The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin

JOURNAL OF VIROLOGY, Oct. 1996, p. 7013–7029 Vol. 70, No. 10
0022-538X/96/$04.00+0

The Heterosexual Human Immunodeficiency Virus Type 1 Epidemic in Thailand Is Caused by an Intersubtype (A/E) Recombinant of African Origin

FENG GAO,[1] DAVID L. ROBERTSON,[2] SANDRA G. MORRISON,[1] HUXIONG HUI,[1] STEVENSON CRAIG,[1] JULIE DECKER,[1] PATRICIA N. FULTZ,[3] MARC GIRARD,[4] GEORGE M. SHAW,[1,3] BEATRICE H. HAHN,[1,3]* AND PAUL M. SHARP[2]

Department of Medicine[1] and Department of Microbiology,[3] University of Alabama at Birmingham, Birmingham, Alabama 35294; Department of Genetics, University of Nottingham, Queens Medical Centre, Nottingham, United Kingdom[2]; and Laboratory of Molecular Virology, Institut Pasteur, Paris, France 75015[4]

Received 26 March 1996/Accepted 19 June 1996

Since 1989, human immunodeficiency virus type 1 (HIV-1) has spread explosively through the heterosexual population in Thailand. This epidemic is caused primarily by viruses classified as "subtype E", which, on the basis of limited sequence comparisons, appear to represent hybrids of subtypes A (*gag*) and E (*env*). However, the true evolutionary origins of "subtype E" viruses are still obscure since no complete genomes have been analyzed, and only one full-length subtype A sequence has been available for phylogenetic comparison. In this study, we determined full-length proviral sequences for "subtype E" viruses from Thailand (93TH253) and the Central African Republic (90CR402) and for a subtype A virus from Uganda (92UG037). We also sequenced the long terminal repeat (LTR) regions from 16 virus strains representing clades A, C, E, F, and G. Detailed phylogenetic analyses of these sequences indicated that "subtype E" viruses do indeed represent A/E recombinants with multiple points of crossover along their genomes. The extracellular portion of *env*, parts of *vif* and *vpr*, as well as most of the LTR are of subtype E origin, whereas the remainder of the genome is of subtype A origin. The possibility that the discordant phylogenetic positions of "subtype E" viruses in *gag*- and *env*-derived trees are the result of unusual rates or patterns of evolution was also considered but was ruled out on the basis of two lines of evidence: (i) phylogenetic trees constructed for synonymous and nonsynonymous substitutions yielded the same discordant branching orders for "subtype E" *gag* and *env* gene sequences, thus excluding selection-driven evolution, and (ii) multiple crossovers in the viral genome are most consistent with the copy choice model of recombination and have been observed in other documented examples of HIV-1 intersubtype recombination. Thai and CAR "subtype E" viruses exhibited the same pattern of A/E mosaicism, indicating that the recombination event occurred in Africa prior to the spread of virus to Asia. Finally, all "subtype E" viruses were found to contain a distinctive two-nucleotide bulge in their transactivation response (TAR) elements. This feature was present only in viruses which also contained a subtype A 5′ *pol* region (i.e., subtype A viruses or A/D and A/E recombinants), raising the possibility of a functional linkage between the TAR region and the polymerase. The implications of epidemic spread of a recombinant HIV-1 strain to viral natural history and vaccine development are discussed.

Genetic diversity is a hallmark of human immunodeficiency virus type 1 (HIV-1) biology and poses a major obstacle to both drug and vaccine development (6, 32, 54, 69, 76, 84, 95). Several factors are known to contribute to the generation of new variants and influence the speed with which these viruses evolve. One is the error-prone nature of the viral reverse transcriptase (RT), which lacks proofreading functions and introduces substitutions at a rate of approximately 3×10^{-5} nucleotide changes per site per replication cycle (64). A second factor is the high rate of virus production (up to 10^{10} virions per day) and the large numbers of replication cycles (approximately 300 per year) that sustain HIV-1 infection in vivo (12, 38, 83, 110). A third factor is in vivo selection pressure, which, for example, is responsible for the rapid emergence of drug-resistant viral escape mutants (110). Together, these mechanisms generate viral variants at extraordinary rates (12) and represent a major force driving HIV-1 evolution in infected populations worldwide. However, recent evidence suggests that recombination among highly divergent HIV-1 strains occurs quite frequently, indicating that this process may also contribute importantly to HIV-1 diversification (15, 16, 25, 56, 85, 86, 88, 91, 99).

Recombination is a fundamental property of retroviruses resulting from their dimeric RNA genome and an RT that can switch between templates during proviral DNA synthesis (10, 31, 40, 41, 47, 105). In fact, it has been suggested that recombination is a necessary component of the viral replication strategy, since breaks in either of the two RNA molecules are common (the forced copy choice model proposed by J. Coffin [10]). In this process, the RT switches between heterozygous RNA templates as DNA synthesis proceeds. As the number of crossovers increases, so does the complexity of the resulting hybrid genome. Until recently, recombination in HIV-1 was thought to involve only the closely related members of the quasispecies that evolve within infected individuals over time (14, 33, 39, 90, 108). However, analysis of sequences submitted to the HIV-1 database indicates that recombination also occurs between highly divergent strains of HIV-1 (34, 91, 99), thus

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

raising the possibility that HIV-1 variants can acquire biologically important traits through exchange of genetic material.

Globally circulating strains of HIV-1 have been classified on the basis of their phylogenetic relationships into genetic groups (termed M and O) and subtypes (A to I) (51, 61, 75, 96). The great majority of these viruses appears to have evolved in a clonal fashion, because evolutionary trees derived from different regions of their genomes exhibit consistent branching orders (75, 96). However, recent phylogenetic studies of HIV-1 group M viruses have identified a surprising number of strains which cluster in different sequence subtypes depending on which part of the genome is analyzed (16, 25, 56, 58, 85, 86, 88). This finding, along with the fact that all of these viruses originated from geographic regions where the same divergent sequence subtypes are known to cocirculate, strongly suggests that these viruses are the product of recombination events. Consistent with the copy choice and other models of recombination (11), most of these hybrid viruses exhibit complex genome structures characterized by multiple crossovers (85, 86). Virus strains representing all of the major HIV-1 group M subtypes have been found to be involved in these recombination events (86). Intersubtype recombinant viruses have been identified in primary patient material (25, 88), in sequential isolates obtained from infected individuals over time (25, 93), and in heterosexual transmission pairs (56, 88, 93). Taken together, these results suggest that intersubtype recombination may be a relatively common occurrence in populations experiencing epidemic spread and intermixing of multiple sequence subtypes (97).

One of the most interesting and clinically important examples of potential HIV-1 recombinants are the so-called "subtype E" viruses, which are most prevalent in Thailand and neighboring countries in Southeast Asia (111, 112). Having been introduced into Thai populations relatively late in the global epidemic (between 1988 and 1989), these viruses spread rapidly and through largely heterosexual transmission routes (46, 52, 77, 79, 100, 109). Of the 750,000 individuals estimated to be infected with HIV-1 in Thailand to date, about 90% are thought to harbor members of "subtype E" (13, 111). At the genetic level, Thai "subtype E" viruses are all highly related to each other, presumably reflecting a founder effect (67, 80). They are also related to the more heterogeneous "subtype E" viruses from the Central African Republic (CAR viruses) and Cameroon (72, 78), suggesting that the African viruses represented the source of infection for the Thai epidemic (72, 73). Evidence that both Thai and CAR "subtype E" viruses may represent recombinants has been derived from phylogenetic studies of *gag* and *env* regions (62, 67, 72, 96). In the extracellular portion of gp120 and gp41, "subtype E" viruses have long been known to cluster as a distinct group, a finding which led to their initial classification as an independent subtype (74). In contrast, in the *gag* region, all "subtype E" viruses thus far characterized fall within the subtype A radiation (62). The simplest explanation for these discordant branching orders is that "subtype E" viruses are the descendants of a hybrid HIV-1, generated by recombination between members of subtypes A and E (25, 86, 96). However, the absence of a *gag* subtype E equivalent, as well as the lack of sequence information in regions other than *gag* and *env*, have left the phylogenetic origins of "subtype E" viruses open to question.

To elucidate the evolutionary history of "subtype E" viruses, we have cloned and characterized full-length proviral sequences from two isolates, one from Thailand (93TH253) and the other (90CR402) from the Central African Republic. Using a variety of methodologies, we have investigated their phylogenetic subtype assignments in different parts of the genome and have examined whether their discordant branching orders for *gag* and *env* gene sequences could be the result of unusual patterns of gene evolution rather than recombination. Our results indicate that "subtype E" viruses do indeed represent A/E intersubtype recombinants. Moreover, comparison of the CAR and Thai strains indicates that they are derived from a common A/E recombinant ancestor, presumably originating in Central Africa, where both subtype A and the parental (nonrecombinant) subtype E must have coexisted (72, 78). These data thus indicate that intersubtype recombination can generate virulent HIV-1 strains which have the capacity to significantly influence the course of the global AIDS epidemic.

MATERIALS AND METHODS

Virus isolates. All virus isolates used for sequence comparisons in this study have been described previously (72, 113, 114) and are listed according to World Health Organization nomenclature, denoting the year of isolation, the country of origin (two-letter code), and the isolate number (50). 92RW020, 92UG031, 92UG037, 92BR025, 93MW959, 93MW960, 93MW965, 92TH006, 92TH022, 93TH976, 93TH966, 93BR019, 93BR020, and 93BR029 were obtained from World Health Organization and National Institutes of Allergy and Infectious Diseases-sponsored Networks for Virus Isolation and Characterization and represent primary HIV-1 isolates from asymptomatic individuals early in infection (25, 26, 113). 92UG975 and 93RU131 were derived from two AIDS patients with subtype G infections who originated from Uganda and Russia, respectively (5, 45). 90CR402 (previously also termed CAR-E 4002) was obtained from an African man who presented with lymphadenopathy, diarrhea, severe weight loss, and recurrent respiratory infections at the University Hospital in Bangui, Central African Republic (72). 93TH253 (previously termed CMU010 or 302053) was isolated from a Thai patient with end-stage AIDS who was hospitalized at Chiang Mai University Medical Center (114). Subtype designation was made on the basis of partial or complete *env* gene sequences (72, 113, 114). All isolates were established and subsequently propagated by cocultivation with normal donor lymphocytes (72, 113, 114), except for 90CR402, which was passaged in chimpanzee peripheral blood mononuclear cells (PBMCs) (22), and 93TH253, which was adapted to growth in immortalized human T-cell lines (unpublished data). For lambda phage cloning, isolates 90CR402 and 93TH253 were expanded in human PBMCs and H9 cells, respectively.

Southern blot analysis. Productively infected 90CR402 (PBMC) and 93TH253 (H9) cultures were amplified by addition of uninfected target cells, monitored for RT activity, and harvested (2×10^8 cells) at peak virus replication. Cells were washed extensively in Hanks balanced salt solution, lysed in 10 ml of lysing buffer (10 mM Tris-HCl [pH 7.5], 5 mM EDTA, 0.5% sodium dodecyl sulfate) containing proteinase K (100 μg/ml), and incubated at 37°C overnight. High-molecular-weight DNA was extracted twice with phenol-chloroform (1:1) and once with chloroform, precipitated with 2 volumes of absolute ethanol, and dissolved in TE buffer (10 mM Tris-HCl [pH 7.6], 1 mM EDTA). Then 10 μg of culture DNA was digested with restriction endonucleases, separated by electrophoresis on a 0.7% agarose gel, blotted onto nitrocellulose filters, hybridized to a ^{32}P-labeled full-length HIV-1 probe (BH10 [35]), and exposed to X-ray film as described previously (63). The resulting autoradiogram was used to estimate the relative abundance of integrated and unintegrated viral DNA and to identify restriction enzymes suitable for lambda phage cloning (63).

Lambda phage cloning of full-length HIV-1 proviruses. Genomic lambda phage libraries were constructed and screened as described previously (29, 59, 98). Briefly, high-molecular-weight DNA from productively infected cultures was digested with an appropriate cloning enzyme (*Bam*HI for isolate 90CR402; *Eco*RI for isolate 93TH253), fractionated by sucrose gradient centrifugation to enrich for fragments 15 to 22 kb in length, and ligated into purified arms of λDASHII (Stratagene, La Jolla, Calif.). Ligation products were packaged in vitro (Gigapack II Gold; Stratagene), titered, and plated on LE392 cells. Recombinant phage clones (20,000 plaques per plate) were screened with a full length HIV-1 probe (BH10 [35]). Positive phage recombinants were plaque purified, and their restriction maps were determined by multiple enzyme digestions. Phage clones containing complete (integrated) HIV-1 proviruses were digested with restriction enzymes, and their inserts were subcloned into appropriate plasmid vectors (digestion of λ93TH253.3 with *Eco*RI released a 15.2-kb insert which was subcloned into pTZ18R; digestion of λ90CR402.1 with *Bam*HI released a 17.8-kb insert which was subcloned into pSP72).

DNA transfection and viral infectivity studies. Plasmid subclones of lambda phage-derived (full-length) HIV-1 proviruses were transfected into 293T or COS-1 cells by the calcium phosphate precipitation method (3). Sixty hours posttransfection, culture supernatants were analyzed for RT activity, filtered, and used to infect CD4-bearing target cells (primary PBMCs and T-cell lines). Cultures were monitored weekly for RT activity and maintained for 2 months. Transfection-derived virus preparations were also examined by Western blot (immunoblot) analysis as described previously (60).

Amplification of near-complete HIV-1 proviruses using long-PCR methods. Amplification of a near-full-length provirus from isolate 92UG037 was performed by using the long-PCR procedure (GeneAmp XL kit; Perkin-Elmer Cetus, Foster City, Calif.) according to the manufacturer's instructions. Primers were designed to amplify HIV-1 proviral sequences between the tRNA primer binding site (primer UP1A, 5′-AGTGGCGCCCGAACAGG-3′) and the poly(A) signal in the 3′ long terminal repeat (LTR) (primer LOW2, 5′-TGAG GCTTAAGCAGTGGGTTTC-3′). Cycling conditions included a hot start (2 min, 94°C), then 20 cycles of denaturation (94°C) for 30 s and extension (68°C) for 10 min, and then 17 cycles of denaturation (94°C) for 30 s and extension (68°C) for 10 min, with 15-s increments per cycle. PCR products were visualized by agarose gel electrophoresis and subcloned into pCRII by T/A overhang. Transformations were performed in TG-1 cells. Colonies were screened by restriction enzyme digestion for full-length inserts.

Sequencing of HIV-1 proviruses. Nucleotide sequences of full-length (or near-full-length) HIV-1 proviruses were determined by the shotgun sequencing approach as described previously (65). Briefly, recombinant phage or plasmid inserts containing a complete provirus were purified by gel electrophoresis and sonicated (model XL2020 sonicator; Heat Systems Inc., Farmingdale, N.Y.) to generate randomly sheared DNA fragments 600 to 1,000 bp in length. Following isolation and purification by gel electrophoresis, fragments were end repaired by using T4 DNA polymerase and Klenow enzyme (65), and blunt-ended DNA fragments were then ligated into M13 or pTZ18R cleaved with *Sma*I and dephosphorylated with calf intestinal alkaline phosphatase to reduce vector self-ligation. pTZ18R colonies or M13 phage plaques were screened by hybridization with a full-length HIV-1 probe (BH10 [35]). Positive colonies or plaques were picked, and vectors were grown by using standard methods (63). Sequence analysis was performed by using cycle sequencing and dye terminator methodologies on an automated DNA Sequenator (model 373A; Applied Biosystems, Inc.) according to the manufacturer's recommendations. On average, 200 shotgun clones per provirus were sequenced. Proviral contigs were assembled from individual sequences by using the Sequencher program (Gene Codes Corporation, Ann Arbor, Mich.). Small gaps were filled in by direct sequencing of the cloned provirus, using strain-specific sequencing primers. Sequences were determined for both strands of DNA.

Sequence analysis. Sequences were analyzed by using Eugene (Baylor College of Medicine, Houston, Tex.), PIMA (101, 102), and MASE (17). Secondary RNA structures of transactivation response (TAR) elements were predicted and plotted by using the programs MFold and PlotFold from the University of Wisconsin Genetics Computer Group software package (version 8).

PCR amplification of LTR sequences. A 670-bp fragment spanning most of the viral LTR was amplified from cultured PBMC DNA of isolates representing subtypes A (92RW020, 92UG031, and 92UG037), C (92BR025, 93MW959, 93MW960, and 93MW965), E (92TH006, 92TH022, 93TH976, and 93TH966), F (93BR019 and 93BR029), and G (92RU131 and 92UG975). PCR amplifications were performed with primers LTR A (5′-CAAGGATCCTTCCCTGATTGGC AGAACTAC-3′) and PBS B (5′-CTTAATACTGACGCTCTCGCACCCAT-3′) in the first round and LTR A and PBS D (5′-CTCTCCTTCTAGCCTCCGCTAGTC-3′) in the second round in a seminested approach (denaturing at 94°C for 1 min, annealing at 55°C for 1 min, and extension at 72°C for 1 min; 30 cycles). Amplified fragments were sequenced directly (without prior subcloning), using cycle sequencing and dye terminator methodologies as described above.

Phylogenetic tree constructions. Phylogenetic relationships of newly derived viral sequences were estimated from comparisons with those of previously reported representatives of HIV-1 group M. Nucleotide sequences were aligned by using CLUSTAL (37), with minor manual adjustments, bearing in mind the predicted protein sequences. Pairwise evolutionary distances were estimated by using Kimura's two-parameter method (49) to correct for superimposed substitutions; sites where there was a gap in any sequence and ambiguous areas in the alignments were excluded from all comparisons. Distances were also calculated for synonymous and nonsynonymous substitutions separately, using Li's method (57). Phylogenetic trees were constructed by the neighbor-joining method (89), and their reliability was estimated from 1,000 bootstrap replicates (18). These methods were implemented by using CLUSTAL W (106) and programs from the PHYLIP package (19).

Nucleotide sequence accession numbers. GenBank accession numbers for the full-length HIV-1 proviral sequences reported in this study are U51188 (90CR402.1), U51189 (93TH253.1), and U51190 (92UG037.1). LTR sequences are available under accession numbers U51282 to U51297.

RESULTS

Lambda phage cloning of full-length "subtype E" proviruses from Thailand and the Central African Republic. Full-length "subtype E" proviral genomes were cloned to investigate their genomic organization and phylogenetic origins and to generate reference reagents useful in vaccine development. For this purpose, we selected two isolates, one (93TH253) representative of strains currently spreading epidemically in Thailand (114) and the other (90CR402) representative of strains from Central Africa that are believed to have been the source of the Thai epidemic (72). In addition, 90CR402 was chosen because it constitutes a highly cytopathic strain which grows to high titers in both human and chimpanzee PBMCs (3a, 22) and thus has been used as a challenge strain in chimpanzee vaccination and superinfection studies (23). Both isolates were initially classified as "subtype E" on the basis of V3 loop sequences (72, 114).

We selected a lambda phage cloning approach to obtain full-length "subtype E" genomes because this method allows the cloning of complete proviruses as single genomic units (29, 59, 63, 98) and avoids the possibility of PCR artifacts due to misincorporation of nucleotides or other polymerase-related errors. Lambda phage cloning requires higher proviral loads than PCR approaches and thus depends on efficient viral growth in culture. Figure 1A depicts a Southern blot of isolates 90CR402 and 93TH253 grown in PBMCs and H9 cells, respectively, and illustrates the low quantities of proviral DNA that we routinely observed in "subtype E"-infected PBMC cultures. Of several "subtype E" isolates analyzed, 90CR402 replicated by far the best in primary PBMC cultures. Three other Thai isolates (including 93TH253) replicated poorly despite multiple passages and frequent addition of uninfected target cells. Because the proviral DNA content in these cultures was insufficient for lambda phage cloning purposes (data not shown), we selected H9-propagated 93TH253, along with PBMC-grown 90CR402, for the construction of genomic libraries.

Restriction enzyme analysis of proviral DNA identified *Bam*HI and *Eco*RI as suitable (no-cutter) cloning enzymes for 90CR402 and 93TH253, respectively (data not shown). Genomic DNA was digested to completion and ligated into purified arms of λDASHII. Libraries were titered and screened with a full-length (subtype B) HIV-1 probe (35). For 93TH253, screening of 2×10^5 recombinant phage clones was sufficient to yield three positive recombinants (only one clone, 93TH253.3, was further analyzed). For 90CR402, 1.2×10^7 plaques (600 plates with 20,000 plaques per plate) had to be screened to identify a single positive HIV-1 recombinant (termed 90CR402.1). Phage clones were plaque purified and analyzed by restriction enzyme analysis, which indicated the positions of the proviruses relative to their flanking cellular sequences (Fig. 1B).

Biological activity of lambda phage-derived "subtype E" proviruses. To test their replication competence, plasmid subclones containing the full-length 90CR402.1 and 93TH253.3 genomes were transfected into 293T cells and COS-1 cells. Positive RT activity in the supernatants of transfected cell cultures indicated that both clones were expressing *gag*, *tat*, *rev*, and *pol* genes. Western blot analysis confirmed the presence of *gag* and *pol* gene products but failed to confirm the expression of envelope glycoproteins in supernatants from cells transfected with 93TH253.3 (data not shown). Subsequent cocultivation of transfected cells with normal donor PBMCs and immortalized T-cell lines did not result in the establishment of productively infected cultures for either clone. These results indicated that both 90CR402.1 and 93TH253.3 proviruses are replication defective.

Sequence analysis of "subtype E" proviruses. Both 90CR402.1 and 93TH253.3 proviruses were sequenced in their entirety. Analysis of their genomic organization revealed the presence of complete 5′ and 3′ LTRs and nine potential open reading frames corresponding to *gag*, *pol*, *vif*, *vpr*, *tat*, *rev*, *vpu*, *env*, and *nef* genes (Fig. 1C). There were no major deletions or rearrangements in either provirus, but in-frame stop codons were found in the *vif* gene of 90CR402.1 and the *env* gene (gp41-encoding portion) of 93TH253.3. Each of these mutations

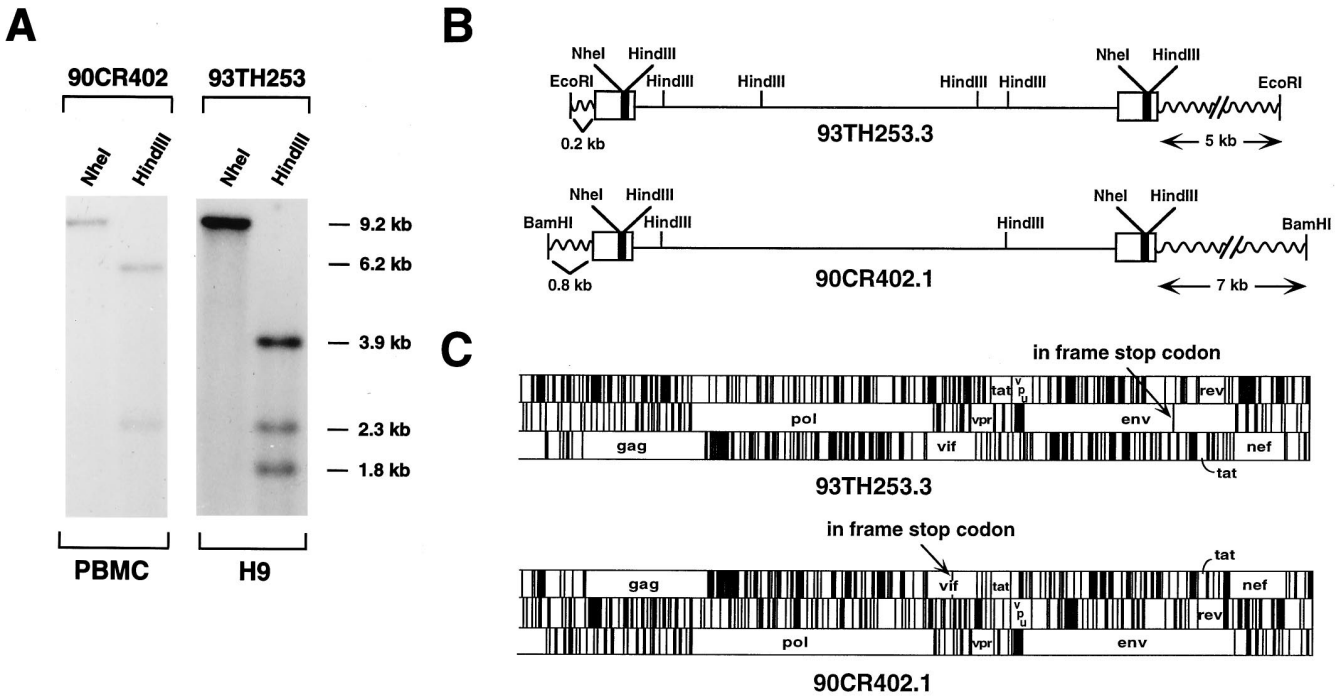

FIG. 1. Molecular cloning of full-length "subtype E" proviruses. (A) Southern blot of PBMC and H9 cultures productively infected with isolates 90CR402 and 93TH253, respectively. Total genomic DNA was digested with *Nhe*I and *Hin*dIII (both of which cleave the viral LTR) and probed with a full-length HIV-1 clone. The size of the resulting restriction fragments is indicated. (B) Schematic representation of full-length lambda phage clones containing the 93TH253.3 and 90CR402.1 genomes, respectively. Restriction enzyme cleavage sites used for cloning (*Bam*HI and *Eco*RI) and Southern blot analysis (A) are shown. Cellular sequences flanking the integrated provirus are denoted by wavy lines. (C) Organization of the "subtype E" genome. Open reading frames were identified using the program FRAMES of the University of Wisconsin Genetics Computer Group package. Stop codons, including two within coding regions, are indicated by vertical lines.

could explain the failure of these clones to establish a productive infection in primary PBMCs. 90CR402.1 also failed to replicate in T-cell lines known to be permissive for *vif*-deficient viruses (60), e.g., SupT1 and CEMx174 (data not shown). This may be due to a restricted cell tropism mediated by the viral envelope gene products rather than additional mutations, since primary isolates frequently fail to infect immortalized T-cell lines. Both 93TH253.3 and 90CR402.1 also lacked the *vpu* initiation codon, which is not required for virus replication; mutations at this codon are frequently seen after in vitro culture (29, 59, 75).

Phylogenetic analyses of "subtype E" *env* and *gag* regions. The phylogenetic relationships of the new "subtype E" viruses were first examined in evolutionary trees of *gag* and *env* gene sequences (Fig. 2). As expected, 90CR402.1 and 93TH253.3 clustered with "subtype E" viruses from Thailand in both genomic regions. Furthermore, their phylogenetic positions relative to the other "subtype E" viruses were consistent with their geographic origins. 93TH253.3 fell just outside the cluster of previously characterized Thai viruses, while 90CR402.1 branched well outside the same cluster. The overall extent of divergence among the Thai viruses (including 93TH253.3) was relatively low, consistent with the founder effect previously proposed for Thai "subtype E" isolates (67, 80). In contrast, the divergence between the Central African strain 90CR402.1 and the Thai "subtype E" viruses was much greater and comparable to that typically seen between epidemiologically unrelated strains within other clades (73, 96). These results thus support the conclusion that "subtype E" viruses have evolved in Central Africa for some period of time prior to their introduction into the Thai population (72, 78).

As in previous analyses (62, 67), the new "subtype E" viruses formed a distinct clade in the *env* tree but fell within the subtype A radiation in the *gag* tree (Fig. 2). This difference in phylogenetic position is quite remarkable and contrasts with that of two subtype A viruses, U455 and DJ258, which are shown in the same trees (Fig. 2). These latter two viruses, although themselves quite divergent, nonetheless occupy similar positions within clade A in both *gag* and *env* regions. In the *gag* tree, "subtype E" viruses, U455, and DJ258 are all approximately equidistantly related; i.e., the three branches leading to the "subtype E" viruses, the DJ258 group (which includes, for example, CI51), and the U455 cluster (which includes, for example, VI59), form a trichotomy. In contrast, in the *env* tree "subtype E" viruses branch off far outside the A clade and long before the divergence of U455 and DJ258. This difference in phylogenetic position of "subtype E" *gag* and *env* sequences is quite striking and highly suggestive of recombination (85, 86).

Diversity and bootstrap plots. Except for viruses of subtype B, sequence information outside of *gag* and *env* regions is extremely limited. Only six complete non-subtype B viruses have been sequenced thus far: one subtype A strain (U455), three subtype D strains (NDK, Z2Z6, and ELI), and two intersubtype recombinants (MAL and IBNG; see below) (75). Full-length reference sequences are not available for the other subtypes, thus making subtype assignments in LTR, *pol*, or accessory gene regions problematic. To characterize the newly derived "subtype E" viruses in regions outside the *gag* and *env* regions, we therefore examined which parts of their genomes were distantly, and which were closely, related to subtype A. This was done by aligning the sequences of 93TH253.3 and U455 (as representatives of "subtype E" and subtype A) and plotting their divergence along the genome (Fig. 3A). Distance values were calculated for a window of 500 nucleotides which

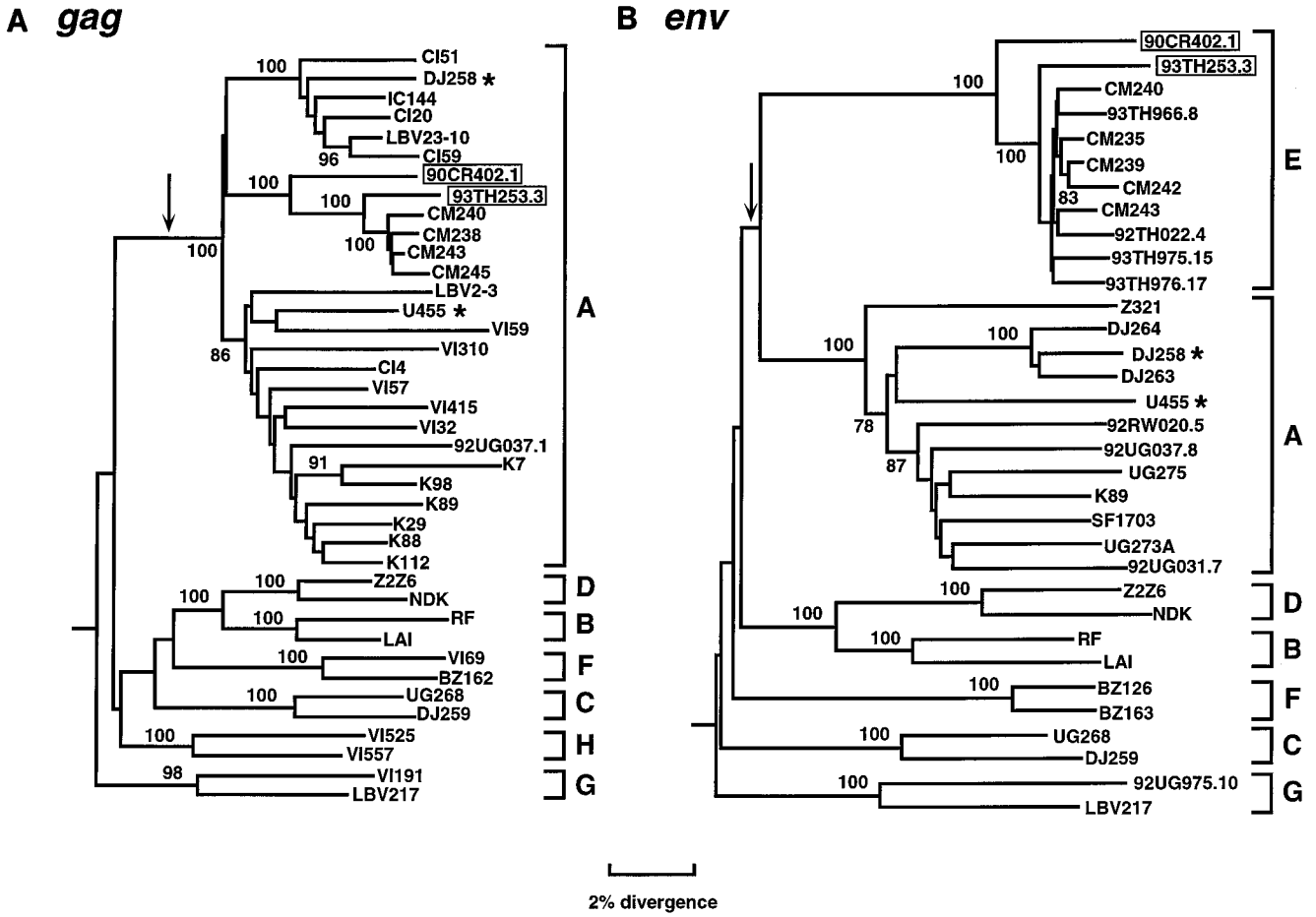

FIG. 2. Phylogenetic relationships of "subtype E" *gag* (A) and *env* (B) sequences in comparison to representatives of other HIV-1 group M subtypes (see the text for a description of the methodology). Horizontal branch lengths are drawn to scale (the scale bar represents 0.02 nucleotide substitutions per site); vertical separation is for clarity only. Values at nodes indicate the percentage of bootstraps in which the cluster to the right was found (bootstrap values of 75% and higher are shown). An arrow indicates the branch that leads to the A/E cluster in *gag* and *env* phylogenies. Asterisks denote two divergent subtype A viruses (U455 and DJ258) referred to in the text. Brackets at the right represent the major sequence subtypes of HIV-1 group M. The 3′ end of *env* (encoding the membrane-spanning and intracytoplasmatic domains) was excluded from the analysis since "subtype E" and "subtype G" viruses are known to be recombinant in this region (25). Trees were rooted by using $SIV_{CPZ}GAB$ as an outgroup.

was moved in steps of 10 nucleotides (the "E"-A distance). For comparison, similar plots were also generated for 93TH253.3 and LAI (the "E"-B distance), as well as for 93TH253.3 and 90CR402.1 (the "E"-"E" distance). Importantly, all sequences were aligned according to their predicted amino acid sequences, and distance values were calculated after removal of all gaps from the alignments.

As expected, the extent of divergence between the two "subtype E" viruses (90CR402.1 in Fig. 3A) varied along the genome, with the envelope region reaching the highest values. Predictably, the distance between 93TH253.3 and the clade B virus LAI was much higher, but more importantly, it generally followed the variation in distance between the two "subtype E" strains (i.e., the same regions exhibited proportionately higher or lower levels of divergence). In contrast, the extent of divergence between 93TH253.3 and U455 varied disproportionately along the genome. For the most part, this "E"-A distance was intermediate (i.e., the U455 plot fell between the 90CR402.1 and the LAI plots), which is consistent with the evolutionary relationships seen in *gag* (Fig. 2A), where "subtype E" viruses fall within subtype A. However, there were regions where the "E"-A distance matched or even exceeded the "E"-B distance (i.e., the U455 plot moved closer to or even above the LAI plot). This reflects the evolutionary relationships in *env* (Fig. 2B), where "subtype E" viruses are distinct from subtype A. Thus, noting the points in Fig. 3A at which the "E"-A distance increased or decreased relative to the "E"-"E" and "E"-B distances allowed the tentative identification of breakpoints in the putative E/A recombinants. For example, at position 5500 (the 5′ end of *env*), the LAI plot fell whereas the U455 plot did not, and both plots crossed around site 5800. Bearing in mind the window size of 500 nucleotides, this finding suggests that a crossover occurred in the region from positions 5500 to 5800. Similarly, at position 4700 (within *vif*), the U455 plot rose relative to the 90CR402.1 plot, suggesting that in this region 93TH253.3 is E-like rather than A-like.

To augment these analyses, we examined further the phylogenetic positions of the two "subtype E" viruses in relationship to U455. Trees were constructed for 90CR402.1, 93TH253.3, U455, two subtype B strains (LAI and RF), and two subtype D strains (NDK and Z2Z6), using simian immunodeficiency virus $SIV_{CPZ}GAB$ as an outgroup; then the magnitude of the bootstrap value supporting the clustering of "subtype E" viruses with subtype A (i.e., the bootstrap value for the branch marked

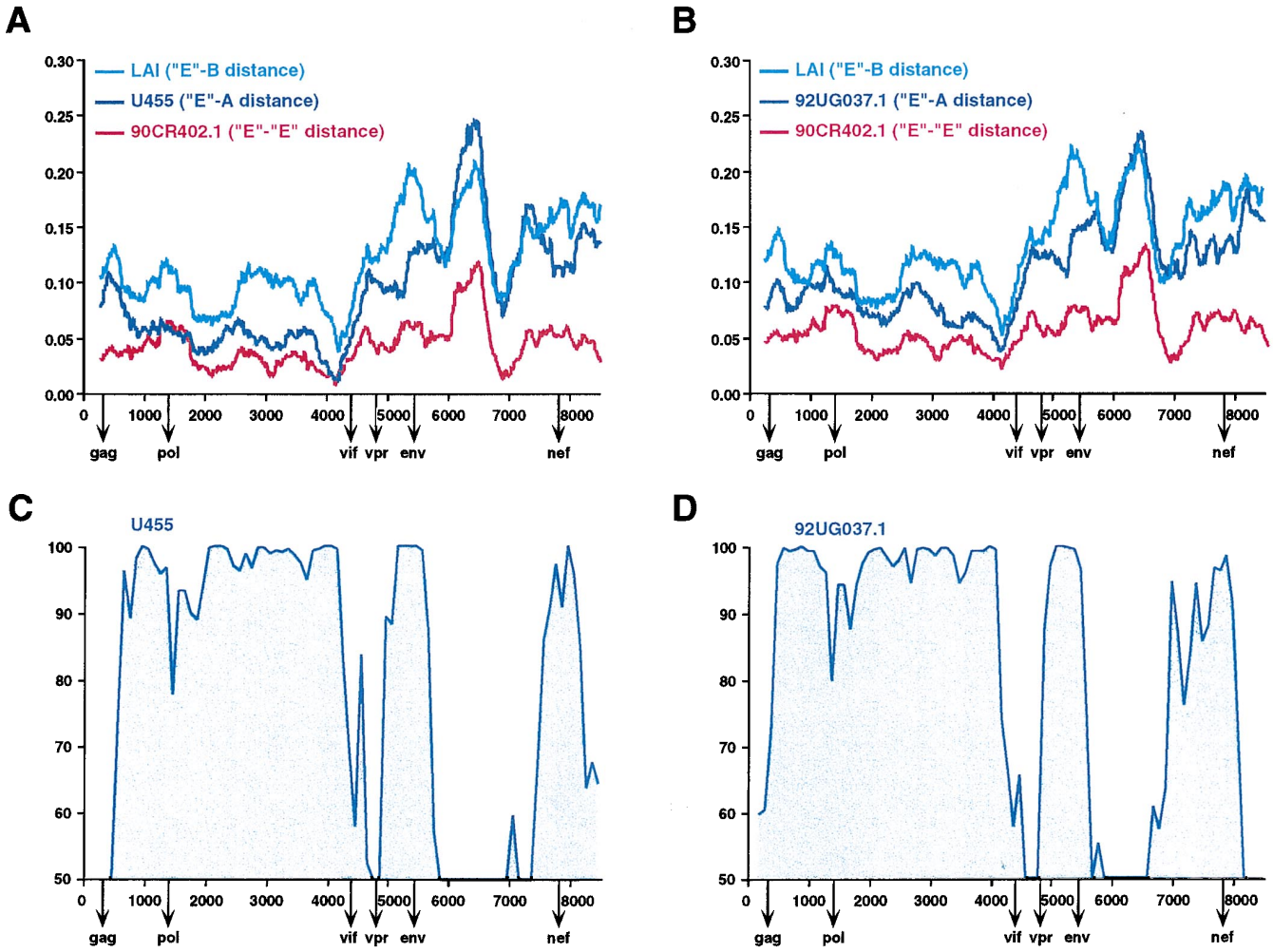

FIG. 3. Diversity and bootstrap plots comparing the sequence relationships among members of subtypes A, "E," and B. (A) Sequence divergence between 93TH253.3 and LAI ("E"-B distance), 93TH253.3 and U455 ("E"-A distance), and 93TH253.3 and 90CR402.1 ("E"-"E" distance). Distance values were calculated for a window 500 nucleotides moved in steps of 10 nucleotides. The *x* axis indicates the nucleotide positions along the alignment (gaps were stripped and removed from the alignment). The beginning of *gag*, *pol*, *vif*, *vpr*, *env*, and *nef* open reading frames are shown. The *y* axis denotes the distance between the viruses compared (0.05 = 5% divergence). (B) Diversity plot generated as for panel A but with 92UG037.1 used in place of U455 as the reference sequence for subtype A. (C) Bootstrap plots illustrating the likelihood of clustering of "subtype E" viruses (90CR402.1 and 93TH253.3) with subtype A. Trees were constructed for 90CR402.1, 93TH253.3, U455, two subtype B strains (LAI and RF), and two subtype D strains (NDK and Z2Z6), using $SIV_{CPZ}GAB$ as an outgroup. The magnitude of the bootstrap value supporting the clustering of "subtype E" viruses with subtype A is reported for windows of 500 nucleotides moved in steps of 100 nucleotides. Regions of the "subtype E" genome which are of subtype A origin exhibit very high (>90%) bootstrap values. Regions of subtype E origin are characterized by very low (<50%) values. The *x* axis is as in panel A; the *y* axis indicates the percentage of bootstrap replicates supporting the clustering of "subtype E" viruses with subtype A. (D) Bootstrap plot as in panel C except using 92UG037.1 as a subtype A reference sequence in place of U455.

by an arrow in Fig. 2) was examined. On the basis of the phylogenies depicted in Fig. 2, we expected regions of "subtype E" viruses which were of subtype A origin to yield very high (>90%) bootstrap values. Alternatively, regions of true subtype E origin were expected to have very low (<50%) values. Figure 3C depicts the results of more than 80 such phylogenetic analyses, each performed on a window of 500 nucleotides, moved in steps of 100 nucleotides (this approach is similar to the "bootscanning" method of Salminen et al. [92]). Very high bootstrap values (>90%) are found for much of the "subtype E" genome, indicating that *gag*, *pol*, *vpr*/*vpu*, and 5′ *nef* regions actually derive from subtype A. The plot also illustrates regions where "subtype E" viruses did not cluster consistently with U455, which, as expected, included the envelope. However, there was a second well-defined area with very low bootstrap values (between positions 4200 and 4900), suggesting that sequences in the *vif* region also represent subtype E.

Recombination breakpoint analysis using a second subtype A reference sequence. Because all conclusions drawn from Fig. 3A and C depend on the assumption that U455 is itself not a recombinant, it was necessary to evaluate the diversity and bootstrap plots by using a second (independent) subtype A reference strain. We thus used long-PCR methods (94) to amplify a near-full-length genome of 92UG037 [between the tRNA primer binding site and the poly(A) signal of the 3′ LTR]. This isolate has previously been reported to represent subtype A (25). The LTR region, which was not included in the long PCR fragment, was independently amplified by regular PCR. Sequence analysis of both fragments allowed the reconstruction of the entire proviral sequence, which was then used to map potential recombination breakpoints in 90CR402.1 and 93TH253.3, as described above (a more detailed description of the 92UG037.1 sequence is provided elsewhere [24]). Diversity and bootstrap plots using 92UG037.1 as the subtype A refer-

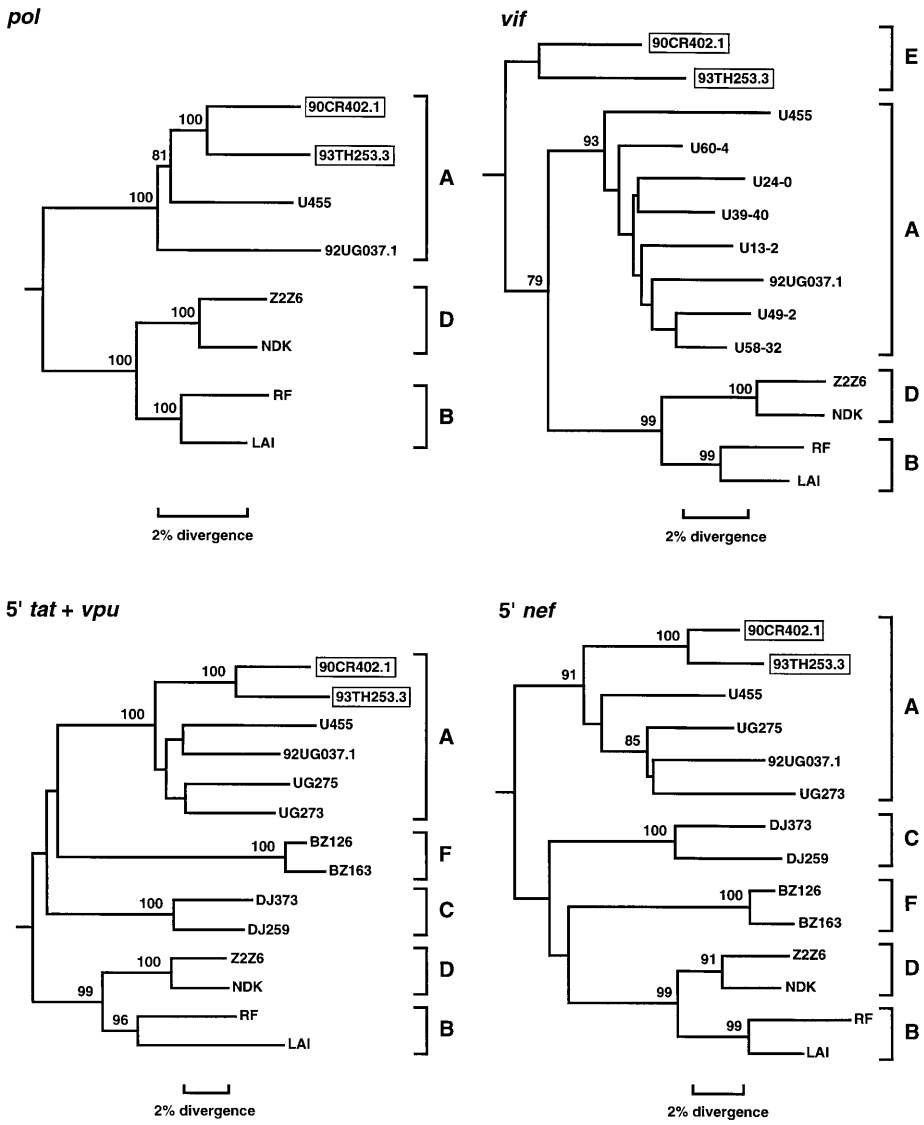

FIG. 4. Phylogenetic relationships of the newly derived "subtype E" viruses in *pol*, *vif*, 5′ *tat* plus *vpu*, and 5′ *nef* regions. See the text and the legend for Fig. 2 for details.

ence sequence were very similar to those using U455 (Fig. 3, compare panels A and C with panels B and D), suggesting that neither U455 nor 92UG037.1 was recombinant. Again, bootstrap values were low in *vif* and *env* regions, although the length of the *env* portion differed slightly between the two plots. In the U455 bootstrap plot, this area extended from positions 5800 to 7500, while the same region mapped between position 5500 and 7000 in the 92UG037.1 plot (compare Fig. 3C and D). The exact points of crossover delineating the subtype E *env* region are thus uncertain.

Phylogenetic analysis of "subtype E" *pol*, *vif*, *tat*/*vpu*, and 5′ *nef* regions. To further characterize the putative A and E portions within the "subtype E" genome, we determined phylogenies for partial gene sequences including as many representatives of additional subtypes as were available from the most recent HIV sequence database release (75). As suggested by the diversity and bootstrap plots, *pol* sequences of the "subtype E" viruses clustered within subtype A (Fig. 4). This was the case for the entire *pol* regions of both 90CR402.1 and 93TH253.3 and was supported by 100% bootstrap values. Sequences from 5′ *tat*/*vpu* and 3′ *nef* domains of "subtype E" also clustered well within the subtype A radiation, again confirming results from the diversity and bootstrap plots. Subtype assignment in the *vif* region, however, was more difficult. From the diversity and bootstrap plots, we expected that "subtype E" viruses would not cluster within subtype A in this region. This was indeed the case. In a tree of *vif* sequences downstream of the *vif*/*pol* overlap (not shown), "subtype E" sequences fell outside the subtype A radiation, although not as far outside as previously seen for *env* sequences. Moreover, inspection of the bootstrap plots indicated a small additional A peak at position 4500 in both Fig. 3C and Fig. 3D, suggesting that within this region "subtype E" viruses are multiply mosaic. To investigate this possibility, we constructed phylogenetic trees from *vif* sequences excluding short regions (of 100 bp) at different positions in the alignment. These analyses yielded one tree (derived from sequences between positions 57 and 579 of the *vif* alignment but excluding positions 351 to 450) in which "sub-

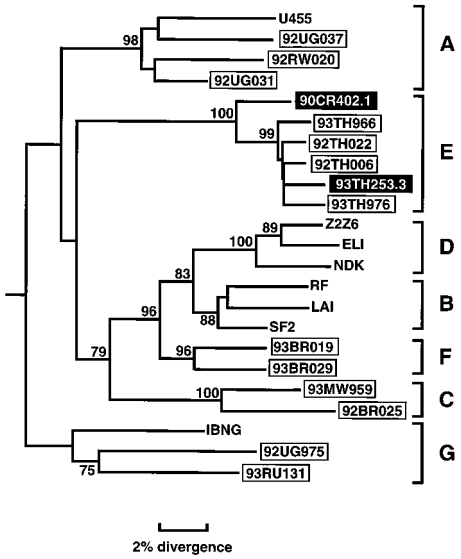

FIG. 5. Phylogenetic relationships of LTR sequences. A tree was constructed from partial LTR sequences, including most of U3 and R (positions 93 to 548 of the LTR alignment). Newly generated reference sequences for subtypes A to G are shown in boxes (sequences were determined without interim cloning directly from PCR amplification products).

type E" viruses fell significantly outside subtype A (Fig. 4, *vif*), thus strongly suggesting that *vif* contains an additional short A portion just before the *vpr* overlap.

Phylogenetic analyses of LTR sequences. To examine the subtype assignment of the "subtype E" LTR region, it was necessary to generate reference sequences for several different HIV-1 group M subtypes for which such information was unavailable. Using regular PCR, we amplified a 670-bp fragment spanning most of U3, R, and U5 from a panel of primary isolates previously reported to represent subtypes A, C, E, F, and G on the basis of *env* sequences (25, 26). A total of 16 new viral LTR sequences (in addition to 93TH253.3 and 90CR402.1) were then determined (also see Fig. 8) and phylogenetically analyzed. Thirteen of these sequences are shown in Fig. 5. Amplification products were sequenced directly without interim cloning, and a phylogenetic tree was constructed from 5′ LTR sequences (between positions 93 and 548 of the alignment [Fig. 5]). Reference sequences for subtype B and D viruses were obtained from the database. Importantly, this tree depicts an overall topology for the various HIV-1 group M subtypes that is very similar to topologies previously obtained from sequences of coding regions (e.g., compare Fig. 5 with *gag* and *env* trees shown in Fig. 2). That is, viruses known to cluster together in a particular subtype according to *gag* or *env* (e.g., viruses belonging to subtype C or F) formed the same distinct groupings in the LTR tree. Even the closer relationship of subtypes B and D viruses previously reported for *gag*, *pol*, or *env* phylogenies (96) is represented in the LTR tree. These results suggest that LTR sequences can be used reliably for subtype classification.

As shown in Fig. 5, "subtype E" LTR sequences formed a distinct lineage, separate from all other clades, including A, and similar to the one identified for "E" *env* sequences (Fig. 2B). Importantly, all six of the "subtype E" viruses analyzed (90CR402.1 and 93TH253.3 as well as four additional isolates from Thailand) fell into one cluster, with Thai strains again grouping most closely (Fig. 2A). These results indicate that the LTR region of "subtype E" viruses is not derived from subtype A but instead likely of true subtype E origin, as are *env* and *vif* regions. Phylogenetic trees from 5′ and 3′ *nef* regions suggested that the 5′ boundary of this E portion coincides with the beginning of the LTR (data not shown). The 3′ boundary could not be determined, because in trees from 3′ LTR sequences the various subtypes no longer formed distinct monophyletic groupings, thus making subtype classification impossible (data not shown). However, the ambiguous region was very small, including only the 3′ half of R, U5, and a portion of the untranslated leader sequence.

Summary of phylogenetic and breakpoint analyses. Combining the results of Fig. 2 to 5, we have generated a composite representation of the "subtype E" genome structure (Fig. 6). This composite suggests that "subtype E" viruses are complex hybrids of subtypes A and E. There are eight apparent points of crossover, delineating eight alternating fragments of subtypes A and E. Most of the LTR, the portion of *env* which encodes gp120 and the extracellular domain of gp41, and the 3′ half of *nef* are comprised of subtype E. The remainder of the genome, including all of *gag* and *pol*, *vpu*, as well as both exons of *tat* and *rev*, derive from subtype A. The *vif* and *vpr* regions are multiply mosaic, which is consistent with frequent template switches during the recombination event (10, 11). Some regions and breakpoints, including the 3′ LTR and the 5′ and 3′ boundaries of the subtype E envelope domain (Fig. 6), could not be defined. Finally, both Thai and CAR "subtype E" strains exhibited the same pattern of mosaicism, indicating that they are derived from a common A/E recombinant ancestor.

Is the hybrid appearance of "subtype E" viruses an artifact? The discordant phylogenetic positions of the different subtype "E" genomic regions are most simply explained as the result of recombination between ancestral subtype A and E viruses. However, the lack of an example of a full-length (nonrecombinant) subtype E virus has led to speculation that the discordant phylogenetic positions of *gag* and *env* sequences could be due to large differences in the rates and/or patterns of evolution in these genes. Since most of the "subtype E" genome is

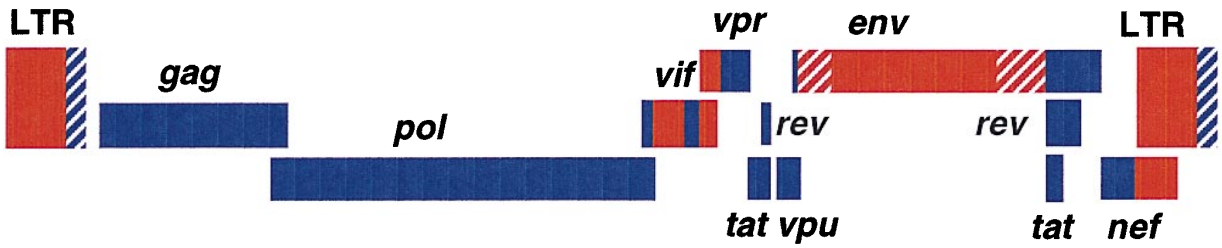

FIG. 6. Inferred structure of the "subtype E" genome. A composite picture summarizing the results of diversity and bootstrap plots, as well as phylogenetic tree analyses, is shown (see Fig. 2 to 5). Regions of subtype A and subtype E origin are indicated in blue and red, respectively; uncertain breakpoints (in *env*) and regions that could not be subtyped (3′ LTR) are represented by hatched boxes.

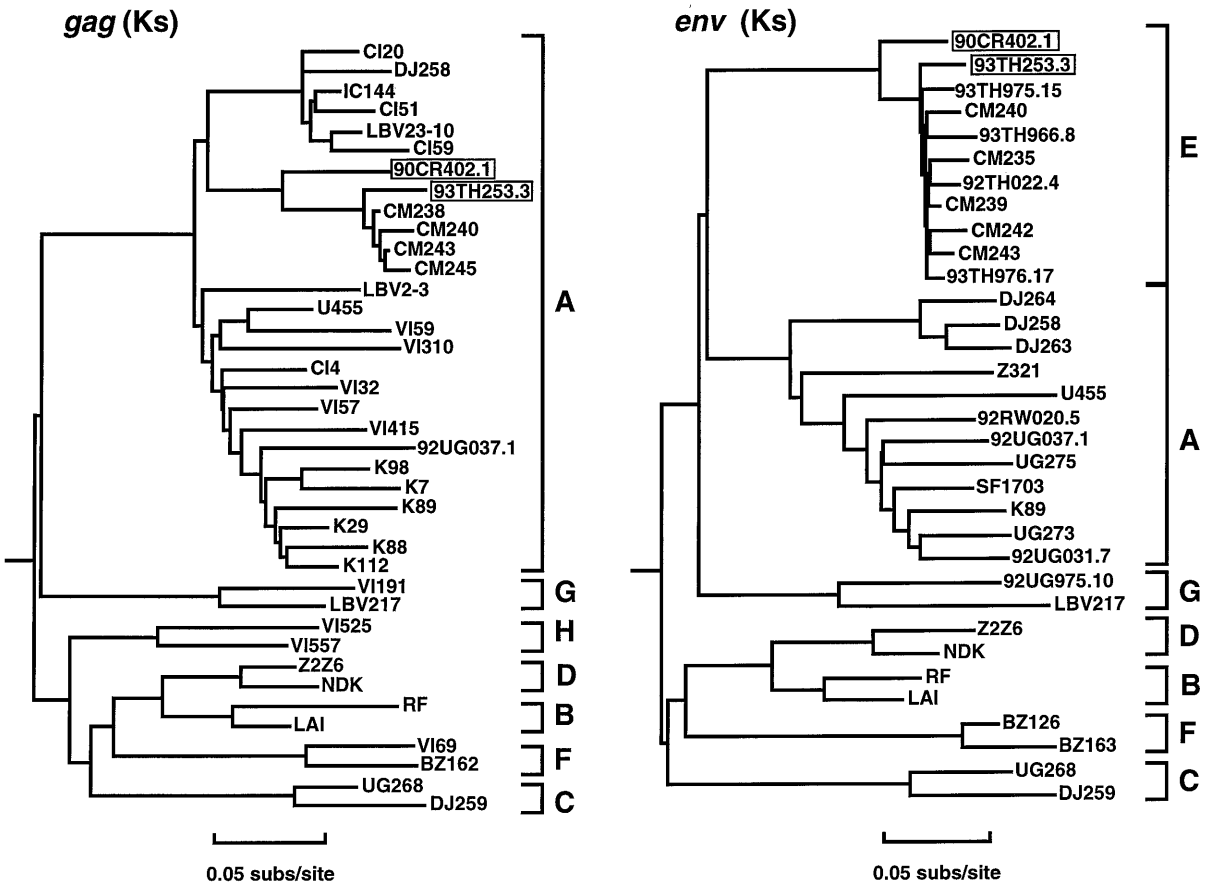

FIG. 7. Phylogenetic analyses for *gag* and *env* sequences based on numbers of synonymous substitutions. Trees were midpoint rooted by using RETREE from the PHYLIP (19) package (see the legend to Fig. 2 and Results for more details).

actually A-like, we would have to infer that it is the *env* gene of "subtype E" viruses that has evolved in an extraordinary fashion.

For the *env* sequences of "E" viruses to fall so far outside the A clade in the phylogenetic tree, one would have to imagine that the "E" *env* genes have evolved extremely fast. However, a simple acceleration of evolutionary rate in "subtype E" *env* would not alone be sufficient to move this cluster of viruses outside the subtype A clade. Rather, such an acceleration would just be expected to lead to longer branches (as has been reported for a strain of HIV-2 that had undergone G-to-A hypermutation [27]). To move the "E" lineage to its present position outside of the A clade would require that a substantial fraction of the substitutions in "E" viruses subsequent to their divergence from subtype A was convergent on the other subtypes. Such a degree of sequence convergence is unprecedented. Nevertheless, it could be argued that HIV-1 *env* sequences are under extraordinary selection pressures and that therefore such a degree of convergence is possible. However, if that were the case, then selection should primarily target the protein sequence and should have little (if any) effect on synonymous substitutions. To examine this possibility directly, we performed separate phylogenetic analyses for *gag* and *env* sequences based on numbers of synonymous and nonsynonymous substitutions. The phylogenetic trees obtained for synonymous changes (Fig. 7) were very similar to those obtained for all sequence substitutions (Fig. 2). In particular, the discrepancy between the phylogenetic positions of "subtype E" sequences in *gag* and *env* trees was just as striking (Fig. 7). Phylogenetic trees based on nonsynonymous changes were also very similar (not shown). Thus, all nucleotide substitutions in subtype E *env* sequences, whether silent or amino acid replacing, yielded the same phylogenetic relationships to the other subtypes. On the basis of these results, we rule out the possibility that disproportionate selection-driven evolution can explain the discordant phylogenetic position of "subtype E" *env* gene sequences and thus conclude that "subtype E" viruses represent intersubtype recombinants.

Analysis of regulatory regions in the "subtype E" LTR. Phylogenetic analyses of LTR sequences (Fig. 5) revealed separate clustering of subtype A and "E" viruses, indicating that in this part of the genome, "subtype E" viruses were distinct from subtype A (Fig. 6). Because the LTR plays an essential role in the regulation of viral gene expression (for reviews, see references 28 and 44), we were interested to see if there were subtype-specific differences in regions of known regulatory function. Figure 8 depicts an LTR sequence alignment in the core enhancer/promoter domain, spanning the region between NF-κB and TAR elements. This alignment revealed several sequence motifs in transcription factor binding regions that were unique to members of "subtype E". For example, all "subtype E" viruses exhibited sequence changes in the distal (5′) NF-κB site (5′-AGGACTTCC-3′) which differed from the consensus (5′-GGGRNNYYCC-3′ [68]) by both a transition at the 5′ end and a single nucleotide deletion. None of the other viruses exhibited these types of changes; instead, most strains

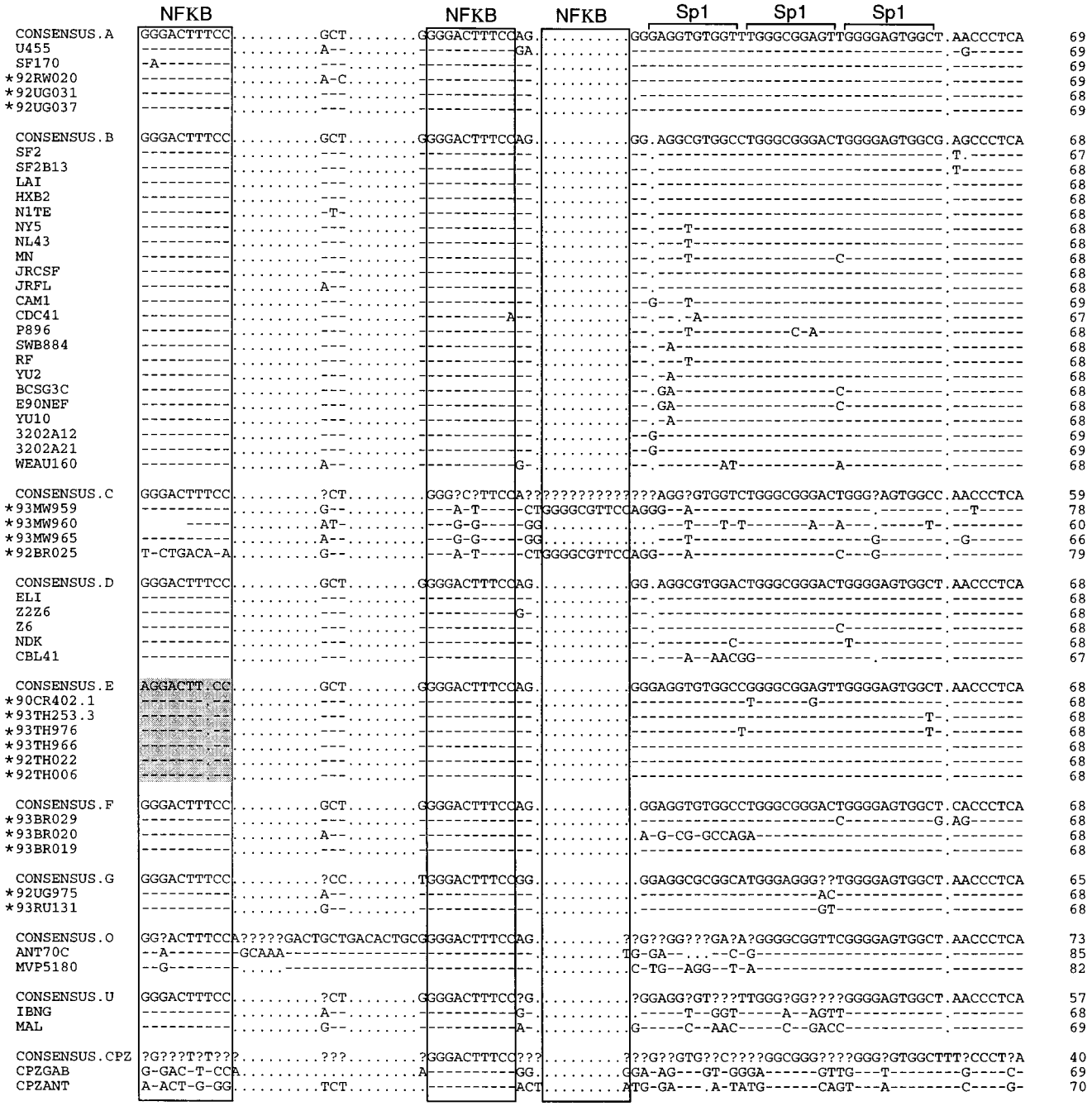

FIG. 8. Alignment of HIV-1/SIV_{CPZ} LTR sequences surrounding the core enhancer region. Newly derived LTR sequences (indicated by asterisks) are compared with previously published HIV-1 and SIV_{CPZ} sequences from the database (75). Putative binding sites for transcription factors NF-κB, Sp1, and TATA, as well as functional motifs within the TAR domain, are indicated. Subtype-specific sequence motifs referred to in the text are shaded. Periods indicate gaps introduced to optimize the alignments. Sequences within a subtype are compared with a consensus sequence. Question marks in the consensus sequence indicate sites at which fewer than 50% of the viruses have the same nucleotide.

contained at least two, and sometimes three, copies of the consensus NF-κB domain. Subtype-specific sequence changes were also observed in the TATA domain, in which "subtype E" viruses exhibited TAAAA. All other strains of HIV-1 or SIV_{CPZ} displayed the sequence TATAA.

Sequence changes were also noted in the TAR element, which is known to form a stable RNA stem-loop structure at the 5′ end of the viral RNA essential for Tat-mediated transactivation of HIV-1 gene expression (20, 28, 44, 71, 87). In contrast to most HIV-1 and SIV_{CPZ} strains, "subtype E" viruses were found to contain only two (instead of three) nucleotides in the bulge of this structure (Fig. 8 and 9). Interestingly, the only other HIV-1 strains which contained such a two-nucleotide bulge were members of subtype A (all five isolates analyzed) and $HIV\text{-}1_{MAL}$, a recombinant composed of subtypes A and D (1, 85). This finding raised the possibility that

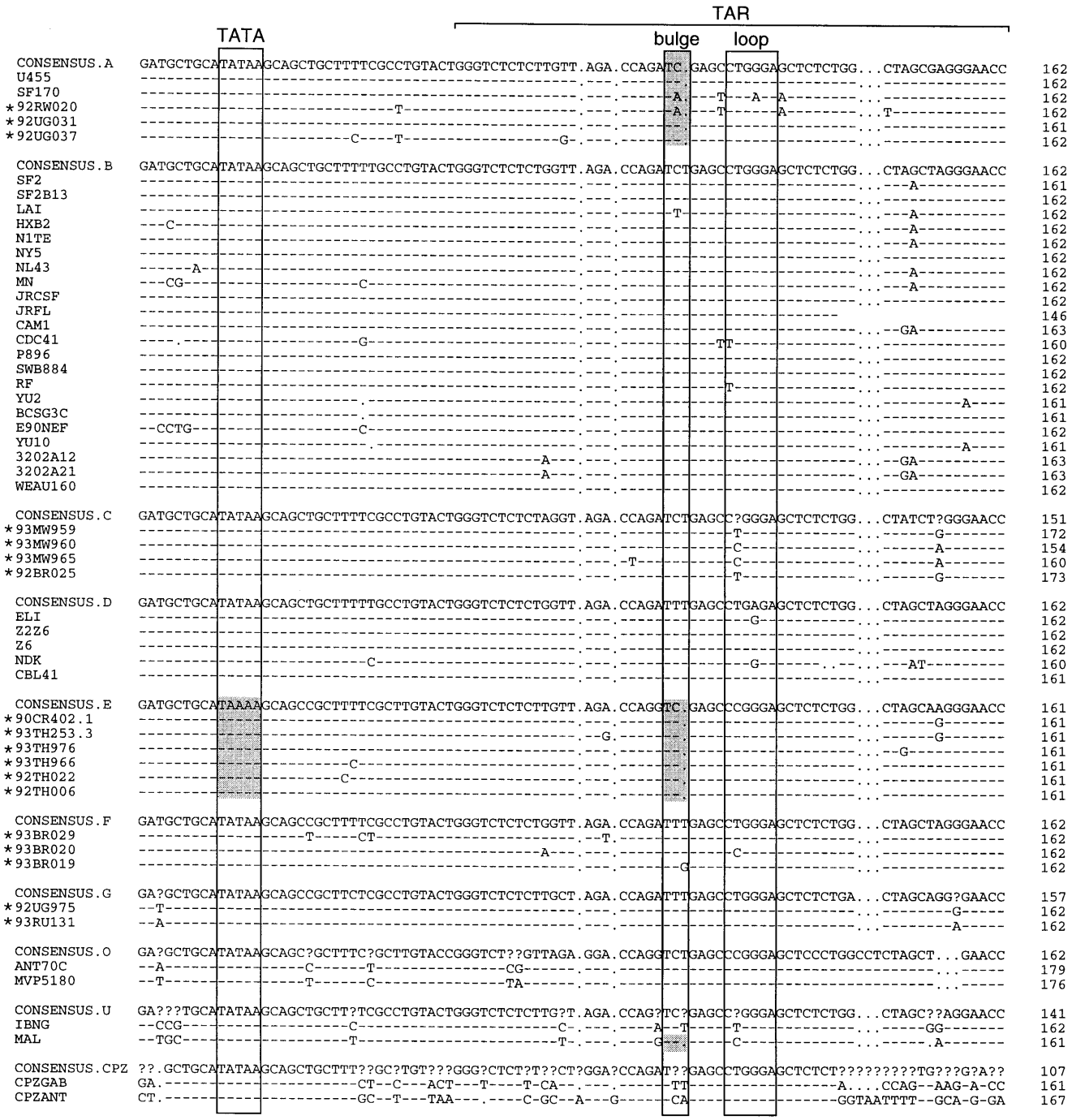

FIG. 8—*Continued.*

the two-nucleotide bulge was functionally linked to other viral genes common to "subtype E" or A viruses. To examine this possibility more closely, we compared the predicted secondary structure of the "subtype E" TAR region (90CR402.1) with those of subtype B (HXB2) and subtype A (U455) (Fig. 9). This analysis revealed that the observed sequence changes did not appear to affect the predicted stem-loop structure, since the calculated energy values were very similar for all three viruses (see the legend to Fig. 9). Moreover, regions of known functional importance (28, 44, 87), such as the uridine residue at position +23 and the composition of the stem between the bulge and the terminal loop, were highly conserved. However, there were differences between subtype A and E TAR sequences, which included substitutions in the stem as well as in the terminal loop (circled in Fig. 9). Importantly, all six "subtype E" strains contained these same substitutions.

Given the high degree of sequence and structural conservation of TAR among the various HIV-1 and SIV_{CPZ} strains (Fig. 8), the lack of a third nucleotide in the TAR bulge of "subtype E" and A viruses is quite remarkable. Subtype A and

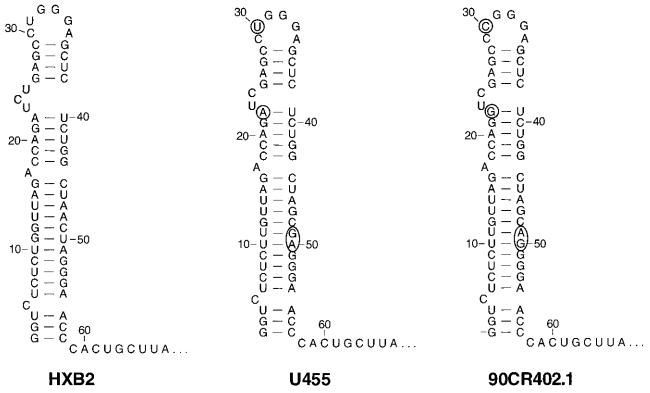

FIG. 9. Secondary structure comparison of TAR RNA sequences from members of subtype B (HXB2), subtype A (U455), and "subtype E" (90CR402.1). Secondary structures were predicted using the MFold and PlotFold programs of the University of Wisconsin Genetics Computer Group software package. The predicted free energy values are −25.5 kcal (1 cal = 4.184 J)/mol for HXB2, −25.2 kcal/mol for U455, and −24.3 kcal/mol for 90CR402.1. The nucleotides comprising the TAR bulge in the different viruses are shaded. Nucleotide differences which distinguish subtype A and subtype E TAR sequences are circled.

"E" TAR elements are obviously functional, since the sequence changes in the bulge are present in five of five subtype A and six of six "subtype E" viruses. Moreover, it is intriguing that the only viruses known to have a two-nucleotide TAR bulge, apart from members of subtype A, are MAL and members of "subtype E," all of which are recombinants containing genomic regions derived from subtype A. Because most of the "subtype E" LTR is clearly not of subtype A origin but likely represents subtype E (Fig. 5), the lack of the third nucleotide in the TAR bulge could represent a compensatory change, i.e., a deletion that occurred after recombination brought together a subtype E LTR with subtype A coding regions. This, in turn, would suggest that a functional link exists between the two nucleotide bulge and a subtype A protein. The most obvious candidate for such a protein would of course be Tat, since this protein is known to bind the bulge region of TAR directly (28, 44). Moreover, both in subtype A and "E" viruses, *tat* gene sequences are derived from subtype A (both first and second exons [Fig. 6]). However, this hypothesis is not consistent with the presence of a two-nucleotide bulge in MAL (Fig. 8), which contains a subtype D *tat* gene (85), or the presence of a three-nucleotide bulge in another HIV-1 strain, termed IBNG, whose sequence was recently reported to represent subtype A (75).

To investigate further whether there is a correlation between the presence of a two-nucleotide TAR bulge and a particular genomic region different from *tat* in subtype A viruses, we analyzed the sequences of MAL and IBNG more closely. In particular, we were interested in coding regions where MAL was A-like (because it contains the two-nucleotide bulge) but IBNG was not (because it contains a three-nucleotide bulge). Bootstrap plots were generated similar to the ones shown in Fig. 3, except that this time MAL and IBNG were included in the tree analyses and their phylogenetic positions relative to representatives of subtype A (U455 and 90UG037.1) were examined (Fig. 10). This analysis confirmed that both IBNG and MAL contained sequences of subtype A origin and, in addition, revealed that IBNG was recombinant. As shown in Fig. 10A, most of the IBNG genome appears to derive from subtype A. However, bootstrap values were very low in 3′ *gag*, 5′ *pol*, *vif*, and parts of *env*, strongly suggesting recombination with another subtype (this may be subtype G, since IBNG groups with subtype G viruses in the LTR [Fig. 5]). Similarly, MAL was found to contain subtype A sequences in the 5′ *gag*, 3′ *pol*, and 5′ *nef* regions (Fig. 10B), in agreement with previous results (85). However, bootstrap plots revealed one additional small subtype A peak in the 5′ *pol* domain that had previously gone unrecognized. The remainder of the MAL genome, except for the middle of *pol*, which cannot be classified (85), represented subtype D (bootstrap plots evaluating the position of MAL relative to subtype D viruses were performed but are not shown). Taken together, these analyses identified only a single region where MAL clustered in subtype A but IBNG did not (horizontal arrows in Fig. 10A and B). This region mapped to a 5′ *pol* domain which encodes parts of the protease and RT. Thus, if there is indeed a functional link between a two-nucleotide TAR bulge and a subtype A protein, we would predict the latter to be encoded by the 5′ region of the *pol* gene.

DISCUSSION

"Subtype E" viruses represent A/E intersubtype recombinants. In this study, we used a variety of complementary analytical strategies to explore the genetic identity and evolutionary origins of "subtype E" viruses. Using diversity and bootstrap plots in combination with phylogenetic tree analyses, we found that most of the "subtype E" genome, including the *gag*, *pol*, *tat*, *rev*, *vpu*, and 5′ *nef* regions, is actually composed of subtype A-derived sequences. However, the extracellular portion of *env*, parts of *vif* and *vpr*, and most of the LTR are not of subtype A origin. Unless there have been additional recombination events with unidentified subtypes during the evolution of these viruses, all of the latter genomic regions must be derived from subtype E. *vif* and *vpr* regions were found to be multiply mosaic, which is consistent with template switches of the RT during proviral DNA synthesis and thus highly suggestive of recombination (10, 11). Nevertheless, because it had been suggested that the discordant phylogenies between different "subtype E" genes may be the result of unusual patterns of evolution, we examined this question by constructing separate phylogenetic trees for synonymous and nonsynonymous substitutions. We found no differences in the relative positions of "subtype E" *gag* and *env* gene sequences. Thus, there is no evidence for selection-driven convergent evolution as an explanation for the discordant branching orders of "subtype E" *gag* and *env* sequences. Finally, we identified subtype-specific sequence motifs in the viral LTR (in NF-κB and TATA domains) that distinguished the "E" clade viruses from all others, including members of subtype A. Based on these findings, we conclude that "subtype E" viruses are in fact recombinants of subtypes A and E. Similar conclusions have recently been reached by other investigators (8) who analyzed a second Thai "subtype E" strain (CM240). However, our findings of at least eight points of recombination crossover, the characterization of a second full-length subtype A reference sequence, separate analyses of synonymous and nonsynonymous nucleotide substitutions, and the identification of signature sequences in the "subtype E" LTR distinguish the present report.

The timing and circumstances of the recombination event that generated the subtype A/E viruses are unknown. However, several lines of evidence suggest that it took place in Central Africa. Thai and CAR strains (93TH253.3 and 90CR402.1) analyzed in this study exhibited the same pattern of mosaicism, indicating that they are derived from the same recombinant ancestor. Since subtype A viruses have not been identified in Thailand (despite extensive sampling), the recombination event most likely occurred in Central Africa, where

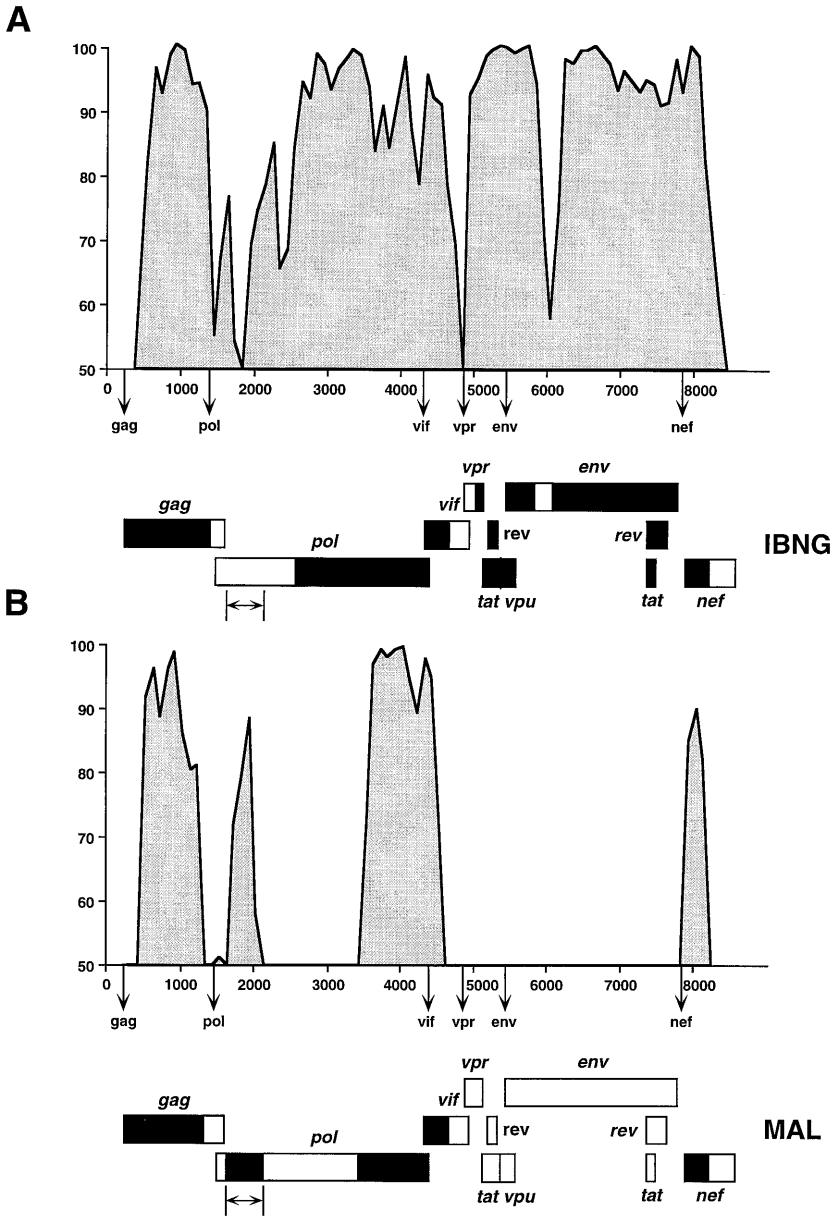

FIG. 10. Bootstrap plots depicting the relationship of IBNG (A) and MAL (B) to subtype A viruses. Regions within the IBNG and MAL genome that are of subtype A origin (as determined from the bootstrap plots) are shown as filled areas. A horizontal arrow denotes the 5′ *pol* domain which correlates with the size of the TAR bulge. See the legend to Fig. 3 and Results for more details.

members of both subtype A and "E" are known to exist (72, 78). At some point after that recombination event, a descendant virus was introduced into the Thai population, where it then spread with extraordinary rapidity (111, 112). The fate of the nonrecombinant (ancestral) subtype E parental strain remains unknown. It is possible that viruses of this subtype still exist but have not yet been identified. Alternatively, the original subtype E lineage may have become extinct, with its only remaining vestige residing in the present day A/E recombinants.

The realization that "subtype E" viruses represent A/E recombinants has obvious practical implications for subtype nomenclature and the identification of new subtypes. In the past several years, subtype classification has become a critical tool for molecular epidemiologists since it provides a means to monitor the geographic distribution, prevalence, and intermixing of globally circulating HIV-1 variants (113). In addition, subtype classification has served as the basis for the identification of intersubtype recombinants (34, 91, 99), thereby alerting the scientific community to the existence of coinfection of individuals by more than one virus strain. It thus defeats the purpose to classify hybrid viruses as distinct subtypes, even if they all exhibit the same pattern of mosaicism. For this reason, we propose to reclassify "subtype E" viruses as A/E recombinants. The term "subtype E" should be reserved for genomic regions within A/E viruses which are of true subtype E origin. This will also avoid unnecessary confusion should nonrecombinant examples of subtype E eventually be identified. The

example of the "subtype E" recombinants also illustrates a problem inherent in the designation of new subtypes (several candidates have recently been identified [55, 75]). Ideally, a new subtype should be designated only after a representative isolate has been sequenced in its entirety and its genome structure has been examined for evidence of mosaicism. Unfortunately, the amount of work involved and the lack of reference sequences even for the currently known subtypes make this impractical, at least at the present time. However, at a minimum, analysis of full-length *gag* and *env* sequences should be required for the designation of any new subtype candidate. This would exclude most hybrid viruses, since the majority contain at least one crossover within these regions or an odd number of crossovers between *gag* and *env* such that these sequences fall into different subtypes (86). It is also probably desirable to have such sequence data for several epidemiologically unlinked representatives, thus reserving subtype names for strains of epidemiological significance.

Full-length reference sequences are needed for the phylogenetic classification of HIV-1. Since intersubtype recombinants are being recognized with increasing frequency (15, 16, 25, 56, 85, 86, 88, 91, 99), HIV-1 subtype reference sequences for regions outside of *gag* and *env* are critically needed to examine newly derived viruses for evidence of mosaicism. For example, in this study we were unable to determine the subtype assignment of the regions in IBNG which did not cluster with subtype A. Similarly, classification of "subtype E" LTR sequences was possible only after appropriate reference sequences were generated for phylogenetic analyses. Finally, a second subtype A reference sequence was required to confirm the breakpoint analyses from diversity and bootstrap plots for the "subtype E" genome. To begin to define the true impact of intersubtype recombination on global HIV variation, it is thus necessary to derive full-length reference sequences for all HIV-1 group M subtypes.

There are several approaches to generate full-length HIV-1 genomes, including lambda phage cloning, long PCR, or PCR amplification of proviral halves. All of these have shortcomings, particularly if the product is intended as a biological reference strain. Lambda phage techniques allow cloning of viral genomes as single genomic units and are not subject to PCR misincorporations or other polymerase-induced artifacts (including recombination). However, this method depends on efficient virus growth in tissue culture or requires screening of vast numbers of recombinant phage clones, as was the case for 90CR402.1. Adaptation of viruses to growth in immortalized T-cell lines can overcome low viral titers but may also select for variants with altered biological properties (53, 70). Thus, unless an isolate replicates with reasonable efficiency in primary target cells, long PCR and PCR amplification of proviral halves are probably more appropriate methods of deriving full-length proviruses. Nevertheless, biological reference reagents are still best derived by lambda phage approaches, even if isolated proviruses require subsequent reconstruction to yield a biologically active genome (we are currently restoring the *vif* gene in 90CR402.1 in an attempt to generate a replication-competent reference clone for this primary "subtype E" isolate).

A/E recombinants from Thailand and the Central African Republic exhibit unique sequence changes in the core enhancer/promoter region of their LTR. Inspection of the core enhancer/promoter sequences among members of all HIV-1 groups and subtypes revealed a number of interesting features, particularly for members of "subtype E". For example, all "E" viruses exhibited unique sequence changes in their 5′ (distal) NF-κB sites, differing from the consensus both in a nucleotide substitution and a single base pair deletion. It is possible that these changes reduce the binding of the NF-κB transcription factor and thus influence gene expression. Alternatively, there may not be any appreciable effect on NF-κB binding, since there are enhancer domains in the human genome which contain comparable changes in their NF-κB domain (for a review, see reference 68). Without direct experimentation, it is thus impossible to predict whether the observed sequence changes in "subtype E" NF-κB or TATA regions have biological consequences. Nevertheless, the sequence changes are present in all isolates, including the distantly related 90CR402.1 strain, and thus represent subtype specific signatures of "subtype E" viruses.

LTR sequence alignments also revealed that a subset of HIV-1 strains exhibited an unusual TAR element. Figure 8 shows that all isolates of subtype A and "subtype E," and one A/D hybrid strain of HIV-1 (MAL), contain a TAR sequence predicted to have a two-nucleotide bulge. All other HIV-1 strains (including members of group O) as well as all known SIV_{CPZ} isolates contain a three-nucleotide TAR bulge. This extent of conservation among otherwise highly divergent viruses strongly suggests that changes in the size of the TAR bulge are generally not tolerated. Moreover, since the only viruses shown to have a two-nucleotide bulge were either members of subtype A or recombinants containing genomic regions derived from subtype A, we speculated that there might be a functional link between the two-nucleotide bulge and a subtype A protein. A recombination event bringing together the gene for such a subtype A protein with a non-subtype A LTR might be expected to select for a compensatory change (i.e., deletion of the third bulge nucleotide) to restore optimal interaction. A candidate coding region was identified in the 5′ *pol* gene by mapping IBNG (Fig. 10).

Recent experimental evidence (36) supports the possibility of a TAR-polymerase interaction. Expressing TAR mutant HIV-1 proviruses in the presence of adenovirus E1A and E1B proteins (which complement for the lack of transcriptional activation by Tat), Harrich et al. identified several TAR point mutants which did not substantially alter viral gene expression but led to greatly diminished levels of strong stop DNA synthesis (36). These results suggest that TAR, in addition to playing an important role in Tat-associated transactivation, is essential for initial steps in HIV-1 reverse transcription. Moreover, the structure of the TAR stem-loop was found to be important for this activity (36). TAR thus appears to interact (directly or indirectly) with components of the initiation complex, which includes the p66-p51 RT heterodimer. These findings are consistent with our observation of an association between the TAR bulge size and the phylogenetic subtype of the corresponding 5′ *pol* region. It is also consistent with the mosaic genome structure of SIV infecting West African sabaeus monkeys (43). While these viruses cluster with SIV_{AGM} from other green monkey species in trees derived from the 3′ halves of their genomes, they group with the HIV-2/SIV_{SM} lineage in trees based on the 3′ half of *gag* and the 5′ half of *pol*. Most interestingly, sabaeus monkey viruses also contain a duplication of their TAR element, a feature unique to members of the HIV-2/SIV_{SM} lineage (4, 21). Thus, in both HIV-1 and SIV_{AGM}/SIV_{SM} hybrids, there is an association between the structure of the TAR element and the phylogenetic origin of the 5′ *pol* region. Together with the experimental results by Harrich et al. (36), these findings argue for a functional interaction between RT and TAR and suggest a preference for a two-nucleotide TAR bulge by the subtype A RT.

Intersubtype recombination represents an important new source of HIV-1 variation. Recombination provides the opportunity for evolutionary leaps, as its genetic consequences are

far more dramatic than the steady accumulation of individual mutations. In this context, it is of interest that the "subtype E" epidemic in Thailand has a number of unique features. Most importantly, "subtype E" viruses are spreading with extraordinary rapidity (111, 112). Introduced in the late 1980s along with subtype B viruses, "E" viruses comprise the great majority of HIV-1 infections in Thailand today (13, 81, 111, 112). Moreover, there is evidence of increasing spread of "subtype E" viruses to surrounding countries (112) as well as to the western hemisphere (2, 7). Finally, "subtype E" viruses are transmitted largely through heterosexual contact (46, 52, 66, 77, 79, 82). Although this may be due in part to epidemiological factors, it may also be the result of particular viral properties. The recent report of preferential growth of "subtype E" viruses in Langerhans cells (103), which are believed to constitute the first line of target cells following mucosal transmission (104), could explain the (preferential) epidemic spread of "subtype E" viruses in Thailand. It is presently unknown whether the mosaic genome structure of "subtype E" viruses has any relation to these unusual biological properties. However, examples of recombination events in HIV and other retroviruses indicate that this process can generate viral variants with significantly altered biological properties (9, 30, 48, 107). Thus, a link between the hybrid origin of "subtype E" viruses and their success as an epidemic pathogen must be considered.

Finally, it should be emphasized that intersubtype recombination appears to represent a relatively recent phenomenon. The various HIV-1 subtypes must have evolved in relative isolation from each other for some time, or their members could not have accumulated the number of distinctive changes that now characterize them as discrete subtypes. Then, with epidemic spread and mixing of clades, the opportunity for coinfection and recombination was created. A recent example of this occurred in Brazil, where subtype B and F recombinants have been generated (25, 88). With the population density of the different subtypes still increasing in many countries (42, 75, 78, 113), additional recombinants are likely to be generated. The widespread dissemination and obvious virulence of "subtype E" viruses indicate that intersubtype recombination can generate potent pathogens. As correlates of immune protection arising from natural infection and vaccines are identified, it will be critically important to assess the effects of viral recombination on virus-host interaction. In this context, continuing surveillance of HIV-1 variation worldwide will be essential.

ACKNOWLEDGMENTS

We thank R. B. Gaynor for unpublished data concerning the role of the HIV-1 TAR element in reverse transcription; K. A. Jones for analysis of LTR sequences; A. Frankel, B. R. Cullen, and C. D. Morrow for helpful discussions; J. K. Carr, M. Salminen, and F. E. McCutchan for exchanging unpublished sequence data; B. G. Weniger for information on the Thai epidemic; Y.-L. Chen, B. Jian, and C. Carruthers for excellent technical assistance; the members of the WHO and NIAID Networks for HIV Isolation and Characterization for continuing collaborative interactions; and J. B. Wilson for artwork and preparation of the manuscript.

This work was supported by grants from the National Institutes of Health (NO1 AI35170 and RO1 AI25291), by shared facilities of the UAB Center for AIDS Research (DNA Sequencing Core; P30 AI27767), and by the Birmingham Veterans Administration Medical Center.

REFERENCES

1. **Alizon, M., S. Wain-Hobson, L. Montagnier, and P. Sonigo.** 1986. Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. Cell **46:**63–74.
2. **Artenstein, A. W., J. Coppola, A. E. Brown, J. K. Carr, E. Sanders-Buell, E. Galbarini, J. R. Mascola, T. C. VanCott, P. Schonbrood, and F. E. McCutchan.** 1995. Multiple introductions of HIV-1 subtype E into the western hemisphere. Lancet **346:**1197–1198.
3. **Ausubel, F. M., R. Brent, R. E. Kinston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl. (ed.).** 1987. Current protocols in molecular biology. John Wiley & Sons, New York.

3a. **Barre-Sinoussi, F., M.-C. Georges-Courbot, P. N. Fultz, E. Muchmore, E. van der Ryst, S. Saragosti, G. Dubreuil, A. Georges, and M. Girard.** Characterization and titration of an HIV-1 clade E chimpanzee challenge stock. Submitted for publication.

4. **Berkhout, B.** 1992. Structural features in TAR RNA of human and simian immunodeficiency viruses: a phylogenetic analysis. Nucleic Acids Res. **20:** 27–31.
5. **Bobkov, A., R. Cheingsong-Popov, M. Garaev, A. Rzhaninova, P. Kaleebu, S. Beddows, M. H. Bachmann, J. I. Mullins, J. Louwagie, W. Janssens, G. van der Groen, F. McCutchan, and J. Weber.** 1994. Identification of an *env* G subtype and heterogeneity of HIV-1 strains in the Russian Federation and Belarus. AIDS **8:**1649–1655.
6. **Bolognesi, D. P., L. Corey, S. H. Vermund, and D. F. Hoth.** 1994. HIV vaccine development: a progress report. Ann. Intern. Med. **8:**603–611.
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 USA. Lancet **346:**1198–1199.
8. **Carr, J. K., M. O. Salminen, C. Koch, D. Gotte, A. W. Artenstein, P. A. Hegerich, D. St. Louis, D. S. Burke, and F. E. McCutchan.** 1996. Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. J. Virol. **70:**5935–5943.
9. **Chakrabarti, R., F. M. Hofman, R. Pandep, L. E. Mathes, and P. Roy-Burman.** 1994. Recombination between feline exogenous and endogenous retroviral sequence generates tropism for cerebral endothelial cells. Am. J. Pathol. **144:**348–358.
10. **Coffin, J. M.** 1979. Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. J. Gen. Virol. **42:**1–26.
11. **Coffin, J. M.** 1990. Retroviridae and their replication, p. 1437–1500. *In* B. N. Fields and D. M. Knipe (ed.), Fields virology. Raven Press, New York.
12. **Coffin, J. M.** 1995. HIV population dynamics *in vivo*: implifications for genetic variation, pathogenesis and therapy. Science **267:**483–489.
13. **Cohen, J.** 1995. Thailand weighs AIDS vaccine tests. Science **270:**904–907.
14. **Delassus, S., R. Cheynier, and S. Wain-Hobson.** 1991. Evolution of human immunodeficiency virus type 1 *nef* and long terminal repeat sequence over 4 years in vivo and in vitro. J. Virol. **65:**225–231.
15. **Diaz, R. S., E. C. Sabino, A. Mayer, J. W. Mosley, M. P. Busch, and The Transfusion Safety Study Group.** 1995. Dual human immunodeficiency virus type 1 infection and recombination in a dually exposed transfusion recipient. J. Virol. **69:**3273–3281.
16. **Douglas, N. W., A. I. Knight, A. Hayhurst, W. Y. Barrett, M. J. Kevany, and R. S. Daniels.** 1996. An efficient method for the rescue and analysis of functional HIV-1 *env* genes: evidence for recombination in the vicinity of the *tat/rev* splice site. AIDS **10:**39–46.
17. **Faulkner, D. M., and J. Jurka.** 1988. Multiple aligned sequence editor (MASE). Trends Biochem. Sci. **13:**321–322.
18. **Felsenstein, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:**783–791.
19. **Felsenstein, J.** 1992. PHYLIP (Phylogeny Inference Package), 3.5c ed. Department of Genetics, University of Washington, Seattle, Wash.
20. **Feng, S., and E. C. Holland.** 1988. HIV-1 *tat* trans-activation requires the loop sequence within tar. Nature (London) **334:**165–167.
21. **Fenrick, R., M. H. Malim, J. Hauber, S.-Y. Le, J. Maizel, and B. R. Cullen.** 1989. Functional analysis of the Tat *trans*-activator of human immunodeficiency virus type 2. J. Virol. **63:**5006–5012.
22. **Fultz, P. N.** Personal communication.
23. **Fultz, P. N., Q. Wei, J. Stallworth, F. Barre-Sinoussi, A. Deslandres, and M. Girard.** 1995. Superinfection of chimpanzees with HIV-1 strains representing the same or different subtypes, p. 257–262. *In* M. Girard and B. Dodet (ed.), Retroviruses of human AIDS and related animal diseases. Dixieme Colloque des Cent Gardes. Pasteur Mérieux, Marnes-la-Coquette, France.
24. **Gao, F., and B. H. Hahn.** Unpublished data.
25. **Gao, F., S. G. Morrison, D. L. Robertson, C. L. Thornton, S. Craig, G. Karlsson, J. Sodroski, M. Morgado, B. Galvao-Castro, H. von Briesen, S. Beddows, J. Weber, P. M. Sharp, G. M. Shaw, B. H. Hahn, and WHO and NIAID Networks for HIV Isolation and Characterization.** 1996. Molecular cloning and analysis of functional envelope genes from human immunodeficiency virus type 1 sequence subtypes A through G. J. Virol. **70:**1651–1667.
26. **Gao, F., L. Yue, S. Craig, C. L. Thornton, D. L. Robertson, F. E. McCutchan, J. A. Bradac, P. M. Sharp, and B. H. Hahn.** 1994. Genetic variation of HIV type 1 in four World Health Organization-sponsored vaccine evaluation sites: generation of functional envelope (glycoprotein 160) clones representative of sequence subtypes A, B, C, and E. AIDS Res. Hum. Retroviruses **10:**1359–1368.
27. **Gao, F., L. Yue, A. T. White, P. G. Pappas, J. Barchue, A. P. Hanson, B. M. Greene, P. M. Sharp, G. M. Shaw, and B. H. Hahn.** 1992. Human infection

by genetically diverse SIV related HIV-2 in West Africa. Nature (London) **358:**495–499.
28. **Gaynor, R.** 1992. Cellular transcription factors involved in the regulation of HIV-1 gene expression. AIDS **6:**347–363.
29. **Ghosh, S. K., P. N. Fultz, E. Keddie, M. S. Saag, P. M. Sharp, B. H. Hahn, and G. M. Shaw.** 1993. A molecular clone of HIV-1 tropic and cytopathic for human and chimpanzee lymphocytes. Virology **194:**858–864.
30. **Golovkina, T. V., A. B. Jaffe, and S. R. Ross.** 1994. Coexpression of exogenous and endogenous mouse mammary tumor virus RNA in vivo results in viral recombination and broadens the virus host range. J. Virol. **68:**5019–5026.
31. **Goodrich, D. W., and P. H. Duesberg.** 1990. Retroviral recombination during reverse transcription. Proc. Natl. Acad. Sci. USA **87:**2052–2056.
32. **Graham, B. S., and P. F. Wright.** 1995. Candidate AIDS vaccines. N. Engl. J. Med. **333:**1331–1339.
33. **Groenink, M., A. C. Andeweg, R. A. M. Fouchier, S. Broersen, R. C. M. van der Jagt, H. Schuitemaker, R. E. Y. de Goede, M. L. Bosch, H. G. Huisman, and M. Tersmette.** 1992. Phenotype-associated *env* gene variation among eight related human immunodeficiency virus type 1 clones: evidence for in vivo recombination and determinants of cytotropism outside the V3 domain. J. Virol. **66:**6175–6180.
34. **Hahn, B. H., D. L. Robertson, and P. M. Sharp.** 1995. Intersubtype recombination in HIV-1 and HIV-2, p. III-22–III-29. *In* G. Myers, B. Korber, S. Wain-Hobson, K.-T. Jeang, L. E. Henderson, and G. N. Pavlakis (ed.), Human retroviruses and AIDS 1995: a compilation and analysis of nucleic acid and amino acid sequences. Los Alamos National Laboratory, Los Alamos, N.Mex.
35. **Hahn, B. H., G. M. Shaw, S. K. Arya, M. Popovic, R. C. Gallo, and F. Wong-Staal.** 1984. Molecular cloning and characterization of the HTLV-III virus associated with AIDS. Nature (London) **321:**166–169.
36. **Harrich, D., C. Ulich, and R. B. Gaynor.** 1996. A critical role for the TAR element in promoting efficient human immunodeficiency virus type 1 reverse transcription. J. Virol. **70:**4017–4027.
37. **Higgins, D. G., and P. M. Sharp.** 1989. Fast and sensitive multiple sequence alignments on a microcomputer. Comput. Appl. Biosci. **5:**151–153.
38. **Ho, D. D., A. U. Neumann, A. S. Perelson, W. Chen, J. M. Leonard, and M. Markowitz.** 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature (London) **373:**123–126.
39. **Howell, R. M., J. E. Fitzgibbon, M. Noe, Z. Ren, D. J. Gocke, T. A. Schwartzer, and D. T. Dubin.** 1991. *In vivo* sequence variation of the human immunodeficiency virus type 1 *env* gene: evidence for recombination among variants found in a single individual. AIDS Res. Hum. Retroviruses **7:**869–876.
40. **Hu, W.-S., and H. M. Temin.** 1990. Retroviral recombination and reverse transcription. Science **250:**1227–1233.
41. **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.
42. **Jameel, S., M. Zafrullah, M. Ahmad, G. S. Kapoor, and S. Sehgal.** 1995. A genetic analysis of HIV-1 from Punjab, India reveals the presence of multiple variants. AIDS **9:**685–690.
43. **Jin, M. J., H. Hui, D. L. Robertson, M. C. Muller, F. Barre-Sinoussi, V. M. Hirsch, J. S. Allan, G. M. Shaw, P. M. Sharp, and B. H. Hahn.** 1994. Mosaic genome structure of simian immunodeficiency virus from West African green monkeys. EMBO J. **13:**2935–2947.
44. **Jones, K. A., and B. M. Peterlin.** 1994. Control of RNA initiation and elongation at the HIV-1 promoter. Annu. Rev. Biochem. **63:**717–743.
45. **Kaleebu, P., A. Bobkov, R. Cheingsong-Popov, P. Bieniasz, M. Garaev, and J. Weber.** 1995. Identification of HIV-1 subtype G from Uganda. AIDS Res. Hum. Retroviruses **11:**657–659.
46. **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. L. 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.
47. **Katz, R. A., and A. M. Skalka.** 1990. Generation of diversity in retroviruses. Annu. Rev. Genet. **24:**409–445.
48. **Kellam, P., and B. A. Larder.** 1995. Retroviral recombination can lead to linkage of reverse transcriptase mutations that confer increased zidovudine resistance. J. Virol. **69:**669–674.
49. **Kimura, M.** 1980. A simple method for estimating evolutionary rates of base substitutions through comparitive studies of nucleotide sequences. J. Mol. Evol. **16:**111–120.
50. **Korber, B. T. M., S. Osmanov, J. Esparza, G. Myers, and WHO Network for HIV Isolation and Characterization.** 1994. The World Health Organization Global Programme on AIDS Proposal for Standardization of HIV Sequence Nomenclature. AIDS Res. Hum. Retroviruses **10:**1355–1358.
51. **Kostrikis, L. G., E. Bagdades, Y. Cao, L. Zhang, D. Dimitriou, and D. D. Ho.** 1995. Genetic analysis of human immunodeficiency virus type 1 strains from patients in Cyprus: identification of a new subtype designated subtype I. J. Virol. **69:**6122–6130.
52. **Kunanusont, C., H. M. Foy, J. K. Kreiss, S. Rerks-Ngarm, P. Phanuphak, S. Raktham, C.-P. Pau, and N. L. Young.** 1995. HIV-1 subtypes and male-to-female transmission in Thailand. Lancet **345:**1078–1083.
53. **Kusumi, K., B. Conway, S. Cunningham, A. Berson, C. Evans, A. K. N. Iversen, D. Colvin, M. V. Gallo, S. Coutre, E. G. Shpaer, D. V. Faulkner, A. deRonde, S. Volkman, C. Williams, M. S. Hirsch, and J. I. Mullins.** 1992. Human immunodeficiency virus type 1 envelope gene structure and diversity in vivo and after cocultivation in vitro. J. Virol. **66:**875–885.
54. **Larder, B. A., P. Kellam, and S. D. Kemp.** 1993. Convergent combination therapy can select viable multidrug-resistant HIV-1 *in vitro*. Nature (London) **365:**451–453.
55. **Leitner, T., A. Alaeus, S. Marquina, E. Lilja, K. Lidman, and J. Albert.** 1995. Yet another subtype of HIV type 1? AIDS Res. Hum. Retroviruses **11:**995–997.
56. **Leitner, T., D. Escanilla, S. Marquina, J. Wahlberg, C. Brotstrom, 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.
57. **Li, W.-H.** 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. **36:**96–99.
58. **Li, W.-H., M. Tanimura, and P. M. Sharp.** 1988. Rates and dates of divergence between AIDS virus nucleotide sequences. Mol. Biol. Evol. **5:**313–330.
59. **Li, Y., J. C. Kappes, J. A. Conway, R. W. Price, G. M. Shaw, and B. H. Hahn.** 1991. Molecular characterization of human immunodeficiency virus type 1 cloned directly from uncultured human brain tissue: identification of replication-competent and -defective viral genomes. J. Virol. **65:**3973–3985.
60. **Liu, H., K. X. Wu, M. Newman, G. M. Shaw, B. H. Hahn, and J. C. Kappes.** 1995. The Vif protein of human and simian immunodeficiency viruses is packaged into virions and associates with viral core structures. J. Virol. **69:**7630–7638.
61. **Loussert-Ajaka, I., M.-L. Chaic, B. T. M. Korber, F. Letournearu, E. Gomas, E. Allen, T.-D. Ly, F. Brun-Vezinet, F. Simon, and S. Saragosti.** 1995. Variability of human immunodeficiency virus type 1 group O strains isolated from Cameroonian patients living in France. J. Virol. **69:**5640–5649.
62. **Louwagie, J., F. E. McCutchan, M. Peeters, T. P. Brennan, E. Sanders-Buell, G. A. 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.
63. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 269–295. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
64. **Mansky, L. M., and H. M. Temin.** 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. J. Virol. **69:**5087–5094.
65. **Martin-Gallardo, A., J. Lamerdin, and A. Carrano.** 1994. Shotgun sequencing, p. 37–41. *In* M. D. Adams, C. Fields, and J. C. Venter (ed.), Automated DNA sequencing and analysis. Academic Press, London.
66. **Mastro, T. D., G. A. Satten, T. Nobkesorn, S. Sangkharomya, and I. M. Longini, Jr.** 1994. Probability of female-male transmission of HIV-1 in Thailand. Lancet **343:**204–207.
67. **McCutchan, F. E., P. A. Hegerich, T. P. Brennan, P. Phanuphak, P. Singharaj, A. Jugsudee, P. W. Berman, A. M. Gray, A. K. Fowler, and D. S. Burke.** 1992. Genetic variants of HIV-1 in Thailand. AIDS Res. Hum. Retroviruses **8:**1887–1895.
68. **Miyamoto, S., and I. M. Verma.** 1995. Rel/NF-κB/IκB story. Adv. Cancer Res. **66:**255–292.
69. **Moore, J., and R. Anderson.** 1994. The WHO and why of HIV vaccine trials. Nature (London) **372:**313–314.
70. **Moore, J. P., and D. D. Ho.** 1995. HIV-1 neutralization: the consequences of viral adaptation to growth on transformed T cells. AIDS **9:**S117–S136.
71. **Muesing, M., D. H. Smith, C. Cabradilla, C. Benton, L. Lasky, and D. Capon.** 1985. Nucleic acid structure and expression of human AIDS retrovirus. Nature (London) **313:**450–458.
72. **Murphy, E., B. Korber, M.-C. Georges-Courbot, B. You, A. Pinter, D. Cook, M.-P. Kieny, A. Georges, C. Mathiot, F. Barre-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:**997–1007.
73. **Myers, G.** 1994. HIV: between past and future. AIDS Res. Hum. Retroviruses **10:**1317–1324.
74. **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.
75. **Myers, G., B. Korber, B. H. Hahn, K.-T. Jeang, J. W. Mellors, F. E. McCutchan, L. E. Henderson, and G. N. Pavlakis.** 1995. Human retroviruses and AIDS 1995: a compilation and analysis of nucleic acid and amino acid sequences. Los Alamos National Laboratory, Los Alamos, N.Mex.
76. **Najera, I., D. D. Richman, I. Olivares, J. M. Rojas, M. A. Peinado, M. Perucho, R. Najera, and C. Lopez-Galindez.** 1994. Natural occurrence of

drug resistance mutations in the reverse transcriptase of human immunodeficiency virus type 1 isolates. AIDS Res. Hum. Retroviruses **10:**1479–1488.

77. **Nelson, K. E., D. D. Celentano, S. Suprasert, N. Wright, S. Eiumtrakul, S. Tulvatana, A. Matanasarawoot, P. Akarasewi, S. Kontolbutra, S. Romyen, N. Sirisopana, and C. Theetranont.** 1993. Risk factors for HIV infection among young adult men in Northern Thailand. JAMA **270:**955–960.
78. **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.
79. **Nopkesorn, T., T. D. Mastro, S. Sangkharomya, M. Sweat, P. Singharaj, K. Limpakarnjanarat, H. D. Gayle, and B. G. Weniger.** 1993. HIV-1 infection in young men in northern Thailand. AIDS **7:**1233–1239.
80. **Ou, C.-Y., Y. Takebe, C.-C. Luo, M. Kalish, W. Auwanit, C. Bandea, N. de la Torre, J. L. Moore, G. Schochetman, S. Yamazaki, 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.
81. **Ou, C.-Y., Y. Takebe, B. G. Weniger, C.-C. Luo, M. L. Kalish, W. Auwanit, S. Yamazaki, H. D. Gayle, N. L. Young, and G. Schochetman.** 1993. Independent introductions of two major HIV-1 genotypes into distinct high-risk populations in Thailand. Lancet **341:**1171–1174.
82. **Pau, C. P., S. Lee-Thomas, W. Auwanit, J. R. George, C.-Y. Ou, B. S. Parekh, T. C. Granade, D. L. Holloman, S. Phillips, G. Schochetman, K. Limpakarnjanarat, N. L. Young, Y. Takebe, H. D. Gayle, and B. G. Weniger.** 1992. Highly specific V3 peptide enzyme immunoassay for serotyping HIV-1 specimens from Thailand. AIDS **7:**337–340.
83. **Perelson, A. S., A. U. Neumann, M. Markowitz, J. M. Leonard, and D. D. Ho.** 1996. HIV-1 dynamics *in vivo*: virion clearance rate, infected cell lifespan, and viral generation time. Science **271:**1582–1586.
84. **Richman, D. D.** 1995. Protease uninhibited. Nature (London) **374:**494.
85. **Robertson, D. L., B. H. Hahn, and P. M. Sharp.** 1995. Recombination in AIDS viruses. J. Mol. Evol. **40:**249–259.
86. **Robertson, D. L., P. M. Sharp, F. E. McCutchan, and B. H. Hahn.** 1995. Recombination in HIV-1. Nature (London) **374:**124–126.
87. **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:**1364–1373.
88. **Sabino, E. C., E. G. Shpaer, M. G. Morgado, B. T. M. Korber, R. S. Diaz, V. Bongertz, S. Cavalcante, B. Gälvao-Castro, J. I. Mullins, and A. Mayer.** 1994. Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiologically linked individuals from Brazil. J. Virol. **68:**6340–6346.
89. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
90. **Sala, M., E. Pelletier, and S. Wain-Hobson.** 1995. HIV-1 gp120 sequences from a doubly infected drug user. AIDS Res. Hum. Retroviruses **11:**653–655.
91. **Salminen, M. O., J. K. Carr, D. S. Burke, and F. E. McCutchan.** 1995. Genotyping of HIV-1, p. III-30–III-34. *In* G. Myers, B. Korber, S. Wain-Hobson, K.-T. Jeang, L. E. Henderson, and G. N. Pavlakis (ed.), Human retroviruses and AIDS 1995: a compilation and analysis of nucleic acid and amino acid sequences. Los Alamos National Laboratory, Los Alamos, N.Mex.
92. **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.
93. **Salminen, M. O., J. K. Carr, D. L. Robertson, P. Hegerich, D. Gotte, C. Koch, E. Sanders-Buel, F. Gao, P. M. Sharp, B. H. Hahn, D. S. Burke, and F. E. McCutchan.** Unpublished data.
94. **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.
95. **Shafer, R. W., and T. C. Merigan.** 1995. Editorial: new virologic tools for the design and analysis of clinical trials. J. Infect. Dis. **171:**1325–1328.
96. **Sharp, P. M., D. L. Robertson, F. Gao, and B. H. Hahn.** 1994. Origins and diversity of human immunodeficiency viruses. AIDS **8:**S27–S42.
97. **Sharp, P. M., D. L. Robertson, and B. H. Hahn.** 1995. Cross-species transmission and recombination of AIDS viruses. Philos. Trans. R. Soc. London Ser. B **349:**41–47.
98. **Shaw, G. M., B. H. Hahn, S. K. Arya, J. E. Groopman, R. C. Gallo, and F. Wong-Staal.** 1984. Molecular characterization of human T-cell leukemia (lymphotropic) virus type III in the acquired immune deficiency syndrome. Science **226:**1165–1171.
99. **Siepel, A. C., and B. T. Korber.** 1995. Scanning the database for recombinant HIV-1 genomes, p. III-35–III-60. *In* G. Myers, B. Korber, S. Wain-Hobson, K.-T. Jeang, L. E. Henderson, and G. N. Pavlakis (ed.), Human retroviruses and AIDS 1995: a compilation and analysis of nucleic acid and amino acid sequences. Los Alamos National Laboratory, Los Alamos, N.Mex.
100. **Sittitrai, W., and T. Brown.** 1994. Risk factors for HIV infection in Thailand. AIDS **8**(suppl. 2)**:**S143–S153.
101. **Smith, R. F., and T. F. Smith.** 1990. Automatic generation of primary sequence patterns from sets of related protein sequences. Proc. Natl. Acad. Sci. USA **87:**118–122.
102. **Smith, R. F., and T. F. Smith.** 1992. Pattern-induced multi-sequence alignment (PIMA) algorithm employing secondary structure-dependent gap penalties for use in comparative protein modeling. Protein Eng. **5:**35–41.
103. **Soto-Ramirez, L. E., B. Renjifo, M. F. McLane, R. Marlink, C. O'Hara, R. Sutthent, C. Wasi, P. Vithayasai, V. Vithayasai, C. Apichartpiyakul, P. Auewarakul, V. Pena Cruz, D.-S. Chui, R. Osathanondh, K. Mayer, T.-H. Lee, and M. Essex.** 1996. HIV-1 langerhans' cell tropism associated with heterosexual transmission of HIV. Science **271:**1291–1293.
104. **Spira, A. I., P. A. Marx, B. K. Patterson, J. Mahoney, R. A. Koup, S. M. Wolinsky, and D. D. Ho.** 1996. Cellular targets of infection and route of viral dissemination after an intravaginal inoculation of simian immunodeficiency virus into rhesus macaques. J. Exp. Med. **183:**215–225.
105. **Stuhlmann, H., and P. Berg.** 1992. Homologous recombination of copackaged retrovirus RNAs during reverse transcription. J. Virol. **66:**2378–2388.
106. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W—improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
107. **Tumas, K. M., J. M. Poszgay, N. Avidan, S. J. Ksiazek, B. Overmoyer, K. J. Blank, and M. B. Prystowsky.** 1993. Loss of antigenic epitopes as the result of *env* gene recombination in retrovirus-induced leukemia in immunocompetent mice. Virology **192:**587–595.
108. **Vartanian, J.-P., A. Meyerhans, B. Asjo, and S. Wain-Hobson.** 1991. Selection, recombination, and G→A hypermutation of human immunodeficiency virus type 1 genomes. J. Virol. **65:**1779–1788.
109. **Wasi, C., B. Herring, S. Raktham, S. Vanichseni, T. D. Mastro, N. L. Young, H. Rübsamen-Waigmann, H. von Briesen, M. L. Kalish, C.-C. Luo, C.-P. Pau, A. Baldwin, J. I. Mullins, E. L. Delwart, 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 the heteroduplex mobility assay: evidence of increasing infection with HIV-1 subtype E. AIDS **9:**943–949.
110. **Wei, X., S. K. Ghosh, M. E. Taylor, V. A. Johnson, E. A. Emini, P. Deutsch, J. D. Lifson, S. Bonhoeffer, M. A. Nowak, B. H. Hahn, M. S. Saag, and G. M. Shaw.** 1995. Viral dynamics in HIV-1 infection. Nature (London) **73:**117–122.
111. **Weniger, B. G., K. Limpakarnjanarat, K. Ungchusak, S. Thanprasertsuk, K. Choopanya, S. Vanichseni, T. Uneklabh, P. Thongcharoen, and C. Wasi.** 1991. The epidemiology of HIV infection and AIDS in Thailand. AIDS **5**(suppl. 2)**:**S71–S85.
112. **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.
113. **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:**1327–1343.
114. **Yu, X.-F., Z. Wang, C. Beyrer, D. D. Celentano, C. Khamboonruang, E. Allen, and K. Nelson.** 1995. Phenotypic and genotypic characteristics of human immunodeficiency virus type 1 from patients with AIDS in northern Thailand. J. Virol. **69:**4649–4655.