Identification of a major restriction in HIV-1 intersubtype recombination

Mario P. S. Chin*, Terence D. Rhodes*[†], Jianbo Chen*, William Fu*, and Wei-Shau Hu**

*HIV Drug Resistance Program, National Cancer Institute, Frederick, MD 21702; and [†]Department of Microbiology, Immunology, and Cell Biology, West Virginia University, Morgantown, WV 26505

Edited by John M. Coffin, Tufts University School of Medicine, Boston, MA, and approved May 11, 2005 (received for review March 27, 2005)

Genetic recombination increases diversity in HIV-1 populations, thereby allowing variants to escape from host immunity or antiviral therapies. In addition to the currently described nine subtypes of HIV-1, many of the circulating strains are intersubtype recombinants. In this study, we determined the recombination rate between two HIV-1 subtype C viruses and between a subtype B virus and a subtype C virus during a single round of virus replication. Although HIV-1 subtype C recombines at a high rate, similar to that of HIV-1 subtype B, the recombination rate between a subtype B virus and a subtype C virus is much lower than the intrasubtype recombination rate. A 3-nt sequence difference in the dimerization initiation signal (DIS) region between HIV-1 subtypes B and C accounts for most of the reduction of intersubtype recombination. By matching the DIS sequences, the B/C intersubtype recombination rate was elevated 4-fold; by introducing mismatches in the 3-nt sequences, the B/B intrasubtype recombination rate was reduced 4-fold. Further analyses showed that the intermolecular template-switching frequency was unaffected by the sequence identity of the DIS region. These results support the hypothesis that mismatched sequences in the DIS region alter the formation of heterozygous virions, thereby lowering the observable recombination rate. Here, we present the discovery of a major restriction in HIV-1 intersubtype recombination. These results have important implications for virus evolution, the mechanism of HIV-1 RNA packaging, high negative interference in recombination, and the generation of circulating intersubtype recombinants within the infected population.

subtype | RNA dimerization

One of the hallmarks of HIV-1 is its high genetic variability, which is demonstrated by the large number of variants circulating in the human population (1). HIV-1 consists of three phylogenetically distinct groups: M, O, and N (2). Among the three HIV-1 groups, group M dominates the global epidemic with at least nine subtypes, 16 circulating recombinant forms, and other strains that are currently being identified (1). (Additional information on recombinant forms is available at www.hiv.lanl.gov.) The most studied form of HIV-1 is subtype B, which circulates predominantly in the North and South Americas, Western Europe, Australia, and Japan (3, 4). The most prevalent HIV-1 subtype in the global epidemic, however, is subtype C, which is particularly dominant in Ethiopia (5), South Africa (6), Tanzania (7), China (8, 9), and India (10).

Genetic recombination is one of the mechanisms that generates the rapid diversification of HIV-1 populations by reassorting mutations generated by reverse transcriptase (RT) (11–15). Recombination among genetically distinct strains and between subtypes of HIV-1 is being identified with increasing frequency in the global epidemic (4, 16). HIV-1 recombinants were estimated to contribute to 10–40% and 10–30% of the infections in Africa and Asia, respectively (4). Many of the intersubtype recombinants circulate with high prevalence in certain geographic regions, such as A/E recombinants in Thailand and B/C recombinants in parts of Southeast Asia and China (17–19).

Frequent retroviral recombination occurs during reverse transcription (20). Although recombination could occur during infection of all retroviral particles, only virions that contain two RNAs encoding different genetic information could generate a genotypically different recombinant (20). Therefore, in order for HIV-1 intersubtype recombination to occur, the RNAs of the two subtypes of viruses must be copackaged into the same virion and recombination must occur during reverse transcription. We and others (13–15, 21–25) have examined HIV-1 recombination with subtype B-based vectors. Although there is ample evidence among the circulating variants (1) and cell culture experiments (26, 27) to indicate that intersubtype recombination could occur, very little is known about the rates and limitations of recombination between different HIV-1 subtypes. We sought to determine whether there are barriers to HIV-1 intersubtype recombination and to define such barriers if they exist. In this study, we examined the recombination potential between HIV-1 subtypes B and C and found that these recombination events occur at a much lower rate than those from two subtype B (24) or two subtype C viruses. We further delineated the cause of lower intersubtype recombination rates and found that sequence differences in a region proposed to be important to RNA dimerization is primarily responsible for the lower rate. These findings reveal a major restriction to intersubtype recombination and have significant implications for the evolution of the viral populations.

Materials and Methods

Plasmid Constructions. Subtype C-based vectors were derived from pMJ4, which contains the full-length, infectious HIV-1 subtype C molecular clone (28) and was obtained from the National Institutes of Health AIDS Research and Reference Reagent Program. To facilitate cloning, an *Nde*I site and an *Xho*I site in the pMJ4 plasmid backbone were eliminated and an *Xho*I site was added in *nef* by standard cloning procedures to generate pSMC-C-KONX.NX. For clarity, pON-H0 and pON-T6 (24) are referred to here as pBH0 and pBT6, respectively. The *Nde*I-to-*Xho*I fragment from pBH0 or pBT6 was inserted into *Nde*I- and *Xho*I-digested pSMC-C-KONX.NX to generate pCH0 or pCT6, respectively.

Chimeric vectors were constructed based on pHIV-BH0, which is identical to pBH0 except for a 1-kb deletion in the flanking host genome of the plasmid backbone. Plasmid pBH0.Cp-3Cgag was constructed by replacing the *Sfo*I-to-*Spe*I fragment of pHIV-BH0 with that of pMJ4. Overlapping PCR was used to generate *Sfo*I-to-*Spe*I DNA fragments that contained hybrid sequences, which replaced their counterparts in

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: DIS, dimerization initiation signal; HSA, heat-stable antigen; IRES, internal ribosomal entry site; moi, multiplicity of infection; PBS, primer binding site; RT, reverse transcriptase.

[†]To whom correspondence should be addressed at: HIV Drug Resistance Program, National Cancer Institute, P.O. Box B, Building 535, Room 336, Frederick, MD 21702. E-mail: whu@ncicrf.gov.

© 2005 by The National Academy of Sciences of the USA

We envisioned two possible mechanisms to explain the cause of the low B/C intersubtype recombination rate. The first possible mechanism is that subtype C virus recombines at a rate much lower than that of subtype B virus. Thus, the lower B/C intersubtype recombination rate might simply reflect the average of the high subtype B rate and low subtype C rate. The second possible mechanism is that subtype C virus can recombine as efficiently as subtype B virus, but a restriction(s) exists between these two viruses to reduce the intersubtype recombination rates.

Sequence Differences in the Packaging Signals Affect B/C Intersubtype Recombination Rates. Detectable recombination occurs during reverse transcription of a virus containing two different viral RNAs. Because both subtype B and C vectors generate similar titers from a particular producer cell line and both recombine frequently, we hypothesized that subtype B and C RNAs are not copackaged efficiently, thereby reducing the production of heterozygous viruses and resulting in the lower intersubtype recombination rate. Because the packaging signal in the viral genome dictates the efficiency of viral RNA packaged by viral protein, the packaging signal is likely to play a role in restricting intersubtype recombination. The major RNA packaging signal in HIV-1 includes four stem-loop structures from downstream of primer binding site (PBS) to the 5' end of *gag*

To further define the region responsible for restricting B/C intersubtype recombination, we constructed two additional vectors that contained the same viral sequences as pBH0, except with the following differences: vector pBH0.Cp-3 contains subtype C sequences from the PBS to the AUG of *gag*, and pBH0.Cgag contains the first 327 nt of the subtype C *gag* from the *gag* AUG to part of CA (Fig. 1c). We then measured the recombination rate between the subtype C vector CT6 and these two vectors in the aforementioned one-round recombination assay. We found that 3.9% of the infection events yielded the GFP⁺ phenotype by viruses harvested from BH0.Cp-3 and CT6 producer cells, whereas only 0.8% of the infection events generated the GFP⁺ phenotype by viruses harvested from BH0.Cgag and CT6 producer cells (Fig. 2b). Therefore, sequences in the stem-loop structure of the 5' UTR, not the 5' *gag* region, affect the intersubtype recombination rates.

Mismatched Sequences in the Subtype B and Subtype C Dimerization Initiation Signal (DIS) Serve as a Major Restriction to HIV-1 Recombination. To further delineate the sequence(s) affecting B/C intersubtype recombination, we compared the sequence from the PBS to the AUG of *gag* in subtype B and C vectors (Fig. 3). This region of the RNA is thought to form four stem-loop structures termed SL1–SL4; SL1–SL3 are located in the UTR, whereas SL4 is located at the 5' end of *gag*. Mutation analyses suggested that SL1–SL4 are all important for viral RNA packaging (32, 33). Additionally, SL1 contains a palindrome at the loop region. This palindromic sequence was proposed to be used by the two copackaged RNAs to form base pairs and serve as a signal for RNA dimer initiation, hence the term DIS (34–36).

B: CAGTGGCGCCCGAACAGGGACTTGAAAGCGAAAGTAAAGCCAGA
C: CAGTGGCGCCCGAACAGGGACTTGAAAGCGAAAGTAAAGCCAGA

PBS DIS

B: GGAGATCTCTCGACGCAGGACTCGGCTTGCTGAAECGCGACCGG
C: GGAGGCTCTCTCGACGCAGGACTCGGCTTGCTGAAETGCACTCGG

SL1

B: CAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAAAAT--TT
C: CAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAAAATTTATT

SL2

B: TGA CTAGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCGTC
C: TGA CTAGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCGTC

SL3 SL4

→ gag

Fig. 3. Alignment of a portion of the packaging signals from HIV-1 subtype B (NL4-3) and C (MJ4) DNA sequences. Mismatches are shown in boldface; sequences for PBS, SL1, SL2, SL3, and SL4 are underlined; the start codon of *gag* is shown in italic and boldface; and DIS regions are shown with boxes.

The subtype B and C vectors that we used differ only in 11 nt in the region from PBS to the AUG of *gag*. Among the 11 nt, four mismatches are located between PBS and SL1, three mismatches are located at or adjacent to DIS, and the other four mismatches are located between SL2 and SL3. Because of the proposed function of the DIS region, we tested whether the sequence differences in the 3 nt within the DIS region affect the B/C intersubtype recombination. We generated pBH0.Cdis (Fig. 1c), which has the same viral sequences as pBH0 except that the 3-nt sequence in the DIS region was changed from subtype B to subtype C sequences (GCGCGCA to GTGC4CT). We generated cell lines containing BH0.Cdis and CT6 and performed the recombination assay. We observed that 3.6% of the infection events yielded the GFP⁺ phenotype (Fig. 2c). These results indicated that sequence identity in the DIS region is critical to the B/C intersubtype recombination rate and that the 3-nt sequence difference is responsible for most, if not all, of the augmentation in the recombination rates when we altered the packaging signals.

By matching the 3-nt differences in the DIS region, we were able to significantly elevate the B/C intersubtype recombination rate (4-fold). If the mismatched sequences in the DIS region placed a restriction on intersubtype recombination, then introducing mismatches into the DIS region of vectors from the same subtype should also reduce the recombination rate. To test this prediction, we generated virus-producing cell lines containing BH0.Cdis (Fig. 1c) and BT6 (Fig. 1a); these two vectors contain identical HIV-1-derived sequences except for the 3 nt in the DIS region. After the single round recombination assay, our results indicated that 1.7% of the infection events generated the GFP⁺ phenotype (Fig. 2c). Compared with the recombination rate between BH0 and BT6 (Fig. 2a) (24), the recombination rate decreased 4-fold when the 3 nt of the DIS region was changed to subtype C sequences. Therefore, our results indicate that DIS sequence mismatches alter HIV-1 recombination rates and serve as a major restriction in B/C intersubtype recombination.

Distributions of GFP⁺ Cells in HSA⁺ or Thy⁺ Cell Populations Provide Insights to the Mechanism of the Restriction in B/C Intersubtype Recombination. In this study, we observed that the 3-nt mismatch in the DIS region can have a 4-fold impact on the recombination rate. We proposed two hypotheses to explain how the DIS mismatches affect recombination rates. The first hypothesis suggests that mismatches affect the formation of heterozygous virions, thereby reducing the number of observable recombination events. The second hypothesis suggests that the mismatches affect the overall ability of RT to switch between the two copackaged RNA genomes (intermolecular template switching). The location of the mismatch leads us to favor the first hypothesis, i.e., the heterozygous virion formation is affected. However, based on our data thus far, we

Fig. 4. Distribution of GFP⁺ cells in Thy⁺ or HSA⁺ cell populations. (a) Generation of virus with Thy⁺GFP⁺ or HSA⁺GFP⁺ phenotypes by recombination. (b) Distribution of GFP⁺ cells in HSA⁺ or Thy⁺ cell populations generated from different recombination vectors. Gray bars represent the percentage of GFP⁺ cells in the HSA⁺ population, and black bars represent the percentage of GFP⁺ cells in the Thy⁺ population; means and standard deviations are shown. Data from at least four sets of experiments were summarized; the ratios of GFP⁺ in Thy⁺ to GFP⁺ in HSA⁺ in different vector pairs are 2.2, 2.8, 2.0, 2.5, 2.3, 2.9, and 2.5 from left (BT6/CH0) to right (BT6/BH0.Cdis), respectively.

could not rule out the second hypothesis. To directly test the second hypothesis, we examined the distribution of the GFP⁺ cells in HSA⁺ or Thy⁺ cell populations to estimate the intermolecular template switching frequency.

In our system, two viral vectors were used in each recombination assay: The first vector encodes *hsa* and *gfp* with mutations 15 nt downstream of the AUG; the second vector encodes *thy* and *gfp* with a mutation 603 nt downstream of the AUG (Fig. 1a). In order for a progeny DNA to encode functional *gfp*, RT must copy the 5' end of the *gfp* in the *thy*-containing vector to avoid the inactivating *gfp* mutations in the *hsa*-containing vector (Fig. 4a). If RT switches templates during copying of the internal ribosomal entry site (IRES) sequences, the resulting progeny could express HSA and GFP (Fig. 4a). If RT does not switch templates during the copying of IRES, then the resulting progeny could express Thy and GFP. When using two subtype B vectors containing the same markers and mutations, we previously compared the ratio of GFP⁺ cells in the HSA⁺ or Thy⁺ cell population and found ≈2- to 3-fold more GFP⁺ cells in Thy⁺ than in HSA⁺ cell populations (24). This distribution is consistent with the measured recombination rates (24). Furthermore, if more distance was provided for RT to switch template by using other *gfp* mutations to generate a longer distance between the 5' mutation of *gfp* and *hsa* or *thy*, we observed an increased ratio of GFP⁺ in HSA⁺ cells compared with GFP⁺ in Thy⁺ cells (24). Therefore, the distribution of GFP⁺ cells in HSA⁺ or Thy⁺ cell populations can be used as an indicator for intermolecular template switching.

Our analyses of the GFP⁺ cells in HSA⁺ or Thy⁺ cell populations are illustrated in Fig. 4b; although the relative GFP⁺ cells fluctuated along with the observed recombination rates, all of the experimental groups had similar trends in terms of relative ratios of GFP⁺ in

Thy⁺ cells and GFP⁺ in HSA⁺ cells. For example, when two subtype C vectors were used (CH0 and CT6), the ratio of GFP⁺ in Thy⁺ cells was $\approx$ 2- to 3-fold higher than that of GFP⁺ in HSA⁺ cells; these values are also consistent with our previous observation with two subtype B vectors. In the recombination pair BH0.Cdis and BT6, the recombination rates were reduced 4-fold when the 3-nt mismatches were introduced. If the 3-nt mismatched sequence caused a 4-fold reduction in the overall intermolecular template-switching frequency, then such reduction would be reflected during copying the IRES. Instead of observing 2- or 3-fold more GFP⁺ in Thy⁺ cells, we should then have observed 8- to 15-fold more GFP⁺ in Thy⁺ cells than GFP⁺ in HSA⁺ cells. However, our results indicated similar ratios of GFP⁺ cells in Thy⁺ or HSA⁺ cell populations in all experimental groups. Therefore, the distributions of the Thy and HSA markers do not support the hypothesis that the 3-nt mismatch in the DIS region causes a reduction in the overall intermolecular template-switching frequency. Taken together, our results strongly suggest that sequence differences in the DIS region affect the heterozygous virion formation, causing the reduced recombination rates.

Discussion

Genetic recombination reassorts the mutations generated by RT and increases variation in the HIV-1 population. We have previously shown that HIV-1 subtype B recombines very frequently and sequences separated by 1.3 kb segregate as unlinked markers in one round of viral replication (15). In this study, we delineated the HIV-1 subtype C recombination rate with markers separated by 588 nt and observed that HIV-1 subtype B and C have similar recombination rates. The similarity of these two intrasubtype rates suggests that the high rate of recombination is a common feature among HIV-1 group M.

Although HIV-1 subtypes B and C can each recombine at a high rate, we observed that recombination between subtype B and C vectors was restricted and occurred at a much lower rate. The possible mechanisms and implications of the restriction for intersubtype recombination are discussed below.

Mechanisms and Implications of the Effect of Mismatched Sequences in the DIS Region on HIV-1 Recombination. In this study, we have demonstrated that a 3-nt sequence difference in the DIS region of HIV-1 subtype B and C is responsible for most of the differences in the intersubtype and intrasubtype recombination rates. A 4-fold difference in recombination rate can occur depending on whether the DIS sequences are matched between different pairs of vectors. These results demonstrate that mismatched DIS sequences affect the ability of HIV-1 to exchange genetic information.

Currently, it is unclear whether one viral RNA dimer or two monomeric viral RNAs are packaged by HIV-1 proteins during virus assembly (33, 37). Our data indicate that the mismatched sequences in the DIS region affected the formation of functional heterozygous virions and not the overall intermolecular template-switching frequencies. These findings strongly support the hypothesis that one viral RNA dimer is packaged by the viral proteins. Under this hypothesis, it is likely that the mismatched sequences in the DIS region affect dimer formation of subtype B and C RNAs and reduce the total number of heterozygous virions, thereby lowering the observable recombination events. Although we favor the dimer-RNA-packaging hypothesis, we cannot rule out the alternative hypothesis that two copies of monomeric viral RNA are packaged. Under this hypothesis, heterozygous virions are formed at a frequency expected from random RNA copackaging; however, the mismatched sequences in the DIS region cause a dimerization defect in the virion and hamper recombination between these copackaged RNAs. We have explored the alternative hypothesis by examining RNA dimer stability. We reasoned that the dimer structures in a majority of the heterozygous virions would have to

be affected to cause a 9-fold decrease in recombination rate. If the affected heterozygous virions contain monomeric RNAs or have a lower RNA dimer thermostability, we might have been able to detect these changes biochemically by using nondenaturing gel electrophoresis and hybridization analyses. Our preliminary results indicated that virion RNAs produced from cells infected with only subtype B vectors, only subtype C vectors, or both subtype B and subtype C vectors were similar to each other in our analyses (data not shown). We did not observe an increased amount of monomers or a significantly decreased thermostability of the dimers using virions generated from cells infected with both subtype B and C vectors. These data do not support the alternative hypothesis. To further study these two hypotheses, other experimental approaches are needed to characterize the effect of the mismatched sequences in the DIS region on virion RNA copackaging and heterozygous formation. Sequence mismatch in the dimer linkage signal of murine leukemia virus affects crossover in that region (38); it would be interesting to find out whether the overall recombination rate in murine leukemia virus is also altered by mismatch in dimer linkage signal.

Reduced Formation of Functional Heterozygous Virions Could Cause High Negative Interference of HIV-1 Recombination. In our previous report, we demonstrated that the frequency of a double-crossover event is similar to the expected rate calculated from those of two single recombination events (24). These results indicated that crossover events are not correlated. Another report (25) suggested that HIV-1 recombination may exhibit high negative interference, which is defined as multiple crossover events occurring more frequently than expected from the single crossover event rates. These two observations appeared to be contradictory at first glance. However, our conclusion that heterozygous virion formation is affected can easily explain both observations. Our data suggest that cells infected with a subtype B and a subtype C virus may generate fewer heterozygous virions than cells infected with two subtype B or two subtype C viruses; however, all of the viruses have the same overall intermolecular template switching frequency. If this scenario were true, then the observed recombination event (GFP⁺) would be reduced because of the lower numbers of B/C heterozygous virions. However, because the overall intermolecular template switching frequency did not change, progeny generated from the heterozygous virion would have the same average number of crossover events. If we used the measured recombination rate (percentage of GFP⁺) to predict the number of crossover events in the recombinant, we would observe more crossover events than expected from the measured recombination rate, hence creating high negative interference in recombination. This idea is demonstrated in our analyses shown in Fig. 4*b*; in all experimental groups, we observed similar ratios of GFP⁺ cells in HSA⁺ or Thy⁺ cell populations, suggesting similar frequencies of intermolecular template switching when copying the IRES. However, the frequency of GFP⁺ generation can be as different as 9-fold. Therefore, if we used the frequency of the GFP⁺ phenotype to predict the likelihood of an observable second crossover event, we would have concluded that HIV-1 recombination has high negative interference. Therefore, high negative interference of HIV-1 recombination can occur by means of inefficient formation of heterozygous virions and not by correlated crossover events.

Restrictions in Intersubtype Recombination and Their Implications in Subtype Evolution and the Generation of Circulating Intersubtype Recombinants. In this report, we have identified a major restriction to intersubtype recombination: the sequence differences in DIS. Certain group M subtypes share the same DIS sequences; for example, B and D both have the GCGCGC sequence and A, C, F, G, H, and J have the GTGCAC sequence (37). Subtypes containing different DIS are likely to encounter the intersubtype recombination restriction described in this report. It is unclear

whether strains with matching DIS will have intersubtype recombination rates similar to the intrasubtype rates. It is possible that other undefined factors that can restrict HIV-1 recombination exist. Even in the B/C system that we examined, other minor restrictions might exist. For example, changing sequences in the DIS region could restore most of the recombination rate but not to the same level as the intrasubtype recombination rate. Other differences between the subtype B and C genomes could account for the additional 2-fold difference in recombination rate. Additionally, our system does not measure the effects of sequence identity. Other approaches are needed to study such effects on intersubtype recombination.

It has been hypothesized that a single zoonosis event transmitted chimpanzee-derived simian immunodeficiency virus to the human population and generated what is now the M group HIV-1; founder effects and geographic separation subsequently played an important role in the derivations of the various subtypes (39–41). There are regions in the world where multiple subtypes circulate in the human population (4, 42); although intersubtype recombinants have emerged in these regions, individual subtypes remain. Multiple factors probably affect the maintenance of the subtype and emergence of a recombinant strain, such as the compatibilities of various cis- and/or trans-acting elements, and the replication fitness of the virus. However, knowing that major restriction(s) exist to curb intersubtype recombination, we speculate that these restrictions could also play a role in reducing the interactions of the various subtypes, thereby contributing to the maintenance of these subtypes.

Currently, there are at least 16 identified circulating recombinant forms in the human population. (Data on these recombinants are available at www.hiv.lanl.gov.) Of the circulating recombinant forms, two strains are B/C recombinants identified

in Southeast Asia and China (18, 19). The circulating subtype B and subtype C strains in those geographic regions have sequence differences from the laboratory strains that we used in this assay. However, sequence analyses indicated that these field strains have the same DIS mismatches, suggesting that the major restriction we identified can be extended to the intersubtype recombination of these field strains. Despite the restriction(s) in intersubtype recombination, there are still many circulating recombinant forms being generated. It is important to note that HIV-1 recombines at an exceedingly high rate. Although we have demonstrated that intersubtype recombination occurs at a lower rate than intrasubtype recombination, we emphasize that the “lower” rate is at the same range as the recombination rate measured in gammaretroviruses, such as murine leukemia virus or spleen necrosis virus (20, 43). Therefore, even with this reduction, multiple intersubtype recombinant forms of HIV-1 can still be generated and circulate in the human population.

In this report, we demonstrated that HIV-1 subtype C, similar to subtype B, recombines frequently. Anti-HIV-1 treatment regimens are now being launched in many parts of the world where HIV-1 subtype C is prevalent or multiple subtypes are circulating. The genetic power of the virus should be considered when designing treatment regimens. Defining the major restriction in intersubtype recombination allowed us to understand the limits and molecular mechanisms of HIV-1 recombination. Through these studies, we hope to develop strategies to curb HIV-1 evolution and facilitate the treatment of this pathogen.

We thank Anne Arthur for expert editorial help; Vinay K. Pathak for discussion throughout the project; Alan Rein, John M. Coffin, Vinay K. Pathak, and Olga Nikolaitchik for critical reading of the manuscript; and Ben Boyle for pMJ4 sequence analysis. This project is supported by the HIV Drug Resistance Program, National Cancer Institute.

1. Leitner, T., Foley, B., Hahn, B., Marx, P., McCutchan, F., Mellors, J. W., Wolinsky, S. & Korber, B. (2003) *HIV Sequence Compendium 2003* (Theor. Biol. Biophys. Group, Los Alamos Natl. Lab., Los Alamos, NM).
2. Robertson, D. L., Anderson, J. P., Bradac, J. A., Carr, J. K., Foley, B., Funkhouser, R. K., Gao, F., Hahn, B. H., Kalish, M. L., Kuiken, C., *et al.* (2000) *Science* **288**, 55–56.
3. Osmanov, S., Pattou, C., Walker, N., Schwardlander, B. & Esparza, J. (2002) *J. Acquired Immune Defic. Syndr.* **29**, 184–190.
4. Takebe, Y., Kusagawa, S. & Motomura, K. (2004) *Pediatr. Int.* **46**, 236–244.
5. Abebe, A., Pollakis, G., Fontanet, A. L., Fisseha, B., Tegbaru, B., Kliphuis, A., Tesfaye, G., Negassa, H., Cornelissen, M., Goudsmit, J. & Rinke de Wit, T. F. (2000) *AIDS Res. Hum. Retroviruses* **16**, 1909–1914.
6. Van Harmelen, J. H., Van der Ryst, E., Loubser, A. S., York, D., Madurai, S., Lyons, S., Wood, R. & Williamson, C. (1999) *AIDS Res. Hum. Retroviruses* **15**, 395–398.
7. Renjifo, B., Chaplin, B., Mwakagile, D., Shah, P., Vannberg, F., Msamanga, G., Hunter, D., Fawzi, W. & Essex, M. (1998) *AIDS Res. Hum. Retroviruses* **14**, 635–638.
8. Luo, C. C., Tian, C., Hu, D. J., Kai, M., Dondero, T. & Zheng, X. (1995) *Lancet* **345**, 1051–1052.
9. Neild, P. J. & Gazzard, B. G. (1997) *Lancet* **350**, 963.
10. Sahni, A. K., Prasad, V. V. & Seth, P. (2002) *Int. J. STD AIDS* **13**, 115–118.
11. Negroni, M. & Buc, H. (2001) *Annu. Rev. Genet.* **35**, 275–302.
12. Hu, W. S., Rhodes, T., Dang, Q. & Pathak, V. (2003) *Front. Biosci.* **8**, d143–d155.
13. Jetzt, A. E., Yu, H., Klarmann, G. J., Ron, Y., Preston, B. D. & Dougherty, J. P. (2000) *J. Virol.* **74**, 1234–1240.
14. Onafuwa, A., An, W., Robson, N. D. & Telesnitsky, A. (2003) *J. Virol.* **77**, 4577–4587.
15. Rhodes, T., Wargo, H. & Hu, W. S. (2003) *J. Virol.* **77**, 11193–11200.
16. Robertson, D. L., Sharp, P. M., McCutchan, F. E. & Hahn, B. H. (1995) *Nature* **374**, 124–126.
17. Gao, F., Robertson, D. L., Morrison, S. G., Hui, H., Craig, S., Decker, J., Fultz, P. N., Girard, M., Shaw, G. M., Hahn, B. H. & Sharp, P. M. (1996) *J. Virol.* **70**, 7013–7029.
18. Piyasirisilp, S., McCutchan, F. E., Carr, J. K., Sanders-Buell, E., Liu, W., Chen, J., Wagner, R., Wolf, H., Shao, Y., Lai, S., *et al.* (2000) *J. Virol.* **74**, 11286–11295.
19. Motomura, K., Kusagawa, S., Kato, K., Nohtomi, K., Lwin, H. H., Tun, K. M., Thwe, M., Oo, K. Y., Lwin, S., Kyaw, O., *et al.* (2000) *AIDS Res. Hum. Retroviruses* **16**, 1831–1843.
20. Hu, W. S. & Temin, H. M. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 1556–1560.
21. Clavel, F., Hoggan, M. D., Willey, R. L., Strebel, K., Martin, M. A. & Repasko, R. (1989) *J. Virol.* **63**, 1455–1459.
22. Galetto, R., Moumen, A., Giacomoni, V., Veron, M., Charneau, P. & Negroni, M. (2004) *J. Biol. Chem.* **279**, 36625–36632.
23. Levy, D. N., Aldrovandi, G. M., Kutsch, O. & Shaw, G. M. (2004) *Proc. Natl. Acad. Sci. USA* **101**, 4204–4209.
24. Rhodes, T. D., Nikolaichik, O., Chen, J., Powell, D. & Hu, W. S. (2005) *J. Virol.* **79**, 1666–1677.
25. Zhuang, J., Jetzt, A. E., Sun, G., Yu, H., Klarmann, G., Ron, Y., Preston, B. D. & Dougherty, J. P. (2002) *J. Virol.* **76**, 11273–11282.
26. St. Louis, D. C., Gotte, D., Sanders-Buell, E., Ritchey, D. W., Salminen, M. O., Carr, J. K. & McCutchan, F. E. (1998) *J. Virol.* **72**, 3991–3998.
27. Quinones-Mateu, M. E., Gao, Y., Ball, S. C., Marozsan, A. J., Abraha, A. & Arts, E. J. (2002) *J. Virol.* **76**, 9600–9613.
28. Ndung'u, T., Renjifo, B. & Essex, M. (2001) *J. Virol.* **75**, 4964–4972.
29. Freed, E. O., Englund, G. & Martin, M. A. (1995) *J. Virol.* **69**, 3949–3954.
30. Jewell, N. A. & Mansky, L. M. (2000) *J. Gen. Virol.* **81**, 1889–1899.
31. Clever, J., Sasseti, C. & Parslow, T. G. (1995) *J. Virol.* **69**, 2101–2109.
32. Berkowitz, R., Fisher, J. & Goff, S. P. (1996) *Curr. Top. Microbiol. Immunol.* **214**, 177–218.
33. Russell, R. S., Liang, C. & Wainberg, M. A. (2004) *Retrovirology* **1**, 23.
34. Muriaux, D., Girard, P. M., Bonnet-Mathoniere, B. & Paoletti, J. (1995) *J. Biol. Chem.* **270**, 8209–8216.
35. Laughrea, M. & Jette, L. (1994) *Biochemistry* **33**, 13464–13474.
36. Skripkin, E., Paillart, J. C., Marquet, R., Ehresmann, B. & Ehresmann, C. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 4945–4949.
37. Paillart, J. C., Shehu-Xhilaga, M., Marquet, R. & Mak, J. (2004) *Nat. Rev. Microbiol.* **2**, 461–472.
38. Mikkelsen, J. G., Rasmussen, S. V. & Pedersen, F. S. (2004) *Nucleic Acids Res.* **32**, 102–114.
39. Hahn, B. H., Shaw, G. M., De Cock, K. M. & Sharp, P. M. (2000) *Science* **287**, 607–614.
40. Gao, F., Bailes, E., Robertson, D. L., Chen, Y., Rodenburg, C. M., Michael, S. F., Cummins, L. B., Arthur, L. O., Peeters, M., Shaw, G. M., *et al.* (1999) *Nature* **397**, 436–441.
41. Sharp, P. M., Bailes, E., Robertson, D. L., Gao, F. & Hahn, B. H. (1999) *Biol. Bull.* **196**, 338–342.
42. Peeters, M., Toure-Kane, C. & Nkengasong, J. N. (2003) *Aids* **17**, 2547–2560.
43. Anderson, J. A., Bowman, E. H. & Hu, W. S. (1998) *J. Virol.* **72**, 1195–1202.