

Caracterización y modelado computacional de la recombinación en secuencias de ADN en el virus del SIDA (VIH-1)

Antonio Carvajal-Rodríguez(1,2), Keith A. Crandall(1) and David Posada(2)
(1)Department of Integrative Biology, BYU, Utah (USA). (2)Departamento de Bioquímica, Genética e Inmunología. Universidad de Vigo (España)

Introducción

El virus de la inmunodeficiencia humana tipo 1 (VIH-1) pertenece a la familia *Retroviridae* y se encuentra entre los patógenos humanos más variables genéticamente(1). Esta gran variabilidad puede explicarse como consecuencia de tres factores(2):

1. La elevada tasa de mutación debida a la frecuente incorporación de bases erróneas por parte de la enzima transcriptasa inversa.
2. La elevada producción de virus en cada generación.
3. La elevada tasa de recombinación genética.

En el presente trabajo nos vamos a centrar en la recombinación genética, que se puede definir como el intercambio genético entre secuencias homólogas de ADN provenientes de dos virus distintos. La recombinación desempeña un papel fundamental a la hora de generar variabilidad genética en el VIH-1, permitiendo al virus escapar de vacunas y/o medicamentos. Veremos además la importancia que tiene el conocer la frecuencia con que la recombinación ocurre a la hora de poder diseñar vacunas eficaces y probarlas.

En el VIH-1, para que la recombinación genere nueva variabilidad una célula ha de ser infectada por dos provirus diferentes, ya sea simultánea (infección dual o múltiple) o secuencialmente (superinfección), de manera que se produzca el encapsulamiento de un transcrito de ARN de cada provirus en un mismo virión heterocigoto. Debido a su forma de operar, saltando entre distintas hebras, la enzima transcriptasa inversa (que copia el ARN viral en DNA) favorece una elevada frecuencia de recombinación. Después de la infección de una nueva célula (Figura 1), la transcriptasa inversa generará una nueva molécula de ADN retroviral que será recombinante de los dos genomas parentales. En individuos que hayan sido infectados por virus genéticamente diferentes se pueden generar así combinaciones genómicas originales y muy diferenciadas.

Por tanto, para que se genere una nueva partícula vírica recombinante portadora de nuevas combinaciones genéticas deben ocurrir dos procesos diferentes, en primer lugar una célula debe ser co-infectada por al menos dos partículas víricas y en segundo lugar, durante el proceso de replicación del virus la enzima transcriptasa inversa debe copiar ADN procedente de los distintos virus en una misma secuencia, formando un mosaico genético (Figura 2). Debido al modo en que la replicación ocurre en los retrovirus, este último proceso sucede de manera frecuente.

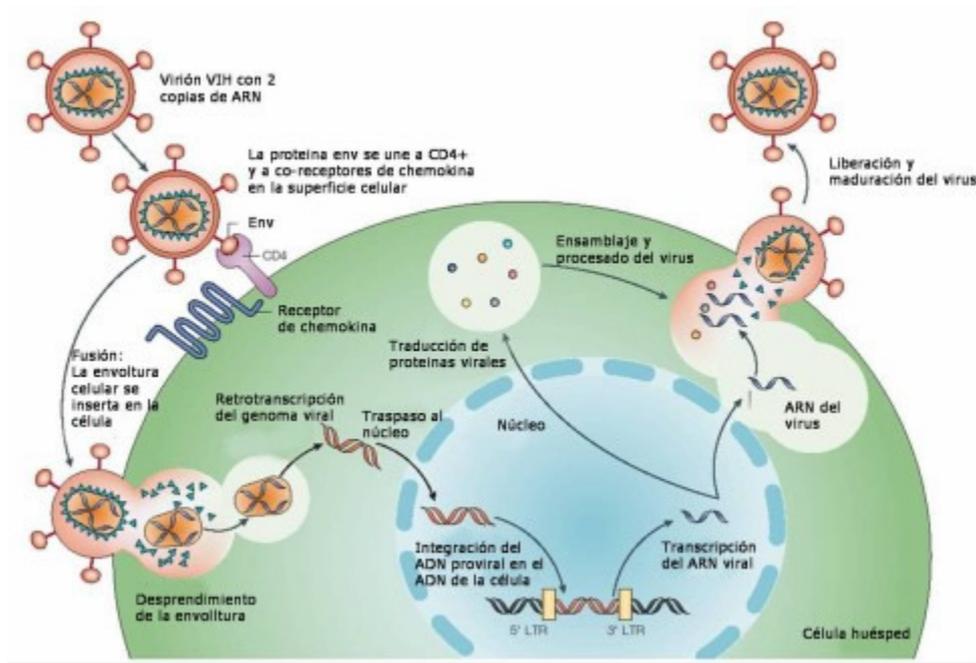


Figura 1. Aspectos clave del ciclo de vida del VIH-1. Aunque el VIH-1 puede infectar distintos tipos de células, el síndrome de inmunodeficiencia adquirida (SIDA) es el resultado de la pérdida de CD4+ que son un elemento clave del sistema inmunológico humano. El gen env codifica las proteínas de la cubierta externa del virus, el gen gag codifica las proteínas de la cápside interna del virus y el gen pol codifica las enzimas usadas en la replicación del virus. Figura modificada a partir de Rambaut A, Posada D, Crandall KA and Holmes EC. 2004. The causes and consequences of HIV evolution. Nature Reviews Genetics 5: 52-61.

Recientemente, se ha demostrado que la doble infección por VIH-1 es frecuente en cultivos de células T(3). Asimismo, la recombinación es también un hecho muy frecuente en el VIH-1 como se demuestra en diferentes estudios en los que se mide directamente mediante técnicas como la citometría de flujo(4). De hecho, los recombinantes son comunes entre las distintas cepas del VIH-1(5), siendo más fácil de detectar la recombinación que ocurre entre miembros de diferentes subtipos (ver más abajo). Algunas formas del virus conocidas como formas recombinantes en circulación (CRFs de sus siglas en inglés) tienen genomas mosaicos que derivan de cinco o más variantes genéticamente distintas. Esto indica la ocurrencia tanto de infección doble como frecuente recombinación(4). Probablemente la recombinación también es frecuente entre miembros del mismo subtipo pero en este caso es más difícil detectarla debido a que los fragmentos genómicos intercambiados pueden ser muy parecidos.

En el VIH-1 se distinguen tres linajes filogenéticos conocidos como grupos M, N y O(6 7) cada uno de los cuales parece haber surgido de manera independiente en África central a partir de infecciones procedentes de cepas de virus de inmunodeficiencia simia de chimpancé. Los virus del grupo M son los principales responsables de la epidemia actual en tanto que el grupo O es de incidencia mucho menor y generalmente reducida al área del centro-oeste de África. Asimismo las cepas del grupo N son raras, con localizaciones en Camerún y oeste del África ecuatorial. El grupo M se puede subdividir en 9 subtipos, A-D, F-H, J y K, y al menos 18 formas recombinantes circulantes (Figura 2). Estas agrupaciones filogenéticas de las cepas de VIH-1, subtipos y formas recombinantes son importantes porque sirven como registro histórico de las epidemias de VIH-1 habidas, resaltando además las diferencias en la distribución geográfica y las áreas donde las distintas variantes son endémicas.

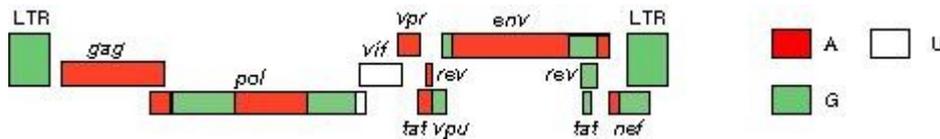


Figura 2. CRF03_AB representa una forma recombinante con dos subtipos. Es frecuente en Rusia y Ucrania principalmente entre usuarios de drogas inyectadas. Imagen tomada de <http://hiv-web.lanl.gov/content/hiv-db/CRFs/CRFs.html>

Implicaciones de la elevada frecuencia de recombinación en VIH-1

La consideración del papel de la recombinación como generador de diversidad a nivel global en el VIH-1 ha ido en aumento en los últimos años. Las formas recombinantes parecen ser mucho más comunes, extendidas a nivel geográfico y diversas de lo que inicialmente se pensaba(7). Además, es posible relacionar la aparición de nuevas formas recombinantes con un fuerte impacto en la salud pública(8), ya que:

1. La recombinación genera una gran variabilidad del virus, lo cual favorece el incremento de la complejidad general de la pandemia del VIH-1.
2. En redes sexuales donde co-circulan diferentes subtipos del VIH-1 los recombinantes surgen muy frecuentemente.

Como consecuencia del efecto de la recombinación en las células de un paciente infectado, podrían emerger virus genéticamente diferentes, variantes resistentes a vacunas sencillas o múltiples o virus con tropismo alterado. Por ello, es importante detectar y estimar la ocurrencia de recombinación de manera precisa porque su cuantificación tiene implicaciones potenciales en cuanto al desarrollo de las vacunas y a la resistencia a medicamentos. Otro aspecto a tener en cuenta se refiere a la comprensión del origen y la historia del virus. En este caso los métodos filogenéticos son una poderosa herramienta para estimar parámetros evolutivos (como tasa de mutación, tamaño de población eficaz, etc) a partir de muestras de ADN. Sin embargo si ignoramos la recombinación, estos métodos pueden producir estimas erróneas(9) (10). Por tanto, son necesarias estimas más precisas de la frecuencia y diversidad de la recombinación dentro y entre subtipos para estimar las pautas temporales de evolución del virus y su incidencia. Estos estudios son de especial relevancia para la planificación e interpretación de los resultados en los ensayos de vacunas. Pero no sólo es importante la obtención de estimas empíricas precisas, si no que además es importante desarrollar métodos analíticos que contemplen la presencia de recombinación.

En resumen, uno de los mayores obstáculos para el tratamiento del VIH-1 es su gran variación genética. La alta tasa de recombinación contribuye de forma significativa a generar esta variación, dificultando la efectividad de los tratamientos. Debido a la recombinación, se pueden generar de manera muy rápida variantes resistentes a drogas múltiples(4). De hecho, con las elevadas tasas de replicación del VIH-1, en individuos infectados la recombinación puede organizar y recombinar el genoma entero en períodos cortos de tiempo. Esta habilidad del VIH-1 para evolucionar representa un gran desafío para su tratamiento y para el desarrollo de vacunas. Es por ello muy importante definir y entender los medios que usa el virus para generar variación, de forma que podamos estimar de mejor manera el potencial evolutivo y las limitaciones del VIH-1. Esto permitirá a su vez un mejor diseño de estrategias terapéuticas para el tratamiento de la infección por VIH-1. Teniendo en cuenta todo lo anterior cobra sentido el uso de herramientas computacionales tanto para detectar como para cuantificar la ocurrencia de recombinación.

Detección de la recombinación a nivel poblacional

La recombinación en el VIH-1 se puede medir de manera directa o indirecta. Ambos tipos de métodos se complementan. Los métodos directos miden tasa de recombinación celular e implican detectar los recombinantes mediante citometría de flujo dentro de las propias células infectadas. Estos métodos aportan información muy interesante y directa, y de hecho recientemente se ha estimado de esta manera que la tasa de recombinación del VIH es de un orden de magnitud mayor que en otros retrovirus(4,11,12,13). Por otro lado, los métodos indirectos miden tasa de recombinación poblacional y permiten contrastar, a partir de muestras de secuencias de ADN, la ocurrencia de recombinación, identificar las secuencias parentales y las recombinantes y localizar los puntos de recombinación. En general se puede decir que los métodos indirectos se implementan en programas de computador que aplican uno o varios algoritmos que tratan de medir la "huella" dejada por la recombinación en los genomas virales. De manera muy genérica, es posible calificar los diversos métodos indirectos en cinco categorías generales no exclusivas: similitud, distancia, filogenéticos, de compatibilidad y de distribución de sustituciones nucleotídicas. En lo que sigue mencionaremos brevemente cada una de estas categorías. Para una revisión más en profundidad pueden consultarse otras referencias(14, 15, 16).

- (a) Métodos de similitud: Son aplicables sólo a regiones codificadoras. Se basan en la comparación del número de sustituciones sinónimas en regiones variables y no variables.
- (b) Métodos de distancia: Buscan alteraciones en los patrones de distancia entre secuencias. Son muy eficientes computacionalmente porque no es necesario reconstruir la filogenia de las secuencias.
- (c) Métodos filogenéticos: Estiman las filogenias usando distintas partes de un genoma. La recombinación se manifestará como discordancia entre las topologías obtenidas. Si no hay recombinación la topología obtenida deberá ser la misma independientemente de la región utilizada para reconstruir la filogenia. Son los métodos más usados.
- (d) Métodos de compatibilidad: Contrastan la existencia de incongruencia filogenética pero usando un análisis de sitio a sitio en lugar de por regiones genómicas. En este caso no es necesaria la filogenia de la secuencia bajo estudio.
- (e) Distribución de sustituciones: Se estudian agrupamientos significativos de sustituciones o ajustes a distribuciones estadísticas. Incluyen los métodos que usan la teoría de coalescencia (ver más adelante).

El funcionamiento de los distintos métodos que estiman recombinación ha sido evaluado mediante el análisis de datos generados por ordenador(14) y también con datos empíricos(17). Estos estudios se han centrado en la detección de la presencia de recombinación y no tanto en la identificación de las secuencias parentales y las recombinantes, ni tampoco en la localización de los puntos específicos de recombinación y no dan por tanto una idea completa de nuestra capacidad de detectar todos los aspectos de la recombinación. En general, La habilidad de detectar recombinación de los distintos métodos depende de factores como la tasa de recombinación, la diversidad genética de la muestra y el grado de variación entre sitios en la tasa de sustitución nucleotídica. La mayoría de los métodos tienen problemas a la hora de detectar eventos de recombinación raros, especialmente cuando la divergencia es baja. Por otro lado, los métodos que usan patrones de sustitución o incompatibilidad son más potentes que los basados en incongruencia filogenética. Además parece que algunos métodos que no se basan en el principio de incompatibilidad tiene problemas para detectar recombinación cuando las filogenias han evolucionado con poblaciones actualmente en expansión (muchos individuos en el presente pero no tantos en el pasado).

Una vez que se detecta la recombinación surge la cuestión de si es un verdadero evento de recombinación in vivo o por el contrario puede haberse producido in vitro durante la reacción en cadena de la polimerasa (PCR por sus siglas en inglés) usada para amplificar la región deseada del genoma. Esto es de especial importancia cuando se trata de cadenas largas

amplificadas mediante PCR tal y como se ha demostrado para secuencias del VIH-1(15). El análisis filogenético nos permite discriminar entre estas dos situaciones. En la recombinación in vivo esperamos diferencias entre el número de mutaciones acumuladas tras el evento de recombinación dependiendo si este es reciente o no. En la recombinación debida a PCR no esperamos estas diferencias.

En resumen, no hay un estimador óptimo de recombinación. Para maximizar las posibilidades de detectar recombinación cuando está presente, evitando al mismo tiempo inferirla cuando no lo está, es aconsejable calcular la cantidad de divergencia y heterogeneidad en las tasas de sustitución de los datos. Es entonces cuando podremos decidir entre unos u otros métodos para estimar recombinación.

Estima de la tasa de recombinación poblacional

Como hemos visto, la recombinación tiene un gran efecto a la hora de generar diversidad genética. Es importante entender este papel en relación al que desarrolla la mutación, y para ello necesitamos no sólo detectar sino estimar de manera precisa la tasa de recombinación. Definimos el parámetro de recombinación poblacional como $\rho = 4N_e r$, siendo N_e el tamaño de población eficaz y r la tasa de recombinación por sitio y generación. Así mismo se define el parámetro de mutación poblacional (diversidad genética) como $\theta = 4N_e \mu$, siendo μ la tasa de mutación por sitio y por generación. Si logramos estimar de modo preciso ambos parámetros podremos definir la tasa relativa de recombinación respecto a la de mutación como $\epsilon = r / \mu$. Al igual que en el caso de θ , hay una gran variedad de métodos para estimar ρ . En general, los estimadores de recombinación usan dos aproximaciones, o bien tratan de cuantificar la recombinación mediante algún estadístico sumario (por ejemplo, la varianza observada en el número de diferencias en las comparaciones entre pares nucleotídicos)(18 19) o bien estiman la tasa de recombinación usando todos los datos mediante la teoría de máxima verosimilitud para obtener una estima conjunta de θ y ρ (20 21 22 23). La ventaja de estos últimos métodos es que usan toda la información de la muestra. La desventaja es que son, desde el punto de vista computacional, muy intensivos. Se han propuesto también soluciones de compromiso que usan subconjuntos de las secuencias(24) o bien combinan ambas aproximaciones(25). Otros métodos usan el marco Bayesiano para estimar los parámetros(26) (27). También se ha propuesto recientemente(28) una aproximación alternativa que asume un modelo de infinitos sitios y usa pares de sitios polimórficos (en lugar de toda la secuencia) y utiliza los métodos de máxima verosimilitud para analizarlos. Este método ha sido extendido(29) para permitir un modelo con finitos sitios de modo que se permite mutación recurrente. Sin embargo no se permiten modelos complejos de sustitución nucleotídica ni variación en la tasa de sustitución entre sitios, aunque estas dos carencias no parecen afectar a las estimas de recombinación en datos simulados bajo modelos complejos y tasa de sustitución heterogénea (datos no publicados). Todos estos métodos asumen tamaños constantes de población, independencia entre sitios, evolución neutra y modelos de infinitos sitios (con la excepción del método de McVean(29)). Sin embargo difieren considerablemente en términos del muestreo poblacional, nivel de polimorfismo nucleotídico y número y tipo de posiciones nucleotídicas a considerar.

Eficiencia de los estimadores de recombinación

En general la eficiencia de los estimadores de recombinación depende de la diversidad genética (θ). De este modo, la mayoría de los métodos fallan cuando la diversidad genética es baja(15). Entre los que mejor funcionan se hallan dos métodos que se corresponden con estrategias de máxima verosimilitud (muy costosas desde el punto de vista computacional) que se distinguen en la forma de recorrer el espacio de las genealogías posibles(21,22). Ambos se basan en la teoría de coalescencia, que es una teoría de procesos estocásticos que construye diversas posibles historias (genealogías) de las secuencias muestreadas desde el presente hacia el pasado hasta llegar al ancestro común más reciente (MRCA de sus siglas en inglés). Finalmente se seleccionará la historia y los valores de los parámetros que mejor se ajusten a los datos. El estudio y comparación de los diversos métodos para estimar

recombinación está en sus comienzos pero es un campo vital porque cuanto mejor entendamos cuán bien funcionan los métodos existentes más fácil es desarrollar nuevos y mejores métodos. Por tanto hace falta más investigación para entender cuan robustos son los distintos estimadores a violaciones de la teorías y modelos poblacionales que asumen.

El futuro. Importancia de las herramientas computacionales en la lucha contra el SIDA y otras pandemias.

Es posible que la tasa de recombinación en el VIH-1 (y en otros virus de ARN) sea un carácter sometido a selección. De hecho la recombinación puede resultar en incrementos directos de eficacia biológica, por ejemplo juntando en un mismo genoma mutaciones distintas de resistencia a medicamentos(30). Así mismo, la reorganización que la recombinación produce puede permitir a los virus librarse de la acumulación de mutaciones deletéreas. Es de esperar que si la recombinación estuviese favorecida selectivamente en el VIH-1 lo mismo ocurriría para la co-infección, puesto que ésta es un requisito previo para que la recombinación genere genomas mosaicos. En este sentido son interesantes resultados recientes que apoyan una ocurrencia de infección múltiple superior a la esperada por azar(3) así como la ocurrencia de recombinación tras superinfección por VIH-1(31). Es por tanto una tarea para el futuro el determinar la variación en las tasas de recombinación en virus e investigar si esta variación está correlacionada con otros efectos biológicos como la patogenicidad. Si la recombinación es un carácter sometido a selección, ¿cómo afecta esto a los modelos utilizados para estimar recombinación?. Por otro lado, esté o no esté la recombinación sometida directamente a selección, otra tarea pendiente que necesita atención es estudiar si la selección natural puede sesgar las estimas de las tasas de recombinación. Por ejemplo el parámetro θ que se utiliza en el modelo de coalescencia con recombinación es una medida de diversidad genética que asume exclusivamente evolución neutral.

BIBLIOGRAFÍA

1. Wain-Hobson S. 1993. The fastest genome evolution ever described: VIH variation in situ. *Curr. Opin. Genet. Dev.* 3: 878-883.
2. Magiorkinis G, Paraskevis D, Vandamme A-M y col. 2003. In vivo characteristics of human immunodeficiency virus type 1 intersubtype recombination: determination of hot spots and correlation with sequence similarity. *Journal of general virology* 84: 2715-2722.
3. Dang Q, Chen J, Unutmaz D y col. 2004. Nonrandom VIH-1 infection and double infection via direct and cell-mediated pathways. *Proc. Natl. Acad. Sci. USA.* 101: 632-637.
4. Rhodes TD, Nikolaitchik O, Chen J y col. 2005. Genetic recombination of human immunodeficiency virus type 1 in one round of viral replication: effects of genetic distance, target cells, accessory genes, and lack of high negative interference in crossover events.
5. Quinones-Mateu ME y Arts EJ. 1999. Recombination in VIH-1 : Update and implications. *AIDS Reviews* 1: 89-100.
6. Kalish ML, Robbins KE, Pieniazek D y col. 2004. Recombinant viruses and early global VIH-1 epidemic. *Emerging infectious diseases* 10: 1227-1234.
7. Nájera R, Delgado R, Pérez_Álvarez L y col. 2002. Genetic recombination and its role in the development of the VIH-1 pandemic. *AIDS* 16: S3-S16.
8. Yábar CV. 2003. Eventos moleculares, genéticos e inmunológicos durante la interacción VIH-hombre. *Rev. Peru. Med. Exp. Salud Pública.* 20 (2): 107-115.
9. Schierup MH and Hein J. 2000a. Consequences of recombination on traditional phylogenetic analysis. *Genetics* 156:879-891.
10. Schierup MH and Hein J. 2000b. Recombination and the molecular clock. *Mol. Biol. Evol.* 17: 1578-1579.
11. Levy DN, Aldrovandi GM, Kutsch O y col. 2004. Dynamics of VIH-1 recombination in its natural target cells. *Proc. Natl. Acad. Sci. USA.* 101: 4204-4209.
12. Onafuwa A, An W, Robson ND y col. 2003. Human immunodeficiency virus type 1 genetic recombination is more frequent than that of Moloney murine leukemia virus despite similar template switching rates. *J. Virol.* 77: 4577-4587.
13. Rhodes T, Wargo H y Hu WS. 2003. High rates of human immunodeficiency virus type 1 recombination: near-random segregation of markers one kilobase apart in one round of viral replication. *J. Viral.* 77: 11193-11200.
14. Posada D y Crandall K. 2001. Performance of methods for detecting recombination from DNA sequences: computer simulations. *Proc. Natl. Acad. Sci. USA* 98: 13757-13562.
15. Posada D, Crandall KA y Holmes EC. 2002. Recombination in evolutionary genomics. *Annu. Rev. Genet.* 36: 75-97.
16. Stumpf P.H. y McVean G. 2003. Estimating recombination rates from population-genetic data. *Nat Rev Genet.* 4: 959-968.
17. Posada D. 2001. Unveiling the molecular clock in the presence of recombination. *Molecular Biology and Evolution* 18: 1976-1978.
18. Hudson RR. 1987. Estimating the recombination parameter of a finite population model without selection. *Genetical Research Cambridge* 50: 245-250.
19. Wakeley J. 1997. Using the variance of pairwise differences to estimate the recombination rate. *Genetical Research Cambridge* 69: 45-48.
20. Griffiths RC y Marjoran P. 1996. Ancestral inference from samples of DNA sequences with recombination. *J. Comput. Biol.* 3: 479-502.
21. Kuhner MK, Yamato J, y Felsenstein J. 2000. Maximum likelihood estimation of recombination rates from population data. *Genetics* 156: 1393-1401.
22. Fearnhead P y Donnelly P. 2001. Estimating recombination rates from population genetic data. *Genetics* 159: 1299-1318.
23. Awadalla P. 2003. The evolutionary genomics of pathogen recombination. *Nat Rev Genet.* 4(1):50-60.
24. Hey J and Wakeley J. 1997. A coalescent estimator of the population recombination rate. *Genetics* 145: 833-846.
25. Wall JD. 2000. A comparison of estimators of population recombination rate. *Mol. Biol. Evol.* 17: 156-163.
26. Falush D, Kraft C, Taylor NS y col. 2001. Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size and minimal age. *Proc. Natl. Sci. USA* 98: 15056-15061.

27. Nielsen R. 2000. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics* 154: 931-942.
28. Hudson RR. 2001. Two-locus sampling distributions and their application. *Genetics* 159: 1805-1817.
29. McVean G, Awadalla P y Fearnhead P. 2002. A coalescent based-method for detecting and estimating recombination from gene sequences. *Genetics* 160: 1231-1241.
30. Moutouh L, Corbeil J y Richman DD. 1996. Recombination leads to the rapid emergence of VIH-1 dually resistant mutants under selective drug pressure. *Proc. Natl. Acad. Sci. USA* 93: 6106-6111.
31. Fang G, Weiser B, Kuiken C y col. 2004. Recombination following superinfection by HIV-1. *AIDS* 18(2): 153-160.