JOURNAL OF VIROLOGY, Mar. 2006, p. 2472–2482 Vol. 80, No. 5
0022-538X/06/$08.00+0 doi:10.1128/JVI.80.5.2472–2482.2006

Extensive Recombination among Human Immunodeficiency Virus Type 1 Quasispecies Makes an Important Contribution to Viral Diversity in Individual Patients

Charlotte Charpentier,[1]† Tamara Nora,[1]† Olivier Tenaillon,[2] François Clavel,[1] and Allan J. Hance[1]*

INSERM U 552[1] and INSERM U 722,[2] Université Paris 7—René Diderot, Faculté de Médecine Xavier Bichat, Paris, France[2]

Received 4 August 2005/Accepted 7 December 2005

Although recombination during human immunodeficiency virus type 1 (HIV-1) replication in vitro and in vivo has been documented, little information is available concerning the extent that recombination contributes to the diversity of HIV-1 quasispecies in the course of infection in individual patents. To investigate the impact of recombination on viral diversity, we developed a technique that permits the isolation of contemporaneous clonal viral populations resulting from single infectious events by plasma-derived viruses, thereby permitting the assessment of recombination throughout the viral genomes, including widely separated loci, from individual patients. A comparison of the genomic sequences of clonal viruses from six patients, including patients failing treatment with antiretroviral therapy, demonstrated strong evidence for extensive recombination. Recombination increased viral diversity through two distinct mechanisms. First, evolutionary bottlenecks appeared to be restricted to minimal segments of the genome required to obtain selective advantage, thereby preserving diversity in adjacent regions. Second, recombination between adjacent gene segments appeared to generate diversity in both *pol* and *env* genes. Thus, the shuffling of resistance mutations within the genes coding for the protease and reverse transcriptase, as well as recombination between these regions, could increase the diversity of drug resistance genotypes. These findings demonstrate that recombination in HIV-1 contributes to the diversity of viral quasispecies by restricting evolutionary bottlenecks to gene segments and by generating novel genotypes in *pol* and *env*, supporting the idea that recombination may be critical to adaptive evolution of HIV in the face of constantly moving selective pressures, whether exerted by the immune system or antiretroviral therapy.

Human immunodeficiency virus type 1 (HIV-1) encapsidates two RNA genomes in progeny viral particles (30, 41) and, during the ensuing replicative cycle, recombination between the two genomes can occur (3, 8, 28, 35). To the extent that the two RNA genomes are distinct, recombination generates viral DNA coding for genetic information different from that of the two parental strains (9, 12, 16, 17). Numerous studies performed in vitro have documented the occurrence of recombination during HIV-1 replication (6, 27, 38, 39, 46, 47, 54). Although these studies have found that the rate of recombination can be influenced by the nature of the target cell and that certain genomic regions may be hot spots for recombination, there is general agreement that recombination occurs frequently throughout the genome (2 to 20 events/genome/replicative cycle) and has the capacity to rapidly shuffle genomic segments from parental viruses (21, 27, 46, 54). Numerous viral strains derived through recombination between distinct HIV-1 subtypes have been identified in infected individuals, indicating that recombination also occurs in vivo (26, 42, 49), and studies performed using limiting-dilution PCR suggest that such events are occurring frequently (51; F. Maldarelli, M. Kearney, S. Palmer, M. Polis, J. Mican, R. Stephens, D. Rock, J. Mellors, and J. Coffin, Abstr. 12th Conf. Retroviruses Opportunistic Infect., abstr. 667, 2005).

The advantages that accrue through recombination and the impact of recombination on HIV-1 pathogenesis in individual patients are not clearly defined (45). The shuffling of polymorphisms found in distinct viral quasispecies could play a role in generating viral diversity (33, 45). The ability to maintain extensive diversity may be extremely important for viral pathogenesis, because it ensures the availability of viral quasispecies able to escape changes in the selective pressures exerted by the immune response or by antiretroviral therapy. Indeed, results obtained with other models support the idea that diversity generated by recombination can be beneficial for adaptation to changing evolutionary pressures (15, 40).

In addition to generating diversity, viral recombination could also be useful in preserving existing diversity. In the course of infection in an individual patient, diversity can be threatened by evolutionary bottlenecks. Although reflective of an initial selective advantage, the emergence of a genetically homogeneous population could prove disadvantageous, because the progeny might later become susceptible to elimination by an immune response focused against shared antigenic determinants. An evolutionary bottleneck resulting from the emergence of a unique viral species with high resistance to antiretroviral agents would pose this risk. If, however, during or

* Corresponding author. Mailing address: INSERM U 552, Hôpital Bichat—Claude Bernard, 46, rue Henri Huchard, 75018 Paris, France. Phone: 33-1-40-25-63-55. Fax: 33-1-40-25-63-70. E-mail: hance@bichat.inserm.fr.

† C.C. and T.N. contributed equally to this work.

TABLE 1. Clinical characteristics of patients at the time of study and summary of the number of clones obtained[a]

Patient	Age/sex	Treatment	Result					
			Viral load (log_{10})	No. of CD4 T cells/μl	Dilution of plasma	No. of clones obtained	% of infected wells	Infectivity (%)
1	41/M	3TC DDI TDF TPV/RTV APV	5.33	8	1/4	25	7	0.008
2	43/M	3TC RTV NFV	7.12	7	1/40	4	15	0.005
					1/160	14	5	0.006
					1/400	1	2	0.006
3	39/M	3TC TDF DDI TPV/RTV APV	5.39	11	1/4	12	3	0.003
4	37/F	3TC TDF ABC APV	5.68	104	1/4	56[b]	14	0.012
5	35/M	3TC TDF TPV/RTV	6.36	8	1/16	34	16	0.016
6	42/M	None	6.22	151	1/4	13	7	0.002
					1/8	6[b]	3	0.003

[a] Abbreviations for antiretroviral drugs are as follows: ABC, abacavir; APV, amprenavir; 3TC, lamivudine; NFV, nelfinavir; DDI, didanosine; RTV, ritonavir; TDF, tenofovir; TPV, tipranavir. M, male; F, female. Infectivity was calculated as follows: (no. of infectious events observed)/(no. of plasma viral RNA copies used/2).
[b] Only selected clones from this patient were sequenced.

subsequent to emergence, such strains could recombine at high rates with preexisting strains, much viral diversity could be maintained in regions outside those responsible for the bottleneck.

Despite its potential importance, the role of recombination in generating and preserving viral diversity in vivo has been difficult to define for technical reasons. The analysis of sequences obtained by amplification and cloning of viral genomes from plasma cannot address this question, because recombination is known to occur during amplification by PCR (23, 32), and individual sequences can be over- or underrepresented in such bulk amplification products due to differences in the efficiency of amplification of individual sequences with a given pair of primers. Recombination during amplification can be avoided by performing limiting-dilution PCR, but this technically demanding approach also has certain liabilities. In particular, the possibility of recombination occurring during the reverse transcription reaction is difficult to exclude, some sequences may not be amplified using any given pair of primers and therefore would be lost from the analysis, and the amplification of a single fragment containing widely separated genomic regions (e.g., *pol* and *env*) is not feasible.

To further evaluate the extent that recombination contributes to the generation and preservation of the diversity of HIV-1 in the course of infection, we developed a technique that permits the isolation of contemporaneous clonal viral populations resulting from single infectious events by plasma-derived viruses and which therefore allows the assessment of recombination between distant regions of viral genomes from individual patients. Using this technique, we generated such clones from patients experiencing virological treatment failure, and sought evidence that (i) high rates of recombination occur throughout the viral genome in vivo, (ii) the restriction of diversity resulting from evolutionary bottlenecks in regions coding for the enzymes targeted by antiretroviral drugs need not extend to other genomic regions, and (iii) the rearrangement between gene segments directly contributes to genetic diversity.

MATERIALS AND METHODS

Patient population. Clonal viral populations were obtained from six HIV-1-infected patients; clinical information for these patients is summarized in Table 1. Because we were interested in evaluating patients in whom viruses may have been subjected to evolutionary bottlenecks, five of the patients we evaluated were failing treatment with antiretroviral agents. The treated patients had initially received nucleoside analog reverse transcriptase (RT) inhibitors only and had developed resistance to these agents prior to treatment with regimens that included protease inhibitors and nonnucleoside RT inhibitors. A decrease in viral load of >50% resulting from antiretroviral treatment was not documented for any patient, despite the fact that viral load had been regularly monitored over the interval of time the patients received treatment with protease inhibitors.

Clonal viral populations. To obtain clonal viral populations, the following protocol was used: MT4 cells expressing CCR5 and CXCR4 receptors (MT4-R5, obtained from O. Schwartz, Pasteur Institute), were maintained in RPMI 1640 medium containing 10% fetal bovine serum, 100 U/ml penicillin G, and 100 μg/ml streptomycin (complete medium). On the day of infection, cultures were resuspended at 2×10^6 cells/ml in complete medium containing 1% (vol/vol) dimethyl sulfoxide, and 0.25-ml aliquots were distributed in 24-well plates. Plasma was diluted (usually 1:4) in complete medium containing (final concentration) 1% dimethyl sulfoxide and 2 μg/ml DEAE-dextran, and 0.25 ml of this mixture was added to each well. The plates were centrifuged ($860 \times g$, 2 h, 22°C), and cultured for 4 h at 37°C to permit viral entry. Cells were recovered and washed once, and 200-μl aliquots containing 2×10^4 cells were distributed into 96-well plates. The cultures were maintained at 37°C in 5% CO_2 and were passaged with a 1:10 dilution every 7 days for up to 40 days. Cultures were inspected frequently by light microscopy, and when patent cytopathic changes were observed, the culture supernatant and the cell pellet from infected wells were recovered separately and frozen at −80°C. If viral replication was observed in >20% of the wells, the experiment was repeated after further dilution of the plasma.

Sequencing of viral genomes. DNA was extracted from infected cell pellets using a QIAamp Viral DNA mini kit (QIAGEN, Valencia, CA) and used to amplify four segments of proviral DNA, two adjacent segments in *gag-pol* (total of ≅1,300 bp) and two adjacent segments in *env* (total of ≅1,300 bp). For several patients, the testing of several sets of primers was required to identify pairs that amplified DNA from all clones of a given patient. In difficult cases, a fragment surrounding the area of interest was cloned and sequenced, permitting the design of suitable primers for amplification of proviral DNA. Amplification products were directly sequenced in both directions using an ABI automated sequencing platform (Applied Biosystems, Foster City, CA). All chromatograms were visually inspected to exclude the presence of sequences with ambiguous or polymorphic bases. Sequences were aligned using CLUSTAL X (version 1.81), and alignments in regions with insertions were verified manually.

Detection of multiple infectious events. To evaluate the relationship between the proportion of infected wells in limiting-dilution experiments and the proportion of wells infected with more than one virus, a suspension containing approximately equal numbers of two distinct clonal viruses (A and B) was prepared, and MT4-R5 cells were infected with serial dilutions of this suspension. DNA was isolated from cell pellets from infected wells, and a 670-bp fragment of RT and/or a 579-bp fragment of *env* were amplified by PCR. Amplified fragments were purified and digested with PstI (RT) or HindIII (*env*), which recognized sites present only in the fragments from virus A or virus B, respectively, thereby allowing the identification of wells infected with one or both viruses. This approach could reliably detect both sequences in mixtures in which the minority

TABLE 2. The proportion of wells infected with single or multiple viruses depends on the proportion of infected wells, as predicted by the Poisson distribution[a]

Amount of virus (pg p24 antigen)	Infected wells/total wells	No. of infected wells tested	% of total wells with indicated result				
			Not infected	Infected with single strain		Infected with both strains	
				Predicted[b]	Observed[c]	Predicted[b]	Observed
0.45	65/96	9	32	49	53	18	15
0.15	77/267	39	71	27	27	2	2
0.075	25/156	23	84	15	16	1	0
0.045	4/108	4	96	4	3	<1	<1

[a] MT4-R5 cells were infected with various amounts of a suspension containing two distinct viruses (strains A and B) and distributed into 96-well plates. Following culture, infected wells were identified by the presence of cytopathic effects, and DNA was isolated from the cell pellet. The genomic regions corresponding to RT and/or *env* were amplified by PCR, and the presence of a single strain or both strains was determined using the restriction endonuclease digestion as described in Materials and Methods. Of the 68 wells tested that were infected with a single strain, 39 and 29, respectively, were infected with strains A and B, indicating that the viruses were present in approximately equal numbers.

[b] Based on the Poisson distribution. The values are corrected for the fact that 1/2, 3/4, and 7/8 of wells infected with 2, 3, or 4 viruses, respectively, would be infected with both strains.

[c] Calculated as follows: (wells tested that were infected with single strain)/(total wells tested) × (% of infected wells).

sequence represented 10% of total viral DNA. Exploring the two regions using DNA from the same infected well always gave concordant results (data not shown).

Evaluation of the evolution of viral sequences during culture. To examine whether the viral sequence, as determined by bulk sequencing, was subject to change during culture, MT4-R5 cells were infected under limiting-dilution conditions with viruses obtained by transfection of HeLa cells with an infectious proviral molecular clone derived from pNL4-3, but carrying protease resistance mutations M46I and I54V and V82A or I54V and V82A, combinations of mutations known to severely impair viral replicative capacity (29). DNA from 36 infected wells was extracted, proviral DNA encompassing the protease region was amplified, and the bulk products were sequenced.

Assessment of recombination. Phylogenetic trees were obtained by maximum likelihood using the DNAML program in the Phylip package (version 3.6). The congruence between phylogenetic trees was evaluated using the incongruence length difference (ILD) test as implemented in PAUP* (version 4.0).

Rmin was determined using the RecMin software designed by Myers and Griffiths (34). The output of this program includes the position of gene segments with obligate recombination events (in the absence of homoplasy), permitting an assessment of the distribution of recombination events.

To estimate recombination rates, the Pairwise program in the LDhat package was used (31). This program is a modification of that proposed by Hudson (18). It estimates recombination rates by combining the coalescent likelihoods of pairwise comparisons of segregating sites and has been extended to take into account the occurrence of recurrent mutations at a single site. In implementing the test, the population mutation parameter ($\theta = 4N_e\mu$, where N_e is the effective population size and μ is the per-site mutation rate) was determined for each sample set (52), and confidence limits for θ were determined by coalescent simulation using DnaSP (48). Likelihood matrices were then constructed for each sample using the estimated value of θ for population recombination parameters ($\rho = 4N_e r$, where r is the per-site recombination rate) ranging from 0 to 1,000. To assess confidence intervals of ρ, values whose log probability was ± 2 of the value with maximum probability were determined.

Assessment of nucleotide divergence. To evaluate nucleotide divergence between sequences for a given segment of the viral genome, the number of synonymous substitutions per potential synonymous site (dS) was calculated pairwise for all clones by using the modified Nei-Gojobori method (36). In this analysis, a transition/transversion ratio of 2.0 was used, the Jukes-Cantor correction was applied, and gaps, when present, were handled by pairwise deletion. The mean distance of each clone relative to all other clones from the same patient and the overall mean divergence were determined. We were interested in comparing the nucleotide divergence of different regions of the viral genome for clones from the same patient. It is known, however, that a variety of factors, such as codon usage, amino acid composition, and RNA secondary structure, can influence dS, independent of differences in mutation rate (1). To control for such differences, a training set of 11 HIV-1 subtype B sequences from the Los Alamos database was used to identify regions in *gag*, protease, RT, and *env* which met the following criteria: (i) all were coding regions, (ii) insertions and deletions did not prevent unambiguous alignment, (iii) mean dS was similar among the regions. The following regions were identified (positions correspond to the HXB2 reference strain): *gag*, 1963 to 2160; protease, 2253 to 2327 and 2412 to 2486; RT, 2865 to 3203; envelope C1, 6324 to 6617; envelope C3, 7110 to 7379. The overall similarity in dS of these regions was confirmed (see Results) by analyzing all HIV-1 subtype B viruses in the Los Alamos alignment database for which sequences of all regions were available ($n = 43$). These regions were then evaluated for all clones from each patient.

Evaluation of viral tropism. The tropism of envelope sequences was evaluated using the algorithm described by Jensen et al. (20). Variants with predicted X4 tropism were detected for 3 of 6 patients (patient 1, 5 of 25 clones; patient 3, 5 of 12 clones; patient 4, 1 of 19 clones). For a given patient, the sequence of the V3 region was identical in the different clones. The tropism of representative clones bearing these V3 sequences was evaluated by generating recombinant viruses expressing the envelopes (gp120 plus extracellular domain of gp41) and evaluating their ability to infect U373-CD4-CCR5 cells and U373-CD4-CXCR4 cells in a single-cycle colorimetric assay. The recombinant viruses from patients 3 and 4 showed exclusive X4 tropism. The recombinant viruses from patient 1 had low infectivity in this assay, and predicted tropism could not be confirmed.

Data presentation and statistical methods. All viral nucleotide positions correspond to the HXB2 reference strain. Resistance mutations are as defined by the IAS-USA Drug Resistance Mutations Group (22). Statistical comparisons of diversity (dS) were made using the Kruskal-Wallis test; posttest comparisons, performed only if P was <0.05, were made using the Dunn test.

RESULTS

Isolation of clonal viruses from plasma of patients infected with HIV-1. To generate clonal viral populations resulting from single infectious events by plasma-derived viruses, we developed the novel limiting-dilution technique described in Materials and Methods. For these studies, it was important that viruses were derived from single infectious events. Limiting the analysis to experiments where <20% of wells were infected reduces the probability of there being wells infected with more than one virus. Nevertheless, in an experiment where 20 of 100 wells are infected, the Poisson distribution predicts that 18 wells would be infected with a single virus and 2 wells would be infected with two viruses. Thus, it was important to demonstrate that multiply infected wells were not occurring more frequently than predicted by chance alone and, above all, that cells infected with more than a single virus could be identified and excluded from analysis. To study this question, clonal viruses derived from two different clinical specimens were mixed in equal proportions, cells were infected with dilutions of this mixture using our protocol, and infected wells were evaluated for the presence of the two strains by restriction analysis (Table 2). The proportion of infected wells decreased with increasing dilutions of the mixture. As expected, the num-

ber of wells infected with both strains was high when the proportion of infected wells was high, but decreased in strict accordance with the Poisson distribution as the percentage of infected wells decreased. Importantly, when viral DNA from wells infected with both strains was amplified and sequenced, such doubly infected wells were easily identified in all cases (six of six wells) by the presence of ambiguous bases at all positions where the sequences of the two clones were different, providing an easy means to exclude from analysis wells infected with more than one virus.

It was also important to demonstrate that the DNA sequences for the clonal viral populations, obtained by bulk sequencing of proviral DNA, did not evolve during culture in vitro. To test this, viruses with a known sequence were produced by transfection using infectious proviral molecular clones derived from pNL4-3 but carrying protease resistance mutations known to impair viral fitness. Clonal viral populations were generated using our technique, and the protease region was sequenced for 36 clones. Without exception, these sequences were identical to those of the original plasmid (14,256 total bases). In particular, no evidence of reversion of the resistance mutations was observed despite the strong replicative disadvantage for these viruses in the absence of drug.

Clonal viruses generated from plasma of HIV-1-infected patients. Clonal viral populations were generated from five chronically infected patients who were failing antiretroviral therapy (Table 1, patients 1 to 5), as well as one treatment-naïve patient (patient 6). For all plasma samples studied, the large majority of infectious events were detected between 13 and 29 days after initiation of culture (81% of 165 total clonal viruses). To characterize the genome of clonal viruses, DNA was extracted from infected cell pellets, and genomic regions of ≅1,300 bp in *gag-pol* and *env* were amplified from proviral DNA and sequenced. For the 127 clonal viruses evaluated (25, 19, 12, 19, 34, and 18 clones, respectively, for patients 1 to 6), base calls were unambiguous at all positions sequenced. In the course of the studies, proviral DNA from only two infected wells produced sequences in which ambiguous bases were identified, and these samples were excluded from analysis. Clonal viruses from a single patient grouped closely together in phylogenetic trees but were widely separated from the viruses obtained from other patients or laboratory strains, regardless of the region evaluated (data not shown). A comparison of the results of population-based genotyping of plasma viruses (available for the protease and RT for four patients) and the sequences of the clonal viruses indicated that the clonal viruses were representative of the those in the total population. For positions that were monomorphic in the plasma genotype, the consensus sequence of the clonal viruses was, without exception, identical (3,622 total bases). For the 62 positions that were polymorphic in the plasma genotype, the same polymorphisms were identified among the sequences from clonal viruses in 59 of 62 cases. In contrast, many polymorphisms identified in low proportions among the sequences of the clonal viruses were not detected by bulk genotyping of plasma viruses.

Evidence for recombination in contemporary viruses. Visual inspection of aligned sequences from contemporary clonal viruses revealed numerous examples consistent with recombination. To systematically evaluate the recombination between different regions of the viral genomes, a phylogenetic method was employed (14). For each patient, phylogenetic trees corresponding to different parts of the genome were generated using the maximum likelihood method, and the topologies were compared using the incongruence length difference test. Rejection of the null hypothesis indicates significant differences between the topologies and strongly supports the occurrence of recombination. An example is shown in Fig. 1, in which sequences that clustered together in the phylogenetic tree of the envelope C2-V4 region are indicated by different colors. In the phylogenetic tree for the envelope C1-V2 region of the same viruses, some clones remain closely associated (for example, the pairs of sequences marked by an asterisk). More commonly, however, viruses that are clustered in the C2-V4 tree are dispersed throughout the C1-V2 tree, and the topologies of these two trees are strongly incongruent by the ILD test ($P < 0.001$). Similarly, viruses that are clustered in the C2-V4 tree are also dispersed in the tree comparing the *gag*-protease region of the same viruses (P was <0.001 using the ILD test). As shown in Table 3, phylogenetic trees for different genomic regions were usually incongruent ($P < 0.05$) when viruses from all six patients were evaluated using this approach, a result compatible with extensive recombination between all genomic regions evaluated. For several cases, the analysis was repeated after systematic removal of single sequences or even pairs of sequences, but no such deletions resolved the incongruence between regions, indicating that incongruence was not determined by only a small number of recombination events (data not shown). The ILD test showed significant incongruence between RT and protease for only three of the six patients evaluated. It is noteworthy that the three patients for whom this test was not significant for these regions are the three patients found to have very low diversity in the protease region (see below), and limited diversity is known to impair the ability of the ILD test to detect recombination (44).

To calculate a lower bound for the minimum number of recombination events necessary to construct a history of the sample sequences in the absence of homoplasy, (Rmin), a modification of Hudson's four-gamete test was used (34). Consistent with the idea that recombination had occurred frequently, Rmin was elevated for all patients, with values of 49, 24, 33, 28, 47, and 33, respectively, for patients 1 to 6. The positions of these putative recombination events were scattered throughout the portions of the HIV genome sequenced.

Estimations of the rates of mutation and recombination. To estimate the contribution of mutation and recombination to sequence diversity, the coalescent-based technique described by McVean et al. (31) was used to determine the population recombination parameter ($\rho = 4N_e r$), where r is the per-site recombination rate per generation and N_e is the effective population size. By comparison with the population mutation parameter ($\theta = 4N_e\mu$), where μ is the per-site mutation rate per generation, the relative rate of recombination compared to mutation (r/μ) can be determined. As shown in Fig. 2, the population mutation parameters were not strikingly different in a comparison of the six patients (median, 0.021; range, 0.018 to 0.032). Considerably greater variation was seen in the population recombination parameters (median, 0.03; range, 0.002 to 0.11). Thus, the rates of recombination relative to the rates of mutation were 4.1, 1.8, 0.8, 0.5, 5.7, and 0.1, respectively, for

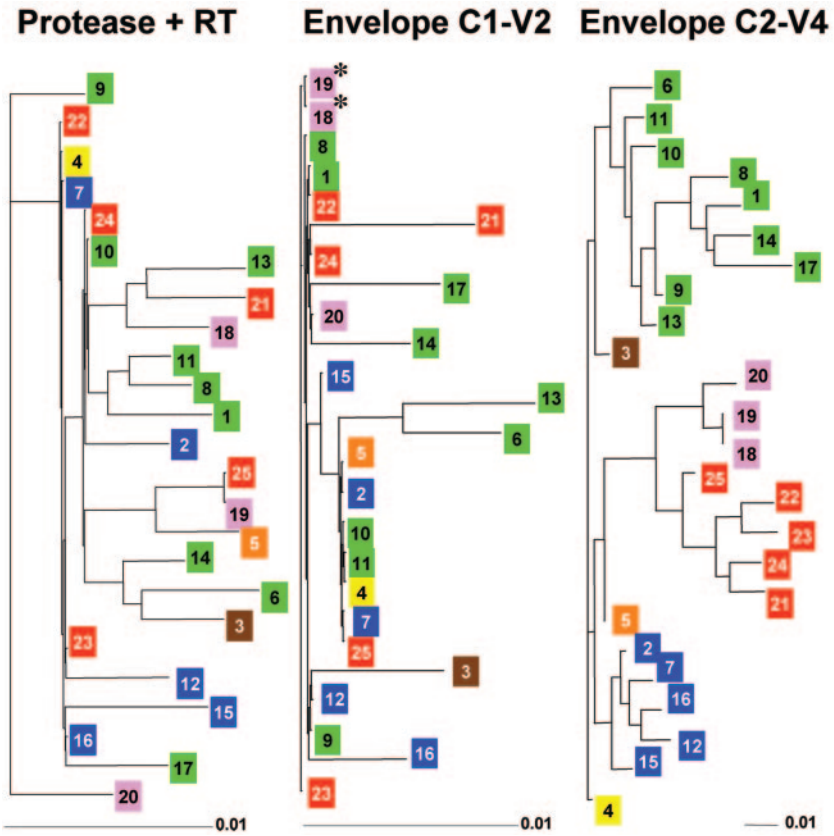

FIG. 1. Phylogenetic trees corresponding to different parts of the genome of clonal viruses from a single patient are incongruent. Genomic regions corresponding to protease + RT (2253 to 3338), C1-V2 envelope (6221 to 6871), and C2-V4 envelope (6872 to 7567) for 25 clonal viruses obtained from patient 1 were sequenced, and phylogenetic trees were generated using the maximum-likelihood method. Clonal viruses that clustered together in the C2-V4 envelope tree are highlighted with the same color. Although certain clones remain clustered in the other trees (e.g., those marked with an asterisk in the middle panel), clonal viruses that cluster together in one tree are generally dispersed throughout the other trees. The trees were incongruent using the ILD test ($P < 0.001$ for all pairwise comparisons).

the six patients. The recombination parameter did not correlate with nucleotide diversity. No obvious correlations were observed between the clinical characteristics of the patients and the population recombination parameter, although this parameter was considerably lower for the single untreated patient studied. More sequences were available from the two patients with the highest population recombination parameter (patients 1 and 5), but when analyses were repeated using randomly selected subsets of these sequences similar in number to those available for the other patients, no significant effect on the recombination parameter was observed, and the recombination parameter remained elevated (data not shown).

In an effort to conform to the assumption of neutral evolution in this model, analyses in which only silent mutations were

TABLE 3. Evaluation of the congruence between phylogenetic trees obtained using nucleotide sequences of clonal viruses from different regions of the viral genome[a]

Regions being compared[b]	*P* value for each comparison for the indicated patient					
	1	2	3	4	5	6
gag + protease versus RT	**0.044**	0.135	**0.004**	**0.002**	**0.001**	**0.001**
Protease versus RT	0.115	0.936	**0.013**	0.146	**0.001**	**0.005**
gag-pol versus *env* C1-V4	**0.001**	**0.001**	**0.022**	**0.003**	**0.001**	**0.001**

[a] Phylogenetic trees for the indicated genomic regions were generated using the maximum likelihood method and were compared using the incongruence length difference test. For each patient, *P* values for each comparison are shown; a *P* value of <0.05 (boldface type) indicates that the phylogenetic trees were not congruent.

[b] The nucleotide sequences used to generate trees were as follows: *gag* + protease, 1930 to 2549; RT, 2670 to 3308; protease, 2253 to 2549; *gag-pol*, 1930 to 3308; *env* C1-V4, 6221 to 7567.

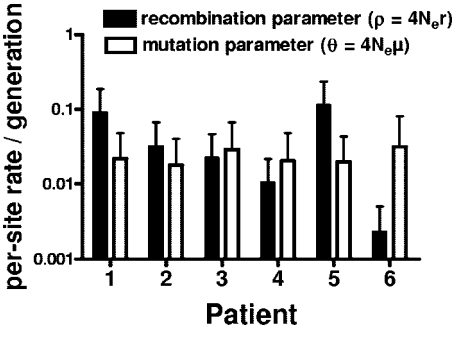

FIG. 2. Contribution of mutation and recombination to sequence diversity. For each patient, the population recombination parameter and the population mutation parameter were determined using the coalescent-based approach described by McVean et al. (31).

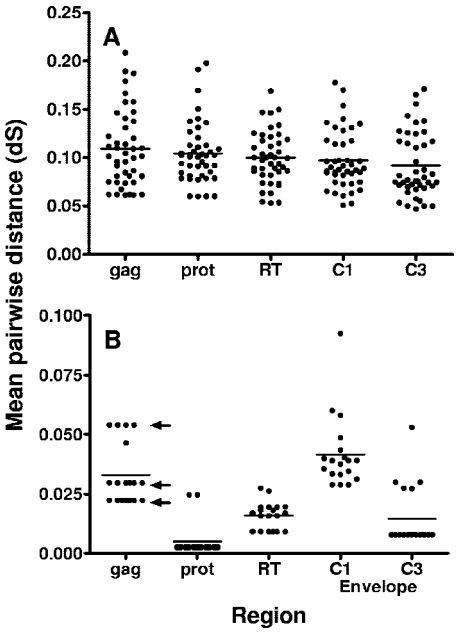

FIG. 3. Sequence diversity in different portions of the viral genome. The dS was determined for selected regions in genes coding for the C-terminal portion of *gag* (gag), protease (prot), RT, envelope C1 (C1), and envelope C3 (C3) using the modified Nei-Gojobori method (36). Each point represents the average pairwise distance of each sequence to all other sequences analyzed, and the bar represents the overall mean divergence. Panel A shows the analysis of all HIV-1 subtype B viruses in the Los Alamos alignment database for which sequences of all regions were available ($n = 43$), and panel B shows results of the analysis of the 19 clonal viruses from patient 4. The small arrows in panel B indicate groups of clones whose *gag* sequences were similar at all silent positions but whose sequences differed from those of the other groups by 3 or 4 silent mutations.

considered were performed. After correction for the number of synonymous sites present in the sequence (≅22%), the population mutation parameters was somewhat higher than those obtained when the entire nucleotide sequence was considered but remained similar among the patients studied (median, 0.047; range, 0.033 to 0.053). Restriction of the analysis to silent mutations had only a small effect on the population recombination parameters (median, 0.034; range, 0.008 to 0.123). In this analysis, the rates of recombination relative to the rates of mutation were 2.0, 0.7, 0.7, 1.0, 2.7, and 0.2, respectively, for the six patients. We conclude that the rate of recombination among contemporary viruses is often high and can occur at a rate that approaches or surpasses the mutation rate.

Diversity in different genomic regions. To evaluate sequence divergence between clones from a given patient, the dS was determined for different regions in the genome. In this analysis, the mean distance of each clone relative to all other clones from the same patient and the overall mean divergence were determined. To permit comparisons between different regions in the genome, this analysis was restricted to segments in the genome for which the extent of diversity due to silent mutations was shown to be similar when sequences of viruses infecting 43 unrelated patients were evaluated using the same approach (Fig. 3A). As shown in Fig. 3B, the extent of diversity due to the presence of synonymous substitutions could be quite different in different portions of the genome for viruses from the same patient. When the patterns of sequence divergence in different portions of the genome for all the patients are compared, however, several common features emerge (Fig. 4). In particular, sequence diversity in the protease region was often lower than that observed in other regions, and it was strikingly low in three cases (patients 1, 2, and 4). This pattern was not observed for some patients, including the individual who had not received antiretroviral therapy (patient 6). As shown above, the diversity observed in the protease region was not different from that of other regions when the sequences from 43 heterologous viruses from the Los Alamos database were compared (Fig. 3A). Similarly, the mean pairwise diversity between the consensus sequences for the six patients was, as expected, considerably higher than that seen between clones from a single patient, but no significant difference was found in a comparison of the different genomic regions (data not shown). These findings argue against the existence of selective pressure operating on synonymous sites in the segment of the protease that reduces diversity that was analyzed in these studies. Because all the patients for whom reduced diversity in the protease region was observed had been treated with protease inhibitors, the results are compatible with the possibility that this region of the genome had passed through one or more bottlenecks associated with the emergence of viruses carrying resistance mutations. To further test this idea, a similar analysis was performed using sequences in the protease region obtained from the same patient prior to and during the evolution of protease resistance (patient C in reference 5). Phylogenetic analysis suggested the occurrence of bottlenecks associated with the emergence of viruses carrying the I54V and A71V mutations between 3 and 23 months and the fixation of the I93L mutation between 24 and 34 months. As shown in Fig. 5, considerable diversity was observed in the protease region prior to and immediately following treatment with protease inhibitors but, at later times, a significant decrease in the diversity in the protease region was observed. Taken together, these findings suggest that evolutionary bottlenecks can limit diversity in the protease region but that such reductions need not extend to adjacent genomic regions.

Less striking reductions in diversity were also observed in RT segments in some patients, consistent with the prior occurrence of bottlenecks in this region (Fig. 4). All the patients studied had developed resistance to RT inhibitors prior to receiving protease inhibitors, and it is possible that the additional time since the occurrence of treatment-related bottlenecks had been sufficient for more silent mutations to have accumulated in the RT, thereby explaining the somewhat greater diversity seen in comparisons of RT and protease. Evidence for restriction in diversity in portions of the *env* gene was also observed. Patients could show either significantly greater diversity in the C1 envelope region than the C3 envelope region (patients 4 and 5, $P < 0.001$), or greater diversity in the C3 region than in the C1 region (patient 3, $P < 0.001$).

Factors influencing apparent diversity between clones. Although dS is commonly used as a measure of "evolutionary distance" resulting from the accumulation of synonymous mutations, a second factor that influenced this parameter in our data was identified. In the example shown in Fig. 3B, the mean value of dS for the *gag* fragment analyzed was 0.03, suggesting

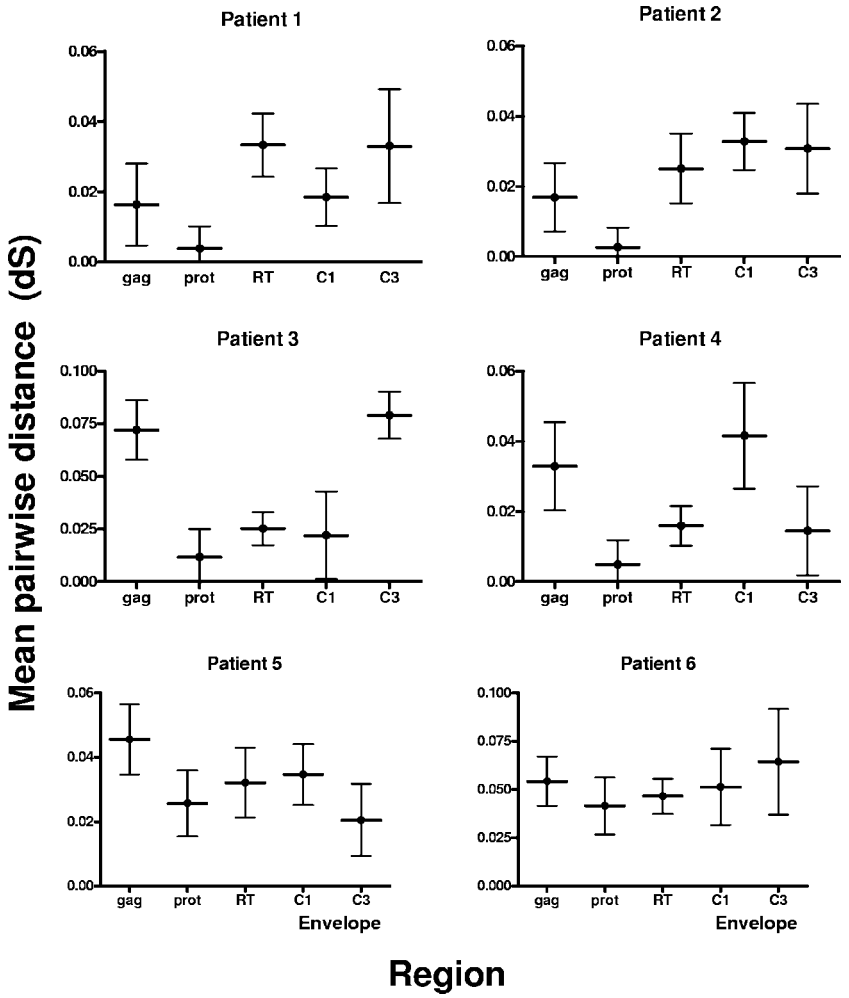

FIG. 4. Patterns of sequence diversity in different portions of the viral genome observed for clonal viruses. For each patient, the dS was determined for selected regions in genes coding for the C-terminal portion of *gag* (gag), protease (prot), RT, envelope C1 (C1), and envelope C3 (C3) as described in the legend to Fig. 3. The result for each of the six patients is presented as the mean ± standard deviation of the average pairwise distance of each clonal virus compared to those of all other clonal viruses obtained from that patient.

moderate diversity. When the distribution of values for each clone relative to all other clones is examined, however, the values fall into three distinct groups with one outlier (Fig. 3B, arrows). Within each of these groups, the nucleotide sequence of the *gag* fragment was identical but differed from that of the other groups by 3 or 4 silent mutations. Thus, the finding of a low value for dS was useful in identifying regions that were essentially homogeneous due to prior bottlenecks, but apparent high diversity often, but not always, reflected situations in which a small number of alternative sequences were used repeatedly by different clones (see below).

Recombination between gene segments contributes to the diversity of *env*. The findings presented above indicate that the diversity within short segments of the HIV-1 genome could be either variable, restricted to a relatively small number of alternative choices, or even essentially homogeneous. Nevertheless, considerable diversity could be maintained by splicing together these segments in novel combinations. To test this idea further, the amino acid sequences of the various envelope domains for clones from a given patient were compared in order to deter-

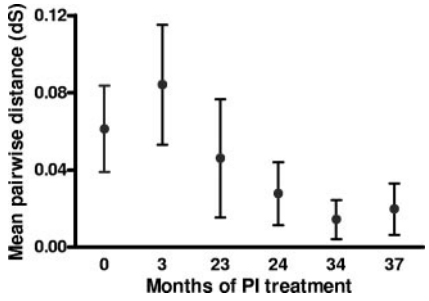

FIG. 5. Selection of protease resistance mutations is associated with a loss of sequence diversity. The dS was determined for the region of the protease gene described in the legend to Fig. 3 by using sequences obtained at different intervals after initiation of treatment with protease inhibitors. Previously published phylogenetic studies indicated that this patient experienced bottlenecks associated with the emergence of viruses carrying the I54V and A71V mutations between 3 and 23 months and fixation of the I93L mutation between 24 and 34 months (5). Results are presented as means ± standard deviations of the average pairwise distances of each sequence compared to those of all other sequences obtained at the same time point.

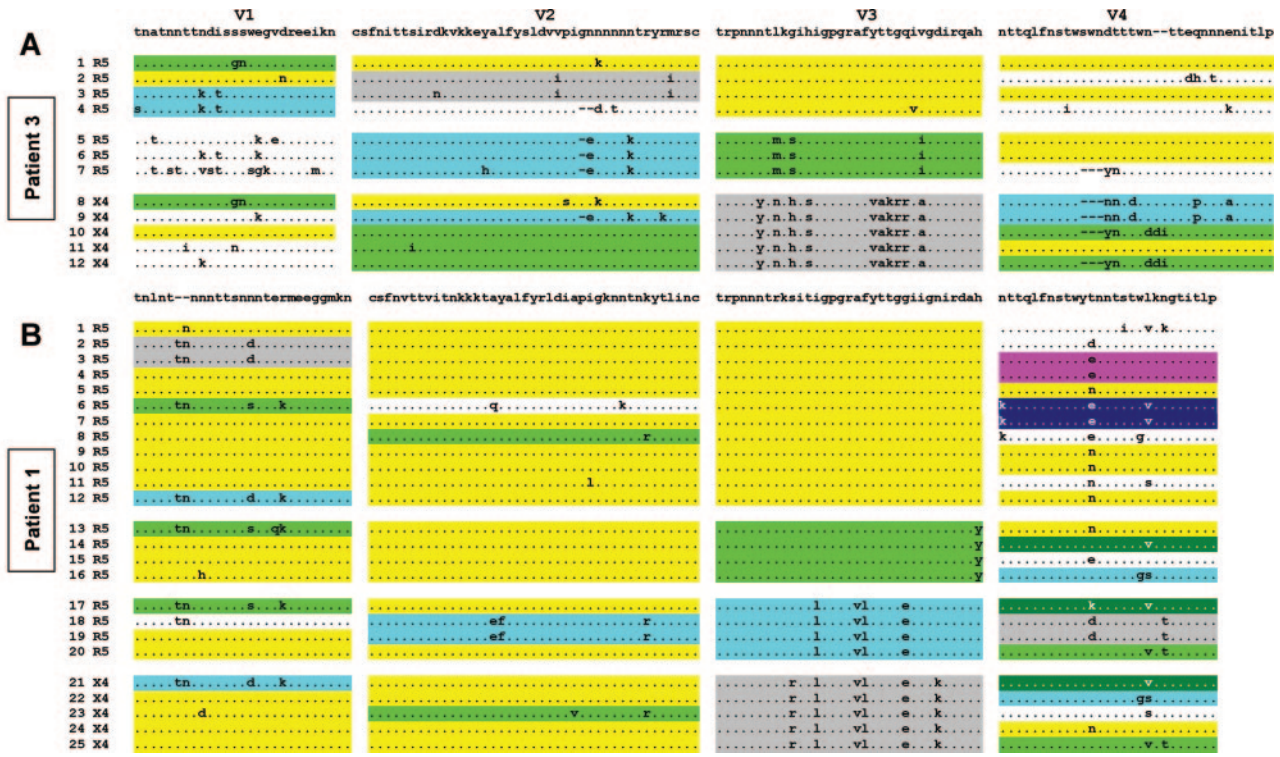

FIG. 6. Recombination between envelope domains contributes to the diversity of *env*. At the top of each panel, the consensus amino acid sequences of the V1, V2, V3, and V4 regions of envelope for the 12 clonal viruses from patient 3 (panel A) and the 25 clonal viruses from patient 1 (panel B) are shown. Only amino acid changes different from the consensus sequence are shown for each clone. For each domain, sequences that are identical or that differ by a single amino acid substitution not identified in another sequence are highlighted with the same color. The tropism of the clonal viruses predicted by the algorithm of Jensen et al. is also indicated.

mine the number of distinct genotypes that were present in each domain and the extent to which clones that shared a similar genotype in one domain did or did not also express similar genotypes in adjacent domains. For this analysis, genotypes that differed by a single amino acid change not identified in another clone were considered to be similar.

The results obtained for patient 3 are shown in Fig. 6A. Three clearly distinct genotypes were observed in the V3 domain, one group with exclusive X4 tropism and two distinct groups with R5 tropism, and the clones have been separated into groups on this basis. The number of distinct sequences within a given domain ranged from two (C2 domain) to nine (V1 domain). For all domains, clones sharing similar genotypes were always present. Some tendency for coevolution of *env* domains was observed for X4 viruses, which shared closely related C3 domains (data not shown) and often used V4 domains not identified in R5 viruses (Fig. 6A). This association was not absolute, however, and one X4 virus (clone 11) used a V4 domain also identified for both of the R5 populations. In contrast, the same V1 or V2 domain could be used by viruses expressing different V3 domains, and the same V3 domain could be associated with a variety of different V1, V2, and C2 domains. As a consequence of this recombinatorial diversity, no two clones expressed the same combination of V1-to-V4 segments.

The results for patient 1 are shown in Fig. 6B. Four distinct genotypes were present in the V3 domain, including one population with predicted X4 tropism. The number of distinct genotypes within a given domain ranged from 4 (V2 and V3 domains) to 13 (V4 domain). For all domains, clones sharing the same genotype were always present. The same V1, V2, or V4 segment could be used by viruses expressing different V3 domains. In this patient, no clear evidence of coevolution between V3 domains and other variable domains was observed for clones with putative X4 tropism. Among the 25 clones evaluated, only three (clones 5, 9, and 10) used the same combination of V1-to-V4 segments. When variability in the C1, C2, C3, and N-terminal portion of C4 was included in this analysis, none of the 25 clones used the same combination of envelope domains.

The results for the other four patients were generally similar to those presented in Fig. 6. The number of distinct genotypes within a given domain ranged from 1 (V3 domain, patients 2, 5, and 6; V1 domain, patient 2) to 12 (C2 domain, patient 5). The proportion of clones using the same combination of envelope segments was 0/19 (patient 2), 0/18 (patient 4), 3/35 (patient 5), and 4/18 (patient 6).

Some, but not all, of the apparent diversity within a given domain appeared to result from recombination events. For example, the V4 genotype of clone 7 from patient 3, although scored as "distinct," could be produced by recombination between clone 1 and clone 10. Many other examples were present and were observed in all domains except for the V3 domain. The V3 domain is short and often showed the least diversity; both of these factors would reduce the likelihood of identifying recombination events. Nevertheless, it is possible that recombination

events that change the amino acid sequence in this domain often have a negative impact on viral infectivity and tend to be underrepresented when infectious viral clones are evaluated.

It is also noteworthy that some clones used a sequence for one domain that was markedly different from that of the other viruses from that patient (e.g., V1 domain in clone 7 from patient 3, and V4 segment in clone 1 from patient 1), but other domains of these viruses were identical to those of other viral clones, suggesting that these "outlier" segments may have been derived from recombination events with minority or archived viral populations.

Drug resistance genotypes and recombination. As shown above, diversity, as measured by the accumulation of silent mutations, was often limited in the protease and could also be low in the RT. Consistent with this finding, only limited diversity in the resistance genotypes was observed. For three treated patients (patients 1, 2, and 4), diversity was restricted to the gain or loss of one or two resistance mutations by single clones. For example, for the 25 clones of patient 1, one clone had gained the F53L mutation in protease, while other clones had lost either the V118I or L100I and K103N mutations in RT. Otherwise, all clones expressed the same 8 protease resistance mutations and the same 11 RT resistance mutations.

For two of the patients studied, two resistance mutations in protease and two resistance mutations in RT were expressed by only a subset of clones. For these four pairs of mutations, some clonal viruses expressed each mutation individually, while others expressed both mutations simultaneously (data not shown). This pattern cannot be explained by the sequential appearance of the mutations alone, and either recombination or the occurrence of more than one mutation or reversion at a single site (homoplasy) must be invoked. Considerable shuffling of protease and RT resistance genotypes was also observed for these two patients. Protease genotypes represented by more than a single clone were usually found in association with several different RT genotypes. Thus, when the protease and RT regions were considered together, the most abundant resistance genotype was expressed by only 25% (patient 3) and 38% (patient 5) of total clonal viruses. Clonal viruses isolated from patient 3 had both R5 and X4 tropism. Clonal viruses in both populations could express the same resistance genotype, and several different genotypes were identified among X4 and R5 viruses. Thus, tropism and resistance genotype did not cosegregate in this patient.

DISCUSSION

We generated clonal viral populations resulting from single infectious events by viruses present in a single specimen of plasma from HIV-1-infected individuals. A comparison of the genomic sequences of these clonal viruses from each of six patients supports the idea that recombination events strongly influence viral diversity during the evolution of infection in a given patient. Our results demonstrate that (i) the overall frequency of recombination is very high; (ii) when bottlenecks occur within a selected region of the genome, recombination may allow the maintenance of diversity in adjacent regions; (iii) shuffling of resistance mutations within the protease and RT and between these regions can generate diversity in resistance genotypes; (iv) recombination between different gene segments also makes an important contribution to diversity in the *env* gene; and (v) recombination between viruses with distinct tropism is clearly occurring.

The analysis of clonal viruses proved to be a useful method to evaluate recombination among contemporaneous viruses from infected patients. Control experiments indicated that under limiting-dilution conditions, most wells had been infected with a single virus, and the wells in which more than a single infectious event had occurred could be identified and excluded after sequence analysis. In studies evaluating the replication of viruses carrying protease resistance mutations that impaired viral replicative capacity, no evidence for the reversion of mutations or the appearance of compensatory mutations was found, suggesting that the time in culture necessary to produce viral clones was not sufficient for variants occurring in vitro to emerge to the extent that their presence was perceptible in bulk amplification products obtained by amplifying proviral DNA. The sequencing of proviral DNA from infected wells in the course of these studies further supported these conclusions. In only 2 of the 129 wells evaluated were ambiguous bases identified in any of the genomic regions that were sequenced. The low frequency of sequences containing ambiguous signals further supports the idea that most wells had been infected with a single virus and that the evolution of viral sequences during culture had little or no impact on the results. By using this approach, any segment of the genome is accessible to study, and evidence of recombination or linkage between distant regions can be evaluated.

Our results confirmed studies performed using limiting-dilution PCR that contemporaneous viruses show evidence of extensive recombination. For the six patients studied here, estimates of the rate of recombination relative to the rate of mutation obtained using a coalescent approach ranged from 0.1 to 5.7, and the recombination rate exceeded the mutation rate for three individuals. Shriner et al., analyzing a different genomic region using sequences obtained by limiting-dilution PCR from a single patient, found that the recombination rate was 5.3-fold higher than the mutation rate (51). The method used in these studies assumes neutral evolution, and extensive homoplasy could result in an overestimation of the recombination rate. Analyses restricted to the evaluation of synonymous mutations, however, gave similar results, supporting the conclusion that this bias did not strongly influence the results. Further analysis of our clones suggested that recombination could have an impact on viral diversity in several distinct but important ways.

Recombination allows regeneration of viral diversity following bottlenecks. The extent of diversity in different portions of the genome, as measured by the accumulation of silent mutations, could be quite variable. In particular, low levels of diversity were often encountered for the protease region in patients receiving antiretroviral treatment. The presence of such bottlenecks is consistent with the idea that during the evolution of drug resistance, the occurrence of specific combinations of resistance mutations that give a decisive fitness advantage is followed by strong selection of viruses carrying this genotype. Previous studies evaluating the impact of treatment-induced bottlenecks occurring in the protease or RT on diversity in other genomic regions have given varying results, with evidence of reduced diversity, transient reductions in diversity, or

no effects having been observed (4, 7, 10, 13, 19, 24, 37, 53). Our results suggest that because of the extensive recombination between viral quasispecies, bottlenecks in one region did not preclude the preservation of considerable diversity in other genomic regions. It should be emphasized, however, that none of the patients we evaluated had marked reductions in viral load resulting from treatment. Thus, during the time necessary for selection of an optimal protease sequence, ample opportunity was available for recombination between other loci. In contrast, patients in whom reductions in *env* diversity has been noted have usually demonstrated marked reductions in viral load prior to the emergence of resistant strains (10, 19, 24, 37), thereby limiting potential partners for recombination. We also observed evidence that viruses had recovered *env* segments from highly divergent strains (possibly from archived viruses or minority populations replicating in protected sites) through recombination, suggesting a mechanism through which *env* diversity could be reconstituted, even in cases where transient bottlenecks had occurred. Findings consistent with this scenario have recently been presented (24). Thus, our results suggest that recombination can help restrict evolutionary bottlenecks to minimal segments of the genome carrying the selective advantage while preserving diversity in adjacent regions.

Recombination contributes to the generation of diversity. Our results also emphasize that recombination plays an important role in the generation of diversity. When individual envelope domains were evaluated, we found that the diversity of a given domain was often limited to a relatively small number of distinct genotypes. For example, no more than four distinct V3 loop sequences were detected for any patient; for several patients, only a single V3 loop sequence was identified. Most patients studied here were known to have been infected with HIV-1 for more than 10 years and therefore had probably passed the time when peak diversity in envelope is seen (50). Recombination between envelope domains, however, was extensive. Thus, even when only limited diversity was observed in individual domains, envelope sequences, when considered in their entirety, showed considerably greater diversity. When all variable and constant envelope regions were included in the analysis, each clone from patients 1 to 4 expressed a unique combination of envelope domains and, although clones using the same combination of envelope domains were observed for patients 5 and 6, these populations represented only 3 of 34 and 4 of 18 total clones, respectively.

Similarly, recombination contributed to the generation of diversity in resistance genotypes, as has previously been observed in vitro (33). Viruses from a given treated patient expressed in common most of the protease and RT resistance mutations identified, consistent with the existence of prior bottlenecks in these regions. In most patients, however, some positions associated with resistance were polymorphic. In cases where such polymorphic sites were expressed by more than a single clone, considerable shuffling of mutations, both within the protease or RT and between these regions, was also observed. Thus, patient 5 had two polymorphic sites in RT and two in protease, but nine distinct resistance genotypes were observed, and the most common genotype was expressed by only 38% of the viral clones studied.

It should be emphasized that our analysis was restricted to viral clones that were infectious for the MT4-R5 target cells. As previously described by others (11, 25, 43), we found that the proportion of viruses in plasma that gave rise to infectious events was extremely low, varying from 0.002 to 0.016% for the six patients studied here. Both viruses using CCR5 and CXCR4 coreceptors were identified, suggesting that tropism did not restrict infectivity in MT4 cells. Numerous other factors may contribute to the apparent low infectivity of plasma viruses, including the presence of lethal mutations in the genome, structural defects in the particles, neutralization by antibodies, and loss of infectivity in the interval between venipuncture and infection of target cells. Viral recombination may also have a deleterious effect on infectivity of plasma viruses by associating incompatible gene segments in individual virions; given the high rates of recombination observed in this study, this factor may play an important role. In this regard, it is noteworthy that among the infectious clonal viruses studied here, no example of recombination in the envelope V3 loop that modified the amino acid sequence of this region was identified, whereas apparent recombination within other envelope segments was frequent. Studies comparing the envelope sequences of infectious clonal viruses and sequences derived by limiting-dilution PCR from the same sample of plasma coupled with evaluation of the functional properties of the corresponding gene segments would help define the extent that recombination may impair the infectivity of viruses in vivo.

The reasons for the prevalence of recombinogenic "sexual" reproduction in biology, a characteristic shared by HIV-1, have not been defined. Currently, two general models are often invoked: recombination could increase the rate of adaptive evolution, and/or improve the efficiency of eliminating the effects damaging mutations (15, 40). Recently, an evaluation of the impact of combinations of drug resistance mutations on fitness in the absence of drug pressure suggested that epistasis between individually deleterious mutations in HIV-1 is positive (antagonistic) (2). Although the epistasis between beneficial mutations was not examined, these findings do not offer support for the second alternative. Our findings demonstrate that recombination in HIV-1 makes an important contribution to the diversity of viral quasispecies, both by restricting evolutionary bottlenecks to gene segments and by generating novel genotypes in *pol* and *env*. This increased diversity resulting from recombination may be critical to adaptive evolution in the face of changes in immune pressure or antiretroviral therapy. Further studies evaluating the contribution of recombination to the sequential evolutionary changes observed for individual patients will be helpful in exploring this possibility.

ACKNOWLEDGMENTS

This work was supported in part by the Agence Nationale de Recherches sur le SIDA (ANRS) grant 2005/002. C.C. is a fellow of the ANRS.

REFERENCES

1. **Adzhubei, A. A., I. A. Adzhubei, I. A. Krasheninnikov, and S. Neidle.** 1996. Non-random usage of 'degenerate' codons is related to protein three-dimensional structure. FEBS Lett. **399:**78–82.
2. **Bonhoeffer, S., C. Chappey, N. T. Parkin, J. M. Whitcomb, and C. J. Petropoulos.** 2004. Evidence for positive epistasis in HIV-1. Science **306:**1547–1550.
3. **Boone, L. R., and A. M. Skalka.** 1981. Viral DNA synthesized in vitro by avian retrovirus particles permeabilized with melittin. II. Evidence for a strand displacement mechanism in plus-strand synthesis. J. Virol. **37:**117–126.

4. **Brown, A. J., and A. Cleland.** 1996. Independent evolution of the *env* and *pol* genes of HIV-1 during zidovudine therapy. AIDS **10:**1067–1073.
5. **Charpentier, C., D. E. Dwyer, F. Mammano, D. Lecossier, F. Clavel, and A. J. Hance.** 2004. Role of minority populations of human immunodeficiency virus type 1 in the evolution of viral resistance to protease inhibitors. J. Virol. **78:**4234–4247.
6. **Chen, J., Q. Dang, D. Unutmaz, V. K. Pathak, F. Maldarelli, D. Powell, and W. S. Hu.** 2005. Mechanisms of nonrandom human immunodeficiency virus type 1 infection and double infection: preference in virus entry is important but is not the sole factor. J. Virol. **79:**4140–4149.
7. **Cleland, A., H. G. Watson, P. Robertson, C. A. Ludlam, and A. J. Brown.** 1996. Evolution of zidovudine resistance-associated genotypes in human immunodeficiency virus type 1-infected patients. J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. **12:**6–18.
8. **Coffin, J. M.** 1979. Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. J. Gen. Virol. **42:**1–26.
9. **Dang, Q., J. Chen, D. Unutmaz, J. M. Coffin, V. K. Pathak, D. Powell, V. N. KewalRamani, F. Maldarelli, and W. S. Hu.** 2004. Nonrandom HIV-1 infection and double infection via direct and cell-mediated pathways. Proc. Natl. Acad. Sci. USA **101:**632–637.
10. **Delwart, E. L., H. Pan, A. Neumann, and M. Markowitz.** 1998. Rapid, transient changes at the env locus of plasma human immunodeficiency virus type 1 populations during the emergence of protease inhibitor resistance. J. Virol. **72:**2416–2421.
11. **Dimitrov, D. S., R. L. Willey, H. Sato, L. J. Chang, R. Blumenthal, and M. A. Martin.** 1993. Quantitation of human immunodeficiency virus type 1 infection kinetics. J. Virol. **67:**2182–2190.
12. **Dixit, N. M., and A. S. Perelson.** 2005. HIV dynamics with multiple infections of target cells. Proc. Natl. Acad. Sci. USA **102:**8198–8203.
13. **Dykes, C., P. Mootsikapun, A. Dexter, L. Berrios, M. Chiulli, R. C. Reichman, and L. M. Demeter.** 2000. Analysis of *env* sequence evolution in human immunodeficiency virus-infected patients receiving therapy with nonnucleoside reverse-transcriptase inhibitors. J. Infect. Dis. **182:**316–320.
14. **Farris, J. S., M. Källersjö, A. G. Kluge, and C. Bult.** 1995. Testing significance of incongruence. Cladistics **10:**315–319.
15. **Goddard, M. R., H. C. Godfray, and A. Burt.** 2005. Sex increases the efficacy of natural selection in experimental yeast populations. Nature **434:**636–640.
16. **Hu, W. S., and H. M. Temin.** 1990. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. Proc. Natl. Acad. Sci. USA **87:**1556–1560.
17. **Hu, W. S., and H. M. Temin.** 1990. Retroviral recombination and reverse transcription. Science **250:**1227–1233.
18. **Hudson, R. R.** 2001. Two-locus sampling distributions and their application. Genetics **159:**1805–1817.
19. **Ibanez, A., B. Clotet, and M. A. Martinez.** 2000. Human immunodeficiency virus type 1 population bottleneck during indinavir therapy causes a genetic drift in the *env* quasispecies. J. Gen. Virol. **81:**85–95.
20. **Jensen, M. A., F. S. Li, A. B. van 't Wout, D. C. Nickle, D. Shriner, H. X. He, S. McLaughlin, R. Shankarappa, J. B. Margolick, and J. I. Mullins.** 2003. Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 *env* V3 loop sequences. J. Virol. **77:**13376–13388.
21. **Jetzt, A. E., H. Yu, G. J. Klarmann, Y. Ron, B. D. Preston, and J. P. Dougherty.** 2000. High rate of recombination throughout the human immunodeficiency virus type 1 genome. J. Virol. **74:**1234–1240.
22. **Johnson, V. A., F. Brun-Vezinet, B. Clotet, B. Conway, D. R. Kuritzkes, D. Pillay, J. Schapiro, A. Telenti, and D. Richman.** 2005. Update of the drug resistance mutations in HIV-1: 2005. Top. HIV Med. **13:**51–57.
23. **Judo, M. S., A. B. Wedel, and C. Wilson.** 1998. Stimulation and suppression of PCR-mediated recombination. Nucleic Acids Res. **26:**1819–1825.
24. **Kitrinos, K. M., J. A. Nelson, W. Resch, and R. Swanstrom.** 2005. Effect of a protease inhibitor-induced genetic bottleneck on human immunodeficiency virus type 1 *env* gene populations. J. Virol. **79:**10627–10637.
25. **Layne, S. P., M. J. Merges, M. Dembo, J. L. Spouge, S. R. Conley, J. P. Moore, J. L. Raina, H. Renz, H. R. Gelderblom, and P. L. Nara.** 1992. Factors underlying spontaneous inactivation and susceptibility to neutralization of human immunodeficiency virus. Virology **189:**695–714.
26. **Leitner, T., D. Escanilla, S. Marquina, J. Wahlberg, C. Brostrom, H. B. Hansson, M. Uhlen, and J. Albert.** 1995. Biological and molecular characterization of subtype D, G, and A/D recombinant HIV-1 transmissions in Sweden. Virology **209:**136–146.
27. **Levy, D. N., G. M. Aldrovandi, O. Kutsch, and G. M. Shaw.** 2004. Dynamics of HIV-1 recombination in its natural target cells. Proc. Natl. Acad. Sci. USA **101:**4204–4209.
28. **Magiorkinis, G., D. Paraskevis, A. M. Vandamme, E. Magiorkinis, V. Sypsa, and A. Hatzakis.** 2003. In vivo characteristics of human immunodeficiency virus type 1 intersubtype recombination: determination of hot spots and correlation with sequence similarity. J. Gen. Virol. **84:**2715–2722.
29. **Mammano, F., V. Trouplin, V. Zennou, and F. Clavel.** 2000. Retracing the evolutionary pathways of human immunodeficiency virus type 1 resistance to protease inhibitors: virus fitness in the absence and in the presence of drug. J. Virol. **74:**8524–8531.
30. **Marquet, R., F. Baudin, C. Gabus, J. L. Darlix, M. Mougel, C. Ehresmann, and B. Ehresmann.** 1991. Dimerization of human immunodeficiency virus (type 1) RNA: stimulation by cations and possible mechanism. Nucleic Acids Res. **19:**2349–2357.
31. **McVean, G., P. Awadalla, and P. Fearnhead.** 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics **160:**1231–1241.
32. **Meyerhans, A., J. P. Vartanian, and S. Wain-Hobson.** 1990. DNA recombination during PCR. Nucleic Acids Res. **18:**1687–1691.
33. **Moutouh, L., J. Corbeil, and D. D. Richman.** 1996. Recombination leads to the rapid emergence of HIV-1 dually resistant mutants under selective drug pressure. Proc. Natl. Acad. Sci. USA **93:**6106–6111.
34. **Myers, S. R., and R. C. Griffiths.** 2003. Bounds on the minimum number of recombination events in a sample history. Genetics **163:**375–394.
35. **Negroni, M., and H. Buc.** 2001. Mechanisms of retroviral recombination. Annu. Rev. Genet. **35:**275–302.
36. **Nei, M., and T. Gojobori.** 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:**418–426.
37. **Nijhuis, M., C. A. Boucher, P. Schipper, T. Leitner, R. Schuurman, and J. Albert.** 1998. Stochastic processes strongly influence HIV-1 evolution during suboptimal protease-inhibitor therapy. Proc. Natl. Acad. Sci. USA **95:**14441–14446.
38. **Nikolenko, G. N., E. S. Svarovskaia, K. A. Delviks, and V. K. Pathak.** 2004. Antiretroviral drug resistance mutations in human immunodeficiency virus type 1 reverse transcriptase increase template-switching frequency. J. Virol. **78:**8761–8770.
39. **Onafuwa, A., W. An, N. D. Robson, and A. Telesnitsky.** 2003. Human immunodeficiency virus type 1 genetic recombination is more frequent than that of Moloney murine leukemia virus despite similar template switching rates. J. Virol. **77:**4577–4587.
40. **Otto, S. P., and T. Lenormand.** 2002. Resolving the paradox of sex and recombination. Nat. Rev. Genet. **3:**252–261.
41. **Paillart, J. C., M. Shehu-Xhilaga, R. Marquet, and J. Mak.** 2004. Dimerization of retroviral RNA genomes: an inseparable pair. Nat. Rev. Microbiol. **2:**461–472.
42. **Peeters, M.** 2000. Recombinant HIV sequences: their role in the global epidemic. *In* C. L. Kuiken, B. Foley, B. Hahn, B. Korber, P. A. Marx, F. McCutchan, J. W. Mellors, J. I. Mullins, J. Sodroski, and S. Wolinsky (ed.), The HIV sequence compendium 2000, LA-UR 01–3860. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, N.Mex.
43. **Piatak, M., Jr., M. S. Saag, L. C. Yang, S. J. Clark, J. C. Kappes, K. C. Luk, B. H. Hahn, G. M. Shaw, and J. D. Lifson.** 1993. High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. Science **259:**1749–1754.
44. **Posada, D., K. A. Crandall, and E. C. Holmes.** 2002. Recombination in evolutionary genomics. Annu. Rev. Genet. **36:**75–97.
45. **Rambaut, A., D. Posada, K. A. Crandall, and E. C. Holmes.** 2004. The causes and consequences of HIV evolution. Nat. Rev. Genet. **5:**52–61.
46. **Rhodes, T., H. Wargo, and W. S. Hu.** 2003. High rates of human immunodeficiency virus type 1 recombination: near-random segregation of markers one kilobase apart in one round of viral replication. J. Virol. **77:**11193–11200.
47. **Rhodes, T. D., O. Nikolaitchik, J. Chen, D. Powell, and W. S. Hu.** 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.
48. **Rozas, J., J. C. Sanchez-DelBarrio, X. Messeguer, and R. Rozas.** 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19:**2496–2497.
49. **Sabino, E. C., E. G. Shpaer, M. G. Morgado, B. T. Korber, R. S. Diaz, V. Bongertz, S. Cavalcante, B. Galvao-Castro, J. I. Mullins, and A. Mayer.** 1994. Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiologically linked individuals from Brazil. J. Virol. **68:**6340–6346.
50. **Shankarappa, R., P. Gupta, G. H. Learn, Jr., A. G. Rodrigo, C. R. Rinaldo, Jr., M. C. Gorry, J. I. Mullins, P. L. Nara, and G. D. Ehrlich.** 1998. Evolution of human immunodeficiency virus type 1 envelope sequences in infected individuals with differing disease progression profiles. Virology **241:**251–259.
51. **Shriner, D., A. G. Rodrigo, D. C. Nickle, and J. I. Mullins.** 2004. Pervasive genomic recombination of HIV-1 in vivo. Genetics **167:**1573–1583.
52. **Watterson, G. A.** 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:**256–276.
53. **Zhang, Y. M., S. C. Dawson, D. Landsman, H. C. Lane, and N. P. Salzman.** 1994. Persistence of four related human immunodeficiency virus subtypes during the course of zidovudine therapy: relationship between virion RNA and proviral DNA. J. Virol. **68:**425–432.
54. **Zhuang, J., A. E. Jetzt, G. Sun, H. Yu, G. Klarmann, Y. Ron, B. D. Preston, and J. P. Dougherty.** 2002. Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. J. Virol. **76:**11273–11282.