

# Evidence of Two Distinct Subsubtypes within the HIV-1 Subtype A Radiation

FENG GAO,<sup>1</sup> NICOLE VIDAL,<sup>2</sup> YINGYING LI,<sup>1</sup> STANLEY A. TRASK,<sup>1</sup> YALU CHEN,<sup>1</sup> LEONDIOS G. KOSTRIKIS,<sup>3</sup> DAVID D. HO,<sup>3</sup> JINWOOK KIM,<sup>4</sup> MYOUNG-DON OH,<sup>4</sup> KANGWON CHOE,<sup>4</sup> MIKA SALMINEN,<sup>5</sup> DAVID L. ROBERTSON,<sup>6</sup> GEORGE M. SHAW,<sup>1,7,8</sup> BEATRICE H. HAHN,<sup>1,7</sup> and MARTINE PEETERS<sup>2</sup>

## ABSTRACT

Members of HIV-1 group M are responsible for the vast majority of AIDS cases worldwide and have been classified on the basis of their phylogenetic relationships into nine roughly equidistant clades, termed subtypes. Although there are no known phenotypic correlates for these genotypes, the disproportionate spread of certain of these lineages has been taken to indicate that subtype-specific biological differences may exist. The subtype nomenclature thus remains an important molecular epidemiological tool with which to track the course of the group M pandemic. In this study, we have characterized HIV-1 strains described previously as unusual subtype A variants on the basis of partial sequence analysis. Six such strains from Cyprus (CY), South Korea (KR), and the Democratic Republic of Congo (CD) were PCR amplified from infected cell culture or patient PBMC DNA, cloned, and sequences in their entirety (94CY017, 97KR004, 97CDKTB48, and 97CDKP58) or as half genomes (97CDKS10 and 97CDKFE4). Distance and phylogenetic analyses showed that four of these viruses (94CY017, 97CDKTB48, 97CDKFE4, and 97CDKS10) were closely related to each other, but quite divergent from all other HIV-1 strains, except for subtype A viruses, which represented their closest relatives. In phylogenetic trees from *gag*, *pol*, *env*, and *nef* regions, the four newly characterized HIV-1 strains formed a distinct sister clade to subtype A, which was as closely related to subtype A as subsubtypes F1 and F2 are to each other. According to current nomenclature rules, this defines a subsubtype, which we have tentatively termed A2. The two other viruses, 97KR004 and 97CDKP58, as well as a full-length HIV-1 sequence from the sequence database (ZAM184), were found to represent complex A2/D, A2/G, and A2/C recombinants, respectively. These results indicate that HIV-1 subtype A is composed of two subsubtypes (A1 and A2), both of which appear to have a widespread geographic distribution. The A2 viruses described here represent the first reference reagents for this new group M lineage.

## INTRODUCTION

**G**LOBALLY CIRCULATING STRAINS of human immunodeficiency virus type 1 (HIV-1) exhibit an extraordinary degree of genetic diversity and have been classified on the basis of their phylogenetic relationships into distinct viral lineages,

termed groups, subtypes, and subsubtypes.<sup>1-3</sup> Groups refer to the distinct HIV-1 lineages M (main), N (non-M/non-O), and O (outlier), which are the result of three independent cross-species (zoonotic) transmissions of chimpanzee viruses (SIVcpz) into the human population.<sup>4,5</sup> Members of these groups are highly divergent from each other, differing on av-

<sup>1</sup>Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama 35294.

<sup>2</sup>Laboratoire Retrovirus, IRD, 34032 Montpellier, France.

<sup>3</sup>Aaron Diamond AIDS Research Center, Rockefeller University, New York, New York 10016.

<sup>4</sup>Department of Internal Medicine, Seoul National University College of Medicine and Clinical Research Institute, Seoul National University Hospital, Seoul 110-744, South Korea.

<sup>5</sup>Department of Infectious Disease, National Public Health Institute, FIN-00300 Helsinki, Finland.

<sup>6</sup>Department of Zoology, University of Oxford, Oxford, OX1 3P3 UK.

<sup>7</sup>Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama 35294.

<sup>8</sup>Howard Hughes Medical Institute, University of Alabama at Birmingham, Birmingham, Alabama 35294.

erage in 30% and 47% of their Gag and Env protein sequences, respectively. Subtypes refer to the major clades within HIV-1 group M, which each probably resulted from the expansion of founder viruses early in the group M epidemic.<sup>6</sup> Nine such clades, termed A–D, F–H, J, and K, have thus far been identified, and these differ on average in 15% and 22% of Gag and Env protein sequences, respectively.<sup>1–3,7,8</sup> Most recently, a subtype classification scheme has been proposed in order to distinguish HIV-1 group M viral lineages that form sister clades to established subtypes, but are not divergent enough to justify a new subtype designation.<sup>2,3,9</sup> Examples are subsubtypes F1 and F2 (Refs. 8 and 9) as well as “subtypes” B and D, which were initially misclassified, but have not been renamed as subsubtypes to avoid confusion with the existing literature.<sup>2,3</sup> Members of the different subsubtypes differ on average in 10% and 17% of their Gag and Env protein sequences, respectively.

The classification of HIV-1s into phylogenetically distinct lineages is not only of academic interest, but has provided a powerful tool with which to track the course of the global AIDS epidemic.<sup>8,10–17</sup> Molecular epidemiological studies have shown that there is considerable heterogeneity in the prevalence and geographic distribution of the various groups, subtypes, and subsubtypes.<sup>6,11,13,16,18,19</sup> For example, infections with groups N and O viruses have been largely limited to persons from West–Central Africa, and on a global scale have accounted for only a small fraction of AIDS cases.<sup>19–21</sup> By contrast, HIV-1 group M strains have infected millions of people worldwide and have spread to more or less every country on the globe.<sup>1,10</sup> Moreover, within group M there is considerable variation in the dissemination of the various subtypes. For example, subtype C viruses have spread to three different continents and are believed to represent about 56% of all circulating group M infections, while subtype J viruses have thus far been described only in a handful of individuals.<sup>7,10,22</sup> Finally, even among subsubtypes there are geographical preferences. Subsubtype F1 has been found in Brazil, Romania, and Cameroon,<sup>8,9,23,24</sup> while subsubtype F2 appears to be restricted to Cameroon.<sup>8,9</sup> The reasons for these differences in geographic dispersal are unknown and could simply reflect epidemiological chance events. However, it is also possible that group- or subtype-specific viral properties, including differences in infectivity, transmissibility, and immunogenicity, play a role. A relationship between group M subtypes and natural resistance against antiretroviral drugs has been observed,<sup>25,26</sup> as well as between subtypes and the efficiency of serological and molecular tests for HIV diagnosis.<sup>27–30</sup> In the absence of natural history data, it will thus be important to continue to survey the global distribution of HIV-1 groups and subtypes.

Most HIV-1 strains group consistently in phylogenetic trees, indicating that they fall into the same group, subtype, or subsubtype regardless of which region of their genomes is analyzed. However, increasingly HIV-1 strains are identified that are mosaic, that is, represent the product of recombination events.<sup>1,31,32</sup> Numerous intersubtype recombinants have been described, some of which have demonstrated their biological fitness by becoming major circulating recombinant forms (CRFs). Examples include CRF01\_AE and CRF02\_AG, which circulate epidemically throughout Southeast Asia and Africa, respectively.<sup>11,12,15</sup> Moreover, recombination events have generated complex hybrids that are composed of multiple group M

subtypes,<sup>31,32</sup> as well as hybrids between HIV-1 groups M and O.<sup>33,34</sup> Such viruses could be envisioned to acquire different biological properties, including drug resistance, altered tropism, and enhanced virulence. Such viruses could also evade serologic detection or even susceptibility to vaccines that are based on one particular virus subtype or group. The frequency by which recombinants are generated and their impact on the global epidemic thus remain an important aspect of molecular epidemiological monitoring.

Subtype A has been estimated to constitute 23% of all circulating group M subtypes and is most prevalent in Africa.<sup>10</sup> Moreover, subtype A portions have been identified in numerous recombinants, including in all but one<sup>35</sup> circulating recombinant form (CRF\_01-CRF\_04)<sup>11,36–39</sup> and both intergroup (M/O) hybrids.<sup>33,34</sup> Thus, both nonrecombinant and recombinant subtype A strains appear to be of molecular epidemiological importance. In this article, we have molecularly characterized HIV-1 strains from Cyprus, South Korea, and the Democratic Republic of Congo (DRC; formerly Zaire) that were previously classified as subtype A variants on the basis of partial sequence analysis (Refs. 16 and 40, and our unpublished data). Phylogenetic analysis of two nearly full-length and two half-genomes revealed the existence of a new HIV-1 group M lineage that clustered as a sister clade to prototypic subtype A viruses. Sequences belonging to this new clade were also found in two newly derived full-length recombinants from the DRC and South Korea, and a previously reported Zambian strain (ZAM184).<sup>40a</sup> These results indicate that subtype A is composed of at least two distinct sublineages. In accordance with a recent HIV-1 nomenclature proposal,<sup>2,3</sup> we propose to call these subsubtypes A1 and A2.

## MATERIALS AND METHODS

### *HIV-1 strains*

The HIV-1 strains characterized in this study were either derived from primary isolates (94CY017 and 97KR004) or amplified from patient peripheral blood mononuclear cells (PBMCs) without interim culture (97CDKTB48, 97CDKFE4, 97CDKS10, and 97CDKP58). Available epidemiological information about the study subjects is summarized in Table 1. Isolate 94CY017 has been reported previously.<sup>40</sup> Briefly, this HIV-1 strain was obtained in 1994 from a 35-year-old female AIDS patient (HO17) from Nicosia, Cyprus. Her virus was characterized as part of a molecular epidemiological study of HIV-1 in Cyprus, which also identified her husband and child as well as the husband's sexual partner as infected with HIV-1. The infection route of the index case is unknown, but may have occurred in the United Kingdom, where the family lived for some time.<sup>40</sup> Isolate 97KR004 was obtained from a 33-year-old female sex worker who lived in Yosoo, a major seaport in South Korea. She was first diagnosed with clinical AIDS in February 1997 (CD4<sup>+</sup> cell count <30 cells/mm<sup>3</sup>) and died in October 1997. She had multiple sexual contacts with sailors from many continents, including Africa. A virus isolate obtained by cocultivation of patient PBMCs with normal donor PBMCs yielded a syncytium-inducing (SI) phenotype on cocultivation with MT-2 cells. Blood samples from subjects 97CD-

TABLE 1. EPIDEMIOLOGICAL INFORMATION FOR HIV-1 STRAINS CHARACTERIZED IN THIS STUDY

<i>HIV-1 strain</i>	<i>City</i>	<i>Country</i>	<i>Year</i>	<i>Preliminary subtype classification</i>	<i>Material</i> <sup>a</sup>	<i>Genomic region</i>	<i>Length (bp)</i>	<i>GenBank Acc. No.</i>	<i>Ref.</i>
94CY017	Nicosia	Cyprus	1994	Acy	Primary isolate	Near full-length genome	9060	AF286237	40
97KR004	Yosoo	South Korea	1997	A variant	Primary isolate	Near full-length genome	9064	AF286239	NA
97CDKTB48	Kinshasa	DRC	1997	A variant	Uncultured PBMCs	Near full-length genome	8972	AF286238	16
97CDKFE4	Kinshasa	DRC	1997	A variant	Uncultured PBMCs	<i>gag/pol</i>	5100	AF286240	16
97CDKS10	Kinshasa	DRC	1997	A variant	Uncultured PBMCs	<i>pol/nef</i>	5323	AF286241	16
97CDKP58	Kinshasa	DRC	1997	A variant	Uncultured PBMCs	Near full-length genome	8843	AF316544	16

*Abbreviations:* DRC, Democratic Republic of Congo; NA, not available.

<sup>a</sup>Genomic DNA was either extracted from uncultured patient peripheral blood mononuclear cells (PBMCs) or from primary PBMC cultures established by cocultivation with normal donor PBMCs.

KTB48, 97CDKFE4, 97CDKS10, and 97CDKP58 were collected as part of a molecular epidemiological survey conducted in April 1997 in the Democratic Republic of Congo (CD).<sup>16</sup> All four individuals were residents of Kinshasa (K). Sample 97CD-KTB48 was derived from a tuberculosis (TB) patient, 97CD-KFE4 from an asymptomatic pregnant woman (FE), 97CDKS10 from an AIDS patient (S), and 97CDKP58 from a female sex worker (P). Blood samples were obtained before the outbreak of the civil war in the DRC in 1997 and the arrival of foreign troops from other African countries. The four HIV-1 strains from Kinshasa thus reflect strains that were endemic in the DRC before 1997.

### PCR amplification and subcloning

Genomic DNA was extracted from short-term PBMC cultures (94CY017 and 97KR004) or from uncultured patient PBMCs (97CDKTB48, 97CDKFE4, 97CDKS10, and 97CDKP58). A near full-length genome of 94CY017 was obtained by long polymerase chain reaction (PCR) amplification using techniques and primer pairs described previously.<sup>39,41</sup> For 97KR004, 97CDKTB48, and 97CDKP58, complete genome sequences were obtained by amplifying four overlapping PCR fragments. For 97CDKFE4 and 96CDKS10, half-genome sequences were amplified spanning *gag/pol* and *pol/nef* regions, respectively. PCR primers and conditions are available on request. PCR products were visualized by agarose gel electrophoresis, purified with a Qiagen PCR purification kit (Valencia, CA), and directly cloned into vector pCR2.1 or pCR XL-TOPO (In Vitrogen, Carlsbad, CA).

### Sequence analysis

Amplification products were sequenced by the primer walking method. Sequences were obtained for both strands of DNA, using an automatic sequencer (ABI 377; Applied Biosystems, Foster City, CA) and contigs were assembled and edited with Sequencer 3.1 (Gene Codes, Ann Arbor, MI). GenBank accession numbers for all newly characterized sequences are listed in Table 1.

### Subtype distance tool

Genetic distance comparisons were performed with the Subtype Distance Tool (SUDI) available at the Los Alamos National Laboratory website (<http://hiv-web.lanl.gov>). This online tool creates plots of pairwise distance values between strains of unknown subtype classification and a set of reference sequences. Query sentences (labeled as U) and reference sequences (labeled with their subtype or subtype designation) are entered as an alignment file, which is then used to construct a neighbor-joining tree using the F84 model of nucleotide substitution.<sup>42</sup> The program generates within-subtype, between-subsubtype, and between-subtypedistances for the reference sequences and indicates which subtype is the closest relative to the query sequences. Pairwise distances are generated by summing up all branch lengths between end-point taxa in the tree. We used SUDI to generate plots independently for *gag*, *pol*, and *env* nucleotide sequences, using the following reference sequences: subtype A (Q23CXCCG, SOSE7253, TZSE8538, UGSE6594, UGSE7535, UGSE8891, 92UG037, U455, and

UG273A); subtype B (HAN, CAM1, JH32, MN, NY5CG, P896, RF, SF2, and YU2); subtype C (94BR025, BW96-BW0502, DJ259A, ETH2220, 301904, 301905, 301999, SE364A, and UG268A2); subtype D (SE365A2, 94UG1141, C971412, UG266A2, UG274A2, 84ZR085, JY1, NDK, and Z2Z6); subsubtype F1 (93BR020.1, BZ126A, BZ163A, and FIN9363); subsubtype F2 (MP255 and MP257); subtype G (DRCBL, HH8793, LBV217, 92NG083, and SE6165); subtype H (90CF056, VI991, and VI997); subtype J (SE91733 and SE92809); subtype K (EQTB11 and MP535); and an outgroup sequence (SIVcpzGAB1). As recommended,<sup>2,3</sup> genetic distances between subtype B and D, as well as subsubtypes F1 and F2, were used as intersubsubtype references.

### Phylogenetic analysis

Phylogenetic relationships of the newly characterized viruses were estimated from sequence comparisons with previously reported HIV-1 subtype and subsubtype reference sequences from the Los Alamos database.<sup>2,3,43</sup> Sequences were added to an existing master alignment, using the profile alignment option of CLUSTAL W (Thompson *et al.*<sup>44</sup>) and adjusted manually when necessary with the alignment editor MASE (Faulkner and Jurka<sup>45</sup>). Sites where there was a gap in any of the sequences, as well as areas of uncertain alignment, were excluded. Pairwise evolutionary distances were estimated by using the Kimura two-parameter model.<sup>46</sup> Phylogenetic trees were constructed by the neighbor-joining method,<sup>47</sup> and the reliability of topologies was estimated by performing bootstrap analysis with 1000 replicates. Phylogenetic relationships were also determined by using the maximum likelihood approach, implemented with the program DNAML from the PHYLIP package.<sup>48</sup> Treecool was used to visualize the phylogenetic trees (<ftp://ftp.cme.msu.edu/pub/RDP/programs/TreeTool/>).

### Diversity plot and bootstrap plot analysis

Sequence alignments were also used for diversity and bootstrap plot analyses as described previously.<sup>39,41</sup> The percent diversity between selected pairs of sequences was determined by moving a window of 400 bp along the genome alignment in 10-bp increments. The diversity values for each pairwise comparison were plotted against the midpoint of the window used along the genome. Bootstrap plots were performed on neighbor-joining trees by using SEQBOOT, DNADIST (with the Kimura two-parameter correction), NEIGHBOR, and CONSENSE from the PHYLIP package<sup>48</sup> for a window of 400 bp moved along the alignment in 50-bp increments. Only a subset of reference sequences was used to reduce computing times. One thousand bootstrap replicates were evaluated for each phylogeny. The program ANALYZE from the bootscanning package<sup>49</sup> was used to examine the clustering of a putative inter-subtype hybrid with representatives of the subtypes presumed to have been involved in the recombination event. The bootstrap values for these sequences were plotted at the midpoint of each window along the genome.

### Exploratory tree analysis

Exploratory tree analysis was performed by the bootstrap plot approach described above, except in this case representatives

of all nearly full-length, nonrecombinant subtype reference sequences were used and the increment for each sliding window was 100 bp. Each neighbor-joining tree was viewed with Treetool.

#### Informative site analysis

To estimate the location and significance of recombination cross-overs, each putative hybrid sequence was compared with a representative of each of the two subtypes inferred to have been involved in the recombination event, and an appropriate outgroup. Recombination breakpoints were mapped by examining the linear distribution of phylogenetically informative sites supporting the clustering of the hybrid with each of the two parental subtypes as previously described.<sup>32,41</sup>

## RESULTS

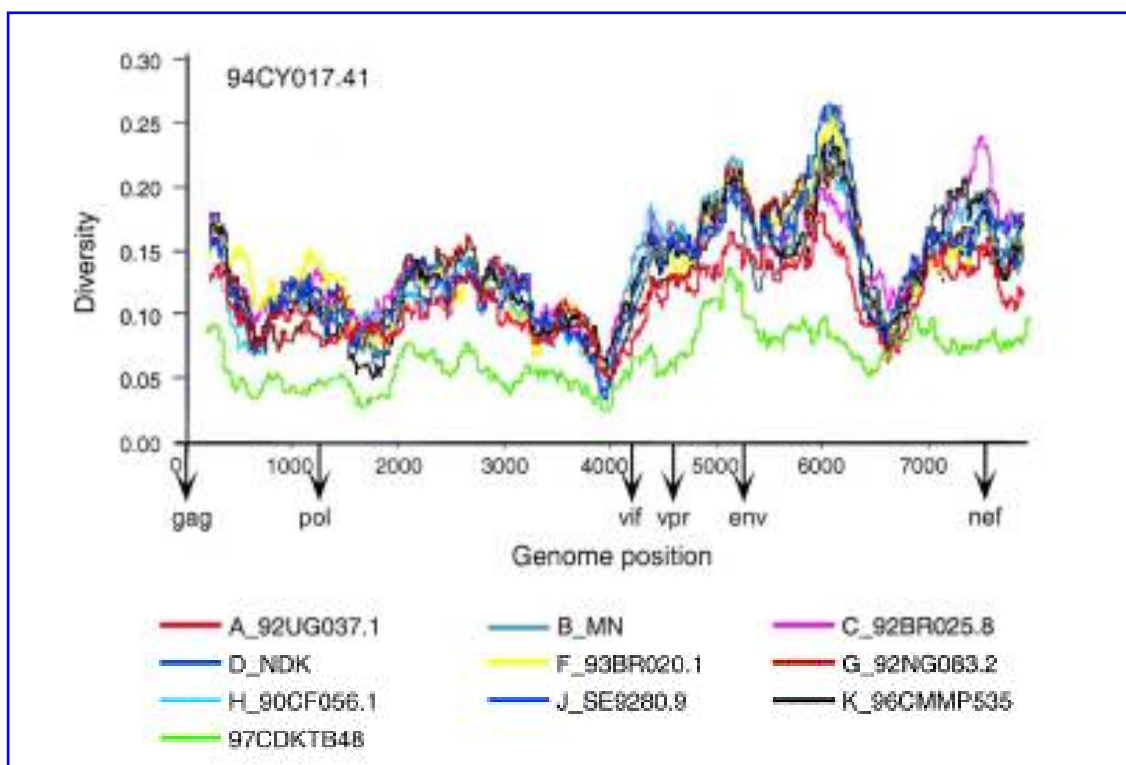
#### Molecular cloning and sequence analysis of six subtype A variants

The purpose of this study was to molecularly characterize HIV-1 strains previously classified as subtype A variants on the basis of partial sequence analysis (Refs. 16 and 40, and our unpublished data). Six such strains were available for study (Table 1): two as primary isolates (94CY017 and 97KR004) and four as uncultured patient PBMCs (97CDKTB48, 97CDKFE4,

97CDKS10, and 97CDKP58). Four of these were completely sequenced, one after amplification and cloning of a nearly full-length genome by long PCR methods (94CY017.41) and three after amplifying and cloning overlapping PCR fragments (97KR004, 97CDKTB4, and 97CDKP58). The remaining two strains were characterized by amplifying and sequencing half-genomes. A 5100-bp fragment spanning *gag* and *pol* was amplified for 97CDKFE4, and a 5323-bp fragment spanning the region between *pol* and *nef* was amplified for 97CDKS10 (amplification of a complete genome equivalent was not possible because of limited patient material). Inspection of the deduced protein sequences of major gene products revealed that all clones contained inactivating mutations due to substitutions, deletions, or insertions. Alignment of Env protein sequences also identified an unusual cysteine residue in the cytoplasmic domain of gp41 (position 774 in the HXB2 Env sequence) of 97CDKTB48, 94CY017.41, 97CDKS10, 97KR004, and 97CDKP58. This additional cysteine residue was also present in three other previously characterized Env protein sequences (ZAM184, ZAM174, and Z321), but was absent from the remaining 210 Env sequences in the sequence database.

#### Identification of a new subsubtype within the HIV-1 subtype A radiation

The newly characterized sequences were first compared with reference sequences from the database, using a diversity plot

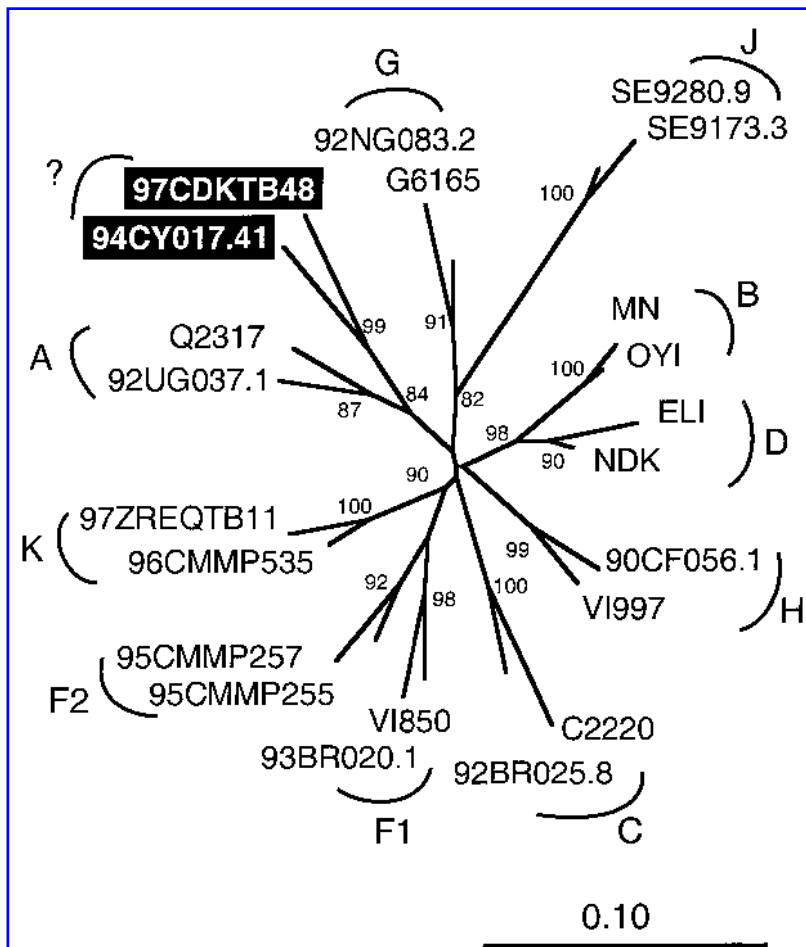


**FIG. 1.** Diversity plots comparing the sequence relationships of two newly characterized nearly full-length subtype A variants (94CY017.41 and 97CDKTB48) with each other and with reference sequences from the Los Alamos HIV Sequence Database. 94CY017.41 is compared with sequences listed at the bottom (reference sequences for subtypes A–D, F–H, J, and K were obtained from the database<sup>1</sup>; all sequences are color coded). Distance values were calculated for a window of 400 bp moved in steps of 10 nucleotides along the genome. The x axis indicates the nucleotide positions along the alignment (gaps were stripped from the alignments). The positions of the start codons of the *gag*, *pol*, *vif*, *vpr*, *env*, and *nef* genes are shown. The y axis denotes the distance between the viruses compared (0.05 = 5% divergence).

approach. This provided a distance measure as well as a first screen for evidence of recombination. Figure 1 depicts an example of such an analysis for the full-length Cyprus isolate 94CY017.41. The results show that over the entire length of its genome, 94CY017.41 is very distant from all database reference sequences, except for A\_92UG037.1, to which it appears to be relatively more closely related. By contrast, 94CY017.41 is closely related to 97CDKTB48, a second newly characterized strain from the DRC, and these two strains exhibit a distance profile typically observed for members of the same subtype. These results suggested that 97CDKTB48 and 94CY017.41 represent nonmosaic members of a new lineage that is relatively more closely related to subtype A than to any of the other subtypes. Similar results were also observed for 97CDKFE4 and 97CDKS10 half-genomic sequences (data not shown). By contrast, 97KR004 and 97CDKP58 exhibited diversity profiles that strongly suggested that they are recombinants. These two full-length sequences were thus analyzed separately (see below).

To further define the phylogenetic position of the new sequences, exploratory tree analysis was performed, using an

alignment that contained the two new full-length "A-like" sequences 97CDKTB48 and 94CY017.41, as well as two reference sequences for each known group M subtype (A, B, C, D, F, G, H, J, and K) and the subsubtypes F1 and F2, respectively. Phylogenetic trees were constructed for windows of 400 bp (moved in 100-bp increments along the alignment) and examined (an example of such a tree spanning the region between nucleotides [nt] 2101 and 2500 is shown in Fig. 2). This analysis showed that in the great majority of regions, the two newly derived viruses clustered together and formed a distinct monophyletic clade. Moreover, this clade formed a sister clade to subtype A, similar to F1 and F2, and B and D clades. In some small regions in *gag* and *pol*, other subtype sequences joined these sister clades; however, even in these regions, 94CY017.41 and 97CDKTB48 were still more closely related to prototypic subtype A sequences than to those from other subtypes. Exploratory tree analyses including the two half-genome sequences showed that 97CDKFE4 and 97CDKS10 also clustered with 94CY017.41 and 97CDKTB48 with significant bootstrap values. Thus, all four of the newly characterized viruses formed

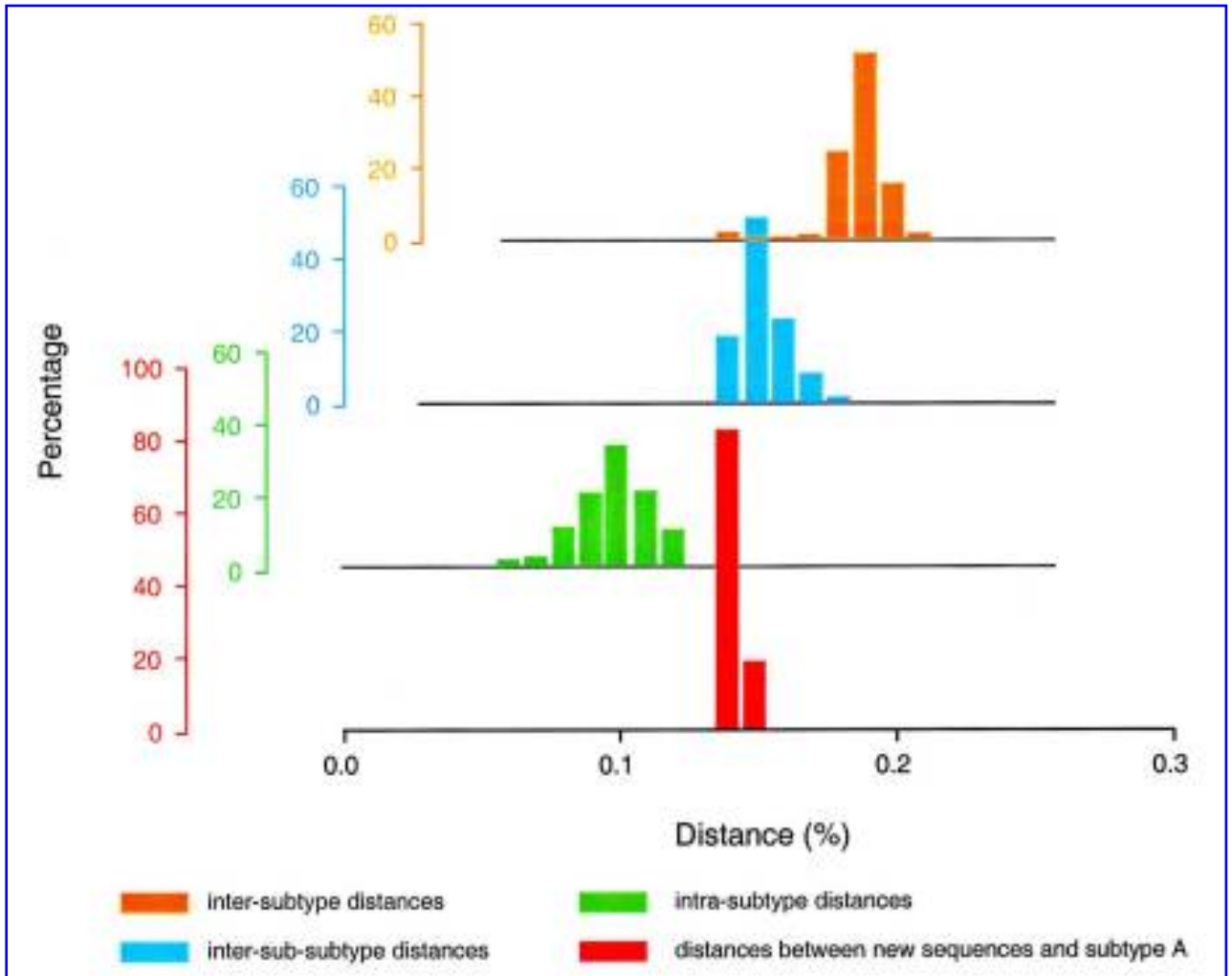


**FIG. 2.** Exploratory tree analysis. Neighbor-joining trees were constructed for 400-bp windows moved in increments of 100 bp along a multiple genome alignment including the HIV-1 sequences indicated. A representative tree from the *pol* region is shown (positions 2101–2500). Brackets identify known subtypes and subsubtypes; a question mark denotes the lineage formed by the newly characterized sequences (boxed). Numbers at the nodes indicate the percentage of bootstrap values with which the adjacent cluster is supported (only values of 80% or higher are shown). Branch lengths are drawn to scale. The scale bar represents 0.10 nucleotide substitution per site.

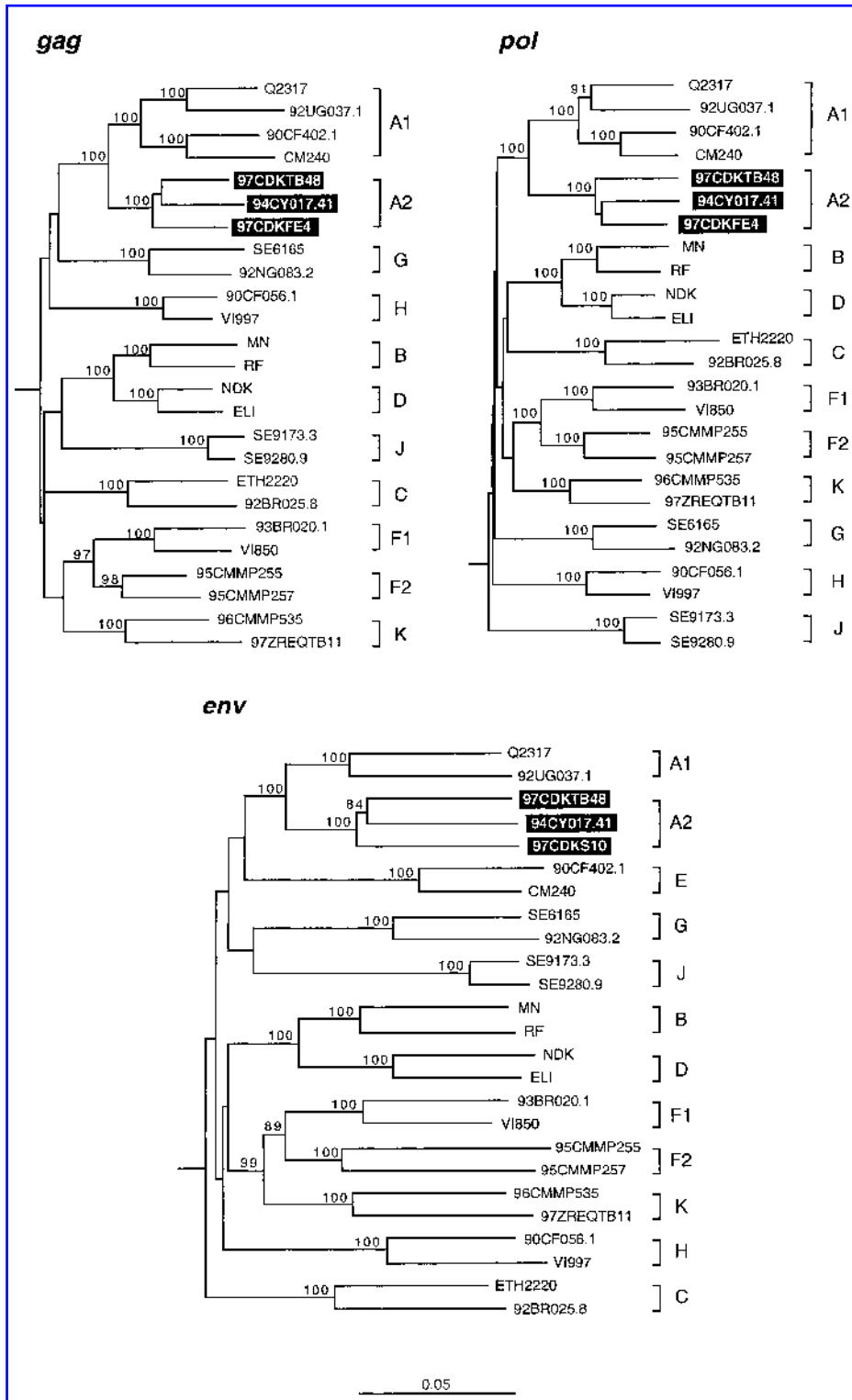
a new lineage within the group M radiation that was relatively more closely related to subtype A.

To determine whether this new lineage represented a new subtype or subsubtype, we used the SUDI (Subtype Distance) program developed specifically for this purpose by investigators at the Los Alamos National Laboratory.<sup>2,3</sup> This subtype distance tool, which is available on the Los Alamos website (<http://hiv-web.lanl.gov>), generates plots of distance measurements between a set of query and reference sequences. Figure 3 depicts an output of this program for *env* nucleotide sequence distances. Consistent with previous reports,<sup>2,3</sup> analysis of the reference set yielded intersubtype distances (plotted in orange) that ranged from 18 to 20%, intersubsubtype distances (plotted in blue) that ranged from 14 to 18%, and intrasubtype distances (plotted in green) that ranged from 5 to 12%. For three newly

derived sequences for which *env* sequences were available (97CDKTB48, 94CY017.41, and 97CDKS10), subtype A was identified as the closest relative. As expected, distances between the three query sequences and subtype A sequences fell within the intrasubtype range (data not shown). However, distance values derived from comparisons between the new sequences and prototypic subtype A sequences fell within the subsubtype range (14 to 15%; plotted in red in Fig. 3). Similar results were also obtained from analysis of *gag* and *pol* sequences, although *pol* distances between the newly characterized and subtype A sequences were slightly higher (i.e., closer to intersubtype distances; data not shown). These data suggested that the newly derived HIV-1 sequences comprised a subsubtype within the subtype A radiation. Of note, SUDI analysis of B/D and F1/F2 genetic distances yielded similar results, except that *env* dis-

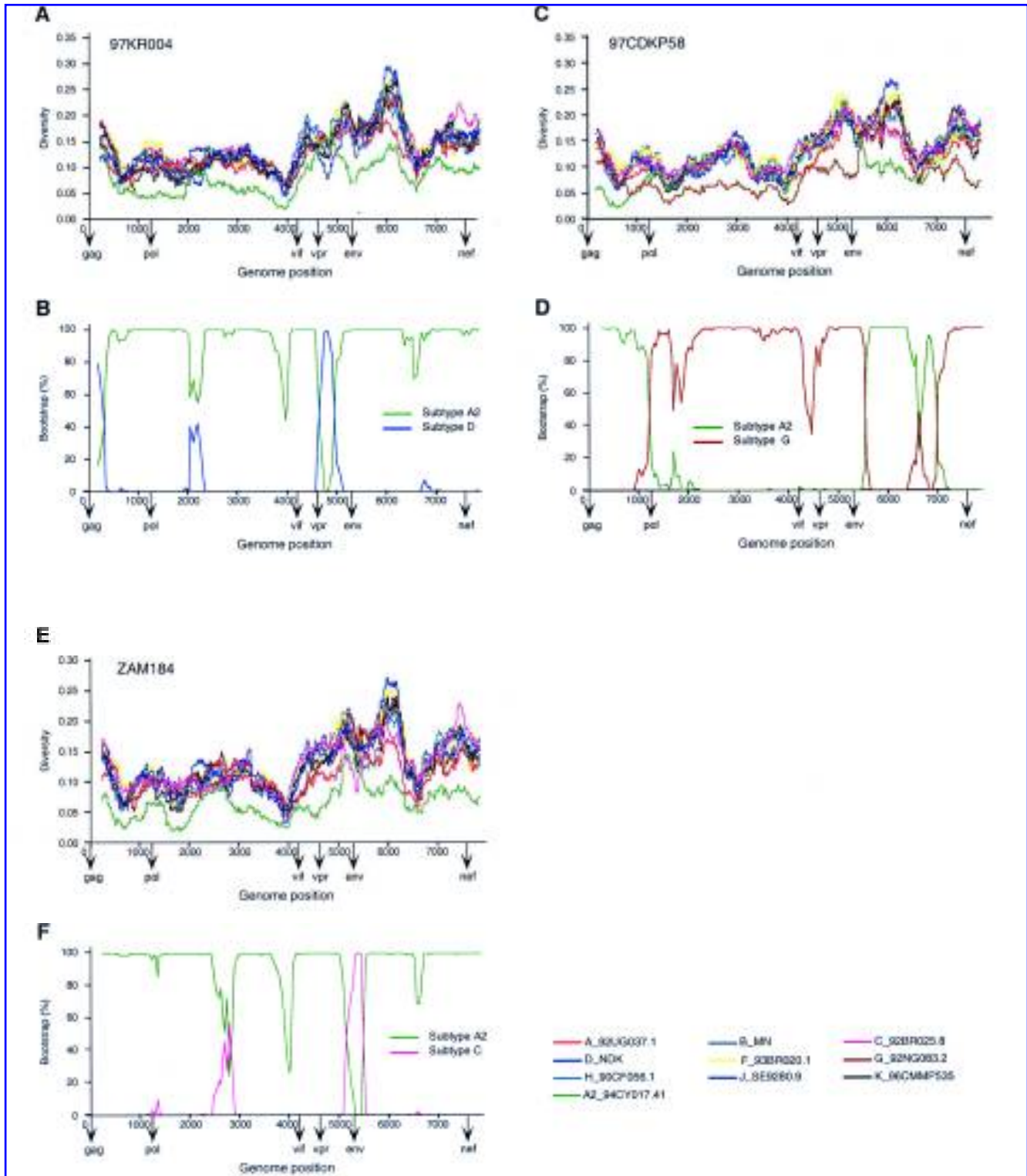


**FIG. 3.** Subtype distance analysis of three newly derived viruses of unknown subtype origin in the envelope region. Intrasubtype (green), intersubsubtype (blue), and intersubtype (yellow) distances were determined from a set of reference sequences that included representatives of subtypes A, C, F, G, H, J, and K as well as subsubtypes B, D, F1, and F2, respectively, using the Subtype Distance (SUDI) program available at the Los Alamos HIV Sequence database website (<http://hiv-web.lanl.gov>). Envelope nucleotide sequences of three newly derived viruses of unknown subtype origin (94CY017.41, 97CDKTB48, and 97CDKS10) were then compared with the corresponding sequences from the same reference set. The subtype sequences representing the closest relatives were identified (in this case subtype A) and the corresponding distances plotted that fell into the subsubtype range (red). The *x* axis indicates nucleotide sequence distances (0.1 = 10%). The *y* axis indicates the percentage of pairwise comparisons that yielded the same nucleotide sequence divergence.



**FIG. 4.** Phylogenetic relationships of newly characterized viruses (highlighted) to representatives of all major HIV-1 group M subtypes in *gag*, *pol*, and *env* regions. Trees were constructed from near full-length *gag* (lacking p17 sequences) as well as full-length *pol* and *env* nucleotide sequences, using the neighbor-joining method and rooted with SIVcpzGAB1 as an outgroup (see text for details). Horizontal branch lengths are drawn to scale (the scale bar represents 0.05 nucleotide substitution 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 denote the major subtypes and subsotypes of HIV-1 group M.





**FIG. 5.** Diversity and bootstrap plots of three recombinants containing fragments of subsubtype A2. The newly derived nearly full-length 97KR004 and 97CDKP58 sequences as well as the previously reported genomic sequence ZAM184<sup>40a</sup> were aligned with HIV-1 subtype reference sequences (gaps were stripped from the alignment). Distance plots of 97KR004 (A), 97CDKP58 (C), and ZAM184 (E) were calculated as described<sup>39,41</sup> (see legend to Fig. 1 for greater detail). Bootstrap plots depict the relationships of 97KR004 (B), 97CDKP58 (D), and ZAM184 (F) to subsubtype A2 (green) as well as representatives of subtypes D (blue), G (brown), and C (pink), respectively. Trees were constructed from the multiple genome alignment, and the bootstrap values supporting the clustering of the tentative recombinants with their parental viruses were plotted for a window of 400 bp moved in increments of 50 bp along the alignment. The *x* axis indicates the nucleotide positions along the alignment (the start codons of the *gag*, *pol*, *vif*, *vpr*, *env*, and *nef* genes are shown). The *y* axis denotes the percentage of bootstrap values. Points of cross-over of the two curves indicate recombination breakpoints.

tances for subtypes B/D were slightly higher, and *gag* and *pol* distances for subtypes B/D and F1/F2 were slightly lower (i.e., closer to the intrasubtype distances; data not shown).

In a final set of experiments, we constructed phylogenetic trees from near full-length *gag* (lacking 400 bp from the 5' end), and complete *pol* and *env* nucleotide sequences of the new viruses as well as a selected group of reference sequences (Fig. 4). This analysis illustrated again that the new sequences formed a tight clade in all three genomic regions, which was supported by 100% bootstrap values. The trees also confirmed that this clade was significantly more closely related to subtype A than to any of the other subtypes, and that the distance between the two sister clades was comparable to distances seen between F1/F2 and B/D clades. According to current nomenclature rules, we have thus tentatively designated this new lineage as sub-subtype A2, with prototypic subtype A viruses now representing sub-subtype A1.

### Identification of A2 recombinant genomes

Since initial analyses suggested that the remaining two full-length sequences 97KR004 and 97CDKP58 were likely mosaic, we used a combination of distance and phylogenetic approaches

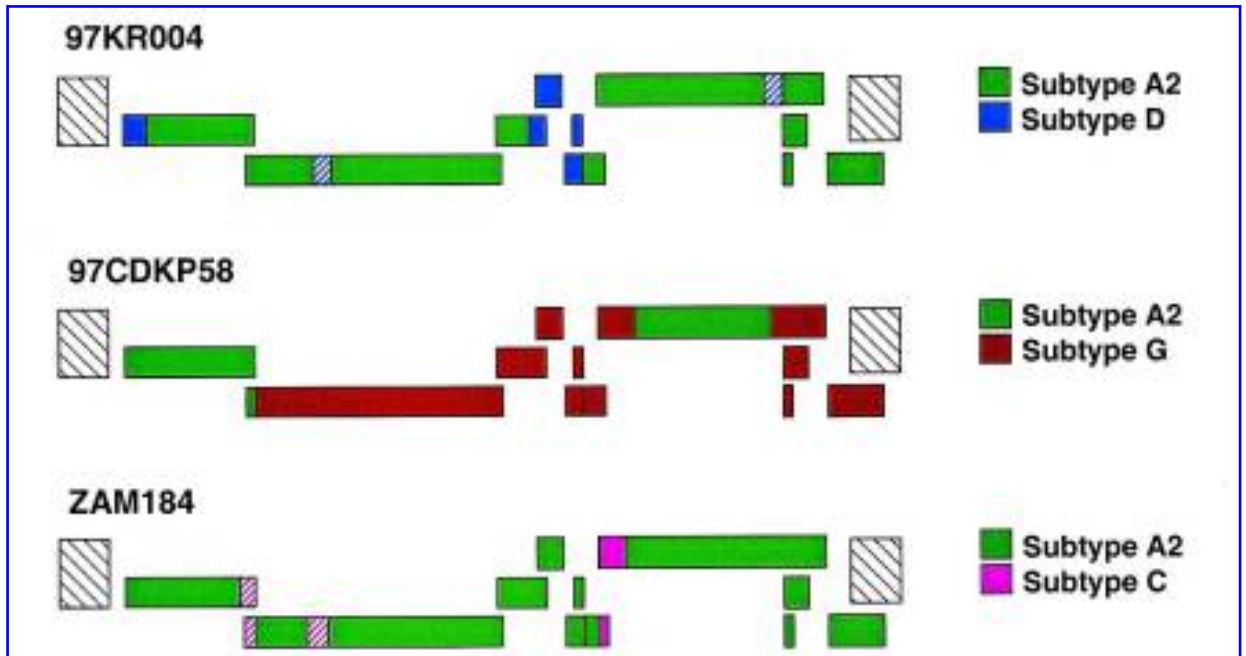
to characterize them in greater detail (Fig. 5). These analyses confirmed that both strains represented complex recombinants containing portions of subsubtype A2 as well as fragments from other subtypes. 97KR004 was found to be almost entirely composed of subsubtype A2. However, three clear-cut cross-overs in the bootstrap plot indicated that it also contained portions of subtype D. Similarly, diversity and bootstrap plot analyses identified 97CDKP58 to be mainly composed of subtype G, but also to contain portions of A2 in the 5' and 3' regions of its genome.

Since full-length nonmosaic reference sequences for subtypes D and G are available, we performed informative site analysis to characterize the location of the recombination breakpoints in greater detail. NDK (subtype D) and 94CY017.41 (subsubtype A2) were used as parental sequences for the analysis of 97KR004, and 94CY017.41 and 92NG083.2 (subtype G) were used for the analysis of 97CDKP58 (90CF056.1 served as an outgroup). By examining the distribution of informative sites that supported the clustering of 97KR004 with either subtype D or subsubtype A2, we identified a total of seven recombination breakpoints (Table 2, top). Two of these defined small subtype D fragments, 143 and 132 bp in length, respectively (indicated by asterisks in Table 2). Because of their short length, they could not be confirmed by phylogenetic tree anal-

TABLE 2. INFORMATIVE SITE ANALYSIS<sup>a</sup>

Sequence	Region	Subtype	Informative sites		
			Subtype A2 (94CY017.41)	Subtype D (NDK)	Outgroup (90CF056.1)
97KR004	1–273	D	2	15	6
	303–2040	A2	53	11	9
	2094–2236	D*	1	5	1
	2254–4568	A2	86	17	17
	4597–4907	D	2	8	0
	4936–6853	A2	83	19	15
	6863–6994	D*	1	7	1
	7047–8046	A2	48	11	17
97CDKP58	1–1324	A2	35	13	5
	1357–5438	G	21	145	28
	5439–6319	A2	45	7	8
	6431–8057	G	12	49	16
ZAM184	1–1109	A2	39	7	4
	1214–1282	C*	0	5	0
	1286–2256	A2	35	3	2
	2304–2441	C*	1	5	0
	2466–5168	A2	100	13	15
	5222–5498	C	2	17	4
5510–8042	A2	116	20	15	

<sup>a</sup>To determine the recombination breakpoints, each putative recombinant was compared with two parental sequences (A2\_94CY017.41 and D\_NDK; A2\_94CY017.41 and G\_92NG083.2; or A2\_94CY017.41 and C\_92BR025.8) and an outgroup (H\_90CF056.1). Recombination breakpoints were mapped by examining the linear distribution of phylogenetically informative sites supporting the clustering of the hybrid with each of the two parental subtypes. Asterisks indicate mosaic regions that could not be confirmed by phylogenetic tree analysis because of their short length.



**FIG. 6.** Schematic representation of the mosaic genome structures of three A2 recombinants. The recombination breakpoints were inferred from bootstrap and informative site analyses. Regions of different subtypes and subsubtypes are color coded. Fragments of uncertain subtype origin are hatched (see text for details). LTR sequences were not analyzed and are thus shown as hatched open boxes.

ysis; thus, their existence remains tentative. The other three breakpoints, however, including one in *gag* and two in the accessory gene region, were well supported by phylogenetic tree analysis, identifying 97KR004 unequivocally as an A2/D recombinant. Similarly, 97CDKP58 was found to contain three breakpoints within its genome by informative site analysis (Table 2, middle). These spanned sufficiently large regions that clustered within their respective subtypes in phylogenetic trees with significant bootstrap values. A schematic representation of the mosaic patterns of these two viruses (with tentative regions hatched) is shown in Fig. 6.

Finally, in a previous study, we had characterized HIV-1 strains that infected a Zambian woman and her spouse.<sup>40</sup> A full-length representative viral sequence (ZAM184) from the index case was found to comprise a complex A/C recombinant with six reported cross-overs.<sup>40a</sup> Interestingly, the subtype A portions of this recombinant (as well as the subtype A portions of additional viruses derived from this couple at different time points) were only distantly related to prototypic subtype A sequences, but were closely related, in both *gag* and *env* regions, to sequences derived from the Cypriot subject HO17, her husband and child, as well as an epidemiologically linked sex partner (a consensus sequence from these four subjects, used for analyses, was termed “Acy” for subtype A variants from Cyprus).<sup>40</sup> This strongly suggested that the subtype A strain that had given rise to the subtype A portions of ZAM184 belonged to subsubtype A2. To address this, we compared ZAM184 with representatives of subsubtypes A1 and A2 independently. As shown in Fig. 5E, the subtype A portions of ZAM184 were most closely related to 94CY017.41 (green curve) and not to 92UG037.1 (red curve). Bootstrap analysis

confirmed this (Fig. 5F), revealing in addition the interspersion of subtype C portions consistent with previous results. Informative site analysis using 94CY017.41 as the A2 reference sequence identified six recombination breakpoints similar to those previously reported (Table 2, bottom). However, again two of these defined subtype C insertions at the *gag/pol* overlap and in the *pol* gene were too short (69 and 138 bp after gap stripping, respectively) to be confirmed by phylogenetic tree analysis. The same results were obtained with two other subtype C parental sequences (data not shown). These data thus confirmed and extended our previous results showing that ZAM184 represents an A2/C recombinant with at least two well-defined recombination cross-overs (Fig. 6).

## DISCUSSION

The purpose of this study was to characterize a group of HIV-1 strains that had previously been identified as subtype A variants. Sequence analysis of two full-length and two half-genomes revealed the existence of a new lineage within the group M radiation; viruses belonging to this lineage formed a tight cluster in all regions of their genomes, but were quite divergent from other HIV-1 strains. Of all known reference sequences, subtype A viruses were the closest relatives, but distance values between them and the subtype A variants exceeded the intrasubtype range. In phylogenetic trees the newly characterized HIV-1 strains formed a distinct sister clade to subtype A, but were not divergent enough to warrant a new subtype designation. Finally, three full-length recombinant genomes were identified to contain portions of this “subtype A-like” lineage.

According to current nomenclature rules,<sup>2,3</sup> this defines a subtype, which we have tentatively termed A2.

To determine to what extent previously reported subtype A viruses are representatives of this lineage, we have screened all full-length HIV-1 *gag* ( $n = 118$ ) and *env* ( $n = 238$ ) sequences available from the Los Alamos database by phylogenetic tree analysis. Interestingly, no other A2-like virus was identified (data not shown). Thus, it appears that the great majority of molecularly characterized subtype A viruses, including all full-length nonmosaic subtype A reference sequences as well as all intersubtype and intergroup recombinants known to contain subtype A sequences (except 97KR004, 97CDKP58, and ZAM184), belong to what should now be referred to as subtype A1. However, it would be premature to reclassify all previously reported subtype A sequences as A1 since there are numerous subgenomic sequences that have not been formally characterized. Indeed, in searching the literature, we found two articles that described subtype A variants that clustered with ZAM184 in trees constructed from partial *env* sequences (V2, C2V3). Both of these subtype A variants (K976 and Q45-CxA) were discovered in the context of independent molecular epidemiological surveys in Kenya.<sup>50,51</sup> They fell within the A2 radiation when analyzed together with our newly derived A2 reference sequences (data not shown). Finally, in a collaborative study with CDC investigators, we most recently reported still another A2 virus (98CM63) in a Cameroonian cohort of HIV-1-seropositive individuals that was identified on the basis of partial *gag*, *pol*, and *env* sequences. 96CM63 clustered with 94CY017.41 with significant bootstrap values in all regions analyzed.<sup>52</sup> Thus, small partial subtype A sequences should either be reanalyzed, or more practically, should remain under the terminology "subtype A" in the database to indicate an as yet undefined subsubtype origin.

The molecular epidemiology of subsubtype A2 is interesting for several reasons. First, with documented A2 infections in Cyprus,<sup>40</sup> South Korea (this study), Cameroon,<sup>52</sup> Zambia,<sup>40a</sup> Kenya,<sup>50,51</sup> and the DRC,<sup>16</sup> it is clear that this subsubtype has already spread to three different continents. However, despite this widespread distribution, the overall prevalence of subsubtype A2, particularly in countries with predominant A1 infections, appears to be exceedingly low. Except for the four strains from the DRC, all other A2 infections represented either single occurrences or a cluster of epidemiologically linked cases. For example, the two Kenyan viruses (K976 and Q45-CxA) were each identified as the only A2 representatives in two independent molecular epidemiological surveys of 17 (Janssens *et al.*<sup>50</sup>) and 22 (Poss *et al.*<sup>51</sup>) unrelated Kenyan individuals in 1990/1992 and 1993, respectively. Similarly, the Cameroonian strain 98CM63 was the only A2 strain among 123 HIV-1 viruses collected in Yaounde, Cameroon in 1998. In this case, 89 other subtype A viruses were identified, 63 of which represented CRF02\_AG (IBNG-like) viruses, while 26 likely represented A1 strains.<sup>52</sup> 94CY017 was 1 of 4 A2 viruses identified in a survey of 25 HIV-1-infected individuals from Nicosia, Cyprus; however, patient histories revealed that these four cases were all epidemiologically linked.<sup>40</sup> Finally, the Korean strain 97KR004 was the only A2 recombinant of 20 HIV-1 strains characterized; however, in this case all other 19 viruses were found to belong to subtype B (our unpublished data). Thus, while the South Korean and Cyprus cases, and possibly also

the Cameroonian infection, could be explained by a more recent arrival of A2 viruses in these countries, the documented presence of A2 sequences in Zambia in 1990 (Salminen *et al.*<sup>40a</sup>), in Kenya between 1990 and 1993 (Refs. 50 and 51), and in the DRC before the onset of the civil war in 1997 (Vidal *et al.*<sup>16</sup>) indicates that this lineage has been circulating in equatorial Africa for a considerable length of time. It will be interesting to determine whether the relative frequency of A2 viruses in these regions has been underestimated, in particular in East Africa, where subsubtype A1 is believed to be predominant, or whether it indeed represents a rather rare lineage. However, even if the latter is the case, the recent arrival of A2 viruses in countries outside Africa should not be ignored. After all, CRF01\_AE viruses that today comprise the most common HIV-1 lineage in Southeast Asia have a similar epidemiological history.<sup>10,36,37</sup>

## ACKNOWLEDGMENTS

We thank Elizabeth Bailes and Cynthia Rodenburg for helpful discussions. This work was supported by grants from the National Institutes of Health (NO1 AI 85338, RO1 AI 533756 and P20 AI 27767, RO1 AI 43868), the Howard Hughes Medical Institute, the Agence National de Recherche sur le SIDA, France (ANRS, Project SIDAK), and the Seoul National University Hospital (03-97-070).

## REFERENCES

1. Kuiken CL, Foley B, Hahn BH, Marx PA, McCutchan F, Mellors JW, Mullins JI, Wolinsky S, and Korber B: *Human Retroviruses and AIDS 1999: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences*. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico, 1999.
2. Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, Funkhouser RK, Gao F, Hahn BH, Kalish ML, Kuiken C, Learn GH, Leitner T, McCutchan F, Osmanov S, Peeters M, Pieniazek D, Salminen M, Sharp PM, Wolinsky S, and Korber B: HIV-1 nomenclature proposal: A reference guide to HIV-1 classification. In: *Human Retroviruses and AIDS 1999: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences* (Kuiken CL, Foley B, Hahn BH, Marx PA, McCutchan F, Mellors JW, Mullins JI, Wolinsky S, and Korber B, eds.). Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico, 2000, pp. 492–505.
3. Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, Funkhouser RK, Gao F, Hahn BH, Kalish ML, Kuiken C, Learn GH, Leitner T, McCutchan F, Osmanov S, Peeters M, Pieniazek D, Salminen M, Sharp PM, Wolinsky S, and Korber B: HIV-1 nomenclature proposal. *Science* 2000;288:55–56.
4. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, Cummins LB, Arthur LO, Peeters M, Shaw GM, Sharp PM, and Hahn BH: Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature* 1999;397:436–441.
5. Hahn BH, Shaw GM, De Cock KM, and Sharp PM: AIDS as a zoonosis: Scientific and public health implications. *Science* 2000;287:607–614.
6. Peeters M and Sharp PM: Genetic diversity of HIV-1: The moving target. *AIDS* 2000;14(Suppl. 3):S129–S140.
7. Laukkanen T, Albert J, Liitsola K, Green SD, Carr JK, Leitner T, McCutchan FE, and Salminen MO: Virtually full-length sequences

- of HIV type 1 subtype J reference strains. *AIDS Res Hum Retroviruses* 1999;15:293-297.
8. Triques K, Bourgeois A, Vidal N, Mpoudi-Ngole, E Mulanga-Kabeya C, Nzilambi N, Torimiro N, Saman E, Delaporte, E and Peeters M: Near-full-length genome sequencing of divergent African HIV type 1 subtype F viruses leads to the identification of a new HIV type 1 subtype designated K. *AIDS Res Hum Retroviruses* 2000;16:139-151.
  9. Triques K, Bourgeois A, Saragosti S, Vidal N, Mpoudi-Ngole E, Nzilambi N, Apetrei C, Ekwalanga M, Delaporte E, and Peeters M: High diversity of HIV-1 subtype F strains in Central Africa. *Virology* 1999;259:99-109.
  10. Esparza J and Bhamarapravati N: Accelerating the development and future availability of HIV-1 vaccines: Why, when, where, and how? *Lancet* 2000;355:2061-2066.
  11. McCutchan FE, Carr JK, Bajani M, Sanders-Buell E, Harry TO, Stoeckli TC, Robbins KE, Gashau W, Nasidi A, Janssens W, and Kalish ML: Subtype G and multiple forms of A/G intersubtype recombinant human immunodeficiency virus type 1 in Nigeria. *Virology* 1999;254:226-234.
  12. McCutchan FE, Salminen MO, Carr JK, and Burke DS: HIV-1 genetic diversity. *AIDS* 1996;10(Suppl. 3):S13-S20.
  13. Louwagie J, McCutchan FE, Peeters M, Brennan TP, Sanders-Buell E, Eddy GA, van der Groen G, Franssen K, Gershy-Damet GM, Deleys R, and Burke DS: Phylogenetic analysis of *gag* genes from 70 international HIV-1 isolates provides evidence for multiple genotypes. *AIDS* 1993;7:769-780.
  14. Toure-Kane C, Montavon C, Faye MA, Gueye PM, Sow PS, Ndoye I, Gaye-Diallo A, Delaporte E, Peeters M, and Mboup S: Identification of all HIV type 1 group M subtypes in Senegal, a country with low and stable seroprevalence. *AIDS Res Hum Retroviruses* 2000;16:603-609.
  15. Montavon C, Toure-Kane C, Liegeois F, Mpoudi E, Bourgeois A, Vergne L, Perret JL, Boumah A, Saman E, Mboup S, Delaporte E, and Peeters M: Most *env* and *gag* subtype A HIV-1 viruses circulating in West and West Central Africa are similar to the prototype AG recombinant virus IBNG. *J Acquir Immune Defic Syndr* 2000;23:363-374.
  16. Vidal N, Peeters M, Mulanga-Kabeya C, Nzilambi N, Robertson D, Ilunga W, Sema H, Tshimanga K, Bongo B, and Delaporte E: Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa. *J Virol* 2000;74:10498-10507.
  17. McCutchan FE: Understanding the genetic diversity of HIV-1. *AIDS* 2000;14(Suppl. 3):S31-S44.
  18. Carr JK, Laukkanen T, Salminen MO, Albert J, Alaeus A, Kim B, Sanders-Buell E, Bix DL, and McCutchan FE: Characterization of subtype A HIV-1 from Africa by full genome sequencing. *AIDS* 1999;13:1819-1826.
  19. Peeters M, Gueye A, Mboup S, Bibollet-Ruche F, Ekaza E, Mulanga C, Ouedrago R, Gandji R, Mpele P, Dibanga G, Koumare B, Saidou M, Esu-Williams E, Lombart JP, Badombena W, Luo N, Vanden Haesevelde M, and Delaporte E: Geographical distribution of HIV-1 group O viruses in Africa. *AIDS* 1997;11:493-498.
  20. Gurtler LG, Hauser PH, Eberle J, von Brunn A, Knapp S, Zekeng L, Tsague JM, and Kaptue L: A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. *J Virol* 1994;68:1581-1585.
  21. Vanden Haesevelde M, Decourt JL, De Leys RJ, Vanderborcht B, van der Groen G, van Heuverswijn H, and Saman E: Genomic cloning and complete sequence analysis of a highly divergent African human immunodeficiency virus isolate. *J Virol* 1994;68:1586-1596.
  22. Rodenburg CM, Li Y, Trask SA, Chen Y, Decker J, Robertson DL, Kalish ML, Shaw GM, Allen S, Hahn BH, Gao F, and the UNAIDS and NIAID Networks for HIV Isolation and Characterization: Near full-length clones and reference sequences for subtype C isolates of HIV-1 from three different continents. *AIDS Res Hum Retroviruses* 2001;17:161-168.
  23. Morgado MG, Sabino EC, Shpaer EG, Bongertz V, Brigido L, Guimaraes MD, Castilho EA, Galvao-Castro B, Mullins JI, Hendry RM, and Mayer A: V3 region polymorphisms in HIV-1 from Brazil: Prevalence of subtype B strains divergent from North American/European prototype and detection of subtype F. *AIDS Res Hum Retroviruses* 1994;10:569-576.
  24. Dumitrescu O, Kalish ML, Kliks SC, Bandea CI, and Levy JA: Characterization of human immunodeficiency virus type 1 isolates from children in Romania: Identification of a new envelope subtype. *J Infect Dis* 1994;169:281-288.
  25. Apetrei C, Descamps D, Collin G, Lousert-Ajaka I, Damond F, Duca M, Simon F, and Brun-Vezinet F: Human immunodeficiency virus type 1 subtype F reverse transcriptase sequence and drug susceptibility. *J Virol* 1998;72:3534-3538.
  26. Descamps D, Apetrei C, Collin G, Damond F, Simon F, and Brun-Vezinet F: Naturally occurring decreased susceptibility of HIV-1 subtype G to protease inhibitors. *AIDS* 1998;12:1109-1111.
  27. Triques K, Coste J, Perret JL, Segarra C, Mpoudi E, Reyes J, Delaporte E, Butcher A, Dreyer K, Herman S, Spadaro J, and Peeters M: Efficiencies of four versions of the AMPLICOR HIV-1 MONITOR test for quantification of different subtypes of human immunodeficiency virus type 1. *J Clin Microbiol* 1999;37:110-116.
  28. Parekh B, Phillips S, Granade TC, Baggs J, Hu DJ, and Respass R: Impact of HIV type 1 subtype variation on viral RNA quantitation. *AIDS Res Hum Retroviruses* 1999;15:133-142.
  29. Murphy G, Belda FJ, Pau CP, Clewley JP, and Parry JV: Discrimination of subtype B and non-subtype B strains of human immunodeficiency virus type 1 by serotyping: Correlation with genotyping. *J Clin Microbiol* 1999;37:1356-1360.
  30. Verrier F, Burda S, Belshe R, Duliege AM, Excler JL, Klein M, and Zolla-Pazner S: A human immunodeficiency virus prime-boost immunization regimen in humans induces antibodies that show interclade cross-reactivity and neutralize several X4-, R5-, and dualtropic clade B and C primary isolates. *J Virol* 2000;74:10025-10033.
  31. 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 Sequence* (Korber B, Foley B, Kuiken C, Leitner T, McCutchan F, Mellors JW, and Hahn BH, eds.). Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico, 1997, pp. III-25-III-30.
  32. Robertson DL, Sharp PM, McCutchan FE, and Hahn BH: Recombination of HIV-1. *Nature* 1995;374:124-126.
  33. Peeters M, Liegeois F, Torimiro N, Bourgeois A, Mpoudi E, Vergne L, Saman E, Delaporte E, and Saragosti S: Characterization of a highly replicative intergroup M/O human immunodeficiency virus type 1 recombinant isolated from a Cameroonian patient. *J Virol* 1999;73:7368-7375.
  34. Takehisa J, Zekeng L, Ido E, Yamaguchi-Kabata Y, Mboudjeka I, Harada Y, Miura T, Kaptu L, and Hayami M: Human immunodeficiency virus type 1 intergroup (M/O) recombination in Cameroon. *J Virol* 1999;73:6810-6820.
  35. Laukkanen T, Carr JK, Janssens W, Liitsola K, Gotte D, McCutchan FE, Op de Coul E, Cornelissen M, Heyndrickx L, van der Groen G, and Salminen MO: Virtually full-length subtype F and F/D recombinant HIV-1 from Africa and South America. *Virology* 2000;269:95-104.
  36. Gao F, Robertson DL, Morrison SG, Hui H, Craig S, Decker J, Fultz PN, Girard M, Shaw GM, Hahn BH, and Sharp PM: The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *J Virol* 1996;70:7013-7029.

37. Carr JK, Salminen MO, Koch C, Gotte D, Artenstein AW, Hegerich PA, St Louis D, Burke DS, and McCutchan FE: Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *J Virol* 1996;70:5935-5943.
38. Liitsola K, Tashkinova I, Laukkanen T, Korovina G, Smolskaja T, Momot O, Mashkileyson N, Chaplinskas S, Brummer-Korvenkontio H, Vanhatalo J, Leinikki P, and Salminen MO: HIV-1 genetic subtype A/B recombinant strain causing an explosive epidemic in injecting drug users in Kaliningrad. *AIDS* 1998;12:1907-1919.
39. Gao F, Robertson DL, Carruthers CD, Li Y, Bailes E, Kostrikis LG, Salminen MO, Bibollet-Ruche F, Peeters M, Ho DD, Shaw GM, Sharp PM, and Hahn BH: 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). *J Virol* 1998;72:10234-10241.
40. Kostrikis LG, Bagdades E, Cao Y, Zhang L, Dimitriou D, and Ho DD: Genetic analysis of human immunodeficiency virus type 1 strains from patients in Cyprus: Identification of a new subtype designated subtype I. *J Virol* 1995;69:6122-6130.
- 40a. Salminen MO, Carr JK, Robertson DL, Hegerich P, Gotte D, Koch C, Sanders-Buell E, Gao F, Sharp PM, Hahn BH, Burke DS, and McCutchan FE: Evolution and probable transmission of intersubtype recombinant human immunodeficiency virus type 1 in a Zambian couple. *J Virol* 1997;71:2647-2655.
41. Gao F, Robertson DL, Carruthers CD, Morrison SG, Jian B, Chen Y, Barré-Sinoussi F, Girard M, Srinivasan A, Abimiku AG, Shaw GM, Sharp PM, and Hahn BH: A comprehensive panel of near-full-length clones and reference sequences for non-subtype B isolates of human immunodeficiency virus type 1. *J Virol* 1998;72:5680-5698.
42. Felsenstein J: *PHYLIP (Phylogenetic Inference Package)*, version 2.51. Department of Genetics, University of Washington, Seattle, Washington, 1984.
43. Carr JK, Foley BT, Leitner T, Salminen MO, Korber BT, and McCutchan F: Reference sequences representing the principal genetic diversity of HIV-1 in the pandemic. In: *Human Retroviruses and AIDS 1998: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences* (Korber B, Kuiken C, Foley B, Hahn BH, McCutchan F, Mellors JW, and Sodrosky J, eds.). Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico, 1998, pp. III-10-III-19.
44. Thompson JD, Higgins DG, and Gibson TJ: CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673-4680.
45. Faulkner DV and Jurka J: Multiple aligned sequence editor (MASE). *Trends Biochem Sci* 1988;13:321-322.
46. Kimura M: A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 1980;16:111-120.
47. Saitou N and Nei M: The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987;4:406-425.
48. Felsenstein J: *PHYLIP (Phylogenetic Inference Package)*, version 3.5c. Department of Genetics, University of Washington, Seattle, Washington, 1992.
49. Salminen MO, Carr JK, Burke DS, and McCutchan FE: Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res Hum Retroviruses* 1995;11:1423-1425.
50. Janssens W, Heyndrickx L, Franssen K, Temmerman M, Leonaers A, Ivens T, Motte J, Piot P, and Van der Groen G: Genetic variability of HIV type 1 in Kenya. *AIDS Res Hum Retroviruses* 1994;10:1577-1579.
51. Poss M, Gosink J, Thomas E, Kreiss JK, Ndinya-Achola J, Mandaliya K, Bwayo J, and Overbaugh J: Phylogenetic evaluation of Kenyan HIV type 1 isolates. *AIDS Res Hum Retroviruses* 1997;13:493-499.
52. Fonjungo PN, Mpoudi EN, Torimiro JN, Alemnji GA, Eno LT, Nkengasong JN, Gao F, Rayfield M, Folks TM, Pieniazek D, and Lal RB: Presence of diverse human immunodeficiency virus type 1 viral variants in Cameroon. *AIDS Res Hum Retroviruses* 2000;16:1319-1324.

Address reprint requests to:

Feng Gao  
 Department of Medicine  
 University of Alabama at Birmingham  
 701 S. 19th Street, LHRB 639  
 Birmingham, Alabama 35294

E-mail: feng@uab.edu

**This article has been cited by:**

1. Catherine A Blish, Rebecca Nedellec, Kishorchandra Mandaliya, Donald E Mosier, Julie Overbaugh. 2007. HIV-1 subtype A envelope variants from early in infection have variable sensitivity to neutralization and to inhibitors of viral entry. *AIDS* **21**:6, 693-702. [[CrossRef](#)]
2. André F Santos, Carlos G Schrago, Ana Maria B Martinez, Raul Mendoza-Sassi, Jussara Silveira, Thatiana M Sousa, Renan B Lengruher, Esmeralda AJM Soares, Eduardo Sprinz, Marcelo A Soares. 2007. Epidemiologic and Evolutionary Trends of HIV-1 CRF31\_BC-Related Strains in Southern Brazil. *JAIDS Journal of Acquired Immune Deficiency Syndromes* **45**:3, 328. [[CrossRef](#)]
3. M. Tagliamonte, N. Vidal, M.L. Tornesello, M. Peeters, F.M. Buonaguro, L. Buonaguro. 2006. Genetic and Phylogenetic Characterization of Structural Genes from Non-B HIV-1 Subtypes in Italy. *AIDS Research and Human Retroviruses* **22**:10, 1045-1051. [[Abstract](#)] [[PDF](#)] [[PDF Plus](#)]
4. Unchalee Visawapoka, Sodsai Tovananubutra, Jeffrey R. Currier, Josephine H. Cox, Carl J. Mason, Monique Wasunna, Mathurose Ponglikitmongkol, William E. Dowling, Merlin L. Robb, Deborah L. Bix, Francine E. McCutchan. 2006. Circulating and Unique Recombinant Forms of HIV Type 1 Containing Subsubtype A2. *AIDS Research and Human Retroviruses* **22**:7, 695-702. [[Abstract](#)] [[PDF](#)] [[PDF Plus](#)]
5. N. Vidal, C. Mulanga, S. Edidi Bazepeo, F. Lepira, E. Delaporte, M. Peeters. 2006. Identification and Molecular Characterization of Subsubtype A4 in Central Africa. *AIDS Research and Human Retroviruses* **22**:2, 182-187. [[Abstract](#)] [[PDF](#)] [[PDF Plus](#)]
6. Mária Mezei, Katalin Balog, Dunja Z. Babič, Gábor Tóth, Gábor Cech, Balázs Vajna, Tamás Tauber, Katja Seme, Janez Tomažič, Ludvik Vidmar, Mario Poljak, János Minárovits. 2006. Genetic Variability of gag and env Regions of HIV Type 1 Strains Circulating in Slovenia. *AIDS Research and Human Retroviruses* **22**:1, 109-113. [[Abstract](#)] [[PDF](#)] [[PDF Plus](#)]
7. André F Santos, Thatiana M Sousa, Esmeralda AJM Soares, Sabri Sanabani, Ana MB Martinez, Eduardo Sprinz, Jussara Silveira, Ester C Sabino, Amílcar Tanuri, Marcelo A Soares. 2006. Characterization of a new circulating recombinant form comprising HIV-1 subtypes C and B in southern Brazil. *AIDS* **20**:16, 2011-2019. [[CrossRef](#)]
8. Stephanie Rainwater, Shannon DeVange, Manish Sagar, J. Ndinya-Achola, Kishorchandra Mandaliya, Joan K. Kreiss, Julie Overbaugh. 2005. No Evidence for Rapid Subtype C Spread within an Epidemic in Which Multiple Subtypes and Intersubtype Recombinants Circulate. *AIDS Research and Human Retroviruses* **21**:12, 1060-1065. [[Abstract](#)] [[PDF](#)] [[PDF Plus](#)]
9. Chunfu Yang, Ming Li, John L. K. Mokili, Jorn Winter, Ndongala M. Lubaki, Kashamuka M. Mwandagaliwa, Mwamba J. Kasali, Atibu J. Losoma, Thomas C. Quinn, Robert C. Bollinger, Renu B. Lal. 2005. Genetic Diversification and Recombination of HIV Type 1 Group M in Kinshasa, Democratic Republic of Congo. *AIDS Research and Human Retroviruses* **21**:7, 661-666. [[Abstract](#)] [[PDF](#)] [[PDF Plus](#)]
10. Inês Bártole, Marta Epalanga, José Bartolomeu, Marlene Fonseca, Ana Mendes, António Gama, Nuno Taveira. 2005. High Genetic Diversity of Human Immunodeficiency Virus Type 1 in Angola. *AIDS Research and Human Retroviruses* **21**:4, 306-310. [[Abstract](#)] [[PDF](#)] [[PDF Plus](#)]
11. Nicole Vidal, Claire Mulanga, Samuel Edidi Bazepeo, Jeanne Kasali Mwamba, Jean-Willy Tshimpaka, Mulowayi Kashi, Nicasia Mama, Christian Laurent, Francois Lepira, Eric Delaporte. 2005. Distribution of HIV-1 Variants in the Democratic Republic of Congo Suggests Increase of Subtype C in Kinshasa Between 1997 and 2002. *JAIDS Journal of Acquired Immune Deficiency Syndromes* **40**:4, 456. [[CrossRef](#)]
12. Manuel Gómez-Carrillo, Jorge F. Quarleri, Andrea E. Rubio, Mauricio G. Carobene, Dario Dileria, Jean K. Carr, Horacio Salomón. 2004. Drug Resistance Testing Provides Evidence of the Globalization of HIV Type 1: A New Circulating Recombinant Form. *AIDS Research and Human Retroviruses* **20**:8, 885-888. [[Abstract](#)] [[PDF](#)] [[PDF Plus](#)]
13. Chunfu Yang, Ming Li, Ya-Ping Shi, Jorn Winter, Anna M. van Eijk, John Ayisi, Dale J. Hu, Richard Steketee, Bernard L. Nahlen, Renu B. Lal. 2004. Genetic Diversity and High Proportion of Intersubtype

- Recombinants among HIV Type 1-Infected Pregnant Women in Kisumu, Western Kenya. *AIDS Research and Human Retroviruses* **20**:5, 565-574. [[Abstract](#)] [[PDF](#)] [[PDF Plus](#)]
14. Eric L. Delwart , Sharyn Orton , Bharat Parekh , Trudy Dobbs , Kenneth Clark , Michael P. Busch . 2003. Two Percent of HIV-Positive U.S. Blood Donors Are Infected with Non-subtype B Strains. *AIDS Research and Human Retroviruses* **19**:12, 1065-1070. [[Abstract](#)] [[PDF](#)] [[PDF Plus](#)]
  15. Alexei E. Masharsky, Nikolai A. Klimov, Andrei P. Kozlov. 2003. Molecular Cloning and Analysis of Full-Length Genome of HIV Type 1 Strains Prevalent in Countries of the Former Soviet Union. *AIDS Research and Human Retroviruses* **19**:10, 933-939. [[Abstract](#)] [[PDF](#)] [[PDF Plus](#)]
  16. Rod S. Daniels , Chun Kang , Dina Patel , Zheng Xiang , Nigel W. Douglas , Natalie N. Zheng , Hae-Wol Cho , Joo-Shil Lee . 2003. An HIV Type 1 Subtype B Founder Effect in Korea: gp160 Signature Patterns Infer Circulation of CTL-Escape Strains at the Population Level. *AIDS Research and Human Retroviruses* **19**:8, 631-641. [[Abstract](#)] [[PDF](#)] [[PDF Plus](#)]
  17. Heungsung Sung, Brian T. Foley, Sun H. Ahn, Young B. Kim, Jeong-Don Chae, Yung O. Shin, Hee I. Kang, Young K. Cho. 2003. Natural Polymorphisms of Protease in Protease Inhibitor-Naive HIV-1 Infected Patients in Korea: A Novel L63M in Subtype B. *AIDS Research and Human Retroviruses* **19**:6, 525-530. [[Abstract](#)] [[PDF](#)] [[PDF Plus](#)]
  18. Carrie L. Young, Dale J. Hu, Robert Byers, Suphak Vanichseni, Nancy L. Young, Robert Nelson, Philip A. Mock, Kachit Choopanya, Robert Janssen, Timothy D. Mastro, Joanne V. Mei. 2003. Evaluation of a Sensitive/Less Sensitive Testing Algorithm Using the bioMérieux Vironostika-LS Assay for Detecting Recent HIV-1 Subtype B' or E Infection in Thailand. *AIDS Research and Human Retroviruses* **19**:6, 481-486. [[Abstract](#)] [[PDF](#)] [[PDF Plus](#)]
  19. Diana D. Huang, Thomas A. Giesler, James W. Bremer. 2003. Sequence Characterization of the Protease and Partial Reverse Transcriptase Proteins of the NED Panel, an International HIV Type 1 Subtype Reference and Standards Panel. *AIDS Research and Human Retroviruses* **19**:4, 321-328. [[Abstract](#)] [[PDF](#)] [[PDF Plus](#)]
  20. Elijah M. Songok, Raphael W. Lihana, Michael K. Kiptoo, Isaiah O. Genga, Rukia Kibaya, Frida Odhiambo, Kana Kobayashi, Yumi Ago, Nicaise Ndembi, Fredrick Okoth, Yoshihide Fujiyama, Jiro Imanishi, Hiroshi Ichimura. 2003. Identification of env CRF-10 among HIV Variants Circulating in Rural Western Kenya. *AIDS Research and Human Retroviruses* **19**:2, 161-165. [[Abstract](#)] [[PDF](#)] [[PDF Plus](#)]
  21. Martine Peeters, Coumba Toure-Kane, John N Nkengasong. 2003. Genetic diversity of HIV in Africa. *AIDS* **17**:18, 2547. [[CrossRef](#)]
  22. Monick Lindenmeyer Guimarães, Aline dos Santos Moreira, Regina Loureiro, Bernardo Galvão-Castro, Mariza Gonçalves Morgado. 2002. High Frequency of Recombinant Genomes in HIV Type 1 Samples from Brazilian Southeastern and Southern Regions. *AIDS Research and Human Retroviruses* **18**:17, 1261-1269. [[Abstract](#)] [[PDF](#)] [[PDF Plus](#)]
  23. Myoung-Don Oh, Sang Won Park, Uiseok Kim, Hong Bin Kim, Young Ju Choe, Euijong Kim, Kangwon Choe. 2002. Determination of Genetic Subtypes of HIV Type 1 Isolated from Korean AIDS Patients. *AIDS Research and Human Retroviruses* **18**:16, 1229-1233. [[Abstract](#)] [[PDF](#)] [[PDF Plus](#)]
  24. Cara Carthel Burns, Lisa M. Gleason, Afsaneh Mozaffarian, Cristina Giachetti, Jean K. Carr, Julie Overbaugh. 2002. Sequence Variability of the Integrase Protein from a Diverse Collection of HIV Type 1 Isolates Representing Several Subtypes. *AIDS Research and Human Retroviruses* **18**:14, 1031-1041. [[Abstract](#)] [[PDF](#)] [[PDF Plus](#)]
  25. Maria A. Papathanasopoulos, Tonie Cilliers, Lynn Morris, John L. Mokili, William Dowling, Deborah L. Birx, Francine E. McCutchan. 2002. Full-Length Genome Analysis of HIV-1 Subtype C Utilizing CXCR4 and Intersubtype Recombinants Isolated in South Africa. *AIDS Research and Human Retroviruses* **18**:12, 879-886. [[Abstract](#)] [[PDF](#)] [[PDF Plus](#)]
  26. Aida Esteves , Ricardo Parreira , Teresa Venenno , Margarida Franco , João Piedade , José Germano de Sousa , Wanda F. Canas-Ferreira . 2002. Molecular Epidemiology of HIV Type 1 Infection in Portugal:



High Prevalence of Non-B Subtypes. *AIDS Research and Human Retroviruses* **18**:5, 313-325. [[Abstract](#)]  
[[PDF](#)] [[PDF Plus](#)]

27. William Dowling, Bohye Kim, Carl Mason, K. Wasunna, Uzma Alam, Lynne Elson, Deborah Birx, Merlin Robb, Francine McCutchan, Jean Carr. 2002. *AIDS* **16**:13, 1809. [[CrossRef](#)]