AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 16, Number 5, 2000, pp. 393–401
Mary Ann Liebert, Inc.

Molecular Epidemiology of HIV Type 1 in a Rural Community in Southwest Uganda

PONTIANO KALEEBU,[1] JAMES WHITWORTH,[1] LAURA HAMILTON,[1] ALLELUIAH RUTEBEMBERWA,[1] FRED LYAGOBA,[1] DILYS MORGAN,[1] MELANIE DUFFIELD,[2] BENON BIRYAHWAHO,[1] BRIAN MAGAMBO,[1] and JON ORAM[2,3]

ABSTRACT

The molecular epidemiology of a population-based cohort in a cluster of 15 villages in southwestern Uganda was investigated by sequencing part of the p24 *gag* gene and performing heteroduplex mobility assays (HMAs) of the V3 region of the *env* gene. Sequence and HMA data, obtained for 69 and 88 proviruses, respectively, showed that the clade A and D viruses were present at a ratio of about 0.67:1. No other clades were detected. Thirteen (22%) of 59 proviruses for which both *gag* and *env* data were obtained appeared to be recombinants. Although both clade A and D viruses were present in 13 of the villages, their distribution was unequal: for example, from *env* data 59% of clade A viruses were found in the eastern villages, compared with only 27% of clade D viruses. Phylogenetic (maximum likelihood) analysis of the p24 *gag* sequences showed a total of five clusters supported by bootstrap resampling values above or close to 75%. Four clusters were sexual partners, but there was no known sexual contact between the persons in the other cluster. The DNA sequences showed between 0.5 and 8.3% divergence from the cohort clade A or D consensus sequences. The sequences were not closely related to those published for other clade A or D proviruses.

INTRODUCTION

THE NUMBER OF UGANDANS infected with HIV-1 has risen steadily[1] since the early recognition of HIV-induced disease in Uganda.[2] There is a wide variation in the prevalence of the virus, ranging from about 5% in rural to about 20% in urban areas. The majority of Ugandan viruses are in clades A and D, with few reports[3–5] of viruses in other clades within the M group of HIV-1. The relative proportions of clade A and D viruses appear to vary in different parts of the country[5–7] and it is possible that the ratio of clade A:D viruses may be changing with time. However, with one exception,[7] existing data are largely confined to hospitals or urban areas, both with high rates of population change, of limited value in molecular epidemiological studies of HIV in Uganda.

In 1989 the Medical Research Council (MRS) Programme on AIDS in Uganda established a population-based cohort, consisting of approximately 5000 adults in a cluster of 15 neighboring villages in rural southwest Uganda. This study population is followed by annual census and serosurveys and in 1990 a random selection of a third of all individuals identified as HIV seropositive were invited to enroll as prevalent cases in a natural history cohort (NHC). All seroconverters identified over the subsequent years were invited to enroll as incident cases and randomly selected, age-stratified HIV-seronegative controls from the general population cohort were also invited to enroll. By the end of 1995 the NHC, which has been described in detail,[8] comprised 93 prevalent cases, 86 incident cases, and 168 seronegative controls, and those living in the study area were invited to attend the study clinic every 3 months. During these visits, participants were given a detailed clinical and physical examination and are encouraged to attend the clinic for medical treatment at any time.

In the early years of the MRC program the main emphasis was on epidemiological and clinical aspects of HIV and AIDS, with limited attempts to characterize the viruses circulating in the study area: for example, little was known of the distribution of viral clades or subtypes in the cohort. The NHC is well suited for studies of the molecular epidemiology of HIV in rural Uganda and we now report the first detailed genetic analysis of the viruses circulating within our study area.

The first objective of the study was to determine the fre-

[1]Medical Research Council Programme for AIDS in Uganda/Uganda Virus Research Institute, Entebbe, Uganda.
[2]Centre for Applied Microbiology and Research, Porton Down, Salisbury, Wilts SP4 0JG, UK.
[3]Westek, Callington, Cornwall PL17 7JT, UK.

quency of viral clades in the study area and to determine the frequency of *gag/env* recombinant viruses. The second objective was to determine the genetic relationships both between the proviruses in the cohort and also between the cohort proviruses and those from other African countries by means of a phylogenetic analysis of *gag* gene DNA sequence data.

MATERIALS AND METHODS

Subjects and blood samples

Blood samples were taken from 136 seropositive members of the NHC who attended the clinic between December 1995 and March 1996. The samples were collected in EDTA tubes, transported to the Uganda Virus Research Institute (UVRI, Entebbe, Uganda), and frozen at −20°C prior to DNA extraction. DNA was extracted from frozen whole blood using either a commercial Puregene kit (Gentra Systems, Minneapolis, MN) or a previously described method[9] with a few modifications. Briefly, blood was thawed slowly on ice and lysis buffer (10 mM Tris [pH 7.6], 10 mM KCl, 10 mM $MgCl_2$ plus 1% Nonidet P-40 [NP-40]) was added to 3 ml or less of whole blood, to give a final volume of 6 ml in a 15-ml centrifuge tube. This was mixed gently and centrifuged at 200 rpm for 10 min. The pellet was washed with 5 ml of lysis buffer, centrifuged as described above, and transferred to an Eppendorf tube. After centrifuging at 13,000 rpm for 20 sec, the deposit was resuspended in 150 μl of lysis buffer containing proteinase K (200 μg/ml) and incubated overnight at 56°C.

After the addition of 50 μl of 5× ANE (10 mM sodium acetate, 100 mM NaCl, 1 mM EDTA, 0.5% sodium dodecyl sulfate [SDS]), DNA was extracted with phenol and chloroform, precipitated with 2 vol of ethanol, dissolved in TE buffer (10 mM Tris-HCl [pH 7.5], 1 mM EDTA), and stored at −20°C.

PCR

To minimize the risk of contamination, well-separated laboratories were used for the preparation of DNA and polymerase chain reaction (PCR) reagents/reaction mixtures: no cloned DNA was used in either laboratory. First- and second-stage PCR and the analyses of PCR products were done in three additional laboratories. About one in five of each set of reaction mixtures was a negative control, containing either DNA from HIV-seronegative persons or distilled water. No PCR products were obtained from any of the negative controls.

Preliminary results showed that primers used to amplify *gag* sequences from clade B viruses were not suitable for use with Ugandan proviruses, almost all of which are in clades A and D. Therefore, consensus clade A and D sequence data for the p24 region of the *gag* gene[10] were used to prepare sets of redundant primers for the amplification of Ugandan proviral DNA. A 520-bp fragment was generated by two-stage PCR, in which the first stage primers were

*gag*1: 5′-CAGC/TCAAAATTA/TC/TCCC/TATAG-3′
*gag*2: 5′-CC/GTCAGCCAAAACC/TCTTGCT-3′

followed by second-stage primers

*gag*3: 5′-GGAGCCACC/TC/GCACAA/GGATC/TTA/GA-3′
*gag*4: 5′-CATGCTGTCATCATTTCC/TTC/TTAA/GTGT-3′

The first-stage reaction mixtures (50 μl) contained 10× PCR buffer II (Perkin-Elmer, Norwalk, CT), 5 μl; 0.078 mM $MgCl_2$; a 200 μM concentration of each dNTP; 10 pmol of primers *gag*1 and *gag*2, and 2.5 units of Amplitaq (Perkin-Elmer). Finally, approximately 0.5–1 μg of DNA was added to the mixtures, which were then thermocycled 35 times at 94°C for 1 min, 55°C for 1 min, and 72°C for 1 min. The first-round products (2 μl) were transferred to the second-stage PCR mixture, which contained the inner (*gag*3 and *gag*4) primer pairs, and thermocyled as described above. For use as DNA-sequencing templates, amplified DNA was purified with a Qiagen kit (Promega, Southampton, UK) and then shipped under ethanol to the Centre for Applied Microbiology (CAMR, Porton Down, Salisbury, UK) for sequence analysis.

For heteroduplex mobility assay (HMA), a 0.5-kb fragment spanning the C2–C3 regions of *env*, including the V3 region, was produced by nested PCR as previously described.[11] Briefly, primers ED5/ED12 were used in first-stage reactions followed by ED31/ED33 as second-stage primers. Replicate samples of amplification products were mixed and denatured with one of the reference samples derived from PCR products of plasmids encoding subtype A–F *env* DNA; after annealing, the samples were loaded onto a 5% polyacrylamide gel, which was electrophoresed at 250 V for 2 hr and 15 min, at constant voltage. Gels were stained with ethidium bromide and the subtype of the test sample determined by comparing the mobilities of heteroduplexes formed with different reference samples.

DNA sequence analysis

All templates were sequenced using primer *gag*3, shown above, as the forward primer and most were also sequenced in the reverse direction, primed by *gag*5 (5′-TTCTAATGTAG-CCC/GCTG-3′). Approximately 0.5 pmol of template DNA was mixed with 4.5 pmol of primer and sequences determined by dye terminator cycling reactions using *Taq* FS in an ABI 377 automated sequencer (Perkin-Elmer) and the Sequence Navigator sequencing programs (Perkin-Elmer).

Phylogenetic analysis

DNA sequences were trimmed to 340 nucleotides and, together with reference sequences from clade A–H and N proviruses,[10] were aligned by the CLUSTAL program.[12] Molecular trees were produced by the fastDNAml program,[13] which is based on the Maximum Likelihood DNAML program[14]; calculations were made with global rearrangements to find the best trees. Bootstrap values were determined for 1000 replicate data sets, using the SEQBOOT, DNADIST, NEIGHBOR, and CONSENSE programs in the PHYLIP 3.52c package of Felsenstein.[15] Additional trees were then produced for clade A sequences from the cohort plus all the clade A sequences present in the Los Alamos/GenBank database in January 1999 and similarly for the clade D sequences from the cohort and the databases. This yielded a total of 36 clade A reference sequences, including 3 from recombinant viruses, which were checked by RIP analysis (see below), and 13 clade D reference sequences. Trees were also produced by the maximum parsimony (PAUP)

TABLE 1. SUMMARY OF COHORT SAMPLES ANALYZED AND GENOTYPE DATA

Gene	Method of analysis	No. of samples: Patient DNA	Amplified	Total	Genotype prevalent (P): incident (I)	Genotype[a]: Clade A (%)	Clade D (%)
gag	Sequencing	109	109	71[b]	33P[b]:38I	25 (36)	44 (64)
env	HMA	118	102	88[c]	32P[c]:56I[c]	37[c] (42)	51[c] (58)

[a]The clade percentages are calculated on the basis of the genotyped samples (69 for *gag* and 88 for HMA).
[b]Includes 66 samples with complete sequence data, 3 samples with incomplete sequence data, and 2 complete sequences that could not be genotyped; see text.
[c]Does not include three samples that gave inconsistent data; see text.

program,[16] using a heuristic search with bootstrap analysis of data sets containing 30 sequences (the maximum allowed by the program). Phylogenetic trees were drawn using the TreeView program.[17]

Analysis of recombinant viruses

All the cohort sequences were examined for the presence of *gag* chimeras by the Recombinant Identification Program (RIP), available online from the Los Alamos web site (http://hiv-web.lanl.gov). The RIP reference sequences were trimmed to match the cohort sequences and the gaps stripped prior to analysis; several window sizes between 80 and 200 nucleotides (nt) were used and the threshold for statistical significance was set at 90%.

Analysis of base substitutions

The Synonymous Nonsynonymous Analysis Program (SNAP) available online from the Los Alamos National Laboratory web site (as given above) was used to analyze pairs of sequences from the four sexual couples for whom sequence data were obtained. SNAP was also used to make pairwise comparisons of all sequences from the cohort and reference samples.

RESULTS

PCR of gag *and* env *regions*

Of 109 DNA samples analyzed, *gag* DNA sequences were amplified from 96 (88%), using the *gag* 1/2 primers followed

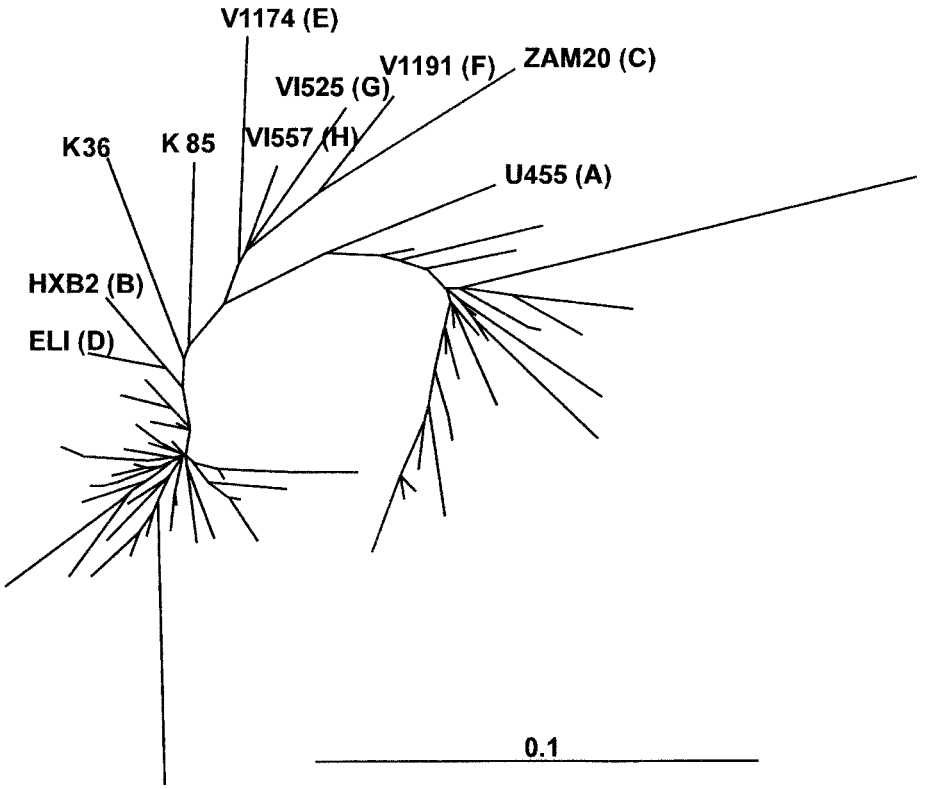

FIG. 1. Newick plot of fastDNAml[13] analysis of p24 *gag* sequence data from the NHC and clade A–H and N references: the tree was plotted using the TreeView program.[17] For greater clarity the NHC sequences are not numbered, except for K36 and K85, neither of which clustered with any of the reference sequences.

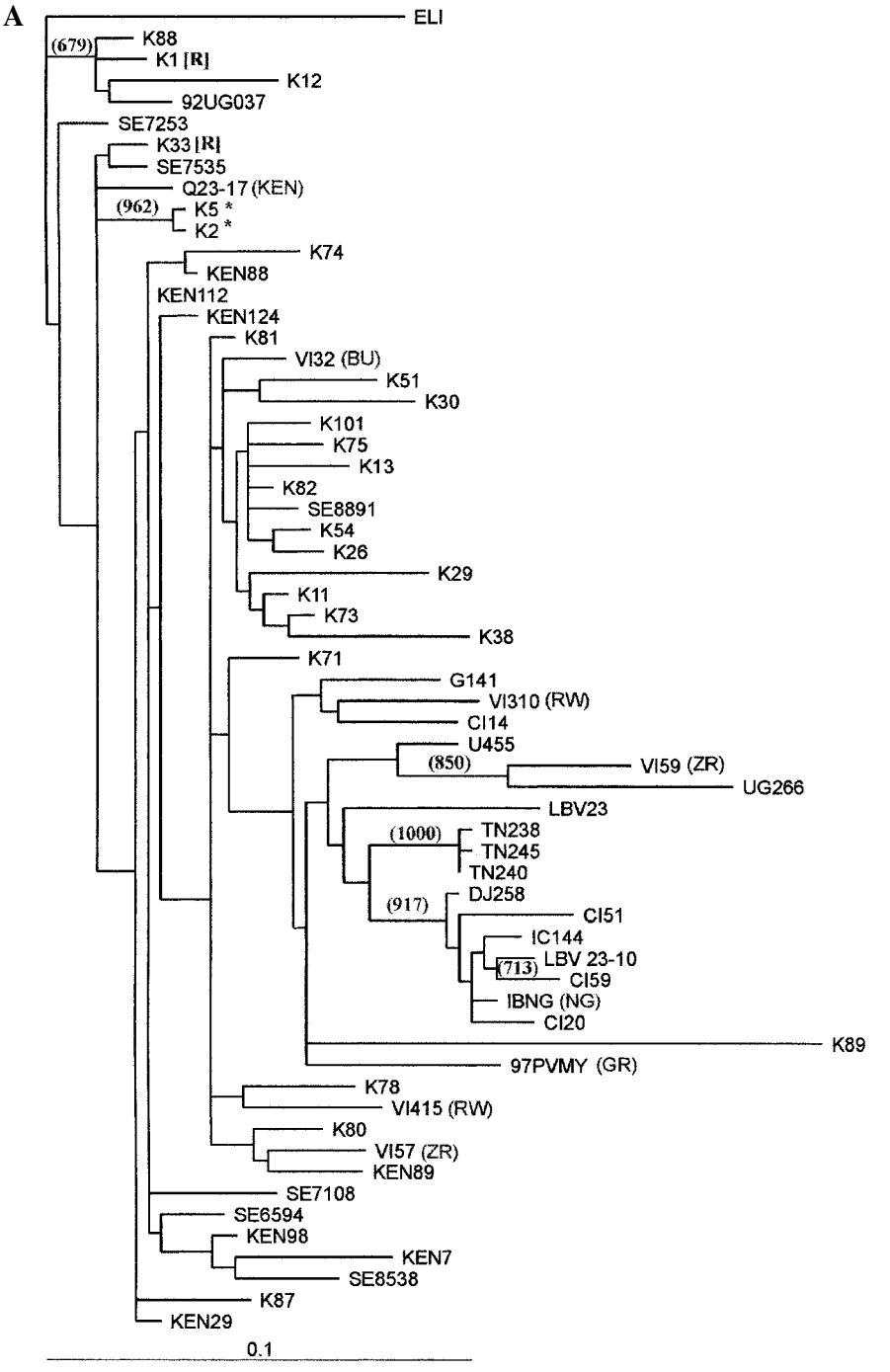

FIG. 2. Phylogenetic analysis of p24 *gag* sequence data from the NHC and reference sequences from the databases. Clade A and D sequences were separately analyzed by the fastDNAml maximum likelihood program. Cohort sequences are denoted by a K prexif; *gag/env* recombinants are indicated as [*R*] and known sexual partners are indicated by pairs of asterisks. The countries of origin of the references sequences are shown as follows, either as prefixes or as suffixes in parentheses: BU, Burundi; CI and IC, Ivory Coast; DJ, Djibuti; G and LBV, Gabon; GR, Greece; KEN, Kenya; NG, Nigeria; RW, Rwanda; SE, Senegal; TN, Thailand; U and UG, Uganda; ZR, Zaire (now the Democratic Republic of Congo). Bootstrap resampling values above or close to 75% from 1000 replicates (see text) are shown in parentheses. (**A**) Clade A sequences, rooted using clade D sequence ELI as an outgroup; (**B**) clade D sequences, rooted using clade A sequence U455 as an outgroup.

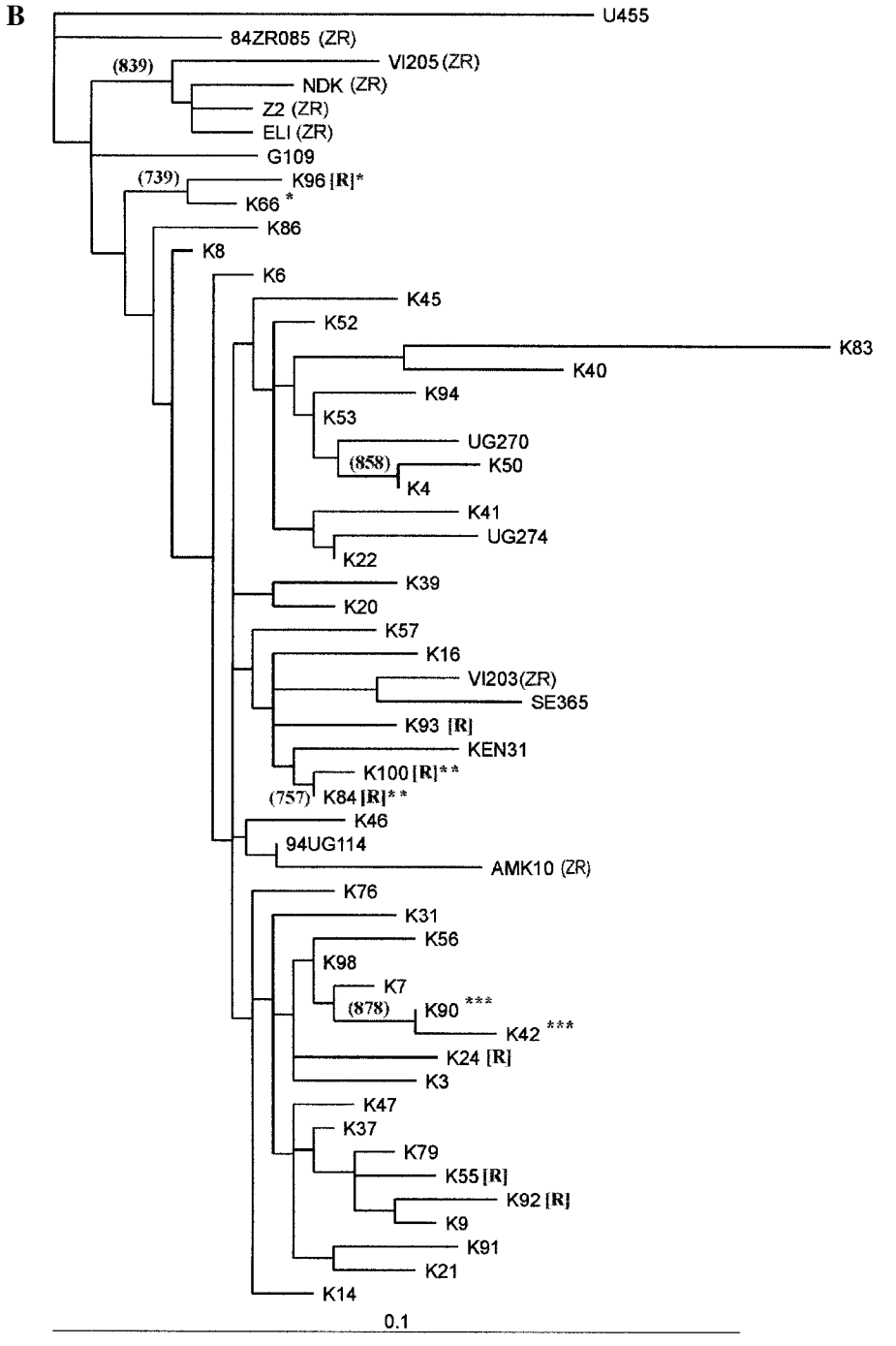

FIG. 2. *Continued.*

by *gag*3/4 primers and an annealing temperature of 55°C: a further 12 samples (11%) were amplified with the same combination of primers with an annealing temperature of 50°C. The remaining sample was amplified with other sets of primers, internal to *gap*3/4, and an annealing temperature of 50°C (data not shown). Of 118 DNA samples analyzed, *env* region DNA sequences were obtained from 99 (84%), using the ED5/12 and ED31/33 primer pairs described above. A further three samples were amplified with ED3/14 as the first primer pair, followed by the ES7/8 primers. Eighty-eight samples gave unambiguous HMA data and a further 3 gave variable results, the amplified DNA comigrating with either clade A or D reference DNAs in different experiments; these samples may have been derived from persons infected with more than one clade of virus but no

other data indicating mixed infections were obtained in this study. Both sets of data—*gag* and *env*—were obtained from a total of 59 DNA samples.

Sequence analysis of the p24 gag region

Sequence data for a 340-nucleotide region of p24 were obtained from 68 (80%) of 85 template samples, which included 7 duplicates obtained from replicate PCRs: the duplicate sequences were either identical or differed by only 1 or 2 nucleotides from the corresponding sequences. More limited data, which were insufficient for phylogenetic analysis but were sufficient to determine proviral genotype, were obtained from a further three templates. The 68 unique templates contained a total of six stop codons resulting from G-to-A substitutions in tryptophan codons: three stop codons were present in provirus K83 and one each in proviruses K26, K30, and K89.

The nucleotide sequence data reported in this article are available from the EMBL, GenBank, and DDBJ Nucleotide Sequence Databases under the accession numbers AJ228255–AJ228322.

Genotype frequencies and genomic recombinants

The numbers and characteristics of the samples analyzed and the results obtained from them are summarized in Table 1. The data obtained by HMA of the V3 *env* region show that 42% of the 88 genotyped proviruses had the clade A genotype and that 58% had the clade D genotype. The ratios obtained for the 69 proviruses genotyped by DNA seuqence and phylogenetic analyses of the p24 *gag* sequences were 36% clade A and 64% clade D. Although no other clades were identified in this cohort, two proviruses could not be genotyped as their sequences did not cluster with any of the references used in the phylogenetic analysis (see below). The clade A and D consensus sequence (data not shown) differed at 28 of 340 (8.2%) positions.

Of the 59 subjects from whom both *gag* p24 and *env* V3 data were obtained, 16 (27%) were p24 *gag* and V3 *env* clade A, 30 (51%) were p24 *gag* and V3 *env* clade D, and 13 (22%) appeared to be clade A/D recombinants. Ten of these were clade D in the p24 *gag* region and clade A in V3 *env*: the other 3 were clade A in p24 *gag* and clade D in V3 *env*. Both partners of a known sexual couple, K84/K100 (see below), were infected with D/A recombinant viruses.

Phylogenetic analysis of gag *sequences*

A Newick plot of a tree obtained by fastDNAml analysis of the *gag* sequences obtained from 68 members of the cohort plus representative sequences from proviruses in clades A to H is shown in Fig. 1. For the sake of clarity the cohort samples are not numbered, with the exception of K36 and K85, which did not cluster with any of the reference sequences (see below). Twenty-five of the 68 sequences clustered with the clade A reference sequence (U455) and 41 with the clade D reference sequence (ELI): as expected from the well-known *gag* sequence similarities between clade B and D proviruses, the clade B reference sequence (HXB2) clustered with the ELI sequence. Several of the clade A and clade D cohort proviruses had diverse sequences that are identified in Fig. 2A and B. A 450-nt region of the p24 region of the *gag* gene of the two provirus sequences that could not be genotyped (K36 and K85) were tested for interclade recombination by RIP analysis. Both gave ambiguous results (data not shown). The K36 sequence appeared to contain several clade A/D recombination sites but as threshold significance was not obtained over the greater part of the sequence, the evidence for recombination within the K36 *gag* gene is weak. The K85 sequence gave ambiguous RIP data, showing few similarities with the reference sequences, regardless of window size. Therefore, the p24 *gag* region of both of these proviruses appeared to be unusual and could repay more extended sequence analyses.

Additional fastDNAml and neighbor-joining trees were produced for clade A and clade D sequences from cohort proviruses plus all the clade A or D reference p24 *gag* sequences currently available in the databases. The fastDNAml trees are shown in Fig. 2A and B as phylograms for clade A and clade D sequences, respectively: the Fig. 2A tree (clade A) is rooted using ELI (clade D) as an outgroup and the FIg. 2B tree (clade D) is rooted using U455 (clade A) as an outgroup. Sequences from *gag/env* recombinant proviruses are indicated by an [R], and bootstrap values above or close to 75% are shown in parentheses.

As shown in the Fig. 2A phylogram, the cohort clade A sequences were obtained in several groups. One cluster comprising cohort sequences K1, K12, and K88 plus the Ugandan reference sequence 96UG037 had a bootstrap value of 68%, which is possibly significant, and although some of the cohort sequences appeared to cluster with sequences from other African countries, e.g., Senegal, Ivory Coast, Rwanda, Burundi, and Kenya, none of these clusterings was supported by bootstrap values above 52% and they are probably not significant. None of the cohort sequences appeared to be related to those from other countries, including Gabon (G), the Central African Republic (LBV), or Nigeria (IGNB). Within the cohort, only one clade A cluster, containing sequences K2 and K5, was supported by a high bootstrap value (96%). These proviruses were obtained from known sexual partners and their sequences were similar, differing by one synonymous (S) and one nonsynonymous (NS) amino acid-changing base substitution. This was the only clade A virus-infected couple from which sequence data were obtained for both partners.

The clade D phylogram (Fig. 2B) shows a similar lack of similarity between the cohort sequences and the reference sequences from various African countries, including Zaire (Democratic Republic of Congo), Gabon, Senegal, and Kenya. More surprisingly, they shared little similarity with reference sequences from other parts of Uganda (UG270, UG274, and 94UG114). Three of the four clade D clusters with high bootstrap values were of pairs of sequences from sexual partners—K42 and K90, K66 and K96, and K84 and K100. These three pairs of sequences were derived from the only known sexual couples from which both sets of clade D sequence data were obtained. The fourth cluster comprising sequences K4 and K50 also had a high bootstrap value, but there was no known sexual contact between these persons, who were 50- and 62-year-old women, but both could have been infected by the same man, who was either not included in the study group or from whom no sequence data were obtained. The K42/K90 sequences differed by 2 S and 2 NS base substitutions, the K66/K96 sequences differed by 3 S and 4 NS substitutions, and the K84/K100 sequences differed by 1 S and 1 NS substitution. The K4/K50 sequences differed by 3 S and 1 NS substitution.

Similar trees were produced by the maximum parsimony (PAUP) method, which also gave bootstrap values over 50% for the five clusters identified by maximum likelihood/bootstrap analysis (data not shown). The PAUP/bootstrap analysis showed an additional clade A cluster, K11/K38/K73, and another clade D cluster, K40/K83; the neighbor-joining/bootstrap values of both of these clusters were about 50% and, as the sequences came from persons with no known sexual contact, their significance is uncertain.

Absence of recombination within the gag *sequences*

As shown in Figs. 1 and 2A and B, some of the cohort clade A and clade D sequences were diverse, differing by between 0.5 and 8.3% for clade A and between 0.5 and 7.9% for clade D proviruses from the respective consensus sequences. To determine whether any of the sequences, especially the outliers, were recombinants within the p24 *gag* gene, all the sequences were analyzed by the RIP. With the possible exception of K36 (see above) no evidence was obtained for recombination within the sequenced region and in almost all cases the match with the clade A or D reference sequence was above the threshold of statistical significance for the greater part of the region (data not shown). The RIP data also confirm the clade assignments shown in Figs. 1 and 2.

Distribution of proviral clades within the study area

The distribution of proviral clades in the study area was determined from the cohort records, which include details of where the participants normally live. The data for V3 *env* (HMA) and p24 *gag* (DNA sequence) are shown in Fig. 3A and B, respectively. The numbers of clade A and D proviruses obtained from each village are shown as the first and second numbers in each pair. Both the V3 *env* and *gag* data show a skew of the clade A proviruses toward the seven eastern villages and of the clade D proviruses toward the eight western villages. Thus, from the V3 *env* data, 22 (59%) of 37 clade A proviruses were obtained from inhabitants of the eastern villages and 37 (73%) of the 51 clade D proviruses were from inhabitants of the western villages. From the *gag* sequence data, 14 (56%) of the 25 clade A proviruses were obtained from people living in the eastern villages and 30 (68%) of the 44 clade D proviruses were obtained from those in the western villages. Therefore there seems to be an unequal distribution of clades in the study area. However, since the distribution between western and eastern villages was made after the data were collected, formal statistical testing is not appropriate.

DISCUSSION

One of our main objectives was to determine the frequencies of viral clades and of *gag/env* recombinant viruses in the study area. Data from both DNA sequence analysis of proviral *gag* and HMA of *env* gene sequences showed a clade A:D ratio of about 0.67:1. A similar clade A:D ratio has been found in a cohort in an adjacent part of Uganda[7] but other studies[5,6] reported a much higher proportion of clade A proviruses in other

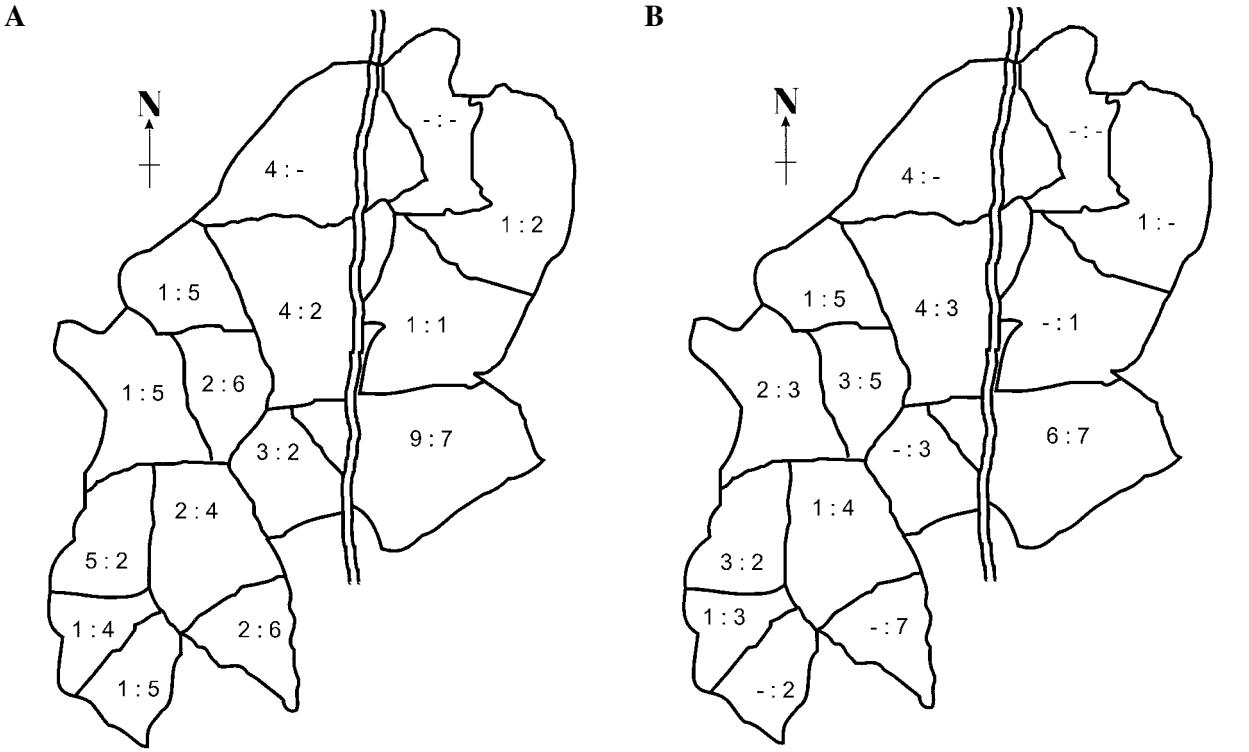

FIG. 3. Village distribution of proviral clades in the NHC study area. The clade A and D proviruses in each village are shown as the first and second numbers, respectively, in each pair. The principal road runs north–south through the center of the area and is shown as a double line. (**A**) V3 *env* data from HMA; (**B**) p24 *gag* sequence data.

parts of Uganda. Rayfield *et al.*[5] found a clade A:D ratio of 1.36:1 in 594 subtyped Ugandan samples, the great majority (93%) of which came from the Kampala/Entebbe area, which is about 120 km from our study area. At present it is not known whether the clade A:D ratio is changing with time, which could indicate differences in the rate of infection within a community or influx from outside. Except in the northernmost villages, both clades were present in all parts of our study area but there was a marked skew of clade D viruses toward the people living in the eastern villages that border the principal road. The distribution of clades within the study area will be investigated in future studies to determine whether there is an influx or more rapid spread of either clade.

In contrast to reports[3,4,6] of other HVI-1 clades in Uganda, none was found in this study or in a study of a neighboring area.[7] However, as two of our cohort sequences did not cluster with any of the reference sequences, they could not be genotyped.

Our analysis of recombination frequency in HIV appears to contain the largest number of samples to date. Of the 59 samples for which both *gag* and *env* data were obtained, 13 (22%) appeared to be recombinants between clade A and D viruses. A similar proportion (20%) of clade A/D recombinants was obtained[18] from a smaller number of Ugandan proviruses in the UNAIDS sample collection. These observed recombination frequencies are considerably higher than the 9% frequency obtained in an analysis[19] of 114 entire genomes in the database. Most of the recombinant viruses in the database originated in Africa and 11 of 12 full-length hybrids examined[20] included segments of clade A viruses, which are widely distributed in Africa. Although one of our *gag/env* recombinants gave an ambiguous result, we obtained no firm evidence for A/D recombination within the relatively short *gag* sequences.

The possibility that many infected cohort members were doubly infected with both clade A and D viruses appears to be unlikely, as a close examination of our sequencing electrophoretograms failed to reveal a mixture of nucleotides at positions where the clade A and D sequences differ. In addition, repeat sampling, PCR, and anlaysis of some of the recombinants confirmed the initial clade assignments for the *gag* and *env* regions. However, repeat HMA analysis of 3 of 91 samples gave conflicting clade A or D heteroduplex mobilities, possibly indicating that these persons were dually infected.

We found that both members of one pair of sexual partners were infected with recombinant viruses, indicating that the recombinants had been transmitted between partners. There have been several reports of the heterosexual[21] or maternal[22] transmission of recombinant viruses between Africans but the relative efficiency of transmission or rates of disease progression of recombinant viruses, compared with nonrecombinants, are not known. As has been noted by others,[19,23] the presence of a high proportion of recombinant viruses could be a complicating factor in vaccine development and assessment: recombination may also have played an important role in the evolution of HIV-1.[24]

We obtained phylogenetic data from 68 persons, a relatively low proportion of the infected population, which can be calculated, assuming a total adult population of 5000 and an 8% HIV prevalence rate,[25] to be about 400 persons.

Since only about 20% of the proviral "pool" was analyzed and we do not have detailed lifetime sexual histories it is not possible to make a detailed molecular epidemiological analysis of HIV in the study population. However, it is of interest to compare our data with a phylogenetic analysis of p17 *gag* sequences from 52 persons from three separate sexual networks in an adjacent part of southwest Uganda,[7] which found no significant correlation between the molecular phylogenetic and sexual contact data. In contrast, we were able to document sexual contact between four of the five pairs of individuals infected with viruses encoding *gag* sequences that clustered with bootstrap resampling values above or close to 75%. Indeed, all four pairs of *gag* sequences from known sexual partners clustered with high bootstrap values. Our cohort, although located in a neighboring area, is drawn from a less mobile rural population with less sexual mixing than those studied by Yirrell *et al.*[7] and this may explain the difference in the results obtained. Like Yirrell *et al.*,[7] we also found a cluster where there was no known sexual connection between the persons concerned. Our data do not support the suggestion[7] that the dynamics of viral spread in Ugandan cohorts are different from those experienced in HIV-infected populations in the developed world.

ACKNOWLEDGMENTS

We gratefully acknowledge the encouragment of Dr. S.D.K. Sempala, Director of UVRI, who facilitated these studies; and the help of Mike Scawen and Andrew Featherstone at CAMR and Rod Page at the University of Glasgow, who assisted with the statistical analysis, DNA sequencing, and phylogenetic analyses. The study was supported by the MRC, the Department for International Development, CAMR, and Westek.

REFERENCES

1. World Health Organization: World Health Organisation Global AIDS statistics. Provisional working estimates of adult HIV prevalence as of the end of 1994 by country. AIDS Care 1996;8:249–254.
2. Serwadda D, Mugerwa RD, Sewenkambo NK, Lwegaba A, Carswell JW, Kirya GB, Bayley AC, Downing RG, Tedder RS, Clayden SA, *et al.*: Slim disease: A new disease in Uganda and its association with HTLV-III infection. Lancet 1985;ii:849–852.
3. Bruce CB, Featherstone ASR, Smith JD, Clegg JCS, and Oram JD: Presence of multiple genetic subtypes of human immunodeficiency virus type 1 proviruses in Uganda. AIDS Res Hum Retroviruses 1994;10:1543–1550.
4. Kaleebu P, Bobkov A, Cheingsong-Popov R, Bieniasz P, Garaev M and Weber J: Identification of HIV-1 subtype G from Uganda. AIDS Res Hum Retroviruss 1995;11:657–659.
5. Rayfield MA, Downing RG, Baggs J, Hu DJ, Pieniazek D, Luo C-C, Biryahwaho B, Otten RA, Sempala SDK, Dondero TJ, and the HIV Variant Working Group: A Molecular epidemiological survey of HIV in Uganda. AIDS 1998;12:521–527.
6. Buonaguro L, Del Huadio E, Monaco M, Greco D, Corti P, Beth-Giraldo F, Buonaguro FM, and Giraldo G: Heteroduplex mobility assay and phylogenetic analysis of V3 region sequences of human immunodeficiency virus type 1 isolates from Gulu, northern Uganda. J Virol 1995;69:7971–7981.

7. Yirrell DL, Pickering H, Palmarini G, Hamilton L, Rutebemberwa A, Biryahwaho B, Whitworth J, and Leigh-Brown AJ: Molecular epidemiological analysis of HIV in sexual networks in Uganda. AIDS 1998;2:285–290.
8. Morgan D, Maude GH, Malamba SS, Okongo MJ, Wagner HU, Mulder DW, and Whitworth JA: An HIV-1 natural history cohort and survival times in rural Uganda. AIDS 1997;11:633–640.
9. Sambrook J, Fritsch EF, and Maniatis T: *Molecular Cloning: A Laboratory Manual,* 2nd ed. Cold Spring Harbor Press, Cold Spring Harbor, New York, 1989.
10. Myers G, Korber B, Hahn BH, Jeang K-T, Mellors JW, McCutchan FE, Henderson LE, and Pavlakis GN, eds.: *Human Retroviruses and AIDS 1995: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences.* Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico, 1995.
11. Delwart EL, Herring B, Rodrigo B, and Mullins JI: Genetic subtyping of human immunodeficiency virus using heteroduplex mobility assay. PCR Methods Appl 1994;4:5202–5216.
12. Higgins DG, Bleasby AJ, and Fuchs R: CLUSTAL V: Improved software for multiple sequence alignment. Comput Appl Biosci 1992;8:189–191.
13. Olsen G, Matsuda H, Hagstrom R, and Overbeek R: fastDNAml: A tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. Comput Appl Biosci 1994;10:41–48.
14. Felsenstein J: Evolutionary trees from DNA sequences: A maximum likelihood approach. J Mol Evol 1981;17:368–376.
15. Felsenstein J: PHYLIP—phylogeny inference package (version 3.2). Cladistics 1989;5:164–166.
16. Swofford DL: *PAUP* (*Phylogenetic Analysis Using Parsimony*). Illinois Natural History Survey, Champaign, Illinois, 1991.
17. Page RDM: TreeView: An application to display phylogenetic trees on personal computers. Comput Appl Biosci 1996;12:357–358.
18. Cornelissen M, Kampinga G, Zorgdrager F, Goudsmit J, and the UNAIDS Network for HIV Isolation and Characterization: Human immunodeficiency virus type 1 subtypes defined by *env* show high frequency of recombinant *gag* genes. J Virol 1996;70:8209–8212.
19. Robertson DL, Sharp PM, McCutchan FE, and Hahn BH: Recombination in HIV-1. Nature (London) 1995;374:124–126.
20. Robertson DL, Gao F, Hahn BH, and Sharp PM: Intersubtype recombinant HIV-1 sequences. In *Human Retroviruses and AIDS 1997: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences* (Korber B, Hahn B, Foley B, Mellors JW, Leitner T, Myers G, McCutchan F, and Kuiken CL, eds.). Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico, 1997, pp. III-25–III-30.
21. Salminen MO, Carr JK, Robertson DL, Hegerich P, Gotte D, Koch C, Sanders-Buell E, Gao F, Sharp PM, Hahn BH, Burke DS, and McCutchen FE: Evolution and probable transmission of intersubtype recombinant human immunodeficiency virus type 1 in a Zambian couple. J Virol 1997;71:2647–2655.
22. Kampinga GA, Simonon A, Van De Perre P, Karita E, Msellati P, and Goudsmit J: Primary infections with HIV-1 of women and their offspring in Rwanda: Findings of heterogeneity at seroconversion, coinfection, and recombinants of HIV-1 subtypes A and C. Virology 1997;27:63–76.
23. Moore J and Anderson R: The WHO and why of HIV vaccine trials. Nature (London) 1994;372:313–314.
24. McCutchan FE, Salminen MO, Carr JK, and Burke DS. HIV-1 genetic diversity. AIDS 1996;10(Suppl. 3):S13–S20.
25. Mulder DW, Nunn AJ, Wagner UH, Kamali A, and Kengeya-Kayondo JF. HIV-1 incidence and HIV-1 associated mortality in a rural Ugandan population. AIDS 1994;8:87–92.

Address reprint requests to:
Pontiano Kaleebu
Medical Research Council Programme for AIDS in Uganda
Uganda Virus Research Institute
P.O. Box 49
Entebbe, Uganda

E-mail: mrc@starcom.co.ug