Pol gene diversity of five human immunodeficiency virus type 1 subtypes: evidence for naturally occuring mutations that contribute to drug resistance, limited recombination patterns and common ancestry for subtypes B and D

Cornelissen, M.T.E.; van den Burg, R.; Zorgdrager, F.K.; Lukashov, V.V.; Goudsmit, J.

Publication date
1997

Published in
Journal of Virology

JOURNAL OF VIROLOGY, Sept. 1997, p. 6348–6358 Vol. 71, No. 9
0022-538X/97/$04.00+0

pol Gene Diversity of Five Human Immunodeficiency Virus Type 1 Subtypes: Evidence for Naturally Occurring Mutations That Contribute to Drug Resistance, Limited Recombination Patterns, and Common Ancestry for Subtypes B and D

MARION CORNELISSEN,* REMCO VAN DEN BURG, FOKLA ZORGDRAGER, VLADIMIR LUKASHOV, AND JAAP GOUDSMIT

Department of Human Retrovirology, Academic Medical Centre, Amsterdam, The Netherlands

Received 25 March 1997/Accepted 27 May 1997

Naturally occurring mutations in the polymerase gene of human immunodeficiency virus type 1 (HIV-1) have important implications for therapy and the outcome of clinical studies. Using 42 virus isolates obtained from the UNAIDS sample collection, we analyzed the protease (99 amino acids [aa]) and the first 297 aa of reverse transcriptase (RT) coding regions. Based on the V3 sequence analysis, the collection includes subtype A (n = 5), subtype B (n = 12), subtype C (n = 1), subtype D (n = 11), and subtype E (n = 13) viruses. Of the 42 protease genes, 37 contained naturally occurring mutations at positions in the gene that contribute to resistance to protease inhibitors (indinavir, saquinavir, ritonavir, and nelfinavir) in clade B isolates. The phenotypic effect of these substitutions in non-B isolates is unclear. The 5′half RT coding region of the 42 isolates was found to be less variable, although 19 of the 42 RT sequences contained amino acid substitutions known to contribute to nucleoside and/or nonnucleoside drug resistance. Since the virus isolates were obtained in 1992, it is unlikely that the infected subjects received protease inhibitors, but we found evidence that one subject acquired a zidovudine (AZT)-resistant HIV-1 strain from a contact who had received AZT. Phylogenetic analysis identified five subtype *pol* clusters: A, B, C, D, and A′. Comparison of *env* and *pol* sequences of the same viruses showed no more recombination events than were already identified on the basis of *gag*/*env* comparison (M. Cornelissen, G. Kampinga, F. Zorgdrager, J. Goudsmit, and the UNAIDS Network for HIV Isolation and Characterization, J. Virol. 70:8209–8212, 1996). In one of the known recombinants, a crossover site between subtypes A and C could be identified, and in another, a crossover site could not be identified due to lack of a reference subtype F *pol* sequence. We analyzed the *ds*/*da* ratio of *gag*, *pol*, and *env* sequences of 35 isolates, excluding the recombinants. Our analysis showed that *gag* and *pol* are subjected to purifying selection with an average ds/da ratio above 1, independent of the subtype and in contrast with V3 (ds/da ≈ 1). Based on the low ds/da ratio of the intergroup analysis of A/E and B/D *gag* and *pol* sequences, we analyzed the evolutionary relation between subtypes B and D in more detail by constructing separate phylogenetic trees for synonymous and nonsynonymous substitutions. Our analysis suggests a common ancestry for subtypes B and D that is distinct from that of subtypes A and E.

Human immunodeficiency virus type 1 (HIV-1) causes AIDS in epidemic waves on all continents. Several factors are known to contribute to the generation of new virus variants and to influence the speed with which these viruses evolve. One is the error-prone nature of the viral reverse transcriptase (RT), which lacks proofreading functions, resulting in nucleotide substitutions, deletions, and insertions. A rate of 3.4×10^{-5} misincorporation per replication cycle in vivo has been suggested (30). A second factor is the high rate of virus production (up to 10^{10} virions per day) and the large number of replication cycles (approximately 300 per year) that sustains HIV-1 infection in vivo (6, 15, 44). The third and most important factor is the rapid selection for viruses of distinct fitness due to immune pressure, coreceptor selection, and/or effective antiviral drugs (44). Recently, recombination has been described as a fourth factor that contributes significantly to HIV-1 diversity (9, 11, 12, 41, 42).

Recombination is a fundamental property of retroviruses because of their dimeric RNA genome and an RT that can switch between templates during proviral DNA synthesis (5, 17, 18). However, the finding of coinfection or superinfection with multiple HIV-1 strains is rare, despite clear evidence of intersubtype recombination (19). Analysis of sequences submitted to the HIV-1 database indicates that recombination occurs more frequently than was previously assumed. Among the 114 published sequences of viruses, at least 10 virus genomes appear to be recombinants with recombinations in the *gag* and/or *env* sequences (41). Using the UNAIDS sample collection, a repository for HIV strains of four World Health Organization WHO-sponsored sites for HIV vaccine efficacy trials, we found an even higher proportion of mosaic HIV-1 genomes (9). In this study, 23 of 53 isolates are intersubtype recombinants. Of these 23 recombinants, 19 had a *gag* gene fragment from one subtype and an *env* fragment from another (B^{env}/C^{gag}, A^{env}/C^{gag}, D^{env}/A^{gag}, and E^{env}/A^{gag}). In the remaining four recombinant viruses, crossover sites in *gag* are identified. Among these intersubtype recombinants are 15 *env*-

* Corresponding author. Department of Human Retrovirology, Academic Medical Centre, University of Amsterdam, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands. Phone: (31-20) 5664853. Fax: (31-20)6916531. E-mail: M.I.Cornelissen@amc.uva.nl.

defined subtype E viruses, which are prevalent in Thailand. Thai subtype E viruses are highly related to each other, presumably reflecting a founder effect (31, 36). Phylogenetic analysis clustered the Thai *gag* sequences as an outgroup of subtype A, suggesting that these viruses may belong to a distinct A′ cluster (9).

The full-length proviral "subtype E" virus sequence shows at least eight points of recombination crossover (13). Most of the long terminal repeat, a 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*, and both exons of *tat* and *rev*, are derived from subtype A.

Currently, few sequence data are available for the *polymerase* (*protease* and *RT*) genes of different subtypes, which is surprising since RT and protease (Prot) are the most important targets of antiretroviral therapy. However, multiple mutations in the *pol* (*prot* and *RT*) gene that confer drug resistance in a subtype B background have been mapped (3, 4, 7, 25, 34, 39). By 1985, Mitsuya et al. (32) had already shown that zidovudine (AZT) inhibits HIV replication. Subsequently, the antiretroviral effect of the drug was shown in patients, and resistant isolates were obtained. New dideoxynucleotide and nonnucleoside analogs were later developed, and specific resistant mutations in RT were identified. Since 1995, Prot inhibitors have entered clinical studies. Analysis of viral isolates grown in cell culture and isolates from treated infected patients have demonstrated that approximately 20 of the 99 amino acids (aa) of Prot can undergo mutations when selection is applied by various Prot inhibitors (8). The prevalence in drug-naive patients of resistant mutations of subtype B have been described for the *prot* and *RT* genes (23, 26, 33, 35).

In this study, we present sequence analyses of a part of the *pol* gene, which codes for the entire Prot and half of RT, of viruses isolates from Africa (Rwanda and Uganda), South America (Brazil), and Asia (Thailand).

MATERIALS AND METHODS

Samples. Infectious virus stock and homologous plasma samples were collected by the World Health Organization (WHO, now named UNAIDS) network for isolation and characterization of HIV. Blood samples were collected from 53 HIV-1-infected individuals in primary sites in Brazil, Rwanda, Thailand, and Uganda. Virus stocks were produced by cocultivation of 5×10^5 patient peripheral blood mononuclear cells with 12×10^6 phytohemagglutinin-stimulated donor peripheral blood mononuclear cells. p24 antigen and/or RT assays were used to monitor virus replication. The culture incubations were terminated when positive, and culture supernatant was filtered, divided into aliquots, and frozen in liquid nitrogen. Due to insufficient aliquots of plasma samples, we used the frozen virus supernatant stock for this study. In total, 42 virus stocks were analyzed: 13 from Brazil, 12 from Uganda, 13 from Thailand, and 4 from Rwanda. The samples are listed according to WHO/UNAIDS nomenclature, denoting the year of isolation, the country of origin (two-letter code), and the isolate number (see Table 1) (22).

PCR. Nucleic acids were isolated by the procedure described by Boom et al. (2). A 5-μl sample of virus supernatant stock was added to lysis buffer containing guanidine thiocyanate, Triton X-100, and EDTA. Under such high-salt conditions, nucleic acids bind to silicon dioxide particles, which act as the solid phase. After washing and elution from the solid phase with 60 μl of sterile H_2O, 10 μl of the eluate was used in a reverse transcription reaction with the antisense primer 3′half RT (5′TGACCCATCAAAAGACTTAATAGCAGAAATA, HxB2 positions 3505 to 3535). The buffers, avian myeloblastosis virus RT units, and incubation time were identical to those described previously (10). After incubation for 45 min at 42°C, we added a PCR mixture containing the sense primer 5′Prot FM (5′CAAGGGAAGGCCAGGGAATTT, HxB2 positions 2111 to 2130), PCR buffer, deoxynucleoside triphosphates, 2.5 mM $MgCl_2$, and 2 U of *Taq* polymerase (Perkin-Elmer Cetus). After incubation for 5 min at 95°C, the reaction mixture was subjected to 35 cycles of amplification in a type 480 DNA thermal cycleγ (Perkin-Elmer Cetus). The cycling profiles were 95°C for 1 min, 55°C for 1 min, and 72°C for 2 min. A nested PCR was performed to obtain enough material for direct sequencing. Besides, the first PCR product of 1,424 bp was too large to sequence at once. Three nested PCRs were designed, named fragments A, B, and C, with overlapping sequences. One-twentieth of the first PCR product was amplified for 25 cycles with the sense primer 5′SP6-prot/RT (5′CTTTAACTTCCCTCAGATCACT, HxB2 positions 2242 to 2263) and the antisense primer 3′T7prot (5′CCTATTGAAACTGTACCAGTA, HxB2 positions 2558 to 2578) to generate a 336-bp fragment. Designated fragment A, it contained the encoding information for the entire Prot (99 aa) and the Prot/RT proteolytic cleavage site. Fragment B, a 543-bp fragment, was generated with the sense primer 5′SP6 P66/out (5′GACCTACACCTGTCAACATAAT, HxB2 positions 2484 to 2505) and the antisense primer 3′end protT7 (5′TGGAAAG GATCACCAGCAATATT, HxB2 positions 3005 to 3027) and had a 94-bp overlap with fragment A. Fragment C, a 487-bp fragment, was generated with the sense primer 5′SP6 P66 (5′AGATATCAGTACAATGTGTT, HxB2 positions 2975 to 2994) and the antisense primer 3′half Pol (5′AAGCAGAG CTAGAACTGGCAGA, HxB2 positions 3441 to 3462). This fragment had a 52-bp sequence overlap with fragment B. All three nested PCRs worked with a final 2.4 mM $MgCl_2$ concentration. The presence of amplified material was verified by electrophoresis on 1% agarose gels stained with ethidium bromide.

DNA sequencing. The PCR fragments were directly sequenced on both strands. The nested primer pairs were extended with an SP6 primer sequence (5′GATTTAGGTGACATATAG) for the sense primers and with a T7 sequence (5′TAATACGACTCACTATAGGG) for the antisense primers to enable direct sequencing with universal primers. Sequencing was performed with *Taq* dye primers (Applied Biosystems, Foster City, Calif.) and the Thermo Sequenase fluorescence-labelled primer cycle-sequencing kit (Amersham International plc, Litle Chalfont, England). The sequence products were analyzed on an automatic sequencer (Applied Biosystems DNA sequencer model 370A and 373A stretch). In total, 1,176 nucleotides (nt), encoding 392 aa, were sequenced, including 297 aa of RT and 95 aa of Prot (the first 4 aa of Prot were part of the 5′ nested sense primer). This sequencing method allowed the accurate identification of the nucleotide sequence, but when a polymorphic nucleotide position was illegible, an X was introduced in the nucleotide sequence. In the amino acid sequence, a polymorphic position is marked with an exclamation point. The direct sequence of sample RW08 fragment C was illegible, and this PCR product was cloned into a TA cloning system (Invitrogen, San Diego, Calif.). Six colonies were sequenced, and a consensus sequence was derived. The consensus was the majority consensus containing at each position the nucleotide that occurred most frequently in the clonal sequences.

Alignment of the sequence was straightforward and was performed manually. Synonymous and nonsynonymous nucleotide p distances (ds and da, respectively) were calculated with the MEGA program (24). All statistical calculations were done with SPSS/PC+ software (version 5.0; SPSS Inc., Chicago, Ill.). Phylogenetic analysis of all sequences was done by the neighbor-joining method as implemented in the MEGA program (24). The distance matrix was generated by Kimura's two-parameter model (21). The statistical strength of the neighbor-joining method was assessed by bootstrap resampling (1,000 data sets).

Nucleotide sequence accession numbers. The *pol* sequences reported in this study have been assigned GenBank accession no. AF009369 to AF009410. The *gag* (9) sequences were given accession no. U86530 to U86581. For the *env* sequences see reference 10. These sequences were used to examine the evolutionary relationship among different phylogenetic trees.

RESULTS

Protease gene. To extend our recombination study of the UNAIDS collection virus set, we sequenced the *prot* (encoding a 99-aa sequence) and *RT* (encoding a 297-aa sequence) genes of 42 viruses, including 13 subtype E viruses. Figure 1A shows the deduced amino acid sequences of the *prot* gene. The first 4 aa, PQIT, are part of the 5′ sequence primer and are excluded from the figure. The sequences are aligned against a consensus derived by assigning to each position the amino acid most frequently found in the individual isolates. For the entire data set, the amino acids that define the cleavage sites at either end of the protein are absolutely conserved. Only the *prot* coding domain is shown in Fig. 1A, but the amino acids immediately downstream of the cleavage site that release the mature Prot are similarly conserved (Fig. 1B). Also conserved in all subtypes are the sensitive domains identified by Loeb et al. (28), including the DTGA, active-site triad characteristic of aspartic proteases (positions 22 to 33), the flap region (positions 47 to 52), and the hydrophobic core of the molecule (positions 74 to 84).

Articles reporting the prevalence of mutations in HIV-1 isolates from individuals untreated with protease inhibitors described the presence of resistance-associated mutations in subtype B isolates (1, 23, 26). Here we examine the *pol* sequences for naturally occurring amino acid changes that are

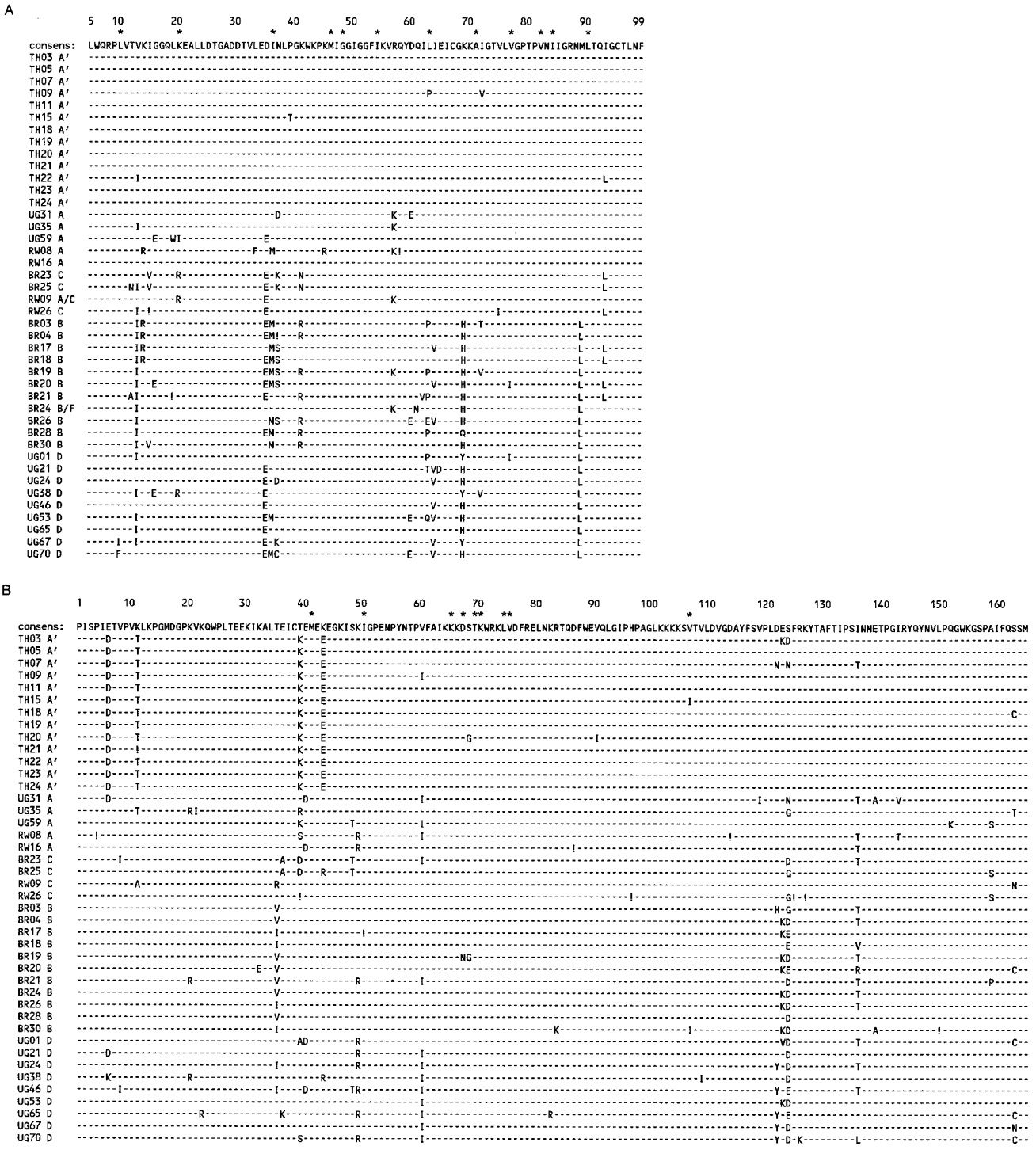

FIG. 1. Deduced amino acid sequences of the Prot (99 aa) and 5′half of RT (297 aa). Each individual sequence is aligned against a consensus, which is derived by assigning to each position the deduced amino acid most frequently found. Dashes indicate identity with the consensus; exclamation points indicate positions where the direct sequences analysis showed polymorphism; asterisks indicate positions in the gene associated with drug resistance. The samples are listed according to subtype, which is denoted after the sample name. (A) Prot sequence. The first 4 aa, PQIT, are part of the 5′ sequence primer and are excluded from this alignment. (B) RT aa sequence starting with position 1, proline.

known to contribute to drug resistance in subtype B isolates. The phenotypic significance of these mutations in isolates other than subtype B isolates is uncertain. Five of the previously described amino acid changes (L10F/I, K20R, M36I, L63P, and V77I) observed within the enzymes that are known to contribute to drug resistance (asterisks in Fig. 1A) were found to occur as natural polymorphisms in isolates of Prot-inhibitor-naive patients. These latter substitutions are not at

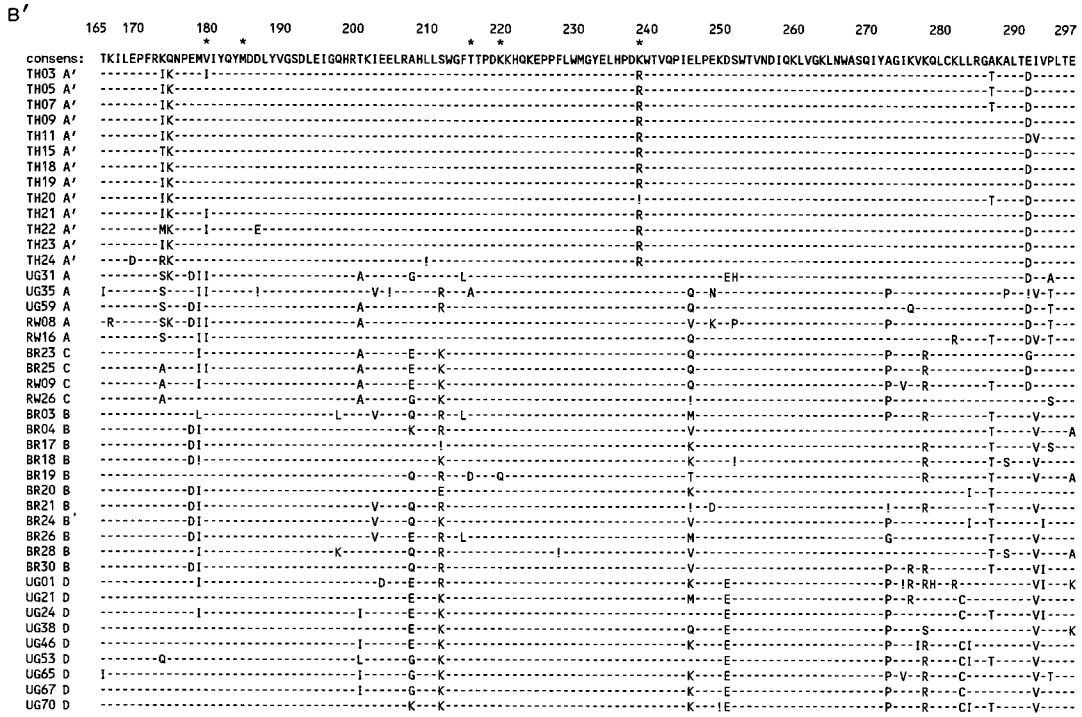

FIG. 1—*Continued.*

the critical positions in mediating the expression of virus resistance. These polymorphisms are more likely to play compensatory roles that permit the accommodation of mutations that directly affect resistance. L10I contributes to indinavir and saquinavir resistance, K20R contributes to indinavir and ritonavir resistance, M36I contributes to ritonavir and nelfinavir resistance, L63P contributes to indinavir resistance, and V77I contributes to nelfinavir resistance. Besides the amino acid substitutions known to contribute to resistance, a few unusual substitutions were observed: L10F, K20I, and L63T/Q/E. In eight isolates, at least two of the following polymorphisms were observed: TH09 and BR21 (M36I and L63P), UG59 (K20I and M36I), BR23, UG38 and RW09 (K20R and M36I), UG21 (M36I and L63T), and UG67 (L10I and M36I). UG01 contained three of these polymorphisms (M36I, L63P, and V77I). The M36I substitution, known to contribute to ritonavir and nelfinavir resistance, was found in 30 of the 42 (71%) isolates examined. A Met at position 36 may represent a subtype B specific substitution, since it was observed in 9 of 10 subtype B isolates but only in 3 of 32 non-B isolates. The frequently observed amino acid polymorphism at position L63P (23) also appeared more common in a subtype B background, appearing in 4 of 10 subtype B isolates compared with 2 of 32 non-B isolates.

***RT* gene.** All *RT* genes show a full-length open reading frame, with unique substitutions along the entire region compared with the consensus sequence (Fig. 1B and B′). Some substitutions can be classified as subtype specific amino acids, e.g., 43E, 173I, and 238R for subtype A′; 35V for subtype B; and 250 E for subtype D.

Crystallographic studies on RT have shown that mutations affecting codons related to resistance to RT inhibitors are clustered in two regions on RT (aa 41 to 75 on one side and 184 to 219 on the other site), suggesting that resistance could be affected by altering interactions between RT and the template. The mutation-affected codons related to resistance are marked in Fig. 1B and B′. In two isolates, we observed mutations at codon 215 that are related to AZT resistance. In BR19, amino acid replacements D67N, T215D, and K219Q were found, and in UG35, Thr was replaced by Ala at codon 215. The AZT resistance properties of Asp and Ala at codon 215 is still unknown, although an Asp was observed 20 months after seroconversion in the recipient of a conferred AZT-resistant-transmission case (14). Additional mutations critical to the loss of sensitivity to nonnucleoside analogs have also been identified. TH20 of subtype A′ and BR30 of subtype B showed substitution V108I, which has been associated with resistance to pyridinones. We found substitutions at other positions that are critical for the loss of sensitivity to the nonnucleoside analogs. At position 179, V179D is associated with resistance to pyridinones, and we observed V179I in eight samples; all subtype A′ *pol* sequences contained a K238R substitution instead of the K238T, a mutation resistant for ateviridine (BHAP), another nonnucleoside. One mutation, D187E, was unexpected since it lies in the conserved YMDD motif thought to form part of the active site of the RT polymerase domain. Mutagenesis studies have shown that mutations at this site can severely impair RT function (25). However, high-level resistance to lamuvidine (3TC) has been documented to occur after substitution of Val for Met in the YMDD motif. The D-to-E substitution in isolate TH22 is a conserved amino acid substitution, and because the virus grows very well, this substitution is highly unlikely to change the viability.

Phylogenetic analyses. A primary objective of this study was to extend our previous work (9) by analyzing the newly determined *pol* sequences for evidence of mosaicism generated by intersubtype recombination (40, 41). Phylogenetic trees were constructed from *prot* and half of the *RT* nucleotide sequences, a total of 1,146 nt. Except for viruses of subtype B, sequence information outside of the *gag* and *env* regions is extremely

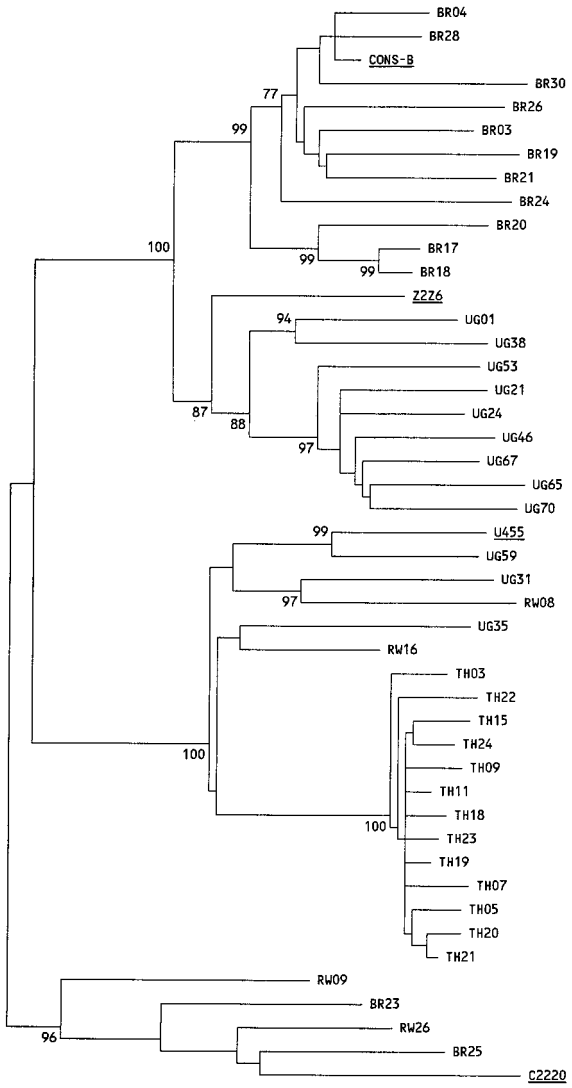

FIG. 2. Results of the phylogenetic analysis of the *pol* gene (1,176 nt) sequences by the neighbor-joining method as part of the MEGA program (24) from 42 virus isolates. Subtype-specific reference *pol* sequences were taken from the data bank: U455 for subtype A, C2220 for subtype C, and Z2Z6 for subtype D. Only for subtype B was a consensus sequence, available as a reference. Values of nodes indicate the percentage of bootstraps in which the cluster to the right was found (1,000 data sets).

limited. Figure 2 shows the result of a phylogenetic analysis including *pol* sequences derived from a subtype A strain (U455), subtype D strain (Z2Z6), consensus B, and the recently sequenced subtype C (C2220) (4). Importantly, this tree depicts the overall topology previously obtained with the *gag* sequences. Four sequence clusters that correspond to HIV-1 subtypes A, B, C, and D can be identified (bootstrap, >87%).

All 13 *env* E samples were very closely related and were placed as an outgroup of subtype A. Phylogenetic analysis of *gag* sequences of these isolates had shown an outgroup position identical to subtype E *env* viruses. Based on that observation, we suggested that subtype E *gag* and now *pol* sequences belong to a distinct variant subtype A cluster, named A′.

Recombination can be detected when different genes, or different regions within the same gene, are placed by phylogenetic analysis into different sequence subtypes or become an outlier within the established subtypes. To investigate possible recombination events for this set of *pol* sequences, neighbor-joining trees based on various parts of *pol* sequences (e.g., the *prot* and *RT* parts) were constructed. The *prot* tree had an overall topology very similar to the topology of Fig. 2, with two exceptions, BR24 and RW09 (data not shown). The *gag* gene of BR24 was recently described as a mosaic of subtype B and F. A crossover site was localized in the P17 part of the *gag* gene. The *RT* gene appeared to be subtype B, but the *prot* gene fell as outlier to subtype B. A more detailed analysis of the BR24 *pol* sequence is not possible due to the limited reference sequences available, especially for subtype F. Isolate RW09 was shown to be a mosaic of subtypes A and C for *gag* (9). The RW09 *prot* sequence (285 nt) clustered with subtype A, while the *RT* part and the total sequence appeared to be subtype C (data not shown). Breakpoints between genomic regions can be identified by a method adapted from the work of Robertson et al. (40). This uses four sequences at a time, i.e., the putative recombinant sequence, the consensus of the two putative subtype parents, and an outgroup. Consensus *pol* sequences for several different subtypes are not available. To investigate RW09, we used BR25 as a parental sequence of subtype C. The other parental sequence, consensus A, was made from the subtype A classified *pol* sequences determined in this study. Of course, the RW09 *pol* sequence was not included in the consensus A sequence. A crossover site was identified ($P < 0.001$ [chi-square test]) between nt 414 and 464, counting the first encoding nucleotide of *prot* as nt 1. The other five *gag* recombinant isolates were examined by the same method and yielded no evidence for additional recombination events in the *pol* sequence of these isolates. All phylogenetic and breakpoint analyses are summarized in Table 1.

Synonymous versus nonsynonymous substitutions. Many molecular evolutionary analyses rely on distinguishing between synonymous (ds) and nonsynonymous (da) differences in DNA sequences. The ds/da ratio provides some indication of evolutionary processes in a particular gene or region. A ds/da ratio close to 1 is expected in a purely neutral model in which there is no selection for or against any changes. A ds/da ratio of more than 1 generally means that the proteins have evolved under negative or purifying selection, because changes are deleterious and therefore eliminated. A ds/da ratio of less than 1 has been proposed to indicate forms of selection that lead to stable polymorphism, and such ratios have been observed in the variable region of gp120 of HIV-1 (16, 29, 43).

Although not entirely full-length genomic sequences, our *pol* sequences, together with their *gag* (9) and *env* (10) sequences, form a unique set of almost entirely genomic sequences (see Materials and Methods for the GenBank accession numbers of these sequences). To measure the selective constraints on the three sequence fragments, we investigated the proportion of synonymous and nonsynonymous substitutions. Chimeric isolates were excluded from this part of the study. The subtype of most included HIV-1 isolates ($n = 20$) was found to be congruent for *gag*, *pol*, and *env*. However, all isolates of subtype E *env* ($n = 13$) cluster with subtype A in *gag* and *pol*; UG35 and UG59 are subtype D in *env* but cluster with subtype A in *gag* and *pol*. (The last two were included for statistical reasons.) Altogether, 3 subtype A, 9 subtype B, 10 subtype D, and 13 subtype E isolates were analyzed for the envelope fragment; 5 subtype A, 9 subtype B, 8 subtype D, and 13 subtype A/E isolates were analyzed for the *gag* and *pol* fragments. (Subtype C sequences were not included, because BR25 was our only full-length subtype C genome.) Nearly all pairwise comparisons between the sequences intra- and inter-phylogenetic groups showed ds/da ratios of more than 1.0 (range, 0.82 to

TABLE 1. V3, *gag* p17/p24, and *prot*/half-*RT* genetic subtyping of HIV-1 from samples collected by the UNAIDS

Country of origin and sample	V3 genotype	*gag* p17 genotype	*gag* p24 genotype	*prot*/half-*RT* genotype
Brazil				
92BR003	B	B	B	B
BR04	B	B	B	B
BR17	B	B	B	B
BR18	B	B	B	B
BR19	B	B	B	B
BR20	B	B	B	B
BR21	B	B	B	B
BR23	**B**	**C**	**C**	**C**
BR24	**B**	**B**	**F**	**?/B**
BR25	C	C	C	C
BR26	B	B	B	B
BR28	B	B	B	B
BR30	B	B	B	B
Thailand				
***n* = 13**	**E**	**A**	**A**	**A**
Rwanda				
92RW008	A	A	A	A
RW09	**A**	**C**	**C**	**A/C**
RW16	A	A	A	A
RW26	**A**	**A**	**C**	**C**
Uganda				
92UG001	D	D	D	D
UG21	**D**	**D/A**	**A**	**D**
UG24	D	D	D	D
UG31	A	A	A	A
UG35	**D**	**A**	**A**	**A**
UG38	D	D	D	D
UG46	D	D	D	D
UG53	D	D	D	D
UG59	**D**	**A**	**A**	**A**
UG65	D	D	D	D
UG67	D	D	D	D
UG70	D	D	D	D

11.26) (Table 2). The da values of the V3 region were significantly higher, both intra- and inter-phylogenetic groups, than were those of *gag* and *pol*, whereas the ds values were similar for all intra-group comparisons and, although higher, in the same order for all inter-group comparisons.

Only the intragroup ds and da values of *env* and *gag* of subtype E were significantly lower than for the other phylogenetic groups; nevertheless, the average ds/da ratios of *env* and *gag* are comparable to those in the other phylogenetic groups. The substitutions in the *pol* region of subtype E viruses were mainly synonymous, resulting in a very high average ds/da ratio (ds/da = 11.26), suggesting that this protein fragment evolved under purifying selection. Between subtypes A/E and B/D, we noted significantly lower ds values in the *gag* and *pol* fragments but not in the V3 region (Table 2). The *env* subtype E group clustered with *pol* and *gag* sequences (7) as an outgroup of subtype A. Therefore, the intergroup ds comparisons of subtype A and E indicate a similar pattern of evolution from a common ancestor. If this is the case, B and D also have a common ancestor, different from the A/E ancestor. To test this hypothesis, we performed separate phylogenetic analysis for *gag*, *pol*, and *env* sequences, based on the numbers of synonymous and nonsynonymous substitutions. The phylogenetic trees based on synonymous changes in the *pol* and *gag* sequences (Fig. 3A and C) were very similar to those for all sequence substitutions (Fig. 2). In particular, the high bootstrap value (bootstrap value, 1,000) between the phylogenetic positions of A/E and B/D sequences was very striking. Phylogenetic trees based on nonsynonymous changes were less convincing for *pol* sequences (Fig. 3B) but were also very similar for the *gag* sequences (Fig. 3D). However, the overall morphology of the *pol* tree, whether based on silent or amino acid replacement, was comparable to the tree based on all nucleotide substitutions. Phylogenetic trees obtained for synonymous substitutions for *env* (Fig. 3E) again showed a clustering of subtype B and D sequences. Although the subtype B and subtype D sequences were placed in separate clusters, the bootstrap value showed a borderline significance for the B cluster (bootstrap value, 82) and, due to the interspersed position of UG65, no significant bootstrap value for the D cluster. The B/D clustering was not observed in the tree based on nonsynonymous substitution (Fig. 3F) or in the tree obtained for all sequences (reference 10 and data not shown), indicating a proportionate selection-driven evolution from a common ancestor for subtypes B and D.

Therefore, sequences derived from subtypes B and D showed the same evolutionary relation in the *gag*, *pol*, and *env* sequences as did the *gag* and *pol* sequences of subtype A and E, but in all four subtypes, *env* sequences showed independent evolution.

TABLE 2. Proportion of synonymous and nonsynonymous substitutions[a]

Subgroups	*env*			*gag*			*pol*		
	ds	da	ds/da	ds	da	ds/da	ds	da	ds/da
Intragroup									
A/A	0.1778	0.1234	1.44	0.1216	0.0535	2.26	0.1554	0.0291	5.48
B/B	0.1152	0.1133	1.06	0.1183	0.0429	2.91	0.1197	0.0201	6.38
D/D	0.1333	0.1693	0.82	0.1222	0.0412	3.22	0.1136	0.0238	5.10
E/E	0.0309	0.0291	1.08	0.0232	0.0129	2.13	0.0366	0.0046	11.26
Intergroup									
A/B	0.3352	0.1795	1.91	0.3115	0.0864	3.62	0.3085	0.0416	7.53
A/D	0.3098	0.2514	1.25	0.3043	0.0896	3.41	0.3128	0.0441	7.20
A/E	0.2793	0.1890	1.49	**0.1893**	**0.0614**	3.15	**0.1820**	**0.0276**	6.73
B/D	0.2427	0.2425	1.04	**0.2105**	**0.0564**	3.80	**0.2124**	**0.0320**	6.72
B/E	0.2507	0.1939	1.30	0.3441	0.0778	4.47	0.3225	0.0373	8.76
D/E	0.2670	0.2283	1.19	0.3259	0.0842	3.89	0.3270	0.0397	8.29

[a] Boldface type indicates lower values than average.

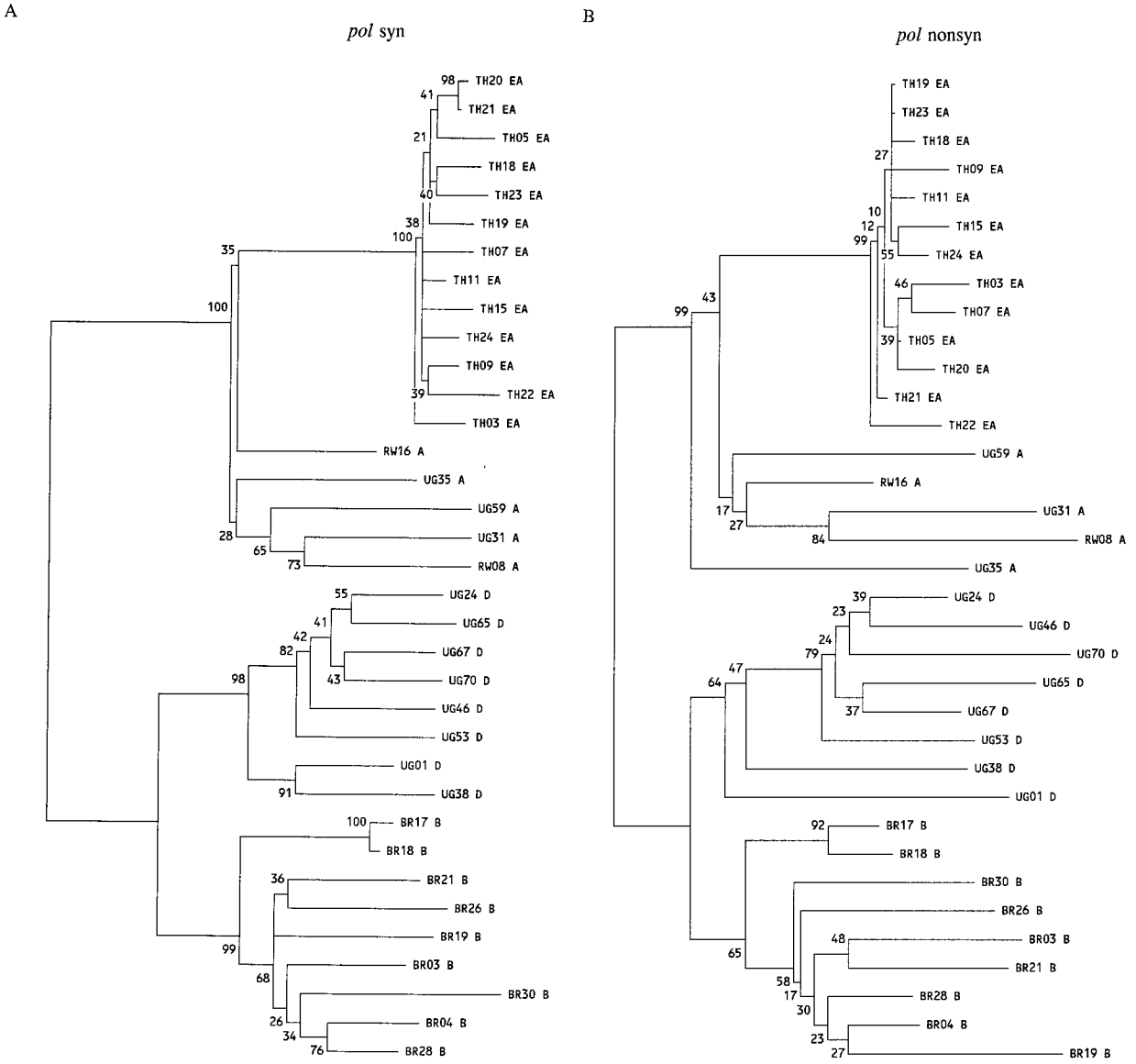

FIG. 3. Phylogenetic analysis for *pol* (A and B), *gag* (C and D), and *env* (E and F) sequences based on numbers of synonymous and nonsynonymous substitutions, respectively. The trees were constructed by the neighbor-joining method as part of the MEGA program (24). Bootstrap values are indicated (1,000 data sets) at the root of the cluster.

DISCUSSION

Mutations contributing to drug resistance. In this study, we determined the natural polymorphism in non-B HIV isolates of the *prot* and the first 297 aa of the *RT* gene. Several of these polymorphisms in subtype B isolates are associated with various Prot and RT inhibitors. Although most do not individually confer significant levels of resistance, these polymorphism can modify the effects of drug-selected mutations, possibly by compensatory changes in the enzyme backbone. The prevalence of these polymorphisms in the HIV-1 isolates infecting an individual suggests that the outcome of drug treatment could be adversely affected in these patients. The level of genetic diversity from untreated patient isolates has been previously studied (1, 26, 33, 35), but in contrast to these reports, all our data were obtained by direct sequencing of PCR products, not by sequencing of individual clones. A sequence derived by a direct analysis is a composition of the nucleotides most frequently observed at each position. It excludes random mutations introduced by the polymerase during the amplification step; therefore, more mutations may exist among our analyzed isolates, as a minority of the population, which were not observed by the direct sequence analysis. The work by Kozal et al. (23) is also a population base sequencing, although these authors analyzed some individual clones derived from two patient samples exhibiting multiple polymorphisms in the *prot* genes to confirm that the observed mutations were linked, since the effect of drug sensitivity is dependent upon multiple polymorphisms occurring in the same genome. Experiments are in progress to study the phenotypic effect of our observed mutations in the non-clade B isolates. However, the main objective of our study was to analyze intrasubtype recombination events in the *pol* gene and to identify possible crossover sites. The high quality of direct sequences makes this possible, but to extend the prevalence of the resistant substitutions in primary

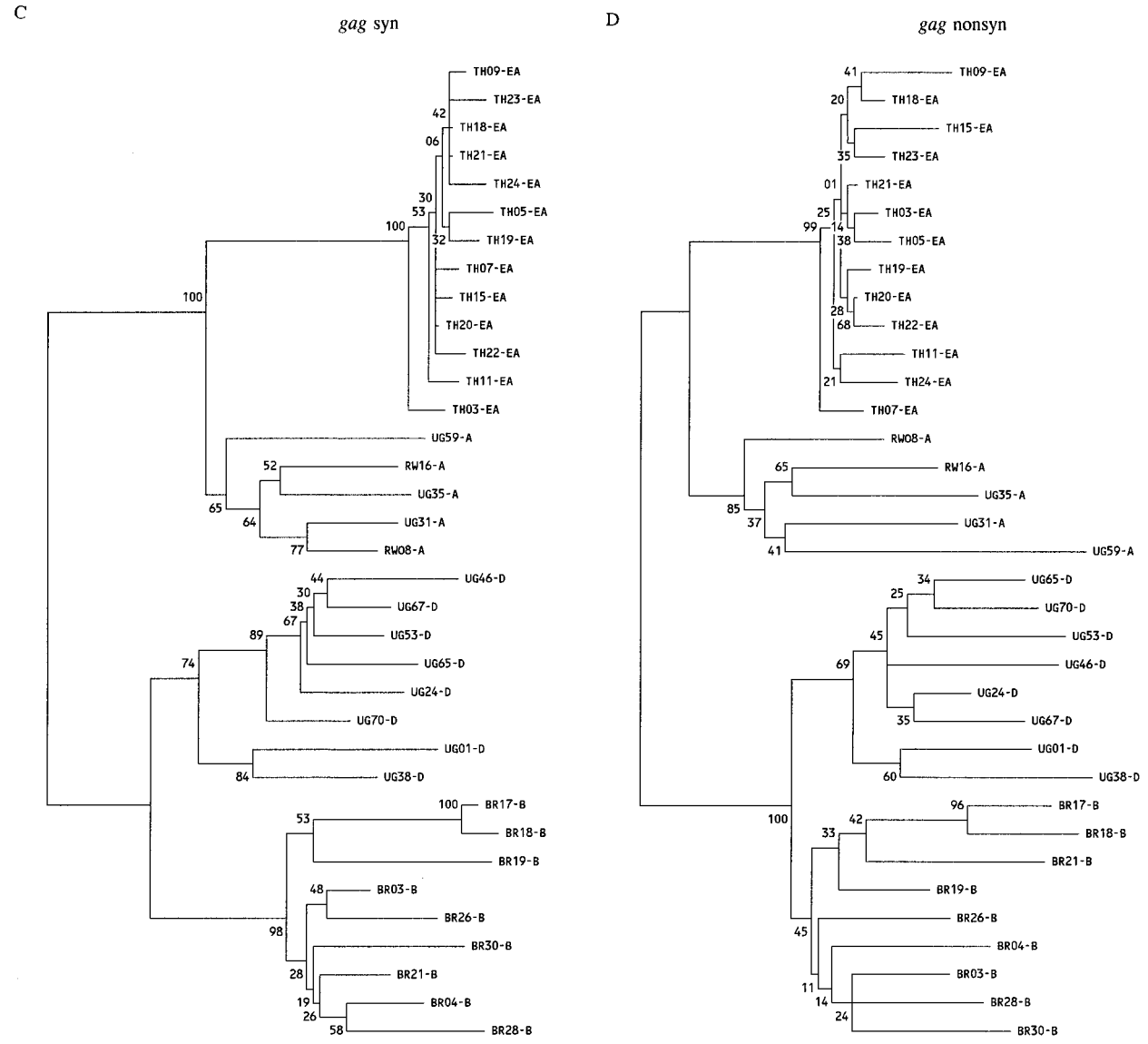

FIG. 3—*Continued.*

isolates, the PCR products must be cloned and sequenced. In addition, plasma can be used to study the in vivo quasispecies, without culture artifacts.

Our data indicate more variability in the *prot* gene than in the 5′ half of *RT*, suggesting that the Prot enzyme is sufficiently flexible to tolerate substitutions in critical residues. Finally, our study showed that populations of virus variants with mutations contributing to drug resistance exist in untreated patients infected with isolates other than subtype B isolates. Since these isolates were obtained in 1992, before the ready availability of Prot inhibitors, the subjects are unlikely to have received these agents surreptitiously; however, we cannot exclude that they acquired AZT-resistant HIV-1 strains from a contact who received AZT. Moreover, untreated individuals have partially AZT-resistant viruses with mutations at codon 70 but not viruses with mutations at codons 41 and 215 (33, 35). This suggests that the K70R mutation is most easily created (by an A-to-G transition), whereas the M41L and T215F/Y must have reduced fitness in an environment without AZT. In this study, Br19 showed a master sequence which contained mutations conferring AZT resistance at codon 67 and 219 and also a mutation at codon 215 (Asp) for which the AZT resistance-conferring properties are unknown. Recently, a patient with a primary infection with a strain having a AZT resistance mutation at codon 215 was monitored during infection (>20 months) in the absence of AZT treatment (14). Sequence analysis at 20 months showed a Ser and an Asp at codon 215, instead of the initial Tyr. Of the three virus subpopulations in this patient, with Tyr, Asp, and Ser at the 215 position, the relative fitness of the Asp- variant was calculated to be 10 to 25% higher than that of the initial Tyr variant (13a). Therefore, it is very likely that BR19 is infected by an AZT-resistant virus population and that Asp at codon 215 is due to selective disadvantage, or fitness loss, of an AZT-resistant virus population. In the other case, UG35, an Ala was found at codon 215. Again, the AZT resistance-conferring properties are unknown, but since no additional AZT-resistant mutations were observed, this critical substitution could result from the occur-

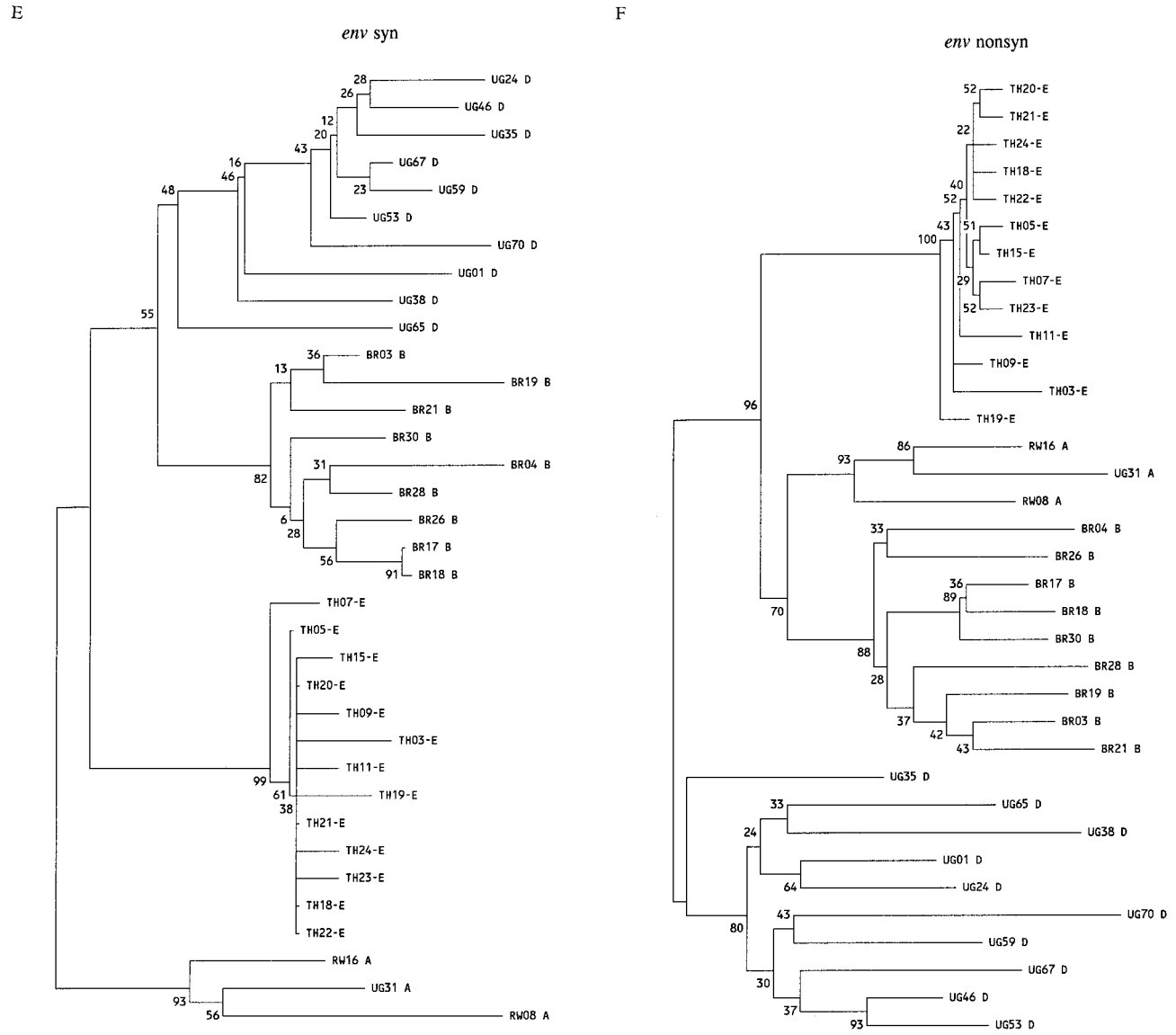

FIG. 3—*Continued.*

rence of random mutations in the absence of selection by drug treatment.

Recombination. Recombinant HIV-1 isolates were initially identified when phylogenetic trees were constructed from sequential segments of available HIV-1 sequences (41). It became apparent that 10 to 20% of the sequences of sufficient length for analysis contained interspersed segments of genetic material from two different genotypes. In this study, we used newly derived *pol* sequences to analyze possible intersubtype recombination events. Phylogenetic analysis showed that the 42 HIV-1 isolates clustered into four groups. Based on the included reference sequences, the four groups could be classified as subtypes A, B, C, and D. Within this set, 8 isolates were chimeras and 13 had a subtype E *env*. These were analyzed in more detail by a variety of phylogenetic analyses and by a modified method of Robertson et al. (40). In only two isolates, BR24 and RW09, did we find evidence for a recombination event in the *pol* gene. In all other isolates, the Gag and Pol sequences exhibit identical subtype classification, with one exception. The exception, UG21, must have a crossover site between $p24^{gag}$ and the start codon of protease. Interestingly, the discordant subtype classification observed between the *env* and *gag* sequences was also seen for the *env* and *pol* sequences (Table 1). This suggests that selection is based on particularly beneficial combinations of *gag* and *pol* belonging to the same subtype. More full-length genome sequence analyses must be done to determine the value of this conclusion. Recently, it was reported that extensive regions of *pol* are required for efficient HIV-1 polyprotein processing and particle maturation (38). Additionally, cotransfection of a deletion mutant with wild-type provirus led to a marked reduction in the virus titer of infectious virus, suggesting that a chimeric *gag-pol* precursor interferes with viral fitness. Taking these results together, we observed no additional chimeric viruses by sequencing the *pol* gene, confirming the observation that analysis of *gag* and *env* sequences is enough to exclude most hybrid viruses (41). The majority contain at least one crossover within these regions, so that these sequences fall into different subtypes.

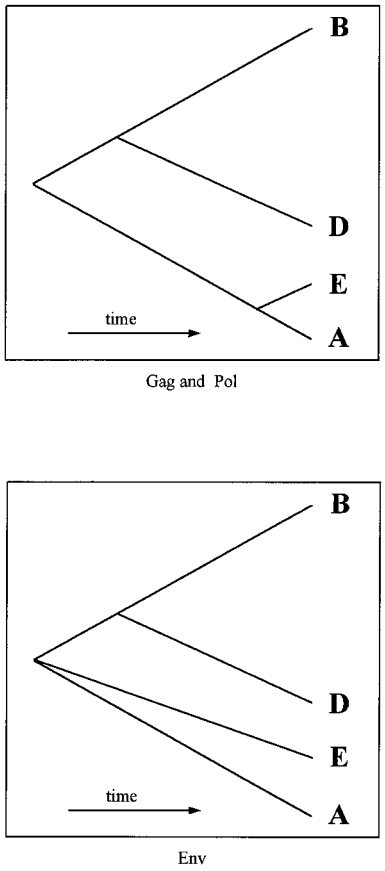

FIG. 4. Model which describes the evolutionary relationship between subtypes B and D and between subtypes A and E in three different gene fragments. This model is based primarily on the results as shown in Fig. 3, the synonymous and nonsynonymous tree analysis. The B/D group branched off before the A/E recombination took place.

Evolution. The present study allows a comparison of the evolution of various HIV-1 subtypes in unrelated asymptomatic patients. We therefore estimated the rates of synonymous and nonsynonymous substitutions per site in the *gag*, *pol*, and *env* genes of 35 isolates collected by UNAIDS. Our analysis indicates that *gag* and *pol* are subject to purifying selection with an average ds/da ratio above 1, independent of the subclassification, and that the ds/da ratios of *pol* fragments were higher than those of *gag* fragments. Given the relatively low natural variation in the *RT* gene, this suggests that *RT* genes are relatively conserved evolutionarily. In the case of V3, ds/da ratios of about 1 were observed, but a positive Darwinian selection (ds/da <1) favoring diversity at the amino acid level has been described (16, 29, 43). This is consistent with the observation that for V3, some nonsynonymous substitutions are related to antibody enhancement, the biological phenotype of the virus, and resistance to neutralization. We compared HIV-1 isolates from unrelated individuals instead of studying the evolution of the virus within one individual, as was done in the above studies. Because of differences in clinical stage and/or the selective environment, our group approach would be statistically more meaningful than individual comparisons. Moreover, selectively favored nonsynonymous substitutions become saturated over time, resulting in a ds/da ratio of about 1.

Based on the low ds/da ratio of the intergroup analysis of A/E and B/D *gag* and *pol* sequences, we analyzed the evolutionary relationship between B and D in more detail, constructing separate phylogenetic trees for synonymous and nonsynonymous substitutions. The evolutionary relation between A and E has been studied extensively (13) by authors who examined whether the discordant phylogenies might be the result of unusual patterns of evolution. Based on the phylogenetic trees for synonymous and nonsynonymous substitutions, they found no evidence for selection-driven convergent evolution. The phylogenetic analysis of the *env* sequences in the present study reached the same conclusion, namely, that the *env* subtype E isolates and subtype A isolates represent true monophyletic groups. The most striking finding from the phylogenetic analysis based on synonymous and nonsynonymous substitutions was the unexpected clustering of subtypes B and D in the synonymous *env* tree. This clustering observed in *env* was also found for the *gag* and *pol* sequences. In the *gag* and *pol* tree, based on all sequences, the B and D cluster showed an identical epidemiological relation as observed for the A/E subtypes. It also became clear that subtype E strains are, in fact, recombinants of a subtype A strain and a still unidentified subtype E strain, with at least eight points of recombinant crossover. The small branches in the phylogenetic analysis and the small mean nucleotide difference (9) suggest a recent recombination event between subtype A and the parental subtype E. This is consistent with epidemiological findings that widespread HIV transmission in Thailand had occurred only since 1988 (37). As yet, there is no clear evidence of intersubtype recombination between B and D, but they have shown a close relationship throughout the genome (20). Leitner et al. (27) discussed an unrooted maximum-likelihood tree, based on $p17^{Gag}$ sequences from the database, in which subtypes B and D were not clearly separated. Subtype D was divided into two sequence clusters that were separated by the subtype B cluster.

In Fig. 4, we outline the evolution of the *gag*/*pol* and *env* gene fragments. The timing of the recombination events that generated the subtype A/E viruses are unknown (13). However, our results contain evidence that the B/D lineage branched off before the A/E recombination events. The ancestor HIV-1 may have become extinct, with its remaining vestige residing in the current A/E recombinant and B and D subtypes. On the basis of all nucleotide substitutions and nonsilent replacements, subtypes B and D are entirely distinguishable in the *env* gene. Only the silent substitutions, which are time related, have shown an epidemiological linkage. This observation can be explained by a simple acceleration of the evolutionary rate in the *env* part, compared with the rest of the genome. The ds/da ratio of the different regions confirms this statement.

ACKNOWLEDGMENTS

We thank Lucy Phillips for editorial assistance and the WHO network on HIV-1 isolation and characterization for providing HIV-1 isolates to the international repository.

REFERENCES

1. **Barrie, K. A., E. E. Perez, S. L. Lamers, W. G. Farmerie, B. M. Dunn, J. W. Sleasman, and M. M. Goodenow.** 1996. Natural variation in HIV-1 protease, Gag p7 and p6, and protease cleavage sites within Gag/Pol polyproteins: amino acid substitutions in the absence of protease inhibitors in mothers and children infected by human immunodeficiency virus type 1. Virology **219:** 407–416.
2. **Boom, R., C. J. A. Sol, M. M. M. Salimans, C. L. Jansen, P. M. E. Wertheim-van Dillen, and J. van der Noordaa.** 1990. Rapid and simple method for purification of nucleic acids. J. Clin. Microbiol. **28:**495–503.
3. **Boucher, C. A. B., E. O'Sullivan, J. W. Mulder, C. Ramautarsing, P. Kellam, G. Darby, J. M. A. Lange, J. Goudsmit, and B. A. Larder.** 1992. Ordered appearance of zidovudine (AZT) resistance mutations during treatment. J. Infect. Dis. **165:**105–110.
4. **Carr, A., and D. A. Cooper.** 1996. HIV protease inhibitors. AIDS **10**(Suppl. A)**:**S151–S157.

5. **Coffin, J. M.** 1979. Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. J. Gen. Virol. **42:**1–26.
6. **Coffin, J. M.** 1995. HIV population dynamics in vivo: implications for genetic variation, pathogenesis and therapy. Science **267:**483–489.
7. **Condra, J. H., D. J. Holder, W. A. Schleif, O. M. Blahy, R. M. Danovich, L. J. Gabryelski, D. J. Graham, D. Laird, J. C. Quintero, A. Rhodes, H. L. Robbins, E. Roth, M. Shivaprakash, T. Yang, J. A. Chodakewitz, P. J. Deutsch, R. Y. Leavitt, F. E. Massari, J. W. Mellors, K. E. Squires, R. T. Steigbigel, H. Teppler, and E. A. Emini.** 1996. Genetic correlates of in vivo viral resistance to indinavir, a human immunodeficiency virus type 1 protease inhibitor. J. Virol. **70:**8270–8276.
8. **Condra, J. H., W. A. Schleif, O. M. Blahy, L. J. Gabryelski, D. J. Graham, J. C. Quintero, A. Rhodes, H. L. Robbins, E. Roth, M. Shivaprakash, D. Titus, T. Yang, H. Teppler, K. E. Squires, P. J. Deutsch, and E. A. Emini.** 1995. In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors. Nature **374:**569–571.
9. **Cornelissen, M., G. Kampinga, F. Zorgdrager, J. Goudsmit and UNAIDS Network for HIV Isolation and Characterization.** 1996. Human immunodeficiency virus type 1 subtypes defined by env show high frequency of recombinant *gag* genes. J. Virol. **70:**8209–8212.
10. **de Wolf, F., E. Hogervorst, J. Goudsmit, E. M. Fenyö, H. Rübensamen-Waigmann, H. Holmes, B. Galvao-Castro, E. Karita, C. Wasi, S. D. K. Sempala, E. Baan, F. Zorgdrager, V. Lukashov, S. Osmanov, C. Kuiken, M. Cornelissen and WHO Network on HIV-1 Isolation and Characterization.** 1994. Syncytium inducing (SI) and non-syncytium (NSI) capacity of human immunodeficiency virus type 1 (HIV-1) subtypes other than B: phenotypic and genotypic characteristics. AIDS Res. Hum. Retroviruses **11:**1387–1400.
11. **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.
12. **Gao, F., S. G. Morrison, D. L. Robertson, C. L. Thornton, S. Craig, G. Karlsson, J. Sodroski, M. Morgado, B. Galvao-Castro, H. Briesen von, S. Beddows, J. Weber, P. M. Sharp, G. M. Shaw, B. H. Hahn, and the WHO and NIAID Network 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.
13. **Gao, F., D. L. Robertson, S. G. Morrison, H. Hui, S. Craig, J. Decker, P. N. Fultz, M. Girard, G. M. Shaw, B. H. Hahn, and P. M. Sharp.** 1996. The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origen. J. Virol. **70:**7013–7029.

13a. **Goudsmit, J.** Unpublished data.

14. **Goudsmit, J., A. de Ronde, D. D. Ho, and A. Perelson.** 1996. Human immunodeficiency virus fitness in vivo: calculations based on a single zidovudine resistance mutation at codon 215 of reverse transcriptase. J. Virol. **70:**5662–5664.
15. **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 **373:**123–126.
16. **Holmes, E. C., L. Q. Zang, P. Simmonds, C. A. Ludlam, and A. J. Leigh Brown.** 1992. Convergent and divergent sequence evolution in the surface envelope glycoprotein of the human immunodeficiency virus type 1 within a single patient. Proc. Natl. Acad. Sci. USA **89:**4835–4839.
17. **Hu, W.-S., and H. M. Temin.** 1990. Retroviral recombination and reverse transcription. Science **250:**1227–1233.
18. **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.
19. **Kampinga, G. A., A. Simonon, P. van de Perre, E. Karita, P. Msellati, and J. Goudsmit.** 1997. Primary infections with HIV-1 of women and their offspring in Rwanda: findings of heterogeneity at seroconversion, coinfection, and recombinants of HIV-1 subtypes A and C. Virology **227:**63–76.
20. **Karlsson, G. B., F. Gao, J. Robinson, B. Hahn, and J. Sodroski.** 1996. Increased envelope spike density and stability are not required for the neutralization resistance of primary human immunodeficiency viruses. J. Virol. **70:**6136–6142.
21. **Kimura, M.** 1980. A simple method for estimating evolutionary rates of base substitution through comparative studies of nucleotide sequences. J. Mol. Evol. **16:**111–120.
22. **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 **11:**1355–1358.
23. **Kozal, M. J., N. Shah, N. Shen, R. Yang, R. Fucini, T. C. Merigan, D. D. Richman, D. Morris, E. Hubbell, M. Chee, and T. R. Gingeras.** 1996. Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays. Nat. Med. **2:**753–759.
24. **Kumar, S., K. Tamura, and M. Wei.** 1993. MEGA: molecular evolutionary genetics analysis, version 1.0. Institute of Molecular Evolutionary Genetics, the Pennsylvania State University, University Park, Pa.
25. **Larder, B. A., D. J. M. Purifoy, K. L. Powell, and G. Darby.** 1987. Site-specific mutagenesis of AIDS virus reverse transcriptase. Nature **327:**716–717.
26. **Lech, W. J., G. Wang, Y. L. Yang, Y. Chee, K. Dorman, D. McCrae, L. C. Lazzeroni, J. W. Erickson, J. S. Sinsheimer, and A. H. Kaplan.** 1996. In vivo sequence diversity of the protease of human immunodeficiency virus type 1: presence of protease inhibitor resistance variants in untreated subjects. J. Virol. **70:**2038–2043.
27. **Leitner, T., D. Escanilla, S. Marquina, J. Wahlberg, C. Brostrom, H. B. Hansson, M. Uhlen, and J. Albert.** 1995. Biological and molecular characterization of subtype D, G, and A/D recombinant HIV-1 transmissions in Sweden. Virology **209:**136–146.
28. **Loeb, D. D., R. Swanstrom, L. Everitt, M. Manchester, S. E. Stamper, and C. A. Hutchison, III.** 1989. Complete mutagenesis of the HIV-1 protease. Nature **340:**397–400.
29. **Lukashov, V. V., C. L. Kuiken, and J. Goudsmit.** 1995. Intrahost human immunodeficiency virus type 1 evolution is related to length of the immunocompetent period. J. Virol. **69:**6911–6916.
30. **Mansky, L. M., and H. M. Temin.** 1996. 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.
31. **McCutchan, F. E., P. A. Hegerich, T. P. Brennan, P. Phanuphak, P. Singharaj, A. Jugsudee, P. W. Berman, A. M. Gary, A. K. Fowler, and D. P. Burke.** 1992. Genetic variants of HIV-1 in Thailand. AIDS Res. Hum. Retroviruses **8:**1887–1895.
32. **Mitsuya, H., K. J. Weinhold, P. A. Furman, M. H. St. Clair, S. Nusinoff-Lehrman, R. C. Gallo, D. P. Bolognesi, D. W. Barry, and S. Broder.** 1985. 3′-Azido-3′-deoxythymidine (BW A509U): an antiviral agent that inhibits the infectivity and cytopathic effect of human T-lymphotropic virus type III/lymphadenopathy-associated virus in vitro. Proc. Natl. Acad. Sci. USA **82:**7096–7100.
33. **Mohri, H., M. K. Singh, W. T. Ching, and D. D. Ho.** 1993. Quantitation of zidovudine-resistant human deficiency virus type 1 in the blood of treated and untreated patients. Proc. Natl. Acad. Sci. USA **90:**25–29.
34. **Molla, A., M. Korneyeva, Q. Gao, S. Vasavanonda, P. J. Schipper, H.-M. Mo, M. Markowitz, T. Chernyavskiy, P. Niu, N. Lyons, A. Hsu, R. Granneman, D. D. Ho, C. A. B. Boucher, J. M. Leonard, D. W. Norbeck, and D. J. Kempf.** 1996. Ordered accumulation of mutations in HIV protease confers resistance to ritonavir. Nat. Med. **2:**760–766.
35. **Najera, I., A. Holguin, M. E. Quinones-Mateu, M. A. Munoz-Ferandez, R. Najera, C. Lopez-Galindez, and E. Domingo.** 1995. *pol* gene quasispecies of human immunodeficiency virus: mutations associated with drug resistance in virus from patients undergoing no drug therapy. J. Virol. **69:**23–31.
36. **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. 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.
37. **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 introduction of two major HIV-1 genotypes into distinct high-risk populations in Thailand. Lancet **341:**1171–1174.
38. **Quillent, C., A. M. Borman, S. Paulous, C. Duaguet, and F. Clavel.** 1996. Extensive regions of pol are required for efficient human immunodeficiency virus polyprotein processing and particle maturation. Virology **219:**29–36.
39. **Richman, D. D.** 1990. HIV and other human retroviruses, p. 581–640. *In* G. J. Galasso, R. J. Whitley, and T. C. Merigan (ed.), Antiviral agents and viral diseases of man. Raven Press, New York, N.Y.
40. **Robertson, D. L., B. H. Hahn, and P. M. Sharp.** 1995. Recombination in AIDS viruses. J. Mol. Evol. **40:**249–259.
41. **Robertson, D. L., P. M. Sharp, F. E. McCutchan, and B. H. Hahn.** 1995. Recombination in HIV-1. Nature **374:**124–126.
42. **Sabino, E. C., E. G. Shpear, M. G. Morgado, B. T. M. Korber, R. S. Diaz, V. Bongertz, S. Cavalcante, B. Galvao-Castro, J. I. Mullins, and A. Mayer.** 1994. Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtype B and F in two epidemiologically linked individuals from Brazil. J. Virol. **68:**6340–6346.
43. **Strunnikova, N., S. C. Ray, R. A. Livingston, E. Rubalcaba, and R. P. Viscidi.** 1995. Convergent evolution within the V3 loop domain of human immunodeficiency virus type 1 in association with disease progression. J. Virol. **69:**7548–7558.
44. **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 human immunodeficiency virus type 1 infection. Nature **373:**117–122.