Human Immunodeficiency Virus Type 1 Genetic Evolution in Children with Different Rates of Development of Disease

SHANHTI GANESHA, RUTHE E. DICKOVER, BETTE T. M. KORBER, YVONNE J. BRYSON, AND STEVEN M. WOLINSKY

Department of Medicine, Northwestern University Medical School, Chicago, Illinois 60611; Department of Pediatrics, University of California, Los Angeles, California 90095; and Theoretical Division, T-10, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Received 24 May 1996/Accepted 16 September 1996

The rate of development of disease varies considerably among human immunodeficiency virus type 1 (HIV-1)-infected children. The reasons for these observed differences are not clearly understood but most probably depend on the dynamic interplay between the HIV-1 quasispecies virus population and the immune constraints imposed by the host. To study the relationship between disease progression and genetic diversity, we analyzed the evolution of viral sequences within six perinatally infected children by examining proviral sequences spanning the C2 through V5 regions of the viral envelope gene by PCR of blood samples obtained at sequential visits. PCR product DNAs from four sample time points per child were cloned, and 10 to 13 clones from each sample were sequenced. Greater genetic distances relative to the time of infection were found for children with low virion-associated RNA burdens and slow progression to disease relative to those found for children with high virion-associated RNA burdens and rapid progression to disease. The greater branch lengths observed in the phylogenetic reconstructions correlated with a higher accumulation rate of nonsynonymous base substitutions per potential nonsynonymous site, consistent with positive selection for change rather than a difference in replication kinetics. Viral sequences from children with slow progression to disease also showed a tendency to form clusters that associated with different sampling times. These progressive shifts in the viral population were not found in viral sequences from children with rapid progression to disease. Therefore, despite the HIV-1 quasispecies being a diverse, rapidly evolving, and competing population of genetic variants, different rates of genetic evolution could be found under different selective constraints. These data suggest that the evolutionary dynamics exhibited by the HIV-1 quasispecies virus populations are compatible with a Darwinian system evolving under the constraints of natural selection.

Pediatric human immunodeficiency virus type 1 (HIV-1) infection is typified by a bimodal pattern of disease progression (9, 24), with substantial differences in the rates of development of illness and survival times (6, 7, 31, 38, 67, 70, 71). Approximately 20% of perinatally infected infants have accelerated immunological deterioration with a precipitously declining CD4 T-cell count (6, 21) after adjusting for the normal physiological T-cell loss during early infancy (87). Rapid deterioration with clinical manifestations of profound immunodeficiency, failure to thrive, and developmental delay or regression in intellectual and developmental milestones develop within the first 2 years of life (24, 71). For these children, there is high early mortality. Of the remaining 80% of children with perinatal HIV-1 infection, a relatively slower development of disease is observed (6, 31, 38). A minority of these children have survived into adolescence with limited clinical sequelae of HIV-1 disease (33). Nevertheless, the average survival time for even these children is less than that of infected adults (67).

A number of underlying determinants have been suggested to be responsible for the variability in clinical outcomes. These include the different routes of perinatal HIV-1 transmission (8, 16, 60), the viral burden (3, 7, 10, 15, 20–22, 30, 56, 85), the biological phenotype (19, 52, 53, 65, 66, 69, 76, 81–83, 86), and the nascent immune response (2, 34, 50, 51, 54, 63, 64). Transmission and persistence of a viral form with attenuated properties have been associated with long-term survival (10, 16, 32, 39, 40, 55). Conversely, accelerated immunological deterioration has been associated with infection in utero, provisionally defined by the isolation of a virus or the enzymatic amplification of a target viral sequence within the first 2 days of life (8), and transmission or emergence of a viral form with great pathogenic potential (69, 76, 81–84). Furthermore, since infection may occur during the ontogeny of the fetal and newborn immune system (50, 64), progenitor cells might also present a fertile milieu for HIV-1 replication by providing a large number of T-cell precursors as target cells and a paucity of effector cells in the T-cell repertoire to effectively control viral replication (63). Each of these potential determinants can affect the rate of development of disease by modifying the dynamic interplay between the selective immune constraints imposed by the host and the HIV-1 quasispecies virus population (18, 49, 73, 90, 92).

To assess the relationship between disease progression and genetic diversity, we analyzed the evolution of viral sequences within six perinatally infected children by examining proviral sequences spanning the C2 through V5 regions of the viral envelope gene by PCR of blood samples obtained at sequential visits. Product DNAs from four time points per child were cloned, and 10 to 13 clones from each sample were sequenced. Phylogenetic reconstructions of these sequences showed distinct clusters corresponding to each child, with greater genetic diversity observed for children with slow than with rapid development of disease, that correlated with a higher accumulation rate of nonsynonymous base substitutions per potential nonsynonymous site relative to the accumulation rate of syn-
obtained from the blood of each child is annotated on the abscissa of each graph. For child A, T-cell counts are presented as absolute numbers rather than age-corrected values (87). The time in weeks after birth when sequencing was performed from samples HIV-1 RNA levels were measured by a quantitative reverse transcription-PCR assay (Roche Molecular Systems, Inc.) and are expressed per milliliter of plasma. The T-cell counts are presented as absolute numbers rather than age-corrected values (87). The time in weeks after birth when sequencing was performed from samples obtained from the blood of each child is annotated on the abscissa of each graph. For child A, I = 8 weeks, II = 32 weeks, III = 54 weeks, and IV = 72 weeks. For child B, I = 4 weeks, II = 68 weeks, III = 176 weeks, and IV = 228 weeks. For child C, I = 2 weeks, II = 16 weeks, III = 62 weeks, and IV = 124 weeks. For child D, I = 3 weeks, II = 82 weeks, III = 104 weeks, and IV = 156 weeks. For child E, I = 20 weeks, II = 64 weeks, III = 130 weeks, and IV = 191 weeks. For child F, I = 0 weeks, II = 50 weeks, III = 82 weeks, and IV = 200 weeks. Specific antiretroviral therapies and their duration are highlighted within the graph: ZDV is represented in purple; DDI is represented in pink; ACTG protocol 152 is represented in orange; ACTG protocol 190 is represented in magenta; and nevirapine is represented in brown.

FIG. 1. Summary of CD4 and CD8 T-cell subset counts and quantitative HIV-1 virion-associated RNA for each of the six perinatally infected children (A through F). CD4 T-cell counts per milliliter (blue), CD8 T-cell counts per milliliter (green), and viral RNA copies per milliliter (red) are annotated on the ordinate. The time after birth in weeks is annotated on the abscissa. CD4 and CD8 T-cell counts were measured by a flow-activated cytometric assay and expressed per microliter of blood.

anonymous base substitutions per potential synonymous site. Thus, the observed difference in the rates of accumulation of mutations was probably due to positive selection for change rather than a difference in replication kinetics. These dynamic interactions between the virus and the host environment are consistent with adaptive evolution.

MATERIALS AND METHODS

Study subjects. Six children with perinatal HIV-1 infection, from the University of California Los Angeles Perinatal Transmission Study, were selected for study on the basis of the difference in their age-adjusted CD4 T-cell count, the HIV-1 virion-associated RNA burden, and the presence or absence of clinical manifestations of immunodeficiency. At each interval visit, clinical data from an interim history and physical examination and laboratory data from appropriate laboratory tests were obtained. Serial peripheral blood samples were drawn in an acid-citrate-dextrose buffer and separated into plasma and mononuclear cell fractions by Ficoll-Hypaque (Pharmacia, Inc.) discontinuous density gradient centrifugation. Plasma was stored at −80°C, and the cellular fraction was cryopreserved in liquid nitrogen vapor phase for future analysis. Informed consent was obtained under the guidelines of the University of California Los Angeles Human Subjects Protection Committee.

Figure 1 shows the CD4 and CD8 T-cell counts, plasma HIV-1 RNA burden, and the time and duration of antiretroviral therapy in relation to the time of parturition. Two children (child A and child B) had a precipitous decline in their CD4 T-cell percentage to values below the normal age-adjusted range with rapid progression to AIDS, one child (child C) had a CD4 T-cell percentage with an initial transient fall followed by recovery within the age-adjusted range temporally associated with the introduction of antiretroviral therapy, and three children (child D, child E, and child F) had a relatively stable CD4 T-cell percentage within the age-adjusted range with a slowly progressive disease course. Two of the last three children (child D and child E) also had a change in the CD4 T-cell percentage that was temporally associated with the beginning of antiretroviral therapy. Four of the six children (child A, child B, child C, and child F) were found to be infected at birth. No mother had received intrapartum antiretroviral therapy.

Child A, a Hispanic male born in June 1991, was found to be HIV-1 infected at birth as defined by positive DNA and RNA PCR tests, a positive immune complex dissociated p24 assay, and recovery of virus from peripheral blood mononuclear cells (PBMCs) cocultured with phytohemagglutinin (PHA)-stimulated normal donor PBMCs at 2 days of age. The infant showed the onset of clinical signs and symptoms of HIV-1 disease by 5 weeks of age; these progressed through bacterial meningitis, pneumonia, and sepsis and a delay in developmental milestones by 20 weeks of age. Clinical AIDS developed despite initiation of zidovudine (ZDV) therapy at 12 weeks of age and addition of dideoxynosine (DDI) at 34 weeks of age. Child A died of Pneumocystis carinii pneumonia at 74 weeks of age.

Child B, a Caucasian male born in April 1990, was found to be HIV-1 infected at birth by positive DNA and RNA PCR tests, a positive immune complex dissociated p24 assay, and virus isolation after coculture of PBMCs with PHA-stimulated normal donor PBMCs. He developed bacterial pneumonia at 3 weeks of age. At 12 weeks of age, he showed the onset of clinical signs and symptoms of HIV-1-related diseases that included persistent Candida albicans infection, parotitis, pneumonitis, and hepatosplenomegaly. ZDV therapy was initiated at 28 weeks of age. Clinical AIDS developed at 36 weeks of age with the onset of cardiomegaly and a delay in developmental milestones. Because in vitro ZDV resistance was found at 91-weeks of age, treatment was switched to DDI. Despite this change in therapy, child B has continued to fare poorly at 5 years of age.

Child C, a Hispanic female born in October 1992, was found to be HIV-1-infected at birth by positive DNA and RNA PCR tests and by recovery of a virus from PBMCs cocultured with PHA-stimulated normal donor PBMCs. She had a rapid onset of clinical signs and symptoms of HIV-1-related disease, including hepatosplenomegaly and generalized lymphadenopathy, and a precipitously fall-
ing CD4 T-cell count by 3 weeks of age. Nevirapine was initiated at 15 weeks of age as part of a clinical trial. This resulted in a significant drop in the HIV-1 virion-associated RNA burden in plasma, an increase in her CD4 T-cell counts, and a resolution in her symptoms by 20 weeks of age. ZDV therapy was added at 22 weeks of age. Both drugs were withdrawn at 140 weeks of age because of a parental decision. Child C has continued to be asymptomatic with a normal CD4 T-cell count and growth within the normal age-adjusted range at 7 years of age.

Child F, an African-American male born in October 1990, was found to be HIV-1 infected at 8 weeks of age by positive DNA and RNA PCR tests and by recovery of the virus from PBMCs cocultured with PHA-stimulated normal donor PBMCs. He was asymptomatic with a normal CD4 T-cell count until the late onset of nonspecific symptoms, including lymphadenopathy and hepatosplenomegaly, at 60 weeks of age. At 96 weeks of age, he entered into the ACTG152 protocol, at which he had a 1.5-log-unit drop in his HIV-1 virion-associated RNA burden. Child F has remained asymptomatic at 5 years of age.

Cells and plasma. Sequential samples of cryopreserved PBMCs stored in 10% dimethyl sulfoxide–30% fetal bovine serum with contemporary samples of frozen plasma were available for analysis from child A at 8, 32, 54, and 72 weeks after birth; child D at 3, 82, 104, and 156 weeks after birth; child E at 20, 50, 130, and 191 weeks after birth; and child F at 0, 50, 82, and 200 weeks after birth. The serial time points are indicated by Roman numerals I through IV for each child in the accompanying figures (for a full description, see the legend to Fig. 1).

Virus culture. PBMCs from each child at each time point were serially diluted (10×) and cocultured with PBMCs as previously described by the ACTG Virology Technology Advisory Committee (1a). Culture supernatant was monitored for HIV-1 p24 antigen production on days 7, 14, and 21 by a commercial enzyme-linked immunosorbent assay (ELISA). A culture was considered positive when the optical density value was above a cutoff of 75 pg/ml. Viruses recovered from the culture supernatant of the first positive well was expanded by a single passage in PHA-stimulated normal donor PBMCs and, their titer was measured to determine the 50% tissue culture infective dose (TCID50) for each isolate.

Nucleic acid quantification. Cell-associated proviral DNA was quantified by a commercial noncompetitive PCR assay as described previously (30). Virion-associated HIV-1 RNA was measured in a 200-μl aliquot of plasma by the Amplicor HIV-1 monitor assay (Roche Molecular Systems, Inc.) as described previously (22).

In vitro enzymatic amplification. Lysate from approximately 1.5 × 10⁶ cells containing a minimum of 20 HIV-1 input molecules was used for enzymatic amplification of DNA by PCR in buffers (2.0 mM MgCl₂, 50 mM KCl, 10 mM Tris-HCl [pH 8.3], 100 μg of nuclelease-free bovine serum albumin) with 20 mM each dATP, dCTP, dGTP, and dTTP, 20 pmol each of the sense and antisense oligonucleotide primers, and 0.5 U of Taq DNA polymerase in a total reaction volume of 100 μl. The reaction tube was inserted into a well of a Perkin-Elmer/Corbett Genescan PCR cycle programmable up to 95°C for 2 min; then two cycles of denaturation at 94°C for 2 min, annealing at 50°C for 30 s, and extension at 72°C for 2 min; and then 30 cycles of denaturation at 94°C for 1 min, annealing at 50°C for 30 s, and extension at 72°C for 1 min 45 s. A final extension at 72°C for 7 min was added to the last cycle. A 5-μl aliquot was again amplified in a 100-μl reaction mixture containing 20 μM each inner oligonucleotide pair with the same cycling profile as above. Specific precautions to avoid template and amplified product DNA carryover were in effect at all times (46). Stringent quality control measures include physical separation of processing areas and use of positive-displacement pipettes. To screen for potential contamination of product DNA, the viral sequences were analyzed by cross-comparisons of all PCR product sequences included in the study as well as comparisons with other viral sequences generated in the laboratory and with common laboratory strains by using a neighbor-joining algorithm (41).

Molecular cloning. A 20-μl aliquot from the final inner nested amplification was resolved by electrophoresis on a 1.0% agarose gel to screen for the appropriate 380-bp PCR product. The remainder of the gel was visualized by ethidium bromide staining. The purified 380-bp PCR product was inserted into vector pGEM T (Promega, Inc.) by using the principles of TA cloning. A 2-μl aliquot of the ligated product was used to transform competent JM109 bacterial cells. Transformed colonies were color selected on ampicillin–isopropyl-β-D-thiogalactopyranoside (IPTG)–5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) agar plates. White colonies were chosen at random and amplified in culture. A small-scale plasmid preparation was performed to recover bacterial DNA. A portion of the plasmid DNA was digested with appropriate restriction endonucleases, and the insert was screened for appropriate size and relative quantity by ethidium bromide resolution on a 1% agarose gel.

DNA sequencing. The double-stranded plasmid DNA was sequenced in both the forward and reverse directions with the M13-21 universal primer, the M13 reverse primer, and an HIV-1 envelope internal primer (KK54) with Dye-Deoxy terminators (Applied Biosystems, Inc.). Approximately 1 μg of double-stranded DNA template was added to 80 μl Tris-HCl (pH 8.3)–2 mM MgCl₂–20 μM each dATP, dCTP, dGTP, and dTTP, 20 pmol each of the sense and antisense oligonucleotide primer, and 0.5 U Taq DNA polymerase. The total 20-μl reaction volume was overlaid with light white mineral oil to prevent evaporation. DNA sequencing was performed in an automated thermal cycler programmed for 25 cycles of denaturation at 98°C for 10 s, primer annealing at 50°C for 30 s, and extension at 72°C for 1 min. The DNA sequencing reaction mixtures were resolved by electrophoresis on a 10% polyacrylamide gel and analyzed with an Applied Biosystems 373A automated sequencing system.

Data analysis. Comparisons were done based on various breakdowns of the available PCR product sequence data sets. Initial amino acid alignments were generated with PIMA (74, 75), an algorithm that incorporates a gap penalty and accounts for amino acid side chain chemistry. Both DNA and protein alignments were manipulated with the multiple alignment sequence editor (MASE) (26). Simple sequence similarity comparisons were made with MASE after removing positions in the alignment in which gaps had been inserted to maintain the alignment. These were calculated as Hamming distances, or (1 − s) × 100, where s is the fraction of shared sites in two aligned nucleotide sequences. Phylogenetic reconstructions were generated by using neighbor joining in the PHYLIP program (28, 28a), with a Kimura two-parameter distance matrix and a ratio of transition to transversion of 1.3. The B clade consensus sequence, defined as the most common nucleotide in a given position, was used as an out group (58). To screen for potential cross-contamination, BLAST (4) was used to compare nucleotide sequences from each of the six children with sequences in the viral subsection of GenBank. The proportion of synonymous substitutions per potential synonymous site and the proportion of nonsynonymous substitutions per potential nonsynonymous site were calculated by the methods of Nei and Gojobori (95) and Wilcoxon signed-rank and a signed-sum tests (78) were used to assess the significance of linked and unlinked synonymous and nonsynonymous scores, respectively.

Nonsynonymous sequence accession numbers. The viral sequences have been submitted to GenBank under accession numbers U47950 through U48211.
HIV-1 sequences. Figure 2 shows the alignments of the deduced amino acid translations of the nucleotide sequences. In most of the viral sequences, the coding potentials of the envelope open reading frames was maintained. Potential inactivating mutations were observed in the two children who had rapid progression to AIDS (three frame shifts and one in-frame stop codon) and the two children who had slow CD4 T-cell decline (three in-frame stop codons). There were seven inactivating mutations in 120,000 bases sequenced, of which four were observed in the V3 region and three were observed in the V4 region loops. The four cysteine residues at positions 296 (INCTR), 331 (AHÇNI), 418 (LPCR1), and 445 (IRÇSN) (numbered according to the HIV-1 HXB2 R clone [58]) involved in disulfide bridge formation of the V3 and V4 region loops were generally highly conserved (56). Five viral sequences had amino acid changes at position 296 (DIII.6 [child D; time point III; clone 6]; FIV.12), 331 (DIII.14, EIV.2), or 445 (EIII.45). Because mutations in these positions can have drastic structural consequences for the conformation of the envelope protein, these changes are also likely to represent inactivating substitutions.

Length polymorphism was observed in the V4 and V5 regions of the envelope gene in the sequence sets from each of the children with the lower HIV-1 virion-associated RNA burdens and slower progression to disease. In these polymorphic regions, duplications or deletions of potential N-linked glycosylation sites (defined by an N-X-T or N-X-S sequon) were commonly observed. In contrast, limited length polymorphism was observed in the V4 and V5 regions of the viral sequences derived from the children with rapid progression to AIDS. Although the N-X-T or N-X-S sequon persisted at a number of positions over time, the loss or gain of other potential N-linked glycosylation sites was also found to occur. While the potential N-linked glycosylation site proximal to the 5′ cysteine of the V3 loop was initially absent in the viral sequences from child F, this site emerged at a later time of sampling. In contrast, this potential N-linked glycosylation site existed in child C at the earliest time of sampling but was replaced to some extent at a later time of sampling.

Net amino acid charge in the V3 loop and viral phenotype. The net charge on the protein regions of the entire V3 loop and a 15-amino-acid fragment encompassing the tip of the loop (4 amino acids) and the flanking N-terminal (4 amino acids) and C-terminal (7 amino acids) sides were calculated for the entire data set as described previously (29, 42, 52, 72, 89). No viral sequences carried the positively charged amino acids in the V3 region that have been associated with a syncytium-inducing phenotype. These data are consistent with the biological phenotype of the first-passage virus culture stocks from each child covering the course of infection, because all the propagated viruses exhibited a nonsyncytium-inducing phenotype in cultures of PBMCs and the established MT-2 transformed T-cell line.

Viral sequence variation among children. To construct evolutionary relationships among the six children, phylogenetic analyses of all viral sequences were performed by using a Kimura two-parameter distance matrix fed into a neighboring tree algorithm (28). Figure 3 shows that distinct clusters of viral sequences corresponding to each child were found. Furthermore, the viral sequences from each child clustered relative to a background set of B clade sequences of the M group (58). Identity or near identity was not found when the viral sequences from these children were compared with sequences of other viruses and molecular clones used in the laboratory (data not shown).

To further check the integrity of the viral sequences and exclude the possibility of contamination, sequences of approximately 650 nucleotides from each of the six children were compared to the viral subsection of GenBank by using BLAST (4). The regions with the highest scores ranged from 300 to 580 bases and had between 84 and 92% sequence identity, suggesting that there was no contamination with preexisting laboratory strains or published HIV-1 sequences. These analyses demonstrate the absence of PCR product contamination for the viral sequences within this data set (41).

Viral sequence variation within each child. Viral sequence similarity expressed as histograms showed that the sequence similarity within a child was greater for children with a relatively stable CD4 T-cell count than for children with a precipitously declining CD4 T-cell number (Fig. 4). The median similarity of all pairwise sequence comparisons within a child was 97%. While the viral sequences from the children who had slow development of disease varied up to 10% in the V3 to V5 regions over time, the viral sequences from the children with a rapid progression to AIDS varied by no more than 5%.

Unlike the observation of others (47), viral sequences obtained within 8 weeks after birth formed a relatively conserved population of viral genotypes, a finding compatible with the observations of ourselves (91) and others (1, 57). The ranges of intraclitide nucleotide similarities were within this sampling time: 96.3 to 100% (median, 98.8%) for child A, 97.2 to 99.8% (median, 98.8%) for child B, 99.0 to 100% (median, 99.5%) for child C, 98.7 to 100% (median, 99.4%) for child D, and 98.7 to 100% (median, 99.4%) for child F. No samples from child E were available within this time frame.

Among these HIV-1-infected children, there was a range of viral sequence diversity within each child over time that was associated with the rate of development of disease. The ranges of nucleotide distances within a child were for the sampling time between 120 and 180 weeks: 96.8 to 99.1% (median, 97.7%; median, 97.7%) for child B, 95.9 to 99.7% (mean, 98.0%; median, 97.7%) for child C, 92.8 to 99.4% (mean, 96.2%; median, 96.8%) for child D, and 90.7 to 99.4% (mean, 97.2%; median, 98.5%) for child E. No samples from child A and child F were available within this time frame. Child F, however, had a range of nucleotide distances of 91.9 to 99.4% (mean, 95.4%; median, 95.6%) for a sample obtained at 200 weeks after birth.

Phylogenetic analyses of sequence sets within each child. Two consistent topological patterns were observed in the phylogenetic reconstruction of genetic lineages of sequences from the children with rapid and slow CD4 T-cell loss over time. For the viral sequences obtained from the blood of those children with slow development of disease (Child D, child E, and child F), greater branch lengths were found relative to those found...
FIG. 2—Continued.
Distinct clusters of viral sequences corresponding to each child were found, sequences from the six children were performed by using a Kimura two-parameter distance matrix led into a neighbor-joining tree construction algorithm. Distinct clusters of viral sequences corresponding to each child were found, indicating absence of PCR product cross-contamination.

For children with a rapid progression to AIDS (Child A, child B, and child C). These differences were observed when sequences were compared at contemporary time points about 75 weeks after delivery.

Within this period, the medians of the neighbor-joining branch lengths to the ancestral node, with the B clade consensus sequence as an out group for these children, were 0.009 (mean, 0.009; interquartile range, 0.006 to 0.010) for child A; 0.016 (mean, 0.015; interquartile range, 0.013 to 0.016) for child B, 0.014 (mean, 0.015; interquartile range, 0.011 to 0.018) for child C; 0.018 (mean, 0.016; interquartile range, 0.009 to 0.022) for child D, 0.036 (mean, 0.035; interquartile range, 0.035 to 0.037) for child E, and 0.022 (mean, 0.022; interquartile range, 0.022 to 0.023) for child F. Therefore, the greater branch lengths were due to a higher rate of accumulation of mutations rather than a longer survival time for children with slower development of disease.

To monitor viral variation and evolutionary relationships over time, phylogenetic analyses of viral sequences within each child were used (Fig. 5). For the children who had a relatively low HIV-1 virion-associated RNA burden (child D, child E, and child F), the viral sequences showed distinctive clustering with sampling time. A pattern of viral sequence changes was observed whereby a specific variant was represented as the predominant form at each time point. This observation was supported by bootstrap proportions of greater than 50 of 100 bootstrap replicates as annotated at the appropriate branch point in the neighbor-joining phylogenetic reconstructions. Thus, for these children, there was a progressive shift in the viral populations over time. In contrast, the phylogenetic reconstructions for the children with the higher HIV-1 virion-associated RNA burdens (child A, child B, and child C) generally showed a relatively conserved, monophyletic population of viral genotypes with intermingling of sequences from different time points. Of note, two children (child B and child C) had a distinct cluster of viral sequences temporally associated with the second time point that was sampled. All viral sequences before and after this sampling time clustered as a monophyletic population. Overall, these findings were consistently observed when maximum-likelihood (28) or parsimony (80) methods were used for the confirmatory phylogenetic analyses.

The distribution of distinguishable variants might also be modified by the administration of a potent antiretroviral agent that destabilizes the equilibrium and by allowing selection of a molecule that encodes a mutation which confers antiretroviral resistance (14, 17). In the phylogenetic analyses of sequence sets for child A and child B, both of whom had a rapid progression to disease, no emergence of a distinct population of viral genotypes was temporally correlated with the introduction of antiretroviral therapy. For child C, however, greater branch lengths were observed for viral sequences from the fourth sample time which, coincidentally, did follow the introduction of a relatively potent antiretroviral therapeutic regimen. Like child C, child E and child F had greater branch lengths with a distinct clustering pattern after the introduction of antiretroviral therapy coincident with a precipitous decline in the HIV-1 virion-associated RNA burden in plasma. Thus, the change in the rate of progression to disease observed for child C and, to a lesser extent, child E and child F may have been the consequence of altered pressure for change provided by the introduction of the antiretroviral therapy.

To screen for evidence of convergent evolution within the V3 region that associated with the rate of progression to disease, the deduced amino acid sequences in this region for each of the six children were compared. In contrast to earlier studies that found an arginine at position 11, a threonine at position 13, a valine at position 19, and a lysine at position 32 to correlate with the development of AIDS (12, 43–45, 79), these specific amino acid changes at these positions were not found in this data set. These data suggest an absence of convergent evolution for children who had a rapid progression to AIDS, in contrast to the observations of others (37, 79).

**Accumulation rates of synonymous and nonsynonymous substitutions.** The accumulation rates of synonymous substitutions per potential synonymous site and nonsynonymous substitution per potential nonsynonymous site (59) in C2 to V5 of env were compared to screen for positive selection for change in these regions. Table 1 shows the accumulation rates of synonymous and nonsynonymous substitutions compared to the consensus of the viral sequences derived from the first time point. This was considered to be a reasonable estimate of the viral sequence that is predominantly represented at the time of primary infection. In general, the accumulation rates of nonsynonymous substitutions were greater in children who had a relatively stable CD4 T-cell count. These differences in accumulation were also at comparable sampling times. When all comparisons were made within each time point, progressive trends were not clear (Table 1); this is expected, as the quasi-species virus population may have undergone a transition in the period before sampling that dictates the observed sequence variation within a sample.

The ratios of synonymous to nonsynonymous base substitutions ($d_s/d_*$) for these six children at the same time points suggest that there is positive selection for change in the C2 to V5 regions. The lower $d_s/d_*$ ratios for children who had slow progression to disease suggest that the C2 through V5 regions in these children are under greater selective pressure for change than is the same region amplified from blood obtained from children who had rapid progression to AIDS. The num-
number of synonymous substitutions per week and the number of nonsynonymous substitutions per week were calculated for each child and statistically compared by using a Wilcoxon signed-rank test to account for linkage of the synonymous and nonsynonymous scores (Table 2). For the combined results for the two more rapidly progressing children, the synonymous substitutions accumulated at a significantly higher rate than did the nonsynonymous substitutions when analyzed by the Wilcoxon signed-rank test ($P = 3 \times 10^{-5}$). In contrast, for the combined results of the three children who had slow progression to disease, the distribution of rates was higher for the nonsynonymous than the synonymous substitutions ($P = 0.0033$). The one child who had a change in the rate of progression to disease after the introduction of potent antiretroviral therapy (child C) also had a higher distribution of rates of nonsynonymous substitutions, similar to that of the three children who had slow progression to disease. Overall, contrasting the combined rates from the two children with rapid progression to disease with those from the three children with slow development of disease by using a Wilcoxon rank sum test, the rates of accumulation were statistically higher for both synonymous ($P = 0.005$) and nonsynonymous ($P < 0.00001$) substitutions in children with slow compared with rapid progression to disease.

**DISCUSSION**

In this study, we characterized the viral sequence variation over time in the mononuclear cells from blood obtained from six perinatally infected children with different rates of progression to HIV-1 disease. Three of these six children (child D, child E, and child F) had a CD4 T-cell percentage that remained within the age-adjusted range, a relatively low HIV-1 virion-associated RNA burden, and slow development of disease. Only one of these three children (child F) was found to be HIV-1 infected at birth. Two additional children who were found to be infected at birth (child A and child B) had a precipitously declining CD4 T-cell percentage well below the age-adjusted range, a relatively high viral burden, and rapid development of clinical signs and symptoms of HIV-1 disease. The remaining child who was also diagnosed to be HIV-1-infected at birth (child C) had clinical and virological parameters and a rate of development of disease that although initially quite rapid, came to resemble the clinical profile of the children with slow progression to disease upon initiating antiretroviral therapy.

Whereas all the children had a relatively conserved monophyletic population of genotypic variants at the sampling time closest to delivery (1, 91), great genetic variation accrued over
FIG. 5. Phylogenetic reconstructions of evolutionary relationships within each child. The deduced amino acid sequences are annotated as described in the legend to Fig. 3. There was substantial intermingling of viral sequences from different time points, with generally low bootstrap values representing different clusters. The phylogenetic reconstructions shown are neighbor-joining trees with bootstrap proportions of greater than 50 of 100 bootstrap replicates (shown at appropriate branch points) (35). The ranges of the neighbor-joining branch lengths to the ancestral node by using the B clade consensus sequence as an out group were 0.00352 to 0.13619 (mean, 0.010538; median, 0.006024; interquartile range, 0.00309 to 0.03322) for child A, 0.01124 to 0.1052 (mean, 0.022306; median, 0.01927; interquartile range, 0.01318 to 0.0285) for child B, 0.00108 to 0.08498 (mean, 0.015042; median, 0.01267; interquartile range, 0.00472 to 0.02176) for child C, 0.00349 to 0.05932 (mean, 0.022169; median, 0.01642; interquartile range, 0.00515 to 0.04951) for child D, 0.01354 to 0.09355 (mean, 0.035794; median, 0.036045; interquartile range, 0.02094 to 0.03985) for child E, and 0.01836 to 0.07649 (mean, 0.03048; median, 0.02906; interquartile range, 0.02094 to 0.03805) for child F. The average branch lengths per month were 0.0004286 (over 74 months) for child A, 0.0003852 (over 212 months) for child B, 0.0004221 (over 124 months) for child C, 0.0005569 (over 191 months) for child D, and 0.0006024 (over 200 months) for child F. For each child, the sample time points are highlighted in blue (I), green (II), red (III), and purple (IV).
time among the children who had slow development of disease. This was evidenced by the greater branch lengths observed for the phylogenetic reconstructions. In contrast, a relatively conserved population of genotypic variants was observed for the children who had a rapid rate of progression to AIDS. The differences between these two groups were apparent at comparable time points after delivery. Thus, consistent with our previous observations for infected adults (91), genetic variation was not a consequence of the difference in the length of survival that existed among these children but, rather, a result of a higher rate of accumulation of mutations.

Viral diversification and evolution were monitored over time by phylogenetic analyses for each child. Distinctive clustering of viral sequences with sampling time was observed for the viral sequences obtained from children who had slow development of disease (child D, child E, and child F). These progressive shifts in the virus population were not observed for the viral sequences from children with rapid progression to disease. These children had a relatively conserved, monophyletic population of genotypic variants that persisted over time. Two of these three children (child B and child C) also had a distinct clustering of viral sequences at the second sample time after delivery. Since the sampled CD4 T-cell population is a reservoir of both expressed and unexpressed proviral molecules (14), intermingling of some viral sequences from different sample times is expected. Rather than observing an increase in the genetic diversity before a precipitous decline in the CD4 T-cell count, hypothesized to be a driving force in progression to AIDS (62), we observed a relatively conserved HIV-1 quasispecies virus population at this time.

Potential selective forces that introduce a genetic bottleneck might also modify the distribution of distinguishable variants in the HIV-1 quasispecies virus population (14, 25, 36) by accelerating the accumulation of deleterious mutations in the absence of recombinogenic events (27, 61). The higher rate of accumulation of nonsynonymous substitutions observed in children who had slow development of disease (Child D, child E, and child F) relative to the children who had rapid progression to AIDS (child A and child B) is consistent with positive selection for change in the envelope region. This finding might be an artifact of the available sampling times or some real phenomena whereby the substitutions occur at a higher frequency and therefore are perpetuated over time (23). Both the humoral and cell-mediated immune response have been postulated to be responsible for the observed differences in selection pressure. Previous studies of the extent and breadth of the humoral immune response associated with the rate of progression to disease, however, have been inconclusive. Although some studies have shown that a vigorous humoral immune response correlates with long-term survival (10, 48), other studies have not supported this finding (92). The cytotoxic T-lymphocyte response has also been proposed as a selective force that drives genetic variation (2, 11, 13, 50, 51, 54, 66, 88, 92). Amino acid substitutions in unambiguously defined human leukocyte antigen class I-restricted epitopes have been suggested as potential harbingers of antigenic escape.

Administration of a potent antiretroviral agent might also modify the distribution of distinguishable variants by destabilizing the equilibrium and allowing the selection of a molecule that encodes a mutation that confers antiretroviral resistance (14, 17). The rapid emergence of these mutations suggests that such mutant viruses preexist in the HIV-1 quasispecies virus population (14). As a consequence of altered pressure for change, these mutant viruses emerge as a distinct population of genetic variants. For three of these six children, a precipitous decline in the plasma HIV-1 virion-associated RNA burden

FIG. 5—Continued.
and a distinct clustering of viral sequences were observed after the introduction of antiretroviral therapy. These data suggest a temporal relationship between the initiation of treatment, a clinical response, and a rapid shift in the HIV-1 quasispecies virus population; however, because of the limitations in our data set, it is not possible to establish causality.

The relative fitness of the HIV-1 quasispecies virus population may also be modified by changes in replication kinetics and cell tropism. Numerous previous studies have documented the effect of changes in these parameters and the outcome of disease (56, 67). The presence of lymphocytotropic, syncytium-inducing viruses has been associated with a high rate of progression to disease (5, 15, 19, 52, 65, 68, 69, 82). In contrast, a lower rate of development of disease has been associated with non-syncytium-inducing viruses (10, 68). In this study, a non-syncytium-inducing virus was isolated from the peripheral blood of each child at each of the sampled time points. Thus, consistent with our earlier study (92), we did not observe a difference in viral cytopathogenicity, at least by current measurements, that was associated with the rate of progression to disease. Furthermore, evidence of convergent evolution of the viral sequences from the children who had rapid progression to AIDS was lacking. Thus, specific amino acids at positions 11, 13, 19, and 32 of the V3 loop, previously considered important for convergent evolution, were not independently conserved among the different viral sequences, in contrast to the observations of others (37, 79). These data are most compatible with the lack of divergence rather than the advent of convergent evolution.

In conclusion, these data are most compatible with the hypothesis that the critical determinant of progression to HIV-1 disease is the dynamic interplay between the HIV-1 quasispecies virus population and the selective constraints imposed by the host environment (18, 36, 92). Since the RNA virus quasispecies are a diverse, rapidly evolving population of competing molecules, different selective constraints would necessitate adaptation and, thereby, produce different rates of evolution (14, 25, 36, 92). Thus, genetic variation over time and evolutionary adaptation would be expected if a variant of relatively low fitness were replicating under significant selective constraints. Conversely, genetic variation over time and evolution would be minimal if a variant of relatively greater fitness were replicating under insignificant selective constraints. Between these two postulated extremes of the virus-host interaction are the different permutations of potential interactions between the virus and the host environment. Thus, the observation that the extent of the genetic complexity of the prevalent virus population sampled from blood was inversely related to the

### Table 1. Accumulation rates of synonymous and nonsynonymous substitutions compared to the consensus of the viral sequences derived from the first time point

<table>
<thead>
<tr>
<th>Child</th>
<th>Sampling time (wk after birth)</th>
<th>Median ds (IR)</th>
<th>Median da (IR)</th>
<th>Median ds/dn (IR)</th>
<th>Intra-time-point median ds/dn (IR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>8</td>
<td>1.5 (0.4–2.2)</td>
<td>0.4 (0.2–0.6)</td>
<td>3.7 (3.2–3.8)</td>
<td>3.8 (3.2–5.5)</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>0.7 (0.2–1.0)</td>
<td>0.6 (0.4–0.8)</td>
<td>1.2 (0.2–3.2)</td>
<td>1.9 (1.0–5.5)</td>
</tr>
<tr>
<td></td>
<td>54</td>
<td>0.0 (0.0–0.0)</td>
<td>0.6 (0.4–0.6)</td>
<td>0.0 (0.0–0.0)</td>
<td>1.1 (0.7–1.8)</td>
</tr>
<tr>
<td></td>
<td>72</td>
<td>1.5 (1.1–2.2)</td>
<td>0.6 (0.5–0.8)</td>
<td>2.3 (1.6–3.8)</td>
<td>3.2 (2.5–4.7)</td>
</tr>
<tr>
<td>B</td>
<td>4</td>
<td>0.8 (0.0–1.5)</td>
<td>0.4 (0.2–0.8)</td>
<td>2.8 (0.7–3.8)</td>
<td>2.5 (1.6–3.8)</td>
</tr>
<tr>
<td></td>
<td>68</td>
<td>0.8 (0.4–2.7)</td>
<td>1.2 (0.7–2.3)</td>
<td>1.2 (0.4–1.5)</td>
<td>1.2 (0.8–1.8)</td>
</tr>
<tr>
<td></td>
<td>170</td>
<td>4.1 (3.0–5.8)</td>
<td>2.1 (1.8–2.4)</td>
<td>1.8 (1.6–2.9)</td>
<td>3.3 (2.0–4.8)</td>
</tr>
<tr>
<td></td>
<td>228</td>
<td>3.0 (3.0–3.0)</td>
<td>1.7 (1.5–1.8)</td>
<td>1.8 (1.4–2.0)</td>
<td>2.6 (1.9–3.8)</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>0.0 (0.0–0.0)</td>
<td>0.2 (0.0–0.4)</td>
<td>0.0 (0.0–0.5)</td>
<td>1.9 (1.1–2.2)</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>1.9 (1.2–3.2)</td>
<td>1.6 (1.3–1.7)</td>
<td>1.4 (0.7–1.7)</td>
<td>2.9 (1.3–7.6)</td>
</tr>
<tr>
<td></td>
<td>62</td>
<td>0.6 (0.0–0.7)</td>
<td>0.7 (0.2–1.1)</td>
<td>0.5 (0.0–1.7)</td>
<td>1.0 (0.6–1.8)</td>
</tr>
<tr>
<td></td>
<td>124</td>
<td>0.7 (0.0–0.7)</td>
<td>0.8 (0.4–2.7)</td>
<td>0.3 (0.0–0.7)</td>
<td>0.6 (0.4–0.9)</td>
</tr>
<tr>
<td>D</td>
<td>3</td>
<td>0.4 (0.0–0.8)</td>
<td>0.4 (0.3–0.4)</td>
<td>1.9 (0.2–4)</td>
<td>1.9 (1.3–3.1)</td>
</tr>
<tr>
<td></td>
<td>82</td>
<td>1.1 (0.8–2.1)</td>
<td>2.3 (0.6–3.0)</td>
<td>0.9 (0.5–1.1)</td>
<td>0.9 (0.6–1.3)</td>
</tr>
<tr>
<td></td>
<td>104</td>
<td>1.5 (0.8–2.0)</td>
<td>2.6 (1.3–3.5)</td>
<td>0.4 (0.2–1.0)</td>
<td>0.9 (0.5–1.1)</td>
</tr>
<tr>
<td></td>
<td>156</td>
<td>3.1 (1.9–4.0)</td>
<td>5.6 (3.2–5.9)</td>
<td>0.6 (0.5–1.0)</td>
<td>1.0 (0.7–1.7)</td>
</tr>
<tr>
<td>E</td>
<td>20</td>
<td>1.9 (0.7–3.2)</td>
<td>1.3 (0.8–2.2)</td>
<td>1.0 (0.8–1.8)</td>
<td>1.1 (0.8–1.8)</td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>3.0 (1.5–3.5)</td>
<td>3.6 (3.1–4.2)</td>
<td>0.6 (0.4–1.1)</td>
<td>0.9 (0.7–1.2)</td>
</tr>
<tr>
<td></td>
<td>130</td>
<td>4.1 (2.2–4.5)</td>
<td>4.5 (4.2–4.6)</td>
<td>0.8 (0.5–1.0)</td>
<td>1.9 (1.3–3.2)</td>
</tr>
<tr>
<td></td>
<td>191</td>
<td>6.0 (5.6–6.7)</td>
<td>6.5 (6.1–6.5)</td>
<td>1.0 (0.8–1.1)</td>
<td>1.9 (1.3–2.7)</td>
</tr>
<tr>
<td>F</td>
<td>0</td>
<td>0.4 (0.0–0.8)</td>
<td>0.3 (0.0–0.4)</td>
<td>0.0 (0.0–2.8)</td>
<td>1.9 (1.2–3.8)</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>3.3 (3.1–3.3)</td>
<td>3.1 (2.4–3.2)</td>
<td>1.0 (1.0–1.3)</td>
<td>0.9 (0.5–1.7)</td>
</tr>
<tr>
<td></td>
<td>82</td>
<td>3.1 (2.3–3.1)</td>
<td>3.7 (3.7–3.7)</td>
<td>0.6 (0.6–0.8)</td>
<td>2.5 (1.9–3.8)</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>4.4 (3.2–4.6)</td>
<td>5.9 (4.9–6.1)</td>
<td>0.7 (0.6–0.8)</td>
<td>0.9 (0.7–1.3)</td>
</tr>
</tbody>
</table>

Notes:
- Time point comparisons of the accumulation of nonsynonymous substitutions per potential nonsynonymous site (ds) and synonymous substitutions per potential synonymous site (da) and the ratio of synonymous to nonsynonymous base substitutions (ds/dn) (59) within each child. The consensus sequence for each time point sampled was determined for each of the six children. The median numbers of nonsynonymous substitutions per potential nonsynonymous site and synonymous substitutions per potential synonymous sites relative to the original consensus sequence and their interquartile ranges (IR) were then calculated for each sample time point. Any viral sequence that lacked a nonsynonymous substitution (i.e., ds = 0) was not included in the calculation.
- One sequence was identical to the consensus, and one had two synonymous base changes.
- Two sequences were identical to the consensus.
- One sequence was identical to the consensus, and one had a single synonymous base change.
- Two sequences were identical to the consensus, and two had single synonymous base changes.

**References:**
5, 15, 19, 52, 65, 68, 69, 82.
A one-sided test was used to test the hypothesis. The evolution of the HIV-1 quasispecies virus population under survival reflects the rate of development of disease complexity of the HIV-1 quasispecies virus populations have a paradigm. These patterns of viral replication and measured relative rate of progression to disease can be explained by this paradigm. These patterns of viral replication and measured relative rate of progression to disease can be explained by this paradigm.

ACKNOWLEDGMENTS

We thank Mary Ann Dillon, Kevin Kunstman, and David McDonald for assistance with the data analysis. We thank Richard Steihm, Paul Krosgstad, Margaret Keller, and Audra Devkis for providing clinical samples and follow-up. This work was supported by Public Health Service grants HD-31756 and HD-306291 from the National Institute of Child Health and Human Development, R-143994-0 from a National Institutes of Health, and the Pediatric AIDS Foundation, and a gift from an anonymous foundation.

REFERENCES

2. AIDS Clinical Trials Group Virology Technology Advisory Committee. 1994. ACTG virology manual for HIV clinical trials. Division of AIDS, National Institute for Allergy and Infectious Diseases, Bethesda, Md.

TABLE 2. Wilcoxon rank sum test of the synonymous and nonsynonymous substitution scores

<table>
<thead>
<tr>
<th>Substitution type and child</th>
<th>No. of substitutions/wk</th>
<th>Minimum</th>
<th>Median</th>
<th>Mean</th>
<th>Maximum</th>
<th>z value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Synonymous</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0.0001958</td>
<td>0.0007882</td>
<td>0.0009964</td>
<td>2.1695</td>
<td>0.1050</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0.0001713</td>
<td>0.0007566</td>
<td>0.007566</td>
<td>2.9224</td>
<td>0.0017</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0.0000601</td>
<td>0.0005253</td>
<td>0.003713</td>
<td>-1.3758</td>
<td>0.1956</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>0.0001553</td>
<td>0.0005553</td>
<td>0.007605</td>
<td>-1.5384</td>
<td>0.0240</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>0.0003453</td>
<td>0.0005140</td>
<td>0.001870</td>
<td>-1.6143</td>
<td>0.0496</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>0</td>
<td>0.0003731</td>
<td>0.0011630</td>
<td>0.015340</td>
<td>-1.6116</td>
<td>0.0465</td>
<td></td>
</tr>
<tr>
<td><strong>Nonsynonymous</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0.0001150</td>
<td>0.0002528</td>
<td>0.001708</td>
<td>2.1820</td>
<td>0.0954</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0.0001184</td>
<td>0.0003882</td>
<td>0.002191</td>
<td>2.9534</td>
<td>0.00983</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0.0002315</td>
<td>0.0005901</td>
<td>0.002975</td>
<td>-1.6363</td>
<td>0.0676</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>0.0003473</td>
<td>0.0004425</td>
<td>0.002008</td>
<td>-1.5259</td>
<td>0.0635</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>0.0001991</td>
<td>0.0003597</td>
<td>0.0004500</td>
<td>0.001870</td>
<td>2.1695</td>
<td>0.0150</td>
<td></td>
</tr>
</tbody>
</table>

* The numbers of synonymous and nonsynonymous substitutions per week were calculated for each child and statistically compared using a Wilcoxon signed-rank test that accounts for the linkage of the synonymous and nonsynonymous scores for a given sequence compared to the first point consensus (78). The minimum, maximum, median, and mean values for the respective sample sets, as well as the calculated P value for the corresponding correction value (z), are provided. A one-sided test was used to test the hypothesis.


