



Multiple Barriers to Recombination between Divergent HIV-1 Variants Revealed by a Dual-Marker Recombination Assay

Olga A. Nikolaitchik[†], Andrea Galli[†], Michael D. Moore,
Vinay K. Pathak and Wei-Shau Hu^{*}

HIV Drug Resistance Program, National Cancer Institute-Frederick, P.O. Box B, Building 535, Room 336, Frederick, MD 21702, USA

Received 17 November 2010;
received in revised form
18 January 2011;
accepted 27 January 2011
Available online
3 February 2011

Edited by J. Karn

Keywords:

GFP;
Gag;
group M;
group O;
reconstitution

Recombination is a major force for generating human immunodeficiency virus type 1 (HIV-1) diversity and produces numerous recombinants circulating in the human population. We previously established a cell-based system using green fluorescent protein gene (*gfp*) as a reporter to study the mechanisms of HIV-1 recombination. We now report an improved system capable of detecting recombination using authentic viral sequences. Frameshift mutations were introduced into the *gag* gene so that parental viruses do not express full-length Gag; however, recombination can generate a progeny virus that expresses a functional Gag. We demonstrate that this Gag reconstitution assay can be used to detect recombination between two group M HIV-1 variants of the same or of different subtypes. Using both *gfp* and *gag* assays, we found that, similar to group M viruses, group O viruses also recombine frequently. When recombination between a group M virus and a group O virus was examined, we found three distinct barriers to intergroup recombination. First, similar to recombination within group M viruses, intergroup recombination is affected by the identity of the dimerization initiation signal (DIS); variants with the same DIS recombined at a higher rate than those with different DIS. Second, using the *gfp* recombination assay, we showed that intergroup recombination occurs much less frequently than intragroup recombination, even though the *gfp* target sequence is identical in all viruses. Finally, Gag reconstitution between variants from different groups is further reduced compared with green fluorescent protein, indicating that sequence divergence interferes with recombination efficiency in the *gag* gene. Compared with identical sequences, we estimate that recombination rates are reduced by 3-fold and by 10- to 13-fold when the target regions in *gag* contain 91% and 72–73% sequence identities, respectively. These results show that there are at least three distinct mechanisms preventing exchange of genetic information between divergent HIV-1 variants from different groups.

Published by Elsevier Ltd.

^{*}Corresponding author. E-mail address: Wei-Shau.Hu@nih.gov.

[†] O.A.N. and A.G. contributed equally to this work.

Abbreviations used: DIS, dimerization initiation signal; RT, reverse transcriptase; SIV, simian immunodeficiency virus; HSA, heat-stable antigen; MOI, multiplicity of infection; CA, capsid; HIV-1, human immunodeficiency virus type 1; GFP, green fluorescent protein; IRES, internal ribosome entry site; LTR, long terminal repeat; PE, phycoerythrin.

Introduction

Our current understanding is that independent zoonotic transmission events introduced different simian immunodeficiency viruses (SIVs) into the human population to give rise to four groups of human immunodeficiency virus type 1 (HIV-1): groups M, N, O and P.¹ Group M viruses are most closely related to SIVcpzptt.² Group N appears to be a recombinant between an ancient group M virus and another SIVcpzptt strain;² whereas groups O and P are thought to be derived from different strains of SIVgor.^{3,4}

The four groups of HIV-1 viruses have an uneven distribution across the human population.⁵ The vast majority of infections in the AIDS pandemic are caused by group M viruses; whereas infections caused by the other three groups of viruses are mainly restricted to regions in Central and in West Africa.⁶ Although it is thought to be transmitted into the human population via a single zoonotic event, the current group M consists of a large number of genetically different variants.⁷ Therefore, group M viruses are further divided into subtypes and sub-subtypes, as well as into stable circulating recombinant forms and unique recombinant forms.^{8,9} Multiple mechanisms contribute to the generation of high diversity in the HIV-1 population to allow the divergence of group M viruses.¹⁰ HIV-1 replicates at high rates in many infected people, generating high viral loads. Additionally, as a member of the Retroviridae family, one of the essential steps of HIV-1 replication is reverse transcription, which is carried out by the virally encoded reverse transcriptase (RT).^{11,12} RT does not have the traditional proofreading function present in cellular DNA polymerases; as a result, RT-mediated DNA synthesis is error prone.¹³ Furthermore, this process has high recombination potential.¹⁴ These features of HIV-1 replication provide the basis for generating viral diversity.¹⁰

One of the hallmarks of HIV-1 replication is frequent recombination.¹⁵ Recombination can occur during DNA synthesis when RT switches between the two copackaged RNAs and uses portions of each RNA as templates for DNA synthesis. If the two copackaged RNAs contain different genetic information, a recombinant distinct from either parent can be generated.¹⁶ Many factors are known to affect recombination between two HIV-1 variants.^{17,18} RNA copackaging is a prerequisite for retroviral recombination; therefore, the frequency at which RNAs generated from different HIV-1 variants can be copackaged into the same particle influences their recombination potential.¹⁹ We have observed that variants with different dimerization initiation signal (DIS) sequences recombine less because their RNAs are copackaged together less frequently.^{20–22} Furthermore, the RNA trafficking pathways can affect

the recombination potential of two viruses,²³ and it has been shown that diversity of nucleotide sequences can affect the template-switching frequency and, hence, the recombination rate.^{24–27}

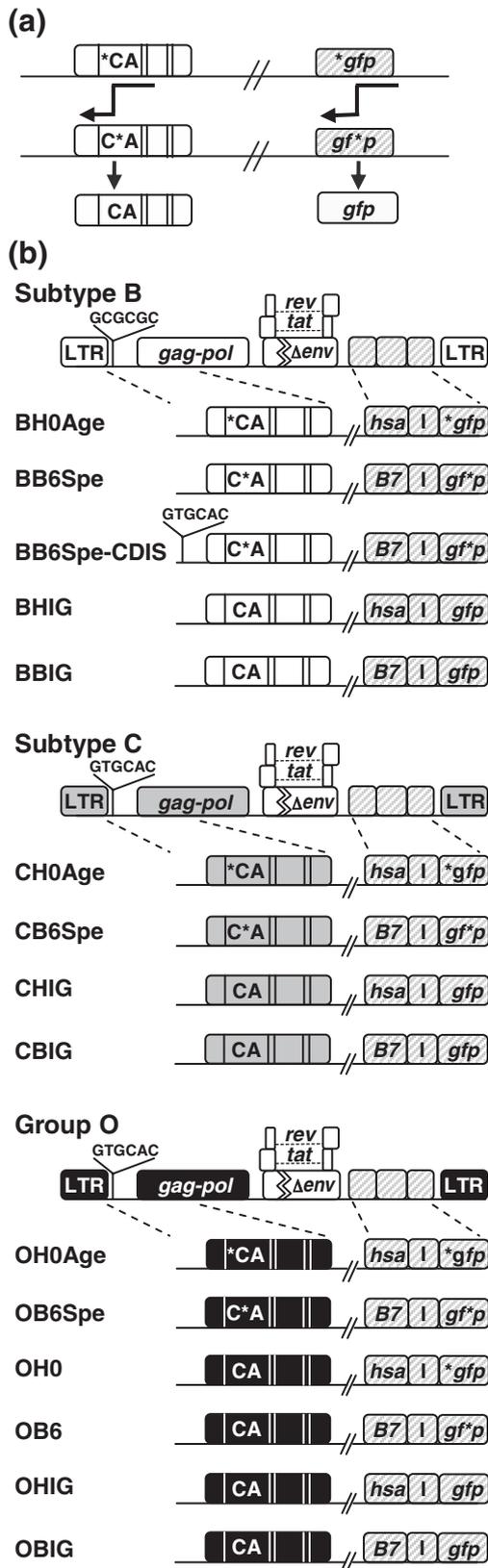
Most of the current studies of HIV-1 recombination focus on group M viruses. Recombination rates of other groups of HIV-1 have not been studied. Dual infection of group O and group M HIV-1 variants have been reported,²⁸ furthermore, intergroup recombinants had been reported.^{28,29} We sought to determine the recombination rate of group O viruses and to explore the potential and the barriers for intergroup HIV-1 recombination. To this end, we first establish a recombination assay that can reflect the nature of nucleic acid sequence diversity among variants from different groups. Previously, we employed a system to study HIV-1 recombination using near-full-length viruses with marker genes inserted in the *nef* gene.¹⁴ In this system, two near-full-length viruses were used, each encoding a surface marker gene and a mutated green fluorescent protein gene (*gfp*). The inactivating mutations in the *gfp* differ in the two HIV-1 vectors; during reverse transcription, recombination can occur to reconstitute a *gfp* gene encoding a functional protein, which can be detected by flow cytometry. The advantage of the system is the ability to easily score a very large number of infection events using flow cytometry and to directly determine recombination rates. However, the *gfp* gene does not reflect some of the characteristics of the viral genome undergoing recombination, such as sequence diversity.

In this report, we developed an improved dual-marker system that can simultaneously detect recombination in the *gfp* gene and in the viral genome, using Gag expression as a marker. Using this system, we examined recombination among group M variants, between two group O viruses and between group M and group O viruses. We found that group O viruses recombine at high rates similar to those of group M viruses. Additionally, group O and group M viruses recombine at low rates even when they carry the same DIS; we have identified distinct mechanisms that contribute to the low intergroup recombination potential.

Results

Strategy to measure recombination using authentic HIV-1 target sequences

We sought to establish an *in vivo* system capable of detecting and quantifying recombination in a HIV-1 population using authentic viral sequences. To achieve this, we modified the existing system to allow the detection of Gag only after a recombination event. Frameshift mutations were introduced



into the capsid (CA) region of *gag* at different positions so that neither parental vector can express a full-length Gag. However, recombination can reconstitute a *gag* encoding a functional polyprotein, which can be detected by antibody staining and flow cytometry (Fig. 1a).

To verify the feasibility of this strategy, we tested first whether Gag expression can be detected reliably by α -p24 antibody staining and flow cytometry. HIV-1 vectors used in the green fluorescent protein (GFP) recombination assay contain near-full-length genomes, including all *cis*-acting elements essential for virus replication, and encode *gag-pol*, *tat* and *rev*. Additionally, a surface marker gene (*hsa* or *B7*), followed by an internal ribosome entry site (IRES) and *gfp*, was inserted in the *nef* region (Fig. 1b). Control viral vectors (BBIG, BHIG, CBIG or CHIG) were transfected into 293T cells along with pNLIII (AD8), an HIV-1 Env expression construct; viruses were harvested and used to infect Hut/CCR5 cells, which were processed and examined by flow cytometry (Fig. 2a). Both subtype B and subtype C Gag can be detected efficiently by this procedure as most of the marker-expressing cells are also positive in Gag expression by antibody staining (Fig. 2b).

We then introduced inactivating mutations in the CA region of *gag* to abolish Gag expression detectable by α -p24 antibody staining and flow cytometry. After testing several mutants (data not shown), we identified two suitable *gag* mutants. The Age mutant contains a +1 frameshift mutation located near the 5' end of the CA-encoding region, and the Spe mutant contains a 4-nt insertion in *gag* near the center of CA (Fig. 1b). The distance between the two mutations is 297 nt, and neither mutant exhibits detectable Gag expression (see below).

Fig. 1. Experimental strategy and HIV-1 vectors used in the recombination assays. (a) Schematic representation of recombination in *gfp* and *gag* genes. Template switching between inactivating mutations within a gene can generate a recombinant expressing functional GFP and/or Gag containing the CA domain. Slanted double lines indicate parts omitted for simplicity. Direction of DNA synthesis is indicated by arrows. (b) General structures of vectors used in this study. All vectors express functional Tat and Rev and carry a deletion in the *env* gene. White, gray and black boxes represent subtype B, subtype C and group O sequences, respectively. Striped boxes indicate marker genes inserted in the *nef* reading frame. The asterisk denotes inactivating mutation. DIS sequences, GCGCGC or GTGCAC, are indicated. The name of each construct is descriptive of its structure: the first letter refers to the group or subtype of the viral sequence; the second letter indicates the surface marker gene (H is for *hsa* and B is for *B7*); numbers mark the position of the inactivating mutation in GFP (0 is for start of the gene and 6 is for 603 nt downstream); and Age or Spe refers to the frameshift mutation toward the 5' end or the 3' end of CA, which are indicated as *CA or C*A, respectively.

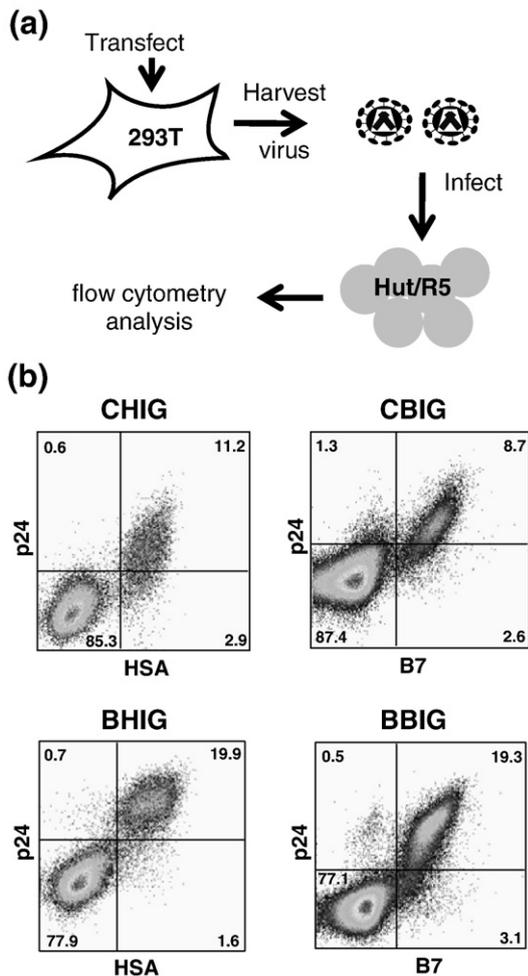


Fig. 2. Detection of Gag expression in infected cells by α -p24 antibody staining and flow cytometry. (a) Experimental protocol used to generate infected cells for Gag detection. Briefly, constructs were used to transfect 293T cells, and viruses were harvested and used to infect target Hut/CCR5 cells, which were processed and analyzed by flow cytometry. (b) Detection of Gag expression in cells infected with subtype C viruses (upper panels) and subtype B viruses (lower panels). Detection of HSA or B7 marker is shown in the x-axis, whereas detection of CA (p24) is shown in the y-axis.

Comparison of HIV-1 recombination detected by Gag or GFP expression

To measure HIV-1 recombination using the Gag reconstitution strategy, we generated cell lines containing two different HIV-1 proviruses: one carrying the *hsa* marker gene and upstream mutations in both *gag* and *gfp* genes and the other carrying the *B7* marker gene and downstream mutations in both *gag* and *gfp* genes (Fig. 1b). These cell lines were established by sequentially infecting 293T cells with HIV-1 vectors at low multiplicities of infection (MOIs) (Fig. 3a). Infected

cells were enriched by several rounds of cell sorting; each final cell line consisted of a pool of more than 100,000 independently infected cells, and more than 97% of the cells expressed heat-stable antigen (HSA) and B7 markers (Fig. 3b, top row). However, these producer cells do not express Gag or GFP (Fig. 3b, middle and lower panels, respectively), confirming that the inactivating mutations indeed abolished the expression of these genes. As these cell lines do not express functional Gag/Gag-Pol or Env, to measure recombination rate, we transduced the producer cells with pSYNGP and pNLIII(AD8), plasmids that express codon-optimized HIV-1 *gag-pol* and HIV-1 *env*, respectively. Viruses were harvested and used to infect Hut/CCR5 cells, which were stained with antibodies and analyzed by flow cytometry. An example of the flow cytometry analyses is shown in Fig. 3c. Virus infection was monitored by the expression of HSA or B7, and recombination was detected by GFP and Gag expression. The number of infected cells was converted to MOI; recombination rate was calculated by dividing the GFP or Gag MOI by the total infection MOI. To ensure the accuracy of measured recombination rates, we only included data points that have >1000 GFP⁺ cells or >500 p24⁺ cells above the background in the negative control (mock-infected cells or uninfected cells).

We first measured recombination between two subtype C viruses, CH0Age and CB6Spe (Fig. 1b). Gag expression was detected in $3.2 \pm 0.1\%$ of infection events, whereas GFP expression was observed in $5.3 \pm 0.6\%$ of infection events (Fig. 4, marked C-C). The genetic distance between the *gag* mutations was shorter than that of the *gfp* mutations (297 nt versus 588 nt), which is likely to account for the lower rate. We also measured the frequencies of GFP and Gag reconstitution between two subtype B viruses, BH0Age and BB6Spe (Fig. 1b), and observed similar rates of recombination, $3.2 \pm 0.3\%$ and $5.1 \pm 0.3\%$ for Gag and GFP, respectively (Fig. 4, marked B-B). In these experiments, the *gag* target sequences between the two vectors are identical; thus, to examine intersubtype recombination *gag* rates, we employed two viruses with different *gag* sequences.

The recombination rates between a subtype C virus and a subtype B virus, CH0Age and BB6Spe, are much lower in both markers (Fig. 4, C-B); Gag and GFP expressions were detected in $0.4 \pm 0.1\%$ and $1.2 \pm 0.2\%$ of infection events, respectively. Subtype B HIV-1 and subtype C HIV-1 contain different DIS sequences, which results in inefficient copackaging of RNAs from these two viruses and a reduced recombination potential. We have previously shown that copackaging of subtype B and subtype C RNAs can be improved by changing the DIS of subtype B virus to that of subtype C.²⁰ Therefore, we also examined recombination between a subtype C virus and a subtype B virus containing the subtype C DIS, namely, CH0Age and BB6Spe-CDIS (Fig. 4,

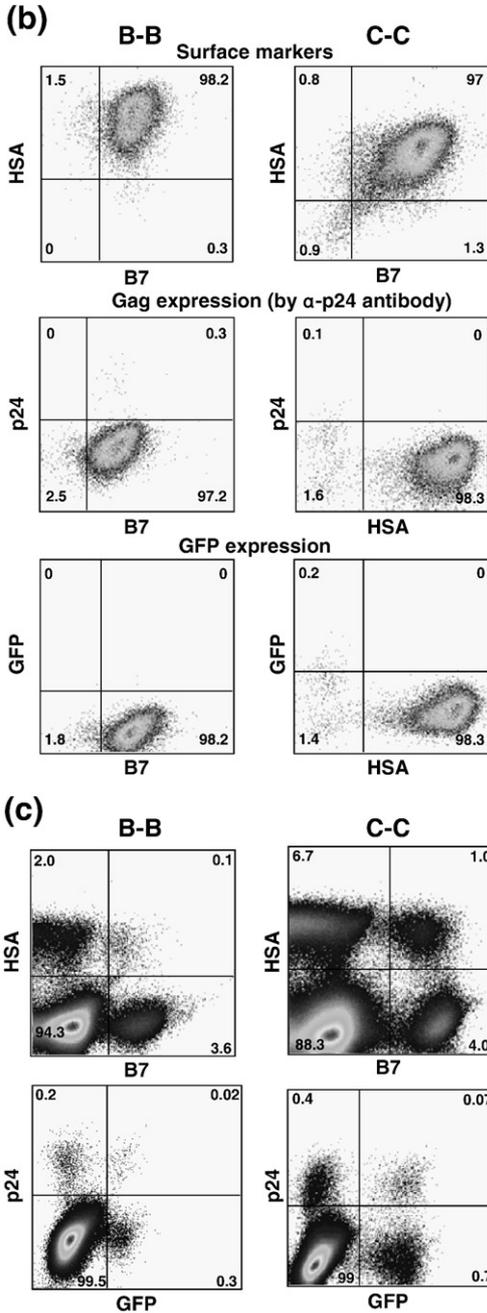
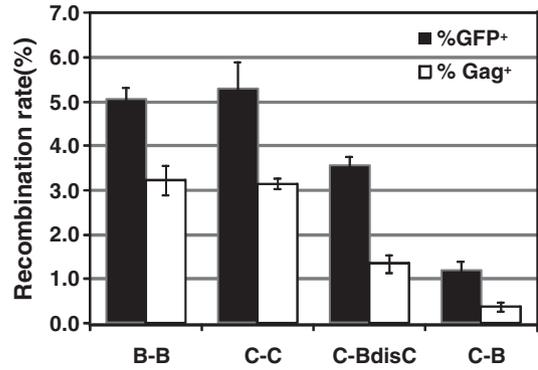
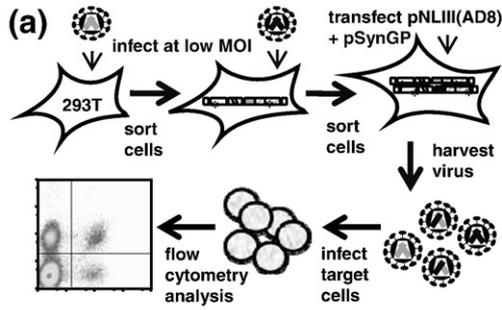


Fig. 4. Recombination of group M viruses measured by *gfp* or *gag* marker gene reconstitution. The *y*-axis indicates the percentage of recombination events measured by GFP⁺ or Gag⁺ among infection events. Black bars indicate the percentage of GFP⁺ events; white bars, the percentage of Gag⁺ events measured by α -p24 antibody staining. Results from three to seven independent experiments were averaged; error bars indicate standard deviations.

C-BdisC); in this case, $1.3 \pm 0.2\%$ and $3.6 \pm 0.2\%$ of infection events had Gag and GFP expressions, respectively. These experiments demonstrated a range of frequencies (3.1–0.4%) at which Gag expression was restored in the viral population. Viral elements, such as DIS, that are known to affect the recombination rates detected by GFP also altered the recombination rates measured by Gag expression. The *gfp* sequences in all of the vectors are identical; thus, GFP reconstitution can be used as an indicator for the recombination potential of these two viruses, taking into accounts issues such as RNA copackaging efficiency. By comparing the Gag and GFP reconstitution frequencies, we can estimate the effect of *gag* sequence divergence on recombination. For example, GFP reconstituted ~ 1.6 -fold to 1.7 -fold more efficiently than Gag when the *gag* target sequences between the two inactivating mutations were identical (Fig. 4, C-C and B-B); GFP reconstituted 2.8-fold more frequently than Gag when the *gag* sequences were derived from different subtypes of HIV-1 (Fig. 4, C-BdisC). Thus, target sequence divergence probably causes lower GFP/Gag reconstitution ratios in intersubtype recombination. Taken together, our results established that Gag can be used as a marker to monitor both intrasubtype and intersubtype HIV-1 recombination.

Fig. 3. Measurements of recombination among group M viruses. (a) Experimental protocol used to establish dual-infected virus producer cell line and to measure recombination. (b) Expression of surface markers, Gag and GFP in selected producer cell lines containing subtype B and subtype C proviruses. Gag expression was detected by α -p24 antibody staining. (c) Representative flow cytometry analyses of target Hut/CCR5 cells from recombination experiments.

Genetic recombination of group O HIV-1

Unlike group M viruses, group O HIV-1 is thought to be originated from SIVgor.³ To determine the recombination rate of group O HIV-1, we generated group O vectors, pOH0 and pOB6 (Fig. 1b), based on the molecular clone CMO2.5.³⁰ We then generated two producer cell lines using the same procedure described in Fig. 3a. More than 97% of cells in the final cell lines expressed both surface markers, and no GFP expression was detected (data not shown). We measured recombination rates using the standard protocol illustrated in Fig. 3a and found that GFP expressed in $4.8 \pm 0.2\%$ of the infection events (data not shown in the figures). Therefore, similar to group M viruses, group O HIV-1 also recombines at high frequency.

Recombination between group M and group O viruses

Next, we assessed whether the Gag recombination assay could be used to measure recombination between highly divergent strains, such as group M and O viruses. To ensure that the antibody we used could detect group O Gag, we infected Hut/CCR5 cells with control viruses OHIG or OBIG (Fig. 1b) using the protocol described in Fig. 2a. Infected cells were stained with α -p24 antibody and antibody against HSA or B7 and were examined using flow cytometry. As shown in Fig. 5a, Gag expression from group O viruses could be detected efficiently using the α -p24 antibody.

To examine recombination between group M and group O viruses, we generated group O viruses with modified *gag* genes. Group O vectors with either Age or Spe frameshift mutation in the *gag* gene do not express detectable levels of Gag (data not shown). Using the protocol described in Fig. 3a, we developed a producer cell line that contained OH0Age and OB6Spe proviruses; more than 91% of these cells expressed both HSA and B7 markers, but they did not have detectable levels of Gag or GFP (Fig. 5b). Using the procedure described in Fig. 3a, we measured the recombination rates; $4.6 \pm 0.5\%$ and $2.2 \pm 0.3\%$ of the infection events had GFP and Gag expressions, respectively (Fig. 5c).

We then examined recombination between a group O virus and a subtype B virus using two cell lines: one cell line containing OH0Age and BB6Spe and the other cell line containing OB6Spe and BH0Age. The target sequences in *CA* and *gfp* in these two cell lines are the same, and similar results were obtained; therefore, we combined the results for simplicity (Fig. 5c, O-B). GFP expression was detected in $0.4 \pm 0.1\%$ of the infection events, whereas the percentage of events with Gag expression could not be determined accurately due to the limited number of positive cells. Group O viruses,

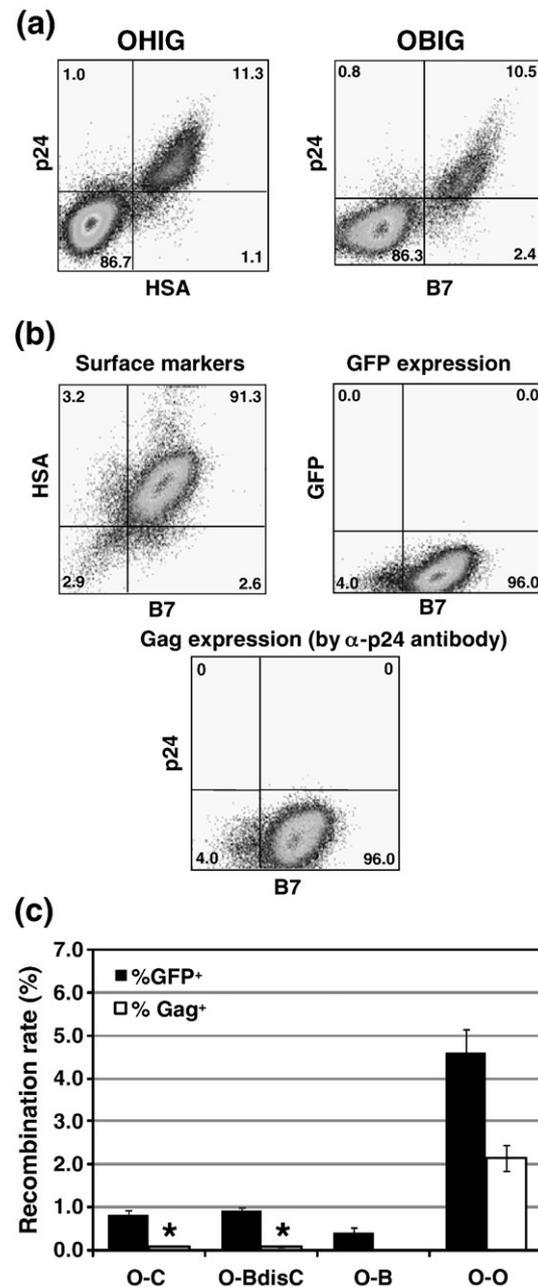


Fig. 5. Measurements of recombination rate between group M and group O viruses. (a) Detection of the group O Gag expression in cells. Target Hut/CCR5 cells were infected with the control group O viruses and were stained with α -p24 and anti-surface marker antibodies. (b) Expression of surface markers, Gag and GFP in producer cell line containing two group O proviruses with *gag* mutations. (c) Intergroup recombination rates. Black and white bars indicate % GFP⁺ and % Gag⁺ cells in infected population, respectively. Except for those marked by *, results from three to six independent experiments were averaged; error bars indicate standard deviations. Each of the two results marked by * is from a single data point, although three (O-C) and four (O-BdisC) repetitions were performed, each yielding sufficient numbers of GFP⁺ cells to calculate GFP recombination rate.

including the molecular clone we used, have the same DIS as subtype C HIV-1, which is different from the DIS sequence of most subtype B variants. To examine whether carrying the same DIS could improve intergroup HIV-1 recombination, we generated a cell line containing a group O virus (OH0Age) and a subtype B virus carrying a subtype C DIS sequence GTGCAC (BB6Spe-CDIS), and we detected an increase in GFP⁺ phenotype to $0.9 \pm 0.1\%$ of the infection events (Fig. 5c, O-BdisC) compared to the 0.4% GFP⁺ frequency when the DIS sequences were different. Additionally, we generated two cell lines containing one group O virus and one subtype C virus: one cell line containing OH0Age and CB6Spe and one cell line containing CH0Age and OB6Spe (combined results shown as O-C in Fig. 5c). Similarly, GFP expression was detected in $0.8 \pm 0.1\%$ of the infection events.

Results shown in Fig. 5c include 16 sets of data that measure GFP reconstitution between a group M virus and a group O virus; most of these experiments scored >2000 GFP⁺ events. If *gag* reconstitution occurred at similar rates among viruses of different groups as that among different subtypes, we should have been able to score >500 p24⁺ cells in most of these experiments. Instead, we were only able to measure two data points: one in O-BdisC (0.07%) and the other in O-C (0.08%), despite the fact that three and four sets of experiments were performed, respectively, each yielding sufficient GFP⁺ cells; these are shown in Fig. 5c with asterisks to indicate that only one datum point was collected. These Gag reconstitution results reveal that, in addition to the barrier revealed by the *gfp* assay, the sequence divergence between group O and group M further prevented efficient recombination in the *gag* gene. Taken together, these results indicated that although matching the DIS sequences did improve the recombination rate between group O and group M viruses, intergroup recombination occurs less frequently than intragroup recombination, indicating that different groups of HIV-1 have a reduced ability to exchange genetic information.

Discussion

In this report, we described an improved experimental system to measure recombination rates between different HIV-1 isolates using authentic viral sequences in the *gag* gene in addition to our previously established *gfp* gene. Our findings showed that the *gag* reconstitution assay can be used to measure recombination between two variants within the same subtype, between viruses from two different subtypes and even between two viruses from different groups. Using this assay, we detected a low level of recombination between a variant from group O HIV-1 and a variant from

group M HIV-1, but we only obtained limited Gag reconstitution during these intergroup recombination experiments, indicating that intergroup recombination occurs much less frequently than intragroup recombination.

Currently, the Gag reconstitution assay has a smaller dynamic range compared with that of the previous GFP assay, as the genetic distance separating the two *gag* mutations is shorter than that of *gfp*. Although the *gag* gene of HIV-1 is approximately 1.5 kb in length, the assay was constrained to a 297-bp separation of the mutations due to the method of detection and a feature of the HIV biology. We used a commercially available α -p24 antibody that is conjugated with a fluorescent dye and is known to be able to detect full-length Gag. The translation of HIV-1 Gag can be initiated from the start codon of *gag* in matrix or from an AUG in CA, using an IRES.^{31,32} As the mutations have to abolish Gag expression from both translation start sites, this severely limited the potential locations at which mutations can be placed. It is foreseeable that, in the future, the dynamic range of the Gag reconstitution assay can be lengthened significantly if a suitable antibody targeting NC or p6 is available that can reliably detect unprocessed Gag polyprotein.

In this report, we examined for the first time the recombination potential of group O viruses and found that group O viruses, similar to group M viruses, recombine frequently. Although the members of group M viruses can contain different DIS sequences,^{33,34} to our knowledge, all group O variants contain the DIS sequence GTGCAC, which is identical with that of most subtype C viruses. Although RNAs containing different DIS sequences can be copackaged into the same particles, it occurred at significantly reduced rate than that of RNAs with the same DIS sequences among group M variants. Our GFP reconstitution assay revealed that the DIS sequences also affect the intergroup recombination rates. However, recombination between a group O virus and a group M virus is approximately fourfold lower than that of two group M variants, even when the viruses have the same DIS sequences (Fig. 5, O-C and O-BdisC versus Fig. 4, C-BdisC; $p < 0.001$, two-way ANOVA). These results suggest that factors other than DIS contribute to reduce recombination rates.

Group O and group M viruses have reduced sequence similarity. Template switching by RT is known to be reduced between two regions with low sequence identity; however, the *gfp* marker gene used to monitor recombination has the same sequence in the group O and group M vectors. Therefore, sequence diversity should not directly affect template switching in *gfp*, but was nonetheless capable of reducing the recombination rate in this region. One possible mechanism is that sequence differences may reduce copackaging of group O and

group M RNAs despite their carrying the same DIS, as additional sequence compatibility is required for efficient copackaging into the same virion. Alternatively, the conformation of the two copackaged RNAs within the core may differ between the viruses, thereby hampering RT switching from one RNA to the other during DNA synthesis. Further examination will be required to distinguish between these two possibilities.

Although the sequence of the *gfp* gene in all vectors tested was identical, their *gag* sequences differed. In this case, sequence diversity does play a direct role in reducing Gag reconstitution during recombination. Intragroup recombination with matched DIS sequences (Fig. 4, C-BdisC) was found to be approximately threefold more efficient in GFP than in Gag when the two strains we used have 86% sequence identity in *gag* and 91% sequence identity in the 297-nt target region. In intergroup recombination, the molecular clones we used have 69% sequence identity in *gag* (both O-B and O-C) and 73% (O-B) and 72% (O-C) sequence identities in the 297-nt target region for recombination measurement. Our inability to measure an accurate Gag reconstitution rate for intergroup recombination, despite large number of GFP⁺ events scored in these experiments, clearly indicates a much reduced intergroup *gag* recombination rate compared with the intragroup rate. Our limited data sets (two measurements) estimated that GFP is reconstituted 10-fold to 13-fold more efficiently than Gag even when the two viruses have the same DIS (0.83–0.93% compared with 0.07–0.08%, Fig. 5c, O-C and O-BdisC), indicating that sequence divergence further reduced intergroup recombination in the *gag* gene. Taken together, these results show that sequence divergence affects intergroup recombination in three distinct ways: a copackaging mechanism mediated by DIS, an indirect mechanism that decreases recombination of homologous sequences (*gfp*) by affecting copackaging independent of DIS or accessibility of the templates and a direct mechanisms reducing template switching in divergent regions (*gag*).

Compared with group M virus infection, group O HIV-1 infection is limited both in the number of infected individuals and in its geographical distribution. Dual infection of group O and group M viruses has been observed,²⁸ and recombinants between them have been isolated from the human population.^{28,29} Our findings indicate that group O and group M variants recombine far less frequently than two viruses of the same group. Previously, we have shown that many B/F intersubtype recombinants have difficulty completing the virus life cycle and were eliminated during virus replication.²⁷ Considering that the genetic distance between group O and group M viruses is greater than that of two subtype variants, it is most likely that many

of the group O and group M recombinants are not viable and would be eliminated through purifying selection. Therefore, O/M recombinants have to overcome multiple barriers to emerge in the human population. The observations of replication-competent O/M recombinants demonstrate the power of HIV-1 recombination and the ability of the HIV-1 genome to evolve.

Materials and Methods

Plasmid construction and nomenclature

The names of all plasmids used in this study begin with “p”, whereas the names of the viruses derived from these plasmids do not. All of the modified HIV-1-based vectors used in this study contain *cis*-acting elements essential for virus replication: *gag-pol*, *tat* and *rev*. Additionally, two marker genes were inserted into the *nef* gene; the first marker gene is either a mouse CD24 HSA gene (*hsa*) or a human CD80 gene (*B7*) and the second marker gene is a GFP gene (*gfp*). The translation of GFP is facilitated by the encephalomyocarditis virus IRES. For simplicity, the first letter of the vector name refers to the group or subtype of the viral sequence; the second letter indicates the surface marker gene (H is for *hsa* and B is for *B7*); numbers mark the position of the inactivating mutation in GFP (0 is for start of the gene and 6 is for 603 nt downstream). For example, pBH0 refers to a plasmid containing subtype B HIV-1-modified genome expressing HSA and having a mutated *gfp* gene with a frameshift at the 5' end of the gene. The *gag* genes of many vectors contain a frameshift mutation in the CA region; either a +1 frameshift at the N-terminus of CA (referred to as Age) or a +4 frameshift at the C-terminus of CA (referred to as Spe).

For clarity, the previously described subtype B vectors pON-fHIG and pBIG are referred to as pBHIG and pBBIG;^{14,23} these two vectors contain functional *gag* and *gfp* and are used as expression and staining controls.

Construct pBH0Age was modified from pON-H0 by adding a +1 frameshift at nucleotide position 1215 (pNL4-3 numbering).³⁵ Plasmid pBB6Spe was modified from pT6-RRE-Spe by adding a +4 frameshift at position 1508 (pNL4-3 numbering) and by replacing the *thy1.2* marker gene with *B7*.³⁵ Plasmid pBB6Spe-disC was generated by replacing endogenous DIS (GCGCGC) sequence in pBB6Spe with subtype C DIS (GTGCAC).

Subtype C construct pCH0Age was modified from pCH0 by adding a +1 frameshift at position 1228 (pMJ4 numbering),²⁰ whereas pCB6Spe was modified from pCH0 by adding a +4 frameshift at the 1522 position (pMJ4 numbering) and by replacing the DNA fragments containing *hsa*-IRES-**gfp* with fragment containing *B7*-IRES-**gfp* from pBB6Spe. Plasmids pCHIG and pCBIG were modified from pCH0 by replacing a DNA fragment containing *hsa*-IRES-**gfp* with DNA fragments containing either *hsa*-IRES-**gfp* or *B7*-IRES-**gfp* from pBHIG and pBBIG, respectively.

Group O vectors were derived from pCMO2.5 (accession number AY623602),³⁰ a generous gift from Dr. Hans-Georg Krausslich. Plasmid pOH0 was generated by first replacing the long terminal repeat (LTR)-*gag-pol* region of

pON-H0 with that of pCMO2.5 to generate pCMO2.5-X-HO. This was carried out by digesting pCMO2.5 with BsrBI and NdeI restriction enzymes, and the fragment containing LTR-*gag-pol* was cloned into the NdeI-digested and ZraI-digested pON-H0. An XhoI site was introduced into the *nef* reading frame of pCMO2.5, and the *nef*-LTR region of the resulting plasmid was used to replace the equivalent region in pCMO2.5-X-HO using XhoI and NgoMIV restriction sites. In plasmid pOH0Age, a +1 frameshift was introduced at nucleotide 1260 (ANT70 numbering) of the pOH0. Plasmid pOB6 was generated by replacing the DNA fragment containing the *hsa*-IRES-**gfp* region of pOH0 with a DNA fragment containing *B7-IRES-gfp** from pBB6Spe. Plasmid pOB6Spe was generated by introducing the +4 frameshift at nucleotide 1553 (ANT70 numbering) into pOB6. Control plasmids pOHIG and pOBIG were cloned by replacing the *hsa*-IRES-**gfp* marker cassette in pOH0 with *hsa*-IRES-*gfp* or *B7-IRES-gfp* fragment from pBHIG and pBBIG, respectively.

Standard molecular cloning techniques were used to construct plasmids.³⁶ The general structure of all plasmids was mapped by restriction enzyme digestion. In addition, to ensure the absence of inadvertent mutations introduced by PCR amplification, we verified all plasmid DNA fragments derived from PCR by sequencing.

Cell culture, transfection, infection and generation of producer cell lines

Human embryonic kidney cell line 293T and human T cell line Hut/CCR5 were maintained in Dulbecco's modified Eagle's medium and Roswell Park Memorial Institute-1640 medium, respectively. Both media were supplemented with 5% fetal calf serum, 5% calf serum, penicillin (50 U/ml) and streptomycin (50 U/ml). Additionally, puromycin (1 µg/ml) and G418 (500 µg/ml) were added to the medium used for Hut/CCR5 cells. All cultured cells were maintained at 37 °C with 5% CO₂.

To generate viral stocks for the production of cell lines, we transfected 293T cells using 10 µM 25-kDa branched polyethyleneimine (Sigma) with viral constructs, pCMV-G and pSYNGP,^{37,38} which express the G protein from vesicular stomatitis virus and codon-optimized HIV-1 Gag/Gag-Pol, respectively. Viruses were harvested 48 h posttransfection, clarified through a 0.45-µm-pore-size filter and used immediately or stored at -80 °C prior to infection. The producer cell lines were generated by sequential infection of low-passage 293T cells by two parental viruses at a low MOI of less than 0.1. Dually infected cells were enriched by several rounds of cell sorting so that more than 91% of the cells expressed markers encoded by both vectors. Producer cell lines were also stained with anti-p24 antibodies to examine Gag/CA expression.

For recombination experiments, producer cells were transfected using TransIT®-LT1 (Mirus) or FuGENE HD (Roche) with pSYNGP that expresses codon-optimized HIV-1 Gag/Gag-Pol and pIINL(AD8)env,³⁷ which expresses HIV-1 Env. Viruses were harvested 48 h posttransfection, clarified through 0.45-µm-pore-size filter and used to infect target Hut/CCR5 cells by spinoculation in 12-well plates (1 × 10⁶ cells/well) or in 6-well plates (2 × 10⁶ cells/well) at 1200g for 1.5–2 h at 25 °C. Target cells

were stained and processed 3 days postinfection and were analyzed by flow cytometry.

Flow cytometry analyses and cell sorting

The following fluorescently labeled antibodies were used in various combinations in this study: phycoerythrin (PE)-conjugated α-mCD24(HSA) antibody (BD Biosciences), biotin-conjugated α-mCD24(HSA) antibody (BD Biosciences) in combination with PerCP-conjugated streptavidin (BD PharMingen), allophycocyanin-conjugated α-hCD80(B7) MEM-233 antibody (Novus Biologicals), Alexa Fluor®647-conjugated α-hCD80(B7) antibody (BioLegend) and PE-conjugated α-p24 KC57-RD1 antibody (Beckman Coulter).

The following protocol was used to stain cells for flow cytometry analyses: 2 × 10⁶ cells were washed twice in a 3 ml of phosphate-buffered saline containing 2.5% fetal calf serum. The resulting cells were incubated with biotin-conjugated α-HSA antibodies (at 1:500 dilution), washed twice and incubated with Alexa Fluor®647 (or allophycocyanin)-conjugated α-B7 antibodies (at 1:100 dilution) and streptavidin-PerCP (at 1:500 dilution). Cells were then washed twice and fixed with 4% formaldehyde, washed twice and incubated with PE-conjugated α-p24 antibodies (at 1:40 dilution in phosphate-buffered saline containing 0.5% saponin and 2.5% fetal calf serum) and washed twice and fixed with 2% formaldehyde. All incubations with antibodies and formaldehyde were carried in a 200-µl volume for 30 min at 25 °C in the dark.

Flow cytometry analyses were performed on a FACSCalibur system (BD Biosciences), and the data obtained were analyzed using the FlowJo software (Tree Star). For each recombination experiment, uninfected target cells were used as negative controls. The background levels in negative controls were generally very low and were subtracted from the values in the positive samples during data analysis. Cell sorting was performed on ARIA II system (BD Biosciences).

Acknowledgements

We thank Hans-Georg Krausslich for group O molecular clone. This research was supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research.

References

1. Sharp, P. M. & Hahn, B. H. (2010). The evolution of HIV-1 and the origin of AIDS. *Philos. Trans. R. Soc. London, Ser. B*, **365**, 2487–2494.
2. Gao, F., Bailes, E., Robertson, D. L., Chen, Y., Rodenburg, C. M., Michael, S. F. *et al.* (1999). Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature*, **397**, 436–441.
3. Van Heuverswyn, F., Li, Y., Neel, C., Bailes, E., Keele, B. F., Liu, W. *et al.* (2006). Human immunodeficiency viruses: SIV infection in wild gorillas. *Nature*, **444**, 164.

4. Plantier, J. C., Leoz, M., Dickerson, J. E., De Oliveira, F., Cordonnier, F., Lemeé, V. *et al.* (2009). A new human immunodeficiency virus derived from gorillas. *Nat. Med.* **15**, 871–872.
5. McCutchan, F. E. (2006). Global epidemiology of HIV. *J. Med. Virol.* **78**, S7–S12.
6. Van Heuverswyn, F. & Peeters, M. (2007). The origins of HIV and implications for the global epidemic. *Curr. Infect. Dis. Rep.* **9**, 338–346.
7. Thomson, M. M., Perez-Alvarez, L. & Najera, R. (2002). Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy. *Lancet Infect. Dis.* **2**, 461–471.
8. Leitner, T., Korber, B., Daniels, M., Calef, C. & Foley, B. (2005). HIV-1 subtype and circulating recombinant form (CRF) reference sequences. In *HIV Sequence Compendium* (Leitner, T., Foley, B., Hahn, B. H., Marx, P. A., McCutchan, F. E., Mellors, J. W., Wolinsky, S. & Korber, B., eds), pp. 41–48, Los Alamos National Laboratory, Los Alamos, NM.
9. Robertson, D. L., Anderson, J. P., Bradac, J. A., Carr, J. K., Foley, B., Funkhouser, R. K. *et al.* (2000). HIV-1 nomenclature proposal. *Science*, **288**, 55–56.
10. Coffin, J. M. (1995). HIV population dynamics *in vivo*: implications for genetic variation, pathogenesis, and therapy. *Science*, **267**, 483–489.
11. Temin, H. M. & Mizutani, S. (1970). RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature*, **226**, 1211–1213.
12. Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature*, **226**, 1209–1211.
13. Mansky, L. M. & Temin, H. M. (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**, 5087–5094.
14. Rhodes, T. D., Nikolaitchik, O., Chen, J., Powell, D. & Hu, W. S. (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. *J. Virol.* **79**, 1666–1677.
15. Robertson, D. L., Sharp, P. M., McCutchan, F. E. & Hahn, B. H. (1995). Recombination in HIV-1. *Nature*, **374**, 124–126.
16. Hu, W. S. & Temin, H. M. (1990). Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proc. Natl Acad. Sci. USA*, **87**, 1556–1560.
17. Galetto, R. & Negroni, M. (2005). Mechanistic features of recombination in HIV. *AIDS Rev.* **7**, 92–102.
18. Onafuwa-Nuga, A. & Telesnitsky, A. (2009). The remarkable frequency of human immunodeficiency virus type 1 genetic recombination. *Microbiol. Mol. Biol. Rev.* **73**, 451–480, Table of Contents.
19. Moore, M. D. & Hu, W. S. (2009). HIV-1 RNA dimerization: it takes two to tango. *AIDS Rev.* **11**, 91–102.
20. Chin, M. P. S., Rhodes, T. D., Chen, J., Fu, W. & Hu, W. S. (2005). Identification of a major restriction in HIV-1 intersubtype recombination. *Proc. Natl Acad. Sci. USA*, **102**, 9002–9007.
21. Moore, M. D., Fu, W., Nikolaitchik, O., Chen, J., Ptak, R. G. & Hu, W. S. (2007). Dimer initiation signal of human immunodeficiency virus type 1: its role in partner selection during RNA copackaging and its effects on recombination. *J. Virol.* **81**, 4002–4011.
22. Chen, J., Nikolaitchik, O., Singh, J., Wright, A., Bencsics, C. E., Coffin, J. M. *et al.* (2009). High efficiency of HIV-1 genomic RNA packaging and heterozygote formation revealed by single virion analysis. *Proc. Natl Acad. Sci. USA*, **106**, 13535–13540.
23. Moore, M. D., Nikolaitchik, O. A., Chen, J., Hammarskjöld, M. L., Rekosh, D. & Hu, W. S. (2009). Probing the HIV-1 genomic RNA trafficking pathway and dimerization by genetic recombination and single virion analyses. *PLoS Pathog.* **5**, e1000627.
24. An, W. & Telesnitsky, A. (2002). Effects of varying sequence similarity on the frequency of repeat deletion during reverse transcription of a human immunodeficiency virus type 1 vector. *J. Virol.* **76**, 7897–7902.
25. Baird, H. A., Gao, Y., Galetto, R., Lalonde, M., Anthony, R. M., Giacomoni, V. *et al.* (2006). Influence of sequence identity and unique breakpoints on the frequency of intersubtype HIV-1 recombination. *Retrovirology*, **3**, 91.
26. Dang, Q. & Hu, W. S. (2001). Effects of homology length in the repeat region on minus-strand DNA transfer and retroviral replication. *J. Virol.* **75**, 809–820.
27. Galli, A., Kearney, M., Nikolaitchik, O. A., Yu, S., Chin, M. P. S., Maldarelli, F. *et al.* (2010). Patterns of human immunodeficiency virus type 1 recombination *ex vivo* provide evidence for coadaptation of distant sites, resulting in purifying selection for intersubtype recombinants during replication. *J. Virol.* **84**, 7651–7661.
28. Yamaguchi, J., Bodelle, P., Vallari, A. S., Coffey, R., MacArthur, C. P., Schochetman, G. *et al.* (2004). HIV infections in northwestern Cameroon: identification of HIV type 1 group O and dual HIV type 1 group M and group O infections. *AIDS Res. Hum. Retroviruses*, **20**, 944–957.
29. Takehisa, J., Zekeng, L., Ido, E., Yamaguchi-Kabata, Y., Mboudjeka, I., Harada, Y. *et al.* (1999). Human immunodeficiency virus type 1 intergroup (M/O) recombination in Cameroon. *J. Virol.* **73**, 6810–6820.
30. Tebit, D. M., Zekeng, L., Kaptué, L., Gürtler, L., Fackler, O. T., Keppler, O. T. *et al.* (2004). Construction and characterization of an HIV-1 group O infectious molecular clone and analysis of *vpr*- and *nef*-negative derivatives. *Virology*, **326**, 329–339.
31. Buck, C. B., Shen, X., Egan, M. A., Pierson, T. C., Walker, C. M. & Siliciano, R. F. (2001). The human immunodeficiency virus type 1 *gag* gene encodes an internal ribosome entry site. *J. Virol.* **75**, 181–191.
32. Poon, D. T., Chertova, E. N. & Ott, D. E. (2002). Human immunodeficiency virus type 1 preferentially encapsidates genomic RNAs that encode Pr55^{Gag}: functional linkage between translation and RNA packaging. *Virology*, **293**, 368–378.
33. St Louis, D. C., Gotte, D., Sanders-Buell, E., Ritchey, D. W., Salminen, M. O., Carr, J. K. & McCutchan, F. E. (1998). Infectious molecular clones with the nonhomologous dimer initiation sequences found in

- different subtypes of human immunodeficiency virus type 1 can recombine and initiate a spreading infection *in vitro*. *J. Virol.* **72**, 3991–3998.
34. Hussein, I. T., Ni, N., Galli, A., Chen, J., Moore, M. D. & Hu, W. S. (2010). Delineation of the preferences and requirements of the human immunodeficiency virus type 1 dimerization initiation signal by using an *in vivo* cell-based selection approach. *J. Virol.* **84**, 6866–6875.
 35. Nikolaitchik, O., Rhodes, T. D., Ott, D. & Hu, W. S. (2006). Effects of mutations in the human immunodeficiency virus type 1 *gag* gene on RNA packaging and recombination. *J. Virol.* **80**, 4691–4697.
 36. Sambrook, J. & Russel, D. W. (2001). *Molecular Cloning: A Laboratory Manual*, 3rd edit. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
 37. Yee, J. K., Miyanochara, A., LaPorte, P., Bouic, K., Burns, J. C. & Friedmann, T. (1994). A general method for the generation of high-titer, pantropic retroviral vectors: highly efficient infection of primary hepatocytes. *Proc. Natl Acad. Sci. USA*, **91**, 9564–9568.
 38. Kotsopoulou, E., Kim, V. N., Kingsman, A. J., Kingsman, S. M. & Mitrophanous, K. A. (2000). A Rev-independent human immunodeficiency virus type 1 (HIV-1)-based vector that exploits a codon-optimized HIV-1 *gag-pol* gene. *J. Virol.* **74**, 4839–4852.