

Sequence Note

Identification of Ugandan HIV Type 1 Variants with Unique Patterns of Recombination in *pol* Involving Subtypes A and D

SUSAN H. ESHLEMAN,¹ MATTHEW J. GONZALES,² GRAZIELLA BECKER-PERGOLA,¹
SHAWN C. CUNNINGHAM,¹ LAURA A. GUAY,¹ J. BROOKS JACKSON,¹ and ROBERT W. SHAFER²

ABSTRACT

Most HIV-1 infections in Uganda are caused by subtypes A and D. The prevalence of recombination and the sites of specific breakpoints between these subtypes have not been reported. HIV-1 *pol* sequences encoding protease (amino acids 1–99) and reverse transcriptase (amino acids 1–324) from 102 pregnant Ugandan women were analyzed by the Recombinant Identification Program, SimPlot, and examination of phylogenetically informative sites to identify sites of recombination between sequence segments belonging to different subtypes. Thirteen percent (13 of 102) of the *pol* sequences contained strong evidence of recombination between subtypes A and D. At least nine different patterns of recombination were observed. Five women infected with a recombinant virus transmitted the recombinant virus perinatally. In this population-based study, intersubtype recombinants were common. The large number of different types of *pol* recombinants identified suggests that recombination occurs readily in the *pol* region. Perinatal transmission of the recombinant viruses demonstrates their evolutionary stability.

DURING ITS SPREAD among humans, the main (M) group of human immunodeficiency virus type 1 (HIV-1) has developed an extraordinary degree of genetic diversity.^{1–4} To genetically classify global group M HIV-1 isolates, researchers initially developed a subtyping system of HIV-1 based on *env* and *gag* sequences. However, it was soon realized that certain isolates belonged to different subtypes depending on the region of the genome analyzed. In 2000, a more comprehensive system based on complete genome analysis categorized viruses into nine pure subtypes (A, B, C, D, F, G, H, J, and K), six circulating recombinant forms (CRFs), and incidental viral variants.^{2,3}

In Uganda, most HIV-1 infections are caused by subtypes A and D.^{5,6} In a previous study of 27 HIV-1-infected volunteer blood donors in Uganda, we found intersubtype recombination between *env* and *pol* (protease) in six individuals and between *gag* and *pol* in five individuals.⁷ In a separate study, we identified dual infection with subtypes A and D in two pregnant Ugandan women⁸; in one case, both subtypes were transmitted perinatally to the infant.⁹ We have also previously described

the subtyping and the presence of drug resistance mutations in a 1269-bp region of *pol* in HIV-1 from 102 pregnant Ugandan women enrolled in a trial of antiretroviral therapy to prevent perinatal HIV-1 transmission. In that study, 50 (49%) women had subtype A viruses, 35 (34%) had subtype D viruses, 4 (4%) had subtype C viruses, and subtyping was inconclusive for 13.¹⁰ Now, we examine each of the 102 sequences for evidence of recombination, including the 13 that were not subtyped in the previous study, and report the prevalence and sites of recombination in viruses from this representative patient population.

SAMPLES USED FOR ANALYSIS

Plasma samples used for analysis were obtained from antiretroviral drug-naïve pregnant women enrolled in the clinical trial HIVNET 012 (Kampala, Uganda) and who were treated perinatally with nevirapine (NVP).^{11,12} The HIVNET 012 study protocol was reviewed and approved by institutional review boards in Uganda and the United States, and informed consent

¹Department of Pathology, Johns Hopkins Medical Institutions, Baltimore, Maryland 21205.

²Division of Infectious Diseases, Stanford University, Stanford, California 94305.

was obtained from all women before enrollment. All the women were antiretroviral drug naive at the time of enrollment. Women received a single dose of NVP during labor and delivery and infants received a single dose of NVP within 72 hr of birth. Samples used for analysis were collected 6–8 weeks after delivery (6–8 weeks after NVP administration) from 102 women and 5 HIV-1-infected infants. Samples collected from nine women before NVP administration were also analyzed. The subtype and presence of drug resistance mutations in protease and first 324 amino acids of reverse transcriptase (RT) before and after prophylaxis have already been described.¹³

SEQUENCE ANALYSIS

Two programs, the Recombinant Identification Program (RIP; <http://hiv-web.lanl.gov>)¹⁴ and SimPlot¹⁵ were used to identify sites of recombination. RIP was used to compute the genetic distance between short segments (200 bp) of the Ugandan sequences and the aligned segments of a panel of reference sequences of known subtype. SimPlot was used to determine the phylogenetic relationship of each sequence segment to aligned segments of reference sequences, using bootstrap resampling (bootscanning). One hundred bootstrap replicates generated by the neighbor-joining method were obtained, using both 200- and 400-bp segments (sliding window size) overlapping by a step of 10 bp.

The reference sequences used for both RIP and SimPlot were obtained from the HIV Sequence Database at the Los Alamos National Laboratory (<http://hiv-web.lanl.gov>). Both RIP and SimPlot identified putative positions of recombination by plotting either genetic similarity values (RIP) or percentage of permuted trees (SimPlot) along the length of each sequence. Sequences were considered to have strong evidence of recombination if both RIP and SimPlot identified sites of recombination using a window size of 400 with a maximum bootstrap percentage of at least 90%. Sequences were considered to have weak evidence of recombination if either RIP or SimPlot identified sites of recombination using a window size of 200 with a maximum bootstrap value of >50% but <90%.

To identify phylogenetically informative nucleic positions, we aligned 10 previously published nonrecombinant subtype A sequences (AF004885, AF069672, AF069670, AF069671, AF107771, M62320, U51190, AF193275, AF069669, and AF069673) and 6 nonrecombinant subtype D sequences (K03454, M27323, U88822, M22639, U88824, and AF133821). Positions at which the subtype A sequence always differed from the subtype D sequence (e.g., the subtype A alignment had all C's and the subtype D alignment had all T's) were considered informative. Positions at which either subtype A or D contained a nucleotide not present in the other alignment were considered potentially informative (e.g., the subtype A alignment had all C's and the subtype D alignment had C's and at least two T's).

To further validate putative sites of recombination, aligned segments were then analyzed by PAUPSearch as part of the Wisconsin Package version 10.1 (Genetics Computer Group, Madison, WI). Neighbor-joining trees using the Kimura two-parameter nucleotide substitution model were created. To determine the reliability of the three topology, bootstrap resampling was performed 500 times for each segment.

PREVALENCE OF RECOMBINATION

Thirteen of 102 sequences were found to have strong evidence of recombination by RIP and by the most conservative bootscan setting, which used a window size of 400 bp. For each of these 13 sequences, the percentage of permuted trees (i.e., bootstrap value) in at least one window was >98%. Phylogenetic trees constructed by splitting these sequences at putative sites of recombination and aligning them with reference sequences confirmed the presence of a second subtype. The bootscan results for these 13 isolates are shown in Fig. 1.

Among the 102 sequences analyzed, 15 sequences were found to have weak evidence of recombination by RIP and SimPlot (bootscanning), using a window size of 200 bp. The bootscan support for recombination in these 15 sequences was low; the percentage of permuted trees never exceeded 84% (range, 50–84%). Six sequences had evidence of recombination only by bootscanning with a window size of 200 bp, 20 sequences had evidence of recombination only by RIP, and 48 sequences had no evidence of recombination by either RIP or bootscanning.

To determine whether RIP or SimPlot identifies sites of recombination by chance in isolates not believed to be recombinants, we applied both programs to a set of 63 *pol* genes of previously published completely sequenced isolates obtained outside of Uganda. The 26 sequences included the following subtypes: 26 subtype B, 13 subtype C, 10 CRF01_AE (in which recombination has been reported only within *env*), 4 subtype A, 4 subtype D, 4 subtype F, and 2 subtype G sequences. None of these sequences had strong evidence of recombination. However, 33 of 63 sequences were found to have weak evidence of recombination by RIP or by bootscanning with a window size of 200 bp. Only 4 of the 33 sequences had evidence of recombination by bootscanning with a window size of 400 bp, including two subtype D isolates with subtype B peaks having bootstrap support of 26 and 29%, one subtype F isolate with a subtype B peak of 48%, and one subtype G isolate with a subtype A peak of 80%.

PATTERNS OF RECOMBINATION

Among the 13 isolates with strong evidence of recombination, there appeared to be 9 different recombination patterns in the RT gene (Fig. 1). Three isolates (549, 760, and 800) shared evidence of recombination between codons 211 and 240. Two isolates (151 and 479) shared evidence of recombination between codons 117 and 120. Two isolates (414 and 843) shared evidence of two sites of recombination between codons 31–38 and codons 214–227. The mean interisolate divergence among these three groups of isolates ranged between 3.2 and 4.1%. None of the women in this study, including those who had viruses sharing evidence of recombination at similar sites, are known to be epidemiologically related.

The alignment of previously published nonrecombinant subtype A and D sequences revealed that 99 of the first 972 nucleotides (codons 1–324) in the RT were either informative (16 nucleotide positions) or potentially informative (83 nucleotide positions) as defined above. Among the 13 isolates with strong evidence of recombination, a median of 34 positions of the possible 99 were informative in that they could be used to inde-

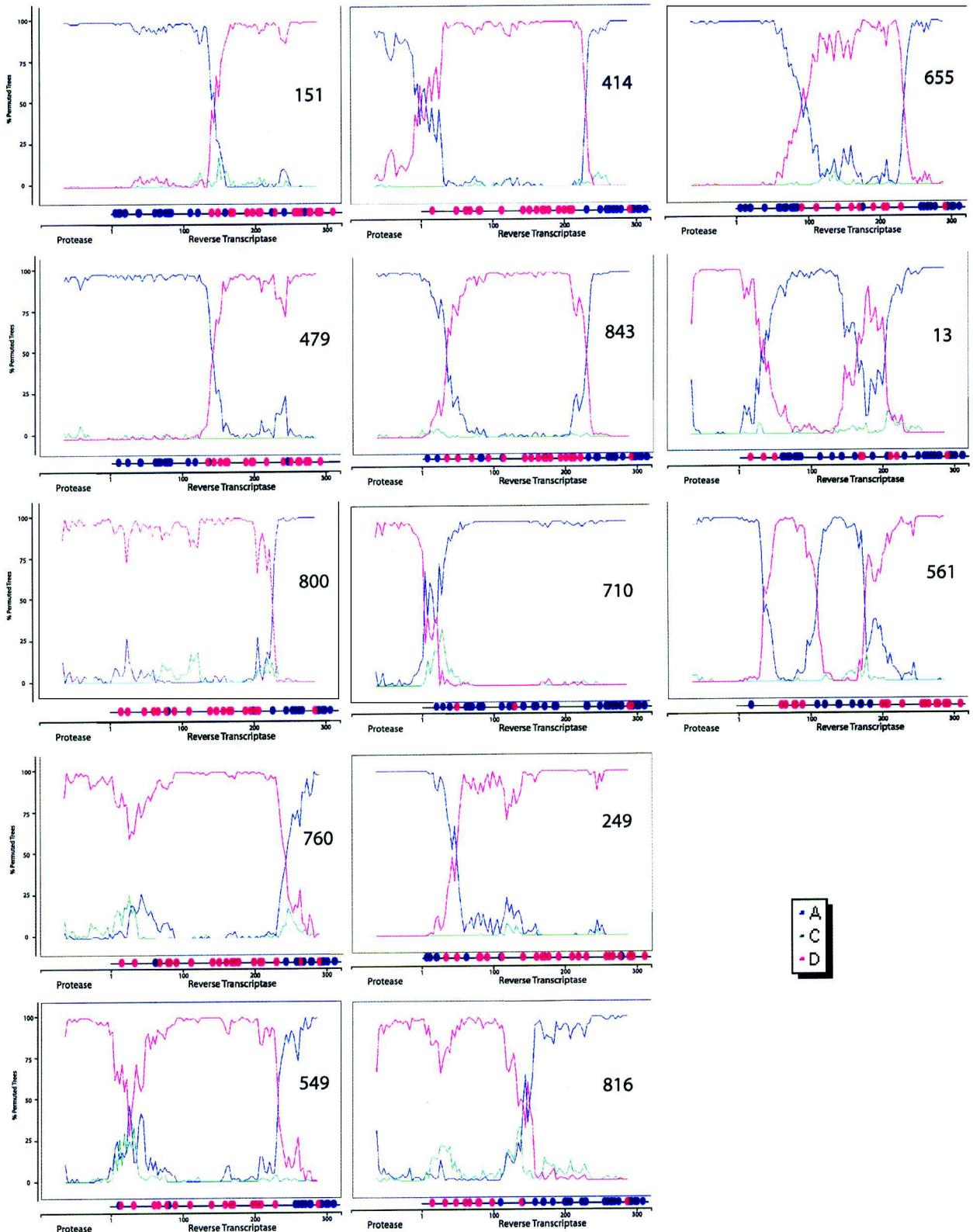


FIG. 1. Bootscanning results for 13 HIV-1 viruses with strong evidence of recombination. The percentage of permuted trees (y axis) clustering with each subtype is shown along the length of the sequence (x axis). Bootscan results were originally obtained by using reference sequences of each of the nine pure subtypes and CRFs 1 and 2. The bootscan analysis was redone, using the two subtypes involved in recombination, subtypes A (blue) and D (pink), as well as one control subtype, subtype C (green). Isolates sharing similar patterns of recombination (151/479, 414/843, and 549/760/800) are grouped. The informative positions for each sequence and their assignments to either subtype A or subtype D are shown beneath the x axis.

pendently assess the classification of sequence segments by SimPlot. The informative positions for each sequence and their assignments to either subtype A or subtype D are shown beneath the ordinate axis for each plot in Fig. 1. For a median of 31 (91%) of the 34 informative positions per sequence, the assignment of the informative sites was consistent with SimPlot bootscanning results. The density of informative sites was insufficient to further refine the breakpoints as determined by the SimPlot analysis.

STABILITY OF RECOMBINANTS

The sequences described above were obtained from plasma samples collected from women 6–8 weeks after administration of NVP for prevention of HIV-1 mother-to-child transmission. Samples collected before NVP administration were available for 9 of the 13 women who had HIV-1 isolates with strong evidence of recombination. HIV-1 sequences from each of the nine isolates had the same evidence of recombination observed at the later time point. Five of the 13 women with recombinant HIV-1 viruses transmitted their infection perinatally despite NVP prophylaxis. Samples from these infected infants were found to have nearly identical sequences with similar breakpoints.

IMPLICATIONS OF RECOMBINATION

HIV-1, like other retroviruses, is recombinogenic as a consequence of its dimeric genome and an RT that can switch between templates during proviral DNA synthesis. Hybrid genomes can be generated when two parental genomes are co-packaged within the same cell. A review has described 11 CRFs, defined as completely sequenced recombinant isolates infecting at least two epidemiologically unrelated individuals. Two of these CRFs (CRF01_AE and CRF02_AG) are highly prevalent and play a major role in the HIV-1 pandemic.

CRFs are most likely to occur in Africa, where there are multiple circulating subtypes, high rates of infection and, in some cases, dual infection. HIV-1 subtypes A and D are common, making it likely that some individuals will be dually infected and at risk for developing recombinant viruses. Systemic sampling of HIV-1 from 102 epidemiologically unrelated Ugandan women provides a minimum estimate of intersubtype recombination in the population studied. Although 13% of the women had strong evidence of being infected with recombinant viruses, the actual prevalence that would be revealed by complete genome sequencing of the isolates may in fact be higher.

The fact that nine of the women had the same recombinant identified at two points in time, and that five of the women transmitted a recombinant virus perinatally, confirms that these recombinant isolates do not represent polymerase chain reaction (PCR) artifacts and also demonstrates the stability of these variants *in vivo*. The fact that three different recombinant forms were present in at least two women suggests that these recombinants may be common variants in parts of Uganda. These variants, however, cannot be considered CRFs because their whole genomes were not sequenced.

Although previous studies have shown that hybrid *pol* proteins can be formed, the multiple patterns of recombination ob-

served in the 13 isolates in this study suggest that many parts of the RT gene are possible sites for recombination. Our analysis also highlights some of the difficulties faced by the programs used to detect recombination within *pol*. The *pol* gene is less variable than *env* and parts of *gag* and therefore has fewer phylogenetically informative sites. Applying such programs as RIP and SimPlot without examining the bootstrap support for putative sites of recombination often shows weak evidence for recombination in isolates that are probably not recombinant. Strong evidence of recombination in *pol* should therefore require using a larger window size (e.g., 400 bp) and should be supported by high bootscan support (90–100%) in regions of recombination.

Intersubtype recombination provides a powerful mechanism for adaptive HIV-1 evolution. Such recombination within the *pol* gene may increase the rapidity with which multidrug-resistant HIV-1 isolates are generated. The use of antiretroviral drug therapy in regions with a high prevalence of HIV-1 infection should be coupled with efforts to prevent ongoing HIV-1 transmissions to reduce the risk of dual infection and the generation of new recombinant forms. Intersubtype recombination also complicates efforts aimed at classifying sequenced isolates, particularly in the more conserved regions of the HIV-1 genome such as those encoding the molecular targets of HIV-1 therapy.

SEQUENCE DATA

GenBank accession numbers for the 102 sequences from women 6–8 weeks after NVP administration are AF188065–AF188166. GenBank accession numbers for the 13 recombinant sequences are as follows: 151, AF388069; 414, AF388141; 655, AF388110; 479, AF388086; 843, AF388131; 561, AF388093; 549, AF388149; 710, AF388156; 13, AF188135; 760, AF388123; 816, AF388160; 800, AF388128; 249, AF388073.

ACKNOWLEDGMENTS

The authors thank Drs. Francis Mmiro and Philippa Musoke (Makerere University, Kampala, Uganda) for providing the samples used for analysis, and thank Estelle Piwovar-Manning, Constance Ducar, and the laboratory staff in Uganda for assistance with sample processing. The authors acknowledge the assistance of Melissa Allen (Protocol Specialist, Family Health International), and also thank Eric Shulse and the Applied Biosystems Genotyping Team for providing reagents used in this study.

Sources of support: This work was supported by (1) the Elizabeth Glaser Pediatric AIDS Foundation, (2) the HIV Network for Prevention Trials (HIVNET) and sponsored by the U.S. National Institutes of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Department of Health and Human Services (DHHS), through contract N01-AI-35173 with Family Health International, contract N01-AI-45200 with Fred Hutchinson Cancer Research Center, and subcontracts with JHU/Makerere University NOI-AI-35173-417, (3) the HIV Prevention Trials Network (HPTN) sponsored by the NIAID, National Institutes of Child Health and Human Development (NICHD), National Institute on Drug Abuse, National Institute of Mental Health, and Office of AIDS Research, of the

NIH, DHHS (U01-AI-46745 and U01-AI-48054), (4) the Pediatric and Adult AIDS Clinical Trials Group (NIH, Division of AIDS, NIAID), and (5) R29 34348 (NIH, Division of CH/HD). Reagents for HIV genotyping were provided by Applied Biosystems (Foster City, CA).

REFERENCES

1. Peeters M and Sharp PM: Genetic diversity of HIV-1: The moving target. *AIDS* 2000;14:S129–S140.
2. Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, Funkouser RK, Gao F, Hahn BH, Kalish ML, Kuiken C, Learn GH, Leitner T, McCutchan F, Osmanov S, Peeters M, Pieniazek D, Salminen M, Sharp PM, Wolinsky S, and Korber B: HIV-1 nomenclature proposal: A reference guide to HIV-1 classification. In: *Human Retroviruses and AIDS: A compilation and Analysis of Nucleic and Amino Acid Sequences* (C.L. Kuiken *et al.*, eds.). Los Alamos National Laboratory, Los Alamos, New Mexico, 1999, pp. 492–505.
3. Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, Funkouser RK, Gao F, Hahn BH, Kalish ML, Kuiken C, Learn GH, Leitner T, McCutchan F, Osmanov S, Peeters M, Pieniazek D, Salminen M, Sharp PM, Wolinsky S, and Korber B: HIV-1 nomenclature proposal. *Science* 2000;288:55–65.
4. Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A, Hahn BH, Wolinsky S, and Bhattacharya T: Timing the ancestor of the HIV-1 pandemic strains. *Science* 2000;288:1789–1796.
5. Rayfield MA, Downing RG, Baggs J, Hu DJ, Pieniazek D, Luo CC, Biryahwaho B, Otten RA, Sempala SDK, and Dondero TJ: A molecular epidemiologic survey of HIV in Uganda. HIV Variant Working Group. *AIDS* 1998;12:521–527.
6. Hu DJ, Baggs J, Downing RG, Pieniazek D, Dorn J, Fridlund C, Biryahwaho B, Sempala SD, Rayfield MA, Dondero TJ, and Lal R: Predominance of HIV-1 subtype A and D infections in Uganda. *Emerg Infect Dis* 2000;6:609–615.
7. Becker-Pergola G, Kataaha P, Johnston-Dow L, Fung S, Jackson JB, and Eshleman SH: Analysis of HIV-1 protease and reverse transcriptase in antiretroviral drug naive Ugandan adults. *AIDS Res Hum Retroviruses* 2000;16:807–813.
8. Becker-Pergola G, Mellquist JL, Guay L, Mmiro F, Ndugwa C, Kataaha P, Jackson JB, and Eshleman SH: Identification of diverse HIV-1 subtypes and dual HIV-1 infection in pregnant Ugandan women. *AIDS Res Hum Retroviruses* 2000;16:1099–1104.
9. Mellquist JL, Becker-Pergola G, Gu J, Guay L, Himes L, Kataaha P, Mmiro F, Ndugwa C, Jackson JB, and Eshleman SH: Dual transmission of subtype A and D HIV-1 viruses from a Ugandan woman to her infant. *AIDS Res Hum Retroviruses* 1999;15:217–221.
10. Eshleman SH, Becker-Pergola G, Deseyve M, Guay LA, Mracna M, Fleming T, Cunningham S, Musoke P, Mmiro F, and Jackson JB: Impact of HIV-1 subtype on women receiving single dose NVP prophylaxis to prevent HIV-1 vertical transmission (HIVNET 012). *J Infect Dis* 2001;184:914–917.
11. Guay LA, Musoke P, Fleming T, Bagenda D, Allen M, Nakabiito C, Sherman J, Bakaki P, Ducar C, Deseyve M, Emel L, Mirochnick M, Fowler MG, Mofenson L, Miotti P, Dransfield K, Bray D, Mmiro F, and Jackson JB: Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-infant transmission of HIV-1 in Kampala, Uganda: HIVNET-012 randomised trial. *Lancet* 1999;354:795–802.
12. Jackson JB, Musoke P, Fleming T, Guay L, Bagenda D, Allen M, Nakabiito C, Sherman J, Bakaki P, Owor M, Ducar C, Deseyve M, Emel L, Mirochnick M, Fowler MG, Mofenson L, Miotti P, Gigliotti M, Bray D, and Mmiro F: Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-child transmission of HIV-1 in Kampala, Uganda: 18 months follow-up of the HIVNET 012 randomised trial. *Lancet* 2002 (in press).
13. Eshleman SH, Mracna M, Guay LA, Deseyve M, Cunningham SC, Mirochnick M, Musoke P, Fleming T, Fowler MG, Mofenson LM, Mmiro F, and Jackson JB: Selection and fading of resistance mutations in women and infants receiving nevirapine to prevent HIV-1 vertical transmission (HIVNET 012). *AIDS* 2001;15:1951–1957.
14. Siepel AC, Halpern AL, Macken C, and Korber BT: A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences. *AIDS Res Hum Retroviruses* 1995;11:1413–1416.
15. Salminen MO, Carr JK, Burke DS, and McCutchan FE: Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res Hum Retroviruses* 1995;11:1423–1425.

Address reprint requests to:

Susan H. Eshleman
 Department of Pathology
 The Johns Hopkins Medical Institutions
 Ross Building 646
 720 Rutland Avenue
 Baltimore, Maryland 21205

E-mail: seshlem@jhmi.edu