HIV Proviral Burden, Genetic Diversity, and Dynamics in Viremic Controllers Who Subsequently Initiated Suppressive Antiretroviral Therapy

F. Harrison Omondi,a,b Hanwei Sudderuddin,b Aniqa Shahid,a,b Natalie N. Kinloch,a,b Bradley R. Jones,b,c Rachel L. Miller,b,c Olivia Tsai,a Daniel MacMillan,b Alicja Trocha,d Mark A. Brockman,a,e Chanson J. Brumme,b,f Jeffrey B. Joy,b,c,f Richard Liang,b Bruce D. Walker,d Zabrina L. Brummea,b

aFaculty of Health Sciences, Simon Fraser University, Burnaby, British Columbia, Canada
bBritish Columbia Centre for Excellence in HIV/AIDS, Vancouver, British Columbia, Canada
cBioinformatics Program, University of British Columbia, Vancouver, British Columbia, Canada
dRagon Institute of MGH, MIT and Harvard, Cambridge, Massachusetts, USA
eDepartment of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, Canada
fDepartment of Medicine, University of British Columbia, Vancouver, British Columbia, Canada

F. Harrison Omondi and Hanwei Sudderuddin contributed equally to this article. F. Harrison Omondi is listed first as he wrote the first draft of the manuscript.

ABSTRACT Curing HIV will require eliminating the reservoir of integrated, replication-competent proviruses that persist despite antiretroviral therapy (ART). Understanding the burden, genetic diversity, and longevity of persisting proviruses in diverse individuals with HIV is critical to this goal, but these characteristics remain understudied in some groups. Among them are viremic controllers—individuals who naturally suppress HIV to low levels but for whom therapy is nevertheless recommended. We reconstructed within-host HIV evolutionary histories from longitudinal single-genome amplified viral sequences in four viremic controllers who eventually initiated ART and used this information to characterize the age and diversity of proviruses persisting on therapy. We further leveraged these within-host proviral age distributions to estimate rates of proviral turnover prior to ART. This is an important yet understudied metric, since pre-ART proviral turnover dictates reservoir composition at ART initiation (and thereafter), which is when curative interventions, once developed, would be administered. Despite natural viremic control, all participants displayed significant within-host HIV evolution pretherapy, where overall on-ART proviral burden and diversity broadly reflected the extent of viral replication and diversity pre-ART. Consistent with recent studies of noncontrollers, the proviral pools of two participants were skewed toward sequences that integrated near ART initiation, suggesting dynamic proviral turnover during untreated infection. In contrast, proviruses recovered from the other two participants dated to time points that were more evenly spread throughout infection, suggesting slow or negligible proviral decay following deposition. HIV cure strategies will need to overcome within-host proviral diversity, even in individuals who naturally controlled HIV replication before therapy.

IMPORTANCE HIV therapy is lifelong because integrated, replication-competent viral copies persist within long-lived cells. To cure HIV, we need to understand when these viral reservoirs form, how large and genetically diverse they are, and how long they endure. Elite controllers—individuals who naturally suppress HIV to undetectable levels—are being intensely studied as models of HIV remission, but viremic controllers, individuals who naturally suppress HIV to low levels, remain understudied even though they too may hold valuable insights. We combined phylogenetics and mathematical modeling to reconstruct proviral seeding and decay from infection to therapy-mediated suppression in four viremic controllers. We recovered diverse

Citation Omondi FH, Sudderuddin H, Shahid A, Kinloch NN, Jones BR, Miller RL, Tsai O, MacMillan D, Trocha A, Brockman MA, Brumme CJ, Joy JB, Liang R, Walker BD, Brumme ZL. 2021. HIV proviral burden, genetic diversity, and dynamics in viremic controllers who subsequently initiated suppressive antiretroviral therapy. mBio 12:e02490-21. https://doi.org/10.1128/mBio.02490-21.

Invited Editor Mary F. Kearney, National Cancer Institute at Frederick.
Editor Thomas E. Smithgall, University of Pittsburgh School of Medicine
Copyright © 2021 Omondi et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.
Address correspondence to Zabrina L. Brumme, zbrumme@sfu.ca.
Received 20 August 2021
Accepted 8 October 2021
Published 16 November 2021
proviruses persisting during therapy that broadly reflected HIV’s within-host evolutionary history, where the estimated half-lives of the persistent proviral pool during untreated infection ranged from <1 year to negligible. Cure strategies will need to contend with proviral diversity and between-host heterogeneity, even in individuals who naturally control HIV.

KEYWORDS genetic diversity, proviral burden and dynamics, proviral half-life, viremic controllers, human immunodeficiency virus

Like all retroviruses, HIV integrates its genome into that of its host cell. Most infected cells die—or are eliminated by the immune system—within a day or two of infection, but a small number persist long-term, even during suppressive antiretroviral therapy (ART) (1–3). While most of these long-lived cells harbor HIV proviruses with large deletions or other genetic defects (4–6), a minority harbor replication-competent proviruses that could reactivate at any time. It is for this reason that ART needs to be maintained for life. Achieving ART-free HIV remission (“functional cure”) or a complete cure (eradication) (7) will therefore require permanently inactivating (8, 9) or eliminating (10, 11) these HIV reservoirs, respectively. Characterizing the burden, genetic diversity, and longevity of the proviruses that persist within the host can advance us toward these goals.

HIV elite controllers—rare individuals who spontaneously control viremia to undetectable levels without ART—are being intensely studied as natural models of HIV remission (12). They may also represent a group in which a complete cure might be more easily achieved (13, 14), in part because their HIV reservoirs are smaller and less genetically diverse (8, 12, 15). Proviral burden and diversity, however, are less well characterized in viremic controllers—individuals who naturally suppress HIV to low levels (normally defined as <2,000 HIV RNA copies/ml in plasma) (16, 17), but who are nevertheless recommended to initiate ART under current guidelines (18). In particular, this group could yield insights into persisting proviral composition in the context of natural yet incomplete HIV control.

More broadly, our understanding of within-host proviral dynamics (i.e., the rates of proviral deposition and subsequent decay) during untreated infection remains incomplete for HIV controllers in general. This is in part because longitudinal studies dating back to infection are rare in this group, and because it is challenging to isolate HIV from samples with low or undetectable viral loads. Understanding these dynamics is nevertheless important because proviral longevity during untreated infection dictates reservoir composition at ART initiation (and thereafter): if the persisting proviral pool turned over slowly pre-ART, then HIV sequences seeded into it during early infection would have a high likelihood of persisting for long periods, but if the pool turned over rapidly, its composition would shift toward recently circulating HIV sequences. As cure or remission interventions would only be administered during suppressive ART, it is critical to understand the dynamics that shape the proviral pool up to this point.

Our understanding of within-host proviral dynamics has recently been enriched by the development of phylogenetic approaches to analyze persisting proviral diversity in context of HIV’s within-host evolution prior to ART (19–23). A recent study that recovered replication-competent HIV sequences during therapy revealed that the majority represented viral variants that circulated close to therapy initiation, though more ancestral sequences, some dating back to transmission, were also recovered (21). Studies employing similar approaches to “date” the overall pool of proviruses persisting on ART, which comprise both genetically intact and defective sequences, have confirmed that these often span the individual’s whole pre-ART history, where individuals differ in the extent to which their proviral pool is skewed toward sequences that integrated around ART initiation (19, 20, 22, 23). Because viremic controllers naturally limit viral replication to low levels, one could hypothesize that their persisting proviral pools may be smaller and less diverse than those of noncontrollers, which could potentially make
them more responsive to remission or cure interventions, but on-ART proviral diversity has not been investigated in this group in the context of HIV’s full within-host evolutionary history.

The age distributions of proviruses sampled shortly after ART can also be used to calculate within-host rates of proviral turnover during untreated infection (19, 23): this is because, at the time of sampling, the majority of proviral turnover would have already occurred prior to therapy. Recent data from noncontrollers suggest that proviral half-lives during untreated infection average less than a year (19, 24), although another study estimated an average of 25 months (23). Regardless, these estimates are far shorter than the published rates of proviral decay on ART, which are estimated as ~4 years for the replication-competent reservoir (25) and >10 years for the proviral pool at large (26). This has led to the hypothesis that ART dramatically slows the rate of proviral turnover in vivo, thereby stabilizing the proviral pool for long-term persistence (21, 27, 28). Pre-ART proviral dynamics, however, have not been investigated in HIV controllers. To address these knowledge gaps, we combined single-genome sequencing, proviral quantification, phylogenetic analysis, and mathematical modeling to characterize proviral burden, diversity, and dynamics in four viremic controllers from HIV infection to therapy-mediated suppression.

RESULTS

Participant characteristics and sampling. We studied four viremic controllers. Three broadly maintained plasma viral loads (pVL) of <2,000 copies of HIV RNA/ml during untreated infection (participants 1, 2, and 3), while one maintained pVL of <2,000 copies of HIV RNA/ml for 3 years before losing control (participant 4). Participants initiated ART a mean of 6.1 years (range, 3.8 to 8.4 years) following their estimated date of infection (EDI). An average of 13 pre-ART plasma samples per participant (range, 10 to 17 samples), which spanned a mean period of 4.5 years (range, 2.5 to 6.7 years) prior to therapy, were available for HIV RNA sequencing, where the first was sampled a mean of 20.5 months (range, 15 to 26 months) following the EDI (Fig. 1; Table 1). In addition, peripheral blood mononuclear cells (PBMC) were available at two time points on suppressive ART (10 million cells per time point). Due to limited cell numbers, the first PBMC sample, taken an average of 18.7 months (range, 9 to 38 months) after ART, was
| Participant | Clinically estimated infection date | Diagnosis date | First plasma viral load, copies/ml (log_{10}) | No. of plasma HIV sequences | ART initiation date | Total proviral burden on ART (% genomically intact) | Proviral sequencing date(s) | No. of proviral sequences (distinct: no.; %) |
|-------------|--------------------------------------|----------------|-----------------------------------------------|----------------------------|-------------------|-----------------------------------------------|-----------------------------|------------------------------------------|
| 1           | Apr 2006                             | Feb 2007       | 1,615 (3.21)                                 | 8                          | 1 Aug 2010       | 215 (51)                                     | 13 Sept 2012                | 48 (40; 83)                              |
| 2           | Dec 2005                             | Jan 2006       | 977 (2.99)                                   | 9                          | 1 May 2014       | 580 (10)                                     | 8 Sept 2015                 | 66 (36; 55)                              |
| 3           | Mar 2008                             | Jun 2008       | 97 (1.99)                                    | 7                          | 1 Jan 2012       | 44 (31)                                      | 14 Jul 2016                 | 24 (12; 50)                              |
| 4           | Mar 2005                             | Nov 2005       | 383 (2.58)                                   | 12                         | 1 Feb 2013       | 778 (94)                                     | 18 Nov 2013                 | 1 full genome                            |
|             |                                      |                |                                               |                            |                   |                                               |                             | 20 Oct 2014                 | 128 (119; 92)                            |

*First detectable sampled plasma viral load, expressed as HIV RNA copies/ml plasma.

*Total number of plasma time points from which HIV RNA nef sequences were successfully amplified.

*"Distinct" refers to sequences that were observed only once in that compartment. Reported numbers exclude defective, hypermutated, and putative within-host recombinant HIV sequences.

*As measured by the intact proviral DNA assay and expressed as HIV copies/10^6 CD4+ T cells.
used to quantify total and genomically intact proviral burden in CD4\(^+\) T cells (29). The second, taken an average of 28.8 months (range, 16 to 54 months) after ART, was used for proviral sequencing and integration date inference (20).

**Quantifying total and genetically intact proviral burden.** We quantified total and genomically intact proviral burden in CD4\(^+\) T cells using the intact proviral DNA assay (IPDA). This is a duplexed droplet digital PCR (ddPCR) assay that targets two HIV regions, the packaging signal (\(\Psi\)) near the 5’ end of the viral genome and the Rev responsive element (RRE) within envelope (\(env\)), which together have high predictive power to distinguish genomically intact proviruses from those with large deletions and/or extensive hypermutation, which typically dominate in vivo (29). Proviruses yielding double (\(\Psi\) and \(env\)) positive signals are inferred to be genetically intact, while those yielding only single positive signal are classified as defective.

We observed marked interindividual differences in total proviral burden. Participant 3 harbored the fewest proviruses (44 HIV DNA copies/10\(^6\) CD4\(^+\) T cells), while participant 4, the individual who eventually lost control, harbored the most (778 HIV DNA copies/10\(^6\) CD4\(^+\) T cells) (Fig. 1; Table 1). The percentages of genomically intact proviruses also differed markedly, with participant 2 harboring the lowest (10% intact proviruses) and participant 4 the highest (94% intact). The latter is remarkable though not without precedent (30). Participant 4’s large overall and intact proviral burden allowed us to isolate one near-full-length intact provirus from this time point (Fig. 1) for integration date inference.

**Single-genome plasma HIV RNA and proviral sequencing.** We characterized HIV RNA nef sequences from longitudinal pre-ART plasma and proviral sequences from the second on-ART PBMC sample, by single-genome amplification. HIV nef was sequenced because of its richness in phylogenetic signal despite its relatively short length (a shorter amplicon was critical as viral loads were low for most samples), because nef is representative of within-host HIV diversity and evolution relative to the rest of the viral genome (20), and because it is the gene most likely to be intact in proviruses persisting on ART (29, 31). Despite low viral loads, we isolated 546 HIV RNA nef sequences from pre-ART plasma (range, 67 to 245 sequences/participant), where an average of 66.2% of these (range, 50.0 to 89.6%) were distinct within each participant (Table 1). We further isolated 267 intact proviral nef sequences during suppressive ART (range, 24 to 129 sequences/participant), where an average of 69.8% (range, 50.0 to 91.5%) were distinct. All participants had HIV subtype B, and their sequences were monophyletic, with no evidence of coinfection or superinfection (Fig. 2). There were multiple instances where we recovered a proviral sequence on ART that was identical to a pre-ART plasma sequence, consistent with the long-term persistence of HIV sequences in vivo.

**Within-host HIV evolutionary reconstruction and proviral dating.** We next characterized the diversity of proviruses persisting on ART and inferred their ages phylogenetically (20). For each participant, we inferred a maximum likelihood phylogeny relating their plasma HIV RNA and proviral sequences, where the root represented the inferred transmitted founder event. We then fit a linear model relating the root-to-tip distances of distinct pre-ART plasma HIV RNA sequences to their collection dates, where the slope of this line represents the within-host HIV nef evolutionary rate under a strict molecular clock assumption, and the x-intercept represents the phylogenetically estimated infection date. We observed statistically significant within-host HIV evolution in plasma in all participants pre-ART, as defined by increasing divergence from the root, where the 95% confidence interval (CI) of the phylogenetically estimated infection date captured the participant’s clinically estimated one in all cases (Table 2). We next describe each participant’s results in detail.

Participant 1 (EDI, April 2006) maintained viremic control, except for two plasma viral load measurements of 8,075 (3.9 log\(_{10}\)) and 2,685 (3.4 log\(_{10}\)) copies/ml in February and October 2008, respectively, prior to initiating ART in August 2010 (Fig. 3A). Their proviral load on suppressive ART was 215 HIV copies/10\(^6\) CD4\(^+\) T cells, with 51% genetically intact (Fig. 1; Table 1). We isolated 104 intact HIV RNA nef sequences from 8 pre-ART plasma time points spanning 2.4 years, of which 52 (50%) were distinct, and 48
proviral nef sequences sampled approximately 2 years post-ART, of which 40 (83%) were distinct (Table 1). The recovery of identical proviral sequences is consistent with clonal expansion of CD4+ T cells harboring integrated HIV (32–34), though we cannot definitively identify these as clonal since we only performed subgenomic HIV sequencing. Within-host phylogenetic analysis revealed increasing plasma HIV RNA divergence from the inferred root over time, along with nonsynonymous substitutions that typify within-host HIV evolution (Fig. 3B). Proviruses sampled on ART were interspersed throughout the phylogeny, except within the clade closest to the root, which comprised the earliest sampled plasma HIV RNA sequences. The linear model relating the root-to-tip distances of distinct pre-ART plasma HIV sequences to their collection dates yielded a pre-ART nef evolutionary rate of $3.22 \times 10^{-5}$ substitutions per nucleotide site per day (Fig. 3C; Table 2). Based on this, the oldest sampled provirus was estimated to

| Participant | $\Delta$AIC | Clinically estimated infection date | Phylogenetically estimated root date (95% CI) | Within-host HIV evolutionary rate$^b$ |
|-------------|--------------|------------------------------------|-----------------------------------------------|--------------------------------------|
| 1           | 21.1         | Apr 2006                            | Feb 2007 (Dec 2004 to Apr 2009)               | $3.22 \times 10^{-5}$                |
| 2           | 37.8         | Dec 2005                            | Jan 2003 (Feb 1999 to Dec 2006)              | $1.17 \times 10^{-5}$                |
| 3           | 39.2         | Mar 2008                            | Oct 2006 (May 2004 to Feb 2009)              | $1.16 \times 10^{-5}$                |
| 4           | 30.3         | Mar 2005                            | June 2005 (Nov 2003 to Jan 2007)             | $1.28 \times 10^{-5}$ (control era)  |
| 167.2       |              |                                    |                                               | $5.35 \times 10^{-5}$ (post-control era) |

$^a$AIC, Akaike information criterion. A $\Delta$AIC of $\approx 10$ was considered evidence of a molecular clock signal (see Materials and Methods).

$^b$Expressed in estimated substitutions per nucleotide site per day.
FIG 3 Participant 1. (A) Clinical history and sampling timeline. Throughout all figures, circles denote plasma HIV RNA sampling and diamonds denote HIV DNA sampling. Shading represents ART. Undetectable viral loads are shown as 20 (1.3 log_{10}) copies/ml. (B) Maximum (Continued on next page)
have integrated in April 2008, 4 years prior to sampling (Fig. 3D). Though the 95% CIs around the proviral date estimates are wide, these data nevertheless suggest that proviruses seeded throughout infection persisted during ART.

Participant 2 (EDI, December 2005) maintained pVL of <2,000 (3.3 log_{10}) copies/ml, except for four measurements slightly above this: one in June 2010 and three in mid/late 2013. The participant initiated ART in May 2014 (Fig. 4A). Their total proviral load on ART was 580 HIV copies/10^6 CD4^+ T cells, with only 10% intact, the smallest intact proportion of all participants (Fig. 1; Table 1). We recovered 67 HIV RNA nef sequences from 9 pre-ART plasma time points spanning more than 4 years, of which 60 (90%) were distinct, and 66 intact proviral nef sequences after 1.2 years on ART, of which 36 (55%) were distinct. The sequences closest to the root of this participant’s phylogeny were proviruses sampled on ART, consistent with these having integrated shortly after transmission (Fig. 4B). Proviruses also interspersed throughout almost all descendant clades. Based on this individual’s calculated pre-ART nef evolutionary rate of 1.17 × 10^{-5} substitutions per nucleotide site per day, proviruses sampled on ART were inferred to have integrated on dates that spanned the individual’s entire pre-ART history (Fig. 4D). One particular proviral sequence, recovered 26 times, was estimated to have integrated 6 years prior to sampling.

Participant 3 (EDI, March 2008) continuously maintained pVL of <1,000 copies/ml before initiating ART in January 2012. Thereafter, pVL remained suppressed, except during a treatment interruption when a pVL of 1,610 (3.2 log_{10}) copies/ml was recorded (November 2013), but unfortunately no plasma was available from that event from which to isolate HIV sequences. This participant had the smallest proviral burden of all (44 HIV copies/10^6 CD4^+ T cells), with an estimated 31% genetically intact (Fig. 1; Table 1). We recovered 130 pre-ART plasma HIV nef sequences (55% distinct) from 7 pre-ART time points, where early plasma time points more frequently yielded identical HIV sequences, consistent with limited viral diversity following infection. We recovered only 24 proviral sequences (50% distinct) on ART, consistent with the small proviral burden. Despite maintenance of pVL of <1,000 copies/ml pre-ART, the within-host phylogeny displayed significant molecular clock signal (1.16 × 10^{-5} substitutions per nucleotide site per day), where similar to participant 2, proviruses sampled on ART interspersed throughout the tree and represented those closest to the root (Fig. 5B and C). The earliest proviruses dated to around transmission, while the latest dated to the viremic episode that occurred during the treatment interruption; the narrower 95% confidence intervals around these point estimates allow us to say with more certainty that this individual’s proviral pool spans a wide age range (Fig. 5D). One sequence, recovered 9 times on ART and representing 38% of recovered proviruses in this individual, dated to around ART initiation.

Participant 4 (EDI, March 2005) initially controlled pVL to <1,000 log_{10} copies/ml for 3 years, but afterwards lost control, where pVL reached a maximum of 65,500 (4.8 log_{10}) copies/ml before initiating ART in February 2013 (Fig. 6A). Two linear models were calibrated, with one each for the “control” and “post-control” periods, where the former was used to phylogenetically verify the clinically estimated infection date. We recovered 245 plasma HIV nef sequences from 12 pre-ART time points (71% distinct),

FIG 3 Legend (Continued)
likelihood within-host phylogeny and corresponding amino acid highlighter plot. In the phylogeny, colored circles denote distinct pre-ART plasma HIV RNA sequences and red diamonds indicate distinct proviral sequences sampled during suppressive ART. Sequences that were recovered repeatedly are shown as open gray circles (for plasma HIV RNA) and diamonds (for proviruses) adjacent to the relevant tip. The root represents the inferred most recent common ancestor of the data set, representing the phylogenetically inferred transmitted founder virus event. The highlighter plot is ordered according to the phylogeny and depicts amino acid substitutions with respect to the reference. (C) HIV sequence divergence-versus-time plot. The blue dashed line represents the linear model relating the root-to-tip distances of distinct pre-ART plasma HIV RNA sequences (colored circles) to their sampling times. This model is then used to convert the root-to-tip distances of distinct proviral sequences sampled during ART (red diamonds) to their original integration dates. The slope of the regression line, which represents the inferred within-host evolutionary rate (ER) in estimated substitutions per nucleotide site per day, is shown at the bottom right. Faint gray lines denote the ancestral relationships between HIV sequences. (D) Integration date point estimates (and 95% confidence intervals) for distinct proviral sequences recovered from this participant.
FIG 4 Participant 2. The panels are as described in the legend to Fig. 3.
FIG 5  Participant 3. The panels are as described in the legend to Fig. 3.
FIG 6  Participant 4. The panels are as described in the legend of Fig. 3, with the following additions. In panels B and D, the red diamond indicated by an asterisk represents the genomically intact HIV sequence isolated from the 2013 sample. In panel C, two linear models were (Continued on next page)
where, like participant 3, identical sequences were most frequently recovered from early time points (Fig. 6B). We also recovered 129 proviral sequences during ART, a remarkable 92% of which were unique, indicating substantial proviral diversity in this individual. The linear model inferred from the viremic control period yielded strong molecular clock signal despite low viremia during this time, but no proviral sequences were recovered that interspersed with these very early sequences (Fig. 6B and C). All recovered proviruses were therefore “dated” using the linear model representing the post-control period, yielding integration dates spanning this whole period (Fig. 6C). Of note, the near-full-length genetically intact provirus isolated during ART dated to 2011.

Correlates of proviral burden and diversity during ART. Early ART limits HIV reservoir size (35–37), and positive correlations have been observed between the duration of uncontrolled viremia and both HIV reservoir size and diversity in noncontrollers (35, 37–40). We therefore investigated the relationship between proviral burden, HIV diversity, and cumulative viral load in our cohort. We observed a strong correlation between overall HIV genetic diversity in plasma pre-ART and proviral diversity on ART (Spearman’s $r = 1$; $P = 0.08$) (Fig. 7A). Furthermore, cumulative viral load, measured as log$_{10}$ viremia copy days during untreated infection, correlated strongly with both total proviral burden as measured by the IPDA and proviral diversity on ART (both Spearman’s $r = 1$; $P = 0.08$) (Fig. 7B and C).

Inferring proviral turnover pre-ART. (i) Mathematical modeling approach. The decay rates (half-lives) of the proviral pool during untreated infection can be inferred from the age distributions of proviruses sampled on ART; this is because at the time of proviral sampling, the bulk of the proviral turnover had already occurred prior to therapy (19, 23, 24). We estimated pre-ART proviral half-lives in two ways: by adapting an existing mathematical model and by estimating these rates directly from the data.

To do the former, we modified a dynamical mathematical model of HIV infection that describes within-host cell and virus concentrations over time within active and latent compartments (see Materials and Methods and reference 23). The model assumes that HIV sequences enter the reservoir at a rate proportional to their abundance in plasma, allowing us to model within-host proviral deposition in a personalized way. As our viral

FIG 6 Legend (Continued)

fit to the data, encompassing the viremic control (blue dashed line) and the “loss-of-control” (green dashed line) periods. Corresponding evolutionary rates are shown in the bottom right corner in matching colors. The regression for the control period was performed using plasma time points 14 June 2006 (i.e., 2006-06-14) to 25 September 2008, while post-control period regression was performed using plasma time points 29 January 2009 to 14 May 2012.
load data did not capture acute-phase dynamics, we merged each participant’s available data with acute-phase dynamics estimated from the (limited) literature on HIV controllers (41–43) (see Materials and Methods and see Table S1 in the supplemental material) to model their proviral deposition. In a subsequent step, we then allowed these proviruses to decay at various constant, exponential rates up to the participant’s sampling date, yielding predictions of what the proviral age distribution would be, at time of sampling, for each decay rate tested.

**FIG 8** Pre-ART proviral decay rate estimates. (A to D) Viral load histories, with engrafted “peak viremia” kinetics from the literature (shown as black diamonds), for participants 1 (A), 2 (B), 3 (C, where the dotted line represents the viremic episode after ART that serves as the reference point in this analysis), and 4 (D). (E to H) Phylogenetically determined proviral ages and predicted proviral age distributions under different rates of proviral decay for participants 1 to 4. Blue histograms denote the proportion of each participant’s distinct proviruses that dated to each year prior to ART initiation, data that are derived from the integration date point estimates in Fig. 3D, 4D, 5D, and 6D, respectively. The gray and dashed black lines predict what the proviral age distributions would be at the time of sampling had proviral decay subsequently occurred under half-lives of 140 or 44 months following deposition (these half-lives represent published on-ART estimates of total DNA [26] and replication-competent reservoir [25] decay, respectively). The green line represents the model-predicted proviral age distributions under the pre-ART proviral decay rate that best fit each participant’s observed data. (I to L) Best-fit half-lives estimated directly from each participant’s observed proviral age distributions using a Poisson generalized linear model (red line) along with 95% confidence intervals (dotted lines).
The participants’ reconstructed viral load dynamics are shown in Fig. 8A to D, while the model-predicted proviral compositions under different decay rates, depicted as the proportion of remaining proviruses that date to each year prior to ART (or in the case of participant 3, to the viremic episode that occurred during treatment interruption), are shown in Fig. 8E to H. For participants 1 to 3, the model predicts that the bulk of the reservoir would have been deposited in the first year of infection (Fig. 8E to G, yellow lines); this is because the estimated acute-phase peak viral load is nearly 2 log10 higher than the subsequent setpoint. For participant 4, the model predicts that the bulk of the reservoir would have been deposited during the post-control period, because viral load was sustained at high levels during that time (Fig. 8H, yellow line).

If we allow these deposited proviruses to decay at half-lives of 140 months (26) and 44 months (25), where these represent published estimates of DNA and replication-competent reservoir decay on ART, respectively, the model predicts that, at the time of sampling, participant proviral distributions would be skewed toward slightly younger ages (gray and black dotted lines). These predicted proviral distributions fit the participants’ observed data (shown as the blue bars) poorly, and participants 1 and 3 particularly so. As a last step, we identified the pre-ART proviral half-life that best fit each participant’s proviral distribution (green line). Participant 1’s was the shortest of all, only 0.41 years (95% CI, 0 to 1.03) (Fig. 8E). Estimated proviral half-lives for participants 3 and 4 were intermediate, while participant 2’s was nearly 9 years, with an upper 95% CI extending to 77 years.

While this model clearly illustrates how faster decay rates shift proviral distributions toward younger ages, it assumes that within-host cell and viral dynamics behave as parameterized by the model. The model-produced best-fit half-lives were also somewhat sensitive to imputed acute-phase viremia dynamics, with higher, longer peaks yielding the shortest half-lives (Table S1). The half-life estimates for participant 3, who controlled viremia the most successfully, were the most sensitive to these imputations: using a low, short peak for example yielded a half-life estimate of 3.04 years (95% CI, 0 to 18.08), nearly three times longer than that estimated in the primary analysis.

(ii) Direct approach. Recognizing the limitations of the dynamical model and our lack of capture of the participants’ actual acute-phase dynamics, we also estimated pre-ART proviral half-lives directly from observed proviral age distributions using a Poisson generalized linear model, an approach that does not incorporate clinical history information. The 95% CI around the half-life estimates produced by this method overlapped those produced by the dynamical model in all cases (Fig. 8I to L). The point estimates for participants 1 and 2 were also very consistent, with the former again exhibiting the shortest proviral half-life of 0.63 years (95% CI, 0.46 to 0.97) and the latter exhibiting the longest of 23.74 years (95% CI, 4.44 to 47.74). Participant 4’s estimated pre-ART proviral half-life was 0.78 years (95% CI, 0.65 to 0.96), slightly shorter than that estimated with the dynamical model, while participant 3’s was 4.79 years (95% CI, 1.44 to 12.37), which was comparable to the dynamical model-derived estimate when using a low, short inferred acute-phase peak viral load (Table S1). Broadly, however, the proviral pools that were skewed toward ART initiation produced short (<1 year) half-life estimates (participants 1 and 4), while those that dated more uniformly throughout infection produced substantially longer half-lives (participants 2 and 3).

Sensitivity analyses with alternate within-host phylogenies. We now present sensitivity analyses that test the robustness of our findings. In our primary analysis, we estimated proviral ages using a phylogeny inferred using the best-fit nucleotide substitution model. Each data set however yielded models with similar predictive power (see Table S2 in the supplemental material), so we recomputed proviral ages and pre-ART half-lives (using the Poisson generalized linear model) from a phylogeny inferred using each participant’s “next-best-fit” model. Results were broadly consistent with the primary findings: participants 1 and 4 again showed proviral distributions that were skewed to dates near ART, yielding pre-ART half-lives of <1 year (see Fig. S1 in the supplemental material). Participant 2’s alternative phylogeny dated a slightly larger proportion of proviruses to around ART initiation, yielding a shorter half-life than that estimated from the primary analysis.
analysis. Nevertheless, both participant 2’s and 3’s proviral distributions maintained an overall “flat” age distribution, yielding the longest estimated half-lives in the cohort. This suggests that our findings are not unduly influenced by the phylogenetic inference step.

**Inferring proviral age distributions and pre-ART half-lives without a molecular clock.** In our primary analysis, we estimated proviral ages from root-to-tip phylogenetic distances by linear regression (20, 22). This assumes a strict molecular clock, which may not hold over long time periods (44), so we reanalyzed our data using two approaches that do not rely on this assumption. The first was “nearest-neighbor” dating (21), where each provirus is assigned to the sampling date of the pre-ART plasma sequence closest to it in the tree (for consistency, the original tree was used). A limitation of this approach is that proviruses can only be assigned to dates when plasma HIV sequences were sampled. Nevertheless, the results were broadly comparable to the original ones (see Fig. S2, left side, in the supplemental material); participants 2 and 3 showed “flat” proviral age distributions and long estimated proviral half-lives, while participants 1 and 4 showed skewed proviral age distributions and proviral half-lives under 1 year.

We also used least-squares dating (LSD) (45), an approach that aims to minimize the variance between the tree branch lengths and sample dates, that has also recently been used to infer proviral sequence ages (46). We used the original rooted tree; here it may be useful to reiterate that the root represents the location that maximizes the (Spearman’s) correlation between plasma HIV root-to-tip distances and sampling times and thus does not rely on a strict clock. LSD also incorporates tree topology information (as opposed to just root-to-tip distances) and allows variable evolutionary rates over the tree’s edges. Overall, the LSD-inferred 95% CI were generally narrower and the within-host evolutionary rates slightly slower than those inferred using linear regression, but the estimated proviral ages and half-lives were again comparable to the original observations (Fig. S2, right side). Taken together, these observations indicate that our findings are robust to a strict clock assumption.

**Accounting for identical sequences in proviral half-life estimates.** When estimating rates of pre-ART proviral turnover, we excluded identical proviral sequences under the assumption that these arose through clonal expansion rather than through independent integration events. Subgenomic HIV sequencing, however, cannot conclusively classify identical sequences as clonal across the whole HIV genome (47), so we reanalyzed our data using all sequences collected (Table 1). Results were again highly consistent with the original data (see Fig. S3 in the supplemental material). Participant 1’s estimated proviral half-life when including 8 identical sequences was 0.61 years (95% CI, 0.46 to 0.9), which was very similar to the original estimate of 0.63 years (95% CI, 0.46 to 0.97). Similarly, participant 2’s estimated half-life after including 30 identical sequences, the majority of which dated to a lineage that circulated 6 years prior to ART, was zero (i.e., no decay), which is comparable to the original estimate of 23.74 years (95% CI, 4.44 to ∞). Our findings are therefore robust to the inclusion or exclusion of identical sequences.

**DISCUSSION**

Though modest in size, our longitudinal cohort of ART-treated viremic controllers allowed us to characterize proviral burden, diversity, and dynamics in this understudied group. Our results indicate that the persisting proviral pools in these individuals can share considerable similarities in terms of size, diversity, and turnover with those of noncontrollers on ART.

Within-host HIV evolutionary studies have revealed that most infections are initiated by a single transmitted founder virus that gives rise to increasingly diverse and divergent descendants (44, 48–53). The dynamics of within-host HIV evolution in controllers, however, remain somewhat unclear: while some studies performed in shorter time frames reported increases in plasma HIV diversity and/or immune escape over time in elite (54–56) and viremic (57) controllers, others found limited or no evidence of evolution (12, 15, 58). Our first notable observation was that significant within-host HIV evolution occurred in all participants, as evidenced by ongoing divergence from a phylogenetically inferred
root, whose date was consistent with the clinically estimated infection date in all cases. Also indicative of within-host evolution, nonsynonymous substitutions consistent with escape from human leukocyte antigen (HLA) class I-restricted cellular immune responses occurred even in participant 3, for whom pre-ART plasma viral loads never reached above 1,000 HIV RNA copies/ml. In this participant, who expresses HLA-A*03:01 and HLA-B*57:01, 25/25 sequences sampled in 2009 harbored the predicted HLA-B*57:01-restricted epitope AGNNACAW at Nef codons 49 to 57 (59). By 2011, 8/27 sequences harbored AENNAACAW (the changed residue is underlined), which has a predicted >10-fold-reduced binding affinity to B*57:01, a shift in frequency that was statistically significant (Fisher’s exact test, P = 0.0044). Similarly, 25/25 sequences sampled in 2009 harbored the predicted A*03:01-restricted epitope KLVPVEK at Nef codons 144 to 152, but by late 2011, 6/16 sequences harbored KLVPVPEE, which has a predicted >160-fold-reduced binding affinity to A*03:01 (P = 0.0069). Participants’ estimated within-host nef evolutionary rates, which ranged from $1.16 \times 10^{-5}$ to $5.35 \times 10^{-5}$ substitutions per nucleotide site per day, assuming a strict clock, were also comparable to those reported in noncontrollers using similar methods (20). Therefore, despite natural viremia control, significant within-host HIV evolution nevertheless occurs in this group in the absence of therapy.

Our results also revealed that total and genomically intact proviral DNA burdens in ART-treated viremic controllers can be substantial. While few studies have estimated reservoir size in controllers using the IPDA, a recent study reported 20-fold lower frequencies of total and intact proviral DNA in elite controllers (medians of 30 and 1.6 copies/10^6 CD4+ T cells, respectively) compared to noncontrollers on ART (medians of 603 and 37 copies/10^6 CD4+ T cells, respectively) (60). The median total and intact proviral burdens in the present study were 398 and 84 copies/10^6 CD4+ T cells, respectively (or 215 and 58 copies/10^6 CD4+ T cells, respectively, if we exclude participant 4, who lost viremic control), values that were more in line with those in noncontrollers (30, 60, 61). Proviral loads in our study, however, were sampled only ~1.5 years after ART on average. Longitudinal studies of treated viremic controllers on ART will therefore be required to confirm whether, similar to noncontrollers, intact proviruses gradually decay on ART while total proviral DNA remains stable (30, 61, 62).

Our results also yield insights into on-ART proviral diversity and the length of time these sequences had persisted within the host. Consistent with a prior study that included viremic controllers (63), proviruses recovered in the present study were genetically diverse and frequently included archival lineages. Similar to studies of noncontrollers (19–23), on-ART proviral age distributions varied: whereas participants 1’s and 4’s proviral pools were skewed toward sequences that integrated in the years prior to ART, participants 2’s and 3’s proviral pools spanned the infection course more evenly and included proviruses that dated to shortly following infection. Despite these differences, the overall diversity of proviruses persisting on ART reflected the overall plasma HIV diversity generated prior to ART (Fig. 7A). Moreover, the overall level of viral replication pre-ART correlated positively with both proviral burden and diversity on ART (Fig. 7B and C). Our recovery of various numbers of identical sequences also suggests that HIV controllers, like noncontrollers, exhibit differential levels of clonal expansion (31–34, 64). Half of participant 3’s proviruses, for example, were identical to at least one other sequence, consistent with a report describing high identical proviral burdens in viremic controllers (63), while 83% of participant 1’s proviruses were distinct, a phenomenon that has also been reported in viremic controllers (57). Despite this heterogeneity, our observations confirm that diverse proviruses persist in those who control HIV prior to therapy, underscoring the importance of early ART even in this group.

Our study also extends our understanding of proviral turnover during untreated infection (19, 21, 23). While the half-lives of the replication-competent and overall proviral pools on ART are on the order of ~4 and >10 years, respectively (25, 26, 30), recent data suggest that pre-ART proviral half-lives are shorter, with studies returning estimates of 8 months (19, 24) and 25 months (23). Dynamic turnover during untreated infection helps explain why proviral pools are often enriched in sequences that
integrated in the years immediately prior to ART (19, 21, 23) and has led to the hypothesis that ART dramatically slows this rate of turnover, thereby “stabilizing” the proviral pool in its immediate pretherapy state (21, 27). Indeed, participants 1’s and 4’s skewed proviral distributions and short estimated pre-ART proviral half-lives resemble those of “typical” noncontrollers (19, 23, 24), though participants 2’s and 3’s flatter proviral age distributions and markedly longer estimated decay rates emphasize that proviral clearance is not rapid in all persons. While longer pre-ART half-lives have been observed in noncontrollers (20), a recent study by our group identified an inverse relationship between set point pVL and rate of pre-ART proviral turnover (24). Taken together with our observation that 2/4 participants exhibited relatively long pre-ART half-lives, this suggests that pre-ART proviral turnover may be slower in HIV controllers on average, though the data clearly underscore the heterogeneous nature of each individual’s persistent HIV pool.

Our study has some limitations. Our cohort is small due to the challenges of recruiting viremic controllers with known infection dates for longitudinal studies, and all participants were male. Though some aspects of the HIV reservoir likely differ between the sexes (65, 66), a recent study of noncontrollers by our group revealed no significant differences in estimated pre-ART proviral half-lives between men and women, where intriguingly the participant with the “flattest” proviral distribution and longest pre-ART half-life was a female who nearly met the criteria for viremic control (24). This suggests that female controllers may not differ markedly from males in terms of proviral composition and pre-ART turnover, but additional studies are needed to confirm this. As samples were limited and viral loads were low, near-full-genome HIV amplification was not feasible, nor was proviral integration site characterization, so we were not able to investigate genomic location and associated heterochromatin features recently characterized in elite controllers (8) in our cohort. Though nef is appropriate for within-host HIV evolutionary studies (20, 64, 67), sequencing of a subgenomic region does not allow us to classify recovered sequences as intact or defective or definitively classify identical sequences as clonally expanded. Our use of the IPDA to quantify intact (versus defective) proviral burden mitigates this only partially. Our methods also assume that every provirus has an equal probability of being sampled, regardless of whether it is unique or a member of a clonally expanded population (where we crudely estimate this probability to be <0.2%, as the blood collected represented ~0.2% of average total body blood volume, but subsequent DNA extraction and PCR efficiency is <100%). Sample availability also prevented us from investigating the mechanisms underlying participant 4’s loss of viremia control, though our observations that this individual’s (largely genetically intact) proviral pool dated exclusively to their post-control period suggests that this event triggered major increases in reservoir size and diversity, further underscoring the importance of timely ART, even in viremic controllers. Finally, the peak viremia kinetics that we drew from the literature (12, 41, 55) to use in the dynamical model may not have reflected actual in vivo kinetics, leading to uncertainty in our model-predicted proviral deposition and turnover rate estimates.

In conclusion, despite their natural ability to control HIV to low levels, significant within-host HIV evolution occurred in all participants pre-ART, giving rise to within-host proviral pools whose size and genetic diversity reflected this evolution. Pre-ART proviral dynamics were also heterogeneous: though two participants’ proviral pools were skewed toward recent integration dates, consistent with rapid pre-ART turnover (19, 23, 24). Those of two others spanned a wide age range, consistent with much slower pre-ART proviral turnover. HIV remission and cure strategies will need to overcome within-host proviral diversity, even in individuals who naturally control HIV prior to therapy.

MATERIALS AND METHODS

Study participants. Four viremic controllers who naturally maintained plasma viral loads (pVL) of <2,000 HIV RNA copies/ml for at least 3 years, and for whom an infection date could be reliably estimated, were studied. All participants were male, with a median age of 55 years. Participants 1, 2, and 3...
broadly maintained pVL of <2,000 copies/ml pre-ART, whereas participant 4 lost viremic control prior to ART. The estimated dates of infection (EDI) for participants 1, 2, and 4, were calculated as the midpoint between the last seronegative and first seropositive HIV test dates, while for participant 3, the EDI was determined as March 2008, 3 months prior to diagnosis, due to a documented exposure followed by seroconversion-like illness. Participant nadir CD4$^+$ T-cell counts ranged from 320 to 516 cells/$\mu$m$^3$, while CD4$^+$ T-cell counts at ART initiation ranged from 401 to 626 cells/$\mu$m$^3$. HLA class I types were as follows: A*03:01-A*32:01/B*13:02-B*40:02/C*06:02-C*15:02 for participant 1, A*24:02-A*32:01/B*08:01-B*14:01/ C*07:02-C*08:02 for participant 2, A*03:01-A*03:01/B*14:02-B*57:01/C*06:02-C*08:02 for participant 3, and A*03:01-A*32:01/B*14:02-B*39:01/C*08:02-C*12:03 for participant 4. All participants provided written informed consent. This study was approved by the Massachusetts General Hospital Research Ethics Board, with additional approvals for sample and data analysis approved by the Simon Fraser University and Providence Health Care/University of British Columbia Research Ethics Boards.

**Single-genome HIV RNA and proviral amplification.** Phylogenetic inference of proviral ages requires the longitudinal isolation of HIV RNA sequences pre-ART, alongside proviral DNA sequences sampled on ART, ideally using single-genome approaches. Because plasma viral loads were low and biological material was limited (only 10 million PBMC per time point), full-length HIV amplification was not feasible, so nef was amplified for the reasons described in the Results section. Total nucleic acids were extracted from plasma using the bioMérieux NucliSens EasyMag system (bioMérieux, Marcy-Étoile, France), while genomic DNA was extracted from PBMC using the Purelink genomic DNA kit (Invitrogen). For pre-ART plasma samples destined for HIV RNA nef amplification, 1 ml of sample was extracted, eluted in 60 $\mu$L, and subjected to DNase I digestion (New England Biolabs, Ltd., catalog number M0303S) to minimize the risk of amplifying HIV DNA in these low-pVL samples. For on-ART PBMC samples destined for proviral nef amplification, cells were sparsely aliquoted of 5 million cells before extraction and elution into 60 $\mu$L. HIV nef was amplified by limiting-dilution nested reverse transcription (RT)-PCR (for HIV RNA) or nested PCR (for proviral DNA) using high-fidelity enzymes and sequence-specific primers optimized for amplification of multiple HIV group M subtypes (64, 67). For HIV RNA extracts, cDNA was generated using NxtScript reverse transcriptase (Roche). Next, cDNA as well as genomic DNA extracts were endpoint dilution such that ~25 to 30% of the resulting nested PCRs, performed using the Expand high-fidelity PCR system (Roche), would yield an amplicon. The primers (5′−3′) used for cDNA generation/1st round PCR were Nef6863F_pan (forward); TAGCAGTAGCTGKGKRACAGTAT (reverse); and Nef9536R_pan (reverse). The primers used for 2nd round PCR were Nef8746F_pan (forward); TCCATACCTCCTAGAAGTACAC (forward) and Nef9474R_pan (reverse). Negative amplification controls were included in every run and also confirmed that plasma HIV RNA amplification did not occur in the absence of reverse transcription, indicating that DNase treatment was effective.

Amplicons were sequenced on an ABI 3730xl automated DNA analyzer. Chromatograms were base called using Sequencer v5.0 (Gene Codes) or the custom software RECall (68). nef sequences that contained nucleotide mixtures, hypermutations (identified using Hypermut 2.0 [69]), suspected within-host recombinant sequences (identified using rpdr Beta 95 [70]), or other defects were excluded from phylogenetic analysis, as were duplicate sequences (instead, the latter were added back at the data visualization stage). Within-host plasma and proviral nef sequences were codon aligned using MAFFT v7 (71) implemented in HIV Align (https://www.hiv.lanl.gov/content/sequence/VIRALIGN/viralign.html) and manually edited in AlView v1.18 (72). Maximum likelihood phylogenies were inferred from aligned, gap-striped within-host sequence data sets using W-IQ-TREE (73) following automated model selection with ModelFinder (74) using an Akaike information criterion (AIC) selection criterion (75). Best-fit models are reported in Table S2. Our primary analysis used a phylogenier inferred using each participant’s top-best-fit model, but a sensitivity analysis was performed using each participant’s next-best-fit model (Fig. S1). Patristic (tip-to-tip phylogenetic) distances were extracted using the cophenetic.phylo function from the R package ape (v5.3) (76).

**HIV proviral full-genome amplification and sequencing.** Single-template, near-full-length proviral amplification was performed on DNA extracted from CD4$^+$ T cells by nested PCR using Platinum Taq high-fidelity DNA polymerase (Invitrogen) such that ~25% of the resulting PCRs yielded an amplicon. The 1st round primers were AAATCTCTAGCAGTGGCGCCAGAACAG (forward) and TGGAGGATCTCTAGTTACCATGGGCCT (reverse). The 2nd round primers were GCCTCCACCGCGCTGAGAARCAAG (forward) and GCCTCCACCTCCTCCCTCTGAGAAGAASKCC (reverse). Negative amplification controls were included in every run and also confirmed that plasma HIV RNA amplification did not occur in the absence of reverse transcription, indicating that DNase treatment was effective.

Amplicons were sequenced on an ABI 3730xl automated DNA analyzer. Chromatograms were base called using Sequencer v5.0 (Gene Codes) or the custom software RECall (68). nef sequences that contained nucleotide mixtures, hypermutations (identified using Hypermut 2.0 [69]), suspected within-host recombinant sequences (identified using rpdr Beta 95 [70]), or other defects were excluded from phylogenetic analysis, as were duplicate sequences (instead, the latter were added back at the data visualization stage). Within-host plasma and proviral nef sequences were codon aligned using MAFFT v7 (71) implemented in HIV Align (https://www.hiv.lanl.gov/content/sequence/VIRALIGN/viralign.html) and manually edited in AlView v1.18 (72). Maximum likelihood phylogenies were inferred from aligned, gap-striped within-host sequence data sets using W-IQ-TREE (73) following automated model selection with ModelFinder (74) using an Akaike information criterion (AIC) selection criterion (75). Best-fit models are reported in Table S2. Our primary analysis used a phylogenier inferred using each participant’s top-best-fit model, but a sensitivity analysis was performed using each participant’s next-best-fit model (Fig. S1). Patristic (tip-to-tip phylogenetic) distances were extracted using the cophenetic.phylo function from the R package ape (v5.3) (76).

**Within-host phylogenetic inference and proviral age reconstruction.** We used a published within-host phylogenetic approach to infer proviral sequence ages (20). Briefly, for each participant, we inferred a maximum likelihood phylogeny relating within-host longitudinal pre-ART plasma HIV RNA and proviral sequences sampled on ART. We then exhaustively rerooted each tree to identify the root location that maximizes the (Spearman’s) correlation between the root-to-tip distances and collection dates of the pre-ART plasma HIV RNA sequences. Here, the root represents the inferred most recent common ancestor (MRCA) of the within-host data set (i.e., the inferred transmitted founder virus). We then fit a linear model relating the collection dates of the plasma HIV RNA sequences to their divergence from the root, where the slope represents the average host-specific rate of HIV evolution prior to ART and the y-intercept represents the phylogenetically estimated MRCA date. For participant 4, two regression lines were fit: one each for their “viremic control” and “post-control” eras.
The integration dates of the proviral sequences sampled during ART, along with their 95% confidence intervals (CI), were then estimated from their divergence from the root using the linear regression. We assessed molecular clock and model fit using a delta Akaike information criterion (ΔAIC), computed as the difference between the AIC of the null model (no evidence of within-host evolution—i.e., a zero slope) and the AIC of the linear regression. A within-host phylogeny was deemed to have a sufficient molecular clock signal if the linear regression produced a ΔAIC of ≥10, (which corresponds to a P value of 0.00053 when using a log-likelihood ratio test), and the 95% confidence interval of the estimated root date contained or preceded the first sampling point. The proviral dating framework was available as a web service at https://bblab-hivresearchtools.ca/django/tools/phylodating, with open source code available at https://www.github.com/cfe-lab/phylodating. As described in the Results, sensitivity analyses were also performed where proviruses were phylogenetically “dated” using two published methods that do not rely on a strict molecular clock assumption (Fig. S2).

**Intact proviral DNA assay.** Each participant’s first PBMC sample on ART was used to estimate total and intact proviral DNA burdens using the intact proviral DNA assay (IPDA) (29). As this method reports proviral burdens in terms of HIV copies per million CD4+ T cells, we first isolated CD4+ T cells from 10 million PBMC by negative selection using the EasySep Human CD4+ T cell enrichment kit (STEMCELL Technologies, catalog number 19052). This yielded a mean of 1.6 million (range, 1.3 to 2.0 million) CD4+ T cells per participant, which were then extracted using the QIAamp DNA minikit (Qiagen) with precautions to minimize shearing. In the IPDA, HIV and human DNA quantification reactions (where the latter target the human RPP30 gene) are conducted in parallel, where copies are normalized to the quantity of input DNA and subsequently corrected for DNA shearing. In each ddPCR reaction, 7 ng (RPP30) or a median of 588 ng (interquartile range [IQR] 488 to 697 ng) (HIV) of genomic DNA was combined with ddPCR Supermix for Probes (no dUTPs; Bio-Rad), primers (final concentration, 900 nM; Integrated DNA Technologies), probe(s) (final concentration, 250 nM; Integrated DNA Technologies), Xho restriction enzyme (New England Biolabs), and nuclelease-free water. The human RPP30 primer and probe sequences (5'—3') are as follows: RPP30 forward primer, GATTTGGACCTGCGAGCG; RPP30 probe, VIC-CTGACCTGAAGGCTCT-GBNFQ; RPP30 reverse primer, GCCGCCTCTCCTACCAAGT; RPP30 shear forward primer, CATTGGCTGCTCTTGGG; RPP30 shear probe, 6-carboxyfluorescein (FAM)-AAGGACAGGTTCTACTTAG-MGBNFQ; and RPP30 shear reverse primer, CATGCAAAAGAGGAAGCGGC. The default (29) HIV primers and probes were as follows: HIV forward primer, CAGGACTCCTGCTGAAG; HIV probe, FAM-TTTTCGCTACTCAGG-MBNFQ; and HIV reverse primer, GCACCATCTCCTCTCCTTAGC; env forward primer, AGTGGTCAGAGAAAAAGAGC; env probe, VIC-CCTGGGTTCTGGA-MBNFQ; anti-hypermutant env probe, CCTTAGGTCTTACTAGGAC-MBNFQ; and env reverse primer, GTCTGGCCTACTGACTG.

The published HIV primers/probes, however, can sometimes fail to detect autologous HIV sequences due to naturally occurring polymorphism (79, 80), thus requiring autologous primers/probes. For this reason, targeted HIV sequencing of the IPDA measurement, and autologous primers/probes were substituted where necessary. The substituted primer and probe sequences (5'—3') were as follows: for participant 2, HIV probe, FAM-TTTTCGCTACTCACAGT; for participant 3, HIV probe, FAM-ATGCGCTACTCAGGAT; for participant 1, HIV probe, FAM-AATTGGCTACTCAGGAT and FAM-AATTGCGCTACTCAGG; for participant 1, env forward primer, GGGTCAGAGAAAAAGGAGC, and env reverse primer, GTCAGGCGACGGCAGGC; and for participant 3, env reverse primer, GCTAGGGCTACGGCAGAT. Droplets were prepared using the Automated Droplet Generator and cycled as previously described (29). Droplets were analyzed on a QX200 Droplet Reader (Bio-Rad) using QuantaSoft software (Bio-Rad, version 1.7.4), where the results of a minimum of 8 technical replicates, comprising a median of 889,420 cells (IQR, 501,339 to 721,871 cells) assayed in total, were merged prior to analysis. Intact HIV-1 copies (9 and env double-positive droplets) were corrected for DNA shearing based on the frequency of RPP30 and RPP30-shear double-positive droplets. The median DNA shearing index, which measures the proportion of sheared DNA in a sample, was 0.40 (IQR, 0.39 to 0.43), which is in line with acceptable levels in this assay (29, 79).

**Inference of pre-ART proviral half-lives.** (i) Dynamical mathematical model. In order to infer pre-ART proviral half-lives, each participant’s sampled proviruses were grouped into “bins” by their estimated year of integration (one “bin” per year preceding ART initiation). For simplicity, proviruses whose integration date point estimate fell after the ART initiation date (or in the case of participant 3, the date of the last viremic episode), were assigned to the ART initiation (or viremic episode) date (all proviruses in this category had 95% confidence intervals that overlapped this date). In the primary analysis, proviral half-life inference was only performed on distinct proviral sequences under the assumption that “rePLICATE” sequences arose through clonal expansion and not through individual integration events, but a sensitivity analysis was performed, including all proviral sequences (Fig. S3).

We estimated host-specific pre-ART proviral half-lives two ways: using a dynamical mathematical model (23) and directly from the data. To do the former, we implemented a published mathematical model that describes each individual’s untreated HIV infection via a set of ordinary differential equations (ODE) that describe the dynamics of the concentrations of cells and virus in each individual over time (23). The model includes susceptible target cells, actively and latently infected cells that can produce viable virus, actively and latently infected cells that cannot produce viable virus, the virus itself, and an immune response. The model assumes that HIV sequences enter the latent proviral pool at a rate proportional to viral burdens in terms of HIV copies per million CD4+ T cells, we

Let S represent the susceptible compartment. Let A0 and L0 represent actively and latently infected cells that are productively infected and produce viable virus; likewise let A1 and L1 represent actively
and latently infected cells that are unproductively infected and cannot produce viable virus. Let $V$ represent viremia, and let $E$ represent the adaptive immune response. The following dynamical system describes within-host HIV infection kinetics:

$$
\begin{align*}
\dot{S} &= \alpha_S - \delta_S S - \beta SV \\
\dot{A}_F &= (1 - \lambda) \tau \beta SV - \delta_A A_F - \kappa A_S E \\
\dot{A}_U &= (1 - \lambda)(1 - \tau) \beta SV - \delta_A A_U - \kappa A_U E \\
\dot{L}_F &= \lambda \tau \beta SV \\
\dot{L}_U &= \lambda (1 - \tau) \beta SV \\
\dot{E} &= \alpha_E + \omega_E \frac{(A_F + A_U)}{E + E_0} - \delta_E E \\
\dot{V} &= \pi A_F - \gamma V - \beta SV
\end{align*}
$$

The following parameters were used: susceptible cell creation rate, $\alpha_S = 70$ cells $\mu l^{-1} day^{-1}$; susceptible cell death rate, $\delta_S = 0.2 day^{-1}$; viral infectivity, $\beta = 10^{-4}$ viral RNA copies cell$^{-1}$ day$^{-1}$; probability of productively infectious virions, $\tau = 0.05$; probability of latency, $\lambda = 10^{-4}$; actively infected cell death rate, $\delta_A = 0.8 day^{-1}$; viral burst size, $\pi = 50,000$ viral RNA copies cell$^{-1}$ day$^{-1}$; viral clearance rate, $\gamma = 23 day^{-1}$; initial adaptive precursor frequency, $\alpha_E = 10^{-4}$ cells $\mu l^{-1} day^{-1}$; adaptive immune killing rate, $\kappa = 0.3 \mu l^{-1} day^{-1}$; adaptive immune recruitment rate, $\omega = 1.6 day^{-1}$; adaptive immune clearance rate, $\delta_E = 0.002 day^{-1}$; and adaptive immune 50% saturation constant (i.e., the number of infected cells required for a half-maximal cytolytic expansion rate [81]), $E_{50} = 250$ cells $\mu l^{-1}$.

The following initial conditions were used as per the original publications (23, 81). The initial viral load value was $V(0) = 0.03$ viral RNA copies per $\mu l$, which is the detectable plasma viral load limit of a typical assay when converted from the conventional $30$ HIV RNA copies/ml to viral RNA copies per $\mu l$. Also let $I_0 = V(0) / \pi = 1.38 \times 10^{-5}$; this represents a quasistatic approximation for the infected cells.

$$
\begin{align*}
S(0) &= \alpha_S / \delta_S = 70 / 0.2 = 350 \\
A_F(0) &= I_0 \tau (1 - \lambda) = 6.89931 \times 10^{-7} \\
A_U(0) &= I_0 (1 - \tau)(1 - \lambda) = 1.310869 \times 10^{-5} \\
L_F(0) &= I_0 \tau \lambda = 6.9 \times 10^{-11} \\
L_U(0) &= I_0 (1 - \tau) \lambda = 1.311 \times 10^{-9} \\
E(0) &= \alpha_E / \delta_E = 0.0001 / 0.002 = 0.05 \\
V(0) &= 30 / 1,000 = 0.03
\end{align*}
$$

Note that $S(0)$ and $E(0)$ are the equilibrium values for the system in the absence of virus.

As described above, reservoir creation is assumed to be proportional to plasma viral load over time, with the probability of a virus entering the latent pool (defined by $\lambda$ above) remaining constant over time. Unfortunately, peak viremia was not captured in our participants’ measured viral load data. Therefore, to recreate each participant’s in vivo plasma viral dynamics as realistically as possible, we merged each participant’s available pVL data to acute-phase dynamics observed in viremic controllers (41), and we also performed sensitivity analyses using peak viremia kinetics observed in elite controllers (42, 43) (Table S1). Using each participant’s reconstructed viral load history and the above equations, we modeled proviral deposition into their latent pool. We then grouped the resulting proviruses by their year of creation.

In a separate step, we then allowed each group of latent proviruses to decay exponentially, up until each participant’s proviral sampling date, under half-lives ranging from 30 to 6,000 days in increments of 30 days. This produced a series of 200 predicted proviral distributions per participant that represented what proportion of proviruses would still remain from each creation year, assuming decay at the stated rate. For context, we also applied decay rates of 44 and 140 months, which represent the half-lives of the replication-competent reservoir (25) and total proviral pool (26) during suppressive ART, respectively. Note that the model assumes that, after ART, the overall size of the proviral pool decreases but
the relative proportions of sequences in each group remain the same; as such, model predictions would be the same regardless of when the proviral pool is sampled on ART. Finally, we identified the best-fit proviral decay rate to each participant’s observed proviral age distribution by maximum likelihood and used standard theory to identify 95% confidence bounds.

(ii) Poisson linear model. To estimate the reservoir half-life directly from the data, we again grouped each participant’s observed proviral sequences by their age, in years, relative to ART initiation (or for participant 3, the date of their last viremic episode). We will refer to the bin containing proviruses from \( t - 1 \) to \( t \) years old as “bin \( t \),” where we make the simplifying assumptions that all proviruses in bin \( t \) are exactly \( t \) years old. We then applied a Poisson generalized linear model with the canonical natural logarithm link function to the binned counts, using the age of the bin \( t \) as the predictor. This choice can be justified as follows: Let \( t_{1/2} \) be the proviral half-life. Assuming the participant’s proviral reservoir decays at an exponential rate, we would expect that the size of bin \( t \) would be approximately \( C \exp(-\theta t) \), where \( C \) represents the initial size of the age bin as if there was no decay, and \( \theta = \ln(2)/t_{1/2} \).

Now, we assume that every provirus has the same small independent probability \( p \) of being sampled. If so, we would then expect the number of observed proviruses from bin \( t \) to be binomially distributed with \( C \exp(-\theta t) \) trials and success probability \( p \). Assuming that \( p \) is small, we can approximate the distribution with a Poisson distribution with parameter

\[
\lambda = Cp \exp(-\theta t) = \exp(D - \theta t)
\]

where \( D = \ln(Cp) \), or in other words,

\[
\ln \lambda = D - \theta t
\]

which is precisely the setup required for a Poisson generalized linear model with natural logarithm link function.

Statistical analyses. Spearman’s correlation was used to assess the relationship between continuous parameters. Statistical analysis was performed in Prism (version 9) software.

Data availability. GenBank accession numbers for sequences reported in this study are MW781931 to MW782197 (proviral DNA) and MW782198 to MW782743 (plasma HIV RNA).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, EPS file, 0.8 MB.
FIG S2, EPS file, 1.3 MB.
FIG S3, EPS file, 1.6 MB.
TABLE S1, DOCX file, 0.02 MB.
TABLE S2, DOCX file, 0.02 MB.

ACKNOWLEDGMENTS

We gratefully thank the study participants, without whom this research would not be possible. We also thank Daniel Reeves, the creator of the dynamical mathematical model, for helpful correspondence. We are also grateful to Daniel Worrall for assistance with clinical data access.

This work was supported in part by the Canadian Institutes of Health Research (CIHR) through a project grant (PJT-159625 to Z.L.B. and J.B.J.) and a focused team grant (HB1-164063 to Z.L.B. and M.A.B.). This work was supported in part by the National Institutes of Health (NIH) under award number NIHA127029 (to Z.L.B.) and the Martin Delaney “BELIEVE” Collaboratory (NIH grant 1UM1AI26617 to Z.L.B. and M.A.B.), which is supported by the following NIH cofunding and participating institutes and centers: NIAID, NCI, NICHD, NHLBI, NIDA, NIMH, NIA, FIC, and OAR. This work was supported by the Martin Delaney “REACH” Collaboratory (NIH grant 1-UM1AI164565-01 to Z.L.B. and M.A.B.), which is supported by the following NIH cofunding institutes: NIMH, NIDA, NINDS, NIDDK, NHLBI, and NIAID. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. F.H.O. is supported by a Ph.D. fellowship from the Sub-Saharan African Network for TB/ HIV Research Excellence (SANTHE), a DELTAS Africa Initiative (grant number DEL-15-006). The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)’s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa’s Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust (grant number 107752/Z/15/Z) and the United Kingdom Government. The views expressed...
in this publication are those of the authors and not necessarily those of AAS, the NEPAD Agency, the Wellcome Trust, or the United Kingdom Government. A.S. and B.R.J. are supported by CIHR Doctoral Research Awards. N.N.K. is supported by a CIHR Vanier Doctoral Award. R.L.M. is supported by a CIHR MSc award. M.A.B. holds a Canada Research Chair, Tier 2, in Viral Pathogenesis and Immunity. Z.L.B. is supported by a Scholar Award from the Michael Smith Foundation for Health Research.

REFERENCES

1. Finzi D, Herrmannova M, Pierson T, Carruth LM, Buck C, Chaisson RE, Quinn TC, Chadwick K, Margolick J, Brookmeyer R, Gallant J, Markowitz M, Ho DD, Richman DD, Siliciano RF. 1997. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. Science 278:1295–1300. https://doi.org/10.1126/science.278.5341.1295.

2. Wong JK, Haezreh M, Gunthard HF, Havlir DV, Ignacio CC, Spina CA, Richman DD. 1997. Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. Science 278:1291–1295. https://doi.org/10.1126/science.278.5341.1291.

3. Chun TW, Stuyver L, Mizell SB, Ehler LA, Micans JAM, Baseler M, Lloyd AL, Nowak MA, Fauci AS. 1997. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. Proc Natl Acad Sci U S A 94:13193–13197. https://doi.org/10.1073/pnas.94.24.13193.

4. Sanchez G, Xu X, Chermann JC, Hirsch I. 1997. Accumulation of defective viral genomes in peripheral blood mononuclear cells of human immunodeficiency virus type 1-infected individuals. J Virol 71:2233–2240. https://doi.org/10.1128/JVI.71.3.2233-2240.1997.

5. Bruner KM, Murray AJ, Pollack RA, Soliman MG, Laskey SB, Capoferri AA, Lai J, Strain MC, Lada SM, Hoh R, Ho YC, Richman DD, Deeks SG, Siliciano JD, Siliciano RF. 2016. Defective proviruses rapidly accumulate during acute HIV-1 infection. Nat Med 22:1043–1049. https://doi.org/10.1038/nm.4156.

6. Imamichi H, Deeks SG, Walker BD, Lichterfeld M, Yu XG. 2020. Distinct viral reservoirs towards an HIV cure 2016. Nat Med 22:839–850. https://doi.org/10.1038/s41598-016-00415-z.

7. Deeks SG, Lewin SR, Ross AL, Ananworanich J, Garcia S, Rodriguez C, del Romero J, Barouch DH, Behrens G, Cavazzana M, Chen Z, Cohen EA, Corbett GM, Chomont N, Douek D, Lifson JD, Lo YR, Kortuzes D, Margolis D, Mellors J, Persaud D, Tucker JD, Barre-Sinoussi F, Alter G, Auerbach J, Autran B, Bazzoni F, Cato S, Cheever SK, Siliciano RF. 2016. Defective HIV-1 proviruses produce novel protein-coding RNA species in HIV-infected patients on combination anti-retroviral therapy. Proc Natl Acad Sci U S A 113:8783–8788. https://doi.org/10.1073/pnas.1609057113.

8. Li C, Mori I, Santurini D, Cellaro S, Di Giorgio S, Gobbetti E, Ho PL, Heneine W, Kim CJ, Lebovics N, Markowitz M, Miller A, Nowak MA, Fauci AS, Lanier LL, Montaner J, O’Shea K, Palella F, Perno C, Power M, Richman DD, Siemiatycki J, Standley H, Siegal F, Siliciano RF, Siliciano JM, Wainberg MA, Weidner G, Williams PL, Wolinsky SM. 2019. Permanent control of HIV-1 latency after allogeneic HCT: results of a randomized clinical trial. Nature 575:486–492. https://doi.org/10.1038/s41586-019-0362-9.

9. Chun TW, Stuyver L, Mizell SB, Ehler LA, Micans JAM, Baseler M, Lloyd AL, Nowak MA, Fauci AS. 1997. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. Proc Natl Acad Sci U S A 94:13193–13197. https://doi.org/10.1073/pnas.94.24.13193.

10. Sanchez G, Xu X, Chermann JC, Hirsch I. 1997. Accumulation of defective viral genomes in peripheral blood mononuclear cells of human immunodeficiency virus type 1-infected individuals. J Virol 71:2233–2240. https://doi.org/10.1128/JVI.71.3.2233-2240.1997.

11. Bruner KM, Murray AJ, Pollack RA, Soliman MG, Laskey SB, Capoferri AA, Lai J, Strain MC, Lada SM, Hoh R, Ho YC, Richman DD, Deeks SG, Siliciano JD, Siliciano RF. 2016. Defective proviruses rapidly accumulate during acute HIV-1 infection. Nat Med 22:1043–1049. https://doi.org/10.1038/nm.4156.

12. Deeks SG, Lewin SR, Ross AL, Ananworanich J, Benkirane M, Cannon P, Nowak MA, Fauci AS. 1997. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. Proc Natl Acad Sci U S A 94:13193–13197. https://doi.org/10.1073/pnas.94.24.13193.

13. Bruner KM, Murray AJ, Pollack RA, Soliman MG, Laskey SB, Capoferri AA, Lai J, Strain MC, Lada SM, Hoh R, Ho YC, Richman DD, Deeks SG, Siliciano JD, Siliciano RF. 2016. Defective proviruses rapidly accumulate during acute HIV-1 infection. Nat Med 22:1043–1049. https://doi.org/10.1038/nm.4156.

14. Chun TW, Stuyver L, Mizell SB, Ehler LA, Micans JAM, Baseler M, Lloyd AL, Nowak MA, Fauci AS. 1997. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. Proc Natl Acad Sci U S A 94:13193–13197. https://doi.org/10.1073/pnas.94.24.13193.

15. Bruner KM, Murray AJ, Pollack RA, Soliman MG, Laskey SB, Capoferri AA, Lai J, Strain MC, Lada SM, Hoh R, Ho YC, Richman DD, Deeks SG, Siliciano JD, Siliciano RF. 2016. Defective proviruses rapidly accumulate during acute HIV-1 infection. Nat Med 22:1043–1049. https://doi.org/10.1038/nm.4156.

16. Imamichi H, Deuar RL, Adelsberger JW, Rehm CA, O’Doherty U, Paxinos EE, Fauci AS, Lane HC. 2016. Defective HIV-1 proviruses produce novel protein-coding RNA species in HIV-infected patients on combination anti-retroviral therapy. Proc Natl Acad Sci U S A 113:8783–8788. https://doi.org/10.1073/pnas.1609057113.

17. Deeks SG, Lewin SR, Ross AL, Ananworanich J, Benkirane M, Cannon P, Nowak MA, Fauci AS. 1997. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. Proc Natl Acad Sci U S A 94:13193–13197. https://doi.org/10.1073/pnas.94.24.13193.

18. Li C, Mori I, Santurini D, Cellaro S, Di Giorgio S, Gobbetti E, Ho PL, Heneine W, Kim CJ, Lebovics N, Markowitz M, Miller A, Nowak MA, Fauci AS, Lanier LL, Montaner J, O’Shea K, Palella F, Perno C, Power M, Richman DD, Siemiatycki J, Standley H, Siegal F, Siliciano RF, Siliciano JM, Wainberg MA, Weidner G, Williams PL, Wolinsky SM. 2019. Permanent control of HIV-1 latency after allogeneic HCT: results of a randomized clinical trial. Nature 575:486–492. https://doi.org/10.1038/s41586-019-0362-9.

19. Chun TW, Stuyver L, Mizell SB, Ehler LA, Micans JAM, Baseler M, Lloyd AL, Nowak MA, Fauci AS. 1997. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. Proc Natl Acad Sci U S A 94:13193–13197. https://doi.org/10.1073/pnas.94.24.13193.

20. Chun TW, Stuyver L, Mizell SB, Ehler LA, Micans JAM, Baseler M, Lloyd AL, Nowak MA, Fauci AS. 1997. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. Proc Natl Acad Sci U S A 94:13193–13197. https://doi.org/10.1073/pnas.94.24.13193.

21. Chun TW, Stuyver L, Mizell SB, Ehler LA, Micans JAM, Baseler M, Lloyd AL, Nowak MA, Fauci AS. 1997. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. Proc Natl Acad Sci U S A 94:13193–13197. https://doi.org/10.1073/pnas.94.24.13193.
select for rare Gag variants associated with reduced viral replication capacity and strong cytotoxic T-lymphocyte recognition. J Virol 83: 2743–2755.  https://doi.org/10.1128/JVI.02265-08.

57. de Azevedo SSD, Caetano DG, Côrtes FS, Teixeira SL, dos Santos Silva K, Hoagland B, Grinsztejn B, Veloso VG, Morgado MG, Bello G. 2017. Highly divergent patterns of genetic diversity and evolution in proviral quasispecies from HIV controllers. Retrovirology 14:29.  https://doi.org/10.1186/s12977-017-0231-7.

58. Wang B, Mikhail M, Dyer WB, Zauders JJ, Kelleher AD, Sakensa NK. 2003. First demonstration of a lack of viral sequence evolution in a nonprogressor, defining replication-incompetent HIV-1 infection. Virology 312: 135–150.  https://doi.org/10.1006/viro.2002.0966.

59. Reynisson B, Alvaredo P, Peters B, Nielsen M. 2020. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res 48:W449–W454.  https://doi.org/10.1093/nar/gkaa379.

60. Kwaak AK, GarliSS CC, Ritter KD, Laird GM, Blankson JN. 2020. Elite suppressors have low frequencies of intact HIV-1 proviral DNA. AIDS 34:641–643.  https://doi.org/10.1097/QAD.0000000000002474.

61. Falcinelli SD, Kilpatrick KW, Read J, Murtagh R, Allard B, Ghofrani S, Kirchheimer J, James KS, Stueiieke E, Baker C, Kuruc JD, Eron JJ, Hudgens MG, Gay CL, Margolis DM. Archin NM. 2021. Longitudinal dynamics of intact HIV proviral DNA and outgrowth virus frequencies in a cohort of individuals receiving antiretroviral therapy. J Infect Dis 224:92–100.  https://doi.org/10.1093/infdis/jiaa718.

62. Gandhi RT, Cytorczyk JC, Bosch RJ, Mar H, Laird GM, Martin A, Collier AC, Riddler SA, Macatangay BJ, Rinaldo CR, Eron JJ, Siliciano JD, McMahan DK, Mellors JW, Hogg E, LeBlanc R, ScIeo C, Palm D, Gandhi M, Fletcher C, Ponoaly A, Awekka F, Halvaa L, Dragan J, Joseph J, Lagattuta R, Lin L, Pederson S, Robertson K, Rubin L, Smith D, Suderuddin H, Shahid A, Joy JB, Brumme ZL, Routy JP. 2019. HIV diversity and genetic compartmentalization by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res 48:W449–W454.  https://doi.org/10.1093/nar/gkaa379.

63. Betrò RT, Jüni TF, Chui CKS, Chu AL, Wynhoven B, Hall TA, Trevino C, Shafer RW, Harrigan PR. 2012. Activating HIV drug resistance genotyping with RECall, a freely accessible sequence analysis tool. J Clin Microbiol 50:1936–1942.  https://doi.org/10.1128/JCM.00689-11.

64. Rose PP, Korber BT. 2000. Detecting hypermutations in viral sequences with an emphasis on G—A hypermutation. Bioinformatics 16:400–401.  https://doi.org/10.1093/bioinformatics/16.4.400.

65. Martin DP, Murrell B, Golden M, Khoosal A, Mulhine B. 2015. RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol 1:vrev003.  https://doi.org/10.1093/virev/vrev003.

66. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780.  https://doi.org/10.1093/molbev/mst010.

67. Larsson A. 2014. AllView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30:3276–3278.  https://doi.org/10.1093/bioinformatics/btu531.

68. Trifénopoulou J, Nguyen LT, von Haeseler A, Minh QG. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res 44:W232–W235.  https://doi.org/10.1093/nar/gkw256.

69. Kalyaanamoorthy S, Minh QG, Wong TKF, Von Haeaseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589.  https://doi.org/10.1038/nmeth.4285.

70. Akaike H. 1974. A new look at the statistical model identification. IEEE Trans Automat Contr 19:716–723.  https://doi.org/10.1109/TAC.1974.1100705.

71. Paradis E, Schliep K. 2019. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526–528.  https://doi.org/10.1093/bioinformatics/bty633.

72. Li B, Gladden AD, Alfett M, Kaldor JM, Cooper DA, Kelleher AD, Allen TM. 2007. Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution. J Virol 81:193–201.  https://doi.org/10.1128/JVI.01231-06.

73. Hunt M, Gall A, Ong SH, Brenner J, Ferns B, Gouldier P, Nastouli E, Keane JA, Kellam P, Otto TD. 2015. IVA: accurate de novo assembly of RNA virus genomes. Bioinformatics 31:2374–2376.  https://doi.org/10.1093/bioinformatics/btv120.

74. Kinloch NN, Ren Y, Cancio Alberto WD, Dong W, Khadka P, Huang SH, Mota TM, Wilson A, Shahid A, Kirby D, Harris M, Kovacs C, Benko E, Ostorowski MA, Del Rio Estrada PM, Wimpelberg A, Cannon C, Hardy W, Maclaren L, Goldstein H, Brumme C, Lee QQ, Lynch RM, Brumme ZL, Jones RB. 2021. HIV-1 diversity considerations in the application of the Intact Proviral DNA Assay (IPDA). Nat Commun 12:165.  https://doi.org/10.1038/s41467-020-20442-3.

75. Simonetti FR, White JA, Tumiotto C, Ritter KD, Cai M, Gandhi RT, Deeks SG, Howell BJ, Montanier L, Blankson JN, Martin A, Laird GM, Siliciano RF, Mellors JW, Siliciano JD. 2020. Intact proviral DNA assay analysis of large cohorts of people with HIV provides a benchmark for the frequency and composition of persistent viral DNA. Proc Natl Acad Sci U S A 117: 18692–18700.  https://doi.org/10.1073/pnas.2006816117.

76. Reeves DB, Peterson CW, Kiem H-P, Schiffer JT. 2017. Autologous stem cell transplantation disrupts adaptive immune responses during rebound simian/human immunodeficiency virus viremia. J Virol 91:95–112.  https://doi.org/10.1128/JVI.00095-17.