Longitudinal Quasispecies Analysis of Viral Variants in HIV Type 1 Dually Infected Individuals Highlights the Importance of Sequence Identity in Viral Recombination

Rebecca L.R. Powell,¹ Lynchy Lezeau,² Thompson Kinge,³ and Phillipe N. Nyambi^{4,5}

Abstract

Little is known regarding the likelihood of recombination between any given pair of nonidentical HIV-1 viruses *in vivo*. The present study analyzes the HIV-1 quasispecies in the C1C2 region of *env*, the *vif-vpr-vpu* accessory gene region, and the reverse transcriptase region of *pol*. These sequences were amplified from samples obtained sequentially over a 12- to 33-month period from five dually HIV-1-infected subjects. Analysis of an average of 248 clones amplified from each subject revealed no recombinants within the three loci studied of the subtype-discordant infecting strains, whose genetic diversity was >11% in *env*. In contrast, two subjects who were initially coinfecting by two subtype-concordant variants with genetic diversity of 7.4% in *env* were found to harbor 10 unique recombinants of these strains, as exhibited by analysis of the *env* gene. The frequent recombination observed among the subtype-concordant strains studied herein correlates with prior sequence analyses that have commonly found higher rates of recombination at loci bearing the most conserved sequences, demonstrating an important role for sequence identity in HIV-1 recombination. Viral load analysis revealed that the samples studied contained an average of 8125 virus copies/ml (range, 882–31,626 copies/ml), signifying that the amount of viral RNA in the samples was not limiting for studying virus diversity. These data reveal that recombination between genetically distant strains may not be an immediate or common outcome to dual infection *in vivo* and suggest critical roles for viral and host factors such as viral fitness, virus diversity, and host immune responses that may contribute to limiting the frequency of intersubtype recombination during *in vivo* dual infection.

Introduction

SEVERAL *IN VITRO* STUDIES HAVE BEEN PERFORMED OVER the past decade in an attempt to better understand the mechanisms of HIV-1 recombination and the formation of unique recombinant forms (URFs).^{1–9} Among such studies includes one published by Baird *et al.*, which examined the C1C4 region of *env* for recombination between two discordant HIV-1 subtypes, A and D, in cell culture demonstrating an abundance of recombinants and revealing recombination breakpoints occurring more frequently in the constant than in the variable regions of the viral envelope.⁸ Subsequent *in vitro* studies with these discordant strains also revealed that factors such as replicative fitness contribute to the frequency at which two viral strains recombine.⁷ Furthermore, analyses of several

recombinant viruses have revealed that recombination appears to occur most frequently in the more conserved regions of the envelope and in the peripheries of the *env* gene, as well as in other conserved regions such as the reverse transcriptase (RT)-RNase region of *pol* and the *vif-vpr-vpu* loci.^{5,8,10} Taken together, the lack of recombination occurring in the relatively variable regions of the viral genome and the high frequency of recombination described within relatively conserved sequences suggest a role for sequence identity in enhancing the frequency of viral recombination during dual infection.

Clearly, it is of critical importance to study actual dual infections *in vivo*, wherein, in addition to the sequence identities of the dually infecting viruses, factors such as viral replicative fitness and host immune pressure likely affect the frequency of HIV-1 recombination. In a recent study by Piantadosi *et al.*,

¹Department of Microbiology, New York University School of Medicine, New York, New York.

²Brooklyn College, The City University of New York, Brooklyn, New York.

³Ministry of Public Health, Yaoundé, Cameroon.

⁴Department of Pathology, New York University School of Medicine, New York, New York.

⁵Veterans Affairs New York Harbor Healthcare Systems, New York, New York.

of seven dually infected Kenyan women, only one individual was found to harbor a virus that was a recombinant of the parental viruses (subtypes A and D) in the V1V5 region of *env*.³ This virus was found in the minority of sequences at 73 days postinfection but absent for the next 4.5 years until its reappearance as a 25% proportion of *env* sequences.³ In a study by Gerdhart *et al.* that analyzed the *vpu/gp120* region of *env* sequences from specimens obtained at 3-month intervals from a subject triply infected by two strains of subtype A and a subtype C virus, exhibiting symptoms of late-stage disease, *env* analysis identified several URFs; however, these recombinants always comprised a small minority (<1%) of the viral quasiespecies in the individual at each of the time points analyzed.¹¹

To best examine the role of sequence identity in the generation of recombinants, individuals dually infected with concordant as well as discordant HIV-1 subtypes must be studied. Few studies have examined the emergence and evolution of recombinants at frequent intervals following dual infection in their hosts, which would best identify recombinants as they appear and disappear over the course of infection. In a recent study, we determined the frequency of dual HIV-1 infection occurring in Cameroon, West Central Africa, where diverse HIV-1 subtypes cocirculate.¹² Our analysis of the p7p24 region of *gag* amplified from patient plasma obtained at 3- to 6-month intervals over 3–4 years revealed a dual infection rate of 16% occurring in Cameroon. The present study analyzes the quasiespecies dynamics of the viruses dually infecting five of these subjects at three genomic loci, including the C1C2 region of *env* (~11,00 bp at the 5' end of *env*), the RT region of *pol*, and the *vif-vpr-vpu* (VVF) accessory gene region (~1500 bp and ~1300 bp, respectively). These three loci were selected based on the relatively higher frequency of recombination expected within, as found and predicted by *in vitro* studies and *in silico* models.^{2,5,7,8,10}

Materials and Methods

Study subjects

Blood samples were collected at 3-month intervals over a 3- to 4-year period from five asymptomatic, antiretroviral drug-naïve, and chronically and dually HIV-1-infected individuals in Cameroon. Of the five subjects, three were female and two were male. The three females included CMNYU107, 23 years old, CMNYU124, 35 years old, and CMNYU129, 43 years old. The two males were CMNYU6518, 22 years old, and CMNYU6544, 36 years old. All subjects declared heterosexual contact/multiple partners as their most likely mode of HIV-1 infection.

Prior *gag* analysis found subjects CMNYU107 and CMNYU6518 to be epidemiologically linked based on their initial time point samples.¹² Both subjects were initially infected by the same virus (CRF01_AE in *gag*), with each subject becoming dually infected by a different strain. For subject CMNYU107, the dually infecting strain (subtype G in *gag*) virtually overtook the quasiespecies at month 21, after which time this virus could no longer be identified. For subject CMNYU6518, the dually infecting strain (CRF02_AG in *gag*) appeared at month 3 and appeared to take over the quasiespecies completely for the remainder of the study. Subjects CMNYU124 and CMNYU129 were also shown to be infected by epidemiologically linked viruses. Initially, CMNYU124

was infected by a subtype F2 virus (in *gag*), and identified as dually infected by a CRF02_AG virus (in *gag*) at month 9. Conversely, CMNYU129 was shown to be initially infected by a CRF02_AG virus (in *gag*), and identified as dually infected at month 9 by a subtype F2 virus (in *gag*). In this case, both the CRF02_AG and F2 viruses appeared directly related. In both cases, the dually infecting strain appeared only at the single time point, after which time it could no longer be identified. Lastly, subject CMNYU6544 was found to be initially infected by a subtype F2 virus (in *gag*), and identified as dually infected at month 12 by a CRF02_AG virus (in *gag*). Analysis of the *gag* gene in our previous study did not identify the coexistence of the dually infecting strains in any of these five subjects.¹²

CD4 cell counts and viral loads

CD4 cell counts were measured by FACSCount (Becton Dickinson, Mountain View, CA) at each sampling time point. Supplemental Table 1 presents a summary of these data (Supplemental Table 1; see www.liebertonline.com), obtained from samples for which sequence analysis was also performed.

The viral load of each sample was determined by the Versant HIV RNA 3.0 Assay (bDNA; Siemens, IL), as recommended by the manufacturer. Viral loads were available for 19/29 samples analyzed, as summarized in Supplemental Table 1.

PCR and sequence analysis

Plasma was obtained by Ficoll-Hypaque gradient centrifugation of whole blood. Viral RNA was extracted from plasma using the QIAamp Viral RNA Mini kit (Qiagen Inc., Valencia, CA). Between 0.1 and 10 μ l of RNA (depending on viral load) was used for one-tube reverse transcriptase polymerase chain reaction (RT-PCR) using the Superscript One-Step RT-PCR for Long Templates kit (Invitrogen, Carlsbad, CA) to amplify the C1C2 region of *env* (HxB2 location 6207–7342), the RT region of *pol* (HxB2 location 2612–4159), and/or the VVF accessory gene region (HxB2 location 5004–6297). One microliter of first-round product was then used in a nested PCR using the Platinum PCR SuperMix High Fidelity system (Invitrogen, Carlsbad, CA). The C1C2 region was amplified using the outer primers ENV A 5' AAAAGGCTTA GGCATCTCCTATGGCAGGAAGAA 3' (forward) and tES8 5' TGAGGGACAATTGGAGAAGTG 3' (reverse), and inner primers ENV B 5' AGAAAGAGCAGAAGACAGTGGCA 3' (forward) and C3rev 5' GTGTTGTAATTCTARGTC 3' (reverse). The amplification protocol for the one-tube RT-PCR was one cycle of 50°C for 30 min and 94°C for 2 min, followed by 40 cycles of 94°C for 15 s, 50°C for 30 s, and 68°C for 2 min, ending with a single extension cycle at 72°C for 7 min. For the second-round PCR, the above protocol was used, beginning with the 94°C denaturation, with a 1 min extension, 55°C annealing temperature, for 35 cycles. The VVF region was amplified using the outer primers 5' GGTATTACAGGGA CAGCAGAG 3' (forward) and 5' GGAAAAATAACATGGT AGA(G/A)CA 3' (reverse), and inner primers 5' GTAGTAC CAAGAAGAAAAGC 3' (forward) and 5' C(C/T)(C/T) TATTTTGTGCATCAGATGC 3' (reverse). The amplification protocols were as listed above, with a 1.5 min extension in the second-round PCR. The *pol* region was amplified using primers described previously, and the same cycling conditions as listed above for the VVF region.¹³ In most cases,

amplicons from at least three RT-PCRs of each region were cloned separately into *TopoTA* cloning vectors and plasmids were transformed into chemically competent *Escherichia coli* cells (Invitrogen, Carlsbad, CA). Positive colonies were selected and sequenced at the 5' and 3' ends with universal T7 and T3 primers and additional primer walking if necessary. Sequences were assembled using the Staden software package.¹⁴ The sequences described in this study have been deposited in GenBank and are available under the accession numbers FJ829078–FJ829241, GQ222716–GQ223057, GQ432962–GQ433345, and GQ432775–GQ432961.

Sequences were automatically aligned with identical regions of reference sequences of all available HIV-1 group M subsubtypes and circulating recombinant forms (CRFs) from the Los Alamos HIV Sequence Database (LANLHSD) using MAFFT, and manually cropped using Seaview.^{15,16} Phylogenetic analyses were performed using MEGA 4.¹⁷ Pairwise evolutionary distances were estimated using Kimura's two-parameter method, and phylogenetic trees were constructed by the neighbor-joining method.^{18,19} The reliability of tree topologies was estimated by bootstrap analysis (1000 replicates).²⁰ Sequences were further analyzed for recombination breakpoints with SimPlot, version 3.5.1, using a 250-bp window moving along the alignment with increments of 30 bp.²¹

Results

Analysis of viral sequences from subjects CMNYU107 and CMNYU6518

Analysis of C1C2env. One hundred and seventeen clones from subject CMNYU107 and 59 clones from subject CMNYU6518 were studied from samples obtained over 30 and 15 months, respectively. All clones from CMNYU107 amplified from samples obtained at months 0, 9, 15, 27, and 30 were found to cluster closest to the CRF36_cpx references ("Strain 1," bootstrap value, 89%; Fig. 1). In contrast, all clones from the sample obtained at month 21 clustered with subtype G, consistent with previous data indicating this study subject was dually infected at this time point.¹² The average distance among sequences obtained at month 21 was 1.4%.

All clones amplified from samples taken from CMNYU6518 at month 0 clustered among the CRF36_cpx-related branches of CMNYU107 sequences ("Strain 1"; Fig. 1), demonstrating these subjects to be epidemiologically related, as previously deduced.¹² Clones amplified from samples obtained from CMNYU6518 over the 12-month period after month 0 ("Strain 2") clustered in a distinct branch within the CRF36_cpx radiation, separated by an average distance of ~11.5% from Strain 1. The average distance among the Strain 2 clones, obtained over a 15-month period, was ~1.9%.

Within the Strain 1 branch, it was evident that the quaspecies were composed of variants that grouped into two subclusters (bootstrap value, 100%; Fig. 1). Whereas the average genetic distance between the two subclusters was 4.4%, certain sequences of each group were separated by as much as an 8.6% distance. Notably, members of each of these subclusters could be found contemporaneously at months 9, 15, and 27 in CMNYU107, and at month 0 in CMNYU6518.

Sequences were further analyzed by bootscanning. Initially, each sequence was queried against all group M subtypes and CRFs. All sequences of Strain 1 and Strain 2 were found to be a URF, each possessing an identical breakpoint

structure, which combined CRFs 01_AE and 36_cpx, with the C1 and C2 regions of *env* each harboring a breakpoint (Fig. 2). The region clustering with CRF36_cpx comprised positions 350–750 of the alignment (corresponding to HxB2 positions 6531–6759). All CMNYU107 month 21 clones were found to cluster with subtype G throughout the length of the alignment (data not shown).

To determine if recombination had occurred between the initial and dually infecting strains, each Strain 2 sequence was queried against the most genetically distant sequences of Strain 1 (CMNYU107, month 15, clone 20, and CMNYU107, month 30, clone 20; genetic distance, 8.6%), which were employed as separate reference groups named Strain 1A and Strain 1B alongside the standard Group M reference set. None of the Strain 2 sequences was found to be recombinant with Strain 1A or Strain 1B (data not shown).

The extent of genetic distance among some Strain 1 sequences suggested that distinct viral variants had dually infected these subjects prior to the dual infections identified previously. Intriguingly, the range of distances among some of these initial Strain 1 viruses pointed to the possibility that this branch of sequences actually consisted of a variety of recombinants of these distinct variants. To further investigate the Strain 1 quaspecies, another round of breakpoint analysis was performed, wherein the most genetically distant sequences from the Strain 1 branch were used as unique references. All other sequences from Strain 1 were queried against these "parent" sequences in addition to the group M set.

It was found that 10 URFs could be identified among the sequences of Strain 1, each possessing breakpoint patterns combining Strains 1A and 1B (Fig. 3). Recombinant Group 1 ($n=2$), identified in CMNYU107 at month 27 and CMNYU6518 at month 0, possessed a single breakpoint from Strain 1A to Strain 1B, approximately 150 bp into the alignment (position 6340 in HxB2 numbering). Group 2 ($n=2$), identified in CMNYU107 at month 27, possessed a breakpoint from Strain 1B to 1A, at a very similar position to Group 1. Group 3 ($n=16$), identified in CMNYU107 at month 0, recombined from Strain 1B to 1A at approximately position 200, and back to 1B at position 300, approximately (positions 6390 and 6445, respectively, in HxB2 numbering). Group 4 ($n=1$) identified in CMNYU6518 at month 0 is a recombinant of Strain 1A to 1B, at position 300 approximately. Group 5 ($n=4$), identified in CMNYU107 at months 9 and 15, possessed a breakpoint from Strain 1B to 1A, also approximately at position 300. Group 6 ($n=4$), identified in CMNYU107 at months 9 and 30, and in CMNYU6518 at month 0, possessed two breakpoints, from 1B to 1A at position 300, as well, and from 1A to 1B at position 450, approximately (position 6620 in HxB2 numbering). Group 7 ($n=1$), identified in CMNYU107 at month 15, possessed a breakpoint from Strain 1B to 1A at position 400, approximately (position 6585 in HxB2 numbering). Group 8 ($n=1$), identified in CMNYU107 at month 9, recombined from Strain 1A to 1B at position 725 in the alignment, approximately (position 6715 in HxB2 numbering). Group 9 ($n=1$), identified in CMNYU107 at month 15, possessed three breakpoints, two of which were present in other recombinant groups: from 1B to 1A at position 300, back to 1B at position 450, and from 1B to 1A at position 800 (position 6783, in HxB2 numbering). Group 10 ($n=2$), identified in CMNYU107 at month 9, possessed three breakpoints, all of which could be found in other recombinant groups: from 1B

FIG. 1. Phylogenetic analysis of C1C2_{env} (HxB2 location 6207–7342) quasispecies clones amplified from CMNYU107 and CMNYU6518 plasma samples obtained over 30 and 15 month studies, respectively. Analysis was performed using the neighbour-joining method with Kimura's two-parameter method and bootstrap analysis (1000 replicates). Bootstrap values above 70% are shown. All reference subsubtypes and CRFs, in addition to all analyzed sequences, were initially used to construct the trees; most references and study subject sequences have been omitted here for clarity. Study sequences are identified by the CMNYU prefix, followed by subject number, study month number (in parentheses), and clone number. Closed circles highlight sequences obtained from CMNYU107 and open circles highlight those from CMNYU6518.

FIG. 2. Breakpoint analysis of the CRF36_cpx-related URF identified in both CMNYU107 and CMNYU6518. Bootstrapping was performed using SimPlot 3.5.1 configured with 1000 bootstrap replicates, 250-bp window, and a step size of 30 bp. The x-axis shows aligned nucleotides of the sequence analyzed and the y-axis shows the bootstrap (significance) value. Genomic regions are indicated at the bottom of the plot.

to 1A at position 150, from 1A to 1B at position 300, and from 1B to 1A at position 800. Importantly, nonrecombinant Strain 1A ($n = 27$) was identified in CMNYU107 at months 9, 15, and 27, and in CMNYU6518 at month 0, whereas nonrecombinant Strain 1B ($n = 34$) was identified in CMNYU107 at months 15, 27, and 30, and in CMNYU6518 at month 0. The remainder of the clones analyzed contained regions that could not be deciphered as either 1A or 1B with significance ($>60\%$ bootstrap), and thus were not classified as pure Strain 1A, 1B, or any of the above recombinant groups. Confirmatory reanalysis using contemporaneous sequences of Strains 1A and 1B amplified from the CMNYU6518 month 0 sample yielded almost identical breakpoint patterns for all sequences (data not shown).

Removing the recombinants from distance analysis, the average distance between Strain 1A and 1B was 7.4%. The average distances between Strain 1A and Strain 2 and between Strain 1B and Strain 2 were 11.6% and 12.4%, respectively. To assess whether these variants could be considered epidemiologically unrelated, these genetic distances were compared to the intra- and intersubsubtype distances of the HIV-1 group M reference subsubtypes, as annotated in the LANLHSD 2007 Subtype Reference Set. Between two and four unrelated references of each subtype were inputted into the Subtyping Distance Tool (SUDI) provided by the LANLHSD (<http://hiv.lanl.gov/content/hiv-db/SUDI/sudi.html>), providing a histogram of the intra- and intersubsubtype distances to which the distances among the variants of the novel strain identified herein could be compared.

SUDI analysis found the intrasubsubtype distances for the C1C2 locus to be between $\sim 6\%$ and $\sim 14\%$ (Fig. 4). As such, the distance between Strains 1A and 1B, at 7.4%, falls within this range, suggesting these variants represent epidemiologically unrelated viruses. The distances separating Strains 1A and 1B from Strain 2 were found to fall at the intrasubsubtype–intersubsubtype junction, clearly demonstrating their epidemiological unrelatedness (Fig. 4). It can be concluded, therefore, that these three HIV-1 variants, though possessing the same breakpoint structure, are not directly related, but rather may represent three unlinked members of an undefined CRF.

Analysis of vif-vpr-vpu (VVV). The sequences of 210 clones were analyzed (Supplemental Fig. 1A; see www.liebertonline.aid). All clones amplified from samples obtained from CMNYU107 at months 0, 9, and 30 ($n = 72$) clustered with all clones amplified from samples obtained from CMNYU6518 at month 0 ($n = 25$) and 10/22 clones obtained at month 15 in the CRF01_AE radiation. The average genetic distance between these clones was $\sim 2\%$. All clones amplified from CMNYU107 at month 21 clustered in a unique branch with CRF14_BG within the subtype G radiation (distance between clones, $\sim 1.4\%$). All clones amplified from CMNYU6518 at months 3 and 6 ($n = 28$) and 12/22 clones amplified from the month 15 sample clustered separately in the CRF02_AG radiation (distance between clones, $\sim 1.2\%$). Each VVV clone was subjected to bootscanning as previously described. It was found that all sequences that clustered with

FIG. 3. Recombinants of CMNYU107/CMNYU6518 Strain 1A and 1B identified among the quasispecies of both study subjects. Bootsscan was performed as described above. The most genetically distant members of Strains 1A and 1B were also used as separate reference groups, in addition to a reference group of sequences from CMNYU107 month 21 (subtype G) and a reference group of sequences from CMNYU6518 Strain 2. Strains 1A and 1B are represented by light and dark gray lines, respectively.

FIG. 4. Histograms of the intra- and intersubsubtype distances of the HIV-1 group M reference subsubtypes, as annotated in the Los Alamos HIV Sequence Database (LANLHSD) 2007 Subtype Reference Set. Two to four unrelated references of each subtype were input into the Subtyping Distance Tool (SUDI) provided by the LANLHSD. The average genetic distances between CMNYU107/CMNYU6518 Strains 1A and 1B, Strains 1A and 2, and Strains 1B and 2 are indicated on the plot by the dotted, solid, and dashed arrows, respectively. A group O sequence was used as the outgroup.

CRF01_AE were a URF of CRF01_AE and CRF36_cpx, with an unclassifiable region of ~180 bp in *vif*. Each sequence possessed an ~163-bp region of CRF36_cpx between positions 975 and 1150 of the alignment, approximately, encompassing parts of *vpr* and *tat* exon 1 (Fig. 5). All clones that clustered with CRF02_AG clustered with this CRF throughout the length of the alignment. All CMNYU107 clones from the month 21 samples clustered along the length of the alignment with subtype G, although only with significance with the subtype G found in CRF14_BG. No recombinants of these three strains were identified.

Analysis of RTpol. The sequences of 89 clones were analyzed (Supplemental Fig. 1B). All clones amplified from CMNYU107 at months 15 and 30 ($n = 43$) clustered closest to the CRF01_AE references (average distance between sequences, ~1.5%). Bootscanning found this strain to be a URF of CRF01_AE and CRF36_cpx. All clones possessed the same structure: the first 650 nt of the alignment clustered closest to CRF01_AE (corresponding to positions 2619–3263 in HxB2 numbering), with the following 400 nt clustering with CRF36_cpx (corresponding to positions 3356–3783 in HxB2), and the final ~500 nt clustering with CRF01_AE (Fig. 5). All clones amplified from the CMNYU107 month 21 sample clustered separately, consistent with the above analysis, in the subtype G radiation closest to the CRF14_BG reference (average distance between sequences, ~1.1%). These clones clustered with CRF14_BG (subtype G) throughout the length of the alignment. All clones amplified from samples obtained from CMNYU6518 at months 6 and 15 clustered with the CRF02_AG references, with an average distance between sequences of ~1.3%. Bootscanning found this strain to be a URF of CRF02_AG and subsubtype A1. Most of the sequence clustered with CRF02_AG; however, an ~200-nt fragment approximately 750 nt into the alignment (corresponding to positions 3356–3542 in HxB2) clustered most significantly with A1 references, and bootscan of the final ~400 nt of this strain could not discern whether the relationship to A1 or CRF02_AG was more significant (Fig. 5). *RTpol* information could not be obtained for CMNYU6518 at month 0, thus the initial strain in *pol* was not determined.

Analysis of viral sequences from subjects CMNYU124 and CMNYU129

Analysis of C1C2env. The sequences of 118 clones from subject CMNYU124 and 187 clones from subject CMNYU129 were analyzed from samples obtained over 18 and 33 months, respectively (Supplemental Fig. 2A; see www.liebertonline.com/aid). All clones amplified from samples obtained from CMNYU124 at months 0, 12, 15, and 18 ($n = 95$) and all clones amplified from the CMNYU129 sample obtained at month 9 ($n = 24$) clustered together with the subsubtype F2 references (average genetic distance, ~1.5%). All clones ($n = 23$) amplified from the CMNYU124 month 9 sample clustered closest to the CRF02_AG references with the CMNYU129 clones amplified from the month 0, 15, 18, 21, 27, 30, and 33 samples, with an average distance of ~2.6% between clones. These data were consistent with previous analyses that identified the dual infections at these time points, which also found them to be epidemiologically linked.¹² Each clone clustered throughout the length of the alignment with the subtype predicted by the neighbor-joining tree (i.e., F2 or CRF02_AG), and no recombinants of the dually infecting strains were present.

Analysis of VVV. The sequences of 131 clones were analyzed (Supplemental Fig. 2B). All CMNYU124 clones from the month 0 and month 18 ($n = 46$) samples clustered with all CMNYU129 clones amplified from the month 9 sample ($n = 10$), in the subsubtype F2 radiation (distance between clones, ~1.1%). Conversely, all clones ($n = 24$) amplified from the CMNYU124 month 9 sample clustered with the CMNYU129 clones amplified from months 0, 18, and 33 ($n = 51$) in the CRF02_AG radiation (distance between clones, ~1.2%). All clones clustered throughout the length of the alignment with the subtype predicted by the phylogenetic analysis, and none was shown to be recombinants of the dually infecting strains (Fig. 5).

Analysis of RTpol. The sequences of 98 clones were analyzed (Supplemental Fig. 2C). It was found that all clones amplified from CMNYU124 at months 0 and 15 ($n = 28$)

FIG. 5. Summary of the quasiespecies identified at each time point studied. Pattern variations of colors representing one subtype signify epidemiological unrelatedness. ND, not done.

clustered with the subsubtype F2 references, with an average distance among clones of $\sim 1.0\%$. All clones amplified from the month 9 sample ($n = 13$) clustered with the clones amplified from CMNYU129 at months 0, 15, and 33 ($n = 47$) in the CRF02_AG radiation (average distance, $\sim 1.1\%$). In contrast to the *env* and VVV analysis, the sequences analyzed from the CMNYU129 month 9 sample clustered closely with the CRF02_AG group, demonstrating that analysis of the RTpol locus alone would not have revealed the CMNYU129 dual infection. All sequences clustered throughout the length of the alignment closest to the subtype predicted by the neighbor-joining tree. None was shown to be recombinants of the dually infecting strains (Fig. 5).

Analysis of viral sequences from subject CMNYU6544

Analysis of C1C2env. The sequences of 76 clones amplified from samples obtained over 12 months were analyzed (Supplemental Fig. 3A; see www.liebertonline.com/aid). All clones ($n = 47$) from the month 0, 3, and 9 samples clustered in the subtype A radiation (average distance, $\sim 1.5\%$). Bootscan analysis found each clone to be a URF of F2 and CRF02_AG, with a breakpoint from CRF02_AG to F2 occurring at approximately position 625 in the alignment (Fig. 5). All clones ($n = 29$) amplified from the final sample, taken at month 12, clustered in a single branch in the CRF02_AG radiation with an average distance between clones of 1.8% and with $\sim 16\%$ distance from the initial strain. This was consistent with previous analysis that identified the dual infection at month 12.¹² These clones clustered along the length of the alignment closest to the CRF02_AG references, although some regions did not cluster with significance (data not shown). Bootscanning was repeated using the month 0–9 sequences as a reference group. It was found that the month 12 sequences clustered most significantly with the month 0–9 references in the first ~ 400 bp of the alignment; however, even the month 9 clones alone as a reference could not generate the high bootstrapping expected for recombination having occurred within the previous 3 months (Supplemental Fig. 3B). As such, this section of the alignment was used to construct a neighbor-joining tree to determine the relatedness of these strains at this locus. This analysis revealed an $\sim 10\%$ genetic distance between each strain infecting CMNYU6544 at this ~ 400 bp region, which was almost equal to that separating these strains at this locus from the CRF02_AG references, demonstrating that the second strain was not a recombinant of the dually infecting viruses (Supplemental Fig. 3C).

Analysis of VVV. The sequences of 83 clones were analyzed (Supplemental Fig. 3D). All clones ($n = 37$) from the month 0 and 3 samples clustered in the CRF02_AG radiation. Each clone was a URF of F2 and CRF02_AG. All clones harbored a breakpoint from F2 to CRF02_AG at position 600 of the alignment, approximately 400 bp into *vif* (Fig. 5). Six clones amplified from the month 6 sample clustered closely with the month 0/3 group, and were similarly found to be the same F2-02_AG recombinant. Sixteen clones amplified from the month 6 sample clustered with all 24 clones amplified from the month 12 sample (average distance, $\sim 1.3\%$), which clustered distinctly in the CRF02_AG radiation ($\sim 10\%$ distance from the initial strain). These clones were purely CRF02_AG and nonrecombinant with the initial strain.

Analysis of RTpol. The sequences of 73 clones were analyzed (Supplemental Fig. 3E). All clones amplified from samples obtained at months 0, 6, and 9 ($n = 55$) clustered with subsubtype F2, with an average distance between sequences of $\sim 1.2\%$. All clones amplified from the month 12 sample clustered separately, with the CRF02_AG references (average distance between clones, 0.6%). All clones clustered along the length of the alignment closest to the subtype predicted by the neighbor-joining tree, and no recombinants of the dually infecting strains were identified.

Correlation of sequence similarity with the frequency of recombination

Several studies have found recombination between non-identical HIV-1 strains to occur most frequently at loci possessing high sequence similarity.^{5,7,8,22} The present analysis demonstrates a tendency for frequent recombination to occur *in vivo* after a dual HIV-1 infection only when the dually infecting strains are sufficiently similar genetically—i.e., possess a genetic distance from each other that falls within the intrasubtype spectrum when compared to the distances among the HIV-1 reference subsubtypes and CRFs. This finding was exemplified by study subjects CMNYU107 and CMNYU6518, whose subtype-concordant dually infecting viruses generated between nine and three detectable recombinants in *env*, respectively. These viruses were $\sim 7.4\%$ distant in C1C2env, a distance that appeared “permissive” for frequent recombination; however, the viruses infecting subject CMNYU6544, which were as little as $\sim 10\%$ distant in the C1C2env and VVV regions, did not generate any detectable recombinants in the loci studied, despite appearing to coexist in the month 6 sample. Furthermore, the remaining dually infecting viruses, possessing ~ 12 – 24% distance in C1C2env, showed no recombination within the loci studied. As such, it is possible that a genetic limit to frequent, detectable recombination exists around ~ 7.5 – 10% distance between infecting strains.

Viral loads and the generation of recombinants

For recombination to occur, a cell must be infected by both viral variants. As high viral loads will likely result in the infection of more cells than if viral loads are low, we investigated whether the lack of recombination between subtype-discordant strains may have been due to lower viral loads in these subjects. The viral load data, presented in Supplemental Table 1, reveal that there were sufficient amounts of viral RNA in the samples analyzed to adequately detect the viral quasiespecies. The samples obtained from subject CMNYU107 (dually infected by concordant subtypes with genetic distance of 7.4%) contained significantly more viral RNA (mean viral copies/ml, 17,356) than those obtained from subjects infected only by subtype-discordant viruses (mean viral copies/ml, 3710). This could indicate that the high degree of recombination identified in CMNYU107 may have been due to a high viral load alone, and not to the sequence similarity of the dually infecting viruses. However, it should be noted that no recombinants of the subtype-discordant virus also infecting CMNYU107 by month 21 were identified, despite these high viral loads. Similarly, subject CMNYU6518 exhibited a significantly high viral load in the final sample (month 15), yet no recombination between the discordant strains infecting this

subject was identified. It remains to be seen whether higher viral loads correlate with the nature of recombination within a dually infected host.

Discussion

In the present study, through the longitudinal analysis at three genetic loci of five subjects dually infected by viruses with varying degrees of genetic discordance, it was shown that frequent, detectable recombination is not a certain or immediate outcome of such infections, specifically those involving discordant HIV-1 subtypes. These data reveal frequent recombination occurring in the two individuals dually infected by subtype-concordant viruses, whereas no recombination between the subtype-discordant viruses was detected within the loci studied.

Several viral and host factors could favor the formation of recombinants among the subtype-concordant dually infected subjects while disfavoring recombination between the subtype-discordant strains. Such factors could include viral diversity, viral fitness, and host immune responses. The subtype-concordant viruses coidentified in CMNYU107 and CMNYU6518 suggest that unlike the subtype-discordant viruses, both parent viruses were able to persist in these hosts, facilitating the generation of recombinants. Because the subtype-concordant viruses share a similar genomic composition, it is possible that they also possess similar viral fitness and similar immunogenicities. Such similar phenotypes would likely allow these viruses to coexist, as neither would possess any particular advantage over another in the host. Furthermore, the similar sequences of these viruses would undoubtedly facilitate genomic dimerization as well as successful recombination during virus replication in dually infected cells.^{5,7,8,24,25}

Unlike the subtype-concordant dual infections, the subtype-discordant dual infections did not reveal recombinants based on the genomic regions analyzed. Several factors could have disfavored the formation of recombinants. Sequence similarity among two viral strains has been shown to be critical for dimerization and recombination.^{24,25} Therefore, discordant sequences will dimerize poorly compared to sequences that are more similar. As the genetic distance between the subtype-discordant strains was significantly greater in the loci studied than that of the concordant strains, it may be that the subtype-discordant genomes are poorly compatible for dimerization, which would lead to a scarcity of recombinants notwithstanding the dual infection of single cells.^{24,25} Currently, study of these dimerization signals is underway, and will be reported elsewhere. Furthermore, the genome-wide variability between the sequences is highly likely to have adversely affected successful recombination of these genomes even if they were able to properly dimerize.^{5,7,8} As well, it may be that any intersubtype recombinants that might have been generated in these infections were not detected due to a lack of fitness in the host, which precluded their expansion and persistence. Such short-lived recombinants may have been unfit as a result of host immune pressure on certain epitopes possessed by these viruses, which one or both of their parental viruses lacked, thus leading to their rapid disappearance. Currently, analyses of the comparative replicative capacities of the dually infecting strains as well as the relative immunity of each host against their respective

infecting viruses are underway and will be reported elsewhere. Undoubtedly, it is possible that recombination may have occurred in regions outside those analyzed presently, although with the exception of frequent recombination at the 3' end of *env*, such recombination was predicted to be relatively less likely by previous studies.

It is reasonable to question the influence of virus load on virus recombination in dually infected patients as the viral loads of the samples obtained from subjects CMNYU124, CMNYU129, and CMNYU6544, each infected by subtype-discordant viruses, tended to be relatively lower, on average less than 5000 viral copies/ml, compared to the virus load in the subtype-concordant dually infected patients. It is possible that the likelihood of generating recombinant viruses was lowered in such subjects, as low virus copies would reduce the chances of single cells being dually infected. However, as the viral loads for CMNYU107 were much higher, and no recombination between the discordant strains infecting this subject was detected, the data so far do not permit any definite conclusions as to the effect of virus load on recombination in these subjects.

Additionally, it may be that insufficient sampling of the quasispecies and/or physical compartmentalization of certain strains may have accounted, in part, for the lack of intersubtype recombinants identified. It may also have been that the use of standard RT-PCR and cloning in the present analysis and not single-genome amplification could have resulted in resampling of the RNA and/or PCR-mediated recombination; however, resampling is highly unlikely to have occurred due to a high amount of input viral RNA, as determined by viral load data and evidenced by the fact that rarely were two sequences at any of the loci identical. This potential was also minimized by cloning from multiple RT-PCRs.²⁶ Furthermore, the use of highly processive, error-proof polymerases in all likelihood avoided any potential PCR-mediated recombination.^{27,28}

Two comprehensive studies of subtype-discordant dual infections have been conducted, each on a single Kenyan commercial sex worker (CSW). Both were analyzed at 3-month intervals, and shown to have generated some recombination between the subtype-discordant strains; yet neither demonstrated the high degree of recombination found for *in vitro* studies. In the first case, the quasispecies was studied extensively in *gag* and *gp41/nef* using PBMC DNA over 30 months. In addition to the ACD and AC parental URF strains, three intersubtype recombinants were detected in *gag* and four were detected in *gp41/nef* over the course of the study, although often as only one or two clones at a single time point.⁹ Similarly, the second study included an in-depth analysis of the quasispecies infecting a subject exhibiting symptoms of late-stage disease in *gag*, *vpv/gp120*, and *gp41/nef* over 12 months. This subject was triply infected by two strains of subtype A and a subtype C virus. In *gag*, two recombinants were detected. In the *env* analysis, nine URFs were identified, including four intersubtype recombinants at the C2, V3, and C4 regions of *env*. However, these URFs only comprised the minority (<1%) of the viral population in the host at each sampling time.¹¹ At *gp41/nef*, three AC recombinants were found, although their proportions were again relatively low throughout the study. Thus, the lack of detectable recombinants among the subtype-discordant dual infections based on the viral genomic regions analyzed herein

(C1C2env, vif-vpr-vpu, and RTpol) may not be inconsistent with previous *in vivo* studies, which although they have identified recombinants, generally comprise the minority of the quasiespecies, as previously described. Furthermore, the genomic regions analyzed for recombination in the studies included the gp41/nef, gag, or the whole gp120, which are not entirely identical to those analyzed in the present study.^{9,11} This could, in part, contribute to some discrepancies in the recombinants identified among these various studies.

Recently, an American cohort of HIV⁺ women was assayed for dual subtype B infection and recombination.¹ This analysis examined the protease-RT region of *pol* and the C2V5 region of *env* amplified from two to four samples from 58 women at 6- to 24-month intervals, and showed findings similar to the present study with regard to temporal fluctuation of the dually infecting strains, in that of the 18 cases in which dual infection was identified at a single interim time point, 16 found no evidence of the second strain in subsequent samples, and in three cases, the initial strain was almost overtaken by the second strain. Furthermore, it was found that among the 31% who were dually infected, recombination had occurred between the *pol* and *env* loci in nearly all individuals, and recombination within the loci studied had occurred in three individuals. Notably, these instances were *intrasubtype* dual infections, as was the case for our dually concordant subtype infected patients, CMNYU107 and CMNYU6518. Thus, this relatively high rate of intrasubtype recombination is not surprising; however, even infection by subtype-concordant viruses in these cases did not yield the huge arrays of recombinants predicted by *in vitro* analyses.^{9,11,29,30}

The present in-depth analysis of five dually HIV-1-infected subjects employed frequent longitudinal sampling and analysis of genomic loci predicted to be common areas for recombination by *in vitro* studies. This analysis is the first of its kind to focus on dually infected members of the general HIV⁺ population in Africa, and has demonstrated that although recombination between subtype-discordant strains was uncommon, recombinants of relatively similar strains were easily detected. By the end of the present study, the subjects remained asymptomatic, with no sample studied containing >32,000 virus copies/ml, despite a mean CD4 count of 312 cells/ml among the final samples. As these infections progress and host defenses weaken, it is possible that continued follow-up will reveal more intersubtype recombinants. Such tracking will be of critical importance to understanding the genesis of HIV-1 recombinants and to properly monitor the ever-increasing genetic diversity of the pandemic.

Acknowledgments

The authors are grateful to the individuals who have donated their blood for these studies and wish to acknowledge the continued support of the Ministry of Public Health, Cameroon. Supported by Grants AI47053 and AI083142 from the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Fogarty International Center (TW01409), Center for AIDS Research (AI027742-17), and funds from the Department of Veterans Affairs (Merit Review Award and the Research Enhancement Program).

Author Disclosure Statement

No competing financial interests exist.

References

1. Templeton AR, Kramer MG, Jarvis J, *et al.*: Multiple-infection and recombination in HIV-1 within a longitudinal cohort of women. *Retrovirology* 2009;6:54.
2. Archer J, Pinney JW, Fan J, *et al.*: Identifying the important HIV-1 recombination breakpoints. *PLoS Comput Biol* 2008;4(9):e1000178.
3. Piantadosi A, Chohan B, Chohan V, McClelland RS, and Overbaugh J: Chronic HIV-1 infection frequently fails to protect against superinfection. *PLoS Pathog* 2007;3(11):e177.
4. Kozaczynska K, Cornelissen M, Reiss P, Zorgdrager F, and van der Kuyl AC: HIV-1 sequence evolution *in vivo* after superinfection with three viral strains. *Retrovirology* 2007;4:59.
5. Fan J, Negroni M, and Robertson DL: The distribution of HIV-1 recombination breakpoints. *Infect Genet Evol* 2007;7(6):717–723.
6. Pernas M, Casado C, Fuentes R, Perez-Elias MJ, and Lopez-Galindez C: A dual superinfection and recombination within HIV-1 subtype B 12 years after primoinfection. *J Acquir Immune Defic Syndr* 2006;42(1):12–18.
7. Baird HA, Gao Y, Galetto R, *et al.*: Influence of sequence identity and unique breakpoints on the frequency of intersubtype HIV-1 recombination. *Retrovirology* 2006;3:91.
8. Baird HA, Galetto R, Gao Y, *et al.*: Sequence determinants of breakpoint location during HIV-1 intersubtype recombination. *Nucleic Acids Res* 2006;34(18):5203–5216.
9. McCutchan FE, Hoelscher M, Tovanabutra S, *et al.*: In-depth analysis of a heterosexually acquired human immunodeficiency virus type 1 superinfection: Evolution, temporal fluctuation, and intercompartment dynamics from the seronegative window period through 30 months postinfection. *J Virol* 2005;79(18):11693–11704.
10. Moumen A, Polomack L, Roques B, Buc H, and Negroni M: The HIV-1 repeated sequence R as a robust hot-spot for copy-choice recombination. *Nucleic Acids Res* 2001;29(18):3814–3821.
11. Gerhardt M, Mloka D, Tovanabutra S, *et al.*: In-depth, longitudinal analysis of viral quasiespecies from an individual triply infected with late-stage human immunodeficiency virus type 1, using a multiple PCR primer approach. *J Virol* 2005;79(13):8249–8261.
12. Powell RLR, Urbanski MM, Burda S, King T, and Nyambi PN: High frequency of HIV-1 dual infections among HIV-positive individuals in Cameroon, West-Central Africa. *J Acquir Immune Defic Syndr* 2009;50(1):84–92.
13. Powell RL, Konings FA, Nanfack A, *et al.*: Quasiespecies analysis of novel HIV-1 recombinants of subtypes A and G reveals no similarity to the mosaic structure of CRF02_AG. *J Med Virol* 2007;79(9):1270–1285.
14. Staden R, Beal KF, and Bonfield JK: The Staden package, 1998. *Methods Mol Biol* 2000;132:115–130.
15. Katoh K and Toh H: Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 2008;9(4):286–298.
16. Galtier N, Gouy M, and Gautier C: SEAVIEW and PHYLO-WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput Appl Biosci* 1996;12(6):543–548.
17. Kumar S, Nei M, Dudley J, and Tamura K: MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* 2008;9(4):299–306.
18. Kimura M: A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 1980;16(2):111–120.

19. Saitou N and Nei M: The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987;4(4):406–425.
20. Felsenstein J: Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 1985;39(4):783–791.
21. Felsenstein J: PHYLIP—Phylogeny Inference Package (Version 3.2). *Cladistics* 1989;5:164–166.
22. Ramirez BC, Simon-Loriere E, Galetto R, and Negroni M: Implications of recombination for HIV diversity. *Virus Res* 2008;134(1–2):64–73.
23. Swanson P, de Mendoza C, Joshi Y, *et al.*: Impact of human immunodeficiency virus type 1 (HIV-1) genetic diversity on performance of four commercial viral load assays: LCx HIV RNA Quantitative, AMPLICOR HIV-1 MONITOR v1.5, VERSANT HIV-1 RNA 3.0, and NucliSens HIV-1 QT. *J Clin Microbiol* 2005;43(8):3860–3868.
24. Chin MP, Chen J, Nikolaitchik OA, and Hu WS: Molecular determinants of HIV-1 intersubtype recombination potential. *Virology* 2007;363(2):437–446.
25. Chin MP, Rhodes TD, Chen J, Fu W, and Hu WS: Identification of a major restriction in HIV-1 intersubtype recombination. *Proc Natl Acad Sci USA* 2005;102(25):9002–9007.
26. Liu SL, Rodrigo AG, Shankarappa R, *et al.*: HIV quasispecies and resampling. *Science* 1996;273(5274):415–416.
27. Fang G, Zhu G, Burger H, Keithly JS, and Weiser B: Minimizing DNA recombination during long RT-PCR. *J Virol Methods* 1998;76(1–2):139–148.
28. Meyerhans A, Vartanian JP, and Wain-Hobson S: DNA recombination during PCR. *Nucleic Acids Res* 1990;18(7):1687–1691.
29. Zhuang J, Jetzt AE, Sun G, *et al.*: Human immunodeficiency virus type 1 recombination: Rate, fidelity, and putative hot spots. *J Virol* 2002;76(22):11273–11282.
30. Levy DN, Aldrovandi GM, Kutsch O, and Shaw GM: Dynamics of HIV-1 recombination in its natural target cells. *Proc Natl Acad Sci USA* 2004;101(12):4204–4209.

Address correspondence to:

Phillipe N. Nyambi

Department of Pathology

New York University School of Medicine

c/o V.A. Medical Center

423 East 23rd street, Room 18124N

New York, New York, 10010

E-mail: phillipe.nyambi@nyumc.org