Genetic Diversity and High Proportion of Intersubtype Recombinants among HIV Type 1-Infected Pregnant Women in Kisumu, Western Kenya

CHUNFU YANG,1 MING LI,1 YA-PING SHI,2,3 JORN WINTER,1 ANNA M. VAN EIJK,3 JOHN AYISI,3 DALE J. HU,1 RICHARD STEKETEE,2 BERNARD L. NAHLEN,4 and RENU B. LAL1

ABSTRACT

The high genetic diversity of HIV-1 continues to complicate effective vaccine development. To better understand the extent of genetic diversity, intersubtype recombinants and their relative contribution to the HIV epidemic in Kenya, we undertook a detailed molecular epidemiological investigation on HIV-1-infected women attending an antenatal clinic in Kisumu, Kenya. Analysis of gag-p24 region from 460 specimens indicated that 310 (67.4%) were A, 94 (20.4%) were D, 28 (6.1%) were C, 9 (2.0%) were A2, 8 (1.7%) were G, and 11 (2.4%) were unclassifiable. Analysis of the env-gp41 region revealed that 326 (70.9%) were A, 85 (18.5%) D, 26 (5.7%) C, 9 (2.0%) each of A2 and G, 4 (0.9%) unclassifiable, and 1 (0.2%) CRF02_AG. Parallel analyses of the gag-p24 and env-gp41 regions indicated that 344 (74.8%) were concordant subtypes, while the remaining 116 (25.2%) were discordant subtypes. The most common discordant subtypes were D/A (40, 8.7%), A/D (27, 5.9%), C/A (11, 2.4%), and A/C (8, 1.7%). Further analysis of a 2.1-kb fragment spanning the gag-pol region from 38 selected specimens revealed that 19 were intersubtype recombinants and majority of them were unique recombinant forms. Distribution of concordant and discordant subtypes remained fairly stable over the 4-year period (1996–2000) studied. Comparison of amino acid sequences of gag-p24 and env-gp41 regions with the subtype A consensus sequence or Kenyan candidate vaccine antigen (HIVA) revealed minor variations in the immunodominant epitopes. These data provide further evidence of high genetic diversity, with subtype A as the predominant subtype and a high proportion of intersubtype recombinants in Kenya.

INTRODUCTION

HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1) group M consists of the great majority of HIV-1 viruses that dominate the global AIDS epidemic.1–3 The group M has been divided into nine distinct lineages, termed as subtypes (A–D, F–H, J, and K). More recently, genomes containing sequences derived from two or more subtypes and associated with different populations and geographic distributions have been found.4–6 The main causes of this high variability are the recombination of heterogeneous genomes by coinfection of cells, and the error-prone reverse transcriptase that can switch between templates during proviral synthesis.7,8 There are at least 15 circulating recombinant forms (CRFs) identified based on complete genome sequences derived from at least three epidemiologically unrelated individuals.6 Thus, genotypic analysis has provided a better understanding of the dynamics of viral spread and molecular epidemiology of HIV-1. The greatest genetic diversity of HIV-1 has been found in sub-Saharan Africa where all known HIV-1 subtypes and many of the CRF were identified.4,6 As in many other sub-Saharan countries, the HIV epidemic in Kenya is having a devastating effect on public health. Of Kenyans, 2.5 million were infected with HIV by the end of 2001 and HIV prevalence in the adult population was 15%.9 Molecular epidemiological studies have indicated the presence of diverse HIV-1 subtypes and unique recombinants.10–16 More recent studies have also documented the presence of subsubtype A2 and A2-containing recombi-

1Division of AIDS, STD, and TB Laboratory Research, National Center for HIV, STD and TB prevention and 2Division of Parasitic Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia 30333.
3Center for Vector Biology and Control Research, Kenya Medical Research Institute, Kisumu, Kenya.
4Roll-back Malaria Program, World Health Organization, Geneva, Switzerland.
nants, and many more unique recombinants.\textsuperscript{16–19} However, almost all the molecular epidemiological studies so far have been limited to urban settings where HIV-1 genetic diversity has been well documented in antenatal clinic attendees,\textsuperscript{10,11,13,16} commercial sex workers,\textsuperscript{12,14,15} and blood donors.\textsuperscript{18} There is very little information about the molecular epidemiology of HIV-1 in rural settings where HIV-1 prevalences are among the highest in the country.\textsuperscript{19,20}

Continuing molecular epidemiological studies have proved to be crucial in vaccine development. For example, molecular epidemiological studies in Thailand documented the independent introduction and spread of two different HIV-1 subtypes, B and CRF01\_AE, and the subsequent predominance of CRF01\_AE over time.\textsuperscript{21–23} This molecular epidemiological information played a key role in our understanding of the dynamics of the Thai epidemic and influenced the decision to introduce a bivalent subtype B/E vaccine in the first HIV-1 vaccine efficacy trial in Thailand.\textsuperscript{24} As HIV-1 vaccine trials are underway or are currently in the planning stage in Kenya,\textsuperscript{18,25} and HIV-1 viral evolution is a dynamic process, there is a need to continue monitoring the extent of HIV-1 genetic diversity in the infected population in different parts of the country.

**MATERIALS AND METHODS**

**Study site and blood sample collection**

Women attending an antenatal clinic in New Nyanza Provincial General Hospital, Kisumu, Kenya were enrolled in the study.\textsuperscript{26} The study protocol was approved in 1995 by the institutional review boards of the Kenya Medical Research Institute (KEMRI), the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, and the Academic Medical Center (AMC) at the University of Amsterdam, Amsterdam, The Netherlands; and the participating institutions reviewed the protocol annually. Routine use of zidovudine or NVP for treatment of HIV-1 infection was not the policy of the Kenyan Health Ministry during the study period. Detailed demographic and clinical information about these participants was presented elsewhere.\textsuperscript{26} Blood samples from all mothers were collected at delivery, and plasma and peripheral blood mononuclear cells (PBMCs) were separated, aliquoted, and stored at \(-70^\circ\text{C}\) until laboratory procedures were performed.

**HIV testing**

HIV testing of pregnant women was conducted using two consecutive rapid tests, an initial Serostrip HIV-1/2 test (Saliva Diagnostic Systems, Pte Ltd, Singapore), followed by a confirmatory Capillus HIV-1/HIV-2 test (Cambridge Diagnostics Laboratory, Rockville, MD) on all samples that tested positive by Serostrip. Western blot was performed on all discordant samples.

**Reverse transcriptase-polymerase chain reaction (RT-PCR) and sequencing**

RNA extracts from the Amplicor HIV-1 monitor test version 1.0 were used to amplify the \textit{gag}-p24 and \textit{env}-gp41 regions by RT-PCR. The primers and protocols were described in detail elsewhere.\textsuperscript{27,28} Purified nested PCR products were sequenced using BigDye terminator (Applied Biosystems, Foster City, CA) in a 377 DNA sequencer (Applied Biosystems) following the manufacturer’s protocols. For some specimens with concordant or discordant subtypes in the \textit{gag}-p24 and \textit{env}-gp41 regions, a 2.1-kb fragment (nt 1237–3370, HXB2) spanning the \textit{gag-pol} region was amplified and sequenced. Briefly, two sets of overlapping primers were used to amplify the 2.1-kb fragment. These include P24 #1 (5’ AGYCAAATTAYCCYATAGT, nt 1174–1193, HXB2) and DP11 (5’ CCATTCCCTGGCTTAATTTTAATGGTA, nt 2572–2598), and DP10 (5’ CAACTCCCTCTCAGAAGCGGAGCCG, nt 2198–2223) and RT-p24R1 (5’ TATTTCGTATATAGCTTTTGTAGGGTCA, nt 3506–3536). Primers for the nested PCR are p24

![FIG. 1. Distribution of subtypes and discordant subtypes among 460 women in Kisumu, Kenya. (A) Subtype distribution based on \textit{gag}-p24 (outer pie) and \textit{env}-gp41 (inner pie). A, A2, C, D, G, and U stand for subtypes and unclassifiable, and X represents other minor subtypes (< 1%). (B) Frequency of discordant subtypes (A/A, D/D, C/C, G/G, A2/A2) and discordant subtypes (D/A, A/D, D/C, U/D, C/A, A/C) and all other combinations in the \textit{gag}-p24 and \textit{env}-gp41 regions.](image-url)
#2 (AGRACYTTRAYGCATGGGT, nt 1237–1265) and DP17 (5′ CTAATGGGAAAATTTAAAGT, nt 2538–2557), and DP16 (5′ CCCTAAAATCACTCTTTGGCA, nt 2252–2272) and Rev7 (ATCCCTGGATAAATCTGACTTGCCA, nt 3345–3370). Protocols for amplification and sequencing were essentially the same as described previously, except that the PCR conditions were 94°C for 1 min, 50°C for 1 min, and 72°C for 2 min. All sequencing was done in both directions.

**FIG. 2.** Phylogenetic analysis of 34 gag–pol sequences (A, consensus nt = 1921 bp, 4 sequences with short length were omitted) and 38 env-gp41 sequences (B, nt = 360 bp) from Kenyan women by the neighbor-joining method using subtype reference strains A, A2, C, D, F, G, H, J, and CRFs (CRF01_AE, CRF02_AG), (boxed). The newly characterized sequences shown in black and underlined are URFs and in colors (A1, red; A2, pink; C, yellow; D, blue; G, green) are pure subtypes in either region. Schematic representation of genomic structures of the gag–pol and env-gp41 regions from all 38 women (C). Four are concordant subtypes (three D and one G), 15 are discordant subtypes (seven D/A, five A/D, one each G/D, A/C, and C/A), and 19 were URFs.
Phylogenetic analysis

We aligned the newly derived sequences, along with selected reference sequences representing all subtypes and relevant CRFs, using the CLUSTALW (1.74) multiple-sequence alignment program.\textsuperscript{29,30} After manual adjustments using BioEdit\textsuperscript{31} and stripping all the gap sites, phylogenetic analyses were carried out on 436-bp consensus sequences for gag-p24 and 360-bp consensus sequences for env-gp41 and neighbor-joining trees were constructed using the Phylib 3.5c package.\textsuperscript{32} The stability of the tree nodes was assessed by bootstrap analysis using 1000 replicates. Bootstrap values \( \geq 70\% \) were considered significant.\textsuperscript{33} Genetic distances were calculated with the Kimura’s two-parameter method.\textsuperscript{32}

To screen intersubtype recombinants, we first applied the recombinant identification program (RIP) (http://linker.lanl.gov/RIP/RIPsubmit.html) to the newly derived sequences.

![Image of genomic structures](image)

**FIG. 3.** Genomic structures of the 19 unique recombinant forms from Kisumu, Kenya (A). The subtype structures of the 19 recombinants were determined by RIP and bootscanning analyses as implemented in SimPlot\textsuperscript{34} and were confirmed by subsegment phylogenetic analyses. Subtype designation and bootstrap values \( \geq 70\% \) are indicated.\textsuperscript{33} (B) Phylogenetic analysis of the 16 newly derived recombinants (in red) with published recombinants from Kenya (in black\textsuperscript{18}) and Tanzania (in blue\textsuperscript{35}).
along with subtype consensus sequences. If a potential recombinant was identified, bootscanning analysis was implemented as in SimPlot to locate the breakpoints. After gap stripping of the alignment, we analyzed 500 replicates using a 400-bp window with a 20-bp increment. We repeated the bootscan analysis with only the parental subtypes plus one subtype in order to obtain a clear recombinant breakpoint. After breakpoint identification, each of the segments on the two sides of the breakpoint was subjected to individual phylogenetic analysis.

Identification of amino acid variations in the immunodominant regions

The newly derived nucleotide sequences were translated into amino acid sequences and amino acid sequences were aligned. The aligned amino acid sequences were compared with the subtype A consensus sequence from the HIV database and the Kenyan candidate vaccine antigen (HIVA) and epitope variations were identified.

RESULTS

HIV-1 Subtypes among pregnant women

From the 518 women included in this study, we were able to amplify gag-p24 from 466 samples and env-gp41 sequences from 468 samples; the remaining samples either had negative RT-PCR or had ambiguous sequences that could not be used for further analysis. For the purpose of this study, 460 samples with both gag-p24 and env-gp41 sequences were used for further analysis.

Analysis of the 460 gag-p24 sequences indicated 310 (67.4%) were subtype A, 94 (20.4%) were subtype D, 28 (6.1%) were subtype C, 9 (2.0%) were subtype A2, 8 (1.7%) were subtype G, and the remaining 11 (2.4%) were unclassifiable (U). A similar analysis of the env-gp41 region revealed that 326 (70.9%) were subtype A, 85 (18.5%) were subtype D, 26 (5.7%) were subtype C, 9 (2.0%) each were subtypes A2 and G, 4 (0.9%) were unclassifiable, and 1 (0.2%) was CRF02_AG (Fig. 1).

Parallel phylogenetic analyses in the gag-p24 and env-gp41 regions to identify discordant subtypes revealed that 344 (74.8%) specimens had the same subtype in both the gag-p24 and env-gp41 regions (270 subtype A, 48 subtype D, 14 subtype C, 7 subtype G, and 5 subtype A2), while the remaining 116 (25.2%) had discordant subtypes (40 D/A, 27 A/D, 11 C/A, 8 A/C, and 30 with other subtype combinations) (Fig. 1). Among all subtype combinations (cordnandt-discordants), A/A was by far the most common one, accounting for 58.7% of all women, followed by D/D (10.4%), D/A (8.7%), A/D (5.9%), C/C (3%), C/A (2.4%), A/C (1.7%), G/G (1.5%), U/D (1.3%), A2/A2 (1.1%), and D/C (0.87%); the remaining 20 women (4.3%) had minor discordant subtype combinations (Fig. 1B). These minor subtype combinations include three U/A, two each of A/U, A2/A, C/D, and U/A2, and one each of A/A2, A/G, A/AG, A2/D, A2/U, C/A2, D/G, D/U, and G/D.

Identification of unique recombinant forms

To determine whether the discordant subtypes with mosaic genomes in the gag-p24 and env-gp41 regions represent infections with two different subtypes or with intersubtype recombi-
binants, we sequenced a 2.1-kb fragment (nt 1237–3370, HXB2) spanning the gag–pol region from 33 specimens with discordant subtypes and five specimens with concordant subtypes in the gag-p24 and env-gp41 regions. Phylogenetic analyses were carried out for the entire gag–pol region as shown in Figure 2.

Analysis of five specimens with concordant subtypes revealed that one of them was an intersubtype recombinant (00KE561) comprising subtypes A, CRF01_AE, and U, whereas the remaining four had a single subtype in this region including three subtype D and one subtype G, with respective subtype D and G in the env-gp41 region (Fig. 2). While this limited analysis suggests that concordant subtypes in the gag-p24 and env-gp41 regions may mainly reflect a pure subtype, we cannot rule out the possibility that there may be potential recombinant breakpoints in other regions of the genome that were not studied here. Thus, our analysis represents an underestimation of true recombinants. Analysis of the 33 discordant specimens revealed that 15 (45%) had a single subtype within the gag–pol region, whereas 18 (55%) had crossover breakpoints identifiable within the gag–pol region (Fig. 2C). Of the 15 specimens with a single subtype in the gag–pol region and a different subtype in the env-gp41 region, seven were subtype D/A, five were A/D, and one each G/D, A/C, and C/A. Whether these discordant specimens represent infection with dual subtypes or infection with recombinants with breakpoints within the pol–env region remains to be determined.

A detailed analysis of the gag–pol region from the 19 recombinants revealed that all of them contained portions of the subtype A or D genome (Fig. 3A). Of the 19 recombinants described here, 15 represent URFs with no matches either to published recombinants in GenBank or to each other (Fig. 3A). There were seven AD recombinants (98KE335, 97KE004, 98KE324, 96KE011, 98KE234, 97KE115, and 97018M) and one each of DU (96KE139), CD (99KE570), BD (97KE100), AC (97KE486), A2D (99KE588), A2CD (97KE114), ADU (97KE051), and AEU (00KE561) (Fig. 3A).

Two pairs of the recombinants (AD, 00KE553 and 99KE591; ADG, 98KE456 and 99KE532) were found to have identical genomic breakpoint structures (Fig. 3A). The AD recombinants, 00KE553 and 99KE591, have their 5’ portion of the genome clustered with subtype D and the 3’ portion clustered with subtype A, with high bootstrap values, while the ADG recombinants, 98KE456 and 99KE532, have a more complex genomic structure. The 5’ portion of these recombinants is clustered with subtype D, followed by a large portion of subtype G, then with subtype A, and finally with subtype G again (Fig. 3A). While both sets of these mothers appear to be epidemiologically unlinked based on the data we have collected, there is a possibility that they may be part of an interconnected social network or may share the same partner. Thus, further epidemiological investigation is needed to clarify this.

We also carried out a comparative analysis of URFs identified in the present study with 21 published recombinants from Kenya18 and Tanzania.35 A 2130-bp region (nt 1240–3370, HXB2) from 16 specimens from the present study (98KE234, 97KE100, and 97KE051 were excluded from this analysis due to a smaller piece), 16 sequences from Kenya, and 5 from Tanzania revealed that no two specimens had identical structure, with the exception of two pairs presented here and one pair published before (KER2003 and KSM4001,18 Fig. 3B). These data strongly imply that these distinct URFs represent an independent origin, indicating ongoing generation of newly emerging recombinants in Kenya.

![FIG. 5. Amino acid sequence comparison of gag-p24 (A) and env-gp41 (B) regions with consensus sequences from the HIV database. (A) gag-p24 sequences were aligned and compared with the subtype A sequence from HIVA (a multiepitope vaccine construct derived from Kenyan sequences25). (B) env-gp41 sequences were aligned and compared with the subtype A consensus sequence from the HIV database.](image-url)
Temporal relationship of concordant and discordant subtypes over time

As infections with dual subtypes are the prerequisite for the generation of recombinants, we next examined the presence of these various mosaic viral genome combinations over a 4-year period, from mid-1996 to mid-2000. All subtype combinations were analyzed for mid-1996–1997 (n = 114), mid-1997–1998 (n = 145), mid-1998–1999 (n = 107), and mid-1999–2000 (n = 94). As shown in Figure 4, there appears to be a trend towards an increasing percentage of concordant subtype A/A (56–67%); however, this trend is statistically insignificant (p = 0.098). In addition, frequencies of major discordant subtypes, such as A/D, D/A were around 5–10% (Fig. 4). We also analyzed the relationship between age of the infected women and the type of HIV strains over the 4-year period. Although the mean age of the women infected with different strains was similar, there was a trend toward older women being more likely to be infected with the most common strain (A/A) (p = 0.007, data not shown). More importantly, the proportion of discordant subtypes was highest among the mothers between the ages of 14 and 18 (25/72; 34.7%) compared to mothers between the ages of 19 and 22 (39/168; 23%) or mothers between the ages of 23 and 39 (28/159; 17.6%), however, these differences were not statistically significant (p > 0.05).

Comparative analysis of major immunodominant epitopes within the gag-p24 and env-gp41

We next determined the amino acid sequence variations within the gag-p24 and env-gp41 regions. For gag-p24 analysis, 346 nucleotide sequences were translated into amino acid...
sequences and analyzed for each subtype using a subtype A sequence from the HIVA, a multiepitope candidate vaccine immunogen currently in trials in Kenya. Analysis revealed remarkable conservation, although we did identify a few regions with many variations (Fig. 5A). For instance, analysis revealed that positions 50, 83, 91, 110, 116, 120, 154, and 171 had several substitutions, although the majority of them were conservative changes. The gag-p24 region between aa 15–55, 108–205, and 216–226 contained the majority of CTL epitopes recognized by diverse class I MHCs, whereas T-helper epitopes were scattered throughout the entire gag-p24. Recently, a B-27-restricted escape mutant has been identified in epitope KK10 (aa 131–140), a strongly conserved epitope associated with good clinical outcome among HIV-1-infected persons. A mutation at anchor position p2 (R132T) resulted in the loss of control of viremia and rapid disease progression. While none of the mothers contained R132T, 10 contained R132K and 1 each contained R132Q and R132S. These substitutions at anchor position may impact peptide binding, as arginine at this position is required for peptide binding. Another gag-p24 epitope (2EEKAFSPVE) recognized by newly defined HLA-B*4415 was also conserved, with few specimens revealing a conservative substitution E to D at anchor position p2. Whether mutations identified at the above mentioned positions affect MHC class I or class II recognition remains to be determined.

Similar analysis of the env-gp41 region from 325 specimens revealed sequence conservation with hot spots at aa 588, 619, 620, 624, 640, 641, 644, 648, 662, 668, 671, and 677 (Fig. 5B). The env-gp41 regions between aa 582–593, 680–693, 700–720, 768–780, 794–822, and 830–850 contained the CTL epitopes. Whereas the T-helper epitopes are not well defined, a single epitope located at aa 613–632 has been defined by in vitro immunization strategies. Overall, most of the epitopes were highly conserved, with few amino acid positions having highly variable structure. The single env-gp41 epitope selected in HIVA, 5B4ERYLKDQQLL, revealed variability at positions R585K/S and K588R/Q/G/H. Again, the functional relevance of these substitutions remains to be determined.

**DISCUSSION**

In the present study, we have molecularly characterized 460 HIV-1 strains from antenatal women in rural western Kenya and found that subtype A is the predominant subtype, followed by subtypes D and C, and a large proportion of the infection represents URFs. This result is in agreement with previous molecular epidemiological studies in Kenya. However, there are new aspects of the present study. First, the current study not only described the distribution of HIV-1 strains in rural western Kenya, but also compared the temporal relationship of concordant and discordant subtypes in the gag-p24 and env-gp41 regions over time. Second, the amino acid sequences of the present characterized viral strains were compared with the consensus A sequence from the HIV database or HIVA, the immunogen used in a vaccine trial in Kenya, to identify variations in the major immunodominant epitopes within gag-p24 and env-gp41. Thus, the current study provides new valuable information.

Multiple forms of URFs with many different subtypes harboring a variety of unique recombination breakpoints have been found in Kenya. In the present study, we found that 50% of the circulating strains in 38 selected specimens were URFs with breakpoints identifiable within the gag-pol region. It is likely that the remaining specimens with discordant subtypes in the gag-p24 and env-gp41 regions represent recombinants with breakpoints within the end of the pol to env-gp41 region. The frequency of recombinants identified by our approach is similar to those using full-length sequences from the central part of Kenya. Extrapolation of these discordant subtypes as potential recombinants led us to estimate that at least 25% of the infections in Kisumu represent infections with recombinants. Almost all recombinants from Kenya had some portion of subtype A. The structures of URFs in Kisumu differed from each other, and did not show any similarity to known CRFs or other recombinants. This strongly suggests that new recombinants are arising continually in Kisumu, Kenya. These data corroborate a recent study in which analysis of full-length genomic sequences from 41 persons identified 39% to be URFs. None of the recombinant structures identified in our study revealed any similarity to the sequences derived from Kenyan blood bank specimens; this fact further indicates that new URFs are continually emerging in Kenya. Assuming that the majority of mosaic viruses with distinct recombinant patterns have independent origins, previous and present studies suggest that dual infections are frequent in this population. Thus, continued investigations are necessary to reveal whether any of these URFs become established as CRFs and initiate a new era of epidemic similar to the ones in Thailand (CRF01_AE), and in West Africa (CRF02_AG).

In contrast to previous studies that have simply described the presence of different HIV-1 subtypes, this is one of the few studies to describe the distribution and degree of genetic diversity over a 4-year period, from 1996 to 2000, in East Africa. Analysis revealed that subtype distribution and proportion of recombinants have not changed appreciably over that time. This observation has bearing on future epidemic trends, as it suggests that newly infected women may be infected with a higher proportion of mosaic genomes. Although few countries, especially those with limited resources, have the necessary infrastructure to systematically monitor the epidemic, periodic molecular epidemiological assessments as we have conducted in Kisumu, Kenya, are important, especially in regions in which trials of subtype-specific vaccines are being considered.

The diverging trend of different HIV-1 subtypes, mosaic viruses, and potential recombinants represents a major challenge in the design and testing of HIV vaccines. Both cross-subtype immunity and subtype-specific immune responses have been reported, however, the relative importance of cross-reactive versus subtype-specific immunity that might be elicited by a protective vaccine remains to be seen. Regardless of these differences, many studies have established that HIV-specific cytotoxic T cell responses are important in protection against both HIV infection and disease progression. Thus far, an HIV-1 peptide-based approach and a multi-CTL-epitope expressing construct have elicited a broadly reactive immune response when used alone or in a prime-boost combination strategy with other vaccine candidates. One such candidate antigen, HIVA, is made up of a string of predefined subtype A epitopes encoded as DNA, which aims to prime HIV-specific
immunity. This candidate vaccine antigen is undergoing clinical trials in parts of Kenya and other East African countries. We therefore carried out a comparative analysis of amino acid sequences from the current study to those selected in HIVA. Despite the enormous genetic diversity with multiple subtypes and recombinants, analysis indicated that most immunodominant regions within the gag-p24 and env-gp41 regions were highly conserved. However, there were substitutions at some critical amino acid positions for some of the B14 and B27-restricted epitopes within the gag-p24 region, and the functional relevance of these changes will have to be further investigated.

SEQUENCE DATA

The sequences from this study have been deposited at GenBank with accession numbers AY492752–AY492789 for the gag-pol region, AY492790–AY493198 for the env-gp41 region, and AY492340–AY492751 for the gag-p24 region.

ACKNOWLEDGMENTS

We are grateful to all the participants and the staff of KEMRI, Kisumu and New Nyanza Provincial General Hospital. The authors wish to thank Drs. K. DeCock, L. Slutsker, A. Lal, and Steve McDougal for helpful suggestions and critical review of the manuscript, Ruiguang Song for statistical assistance, and K. Distel for editorial assistance.

REFERENCES

31. Hall T: BioEdit: A user-friendly biological sequence alignment ed-
itor and analysis program for window 96/98/NT. Nucleic Acids Symp Ser No 41, 1999;95–98.


Address reprint requests to:
Chunfu Yang
HIV Immunology and Diagnostic Branch
DASTLR, NCHSTP, CDC, Mail Stop D-12
1600 Clifton Road
Atlanta, Georgia 30333
E-mail: cyangl@cdc.gov