JOURNAL OF VIROLOGY, Dec. 1995, p. 7548–7558 Vol. 69, No. 12
0022-538X/95/$04.00+0


# Convergent Evolution within the V3 Loop Domain of Human Immunodeficiency Virus Type 1 in Association with Disease Progression

NATALY STRUNNIKOVA,[1] STUART C. RAY,[2] ROBERT A. LIVINGSTON,[1] ELIZABETH RUBALCABA,[1] AND RAPHAEL P. VISCIDI[1]*

*Departments of Pediatrics[1] and Medicine,[2] Johns Hopkins University School of Medicine, Baltimore, Maryland 21287*

Received 18 April 1995/Accepted 25 August 1995

**Phylogenetic analysis was used to study in vivo genetic variation of the V3 region of human immunodeficiency virus type 1 in relation to disease progression in six infants with vertically acquired human immunodeficiency virus type 1 infection. Nucleotide sequences from each infant formed a monophyletic group with similar average branch lengths separating the sets of sequences. In contrast to the star-shaped phylogeny characteristic of interinfant viral evolution, the shape of the phylogeny formed by sequences from the infants who developed AIDS tended to be linear. A computer program, DISTRATE, was written to analyze changes in DNA distance values over time. For the six infants, the rate of divergence from the initial variant was inversely correlated with CD4 cell counts averaged over the first 11 to 15 months of life ($r = -0.87$, $P = 0.024$). To uncover evolutionary relationships that might be dictated by protein structure and function, tree-building methods were applied to inferred amino acid sequences. Trees constructed from the full-length protein fragment (92 amino acids) showed that viruses from each infant formed a monophyletic group. Unexpectedly, V3 loop protein sequences (35 amino acids) that were found at later time points from the two infants who developed AIDS clustered together. Furthermore, these sequences uniquely shared amino acids that have been shown to confer a T-cell line tropic phenotype. The evolutionary pattern suggests that viruses from these infants with AIDS acquired similar and possibly more virulent phenotypes.**

Since the discovery of human immunodeficiency virus type 1 (HIV-1), it has been recognized that the virus is characterized by extensive genetic variation and biological heterogeneity (3, 27, 39, 47). Furthermore, it has been suspected that these properties of HIV-1 are related to viral pathogenesis. Numerous studies have shown that HIV-1 strains that are isolated from individuals with AIDS differ in biological properties from those isolated from asymptomatic persons (6, 11, 34, 46). In addition, strains recovered from an individual early during the course of infection differ from those isolated from the same individual at a later time, when disease has developed. Among the properties of HIV-1 that have been shown to change over time in infected individuals are efficiency of viral replication in vitro, differential tropism for macrophages and T-cell lines, and ability to induce syncytium formation in T-cell lines (13, 20, 26, 30, 37, 40, 48). The genetic determinants of these biological properties of HIV-1 have been localized primarily to the V3 region of the *env* gene. Although HIV-1 strains appear to change over time and to acquire properties associated with virulence, how this process occurs is unknown.

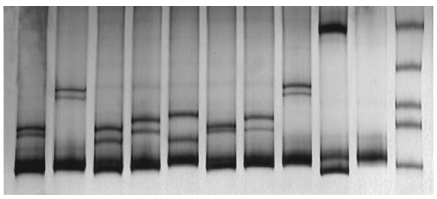
One approach to the study of genetic variation is the use of phylogenetic analysis to reconstruct evolutionary relationships. The methods of molecular phylogeny have been used to assign HIV-1 genomes to six major subtypes (28, 31), to estimate the time of divergence of HIV-1 from a common ancestor and from other lentiviruses (32), and to help resolve questions about the probable source of transmission of HIV-1 (2, 8, 22, 38, 51). Phylogenetic methods have not been applied as extensively to the analysis of sequence data from sequential specimens from HIV-1-infected individuals (21, 44).

We undertook a longitudinal study to characterize the evolution of plasma HIV-1 *env* C2-V3 region cDNA sequences in relation to disease progression in six infants with vertically acquired HIV-1 infection. Longitudinal studies of disease progression in HIV-1-infected infants have two advantages over similar studies of HIV-1-infected adults. Since most infants acquire the infection in utero or during the birthing process, the duration of infection is known with reasonable accuracy and the disease process can be studied from a time close to its inception. Prospective studies are practical because approximately 10 to 25% of HIV-1-infected infants develop AIDS within the first 2 years of life.

One limitation of studies of genetic variation is the high cost of sequencing multiple variants. We used heteroduplex analysis to screen large numbers of clones for those that might contain sequence variants. This approach allowed us to detect rare clones and to determine the frequency of variants while minimizing the amount of sequencing that needed to be done. Having a longitudinal data set allowed us to generate an actual rate of nucleotide substitution. To estimate rates, we designed a computer program, DISTRATE, that calculates and plots DNA distance values against time. This analysis showed that the rate of divergence from the initial variant was inversely correlated with the average CD4 cell count during the first year of life. For nucleotide sequence data, we used phylogenetic methods to determine the genetic lineage of viral variants, and, as expected, found that C2-V3 variants from each infant formed a distinct lineage. We also analyzed inferred amino acid sequences for similarities by using tree-building methods and inspection of multiple sequence alignments. Unexpectedly, analysis of the V3 loop region showed that sequences identi-

* Corresponding author. Mailing address: The Johns Hopkins Hospital, Blalock Room 1111, 600 N. Wolfe St., Baltimore, MD 21287. Phone: (410) 614-1494. Fax: (410) 614-1491. Electronic mail address: rviscidi@welchlink.welch.jhu.edu.

A.




B.

Percent Sequence Dissimilarity Between Clones From Heteroduplex Gel Analysis

| Clone types | A1 | B1 | C1 | D1 | E1 | A2 | D2 | B2 | HIVIIIB |
|---|---|---|---|---|---|---|---|---|---|
| A1 | | 8.2 | 1.8 | 1.4 | 2.9 | 0.4 | 1.4 | 8.2 | 22 |
| B1 | | | 8.6 | 8.2 | 7.2 | 8.2 | 8.2 | 0 | 25 |
| C1 | | | | 2.5 | 2.1 | 1.8 | 2.5 | 8.6 | 21 |
| D1 | | | | | 1.7 | 1.4 | 0 | 8.2 | 19 |
| E1 | | | | | | 2.9 | 1.8 | 7.2 | 20 |
| A2 | | | | | | | 1.4 | 8.2 | 22 |
| D2 | | | | | | | | 8.2 | 19 |
| B2 | | | | | | | | | 25 |
| HIVIIIB | | | | | | | | | |

FIG. 1. Heteroduplex analysis of selected recombinant plasmids containing HIV-1 C2-V3 region gene segment inserts. (A) PCR-amplified DNA fragments of the inserts from individual plasmids were mixed with the homologous PCR-amplified DNA fragment of HIV-$1_{MN}$. The reaction mixture was heated to 95°C for 10 min, incubated at 37°C for 15 min, cooled rapidly on ice, loaded into the well of a 5% neutral Hydrolink polyacrylamide gel, and electrophoresed overnight in a tap water-cooled gel apparatus. Clonotypes were distinguished on the basis of distinct band migration patterns (A to E), and identical clonotypes were enumerated (A1, A2, etc.). IIIB, reaction of the IIIB DNA segment with MN; MN, reaction of MN DNA segment with itself; M, molecular weight marker. (B) DNA inserts were sequenced, and a matrix of all pairwise comparisons based on percent dissimilarity was constructed. Overall, identical clonotypes differed by 0.0 to 2.6% (mean, 1.2%), and in this example, clonotypes differed from each other by 1.4 to 8.6%.

fied at later times from the two infants who developed AIDS clustered monophyletically. Furthermore, these sequences uniquely shared amino acids that have been shown to confer a T-cell line tropic phenotype. One interpretation of this finding is that disease progression is associated with convergent evolution toward a possibly more virulent phenotype.

## MATERIALS AND METHODS

**Study subjects and sample collection.** Blood samples were obtained from HIV-1-infected infants seen at The Johns Hopkins Hospital. After informed consent was obtained from the parents of the infants, blood specimens were collected in acid-citrate-glucose and transported to the laboratory within 6 h of phlebotomy.

**Sample preparation.** Plasma was separated from whole blood by centrifugation at 500 × *g* for 20 min and then for a further 10 min at 2,000 × *g*. To pellet virions, a 50- to 200-μl aliquot of plasma was subjected to ultracentrifugation at 35,000 rpm for 1 h in a Beckman SW 50.1 rotor. Pellets were resuspended in 100 μl of 20 mM Tris-HCl (pH 7.5)–150 mM NaCl–2 mM EDTA containing 1 mg of proteinase K per ml and 0.1% sodium dodecyl sulfate and incubated at 37°C for 1 h. RNA was precipitated with ethanol in the presence of 40 μg of glycogen and 10 ng of 7.5-kb synthetic RNA per ml as a carrier. The precipitate was recovered by ultracentrifugation. The pellets were dried and dissolved in 100 μl of sterile diethylpyrocarbonate-treated water. Five cDNA samples failed to yield an amplification product in the nested PCR. For these blood samples, HIV-1 proviral DNA was amplified from lysates of peripheral blood mononuclear cells. Peripheral blood mononuclear cells were isolated by Ficoll-Hypaque density gradient centrifugation, resuspended ($8 \times 10^6$/ml) in PCR buffer containing 0.5% Tween 20, 0.5% Nonidet P-40, and 100 μg of proteinase K per ml, and incubated overnight at 37°C. An aliquot of cellular material equivalent to $2 \times 10^5$ cells was used in the nested PCR assay.

**Reverse transcription reaction.** For the reverse transcription reaction, a 5-μl aliquot of RNA was heated to 65°C for 10 min and then annealed to 10 pmol of random hexanucleotide primers for 5 min at 65°C. The mixture was adjusted to a final volume of 20 μl of 50 mM Tris-HCl (pH 8.3)–75 mM KCl–3 mM $MgCl_2$–20 mM dithiothreitol containing 0.5 mM deoxynucleoside triphosphates. After addition of 200 U of Moloney murine leukemia virus reverse transcriptase, the reaction was allowed to proceed for 1 h at 37°C.

**Nested PCR amplification of V3 region of *env* gene of HIV-1.** Nested PCR was performed for an approximately 330-bp product, which included the third variable (V3) and flanking regions of the *env* gene, with the following primers (positions correspond to HIV-1 isolate HXB2; accession number K03455): PND-01, 5′-CAGCACAGTACAATGTACACATGGAAT-3′ (bp 6949 to 6975); PND-02, 5′-ATTACAGTAGAAAAATTCCCCTCCAC-3′ (bp 7381 to 7356); PND-03, 5′-TGGCAGTCTAGCAGAAGAAG-3′ (bp 7009 to 7028); and PND-04, 5′-ACAATTTCTGGGTCCCCTCCT-3′ (bp 7338 to 7318). Mixtures for PCR consisted of 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 2.5 mM $MgCl_2$ (PCR buffer) containing each of the four deoxynucleotide triphosphates at 0.2 mM, each primer at 0.25 μM, 2 U of *Taq* polymerase, and a 5-μl aliquot of cDNA in a volume of 50 μl. Cycle conditions were set at 94°C for 15 s, 55°C for 15 s, and 72°C for 30 s. After 30 cycles, the product was extended at 72°C for 10 min. The product of the first PCR was diluted 1:100 in deionized water, and 5 μl of the diluted product was used in the second PCR. Conditions for the second PCR were identical to those for the first PCR.

**Cloning of PCR products.** The PCR products were ligated to Novagen pT7Blue T-vector, which is a thymidine-tailed vector, and transfected into NovaBlue competent cells as recommended by the manufacturer (Novagen, Inc.). The transformed Novablue cells were streaked onto agar indicator plates, and 25 to 45 white colonies were picked, grown overnight in 1-ml liquid broth cultures, and screened for inserts by PCR. For screening, a 5-μl sample of the bacterial culture was amplified through 25 cycles with primers PND-03 and PND-04 and the cycle conditions described above.

**Heteroduplex mobility assay for screening of clones.** Clones were screened for sequence variants by a heteroduplex mobility assay. Heteroduplex analysis is a method for detection of sequence variants based on differences in electrophoretic mobility of hybrids between different sequences (15). Equal amounts of PCR-amplified sequences from the clones and from HIV-$1_{MN}$ were mixed in a final volume of 5 μl containing 0.1 mM EDTA. The mixture was heated to 94°C for 10 min, incubated at 37°C for 15 min, and rapidly cooled on ice. An aliquot of the reaction mixture was loaded into the well of a 5% neutral Hydrolink polyacrylamide gel and electrophoresed overnight in a tap water-cooled gel apparatus. The gel was fixed and stained with silver nitrate (7). Heteroduplexes formed by nonhomologous allelic PCR products were visualized as bands that migrate more slowly on the gel than bands formed by fully complementary homoduplex molecules. The number of clones corresponding to a particular heteroduplex band pattern was recorded, and one clone was arbitrarily selected for sequencing. Sequence analysis of 42 pairs of clones with identical mobility on heteroduplex gels showed that the average nucleotide sequence difference between pairs was 1.2% (range, 0.0 to 2.6%). The agreement between heteroduplex mobility analysis and DNA sequence analysis was accurate enough to allow us to determine the frequency of viral variants by heteroduplex analysis, thereby minimizing the number of clones that had to be sequenced. A representative heteroduplex gel and a comparison of the nucleotide sequences of the variants are shown in Fig. 1.

**Nucleotide sequencing.** Plasmid DNA was isolated from a 30-ml broth culture by the Qiagen Plasmid Midi Kit in accordance with the manufacturer's instructions. DNA templates were sequenced in both directions with Pharmacia Auto-Read Sequencing Kits and a Pharmacia Biotech ALF DNA Sequencer in accordance with the manufacturer's instructions. The sequence data were edited in the SEQMAN program of the DNASTAR software package. Ambiguities were resolved by forming a consensus with sequences generated in the forward and reverse directions. The error for the amplification, cloning, and sequencing procedure was estimated at 1 in 5,580 bases.

**Phylogenetic analysis.** Nucleotide sequences were aligned manually in the ESEE program, version 3.0 (10). Proper alignment required the insertion of a single gap in the 3′ flanking region of all 24 sequences from infant D. This site was stripped from all analyses. A pairwise matrix of evolutionary distances was generated with the DNADIST program (maximum-likelihood option) in the PHYLIP package, version 3.55c (18). The matrix was based on 279 sites, 147 of which varied. A phylogenetic tree was constructed from the distance matrix with the FITCH program on the basis of the topology generated by bootstrap analysis (17). The bootstrap analysis was performed with SEQBOOT (100 resamplings), DNADIST, NEIGHBOR (neighbor joining), and CONSENSE. Phylogenetic analysis was also performed by the maximum-parsimony method with the

TABLE 1. Sampling times and CD4$^+$ cell data for six HIV-1-infected infants

| Infant | Sample no. | Age (days) | Date of birth | No. of CD4$^+$ cells/mm$^3$ | % of CD4$^+$ cells |
|---|---|---|---|---|---|
| A | 1 | 14 | 2/18/93 | 3,929 | 27 |
| | 2 | 65 | | 4,229 | 33 |
| | 3 | 97 | | 2,302 | 17 |
| | 4 | 124 | | 955 | 13 |
| | 5 | 186 | | ND[a] | ND |
| | 6 | 404 | | 1,044 | 15 |
| B | 1 | 22 | 7/5/93 | 3,666 | 47 |
| | 2 | 43 | | 2,918 | 34 |
| | 3 | 66 | | 1,747 | 38 |
| | 4 | 128 | | 1,201 | 36 |
| | 5 | 186 | | 1,126 | 40 |
| C | 1 | 15 | 8/16/92 | 4,040 | 60 |
| | 2 | 162 | | 2,508 | 25 |
| | 3 | 215 | | 1,667 | 25 |
| | 4 | 397 | | 2,672 | 26 |
| | 5 | 463 | | 1,757 | 19 |
| | 6 | 488 | | ND | ND |
| | 7 | 589 | | ND | ND |
| D | 1 | 59 | 1/30/93 | 2,861 | 51 |
| | 2 | 79 | | 1,079 | 16 |
| | 3 | 90 | | 343 | 10 |
| | 4 | 192 | | 49 | 2 |
| | 5 | 212 | | ND | ND |
| | 6 | 248 | | ND | ND |
| | 7 | 305 | | 16 | 0.7 |
| E | 1 | 21 | 10/26/93 | 2,761 | 57 |
| | 2 | 161 | | 2,004 | 31 |
| | 3 | 212 | | 1,305 | 28 |
| | 4 | 240 | | 654 | 17 |
| | 5 | 385 | | 45 | 2 |
| | 6 | 410 | | ND | ND |
| | 7 | 465 | | 6 | 0.4 |
| | 8 | 531 | | ND | ND |
| F | 1 | 2 | 6/13/93 | ND | ND |
| | 2 | 16 | | 860 | 20 |
| | 3 | 25 | | 1,809 | 38 |
| | 4 | 94 | | 2,449 | 25 |
| | 5 | 130 | | 1,317 | 22 |
| | 6 | 196 | | 2,259 | 26 |

[a] ND, not determined.

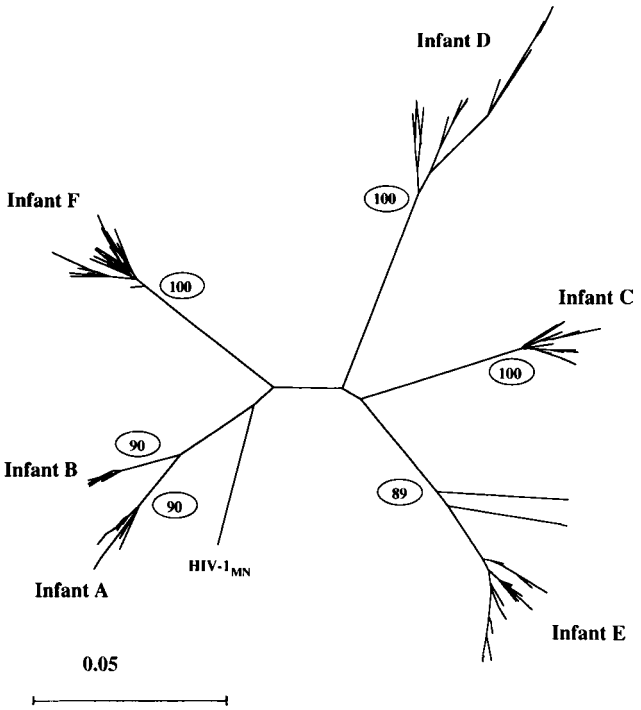



FIG. 2. Phylogenetic tree of 151 HIV-1 *env* C2-V3 nucleotide sequences from six infants with vertically acquired HIV-1 infection sampled serially over a 6- to 15-month period. The tree was generated by the Fitch-Margoliash distance method and drawn with the DRAWTREE program. Branch lengths are drawn to scale (as indicated). The circled numbers at the nodes indicate proportion of support in 100 bootstrap resamplings. The numbers of sequences from infants A, B, C, D, E, and F were 17, 14, 17, 24, 42, and 37, respectively.

DNAPARS program of the PHYLIP package. The trees for each method were rooted with the HIV-$1_{MN}$, HIV-$1_{SF2}$, and HIV-$1_{ELI}$ sequences with equivalent results.

Amino acid sequences were inferred from the set of nucleotide sequence data. Phylogenetic distances were generated from the Dayhoff PAM001 matrix with the PROTDIST program from the PHYLIP package. The matrix was based on 92 sites, 67 of which varied. A phylogenetic tree was constructed from these distances with the FITCH program on the basis of the topology generated by bootstrap analysis. The bootstrap analysis was performed with SEQBOOT (100 resamplings), PROTDIST, NEIGHBOR (neighbor joining), and CONSENSE. Phylogenetic analyses were also performed by the maximum-parsimony method with the PHYLIP program PROTPARS. The trees for each method were rooted with the HIV-$1_{MN}$, HIV-$1_{SF2}$, and HIV-$1_{ELI}$ sequences with equivalent results.

Analyses of subdomains of the C2-V3 region, the V3 loop, and flanking segments were based on the predominant sequence from each time point. The predominant sequence was defined as the majority sequence. When there was no majority sequence, the sequence that matched the consensus sequence for that time point was used. Consensus sequences were based on the most common amino acid in a given position. The frequency of viral variants as determined by heteroduplex analysis was used to determine the majority sequence and to select the most common amino acids. Phylogenetic distances were generated from the Dayhoff PAM001 matrix with the PROTDIST program from the PHYLIP package. A phylogenetic tree was constructed from these distances with the FITCH program. Bootstrap analysis was performed with the SEQBOOT (500 resamplings), PROTDIST (PAM001 matrix), FITCH, and CONSENSE programs. For analysis of the V3 loop, distances were also generated with the PROTDIST program by using the Kimura and categories options, and phylogenetic analyses were also performed by the neighbor-joining and maximum-parsimony methods with the PHYLIP programs NEIGHBOR and PROTPARS, respectively. Amino acid sequences were also analyzed by the adaptation of the PIMA program described by Korber et al. (24). Gap penalty values were not relevant, as our data contained no length polymorphisms in the V3 loop. The PIMA-generated similarity scores were put into a UPGMA clustering program to generate a phenogram. The trees for each method were rooted with the HIV-$1_{MN}$, HIV-$1_{SF2}$, and HIV-$1_{ELI}$ sequences with equivalent results.

**DISTRATE computer program.** The DNA distance matrix and information regarding dates of birth and sample collection and the number of times each variant was detected were used as input to DISTRATE, a computer program which calculates and plots weighted means against time. To calculate weighted means, the distance between each sequence in the first successive sample and each sequence in samples was multiplied by the proportion of each sequence in its respective sample. The proportion was based on the number of clones with the same heteroduplex band pattern as a representative clone, which was selected arbitrarily for sequencing. The weighted mean DNA distance was generated by summing the products.

**Synonymous and nonsynonymous substitution frequencies.** Matrices of synonymous and nonsynonymous frequencies were generated by using the method of Nei and Gojobori (35) and used as input for the DISTRATE computer program described above. Rates were determined from regression lines fitted to the mean value at each time point. The ratio of synonymous to nonsynonymous substitutions was calculated from the entire set of sequences from each patient. For each infant, all pairwise comparisons were made by the method of Nei and Gojobori and then averaged. The values were used to obtain the ratio of synonymous to nonsynonymous substitution frequencies.

**Nucleotide sequence accession numbers.** The sequences reported here have been deposited in the GenBank database under accession numbers U22682 to U22810 and U22834 to U22835.

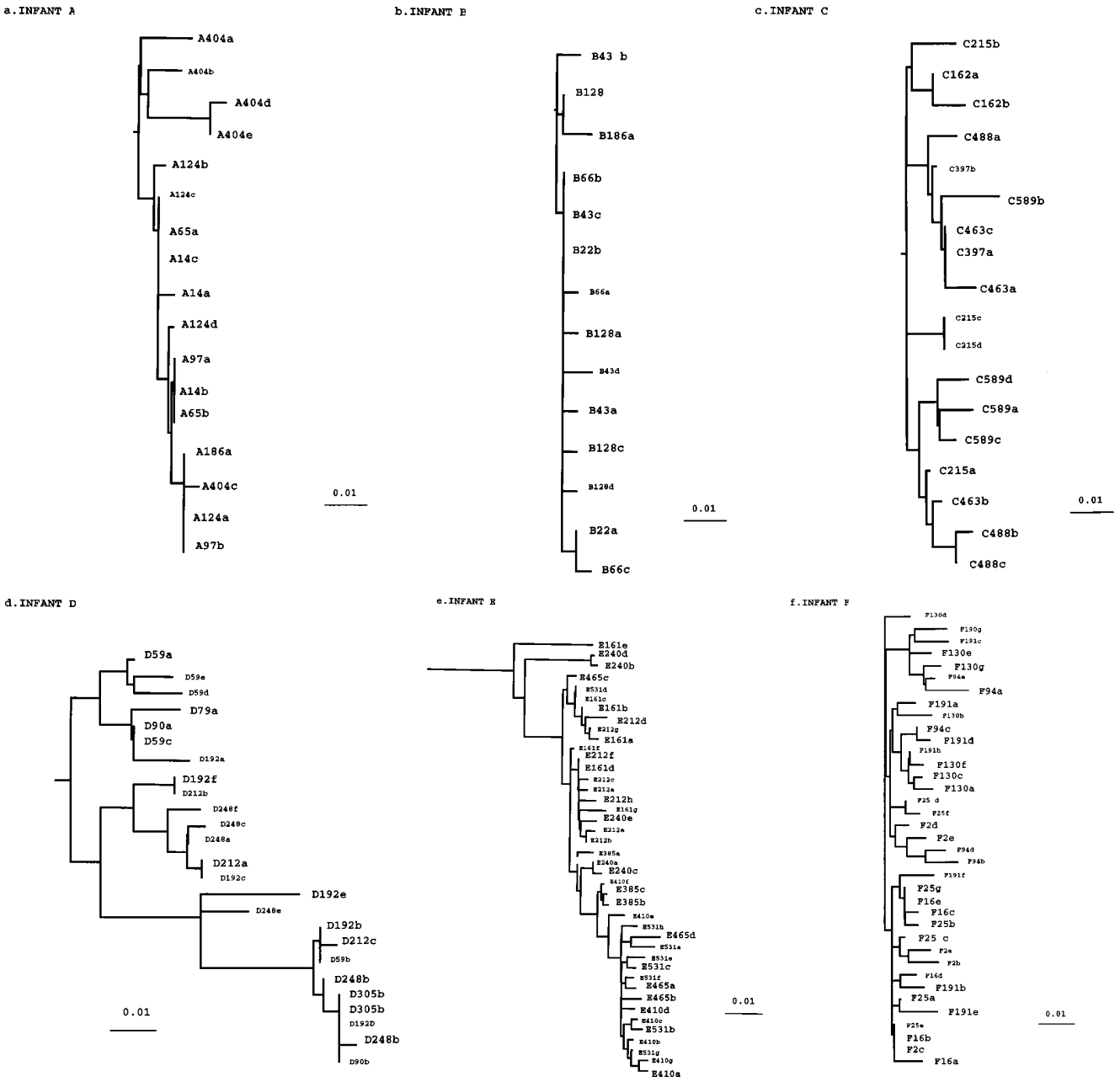



FIG. 3. Phylogenetic trees of HIV-1 *env* C2-V3 nucleotide sequences from six infants. The trees were generated by the Fitch-Margoliash distance method and drawn with the DRAWGRAM program. Branch lengths are drawn to scale (as indicated). Taxa are identified with a letter for the infant, followed by a number for days of life at time of sample collection and a letter for unique sequences from the same sample. A smaller font size was used for taxa that constituted less than 15% of the viral variants in a sample.

## RESULTS

**Characteristics of the study population.** The studies were done with blood samples from six unrelated, HIV-1-infected infants (Table 1). The six infants were born in Baltimore, Md., between August 1992 and July 1993. Two of the infants (D and E) developed AIDS. Infant D was diagnosed with encephalopathy at 11 months of age, and infant E developed *Candida* esophagitis and ecthyema gangrenosum at 16 months of age. Both of these infants had fewer than 50 $CD4^+$ cells per μl by 1 year of age. On the basis of the decline in their $CD4^+$ cell counts, infants A and B had progressed to a more advanced stage of disease than infants C and F. All of the infants received antiviral drug therapy. Peripheral blood samples were obtained at various intervals over a sampling period ranging from 10 to 19 months for five infants and a period of 6 months for infant B. The first sample was obtained from four infants within 2 months of birth and from two infants within 6 months of birth. In total, 37 blood samples were obtained from the six infants, with five to seven samples per infant.

**Nucleotide sequence analysis.** One hundred thirty-one unique viral nucleotide sequences were identified from 37 blood samples. When sequences detected on more than one visit from the same infant were counted, the number of se-

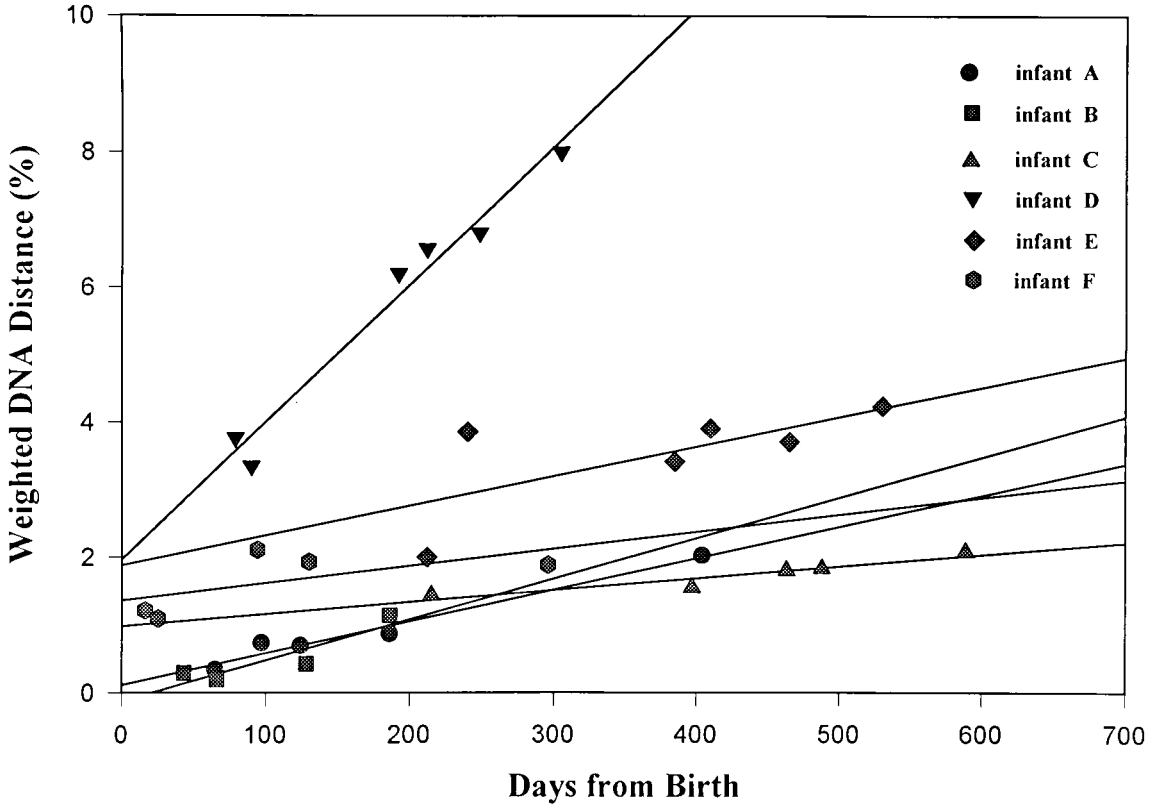



FIG. 4. Weighted mean DNA distance from the first date of collection versus time for six infants with vertically acquired HIV-1 infection. A distance matrix was calculated from the sequence data by using the maximum-likelihood option of the DNADIST program in the PHYLIP package. This matrix and information regarding the dates of collection were then used as input to DISTRATE, a computer program which calculates and plots weighted means against time. Lines were generated by least-squares linear regression.

quences used for analysis was 151. To determine the evolutionary relationship among these viral variants, the nucleotide sequences were subjected to phylogenetic analyses. The phylogenetic tree showed that sequences from each infant formed a monophyletic group which was strongly supported by bootstrap analysis (Fig. 2). The pattern of interinfant viral evolution is consistent with a star phylogeny, as the average nucleotide branch lengths separating the sets of sequences from each infant were similar, with the exception of infants A and B, whose sequences were more closely related to each other than to those of the other infants. The same relationship among the infants' sequences was also found by phylogenetic analysis by using a parsimony method (data not shown). The branches denoting sequences from the two infants who developed AIDS extended further and thus diverged more from the putative ancestral node than those from the asymptomatic infants. The average branch lengths for the infants with AIDS, 0.049 (range, 0.0 to 0.085) and 0.033 (0.0 to 0.096), were greater than those for the asymptomatic infants, 0.013 (0.0 to 0.034), 0.008 (0.0 to 0.018), 0.019 (0.0 to 0.035), and 0.019 (0.0 to 0.042), indicating that viruses from the infants who developed AIDS are genetically more diverse. In contrast to the pattern of multiple distinct lineages for each infant, viruses from the infants who developed AIDS (D and E), but not viruses from asymptomatic infants (A to C and F), tended to form a pattern of sequential dominant lineages with sequences from the earliest time points closest to the ancestral node (Fig. 3).

As another approach to the investigation of evolution of viral variants, we wrote a computer program, DISTRATE, that calculates and plots the weighted DNA distance between sequences from the first sample and each subsequent sample versus time (Fig. 4). The DNA distance values were weighted according to the proportional representation of each sequence variant as determined by heteroduplex mobility assay of multiple clones. Weighting allowed us to include a measure of the frequency of genetic variants, as well as nucleotide sequence composition, in the determination of genetic diversity. For each infant, the slope of the regression line fitted to the DNA distance values was positive, indicating that sequences diverged from the initial variants over time. The rates of divergence for sequences from infants A, B, C, D, E, and F were 1.7, 2.1, 0.6, 7.4, 1.6, and 0.9% per year, respectively. These values correlated with disease progression as measured by serial $CD4^+$ cell counts. The $CD4^+$ cell counts obtained during the study period were plotted against time, and a time-weighted average $CD4^+$ cell count was determined from the area under the curve divided by the time interval (4). The average $CD4^+$ cell counts during the study period were 1,614, 1,717, 2,528, 288, 1,106, and 1,953 cells per $mm^3$ for infants A, B, C, D, E, and F, respectively. The rate of sequence divergence was inversely correlated with the average $CD4^+$ cell count (Pearson's correlation coefficient, $r = -0.87$, $P = 0.024$). Extrapolation of the regression line to the $y$ axis yielded a $y$ intercept value which can be interpreted as a measure of the genetic diversity of the initial virus population. The inferred initial diversity of viruses from infants who developed AIDS (1.9 and 2.0%) was greater than that of viruses from asymptomatic infants (1.1,

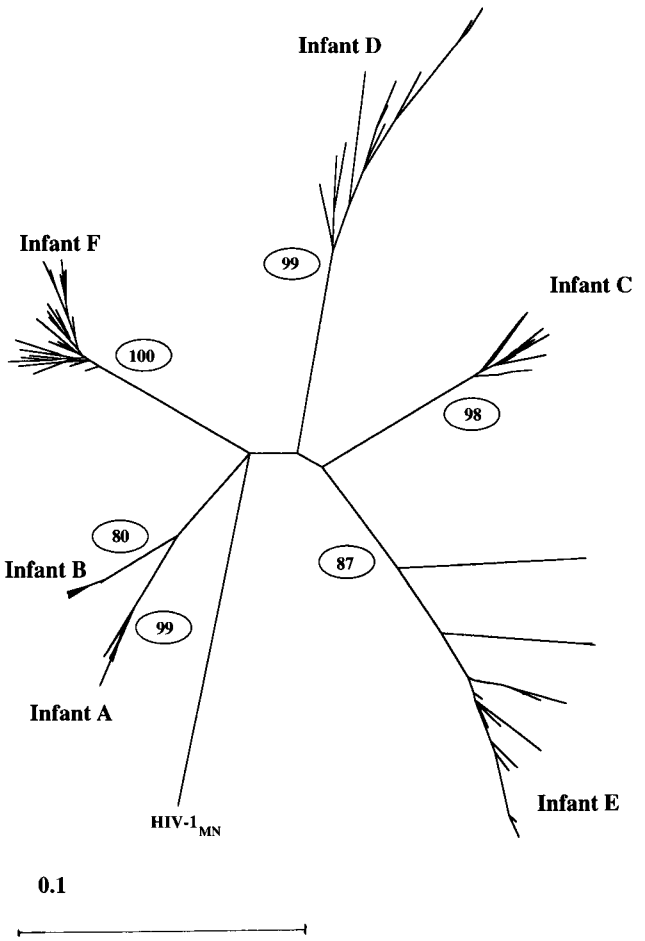



FIG. 5. Phylogenetic tree of 91 HIV-1 *env* C2-V3 amino acid sequences from six infants with vertically acquired HIV-1 infection sampled serially over a 6- to 15-month period. Amino acid sequences were inferred from 131 unique nucleotide sequences. The number of taxa was reduced by eliminating identical sequences obtained from the same infant. The tree was generated by the Fitch-Margoliash method and drawn with the DRAWTREE program. Branch lengths are drawn to scale (as indicated). The circled numbers at the nodes indicate proportion of support in 100 bootstrap resamplings. The numbers of sequences from infants A, B, C, D, E, and F were 8, 8, 15, 16, 23, and 30, respectively.

1.0, 0.9, and 0.0%); however, the difference did not reach statistical significance (*t*-test $P = 0.08$).

We also calculated actual synonymous and nonsynonymous nucleotide substitution rates from similar plots. The rates of nonsynonymous nucleotide substitution were higher for the infants who developed AIDS ($6.5 \times 10^{-2}$ and $2.0 \times 10^{-2}$ substitutions per site per year) than for the asymptomatic infants ($1.4 \times 10^{-2}$, $1.2 \times 10^{-2}$, $0.8 \times 10^{-2}$, and $1.0 \times 10^{-2}$ substitutions per site per year) (*t*-test $P = 0.03$). The synonymous substitution rate could not be determined in all cases by this method because of low regression coefficients (<0.1). Therefore, the ratio of synonymous to nonsynonymous substitutions was calculated by a more conventional method in which all of the sequences from each patient were compared (35). In this type of analysis, the temporal relationship among the sequences is not considered. The mean ratio of synonymous to nonsynonymous substitutions for the six infants was 1.4 (range, 0.6 to 2.3). This ratio is consistent with previous findings for the *env* C2-V3 region and suggests that the *env* gene is under selective pressure (43, 50).

**Inferred amino acid sequence analysis.** To uncover evolutionary relationships among viral variants that might be dictated by protein structure and function, we applied tree-building methods to the analysis of inferred amino acid sequences. After removal of identical sequences obtained from the same infant, there were 91 unique amino acid sequences. Trees constructed from the full-length protein fragment (92 amino acids after stripping of a gap) by using distance and parsimony methods revealed phylogenetic relationships similar to those found by analysis of nucleotide sequence data (Fig. 5).

Because it is a biologically important domain, we also analyzed the V3 loop region (35 amino acids) separately (16, 29, 41, 42). Unexpectedly, the tree generated by the Fitch-Margoliash method showed that V3 loop amino acid sequences detected at later time points from the infants who developed AIDS formed a monophyletic group (Fig. 6). In contrast, sequences obtained at earlier time points from these subjects were distantly related. The monophyletic grouping of late sequences from the infants who developed AIDS recurred in 453 of 500 replicate trees by bootstrap analysis and was seen in trees constructed by other distance methods and a parsimony method (data not shown). Analysis of the 3′ flanking sequences (31 amino acids) and the combined flanking sequences (55 amino acids) did not show significant interinfant clustering of sequences, except for infants A and B, whose full-length amino acid sequences were also significantly related (data not shown).

The C2-V3 region sequences were also analyzed for amino acid similarities by inspection of multiple sequence alignments (Fig. 7). The coding potential of the envelope open reading frame was maintained in all sequences. The two cysteines which define the V3 loop region were preserved in 134 of 136 sequences. The N-linked glycosylation site proximal to the 5′ cysteine of the V3 loop was absent from all sequences obtained in the first 6 months of life from the four asymptomatic infants. This finding is in agreement with the results of Ahmad et al. and Wolinsky et al. (1, 51). For the two asymptomatic infants (A and C) who had blood samples obtained after 6 months of life, the N-linked glycosylation site was present in all or most sequences. In contrast to the asymptomatic infants, the glycosylation site was present in all sequences from the two infants who developed AIDS, including sequences obtained from infant D at 2 months of life.

For each infant, consensus V3 loop sequences were formed from the set of sequences obtained during the first (early) half and second (late) half of the study period (Fig. 8). The early and late amino acid consensus sequences of the infants who developed AIDS differed by six or seven amino acid replacements, while those of the asymptomatic infants differed by zero to one amino acid replacement. The greater number of amino acid replacements in sequences from the infants who developed AIDS is consistent with evidence from the phylogenetic analysis for more rapid viral evolution in these infants. The late consensus sequences of the infants who developed AIDS uniquely shared three amino acids: an arginine in position 11, a valine in position 19, and a lysine in position 32. They also shared a threonine in position 13 that was not present in the consensus sequences of the asymptomatic infants or in the early consensus sequence of infant D. These amino acids have previously been shown by site-directed mutagenesis to be critical for conferring T-cell line tropism on a macrophage tropic strain (12, 25, 49). The proportion of variants containing this motif increased over time, and they constituted 100% of the variants in the last sampling from both infants (Fig. 7).

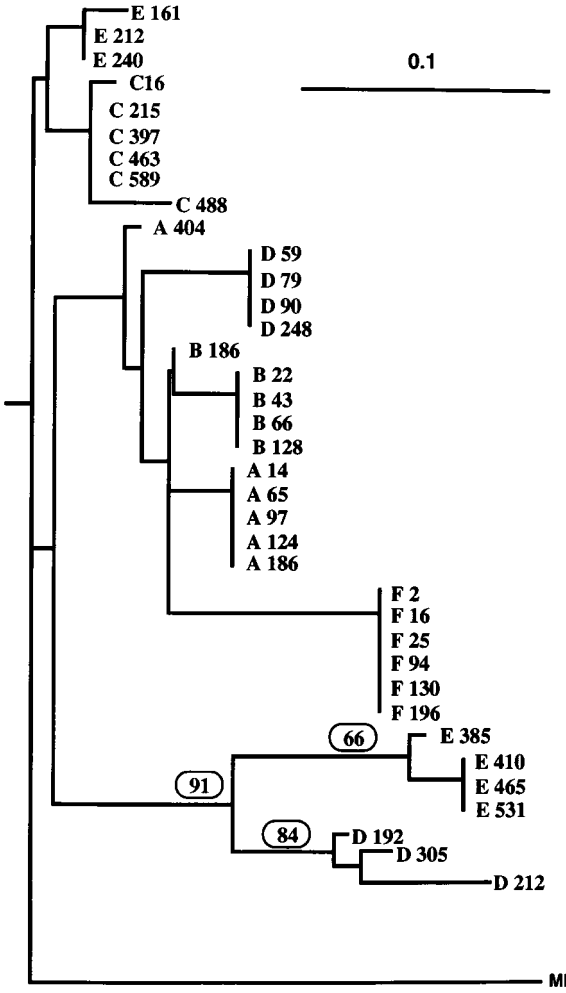



FIG. 6. Phylogenetic tree of 37 predominant HIV-1 *env* V3 loop amino acid sequences from six infants with vertically acquired HIV-1 infection sampled serially over a 6- to 15-month period. Infants D and E developed AIDS, and infants A, B, C, and F were healthy. The predominant sequence was defined as the majority sequence. When there was no majority sequence, the sequence that matched the consensus sequence for that time point was used. Consensus sequences were based on the most common amino acid in a given position. The frequency of viral variants as determined by heteroduplex analysis was used to determine the majority sequence and to select the most common amino acids. The tree was generated by the Fitch-Margoliash method and drawn with the DRAWGRAM program. The circled numbers at the nodes indicate the proportion of support in 500 bootstrap resamplings. Each sequence is identified with a letter for the subject, followed by the age in days at the time of sampling. Horizontal branch lengths are drawn to scale (as indicated) with arbitrary vertical spacing.

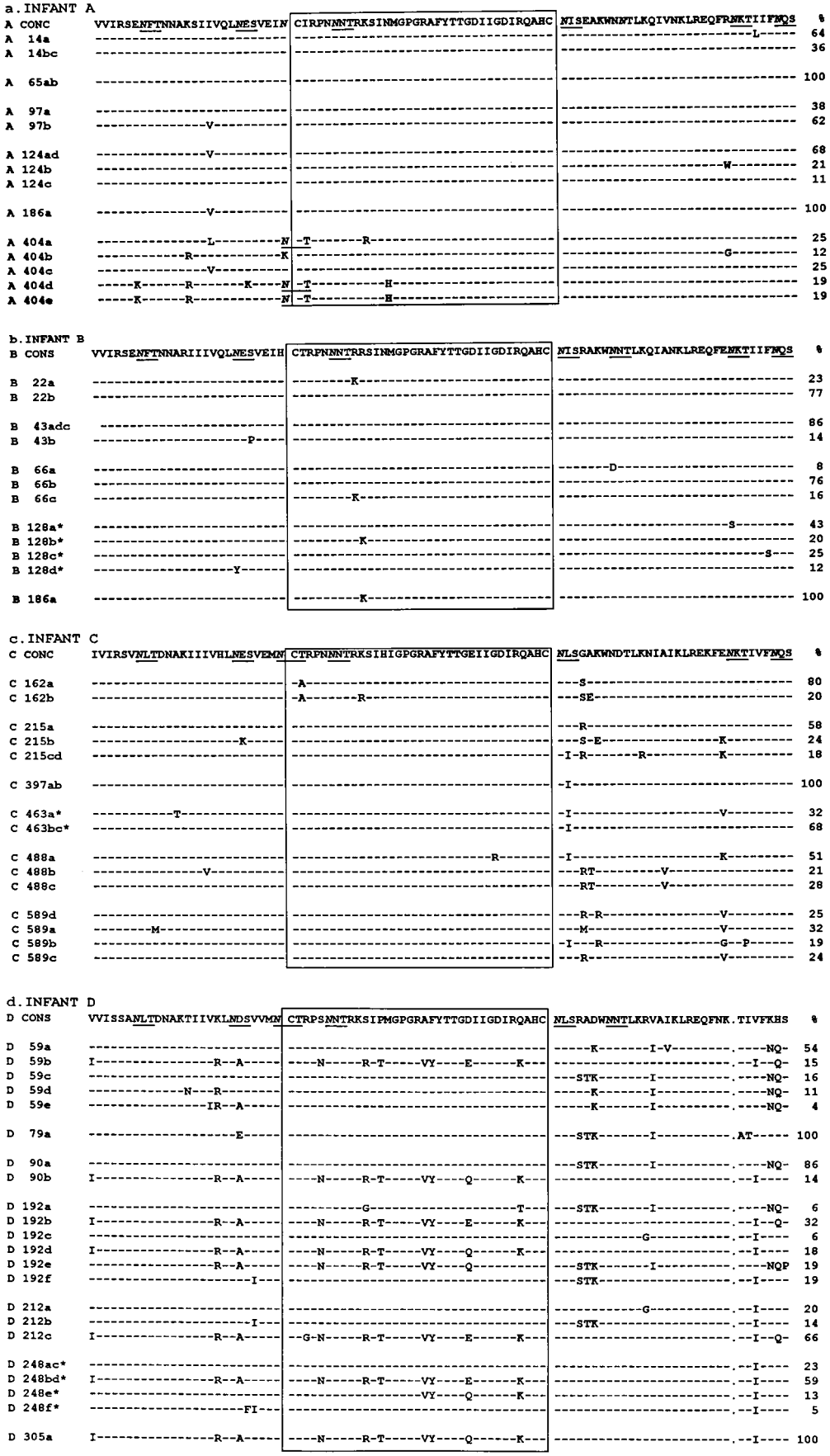
## DISCUSSION

One approach to the study of HIV-1 genetic variation is the use of sequence data to reconstruct evolutionary relationships. We applied these methods to V3 nucleotide and inferred amino acid sequences obtained in a longitudinal study of HIV-1-infected infants. The phylogenetic analysis of nucleotide sequences revealed differences in the modes of inter- and intra-infant viral evolution. The clusters of sequences from the six infants were arrayed in a star-shaped phylogeny, as has been found in general for HIV-1 (28, 31). This pattern of evolution is indicative of a fixed rate of divergence from a common ancestral sequence. There is no direct epidemiological linkage between the infants or their mothers; therefore, the sequence may represent a general ancestral subtype B virus or the ancestor of particular local strains. For intrainfant viral evolution, the shape of the phylogeny tended to be linear. The pattern is most clearly illustrated by the infants who developed AIDS, whose sequences can be ordered in a stepwise pattern with the earliest sequences closest to the ancestral node. The phylogeny is reminiscent of that for influenza A and parainfluenza viruses isolated over many years (5, 9) and suggests a process of sequential replacement of variants, perhaps as a result of periodic escape from an immune response (36) or increasing tropism for a productive host cell type (26, 30, 33, 50). The rapid viral evolution in infants who develop AIDS is most likely due to high rates of viral replication.

Having a set of longitudinal data allowed us to generate an actual rate of nucleotide substitution. To estimate rates, we designed a computer program, DISTRATE, that calculates and plots DNA distance values against time. Several measures of sequence variation over time were derived from the program, including rate of sequence divergence, rates of synonymous and nonsynonymous nucleotide substitution, and inferred initial diversity of the viral population. For most measures, there was a trend toward higher rates for infants whose disease progressed more rapidly; however, studies with larger numbers of infants are necessary to determine if DISTRATE analysis can predict disease progression and provide insights into the genetic mechanisms underlying viral diversity.

The most striking finding from the analysis of inferred amino acid sequences was the unexpected clustering of V3 loop sequences detected at later time points from the two infants who developed AIDS. This clustering was unexpected given the initially diverse genetic background of the infants' viruses and the tendency of these viruses to diverge over time. The observation was robust, since multiple tree-building methods gave the same result and the clustering was supported by bootstrap analysis. There are two possible explanations for the discrepancy between the clustering of viral variants based on genetic lineage and the protein similarity-based clustering of V3 loop sequences. One explanation is that the apparent similarity is due to lack of divergence from an evolutionarily conserved sequence. This explanation assumes that the infants were infected with multiple strains and that the later sequences from the infants who developed AIDS failed to replicate efficiently during the early period of infection. However, recent studies have shown that viruses detected during primary infection are genetically homogeneous (52, 53). In addition, from their inferred phenotype, the later sequences from the infants with AIDS would be expected to replicate efficiently (6, 13, 14, 19, 41). Convergent evolution, or the acquisition of identical amino acids and function in two lineages after they have diverged, could also explain the finding (45). A number of observations support this hypothesis. The viruses from the infants

FIG. 7. Deduced amino acid sequence alignments of the C2-V3 region of the envelope gene for six HIV-1-infected infants. The sequences from each infant (A to F) are aligned with the majority consensus sequence from that infant at the top of each alignment. Amino acids identical to the consensus sequence are indicated by a dash. The dot in the consensus sequence of infant D indicates a gap inserted to maintain alignment with the consensus sequences of the other infants. N-linked glycosylation sites are indicated by underlined letters with the first letter italicized. The V3 loop region is boxed. Each amino acid sequence is identified by a capital letter for the infant, followed by the age in days at the time of sample collection and a lowercase letter for unique DNA sequences from the same sample (in some cases, two or more DNA sequences encoded the same amino acid sequence). Sequences marked by asterisks were obtained by amplification of viral DNA from peripheral blood monocytes. For each sampling, the proportion of clones with identical amino acid sequences is shown to the right of the sequence.


a. INFANT A
b. INFANT B
c. INFANT C
d. INFANT D

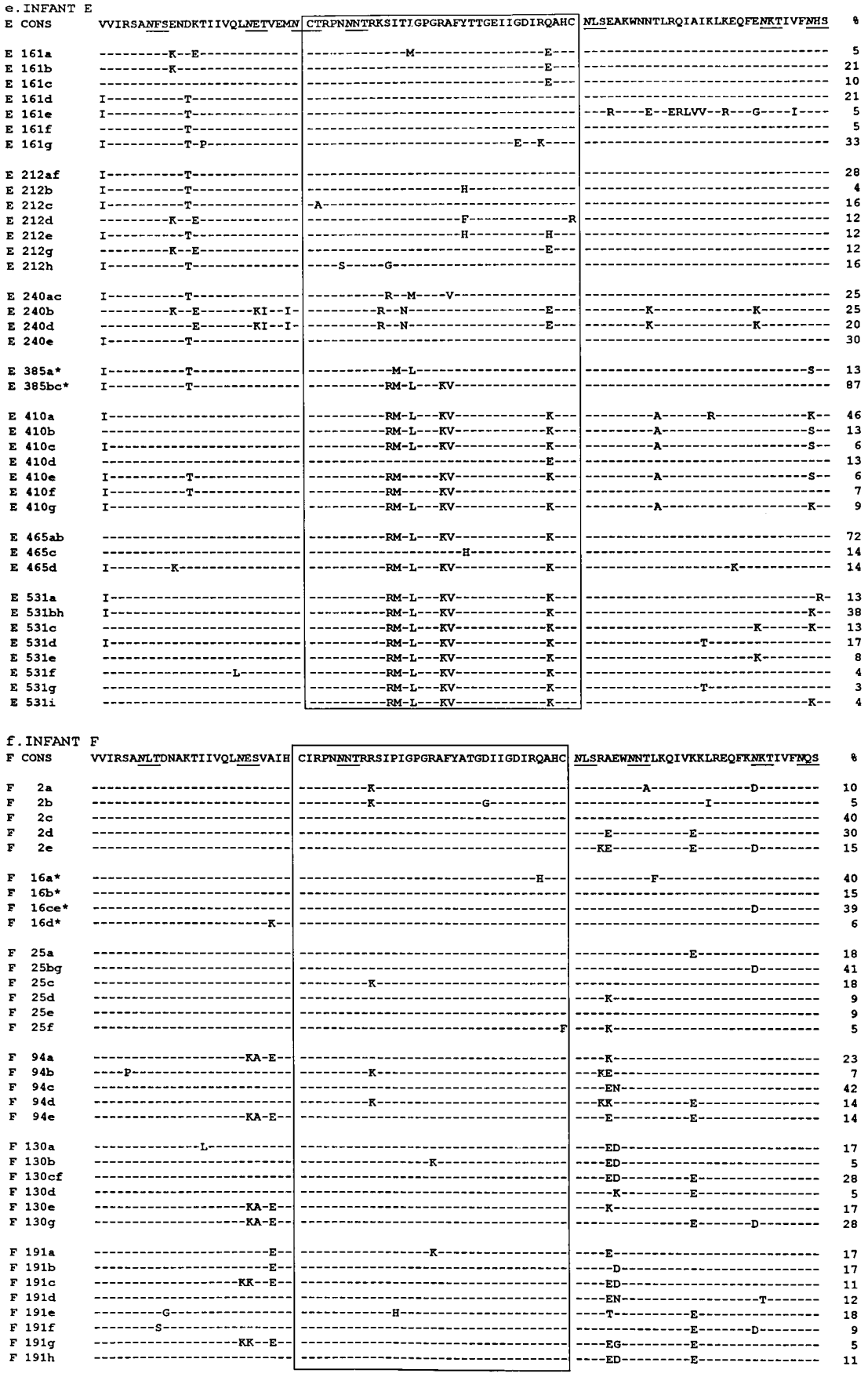



FIG. 7—*Continued.*

| SUBJECT | TIME PERIOD | CONSENSUS SEQUENCE |
|---|---|---|
| A | EARLY | CIRPNNNTRKSINMGPGRAFYTTGDIIGDIRQAHC |
| | LATE | -T---------- |
| B | EARLY | CTRPNNNTRRSINMGPGRAFYTTGDIIGDIRQAHC |
| | LATE | ---------K---------- |
| C | EARLY | C^RPNNNTRKSIHIGPGRAFYTTGEIIGDIRQAHC |
| | LATE | -T---------- |
| D | EARLY | CTRPSNNTRKSIPMGPGRAFYTTGDIIGDIRQAHC |
| | LATE | ----N-----R-T-----VY----Q------K--- |
| E | EARLY | CTRPNNNTRKSITIGPGRAFYTTGEIIGDIRQAHC |
| | LATE | ----------RM-L---KV----------K--- |
| F | EARLY | CIRPNNNTRRSIPIGPGRAFYATGDIIGDIRQAHC |
| | LATE | ---------- |

FIG. 8. Comparison of HIV-1 *env* V3 loop protein consensus sequences from early and late time periods during infection of AIDS-affected (D and E) and healthy (A, B, C, and F) infants. Early and late time periods were defined as the first and second halves of the sampling period, respectively. A consensus sequence based on the most common amino acid in a position was determined from the set of sequences for each time period. The frequency of viral variants as determined by heteroduplex analysis was used to select the most common amino acid. Dashes represent identity with the upper sequence in each pair. Amino acids that are shared between late consensus sequences of the subjects who developed AIDS are in boldface and underlined.

constitute distinct lineages. There was a temporal progression from early, phylogenetically unrelated sequences to late, convergent sequences. Convergence would be expected to occur in a functional domain of a protein and to alter an important biological property of a gene. The V3 loop region of the *env* gene can be considered a functional domain, and the inferred phenotype of the convergent sequences is T-cell line tropic while that of the early sequences from the same infants is macrophage tropic (23, 34, 42, 48). The consensus V3 loop amino acid sequence of the convergent viruses is identical, except for one conservative substitution, to a previously described motif for T-cell line tropic viruses (12, 49). The importance of these sites for protein structure and function is also supported by the observation of Korber et al. that three of the four sites are involved in covariant mutations (25). Finally, the inferred phenotype of the convergent sequences is one that has been associated previously with AIDS (13, 29, 33, 34).

Because most mutations are selectively neutral and the same function can be performed in different ways, convergent evolution is rare. The best example is the stomach lysozyme of two foregut fermenters, the cow and the langur monkey. From a tree analysis of the lysozyme enzymes of these and related species, Stewart et al. (45) have suggested that the langur monkey lysozyme gained sequence similarity to the cow lysozyme. Convergence within the V3 loop of HIV has also been proposed previously. Holmes et al. (21) identified mutations that resulted in the same hexapeptide motif (GPGSAV) in the crown of the V3 loop in viruses from two lineages from a single patient. However, a specific function was not associated with the motif, and for viruses from the same patient, it is difficult to exclude the possibility of a recombination event.

Whether convergent evolution toward the particular V3 loop amino acid motif we observed will occur in other individuals infected with subtype B viruses and in individuals infected with viruses of other subtypes needs to be determined by future studies. However, it is noted that three of the four amino acids characteristic of the convergent sequences have been described in viruses from subtypes A, B, D, E, and U, and two of these amino acids have been described in subtype C viruses (31). Thus, it is possible that the same motif could occur in individuals infected with viruses of other subtypes.

Convergent evolution of HIV-1 variants from infants who develop AIDS lends support to the notion that progression to disease is associated with the emergence of specific viruses. If, as their inferred phenotype suggests, these viruses are also more virulent than those found during the early, asymptomatic period of infection, then convergence would also imply that there are a limited number of ways that HIV-1 can cause AIDS. This, in turn, could have important implications for early identification of patients at risk for progression to disease, as well as for development of vaccines and other therapeutic modalities to prevent AIDS.

## ACKNOWLEDGMENTS

This work was supported by grants R01 AI30916 (R.P.V.) and U01 AI27565 (R.A.L.) from The National Institutes of Health.

We thank Man Nguyen, Christine Chung, and Oliver Laeyendecker for technical assistance, David Schwartz for helpful discussions, Alex Strunnikov for critical comments on the manuscript, and Nancy Hutton and Mary Joyner for clinical collaboration.

## REFERENCES

1. **Ahmad, N., B. M. Baroudy, R. C. Baker, and C. Chappey.** 1995. Genetic analysis of human immunodeficiency virus type 1 envelope V3 region isolates from mothers and infants after perinatal transmission. J. Virol. **69:**1001–1012.
2. **Albert, J., J. Wahlberg, T. Leitner, D. Escanilla, and M. Uhlén.** 1994. Analysis of a rape case by direct sequencing of the human immunodeficiency virus type 1 *pol* and *gag* genes. J. Virol. **68:**5918–5924.
3. **Alizon, M., S. Wain-Hobson, L. Montagner, and P. Sonigo.** 1986. Genetic variability of the AIDS virus. Cell **46:**63–74.
4. **Altman, D. G.** 1991. Practical statistics for medical research, p. 426–433. Chapman & Hall, London.
5. **Ambrose, M. R., S. V. Hetherington, A. S. Watson, R. A. Scroggs, and A. Portner.** 1995. Molecular evolution of the F glycoprotein of human parainfluenza virus type 1. J. Infect. Dis. **171:**851–856.
6. **Asjo, B., J. Albert, A. Karlsson, L. Morfeldt-Mamson, G. Biberfeld, K. Lidman, and E. M. Fenyo.** 1986. Replicative properties of human immunodeficiency virus from patients with varying severity of HIV infections. Lancet **ii:**234–238.
7. **Bassam, B. J., G. Caetano-Anolles, and P. M. Gresshoff.** 1991. Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal. Biochem. **196:**80–83.
8. **Briant, L., C. M. Wade, J. Puel, A. L. Brown, and M. Guyader.** 1995. Analysis of envelope sequence variants suggests multiple mechanisms of mother-to-child transmission of human immunodeficiency virus type 1. J. Virol. **69:**3778–3788.
9. **Buonagurio, D. A., S. Nakada, J. D. Parvin, M. Krystal, P. Palese, and W. M. Fitch.** 1986. Evolution of human influenza A viruses over 50 years: rapid, uniform rate of change in NS gene. Science **232:**980–982.
10. **Cabot, E. L., and A. T. Beckenback.** 1989. Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. Comput. Appl. Biosci. **5:**233–234.
11. **Cheng-Mayer, C., D. Seto, M. Tateno, and J. A. Levy.** 1988. Biologic features of HIV that correlate with virulence in the host. Science **240:**80–82.
12. **Chesebro, B., K. Wehrly, J. Nishio, and S. Perryman.** 1992. Macrophage-tropic human immunodeficiency virus isolates from different patients exhibit unusual V3 envelope sequence homogeneity in comparison with T-cell-tropic isolates: definition of critical amino acids involved in cell tropism. J. Virol. **66:**6547–6554.
13. **Cornelissen, M., G. Mulder-Kampinga, J. Veenstra, F. Zorgdrager, C. Kuiken, S. Hartman, J. Dekker, L. van der Hoek, C. Sol, R. Coutinho, and J. Goudsmit.** 1995. Syncytium-inducing (SI) phenotype suppression at seroconversion after intramuscular inoculation of a non-syncytium-inducing/SI phenotypically mixed human immunodeficiency virus population. J. Virol. **69:**1810–1818.
14. **De Jong, J., J. Goudsmit, W. Keulen, B. Klavor, W. Krone, M. Tersmette, and A. De Ronde.** 1992. Human immunodeficiency virus type 1 clones chimeric for the envelope V3 domain differ in syncytium formation replication capacity. J. Virol. **66:**757–765.
15. **Delwart, E. L., E. G. Shpaer, J. Louwagie, F. E. McCutchan, M. Grez, H. Rubsamen-Waigmann, and J. I. Mullins.** 1993. Genetic relationships determined by a DNA heteroduplex mobility assay: analysis of HIV-1 env genes. Science **262:**1257–1261.
16. **Donaldson, Y. K., J. E. Bell, E. C. Holmes, E. S. Huges, H. K. Brown, and P. Simmonds.** 1994. In vivo distribution and cytopathology of human immunodeficiency virus type 1 showing restricted sequence variability in the V3 loop. J. Virol. **68:**5991–6005.

17. **Felsenstein, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:**783–791.
18. **Felsenstein, J.** 1989. PHYLIP-phylogeny inference package (version 3.2). Cladistics **5:**164–166.
19. **Fouchier, R. A. M., M. Groenink, N. A. Kootstra, M. Tersmette, H. G. Huisman, F. Miedema, and H. Schuitemaker.** 1992. Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. J. Virol. **66:**3183–3187.
20. **Hahn, B. H., G. M. Shaw, M. E. Taylor, R. R. Redfield, P. D. Markham, S. Z. Salahuddin, F. Wong-Staal, R. C. Gallo, E. S. Parks, and W. P. Parks.** 1986. Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. Science **232:**1548–1553.
21. **Holmes, E. C., L. Q. Zhang, P. Simmonds, C. A. Ludlam, and A. J. Leigh Brown.** 1992. Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. Proc. Natl. Acad. Sci. USA **89:**4835–4839.
22. **Holmes, E. C., L. Q. Zhang, P. Simmonds, A. S. Rogers, and A. J. L. Brown.** 1993. Molecular investigation of human immunodeficiency virus (HIV) infection in a patient of an HIV-infected surgeon. J. Infect. Dis. **167:**1411–1414.
23. **Hwang, S. S., T. J. Boyle, H. K. Lyerly, and B. R. Cullen.** 1991. Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. Science **253:**71–73.
24. **Korber, B. T. M., K. MacInnes, R. F. Smith, and G. Myers.** 1994. Mutational trends in V3 loop protein sequences observed in different genetic lineages of human immunodeficiency virus type 1. J. Virol. **68:**6730–6744.
25. **Korber, T. M., R. M. Farber, D. H. Wolpert, and A. S. Lapedes.** 1987. Covariation of mutations in the V3 loop of human immunodeficiency virus type 1 envelope protein: an information theoretic analysis. Proc. Natl. Acad. Sci. USA **90:**7176–7180.
26. **Kuiken, C. L., J. J. De Jong, E. Baan, W. Keulen, M. Tersmette, and J. Goudsmit.** 1992. Evolution of the V3 envelope domain in proviral sequences and isolates of human immunodeficiency virus type 1 during transition of the viral biological phenotype. J. Virol. **66:**4622–4627.
27. **Levy, J. A.** 1993. Pathogenesis of human immunodeficiency virus infection. Microbiol. Rev. **57:**183–289.
28. **Louwagie, J., F. E. McCutchan, M. Peeters, T. P. Brennan, E. Sanders-Buell, G. A. Eddy, et al.** 1993. Phylogenetic analysis of gag genes from 70 international HIV-1 isolates provides evidence for multiple genotypes. AIDS **7:**769–780.
29. **Mammano, F., F. Salvatori, L. Ometto, M. Panozzo, L. Chieco-Bianchi, and A. De Rossi.** 1995. Relationship between the V3 loop and the phenotypes of human immunodeficiency virus type 1 (HIV-1) isolates from children perinatally infected with HIV-1. J. Virol. **69:**82–95.
30. **McNearney, T., Z. Hornickova, R. Markham, A. Birdwell, M. Arens, A. Saah, and L. Ratner.** 1992. Relationship of human immunodeficiency virus type 1 sequence heterogeneity to stage of disease. Proc. Natl. Acad. Sci. USA **89:**10247–10251.
31. **Meyers, G., B. Korber, J. A. Berzofsky, R. F. Smith, and G. N. Pavlakis.** 1992. Human retroviruses & AIDS 1992, p. III2–III4. Los Alamos National Laboratory, Los Alamos, N.Mex.
32. **Meyers, G., K. Maclinnes, and B. Korber.** 1992. The emergence of simian/human immunodeficiency viruses. AIDS Res. Hum. Retroviruses **8:**373–386.
33. **Miedema, F., L. Meyaard, M. Koot, M. R. Klein, M. T. L. Roos, M. Groenink, R. A. M. Fouchier, A. B. Van't Wout, M. Tersmette, P. T. A. Schellekens, and H. Schuitemaker.** 1994. Changing virus-host interactions in the course of HIV-1 infection. Immunol. Rev. **140:**35–69.
34. **Milich, L., B. Margolin, and R. Swanstrom.** 1993. V3 loop of the human immunodeficiency virus type 1 Env protein: interpreting sequence variability. J. Virol. **67:**5623–5634.
35. **Nei, M., and T. Gojobori.** 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:**418–426.
36. **Nowak, M. A., R. M. Anderson, A. R. McLean, T. F. W. Wolfs, J. Goudsmit, and R. M. May.** 1991. Antigenic diversity thresholds and the development of AIDS. Science **254:**963–969.
37. **Oka, S., S. Ida, T. Shioda, Y. Takebe, N. Kobayashi, Y. Shibuya, et al.** 1994. Genetic analysis of HIV-1 during rapid progression to AIDS in an apparently healthy man. AIDS Res. Hum. Retroviruses **10:**271–276.
38. **Ou, C., C. A. Ciesielski, G. Myers, C. I. Bandea, C. Luo, B. T. M. Korber, J. I. Mullins, G. Schochetman, R. L. Berkelman, A. N. Economou, J. J. Witte, L. J. Furman, G. A. Satten, K. A. MacInnes, J. W. Curran, H. W. Jaffe, Laboratory Investigation Group, and Epidemiologic Investigation Group.** 1992. Molecular epidemiology of HIV transmission in a dental practice. Science **256:**1165–1171.
39. **Rubsamen-Waigmann, H., H. Von Briesen, H. Holmes, A. Bjorndal, B. Korber, R. Esser, S. Ranjbar, P. Tomlinson, B. Galvao-Castro, E. Karita, S. Sempala, C. Wasi, S. Osmanov, E.-M. Fenyo, and the WHO Network for HIV Isolation and Characterization.** 1994. Standard conditions of virus isolation reveal biological variability of HIV type 1 in different regions of the world. AIDS Res. Hum. Retroviruses **10:**1401–1408.
40. **Schuitemaker, H., M. Koot, N. A. Kootstra, M. Wouter Dercksen, R. E. Y. De Goede, R. P. Van Steenwijk, J. M. A. Lange, J. K. M. Eeftink Schattenkerk, F. Miedema, and M. Tersmette.** 1992. Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression to disease is associated with a shift from monocytotropic to T-cell-tropic virus populations. J. Virol. **66:**1354–1360.
41. **Shioda, T., J. A. Levy, and C. Cheng-Mayer.** 1991. Macrophage and T-cell line tropisms of HIV-1 are determined by specific regions of the envelope gp120 gene. Nature (London) **349:**167–169.
42. **Shioda, T., J. A. Levy, and C. Cheng-Mayer.** 1992. Small amino acid changes in the V3 hypervariable region of gp120 can affect the T-cell-line and macrophage tropism of human immunodeficiency virus type 1. Proc. Natl. Acad. Sci. USA **89:**9434–9438.
43. **Shpaer, E. G., and J. I. Mullins.** 1993. Rates of amino acid change in the envelope protein correlate with pathogenicity of primate lentiviruses. J. Mol. Evol. **37:**57–65.
44. **Simmonds, P., L. A. Zhang, F. McOmish, P. Balfe, C. A. Ludlam, and A. J. Leigh Brown.** 1991. Discontinuous changes of human immunodeficiency virus (HIV) type 1 *env* sequences in plasma viral and lymphocyte-associated proviral populations in vivo: implications for models of HIV pathogenesis. J. Virol. **65:**6266–6276.
45. **Stewart, C., J. W. Schilling, and A. C. Wilson.** 1987. Adaptive evolution in the stomach lysozymes of foregut fermenters. Nature (London) **330:**401–404.
46. **Tersmette, M., R. A. Gruters, F. De Wolf, R. E. Y. De Goede, J. M. A. Lange, P. T. A. Schellekens, J. Goudsmit, H. G. Huisman, and F. Miedema.** 1989. Evidence for a role of virulent human immunodeficiency virus (HIV) variants in the Pathogenesis of acquired immunodeficiency syndrome: studies on sequential HIV isolates. J. Virol. **63:**2118–2125.
47. **Wain-Hobson, S.** 1992. Human immunodeficiency virus type 1 quasispecies in vivo and ex vivo. Curr. Top. Microbiol. Immunol. **176:**181–193.
48. **Westervelt, P., D. B. Trowbridge, L. G. Epstein, B. M. Blumberg, Y. Li, B. H. Hahn, G. M. Shaw, R. W. Price, and L. Ratner.** 1992. Macrophage tropism determinants of human immunodeficiency virus type 1 in vivo. J. Virol. **66:**2577–2582.
49. **Willey, R. L., T. S. Theodore, and M. A. Martin.** 1994. Amino acid substitutions in the human immunodeficiency virus type 1 gp120 loop that change viral tropism also alter physical and functional properties of the virion envelope. J. Virol. **68:**4409–4419.
50. **Wolfs, T. F. W., J. J. De Jong, H. Van Den Berg, J. M. G. H. Tunagel, W. J. A. Krone, and J. Goudsmit.** 1990. Evolution of sequences encoding the principal neutralization epitope of human immunodeficiency virus 1 is host dependent, rapid, and continuous. Proc. Natl. Acad. Sci. USA **87:**9938–9942.
51. **Wolinsky, S. M., C. M. Wike, B. T. M. Korber, C. Hutto, W. P. Parks, L. L. Rosenblu, K. J. Kunstman, M. R. Furtado, and J. L. Munoz.** 1992. Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. Science **255:**1134–1137.
52. **Zhang, L. Q., P. MacKenzie, A. Cleland, E. C. Holmes, A. J. Leigh Brown, and P. Simmonds.** 1993. Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. J. Virol. **67:**3345–3356.
53. **Zhu, T., H. Mo, N. Wang, D. S. Nam, Y. Cao, R. A. Koup, and D. D. Ho.** 1993. Genotypic and phenotypic characterization of HIV-1 in patients with primary infection. Science **261:**1179–1181.