

An Isolate of Human Immunodeficiency Virus Type 1 Originally Classified as Subtype I Represents a Complex Mosaic Comprising Three Different Group M Subtypes (A, G, and I)

FENG GAO,¹ DAVID L. ROBERTSON,² CATHERINE D. CARRUTHERS,¹ YINGYING LI,¹ ELIZABETH BAILES,³ LEONDIOS G. KOSTRIKIS,⁴ MIKA O. SALMINEN,⁵ FREDERIC BIBOLLET-RUCHE,¹ MARTINE PEETERS,⁶ DAVID D. HO,⁴ GEORGE M. SHAW,^{1,7} PAUL M. SHARP,³ AND BEATRICE H. HAHN^{1*}

Department of Medicine and Microbiology,¹ Howard Hughes Medical Institute,⁷ University of Alabama at Birmingham, Birmingham, Alabama 35294; Laboratory of Structural and Genetic Information, CNRS-EP 91, Marseilles 13402,² and Laboratoire Retrovirus, ORSTOM, Montpellier,⁶ France; Division of Genetics, Queens Medical Centre, University of Nottingham, Nottingham, United Kingdom³; Aaron Diamond AIDS Research Center, The Rockefeller University, New York, New York 10016⁴; and Department of Infectious Diseases, National Public Health Institute, Helsinki 00300, Finland⁵

Received 26 May 1998/Accepted 18 August 1998

Full-length reference clones and sequences are currently available for eight human immunodeficiency virus type 1 (HIV-1) group M subtypes (A through H), but none have been reported for subtypes I and J, which have only been identified in a few individuals. Phylogenetic information for subtype I, in particular, is limited since only about 400 bp of *env* gene sequences have been determined for just two epidemiologically linked viruses infecting a couple who were heterosexual intravenous drug users from Cyprus. To characterize subtype I in greater detail, we employed long-range PCR to clone a full-length provirus (94CY032.3) from an isolate obtained from one of the individuals originally reported to be infected with this subtype. Phylogenetic analysis of C2-V3 *env* gene sequences confirmed that 94CY032.3 was closely related to sequences previously classified as subtype I. However, analysis of the remainder of its genome revealed various regions in which 94CY032.3 was significantly clustered with either subtype A or subtype G. Only sequences located in *vpr* and *nef*, as well as the middle portions of *pol* and *env*, formed independent lineages roughly equidistant from all other known subtypes. Since these latter regions most likely have a common origin, we classify them all as subtype I. These results thus indicate that the originally reported prototypic subtype I isolate 94CY032 represents a triple recombinant (A/G/I) with at least 11 points of recombination crossover. We also screened HIV-1 recombinants with regions of uncertain subtype assignment for the presence of subtype I sequences. This analysis revealed that two of the earliest mosaics from Africa, Z321B (A/G/?) and MAL (A/D/?), contain short segments of sequence which clustered closely with the subtype I domains of 94CY032.3. Since Z321 was isolated in 1976, subtype I as well as subtypes A and G must have existed in Central Africa prior to that date. The discovery of subtype I in HIV-1 hybrids from widely distant geographic locations also suggests a more widespread distribution of this virus subtype, or at least segments of it, than previously recognized.

A molecular epidemiological survey of human immunodeficiency virus type 1 (HIV-1) conducted in Cyprus in 1994 revealed a surprising degree of genetic variation among locally circulating strains (14). Kostrikis and colleagues screened 25 HIV-1-infected individuals attending an outpatient clinic in Nicosia by performing phylogenetic analyses of C2-V3 *env* gene sequences amplified from their uncultured peripheral blood mononuclear cells (PBMCs). This identified representatives of several different HIV-1 group M clades, including subtype A, C, and F strains, that are not commonly found in European populations. Moreover, both members of a heterosexual couple with a history of intravenous drug use and documented travel outside of Cyprus were found to be infected with similar viruses that could not be assigned to any of the previously defined HIV-1 subtypes. These viruses formed an independent lineage roughly equidistant from all other group

M subtypes, and so it was proposed to classify them as members of a new clade, termed subtype I (14).

At about the same time as this initial description of subtype I, it was realized that numerous HIV-1 strains are mosaics of sequences from more than one clade (18, 19). Subsequent confirmation of the widespread occurrence of such hybrid viruses, often with evidence of multiple recombination crossovers along the genome (3, 4, 6, 9, 10, 17, 23), indicated that classification and definition of new subtypes should be based on complete genomic sequences (9, 16, 17). This is particularly important for viruses originating from geographic regions where multiple subtypes cocirculate since these have a high probability of being recombinant. To characterize subtype I in greater detail, we thus cloned a full-length provirus from a short-term-cultured, primary isolate established from one of the two individuals (HO32) infected with this subtype (14). Using primers corresponding to the tRNA primer binding site (5'-TCTCTaccgctGGCGCCCGAACAGGGAC-3', lowercase letters indicate an *Mlu*I site) and the polyadenylation signal in the 3' long terminal repeat (LTR) (5'-ACCAGacgctACAACAGACGGCACACACTACTT-3'), we used long-range PCR

* Corresponding author. Mailing address: Department of Medicine and Microbiology, University of Alabama at Birmingham, 701 S. 19th St., LHRB 613, Birmingham, AL 35294. Phone: (205) 934-0412. Fax: (205) 934-1580. E-mail: bhahn@cordelia.dom.uab.edu.

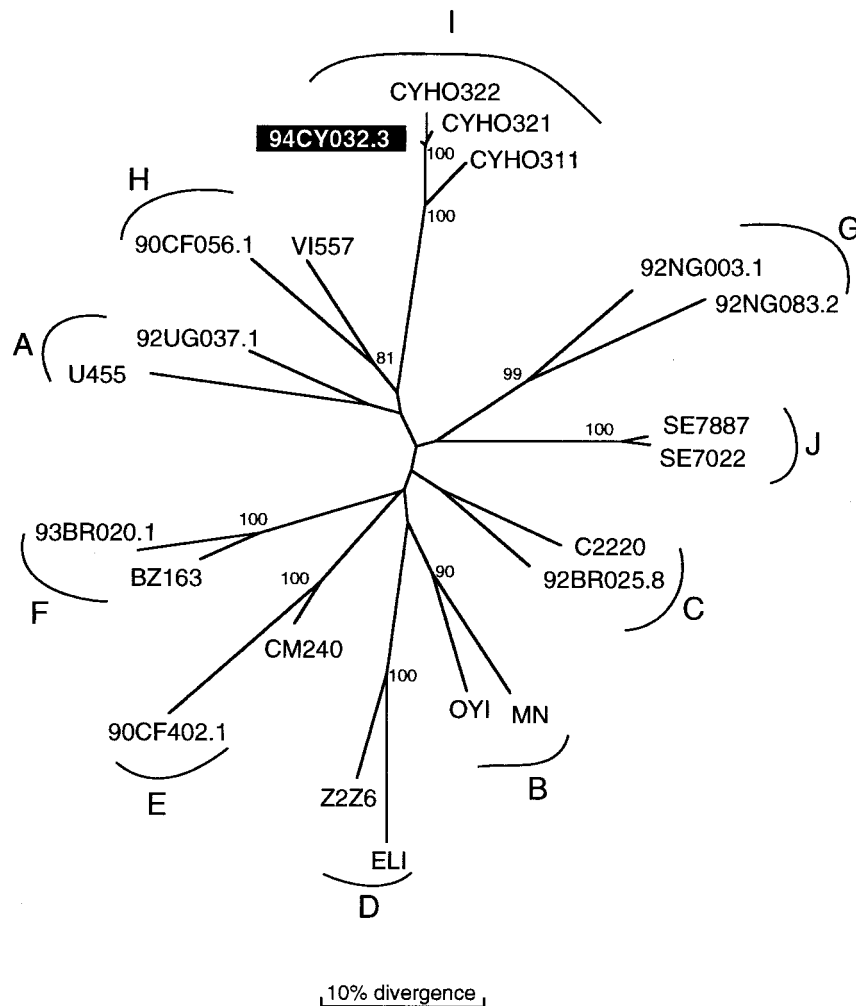


FIG. 1. Confirmation of 94CY032.3 (highlighted) as a subtype I representative in the C2-V3 envelope region. Reference sequences for all known group M subtypes were obtained from the Los Alamos sequence database (13), aligned using CLUSTAL W (26), and adjusted manually using the alignment editor MASE (7). Sites for which there was a gap in any of the sequences were excluded from all sequence comparisons. Pairwise evolutionary distances were estimated by using Kimura's two-parameter method to correct for superimposed substitutions (12). Phylogenetic trees were constructed by the neighbor-joining method (20), and the reliability of topologies was estimated by performing bootstrap analysis using 1,000 replicates (8). Values at nodes indicate the percentage of bootstraps in which the cluster to the right was found (bootstrap values of 80% and higher are shown). Branch lengths are drawn to scale.

to amplify nearly full-length genomic fragments (9, 10, 21) that contained all coding and regulatory regions except for 102 bp of 5' unique LTR sequences (U5). Amplification products were subcloned into a plasmid vector and mapped by restriction enzyme digestion. One clone, termed 94CY032.3, was selected for further analysis. A 694-bp fragment spanning the remainder of the LTR was amplified separately using a semi-nested approach (10).

The complete sequence of 94CY032.3 was determined by the primer-walking approach. Examination of potential coding regions revealed the expected reading frames for *gag*, *pol*, *vif*, *vpr*, *tat*, *rev*, *vpu*, *env*, and *nef* (data not shown). None of the genes contained major deletions, insertions, or rearrangements. However, both *env* and *vif* genes contained single in-frame stop codons. There was also a single-base-pair insertion at position 5199 which caused a frameshift and altered six amino acid residues at the C terminus of the Vpr protein. All other protein domains of known function as well as major regulatory sequences, including the primer binding site, the packaging signal, and major splice sites, appeared to be intact. Similarly, the number, position, and consensus sequences of

promoter and enhancer elements in the 94CY032.3 LTR were indistinguishable from those of most other HIV-1 strains, except for the presence of an unusual TATA sequence (TAAAA), thus far only found in subtype E (A/E) viruses from Thailand and the Central African Republic (4, 10).

To compare 94CY032.3 to previously reported subtype I sequences, we constructed a phylogenetic tree from C2-V3 sequences, including representatives of all 10 known group M subtypes (Fig. 1). As expected, 94CY032.3 clustered most closely with CYHO321 and CYHO322, sequences amplified from uncultured PBMC DNA of the same individual (HO32) from whom the 94CY032 isolate was derived. 94CY032.3 also clustered very closely with CYHO311, a sequence derived from the sexual partner of HO32 (14), strongly suggesting that the two infections were epidemiologically linked. Finally, as observed in the past (14), all subtype I sequences clustered independently, forming a distinct lineage roughly equidistant from all other subtypes, including subtype J (15). These findings thus confirmed the authenticity of the 94CY032.3 clone and validated it as a representative of subtype I in the C2-V3 region of the viral envelope.

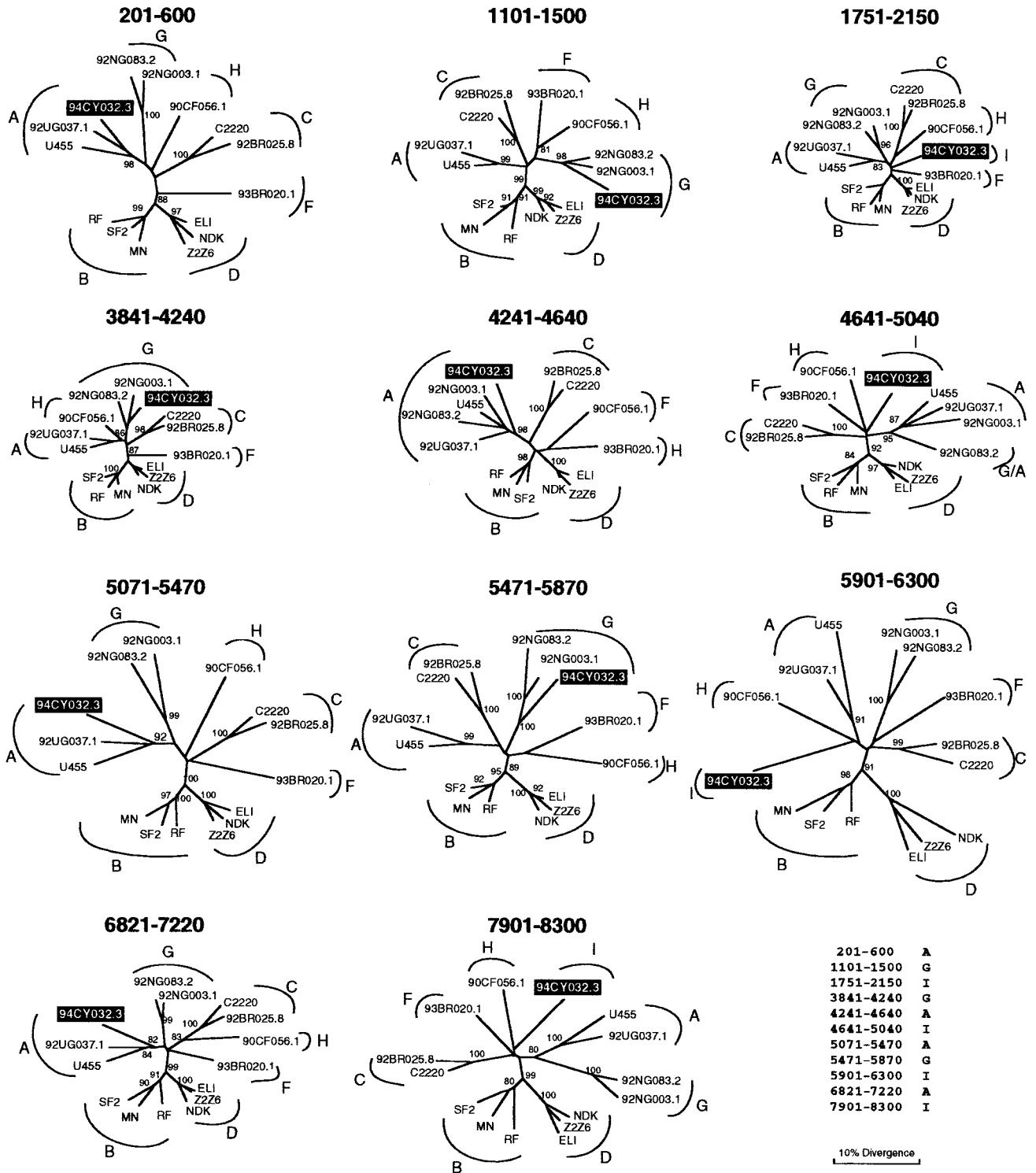


FIG. 2. Exploratory tree analysis. Neighbor-joining trees were constructed for a 400-bp window moved in increments of 10 bp along the multiple genome alignment. Trees depicting discordant branching orders for 94CY032.3 (highlighted) are shown. The position of each tree in the alignment is indicated; subtypes are identified by brackets. Numbers at nodes indicate the percentage of bootstrap values with which the adjacent cluster is supported (only values above 80% are shown). Branch lengths are drawn to scale.

To characterize the remainder of the 94CY032.3 genome, we next performed pairwise sequence comparisons with recently reported nonmosaic reference sequences for subtypes A through H (9, 16) as well as selected intersubtype recombi-

nants (17). We have used this approach in the past to screen newly derived sequences for regions of unusual sequence similarity or dissimilarity that might indicate recombination (9, 10). Briefly, 94CY032.3 was added to a multiple genome align-

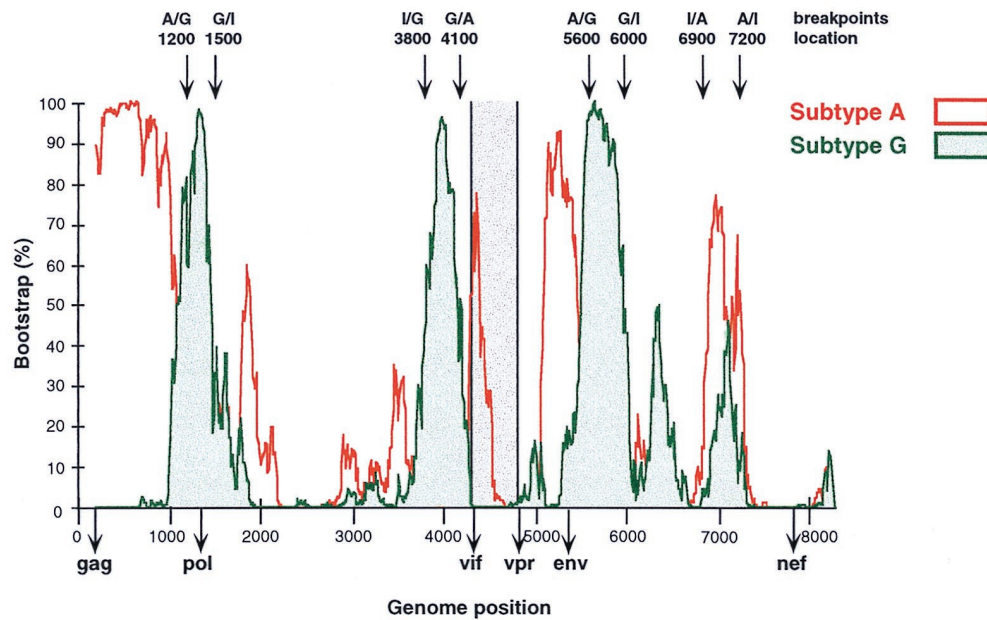


FIG. 3. Bootstrap plot analysis to map recombination breakpoints in 94CY032.3. Bootscanning was performed essentially as described previously (9, 10, 22), plotting the magnitude of the bootstrap value supporting the clustering of 94CY032.3 with 92UG037.1 (subtype A in red) in comparison with that of 94CY032.3 and 92NG083.2 (subtype G in green) for a window of 400 bp moved in increments of 10 bp along the alignment. Regions of subtype A or G origin are identified by very high bootstrap values (>80%). The location of eight recombination crossovers is indicated (see text for details). Breakpoint analysis between positions 4200 and 4800 was not possible due to the recombinant nature of 92NG083.2 (shaded) (9). The beginnings of *gag*, *pol*, *vif*, *vpr*, *env*, and *nef* open reading frames are shown. The y axis indicates the percent bootstrap replicates which support the clustering of 94CY032.3 with representatives of the respective subtypes.

ment (available upon request) which included a total of 28 sequences from the database (13), representing subtypes A (U455 and 92UG037.1), B (LAI, RF, OYI, MN, and SF2), C (C2220 and 92BR025.8), D (NDK, Z2Z3, ELI, 84ZR085.1, and 94UG114.1), F (93BR020.1), and H (90CF056.1) as well as A/C (ZAM184 and 92RW009.6), A/G (92NG083.2, 92NG003.1, Z321, and IBNG), A/D (MAL), A/E (93TH253.3, CM240, and 90CF402.1), and B/F (93BR029.4) recombinants. Simian immunodeficiency virus SIVcpzGAB was included in the alignment to provide an outgroup. All sites with a gap in any of the sequences were removed from the alignment to ensure that all comparisons were made across the same sites. The percent nucleotide sequence diversity between 94CY032.3 and the other viruses was then calculated for sequence pairs by moving a window of 400 bp in steps of 10 bp along the genome. The resulting distance profiles for the various pairwise comparisons were very similar (data not shown), suggesting that 94CY032.3 was roughly equidistant from all other subtypes in most regions of its genome. However, careful inspection of the graphs revealed several small areas of disproportionate sequence similarity involving sequences from subtypes A and G. For example, at the 5' end of *gag* and *vif* and at the 3' and 5' ends of *env*, diversity plots indicated a relative greater similarity of 94CY032.3 to 92UG037.1 and U455. Similarly, at the 3' end of *gag* and the 3' end of *pol*, a relative greater similarity of 94CY032.3 to 92NG083.2 was noticed. Together, these results suggested that 94CY032.3 contained subtype A- and G-like segments, in addition to regions that appeared to be equidistant from the other subtypes. Note that in the absence of a nonmosaic, full-length subtype G genome (9, 16), we used 92NG083.3 as a subtype G reference, although it contains a small A segment in the *vif/vpr* region.

Relative differences in the extent of sequence similarity, as determined by diversity plots (9, 10) or other methods of distance measurement (24), are not always an indicator of recom-

bination, but can reflect variations in the evolutionary rates of the lineages compared. To determine whether 94CY032.3 was truly mosaic, we thus performed an exploratory tree analysis, looking for significantly discordant phylogenetic positions for different parts of its genome (Fig. 2). The multiple genome alignment described above was used, but only three representatives of subtypes B and D were included, and all known recombinants other than the near-full-length subtype G viruses 92NG083.3 and 92NG003.1 were excluded. We constructed unrooted trees for overlapping fragments of 400 bp, moved in 10-bp increments along the alignment. Inspection of the resulting topologies revealed that the branching order of 94CY032.3 changed a total of 10 times, with all of the discordant positions supported by significant bootstrap values. 94CY032.3 alternated between subtype A (Fig. 2, panels 201-600, 4241-4640, 5071-5470, and 6821-7220) and subtype G (Fig. 2, panels 1101-1500, 3841-4240, and 5471-5870), as well as an independent position (Fig. 2, panels 1751-2150, 4641-5040, 5901-6300, and 7901-8300). The 5901-6300 segment corresponds approximately to the C2-V3 region that has served as the basis for the definition of subtype I. It is thus most parsimonious to assume that all four non-subtype A, non-subtype G segments within 94CY032.3 share a common origin and represent subtype I. These analyses indicate that 94CY032.3 is a mosaic of sequences belonging to three different group M subtypes.

To map the boundaries of the putative A, G, and I segments, we performed bootstrap plot analyses as previously described (9, 10, 22), plotting the magnitude of the bootstrap values that supported the clustering of 94CY032.3 with 92UG037.1 (subtype A), as well as that of 94CY032.3 with 92NG083.2 (subtype G). The results of these analyses allowed us to map the approximate location and boundaries of the various subtype A and G segments along the 94CY032.3 genome (Fig. 3). Bearing in mind the window size of 400 nucleotides and considering only peaks of significant bootstrap values (>80%), we identi-

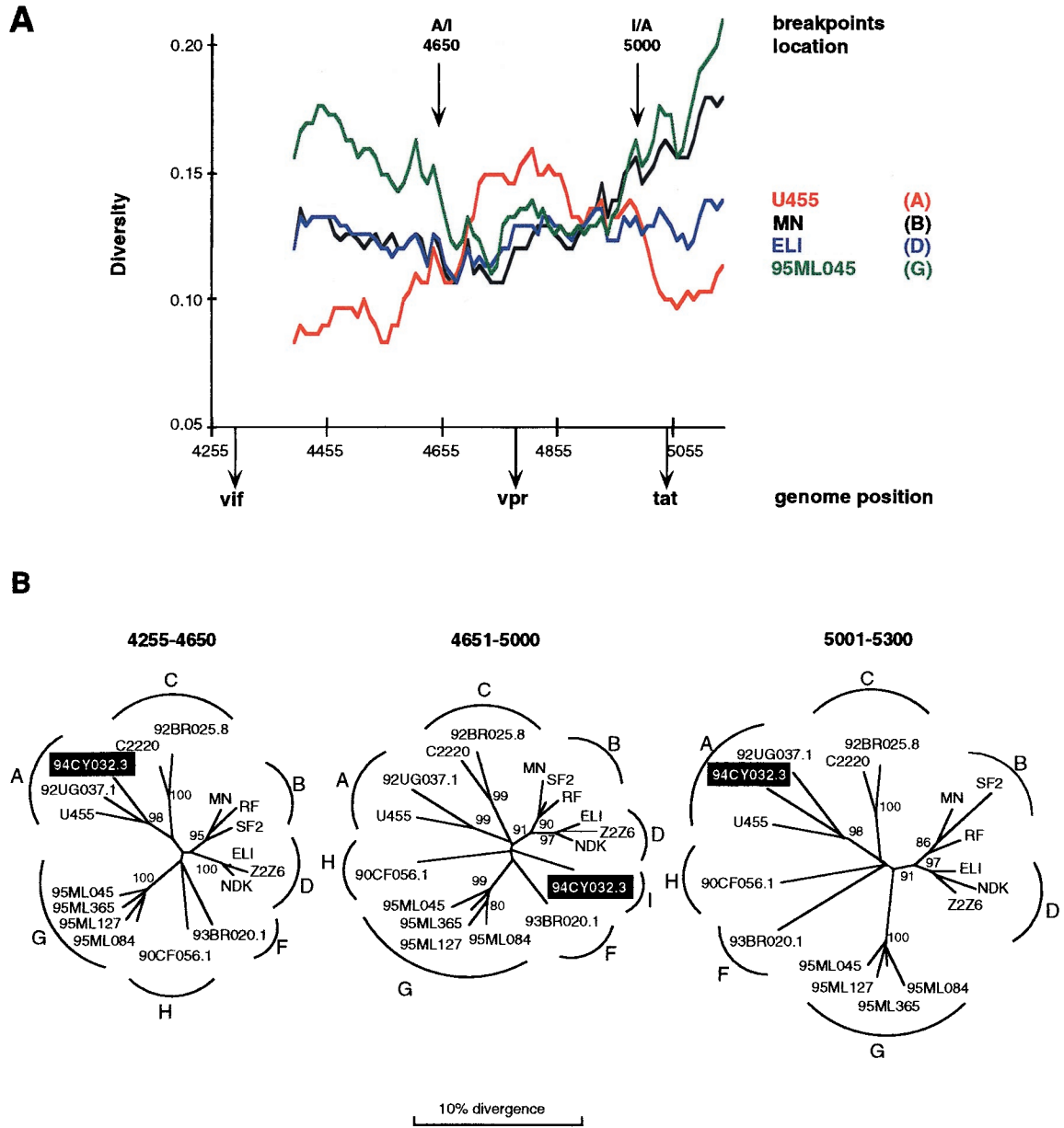


FIG. 4. Recombination breakpoint analysis of 94CY032.3 in the *vif/vpr* region. (A) Diversity plots comparing 94CY032.3 to reference sequences from the database (U455, MN, and ELI) as well as to a recently identified (partial) nonrecombinant subtype G sequence (95ML045) (2). Sequences are color coded. Distance values were calculated for a window of 300 bp moved in steps of 10 nucleotides. The *x* axis indicates the nucleotide positions along the alignment (gaps were stripped and removed from the alignment). The positions of the start codons of the *vif*, *vpr*, and *tat* genes are shown. The *y* axis denotes the distance between the viruses compared (0.05 = 5% difference). (B) Neighbor-joining trees depicting the position of 94CY032.3 in regions flanking the breakpoints identified in panel A. Trees were constructed from the genomic regions indicated. Subtypes are identified by brackets. Four sequences from Mali represent subtype G; these are the only available subtype G reference sequences in this region, since all other characterized subtype G viruses contain A fragments (2, 9, 16, 17). Numbers at nodes indicate the percentage of bootstrap values with which the adjacent cluster is supported; only values above 80% are shown. Branch lengths are drawn to scale.

fied two A/G crossovers around positions 1200 and 5600 and one G/A crossover around position 4100. The bootstrap plots also outlined regions with no peaks or peaks below 80%, which coincided with segments that clustered independently in the exploratory tree analysis and thus were likely in subtype I. Delineating the boundaries of these regions suggested five additional breakpoint positions: G/I at 1500, I/G at 3800, G/I at 6000, I/A at 6900, and A/I at 7200. Because full-length, non-mosaic reference sequences for the parental lineages (G and I) were not available, most of the breakpoints could not be

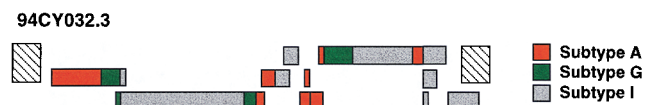


FIG. 5. Inferred structure of the mosaic 94CY032.3 genome. Regions of different subtype origin are color coded. All breakpoints are tentative, since full-length nonrecombinant subtype G and I reference sequences are not available. LTR sequences were not analyzed and are thus shown as hatched boxes.

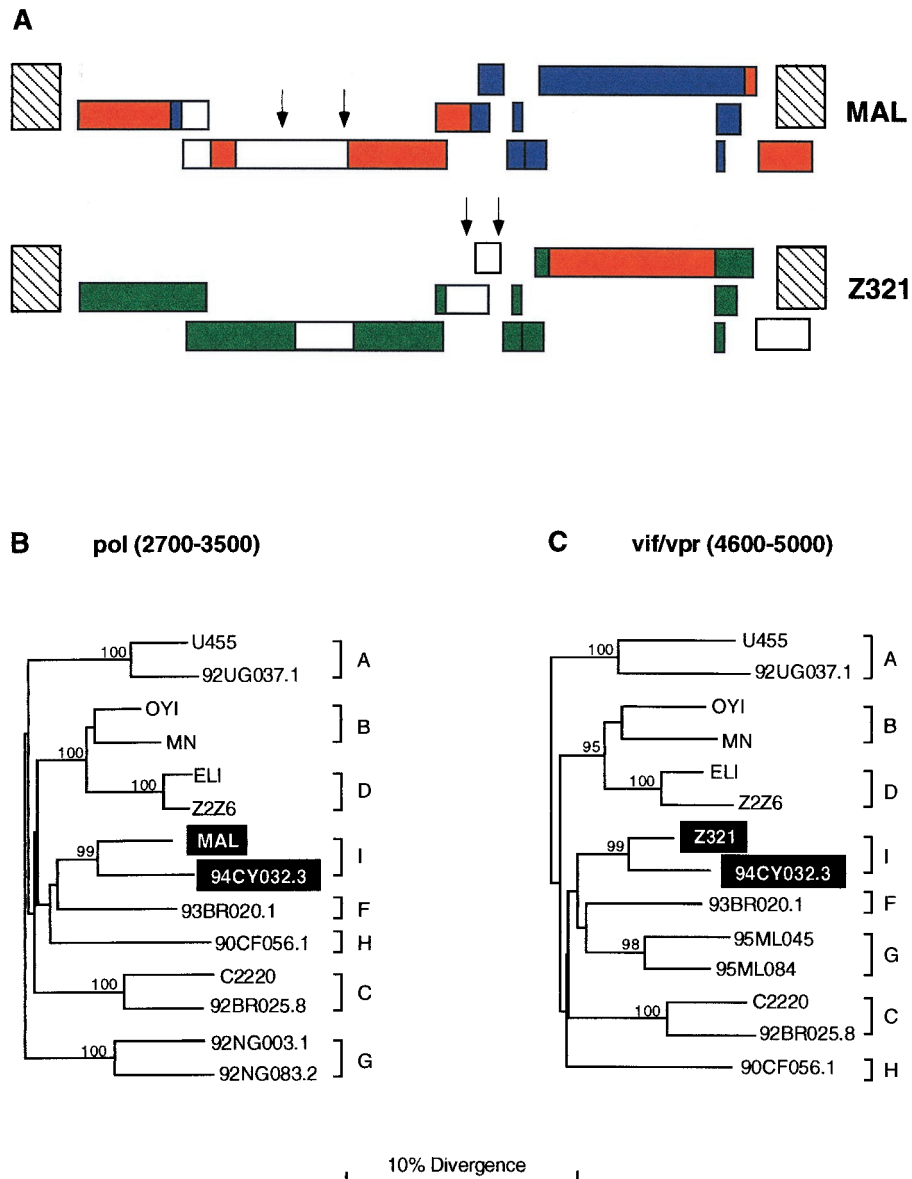


FIG. 6. Screening of intersubtype recombinants containing regions of uncertain subtype assignment for the presence of subtype I sequences. (A) Schematic representation of the recombinant genomes of MAL and Z321. Fragments of subtypes A, D, and G origin are color coded red, blue, and green, respectively; segments of uncertain subtype assignment are shown in white. Arrows denote regions of unusual sequence similarity with 94CY032.3 as determined by diversity plot analysis (data not shown). (B and C) Phylogenetic analysis of the putative subtype I domains of MAL and Z321 delineated in panel A. Neighbor-joining trees of partial *pol* (A) and *vif/vpr* (B) sequences. Horizontal branch lengths are drawn to scale; the scale bar represents 0.1 nucleotide substitutions per site. Vertical separation is for clarity only. Values at the nodes indicate the percent bootstraps in which the cluster to the right was supported; bootstrap values of 80% and higher are shown. Brackets at the right represent the major sequence subtypes of HIV-1 group M. Subtype I clusters are highlighted; trees were midpoint rooted.

mapped with certainty. However, the A/G breakpoints at positions 1200 and 5600 were confirmed by informative site analysis (data not shown). The recombinant nature of 92NG083.2 prohibited reliable breakpoint analysis between positions 4200 and 4800 (9, 16) (highlighted in Fig. 3).

To map the recombination breakpoints in this remaining region, we made use of four recently reported, partial but nonmosaic subtype G sequences from Mali which spanned the *vif/vpr* region and thus bridged the subtype A gap of 92NG083.2 (2). Figure 4A illustrates a set of distance plots that compare 94CY032.3 to one of these newly derived G sequences (95ML045) as well as representatives of subtype A (U455), B (MN), and D (ELI), respectively. Consistent

with the results from the exploratory tree analysis (Fig. 2), 94CY032.3 was disproportionately more closely related to U455 in the 5' and 3' thirds of this fragment, suggesting the presence of subtype A-like segments. However, in the middle of the fragment, 94CY032.3 was clearly equidistant from U455 and the other subtypes, suggesting an independent position. Thus, noting the points at which the "A" distance increased and decreased relative to the other distances allowed us to map the two remaining breakpoints, one near position 4650 and the other near position 5000. Trees constructed from sequences surrounding these two breakpoints (Fig. 4B) confirmed that 94CY032.3 switched position from subtype A (panel 4255-4650) to subtype I (4651-5000) and back to subtype A (5001-

5300). Note that the new subtype G sequences only cover the region between positions 4255 and 5300.

Figure 5 summarizes the results from all phylogenetic analyses, depicting a schematic representation of the mosaic genome structure of 94CY032.3. Segments of different subtype origin are color coded, and there are a total of 10 recombination breakpoints between the 5' end of *gag* and the 3' end of *nef*. LTR sequences were not separately analyzed for mosaicism, but the discordant subtype assignments of the *gag* and *nef* regions necessitate at least one more breakpoint within either the viral LTR or the *gag* leader sequence. Given this extent of mosaic complexity, 94CY032.3 must be the result of multiple recombination events, either in the same or different individuals.

Having identified several fragments of subtype I in 94CY032.3, we next determined whether there was any evidence for its presence in other full-length recombinants from the database. Two known mosaics, MAL (1, 19) and Z321 (5), were of particular interest, because previous analyses had indicated that these viruses contain regions of uncertain subtype assignment (17, 19). For example, MAL has long been known to represent a mosaic of subtypes A and D (color coded as red and blue, respectively, in Fig. 6A) but also contains a sizable *pol* fragment that has defied previous subtype classification (white in Fig. 6A) (17, 19). Similarly, Z321 is a known mosaic of subtypes A and G (red and green, respectively, in Fig. 6A) (5), but a recent reanalysis of its recombination breakpoints also identified regions that could not be assigned to any known subtype (17). To determine whether any of these regions represented subtype I, we again performed distance plot analysis, looking for dips in the diversity profiles of 94CY032.3, with MAL and Z321 as an indication of relatively greater sequence similarity. Indeed, two such dips were identified, one in the *pol* region of MAL and another in the *vif/vpr* region of Z321, both of which coincided with previously unclassified segments of their genomes (data not shown). Subsequent phylogenetic tree analysis confirmed that these regions were indeed of subtype I origin, since MAL and Z321 clustered with very high bootstrap values with the subtype I domains of 94CY032.3 (Fig. 6B and C). However, subtype I did not account for all of the unclassifiable regions in MAL and Z321 (17), and it thus remains unclear whether these represent still other, as yet unidentified, subtypes or regions of multiple breakpoints that cannot be mapped by using current methods.

In summary, we show here that a strain of HIV-1, proposed in 1995 as a prototypic subtype I isolate (14), in fact represents a complex mosaic comprising subtypes A, G, and I, respectively. We also show that two of the oldest known isolates from Africa, MAL isolated in 1984 (1) and Z321 isolated in 1976 (11, 25), contain short segments of sequence closely related to the subtype I domains of 94CY032.3. These findings support the following conclusions. (i) Although initially detected in Cyprus, subtype I must have existed in Africa as early as 1976; it is unknown whether full-length, nonmosaic representatives of subtype I still exist but have not yet been sampled or whether this subtype is represented only by fragments in present-day recombinants. (ii) The ancestry of 94CY032.3 has likely involved multiple recombination events; it remains unclear whether these occurred in Africa and/or in Cyprus, where a number of different subtypes have also been documented (14). (iii) Subtype I, along with subtypes A and G, must have diverged substantially earlier than the 1970s in order to be detectable as a distinct segment in the Z321 genome; this is consistent with the recent molecular characterization of a virus from 1959 which in phylogenetic analyses appears to have postdated the group M radiation (27). (iv) Finally, the finding

of subtype I in several different recombinants, including one from an intravenous drug user (14), suggests that this subtype may be more widespread than previously thought, at least in the form of mosaic genome fragments. It will be interesting to screen additional viruses from drug user populations and their contacts in Cyprus and Greece to determine the current prevalence and geographic distribution of subtype I-containing viruses.

Nucleotide sequence accession numbers. The complete sequence of 94CY032.3 has been deposited at GenBank under accession no. AF049337 and AF049338.

We thank F. E. McCutchan for unpublished sequence information, Y.-L. Chen and B. Jian for excellent technical assistance, and W. L. Abbott for artwork and preparation of the manuscript.

This work was supported by grants from the National Institutes of Health (N01 AI 35170 and R01 AI 25291), by shared facilities of the UAB Center for AIDS Research (DNA Sequencing Core; P30 AI27767), and by the Birmingham Veterans Administration Medical Center.

REFERENCES

1. Alizon, M., S. Wain-Hobson, L. Montagnier, and P. Sonigo. 1986. Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. *Cell* **46**:63-74.
2. Bibollet-Ruche, F., C. Mulanga-Kabeya, J. L. Perret, E. E. Williams, E. Mpoudi-Ngole, S. Saragosti, C. Brengues, E. Delaporte, and M. Peeters. Unpublished data.
3. Bobkov, A., R. Cheingsong-Popov, M. Salminen, F. McCutchan, J. Louwagie, K. Ariyoshi, H. Whittle, and J. Weber. 1996. Complex mosaic structure of the partial envelope sequence from a Gambian HIV type 1 isolate. *AIDS Res. Hum. Retroviruses* **12**:169-171.
4. Carr, J. K., M. O. Salminen, C. Koch, D. Gotte, A. W. Arntstein, P. A. Hegerich, D. St. Louis, D. S. Burke, and F. E. McCutchan. 1996. Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *J. Virol.* **70**:5935-5943.
5. Choi, D. J., S. Dube, T. P. Spicer, H. B. Slade, F. C. Jensen, and B. J. Poiesz. 1997. HIV type 1 isolate Z321, the strain used to make a therapeutic HIV type 1 immunogen, is intersubtype recombinant. *AIDS Res. Hum. Retroviruses* **13**:357-361.
6. Cornelissen, M., G. Kampinga, F. Zorgdrager, J. Goudsmit, and the UNAIDS Network for HIV Isolation and Characterization. 1996. Human immunodeficiency virus type 1 subtypes defined by *env* show high frequency of recombinant *gag* genes. *J. Virol.* **70**:8209-8212.
7. Faulkner, D. M., and J. Jurka. 1988. Multiple aligned sequence editor (MASE). *Trends Biochem. Sci.* **13**:321-322.
8. Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**:783-791.
9. Gao, F., D. L. Robertson, C. D. Carruthers, S. G. Morrison, B. Jian, Y. Chen, F. Barré-Sinoussi, M. Girard, A. Srinivasan, A. G. Abimiku, G. M. Shaw, P. M. Sharp, and B. H. Hahn. 1998. A comprehensive panel of near full-length clones and reference sequences for non-subtype B isolates of human immunodeficiency virus type 1. *J. Virol.* **72**:5680-5698.
10. Gao, F., D. L. Robertson, S. G. Morrison, H. Hui, S. Craig, J. Decker, P. N. Fultz, M. Girard, G. M. Shaw, B. H. Hahn, and P. M. Sharp. 1996. The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *J. Virol.* **70**:7013-7029.
11. Getchell, J. P., D. R. Hicks, A. Srinivasan, J. L. Heath, D. A. York, M. Malonga, D. N. Forthal, J. M. Mann, and J. B. McCormick. 1987. Human immunodeficiency virus isolated from a serum sample collected in 1976 in Central Africa. *J. Infect. Dis.* **156**:833-837.
12. Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**:111-120.
13. Korber B. T. M., B. Foley, C. Kuiken, T. Leitner, F. McCutchan, J. W. Mellors, and B. H. Hahn. 1997. Human retroviruses and AIDS: a compilation and analysis of nucleic acid and amino acid sequences. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, N.M.
14. Kostrikis, L. G., E. Bagdades, Y. Cao, L. Zhang, D. Dimitriou, and D. D. Ho. 1995. Genetic analysis of human immunodeficiency virus type 1 strains from patients in Cyprus: identification of a new subtype designated subtype I. *J. Virol.* **69**:6122-6130.
15. Leitner, T., and J. Albert. 1995. A new genetic subtype of HIV-1, p. III-147-III-150. In G. Myers, B. Korber, B. H. Hahn, K.-T. Jeang, J. W. Mellors, F. E. McCutchan, L. E. Henderson, and G. N. Pavlakis (ed.), *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, N.M.
16. **Leitner, T., B. T. M. Korber, D. L. Robertson, F. Gao, and B. H. Hahn.** 1997. Updated proposal of reference sequences of HIV-1 genetic subtypes, p. III-19–III-24. *In* B. Korber, B. Foley, C. Kuiken, T. Leitner, F. McCutchan, J. W. Mellors, and B. H. Hahn (ed.), *Human retroviruses and AIDS 1997: a compilation and analysis of nucleic acid and amino acid sequences*. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, N.M.
 17. **Robertson, D. L., F. Gao, B. H. Hahn, and P. M. Sharp.** 1997. Intersubtype recombinant HIV-1 sequences, p. III-25–III-30. *In* B. Korber, B. Foley, C. Kuiken, T. Leitner, F. McCutchan, J. W. Mellors, and B. H. Hahn (ed.), *Human retroviruses and AIDS 1997: a compilation and analysis of nucleic acid and amino acid sequences*. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, N.M.
 18. **Robertson, D. L., P. M. Sharp, F. E. McCutchan, and B. H. Hahn.** 1995. Recombination in HIV-1. *Nature* **374**:124–126.
 19. **Robertson, D. L., B. H. Hahn, and P. M. Sharp.** 1995. Recombination in AIDS viruses. *J. Mol. Evol.* **40**:249–259.
 20. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
 21. **Salminen, M. O., C. Koch, E. Sanders-Buell, P. K. Ehrenberg, N. L. Michael, J. K. Carr, D. S. Burke, and F. E. McCutchan.** 1995. Recovery of virtually full length HIV-1 provirus of diverse subtypes from primary virus cultures using the polymerase chain reaction. *Virology* **213**:80–86.
 22. **Salminen, M. O., J. K. Carr, D. S. Burke, and F. E. McCutchan.** 1995. Identification of breakpoints in intergenotypic recombinants of HIV-1 by bootscanning. *AIDS Res. Hum. Retroviruses* **11**:1423–1425.
 23. **Salminen, M. O., J. K. Carr, D. L. Robertson, P. Hegerich, D. Gotte, C. Koch, E. Sanders-Buell, F. Gao, P. M. Sharp, B. H. Hahn, D. S. Burke, and F. E. McCutchan.** 1997. Evolution and probable transmission of intersubtype recombinant human immunodeficiency virus type 1 in a Zambian couple. *J. Virol.* **71**:2647–2655.
 24. **Siepel, A. C., A. L. Halpern, C. Macken, and B. T. M. Korber.** 1995. A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences. *AIDS Res. Hum. Retroviruses* **11**:1413–1416.
 25. **Srinivasan, A., D. York, D. J. Butler, R. Jannoun-Nasr, J. Getchell, J. McCormick, C.-Y. Ou, G. Myers, T. Smith, E. Chen, G. Flaggs, P. Berman, G. Schochetman, and S. Kalyanaraman.** 1989. Molecular characterization of HIV-1 isolated from a serum collected in 1976: nucleotide sequence comparison of recent isolates and generation of hybrid HIV. *AIDS Res. Hum. Retroviruses* **5**:121–129.
 26. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W—improving the sensitivity of progressive multiple sequence alignment through sequence weighing, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
 27. **Zhu, T., B. T. Korber, A. J. Nahmias, E. Hooper, P. M. Sharp, and D. D. Ho.** 1998. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* **391**:594–597.