Virus Population Homogenization following Acute Human Immunodeficiency Virus Type 1 Infection

Gerald H. Learn, 1 David Muthui, 2 Scott J. Brodie, 2 Tuofu Zhu, 2, 3 Kurt Diem, 2
James I. Mullins, 1, 2, 3, 4* and Lawrence Corey 1, 2, 3, 4

Departments of Microbiology, 1 Laboratory Medicine, 2 and Medicine, 3 University of Washington,
Seattle, Washington 98195, and Infectious Diseases Program, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109 4

Received 17 July 2002/Accepted 21 August 2002

Understanding the properties of human immunodeficiency virus type 1 (HIV-1) variants capable of establishing infection is critical to the development of a vaccine against AIDS. Previous studies of men have shown that the HIV-1 env gene is homogeneous early in infection, leading to the suggestion that infection is established by a single transmitted variant. However, we report here that all of eight homosexual men evaluated beginning 3.7 to 9 weeks following onset of symptoms of acute infection harbored diverse virus populations in their blood, with median genetic distances averaging 1.08% in the env C2V5 region and 0.81% in the gag p17 gene. Within another 4.7 to 11 weeks, the variant lineage in env became more homogeneous, while gag sequences continued to diversify. Thus, the homogenization that has been reported to characterize acute infection is actually preceded by the replication of multiple virus variants. This early selective process focuses on viral properties within Env but not Gag p17. Hence, the viral homogeneity observed early in HIV-1 infection results from a selective process that occurs during the establishment of infection.

Protective vaccines should anticipate features of viruses most capable of establishing infection. Hence, a description of the selective events that occur during the earliest stages of human immunodeficiency virus type 1 (HIV-1) infection is important for understanding the requirements for an AIDS vaccine. Most reports to date have shown that infections typically begin with a homogeneous HIV-1 env gene sequence population (16, 30, 32, 33). Multiple hypotheses have been put forth to account for this apparent homogeneity. One is that little infectious virus is normally transmitted, such that descendants of only a single virus establish infection in men infected by homosexual, perinatal, and parenteral contact (5, 12, 20, 27, 32, 33). Recent studies suggest that this is not a likely scenario; women with vaginal exposure tend to have multiple variants replicating early postinfection (19), as if they were exposed to relatively large infectious inocula. Also contrary to this hypothesis is the observation by Cornelissen et al. in two cases of parenteral transmission which shows that a selective outgrowth of syncytium-inducing variants predominated in primary infection but these were subsequently replaced by a homogeneous population of non-syncytium-inducing variants (4). A second hypothesis is that the initial target cells restrict the replication of most viral types. Support for this hypothesis comes from the demonstration of a frequent outgrowth of viruses with a macrophage tropic phenotype and genotypic signature (3, 27). Also in support of the latter hypothesis was the finding by Zhang et al. (32) that while env sequences were homogeneous early in infection, gag sequences had detectable variability, as if selective pressure for particular env characteristics (or linked characteristis) was restricting virus outgrowth in early target cells. However, this homogeneity results whether the route of infection is mucosal or from direct venous injection (25).

While investigating the association of immune response to virus evolution among subjects with acute HIV-1 infection (13), we noted significant genomic variation even in the earliest samples. We systematically evaluated this observation by examining a consecutively enrolled cohort of homosexual men with sexually acquired primary HIV-1 infection. We confirmed the findings of Zhang et al. (32) and demonstrated for the first time that very early after homosexual transmission, multiple distinct env genotypes are capable of replicating, but that soon thereafter another variant population dominates. Our results are therefore consistent with a third hypothesis to account for the early homogeneity of virus population: that multiple strains are transmitted but that their representation is quickly altered as a result of a selective process(es) (whether due to immune responses or the host cell context of infection) that occurs early in infection.

MATERIALS AND METHODS

Study group. The first eight patients seen at the University of Washington Primary Infection Clinic who presented within 66 days (median, 42.5 days) of onset of acute HIV-1 infection were studied. The enrollment criteria, detailed demographic characteristics, and definition of onset of infection of this cohort have been reported (21, 22). In brief, all had an acute symptomatic illness associated with HIV-1 primary viremia that was followed by seroconversion. No participant received antiretroviral therapy during the study period, and each had an absolute CD4+ T-cell count >50/mm3 (median, 590/mm3) and an HIV plasma RNA concentration of >100,000 copies/ml (median, 148,000 copies/ml) at enrollment. Patients were seen at 2-week intervals for the initial 2 months and then at monthly intervals for the following 6 months.

Virus quantitation and genetic analysis. HIV-1 RNA concentrations in plasma were determined by the branched DNA (bDNA) assay (limit of detection of 500 copies/ml; Chiron Corporation, Emeryville, Calif.) as described previously (10, 22). For genetic analysis, DNA was purified from peripheral blood mononuclear cells (PBMCs), or viral RNA was extracted from plasma, reverse tran-
sequences were inferred from nucleotide sequences. The second, nested PCR generated ~625-bp products with internal primers DM7 and DM8 and amplified the C2V5 coding region of gp120.

PCR conditions for p17 gag sequences were similar to those for env except for the use of a higher MgCl₂ concentration (2.0 mM). Primer Gpr.2 (HXB2 1465-1494, 5'-TGTCACCTCTCCCTGGTTCTCCTACTGGC-3') was used for the reverse transcription reaction. The first PCR used Gpr.2 and Gpr.1 (HXB2 593-620, 5'-GATCCCTCGAGACCCCTTGAGCTGG-3') as primers. The nested PCR primers were Gpr.3 (HXB2 763-792, 5'-GACTAGCGGAGGCTATG-3') and Gpr.4 (HXB2 1335-1363, 5'-GGGCGGTGCTCTCCTGTAATGGTCAGA-3') and generated ~572-bp products that encompassed the entire gag p17.

Heteroduplex mobility assays (HMA) and heteroduplex tracking assays (HTA) of HIV-1 env sequences were performed with all patient samples and followed methods described previously (5, 6). For HTA, probes were generated by subjecting PCRs to an additional three cycles of PCR (DM7 and DM8 primers) that included [α-32P]dCTP, as described previously (6), and were used at a 1:100 ratio to probe the PCR product of 50 HIV-1 templates.

PCR products were cloned into the TA vector (Invitrogen, San Diego, Calif.) and selected for sequencing by using procedures to avoid template resampling (24). Sequences were determined with an automated DNA sequencer (Applied Biosystems, Foster City, Calif.) and then edited by using SEQUENCER version 3.0 (Gene Codes Corp., Ann Arbor, Mich.) and evaluated as described by Learn et al. (9). Briefly, sequences were aligned by using CLUSTALW version 1.7 (26) followed by visual inspection and adjustment. Alignment gaps in the nucleotide sequences were introduced between codons where possible. Regions in the alignment that could not be unambiguously aligned were removed from further analyses (9). Pairwise maximum-likelihood nucleotide and protein distances were calculated by using PAUP+ (D. L. Swofford, PAUP 4.0: phylogenetic analysis using parsimony [and other methods], version 4.0b2a, 1999) and PROT-DIST (Kimura option) from version 3.5 of the PHYLIP software package (J. Felsenstein, PHYLIP [phylogeny inference package], version 3.5c, 1993), respectively. Phylogenetic trees and ancestral states at nodes were inferred by using maximum-likelihood phylogenetic methods. Evolutionary models were selected by using the Akaike information criterion (AIC) (1) in Modeltest version 3.06 (17). The parameters of the model for env (TVM + Γ) were as follows: equilibrium nucleotide frequencies, \( f_A = 0.3975, f_C = 0.1651, f_G = 0.1971, f_T = 0.2403; \) shape parameter (\( α \)) of the Γ distribution reflecting site-to-site rate variability of variable sites, \( α = 0.5197; \) R matrix values, \( R_{AC} = 1.6055, R_{AG} = R_{CG} = R_{CT} = 3.2711, R_{GT} = 0.6066, R_{CA} = 0.8723, R_{CG} = 1.0000; \) for gag (HYK + Γ): \( f_A = 0.3596, f_C = 0.1847, f_G = 0.2556, f_T = 0.2001; α = 1.0402; \) transition/transversion ratio Ti/Tv = 2.1638. Ancestral sequences were inferred by using maximum-likelihood estimation for the basal node of the population of sequences from each subject for the gag and env regions studied. These ancestral sequences are taken as the most-recent common ancestors (MRCA) of the infecting viruses for the respective subjects. Pairwise uncorrected synonymous and nonsynonymous nucleotide differences were calculated by using SNAP (http://hiv-web.lanl.gov/content/hiv-db/SNAP/WEBSNAP/SNAP.html) (7, 14). Amino acid sequences were inferred from nucleotide sequences.

Nucleotide sequence accession numbers. The gene sequences determined in this study were deposited in GenBank under accession no. AF418311 to AF418547 and AF418694 to AF418920.

RESULTS

We addressed the origins of the homogeneous virus populations found early in HIV-1 infection through examination of the first eight patients seen at the University of Washington Primary Infection Clinic who presented within 66 days of onset of acute HIV-1 infection (13, 21, 23). Analysis of genetic diversity was assayed at two visits; the first visit was from 26 to 63 days after onset of symptoms (mean = 43.3 days), while the second was at 72 to 137 days (mean = 98 days). HMA of HIV-1 env C2V5 sequences was performed with all patient samples. In four cases, as expected (5), only homoduplexes were found (Fig. 1a, lanes A to H), indicating a largely homogeneous viral sequence population. In the other four cases, however, heteroduplexes were observed at the initial enrollment time points (Fig 1a, lanes A to D), indicating diversity in the viral population. On average, these individuals were first sampled 11 days closer to the onset of symptoms (mean, 39.5 days; range, 26 to 63 days versus a mean of 50.5 days and a range of 42 to 60 days for the other group) and thus to the estimated time of infection.

To evaluate these viral populations in more detail, we generated multiple molecular clones of these fragments and used HMA to identify distinct variants within the initial virus populations (data not shown). Two distinct variants from each subject were used as probes in HTA to evaluate their representation over time. In patient A, an initially abundant sequence substantially decreased in representation by 72 days in both viral RNA in plasma and viral DNA in PBMC (Fig. 1b, first and second lanes of each pair, respectively). However, a second variant was found to increase in representation over time and by 72 days following onset of symptoms had become dominant with little evidence of heterogeneity (Fig. 1c).

Similarly, multiple clones from patients B through D were analyzed by HMA, and distinct variants were used as probes in HTA. In all cases, a variant that was abundant at the first time point rapidly decreased in representation and was a minor species by the next time point examined (Fig. 1b). Furthermore, other variants were again found to increase rapidly in representation over time and become dominant with little evidence of env sequence heterogeneity (Fig. 1c). Thus, multiple HIV-1 variants established infection in these individuals. However, within one to two more months, the populations became overgrown with the descendents of a highly related group or a single variant present within the early mixture.

To detect a heteroduplex mobility shift, and thus viral population diversity, the HMA and HTA used in this study require that the two annealed sequences have mismatches at ≥1 to 2% of the bases, or at least a single insertion/deletion mutation (indel) of any length (6). Finding viral diversity in four of eight patients by HMA suggested the possibility that all eight had diversity that might be detected by the more sensitive means of DNA sequencing. Thus, we extended these observations by sequencing viral env fragments from all eight individuals, including the four that had no evident diversity by HMA. Overall, the median sequence diversity in the C2V5 region at the earliest sampling ranged from 0.40 to 3.02% (average, 1.08%). These levels were on average greater than that expected from either the infidelity inherent to reverse transcription-PCRs (up to about 0.7%) (9) or from evolution from a single infecting strain (about 0.9%/year) (8, 24).

Consistent with the HTA data, the median sequence diversity in env decreased over the next 1 to 2 months in seven of the eight men to a level close to the error rate of the PCR amplification process (range, 0.20 to 1.86%; average, 0.82%). We also measured the divergence of sequences at each time point from the MRCA of each infection (see Fig. 4) and found that env sequences at the later time point were, on average, as close as or closer to this calculated root sequence of the infection (MRCA) for five of the eight subjects (B, C, D, E, and G; 1.26% at the first time point and 1.07% at the second time point).

We next sought to determine whether the emergent homogeneous virus populations resulted from the outgrowth of a single virus or corresponded to a population that shared a
common \textit{env} sequence but was diverse in other regions of the genome. We therefore evaluated another region of the viral genome previously shown to have significant interpatient variability (31, 33), the \textit{gag} p17 coding sequence. The average median variability of \textit{gag} sequences at the first time point examined was slightly lower than in \textit{env} (0.81%; range, 0 to 2.65%). However, the variability of \textit{gag} sequences increased over the same period for five of the eight men (0.26 to 2.66%; average, 1.09%). This increase was seen for subjects B, C, D, E, and H; \textit{env} variability decreased for four of these (B through E) and was unchanged for the fifth. Similarly, the average of the mean distance from the MRCA grew from 0.89 to 1.13% over the same period.

To examine the differences in the evolution of these sequences in the context of natural selection, we used the computer program SNAP (7, 14) to determine the pairwise differences between sequences, both as diversity within sample times and as divergence from MRCA. In general, for \textit{gag}, the distance from MRCA increased from the first visit to the second at both synonymous and nonsynonymous sites; similarly, diversity tended to increase at both classes of sites in \textit{gag}. This pattern is consistent with the expectation that genetic divergence increases over time. For \textit{envC2-V5} the pattern differed: in general, nonsynonymous site divergence from MRCA decreased from the first visit to the second visit while at synonymous sites it increased; nonsynonymous site diversity decreased from visit 1 to visit 2, while the trend at synonymous sites was unclear, decreasing for four patients but increasing for the others.

We used random-effects models (using the subject and the day-by-subject interaction as random effects) to evaluate the diversity at each sample time (Fig. 3), as well as the divergence from the calculated MRCA virus over time (Fig. 4) for both \textit{gag} and \textit{env} sequences. In both cases, there was a significant gene-by-day effect \((P < 0.025\) for diversity, with degree of freedom adjusted for pairwise comparisons; \(P < 0.0001\) for divergence from MRCA), indicating that within these subjects the two genes behaved differently over time. The net effect was a decrease in both diversity and in divergence from MRCA in \textit{env}, while diversity and divergence in \textit{gag} increased.

**DISCUSSION**

Our results confirmed the findings of Zhang et al. (32), which showed higher levels of diversity of \textit{gag} versus \textit{env} sequences in primary HIV-1 infection. However, we demonstrated for the first time that at earlier times after homosexual transmission, multiple distinct \textit{env} genotypes replicate. That divergence from the MRCA decreased over time in \textit{env} indicates that, rather than evolving de novo, the variants that took over the virus population 1 to 2 months later were outgrowths, perhaps clonal with respect to \textit{env}, from the earlier time point. Our results are therefore consistent with a third hypothesis to account for the early homogeneity of the virus population—
that multiple strains are transmitted but that their representa-
tion is quickly altered as a result of a selective process(es) that 
occurs after the establishment of infection. This selection is 
due either to the available host cell milieu prior to the onset of 
detectable immunologic responses or to initial immune re-
spones. No significant differences in neutralizing antibody re-
spone were observed among the subjects (data not shown).

Our results indicate that multiple viruses are transmitted in 
HIV-1 infection at a frequency that is higher than had previ-
ously been estimated. This finding was likely to have been 
obscured in previous studies by the rapid homogenization of 
the virus population we found to occur in the first 3 months of 
infection. Most studies of viral populations in newly HIV-1-
infected people have examined samples taken during subacute

FIG. 2. Phylogenetic reconstruction of HIV-1 env (C2V5 of the gp120 region) (a) and gag p17 (b) sequences. Sequences obtained at the earliest 
time point (filled boxes and circles) and at a later time point (open boxes and circles) are shown (see Fig. 1b for time points used). Arrows indicate 
nodes at which MRCAs were inferred.
stages of infection; usually beginning 4 or more months following seroconversion (28, 29, 33), including studies that have evaluated differences between recently infected men and women (11, 18). Previous work has shown that selective penetration and/or amplification occurs as the virus establishes infection in a new host (33). Because of the apparent homogeneity of these populations, it has been thought that this selection occurred at the local site of entry or at least in the initial stages of growth in cells first contacted within the host (3, 25, 32, 33). Our results refine this view by demonstrating that amplification of a relatively broad population of virus can indeed occur, though not nearly as broad as that likely to be

FIG. 3. Distribution of all pairwise genetic distances (diversity) for patients A through H (panels a to h, respectively) among p17 gag gene sequences (triangles) and C2-V5 env sequences (circles) as a function of number of days after onset of symptoms of acute HIV-1 infection. Regression lines are indicated for gag (solid lines) and env (broken lines).
found in the virus donor (24), and is followed by further homogenizing selection for certain virus features. Although we examined only about 10% of the viral genome, these homogenizing forces appeared to act against env or closely linked sequences, since no evidence of further homogenization of gag gene sequences was evident. Our results are therefore most consistent with the hypothesis that a diverse population of virus is transmitted and grows within the new host. Among this diverse population, those that are most capable of widespread dissemination are selected based on specific features of the virus in the context of the early host immune response, or target cell requirements, resulting in selection for a more refined population with variable as well as conserved sequence elements. This may result from selective outgrowth of variants that have a particular env sequence or be due to intrahost recombination that results in the dissemination of particular env elements across a diverse population of virus backbone sequences. The pattern of increasing synonymous site divergence from the MRCA in env accompanied by decreasing divergence from MRCA at nonsynonymous sites is consistent with purifying selection over the env region examined.

Our data reveal greater complexity in the host-viral interaction early in HIV-1 infection than has been appreciated to date. The process of HIV-1 variant selection that we observed appears before neutralizing antibody responses but at or beyond the time of seroconversion and the appearance of HIV-specific T-cell responses (10, 13). Early HIV-specific cytotoxic T-lymphocyte responses do appear to drive the amount of viral gene diversity in humans (2) and in experimental simian immunodeficiency virus models (15). Delineation of the targeted virus features and selective forces shaping the early virus population should assist development of vaccines that target features of the virus required for efficient infection and dissemination.

FIG. 4. Divergence from the inferred MRCA for patients A through H (panels a to h, respectively) among p17 gag gene sequences (triangles) and C2-V5 env sequences (circles) as a function of number of days after onset of symptoms of acute HIV-1 infection. Regression lines are indicated for gag (solid lines) and env (broken lines).
ACKNOWLEDGMENTS

This work was supported by grants from the U.S. Public Health Service, including to the University of Washington Center for AIDS Research.

We acknowledge A. J. Rossini and D. C. Nickle for help with statistical analyses.

REFERENCES