Selection of HIV Variants with Signature Genotypic Characteristics during Heterosexual Transmission

Manish Sagar,1 Oliver Laeyendecker,3,4 Sandra Lee,2 Jordyn Gamiel,4 Maria J. Wawer,5 Ronald H. Gray,5 David Serwadda,7 Nelson K. Sewankambo,7 James C. Shepherd,4 Jonathan Toma,6 Wei Huang,6 and Thomas C. Quinn3,4

1Brigham and Women’s Hospital and 2Dana Farber Cancer Center, Boston, Massachusetts; 3Laboratory of Immunoregulation, National Institute of Allergy and Infectious Diseases, National Institutes of Health, and 4School of Medicine and 5Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland; 6Monogram Biosciences, South San Francisco, California; and 7Makerere University, Kampala, Uganda

Background. Newly infected subjects acquire a limited number of human immunodeficiency virus type 1 (HIV-1) variants with specific genotypic and phenotypic features from the array of viruses present in a chronically infected transmitting partner.

Methods. We examined HIV-1 envelope sequences from the earliest available serum sample after HIV-1 acquisition in 13 newly infected subjects and from their epidemiologically linked HIV-1–infected heterosexual partner. Samples from both members were collected on the same day in the Rakai Community Cohort Study.

Results. Ten couples were infected with subtype D HIV-1, and 3 pairs had subtype A HIV-1. Newly infected subjects acquired a subset of the viruses that were circulating in the transmitting partner; transmitted variants had less diversity and divergence and were more closely related to the ancestral sequences. The majority of signature amino acid differences among donor and recipient sequences were in and immediately following the V3 loop. Envelopes from recipients were significantly shorter and had a lower V3 charge than envelopes from donors, but there was no significant difference in the number of potential N-linked glycosylation sites.

Conclusion. A minority subset of HIV-1 variants with signature genotypes is favored for transmission in this population.

Regardless of the mode of HIV-1 acquisition and the subtype of the virus, newly infected subjects have only a limited number of variants [1–10]. Furthermore, early in infection, most viruses use the CCR5 receptor [5, 6, 11, 12], and this coreceptor restriction limits HIV-1 transmission to subjects with abnormal CCR5 receptor expression [13–15]. Over the course of infection, the virus evolves, leading to the presence of variants with diverse envelope genotypes and an ability to potentially use other coreceptors, such as CXCR4, for host-cell entry [10, 16–18]. Collectively, these observations suggest that certain variants are selected during transmission, presumably on the basis of specific viral properties and host factors. It is essential to elucidate the viral characteristics that confer fitness for transmission in order to develop vaccines or microbicides that specifically target the HIV-1 variants preferentially acquired by a newly infected subject.

Recent studies suggest that viruses with shorter and less glycosylated envelope variable loops were preferentially acquired by heterosexually infected subjects with subtype A and C HIV-1 [1, 19]. Similar signature envelope characteristics, however, were not observed in subjects newly infected with subtype B HIV-1 who primarily acquired their infection through injection drug use and/or homosexual contact [19, 20]. Because selection during transmission occurs during all modes of transmission and with every HIV-1 subtype, a signature viral
phenotype may confer fitness for transmission; different subtypes, however, may have unique genotypic determinants for influencing this viral property. To provide further insight into the selection of specific HIV-1 envelope genotypes and phenotypes during transmission, we examined envelope sequences from 13 newly infected subjects and their epidemiologically linked heterosexual partners in the Rakai Community Cohort Study. The majority of the couples in this cohort are infected with HIV-1 subtype D, for which envelope characteristics early after transmission have not been elucidated.

SUBJECTS, MATERIALS, AND METHODS

Study subjects. Subjects were retrospectively identified from the previously described Rakai Community Cohort Study in the Rakai district of southwestern Uganda between 1994 and 1999 [21, 22]. In brief, consenting participants were offered free voluntary counseling and testing as individuals and as couples, were provided with health education on HIV prevention and free condoms, and had serum collected for HIV-1 serological analysis every 10 months. We investigated 13 newly infected monogamous subjects with their epidemiologically linked partner on the basis of the availability of serum samples and prior confirmation of molecular linkage of sequence data from the Gag and gp41 regions [23]. In the partner with incident infection, the HIV-1 seroconversion date was estimated as the midpoint between the last visit with an HIV−1−seronegative test result and the first visit during which HIV−1 antibody was detected. The study was approved by human subjects—research review boards at the Uganda Virus Research Institute, the AIDS Research Subcommittee of the Ugandan National Council for Science and Technology, Columbia University, and Johns Hopkins University.

Envelope analysis. HIV-1 RNA was isolated from 200−500 μL of the serum samples, and RT-PCR was used to amplify a library of full-length envelope genes as previously described [24]. Eight individual full-length envelope clones were isolated and sequenced from this library of envelope genes from each serum sample [25]. All unique sequences reported in this publication have been submitted to Genbank (accession numbers EU852934–EU853141).

All sequences were aligned using Clustal X and further manually codon aligned using MacClade, version 4.01. Sequences were gap stripped, and neighbor joining phylogenetic analysis with a general time-reversible model and a gamma distribution rate parameter of 0.5 was used to assess relationships among sequences from newly infected subjects and their epidemiologically linked partners. Sequences were examined for intersubtype recombination, using RIP 3.0 (available at: http://www.hiv.lanl.gov/content/sequence/RIP/RIP.html). Genetic diversity in each subject was estimated using a Kimura 2-parameter model with a gamma rate distribution of 0.5, using MEGA, version 3.1 [26]. Heuristically searched maximum likelihood trees were generated for each transmission pair with outgroup sequences of the same subtype from the Los Alamos database, using the best-fit evolutionary model parameters selected by use of Modeltest, version 3.06 [27], and the software package PAUP, version 4.0b2a [28]. Each tree was also examined for segregation of recipient and donor sequences, using the Maddison-Slatkin test [29] as implemented in MacClade, version 4.01. The minimum number of mixing events between the recipient and donor sequences from the tree was compared with the distribution of migration events for 1000 randomly generated trees to estimate statistical significance.

For each sequence, V1-V2 (131−196), V3 (296−331), V1-V4 (131−418), and V1-V5 (131−463) segments corresponding to HXB2 envelope amino acids were defined. The number of amino acids and predicted N-linked glycosylation sites in each segment were determined through manual counting and by using the tool available at http://www.hiv.lanl.gov/content/sequence/GLYCOSITE/glycosite.html, respectively. Recipient and donor sequences in each partnership were examined for signature amino acid differences, using VESPA (available at: http://www.hiv.lanl.gov/content/sequence/VESPA/vespa.html) [30].

Coreceptor use was examined using the Monogram Biosciences Trofile assay as previously described [24]. Briefly, each subject’s library of full-length envelopes was incorporated into HIV-1 pseudoviruses that express the firefly luciferase gene. Tropism of these pseudoviruses was designated by their ability to infect U87/CD4+/CCXCR5 and/or U87/CD4+/CXCR4 cells, as assessed by luciferase expression.

Statistical analysis. Viral diversity and divergence for the sequences from the newly infected subject and the transmitting partner were compared using the Wilcoxon matched-pairs signed rank test. The median value for other sequence characteristics from the 8 recipient and the 8 donor sequences from each partnership were compared using the Wilcoxon matched-pairs signed rank test. To account for the multiple sequences from each subject, we also compared the sequence characteristics in each pair, using the Wilcoxon rank sum test. The Wilcoxon rank sum test, stratified by pair, was further used for aggregate comparisons among all 13 couples. All statistical analyses were done with Intercooled Stata, version 8.0 (Stata), and SAS, version 8.2 (SAS Institute). All P values are for 2-sided tests.

RESULTS

Subjects and partnerships. HIV-1 envelope sequences from the 13 newly infected subjects were analyzed a median of 189 days (range, 142−359 days) after estimated HIV-1 seroconversion (table 1). All 13 transmitting partners had prevalent HIV-1 infection at the time of enrollment and were followed for a minimum of 2 years before the transmission event. Thus, transmission did not occur during the acute phase of the index partner’s
infection. Eight full-length envelope sequences were examined from each subject. Neighbor joining phylogenetic analysis showed that sequences from the newly infected subject and the transmitting partner clustered together with high bootstrap support, confirming the epidemiological linkage (figure 1). Ten couples had subtype D HIV-1, and subjects in the remaining 3 partnerships (120, 368, and 601) were infected with subtype A HIV-1. No sequences from any subject showed evidence of intersubtype recombination.

**Sequence diversity and divergence.** To better examine genetic relationships among the sequences from the partner with incident infection and the transmitting partner, maximum likelihood trees were constructed for each couple (figure 2). Newly infected subjects had a median genetic diversity of 0.7% (range, 0.3%–2.44%), which was significantly lower than the genetic diversity present in the transmitting partner (median, 1.4% [range, 0.4%–4.6%]; P < .001, by the Wilcoxon matched-pairs signed rank test). Recipient sequences (median distance to the recipient ancestor, 12 [range, 6.5–94]) also demonstrated significantly less divergence, compared with donor sequences (median distance to the donor ancestor, 102 [range, 20–165]; P < .001, by the Wilcoxon matched-pairs signed rank test). In each pair, recipient sequences demonstrated significantly separate clustering relative to the donor sequences, as examined by the Maddison-Slatkin test. In each couple, recipient and donor sequences could be separated by a maximum of 2 steps. Significantly less than 5% of the 1000 randomly generated trees from each partnership’s sequences could be separated by ≤2 steps, which suggests that the recipient and donor sequences in each pair showed significantly different clustering (P < .005). The distinct clustering and decreased diversity and divergence suggest that a limited number of minority variants in the donors were preferentially favored for transmission during heterosexual transmission in Rakai, Uganda.

**Sequence characteristics.** To evaluate some of the characteristics of these minority variants, the maximum likelihood trees were used to compute branch lengths to the one common ancestor that linked all the donor and recipient sequences in each couple (figure 2). The median length of the newly infected subjects’ sequence branch to the common ancestor (median of the 13 recipient medians, 116.5 [range, 8–167.5]) was significantly shorter than that for the index partners’ sequence branch (median of the 13 donor medians, 133 [range, 25–175.5]; P = .04, by the Wilcoxon matched-pairs signed rank test). Because we isolated multiple sequences from every subject, we assessed the difference in the distribution of distance to the common ancestor between recipients and donors in each couple. In every pair except couple 108, the newly infected subject showed distributions with smaller distances to the common ancestor, compared with the donor. Recipients had significantly smaller branch lengths to the common ancestor than the donors in aggregate comparisons involving either 13 pairs or 10 pairs with subtype D HIV-1 (P < .001 for both comparisons, by the Wilcoxon rank sum test stratified by couple). Smaller branch lengths to the common ancestor suggest that recipient sequences
were more closely related to ancestral sequences than the majority of viruses circulating in the donor.

Sequences from each partnership were evaluated for envelope length and glycosylation. The median length and the number of predicted N-linked glycosylation sites of the donor and recipient sequences were not significantly different in any segment of the envelope (i.e., V1-V2, V3, V1–V4, and V1–V5) (P > .05, by the Wilcoxon matched-pairs signed rank test). To account for the multiple sequences isolated from each subject, we again compared the distribution of sequence characteristics among the newly infected individual and the transmitting partner within each pair. In general, envelope V1-V2, V1–V4, and V1–V5 lengths in recipient variants were shorter than those in donor variants in a majority of the partnerships. For instance, in only 3 of 13 couples (295, 338, and 605), the transmitting partner had shorter envelope V1–V4 regions than the newly infected subject (figure 3). In aggregate, recipient envelope V1-V2, V1–V4, and V1–V5 sequences were significantly shorter than donor sequences (P = .002, P < .001, and P = .001, respectively, by the Wilcoxon rank sum test stratified by couple). V1-V2, V1–V4, and V1–V5 sequences in newly infected subjects remained significantly shorter than those in their transmitting partners (P = .009, P = .004, and P = .006, respectively, by the Wilcoxon rank sum test stratified by couple), even when the analysis was limited to the 10 partnerships with subtype D HIV-1. Among all couples, however, there was no significant difference in the number of predicted N-linked glycosylation sites between recipient and donor sequences in the V1-V2, V1–V4, or V1–V5 segments (figure 3).

To identify other differences among envelope sequences from newly infected subjects and the viruses circulating in the transmitting partner, we examined sequences for signature amino acid differences, using VESPA [30]. VESPA calculates the frequency of each amino acid at each position in an alignment for the query and the reference set and selects the positions for which the most common character in the query set differs from...
Figure 2. Findings of maximum likelihood tree analysis of donor sequences (blue boxes) and recipient sequences (green boxes), showing that transmitted viruses are minority ancestral variants. See the text for a description of methods and materials. Arrows, most recent common ancestors; targets, donor ancestor nodes; asterisks, recipient ancestor nodes.
that in the background set. A median of 18 signature amino acids (range, 8–50 amino acids) differentiated the consensus donor sequence from the consensus recipient sequence in different partnerships (table 2, which is available only in the electronic edition of the Journal). When all sequences from newly infected subjects were compared with all sequences from the transmitting partners, however, there was no consistent amino acid difference that differentiated recipient and donor envelopes. Thus, signature amino acid differences were specific to each partnership. We hypothesized that, although there were no consistent amino acid differences, signature amino acids may be concentrated in specific regions of the envelope gene. Among all couples, signature amino acid differences were recorded and tallied relative to the HXB2 envelope sequence [31] to identify a region of the envelope gene with a concentration of signature amino acid differences (figure 4). By visual inspection alone, the majority of predicated amino acid differences among the recipient and the donor sequences were in and immediately following the envelope V3 loop and in the heptad region 2 domain of gp41.

Because signature amino acid differences among donor and recipient viruses were concentrated in and around the envelope V3 loop, we examined the V3 segment for specific characteristics known to influence its function. For subtype B HIV-1, the V3 loop charge [32] and the presence of positively charged amino acids at position 11 and 25 [33–35] have been associated with specific coreceptor use and cellular tropism. In all pairs except 108, 295, and 601, recipient sequences had a lower V3 charge, compared with donor envelopes (figure 5). In aggregate, recipient V3 loops had a significantly lower charge, compared with donor V3 sequences (\( P = 0.007 \), by the Wilcoxon rank sum test stratified by couple), and this difference remained significant when the analysis was restricted to the 10 couples with subtype D HIV-1 (\( P = 0.004 \), by the Wilcoxon rank sum test stratified by couple). The consensus donor sequence had a higher charge than the consensus recipient sequence at positions 11 and/or 25 of the V3 loop in only 3 couples (32, 326, and 605), suggesting that the charge differences were not primarily because of amino acid differences at these positions of the V3 loop. Because the

Table 2. Signature amino acid differences among recipients and donors, relative to position on the viral envelope gene.

The table is available in its entirety in the online edition of the Journal of Infectious Diseases.
ability to use the CXCR4 receptor as opposed to the CCR5 receptor has been associated with a higher V3 charge in subtype B HIV-1, we examined coreceptor tropism, using the Trofile assay [24]. Nearly all subjects had viruses that strictly used the CCR5 receptor; those from the newly infected individual from couple 295 showed dual or mixed tropism (table 1). Interestingly, V3 sequences of clones from this recipient had a positively charged amino acid at position 25 and a higher V3 charge, compared with V3 loops in the index partner (figure 5). CXCR4-using virus (X4 or dual) was detected in the virus population of the recipient but not the index partner of couple 295, suggesting that CXCR4-using variants may have arisen de novo or may have been transmitted from a minority of the virus population that was present in the partner. Taken together, our data suggest that the higher V3 charge differences observed in donors did not reflect that they had a greater number of CXCR4-using viruses than newly infected subjects. Indeed, previous studies have suggested that, within subtype D HIV-1, coreceptor use differences are not exclusively associated with envelope V3 loop sequence changes [25].

DISCUSSION

In this study, we examined a minimum of 8 envelope sequences from 13 newly infected subjects and their chronically infected heterosexual partners from retrospectively identified discordant couples in a population-based cohort in the Rakai district of Uganda. Other publications have also investigated sequences from newly infected subjects and their sexual partners [1, 5, 20],

Figure 4. Signature amino acid differences cluster around the V3 loop. Variable loops 1, 2, 4, and 5 were not included in the analysis because subject sequences could not be reliably aligned. The V3 loop and heptad region 2 (HR2) borders are delineated with red lines. Positions of the envelope gp120, gp41, C1-C2, C3-C5, and heptad region 1 (HR1) regions are also shown. See the text for a description of methods and materials.

Figure 5. V3 charges for viruses from donors (D) and recipients (R), showing that transmitted viruses have a lower V3 charge. Boxes, interquartile ranges; horizontal lines, median values.
but our study examined the largest number of discordant couples to date. Furthermore, subjects in this cohort were primarily infected with subtype D HIV-1, which has not been examined in detail previously. Our analyses showed that a limited number of minority HIV-1 variants closely related to ancestral sequences with condensed envelopes and less overall V3 charge are preferentially acquired by the naive host from the transmitting partner. In aggregate, our findings from the Rakai population, in which HIV-1 subtype D is the predominant variant transmitted, suggest that a small number of viruses with specific envelope genotypic characteristics are selected during heterosexual transmission, which accords with findings from other studies [1, 19].

We showed that the transmitted viruses are closely related to the ancestor sequences; this finding has important implications for the relationship between HIV-1 envelope evolution and pathogenesis. The HIV-1 envelope evolves in a host presumably to escape host immune responses and to increase host-cell range [36]. Because newly acquired viruses are more closely related to the ancestor sequences than the majority of viruses that have evolved in a chronically infected host, envelope modifications that occur in a host over the course of infection may decrease a virus’ fitness for transmission. Similar observations in heterosexual transmission pairs with subtype C HIV-1 [1, 37] and from an analysis of longitudinal samples from subjects infected with subtype B HIV-1 [37] suggest that HIV-1 of different subtypes are similarly constrained from productively infecting and disseminating in a naive host by changes that occur in the envelope glycoprotein over the course of infection. Further studies are needed to determine the biological mechanism for the disadvantage that these modifications confer during transmission.

An increase in the number or a shift in the position of the glycosylated residues in the envelope glycoprotein are changes that appear to occur in various HIV-1 subtypes in a host over the course of infection [38, 39]. Less glycosylated HIV-1 variants, however, are favored during heterosexual transmission of HIV-1 subtypes A and C [1, 19]. In our study, transmitted viruses had significantly smaller envelopes but no significant difference in predicted glycosylation sites, compared with variants circulating in the chronically infected partner. This suggests that, among the viruses circulating in Rakai, most of which are subtype D HIV-1, there is preferential selection for variants with compact but not necessarily less glycosylated envelopes. Samples from the newly infected subjects in our study, however, were obtained an average of \(\sim 5\) months after estimated seroconversion, which is longer than the sampling interval of \(\sim 2-4\) months after the last visit with a seronegative test result for individuals infected with subtypes A and C [1, 19]. Given that envelope characteristics, especially the level of glycosylation, can evolve in the early months after HIV-1 infection [10, 38, 40], it is possible that we failed to observe a significant difference in predicted glycosylation between recipient and donor viruses because the sampling interval for our newly infected subjects was longer than that for subjects in other studies. If modifications in the number and type of glycosylation sites are fixed in the virus population before changes in envelope length occur, the longer sampling interval may have precluded us from observing a significant difference in the number of predicted glycosylated sites between recipient and donor viruses.

We also showed that signature amino acid differences among recipient and donor sequences were primarily located in and immediately following the V3 loop. Interestingly, there was high variability in sequences immediately following the V3 loop in HIV-1 subtypes B and C between persons with early infection and those with chronic infection [41]. We further observed that variants in newly infected subjects had less overall V3 charge than HIV variants in the transmitting partner. Previously published subtype B and C envelope sequences from discordant couples did not show similar V3 charge differences [1, 20]. It is possible that HIV-1 subtypes B and C have a limited range of V3 charges among different variants, which may preclude the ability to document significant differences among viruses in the recipient and the donor. Although the proportion of variants that use the CXCR4 coreceptor is greater among HIV-1 subtype D than among other HIV-1 subtypes [25], V3 charge differences among donor and recipients was not associated with differences in coreceptor use.

We identified common sequence characteristics, such as smaller envelopes and less overall V3 charge, by examining aggregate differences among a cohort of discordant partners. These signature characteristics, however, were not observed in every couple in our study. For instance, in couple 108, HIV-1 variants in newly infected subjects did not have any significant differences in envelope length or V3 charge from those of viruses in their corresponding transmitting partners. Furthermore, signature genotypes, such as condensed and less glycosylated envelopes, have not been defined among all newly acquired viruses, particularly subtype B HIV-1 [19, 20]. Selection of a limited number of HIV-1 variants, however, occurs during all routes of HIV-1 acquisition and among all HIV-1 subtypes examined to date [1–7, 9, 10]. Taken together, these observations suggest that, regardless of the subtype, mode of HIV-1 acquisition, or index partners’ HIV-1 variants, only a limited number of viruses are successfully transmitted to a naive host. This selection during transmission may be based on a viral property that confers an advantage during transmission, although different HIV-1 subtypes and/or variants in an individual subject may have unique genotypic means for imparting this phenotypic characteristic.

It remains unclear how the differences in envelope length and V3 charge confer an advantage during transmission. Previous studies have suggested that envelope length may influence viral replicative capacity [39, 42]. In addition, V3 charge differences could affect the affinity of the envelope glycoprotein for the host cell coreceptor, although our studies suggest that it did not influence coreceptor use. Thus, differences in envelope length and
V3 charge may augment the ability of envelopes to bind receptors and initiate host cell fusion, which, if true, may increase the capacity to productively infect early target cells and disseminate from the site of invasion. Further studies are needed to elucidate the mechanism of how the signature envelope genotypes confer an advantage during transmission. Detailed insight into the biological mechanisms of selection during transmission will greatly aid the ability to target the variants favored during transmission with a vaccine or a microbicide.

Acknowledgments

We thank the subjects who contributed samples for these studies as part of the Rakai Health Sciences Project and Simon Frost for providing aligned envelope sequences from transmission pairs with subtype B HIV-1 infection [20].

References

35. Guay LA, Musoke P, Fleming T, et al. Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-


