No evidence of HIV replication in children on antiretroviral therapy

Gert U. Van Zyl,1 Mary Grace Katusiime,1 Ann Wiegand,2 William R. McManus,2 Michael J. Bale,2 Elias K. Halvas,2 Brian Luke,4 Valerie F. Boltz,2 Jonathan Spindler,2 Barbara Laughton,1 Susan Engelbrecht,1 John M. Coffin,6 Mark F. Cotton,3 Wei Shao,4 John W. Mellors,4 and Mary F. Kearney2

1Division of Medical Virology, Stellenbosch University and National Health Laboratory Service (NHLS) Tygerberg, Cape Town, South Africa. 2HIV Dynamics and Replication Program, National Cancer Institute (NCI), Frederick, Maryland, USA. 3Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. 4Advanced Biomedical Computing Center, Leidos Biomedical Research Inc. and Frederick National Laboratories for Cancer Research, Frederick, Maryland, USA. 5Department Pediatrics and Child Health, Tygerberg Children’s Hospital and Family Clinical Research Unit, Stellenbosch University, Cape Town, South Africa. 6Department of Molecular Biology and Microbiology, Tufts University, Boston Massachusetts, USA.

Introduction

Antiretroviral therapy (ART) is known to suppress viremia and prolong survival of HIV-1–infected individuals but does not eliminate the HIV reservoirs that cause viral rebound after treatment is stopped. There has been continued debate about whether the HIV reservoir is maintained through the persistence and proliferation of cells infected before the initiation of ART (1–5), or from ongoing viral replication in potential ART sanctuary sites, such as lymph nodes (LNs) (6–10), with subsequent trafficking of recently infected cells into the blood (10, 11). If ongoing replication in tissues maintains the HIV reservoir, then preventing infection of new cells by developing antiretrovirals that better penetrate sanctuary sites such as LNs would be a high priority. Conversely, if current ART is fully effective at blocking full cycles of viral replication in both tissues and blood, then the elimination of proliferating and long-lived infected cells would be the highest priority for achieving an HIV-1 cure. It is therefore critical that the efficacy of current ART be fully understood to identify the most appropriate curative strategy.

Prior studies addressing the question of ongoing replication have analyzed the evolution of virion RNA or proviral DNA sequences in chronically infected adults as a function of the duration of ART (3, 10, 12–15). Such studies have ascribed the success of ART to its ability to effectively prevent the infection of new cells by reducing the in vivo reproductive rate to less than 1 and effectively halting viral evolution and preventing the associated development of drug resistance (3, 16, 17). The prevention of new rounds of viral replication is thought to result in the rapid decay of short-lived, virus-producing cells and associated plasma viremia, followed by a gradual decay of longer-lived cells (18, 19). These studies, however, have not settled the controversy, mainly because the high baseline genetic diversity in chronically infected adults severely limits the ability to clearly detect evolutionary changes in viral sequences. We therefore turned our attention to perinatally infected infants who were started on ART when HIV diversity was low (20–24), making it simpler to clearly detect mutations that accumulate with viral replication.

To address the issue of ongoing viral replication in patients on current ART regimens, we compared single HIV p6, protease, and reverse transcriptase (p6-PR-RT) sequences obtained from plasma and/or peripheral blood mononuclear cells (PBMCs) near the time of ART initiation and after 7 to 9 years of ART in 10 children who initiated treatment during early infection and with low viral diversity. Eight of the children had viremia suppression on ART for the duration of the study, and two did not have viremia suppression for fifteen and thirty months and thus served as positive controls for HIV replication and evolution in vivo.
The Journal of Clinical Investigation

RESEARCH ARTICLE

Figure 1. Properties of HIV-1 infection in the 10 children studied. HIV sequence diversity was measured as the APD at the indicated time points. Divergence was measured using a test for panmixia, with a P value of less than 0.001 considered significant after correcting for multiple comparisons (8), and by measuring the root-to-tip distances of sequences on phylogenetic trees (longer branches over time indicate viral evolution). A 1-sample t test was used to determine if there was a significant change in APD on ART, and a 2-tailed F test was used to determine whether the root-to-tip slopes were significantly different from 0. All branches are shown as the p-distance in the NJ trees. Baseline HIV sequences were either from PBMC DNA (red triangles) or plasma RNA (red circles), and sequences for long-term ART were all from PBMC DNA (black triangles), sampled at the indicated time points. Blue brackets show sequences with evidence of HIV evolution. Plots to the left of the trees show the plasma HIV RNA viral load (VL) as a function of age. Red and black arrows on the plots indicate the time of collection of samples analyzed by SGS. Red vertical arrows indicate the patient’s age at ART initiation. White symbols indicate samples with an undetectable VL, plotted at the limit of sensitivity of the assay used. The apparent decline in the level of undetectable VL in some samples is due to a switch to an assay with a lower limit of detection. (A and B) Patients with detectable viremia for more than 1 year after the time of the first sample. (C–J) Patients with undetectable viremia for all samples taken after the first sample. The blue arrow indicates a sequence that was omitted in 1 sensitivity analysis, as discussed in the text. Results from a total of 20 samples are included in this figure.

Results

Children identified as HIV-1 RNA positive at birth and initiated on ART (Supplemental Table 1; supplemental material available online with this article; https://doi.org/10.1172/JCI94582DS1) as part of the Children with HIV Early Antiretroviral Therapy (CHER) trial (25) were studied for evidence of HIV-1 replication during ART. We used single-genome amplification and sequencing (SGS) to compare HIV p6-PRT-RT sequences obtained from plasma or PBMCs near the time of ART initiation (baseline) and after 7 to 9 years on ART (long-term ART) in 10 children (Supplemental Table 1). The fragment analyzed is 1.2 kb in length and covers the p6 region of gag as well as protease and the first 700 bp of reverse transcriptase in the pol gene. The sequences from the 10 children are shown in neighbor-joining (NJ), p-distance trees in Figure 1. Eight had undetectable plasma HIV-1 RNA throughout the period of observation and had initiated treatment by 10 months of age (median: 2.45 months; range: 1.8–9.9 months). Two children, serving as controls, did not have plasma viremia suppression on ART for fifteen to thirty months. To look for evidence of ongoing viral replication during ART, HIV populations in samples taken at baseline were compared with the populations present after long-term ART. Three measures of evolution were assessed for significant change over time: (a) HIV-1 genetic diversity, measured as the average pairwise distance (APD), from early infection to long-term infection; (b) genetic divergence from the founder virus(es) using a test for panmixia (3,12,26,27); and (c) root-to-tip distances in maximum-likelihood (ML) trees (3, 12, 27). ML trees were rooted in 2 ways: first, on an outgroup (HIV subtype C consensus; www.hiv.lanl.gov), and second, on the majority sequence of the baseline sample (Table 1).

HIV diversity and evolution in children with unsuppressed viremia. The HIV populations in all children had very low diversity (median APD of 0.20%; range, 0.04%–0.29%) at the first sampling time point (baseline) (Supplemental Table 1 and Table 1) (28), and most had a large rake of identical sequences, which was consistent with a single transmitted/founder virus (29, 30) and provided a low background on which to detect evolution (Figure 1, Figure 2, and Table 1). The 2 children with unsuppressed viremia for 15 and 30 months following the first sample (Figure 1, A and B; patient identifiers [PIDs] ZA001 and ZA002) showed clear evidence of HIV evolution: significant increases in viral diversity and a low probability of panmixia (Figure 1, A and B, and Table 1); significantly longer root-to-tip distances in ML trees (Table 1); and obvious clusters of sequences on longer branches in NJ trees (Figure 1, A and B); indicating the emergence of new viral variants over time. One of the two viremic children also had a population of sequences after eight years on ART that remained identical to the founder virus (Figure 1A; PID ZA001), suggesting that these proviruses are in long-lived, possibly proliferating, cells harboring the original transmitted variant.

HIV diversity and evolution in children on ART with suppressed viremia. The viral populations in the 8 children with fully suppressed viremia for the sampling interval were noticeably different from the viremic children. In all but 1 of the children, there was no evidence of evolution by any method applied (Figure 1, C–J, and Table 1), and none of the trees showed populations of divergent sequences on separate nodes or longer branches after long-term ART compared with those at baseline, consistent with a lack of virus replication and evolution during ART. One of the children (PID ZA003) did show an increase in viral diversity from 0.04% to 0.10% (P = 0.003) (10) over time, but no other test (panmixia or slope of root-to-tip distances in ML trees) showed results consistent with evolution (Figure 1C and Table 1). The NJ tree (Figure 1C) revealed that the difference in diversity in PID ZA003 between the early and late samples was attributable to 1 sequence on a longer branch (indicated with a blue arrow in Figure 1C). When this sequence was removed from the tree (not shown), the diversity difference became nonsignificant (0.04% vs. 0.03%; P = 0.6). The HIV DNA level in this child on long-term ART was very low (1.5 proviruses per million PBMCs); consequently, only 7 sequences were obtained, increasing the likelihood that a single sequence could unduly influence diversity estimates. Of note, all 7 sequences, including the variant on the longer branch, were included in all other analyses. Overall, the test for panmixia and analyses of root-to-tip distances showed no evidence of HIV evolution in any child whose viremia was fully suppressed on ART from the time of its initiation.

Persistence of HIV in cells infected prior to ART through cellular proliferation. Populations of identical proviral DNA sequences were observed after long-term ART in all of the children with suppressed viremia and may have been from the proliferation of infected cells during ART (1, 31). In all cases, some of the identical sequences present after 7 to 9 years of ART were exact matches to the baseline variants. Two children (PID ZA003 and ZA004) had almost no viral diversity at baseline (0.04% and 0.11% APD, respectively), were initiated on ART at approximately 2 months of age, and had HIV DNA levels of only 1.5 and 23.6 copies per million PBMCs after 8 years on ART. Also, the HIV sequences within the infected cells that persisted on ART were essentially identical to those at baseline (i.e., no nucleotide divergence was observed above the background for SGS [ref. 32]). These examples provide
<table>
<thead>
<tr>
<th>Participant group</th>
<th>PID</th>
<th>Sample time point and type</th>
<th>Age at sampling (mo)</th>
<th>Age at ART initiation (mo)</th>
<th>HIV diversity&lt;sup&gt;a&lt;/sup&gt;</th>
<th>P value for change in diversity&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Panmixia&lt;sup&gt;c&lt;/sup&gt;</th>
<th>ML root-to-tip slope (rooted on consensus C) (sub/site/mo)&lt;sup&gt;d&lt;/sup&gt; (×10&lt;sup&gt;–5&lt;/sup&gt;)</th>
<th>P value&lt;sup&gt;e&lt;/sup&gt; (rooted on consensus C)</th>
<th>ML root-to-tip slope (rooted on baseline majority) (sub/site/mo)&lt;sup&gt;d&lt;/sup&gt; (×10&lt;sup&gt;–5&lt;/sup&gt;)</th>
<th>P value&lt;sup&gt;e&lt;/sup&gt; (rooted on baseline majority)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not suppressed on ART for part of sampling period</td>
<td>ZA001</td>
<td>Baseline (PBMCs)</td>
<td>14.5</td>
<td>10.9</td>
<td>0.18%</td>
<td>&lt;0.0001</td>
<td>0.002</td>
<td>3.9</td>
<td>0.0008</td>
<td>3.9</td>
<td>0.0008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Long-term ART (PBMCs)</td>
<td>105.9</td>
<td></td>
<td>0.42%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ZA002</td>
<td>Baseline (plasma)</td>
<td>1.8</td>
<td>17.4</td>
<td>0.10%</td>
<td>&lt;0.0001</td>
<td>&lt;10&lt;sup&gt;–4&lt;/sup&gt;</td>
<td>2.3</td>
<td>0.002</td>
<td>3.3</td>
<td>7 × 10&lt;sup&gt;–7&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Long-term ART (PBMCs)</td>
<td>99.7</td>
<td></td>
<td>0.57%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continuous suppression on ART for duration of sampling period</td>
<td>ZA003</td>
<td>Baseline (plasma)</td>
<td>1.4</td>
<td>1.8</td>
<td>0.04%</td>
<td>0.003</td>
<td>0.3</td>
<td>2.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Long-term ART (PBMCs)</td>
<td>98.5</td>
<td></td>
<td>0.10%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ZA004</td>
<td>Baseline (PBMCs)</td>
<td>12.0</td>
<td>2.7</td>
<td>0.11%</td>
<td>0.6</td>
<td>0.5</td>
<td>−0.075</td>
<td>0.8</td>
<td>−0.035</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Long-term ART (PBMCs)</td>
<td>107.8</td>
<td></td>
<td>0.09%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ZA005</td>
<td>Baseline (PBMCs)</td>
<td>14.3</td>
<td>6.0</td>
<td>0.19%</td>
<td>0.6</td>
<td>0.8</td>
<td>0.29</td>
<td>0.7</td>
<td>0.15</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Long-term ART (PBMCs)</td>
<td>102.4</td>
<td></td>
<td>0.22%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ZA006</td>
<td>Baseline (PBMCs)</td>
<td>14.4</td>
<td>9.0</td>
<td>0.29%</td>
<td>0.6</td>
<td>0.06</td>
<td>0.021</td>
<td>1.0</td>
<td>−0.11</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Long-term ART (PBMCs)</td>
<td>98.3</td>
<td></td>
<td>0.25%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ZA007</td>
<td>Baseline (PBMCs)</td>
<td>15.1</td>
<td>9.9</td>
<td>0.25%</td>
<td>0.3</td>
<td>0.04</td>
<td>−0.55</td>
<td>0.3</td>
<td>−0.24</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Long-term ART (PBMCs)</td>
<td>108.6</td>
<td></td>
<td>0.18%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ZA008</td>
<td>Baseline (PBMCs)</td>
<td>26.1</td>
<td>2.2</td>
<td>0.21%</td>
<td>0.7</td>
<td>0.4</td>
<td>−0.13</td>
<td>0.8</td>
<td>−0.27</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Long-term ART (PBMCs)</td>
<td>83.4</td>
<td></td>
<td>0.19%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ZA009</td>
<td>Baseline (PBMCs)</td>
<td>7.7</td>
<td>2.0</td>
<td>0.24%</td>
<td>0.3</td>
<td>0.4</td>
<td>0.43</td>
<td>0.4</td>
<td>0.62</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Long-term ART(PBMCs)</td>
<td>111.5</td>
<td></td>
<td>0.33%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ZA010</td>
<td>Baseline (PBMCs)</td>
<td>7.3</td>
<td>1.8</td>
<td>0.21%</td>
<td>0.3</td>
<td>0.4</td>
<td>0.0062</td>
<td>1.0</td>
<td>0.63</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Long-term ART(PBMCs)</td>
<td>101.8</td>
<td></td>
<td>0.17%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>HIV population diversity measured by APD. <sup>b</sup>P value for increase in diversity after long-term ART from baseline (t-distribution with baseline APD variance estimated by 1,000 bootstrap repetitions [1-sample t test]). <sup>c</sup>Panmixia is a test for divergence (8). A P value of 10<sup>–3</sup> was considered significant, accounting for multiple comparisons of sequences obtained from each time point. <sup>d</sup>Consensus subtype C sequence from www.hiv.lanl.gov. <sup>e</sup>P values in bold indicate statistical significance. sub, substitutions.
the clearest evidence that HIV did not evolve for 8 years on ART and strongly imply that HIV replication is halted by ART in both the blood and LNs, since it is known that cells continually traffic from the lymphatics into the blood (10, 11). Overall, the analyses for all 8 children indicate that when ART suppresses viremia below the limit of detection of assays used clinically, viral replication is blocked, and only archival variants persist in cells, which may expand through cell proliferation, as has been observed in adults and children (2, 31).

Comparison of NJ and BEAST phylogenetic analyses. To compare our phylogenetic methods with those used by others (10), we also generated trees from the same data set using BEAST (Bayesian evolutionary analysis sampling trees) (http://beast.bio.ed.ac.uk/) (Figure 3). One example is shown in Figure 3B and is compared with a NJ tree of the same sequences (Figure 1D and Figure 3A). The sequences are numbered for cross-reference between all the trees in Figure 3. It is clear that the time-stamped Bayesian method in BEAST using a strict molecular clock (10) forces the appearance of evolution, even in cases in which the populations do not vary beyond the level of random variation expected from SGS sampling. In addition, identical HIV variants cannot be readily identified in trees generated by BEAST using such time-stamped parameters, because all sequences are assigned an “evolutionary distance” by the program, such that sequences sampled later are placed on separate nodes and longer branches than are identical sequences present at baseline (2, 31). This artifactual appearance of evolution occurs because a time stamp, associated with the sample collection date, is used in the algorithm to generate the tree. Thus, the “evolution” in the tree produced by BEAST is largely a result of artifacts from algorithmic assumptions that are not appropriate for visualizing intra-patient HIV evolution, for estimating HIV evolutionary rates in patients on ART, or for drawing conclusions regarding HIV replication during ART.

Simulated NJ tree after 8 years on ART using HIV evolutionary rate. To investigate further the accuracy of a previous report (10), we made NJ trees that contained both the baseline sequences from the same patient (PID ZA004) and the sequences that would be present in the blood after 8 years if evolution on ART had occurred at the rate recently reported (6.24 × 10^-4 substitutions/site/month) (10), with trafficking of newly infected cells from LNs to the blood, as was previously proposed (10, 11) (Figure 3C). Comparison with the actual NJ tree derived from the same sequences (Figure 1D and Figure 3A) revealed obvious topological differences (Figure 3C) and indicated that the evolutionary rate reported in ref. 10 was grossly overestimated. To show that our simulation model accurately reflects the accumulation of mutations in vivo, we performed the same simulation for the children (ZA001 and ZA002) who served as replication controls using evolutionary rates calculated from the change in the root-to-tip distances over time for these children. The simulated NJ trees for each control are shown in Supplemental Figure 1 and have topologies similar to those of the actual trees in Figure 1, A and B. The results of these analyses indicate that the more appropriate method for evaluating the population structure and divergence of HIV populations during ART is NJ tree construction, which does not make any assumptions regarding evolution or impose a time stamp, but only displays the absolute genetic differences among variants in the population.

Discussion
The failure to detect any significant change in the HIV population for 7 to 9 years in children on ART with sustained suppression of viremia provides strong evidence against ongoing replication of HIV-1 either in the blood or in sanctuary sites with subsequent trafficking of infected cells peripherally, as is expected (10). These findings directly refute those purported to show HIV evolution in patients on ART (10) and confirm that the HIV reservoir is likely maintained largely, if not solely, by the persistence and expansion of cells that were infected prior to the initiation of treatment in individuals on ART with sustained suppression of viremia. Our findings are consistent with those of another report that failed to find significant evidence for HIV evolution during ART in children by comparing plasma sequences at baseline with virus that was produced in the supernatants of cultured resting CD4+ T cells (33) after approximately 5 years on ART. The “slow” evolution reported (33) included substitutions that were not above the background from the ex vivo methods used (viral outgrowth followed by bulk PCR and cloning).

Using specimens collected from 2 children with known viral replication during the sampling interval as positive controls for detection of HIV evolution, we were able to demonstrate that HIV replication is readily detectable using measurements of diversity, divergence, and ML or NJ phylogenetics. Using these same methods, we showed that there was no evidence for ongoing cycles of HIV replication during the 7- to 9-year period of ART in children.
who have fully suppressed viremia. We also showed that the use of the time-stamped Bayesian phylogenetic method using a strict molecular clock applied in the program BEAST can generate the appearance of HIV evolution, even in sets of identical sequences and, therefore, is not an accurate method for addressing the question of ongoing HIV replication during ART. It is important to note that all our analyses were conducted on blood samples and, as such, does not replicate exactly the previous report (10). Thus, it is important to continue to investigate HIV diversity and evolution in LN samples, although it is widely accepted that the migration of cells from LNs to the blood would occur over a 7- to 9-year period (10, 11). It should also be noted that most proviruses in samples cells from LNs to the blood would occur over a 7- to 9-year period in LN samples, although it is widely accepted that the migration of 

Figure 3. Comparison of analytical approaches. (A) NJ tree of the sequences from a child (PID ZA004) who started treatment at less than 3 months of age and had full viremia suppression for 8.9 years (Supplemental Table 1). This tree is identical to that shown in Figure 1D for the same child. Red triangles indicate single-genome DNA sequences obtained 9 months after ART initiation. Black triangles indicate single-genome DNA sequences obtained 8 years later. The distance tree shows that the populations did not shift for 8 years on ART; populations at both time points were almost completely homogeneous, with only 8 nucleotide differences in 16,800 bases at baseline and 9 nucleotide differences in 19,200 bases 8 years later. The founder sequences persisted across the sampling time points. (B) The same sequences analyzed by BEAST using a strict molecular clock (as in ref. 10) show divergence over time, despite the identity of most sequences at both time points. Individual sequences are numbered identically in the 2 trees shown in A and B. (C) Simulated NJ distance tree using the estimated HIV evolution rate of 6.24 × 10–4 substitutions/site/month from ref. 10, applying the baseline sequences from the same child as in A and B. The simulated tree shows artifactual evolution, while the actual tree shows no evidence of sequence divergence after 8 years of ART.

Our earlier data and those of others on adults (13–17), combined with the current data on children, are inconsistent with the idea of ongoing viral replication as a mechanism of HIV reservoir maintenance and replenishment in the majority of individuals with sustained suppression of viremia on ART. Consequently, the development of newer antiretroviral agents to better penetrate putative sanctuaries and block ongoing viral replication does not appear to be critical for curing HIV-1 infection. Rather, efforts should continue to focus on the exposure and elimination of immunologically hidden reservoir cells that persist in individuals on ART.

Methods

**HIV-1 quantification and genetic analyses.** HIV plasma RNA levels were determined using the Roche Amplicor HIV Monitor assay, version 1.0 (Roche Diagnostics), or the Abbott Diagnostics RealTime HIV-1
assay (Abbott Laboratories). HIV DNA levels were determined using the integrase cell-associated DNA (iCAD) assay (42). SGS of a portion of p6-RT-RT was performed using plasma virus RNA or PBMC DNA as previously described (3, 29, 32, 42, 43), with modifications for amplification of subtype C templates using the following PCR primers: 1849+ (5′-GAGTGTTGGCTGAGGCAATGAG-3′) and 3410–
5′-CAGTTAGTGTACTATGCTTGTTAGTCTT-3′). Sequences were aligned using ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Sequences that were obviously defective for viral replication, including those that contained stop codons due to hypermutation, were omitted from downstream analyses. Population genetic diversity was calculated as the APD using Molecular Evolutionary Genetics Analysis, version 6.0 (MEGA6) (http://www.megasoftware.net). Shifts in population structure were calculated using a subdivision test for panmixia, with a P value of less than 0.01 as the significance cutoff (as described by the original report and other publications) to account for the high number of comparisons between sequences and nucleotide sites (3, 12, 26). The test was derived from a geographic population structure test proposed by Hudson et al. (26). NJ trees were constructed using MEGA6.

ML tree construction and root-to-tip length calculation. PAUP4.0 (http://paup.sc.fsu.edu/) was used to construct ML trees. The model used in the tree construction was GTR+1+F. Either HIV-1 consensus C or the majority sequence from the baseline of the first time-point sample sequences were set as the outgroup. Root-to-tip length was calculated on the basis of the ML trees with TreeStat, version 1.6.2, in the BEAST package (http://beast.bio.ed.ac.uk/).

BEAST analysis. BEAST, version 1.6.2 (http://beast.bio.ed.ac.uk/), was used to perform time-stamped Bayesian phylogenetic analysis using a strict molecular clock (10). HKY-gamma was used as substitution model. All Markov Chain Monte Carlo (MCMC) chains were run at least 100 million steps to ensure good convergence of the parameters. Tracer, version 1.5, was used to evaluate the quality of BEAST runs (http://beast.bio.ed.ac.uk/tracer).

Simulation of in vivo evolutionary models on sequences. To evaluate the accuracy of the evolutionary rate reported in ref. 10, an in-house program was generated using the reported evolutionary rate in substitutions/site/month and the length of time between the sampling in 1 child (ZA004) to simulate the expected accumulation of diversity that would exist after approximately 8 years on ART if the reported evolutionary rate is correct. Given our observations in the 2 replication controls, the model allowed 50% of the introduced mutations to be randomly distributed, while 50% were limited to mutations at specific sites (reflecting positive selection within cytotoxic T lymphocyte [CTL] epitopes, as observed in the controls). The number of mutations per sequence was determined via a Poisson distribution. The simulations were generated by applying the model on the consensus sequence of the early time point. This model assumes that all mutations at the random sites are equally likely and makes no assumptions about the relative frequencies of transversions versus transitions. The code is written in C++ and is available from the authors upon request.

Statistics. The test for panmixia is described in Achaiz et al. and in Kearney et al. (3, 26). A P value significance cutoff of less than 0.01 was used to account for multiple comparisons as previously described (3). A significant change in APD over time was determined using a 2-tailed, 1-sample t test, where the assumed mean was the APD of the early time point and the sample distribution was the APD of the late time point, with variance estimation performed using 1,000 bootstrap replicates. Significance for the root-to-tip distance analysis was determined by linear regression. The variance of the observed points about this line was used to calculate the P value using a 2-tailed F test based on the null hypothesis that the slope of the line was zero.

Study approval. The CHER trial is registered with ClinicalTrials.gov (NCT00102960). Guardians of all donors provided written informed consent, and the study was approved by the IRB of Stellenbosch University.

Author contributions
MGK, AW, EKH, VFB, and JS performed RNA and DNA SGS, quantified DNA levels, analyzed data, and edited the manuscript. WS, WRM, MJB, and BL analyzed data. BL, SE, and MFC conceived the research and wrote the manuscript. JMC, JWM, MFK, and GUZV conceived the research, analyzed data, and wrote the manuscript.

Acknowledgments
We thank Connie Kinna, Valerie Turnquist, and Susan Toms (HIV Dynamics and Replication Program, NCI, Frederick, MD, USA) for administrative support, and Joseph Meyer (Scientific Publications, Graphics, & Media, NCI at Frederick, Frederick, MD, USA) for graphics support. Support was provide by the NCI (U50–South Africa Program for Collaborative Biomedical Research; 1U01CA200441-01); the National Institute of Mental Health (R01MH105134-01); the South Africa National Research Foundation (CPRR14080184941); the South African Polio Research Foundation; a Leidos contract (HHSN261200800001E) awarded to the University of Pittsburgh; South African MRC Collaborating Centre Funding; and NIH intramural funding to the NCI. JMC was a Research Professor of the American Cancer Society.

Address correspondence to: Mary Kearney, HIV Replication and Dynamics Program, National Cancer Institute at Frederick, 1050 Boyles Street, Building 535, Room 108D, Frederick, Maryland 21702, USA. Phone: 301.846.6796; Email: kearnemym@mail.nih.gov.


