JOURNAL OF VIROLOGY, Sept. 1996, p. 5935–5943 Vol. 70, No. 9
0022-538X/96/$04.00+0

Full-Length Sequence and Mosaic Structure of a Human Immunodeficiency Virus Type 1 Isolate from Thailand

JEAN K. CARR,[1]* MIKA O. SALMINEN,[1] CHRISTINE KOCH,[1]† DEANNA GOTTE,[1] ANDREW W. ARTENSTEIN,[2] PATRICIA A. HEGERICH,[1] DANIEL ST. LOUIS,[1] DONALD S. BURKE,[2] AND FRANCINE E. MCCUTCHAN[1]

Henry M. Jackson Foundation for the Advancement of Military Medicine[1] and Division of Retrovirology, Walter Reed Army Institute of Research,[2] Rockville, Maryland 20850

Received 23 January 1996/Accepted 17 May 1996

Human immunodeficiency virus type 1 isolates of envelope genotype E are contributing substantially to the global pandemic. These strains appear to be mosaics, with the *gag* gene from clade A and the envelope from clade E; the parental clade E strain has not been found. Here we report the first full genomic sequence of one such mosaic virus, isolate CM240 from Thailand. Multiple breakpoints between the two parental genotypes have been found in a CM240 virus. The entire *gag-pol* region and most, if not all, of the accessory genes *vif*, *vpr*, *tat*, *rev*, and *vpu* appear to derive from clade A. The genotype switches to E shortly after the signal peptide of the envelope and back to clade A near the middle of gp41; thus, the portion of the envelope that lies on the cytoplasmic side of the membrane appears to be principally derived not from clade E, as previously thought, but from clade A. Another small segment not belonging to any recognized clade and presumably also contributed by the parental E strain has been found in the long terminal repeat. It may be significant that the implied virion structure resembles a pseudotype virus with the matrix and core from one clade and the outer envelope from another. In the long terminal repeat, differences were observed between CM240 and other clades in the number of NF-κB binding sites, the sequence of the TATA box, and the putative secondary structure of the transactivation response region stem-loop. The mosaic structure of a CM240 virion is suggestive of phenotypic differences which might have contributed to the emergence of this variant.

Two independent reports in 1992 identified a new variant of human immunodeficiency virus type 1 (HIV-1) in Thailand (33, 41). This virus defied classification from the beginning. In the established phylogeny, which at the time included HIV-1 genotypes A through D, the envelope gene sequence of the new variant was equidistant from the other four and was called genotype E (36). Surprisingly, the *gag* gene, encoding the matrix and core proteins of the virus, did not appear to form a new clade, grouping instead with genotype A (31, 37). Such apparent HIV-1 intersubtype mosaicism had been recognized on only one previous occasion, in the Zairian isolate MAL (39); mosaic forms were generally considered to be rare outliers and were not expected to contribute substantially to the global epidemic.

The HIV-1 epidemic in Thailand overturned these predictions definitively. Although the first HIV-1 cases described in Thailand were of genotype B, common in Europe and in the Western Hemisphere, the envelope genotype E viruses, once introduced into northern Thailand and spread through heterosexual contact, virtually swept the country in a few years (49, 52, 55, 56). Genotype B viruses remain the minority strain in Thailand and are even losing ground to envelope genotype E where they initially predominated (28, 54), in injecting drug users. It is now estimated that more than 500,000 people are infected with the envelope clade E virus in Thailand alone, and evidence is accumulating for the wider spread of this genotype in Southeast Asia and in the Western Hemisphere (1, 7, 21, 57).

* Corresponding author. Mailing address: Henry M. Jackson Foundation Research Laboratory, 1600 East Gude Dr., Rockville, MD 20850. Phone: (301) 217-9410. Fax: (301) 762-7460. Electronic mail address: jcarr@hiv.hjf.org.

† Present address: National Institutes of Health, Bethesda, MD 20892.

The origin of the envelope clade E virus has been an important subject of investigation, given its demonstrated potential for the establishment of new epidemics. Paradoxically, the viral strain that contributed the E portion of this virus has not come to light, despite genetic analysis of more than 700 HIV-1 isolates from international locations (9, 27, 30, 37). If the Thailand variant arose through recombination between an A and an E virus, the clade E parental strain either has remained exceedingly rare and geographically circumscribed or has since disappeared. Two countries in central Africa have yielded HIV-1 viruses which also classify as genotype E in envelope and, to the extent that they have been analyzed, are genotype A in *gag* (35, 40). Thus, the putative recombination event(s) probably occurred before the spread of this variant outside of Africa.

Genetic characterization of the envelope clade E viruses has been proceeding vigorously since they were discovered in 1992. The analyses have included a large number of isolates but have been limited to the *gag* and envelope (*env*) genes, which together comprise less than half of the genome. Here (and in the report by Gao et al. [20]), the first complete genomic analyses of these apparently mosaic HIV-1 strains are provided.

MATERIALS AND METHODS

Viral isolate. HIV-1 strain CM240, which is the subject of this report, is a virus which was isolated from a 21-year-old man in northern Thailand. His HIV-1-positive serostatus was determined after selection by lottery for service in the Royal Thai Army in 1990. In a face-to-face interview, his only reported risk factor for HIV-1 infection was contact with female commercial sex workers. He was asymptomatic at the time a blood sample was obtained and transported from Thailand to Rockville, Md., for virus isolation. Peripheral blood mononuclear cells were separated on Ficoll-Hypaque and were cocultivated with phytohemagglutinin-stimulated donor peripheral blood mononuclear cells as described previously (8). DNA from the p24 antigen-positive culture was the source of proviral DNA.

PCR amplification and cloning. Using a recently described procedure for PCR

amplification of virtually full-length HIV-1 genomes (47), we obtained most of the provirus in one continuous segment. The PCR primers were positioned to amplify all but 73 bp of the HIV-1 long terminal repeat (LTR), but the segment, once cloned, proved to have deletion near its 3′ end. A single PCR clone contained 8,193 nucleotides (nt) of the 9,203-nt genome. Smaller, overlapping fragments containing the *nef* and LTR regions of the genome were PCR amplified and cloned separately, using PCR primers as described previously (2, 34). The three separate clones were completely sequenced, and overlapping segments, which corresponded with virtual identity, were used to assemble a full-length genomic sequence.

DNA sequencing. Template DNA for automated sequencing was prepared as described previously (47). Clones were fully sequenced on both strands by using fluorescent dye terminators and an Applied Biosystems (Applied Biosystems Inc., Foster City, Calif.) model 373A DNA sequencer. DNA sequences were assembled by using DNAStar (DNAStar Inc., Madison, Wis.) or Sequencher (Genecodes Inc., Ann Arbor, Mich.) software on Macintosh computers. All sequence ambiguities were resolved, and the final sequence, assembled from the three independent clones, spanned 9,203 nt and represented the entire HIV-1 genome. Previously, *gag* and *env* genes from isolate CM240 were PCR amplified separately and sequenced (our unpublished data, compiled in reference 37), and these genes were used in some analyses.

Analysis. A multiple alignment of the newly derived CM240 full-length genome with available full-length HIV-1 genomes of other clades was generated. Reference isolates included U455 (Uganda, 1990, clade A), MN and NL4-3 (United States, 1986 to 1988, clade B), ELI, NDK, and Z2Z6 (Zaire, 1986 to 1989, clade D), C2220 (Ethiopia, 1986, clade C [46]), and 90CR402 (Central African Republic, 1990, clade E [20]). The multiple alignment was broken into 37 overlapping segments of equal length, and each segment was analyzed separately and by two independent methods. First, the CM240 genome was analyzed for parental genotypes and for breakpoints by bootscanning as described previously (45). Briefly, phylogenetic trees were constructed from each segment by using maximum parsimony (Phylip package, version 3.52c [17]), and the bootstrap value (16) for inclusion of isolate CM240 in clade A was determined. A bootstrap value of 95% or above was considered definitive. For the second analysis, genetic distances between each pair of isolates were calculated by using maximum likelihood (15), and the distance of isolate CM240 to isolates of other clades was considered for each segment. A similar analysis was applied to the envelope gene in an alignment containing additional HIV-1 isolates. The nucleotide positions of breakpoints are designated with respect to reference isolate HIV_{MN} (GenBank accession number M17449).

Nucleotide sequence accession number. The full-length sequence of isolate CM240 is available under GenBank accession number U54771.

RESULTS

The CM240 genome. The full-length sequence of isolate CM240 is presented in Fig. 1. First, the *gag* and *env* genes were analyzed, as these segments are the best described among HIV-1 isolates (37). Table 1 shows a distance matrix comparing *gag* and *env* of CM240 with those of reference strains from other clades. The *gag* and *env* genes of isolates CM240 and 90CR402 were the most closely related, with distances of 6.4 and 7.1%, respectively. The *env* gene of CM240 was virtually equidistant from those of the isolates of clades A, B, C, and D, reiterating this isolate's previous classification as envelope clade E. The *gag* gene was distant from those of clades B, C, and D but was closer to that of clade A than to those of other clades (9.7% versus 14.9 to 16.0% [Table 1]). These relationships confirm that the full-length sequence of CM240, like the shorter sequences of other envelope clade E HIV-1 isolates from Thailand and Africa, appears to be a separate clade by envelope analysis but groups (albeit loosely) with clade A in the *gag* gene. The close similarity of isolates CM240 and 90CR402 in both *gag* and *env* lends support to their inclusion in a single clade.

Previously uncharacterized segments of a CM240 genome were then analyzed (Fig. 1). The organization of the accessory gene regions and the fine structure of the LTR are of particular interest, since these differ among primate lentivirus lineages (12, 23). The organization both of the midgenomic accessory genes and of the *nef/env* region are like those of previously described HIV-1 isolates, with a *vpu* gene instead of *vpx* and with *vpr* overlapping *vif*. The open reading frames for *env* and *nef* are not overlapping, as in other HIV-1 genomes and in distinction to HIV-2 and simian immunodeficiency virus (SIV). Translation of the open reading frames in this CM240 sequence provides an opportunity to evaluate all of the encoded proteins. Figure 1 shows that the reading frames for *gag*, *pol*, *vif*, *vpr*, *tat*, *rev*, *env*, and *nef* encode proteins of the expected lengths and are without obvious inactivating defects. The *vpu* gene is interrupted by an in-frame stop codon at amino acid position 23. The sequences of the *gag* and *env* genes which were previously PCR amplified separately and sequenced (our unpublished data, compiled in reference 37) differ from their counterparts in the full-length sequence by 0.5 and 0.4%, respectively (data not shown). The complete sequence reported here appears to be representative of viruses from isolate CM240 and, with repair of the defect in *vpu* and assembly of the component clones, may provide a full-length, infectious molecular clone for further characterization.

In the LTR, the configuration of the core promoter/enhancer region was analyzed (Fig. 2). Primate lentiviruses typically possess three SP1 binding sites, and isolate CM240 is normal in this regard. In the HIV-1 lineage, available sequences from subtypes A, B, and D have two sites for NF-κB binding upstream of the SP1 sites. Evidence for diversity among HIV-1 clades in the number of NF-κB binding sites emerged with the first sequence from clade C, which possesses an additional site (46). In contrast to clades A, B, C, and D, isolate CM240 has only one canonical NF-κB binding site. Furthermore, the TATA box of this virus is TAAAA as in $HIV\text{-}1_{Z321}$ and the SIV_{AGM} vervet and grivet isolates, instead of the TATAA of other HIV-1, HIV-2, and the SIV_{AGM} sabaeus isolates.

Additional atypical features were observed in the transactivation response (TAR) stem-loop structure of isolate CM240. This structure, located at the 5′ end of the genomic RNA, regulates transcription by forming a binding site for the viral protein Tat and cellular proteins (4, 11, 19, 32). It has been shown that the critical regions for binding are the bulge on the side of TAR and the tip of the loop (4, 18). As shown in Fig. 2, CM240, consistent with the A clade, differs from the B, C, and D clades in having a two-base bulge instead of a three-base bulge. The tip of the TAR loop also differs by one base from those of HIV-1 clades A through D, being CCGGG instead of CTGGG.

Whatever recombination or other processes contributed to the development of this CM240 genome structure, an overall HIV-1-like genome organization was maintained. Some alterations, however, differentiate this isolate (and, conceivably, envelope clade E viruses in general; see reference 20) from the HIV-1 isolates sequenced to date in the promoter/enhancer region, in the TATA box, and in the TAR stem-loop structure.

The mosaic structure of CM240. We next determined which segments of this CM240 genome derived from the two apparent parental genotypes, clade A and clade E. Two methods of analysis were used, and they provided highly congruent results (Fig. 3). Using the maximum likelihood DNA distance algorithm (15), we constructed a matrix of pairwise distances. The between-clade range, that is, the maximum and minimum distance of the only available fully sequenced clade A strain, U455, to isolates of clades B, C, and D, was plotted for each subsegment; the distance of isolate CM240 to isolate U455 was plotted similarly (Fig. 3A). The CM240-U455 distance is smaller than the distance between clades until the beginning of envelope, whereupon the distance widens to a between-clade distance. The distance to U455 narrows after the membrane-spanning domain of gp41 and widens to a between-clade distance near the beginning of the U3 region of the LTR. A small segment comprising the 3′ portion of *vif* and the 5′ portion of

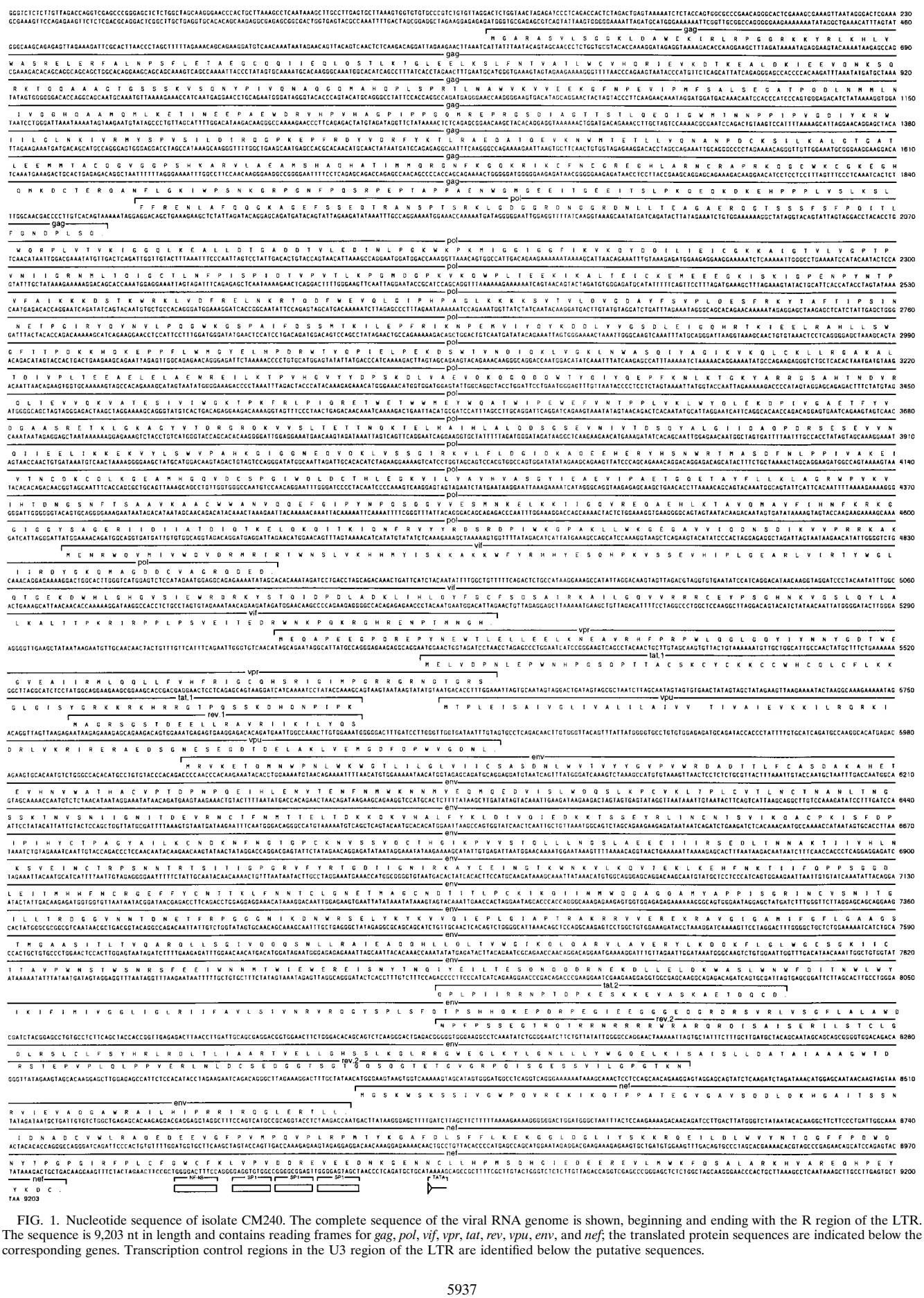

FIG. 1. Nucleotide sequence of isolate CM240. The complete sequence of the viral RNA genome is shown, beginning and ending with the R region of the LTR. The sequence is 9,203 nt in length and contains reading frames for *gag*, *pol*, *vif*, *vpr*, *tat*, *rev*, *vpu*, *env*, and *nef*; the translated protein sequences are indicated below the corresponding genes. Transcription control regions in the U3 region of the LTR are identified below the putative sequences.

TABLE 1. Genetic distances between HIV-1 *gag* and *env* genes

Isolate[a]	Clade	Distance to CM240[b]	
		gag	*env*
90CR402	Envelope E	6.4	7.1
U455	A	9.7	18.7
MN	B	15.2	18.7
C2220	C	16.0	20.5
ELI	D	14.9	20.0

[a] *gag* and envelope (gp160) genes from full-length genomes were extracted and aligned. Reference isolates were from reference 20, from the Los Alamos database (isolates U455, MN, and ELI), or from reference 46.

[b] Maximum likelihood distances (percent difference) between each isolate and isolate CM240.

vpr approached the lower boundary of a between-clade distance and may also be of clade E. However, this assignment is not as definitive as for other regions by this analysis. Thus, most, if not all, of isolate CM240 appears to derive from clade A except the external portion of the envelope (gp120 and the external portion of gp41) and a segment of the LTR.

Figure 3B shows results of bootscanning (45) using maximum parsimony. The bootstrap value (100 iterations) of the node joining isolates CM240 and 90CR402 with the clade A virus U455 is plotted for each segment of the genome. As before, CM240 joins U455 with high bootstrap values until the beginning of envelope, where the value falls and remains low until the membrane-spanning domain. Inclusion in clade A continues until the beginning of the LTR. A slight dip in the bootstrap value occurred in the 5′ portion of *vif*, but it was not as definitive as that seen in the envelope and LTR regions. Thus, a portion of *vif* remains unresolved by this analysis, leaving open the possibility of a small clade E segment in this region. These results are largely in accord with the distance matrix analysis described above.

Examples of the 37 phylogenetic trees from the bootscanning analysis are shown in Fig. 4. The bootstrap values of the nodes establishing whether CM240 and 90CR402 group with U455 are indicated. Bootstrap values of 95% or higher establish inclusion in clade A, while low bootstrap values draw the interpretation that isolates CM240 and 90CR402 cannot be assigned to clade A. The trees shown in Fig. 4 demonstrate that adjacent 500-nt segments yield trees with dramatically differing topologies. The breakpoints between clade A and non-clade A genetic material cannot be mapped precisely by our techniques, but we place the breakpoints within approximately 200 nt of their indicated locations in Fig. 4. A more detailed analysis of the envelope gene, using additional clade A isolates and smaller fragment sizes, confirmed these results (data not shown).

These analyses provide evidence that the CM240 genome has multiple breakpoints between genomic segments from different clades. Breakpoints near the beginning of *env*, at the membrane-spanning domain of gp41, and in the LTR have been identified by two independent methods and are conclusively established. The presence of additional breakpoints in *vif/vpr* is also a possibility, but the segments identified by parsimony and distance methods were disparate, precluding a definitive assignment. With the available data, it is not possible to determine whether the same parental strain (the putative clade E virus) contributed both the envelope and the LTR portions of the genome, but our analyses do establish that both segments are not derived from clades A through D; the simplest interpretation would be that they both come from a single parental strain. The remainder of the genome, about 75% of the genetic material, can be assigned to genotype A by independent methods of analysis; a segment in *vif/vpr* remained unresolved (but see reference 20).

Distribution of breakpoints and virion structure. Interclade recombinant HIV-1 strains are known to arise at appreciable frequency in the global epidemic (24, 29, 43), but for most, complete genomic sequence is unavailable. The detailed description of the mosaic genome found in isolate CM240 affords an opportunity to consider the full structure of the virion and, in turn, to open inquiry into the possible selective advantages associated with the pattern of breakpoints observed.

We have definitively mapped three breakpoints between segments of genetic material from different clades in CM240: one between nt 6350 and 6550, another between nt 8350 and 8550, and a third between nt 9150 and 9350 (nucleotide positions are numbered according to reference strain HIV_{MN}). An interpretation of the CM240 virion structure is presented in Fig. 5. It is striking that the matrix and core structure of the

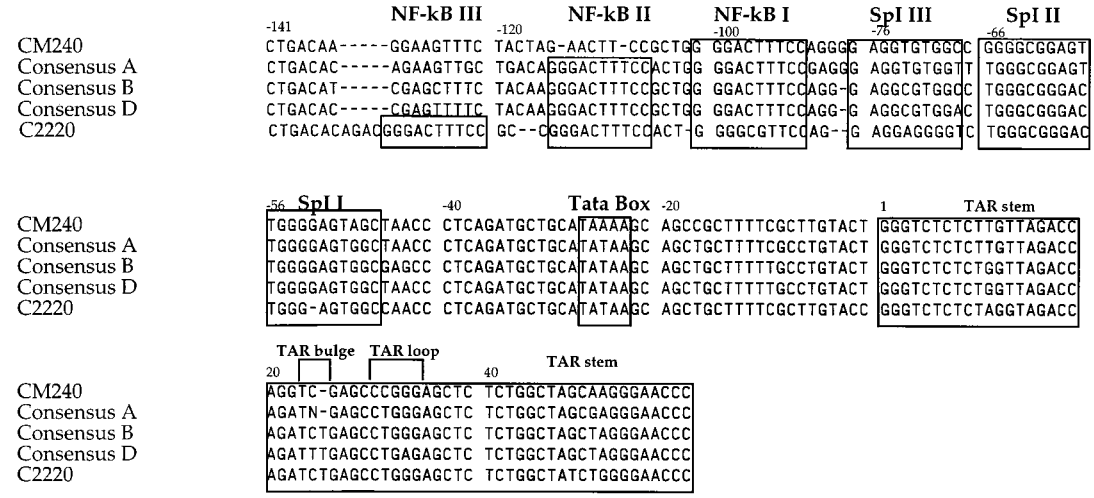

FIG. 2. Nucleotide sequence of the core promoter/enhancer region of the CM240 isolate. The nucleotide sequence of the core promoter/enhancer region of CM240 is shown aligned with the consensus sequences of clade A, clade B, and clade D viruses (39) and the C clade virus C2220 (45). Regulatory features are boxed and labeled above the alignment.

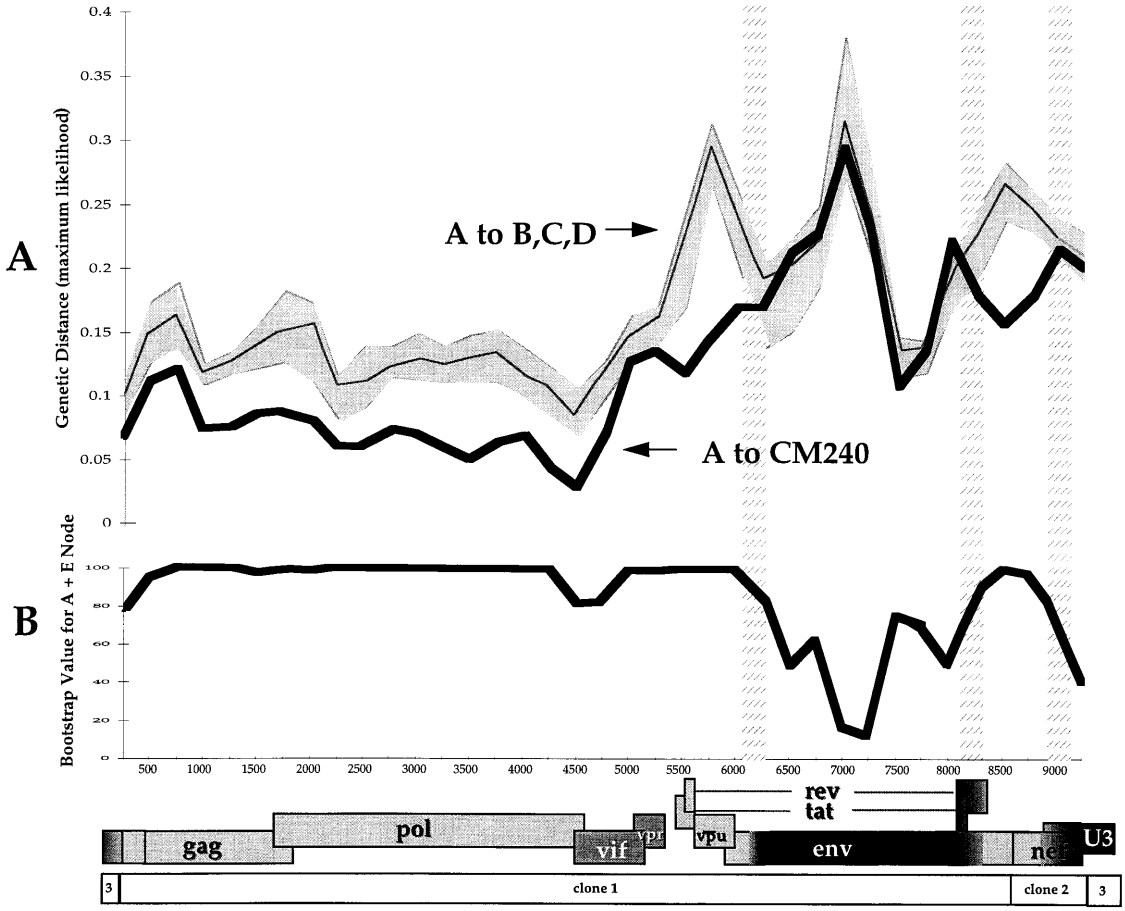

FIG. 3. Analysis of the mosaic structure of CM240. Thirteen full-length HIV-1 isolates (CM240, 90CR402, U455, MN, OY1, CAM1, HXB2, NL4-3, U23487, C2220, ELI, NDK, and Z2Z6) (38) were aligned, and the alignment was sectioned into 500-nt segments with 250-nt overlaps. Each of the resulting 37 segments was analyzed separately. Panel A shows the genetic distance calculated by using maximum likelihood between each pair of isolates. The shaded area indicated the range of distances between clade A (U455) and isolates of clades B, C, and D, and the mean of that distance is plotted in the thin gray line. The solid dark line shows the distance between clade A isolate U455 and isolate CM240. Panel B plots the bootstrap value for the node joining isolates U455, CM240, and 90CR402, out of 100 phylogenetic trees built by using parsimony. The distance value (A) and the bootstrap value (B) for each segment were plotted at the midpoint of the segment. Three vertical lines approximate the hypothesized breakpoints. The panel at the bottom shows a map of the open reading frames in the HIV-1 genome and indicates which segments of the genome of CM240 were derived from the three clones used to assemble the complete sequence.

virus, including Gag, Pol, and the internal portion of gp41, derives from clade A, while the external proteins (gp120 and the portion of gp41 N terminal to the membrane-spanning domain) derive from clade E. The structure is highly reminiscent of those virus pseudotypes that bear the coat of one virus and the core of another (5, 51), often with expanded host range or cellular tropism as a result. It may be of significance that three other recombinant viruses also contain an apparent breakpoint in the cytoplasmic portion of gp41 (isolates KE124, MAL, and VI525 [43]), particularly since these virus isolates are from widely separated locations and from other HIV-1 clades. At present, however, too few completely sequenced, recombinant forms are available to determine whether there are independent occurrences of mosaic genomes approximating the structure of CM240.

DISCUSSION

The HIV-1 strains first identified in the epidemic in Thailand have been known to be of epidemiological significance for some years, but a full genomic analysis has been hampered by technical limitations. Most of the initial isolates were unable to be adapted to continuous growth in T-cell lines, which has previously been important for the recovery of full-genomic clones. The development of long PCR (47) on the one hand and the persistent application of genomic library construction techniques using bacteriophage lambda (20) on the other have yielded the first prototypic strains. It is highly significant that the sequences obtained by different cloning methods and from Asian and African isolates, respectively, are quite similar over the entire genome. Indeed, isolates CM240 and 90CR402 were tightly clustered in 37 of 37 individual phylogenetic trees constructed from overlapping segments of the entire genome (Fig. 4 shows selected examples). Thus, an HIV-1 with a multiply mosaic genomic structure has established significant foci of infection in both Asia and Africa and is continuing to spread globally.

The genetic regulation of the expression of the HIV-1 virus is closely controlled. One of the regulatory factors which control HIV-1 expression is the transactivating host protein NF-κB (reviewed in references 22 and 48). All HIV-1 isolates have binding sites for NF-κB upstream from the TATA box,

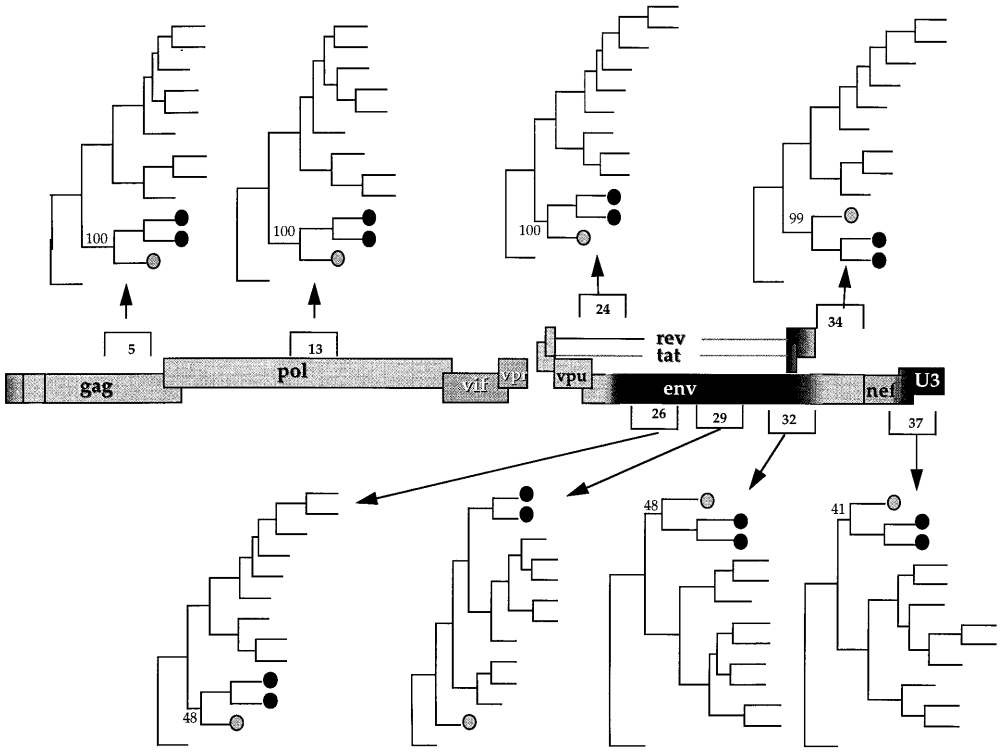

FIG. 4. Changing topology of phylogenetic trees for different segments of the CM240 genome. Selected phylogenetic trees (numbers 5, 13, 24, 26, 29, 32, 34, and 37, respectively) for 500-nt segments of the CM240 genome are shown. These were generated by using maximum parsimony on 100 bootstrapped resamplings of the data set. The clade A isolate, U455, is shown by the gray circle, and the envelope clade E isolates 90CR402 and CM240 are shown by black circles. The other isolates were from clades B, C, and D. All trees were rooted on the single clade C isolate. The bootstrap values for the node, if any, which grouped isolates U455, CM240, and 90CR402 are indicated. The trees interpreted as assigning CM240 to clade A or to clade E are shown above or below the genome, respectively. The map of the HIV-1 genome shows portions of CM240 derived from clade A in gray and those apparently derived from clade E in black, with shading to indicate the range of uncertainty with respect to the exact breakpoints.

but it is becoming clear that there are consistent differences between the clades in the number of such sites present. Previous work has shown that viruses of the C clade have three sites and that the A, B, and D clades have two sites, and it is now shown that the E clade, as represented by CM240, has only one site. The other clear difference in the CM240 transregulatory domain is found in sequence changes which should alter the structure of the TAR stem-loop. This RNA structure at the 5′ end of the viral genome forms a binding site for the viral protein Tat and for host proteins. In contrast to clades B, C, and D, CM240 has a two-base bulge on the side of the TAR stem, the primary binding site for Tat (44). The arginine-rich region of Tat, known to interact directly with the bulge and mediate transactivation (10, 14), has four arginines in CM240 instead of five or six. It is possible that these and other differences have phenotypic implications in terms of host cell specificity, but additional experiments would be needed to establish this directly.

Elements of the mosaic structure of CM240 that are firmly established include (i) the association of *gag*, *pol*, *tat*, *rev*, *vpu*, and the internal portion of the envelope with clade A and (ii) the derivation of the external envelope, in both gp120 and gp41 subunits, from clade E. We were unable to fully interpret the origin of the *nef*/LTR segment, other than establishing that it is not compatible with a clade A origin. It could be from the parental E strain or from another clade not available for comparison or could itself represent a multiply mosaic segment. The region around the *vif*/*vpr* genes was not resolvable by using the currently available clade A viruses, but it showed preliminary evidence suggestive of the presence of non-clade A segments. Availability of more clade A viral sequences will provide more power in this analysis.

The pattern of recombination breakpoints suggest a mechanism for the generation of at least part of the CM240 mosaic virus. Retroviral reverse transcription occurs in a transcription complex containing two viral genomes which are copackaged in the virion (3, 6, 53). Presumably most recombinants result from the copackaging of viral genomes from two different subtypes, forming a heterodimer. Since the 5′ end of the CM240 viral genome is from clade A and the 3′ end is not from clade A, it is possible that during reverse transcription, the strong stop DNA jumped from the 5′ end of a clade A genome to the 3′ end of a non-clade A genome. When this is permuted into the proviral form, it produces a chimeric LTR with R and U5 from clade A and U3 from a different clade, which is what is seen in CM240. Additional strand switching would be needed to generate the other breakpoints seen in CM240.

Other HIV-1 mosaic genomes also contain multiple breakpoints. The first described mosaic isolate, MAL, is a patchwork of genetic material from clades A and D, possessing several apparent breakpoints (43). A partial genomic analysis of other recombinant strains showed that even when only *gag* or *env* genes are considered, many contained multiple breakpoints (43). Thus, the multiple crossover points detected in CM240 may reflect a common mechanism of frequent strand switching by reverse transcriptase (26).

The process of reverse transcription may generate a myriad of recombinants in those who are infected with two different

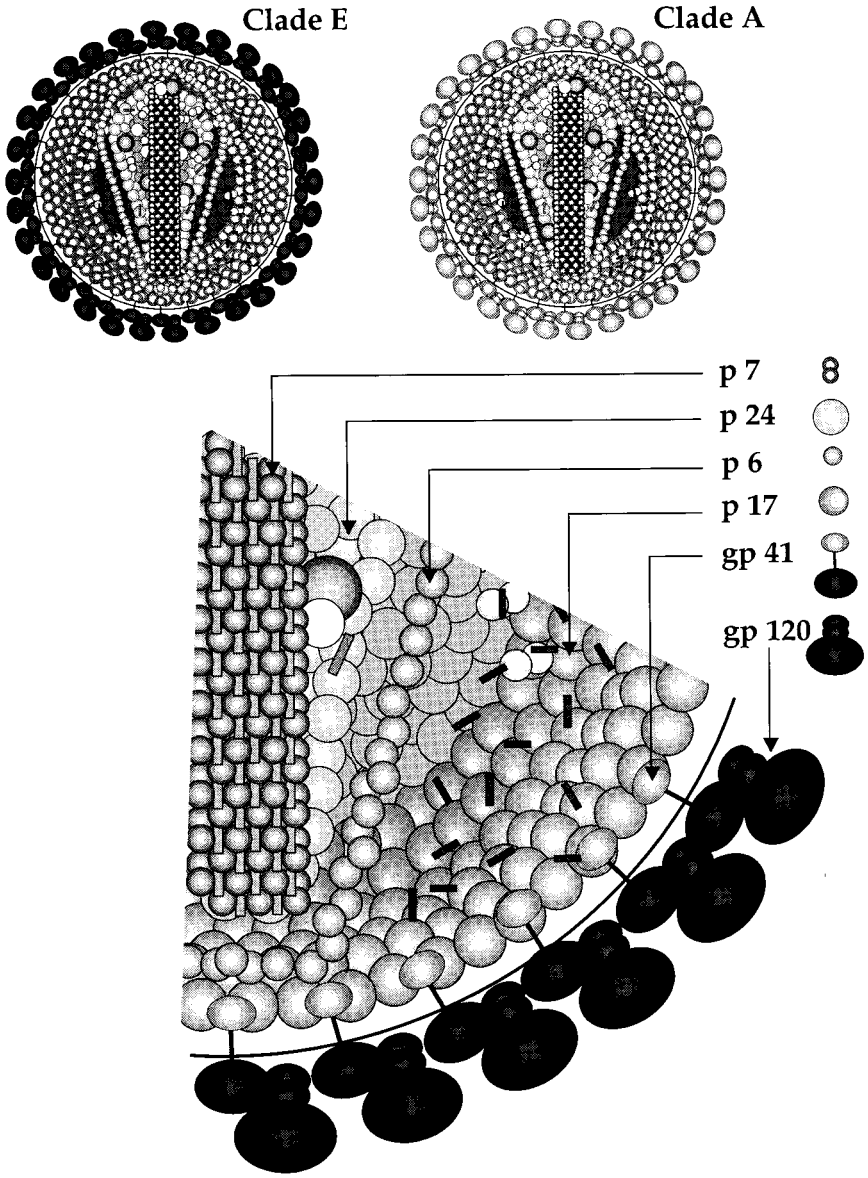

FIG. 5. Proposed structure of the CM240 virion. An overview the HIV-1 viral particle is shown at the top. A clade A virus (right) and portions of CM240 (left) that derive from clade A are depicted in light gray, while those portions that apparently come from clade E are shown in dark gray. In the expanded diagram of CM240, the association of the clade A, internal cytoplasmic portion of the envelope (gp41) with clade A matrix (p17) is shown, along with the hypothetical location of clade A core proteins including p7, p24, and p6. The entire portion of envelope external to the membrane (gp120 and the external domain of gp41) appears to be of a different clade, with the breakpoint positioned, within resolution of the techniques used, at the membrane-spanning domain. Not shown is another portion of non-clade A genetic material in the U3 region of the LTR. The structure may mimic that of viral pseudotypes, with the matrix and core of one virus and the envelope of another.

subtypes, but infectious virus would result only if the progeny were viable. There are many functions of the virus which are known to require interactions between and within different regions of the viral nucleic acid and viral proteins, and probably many more not yet known. It is reasonable to assume that some clade compatibility is required for these interactions. One example of this is the *tat* and *rev* coding regions. The first exons of *tat* and *rev* are in a clade A-derived region, as are the second exons, as a result of a crossover right before the second exons. It is possible that the amino acids encoded by the first exons have to be clade compatible with those encoded by the second exons. It would also be logical to expect compatibility between the viral protein Rev and its target binding site, a stem-loop structure on the Env-encoding RNA called the Rev-responsive element (13). This binding is essential for proper transport of the viral mRNA (25, 50). Interestingly, in this CM240 mosaic the *rev* gene is derived from a clade A lineage while the Rev-responsive element is clearly from a non-clade A lineage. The cross-clade compatibility of Rev binding has not been studied experimentally, but the existence of the CM240 isolate would suggest that there is compatibility, at least between clades A and the parental E.

The pseudotype structure of CM240 is suggestive of yet another aspect of the genetic plasticity of HIV-1. It may be that the exchange of the external envelope between clades confers a selective advantage in some circumstances. This would seem to be an evolutionary strategy far beyond the capability of point mutations, however frequent, to generate variants of

increased fitness. The variable configuration of NF-κB sites, the alteration in the TATA box sequence, and the differences in the TAR bulge and loop may be examples of genetic differences between HIV-1 clades which have the potential to produce significant biological or epidemiological effects. Full genomic analysis of other mosaic HIV-1 genomes will be required determine whether the breakpoints found in CM240 are recurrent and to discern whether other mosaic genomes predict virion alterations having relatively clear-cut functional implications.

With the availability of the components for assembly of a full-length infectious clone, the stage is set for a more complete understanding of the biological properties of the HIV-1 variant that predominates in Thailand. The speed with which these variants established themselves may have resulted from a combination of virologic, behavioral, and demographic factors, but the continuing spread of these strains, both elsewhere in Southeast Asia and more broadly, would suggest that this variant can establish itself successfully in a number of different circumstances. Future study of more viral isolates will reveal whether the mosaic pattern described for CM240 reflects a common adaptive strategy for HIV-1.

ACKNOWLEDGMENTS

We acknowledge Nelson Michael for providing the primers used to clone the LTR and Louis Henderson for the template used to generate Fig. 5. We also thank Feng Gao and Beatrice Hahn for making the African E virus sequence available to us in advance of publication.

This work was supported in part by Cooperative Agreement DAMD17-93-V-3004, between the U.S. Army Medical Research and Materiel Command and the Henry M. Jackson Foundation for the Advancement of Military Medicine.

REFERENCES

1. **Artenstein, A. W., J. Coppola, A. E. Brown, J. K. Carr, E. Sanders-Buell, E. Galbarini, J. R. Mascola, T. C. VanCott, P. Schonbrood, F. E. McCutchan, and D. S. Burke.** 1995. Multiple introductions of HIV-1 subtype E into the Western Hemisphere. Lancet **346:**1197–1198.
2. **Artenstein, A. W., P. A. Hegerich, C. Beyrer, K. Rungreungthanakit, N. L. Michael, and C. Natpratan.** 1996. Sequences and phylogenetic analysis of the *nef* gene from Thai subjects harboring subtype E HIV-1. AIDS Res. Hum. Retroviruses **12:**557–560.
3. **Bender, W., Y.-H. Chien, S. Chattopadhyay, P. K. Vogt, M. B. Gerdner, and N. Davidson.** 1978. High molecular-weight RNAs of AKR, NZB, and wild mouse viruses and avian reticuloendotheliosis virus all have similar dimer structures. J. Virol. **25:**888–896.
4. **Berkhout, B., R. H. Silverman, and K.-T. Jeang.** 1989. Tat transactivates the human immunodeficiency virus through a nascent RNA target. Cell **59:**273–282.
5. **Boettiger, D.** 1979. Animal virus pseudotypes. Prog. Med. Virol. **25:**37–68.
6. **Bowerman, B., P. O. Brown, J. M. Bishop, and H. E. Varmus.** 1989. Retroviral integration: structure of the initial covalent product and its precursor and a role for the viral IN protein. Proc. Natl. Acad. Sci. USA **86:**2525–2529.
7. **Brodine, S. K., J. R. Mascola, P. J. Weiss, S. I. Ito, K. R. Porter, A. W. Artenstein, F. C. Garland, F. E. McCutchan, and D. S. Burke.** 1995. Detection of diverse HIV-1 genetic subtypes in the United States. Lancet **346:** 1198–1199.
8. **Burke, D. S., A. K. Fowler, R. R. Redfield, S. Dilworth, and C. N. Oster.** 1990. Isolation of HIV-1 from the blood of seropositive adults: patient stage of illness and sample innoculum size are major determinants of a positive culture. J. Acquired Immune Defic. Syndr. **3:**1159–1167.
9. **Burke, D. S., and F. E. McCutchan.** Global distribution of HIV-1 clades. V. T. DeVita, Jr., S. Hellman, S. A. Rosenberg, M. E. Essex, A. S. Fauci, and J. S. Freeman (ed.), *In* AIDS: etiology, diagnosis, treatment and prevention, 4th ed., in press. J. B. Lipponcott Co., Philadelphia.
10. **Calnan, B. J., B. Tidor, S. Biancalana, D. Hudson, and A. D. Frankel.** 1991. Arginine-mediated RNA recognition: the arginine fork. Science **252:**1167–1171.
11. **Cullen, B. R.** 1990. The HIV-1 tat protein: and RNA sequence-specific processivity factor. Cell **63:**655–657.
12. **Cullen, B. R., and E. D. Garrett.** 1992. A comparison of regulatory features in primate lentiviruses. AIDS Res. Hum. Retroviruses **8:**387–393.
13. **Daly, T. J., K. S. Cook, G. S. Gray, T. E. Maione, and J. R. Rusche.** 1989. Specific binding of HIV-1 recombinant Rev protein to the Rev-responsive element in vitro. Nature (London) **342:**816–819.
14. **Endo, S.-I., S. Kubota, H. Siomi, A. Adachi, S. Oroszlan, M. Maki, and M. Hatanaka.** 1989. A region of basic amino-acid cluster in HIV-1 tat protein is essential for trans-acting activity and nucleolar localization. Virus Genes **3:**99–110.
15. **Felsenstein, J.** 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17:**368–376.
16. **Felsenstein, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:**783–791.
17. **Felsenstein, J.** 1989. PHYLIP-phylogenetic inference package (version 3.2). Cladistics **5:**164–166.
18. **Feng, S., and E. C. Holland.** 1988. HIV-1 tat trans-activation requires the loop sequence within tar. Nature (London) **334:**165–167.
19. **Frankel, A. D.** 1992. Activation of HIV transcription by Tat. Curr. Opin. Genet. Dev. **2:**293–298.
20. **Gao, F., D. L. Robertson, S. G. Morrison, H. Hui, S. Craig, P. N. Fultz, J. Decker, M. Griard, 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., in press.
21. **Global AIDS News.** 1993. Cambodia faces new threat of AIDS. Newsl. W. H. O. Global Programme AIDS **2:**8.
22. **Grilli, M., J. S. Chiu, and M. J. Lenardo.** 1993. NF-kappaB and Rel: participants in a multi-form transcriptional regulatory system. Int. Rev. Cytol. **143:**1–62.
23. **Guyader, M., M. Emerman, P. Sonigo, F. Clavel, L. Montagnier, and M. Alizon.** 1987. Genome organization and transactivation of the human immunodeficiency virus type 2. Nature (London) **326:**662–669.
24. **Hahn, B. H., D. L. Robertson, F. E. McCutchan, and P. M. Sharp.** 1994. Recombination and diversity of HIV: implications for vaccine development. Neuvième Colloque des Cent Gardes, Paris.
25. **Hammarskjold, M.-L., J. Heimer, B. Hammarskjold, I. Swangman, L. Albert, and D. Rekosh.** 1989. Regulation of human immunodeficiency virus env expression by the rev gene product. J. Virol. **63:**1959–1966.
26. **Hu, W.-S., and H. M. Temin.** 1990. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. Proc. Natl. Acad. Sci. USA **87:**1556–1560.
27. **Janessens, W., L. Heyndrickx, K. Fransen, J. Motte, M. Peeters, J. N. Nkengasong, P. M. Ndumbe, E. Delaporte, J. L. Perret, C. Atende, P. Piot, and G. van der Groen.** 1994. Genetic and phylogenetic analysis of env subtypes G and H in Central Africa. AIDS Res. Hum. Retroviruses **10:**877–879.
28. **Kalish, M. L., A. Baldwin, S. Raktham, C. Wasi, C.-C. Luo, G. Schochetman, T. D. Mastro, N. Young, S. Vanichseni, H. Rubsamen-Waigmann, H. von Briesen, J. I. Mullins, E. Delwart, B. Herring, J. Esparza, W. I. Heyward, and S. Osmanov.** 1995. The evolving molecular epidemiology of HIV-1 envelope subtypes in injecting drug users in Bangkok, Thailand: implications for HIV vaccine trials. AIDS **9:**851–857.
29. **Leitner, T., D. Escanilla, S. Marquina, J. Wahlberg, C. Brostrom, H. B. Hansson, M. Uhlen, and J. Albert.** 1995. Biological and molecular characterization of subtype D, G, and A/D recombinant HIV-1 transmissions in Sweden. Virology **209:**136–146.
30. **Louwagie, J., W. Janssens, J. Mascola, L. Heyndrickx, P. Hegerich, G. van der Groen, F. E. McCutchan, and D. S. Burke.** 1995. Genetic diversity of the envelope glycoprotein from human immunodeficiency virus type 1 isolates of African origin. J. Virol. **69:**263–271.
31. **Louwagie, J., F. McCutchan, M. Peeters, T. P. Brennan, E. Sanders-Buell, G. Eddy, G. van der Groen, K. Fransen, G. M. Gershy-Damet, R. Deleys, and D. S. Burke.** 1993. Phylogenetic analysis of gag genes from 70 international HIV-1 isolates provides evidence for multiple genotypes. AIDS **7:**769–780.
32. **Marciniak, R. A., B. J. Calnan, A. D. Frankel, and P. A. Sharp.** 1990. HIV-1 tat protein trans-activates transcription in vitro. Cell **63:**791–802.
33. **McCutchan, F. E., P. Hegerich, T. Brennan, P. Phanuphak, P. Preecha, T. Achasa, P. Berman, A. K. Fowler, and D. S. Burke.** 1992. Genetic variants of HIV-1 in Thailand. AIDS Res. Hum. Retroviruses **8:**1887–1895.
34. **Michael, N. L., L. D'Arcy, P. K. Ehrenberg, and R. R. Redfield.** 1994. Naturally occurring genotypes of the human immunodeficiency virus type 1 display a wide range of basal and *tat*-induced transcriptional activities. J. Virol. **68:**3163–3174.
35. **Murphy, E., B. Korber, M.-C. Georges-Courbot, B. You, A. Pinter, D. Cook, M.-P. Kieny, A. Geogres, C. Mathiot, F. Barré-Sinoussi, and M. Girard.** 1993. Diversity of V3 region sequences of human immunodeficiency viruses type 1 from the Central African Republic. AIDS Res. Hum. Retroviruses **9:**997–1006.
36. **Myers, G., B. Korber, J. A. Berzofsky, R. F. Smith, and G. N. Pavlakis.** 1992. 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. Mex.
37. **Myers, G., B. Korber, S. Wain-Hobson, K. T. Jeang, L. E. Henderson, and G. N. Pavlakis.** 1994. 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. Mex.
38. **Myers, G., B. Korber, S. Wain-Hobson, K. T. Jeang, L. E. Henderson, and**

G. N. Pavlakis. 1995. 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. Mex.

39. **Myers, G., A. B. Rabson, S. F. Josephs, T. F. Smith, and F. Wong-Staal.** 1988. 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. Mex.
40. **Nkengasong, J. N., W. Janssens, L. Heyndrickx, K. Fransen, P. M. Ndumbe, J. Motte, A. Leonaers, M. Ngolle, J. Ayuk, P. Piot, and G. van der Groen.** 1994. Genetic subtypes of HIV-1 in Cameroon. AIDS **8:**1405–1412.
41. **Ou, C. Y., Y. Takebe, C. C. Luo, M. Kalish, W. Auwanit, C. Bandea, N. de la Torre, H. D. Gayle, N. L. Young, and B. G. Weniger.** 1992. Wide distribution of two subtypes of HIV-1 in Thailand. AIDS Res. Hum. Retroviruses **8:**1471–1472.
42. **Robertson, D. L., B. H. Hahn, and P. M. Sharp.** 1995. Recombination in AIDS viruses. J. Mol. Evol. **40:**249–259.
43. **Robertson, D. L., P. M. Sharp, F. E. McCutchan, and B. H. Hahn.** 1995. Recombination in HIV-1. Nature (London) **374:**124–126.
44. **Roy, S., U. Delling, C. H. Chen, C. A. Rosen, and N. Sonenberg.** 1990. A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated trans-activation. Genes Dev. **4:**1365–1373.
45. **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.
46. **Salminen, M. O., B. Johansson, A. Sonnerborg, S. Ayehunie, D. Gotte, P. Leinikki, D. S. Burke, and F. E. McCutchan.** Full-length sequence of an Ethiopian human immunodeficiency virus type 1 (HIV-1) isolate of genetic subtype C. AIDS Res. Hum. Retroviruses, in press.
47. **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.
48. **Siebenlist, U., G. Franzoso, and K. Brown.** 1994. Structure, regulation and function of NF-kappaB. Annu. Rev. Cell Biol. **10:**405–455.
49. **Sirisopana, N., K. Torugsa, C. J. Mason, L. E. Markowitz, R. A. Michael, D. S. Burke, A. Jugsudee, T. Supapongse, C. Chuenchitra, P. Singharaj, A. E. Johnson, J. G. McNeil, F. E. McCutchan, and J. K. Carr.** 1996. Correlates of HIV-1 seropositivity among young men in Thailand, J. Acquired Immune Defic. Syndr. **11:**492–498.
50. **Sodroski, J., W. C. Goh, C. Rosen, A. Dayton, E. Terwilliger, and W. Haseltine.** 1986. A second post-transcriptional trans-activator gene required for HTLV-III replication. Nature (London) **321:**412–417.
51. **Spector, D. H., E. Wade, D. A. Wright, V. Koval, C. Clark, D. Joquish, and S. A. Spector.** 1990. Human immunodeficiency virus pseudotypes with expanded cellular and species tropism. J. Virol. **64:**2298–2308.
52. **Srinivasan, A., D. York, D. Butler, Jr., 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 to recent isolates and generation of hybrid HIV. AIDS Res. Hum. Retroviruses **5:**121–129.
53. **Stuhlmann, H., and P. Berg.** 1992. Homologous recombination of copackaged retrovirus RNAs during reverse transcription. J. Virol. **66:**2378–2388.
54. **Wasi, C., B. Herring, S. Rakthan, S. Vanichseni, T. D. Mastro, N. L. Young, H. Rubsamen-Waigmann, H. von Briesen, M. L. Kalish, C.-C. Luo, C.-P. Pau, A. Baldwin, J. I. Mullins, E. L. Delwart, B. Herring, J. Esparza, W. L. Heyward, and S. Osmanov.** 1995. Determination of HIV-1 subtypes in injecting drug users in Bangkok, Thailand using peptide-binding enzyme immunoassay and heteroduplex mobility assay: evidence of increasing infection with HIV-1 subtype E. AIDS **9:**843–849.
55. **Weniger, B. G., K. Limpakarnajanarat, K. Ungchusak, S. Thanprasertsuk, K. Choopanya, S. Vanichseni, T. Uneklabh, and P. Thongcharoen.** 1991. The molecular epidemiology of HIV-1 infection and AIDS in Thailand. AIDS **5**(Suppl. 2)**:**S71–S85.
56. **Weniger, B. G., Y. Takebe, C. Y. Ou, and S. Yamazaki.** 1994. The molecular epidemiology of HIV in Asia. AIDS **8**(Suppl. 2)**:**S13–S28.
57. **WHO Network for HIV Isolation and Characterization.** 1994. HIV type 1 variation in World Health Organization-sponsored vaccine evaluation sites: genetic screening, sequence analysis, and preliminary biological characterization of selected viral strains. AIDS Res. Hum. Retroviruses **10:**1327–1344.