Among 46 Near Full Length HIV Type 1 Genome Sequences from Rakai District, Uganda, Subtype D and AD Recombinants Predominate

MATTHEW E. HARRIS,1 DAVID SERWADDA,2 NELSON SEWANKAMBO,2 BOHYE KIM,3 GODFREY KIGOZI,2 NOAH KIWANUKA,2 JAMES B. PHILLIPS,3 FRED WABWIRE,2 MARY MEEHEN,4 TOM LUTALO,5 JAMES R. LANE,3 RANDALL MERLING,3 RON GRAY,6 MARIA WAWER,4 DEBORAH L. BIRX,1 MERLIN L. ROBB,1 and FRANCINE E. MCCUTCHEON3

ABSTRACT

The impact of HIV-1 genetic diversity on candidate vaccines is uncertain. To minimize genetic diversity in the evaluation of HIV-1 vaccines, vaccine products must be matched to the predominant subtype in a vaccine cohort. To that end, full genome sequencing was used to detect and characterize HIV-1 subtypes and recombinant strains from individuals in Rakai District, Uganda. DNA extracted from peripheral blood mononuclear cells (PBMC) was PCR amplified using primers in the long terminal repeats (LTRs) to generate nearly full length genomes. Amplicons were directly sequenced with dye terminators and automated sequencers. Sequences were phylogenetically analyzed and recombinants were detected and mapped with distance scan and bootscan. Among 46 sequences, 54% were subtype D, 15% were subtype A, and 30% were recombinant. All recombinants were individually unique, and most combined subtypes A and D. Subtype D comprised more than 70% of all the HIV-1 genomes in Rakai when both pure subtypes and recombinants were considered. Candidate vaccines based on HIV-1 subtype D would be appropriate for evaluation in Rakai District, Uganda.

INTRODUCTION

HIV-1 is genetically diverse.1–4 A phylogenetic tree of HIV-1 sequences reveals three groups: M (main group), N (new or non-M, non-O), and O (outlier).1,5 Group M represents the vast majority of the infections worldwide and is divided, by phylogenetic analysis, into nine distinct subtypes (A, B, C, D, F, G, H, J, K).1,3,5 Intersubtype diversity ranges from 15% of nucleotides in the structural (gag) gene to 25% in the envelope (env) gene.1 Intrasubtype diversity ranges from 5 to 12%.1

Genetic recombination further contributes to the diversity of HIV-1.3,4,6 Recombination arises when a cell becomes coinfected with two different viruses. A plus-strand RNA from each virus can be copackaged into a new virion, and, after infection of a new cell, reverse transcription can recombine the genetic material from the two RNA strands during proviral DNA synthesis.3 It is difficult to define the contribution of recombination within the context of the primarily HIV-1 subtype B epidemic in the United States. However, the frequency can be defined in regions of the world in which the HIV-1 epidemic has subtype diversity and intersubtype recombination occurs. A circulating recombinant form (CRF) can be identified when three or more sequences from epidemiologically unrelated individuals have recombinant viruses sharing identical recombinant structures.1,5 To date, 14 circulating recombinant forms have been reported (http://hiv-web.lanl.gov).

The distribution of HIV-1 subtypes and recombinants varies geographically.2–4 The greatest diversity of HIV-1 subtypes is observed in West Central Africa where the HIV-1 epidemic is thought to have arisen from one or more zoonotic transmissions.2,4,7–9 The inferred spread of HIV-1 subtypes from West

1Walter Reed Army Institute of Research, Silver Spring, Maryland 20910.
2Makerere University, Kampala, Uganda.
3The Henry M. Jackson Foundation, Rockville, Maryland 20850.
4Columbia University, New York, New York 10032.
5Uganda Virus Research Institute, Entebbe, Uganda.
6Johns Hopkins University, Baltimore, Maryland 21205.
Central Africa has apparently involved a limited number of subtypes. For example, subtype B represents the vast majority of infections in the United States, Europe, and Australia and subtype C predominates in Southern Africa. CRFs represent the predominant viral strain in several countries. Differences between HIV-1 subtypes or human population genetics may have played some part in the global distribution of subtypes.

The genetic diversity of HIV-1 represents one of several challenges in the path to a globally effective vaccine. There are several viruses for which vaccines have been successful that have much less genetic diversity than HIV-1. It is unknown if a vaccine based on one particular sequence or even on a consensus sequence from one HIV-1 subtype will elicit an immune response sufficient to prevent infection by other subtypes or by a recombinant. There is evidence for cross-subtype recognition by both neutralizing antibodies and cytotoxic T cell lymphocytes. These patterns may be due to the random introduction of founder viruses.

Differences between HIV-1 subtypes or human population genetics may have played some part in the global distribution of subtypes. The genetic diversity of HIV-1 represents one of several challenges in the path to a globally effective vaccine. There are several viruses for which vaccines have been successful that have much less genetic diversity than HIV-1. It is unknown if a vaccine based on one particular sequence or even on a consensus sequence from one HIV-1 subtype will elicit an immune response sufficient to prevent infection by other subtypes or by a recombinant. There is evidence for cross-subtype recognition by both neutralizing antibodies and cytotoxic T cell lymphocytes. These patterns may be due to the random introduction of founder viruses.

Materials and methods

Cohort

Thirty-seven samples were derived from volunteers in Rakai district, Uganda enrolled into longitudinal, population-based cohort studies beginning in 1995. Participants in the current analysis were identified from the STD Control for AIDS Prevention trial (1994–1999) and from the Community HIV Epidemiological Research Study, a population-based HIV surveillance project. All volunteers provided written informed consent at each survey round. The studies were reviewed and approved by human subject ethics and safety committees in the United States (Columbia University, Johns Hopkins University, Walter Reed Army Institute of Research, NIH Office of Human Research Protection) and in Uganda (Uganda Virus Research Institute and the AIDS Research Subcommittee of the Uganda Council for Science and Technology). All participants were offered health and HIV prevention education, condoms, and voluntary HIV testing and counseling.

Of the 46 individuals whose HIV-1 strain was evaluated in these studies, 17 were prevalent cases, and 29 seroconverted during the studies. Seroconversion for these incident cases was estimated as the midpoint between the last negative and first positive HIV-1 test, which were conducted approximately every 10 months.

Materials

When a volunteer tested positive for HIV-1, whole blood was drawn at 0 and 2 months. Peripheral blood mononuclear cells (PBMC) were purified from anticoagulated whole blood using CPT (Becton Dickinson, Annapolis, MD) or ADC/Ficoll density centrifugation (Amersham Pharmacia Biotech, Piscataway, NJ). Samples 57128, 57129, 57131, 57132, 57134–57137, and 57139 were anonymously collected, discard samples from a clinic in Rakai. Some samples (57128, 57130, 57143, 57147, K30889, and A07412) were cocultured with phytohemagglutinin (PHA)-stimulated donor PBMC prior to extraction.

Sample processing and amplification

Genomic DNA was extracted from PBMC, derived either from the 0 or 2 month blood draw, using Qiagen’s Blood and Tissue kit. Some samples (57128, 57143, 57147, K30889, and A07412) were cocultured with phytohemagglutinin (PHA)-stimulated donor PBMC prior to extraction.

Sequencing

PCR amplicons were directly sequenced with the exception of samples 57128, 57129, 57131, 57136, 57137, 57139, 57140, 57142, 57143, and 57146. These amplicons were cloned into pCR-XL-TOPO using TOPO TA cloning kits (Invitrogen, San Diego, CA) or in the case of 57143 into pLTR(B) as previously described. One full length clone was selected for sequencing.

Phylogenetic analysis

Sequences were manually aligned to an existing alignment that began at the seventh nucleotide of gag and extended two-thirds into nef. Gaps in the alignment were stripped prior to analysis. Phylogenetic analysis was performed using the Phylip
software package. Neighboring trees with bootstrap (100 iterations) were built using Phylip distance methods with the Kimura two-parameter model. Bootstrapping values greater than or equal to 70% were considered significant. Distance scans using the Kimura two-parameter model were performed with 300 nucleotide windows overlapping by 50 nucleotides and a transition/transversion ratio of 2.0. Bootscan analyses were performed using parsimony with 300 nucleotide windows overlapping 20 nucleotides. Initially, reference sequences of subtypes A, D, C, and J were used for bootscanning. Then distance scans were used to pick the best references for bootscanning of potential recombinants. Breakpoints were estimated as the midpoint of the window where bootstrap values approached 50%.

The reference sequences used in neighboring trees to confirm breakpoints included A1 (92UG037, SE8891), A2 (97CDKT48, 94CY017), F1 (V1850, FIN9363), F2 (95CMMPP555, CM53657), B (LAI, MN), C (97TZ05, 97TZ04), D (94UG1141, MB2059), G (SE6165, HH8793), H (V1991, 90CF056), J (SE92809, SE91733), K (97EQTB11C, 96CMMPP535C), CRF01_AE (CM240, 90CF402), and CRF02_AG (IbNG, DJ264).

**Genetic distribution analysis**

To interpret the overall genetic composition of major HIV-1 structural genes, *gag*, *pol*, *env*, the contribution to subtypes A, C, and D from each of the recombinants was calculated using the breakpoint positions in the alignment and the positions of start and stop codons for each gene. Nonrecombinant and recombinant viruses were then combined to determine the proportions of subtypes A, C, and D in all sequences for each of the major gene products that will be vaccine targets. The signal peptide of *env* was not included in the analysis of *env*. Gaps were not stripped for the analysis of genetic distribution.

**Hypermutation analysis**

All seven nonrecombinant subtype A sequences were analyzed by the Los Alamos Hypermut program (http://hiv-web.lanl.gov/HYPERMUT/hypermut.html) to identify the presence of hypermutation. Subtype A strains 57136 and G03379 were used as reference sequences. Similarly, all 25 nonrecombinant subtype D sequences were analyzed for hypermutation using Hypermut. Subtype D strains A07412 and A03349 served as reference sequences.

**RESULTS**

Viral sequences were obtained from 46 HIV-1 seropositive individuals. Demographic data for these individuals and the subtype of the virus are presented in Table 1. In this analysis, 62% of volunteers were males, and the mean age in years (± standard deviation) was 27 (±8). The mean ages in years for females and males were 24 (±7) and 29 (±8), respectively. The blood draw dates for sequencing occurred in 1998 and 1999. For the incident cases, the mean time (± standard deviation) between seroconversion and collection of PBMC for sequencing was 17 months (±6 months). Thus, the sequences represent a recent composite of the viruses in the Rakai district and include both prevalent and incident infections.

A neighboring tree of the 46 sequences, along with reference sequences of subtypes A, C, and D, was generated (Fig. 1A). Thirty-three of the sequences clustered with the subtype D reference sequences, and four of these (57139, E22831, F27390, and K38855; Fig. 1A) formed noticeably deeper branches in the subtype D cluster. Twelve sequences clustered significantly with subtype A references, and two of these (C06443 and J21953, Fig. 1A) have noticeably deeper branches in the cluster. Sequence B26587 was an outlier to all of these subtypes; none of the sequences shared a significant bootstrap with only subtype C references.

The deeper branches within subtypes A and D and the outlier sequence suggested that these sequences were recombinant, so all sequences were examined for the presence of recombination using bootscanning (data not shown). Fourteen of the 46 sequences were recombinant and are indicated by crosshatching in Figure 1A. Some of them clustered closely within subtype A or D, while others were either deep branches within these subtypes or were outside all three subtypes. The confirmation and structure of the recombinant strains will be described below.

To gain information about the intrasubtype diversity among the nonrecombinant strains, another neighboring tree was constructed excluding the recombinant strains (Fig. 1B). References for subtypes C, D, and A including references for the recently identified A2 subtype were included. The 25 subtype D sequences formed a significant cluster with references from East Africa (bootstrap value of 100%). There were seven nonrecombinant subtype A sequences. None of the sequences, identified as subtype A, clustered with references for subtype A2. C27305 is the most distant of these; it had an unusually long branch length. The translated protein sequences of C27305 showed many premature stop codons in its open reading frames (data not shown). Using the Hypermut program (http://hiv-web.lanl.gov/HYPERMUT/hypermut.html), it was determined that C27305 had an approximately 2-fold greater number of G to A changes relative to the other subtype A sequences from Rakai (data not shown). The distribution of these G to A changes in the four GN dinucleotide contexts (where N is any nucleotide) was skewed to the GG context consistent with hypermutation. In contrast, the other subtype A Rakai sequences had an almost even distribution of G to A changes in the four contexts (data not shown). No nonrecombinant subtype D strains were found to be hypermutated (data not shown). Excluding strain C27305, the mean (± standard deviation) pairwise distances between subtype A strains (n = 6) and between subtype D strains (n = 25) were 0.08 (±0.01) and 0.070 (±0.008), respectively. In comparison, 11 full length subtype A references from Africa had a mean pairwise distances of 0.079 (±0.009) and were intermingled with the Rakai subtype A sequences. Twenty-three of the 25 subtype D sequences and two references formed a significant internal cluster (bootstrap value of 100%). Sequences 57130 and 57143 and several references were outside this internal cluster (Fig. 1B). 57130, 57143, and Ugandan reference 92UG001 were 1–2% more distant relative to the distances between the subtype D Rakai sequences in the internal cluster. However, a bootstrap (100%) clustered 57130, 57143, and 92UG001 with the other Rakai
subtype D sequences. West African references formed a significant cluster (bootstrap value of 100%).

Next the recombinant strains were analyzed in detail. Distance scanning was used to identify the component subtypes and bootscanning was used to locate recombination breakpoints.

Segments of the genome were excised according to these breakpoints and analyzed by neighbor-joining with the bootstrap values greater than 70% were used to assign segments to a particular subtype. The results of these analyses are shown in

---

**Table 1. Demographic Data and the Subtype or Recombinant Type of HIV-1 of Individuals from Rakai District, Uganda**

<table>
<thead>
<tr>
<th>Sample name</th>
<th>Age&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Sex&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Blood draw for sequencing (dd/mm/yy)</th>
<th>Interval between estimated seroconversion and blood draw (months)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Subtype</th>
<th>Genbank accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>98UG57128</td>
<td>n.a.</td>
<td>n.a.</td>
<td>May 1998</td>
<td>Unknown</td>
<td>D</td>
<td>AF484502</td>
</tr>
<tr>
<td>98UG57129</td>
<td>n.a.</td>
<td>n.a.</td>
<td>May 1998</td>
<td>Unknown</td>
<td>AD</td>
<td>AF484503</td>
</tr>
<tr>
<td>98UG57130</td>
<td>25 M</td>
<td></td>
<td>May 1998</td>
<td>Unknown</td>
<td>D</td>
<td>AF484504</td>
</tr>
<tr>
<td>98UG57131</td>
<td>n.a.</td>
<td>n.a.</td>
<td>May 1998</td>
<td>Unknown</td>
<td>D</td>
<td>AF484505</td>
</tr>
<tr>
<td>98UG57132</td>
<td>n.a.</td>
<td>n.a.</td>
<td>May 1998</td>
<td>Unknown</td>
<td>D</td>
<td>AF484506</td>
</tr>
<tr>
<td>98UG57134</td>
<td>n.a.</td>
<td>n.a.</td>
<td>May 1998</td>
<td>Unknown</td>
<td>A</td>
<td>AF484507</td>
</tr>
<tr>
<td>98UG57135</td>
<td>n.a.</td>
<td>n.a.</td>
<td>May 1998</td>
<td>Unknown</td>
<td>A</td>
<td>AF484508</td>
</tr>
<tr>
<td>98UG57136</td>
<td>n.a.</td>
<td>n.a.</td>
<td>May 1998</td>
<td>Unknown</td>
<td>A</td>
<td>AF484509</td>
</tr>
<tr>
<td>98UG57137</td>
<td>n.a.</td>
<td>n.a.</td>
<td>May 1998</td>
<td>Unknown</td>
<td>AD</td>
<td>AF484521</td>
</tr>
<tr>
<td>98UG57139</td>
<td>n.a.</td>
<td>n.a.</td>
<td>May 1998</td>
<td>Unknown</td>
<td>AD</td>
<td>AF484510</td>
</tr>
<tr>
<td>98UG57140</td>
<td>25 F</td>
<td></td>
<td>May 1998</td>
<td>Unknown</td>
<td>D</td>
<td>AF484511</td>
</tr>
<tr>
<td>98UG57142</td>
<td>27 F</td>
<td></td>
<td>May 1998</td>
<td>Unknown</td>
<td>A</td>
<td>AF484512</td>
</tr>
<tr>
<td>98UG57143</td>
<td>40 M</td>
<td></td>
<td>May 1998</td>
<td>Unknown</td>
<td>D</td>
<td>AF484514</td>
</tr>
<tr>
<td>98UG57146</td>
<td>40 F</td>
<td></td>
<td>May 1998</td>
<td>Unknown</td>
<td>D</td>
<td>AF484513</td>
</tr>
<tr>
<td>98UG57147</td>
<td>30 F</td>
<td></td>
<td>May 1998</td>
<td>Unknown</td>
<td>AD</td>
<td>AF484522</td>
</tr>
<tr>
<td>99UGA03349</td>
<td>37 M</td>
<td></td>
<td>14/10/1999</td>
<td>14</td>
<td>D</td>
<td>AF484518</td>
</tr>
<tr>
<td>99UGA07072</td>
<td>40 M</td>
<td></td>
<td>14/09/1999</td>
<td>13</td>
<td>A</td>
<td>AF484478</td>
</tr>
<tr>
<td>99UGA07412</td>
<td>32 M</td>
<td></td>
<td>2/8/99</td>
<td>25</td>
<td>D</td>
<td>AF484477</td>
</tr>
<tr>
<td>99UGB21875</td>
<td>27 M</td>
<td></td>
<td>17/08/1999</td>
<td>27</td>
<td>D</td>
<td>AF484480</td>
</tr>
<tr>
<td>99UGB25647</td>
<td>15 M</td>
<td></td>
<td>6/9/99</td>
<td>11</td>
<td>D</td>
<td>AF484481</td>
</tr>
<tr>
<td>99UGB26587</td>
<td>18 F</td>
<td></td>
<td>7/9/99</td>
<td>28</td>
<td>AD</td>
<td>AF484482</td>
</tr>
<tr>
<td>99UGB32394</td>
<td>20 F</td>
<td></td>
<td>12/7/99</td>
<td>16</td>
<td>D</td>
<td>AF484483</td>
</tr>
<tr>
<td>99UGC06443</td>
<td>26 M</td>
<td></td>
<td>16/08/1999</td>
<td>23</td>
<td>AD</td>
<td>AF484479</td>
</tr>
<tr>
<td>99UGC27305</td>
<td>16 F</td>
<td></td>
<td>4/8/99</td>
<td>17</td>
<td>A</td>
<td>AF484484</td>
</tr>
<tr>
<td>99UGC38442</td>
<td>23 F</td>
<td></td>
<td>2/11/99</td>
<td>Unknown</td>
<td>AD</td>
<td>AF484517</td>
</tr>
<tr>
<td>99UGD23550</td>
<td>20 F</td>
<td></td>
<td>15/07/1999</td>
<td>23</td>
<td>D</td>
<td>AF484485</td>
</tr>
<tr>
<td>99UGD26830</td>
<td>30 M</td>
<td></td>
<td>10/8/99</td>
<td>12</td>
<td>D</td>
<td>AF484486</td>
</tr>
<tr>
<td>99UGE08364</td>
<td>23 M</td>
<td></td>
<td>4/8/99</td>
<td>11</td>
<td>D</td>
<td>AF484487</td>
</tr>
<tr>
<td>99UGE13613</td>
<td>35 M</td>
<td></td>
<td>1/9/99</td>
<td>12</td>
<td>D</td>
<td>AF484515</td>
</tr>
<tr>
<td>99UGE22831</td>
<td>27 M</td>
<td></td>
<td>12/8/99</td>
<td>12</td>
<td>AD</td>
<td>AF484488</td>
</tr>
<tr>
<td>99UGE23438</td>
<td>15 F</td>
<td></td>
<td>15/07/1999</td>
<td>16</td>
<td>D</td>
<td>AF484489</td>
</tr>
<tr>
<td>99UGF03726</td>
<td>25 M</td>
<td></td>
<td>19/10/1999</td>
<td>17</td>
<td>D</td>
<td>AF484519</td>
</tr>
<tr>
<td>99UGF05734</td>
<td>30 M</td>
<td></td>
<td>19/08/1999</td>
<td>15</td>
<td>D</td>
<td>AF484490</td>
</tr>
<tr>
<td>99UGF25926</td>
<td>23 M</td>
<td></td>
<td>13/09/1999</td>
<td>13</td>
<td>AD</td>
<td>AF484491</td>
</tr>
<tr>
<td>99UGF27390</td>
<td>33 F</td>
<td></td>
<td>15/07/1999</td>
<td>23</td>
<td>AD</td>
<td>AF484492</td>
</tr>
<tr>
<td>99UGG03379</td>
<td>42 M</td>
<td></td>
<td>26/08/1999</td>
<td>17</td>
<td>A</td>
<td>AF484493</td>
</tr>
<tr>
<td>99UGG10555</td>
<td>45 M</td>
<td></td>
<td>19/07/1999</td>
<td>24</td>
<td>D</td>
<td>AF484494</td>
</tr>
<tr>
<td>99UGG35093</td>
<td>25 F</td>
<td></td>
<td>14/07/1999</td>
<td>Unknown</td>
<td>D</td>
<td>AF484495</td>
</tr>
<tr>
<td>99UGJ21953</td>
<td>34 M</td>
<td></td>
<td>12/10/99</td>
<td>19</td>
<td>AD</td>
<td>AF484496</td>
</tr>
<tr>
<td>99UGJ27597</td>
<td>35 M</td>
<td></td>
<td>27/07/1999</td>
<td>10</td>
<td>D</td>
<td>AF484497</td>
</tr>
<tr>
<td>99UGJ32228</td>
<td>22 M</td>
<td></td>
<td>16/09/1999</td>
<td>12</td>
<td>D</td>
<td>AF484516</td>
</tr>
<tr>
<td>99UGK09259</td>
<td>30 F</td>
<td></td>
<td>14/09/1999</td>
<td>13</td>
<td>D</td>
<td>AF484498</td>
</tr>
<tr>
<td>99UGK09958</td>
<td>25 M</td>
<td></td>
<td>12/8/99</td>
<td>23</td>
<td>D</td>
<td>AF484499</td>
</tr>
<tr>
<td>99UGK10192</td>
<td>17 M</td>
<td></td>
<td>22/07/1999</td>
<td>17</td>
<td>CD</td>
<td>AF484500</td>
</tr>
<tr>
<td>99UGK30889</td>
<td>19 M</td>
<td></td>
<td>26/08/1999</td>
<td>12</td>
<td>AC</td>
<td>AF484501</td>
</tr>
<tr>
<td>99UGK38855</td>
<td>19 F</td>
<td></td>
<td>1/12/99</td>
<td>9</td>
<td>AD</td>
<td>AF484520</td>
</tr>
</tbody>
</table>

<sup>a</sup>n.a., not available.

<sup>b</sup>Interval is unknown for prevalent cases.
FIG. 1. (A) Phylogenetic analysis of 46 near full length HIV-1 sequences. A distance matrix (Kimura two-parameter) was generated and used to construct a neighbor-joining tree of all 46 near full length sequences and references for subtypes A, C, and D. The stability of the tree topography was tested by bootstrapping. Only those significant bootstraps that show clusters of the greatest number of sequences and references from a single subtype are shown. The patterns in the boxes represent a subtype or recombinant as follows: subtype A, subtype D, AD recombinant, AC recombinant, CD recombinant. The names of the reference sequences are preceded by their subtype assignment, and the clusters are labeled as subtype A, C, or D. The scale bar represents 10% genetic distance. (B) Phylogenetic analysis of 32 nonrecombinant full length HIV-1 sequences. The neighbor-joining tree was generated as in A but includes only the nonrecombinant strains. References for subtypes A, including both subsubtypes, C, and D were included. Not all significant bootstraps are indicated. The names of the reference sequences are preceded by their subtype or subsubtype assignment. The scale bar represents 10% genetic distance.
Figure 2. With the exception of a small region of strain 57139, all regions of the recombinants could be assigned to subtype A, C, or D. Each of the recombinants had a different structure. There were 11 AD recombinants, 2 AC recombinants, and one CD recombinant. Both AC recombinants have subtype C vpu. Among the AD recombinants, there were strains composed almost entirely of subtype A, but others were almost entirely subtype D. Two had only two crossover points while one alternated subtype 11 times across the genome. There was no obvious preferred location for breakpoints.

Of interest, there were three subtype C-containing recombinants but no nonrecombinant subtype C strains. The subtype and breakpoint structures of the recombinant strains did not match the structures of recombinants recently identified in neighboring Kenya\textsuperscript{55} and Tanzania\textsuperscript{56} or reported in GenBank (data not shown).

The overall proportions of subtypes and recombinants are depicted in Figure 3. Subtype D accounted for 54%, subtype A for 15%, and recombinant strains comprised 30%. Most of the recombinants were AD, constituting 86% of the recombinants and 24% of all strains. When considering only samples from individuals with known dates of seroconversion (which were within 1997 and 1998), subtype D was 59%, subtype A was 10%, and recombinants strains were 31%.

The most prominent HIV-1 strains among this sample from Rakai were subtype D and D-containing recombinants, together constituting 80% of the strains. To further inform vaccine development, the subtype composition of major structural genes in the sample was considered. If the coding sequence for a structural gene contained one or more breakpoints in a recombinant strain, we estimated the fraction of the gene that derived from each of the subtypes. When combined with the pure subtypes, the results (Fig. 4) showed no obvious disparity among gag, pol, and env genes. Figure 4 shows the separate contributions of nonrecombinant (solid bars) and recombinant (hatched bars) strains. More than 70% of the genetic material in structural genes comes from subtype D, mostly within nonrecombinant strains. At least half of the genetic material from recombinants was subtype D. Subtype A accounts for about 30% of the genetic material, with about half from nonrecombinant and half from recombinant strains. The trace amount of subtype C material was mostly in gag and pol genes and only in recombinants in each of the three genes.

DISCUSSION

Subtype D is the most important component of HIV-1 strains in Rakai. Nonrecombinant subtype D represents more than half of the viruses sequenced, and subtype D is found in all but two

**FIG. 2.** Structure of the recombinant viruses. The breakpoints of each of the recombinants are depicted in relation to the HIV-1 genome. Each box represents a subgenomic region whose breakpoints were found by bootscanning. The subtype for each region is indicated above each box as well as by the pattern in each box. The patterns are as follows: ■ subtype A, ■ subtype D, □ subtype C, or ■ unclassified. The number in each box is the bootstrap value for a cluster containing the region of interest with references of the indicated subtype (NS = not significant, a bootstrap less than 70%).
recombinants. In the total viral population, subtype D represents 70% or more of the genetic material in three HIV-1 genes that are candidates for inclusion in a vaccine. All but two of the nonrecombinant subtype D sequences radiate from approximately the same point and are roughly equidistant from each other. Two sequences (57130 and 57143) are outliers relative to the other Rakai subtype D sequences based on their position in a phylogenetic tree of nonrecombinant subtypes. However, these two sequences did cluster with the Rakai sequences and East African references rather than with West African subtype D references. A candidate vaccine based on subtype D, using any of the 23 equidistant subtype D sequences reported here, would be an appropriate choice for a monovalent vaccine in the Rakai district.

Subtype A is the other nonrecombinant found in Rakai, representing approximately 15% of strains. When the amount of subtype A is considered in both recombinant and nonrecombinant viruses, it comprises about 30% of the sequence coding for gag, pol, and env. Previous estimates of subtype A by partial genome analyses ranged from 20% to over 50%. The reports above sampled outside of the Rakai district, and it is possible that the proportion of subtype A varies in different parts of Uganda, but underestimation of recombinants will also contribute to the differences.

Recombinants represent 30% of the viruses in Rakai. Partial genome analyses had predicted 11–28% recombinants. Most of the estimates are low relative to our data and are likely due to the examination of only one or two small regions in the genome. Based on our data, it is likely that many of the subtype A fragments detected previously were part of recombinant viruses, especially AD recombinants. A recent study by Yirrell et al., which sequenced within gag and env, estimated 28% recombinants, and 23% AD recombinants, which closely mirrors our data. Their estimates of subtype A and D also are close to our findings. No circulating recombinant forms, or CRFs, have been de-
ected in the Rakai district. The AD recombinants identified in this study are all unique varying greatly in their relative composition of subtypes A and D. The single CD recombinant is distinct from CRF10_CD found in neighboring Tanzania and other known CD recombinants. The two AC recombinants are also unique.

Partial genome analyses previously identified subtypes C24–26,28,29,31,34,38,39,58 and G24,30,39 in Uganda. In these analyses, subtypes C and G were found in either gag or env. Using analyses of multiple genome regions, two studies had evidence for pure subtype C strains and one study had evidence for a pure strain of subtype G. None of our 46 full length sequences from Rakai is a pure subtype C or G, and the Los Alamos database does not have a single submission of a full length subtype C or G from Uganda. We detected subgenomic regions of subtype C, but this represents very little of the genetic material in Rakai. We found subtype C in gag, vpu, and nef but not in env as previously reported. The analysis of full length sequences did not reveal any subtype G regions except where it is not separable from subtype A. Subtype G is found primarily in Central and West Africa.

Two countries neighboring Uganda have epidemics with the same subtypes as Rakai but in different proportions. In southern Tanzania (Myeba), the predominant subtype in nonrecombinant and recombinant viruses is C. This probably reflects introduction from neighbors to the south where subtype C is abundant. In Myeba, there are also nonrecombinant (22%) and recombinant subtype A and recombinant subtype D viruses. Partial genome sequencing reveals subtype A and D in other regions of Tanzania. Subtype C has only recently been observed in the northern areas of Tanzania. Bordering Uganda on the east, Kenya has predominantly subtype A in both nonrecombinants (55%) and recombinants based on a new study of near full length sequences. Subtypes C and D occur as nonrecombinants (2% each) but to a much lesser extent than subtype A, and they are also found in recombinants. A full length subtype G has been found previously from Kenya, but was not found in the new report. With recombinants between A1, A2, and D; A2 and D; A1 and D; A1 and G; A1 and C; A1, C, and D; and C and D, Kenya has more diversity among its recombinants than those in Rakai, Uganda.

In summary, subtype D is the predominant subtype found in nonrecombinant and recombinant viruses in the Rakai district of Uganda. Despite 30% of our sequences being recombinants, 70% of the genetic material in gag, pol, and env is subtype D. Thus, a vaccine based on subtype D has the opportunity to elicit immune responses that are closely related, wholly or in part, to a large majority of viruses in Rakai. The proposed, future vaccine trials for East Africa can be a model for subtype escape and will provide insight into vaccine requirements for broadness of immunogenicity and evidence of cross-protection.

SEQUENCE DATA

The GenBank accession numbers for the 46 sequences are AF484477–AF484522 (see also Table 1).

ACKNOWLEDGMENTS

The authors thank Drs. Jean Carr and William Dowling for their critical reading of the manuscript. This work was supported by a cooperative agreement between the Henry M. Jackson Foundation for the Advancement of Military Medicine and the United States Department of Defense. The views and opinions expressed herein do not necessarily reflect those of the U.S. Army or of the Department of Defense.

REFERENCES


49. Buijtenen J, Treetool, 2.0.2. Ribosomal DNA Database Project, University of Illinois Board of Trustees, Urbana, IL, 1994.


Address reprint requests to:

Matthew E. Harris
Walter Reed Army Institute of Research
Division of Retrovirology
1 Taft Court, Suite 250
Rockville, MD 20850

E-mail: Harris@hivresearch.org