección natural tal como hemos descrito, facilitara una mayor adaptación que si ésta estuviera simplemente mediada por la selección. Esta teoría es conocida como teoría del equilibrio cambiante (Wright 1982). Por ello es importante estudiar las fluctuaciones poblaciones que se dan durante la infección y transmisión viral del VIH-1 para entender mejor como se adapta y evoluciona el virus.

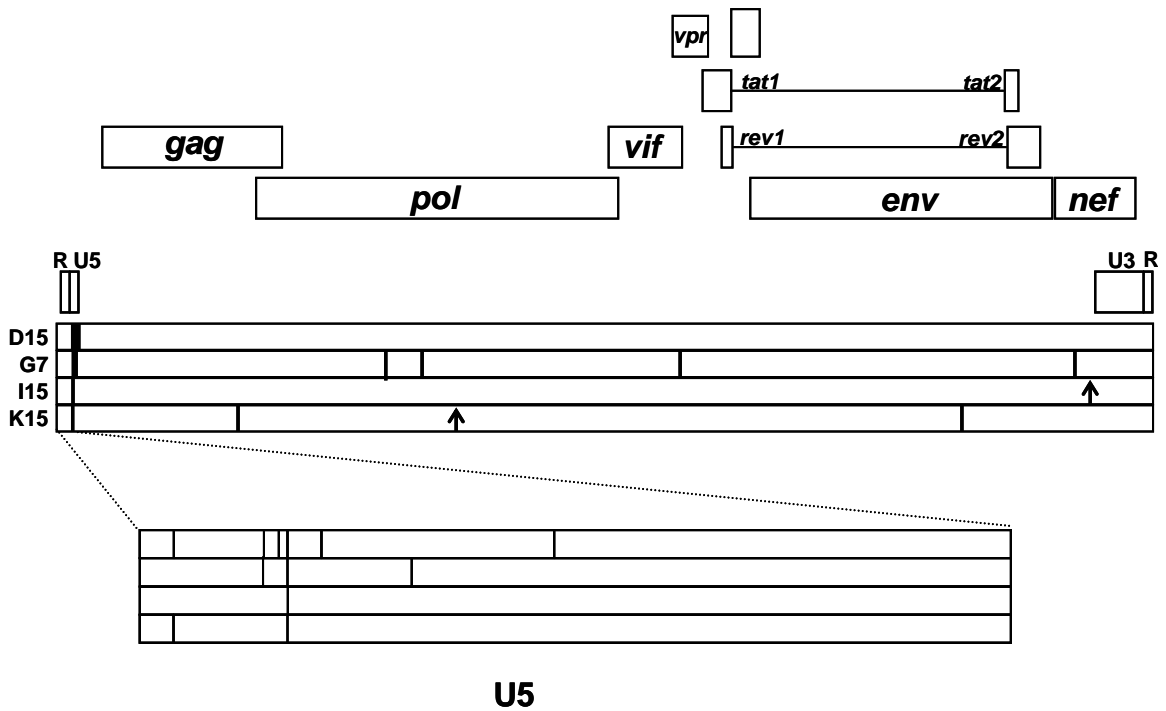


### **1. 3.- ANTECEDENTES.**

Estudios *in vitro* en distintos virus de ARN (FMDV, VSV y  $\Phi 6$ ) han permitido establecer la vigencia de distintas teorías evolutivas, el trinquete de Muller, la Reina Roja, entre otros. En el VIH-1, nuestro grupo ha llevado a cabo distintos trabajos *in vitro* para comprobar parámetros evolutivos de este virus. Se estudió por ejemplo el efecto de la multiplicidad de infección (m.d.i.) en la acumulación de variación genética (Sanchez-Palomino y cols 1993). En el contexto de esta Tesis Doctoral, estudios previos realizados en nuestro laboratorio sobre la evolución de 10 clones virales del VIH-1, sometidos a pases en cuello de botella, detectaron la disminución de su eficacia biológica (Yuste y cols 1999) por medio del trinquete de Muller. Para caracterizar mejor esta disminución, se obtuvo el genoma viral consenso de 4 de estos clones (D15, G7, I15 y K15) antes y después de los cuellos de botella. La comparación mostró la fijación de una serie de sustituciones responsables de la pérdida de eficacia biológica detectada en los clones ((Yuste y cols 2000), ver Figura 1.7). Se observó también que las mutaciones parecían agruparse en la extremo 5' del genoma viral del VIH-1, en concreto en la región de la 5' LTR y el gen *gag*. Para estudiar si estos 4 clones eran capaces de recuperar su eficacia biológica, se realizaron pases seriados en donde se permitía al virus replicar en grandes tamaños poblacionales. Estos pases produjeron como se esperaba (comentado en el apartado 1. 2. 3. 2) la recuperación de eficacia de los clones virales (Yuste y cols 2005). Para determinar qué mutaciones podían ser las responsables de este aumento de eficacia biológica, se realizó la secuenciación de los genomas completos de los pases finales de recuperación de eficacia biológica. Los clones biológicos secuenciados fueron los denominados D15p10, G7p10, I15p10 y K15p10 y fueron publicados en Yuste y cols (2005). Sus genomas se compararon con los genomas finales después de los pases en cuello de botella que habían sido previamente obtenidos (Yuste y cols 2000). Para el clon D15p10 se observaron 6 sustituciones, para el G7p10, 7 sustituciones, para el I15p10, dos y para el clon K15p10, 5 sustituciones. De entre todas las sustituciones encontradas (21 en total, ver Figura 1.7), 13 se produjeron en regiones no codificantes y 8 se produjeron en regiones codificantes. De entre las codificantes, se observaron que 6 fueron sustituciones no sinónimas (75% del total) y 2 fueron sustituciones sinónimas. Las sustituciones no sinónimas que se produjeron fueron por aminoácidos que se encontraban más conservados en las bases de datos de VIH-1 (Leitner y cols 2005). Las transiciones

fueron 3 veces más frecuentes que las transversiones, siendo las sustituciones Adenina (A) → Guanidina (G), Citidina (C) → Timidina (T) y G → A, las más frecuentes (35%, 20% y 15% respectivamente). Se determinó también que el 25% de las sustituciones que se produjeron en el pase final de recuperación eran reversiones a nucleótidos pertenecientes a los clones biológicos iniciales de los pases en cuello de botella (D1, E1, G1, H1, I1 y K1, ver Figura 2.1). Las sustituciones en regiones no codificantes se localizaban mayoritariamente en la región 5'LTR de los distintos clones. Se encontraron 12 sustituciones cercanas al motivo de unión al iniciador de RNAt, denominado "primer binding site" (PBS), en donde se produce el inicio de la replicación del VIH-1, y una en la región U3 del extremo 3'LTR. Algunas de estas sustituciones fueron encontradas en varios de los clones. La sustitución G570A se encontró en los clones D15p10 y K15p10, la sustitución A584G y T584G se encontró en los clones D15p10 y G7p10, y por último la sustitución A587G se observó en los 4 clones en su pase final.

Todos estos resultados sobre la recuperación de eficacia indicaban que pocas mutaciones mediaban el aumento de eficacia (una media de 5 mutaciones por clon). Además, la aparición y fijación de las mismas mutaciones en varios clones virales que habían sido mantenidos independientes hacían suponer la existencia de convergencia. Asimismo era interesante la predominancia de reversiones (25%). A la vista de todos estos resultados se planteó la realización de la presente Tesis Doctoral, para intentar profundizar en el efecto de las fluctuaciones poblacionales (cuellos de botella y pases seriados) sobre el modelado de la variabilidad genética del VIH-1. Por ello se realizó una nueva recuperación de eficacia con nuevos clones pertenecientes a los mismos linajes ya estudiados (D, G, I y K), así como también otros linajes no estudiados previamente (E y H). Se estudió tanto la eficacia, los genomas virales completos, como la clonación molecular de cada uno de los distintos clones para estudiar la variabilidad genética subyacente de la cuasiespecie durante los pases.



**Figura 1.7.** Distribución en el genoma completo del VIH-1 de todas las mutaciones aparecidas durante la recuperación de eficacia biológica. Las barras horizontales debajo del esquema del genoma viral representan los distintos clones biológicos estudiados. Las líneas verticales representan los cambios no sinónimos, mientras que las flechas indican los cambios sinónimos. La región U5 de la región 5'LTR se expandió para apreciar mejor la acumulación de cambios en esta región.

## **OBJETIVOS**

Los objetivos específicos de esta Tesis Doctoral son los siguientes:

- 1) Estudiar la recuperación de la eficacia biológica por medio de pases seriados en gran tamaño poblacional de 6 linajes virales con eficacia biológica disminuída derivados de una población viral del VIH-1.
- 2) Análisis de los genomas virales completos de los 6 linajes virales antes y después de la recuperación de la eficacia biológica, para estudiar el papel de las mutaciones fijadas como posibles responsables de la recuperación.
- 3) Determinación de las cuasiespecies virales de los 6 linajes virales antes y después de la recuperación de la eficacia biológica, para estudiar la implicación de la heterogeneidad de la cuasiespecie viral, y la acumulación de mutaciones sinónimas y no sinónimas en la recuperación de la eficacia biológica de los virus.

## **2.- MATERIALES Y MÉTODOS**

## **2. 1.- CULTIVOS CELULARES.**

### 2. 1. 1.- Células.

Para el cultivo del virus utilizamos las líneas celulares linfoides CD4+ denominadas MT-2 y MT-4 (Miyoshi *y cols* 1981; Yoshida *y cols* 1982). La línea celular MT-2 se uso para las titulaciones de los distintos virus. La línea MT-4 se uso para los pases seriados de los virus y las competiciones.

### 2. 1. 2.- Medios y condiciones del cultivo celular.

Todas las líneas celulares se cultivaron en medio RPMI 1640 (Cambrex) suplementado con suero bovino fetal (FBS) al 10% (PAN Biotech), L-glutamina 2mM, penicilina 100 unidades (U)/ml, estreptomicina 100 µg/ml (Cambrex). A este medio suplementado lo denominaremos a partir de ahora medio completo. Todos los cultivos celulares e infecciones se mantuvieron a una temperatura de 37°C en atmósfera de CO<sub>2</sub> al 5% con una humedad relativa del 95%.

### 2. 1. 3.- Virus e infecciones.

#### 2. 1. 3. 1.- Cálculo de la viabilidad celular.

La viabilidad de los distintos cultivos celulares fue determinada por cuantificación en el microscopio de la proporción de células vivas y muertas en una cámara de Neubauer, tiñéndose las células con azul tripan (Sigma).

#### 2. 1. 3. 2.- Virus s61.

La semilla viral inicial se aisló por cocultivo de linfocitos de sangre periférica del paciente 61/89 en el año 1989 con linfocitos de donante sano en presencia de interleukina 2 (IL-2). Posteriormente esta semilla viral inicial se usó para infectar células MT-4 obteniéndose de esta manera la semilla viral denominada s61. El aislamiento de esta semilla se realizó en el Servicio de Diagnostico y Referencia de Retrovirus (Centro Nacional de Microbiología, Instituto de Salud Carlos III).

2. 1. 3. 3.- Clones virales D15, E11, G7, H12, I15, K15.

Estos clones virales son clones biológicos obtenidos a partir del virus s61 por plaqueo en células MT-4, y posteriormente sometidos a cultivo placa a placa durante 15 pases (Yuste y cols 1999; Yuste y cols 2000). Todos estos clones vieron su eficacia biológica disminuída tras los pases placa a placa. Algunos de ellos se extinguieron antes de los 15 pases, como es el caso de los clones E, G y H en los que sólo se pudieron realizar pases placa a placa hasta los pases 7, 12, 13 respectivamente. De cada uno de estos clones se disponía de una serie de placas aisladas en el último plaqueo que se pudo realizar (Yuste y cols 1999), resuspendidas en 200µl de medio completo. Para la realización del presente experimento se seleccionaron 2 placas de cada clon viral a excepción de los clones E11 y H12 de los cuales sólo se seleccionó una. Teníamos por tanto 6 linajes virales distintos (D, E, G, H, I, y K), y dos réplicas para cada uno de ellos a excepción de los linajes E y H.

Para la obtención de una semilla inicial para el experimento, se infectaron con 100µl de sobrenadante de cada placa en  $1 \times 10^5$  células MT-4, añadiéndose al cabo de 3 días 100µl de medio completo. Cuando se vieron los efectos citopáticos de la infección viral por disgregación de las células MT-4 y la aparición de sincitios (aproximadamente a los 5 días), se añadieron al cultivo  $5 \times 10^5$  células MT-4. Al cabo de 8 días se añadieron  $5 \times 10^6$  células MT-4 en 5ml de medio completo. Y por último a los 10 días se añadieron otras  $5 \times 10^6$  células MT-4 en 5ml de medio completo. En ese momento se controló la replicación viral por determinación de la viabilidad hasta que esta decreció a tan solo el 30% de células vivas, momento en que se recogió el cultivo separando el sobrenadante y las células y se congeló todo a  $-70^\circ\text{C}$ .

2. 1. 3. 4.- Titulación de los virus.

El título los clones iniciales y finales se obtuvo calculando el número de dosis infectivas 50 por ml (TCID<sub>50</sub>/ml). Para ello, se infectaron en placas de cultivo de 96 pocillos  $1 \times 10^5$  células MT-2 por pocillo, con diluciones seriadas de cada clon viral (desde  $10^{-1}$  hasta  $10^{-8}$ ) en 100µl de medio completo. De cada una de las diluciones se realizaron 8 réplicas. A los 3 días de cultivo se añadieron 100µl de medio completo y a los 7 días se observó cada dilución de la placa en busca de efecto citopático (ECP, reconocible por la formación de sincitios, es decir, células gigantes multinucleadas). El



efecto es visible por la formación de sincitios celulares. Se contó el número de pocillos positivos para ECP por dilución y replica en la que aparecen, y obtuvimos el título viral según la fórmula de Spearman-Kärber (Hubert 1984). De cada clon se realizaron 2 réplicas de la titulación.

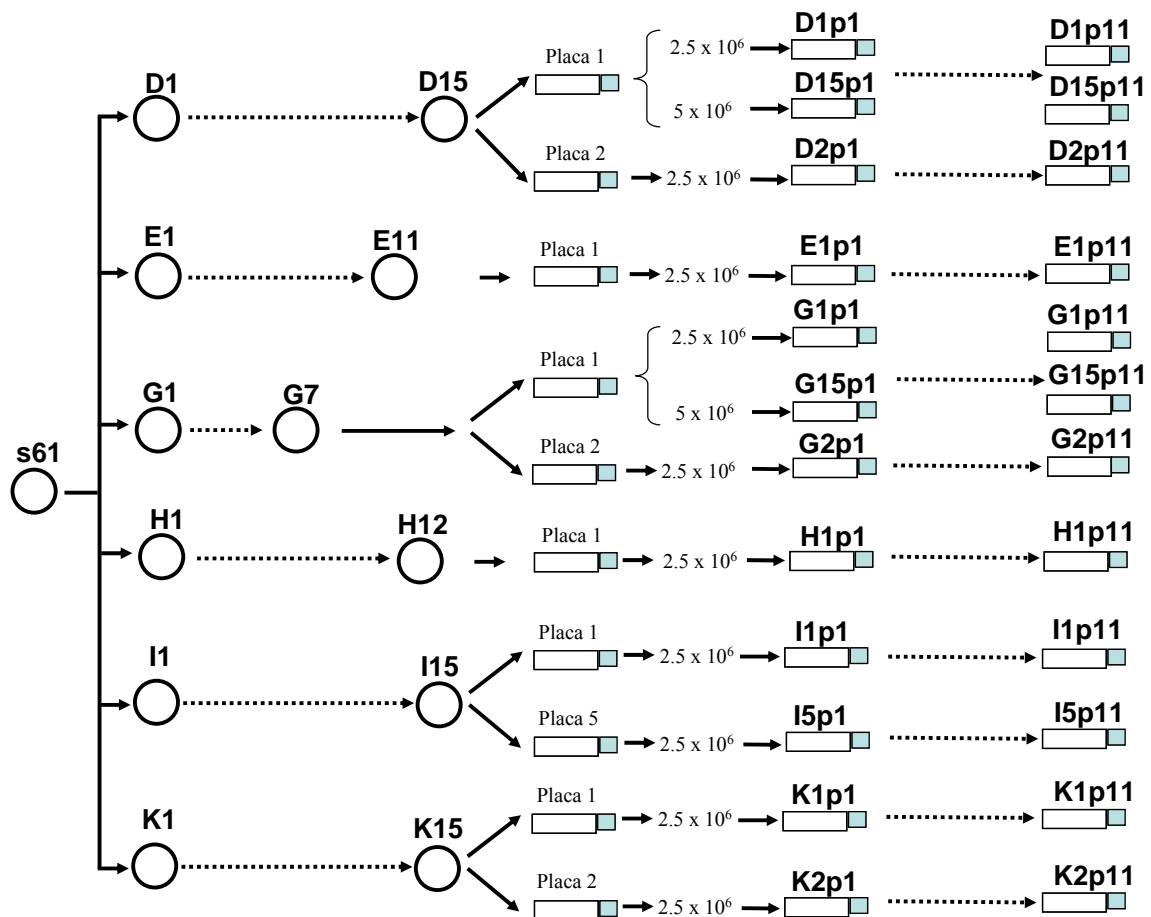
#### 2. 1. 3. 5.- Experimento para la recuperación de la eficacia biológica.

Las semillas de los linajes con sus réplicas se mantuvieron en régimen de cultivo de pases seriados en la línea celular MT-4 a una multiplicidad de infección (m.d.i.) de 0.01. Las infecciones se realizaron en  $2.5 \times 10^6$  células MT-4, a excepción de las infecciones con los clones D1p1 y G1p1, en las que además se infectaron en  $5 \times 10^6$  células (denominados a partir de ahora como D15p1 y G15p11), y de los clones H12 y E11, en las que únicamente se realizó una infección en  $5 \times 10^6$  de células (Figura 2.1). Cuando el efecto citopático era evidente y la viabilidad del cultivo decrecía al 30% se recogía el sobrenadante que se utilizaba para llevar a cabo el siguiente pase. Para ello se preparaba una dilución 1/10 del sobrenadante viral de la infección en 100  $\mu$ l de medio completo y se infectaban con este volumen nuevas células MT-4. De esta manera, se realizaron 11 pases seriados. En el último pase (pase 11), se infectó el doble de la cantidad normal de células ( $5 \times 10^6$  células MT-4) en los clones D1p11, D2p11, G1p11, G2p11, I1p11, I5p11, K1p11 y K2p11 para obtener una cantidad suficiente de cada clon viral para los posteriores análisis. Se recogió el sobrenadante viral por un lado y las células por otro y todo ello se congeló a  $-70^\circ\text{C}$ .

#### 2. 1. 3. 6.- Competiciones virales para el cálculo de la eficacia biológica.

Para determinar la eficacia biológica de los clones, se usó un virus de referencia del laboratorio, el clon viral J1, que tiene una delección de 45 nucleótidos en la región V1-V2 del gen de la envuelta, en comparación al virus de referencia HXB2 (Leitner y cols 2005) entre el nucleótido 6625 al 6670. Como se describirá en el apartado 2. 2. 5, esta delección se usará para detectar y calcular la eficacia biológica de los distintos clones, gracias a la distinta movilidad en geles de acrilamida:bisacrilamida de las variantes virales en cada pase de competición. Para realizar las competiciones se preparó una mezcla 1:1 del virus referencia con cada uno de los clones virales iniciales (pase 1) y finales de los pases seriados (pase 11) a una m.d.i. de 0.001. Con esta mezcla

infectamos  $1 \times 10^5$  células MT-4 en  $100 \mu\text{l}$  de medio completo. Para los clones virales D1p1, D2p1, I1p1, I5p1, y K1p11 se realizó también una competición mezcla en proporción 9 del clon viral frente a 1 del clon de referencia J1. A los 3 días de la infección se añadieron  $100 \mu\text{l}$  de medio completo. La infección se siguió por visualización de sincitios. Normalmente el intervalo entre cada pase de competición fue de 5 ó 6 días dependiendo de lo rápido que avanzara la infección. En general, cuando se producía la disgregación de las células en el cultivo, se procedía a realizar el siguiente pase de competición. Para ello, se preparaba una dilución 1/10 del sobrenadante y se infectaban  $1 \times 10^5$  células MT-4. De esta manera se realizaron cinco pases seriados en competición, realizándose de cada competición una replica. Finalmente, de cada pase de competición se recogió el sobrenadante y las células y todo ello se congeló a  $-70^\circ\text{C}$ .



**Figura 2.1.** Esquema de la obtención de los clones biológicos, y posterior recuperación de eficacia biológica por pases seriados. Los círculos marcan los pases placa a placa realizados anteriormente por Yuste y cols (1999). Las botellas indican los pases seriados de los clones virales del pase 1 al pase 11 realizados para esta Tesis doctoral.

## **2. 2.- MÉTODOS MOLECULARES.**

### 2. 2. 1.- Extracción del ácido desoxirribonucleico (ADN) de los cultivos infectados.

Se partió de  $5 \times 10^6$  células infectadas con el virus y se siguió un método de extracción estándar con fenol:cloroformo (Perucho y cols 1981). Para ello las células se resuspendieron en 1 volumen de solución de lisis A (Tris 10 mM pH 7.5, ácido etilendiaminotetraacético (EDTA) 10 mM, NaCl 0.15 M) a los que se adiciona proteinasa K (200  $\mu$ g/ml) y dodecil sulfato sódico (SDS) al 10%. Tras 1 hora de incubación a 65°C las muestras se dejaron toda una noche a 37°C se añadió 1 volumen de solución de lisis B (Tris 10mM pH 8, EDTA 10 mM; NaCl 0.65 M) y se realizaron tres extracciones con fenol:cloroformo:alcohol isoamílico (25:24:1), seguido de una con cloroformo:alcohol isoamílico (24:1). El ADN contenido en la fase acuosa se precipitó con 2 volúmenes de etanol absoluto, de donde el ADN se recogió y transfirió a un tubo limpio en donde se dejó secando durante 1 hora. Posteriormente el ADN se resuspendió añadiendo agua estéril y se dejó a 37°C durante 1 hora. La lectura de la concentración se realizó en espectrofotómetro a una longitud de onda de 260 nm. El ADN se guardó a -20°C.

### 2. 2. 2.- Extracción de ácido ribonucleico (ARN) de los cultivos infectados.

Para la obtención del ARN viral se empleó el método descrito por Boom y cols (1990), que consiste en la incubación a temperatura ambiente durante 10 minutos de 20  $\mu$ l de sobrenadante de cultivo (con un título medio de virus de  $10^5$  TCID<sub>50</sub>/ml) con una mezcla de 900  $\mu$ l de solución de lisis (120 g de isotiocianato de guanidina (Fluka), 100 ml de Tris-HCl 0.1 M pH 6.4, 22 ml de EDTA 0.2 M pH 8 y 2.6 g de Triton X-100) y 40  $\mu$ l de sílica. Se centrifugó a 14.000 r.p.m. durante 15 segundos y se desechó el sobrenadante. Seguidamente el sedimento obtenido se lavó dos veces con 1 ml de solución de lavado (120 g de isotiocianato de guanidina (Fluka), 100 ml de Tris-HCl 0.1 M pH 6.4), dos veces más con 1 ml de etanol al 70% y una vez con 1 ml de acetona. En cada lavado se centrifugó a 14.000 r.p.m. durante 15 segundos y se desechó el sobrenadante. Tras el último lavado se secó la sílica por incubación a 56 °C durante 10 minutos. El ARN fue eluído de la sílica incubando con 50  $\mu$ l de agua estéril a 56°C

durante 10 minutos. Finalmente se centrifugo a 15.000 r.p.m. durante 2 minutos para separar la sílica del ARN eluído en el agua estéril. El ARN se guardo a -80°C.

### 2. 2. 3.- Amplificación de los genomas virales consenso.

Para la obtención de los genomas virales consenso se partió de cantidades en exceso tanto de ADN como de ARN (al menos 2000 copias del ARN viral) para llevar a cabo las amplificaciones por la reacción en cadena de la polimerasa (PCR). Por ello, entendemos que de las secuencias obtenidas de estas PCRs obtendremos las bases en altas proporciones en la población viral para cada posición nucleotídica dado que partimos de gran cantidad de molde.

#### 2. 2. 3. 1.- A partir de ADN.

Previamente se seleccionaron distintas parejas de cebadores disponibles en el laboratorio para poder realizar la amplificación del genoma viral en 4 fragmentos. Asimismo, se seleccionaron también unos cebadores internos a los anteriores para en caso necesario realizar amplificaciones anidadas.

Estos cebadores se probaron en distintas condiciones de magnesio y gradiente de temperaturas de anillamiento, en un termociclador PTC-200 (MJ Research) con placa de 96 pocillos (0.2 mm) para obtener las mejores condiciones de amplificación. Con estos resultados (Tabla 2.1) se realizaron las amplificaciones de los virus E11 y H12 en los pases finales e iniciales de la recuperación de eficacia biológica (apartado 2.1.3.5).

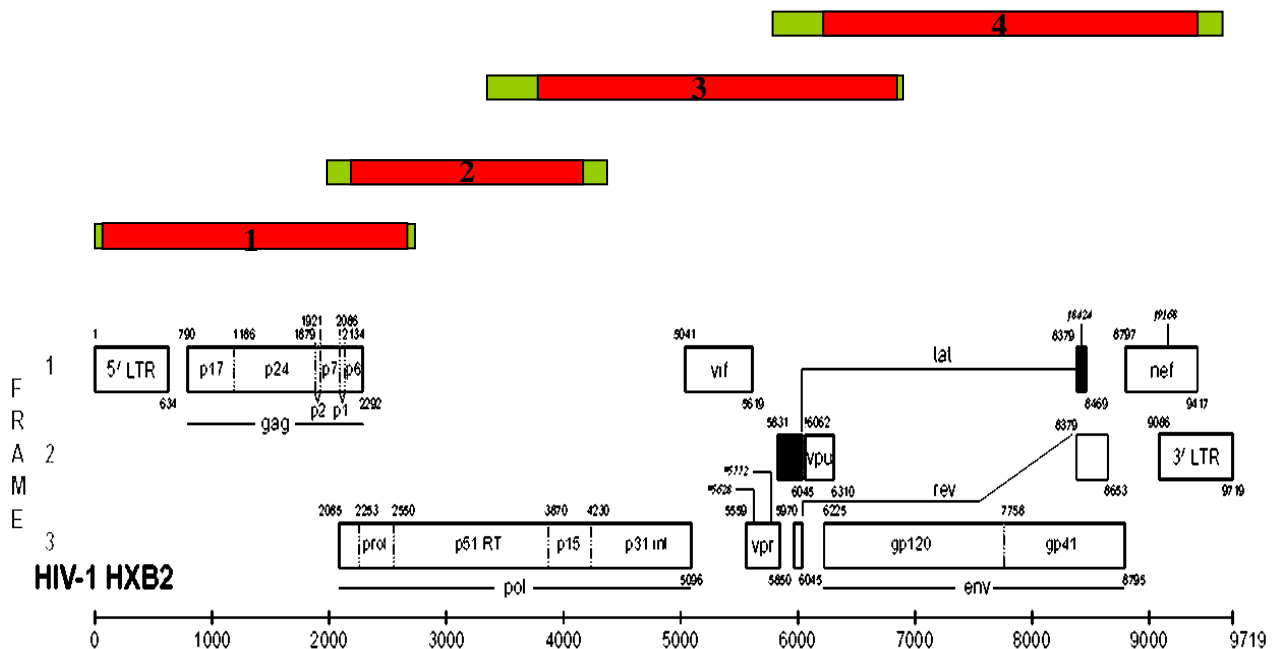
Para la amplificación se usaron los siguientes ciclos de reacción en cadena de la polimerasa (PCR):

- 1) 94°C 5 minutos.
  - 2) 94°C 30 segundos.
  - 3) X°C 30 segundos.
  - 4) 72°C entre 2 y 4 minutos.
  - 5) 72°C 10 minutos.
- } 35 ciclos.

X es el valor de temperatura de anillamiento que mostramos en la Tabla 2.1. El enzima usado fue la Taq Polimerasa (Applied Biosystems, 2.5 unidades (U) de enzima por reacción).

Fragmento	Cebadores	Magnesio(mM)	Temperatura(°C)	Longitud (nt)
1	179 / 147	2	55	2668
Anidado 1	187 / 186R	2	61	2584
2	148 / 53	1.5	55	2372
Anidado 2	PRF1 / 59	1.5	55	1898
3	58 / 52	1.5	55	3580
Anidado 3	385 / 31	1.5	58	3044
4	166 / 216	1.5	61	3742
Anidado 4	99 / 35	1.5	61	3042

**Tabla 2.1.** Descripción de las condiciones usadas para la estrategia de amplificación del genoma completo del VIH-1 a partir de ADN. La descripción detallada de los cebadores esta recogida en la Tabla 2.4.



**Figura 2.2.** Representación de las PCR usadas para secuenciar el genoma completo de los clones virales E11 y H12. Los rectángulos verdes representan las primeras PCR, los rectángulos rojos representan las PCR anidadas. El número del fragmento aparece en cada recuadro. Dibujo del genoma viral del virus HXB2 obtenido de (Leitner y cols 2005).

2. 2. 3. 2.- A partir de ARN.

Los cebadores usados para la amplificación de ARN fueron seleccionados, al igual que en el apartado anterior, de entre los disponibles en el laboratorio. De manera que se pudieron obtener 6 parejas de cebadores para la transcripción inversa (RT) y PCR y 6 parejas de cebadores para realizar las PCRs anidadas, que permitían amplificar el genoma completo viral (Figura 2.3 y Tabla 2.2). Las RT-PCRs se realizaron con el kit comercial OneStep RT-PCR (Qiagen) según las especificaciones del proveedor, y usando el ARN previamente extraído de los clones virales. Se ajustaron los tiempos de elongación a la longitud del fragmento a amplificar y la temperatura de anillamiento fue en todos los casos de 58°C.

<b>Fragmento</b>	<b>Cebadores</b>	<b>Longitud (nt)</b>
1	55 / 109	1743
Anidado 1	55 / 40	1616
2	148 / 53	2372
Anidado 2	PRF1 / 59	2132
3	183' / 25	2410
Anidado 3	385 / 29	1900
4	307 / 31	1240
Anidado 4	294 / 31	1150
5	99 / 197	2327
Anidado 5	34 / 56	2165
6	214 / 216	1244
Anidado 6	214 / 188	1160

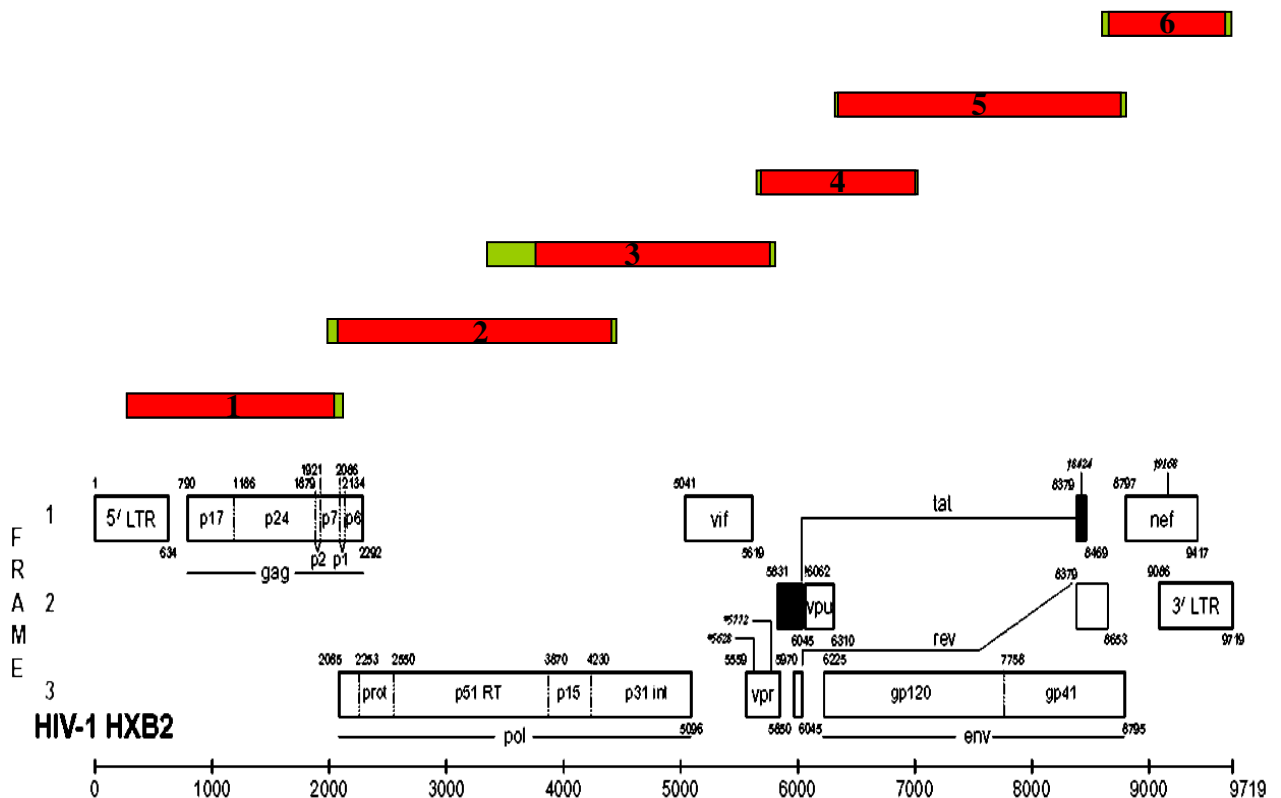
**Tabla 2.2.** Descripción de los cebadores y segmentos amplificados en la estrategia de secuenciación del genoma completo a partir de ARN. La secuencia de los cebadores esta recogida en la Tabla 2.4.

Los ciclos para las RT-PCRs fueron los siguientes:

- 1) 50°C 30 minutos.
  - 2) 94°C 2 minutos.
  - 3) 94°C 30 segundos.
  - 4) 55°C 1 minuto.
  - 5) 72°C entre 1 y 2 minutos 30 segundos.
  - 6) 72°C 10 minutos.
- } 35 ciclos.

Para las PCR anidadas de estas reacciones se uso el enzima Taq polimerasa (Applied Biosystems, 2.5 U/reacción) con 1.5mM de MgCl<sub>2</sub>, y utilizando como molde 1µl de la RT-PCR anterior.

La disposición de los fragmentos en el genoma viral se muestra en la siguiente figura:



**Figura 2.3.** Localización de las distintas PCRs usadas para la amplificación del genoma viral completo de los linajes D, G, I, y K a partir de ARN. Los recuadros verdes representan RT-PCRs, y los recuadros rojos PCRs anidadas. Dentro del recuadro se representa el número de fragmento. Obtenido de (Leitner y cols 2005).

2. 2. 4.- Obtención de las cuasiespecies virales de cada clon biológico en los pases inicial y final.

Para la obtención de las cuasiespecies virales se partió de cantidades en exceso (al menos 2000 copias de ARN viral) del ARN extraído a partir de los sobrenadantes de las infecciones, para así tener una buena representación de la población viral existente. Con este ARN se prepararon varias RT-PCRs (en las mismas condiciones que en el apartado anterior) de todos los clones virales con 3 parejas distintas de cebadores que abarcaban distintas regiones genómicas (ver Tabla 2.3).

<b>Región</b>	<b>Cebadores</b>	<b>Localización en HXB2</b>	<b>Longitud (nt)</b>
Secuencia leader / <i>gag</i>	170 / 174	672 - 1363	692
<i>vpu</i> / <i>env</i> (V1 - V2)	560 / 52	6045-6734	711
<i>env</i> (V3 – V4)	27 / 167	7039 - 7641	600

**Tabla 2.3.** Detalle de los cebadores usados para las RT-PCRS necesarias para obtener la cuasiespecie de las distintas regiones del genoma viral. Los cebadores están descritos en la Tabla 2.4.

Los fragmentos de estas RT-PCRs se clonaron en el kit comercial TA TOPO cloning (Invitrogen), siguiendo las especificaciones facilitadas por el proveedor. Se obtuvieron colonias transformantes que fueron amplificadas directamente con los mismos cebadores usados para la realización de las RT-PCRs para el clonación. Para ello se usó la Taq polimerasa (Applied Biosystems) en las mismas condiciones descritas anteriormente (apartado 2.2.3.2). Seguidamente se resolvieron las amplificaciones de las colonias transformantes en geles de agarosa al 1% y se visualizaban por tinción con bromuro de etidio (BrEt), para detectar aquellas amplificaciones positivas y así poder secuenciarlas (tal como se describe en el apartado 2.2.6).

2. 2. 5.- Ensayo para la determinación de la eficacia biológica de los clones biológicos.

Para estudiar la eficacia biológica de los clones se llevó a cabo el siguiente proceso. Primeramente se extrajo el ARN viral (apartado 2.2.2.) del sobrenadante guardado de cada competición y clon viral. Se realizó una RT-PCR con la pareja de cebadores 99/52 (ver Tabla 2.4) que abarcan la región V1 – V2 del gen de la envuelta



(de la posición nucleotídica 6206 a la 6874 según HXB2). Se cuantificó cada RT-PCR en gel de agarosa 1%, usando el marcador Lambda Hind III (Promega).

Posteriormente, se analizaron estos productos de PCR en un gel de acrilamida:bisacrilamida (30:1) al 8% de 20 cm de largo y 8 mm de espesor, a un voltaje constante de 250V durante 4 horas en tampón tris-ácido bórico-EDTA (TBE) 1X. Una vez finalizada la electroforesis se tiñeron los geles durante 30 minutos en TBE 1x con BrEt a una concentración de 1µl/10ml. La cuantificación de cada variante viral para el cálculo de la eficacia biológica se explicara en el apartado 2. 4. 1. Los geles de realizaron de manera que se resolvían a la vez las dos réplicas de una misma competición y clon viral. En el gel además se añadió una RT-PCR (con las mismas condiciones y cebadores descritos en este apartado) únicamente con el clon viral a estudiar y así como del virus de referencia, J1, de manera que tuviéramos una referencia para detectar la distinta movilidad de las variantes.

#### 2. 2. 6.- Secuenciación.

Los productos de PCR fueron secuenciados empleando el equipo comercial “Big Dye™ Terminator Cycle Sequencing v2.0” (Applied Biosystems, Perkin Elmer, Norwalk, Conn), siguiendo las instrucciones suministradas por el fabricante. Para la secuenciación de los fragmentos amplificados de cada clon viral en la obtención de sus genomas completos, se usaron distintos cebadores tanto sentido como antisentido de los disponibles en el laboratorio. Para la cuasiespecie con los cebadores 170/174 en la región de la secuencia leader/gag se uso para secuenciar el cebador 170; para la cuasiespecie con los cebadores 27/167 en la región *vpu/env* (V1-V2) se uso el cebador 167 y finalmente para la cuasiespecie con los cebadores 560/52 en la región *env* (V3-V4) el cebador 560 (ver Tabla 2.4). La secuenciación se realiza mediante amplificación por PCR asimétrica en presencia de los cuatro dNTPs, marcados cada uno con un colorante fluorescente distinto.

La PCR asimétrica consistió en:

- 1) 94°C 3 minutos.
  - 2) 94°C 30 segundos.
  - 3) 55°C 15 segundos.
  - 4) 60°C 4 minutos.
- } 25 ciclos.

Las reacciones de secuenciación se realizaron en un termociclador PTC-200 (MJ research). Los productos obtenidos fueron analizados en un secuenciador automático ABI PRISM™ modelo 377 (Applied Biosystems, Perkin Elmer, Norwalk, Conn.) por la Unidad de Genómica del ISCIII (Dra A. Ballester).

#### 2. 2. 7.- Inmunoensayo de electroquimiluminiscencia (ECLIA) para la determinación de los niveles de la proteína p24 viral.

Para cuantificar la proteína p24 se analizaron sobrenadantes de las infecciones, tanto del pase inicial como del final de recuperación de eficacia biológica. Se usaron las especificaciones del fabricante (ROCHE). Se partía de 50µl de sobrenadante obtenido de las infecciones virales, los cuales eran incubados con un anticuerpo biotinilado monoclonal específico contra la proteína p24, seguida de una segunda incubación con un anticuerpo frente al anticuerpo específico de p24. El segundo anticuerpo era analizado por la incorporación de micropartículas recubiertas de estreptavidina. Posteriormente, la señal de estreptavidina era detectada por el analizador, que nos proporcionaba una cuantificación de la proteína p24 en nuestra muestra con las unidades en pg/ml.

## **2. 3.- CALCULO DE EFICACIA BIOLOGICA Y ANALISIS DE SECUENCIAS.**

### 2. 3 1.- Cuantificación de los variantes virales en geles de acrilamida y calculo de su eficacia biológica.

La cuantificación se realizó con el programa TINA v2.09a (raytest Isotopenmeßgeräte GmbH). Cada variante viral que se diferenciaba en el gel por movilidad se cuantificaba su densidad óptica (DO). De manera que obteníamos un valor de DO de la banda específica tanto del virus de referencia (J1) como del clon viral. Seguidamente, para los cinco pases de cada competición viral y réplica, se calculaba la proporción del clon viral a estudiar frente al de referencia para cada pase de competición  $t$  ( $R_t$ ). El ratio inicial del virus competido frente al de referencia viene determinado por ( $R_0$ ). Finalmente, el antilogaritmo de la pendiente de regresión,

$$\ln R_t = \ln R_0 + t \ln W$$

nos dará una estimación de la eficacia biológica de nuestro clon viral frente al virus de referencia.

### 2. 3. 2.- Obtención de los genomas completos consenso.

Para la obtención del genoma completo consenso de los virus se usó el programa Seqman del paquete informático DNASTAR v5 (Lasergene), que permite el ensamblaje de las secuencias obtenidos en la secuenciación de cada uno de los fragmentos amplificados, de manera que se obtuvo un alineamiento de todos los fragmentos solapantes, y con ello el genoma completo consenso (9.2kb) de cada clon viral.

### 2. 3. 3.- Alineamientos de secuencias.

Para la generación de los alineamientos, tanto de las distintas secuencias de los genomas completos virales como de los distintos clones moleculares, se usó el programa CLUSTAL X v1.83 (Thompson y *cols* 1997). Los alineamientos fueron editados manualmente con el programa BIOEDIT sequence alignment editor v5.0.9 (Hall 1999; Tippmann 2004).

2. 3. 4.- Heterogeneidad, sustituciones sinónimas, sustituciones no sinónimas y generación de árboles filogenéticos.

Con los alineamientos de todos los clones moleculares de cada región analizada (ver apartados anteriores), se calcularon la heterogeneidad, y las sustituciones sinónimas y no sinónimas. La heterogeneidad se calculó como el promedio de diferencias entre comparaciones clon a clon dentro de cada población viral. Para el cálculo del número de diferencias promedio sinónimas (Sd) y no sinónimas (Nd) entre los distintos clones dentro de cada población viral, se usó el modelo modificado de Nei-Gojobori (Nei y cols 1986). De la misma manera calculamos el número de sustituciones no sinónimas por sitio (dN) y sustituciones sinónimas por sitio (dS).

Con el alineamiento del genoma completo de cada uno de los clones virales se obtuvo el árbol filogenético entre los linajes virales. Primero se determinó el mejor modelo evolutivo por medio del programa Modeltest 3.8 (Posada 2006). Según éste, el mejor modelo era el de Kimura dos parámetros con una proporción de transiciones con respecto a transversiones de 2, y con una distribución gamma de parámetro 0.7. El resto de análisis y cálculos se realizaron con el programa MEGA v3.1 (Kumar y cols 2004).

2. 3. 5.- Obtención de las dS y dN de cada clon viral y pase por medio de una ventana deslizante a lo largo del genoma.

Con los mismos alineamientos de las distintas regiones del genoma viral usados para los cálculos del apartado anterior, se realizó el cálculo de las dS y dN acumulada para cada posición aminoacídica por medio una de ventana deslizante según el programa CRANN (Creevey y cols 2002). El método utilizado está recogido en (Li 1993). El tamaño de ventana deslizante fue el 10% de la longitud de la región en aminoácidos, y el número de posiciones que se desplazaba cada vez la ventana la determinamos en el 50% del tamaño de la ventana deslizante (ejemplo: longitud secuencia 175 aminoácidos, con lo que la longitud de la ventana deslizante será de 17 aminoácidos, y el desplazamiento de la ventana será de 8 aminoácidos cada vez). De esta manera, se obtuvo una serie de valores acumulativos de dS y dN para cada desplazamiento de la ventana.

2. 3. 6.- Análisis estadístico.

Para los análisis estadísticos se usó el programa STATGRAPHICS PLUS v5 (Statistical Graphics Corp) así como también el programa SPSS v12 (SPSS inc). El programa GrapPad v4 (GraphPad software, Inc) se uso para realizar algunas de las representaciones graficas.

<b>Cebador</b>	<b>Secuencia (5' → 3')</b>	<b>Orientación</b>	<b>HXB2</b>
PRF1	CAGCCCCACCAGAAGGC	+	2156-2172
25	GCCAGAATTGGGTGTCGACATAAGCAGAA	-	5773-5800
27	ATAAGCTTGCAGTCTAGCAGAAGAAGA	+	7004-7030
29	GCGGTCGACATAGATATGTTGGGGTAAGCCAT	-	5677-5706
31	GCGCTGCAGGCCTGTGTAATGACTGAGGT	-	6816-6846
34	GTCACAGTCTATTATGGGGTACCTGTGT	+	6330-6357
35	AGCTAGTACCAGTTGAGCCAGAATTCCGC	-	142-172
40	GCGGAATTCCTAAAAAATTAGCCTGTCT	-	2074-2103
52	TAATGTATGGGAATTGGCTCAA	-	6853-6874
53	GGCGAATTCCTAGCCATTGCTCTCCA	-	4284-4310
55	CTGAGCCTGGGAGCTCTCTGGC	+	481-502
56	CCGCAGATCGTCCCAGATAAG	-	8487-8507
58	GCCAGAAAAAGACAGCTGGACTGT	+	3287-3310
59	ATGATTCCTAATGCATATTGTGAGT	-	4042-4066
99	AGAGCAGAAGACAGTGGC	+	6206-6223
109	CCCCAGACCTGAAGCTCTCTT	-	2168-2188
147	TCTTCTGTCAATGGCCATTGTTTAAC	-	2610-2635
148	TTCAATTGTGGCAAAGAAGGGCAC	+	1966-1989
166	TGGAGCCAGTAGATCCTAGACTAGAGCCCT	+	5832-5861
167	TTCTCCAATTGTCCCTCATATCTCCTCCTCCA	-	7634-7665
170	TCTCTAGCAGTGGCGCCCGAACAGGGAC	+	626-653
174	TCTATCCCATTCTGCAGCTTCCTCATTGAT	-	1402-1431
179	TGGAAGGGCTAATTCCTCC	+	1-20
183'	TGTAAACTCCTTAGGGGAAC	+	3387-3406
186R	TTGGGCCATCCATTCCCTGGC	-	2588-2607
187	CACACACAAGGCTACTTCCCT	+	57-77
188	GCCACTCCCCTGTCCCGCCC	-	389-408
197	GGTGGTAGCTGAAGAGGCACA	-	8512-8532
214	AATAGAGTTAGGCAGGGATACTCACCATT	+	8340-8368
216	GAGCTCCCAGGCTCAGATCTGGTCT	-	468-492
294	ATGGCTTAGGGCAACATATCTATG	+	5677-5700
307	AGGGCCACAAAGGGAACCATAC	+	5582-5603
385	TGGGAGTTTGTTAATACCCT	+	3789-3809
560	CTATGGCAGGAAGAAGCCGAGA	+	5968-5989

**Tabla 2.4.** Descripción de los distintos cebadores usados para las distintas amplificaciones realizadas. La orientación de los cebadores, + es sentido, y – es antisentido. La localización en el genoma de cada cebador según el virus de referencia HXB2.

### **3.- RESULTADOS**

### **3. 1.- RECUPERACIÓN DE LA EFICACIA BIOLÓGICA DE LOS CLONES VIRALES.**

#### 3. 1. 1.- Análisis del título y el valor por ECLIA de la proteína viral p24 de los clones virales en el pase inicial y final de la recuperación de eficacia biológica.

Dos variables importantes que nos pueden ayudar a entender como el virus está replicando en un ambiente determinado, son el título viral, calculado como TCID<sub>50</sub>/ml (número de dosis infectivas necesarias para infectar el 50% del cultivo por mililitro), y los niveles de la proteína viral p24 medida por ECLIA en pg/ml (ver apartado 2. 2. 7). En la Tabla 3.1 están representados el valor de título viral y los niveles de la proteína p24. Observamos un aumento estadísticamente significativo entre el título viral inicial y final de los pases de recuperación de eficacia biológica (Wilcoxon W test,  $p=0.009$ ), aunque los clones D15p11, G15p11 y K2p11 tuvieron descensos en sus títulos virales con respecto a sus propios clones iniciales. Este aumento implicaría que los clones biológicos después de los pases seriados están produciendo más partículas virales capaces de generar una infección productiva. También cuantificamos la proteína viral p24. Las diferencias entre los valores de p24 de los pases inicial y final de recuperación de eficacia no fueron significativas (t-test  $p=0.96$ ). La falta de diferencias entre el valor de la proteína viral p24 entre el pase inicial y final, indicaría que la cantidad de virus que se está produciendo en ambos casos es similar.

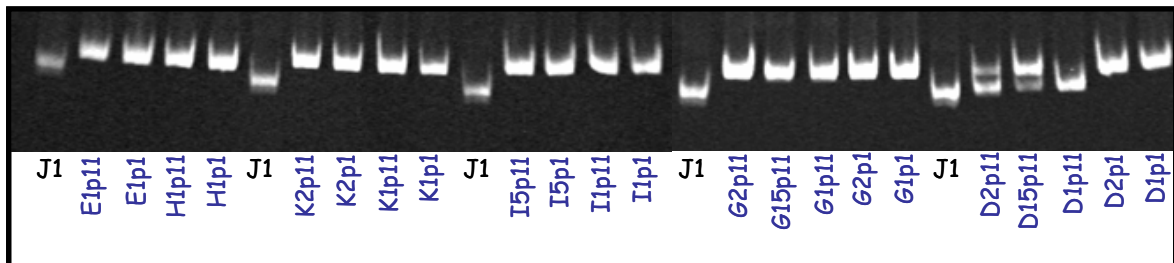


Clon	Título (TCID <sub>50</sub> / ml)	Incremento del título (%)	p24 (10 <sup>5</sup> pg/ml)	Incremento de p24 (%)
D1p1	4,2×10 <sup>5</sup>		174,4	
D1p11	7,5×10 <sup>5</sup>	+78	209,4	+16
D15p11	1,5×10 <sup>5</sup>	-64	89,6	-39
D2p1	3,7×10 <sup>5</sup>		162,2	
D2p11	5,6×10 <sup>5</sup>	+51	176,4	+7
E1p1	3,7×10 <sup>4</sup>		566,4	
E1p11	1,2×10 <sup>6</sup>	+3143	677,6	+18
G1p1	1,2×10 <sup>5</sup>		242,2	
G1p11	1,8×10 <sup>5</sup>	+50	227,2	-5
G15p11	1×10 <sup>5</sup>	-16	270,6	+10
G2p1	7,8×10 <sup>4</sup>		179,2	
G2p1	4,2×10 <sup>5</sup>	+438	427	+110
H1p1	1,3×10 <sup>4</sup>		372,2	
H1p11	3,7×10 <sup>5</sup>	+2746	563,6	+46
I1p1	4×10 <sup>4</sup>		166,8	
I1p11	6,5×10 <sup>5</sup>	+1525	234,8	+32
I5p1	2,5×10 <sup>4</sup>		145,2	
I5p11	9,5×10 <sup>4</sup>	+280	177,2	+17
K1p1	7,5×10 <sup>4</sup>		478,8	
K1p11	1,3×10 <sup>5</sup>	+73	265,8	-41
K2p1	4,9×10 <sup>4</sup>		414,6	
K2p11	4,4×10 <sup>4</sup>	-10	197	-47
Mediana por pases				
Pases 1	6,2×10 <sup>4</sup>		210	
Pases 11	2,7×10 <sup>5</sup>	+335	231	+10

**Tabla 3.1.** Valores de título viral y cuantificación de proteína viral p24 para cada uno de los clones y pase estudiados.

3. 1. 2.- Diferenciación por movilidad electroforética en geles de poliacrilamida de los clones virales iniciales y finales de los pases de recuperación de eficacia biológica.

Para determinar la eficacia biológica primero debíamos comprobar si era posible diferenciar los diferentes clones biológicos en competiciones frente al virus de referencia J1. Para ello, se realizaron geles de acrilamida:bisacrilamida al 8% (ver apartado 2. 2. 5 de Materiales y Métodos), y se determinó la distinta movilidad electroforética de cada clon viral con respecto al de referencia, J1, en la región de la envuelta que comprende los dominios V1 – V2.



**Figura 3.2.** Comparación de la movilidad electroforética de cada clon viral frente al J1.

Como se puede ver en la Figura 3.2, todos los clones virales se podían diferenciar por su distinta movilidad con el clon de referencia J1. En el gel se observa que los clones biológicos D15p11 y D2p11 contienen dos poblaciones virales distintas, una delecionada (banda inferior) y otra no delecionada (banda superior). En el caso del clon viral D1p11, éste sólo tiene la población delecionada. Estas poblaciones delecionadas tienen una movilidad electroforética cercana a la del clon viral J1 en estas condiciones, lo cual podría dificultar la cuantificación en las competiciones de estos clones.

3. 1. 3.- Cálculo de eficacia biológica en los clones virales iniciales y finales.

La eficacia biológica de los clones se determinó por cultivos seriados en competición a proporción 1:1 durante 5 pases con el clon de referencia, J1 (apartado 2.1.3.6). Para el cálculo de la eficacia biológica, se resolvieron las RT-PCRs de los sobrenadantes de cada pase de las competiciones virales en geles de acrilamida:bisacrilamida, tal como se describe en el apartado 2. 4. 1. de Materiales y Métodos. Posteriormente se obtuvieron las densitometrías de las bandas correspondientes de los distintos clones virales en relación al virus J1.

3. 1. 3. 1.- Clones virales del linaje D.

Del linaje D analizamos 2 clones iniciales, D1p1 y D2p1. Estos clones iniciales fueron sometidos a pases seriados (descrito anteriormente). En el caso del D1p1 se realizaron los pases seriados tanto en  $2,5 \times 10^6$  células MT4, dando lugar al clon final D1p11, como en  $5 \times 10^6$  células MT4, dando lugar al clon final D15p11. El clon D2p1 solo fue pasado en  $2,5 \times 10^6$  células MT4, dando lugar al clon final D2p11.

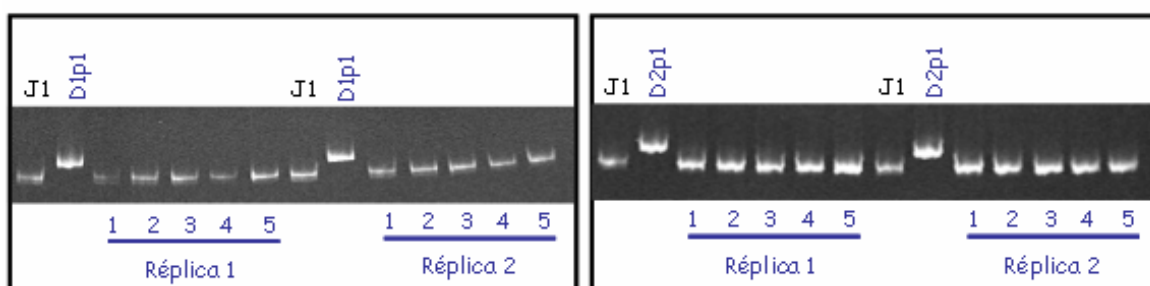
Para determinar la eficacia biológica relativa de cada uno de estos clones biológicos (2 iniciales y 3 finales), se realizaron 5 pases de competición de estos clones frente al clon de referencia J1 (Figura 3.3) en una proporción 1:1. Como se puede observar en la figura 3.3, el clon viral D1p1 está desplazado frente al clon viral J1 en ambas réplicas de la competición en proporción 1:1 realizadas, y no es detectable en ninguno de los 5 pases de la competición. Por el contrario, en los geles de los clones finales (D1p11 y D15p11) vemos como el clon D1p11 ha desplazado al clon J1 desde el primer pase (no se observa al clon J1 en la competición). En el caso del clon D15p11, se observa que es capaz de desplazar al clon J1 desde el pase 1 de competición. Además se observa cómo la población delecionada del D15p11 (la que tiene una movilidad electroforética mayor) desplaza a la población no delecionada. Por lo tanto se ha producido un aumento de la eficacia biológica muy significativo pero no cuantificable entre el pase 1 y pase 11 en el clon D1.

En el caso del clon viral D2p1, observamos como no es capaz tampoco de competir frente al J1 antes de los pases. Mientras que el clon viral D2p11 se observa, como, al igual que en el clon viral D15p11, desplaza al clon J1 desde el primer pase de competición. Se observa también que este clon biológico está formado por dos

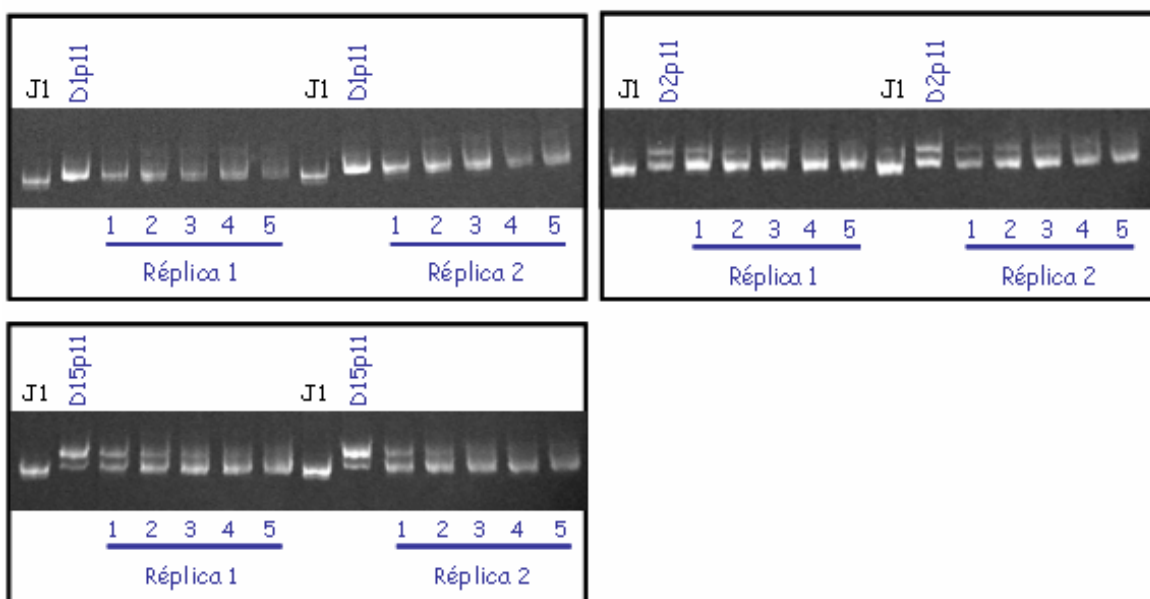
poblaciones, una delecionada (mayor movilidad electroforética) y otra no (menor movilidad electroforética). Tras 5 pases de competición las variantes delecionadas son las que desplazan finalmente a las variantes no delecionadas. En resumen, en este clon se habría producido un aumento de la eficacia biológica con respecto al clon inicial.

## Competiciones 1:1

### Pase 1



### Pase 11



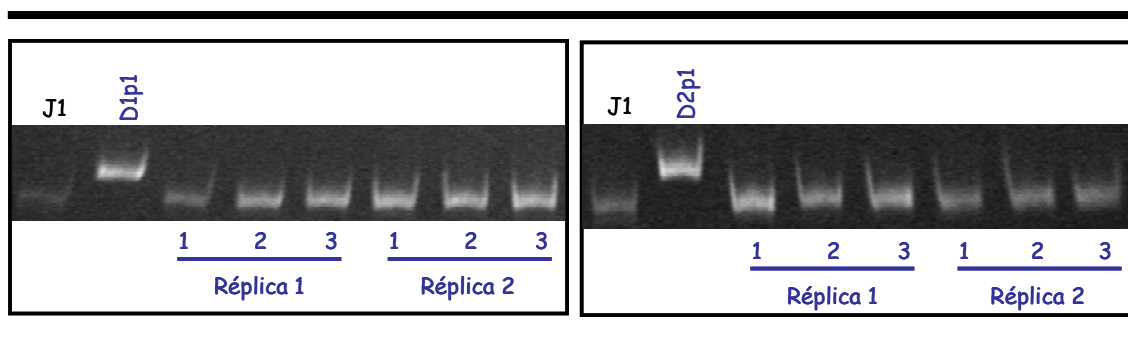
**Figura 3.3.** Geles de acrilamida de las competición 1:1 de los clones biológicos D1p1, D2p1, D1p11, D15p11 y D2p11.

Al no poder detectar los clones iniciales (D1p1 y D2p1) en las competiciones a proporción 1:1, realizamos una nueva competición frente al clon J1 aumentando la proporción 9 veces a favor de los clones iniciales, es decir a una proporción 9:1 (ver

Figura 3.4). En estas competencias, obtuvimos un resultado similar al de las competencias 1:1. Observamos como el clon J1 se impone igualmente a los clones virales. Este resultado indica que los clones iniciales de recuperación tienen una eficacia biológica muy disminuida con respecto a los clones finales.

## Competiciones 9:1

### Pase 1



**Figura 3.4.** Geles de acrilamida de las competición 9:1 de los clones biológicos D1p1 y D2p1.

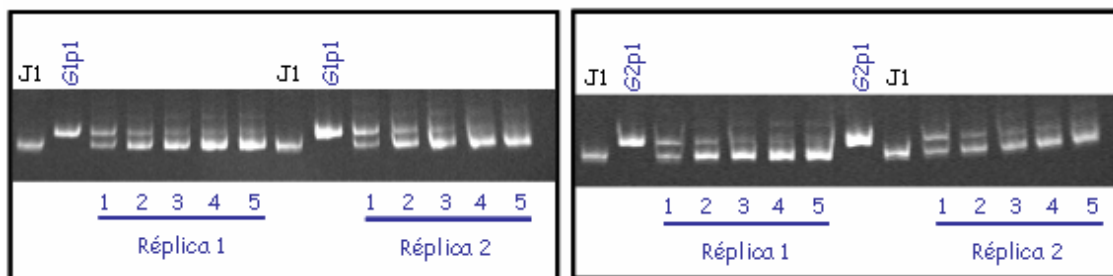
#### 3. 1. 3. 2.- Clones virales del linaje G.

Disponíamos de 2 clones iniciales del linaje G, G1p1 y G2p1. Con el virus G1p1 se realizaron los pases seriados tanto en  $2,5 \times 10^6$  de células MT4, dando lugar al clon final G1p11, como en  $5 \times 10^6$  células MT4, dando lugar al clon final G15p11. El clon G2p1 fue pasado en  $2,5 \times 10^6$  células MT4, dando lugar al clon final G2p11.

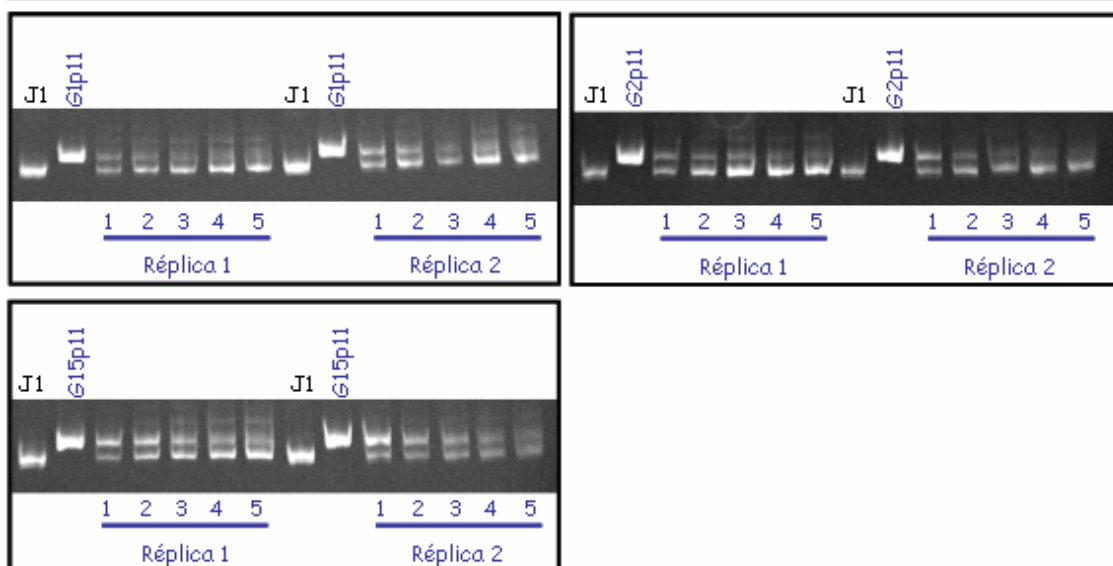
En la competición de los clones iniciales de recuperación de eficacia biológica, los clones G1p1 y G2p1, son detectables hasta el pase 4. En las competencias de los clones finales de recuperación de eficacia biológica, estos son detectables hasta el pase 4 ó 5 de competición (ver Figura 3.5). Esto indica un aumento en la eficacia biológica después de los 11 pases seriados.

## Competiciones 1:1

### Pase 1



### Pase 11



**Figura 3.5.** Geles de acrilamida de las competición 1:1 de los clones biológicos G1p1, G2p1, G1p11, G15p11 y G2p11.

#### 3. 1. 3. 3.- Clones virales del linaje I.

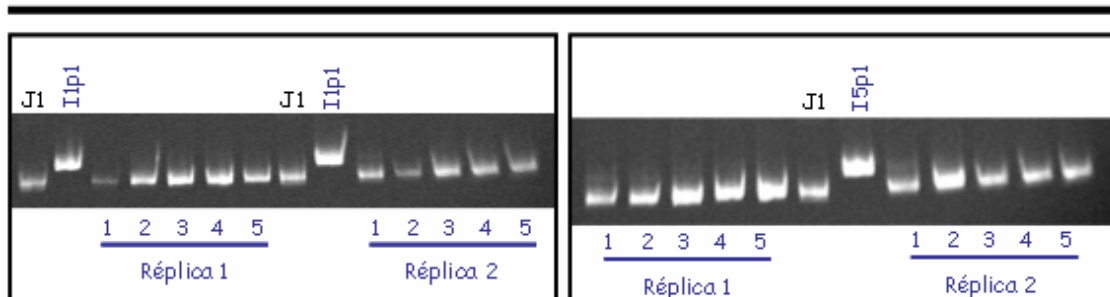
Del linaje I, partíamos de dos clones iniciales I1p1 e I5p1. De ambos se realizaron pases seriados en  $2,5 \times 10^5$  células MT4 obteniendo los clones finales I1p11 e I5p11. Se realizaron competiciones 1:1 tanto de los clones iniciales como finales. De los clones iniciales, I1p1 e I5p1 realizamos también una competición en proporción 9:1 a favor de los clones iniciales, I1p1 e I5p1.

En la competición 1:1 (ver Figura 3.6), los clones iniciales de recuperación de eficacia biológica, los clones I1p1 e I5p1 no son detectables en ninguno de los pases de

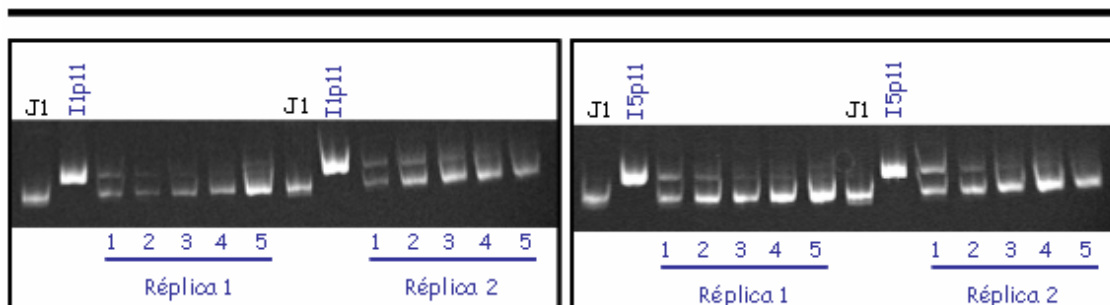
la competición (en ambas réplicas), indicando que habían sido completamente desplazados por el clon J1. Sin embargo, en los clones finales vemos como los clones I1p11 e I5p11, son detectables hasta el pase 4 de competición, indicando que en ambos clones se ha producido un aumento de la eficacia biológica con respecto a los iniciales.

## Competiciones 1:1

### Pase 1



### Pase 11



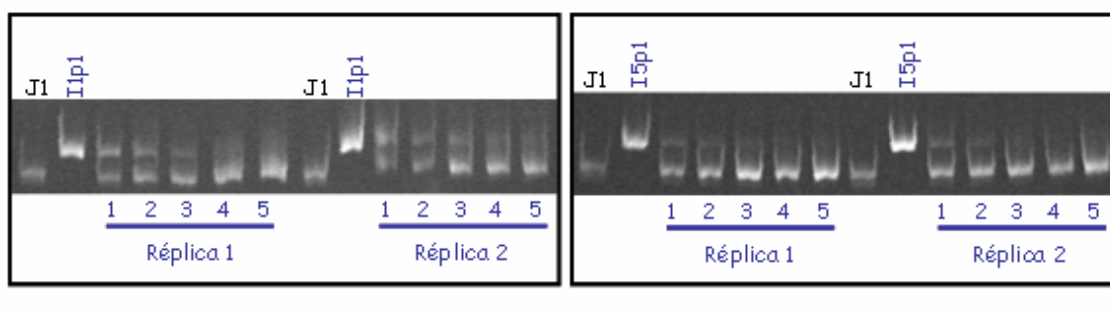
**Figura 3.6.** Geles de acrilamida de las competición 1:1 de los clones biológicos I1p1, I5p1, I1p11 e I5p11.

Dado que no se detectaron los clones iniciales en la competición 1:1, se realizó una competición de estos a proporción 9:1. Observamos de esta manera que los clones iniciales eran detectables hasta el pase 3 de competición (ver Figura 3.7), y por tanto podemos realizar el cálculo de su eficacia biológica relativa al virus J1.

Podemos observar por tanto que tanto el clon I1p1 como el I5p1 se comportan de la misma manera en las competiciones. No parece que ninguno de los dos clones sea más eficaz que el otro.

## Competiciones 9:1

### Pase 1



**Figura 3.7.** Geles de acrilamida de las competición 9:1 de los clones biológicos I1p1 e I5p1.

#### 3. 1. 3. 4.- Clones virales del linaje K.

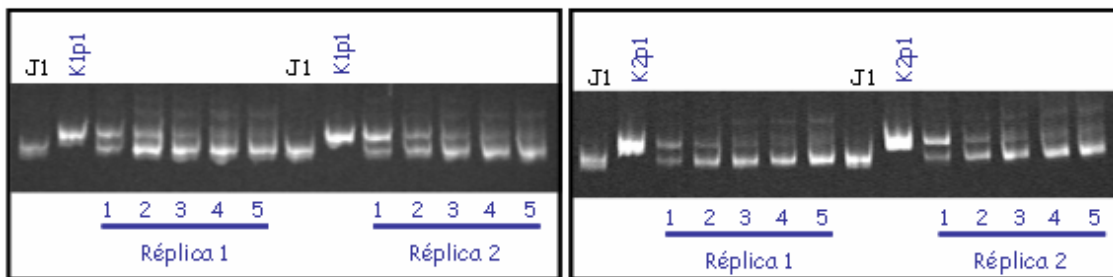
Del linaje K teníamos dos clones iniciales, K1p1 y K2p1. Ambos fueron mantenidos en  $2,5 \times 10^6$  células MT4, dando lugar a los clones finales K1p11 y K2p11. Todos estos clones fueron puestos a competir frente al virus J1 en una proporción 1:1. Además el clon K1p11 fue competido también a una proporción 9:1.

En las competiciones 1:1 de los clones iniciales observamos que el clon K1p1 es detectable hasta el pase 4 de la competición al igual que el clon K2p1. Mientras que en los clones finales, el clon K2p11 es detectable en competición frente al J1 hasta el pase 5 de competición a proporción 1:1 (ver Figura 3.8). Sin embargo, la competición con el clon final K1p11 fue defectuosa, y solamente pudimos analizar la competición 9:1 (ver Figura 3.9), en donde observamos como el clon K1p11 se mantiene en la competición hasta el pase 4. Por ello, se está produciendo un aumento de la eficacia biológica entre pases para el linaje K.

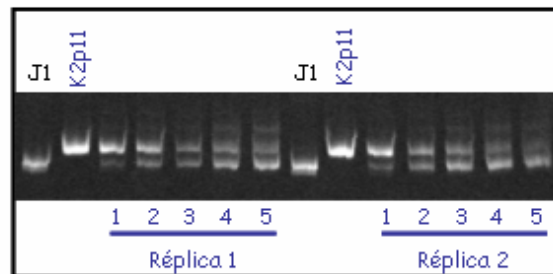


## Competiciones 1:1

### Pase 1



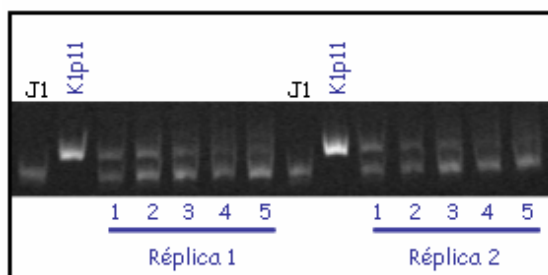
### Pase 11



**Figura 3.8.** Geles de acrilamida de las competición 1:1 de los clones biológicos K1p1, K2p1, K1p11 y K2p11.

## Competiciones 9:1

### Pase 11



**Figura 3.9.** Gel de acrilamida de la competición 9:1 del clon biológico K1p11.

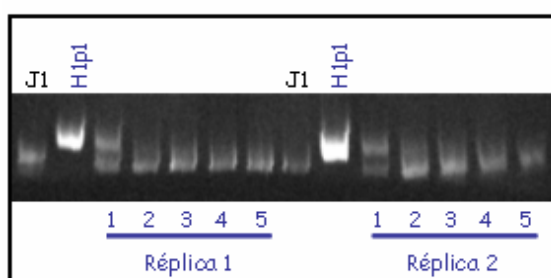
## 3. 1. 3. 5.- Clones virales del linaje H.

Del linaje H sólo se disponía de una placa y por tanto solo obtuvimos el clon inicial denominado H1p1, que fue pasado seriadamente en  $5 \times 10^6$  de células MT4 dando lugar al clon final H1p11. Se realizó una competición a proporción 1:1, pero también tuvimos que realizar la competición del H1p1 a proporción 9:1.

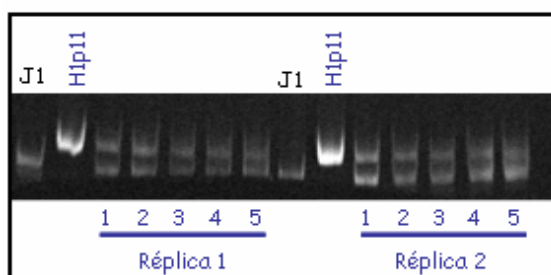
El clon inicial H1p1 es detectable en el pase 1 de la competición, y es desplazado en el pase 2 de competición, no pudiendo por tanto realizar su cálculo de eficacia biológica. Mientras que el clon final H1p11 es detectable hasta el pase 5 de competición frente al clon J1. Esto nos indica un aumento importante de eficacia entre el pase 1 y pase 11 (ver Figura 3.10).

## Competiciones 1:1

### Pase 1



### Pase 11

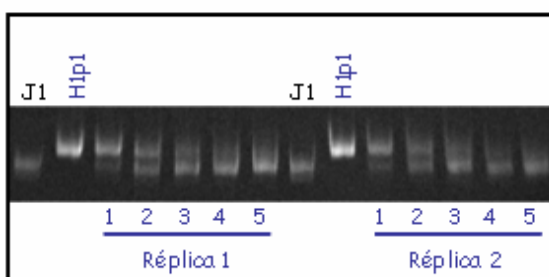


**Figura 3.10.** Geles de acrilamida de las competición 1:1 de los clones biológicos H1p1 e H1p11.

Para tratar de cuantificar la eficacia del clon inicial H1p1, realizamos la competición de éste frente al clon J1 a proporción 9:1 (ver Figura 3.11). Se puede observar que el clon inicial H1p1 es detectable hasta el tercer pase de competición, pudiéndose por tanto realizar el cálculo de su eficacia biológica.

## Competiciones 9:1

### Pase 1



**Figura 3.11.** Gel de acrilamida de la competición 9:1 del clon biológico H1p1.

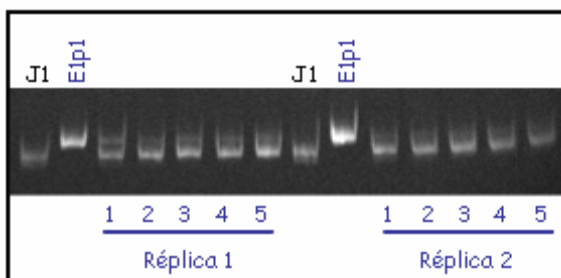
### 3. 1. 3. 6.- Clones virales de linaje E.

Del linaje E, sólo disponía de un clon inicial para la recuperación de eficacia biológica, el clon E1p1. Realizamos los pases seriados de este clon en  $5 \times 10^6$  de células MT4 dando lugar al clon final E1p11. Con estos clones realizamos las competiciones frente al virus J1 en proporción 1:1.

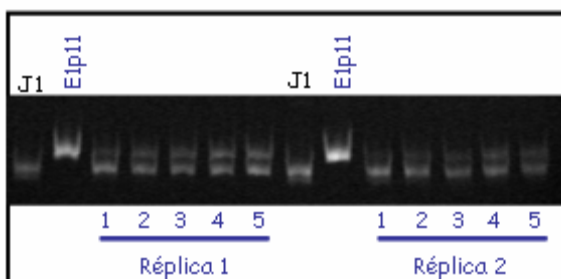
El clon inicial E1p1 es detectable frente al virus J1 hasta el pase 4 de competición. El clon final E1p11 es detectable en la competición frente al clon J1 hasta el pase 5 (ver Figura 3.12). Por tanto en ambos casos se puede determinar la eficacia biológica relativa frente al clon J1, y se comprueba que el clon E ha ganado eficacia biológica tras los pases seriados.

## Competiciones 1:1

### Pase 1



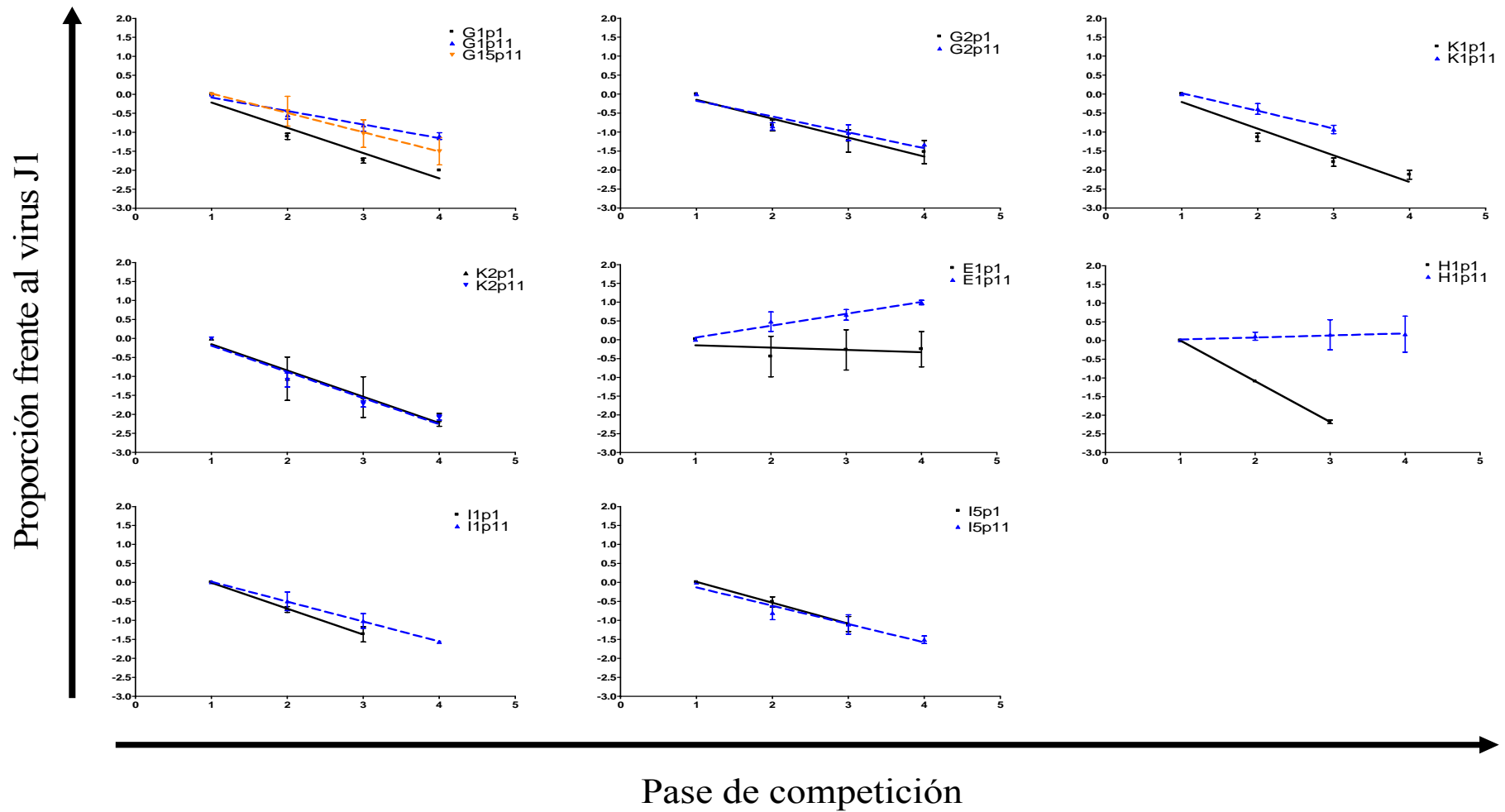
### Pase 11



**Figura 3.12.** Geles de acrilamida de las competiciones 1:1 de los clones biológicos E1p1 y E1p11.

#### 3. 1. 3. 7.- Análisis y cuantificación de las eficacia biológicas iniciales y finales.

Para la cuantificación de la presencia de cada clon en las competiciones, se llevó a cabo una estimación de la densidad óptica de cada banda obtenida en los geles de acrilamida. Se realizaron las regresiones lineales según se describe en el apartado 2. 3. 1 de Materiales y Métodos. En el eje de abscisas se representa el pase de competición y en el eje de ordenadas el valor relativo del clon frente al virus J1. La pendiente de la regresión lineal (ver Figura 3.13) es la eficacia de cada clon viral. Para determinar si las pendientes eran distintas entre los clones biológicos iniciales con respecto a sus clones finales, realizamos una prueba t-student comparando según cada linaje las pendientes de los clones iniciales frente a las de los clones finales respectivamente. Los datos de eficacia biológica de cada clon y pase están recogidos en la Tabla 3.2.



**Figura 3.13.** Representación gráfica de regresiones lineales de los pases iniciales (negro) y los pases finales (azul y naranja, con línea punteada) de recuperación de eficacia biológica. Se indica la media de las dos competencias con su desviación estándar.

## 3. 1. 3. 8.- Eficacias biológicas de los clones biológicos relativas al clon J1.

El valor de eficacia biológica relativo al virus J1 obtenido para cada clon y el incremento obtenido con respecto al clon inicial tras los pases seriados de recuperación se recogió en la Tabla 3.2.

Clon	Competición	Eficacia	Incremento de eficacia (%)	p-valor
D1p1	1:1 / 9:1	— <sup>a</sup>		
D1p11	1:1 / 9:1	— <sup>a</sup>	<b>Si<sup>b</sup></b>	—
D15p11	1:1 / 9:1	— <sup>a</sup>	<b>Si<sup>b</sup></b>	—
D2p1	1:1 / 9:1	— <sup>a</sup>		
D2p11	1:1 / 9:1	— <sup>a</sup>	<b>Si<sup>b</sup></b>	—
E1p1	1:1	0.94		
E1p11	1:1	1.38	<b>47%</b>	<b>0.01</b>
G1p1	1:1	0.52		
G1p11	1:1	0.70	<b>35%</b>	<b>0.004</b>
G15p11	1:1	0.60	15%	0.2
G2p1	1:1	0.61		
G2p11	1:1	0.66	8%	0.43
H1p1	9:1	0.33		
H1p11	1:1	1.05	<b>219%</b>	<b>&lt;0.0001</b>
I1p1	9:1	0.51		
I1p11	1:1	0.59	16%	0.07
I5p1	9:1	0.58		
I5p11	1:1	0.62	7%	0.5
K1p1	1:1	0.5		
K1p11	9:1	0.63	26%	0.07
K2p1	1:1	0.5		
K2p11	1:1	0.5	0%	0.99

**Tabla 3.2.** Eficacias biológicas de cada clon viral en el pase inicial y final de recuperación de eficacia biológica. Todas las eficacias se calcularon relativas al virus de referencia J1. <sup>a</sup> No se pudo determinar la eficacia biológica de estos clones. <sup>b</sup> Aunque no se pudo calcular la eficacia biológica se puede observar de manera cualitativa que estos clones biológicos aumentaron su eficacia biológica entre el pase inicial y

final de recuperación. En negrita se marcan aquellos aumentos de eficacia que resultaron significativos en la comparación de las pendientes.

Del linaje D, en los clones finales teníamos dos poblaciones virales (deleccionadas y no deleccionadas), de manera que no pudimos determinar la eficacia. En los clones iniciales al hacer la competición en la proporción 1:1 no fuimos capaces de detectar estos virus en la competición, y tampoco fuimos capaces de detectarlos si repetíamos la competición desbalanceando la proporción a 9:1; de manera que tampoco pudimos determinar la eficacia biológica en ningún caso. Esto nos indicaba que los clones iniciales D1p1 y D2p1 debían tener una eficacia biológica muy reducida. Pero sin embargo, después de los 11 pases en gran tamaño poblacional, pudimos observar (ver apartado 3.1.3.1) como los clones finales D1p11, D15p1 y D2p11, aunque tenían una población viral mezcla de virus deleccionados y no deleccionados, era la población deleccionada la que desplazaba tanto al clon J1 como a la población no deleccionada. Por lo tanto de una manera cualitativa, los clones D aumentan significativamente su eficacia biológica después de 11 pases en gran tamaño poblacional.

Para determinar si se había producido un aumento de la eficacia biológica después de los pases seriados, realizamos una prueba t-student emparejada entre todos los clones iniciales y todos los clones finales. El resultado de esta prueba dio estadísticamente significativo con un  $p$ -valor=0.042, por tanto en general se estaba produciendo un aumento de eficacia biológica entre pases. Asimismo, realizamos un test de los signos, que nos permite calcular el aumento de eficacia incluyendo al linaje D del que solamente teníamos un valor cualitativo sobre su aumento de eficacia. Se observó que 11 de 12 clones habían aumentado eficacia biológica y el restante, el clon K2p11, no había aumentado su eficacia. Comprobamos que existía un aumento de eficacia estadísticamente significativo con un  $p$ =0.001. Por tanto los pases seriados en grandes tamaños poblacionales promovían el aumento de eficacia de clones con eficacia biológica disminuida.

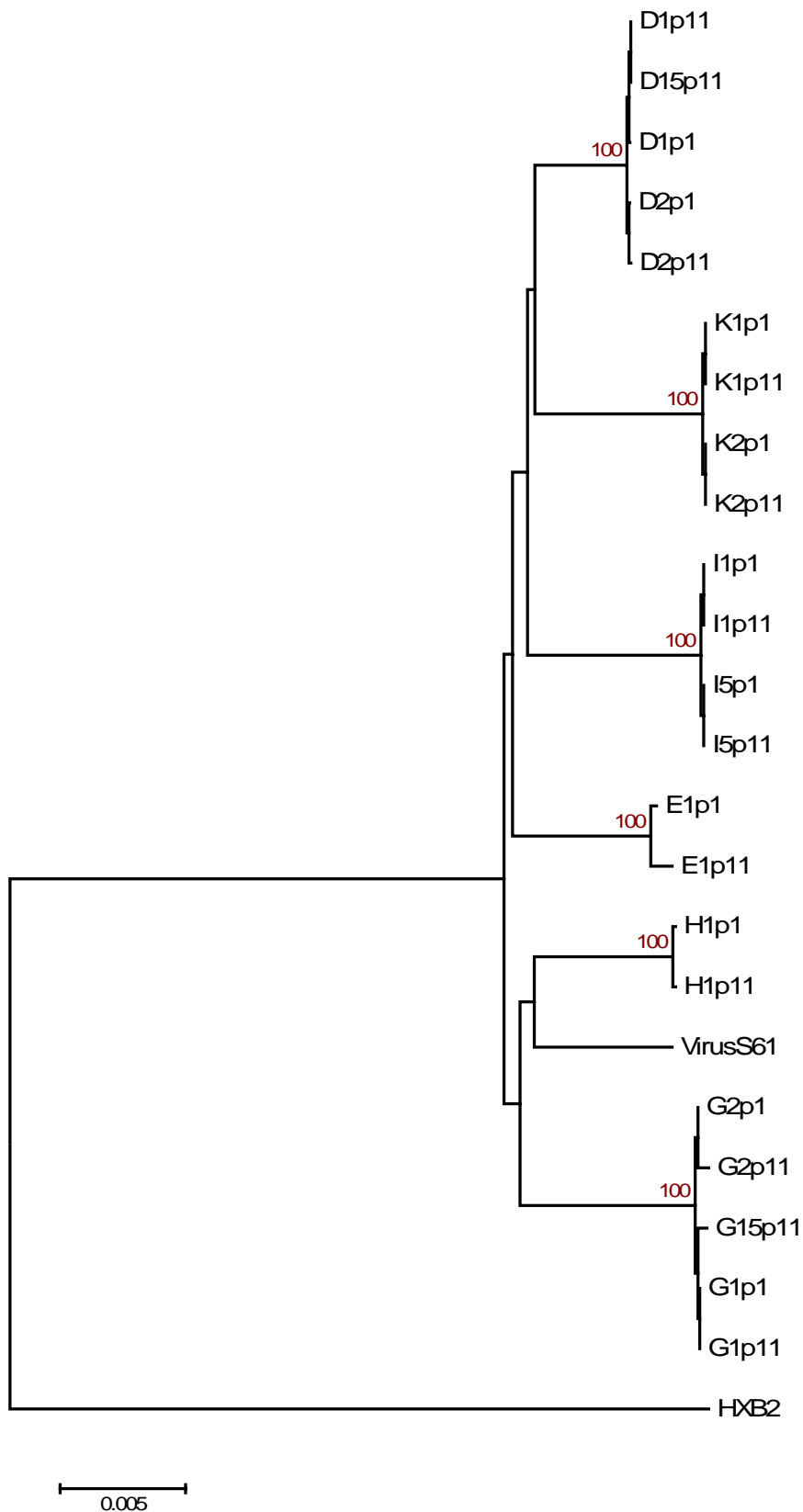
### **3. 2.- COMPARACION DE LOS GENOMAS CONSENSO DE LOS CLONES.**

#### 3. 2. 1.- Árbol filogenético de los clones virales.

Para estudiar qué cambios genéticos eran los responsables del aumento de eficacia biológica tras los pases seriados, nos planteamos la secuenciación de los genomas completos. Se obtuvieron los genomas completos consenso de todos los clones tanto en el pase inicial como en el pase final de recuperación de la eficacia biológica, en total 22 genomas (10 de los pases iniciales y 12 de los pases finales). Los genomas completos comprenden desde el extremo 5'LTR, incluyendo la región U3, R, U5 y la secuencia leader, hasta el gen *nef*.

Con estas secuencias se llevaron a cabo en primer lugar los alineamientos de los genomas virales por medio del programa Clustal X (Thompson y cols 1997), y seguidamente se obtuvieron las relaciones filogenéticas de los clones (ver Figura 3.14) con el programa MEGA v3.1 por el método del vecino más próximo, tal y como se describe en Materiales y Métodos (apartado 2.3.4). Para dar valor estadístico a las diferentes ramas se realizó un re-muestreo con reemplazamiento (bootstrap) con 1000 réplicas. Como se puede observar en el árbol filogenético, cada uno de los clones se mantiene agrupado junto con sus pases y placas, con un alto valor de bootstrap (100%), indicando que no hay contaminación entre muestras y que cada linaje ha mantenido su identidad. La distancia genética promedio entre linajes es del 1%. Otro dato importante es que entre los pases iniciales y finales de cada linaje no se ha producido una gran divergencia, dado que las ramas se encuentran muy próximas. Como mostraremos en apartados posteriores, se han producido pocas sustituciones entre los pases iniciales y finales de recuperación de eficacia biológica.





**Figura 3.14.** Árbol filogenético de los clones virales. El grupo externo del árbol es el virus de referencia HXB2. El virus s61 es el virus parental del cual se obtuvieron todos los linajes virales. En rojo se destacan los valores de bootstrap.

3. 2. 2.- Análisis de las sustituciones fijadas en el genoma viral entre el pase inicial y final de recuperación de eficacia biológica.

Como hemos visto anteriormente, la mayoría de los clones virales aumentaron su eficacia biológica tras los 11 pases seriados. Estos aumentos de eficacia deberían estar correlacionados con la aparición y fijación de mutaciones en el genoma de los clones biológicos finales. Comparando las secuencias de los pases iniciales y finales obtuvimos la relación de mutaciones que habían aparecido entre pases (ver Tabla 3.3). Observamos que la media de sustituciones fijadas entre pases fue muy pequeña con un valor promedio de 1.6 mutaciones por genoma y clon. Incluso se detectaron 5 clones de los 12 estudiados (42%) que no fijaron ninguna mutación entre el pase inicial y final, los clones G1p11, I1p11, I5p11, K1p11 y K2p11. En el resto de clones biológicos la distribución de mutaciones fue la siguiente: Una mutación en los clones D1p11, D15p11, y D2p11, 3 mutaciones en el clon H1p11, 4 mutaciones en los clones G15p11 y G2p11, y 6 mutaciones en el clon E1p11. Si estudiamos el tipo de mutación fijada, podemos observar un alto porcentaje de sustituciones desde el nucleótido G al nucleótido A (ver Figura 3.15), siendo ésta una típica sustitución del VIH-1. Además, las transiciones predominan (74%) frente a las transversiones. Al analizar cuantas reversiones se habían producido, detectamos 3 en total (15% de las sustituciones), 1 tanto en el clon E1p11, como en el G15p11 y el H1p11.

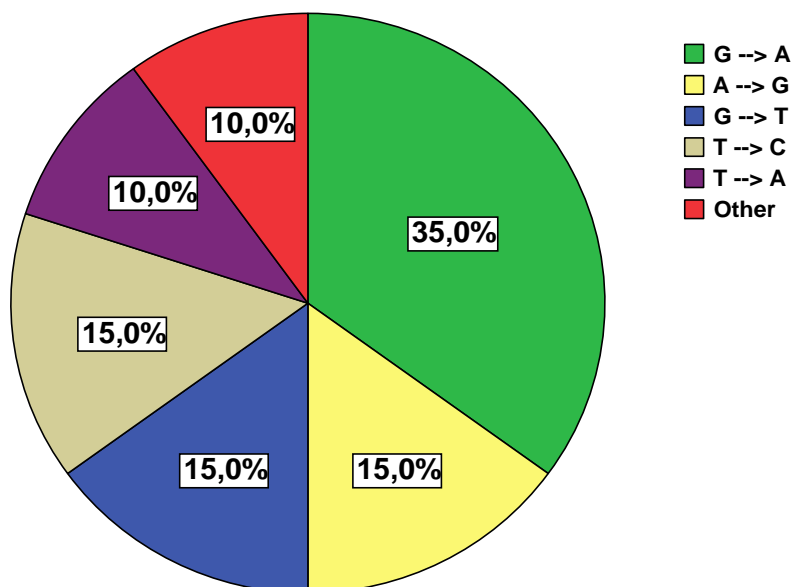
Los clones D1p11, D15p11 y D2p11, fijaron una misma sustitución en la región U5 de la 5'LTR en la posición 593, del nucleótido G al nucleótido T. Este dato es relevante dado que ha ocurrido en las 3 réplicas del linaje D. Excluimos la contaminación dado que se tomaron todas las precauciones posibles para evitarla, además de que existían mutaciones específicas para cada linaje viral que facilitaban la distinción entre ellos (como se observa por el alto valor de bootstrap en la Figura 3.14). Esta mutación además ya fue detectada en el trabajo previo (Yuste y cols 2005). Esta mutación es específica de este linaje viral, dado que el resto de clones mantienen el nucleótido G, por ello podría ser interesante analizarla para futuros estudios sobre la replicación del VIH-1.

En el clon E1p11, las mutaciones que encontramos en la región codificante se localizaban mayoritariamente en el gen *rev*. De todas éstas, la mutación desde el nucleótido T al C en la posición 8651 (T8651C) estaba produciendo la pérdida del codón de terminación del gen *rev* alargando la proteína en 3 aminoácidos más. Todos

estos cambios en *rev* se solapaban con la pauta de lectura del gen de la envuelta, aunque en este gen los cambios eran sinónimos.

En el clon final G15p11 encontramos una acumulación de mutaciones en el gen de la polimerasa, además de una mutación que afectaba tanto al gen *rev* como al gen *tat* (A8402G), produciendo en ambos un cambio no sinónimo. En el clon final G2p11, las mutaciones se localizaban solamente en el gen *pol*. Observamos, que las 4 mutaciones que encontramos se localizaban dos a dos en codones contiguos, generando cada una de estas parejas un cambio no sinónimo en aminoácidos adyacentes. Según la base de datos de los Álamos (Leitner y cols 2005) estas posiciones están muy conservadas, siendo G en vez de A la base más habitual (mayor al 99%).

Una mutación que parece interesante es la producida en el clon biológico H1p11 en la posición 7758. Esta mutación, que es una reversión, se encuentra localizada en el motivo de procesamiento de la gp160 a gp120 y gp41. Es una región compuesta de 7 aminoácidos, y el cambio que se produce podría estar implicado en recuperación de este procesamiento, que se podría haber visto mermado por la introducción del nucleótido A durante los cuellos de botella. También es destacable en este clon la recuperación de la pauta de lectura de la proteína nef, que aunque se ha descrito que *in vitro* es dispensable, sí que se ha observado que favorece una buena replicación viral.



**Figura 3.15.** Porcentaje de cada tipo de sustitución detectada en los pases finales de recuperación de eficacia biológica.

Clon	Número mutaciones	Localización	Cambio base	Cambio aa	Valor <sup>a</sup>
<b>D1p1</b>					
D1p11	1	U5 5' LTR (593)	G → T	NA	-
D15p11	1	U5 5' LTR (593)	G → T	NA	-
<b>D2p1</b>					
D2p11	1	U5 5' LTR (593)	G → T	NA	-
<b>E1p1</b>					
E1p11	6	U3 5' LTR (255)	G → A	NA	-
		U5 5' LTR (572)	T → C	NA	-
		<i>rev</i> (8499)	G → A	R → Q	3
		<i>rev</i> (8511) <b>R</b>	T → C	L → P	3
		<i>rev</i> (8651)	T → C	STOP → Q	
		<i>env gp41</i> (8766)	T → A	D → N	5
<b>G1p1</b>					
G15p11	4	<i>pol</i> (3129) <b>R</b>	A → G	K → E	4
		<i>pol</i> (4421)	T → A	D → E	5
		<i>pol</i> (4842)	G → A	A → T	5
		<i>rev</i> (8402)	A → G <sup>b</sup>	T → A	5
		<i>tat</i>		D → G	4
<b>G2p1</b>					
G2p11	4	<i>pol</i> RT(3144)	G → A	E → K <sup>c</sup>	4
		<i>pol</i> RT(3146)	G → A		
		<i>pol</i> RT(3147)	C → A	Q → K <sup>c</sup>	4
		<i>pol</i> RT(3149)	G → A		
<b>H1p1</b>					
H1p11	3	<i>gag</i> p24(1443)	G → A	M → I	4
		<i>env</i> g120-gp41(7758) <b>R</b>	A → G	T → A	5
		<i>nef</i> (8913)	- → A	Recuperación fase lectura	
<b>Total</b>					
	20				

**Tabla 3.3.** Resumen de las sustituciones detectadas en los genomas virales tras 11 pases seriados. La localización de la sustitución es en relación con el virus de referencia HXB2. NA, indica no aplicable por ser una región no codificante. R indica reversión. <sup>a</sup> El valor de accesibilidad indica lo diferente que es un cambio aminoacídico (Feng y cols 1984) siguiendo un rango, en donde un valor de 6 indicaría que no hay

diferencias, y el 1 un cambio drástico. <sup>b</sup> Esta sustitución está produciendo un cambio de aminoácido en dos proteínas distintas. <sup>c</sup> Indica que ambas sustituciones están implicadas en el cambio aminoacídico.

### **3. 3.- COMPORTAMIENTO DE LAS CUASIESPECIES VIRALES DE LOS CLONES BIOLÓGICOS ANTES Y DESPUES DE LA RECUPERACION DE EFICACIA BIOLÓGICA.**

Al secuenciar y comparar los genomas consenso completos de los distintos clones biológicos, se pudo observar que se había producido una baja fijación de sustituciones entre el pase inicial y final de recuperación. Incluso en algunos casos, como ya describimos en el apartado anterior, detectamos la falta de cambios en los genomas consenso virales. Esto indicaba que debía haber otros factores que pudieran estar afectando la eficacia de los clones virales. Para ello, decidimos estudiar de como la composición y heterogeneidad de las cuasiespecies de los distintos clones podían afectar al aumento en la eficacia biológica.

Para lograr este objetivo, estudiamos las cuasiespecies de cada clon, en las poblaciones iniciales y finales de la recuperación de eficacia biológica. Este estudio se realizó en tres regiones genómicas distintas: *leader/gag*, *vpu/env* (V1-V2) y *env* (V3-V4), ver Materiales y Métodos (apartado 2.2.4). El número de clones, nucleótidos y aminoácidos analizados están descritos en la Tabla 3.4. Por resultados previos del laboratorio habíamos establecido la tasa de mutación debida a la RT-PCR al obtener los clones moleculares en 6 errores por cada 8000 bases amplificadas (I. Olivares, comunicación personal). Tal que si de media cada clon realizado contiene 650 bases, 0.5 mutaciones serían debidas a la RT-PCR.

Se selecciono la región *leader/gag* por los resultados previos en donde la recuperación de eficacia se había localizado preferentemente en la región 5' del genoma (Yuste y cols 2005). La siguiente región seleccionada fue *vpu/env* (V1-V2), ya que nos permitía detectar la posible variabilidad en la región en donde se estaba realizando el calculo de la eficacia biológica, y la inclusión de la *vpu* podría ser interesante a la hora de estudiar la variabilidad en un gen accesorio. La región de la *env* (V3-V4) se usó como control por ser una región que tradicionalmente se ha usado en los estudios evolutivos del VIH-1 y se sabe que es muy variable en los virus estudiados de pacientes.

Región	Pase	Clones	Bases	Aa	Tasa de cambio <sup>a</sup>
leader/gag	1	184 (18)	127328	35144	1.3x10 <sup>-3</sup>
	11	214 (18)	148088	40874	1.6x10 <sup>-3</sup>
vpu / env (V1 – V2)	1	174 (17)	135198	45066	1.6x10 <sup>-3</sup>
	11	225 (19)	174825	58275	2.7x10 <sup>-3</sup>
env (V3 – V4)	1	176 (18)	105600	35024	1.4x10 <sup>-3</sup>
	11	218 (18)	130800	43382	1.6x10 <sup>-3</sup>
Por pase	1	534	232928	115234	
	11	657	278888	142531	
<b>TOTAL</b>		<b>1191</b>	<b>511816</b>	<b>257765</b>	

**Tabla 3.4.** Descripción del número de clones, número de bases y aminoácidos analizados. Entre paréntesis se indica la media de clones analizados por linaje viral. La tasa de cambio se calculó con el total de mutaciones encontradas en cada cuasiespecie comparando con la secuencia consenso de cada clon viral. <sup>a</sup> Expresada por sitio nucleotídico.

Para estudiar si existía algún tipo de sesgo en el número de clones moleculares obtenidos por linaje, pase o región, realizamos un test no paramétrico de Kruskal-Wallis. Según este análisis, determinamos que no había diferencias significativas en el número de clones moleculares analizados tanto entre linajes ( $p=0.547$ ), pases ( $p=0.665$ ) o región ( $p=0.835$ ) analizada. Por tanto, los posteriores análisis con los clones moleculares estudiados no están sometidos a sesgo debido a diferencias en el número de clones obtenidos en cada caso.

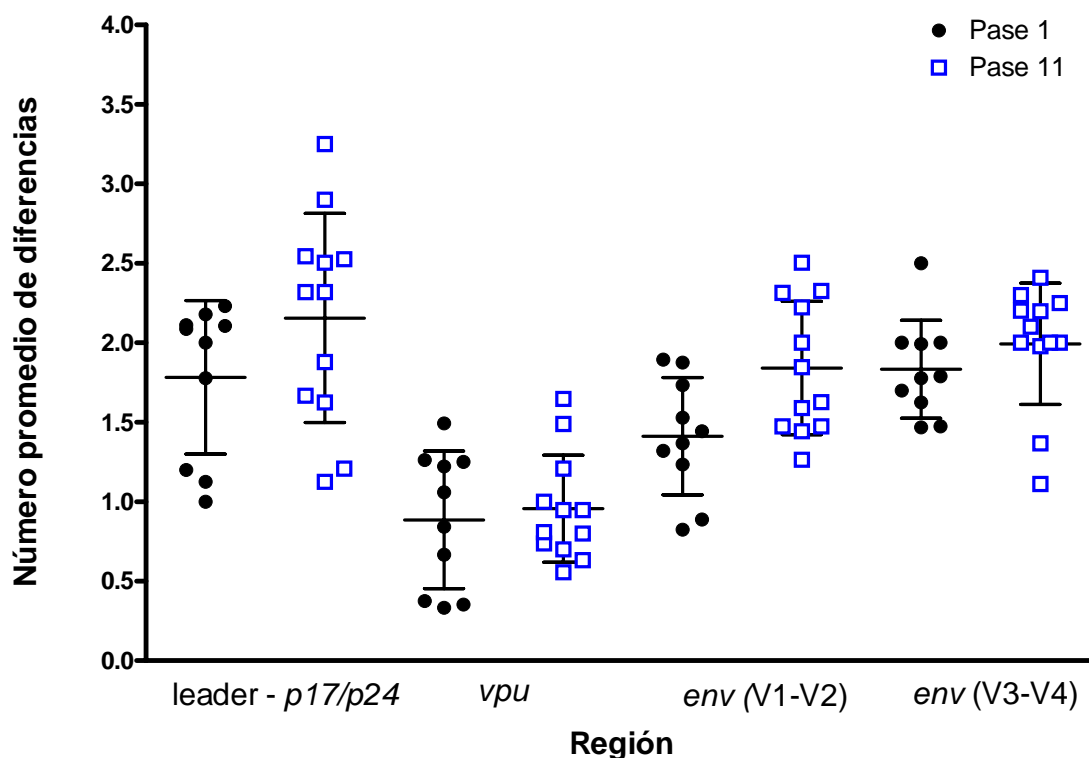
### 3. 3. 1.- Análisis de los cambios de heterogeneidad entre el pase inicial y final de recuperación de eficacia biológica.

Para cada una de las tres regiones del genoma estudiadas, se calculó un valor promedio de heterogeneidad. La región de leader/gag comprende 118 nucleótidos de la secuencia leader, los 396 nucleótidos que componen la proteína p17 (132 aminoácidos) y los 177 primeros nucleótidos de la proteína p24 (59 aminoácidos). Para analizar los clones obtenidos en la región entre vpu/env (V1-V2), esta región fue separada en vpu y env (V1-V2), para facilitar su comparación. Del gen vpu podíamos analizar su secuencia completa, 246 nucleótidos que corresponden a los 83 aminoácidos de su proteína. Del gen env en la región (V1-V2), analizamos los primeros 528 nucleótidos del gen de la

envuelta, es decir, 176 aminoácidos. En la región V3-V4 del gen de la envuelta, analizamos 598 nucleótidos, localizados entre los aminoácido 273 y 473 de la gp120.

Los cálculos se realizaron mediante comparaciones una a una de todas las secuencias que teníamos de la cuasiespecie de cada linaje y dando un valor promedio del número de diferencias encontradas. Las regiones con deleciones en las secuencias se eliminaron de las comparaciones.

Los resultados para la heterogeneidad están detallados en la Figura 3.16.



**Figura 3.16.** Número de diferencias promedio de heterogeneidad. El eje de abscisas representa las distintas regiones genómicas estudiadas. Los círculos negros y los cuadrados azules representan el pase inicial y final de recuperación de eficacia biológica, respectivamente. La línea horizontal representa la media, y las barras de error la desviación estándar.

Para la realización de los análisis estadísticos de la heterogeneidad utilizamos el test estadístico no paramétrico de Kruskal-Wallis, dado que aunque los valores de heterogeneidad seguían una distribución normal, se violaba el supuesto de la igualdad de varianzas, con lo que no podíamos realizar un análisis de varianzas (ANOVA).

En primer lugar analizamos si existían diferencias en el valor de heterogeneidad entre los distintos clones, no encontrando diferencias entre ellos ( $p=0.4$ ). De manera que podíamos hacer los siguientes análisis teniendo en cuenta el total de las secuencias

al no haber encontrado diferencias entre los 6 linajes virales. Comparamos la heterogeneidad de cada linaje viral individualmente entre el pase inicial y final de recuperación de eficacia biológica y no encontramos diferencias estadísticamente significativas (D,  $p=0.64$ , E,  $p=0.15$ , G,  $p=0.247$ , H,  $p=0.15$ , I,  $p=0.79$ , y K,  $p=0.4$ ). Esto nos indicaba que individualmente los linajes no parecían estar aumentando su heterogeneidad tras los pases seriados. Sin embargo, al analizar si globalmente existían diferencias en la heterogeneidad entre el pase inicial y final de recuperación de eficacia, observamos que no había una diferencia estadísticamente significativa, aunque el valor era marginal ( $p=0.06$ ) y por tanto sugería una tendencia hacia el aumento de heterogeneidad al realizar los 11 pases seriados.

Encontramos una diferencia estadísticamente significativa entre las distintas regiones analizadas (LTR/*gag*, gen *vpu*, V1-V2 y V3-V4) con un  $p<0.0001$ . Esto era así tanto si comparábamos solamente las regiones en los pases iniciales o en los finales (ambos  $p<0.0001$ ). Los mayores incrementos de heterogeneidad se detectaron en la región leader-*gag*, tanto en el pase inicial como final, seguido de la región de la envuelta V1-V2, y de la región de la envuelta V3-V4, y finalmente *vpu*. Al comparar cada región individualmente entre el pase inicial y el final, solamente detectamos diferencias estadísticamente significativas en la región V1-V2 ( $p=0.032$ ).

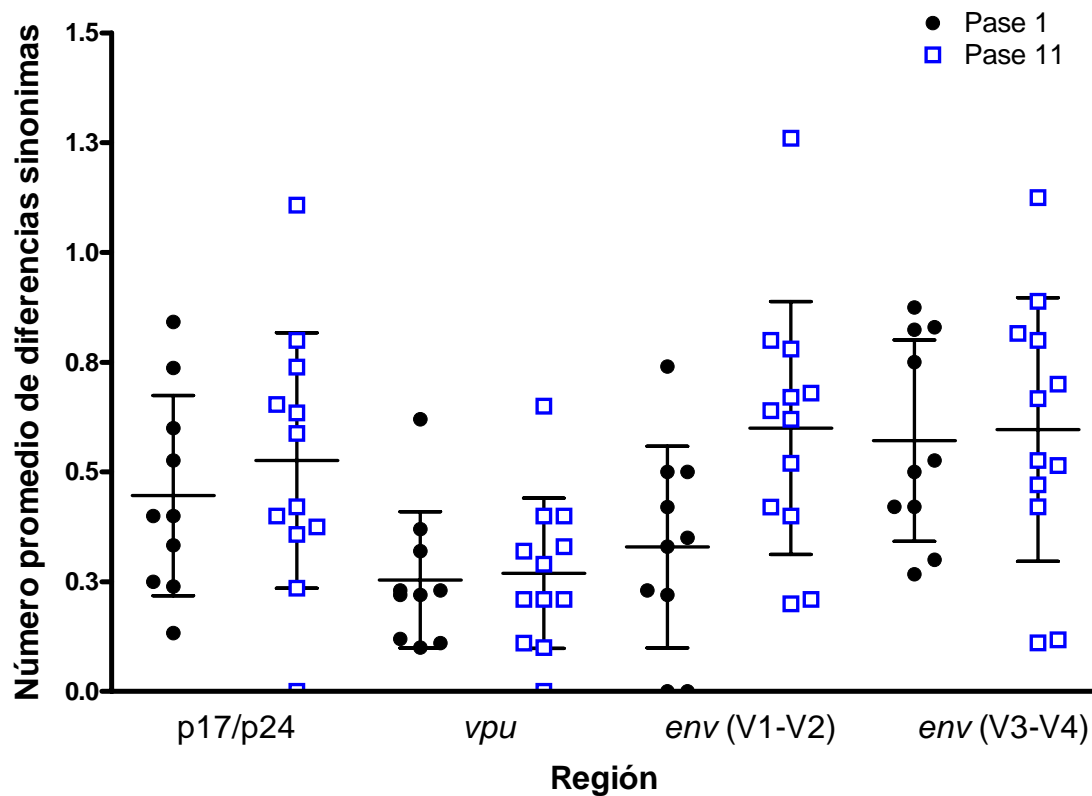
Todos estos resultados nos indican que se estaba produciendo un aumento de heterogeneidad de los clones biológicos finales frente a los iniciales. Sin embargo, no todas las regiones genómicas estudiadas se estaban comportando de la misma manera, destacando entre ellas la región del gen de la envuelta (V1-V2), en donde se estaba produciendo el mayor aumento de heterogeneidad entre los pases de recuperación de eficacia biológica, seguido por la región leader/*gag*, la región de la *env*(V1-V3), y el gen *vpu*, en donde el aumento era más limitado.

### 3. 3. 2.- Análisis de los cambios en las sustituciones sinónimas (Sd) entre el pase inicial y final de recuperación de eficacia biológica.

Usando los mismos clones moleculares obtenidos previamente, analizamos los valores de Sd y Nd de cada uno de los clones, pases y regiones. Así teníamos la secuencia completa de la proteína p17 (192 aminoácidos), y el principio de la proteína p24 (60 aminoácidos) del gen *gag*, la secuencia completa del gen *vpu*, el inicio del gen de la envuelta en el fragmento que comprendía la región V1-V2 (del aminoácido 1 al



170 de la gp120), y finalmente también en el gen de la envuelta la región V3-V4 (del aminoácido 273 al 473 de la gp120). Posteriormente calculamos el número de diferencias promedio de sustituciones sinónimas (Sd, ver Figura 3.17) para cada uno de los clones, regiones y países.



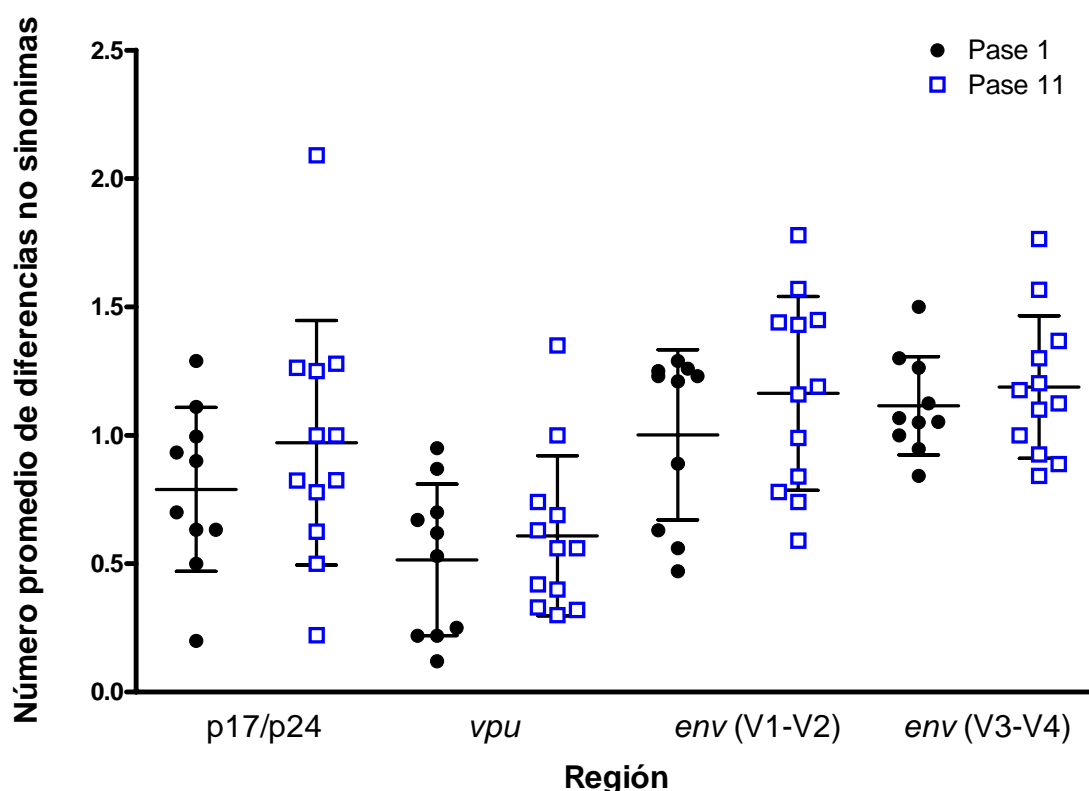
**Figura 3.17.** Representación de los valores de sustituciones sinónimas (Sd) calculadas como número de diferencias promedio entre los clones obtenidos para cada región y pase. Los círculos negros representan el pase 1 (inicial), y los cuadrados azules el pase 11 (final). La línea horizontal representa la media, y las barras de error la desviación estándar.

Los valores Sd se analizaron según el test no paramétrico Kruskal-Wallis, por las mismas razones que para la heterogeneidad (ver apartado anterior), de manera que este será el  $p$ -valor presentado en los análisis estadísticos de este apartado. Como se puede observar en la Figura 3.17, la evolución de los cambios sinónimos era muy limitada tanto en los pases iniciales como en los finales de recuperación de eficacia biológica, no encontramos diferencias estadísticamente significativas entre las Sd de los distintos linajes virales ( $p=0.551$ ). Asimismo, tampoco encontramos diferencias entre los pases ( $p=0.167$ ). Sin embargo cabe destacar la existencia de diferencias estadísticas entre las distintas regiones ( $p<0.0001$ ) en cada uno de los pases. Por tanto, no todas las regiones se están comportando igual a la hora de acumular Sd. Comprobamos que el aumento

entre pase inicial y final de las Sd de la región *env* (V1-V2) era el único aumento significativamente estadístico con un  $p=0.038$ .

### 3. 3. 3.- Análisis de los cambios en las sustituciones no sinónimas (Nd) entre el pase inicial y final de recuperación de eficacia biológica.

De igual manera que con las Sd, obtuvimos el número promedio de diferencias para las Nd de las regiones descritas anteriormente para cada clon, región y pase. Los resultados obtenidos se representan en la Figura 3.18.



**Figura 3.18.** Representación de los valores de sustituciones no sinónimas (Nd) calculadas como número de diferencias promedio entre los clones obtenidos para cada región y pase. Los círculos negros representan el pase 1 (inicial), y los cuadrados azules el pase 11 (final). La línea horizontal representa la media, y las barras de error la desviación estándar.

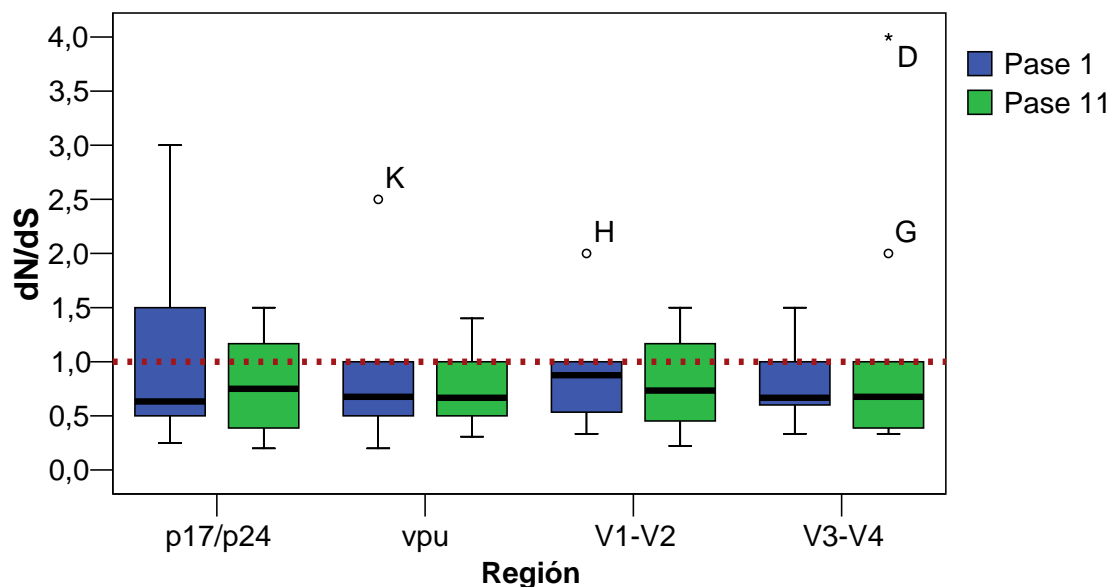
Para los valores Nd, los  $p$ -valores que presentamos son todos referidos al test Kruskal-Wallis. Observamos que no había diferencias estadísticamente significativas entre los distintos linajes virales analizados ( $p=0.263$ ). Tampoco encontramos diferencias estadísticamente significativas entre los valores de Nd entre el pase inicial y final ( $p=0.242$ ). Sin embargo, al igual que con los valores de Sd, encontramos una

diferencia estadística en las Nd de las distintas regiones ( $p < 0.0001$ ), para cada uno de los pases. En este caso, los valores de Nd del gen *vpu* eran menores comparado al resto de las regiones: frente a p17/p24 ( $p = 0.006$ ), frente a la región V1-V2 ( $p < 0.0001$ ) y finalmente frente a la región V3-V4 ( $p < 0.0001$ ).

### 3. 3. 4.- Estudio del cociente dN/dS para las regiones estudiadas.

Para este análisis calculamos los valores de dS y dN para cada clon estudiado. Para determinar la existencia y la dirección de la selección se calculó el cociente entre dN y dS. Un valor de dN/dS mayor a 1, indica selección positiva, debido a una prevalencia de las sustituciones no sinónimas; si por el contrario dN/dS es menor a 1, esto nos indica selección purificadora, es decir, prevalencia de las sustituciones sinónimas. Cuando analizábamos si los valores del cociente eran menores a 1, observamos que en el pase inicial era no significativo ( $p = 0.139$ ). Por tanto los clones están bajo selección neutral (dN/dS=1, Figura 3.19). Mientras que los valores para el pase final, aunque no significativos ( $p = 0.06$ ) tendían a ser menores a 1, y por tanto indicaban selección purificadora.

En la región p17/p24 se observó una homogenización y ligero aumento del cociente. En la región *env* (V1-V2), se observó un ligero descenso en el cociente. También se detectaron linajes cuyos valores del cociente se alejaban de la media de la población; en la región del gen *vpu*, y *env* (V1-V2), se observó una pérdida de estos valores extremos al realizar los 11 pases seriados (linajes K y H, respectivamente). Sin embargo, en el caso de la región *env* (V3-V4) se detectó una aparición de valores extremos al realizar los pases seriados (linajes D y G). En la región del gen *vpu* no se observaron cambios apreciables en el cociente, sugiriendo la muy limitada evolución de este gen en nuestras condiciones experimentales.



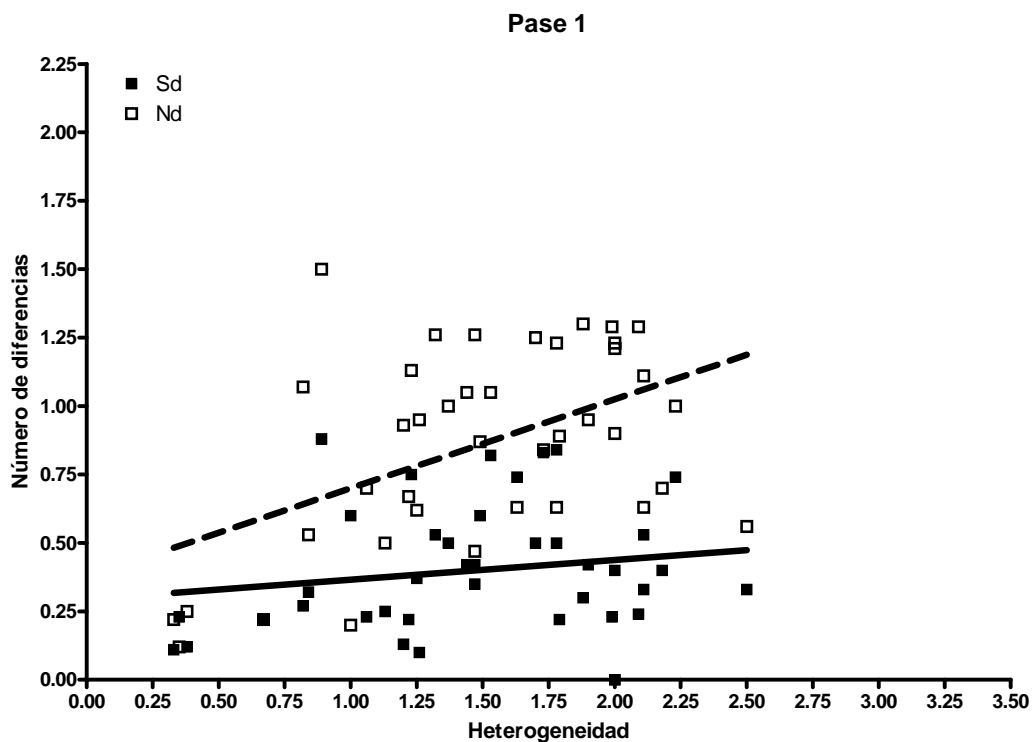
**Figura 3.19.** Gráfica en caja (Boxplot) del cociente dN/dS, para cada una de las regiones estudiadas. La línea punteada roja indica neutralidad. La caja representa el 50% de los valores; la línea horizontal oscura indica la mediana; la barras, tanto la superior como la inferior representan cada una el 25% de datos restante. Los círculos y el asterisco representan valores que se diferencian más de 1.5 ó 3 veces, respectivamente, de la longitud de la caja.

### 3. 3. 5.- Análisis conjunto del aumento de heterogeneidad, Sd, Nd y eficacia.

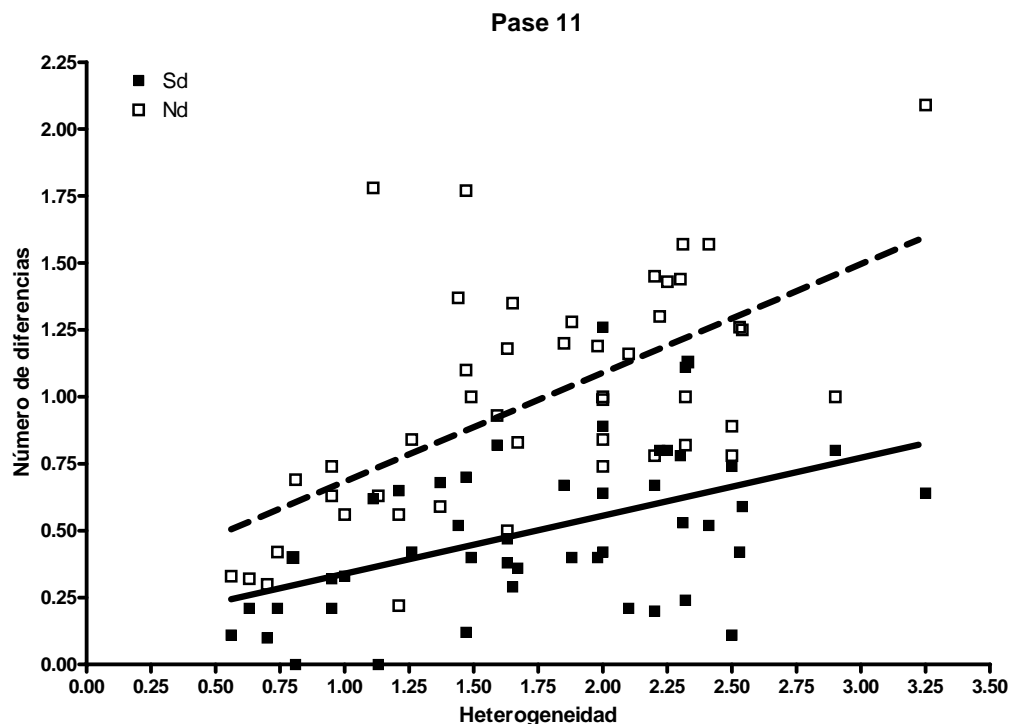
Para estudiar de qué manera estaba relacionado el aumento de heterogeneidad con las sustituciones sinónimas, no sinónimas y la eficacia, realizamos regresiones lineales para cada uno de los pases de recuperación de eficacia biológica. Las correlaciones para Sd y Nd se realizaron tomando individualmente cada valor de heterogeneidad de cada región y clon. Para las correlaciones de heterogeneidad con la eficacia biológica se realizó un promedio de todas las regiones y se le asignó a ese valor el valor de eficacia del clon viral.

## 3. 3. 5. 1.- Correlaciones entre la heterogeneidad y Sd o Nd.

En el pase inicial, no observamos una correlación entre heterogeneidad y Sd. Sin embargo, observamos una correlación positiva entre heterogeneidad y Nd ( $p=0.001$ , ver Figura 3.20), indicando que en el momento de iniciar la recuperación de eficacia los clones que tenían alta heterogeneidad eran los que poseían también un elevado valor de Nd. En el pase final observamos una correlación positiva tanto para las Sd ( $p=0.0006$ ) como para las Nd ( $p<0.0001$ ), indicando que el aumento de heterogeneidad que se observaba era debido tanto a las Sd como a las Nd (ver Figura 3.21), y por tanto apoyando una evolución neutral o purificadora.



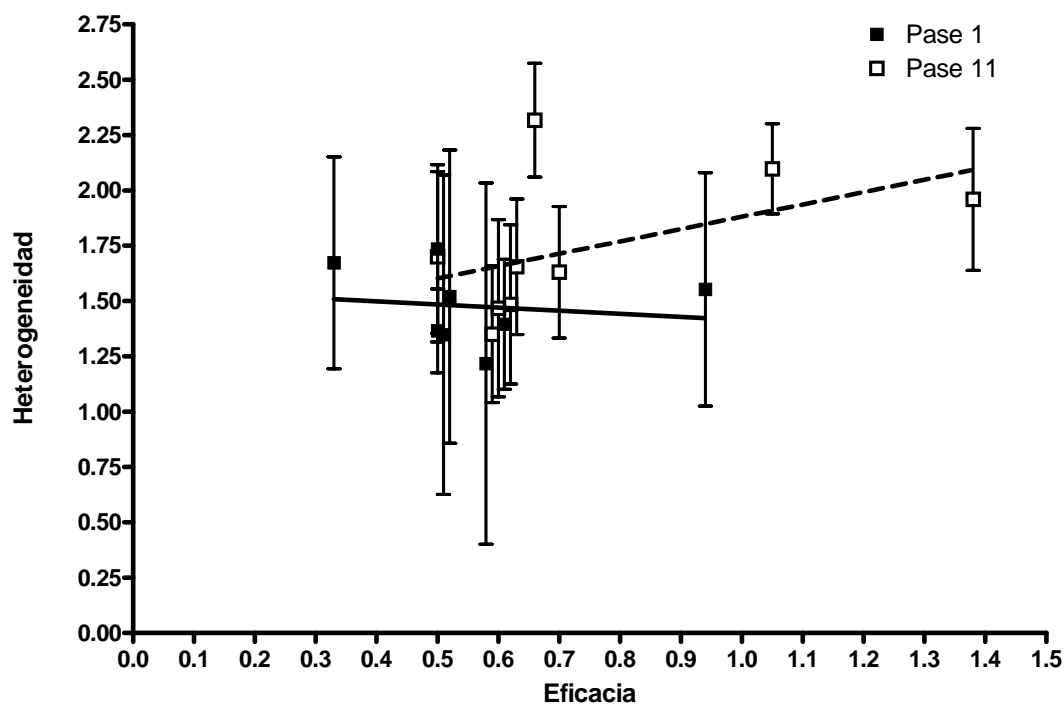
**Figura 3.20.** Regresión lineal de la heterogeneidad frente a Sd (cuadrado lleno) y Nd (cuadrado vacío) en el pase inicial de recuperación de eficacia biológica. La línea continua representa la regresión para Sd y la línea punteada para Nd.



**Figura 3.21.** Regresión lineal de la heterogeneidad frente a Sd (cuadrado lleno) y Nd (cuadrado vacío) en el pase final de recuperación de eficacia biológica. La línea representa la regresión para Sd y la línea punteada para Nd.

### 3. 3. 5. 2.- Correlación entre heterogeneidad y eficacia biológica.

Realizamos la regresión lineal de los valores de eficacia biológica y heterogeneidad tanto para el pase inicial como el pase final de recuperación de eficacia biológica. Los resultados indicaron que ni en el pase inicial ni en el final existía una correlación entre las dos variables estudiadas,  $p=0.75$  y  $p=0.18$ , respectivamente (ver Figura 3.22). En esta regresión no pudimos tener en cuenta los linajes virales D1p11, D15p11 y D2p11, a pesar de que estos linajes fueron los que sufrieron el mayor aumento de eficacia ya que desplazaron totalmente al virus de referencia J1 en los pases de competición. Por ello, de poder incluir la eficacia de estos clones en la regresión para el pase 11 seguramente esta sería significativa, y por tanto si que existiría una correlación entre heterogeneidad y eficacia biológica. De manera que esto sugiere una tendencia a mayores valores de heterogeneidad asociados a mayores valores de eficacia.



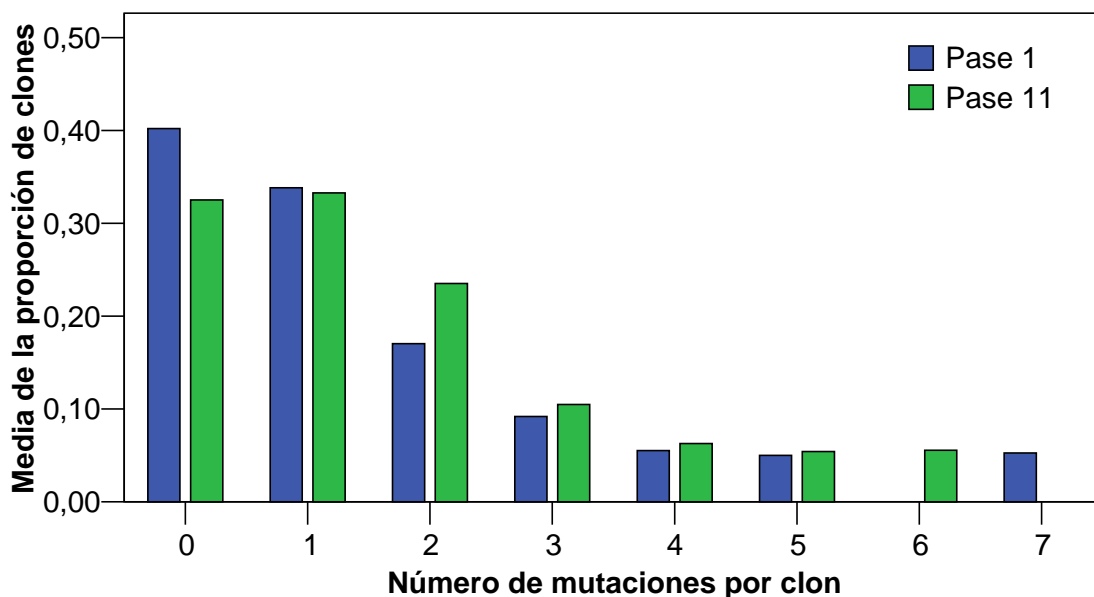
**Figura 3.22.** Regresión lineal de la eficacia biológica frente a la heterogeneidad. Los cuadrados llenos y vacíos representan el pase 1 y pase 11 respectivamente de los pases de recuperación de eficacia biológica.

### 3. 3. 6.- Análisis de la acumulación de mutaciones en los clones moleculares estudiados.

Quisimos analizar si existían diferencias entre el pase inicial y final de recuperación de eficacia biológica con respecto a la acumulación de mutaciones en los clones moleculares. En primer lugar, obtuvimos para cada uno de los linajes virales y pases el valor del número de mutaciones acumuladas en cada clon molecular al comparar con su respectiva secuencia consenso, de manera que categorizamos los clones moleculares según el número de mutaciones que habían acumulado. Una vez obtenido el número de mutaciones por clon molecular, obtuvimos la proporción de cada clon frente al total de clones moleculares analizados en cada caso. Seguidamente, realizamos un análisis no paramétrico según Kruskal-Wallis para detectar diferencias en la distribución de mutantes (ver Figura 3.23). De esta manera, comprobamos que no existían diferencias entre los distintos clones virales ( $p=0.775$ ). Tampoco detectamos diferencias entre las 3 regiones (leader/gag, env V1-V2 y V3-V4) estudiadas ( $p=0.275$ ), ni entre los pases iniciales frente a los finales ( $p=0.936$ ). Analizando más en detalle, encontramos diferencias significativas entre el pase inicial y final entre el número de clones moleculares con 0 y 2 mutaciones ( $p=0.057$  y  $p=0.007$  respectivamente). Esto

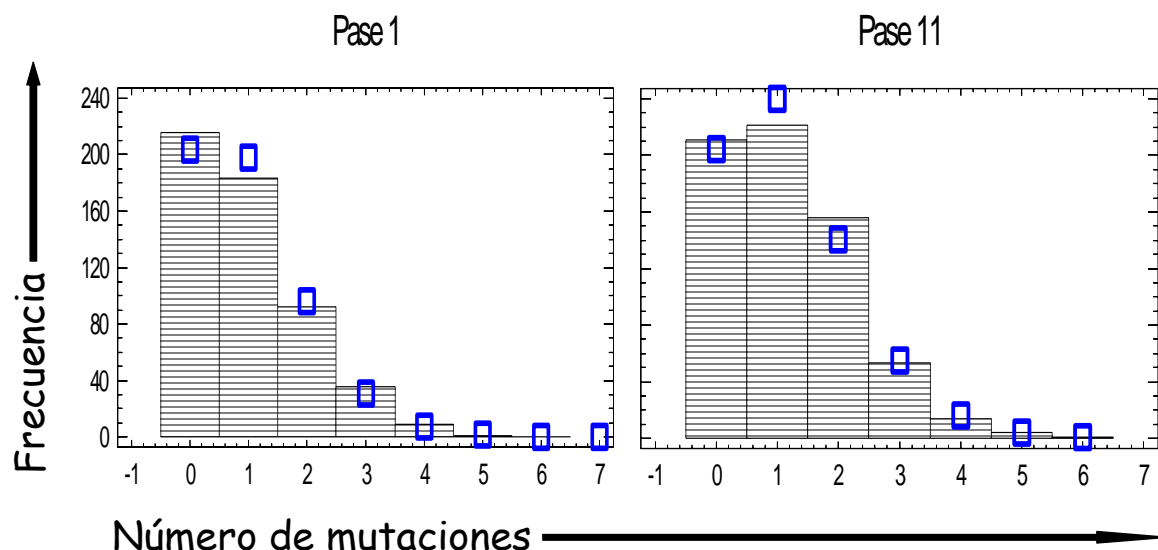
indica que entre el pase inicial y final de recuperación de eficacia biológica se estaba produciendo una disminución en el número de mutantes con 0 mutaciones, y un aumento de clones moleculares con 2 mutaciones.

Una vez analizada de manera general la distribución del número de mutaciones en los distintos clones y en los pases, ajustamos el número de mutaciones encontradas en ambos pases a una distribución Poisson (ver Figura 3.24). Dado que si las mutaciones se están produciendo de manera espontánea sin verse sometidas a selección, simplemente afectadas por la deriva genética, la acumulación de estas podría modelizarse según una distribución Poisson (Haigh 1978). Para el pase inicial el parámetro de la Poisson ( $\lambda_i$ ) estimado era igual a 0.97 y para el pase final ( $\lambda_f$ ) de 1.2. Según esto, observamos que ninguno de los pases se podía diferenciar estadísticamente de su distribución de Poisson asociada (Pase 1,  $p=0.37$ , Pase 11,  $p=0.44$ ).



**Figura 3.23.** Proporción del número de mutaciones por clon y por pase.





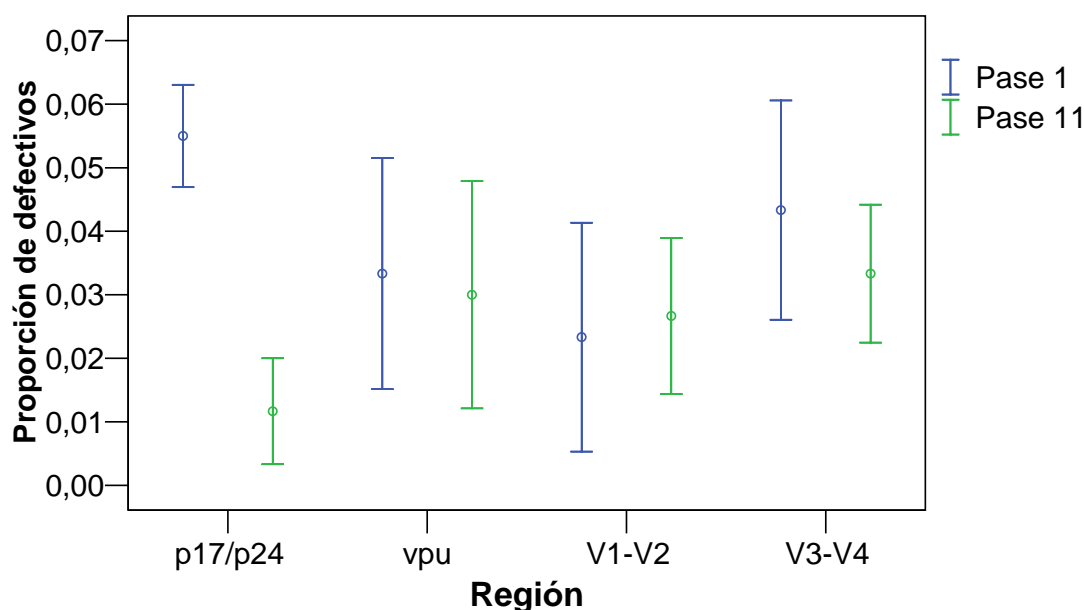
**Figura 3.24.** Ajuste del número de mutaciones a la distribución Poisson. Las barras indican los valores observados, los puntos indican los valores esperados de la distribución Poisson asociada al parámetro de la Poisson estimado con nuestros valores experimentales.

### 3.3.7.- Análisis de clones moleculares con codón de terminación.

Mediante el análisis de las cuasiespecies virales en las distintas regiones codificantes observamos que algunos de los clones moleculares secuenciados tenían codones de terminación prematuros, que podía imposibilitar la síntesis de una proteína completa y funcional en ese clon específico. Según estos datos, quisimos comprobar si el número de clones con codones de terminación variaba entre los pases, y si había diferencias entre las regiones estudiadas. Para ello calculamos la proporción de clones moleculares con codón de terminación agrupando según los distintos linajes (D, E, G, H, I y K), y ponderando por el número de clones moleculares obtenidos en cada caso. Con el valor obtenido realizamos comparaciones estadísticas con el test no paramétrico de Kruskal-Wallis.

Comparamos la proporción de codón de terminación entre los pases iniciales y los finales de recuperación de eficacia biológica de manera global sin diferenciar entre clones ni regiones. Observamos que no había diferencias significativas entre los dos pases seriados ( $p=0.230$ ). Tampoco encontramos diferencias globales entre las distintas regiones estudiadas, p17/p24 del gen *gag*, gen *vpu*, y región V1-V2 y región V3-V4 del gen *env* ( $p=0.69$ ), ni entre los distintos linajes ( $p=0.253$ ). Sin embargo si representábamos los valores promedios de clones defectivos de cada región y pase (ver

Figura 3.25), la región p17/p24 era bastante distinta entre el pase inicial y el final. Así que analizamos esa región y realizamos la comparación entre los distintos pases. Observamos una diferencia significativa entre los dos pases ( $p=0,01$ ) para la región p17/p24 (5% defectivos inicialmente frente a un 1% en el pase final). El resto de las regiones tomadas individualmente no eran significativamente distintas. Por tanto durante los pases seriados se estaban eliminando de la población viral los clones que tenían codones de terminación en la región p17/p24.



**Figura 3.25.** Representación gráfica de la variación de la proporción de codones de terminación (ordenadas) entre las distintas regiones (abscisas) y los distintos pases (en azul el pase 1, y en verde el pase 11). La media esta representada por un círculo, las líneas representan la desviación estándar.

### 3. 3. 8 . - Análisis de la distribución de las sustituciones sinónimas y no sinónimas en las distintas regiones estudiadas.

Se calculó la media del número de posiciones variables para cada región (ver Tabla 3.5) y se observó un incremento del número de posiciones variables entre los pases seriados. Pudimos detectar una diferencia estadísticamente significativa entre los sitios variables de la región *env* (V1-V2) entre pases (prueba t-student,  $p=0.003$ ). En el resto de regiones aunque no existían diferencias estadísticamente significativas se observó una tendencia al aumento en el número de sitios variables, con la excepción del gen *vpu*.

Linaje	Pase	leader/gag	vpu	env (V1-V2)	env (V3-V4)	Total
<b>D</b>						
	1	15	7	11.5	18.5	52
	11	19.6	9.3	19.3	19.3	67.5
<b>E</b>						
	1	20	8	18	14	60
	11	20	9	20	23	72
<b>G</b>						
	1	12.5	8.5	12.5	18	51.5
	11	17.3	7.6	18.3	16.6	59.8
<b>H</b>						
	1	19	9	13	16	57
	11	15	14	20	18	67
<b>I</b>						
	1	17.5	6	8	13.5	45
	11	15	5.5	10.5	15	46
<b>K</b>						
	1	14.5	11	13	14.5	53
	11	20.5	7	18.5	16.5	62.5
<b>Entre linajes</b>						
	1	16.4	8.25	12.6	15.75	<b>53</b>
	11	17.9	8.7	17.8*	18.1	<b>62.5</b>

**Tabla 3.5.** Media de sitios variables por linaje y pase para cada una de las 4 regiones estudiadas. El asterisco indica diferencias estadísticamente significativas entre el número de sitios variables ( $p < 0.05$ ).

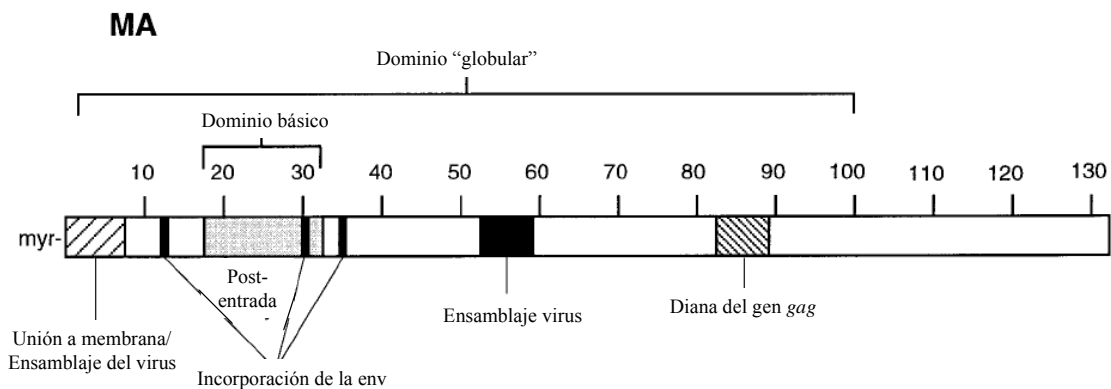
Posteriormente, analizamos si las sustituciones tanto sinónimas como no sinónimas se estaban, o no, distribuyendo aleatoriamente en las distintas posiciones aminoacídicas a lo largo de las regiones estudiadas. Para ello, utilizamos el programa CRANN (Materiales y Métodos apartado 2. 3. 7) y realizamos un análisis por medio de una ventana deslizante a lo largo de la secuencia para obtener el valor acumulativo de las dS y dN a lo largo de la región estudiada.

3. 3. 8. 1.- Región gen *gag* (p17/p24).

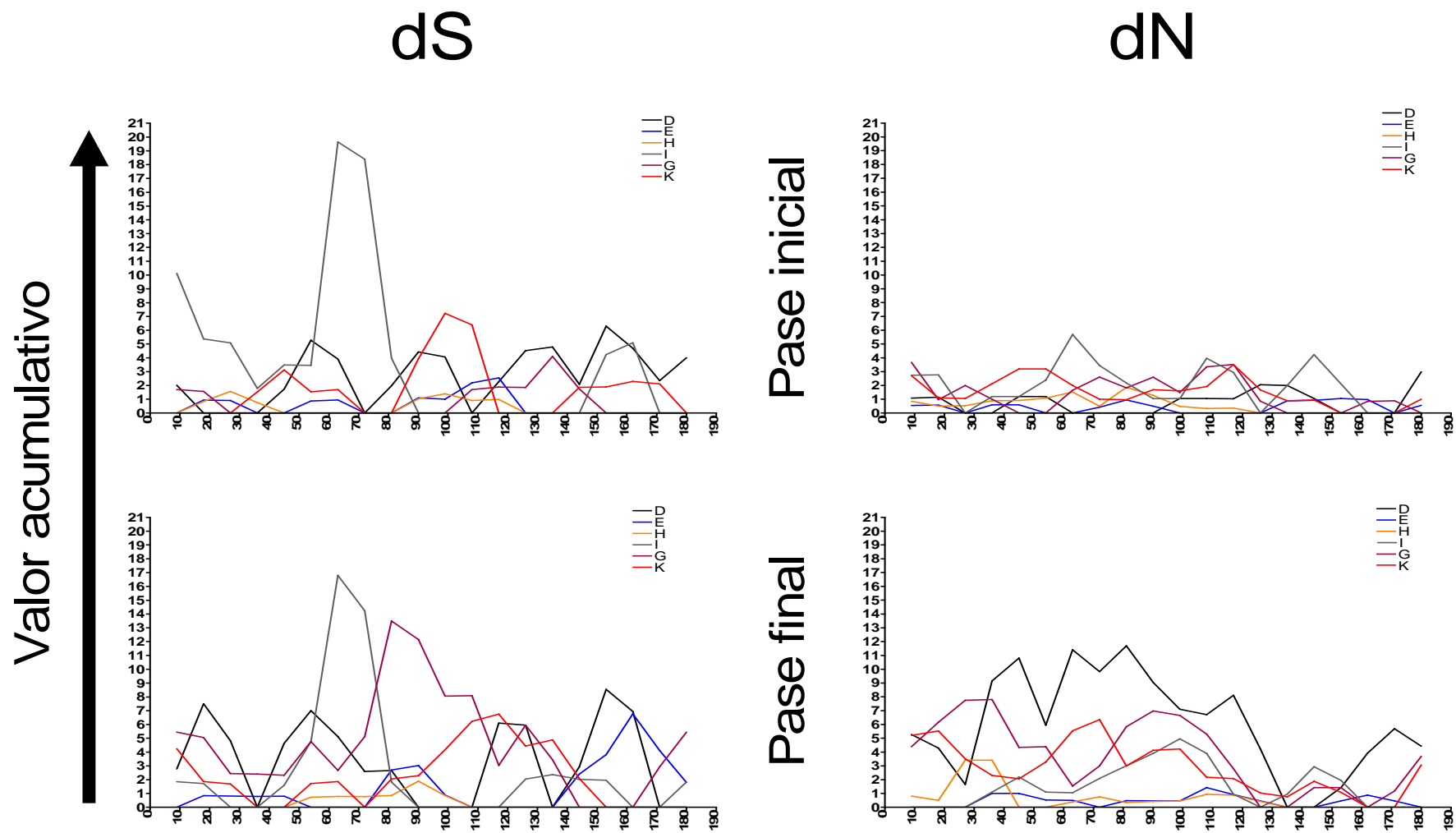
Comparamos la distribución de sustituciones sinónimas y no sinónimas de cada clon en el pase inicial y final de recuperación de eficacia biológica (Figura 3.27). La proteína p17 comprende del aminoácido 1 al 133, y a continuación separado por la señal de procesamiento (del aminoácido 133 al 134), se sitúa la región codificante para la proteína p24, de la que solamente tenemos analizados los primeros 56 aminoácidos, del aminoácido 134 al 190 (Figura 3.24). La proteína p17 tiene varios dominios importantes representados en la Figura 3.26.

Como se puede observar en la Figura 3.27, existen variaciones en la distribución de las sustituciones entre el pase inicial y final. Realizamos la comparación entre el pase inicial y final según el test estadístico no paramétrico de Kruskal-Wallis. Para las sustituciones sinónimas sólo encontramos diferencias entre pases en el linaje E ( $p < 0.0001$ ), para el resto de linajes el test fue no significativo.

Al comparar las sustituciones no sinónimas observamos diferencias estadísticamente significativas entre pases en el linaje D,  $p < 0.0001$ , E,  $p = 0.003$  y el G,  $p = 0.015$ . En el resto de linajes, las diferencias fueron no significativas.



**Figura 3.26.** Representación gráfica de los dominios de la proteína p17 del gen *gag*. Modificado de (Freed 1998).



**Figura 3. 27.** Distribución de los valores acumulativos de las sustituciones sinónimas (izquierda) y no sinónimas (derecha) en la cuasispecie inicial y final de la p17/p24. El eje de abscisas representa la región analizada en aminoácidos por la ventana deslizante.

### 3. 3. 8. 2.- Región gen *vpu*

La proteína *vpu* tiene dos actividades biológicas que se corresponden con dos dominios estructurales, que están bien localizadas y son independientes. La primera región es la el dominio N-terminal hidrofóbico transmembrana (TM) localizado del aminoácido 1 al 28. Y la segunda región es el dominio C-terminal polar citoplasmático localizado entre al aminoácido 29 al 80 (ver Figura 3.28). Comparamos la distribución de las sustituciones entre el pase inicial y final de recuperación de eficacia biológica, al igual que en la región anterior. Para las sustituciones sinónimas encontramos que sólo dos linajes tenían diferencias significativas entre pases, el D,  $p=0.019$  y el G,  $p=0.019$ . El resto de linajes fueron no significativos. Para las sustituciones no sinónimas la situación era similar, encontrando tan solo 3 linajes con diferencias estadísticamente significativas entre los pases, el D,  $p=0.049$ , el G,  $p=0.001$ , y el I,  $p=0.002$ . En el resto de linajes, las diferencias fueron no significativos.

En la Figura 3.28, podemos ver como los linajes D y G son los que más varían entre los pases de recuperación. El resto de clones no presentan cambios significativos, y tampoco alcanzan en ningún caso valores elevados de sustituciones. Hay que destacar también el bajo número de sustituciones sinónimas y no sinónimas en el pase inicial de cada uno de los linajes. Esto, junto con la poca variación encontrada en el pase final, nos indica que esta proteína no está siendo sujeta a una marcada selección para la recuperación de eficacia biológica.

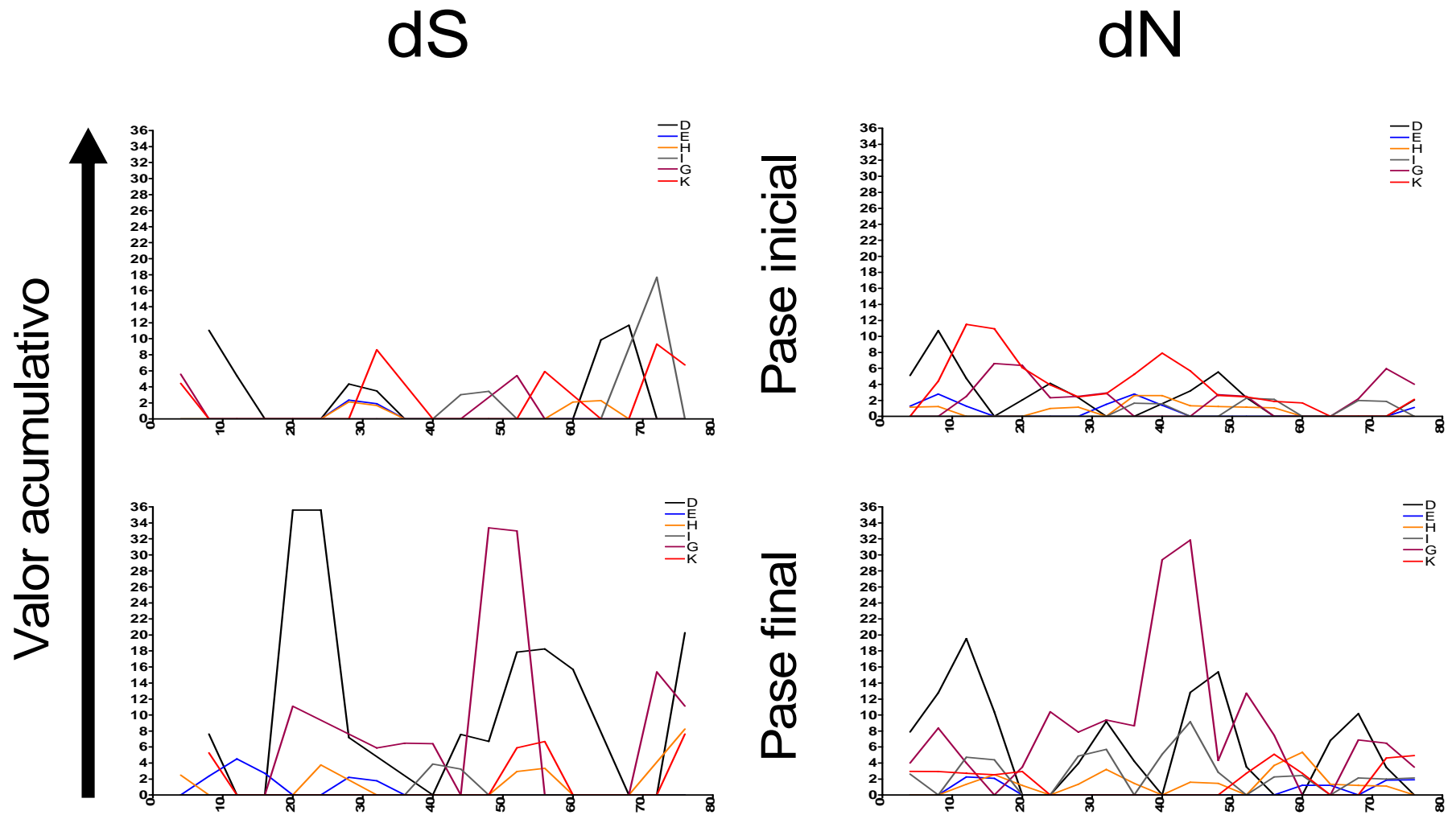
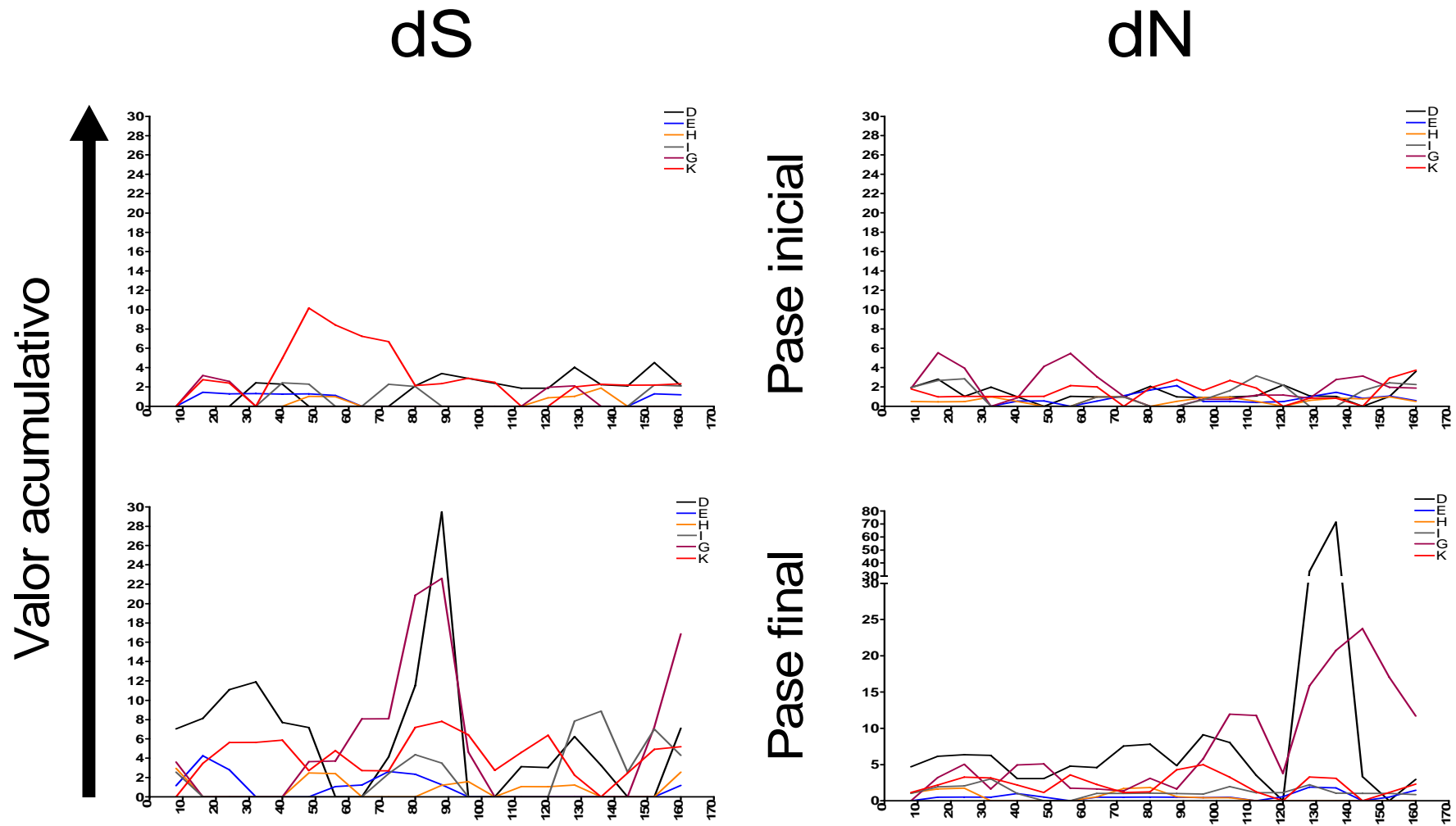


Figura 3. 28. Distribución de los valores acumulativos de las sustituciones sinónimas (izquierda) y no sinónimas (derecha) en la cuasiespecie inicial y final de la proteína vpu. El eje de abscisas representa la región analizada en aminoácidos por la ventana deslizando.

3. 3. 8. 3.- Región gen *env* (V1-V2).

En esta región analizamos el comienzo del gen *env*. En él se encuentran dos motivos importantes: el motivo V1, del aminoácido 130 al 154 (ver Figura 3.29), y el motivo V2, del aminoácido número 155 al 169 (los 14 primeros aminoácidos de los 44 que tiene V2). Comparamos el pase inicial y final de cada uno de los linajes para cada una de las sustituciones. Para las sustituciones sinónimas, en dos linajes se observaron diferencias significativas entre los pases, el D,  $p=0.013$  y el G,  $p=0.011$ . El resto de linajes no tenían diferencias estadísticamente significativas. Para las sustituciones no sinónimas, encontramos tres linajes significativamente distintos, el D,  $p<0.0001$ , el G,  $p=0.002$  y el K,  $p=0.028$ . El resto de linajes no mostraron diferencias estadísticamente significativas.

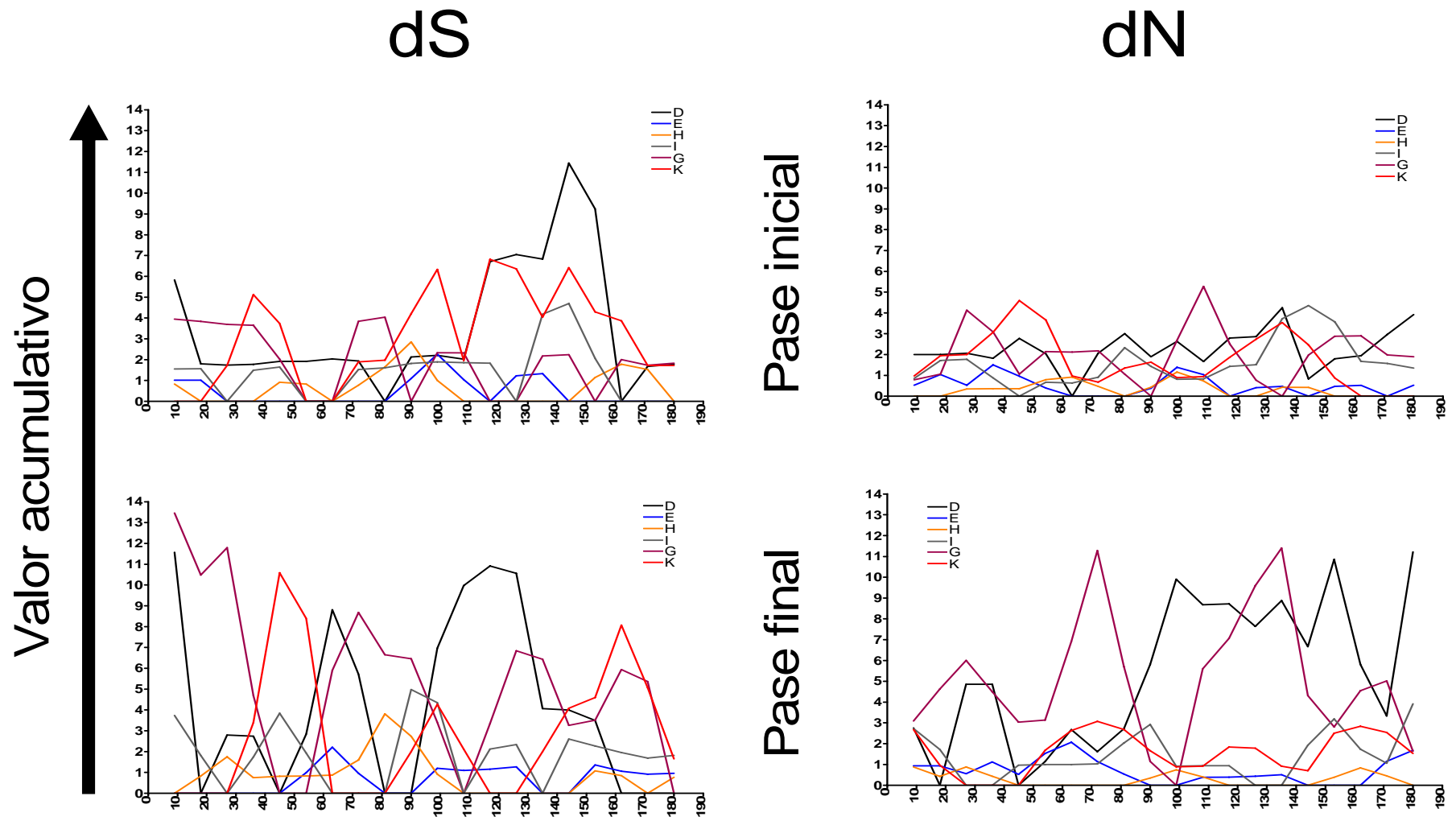




**Figura 3. 29.** Distribución de los valores acumulativos de las sustituciones sinónimas (izquierda) y no sinónimas (derecha) en la cuasiespecie inicial y final de la proteína env (V1-V2). El eje de abscisas representa la región analizada en aminoácidos por la ventana deslizante.

3. 3. 8. 4.- Región gen *env* (V3-V4).

En este fragmento del gen *env* tenemos dos motivos importantes: el motivo V3, de la posición aminoacídica 1 a la 59 (ver Figura 3.30), y el motivo V4, del aminoácido 114 al 142. Al comparar las sustituciones sinónimas entre el pase inicial y final de recuperación de eficacia biológica, un linajes tenía diferencias significativas entre los pases de recuperación de eficacia biológica, el G,  $p=0.003$ . En el resto de linajes no se observaron cambios significativos. Sólo dos linajes fueron significativamente distintos al comparar los pases en las sustituciones no sinónimas, el D,  $p=0.006$  y el G,  $p<0.0001$ . El resto no presentaban diferencias estadísticamente significativos.



**Figura 3. 30.** Distribución de los valores acumulativos de las sustituciones sinónimas (izquierda) y no sinónimas (derecha) en la cuasiespecie inicial y final de la proteína env (V3-V4). El eje de abscisas representa la región analizada en aminoácidos por la ventana deslizante.

3. 3. 9.- Aparición de deleciones en la región V1-V2 del gen de la envuelta después de los pases de recuperación.

Al realizar la clonación molecular para obtener las cuasiespecies virales, detectamos que en el pase final de recuperación de eficacia biológica de varios de los clones biológicos habían aparecido deleciones en la región V1-V2. Esto se correspondía con lo observado en los geles de acrilamida (ver apartado 3.1.3). En total detectamos 6 deleciones distintas de tamaño variable pero localizadas en una región muy concreta del genoma (ver Tabla 3.6).

Clon	Tamaño deleción (nt)	Localización	Frecuencia en la cuasiespecie (%)
D1p11	0	NA	5
	33	6625-6657	95
D15p11	0	NA	63
	12	6639-6650	11
	3	6630-6632	5
	12	6639-6650	
	33	6625-6657	21
D2p11	0	NA	22
	33	6625-6657	67
	42	6621-6662	11
G1p11	0	NA	70
	9	6637-6645	30
G15p11	0	NA	79
	9	6637-6645	16
	15	6637-6651	5

**Tabla 3.6.** Descripción de las distintas deleciones encontradas en la región V1-V2 del gen de la envuelta. Sólo se muestran los linajes y clones con deleciones. La localización de las deleciones según el virus de referencia HXB2 (este virus tiene una deleción también en esta región de la posición 6630 a la 6653 comparado con nuestro virus parental s61). NA, no aplicable.

Es importante destacar que para los dos linajes (D y G) en los que aparecieron deleciones parece haber una deleción mayoritaria para cada uno de ellos: la deleción de 33 nucleótidos para los clones D, y la deleción de 9 nucleótidos para los clones G. Esto ocurrió a pesar de que en ambos casos los clones fueron pasados seriadamente de manera independiente, y podría indicar la importancia de estas deleciones para el aumento de la eficacia biológica en estos clones virales.

## **4.- DISCUSIÓN**

En la presente Tesis Doctoral se han estudiado fluctuaciones poblacionales del VIH-1 *in vitro* y el efecto que éstas tienen sobre la generación de variabilidad genética, la acumulación de mutaciones en el genoma consenso, así como en la eficacia biológica viral. El material de partida fue una serie de clones biológicos que habían sido sometidos a 15 pases en cuellos de botella (en el caso de los clones E, G y H, respectivamente 12, 7 y 13 pases, (Yuste y cols 1999)). Debido a los cuellos de botella, todos estos clones habían visto reducida su eficacia biológica. La causa de ello, al igual que ya se había observado en otros virus de ARN (Chao 1990; Duarte y cols 1992; Clarke y cols 1993; Yuste y cols 1999) había sido la acumulación por deriva genética de mutaciones deletéreas, produciéndose el llamado trinquete de Muller (Muller 1964). El estudio de este mecanismo en el VIH-1 por Yuste y cols (1999), mostró que la disminución de la eficacia biológica para este virus era más drástica que en otros virus de ARN, dado que en el VIH-1 se había llegado en algunos casos a la extinción (Yuste y cols 1999). Asimismo, en estudios con virus de ARN, se había observado que al someter a virus que habían perdido eficacia biológica a pases seriados en gran tamaño poblacional, estos aumentaban su eficacia biológica (Novella y cols 1995; Elena y cols 1998; Escarmis y cols 1999). La ganancia de eficacia se explicaba por el aumento en el tamaño poblacional viral que podía permitir la generación de nueva variabilidad, suficiente para que los virus pudieran acumular mutaciones compensatorias (y reversiones aunque en menor proporción, (Escarmis y cols 1999; Novella y cols 2004)).

En nuestro trabajo anterior (Yuste y cols 2005), usando virus previamente debilitados por pases en cuello de botella, se demostró el aumento de la eficacia biológica del VIH-1 cuando éste era pasado seriadamente en gran tamaño poblacional. Se observó que pocas mutaciones en los genomas consenso eran las responsables de la recuperación de la eficacia biológica de los distintos clones biológicos estudiados, encontrándose una preferencia en la acumulación de mutaciones en la región 5' del genoma viral, y además una alta tasa de reversiones (25%). Algunas de estas mutaciones parecían ser convergencias evolutivas al darse en más de un linaje, aunque los pases de recuperación se habían realizado de manera independiente. Este hecho unido a la falta de información relativa a lo que pudiera estar pasando en la cuasiespecie, y la importancia que puede tener el entender cómo las terapias antiretrovirales y las transmisiones pueden modelar la posterior evolución del VIH-1 *in vivo* es lo que inspiró la realización de esta Tesis Doctoral.

#### **4.1.- Aumento de eficacia biológica, título viral y niveles de la proteína p24.**

El número de clones biológicos con eficacia biológica disminuida analizados en este trabajo fue de 10, procedentes de 6 linajes diferentes (D, E, G, H, I y K). De éstos clones iniciales se derivaron 12 clones finales obtenidos después de 11 pases seriados a una m.d.i. de 0.001. Al final de los pases seriados, se comprobó que se había producido un aumento de eficacia biológica entre el pase inicial y final ( $p=0.042$ ). Este dato corroboraba el aumento de eficacia descrito por Yuste y cols (2005) tras los pases seriados. Sin embargo, el aumento de eficacia biológica no había sido similar para todos los clones biológicos, encontrando diferentes incrementos de eficacia biológica tanto entre clones biológicos pertenecientes distintos linajes virales como a un mismo linaje viral (ver Tabla 4.1). El clon G1p11 había aumentado significativamente eficacia biológica mientras que otras dos replicas realizadas del linaje G no aumentaban significativamente eficacia biológica (G15p11 y G2p11). En los linajes I y K, ambas replicas no mostraron un aumento significativo de eficacia biológica, aunque en el caso del linaje K, el clon K1p11 había aumentado en un 26% su eficacia biológica frente al no aumento del K2p11. De la misma manera, aunque no pudimos cuantificar el aumento de eficacia del linaje D, observamos que todas las replicas de este linaje estaban aumentando eficacia de manera similar (ver apartado 3.2.3.1). En cuanto a las posibles diferencias debido al número de células usado para los pases seriados, observamos que de los 4 clones mantenidos en  $5 \times 10^6$  de células, 3 de ellos aumentaban eficacia significativamente. De estos en el caso del linaje D y G teníamos replicas con pases en  $2.5 \times 10^6$  de células. En el linaje D, no parecía afectar el número de células usado para los pases ( $5$  o  $2.5 \times 10^6$  de células) dado que en ambos casos detectamos aumento de eficacia, mientras que en el linaje G, teníamos que para los pases en  $2.5 \times 10^6$  de células un clon aumentaba eficacia (G1p11) y otro se mantenía sin un aumento apreciable de eficacia (G2p11) al igual que para el clon pasado en  $5 \times 10^6$  de células (G15p11). Por tanto, estos resultados ampliaban el trabajo realizado previamente por Yuste y cols (2005) en cuanto al análisis del aumento de eficacia biológica. Esta variabilidad encontrada en la eficacia biológica se podría explicar por la hipótesis de los paisajes adaptativos, ya que indicaría que cada uno de los distintos linajes y réplicas por deriva se situarían en a distintos puntos del paisaje, y por tanto el aumento de eficacia no tendría por que ser similar para todos ellos. Además, tenemos que tener en cuenta que la generación de mutaciones es un proceso aleatorio y por tanto no tendría por qué suceder de la misma manera en los distintos clones estudiados.



Además de la eficacia biológica relativa, otros parámetros fenotípicos importantes del VIH-1 son el título viral (TCID<sub>50</sub>/ml) y los niveles de la proteína p24 viral (pg/ml). El título viral, según la metodología usada para medirlo, nos proporciona una estimación del número de partículas virales capaces de producir una infección. Por el contrario, la medida por ECLIA de los niveles de la proteína p24 nos proporciona una estimación indirecta del número de partículas virales que tenemos independientemente de si éstas son infecciosas o no, y es por tanto el valor total de partículas que un determinado linaje genera. Según (Marozsan y cols 2004), ni el valor de p24 ni el título viral, son buenas estimaciones de la cantidad de virus infecciosos que tenemos en una preparación. Esto es debido a que estas mediciones se pueden ver influenciadas por la diversidad de secuencia que tenemos entre los distintos subtipos de VIH-1 (entre el 15 y el 30%). Según ellos, la mejor estimación sería la cuantificación de la actividad de la retrotranscriptasa viral, la cual se ve menos afectada por la variabilidad en la secuencia genómica. Aunque este estudio es importante a la hora de comparar parámetros genotípicos entre distintos subtipos, en nuestro caso, no creemos que sea determinante este efecto dado que no encontramos grandes diferencias en la variabilidad entre los distintos linajes estudiados (diferencia promedio del 1%). Según nuestros resultados, el título viral de los clones biológicos aumentaba con un valor estadísticamente significativo ( $p=0.009$ ) en el pase final con respecto al pase inicial de recuperación de eficacia biológica, mientras que los niveles de la proteína p24 no variaba significativamente entre pases ( $p=0.96$ ). Para relacionar ambos valores, calculamos la proporción entre título viral y los niveles de la proteína p24, y comparamos por el test estadístico de Kruskal-Wallis el pase inicial frente al final. La proporción aumentaba en el pase final con respecto al inicial con un  $p$ -valor=0.01, indicando por tanto, un aumento en el número de virus infecciosos. Otro dato que corroboraba el aumento de virus infecciosos era la disminución en el número de clones con codón de terminación después de los 11 pases seriados en las cuasiespecies virales. Esto era especialmente aparente en la región leader/gag ( $p=0.01$ ). La aparición de clones defectivos se ha visto asociado a un bajo valor de título viral (Bernier y cols 1995), ya que éstos pueden limitar la infección al competir con los virus capaces de producir una infección productiva, y disminuir el número de células susceptibles a infección. Nuestros resultados apuntan a que entre el pase inicial y final de recuperación de eficacia biológica, se está seleccionando negativamente de la población a aquellos viriones incapaces de infectar eficaz y productivamente, de manera que esto debería favorecer el

aumento de la eficacia viral. Por tanto, permitiendo al VIH-1 replicar en condiciones de mayor tamaño poblacional se favorece que el virus recupere su capacidad de generar partículas infecciosas y por tanto recuperar su eficacia biológica.

#### 4.2.- Variación en los genomas consenso virales tras los pases seriados.

Además de los parámetros estudiados en el apartado anterior, al estudiar las mutaciones fijadas en los genomas consenso de los clones virales observamos que la recuperación de eficacia biológica estaba mediada por la fijación de pocas mutaciones. Encontramos 20 cambios al comparar los genomas consenso iniciales y finales de los pases seriados de los clones virales (ver Tabla 3.3), siendo el promedio encontrado de mutaciones por clon viral de 1.6. Este dato era inferior al encontrado en el trabajo previo (Yuste y cols 2005) en donde la media era de 5 mutaciones por clon. Del total de mutaciones, un 15% eran reversiones mientras que en el trabajo previo hallamos un 25%. Estos datos indican que las reversiones parecen tener un papel importante a la hora de recuperar eficacia en el VIH-1. Este efecto también se ha detectado en otros virus de ARN (Burch y cols 1999; Escarmis y cols 1999; Novella y cols 2004). Detectamos también una acumulación de mutaciones en la región no codificante 5'LTR del genoma viral, al igual que en el trabajo de Yuste y cols (2005), con un total de 5 mutaciones, el 25% del total de mutaciones encontradas. Estos resultados resaltan la importancia de esta región a la hora de acumular mutaciones *in vitro*, teniendo en cuenta que esta región es solamente un 8% del total del genoma viral. El resto de las 15 mutaciones que observamos tras los pases seriados de los clones se localizaron en las regiones codificantes del genoma viral.

Tres de las cinco mutaciones encontradas en la región 5'LTR parecen ser una convergencia que se estaba produciendo en los 3 clones finales del linaje D (D1p11, D15p11 y D2p11, mutación T593G). Debido a que estos pases eran totalmente independientes y teniendo en cuenta que todo el trabajo había sido llevado a cabo para minimizar posibles contaminaciones, esto podría resaltar la importancia que la región leader parece tener a la hora de aumentar la eficacia biológica *in vitro* al menos en el linaje D. La LTR 5' es una región en donde se regulan muchas de las funciones de replicación y transcripción del virus (Verdin y cols 1995; Al-Harhi y cols 1998; Roebuck y cols 1999; Pereira y cols 2000; Beerens y cols 2002; Marzio y cols 2002; van Opijnen y cols 2004). Esta mutación se encuentra muy cercana al motivo PBS formando parte de una de las zonas del lazo que ayudan a estabilizar esta estructura. Esta mutación

debe ser determinante si se está fijando en estos linajes, dado que el mismo cambio ya fue detectado también en el trabajo previo (Yuste y cols 2005). Todo parece indicar que de alguna manera esta mutación podría estar mediando el aumento de eficacia. También pensamos que las deleciones producidas en el gen de la envuelta después de los pasajes seriados (ver Tabla 3.6) podrían estar mediando la recuperación de eficacia, ya que se ha descrito que deleciones en esta región facilitan el reconocimiento del receptor CD4 por parte de la gp120 (Imamichi y cols 2002). En el trabajo previo de Yuste y cols (2005) se detectó esta misma deleción en el linaje D, por tanto esto también parece ser específico de este linaje a la hora de recuperar eficacia biológica, indicando que en algunos casos las posibilidades para compensar están limitadas (Cuevas y cols 2002; Novella y cols 2004).

En el clon viral E1p11, los cambios en el genoma consenso se localizaron mayoritariamente en el gen *rev*. Se sabe que este gen media el transporte de los mensajeros virales del núcleo al citoplasma (Meyer y cols 1994), además de que su secuencia está muy conservada. Encontramos que la mutación G8499A no se hallaba en la base de datos de los Álamos (Leitner y cols 2005) siendo el nucleótido G el que aparecía en todos los casos. También observamos que la mutación T8651C retardaba la finalización de la traducción de la proteína en 3 aminoácidos, generando por tanto una proteína mayor. Aunque no conocemos más datos sobre la posible compensación que producen todas estas mutaciones en el linaje E para el aumento de eficacia, es de resaltar que la mitad de las mutaciones de este clon se localicen en el gen *rev*.

En el clon G15p11, la mayoría (75%) de las mutaciones se localizaron en el gen *pol*, y estos cambios fueron por los nucleótidos más conservados en las bases de datos (siendo la mutación A3129G una reversión) a excepción de la T4421A, en el nucleótido T era el mayoritario. La última mutación de este clon viral (A8402G), afectaba a dos genes reguladores (*rev* y *tat*), y el nucleótido más conservado era A. Tanto en el clon G1p11 como el G15p11, se detectaron deleciones en la región de los dominios V1-V2. Estas deleciones tal como ya hemos descrito en el caso del linaje D podrían estar favoreciendo un aumento de la eficacia biológica. En el clon viral G2p11, todas las mutaciones se produjeron en el gen *pol*, y solamente la mutación G3146A era de baja ocurrencia en la base de datos. Lo interesante de estos cambios es que las mutaciones se han producido en codones contiguos de manera que se han modificado la primera y tercera posición para poder realizar el cambio de aminoácido encontrado. Esto podría

indicar que estos cambios, por la baja probabilidad de la fijación de mutaciones tan cercanas, estarían posiblemente implicados en el aumento de eficacia biológica.

En el clon viral H1p11, otra mutación que parece interesante es la mutación A7758G, dado que estaba localizada en el primer aminoácido de la gp41 contiguo de la secuencia aminoacídica específica (REKRT → REKRA, subrayada la secuencia específica y en negrita el cambio encontrado) implicada en el procesamiento proteolítico de la glicoproteína gp160 para dar lugar a las proteínas gp120 y gp41. En un estudio previo (Adams y cols 2000) destacan que la mutación a aminoácidos distintos a la alanina (A) resultan en una menor generación de la proteína gp120, y también aunque en menor medida de la proteína gp41, y en una formación retrasada de sincitios. Además esta mutación es una reversión, por lo cual todo parece indicar que este cambio debe ser importante a la hora de recuperar la eficacia biológica del clon viral. En este mismo clon, detectamos también que se recuperaba la pauta de lectura del gen *nef*, que aunque se sabe es dispensable *in vitro* (Anderson y cols 2003), parece ser que su presencia favorece la replicación de los virus pudiendo estar mediando también de alguna manera en la recuperación de eficacia biológica (Joseph y cols 2005).

Como se ha podido comprobar, cada linaje está fijando mutaciones en zonas genómicas distintas, y por tanto recorriendo de manera diferente el camino en el paisaje adaptativo para recuperar su eficacia. El linaje G por ejemplo parece necesitar fijar mutaciones en el gen *pol* para recuperar eficacia dado que de las 8 mutaciones de los dos clones independientes de este linaje 7 son en *pol*, indicando que este linaje necesitaría mutaciones compensatorias en el genoma *pol*, que es el que parece mediar su recuperación de eficacia biológica. Algo similar podríamos decir del linaje D, en donde, como ya hemos descrito, la misma mutación se ha dado en las 3 réplicas finales en la secuencia leader, aparte de las deleciones en la región V1-V2 del gen de la envuelta. El linaje E, sin embargo parece requerir del gen *rev*, junto con algunos cambios en la LTR 5'. Mientras que el linaje H, fija una reversión en la zona de procesamiento de la gp160, y recupera la pauta de lectura de *nef* para recuperar eficacia biológica.

Un resultado interesante fue encontrar que 5 clones (G1p11, I1p11, I5p11, K1p11 y K2p11) después de los pases seriados no presentaban cambios en su genoma consenso. Cuatro de ellos, I1p11, I5p11, K1p11 y K2p11 no habían recuperado eficacia biológica significativamente. Por el contrario, el clon viral G1p11 estaba recuperando eficacia biológica sin fijar ninguna mutación. En el VSV se ha descrito algo similar después de un proceso de recuperación de eficacia biológica (Novella y cols 2004).

Según este estudio, 9 de los 13 virus que habían recuperado eficacia biológica lo hacían sin haber modificado su genoma consenso entre los países seriados. La invariabilidad en los genomas consenso ya había sido descrita previamente en Arenavirus (Grande-Perez y cols 2005) y en Aphthovirus (Gonzalez-Lopez y cols 2005), aunque en estos casos no se buscaba una recuperación de la eficacia biológica sino que estaban forzando a los virus por medio de un mutágeno a acercarse a la catástrofe de error.

Nuestros resultados junto con lo ya descrito en la literatura, nos indican que se pueden producir cambios en la eficacia biológica, dado que todos los virus incrementaron eficacia biológica, de una población viral fijando muy pocas mutaciones e incluso sin fijar cambios en el genoma consenso de estas poblaciones. Esto nos llevo a pensar que otros parámetros podrían jugar un papel importante en la recuperación de la eficacia biológica, entre ellos la heterogeneidad de la cuasiespecie viral. En este sentido hay que tener en cuenta que sobre la variabilidad la selección natural puede actuar para seleccionar mutaciones, aunque posteriormente el proceso de fijación pueda ser lento debido a que la mutación debe de alcanzar una frecuencia suficiente en la población para que esto ocurra (Wilke 2003), aunque también podría depender de la eficacia biológica asociada.

#### **4.3.- Posible efecto de la heterogeneidad de la población viral en la eficacia biológica.**

Según estudios recientes, la heterogeneidad de la cuasiespecie viral puede tener profundos efectos fenotípicos, como por ejemplo en el virus de las paperas o polio (Sauder y cols 2006; Vignuzzi y cols 2006), así como sobre la patogenicidad. Esto explicaría cómo, a pesar de tener pocos o ningún cambio en los genomas consenso de los distintos linajes estudiados, observamos un aumento de su eficacia biológica. Para determinar si la heterogeneidad de la cuasiespecie puede ser un factor responsable del aumento de eficacia en nuestro modelo *in vitro*, tratamos de estimar la heterogeneidad de las poblaciones virales obtenidas tras los cuellos de botella, en donde habíamos observado una reducción de la eficacia viral, y en las poblaciones virales resultantes después de someter a éstas a países seriados. Para ello, estudiamos 4 regiones genómicas distintas, LTR/gag(p17/p24), *vpu*, *env*(V1-V2) y *env*(V3-V4) por el método de clonación molecular. Se había demostrado, que el método de clonación molecular para obtener la estructura poblacional de un virus daba resultados equivalentes a la clonación biológica (Arias y cols 2001), con la ventaja de que de esta manera podíamos detectar

clones con codones de terminación. Obtuvimos así una representación de las poblaciones virales de cada linaje viral (con sus réplicas) y pase. Al comparar todos los clones moleculares obtenidos de todos los linajes no encontramos diferencias significativas entre estos ( $p=0.4$ ). Este resultado era importante porque nos permitía agrupar todos los datos de heterogeneidad de los distintos linajes para realizar comparaciones. Al analizar si existían diferencias significativas entre pases, detectamos una tendencia al aumento de la heterogeneidad de las poblaciones virales tras los pases seriados ( $p=0.06$ ). En nuestros experimentos, al permitir a los virus replicar en tamaños poblacionales mayores (pases seriados) estábamos permitiendo el restablecimiento de la heterogeneidad perdida tras la transmisión (cuellos de botella). Estos resultados de heterogeneidad obtenidos en nuestro modelo *in vitro* pueden tener un paralelo en varias situaciones *in vivo*, por ejemplo se correlacionan con estudios previos que analizaban tanto transmisiones horizontales (Cichutek y cols 1992; McNearney y cols 1992; Pang y cols 1992; Wolfs y cols 1992; Zhang y cols 1993; Zhu y cols 1993; Delwart y cols 2002) como en transmisiones verticales (Wolinsky y cols 1992; Ahmad y cols 1995), en las que se había observado la homogenización de la población viral durante la transmisión. Esta pérdida podía hasta del 99% de la heterogeneidad total después de un evento de transmisión horizontal en VIH-1 ((Edwards y cols 2006) analizando los genes *env* y *gag*. Otro factor que produce *in vivo* la disminución de la variabilidad del VIH-1 es la terapia antiretroviral denominada HAART (highly active anti-retroviral therapy) frente al VIH-1. Esta terapia consiste en una mezcla de 3 ó más drogas, normalmente un inhibidor de la proteasa viral y dos análogos o no análogos de nucleósido que tienen como diana la retrotranscriptasa viral. Durante la terapia HAART se produce una reducción muy drástica en los tamaños poblacionales del VIH-1, llegando a estar por debajo del nivel de detección, pasando de  $10^{8-11}$  a menos de 50 copias/ml en sangre, (Saag 2001). Aunque la terapia haga indetectable la carga viral en sangre, el virus es capaz de acantonarse en reservorios (Crowe y cols 2000; Saksena y cols 2003). De esta manera, el virus seguirá replicándose aunque con unas tasas muy bajas limitando de esta manera la capacidad del mismo de generar variabilidad y evolucionar. Según Shankarappa y cols (1999), la variabilidad genética del VIH-1 aumenta durante el curso natural de la infección desde un punto inicial homogéneo justo después de la transmisión. Estos autores distinguen 5 etapas en la infección normal por parte del VIH-1 desde el momento de la primoinfección hasta el desarrollo de SIDA. Durante estas etapas se produce un aumento de la divergencia con respecto al virus inicial, y un

aumento de la varianza genética. Esta dinámica cambia en las 2 últimas etapas descritas por ellos, en donde se produce un estancamiento de la divergencia y la diversidad, posiblemente por el desarrollo de SIDA. La compartimentalización a la que el VIH-1 se ve sometido también podría ser un factor importante a la hora de generar mayor variabilidad, dado que en las poblaciones aisladas se estaría produciendo una evolución independiente al resto de la población (Sanjuan *y cols* 2004; Borderia *y cols* 2007), de manera que cuando estas variantes recirculasen se produciría un aumento general en la heterogeneidad de la cuasiespecie en el huésped.

En nuestro modelo *in vitro* aunque estamos estudiando la primera etapa de generación de variabilidad, como se refleja en que el número de mutaciones que estamos observando en el análisis de heterogeneidad no era muy elevado, desde un número de diferencias medio de 1.47 en el pase inicial a 1.74 en el final, podemos observar que existe una tendencia al aumento de la heterogeneidad sobre todo en la región de la envuelta en V1 y V2, y en la región de la secuencia leader y la proteína p17. Además de la heterogeneidad de la población estudiamos los valores de Sd y Nd de las distintas regiones analizadas. Encontramos diferencias entre ellas, siendo la región con más sustituciones la *env* (V1-V2), seguida de *gag* (p17/p24), *env* (V3-V4), y finalmente el gen *vpu*, que no variaba prácticamente después de los pases seriados. Esto indicaría que el gen *vpu* no parece estar bajo fuerzas selectivas importantes. Sin embargo la región de *gag* (p17) y *env* (V1-V2), serían más importantes en la recuperación de eficacia biológica.

Dado que en las cuasiespecies virales la selección actúa sobre la población en general y no simplemente sobre los distintos individuos que la componen, esto explicaría por qué un aumento de la heterogeneidad podría estar asociado a un aumento en la eficacia biológica. Esto apoyaría la teoría de las cuasiespecies en la evolución en virus de ARN (Domingo *y cols* 1985) y que había sido postulada anteriormente por Manfred Eigen (Eigen 1971). En el siguiente apartado analizaremos en más detalle esta posibilidad (ver Tabla 4.1 para un resumen de los parámetros analizados).

#### **4.4.- Relación de la heterogeneidad, Sd, Nd, eficacia biológica y la proporción de dN/dS.**

Para entender cuáles eran los tipos de sustituciones predominantes durante la generación de la variabilidad, realizamos correlaciones entre la heterogeneidad y las Sd y Nd, tanto en el pase inicial como en el final. Nuestros resultados muestran que en el

pase inicial sólo existía una correlación positiva significativa con Nd ( $p=0.001$ ) pero no con las Sd, lo que indica que en los pases iniciales la heterogeneidad generada era debida mayoritariamente a las sustituciones no sinónimas. Por el contrario, en el pase final de recuperación de eficacia, detectamos que tanto Sd ( $p=0.0006$ ) como Nd ( $p<0.0001$ ) correlacionaban positivamente con el aumento de la heterogeneidad, por tanto el aumento de heterogeneidad encontrado en el pase final estaría relacionado directamente tanto con un aumento de las sustituciones sinónimas y no sinónimas. Todo esto nos indicaba que en el pase inicial que provenía de los cuellos de botella, la cuasiespecie estaba afectada por la deriva genética y por ello teníamos una mayoría de mutaciones no sinónimas. Por el contrario cuando mantenemos la población en pases seriados en gran tamaño poblacional y permitimos la generación de una mayor variabilidad, observamos que se estaba produciendo tanto Sd como Nd, indicando que las poblaciones ya no se encontraban sometidas sobre todo a deriva genética, sino que por tanto la selección debía estar ahora modelando la evolución de estas poblaciones virales. Para entender mejor como la selección estaba actuando sobre las distintas poblaciones obtuvimos los valores de dN y dS. Con estos valores calculamos la proporción dN/dS que nos indicaría qué sentido tiene la evolución. Aunque no había diferencias estadísticamente significativas entre pases globalmente y entre pases por regiones ( $p>0.05$ ), al analizar si el cociente era menor a 1 en el pase inicial, detectamos que no era significativamente distinto ( $p=0.139$ ), mientras que el pase final aunque seguía sin ser estadísticamente significativo ( $p=0.06$ ) estaba cerca de un valor significativo, que sugería una tendencia hacia una selección purificadora. Este tipo de selección había sido ya detectada previamente para el VIH-1 (Brown y cols 1988; Li y cols 1988; Gojobori y cols 1990; Seibert y cols 1995; Kils-Hutten y cols 2001; Smith y cols 2004; Edwards y cols 2006). Además el dato de tendencia hacia la selección purificadora, corroboraría la pérdida que observamos de clones con codones de terminación en nuestros análisis, y la ganancia de título sin variación de los niveles de la proteína p24 que habíamos observado (discutido en apartados anteriores).

Para estudiar si existía una relación entre el aumento de heterogeneidad con el aumento de eficacia biológica, realizamos una correlación entre heterogeneidad y eficacia biológica, tanto en el pase inicial como en el final. Aunque no existía una correlación estadísticamente significativa entre ellas ( $p=0.75$  y  $p=0.18$ , respectivamente) sí que parecía haber una tendencia al aumento de la eficacia viral y más si teníamos en cuenta que ninguno de los clones del linaje D pudieron ser incluidos (aunque teníamos



evidencias claras de que estaban aumentando su eficacia biológica con los pases). Esto indica que a valores elevados de heterogeneidad tenemos valores elevados de eficacia, lo cual tiene relación con el hecho de que en las cuasiespecies la selección no se realiza sobre un único individuo sino que mas bien sobre toda la cuasiespecie en general, de manera que al aumentar la complejidad favorecemos la aparición de variantes con mayor eficacia biológica que producen un aumento general de la eficacia biológica de la cuasiespecie como conjunto (Domingo y cols 1998; Domingo y cols 2006).

#### **4.5.- Acumulación de mutaciones en la población viral tras los pases seriados.**

Un hecho importante de la generación de la variabilidad es saber de qué manera esta variabilidad se está distribuyendo en la población. Durante la generación de nueva variabilidad y aparición de nuevas mutaciones, éstas se pueden distribuir en principio de dos maneras. Por un lado, todos los distintos genomas están mutando y fijando mutaciones por igual, y por tanto pasaríamos de una predominancia de genomas con 0 mutaciones (iguales al consenso de la cuasiespecie) a una población con predominancia de genomas con valores de 1 mutación en general. La segunda posibilidad, sin embargo, es que las mutaciones se estén fijando en genomas específicos de la cuasiespecie que ya tienen mutaciones previas, de manera que por un lado tendríamos una distribución bimodal de genomas con pocas mutaciones (0 ó 1) y otros pocos genomas con más de 2 mutaciones. Con cualquiera de estas dos posibilidades, al calcular la heterogeneidad poblacional no seríamos capaces de detectar cuál de las dos tendencias es la correcta, ya que al obtener un valor promedio de las diferencias entre genomas, éste sería igual en los dos casos propuestos. Para dilucidar esta cuestión, estudiamos para cada uno de los linajes, región y pase, el número de mutaciones que tenía cada uno de los clones moleculares con respecto a su secuencia consenso. De esta manera tabulamos la proporción de genomas con 0, 1, 2, 3, 4 ó más mutaciones (ver Figura 3.20). Dado que no existían diferencias entre las distintas regiones y linajes en cuanto a la distribución de genomas ( $p > 0.05$ ), agrupamos todas las proporciones entre pase inicial y final de recuperación de eficacia biológica. De esta manera, detectamos que entre los pases se estaba produciendo una disminución de los genomas con 0 mutaciones ( $p = 0.044$ ) y un aumento de los genomas con 2 mutaciones ( $p = 0.031$ ), mientras que las categorías con 1, 3, 4 ó más mutaciones permanecían estables. Por ello, parece que las mutaciones se están acumulando de una manera aleatoria. No detectamos una acumulación específica de mutaciones en genomas concretos de la cuasiespecie, sino que todos los genomas

secuenciados de la cuasiespecie parecen estar acumulando mutaciones durante los pases seriados.

#### **4.6.- Deriva genética y selección en la cuasiespecie viral.**

En los pases en cuello de botella la evolución de la población viral está modelada por la deriva genética debido a que los tamaños poblacionales son pequeños. En contraposición, en los pases seriados en los que los tamaños poblaciones son elevados y en donde la selección puede actuar para modelar la evolución de la población viral (Rouzine y cols 2001). Según este planteamiento, en los pases iniciales deberíamos ser capaces de detectar la aleatoriedad de la fijación de las mutaciones encontradas en los distintos clones moleculares, mientras que en los pases finales debería ser no aleatoria. Para determinarlo, ajustamos el número de mutaciones por clon molecular obtenido previamente a una distribución Poisson. La distribución Poisson expresa la probabilidad de que un número de eventos ocurran en un tiempo fijo si estos eventos ocurren con una tasa media conocida, y son independientes del tiempo desde el último evento. Si la acumulación de mutaciones en los clones moleculares es debida a la deriva genética, el número de mutaciones (eventos) se mantendría constante y se ajustarán a la distribución. Si por el contrario es debida a la selección, los eventos no se ajustarán a la distribución, dado que aunque el número de mutaciones generado debería mantenerse, no así la fijación de éstas. Según esto, los resultados obtenidos no nos permitían descartar que para los pases iniciales las mutaciones se estuvieran acumulando según una distribución Poisson con  $\lambda_i=0.97$  ( $p=0.37$ ). Y lo mismo ocurría para los pases finales ( $p=0.44$ ), aunque se observaba un cambio a mayor en el parámetro de la distribución Poisson  $\lambda_f=1.2$ . Aunque en apartados anteriores habíamos detectado una tendencia hacia una selección purificadora, no habíamos podido descartar neutralidad en la selección, de manera que las mutaciones parecían acumularse en los distintos genomas de una manera aleatoria, por tanto el resultado obtenido con el ajuste a la distribución Poisson nos señala en la misma dirección. Asimismo, no detectamos un aumento en la acumulación de mutaciones en genomas concretos de la población, sino que las mutaciones generadas se estaban distribuyendo de manera similar por toda la población (ver apartado 3.3.6).

#### 4.7.- Análisis de los sitios variables y conservados durante los pases seriados.

Una vez determinada la existencia de clones moleculares que estaban acumulando más sustituciones comparados con otros clones, estudiamos si en las distintas regiones estudiadas, existían diferencias en el número de posiciones nucleotídicas variables o conservadas. Observamos que en todas las regiones a excepción de *vpu*, se estaba produciendo un aumento de regiones variables y una disminución de las regiones conservadas al realizar los pases seriados (ver Tabla 3.5). La región en donde este cambio era más acusado era en el gen de la envuelta (V1-V2), seguido de p17/p24 y el gen de la envuelta (V3-V4). Este resultado está correlacionado con el aumento de heterogeneidad que observamos tras los pases seriados. El porcentaje de posiciones variables en cada una de las regiones era similar, estando entre el 22 y el 29% en los pases iniciales, y entre el 28 al 34% en los pases finales. Por tanto, según este resultado en general entre un 78 y un 66% serían posiciones conservadas. Es lo que se conoce como un genoma antiredundante (Krakauer y cols 2002), esto es, no es posible realizar cambios sin producir un drástico efecto en la eficacia biológica del virus. En otras palabras, los efectos deletéreos de las mutaciones son mucho más exacerbados que en sistemas más redundantes. En concreto como ya se vió previamente (Yuste y cols 1999), la pérdida de eficacia biológica por transmisión en cuellos de botella fue mucho mayor en el VIH-1 (82%), si lo comparamos con otros virus RNA como el FMDV (Escarmis y cols 1996) en el que la pérdida fue del 35%, ó el bacteriófago  $\phi 6$  (Chao 1990), que perdió 22% de eficacia. Esto se podría atribuir al bajo número de rondas de replicación que se dan durante la formación de placas en el VIH-1 comparado con otros virus (Escarmis y cols 2006).

Asimismo también estudiamos como las sustituciones sinónimas y no sinónimas se estaban distribuyendo en las distintas posiciones aminoacídicas de las regiones estudiadas (Figuras de la 3.27 a la 3.30). Lo primero que se observa es que en general entre el pase inicial y final hay un aumento de sustituciones, tanto sinónimas como no sinónimas (si bien este aumento no es igual para todas las regiones, como ya hemos discutido previamente). De las distintas regiones estudiadas, no parecen existir posiciones en donde se estén acumulando específicamente las sustituciones, sino que se dan por igual a lo largo de toda la secuencia. Aunque hay que destacar que algunos linajes como el D y G parecen estar acumulando una mayor cantidad de cambios con respecto al resto de linajes.

	Eficacia	Consenso <sup>b</sup>	Heterogeneidad <sup>a</sup>			
			Leader/p17	vpu	V1-V2	V3-V4
D1p11		Si (1)				
D15p11		Si (1)				
D2p11		Si (1)				
E1p11		Si (6)				
G1p11		No				
G15p11		Si (4)				
G2p11		Si (4)				
H1p11		Si (3)				
I1p11		No				
I5p11		No				
K1p11		No				
K2p11		No				

**Tabla 4.1.** Resumen de cada uno de los clones implicados en la recuperación de eficacia biológica. <sup>a</sup> El código de colores indica en verde, azul y rojo respectivamente, el aumento, estabilidad o descenso en la heterogeneidad calculada para cada una de las regiones estudiadas. <sup>b</sup> Indica si hay o no cambios en la secuencia consenso entre países y entre paréntesis se detalla el número de mutaciones detectadas.

## **5.- CONCLUSIONES**

Tras estudiar la recuperación de eficacia biológica por pases seriados en poblaciones clónales del VIH-1 podemos concluir lo siguiente:

- 1.- Existe un aumento general de la eficacia biológica en los clones estudiados aunque solo en 6 de ellos esta ganancia era estadísticamente significativa.
- 2.- Al analizar los genomas consenso de los distintos clones antes y después de los pases seriados, se detectó una baja fijación de sustituciones (1.6 sustituciones de media), tan solo 7 clones de 12 fijaron mutaciones. En los 5 clones restantes estudiados no se detectó ninguna sustitución.
- 3.- De los 7 clones que fijaron sustituciones sólo 5 de estos recuperaron a su vez eficacia biológica con valores significativos tras los pases seriados. Esto indicaría que en el resto de casos el aumento de eficacia se podría deber a otros factores.
- 4.- Durante los pases seriados se ha producido un aumento del título viral mientras que los niveles de la proteína p24 generada no variaban. Además se estaba produciendo un descenso en el número de clones moleculares con codones de terminación, sobre todo en la región del gen *gag*, lo cual implica que se está produciendo una eliminación de partículas virales defectivas en la cuasiespecie. Por ello, estos datos indican la existencia de fuerzas purificadoras en la recuperación de la eficacia biológica del VIH-1 *in vitro*.
- 5.- La heterogeneidad no está variando por igual en todo el genoma viral, siendo de las regiones analizadas, el inicio del gen *gag*, la proteína p17 e inicio de la proteína p24, y la región V1-V2 del gen de la envuelta las regiones que mayor aumento de heterogeneidad acumulan, mientras que el gen *vpu* es el que menos varía de todas las regiones analizadas.
- 6.- En algunos casos, linajes D y G, la aparición de deleciones en la región V1-V2 del gen de la envuelta podría estar implicado en el aumento de eficacia biológica.
- 7.- Hemos encontrado una correlación entre la heterogeneidad de la cuasiespecie y las sustituciones no sinónimas en los pases iniciales, y con las sinónimas y no sinónimas en los pases finales.

8.- El cociente entre dN/dS en los pases iniciales es igual a 1 indicando una selección neutral, mientras que tras los pases seriados se observa una tendencia hacia selección purificadora.

9.- Hemos comprobado que en los pases finales parece existir una relación entre la ganancia de heterogeneidad y de eficacia, aunque sin llegar a valores estadísticamente significativos debido al limitado número de clones analizados y por los pocos pases seriados realizados. Estos datos sugieren una selección de la población en su conjunto y no de genomas individuales, apoyando, por tanto apoyan la teoría de las cuasiespecies en la evolución viral.

10.- Existe un marcado componente estocástico en la evolución del VIH-1 *in vitro* como se comprueba por la distinta fijación de mutaciones, generación de variabilidad, y eficacia biológica tanto entre réplicas de diferentes linajes, como dentro de las réplicas de un mismo linaje.

## **BIBLIOGRAFIA**



- (1994). "HIV type 1 variation in World Health Organization-sponsored vaccine evaluation sites: genetic screening, sequence analysis, and preliminary biological characterization of selected viral strains. WHO Network for HIV Isolation and Characterization." AIDS Res Hum Retroviruses **10**(11): 1327-43.
- Adams, O., H. Schaal and A. Scheid** (2000). "Natural variation in the amino acid sequence around the HIV type 1 glycoprotein 160 cleavage site and its effect on cleavability, subunit association, and membrane fusion." AIDS Res Hum Retroviruses **16**(13): 1235-45.
- Ahmad, N., B. M. Baroudy, R. C. Baker and C. Chappey** (1995). "Genetic analysis of human immunodeficiency virus type 1 envelope V3 region isolates from mothers and infants after perinatal transmission." J Virol **69**(2): 1001-12.
- Al-Harathi, L. and K. A. Roebuck** (1998). "Human immunodeficiency virus type-1 transcription: role of the 5'-untranslated leader region (review)." Int J Mol Med **1**(5): 875-81.
- Anastassopoulou, C. G. and L. G. Kostrikis** (2005). "Viral correlates of HIV-1 disease." Curr HIV Res **3**(2): 113-32.
- Anderson, J. L. and T. J. Hope** (2003). "Recent Insights into HIV Accessory Proteins." Curr Infect Dis Rep **5**(5): 439-450.
- Arias, A., E. Lazaro, C. Escarmis and E. Domingo** (2001). "Molecular intermediates of fitness gain of an RNA virus: characterization of a mutant spectrum by biological and molecular cloning." J Gen Virol **82**(Pt 5): 1049-60.
- Baltimore, D.** (1970). "RNA-dependent DNA polymerase in virions of RNA tumour viruses." Nature **226**(5252): 1209-11.
- Barre-Sinoussi, F., J. C. Chermann, F. Rey, M. T. Nugeyre, S. Chamaret, J. Gruest, C. Dauguet, C. Axler-Blin, F. Vezinet-Brun, C. Rouzioux, W. Rozenbaum and L. Montagnier** (1983). "Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)." Science **220**(4599): 868-71.
- Bazykin, G. A., J. Dushoff, S. A. Levin and A. S. Kondrashov** (2006). "Bursts of nonsynonymous substitutions in HIV-1 evolution reveal instances of positive selection at conservative protein sites." Proc Natl Acad Sci U S A **103**(51): 19396-401.
- Beerens, N. and B. Berkhout** (2002). "The tRNA primer activation signal in the human immunodeficiency virus type 1 genome is important for initiation and processive elongation of reverse transcription." J Virol **76**(5): 2329-39.
- Berkowitz, R., J. Fisher and S. P. Goff** (1996). "RNA packaging." Curr Top Microbiol Immunol **214**: 177-218.
- Bernier, R. and M. Tremblay** (1995). "Homologous interference resulting from the presence of defective particles of human immunodeficiency virus type 1." J Virol **69**(1): 291-300.
- Borderia, A. V., F. M. Codoner and R. Sanjuan** (2007). "Selection promotes organ compartmentalization in HIV-1: Evidence from *gag* and *pol* genes." Evolution in press.
- Borrow, P., H. Lewicki, X. Wei, M. S. Horwitz, N. Pfeffer, H. Meyers, J. A. Nelson, J. E. Gairin, B. H. Hahn, M. B. Oldstone and G. M. Shaw** (1997). "Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus." Nat Med **3**(2): 205-11.

- Bour, S., U. Schubert and K. Strebel** (1995). "The human immunodeficiency virus type 1 Vpu protein specifically binds to the cytoplasmic domain of CD4: implications for the mechanism of degradation." *J Virol* **69**(3): 1510-20.
- Brander, C. and B. D. Walker** (2003). "Gradual adaptation of HIV to human host populations: good or bad news?" *Nat Med* **9**(11): 1359-62.
- Brown, A. J.** (1997). "Analysis of HIV-1 env gene sequences reveals evidence for a low effective number in the viral population." *Proc Natl Acad Sci U S A* **94**(5): 1862-5.
- Brown, A. L. and P. Monaghan** (1988). "Evolution of the structural proteins of human immunodeficiency virus: selective constraints on nucleotide substitution." *AIDS Res Hum Retroviruses* **4**(6): 399-407.
- Brown, P. O., B. Bowerman, H. E. Varmus and J. M. Bishop** (1989). "Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein." *Proc Natl Acad Sci U S A* **86**(8): 2525-9.
- Bruggeman, J., A. J. Debets, P. J. Wijngaarden, J. A. deVisser and R. F. Hoekstra** (2003). "Sex slows down the accumulation of deleterious mutations in the homothallic fungus *Aspergillus nidulans*." *Genetics* **164**(2): 479-85.
- Bukrinsky, M. I., S. Haggerty, M. P. Dempsey, N. Sharova, A. Adzhubel, L. Spitz, P. Lewis, D. Goldfarb, M. Emerman and M. Stevenson** (1993). "A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells." *Nature* **365**(6447): 666-9.
- Burch, C. L. and L. Chao** (1999). "Evolution by small steps and rugged landscapes in the RNA virus phi6." *Genetics* **151**(3): 921-7.
- Butto, S., C. Argentini, A. M. Mazzella, M. P. Iannotti, P. Leone, P. Leone, A. Nicolosi and G. Rezza** (1997). "Dual infection with different strains of the same HIV-1 subtype." *Aids* **11**(5): 694-6.
- Caballero, A.** (1994). "Developments in the prediction of effective population size." *Heredity* **73** ( Pt 6): 657-79.
- Cao, J., J. McNevin, U. Malhotra and M. J. McElrath** (2003). "Evolution of CD8+ T cell immunity and viral escape following acute HIV-1 infection." *J Immunol* **171**(7): 3837-46.
- Cichutek, K., H. Merget, S. Norley, R. Linde, W. Kreuz, M. Gahr and R. Kurth** (1992). "Development of a quasispecies of human immunodeficiency virus type 1 in vivo." *Proc Natl Acad Sci U S A* **89**(16): 7365-9.
- Clarke, D. K., E. A. Duarte, A. Moya, S. F. Elena, E. Domingo and J. Holland** (1993). "Genetic bottlenecks and population passages cause profound fitness differences in RNA viruses." *J Virol* **67**(1): 222-8.
- Crawford, S. and S. P. Goff** (1985). "A deletion mutation in the 5' part of the pol gene of Moloney murine leukemia virus blocks proteolytic processing of the gag and pol polyproteins." *J Virol* **53**(3): 899-907.
- Creevey, C. J. and J. O. McInerney** (2002). "An algorithm for detecting directional and non-directional positive selection, neutrality and negative selection in protein coding DNA sequences." *Gene* **300**(1-2): 43-51.
- Crowe, S. M. and S. Sonza** (2000). "HIV-1 can be recovered from a variety of cells including peripheral blood monocytes of patients receiving highly active antiretroviral therapy: a further obstacle to eradication." *J Leukoc Biol* **68**(3): 345-50.
- Cuevas, J. M., S. F. Elena and A. Moya** (2002). "Molecular basis of adaptive convergence in experimental populations of RNA viruses." *Genetics* **162**(2): 533-42.

- Chao, L.** (1990). "Fitness of RNA virus decreased by Muller's ratchet." *Nature* **348**(6300): 454-5.
- Chen, M. Y., F. Maldarelli, M. K. Karczewski, R. L. Willey and K. Strebel** (1993). "Human immunodeficiency virus type 1 Vpu protein induces degradation of CD4 in vitro: the cytoplasmic domain of CD4 contributes to Vpu sensitivity." *J Virol* **67**(7): 3877-84.
- Choe, H., M. Farzan, Y. Sun, N. Sullivan, B. Rollins, P. D. Ponath, L. Wu, C. R. Mackay, G. LaRosa, W. Newman, N. Gerard, C. Gerard and J. Sodroski** (1996). "The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates." *Cell* **85**(7): 1135-48.
- Chowers, M. Y., C. A. Spina, T. J. Kwoh, N. J. Fitch, D. D. Richman and J. C. Guatelli** (1994). "Optimal infectivity in vitro of human immunodeficiency virus type 1 requires an intact nef gene." *J Virol* **68**(5): 2906-14.
- Chun, T. W., L. Carruth, D. Finzi, X. Shen, J. A. DiGiuseppe, H. Taylor, M. Hermankova, K. Chadwick, J. Margolick, T. C. Quinn, Y. H. Kuo, R. Brookmeyer, M. A. Zeiger, P. Barditch-Crovo and R. F. Siliciano** (1997). "Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection." *Nature* **387**(6629): 183-8.
- Dayton, A. I., J. G. Sodroski, C. A. Rosen, W. C. Goh and W. A. Haseltine** (1986). "The trans-activator gene of the human T cell lymphotropic virus type III is required for replication." *Cell* **44**(6): 941-7.
- Delwart, E., M. Magierowska, M. Royz, B. Foley, L. Peddada, R. Smith, C. Heldebrant, A. Conrad and M. Busch** (2002). "Homogeneous quasispecies in 16 out of 17 individuals during very early HIV-1 primary infection." *Aids* **16**(2): 189-95.
- Deng, H., R. Liu, W. Ellmeier, S. Choe, D. Unutmaz, M. Burkhart, P. Di Marzio, S. Marmon, R. E. Sutton, C. M. Hill, C. B. Davis, S. C. Peiper, T. J. Schall, D. R. Littman and N. R. Landau** (1996). "Identification of a major co-receptor for primary isolates of HIV-1." *Nature* **381**(6584): 661-6.
- Diaz, R. S., E. C. Sabino, A. Mayer, J. W. Mosley and M. P. Busch** (1995). "Dual human immunodeficiency virus type 1 infection and recombination in a dually exposed transfusion recipient. The Transfusion Safety Study Group." *J Virol* **69**(6): 3273-81.
- Domingo, E., E. Martinez-Salas, F. Sobrino, J. C. de la Torre, A. Portela, J. Ortin, C. Lopez-Galindez, P. Perez-Brena, N. Villanueva, R. Najera and et al.** (1985). "The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance--a review." *Gene* **40**(1): 1-8.
- Domingo, E. and J. J. Holland** (1997). "RNA virus mutations and fitness for survival." *Annu Rev Microbiol* **51**: 151-78.
- Domingo, E., E. Baranowski, C. M. Ruiz-Jarabo, A. M. Martin-Hernandez, J. C. Saiz and C. Escarmis** (1998). "Quasispecies structure and persistence of RNA viruses." *Emerg Infect Dis* **4**(4): 521-7.
- Domingo, E., C. Escarmis, L. Menendez-Arias and J. Holland** (1999). *Viral quasispecies and fitness variation*. San Diego, Academic Press.
- Domingo, E., C. K. Biebricher, M. Eigen and J. Holland** (2001). *Quasispecies and RNA virus evolution: principles and consequences*. Austin, Landes Bioscience.
- Domingo, E., N. Pariente, A. Airaksinen, C. Gonzalez-Lopez, S. Sierra, M. Herrera, A. Grande-Perez, P. R. Lowenstein, S. C. Manrubia, E. Lazaro and C. Escarmis** (2005). "Foot-and-mouth disease virus evolution: exploring pathways towards virus extinction." *Curr Top Microbiol Immunol* **288**: 149-73.

- Domingo, E., C. Escarmis, E. Lazaro and S. C. Manrubia** (2005). "Quasispecies dynamics and RNA virus extinction." *Virus Res* **107**(2): 129-39.
- Domingo, E., V. Martin, C. Perales, A. Grande-Perez, J. Garcia-Arriaza and A. Arias** (2006). "Viruses as quasispecies: biological implications." *Curr Top Microbiol Immunol* **299**: 51-82.
- Doranz, B. J., J. Rucker, Y. Yi, R. J. Smyth, M. Samson, S. C. Peiper, M. Parmentier, R. G. Collman and R. W. Doms** (1996). "A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors." *Cell* **85**(7): 1149-58.
- Draenert, R., S. Le Gall, K. J. Pfafferoth, A. J. Leslie, P. Chetty, C. Brander, E. C. Holmes, S. C. Chang, M. E. Feeney, M. M. Addo, L. Ruiz, D. Ramduth, P. Jeena, M. Altfeld, S. Thomas, Y. Tang, C. L. Verrill, C. Dixon, J. G. Prado, P. Kiepiela, J. Martinez-Picado, B. D. Walker and P. J. Goulder** (2004). "Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection." *J Exp Med* **199**(7): 905-15.
- Drake, J. W., B. Charlesworth, D. Charlesworth and J. F. Crow** (1998). "Rates of spontaneous mutation." *Genetics* **148**(4): 1667-86.
- Duarte, E., D. Clarke, A. Moya, E. Domingo and J. Holland** (1992). "Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet." *Proc Natl Acad Sci U S A* **89**(13): 6015-9.
- Edwards, C. T., E. C. Holmes, D. J. Wilson, R. P. Viscidi, E. J. Abrams, R. E. Phillips and A. J. Drummond** (2006). "Population genetic estimation of the loss of genetic diversity during horizontal transmission of HIV-1." *BMC Evol Biol* **6**: 28.
- Edwards, C. T., E. C. Holmes, O. G. Pybus, D. J. Wilson, R. P. Viscidi, E. J. Abrams, R. E. Phillips and A. J. Drummond** (2006). "Evolution of the human immunodeficiency virus envelope gene is dominated by purifying selection." *Genetics* **174**(3): 1441-53.
- Eigen, M.** (1971). "Selforganization of matter and the evolution of biological macromolecules." *Naturwissenschaften* **58**(10): 465-523.
- Eigen, M. and C. K. Biebricher** (1988). Sequence space and quasispecies distribution. *RNA genetics*. E. Domingo, J. Holland and P. Ahlquist, CRC Press, Inc., Boca Raton, Fla. **3**.
- Elena, S. F., M. Dávila, I. S. Novella, J. Holland, E. Domingo and A. Moya** (1998). "Evolutionary dynamics of fitness recovery from the debilitating effects of Muller's ratchet." *Evolution* **52**: 309-314.
- Elena, S. F., R. Sanjuan, A. V. Borderia and P. E. Turner** (2001). "Transmission bottlenecks and the evolution of fitness in rapidly evolving RNA viruses." *Infect Genet Evol* **1**(1): 41-8.
- Escarmis, C., M. Davila, N. Charpentier, A. Bracho, A. Moya and E. Domingo** (1996). "Genetic lesions associated with Muller's ratchet in an RNA virus." *J Mol Biol* **264**(2): 255-67.
- Escarmis, C., M. Davila and E. Domingo** (1999). "Multiple molecular pathways for fitness recovery of an RNA virus debilitated by operation of Muller's ratchet." *J Mol Biol* **285**(2): 495-505.
- Escarmis, C., G. Gomez-Mariano, M. Davila, E. Lazaro and E. Domingo** (2002). "Resistance to extinction of low fitness virus subjected to plaque-to-plaque transfers: diversification by mutation clustering." *J Mol Biol* **315**(4): 647-61.

- Escarmis, C., E. Lazaro and S. C. Manrubia (2006). Population Bottlenecks in Quasispecies Dynamics. Quasispecies: Concepts and Implications for Virology. E. Domingo, Springer. **299**: 410.
- Essajee, S. M., H. Pollack, G. Rochford, I. Oransky, K. Krasinski and W. Borkowsky (2000). "Early changes in quasispecies repertoire in HIV-infected infants: correlation with disease progression." AIDS Res Hum Retroviruses **16**(18): 1949-57.
- Feng, D. F., M. S. Johnson and R. F. Doolittle (1984). "Aligning amino acid sequences: comparison of commonly used methods." J Mol Evol **21**(2): 112-25.
- Feng, Y., C. C. Broder, P. E. Kennedy and E. A. Berger (1996). "HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor." Science **272**(5263): 872-7.
- Fisher, A. G., M. B. Feinberg, S. F. Josephs, M. E. Harper, L. M. Marselle, G. Reyes, M. A. Gonda, A. Aldovini, C. Debouk, R. C. Gallo and et al. (1986). "The trans-activator gene of HTLV-III is essential for virus replication." Nature **320**(6060): 367-71.
- Freed, E. O., G. Englund and M. A. Martin (1995). "Role of the basic domain of human immunodeficiency virus type 1 matrix in macrophage infection." J Virol **69**(6): 3949-54.
- Freed, E. O. (1998). "HIV-1 gag proteins: diverse functions in the virus life cycle." Virology **251**(1): 1-15.
- Fujiwara, T. and K. Mizuuchi (1988). "Retroviral DNA integration: structure of an integration intermediate." Cell **54**(4): 497-504.
- Gallo, R. C., S. Z. Salahuddin, M. Popovic, G. M. Shearer, M. Kaplan, B. F. Haynes, T. J. Palker, R. Redfield, J. Oleske, B. Safai and et al. (1984). "Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS." Science **224**(4648): 500-3.
- Ganeshan, S., R. E. Dickover, B. T. Korber, Y. J. Bryson and S. M. Wolinsky (1997). "Human immunodeficiency virus type 1 genetic evolution in children with different rates of development of disease." J Virol **71**(1): 663-77.
- Garcia, J. V. and A. D. Miller (1991). "Serine phosphorylation-independent downregulation of cell-surface CD4 by nef." Nature **350**(6318): 508-11.
- Gojobori, T., E. N. Moriyama and M. Kimura (1990). "Molecular clock of viral evolution, and the neutral theory." Proc Natl Acad Sci U S A **87**(24): 10015-8.
- Gonzalez-Lopez, C., G. Gomez-Mariano, C. Escarmis and E. Domingo (2005). "Invariant aphthovirus consensus nucleotide sequence in the transition to error catastrophe." Infect Genet Evol **5**(4): 366-74.
- Gottlieb, M. S., R. Schroff, H. M. Schanker, J. D. Weisman, P. T. Fan, R. A. Wolf and A. Saxon (1981). "Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency." N Engl J Med **305**(24): 1425-31.
- Gottlinger, H. G., T. Dorfman, J. G. Sodroski and W. A. Haseltine (1991). "Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release." Proc Natl Acad Sci U S A **88**(8): 3195-9.
- Grande-Perez, A., G. Gomez-Mariano, P. R. Lowenstein and E. Domingo (2005). "Mutagenesis-induced, large fitness variations with an invariant arenavirus consensus genomic nucleotide sequence." J Virol **79**(16): 10451-9.
- Haigh, J. (1978). "The accumulation of deleterious genes in a population--Muller's Ratchet." Theor Popul Biol **14**(2): 251-67.

- Hall, T. (1999). "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT." *Nucl. Acids. Symp. Ser* **41**: 95-98.
- Heeney, J. L., A. G. Dalgleish and R. A. Weiss (2006). "Origins of HIV and the evolution of resistance to AIDS." *Science* **313**(5786): 462-6.
- Heinzinger, N. K., M. I. Bukinsky, S. A. Haggerty, A. M. Ragland, V. Kewalramani, M. A. Lee, H. E. Gendelman, L. Ratner, M. Stevenson and M. Emerman (1994). "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* **91**(15): 7311-5.
- Ho, D. D., A. U. Neumann, A. S. Perelson, W. Chen, J. M. Leonard and M. Markowitz (1995). "Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection." *Nature* **373**(6510): 123-6.
- Holland, J., K. Spindler, F. Horodyski, E. Grabau, S. Nichol and S. VandePol (1982). "Rapid evolution of RNA genomes." *Science* **215**(4540): 1577-85.
- Holland, J. J., J. C. de la Torre, D. K. Clarke and E. Duarte (1991). "Quantitation of relative fitness and great adaptability of clonal populations of RNA viruses." *J Virol* **65**(6): 2960-7.
- Hu, W. S. and H. M. Temin (1990). "Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination." *Proc Natl Acad Sci U S A* **87**(4): 1556-60.
- Huang, M., J. M. Orenstein, M. A. Martin and E. O. Freed (1995). "p6Gag is required for particle production from full-length human immunodeficiency virus type 1 molecular clones expressing protease." *J Virol* **69**(11): 6810-8.
- Hubert, J. J. (1984). Bioassay. D. Hunt: 65-66.
- Iglesias-Sanchez, M. J. and C. Lopez-Galindez (2005). "Each genomic RNA in HIV-1 heterozygous virus generate new virions." *Virology* **333**(2): 316-23.
- Imamichi, H., T. Igarashi, T. Imamichi, O. K. Donau, Y. Endo, Y. Nishimura, R. L. Willey, A. F. Suffredini, H. C. Lane and M. A. Martin (2002). "Amino acid deletions are introduced into the V2 region of gp120 during independent pathogenic simian immunodeficiency virus/HIV chimeric virus (SHIV) infections of rhesus monkeys generating variants that are macrophage tropic." *Proc Natl Acad Sci U S A* **99**(21): 13813-8.
- Johnson, V. A., F. Brun-Vezinet, B. Clotet, B. Conway, R. T. D'Aquila, L. M. Demeter, D. R. Kuritzkes, D. Pillay, J. M. Schapiro, A. Telenti and D. D. Richman (2003). "Drug resistance mutations in HIV-1." *Top HIV Med* **11**(6): 215-21.
- Joseph, A. M., M. Kumar and D. Mitra (2005). "Nef: "necessary and enforcing factor" in HIV infection." *Curr HIV Res* **3**(1): 87-94.
- Jung, A., R. Maier, J. P. Vartanian, G. Bocharov, V. Jung, U. Fischer, E. Meese, S. Wain-Hobson and A. Meyerhans (2002). "Multiply infected spleen cells in HIV patients." *Nature* **418**(6894): 144.
- Katoh, I., Y. Yoshinaka, A. Rein, M. Shibuya, T. Odaka and S. Oroszlan (1985). "Murine leukemia virus maturation: protease region required for conversion from "immature" to "mature" core form and for virus infectivity." *Virology* **145**(2): 280-92.
- Keightley, P. D. and A. Caballero (1997). "Genomic mutation rates for lifetime reproductive output and lifespan in *Caenorhabditis elegans*." *Proc Natl Acad Sci U S A* **94**(8): 3823-7.
- Kibota, T. T. and M. Lynch (1996). "Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*." *Nature* **381**(6584): 694-6.

- Kils-Hutten, L., R. Cheynier, S. Wain-Hobson and A. Meyerhans** (2001). "Phylogenetic reconstruction of inpatient evolution of human immunodeficiency virus type 1: predominance of drift and purifying selection." *J Gen Virol* **82**(Pt 7): 1621-7.
- Kimura, M.** (1985). "The neutral theory of molecular evolution." *New Scientist* **11**: 41-46.
- Knipe, D. M., P. M. Howley, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizaman and S. E. Straus** (2001). *Fields Virology*, Lippincott Williams & Wilkins.
- Krakauer, D. C. and J. B. Plotkin** (2002). "Redundancy, antiredundancy, and the robustness of genomes." *Proc Natl Acad Sci U S A* **99**(3): 1405-9.
- Kremer, M. and B. S. Schnierle** (2005). "HIV-1 Vif: HIV's weapon against the cellular defense factor APOBEC3G." *Curr HIV Res* **3**(4): 339-44.
- Kumar, S., K. Tamura and M. Nei** (2004). "MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment." *Brief Bioinform* **5**(2): 150-63.
- Leitner, T., B. Foley, B. Hahn, P. Marx, F. McCutchan, J. Mellors, S. Wolinsky and B. Korber** (2005). *HIV Sequence Compendium 2005*, Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, NM, LA-UR 06-0680.
- Li, W. H., M. Tanimura and P. M. Sharp** (1988). "Rates and dates of divergence between AIDS virus nucleotide sequences." *Mol Biol Evol* **5**(4): 313-30.
- Li, W. H.** (1993). "Unbiased estimation of the rates of synonymous and nonsynonymous substitution." *J Mol Evol* **36**(1): 96-9.
- Louwagie, J., F. E. McCutchan, M. Peeters, T. P. Brennan, E. Sanders-Buell, G. A. Eddy, G. van der Groen, K. Fransen, G. M. Gershy-Damet, R. Deleys and et al.** (1993). "Phylogenetic analysis of gag genes from 70 international HIV-1 isolates provides evidence for multiple genotypes." *Aids* **7**(6): 769-80.
- Lu, Y. L., P. Spearman and L. Ratner** (1993). "Human immunodeficiency virus type 1 viral protein R localization in infected cells and virions." *J Virol* **67**(11): 6542-50.
- Lynch, M., R. Burger, D. Butcher and W. Gabriel** (1993). "The mutational meltdown in asexual populations." *J Hered* **84**(5): 339-44.
- Mansky, L. M. and H. M. Temin** (1995). "Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase." *J Virol* **69**(8): 5087-94.
- Marozsan, A. J., E. Fraundorf, A. Abraha, H. Baird, D. Moore, R. Troyer, I. Nankja and E. J. Arts** (2004). "Relationships between infectious titer, capsid protein levels, and reverse transcriptase activities of diverse human immunodeficiency virus type 1 isolates." *J Virol* **78**(20): 11130-41.
- Marzio, G., M. Vink, K. Verhoef, A. de Ronde and B. Berkhout** (2002). "Efficient human immunodeficiency virus replication requires a fine-tuned level of transcription." *J Virol* **76**(6): 3084-8.
- Masur, H., M. A. Michelis, J. B. Greene, I. Onorato, R. A. Stouwe, R. S. Holzman, G. Wormser, L. Brettman, M. Lange, H. W. Murray and S. Cunningham-Rundles** (1981). "An outbreak of community-acquired *Pneumocystis carinii* pneumonia: initial manifestation of cellular immune dysfunction." *N Engl J Med* **305**(24): 1431-8.
- McNearney, T., Z. Hornickova, R. Markham, A. Birdwell, M. Arens, A. Saah and L. Ratner** (1992). "Relationship of human immunodeficiency virus type 1 sequence heterogeneity to stage of disease." *Proc Natl Acad Sci U S A* **89**(21): 10247-51.

- Meyer, B. E. and M. H. Malim** (1994). "The HIV-1 Rev trans-activator shuttles between the nucleus and the cytoplasm." *Genes Dev* **8**(13): 1538-47.
- Miller, M. D., M. T. Warmerdam, I. Gaston, W. C. Greene and M. B. Feinberg** (1994). "The human immunodeficiency virus-1 nef gene product: a positive factor for viral infection and replication in primary lymphocytes and macrophages." *J Exp Med* **179**(1): 101-13.
- Miyoshi, I., I. Kubonishi, S. Yoshimoto, T. Akagi, Y. Ohtsuki, Y. Shiraishi, K. Nagata and Y. Hinuma** (1981). "Type C virus particles in a cord T-cell line derived by co-cultivating normal human cord leukocytes and human leukaemic T cells." *Nature* **294**(5843): 770-1.
- Mukai, T., S. I. Chigusa, L. E. Mettler and J. F. Crow** (1972). "Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*." *Genetics* **72**(2): 335-55.
- Muller, H. J.** (1964). "The relation of recombination to mutational advance." *Mutat. Res.* **1**: 2-9.
- Myers, G.** (1994). "Tenth anniversary perspectives on AIDS. HIV: between past and future." *AIDS Res Hum Retroviruses* **10**(11): 1317-24.
- Nei, M. and T. Gojobori** (1986). "Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions." *Mol Biol Evol* **3**(5): 418-26.
- Nerrienet, E., M. L. Santiago, Y. Foupouapouognigni, E. Bailes, N. I. Mundy, B. Njinku, A. Kfutwah, M. C. Muller-Trutwin, F. Barre-Sinoussi, G. M. Shaw, P. M. Sharp, B. H. Hahn and A. Ayouba** (2005). "Simian immunodeficiency virus infection in wild-caught chimpanzees from cameroon." *J Virol* **79**(2): 1312-9.
- Neumann, A. U., N. P. Lam, H. Dahari, D. R. Gretch, T. E. Wiley, T. J. Layden and A. S. Perelson** (1998). "Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy." *Science* **282**(5386): 103-7.
- Novella, I. S., E. A. Duarte, S. F. Elena, A. Moya, E. Domingo and J. J. Holland** (1995). "Exponential increases of RNA virus fitness during large population transmissions." *Proc Natl Acad Sci U S A* **92**(13): 5841-4.
- Novella, I. S. and B. E. Ebendick-Corpus** (2004). "Molecular basis of fitness loss and fitness recovery in vesicular stomatitis virus." *J Mol Biol* **342**(5): 1423-30.
- Pang, S., Y. Shlesinger, E. S. Daar, T. Moudgil, D. D. Ho and I. S. Chen** (1992). "Rapid generation of sequence variation during primary HIV-1 infection." *Aids* **6**(5): 453-60.
- Pereira, L. A., K. Bentley, A. Peeters, M. J. Churchill and N. J. Deacon** (2000). "A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter." *Nucleic Acids Res* **28**(3): 663-8.
- Perucho, M., M. Goldfarb, K. Shimizu, C. Lama, J. Fogh and M. Wigler** (1981). "Human-tumor-derived cell lines contain common and different transforming genes." *Cell* **27**(3 Pt 2): 467-76.
- Poon, B., K. Grovit-Ferbas, S. A. Stewart and I. S. Chen** (1998). "Cell cycle arrest by Vpr in HIV-1 virions and insensitivity to antiretroviral agents." *Science* **281**(5374): 266-9.
- Posada, D.** (2006). "ModelTest Server: a web-based tool for the statistical selection of models of nucleotide substitution online." *Nucleic Acids Res* **34**(Web Server issue): W700-3.
- Price, D. A., P. J. Goulder, P. Klenerman, A. K. Sewell, P. J. Easterbrook, M. Troop, C. R. Bangham and R. E. Phillips** (1997). "Positive selection of HIV-1



- cytotoxic T lymphocyte escape variants during primary infection." Proc Natl Acad Sci U S A **94**(5): 1890-5.
- Ratner, L., W. Haseltine, R. Patarca, K. J. Livak, B. Starcich, S. F. Josephs, E. R. Doran, J. A. Rafalski, E. A. Whitehorn, K. Baumeister and et al.** (1985). "Complete nucleotide sequence of the AIDS virus, HTLV-III." Nature **313**(6000): 277-84.
- Richman, D. D., T. Wrin, S. J. Little and C. J. Petropoulos** (2003). "Rapid evolution of the neutralizing antibody response to HIV type 1 infection." Proc Natl Acad Sci U S A **100**(7): 4144-9.
- Roebuck, K. A. and M. Saifuddin** (1999). "Regulation of HIV-1 transcription." Gene Expr **8**(2): 67-84.
- Rouzine, I. M., A. Rodrigo and J. M. Coffin** (2001). "Transition between stochastic evolution and deterministic evolution in the presence of selection: general theory and application to virology." Microbiol Mol Biol Rev **65**(1): 151-85.
- Saag, M. S.** (2001). "The impact of highly active antiretroviral therapy on HIV-specific immune function." Aids **15 Suppl 2**: S4-10.
- Saksena, N. K. and S. J. Potter** (2003). "Reservoirs of HIV-1 in vivo: implications for antiretroviral therapy." AIDS Rev **5**(1): 3-18.
- Sala, M., E. Pelletier and S. Wain-Hobson** (1995). "HIV-1 gp120 sequences from a doubly infected drug user." AIDS Res Hum Retroviruses **11**(5): 653-5.
- Sanchez-Palomino, S., J. M. Rojas, M. A. Martinez, E. M. Fenyo, R. Najera, E. Domingo and C. Lopez-Galindez** (1993). "Dilute passage promotes expression of genetic and phenotypic variants of human immunodeficiency virus type 1 in cell culture." J Virol **67**(5): 2938-43.
- Sanjuan, R., F. M. Codoner, A. Moya and S. F. Elena** (2004). "Natural selection and the organ-specific differentiation of HIV-1 V3 hypervariable region." Evolution Int J Org Evolution **58**(6): 1185-94.
- Sanjuan, R., J. M. Cuevas, A. Moya and S. F. Elena** (2005). "Epistasis and the adaptability of an RNA virus." Genetics **170**(3): 1001-8.
- Sattentau, Q. J., J. P. Moore, F. Vignaux, F. Traincard and P. Poignard** (1993). "Conformational changes induced in the envelope glycoproteins of the human and simian immunodeficiency viruses by soluble receptor binding." J Virol **67**(12): 7383-93.
- Sauder, C. J., K. M. Vandeburgh, R. C. Iskow, T. Malik, K. M. Carbone and S. A. Rubin** (2006). "Changes in mumps virus neurovirulence phenotype associated with quasispecies heterogeneity." Virology **350**(1): 48-57.
- Schrofelbauer, B., Q. Yu and N. R. Landau** (2004). "New insights into the role of Vif in HIV-1 replication." AIDS Rev **6**(1): 34-9.
- Schubert, U., S. Bour, A. V. Ferrer-Montiel, M. Montal, F. Maldarell and K. Strebel** (1996). "The two biological activities of human immunodeficiency virus type 1 Vpu protein involve two separable structural domains." J Virol **70**(2): 809-19.
- Schwartz, O., V. Marechal, O. Danos and J. M. Heard** (1995). "Human immunodeficiency virus type 1 Nef increases the efficiency of reverse transcription in the infected cell." J Virol **69**(7): 4053-9.
- Schwartz, O., V. Marechal, S. Le Gall, F. Lemonnier and J. M. Heard** (1996). "Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein." Nat Med **2**(3): 338-42.

- Seibert, S. A., C. Y. Howell, M. K. Hughes and A. L. Hughes** (1995). "Natural selection on the gag, pol, and env genes of human immunodeficiency virus 1 (HIV-1)." *Mol Biol Evol* **12**(5): 803-13.
- Seo, T. K., J. L. Thorne, M. Hasegawa and H. Kishino** (2002). "Estimation of effective population size of HIV-1 within a host: a pseudomaximum-likelihood approach." *Genetics* **160**(4): 1283-93.
- Shankarappa, R., J. B. Margolick, S. J. Gange, A. G. Rodrigo, D. Upchurch, H. Farzadegan, P. Gupta, C. R. Rinaldo, G. H. Learn, X. He, X.-L. Huang and J. I. Mullins** (1999). "Consistent Viral Evolutionary Changes Associated with the Progression of Human Immunodeficiency Virus Type 1 Infection." *J. Virol.* **73**(12): 10489-10502.
- Shaw, R. G., D. L. Byers and E. Darmo** (2000). "Spontaneous mutational effects on reproductive traits of arabidopsis thaliana." *Genetics* **155**(1): 369-78.
- Shriner, D., R. Shankarappa, M. A. Jensen, D. C. Nickle, J. E. Mittler, J. B. Margolick and J. I. Mullins** (2004). "Influence of random genetic drift on human immunodeficiency virus type 1 env evolution during chronic infection." *Genetics* **166**(3): 1155-64.
- Simon, F., P. Maucere, P. Roques, I. Loussert-Ajaka, M. C. Muller-Trutwin, S. Saragosti, M. C. Georges-Courbot, F. Barre-Sinoussi and F. Brun-Vezinet** (1998). "Identification of a new human immunodeficiency virus type 1 distinct from group M and group O." *Nat Med* **4**(9): 1032-7.
- Smith, R. A., D. J. Anderson and B. D. Preston** (2004). "Purifying selection masks the mutational flexibility of HIV-1 reverse transcriptase." *J Biol Chem* **279**(25): 26726-34.
- Spina, C. A., T. J. Kwok, M. Y. Chowes, J. C. Guatelli and D. D. Richman** (1994). "The importance of nef in the induction of human immunodeficiency virus type 1 replication from primary quiescent CD4 lymphocytes." *J Exp Med* **179**(1): 115-23.
- Steinhauer, D. A., J. C. de la Torre and J. J. Holland** (1989). "High nucleotide substitution error frequencies in clonal pools of vesicular stomatitis virus." *J Virol* **63**(5): 2063-71.
- Steinhauer, D. A., E. Domingo and J. J. Holland** (1992). "Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase." *Gene* **122**(2): 281-8.
- Strebel, K., D. Daugherty, K. Clouse, D. Cohen, T. Folks and M. A. Martin** (1987). "The HIV 'A' (sor) gene product is essential for virus infectivity." *Nature* **328**(6132): 728-30.
- Temin, H. M. and S. Mizutani** (1970). "RNA-dependent DNA polymerase in virions of Rous sarcoma virus." *Nature* **226**(5252): 1211-3.
- Thali, M., J. P. Moore, C. Furman, M. Charles, D. D. Ho, J. Robinson and J. Sodroski** (1993). "Characterization of conserved human immunodeficiency virus type 1 gp120 neutralization epitopes exposed upon gp120-CD4 binding." *J Virol* **67**(7): 3978-88.
- Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin and D. G. Higgins** (1997). "The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools." *Nucleic Acids Res* **25**(24): 4876-82.
- Thomson, M. M., L. Perez-Alvarez and R. Najera** (2002). "Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy." *Lancet Infect Dis* **2**(8): 461-71.

- Tippmann, H. F.** (2004). "Analysis for free: comparing programs for sequence analysis." *Brief Bioinform* **5**(1): 82-7.
- van 't Wout, A. B., H. Blaak, L. J. Ran, M. Brouwer, C. Kuiken and H. Schuitemaker** (1998). "Evolution of syncytium-inducing and non-syncytium-inducing biological virus clones in relation to replication kinetics during the course of human immunodeficiency virus type 1 infection." *J Virol* **72**(6): 5099-107.
- van Opijnen, T., R. E. Jeeninga, M. C. Boerlijst, G. P. Pollakis, V. Zetterberg, M. Salminen and B. Berkhout** (2004). "Human immunodeficiency virus type 1 subtypes have a distinct long terminal repeat that determines the replication rate in a host-cell-specific manner." *J Virol* **78**(7): 3675-83.
- Verdin, E. and C. Van Lint** (1995). "Internal transcriptional regulatory elements in HIV-1 and other retroviruses." *Cell Mol Biol (Noisy-le-grand)* **41**(3): 365-9.
- Vignuzzi, M., J. K. Stone, J. J. Arnold, C. E. Cameron and R. Andino** (2006). "Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population." *Nature* **439**(7074): 344-8.
- von Schwedler, U., R. S. Kornbluth and D. Trono** (1994). "The nuclear localization signal of the matrix protein of human immunodeficiency virus type 1 allows the establishment of infection in macrophages and quiescent T lymphocytes." *Proc Natl Acad Sci U S A* **91**(15): 6992-6.
- Wain-Hobson, S., P. Sonigo, O. Danos, S. Cole and M. Alizon** (1985). "Nucleotide sequence of the AIDS virus, LAV." *Cell* **40**(1): 9-17.
- Wei, X., S. K. Ghosh, M. E. Taylor, V. A. Johnson, E. A. Emini, P. Deutsch, J. D. Lifson, S. Bonhoeffer, M. A. Nowak, B. H. Hahn and et al.** (1995). "Viral dynamics in human immunodeficiency virus type 1 infection." *Nature* **373**(6510): 117-22.
- White, J. M.** (1992). "Membrane fusion." *Science* **258**(5084): 917-24.
- Wilke, C. O.** (2003). "Probability of fixation of an advantageous mutant in a viral quasispecies." *Genetics* **163**(2): 467-74.
- Wolf, J. B., E. D. Brodie III and M. J. Wade** (2000). *Epistasis and the evolutionary process*. Oxford, Oxford University Press.
- Wolfs, T. F., G. Zwart, M. Bakker and J. Goudsmit** (1992). "HIV-1 genomic RNA diversification following sexual and parenteral virus transmission." *Virology* **189**(1): 103-10.
- Wolinsky, S. M., C. M. Wike, B. T. Korber, C. Hutto, W. P. Parks, L. L. Rosenblum, K. J. Kunstman, M. R. Furtado and J. L. Munoz** (1992). "Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants." *Science* **255**(5048): 1134-7.
- Wolinsky, S. M., B. T. Korber, A. U. Neumann, M. Daniels, K. J. Kunstman, A. J. Whetsell, M. R. Furtado, Y. Cao, D. D. Ho and J. T. Safrit** (1996). "Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection." *Science* **272**(5261): 537-42.
- Wright, S.** (1932). *The roles of mutation, inbreeding, crossbreeding, and selection in evolution*. Proceedings of the Sixth International Congress on Genetics.
- Wright, S.** (1982). "The shifting balance theory and macroevolution." *Annu Rev Genet* **16**: 1-19.
- Yoshida, M., I. Miyoshi and Y. Hinuma** (1982). "Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease." *Proc Natl Acad Sci U S A* **79**(6): 2031-5.

- Yuan, X., X. Yu, T. H. Lee and M. Essex** (1993). "Mutations in the N-terminal region of human immunodeficiency virus type 1 matrix protein block intracellular transport of the Gag precursor." *J Virol* **67**(11): 6387-94.
- Yuste, E., S. Sanchez-Palomino, C. Casado, E. Domingo and C. Lopez-Galindez** (1999). "Drastic fitness loss in human immunodeficiency virus type 1 upon serial bottleneck events." *J Virol* **73**(4): 2745-51.
- Yuste, E., C. Lopez-Galindez and E. Domingo** (2000). "Unusual distribution of mutations associated with serial bottleneck passages of human immunodeficiency virus type 1." *J Virol* **74**(20): 9546-52.
- Yuste, E., A. V. Borderia, E. Domingo and C. Lopez-Galindez** (2005). "Few mutations in the 5' leader region mediate fitness recovery of debilitated human immunodeficiency virus type 1 viruses." *J Virol* **79**(9): 5421-7.
- Zhang, L. Q., P. MacKenzie, A. Cleland, E. C. Holmes, A. J. Brown and P. Simmonds** (1993). "Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection." *J Virol* **67**(6): 3345-56.
- Zhou, W., L. J. Parent, J. W. Wills and M. D. Resh** (1994). "Identification of a membrane-binding domain within the amino-terminal region of human immunodeficiency virus type 1 Gag protein which interacts with acidic phospholipids." *J Virol* **68**(4): 2556-69.
- Zhu, H., H. Jian and L. J. Zhao** (2004). "Identification of the 15FRFG domain in HIV-1 Gag p6 essential for Vpr packaging into the virion." *Retrovirology* **1**(1): 26.
- Zhu, T., H. Mo, N. Wang, D. S. Nam, Y. Cao, R. A. Koup and D. D. Ho** (1993). "Genotypic and phenotypic characterization of HIV-1 patients with primary infection." *Science* **261**(5125): 1179-81.
- Zhu, T., N. Wang, A. Carr, S. Wolinsky and D. D. Ho** (1995). "Evidence for coinfection by multiple strains of human immunodeficiency virus type 1 subtype B in an acute seroconverter." *J Virol* **69**(2): 1324-7.
- Zhu, T., B. T. Korber, A. J. Nahmias, E. Hooper, P. M. Sharp and D. D. Ho** (1998). "An African HIV-1 sequence from 1959 and implications for the origin of the epidemic." *Nature* **391**(6667): 594-7.