AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 27, Number 3, 2011

DOI: 10.1089/aid.2010.0061



# Molecular Characterization of HIV Type 1 Among HIV-Infected Respondents in a Cohort Being Prepared for HIV Phase III Vaccine Clinical Trials, Western Kenya

Micah Oyaro,[1,2] John Mbithi,[1] Fred Oyugi,[2] Annette Laten,[3] Omu Anzala,[2] and Susan Engelbrecht[3]



## Abstract

Kenya is one of the sub-Saharan African countries affected by HIV-1 infection and AIDS. We investigated HIV-1 genetic diversity in 130 individuals from Busia, Bungoma, and Kakamega in western Kenya as part of an HIV-1 vaccine feasibility study in preparation for Phase III efficacy clinical trials. After RNA extraction the partial *gag* (484 bp) and *env* (1297 bp) regions were amplified and directly sequenced. Phylogenetic analysis was done using MEGA version 4 and recombinants were identified using the jpHMM tool and phylogenetic analysis. HIV-1 sequences were amplified from 122 of the 130 samples, 118 (90.8%) from the *gag* region and 78 (60 %) from the *env* region and 74 samples (56.9%) from both the *gag* and *env* regions. Of these sequenced on both regions, 51.4% were subtype A, 9.4% subtype D, 1.4% subtype C, 4.1% subtype G, and 33.7% were discordant and thus possible recombinants, including A1/C, A1/D, A1/A2, and A2/C. The jpHMM tool indicated a further two samples with CD and BD breakpoints within the *env* gene and one within the *gag* gene (A1C). An additional sample had an A1D breakpoint in the *gag* gene, but the envelope was not amplified. HIV-1 subtype diversity in western Kenya should be considered in vaccines designed for clinical trials in this region and this genetic diversity should be continuously monitored.

## Introduction

Sub-Saharan Africa remains the region most affected by the HIV/AIDS pandemic, but in several regions, including Kenya, the HIV-1 prevalence rate is stable or declining.[1] In spite of this decline, the adult HIV-1 prevalence rate in Kenya is 7.8% with 1.4 million people living with the virus.[2] One of the major features of HIV-1 is the extreme genetic variability that may be reflected in differences in biological characteristics that determine transmissibility, pathogenesis, and immunogenicity. Therefore the genetic variation of HIV-1 and its evolution in time have important implications for the control of the pandemic.

Analysis of the HIV-1 *env* genes of virus strains from different geographic regions reveals that HIV-1 can be divided into three main groups: M (major), O (outlier), and N (non-M, non-O). HIV-1 group M has been further subdivided into genetically equidistant clusters of HIV-1 *env* genes, comprising subtypes A–D, F–H, J, K, and at least 45 circulating recombinant forms (CRFs) and numerous unique recombinant forms (URFs), http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html. Recently a new human immunodeficiency virus, closely related to gorilla simian immunodeficiency virus (SIVgor), has been described and a fourth group, designated P, was proposed.[3]

Previous molecular epidemiological studies done in Kenya mostly used samples from Nairobi.[4–9] A few other studies also investigated southern,[10] northern,[11] and western Kenya,[12] as well as the coastal region.[13,14] All of these studies indicated that HIV-1 subtype A is the most common subtype in Kenya, but that subtypes C, D, G, and recombinant forms were also detected.[4–14]

The complexity of HIV-1 diversity creates major challenges for vaccine design and development strategies. More than one HIV-1 vaccine candidate based on HIV-1 subtype A has been evaluated in Kenya during Phase I and II clinical trials.[15] The western Kenya region, which is targeted for HIV-1 vaccine clinical trials, is close to Uganda, a neighboring country that is dominated by HIV-1 subtypes A and D in almost equal proportions.[16] Our aim was thus to investigate HIV-1 genetic diversity by amplification and sequencing of partial *gag* and *env* genes from 130 HIV-1-positive individuals from different

[1]Kenyatta University, Department of Biological Sciences, Nairobi, Kenya.
[2]Kenya Aids Vaccine Initiative (KAVI), Department of Medical Microbiology, University of Nairobi, Nairobi, Kenya.
[3]Division of Medical Virology, National Health Laboratory Services (NHLS) and University of Stellenbosch, Tygerberg, South Africa.

hospitals in western Kenya. This project was part of an HIV-1 vaccine feasibility study in preparation for Phase III efficacy clinical trials.

## Materials and Methods

### *Study population and sample collection*

Volunteers were recruited from the general population from January to August 2007. They were seeking Voluntary Counseling and Testing (VCT) services in the Busia and Bungoma District hospitals and the Kakamega Provincial Hospital in the western part of Kenya. Following the signing of an informed consent form and counseling, all eligible participants were interviewed according to a standardized questionnaire developed for this study. The questionnaire included demographic characteristics such as age, gender, marital status, nationality, sexual behavior, and occupation. The study was approved by the Kenyatta National Hospital and Research and Ethics Committee (KNHREC).

Eight milliliters of venous blood was collected in EDTA tubes and centrifuged to separate the cells and plasma, which were stored in separate vials. The first plasma aliquot was used to confirm HIV-1 infection using the Vironostika Uniform II kit (Organon Teknika, Boxtel, The Netherlands) according to the manufacturer's instructions. For molecular characterization, the other plasma aliquot was frozen at −70°C until used.

### *PCR amplification and sequencing of the partial gag and env genes*

HIV-1 RNA was extracted from the plasma using the QIAamp Viral RNA kit (Qiagen GmbH, Hilden, Germany) according to manufacturer's instructions and stored at −70°C. Proviral DNA from the buffy coat of two samples was extracted using the QIAamp DNA Blood Mini kit (Qiagen, GmbH, Hilden, Germany) according to the manufacturer's instructions.

Reverse transcriptase polymerase chain reaction (RT-PCR) amplification was performed on a 484-bp fragment of the *gag* gene (HXB2 nucleotides 1237–1721) using the Access-RT kit (Promega, Madison, WI) and methods previously described.[17]

PCR was also done for a 1372-bp *env* region (HXB2 nucleotides 7002–8374) that included the gp120 V3 region up to the gp41 immunodominant region. For prenested RT-PCR, the primers ED5 (5′-ATGGGATCAAAGCCTAAAGCCATGTG-3′)[18] and gp41R1[19] were used. Briefly, we used the Access RT-PCR kit (Promega, Madison, WI) with 5 μl of RNA, 200 μM of each nucleotide, 40 μmol of each primer, 1 mM $MgSO_4$, 5 U each of AMV RT and Tfl DNA polymerase, and AMV/Tfl buffer in a total volume of 50 μl. After reverse transcription of 45 min at 48°C, the reaction was held at 94°C for 2 min, followed by 40 cycles of denaturing the DNA (94°C; 30 s), annealing of primers (58°C; 30 s), and extension of the annealed primers (68°C; 2 min). This was followed by a final extension step of 7 min at 68°C and the PCR product was kept at 4°C. A nested PCR of 1297 bp (HXB2 nucleotides 7002–8299) was done with a Promega GoTaq Flexi kit (Promega, Madison, WI). Template DNA for the nested PCR consisted of 3 μl of reaction product from the first round with 200 μM of each nucleotide, 40 μmol of each primer, ES7[20] and M*env*19R,[17] 1.5 mM $MgCl_2$, 2.5 U of Taq polymerase, and GoTaq buffer made up in a total volume of 50 μl. The PCR cycle method used was similar to that of prenested PCR, except for the primer annealing step at 44°C. The PCR products were visualized using agarose gel electrophoresis.

PCR products were purified using *Exonuclease* (*Exo*l) and Shrimp alkaline phosphatase (SAP) (USB Corporation, Cleveland, OH) at 37°C for 15 min followed by heat inactivation of the enzymes at 80°C for 15 min. These purified products were kept at −20°C until sequencing reactions were done. All PCR products were sequenced on both strands using the BigDye Terminator V3.1 Cycle Sequencing kit and analyzed on an ABI Prism 3130xl automated DNA sequencer (Applied Biosystems, Foster City, CA).

### *Sequence and phylogenetic analysis*

Sequences were analyzed and the overlapping DNA fragments were assembled using Sequencher version 4.8 (Gene Codes Corporation, Ann Arbor, MI). Nucleotide sequences were translated into amino acid sequences and submitted to GenBank using Sequin v7.70 (http://www.ncbi.nlm.nih.gov). Multiple alignments were created with the 2008 HIV-1 reference sequences obtained from the LANL HIV Database (http://hiv-web.lanl.gov) using Clustal X version 2.0.[21] These alignments were checked manually and edited with Geneious Pro version 4.8.3.[22] Neighbor-joining phylogenetic trees were constructed with MEGA version 4[23] using the Kimura two-parameter.[24] The REGA HIV-1 subtyping Tool Version 2.0 (http://bioafrica.net/rega-genotype/html/subtypinghiv.html) and HIV-BLAST (http://www.hiv.lanl.gov/content/sequence/BASIC_BLAST/basic_blast.html) were used to compare subtyping results. To detect recombinants, we used the jumping profile Hidden Markov Model (jpHMM-HIV) tool (http://jphmm.gobics.de)[25,26] and the Recombinant Identification Program, RIP version 3 (http://www.hiv.lanl.gov/content/sequence/RIP/RIP.html). Using the predicted recombination breakpoints, sequence fragments were edited and trimmed with Geneious Pro version 4.8.3.[22] Sequences were then realigned using these fragments and neighbor-joining phylogenetic trees were constructed as before.

## Results

### *Patient demographics and clinical and epidemiological features*

The study group consisted of 130 volunteers and included 26% males ($n = 34$) and 74% females ($n = 96$). The mean age of the males was 36 years and that of the females was 32 years. None of the volunteers included in the study received antiretroviral therapy. The following numbering system was used: 2201, 3301, 4401, where "22" indicates samples from Busia, "33" indicating samples from Bungoma, and "44" indicates samples from Kakamega. Both Kakamega and Busia lie along a major highway to Uganda and all study participants were of Kenyan nationality, except for samples 2202, 2228, and 4436, which were of Ugandan nationality (Table 1).

### *PCR amplification and sequencing*

HIV-1 sequences were amplified and 118 (90.8%) were positive in the *gag* region and 78 (60 %) were positive in the *env* region. Only eight (6.2%) of the samples were negative

Table 1. Patient Information and Assigned HIV-1 Subtypes for Samples Obtained from Western Kenya

| *Sample ID* | *Gender* | *Age* | Gag *subtype* | Env *subtype* | Gag/env *recombinants* |
|---|---|---|---|---|---|
| 2201 | M | 45 | A1 | A1 | |
| 2202 | F | 18 | A1 | Neg | |
| 2203 | F | 25 | A2 | Neg | |
| 2204 | F | 22 | A1 | A1 | |
| 2205 | F | 25 | A1 | Neg | |
| 2206 | F | 39 | A1 | A1 | |
| 2207 | F | 36 | D | Neg | |
| 2208 | F | 40 | A1 | Neg | |
| 2209 | F | 27 | A1 | Neg | |
| 2210 | M | 27 | A1 | Neg | |
| 2211 | F | 21 | D | D | |
| 2212 | M | 21 | D | D | |
| 2213 | F | 24 | A1 D | Neg | A1 (a1-d)/neg |
| 2214 | F | 24 | D | A1 | D/A1 |
| 2215 | F | 18 | C | Neg | |
| 2216 | F | 20 | Neg | Neg | |
| 2217 | M | 30 | D | Neg | |
| 2218 | F | 43 | A1 | A1 | |
| 2219 | M | 56 | A1 C | A1 | A1 (a1-c)/A1 |
| 2220 | F | 38 | G | G | |
| 2221 | F | 25 | A1 | Neg | |
| 2222 | F | 35 | G | G | |
| 2223 | M | 26 | A1 | D | A1/D |
| 2224 | M | 35 | Neg | C | |
| 2225 | F | 33 | D | Neg | |
| 2226 | F | 42 | D | D | |
| 2227 | F | 20 | A1 | A1 | |
| 2228 | F | 24 | A1 | A1 | |
| 2229 | F | 30 | A1 | D | A1/D |
| 2230 | F | 34 | A1 | A1 | |
| 2231 | M | 43 | A1 | A1 | |
| 2233 | M | 43 | A1 | A1 | |
| 2234 | F | 29 | D | Neg | |
| 2235 | F | 23 | Neg | Neg | |
| 2237 | M | 52 | A1 | A1 | |
| 2238 | M | 34 | A1 | C | A1/C |
| 2239 | F | 30 | A1 | A1 | |
| 2240 | F | 39 | D | D | |
| 2241 | F | 22 | D | A1 (a1-d) | D/A1 (a1-d) |
| 2242 | F | 25 | A1 | A1 | |
| 2243 | M | 23 | A1 | A1 | |
| 2244 | F | 32 | D | D | |
| 2245 | F | 47 | A1 | A1 | |
| 2246 | F | 48 | A1 | Neg | |
| 2247 | M | 37 | A1 | A1 | |
| 2248 | M | 50 | A1 | A1 | |
| 2249 | M | 32 | A1 | D | A1/D |
| 2250 | M | 30 | A1 | D | A1/D |
| 2251 | M | 40 | A1 | A1 | |
| 2252 | M | 32 | D | D | |
| 2253 | F | 30 | D | A1 | D/A1 |
| 2254 | F | 54 | A1 | C | A1/C |
| 3302 | F | 27 | C | Neg | |
| 3303 | F | 36 | D | D | |
| 3304 | F | 39 | Neg | A1 | |
| 3306 | F | 36 | A1 | Neg | |
| 3307 | F | 27 | A1 | Neg | |
| 3308 | M | 36 | A1 | C (c-a1-c) | A1/C (c-a1-c) |
| 3309 | F | 48 | C | A1 | C/A1 |
| 3310 | F | 26 | A1 | Neg | |
| 3311 | F | 46 | A1 | A1 | |
| 3312 | F | 33 | A1 | Neg | |
| 3313 | F | 42 | A1 | A1 | |
| 3314 | F | 42 | Neg | Neg | |
| 3315 | M | 31 | D | A1 | D/A1 |

Table 1. Continued

| *Sample ID* | *Gender* | *Age* | Gag *subtype* | Env *subtype* | Gag/env *recombinants* |
|---|---|---|---|---|---|
| 3316 | F | 49 | A1 | D | A1/D |
| 3317 | F | 22 | A2 | C | A2/C |
| 3318 | F | 40 | A1 | A1 | |
| 3319 | M | 38 | A1 | Neg | |
| 3321 | F | 29 | Neg | Neg | |
| 3322 | F | 35 | A1 | A1 | |
| 3323 | F | 29 | Neg | A1 | |
| 3324 | M | 31 | G | G | |
| 3325 | F | 25 | D | Neg | |
| 3326 | F | 24 | A1 | A1 | |
| 3327 | F | 31 | A1 | A1 | |
| 3328 | M | 33 | A1 | D (d-g-d) | A1/D (d-g-d) |
| 3329 | M | 62 | A1 | A1 | |
| 3330 | F | 39 | A1 | Neg | |
| 3331 | F | 27 | A1 | A2 (a2-a1) | A1/A2 (a2-a1) |
| 3332 | F | 45 | A1 | A1 | |
| 3333 | F | 48 | A1 | Neg | |
| 3335 | F | 32 | Neg | A1 | |
| 3336 | F | 36 | A1 | Neg | |
| 3337 | F | 25 | A1 | Neg | |
| 3338 | M | 50 | A1 | Neg | |
| 3339 | M | 67 | D | A1 | D/A1 |
| 3340 | M | 39 | A1 | Neg | |
| 4401 | F | 50 | A1 | A1 | |
| 4402 | M | 34 | A1 | A1 | |
| 4403 | M | 34 | Neg | Neg | |
| 4404 | F | 28 | A1 | Neg | |
| 4405 | F | 57 | A1 | Neg | |
| 4406 | F | 37 | A1 | Neg | |
| 4407 | F | 29 | D | D (d-b) | D/D (d-b) |
| 4408 | F | 37 | D | Neg | |
| 4409 | F | 31 | A1 | Neg | |
| 4410 | M | 37 | A1 | Neg | |
| 4411 | F | 22 | A1 | Neg | |
| 4414 | M | 34 | A1 | Neg | |
| 4415 | F | 38 | A1 | Neg | |
| 4416 | F | 26 | D | A1 | D/A1 |
| 4417 | F | 44 | A1 | A1 | |
| 4418 | F | 27 | D | D (c-d) | D/D (c-d) |
| 4419 | F | 23 | A1 | Neg | |
| 4420 | F | 26 | C | C | |
| 4421 | F | 19 | A1 | A1 | |
| 4422 | M | 42 | A1 | A1 | |
| 4423 | M | 50 | Neg | Neg | |
| 4424 | F | 26 | A1 | A1 | |
| 4425 | F | 49 | C | A1 | C/A1 |
| 4426 | F | 32 | A1 | Neg | |
| 4427 | F | 30 | A1 | A1 | |
| 4429 | F | 37 | A1 | Neg | |
| 4430 | F | 21 | A1 | A1 | |
| 4431 | F | 48 | A1 | Neg | |
| 4432 | F | 29 | A1 | A1 | |
| 4433 | F | 31 | A1 | A1 | |
| 4434 | F | 23 | D | A1 | D/A1 |
| 4435 | F | 29 | C | A1 | C/A1 |
| 4436 | F | 18 | A1 | C (a1-c) | A1/C (a1-c) |
| 4437 | F | 41 | A1 | A1 | |
| 4439 | F | 31 | A1 | A1 | |
| 4440 | F | 40 | A1 | Neg | |
| 4441 | M | 37 | A1 | A1 | |
| 4442 | F | 26 | A1 | Neg | |
| 4445 | F | 34 | D | Neg | |
| 4447 | F | 40 | A1 | Neg | |
| 4449 | F | 33 | Neg | Neg | |
| 4450 | F | 28 | Neg | Neg | |

Table 2. HIV-1 Subtypes Assigned for the *gag* and *env* Regions, Circulating in the Busia, Bungoma, and Kakamega Regions of Western Kenya

| *Region* | *Busia* | *Bungoma* | *Kakamega* | *Total* |
|---|---|---|---|---|
| *gag* | | | | |
| Subtype A1 | 30 (57.7%) | 23 (63.9%) | 29 (69%) | 82 (69.5%) |
| Subtype A2 | 1 (1.9%) | 1 (2.8%) | 0 (0%) | 2 (1.7%) |
| Subtype C | 1 (1.9%) | 2 (5.6%) | 3 (7.1%) | 6 (5.1%) |
| Subtype D | 13 (25.%) | 4 (11.1%) | 6 (14.3%) | 23 (19.5%) |
| Subtype G | 2 (3.8%) | 1 (2.8%) | 0 (0%) | 3 (2.5%) |
| Recombinant | 2 (3.8%) | 0 (0%) | 0 (0%) | 2 (1.7%) |
| Total | 49 (100%) | 31 (100%) | 38 (100%) | 118 (100%) |
| *env* | | | | |
| Subtype A1 | 20 (38.5%) | 14 (38.9%) | 17 (40.5%) | 51 (65.4%) |
| Subtype A2 | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) |
| Subtype C | 3 (5.8%) | 1 (2.8%) | 1 (2.4%) | 5 (6.4%) |
| Subtype D | 10 (19.2%) | 2 (5.6%) | 0 (0%) | 12 (15.4%) |
| Subtype G | 2 (3.8%) | 1 (2.8%) | 0 (0%) | 3 (3.8%) |
| Recombinant | 1 (1.9%) | 3 (8.3%) | 3 (7.1%) | 7 (9.0%) |
| Total | 36 (100 %) | 21 (100%) | 21 (100%) | 78 (100%) |

with both *gag* and *env* PCR. Four samples were *gag* PCR negative, but *env* PCR positive and 44 samples were negative for *env* PCR but positive for *gag* PCR. The *gag* region is more conserved than the *env* region and genetic variation might explain why only 60% of the samples could be amplified with the *env* PCR primer set. In summary, from 130 samples only 74 (56.9%) were successfully sequenced and characterized on both the *gag* and *env* regions. Of these 74 samples, 38 (51.4%) sequences were subtype A, seven (9.4%) sequences were subtype D, one (1.4%) sequence was subtype C, three (4.1%) sequences were subtype G, and 25 (33.7%) sequences were discordant and thus possible recombinants. A summary of all the identified subtypes is indicated in Table 2.

### *Phylogenetic analysis and subtyping the gag gene*

The result of the phylogenetic analysis of the partial *gag* gene is shown in Fig 1. The sequences did not cluster by collection site but were intermixed throughout the tree, indicating multiple introductions into western Kenya. The majority of the samples clustered with HIV-1 subtype A1. Two sequences (samples 2203 and 3317) clustered with subtype A2, three sequences clustered with subtype G (samples 2220, 2222, and 3324), and five sequences (samples 2215, 3302, 3309, 4420, and 4425) clustered with subtype C. The sequence of

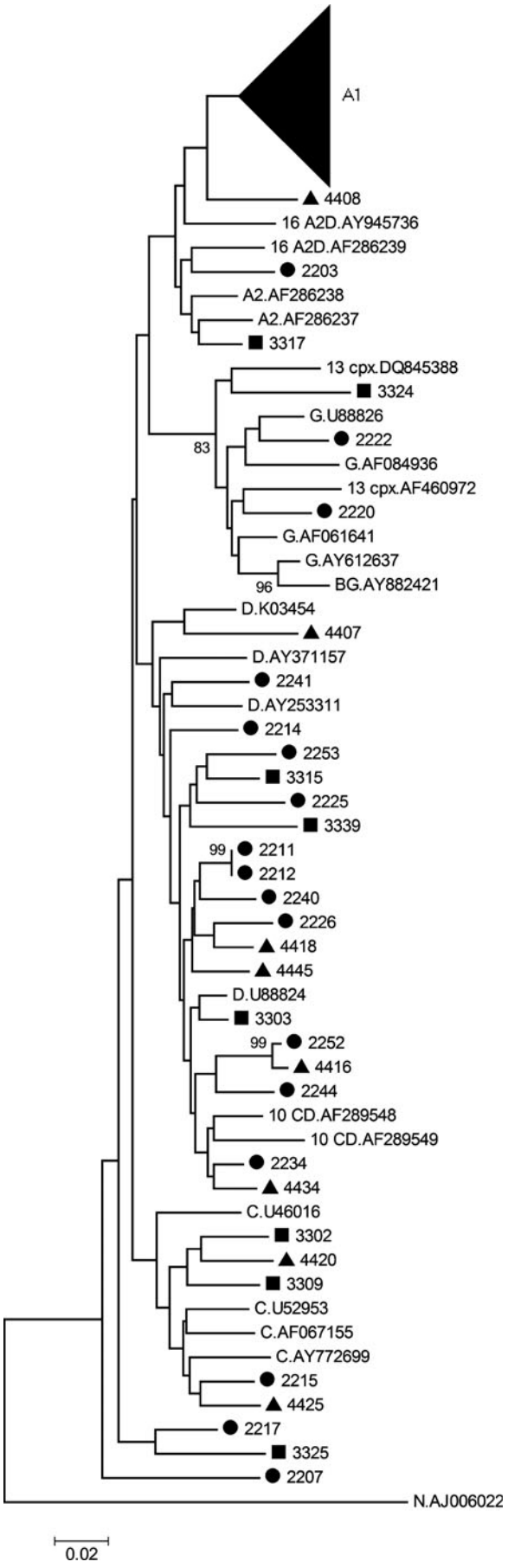


**FIG. 1.** Phylogenetic tree of the HIV-1 *gag* region spanning nucleotides 1237–1721 (HXB2 coordinates). Patient samples from Busia (●), Bungoma (■), and Kakamega (▲) in Kenya were aligned and compared with reference sequences from the Los Alamos HIV database using Clustal X version 2.0[21] and manually checked and edited with Geneious Pro version 4.8.3.[22] Reference sequences are indicated by the subtype and GenBank accession number. The neighbor-joining tree was constructed using the Kimura two-parameter algorithm[24] in MEGA version 4.[23] The bootstrap values of 1000 replicates above 70% are indicated next to the node and the scale bar at the bottom represents the number of base substitutions per site. All positions containing gaps and missing data were eliminated from the dataset.

sample 4435 was too short to include in *gag* phylogenetic analysis, but with blast analysis and the jpHMM HIV tool it was subtyped as C. Nineteen samples were grouped with subtype D. The sequences from samples 2207, 2217, 3325, and 4408 were outliers in the phylogenetic tree and did not cluster with any subtype, but the jpHMM tool indicated that these sequences were subtype D. Although discrepancies between online tools for assigning subtyping/recombinant forms were observed, this could mostly be resolved. The jpHMM analysis was used to indicate the breakpoints of the recombinants and these partial sequences were then used for phylogenetic analysis.

Final analysis of the 118 *gag* sequences indicated that 82 (69.5%) were subtype A1, two (1.7%) were subtype A2, six (5.1%) were subtype C, 23 (19.5%) were subtype D, 3 (2.5%) were subtype G, and two (1.7%) were possible recombinants (Table 1). These possible *gag* recombinant sequences from samples 2213 and 2219 need to be further analyzed. Phylogenetic analysis of the *gag* gene also indicated that sequences from samples 2211 and 2212 clustered closely together with a bootstrap value of 98%. These samples were epidemiologically linked.

### *Phylogenetic analysis and subtyping the env gene*

The complete *env* gene PCR fragment (1297 bp) could not be sequenced for all the samples. Some of the sequence chromatograms were difficult to read due to multiple peaks, specifically in the gp120 variable regions. Thus we had *env* sequences of almost 1297 bp ($n = 52$) and shorter sequences of about 600 bp ($n = 26$). The results of the phylogenetic analysis of the partial *env* gene are shown in Fig. 2.

In these *env* phylogenetic trees, the majority of the sequences clustered with subtype A1. Seven sequences (samples 2224, 2238, 2254, 3308, 3317, 4420, and 4436) clustered with subtype C, three sequences (samples 2220, 2222, and 3324) clustered with subtype G, and sequences from 15 samples clustered with subtype D. The sequence from sample 2241 (Fig. 2a) was typed as A1, but had a long branch. The jpHMM tool indicated that this sequence is an A1D recombinant (Table 2) and this was confirmed with phylogenetic analysis of the subfragments. Another sequence (sample 3331) clustered with subtype A2 (Fig. 2b), with a very long branch and the jpHMM tool also indicated recombination in this sequence. Final analysis of the 78 *env* sequences indicated that 51 (65.4%) sequences were subtype A1, five sequences (6.4%) were subtype C, 12 sequences (15.4%) were subtype D, three sequences (3.8%) were subtype G, and seven sequences (9.0%) were recombinants (Tables 1 and 2).

### *Recombinant analysis using jpHMM*

Recombinants were further evaluated with the jpHMM-HIV tool, a probabilistic generalization of the jumping-alignment approach. Because recombination breakpoints identified by jpHMM were found to be significantly more accurate than breakpoints defined by traditional methods based on comparing single representative sequences,[25,26] we used this approach to identify recombinants. Using this tool and phylogenetic analysis (data not shown), we identified nine unique recombinants (Table 3) with breakpoints within the *env* and *gag* genes. Sequences 2219, 3308, and 4436 were A1C recombinants, 2213 and 2241 were A1D recombinants, 3328 was a DG recombinant, 4418 was a CD recombinant, 3331 was an A1A2 recombinant, and 4407 was a BD recombinant. Bootstrap support for the BD recombinant was low.

Phylogenetic and recombinant analysis indicated that the 74 *gag/env* subtypes included 38 (51.4%) A1/A1, seven (9.4%) D/D, three (4.1%) G/G, four (5.4%) A1/C, six (8.1%) A1/D, three (4.1%) C/A1, seven (9.5%) D/A1, and one (1.4%) each of C/C, A1/A2, and A2/C. The jpHMM tool indicated a further two samples with breakpoints within the *env* gene (4418 and 4407) and another one within the *gag* gene (2219). An additional sample, 2213, had a breakpoint in the *gag* gene, but the envelope was not amplified (Table 3).

## Discussion

We investigated HIV-1 genetic diversity in 130 HIV-1-positive individuals from western Kenya in preparation for Phase III vaccine trials. In this study we detected subtypes A (A1 and A2), C, D, G, and recombinants, including recombinants with breakpoints in the *gag* ($n = 2$) and *env* ($n = 7$) regions. The results found in this study are consistent with the previous studies.

Kenya is bordered by five countries (Tanzania, Uganda, Sudan, Ethiopia, and Somalia) and the distribution of HIV-1 subtypes in these countries is variable. For instance, subtype C and AC recombinants dominate in Somalia and Ethiopia,[27] subtypes C and D are common in Sudan,[28] and subtypes A and D are common in Uganda.[16] In Tanzania the diversity includes subtypes A, C, and D, as well as AC, AD, and CRF10_CD recombinants (http://www.hiv.lanl.gov/content/index).

Although subtype A is the most prevalent subtype in Kenya, the HIV-1 subtype distribution can vary geographically. In Nairobi, Nielson and co-workers[5] detected subtypes A (70.3%), D (20.5%), C (6,9%), G (0.3%), and recombinants (2.2%) by analyzing the partial *env* gene of 320 samples. The Pumvani MTCT cohort in Nairobi was divided into subtypes A (58%), D (20%), and C (1%) by analyzing the partial *gag* and *protease* genes of 130 samples.[6] Analyzing the *integrase* gene of 140 samples from Nairobi, Lihana and co-workers[7] detected subtypes A (64%), D (17%), C (9%), and G (1%). By analyzing the complete genome in 10 samples from Nairobi, 50% of the samples were subtype A and 50% were recombinants.[29]

In the southern part of Kenya, 56% of samples analyzed were subtype A and 40% were recombinants when the complete HIV-1 genome of 41 samples was sequenced.[10] In northern Kenya the partial *env* region of 72 samples was analyzed and subtypes A (50%), C (39%), and D (11%) were detected. This region borders Ethiopia, which is dominated mainly by subtype C, and this study indicated that cross-border movement can influence the circulation of HIV-1 subtypes.[11] In a previous study done in the western part of Kenya on the partial *env* region of 30 samples, it was revealed that subtype A was the most prevalent (67%) strain, followed by subtypes D, C, and G. Twenty-three percent of the samples were recombinants (AD, AC, and CRF10_AD) and unclassified strains, indicating that western Kenya may be a hotspot for recombination.[30] A larger investigation of 460 samples from Kisumu in western Kenya sequenced partial *gag* and *env* regions and 344 samples (75%) were concordant in both regions (subtypes A, 59%; D, 10%; C, 2%; G, 3%) and 25% were discordant, indicating D/A (40, 8.7%), A/D (27, 5.9%), C/A

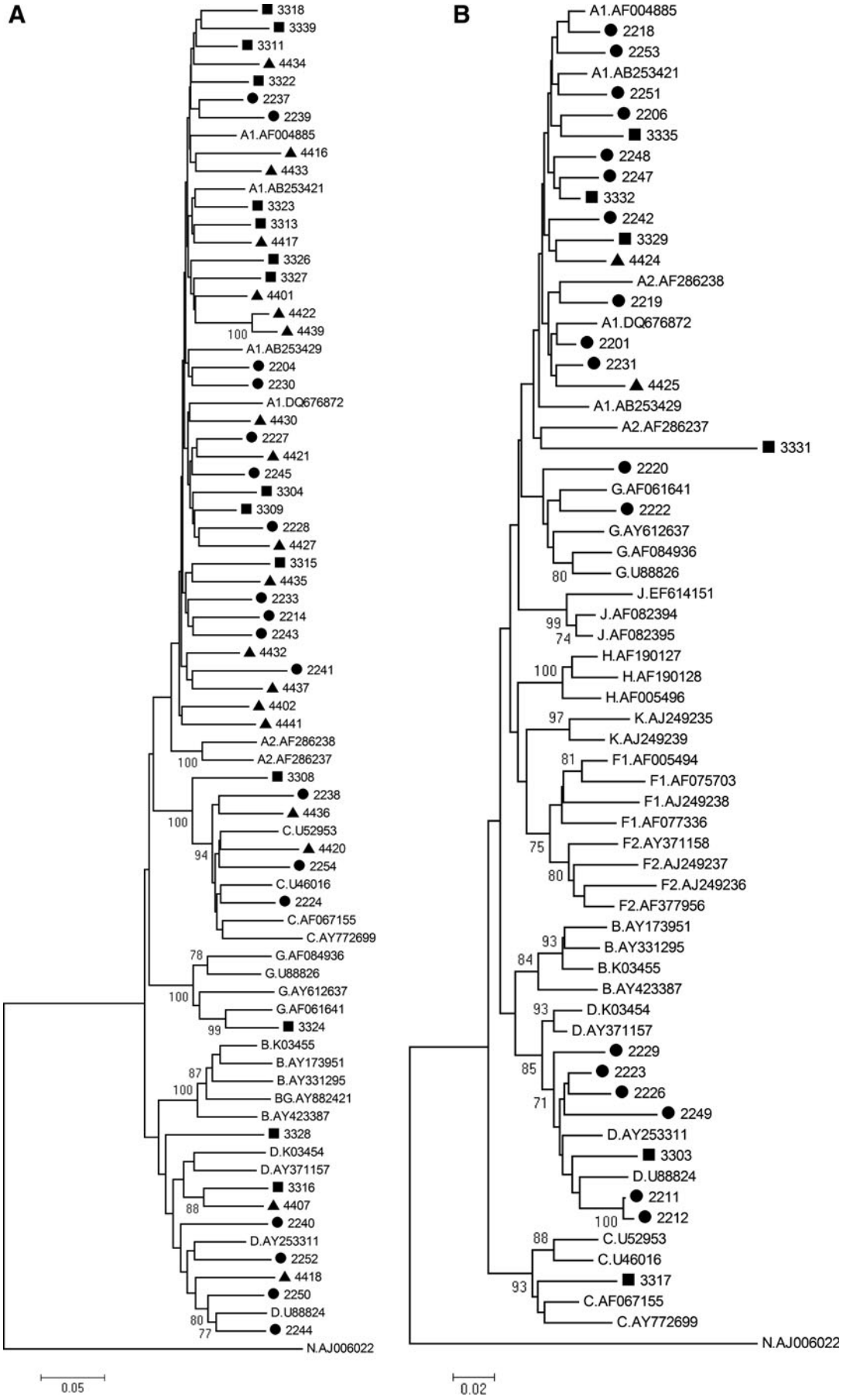

Table 3. Location of Breakpoints in Recombinant HIV-1 Strains from Kenya

| Strain | Fragment start position[a] | Uncertainty region[a] | Breakpoint interval[a] | Fragment end position[a] | Fragment subtype |
|---|---|---|---|---|---|
| 2213 | 1249 | | 1359–1371 | 1360 | D |
| | 1361 | | | 1704 | A1 |
| 2219 | 1249 | | 1593–1634 | 1611 | A1 |
| | 1612 | | | 1704 | C |
| 2241 | 7068 | | 7626–7717 | 7684 | A1 |
| | 7685 | | | 8285 | D |
| 3308 | 7035 | | 7641–7855 | 7823 | C |
| | 7824 | | 8115–8150 | 8122 | A1 |
| | 8123 | | | 8288 | C |
| 3328 | 7035 | | 7299–7315 | 7312 | D |
| | 7313 | | 7610–7673 | 7644 | G |
| | 7635 | | | 8288 | D |
| 3331 | 7652 | | 7896–7979 | 7931 | A2 |
| | 7932 | | | 8288 | A1 |
| 4407 | 7059 | 7947–8058 | | 8058 | D |
| | 8059 | 8059–8240 | | 8240 | B |
| 4418 | 7047 | | 7222–7255 | 7232 | C |
| | 7233 | | | 8288 | D |
| 4436 | 7035 | | 7215–7230 | 7220 | A1 |
| | 7221 | | | 8288 | C |

[a]Numbering according to reference strain HXB2.

(11, 2.4%), and A/C (8, 1.7%) recombinants.[12] Though all these molecular studies have been done in the same region, the data indicate that there are significant differences in the distribution of HIV pure subtypes and recombinant forms. However, most of the studies have obtained samples from women participants. These data may have been different if there was equal participation of both male and female participants.

In this study we sequenced 74 samples in both the *gag* and *env* regions. Concordant results indicated that subtype A was dominant (51.4%), followed by D (9.4%), C (1.4%), and G (4.1%). Twenty-two samples were discordant in the *gag* and *env* regions, indicating possible A1/C ($n=4$), A1/D ($n=6$), C/A1 ($n=3$), D/A1 ($n=7$), A1/A2 ($n=1$), and A2/C ($n=1$) recombinants. A further seven samples had breakpoints in the *env* region and another two in the *gag* region. Detection of recombination in 25 (33.7%) of the samples might underestimate recombination. The recombinants with breakpoints in the *gag* region (strain 2213, gag A1D and 2219, gag A1C) were both from Busia. The *env* sequence from 2219 was a subtype A1 and in 2213 we were unable to amplify the *env* region. The samples with breakpoints in the *env* region were A1C ($n=2$), and one each of A1/A2, A1D, DG, and DC.

Our study has shortcomings as well. We never confirmed our results with full-length sequencing, although similar studies confirmed a good correlation between sequence results and partial gene sequencing. Evidence of intersubtype genes on a single region as described above suggests that full-length sequencing may be a more inclusive technique to map out all recombinations and mutations and therefore offers a true picture of viral diversity, especially in Africa where many subtypes coexist.[31]

The genetic variability of HIV globally presents a major challenge for vaccine developers because immune responses that recognize HIV from one subtype may fail to recognize viruses from other subtypes. Previous HIV vaccine candidates that have been evaluated in Kenya were based on the HIV-1 subtype A, the dominant strain.[4,5,30] Two HIV vaccine candidates were developed in a partnership between the University of Nairobi's Kenya AIDS Vaccine Initiative (KAVI), the Medical Research Council, University of Oxford, and the International AIDS Vaccine Initiative (IAVI). With the evidence of superinfection and recombination,[32,33] a vaccine designed based on the dominant subtype may not offer protection against another subtype. HIV-1 subtype diversity in western Kenya should be considered in vaccines designed for clinical trials in this region and this genetic diversity should be continuously monitored.

The study demonstrates that four HIV-1 pure subtypes (A, C, D, and G) and a high proportion of recombinants of these subtypes are present in western Kenya. These data therefore suggest that a multiclade HIV-1 vaccine with antigenic determinants from all subtypes present in the region may be the best vaccine for future clinical trials in western Kenya. More studies are needed to monitor the molecular evolution of recombinants and the introduction of new viral strains ahead of efficacy clinical trials.

## Sequence Data

The sequences were deposited in GenBank with Accession numbers FJ346340–FJ346535.

## Acknowledgments

A part of the work has been partially presented at AIDS Vaccine 2008, Cape Town, 13–16 October 2008 (AIDS Research and Human Retroviruses 2008; volume 24, Supplement 1).

## Author Disclosure Statement

No competing financial interests exist.

## References

1. UNAIDS, 2009: *AIDS epidemic update 2009.* http://www.unaids.org/en/KnowledgeCentre/HIVData/EpiUpdate/EpiUpdArchive/2009/default.asp.
2. National AIDS Control Council, Office of the President, Kenya, 2008: *UNGASS 2008 Country Report for Kenya.*

**FIG. 2.** Phylogenetic tree of the HIV-1 **(a)** *env* region (1297 bp) and **(b)** *env* region (600 bp). Patient samples from Busia (●), Bungoma (■), and Kakamega (▲) in Kenya were aligned and compared with reference sequences from the Los Alamos HIV database using Clustal X version 2.0[21] and manually checked and edited with Geneious Pro version 4.8.3.[22] Reference sequences are indicated by the subtype and GenBank accession number. The neighbor-joining tree was constructed using the Kimura two-parameter algorithm[24] in MEGA version 4.[23] The bootstrap values of 1000 replicates above 70% are indicated next to the node and the scale bar at the bottom represents the number of base substitutions per site. All positions containing gaps and missing data were eliminated from the dataset.

NACC, Nairobi. http://data.unaids.org/pub/Report/2008/Kenya_2008_country_progress_report_en.pdf.
3. Plantier J-C, Leoz, M, Dickerson JE, *et al.*: A new human immunodeficiency virus derived from gorillas. Nat Med 2009;15:871–872.
4. Poss M, Gosink J, Thomas E, *et al.*: Phylogenetic evaluation of Kenyan HIV type 1 isolates. AIDS Res Hum Retroviruses 1997;13:493–499.
5. Neilson JR, John GC, Carr JK, *et al.*: Subtypes of human immunodeficiency virus type 1 and disease stage among women in Nairobi, Kenya. J Virol 1999;73:4393–4403.
6. Murray MC, Embree JE, Ramdahin SG, *et al.*: Effect of human immunodeficiency virus (HIV) type 1 viral genotype on mother-to-child transmission of HIV-1. J Infect Dis 2000; 181:746–749.
7. Lihana RW, Khamadi SA, Kiptoo MK, *et al.*: HIV type 1 subtypes among STI patients in Nairobi: A genotypic study based on partial pol gene sequencing. AIDS Res Hum Retroviruses 2006;22:1172–1177.
8. Khoja S, Ojwang P, Khan S, *et al.*: Genetic analysis of HIV-1 subtypes in Nairobi, Kenya. PLoS ONE 2008;3:e3191.
9. Lihana RW, Khamadi, S, Lwembe, RM, *et al.*: The changing trend of HIV-1 subtypes in Nairobi. AIDS Res Hum Retroviruses 2009;25:337–342.
10. Dowling WE, Kim B, Mason CJ, *et al.*: Forty-one near full-length HIV-1 sequences from Kenya reveal an epidemic of subtype A and A-containing recombinants. AIDS 2002;16: 1809–1820. Erratum in AIDS 2002;16:2104.
11. Khamadi SA, Ochieng W, Lihana RW, *et al.*: HIV type 1 subtypes in circulation in northern Kenya. AIDS Res Hum Retroviruses 2005;21:810–814.
12. Yang C, Li M, Shi YP, *et al.*: Genetic diversity and high proportion of intersubtype recombinants among HIV type 1-infected pregnant women in Kisumu, western Kenya. AIDS Res Hum Retroviruses 2004;20:565–574.
13. Khamadi S, Lihana RW, Osman S, *et al.*: Genetic diversity of HIV type 1 along the coastal strip of Kenya. AIDS Res Hum Retroviruses 2009;25:919–923.
14. Tovanabutra S, Sanders EJ, Graham SM, *et al.*: Evaluation of HIV type 1 strains in men having sex with men and in female sex workers in Mombasa, Kenya. AIDS Res Hum Retroviruses 2010;26:123–131.
15. Hanke T and McMichael AJ: Design and construction of an experimental HIV-1 vaccine for a year-2000 clinical trial in Kenya. Nat Med 2000;6:951–955.
16. Rayfield MA, Downing RG, Baggs J, *et al.*: A molecular epidemiologic survey of HIV in Uganda. HIV Variant Working Group. AIDS 1998;12:521–527.
17. Swanson P, Devare SG, and Hackett J: Molecular characterization of 39 HIV isolates representing group M (subtypes A-G) and group O: Sequence analysis of gag p24, *pol integrase,* and *env* gp41. AIDS Res Human Retroviruses 2003;19:625–629.
18. Delwart EL., Shpaer EG, Louwagie J, *et al.*: Genetic relationships determined by a DNA heteroduplex mobility assay: Analysis of HIV-1 env genes. Science 1993;262(5137):1257–1261.
19. Pieniazek D, Yang C, and Lal RB: Phylogenetic analysis of gp41 envelope of HIV-1 groups M, N, and O. In: *Human Retroviruses and AIDS 1998: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences* (Korber B, Kuiken CL, Foley B, *et al.*, Eds.). Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.
20. Bachmann MH, Delwart EL, Shpaer EG, *et al.*: Rapid genetic characterization of HIV-1 from four WHO-sponsored vaccine evaluation sites using a heteroduplex mobility assay. AIDS Res Hum Retroviruses 1994;10:1343–1351.
21. Larkin MA, Blackshields G, Brown NP, *et al.*: Clustal W and Clustal X version 2.0. Bioinformatics 2007;23:2947–2948.
22. Drummond AJ, Ashton B, Cheung M, *et al.*: (2009). Geneious v4.8, available from http://www.geneious.com.
23. Tamura K, Dudley J, Nei M, and Kumar S: *MEGA4*: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 2007;24:1596–1599.
24. Kimura M: A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 1980;16:111–120.
25. Zhang M, Schultz A-K, Calef C, *et al.*: jpHMM at GOBICS: A web server to detect genomic recombinations in HIV-1. Nucleic Acids Res 2006;34:W463–465.
26. Schultz AK, Zhang M, Leitner T, *et al.*: A jumping profile hidden Markov model and applications to recombination sites in HIV and HCV genomes. BMC Bioinformat 2006;7:265.
27. Abebe A, Kuiken CL, Goudsmit J, *et al.*: HIV type 1 subtype C in Addis Ababa, Ethiopia. AIDS Res Human Retroviruses 1997;13:1071–1075.
28. Hierholzer M, Graham RR, El Khidir I, *et al.*: HIV type 1 strains from east and west Africa are intermixed in Sudan. AIDS Res Human Retroviruses 2002;18:1163–1166.
29. Land AM, Ball TB, Luo M, *et al.*: Full-length HIV type 1 proviral sequencing of 10 highly exposed women from Nairobi, Kenya reveals a high proportion of intersubtype recombinants. AIDS Res Hum Retroviruses 2008;24:865–872.
30. Songok EM, Lihana RW, Kiptoo MK, *et al.*: Identification of env CRF-10 among HIV variants circulating in rural western Kenya. AIDS Res Hum Retroviruses 2003;19:161–165.
31. Louwagie JW, Janssens J, Mascola L, *et al.*: Genetic diversity of the envelope glycoprotein from human immunodeficiency virus type 1 isolates of African origin. J Virol 1995;69:263–271.
32. Pernas M, Casado C, Fuentes R, *et al.*: A dual superinfection and recombination within HIV-1 subtype B 12 years after primoinfection. J Acquir Immune Defic Syndr 2006;42:12–18.
33. Burke DS: Recombination in HIV: An important viral evolutionary strategy. Emerg Infect Dis 1997;3:253–259.

Address correspondence to:
*Susan Engelbrecht*
*Department of Medical Virology*
*University of Stellenbosch*
*P.O. Box 19063*
*Tygerberg 7505*
*South Africa*

*E-mail:* susanen@sun.ac.za