

RESEARCH ARTICLE

HIV-1 variants are archived throughout infection and persist in the reservoir

Kelsie Brooks¹, Bradley R. Jones², Dario A. Dileria¹, Daniel J. Wilkins^{1,2a}, Daniel T. Claiborne^{1,2b}, Samantha McNally¹, Jill Gilmour³, William Kilembe⁴, Jeffrey B. Joy^{2,5}, Susan A. Allen^{4,6}, Zabrina L. Brumme^{2,7}, Eric Hunter^{1,6*}

1 Emory Vaccine Center, Emory University, Atlanta, Georgia, United States of America, **2** British Columbia Centre for Excellence in HIV/AIDS, Vancouver, British Columbia, Canada, **3** Human Immunology Lab, International AIDS Vaccine Initiative, London, England, United Kingdom, **4** Zambia-Emory HIV Research Project, Lusaka, Zambia, **5** Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada, **6** Department of Pathology & Laboratory Medicine, Emory University School of Medicine, Atlanta, Georgia, United States of America, **7** Faculty of Health Sciences, Simon Fraser University, Burnaby, British Columbia, Canada

^{2a} Current address: Department of Biology, Emory University, Atlanta, Georgia, United States of America

^{2b} Current address: Ragon Institute of MGH, MIT & Harvard, Cambridge, Massachusetts, United States of America

* ehunte4@emory.edu



OPEN ACCESS

Citation: Brooks K, Jones BR, Dileria DA, Wilkins DJ, Claiborne DT, McNally S, et al. (2020) HIV-1 variants are archived throughout infection and persist in the reservoir. *PLoS Pathog* 16(6): e1008378. <https://doi.org/10.1371/journal.ppat.1008378>

Editor: Daniel C. Douek, Vaccine Research Center, UNITED STATES

Received: August 30, 2019

Accepted: February 3, 2020

Published: June 3, 2020

Copyright: © 2020 Brooks et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Sequences are available at Genbank under accession numbers MT194125 - MT195535.

Funding: The research reported in this publication was funded in part by: NIH R01 AI064060 (E.H.), NIH R01 AI051231 (E.H.), NIH R21 AI127029 (J.B. J/Z.L.B.), NIH UM1 AI126617 (Z.L.B.), NIH F31 AI122926 (K.B.), and the Yerkes National Primate Research Center base grant through the Office of Research Infrastructure Programs/OD P510D11132. This work was also funded in part

Abstract

The HIV-1 reservoir consists of latently infected cells that persist despite antiretroviral therapy (ART). Elucidating the proviral genetic composition of the reservoir, particularly in the context of pre-therapy viral diversity, is therefore important to understanding reservoir formation and the persistence of latently infected cells. Here we investigate reservoir proviral variants from 13 Zambian acutely-infected individuals with additional pre-therapy sampling for a unique comparison to the ART-naïve quasispecies. We identified complete transmitted/founder (TF) viruses from seroconversion plasma samples, and additionally amplified and sequenced HIV-1 from plasma obtained one year post-infection and just prior to ART initiation. While the majority of proviral variants in the reservoir were most closely related to viral variants from the latest pre-therapy time point, we also identified reservoir proviral variants dating to or near the time of infection, and to intermediate time points between infection and treatment initiation. Reservoir proviral variants differing by five or fewer nucleotide changes from the TF virus persisted during treatment in five individuals, including proviral variants that exactly matched the TF in two individuals, one of whom had remained ART-naïve for more than six years. Proviral variants during treatment were significantly less divergent from the TF virus than plasma variants present at the last ART-naïve time point. These findings indicate that reservoir proviral variants are archived throughout infection, recapitulating much of the viral diversity that arises throughout untreated HIV-1 infection, and strategies to target and reduce the reservoir must therefore permit for the clearance of proviruses encompassing this extensive diversity.

by the International AIDS Vaccine Initiative (IAVI) and made possible by the support of many donors, including the United States Agency for International Development (USAID). The full list of IAVI donors is available at <http://www.iavi.org>. The contents of this manuscript are the responsibility of the authors and do not necessarily reflect the views of USAID or the US Government. Additional funding was provided by the Center for AIDS Research at Emory University (NIH P30 AI050409), and a project grant from the Canadian Institutes for Health Research (PJT-159625 to J.B.J./Z.L.B.). Z.L.B. is supported by a Michael Smith Foundation for Health Research (MSFHR) Scholar Award. K.B. was supported with an Action Cycling Atlanta (AV200) Fellowship, and E.H. is a Georgia Research Alliance Eminent Scholar. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Despite reducing viremia to levels below the limit of detection in standard assays, effective antiretroviral therapy (ART) does not eradicate cells latently infected with HIV-1. These cells serve as a reservoir for viral rebound if therapy is interrupted; thus, understanding the composition of the reservoir may yield further targets for HIV-1 cure strategies. We have taken a genetic approach to elucidating the reservoir in 13 Zambian subtype C seroconverters who were followed longitudinally through ART initiation and virologic suppression. In five of the 13 individuals, provirus sequences identical to or differing by five or fewer nucleotides from the transmitted/founder virus were detected, indicating archiving and persistence of early infection variants for more than six years following infection. While the majority of proviral variants in latently infected cells were most closely related to plasma virus circulating immediately prior to treatment initiation, additional variants dating to intermediate time points in the infection were also observed. These findings demonstrate that virus is archived during all stages of ART-naïve infection, and these variants persist throughout ART. HIV-1 cure strategies to eliminate the reservoir must address the broad genetic diversity of a within-host proviral quasispecies including variants archived from acute through chronic infection.

Introduction

Although over 23 million individuals living with HIV were receiving ART by the end of 2018, only two have been cured following stem-cell transplantation of HIV-1 resistant cells [1–4]. ART alone is not curative due to the persistence and proliferation of latently infected CD4+ T cells harboring intact but quiescent proviruses unaffected by antiretrovirals that target stages of active viral replication [5–11]. This long-lived and potentially self-renewing population of latently infected cells can serve as a reservoir for viral rebound in the event of treatment cessation [12–16], and efforts to understand the reservoir are therefore essential to HIV-1 cure strategies. Genetic approaches investigating the reservoir have sequenced rebounding virus in HIV-1 patients undergoing treatment interruption, reactivated virus from latently infected cells stimulated *in vitro*, and proviral populations during ART [14–26]. These studies provided critical insights into the complex nature of provirus remaining during treatment, only a small fraction of which is capable of replicating, and a further subset of which reactivates with treatment interruption [22]. The sources and establishment of this reservoir are of considerable interest, and although the reservoir is seeded beginning in very early stages of infection [5, 12, 14, 20, 27–29], few investigations have explored the relationship of the reservoir to early infection, pre-therapy viral variants. The extent to which these variants may persist in the reservoir during virologically suppressive ART is incompletely understood.

Recent studies examining associations between transmitted virus, its descendent quasispecies in chronic ART-naïve infection, and the reservoir include genetic analyses of amplified virus by Brodin et al. [30] and Abrahams et al. [31], particularly from quantitative viral outgrowth assay (QVOA) as in Abrahams et al. [31], while additional work from Jones et al. [32] infers age of latent proviral genes in relation to pre-therapy plasma variants. All three groups describe heterogeneous populations of proviral sequences that do not indicate ongoing evolution during virologically suppressive ART [30–32]. Furthermore, all three studies observe proviral variants that are inferred to be most closely related to sequences circulating in the plasma immediately prior to the start of treatment as well as variants contemporaneous with the earliest pre-therapy sample [30–32]. The frequencies of proviral variants dating to particular pre-therapy eras are distinct in each

study, with Brodin et al. [30] and Abrahams et al. [31] describing 60% and 71%, respectively, of proviral variants during treatment as most closely related to pre-therapy sequences from immediately prior to treatment initiation, while these frequencies are higher than that described in Jones et al. [32]. Given the interpatient variability in proviral population structure present in these studies, it is perhaps unsurprising for discrepancies in findings as well.

In this study, we examined reservoir proviral sequences in the context of pre-therapy plasma HIV-1 RNA diversity and evolution in 13 Zambian seroconvertors. Critically, our reconstruction of within-host HIV-1 evolution includes the inference of the near full-length transmitted/founder virus from single genome amplification, allowing us to investigate the possible long-term persistence of this sequence within the reservoir. Utilizing the phylogenetic approaches developed by Jones et al. [32] and additional analyses to assess the reservoir during short-term ART, our findings indicate that latent proviral diversity broadly reflects plasma HIV-1 RNA diversity during the period of pre-therapy infection. A majority of variants appear most closely related to those circulating in plasma near the time of ART initiation, but the reservoir quasispecies can in some individuals include variants present at the time of transmission, and demonstrates persistence of variants archived throughout ART-naïve infection.

Results

Participant selection and sampling methods

We identified 13 Zambian seroconvertors from the Zambia-Emory HIV Research Project for study according to the following criteria: ART-naïve infection of at least two years, subsequent ART with viremia <50 copies/mL at one or more time points following therapy initiation, and sample availability during treatment (Table 1). All participants received combination ART per country guidelines. We amplified and sequenced a minimum of seven near full-length genomes (NFLGs) by single genome amplification (SGA) from the earliest HIV+ plasma sample available for each participant (seroconversion sample), which was collected a median of 44 days post-estimated date of infection (EDI) (Table 1 and Fig 1). Sequencing was performed using Pacific Biosciences Single Molecule, Real-Time (SMRT) sequencing [33], and the transmitted/founder (TF) virus was inferred as the consensus of the low-diversity NFLGs (S1A Fig). Participant Z1658F was determined to have been infected with two TF viruses (S1 Fig). We additionally utilized SGA and SMRT sequencing to amplify and sequence an approximately 3.6 kb amplicon spanning the *vpu*, *env*, and *nef* genes from plasma samples collected approximately one year following the EDI, as well as from the last pre-therapy time point available (Fig 1). For six individuals, we additionally amplified and sequenced proviral DNA from the last pre-therapy time point to assess the divergence of sequences from the TF virus when collected from the cellular compartment rather than plasma.

For all participants, the *vpu-env-nef* amplicon was also generated by SGA from DNA of peripheral white blood cells collected during treatment; for four individuals, proviral sequences were also sampled at a second, later time point during ART (Table 1 and Fig 1). All sequences analyzed across time points were free of obvious defects such as nonsense mutations, INDELs resulting in disruption of an open reading frame, or in-frame INDELs of more than 90 nucleotides (up to 90 accepted to accommodate Env variable loops). In total, 1,275 sequences, excluding APOBEC hypermutants, were generated and analyzed for all participants and time points (S1 Table).

Infection date estimates

Although participants studied in this investigation were routinely tested for HIV, providing narrow windows for infection between the last HIV negative and first positive tests, estimation

Table 1. Participant details.

Participant ID	Last HIV- Test Date	First HIV+ Test Date	Estimated Date of Infection (EDI) ^a	EDI to SC Sample (days)	EDI to ART Start (years)	ART Regimen	Time on ART to ART Sample 1, 2 (years)
Z1094F	08-Apr-2008	04-Jul-2008	21-May-2008	44	2.75	EFV, TDF, FTC	0.48
Z1123M	14-Mar-2008	12-Apr-2008	02-Apr-2008	10	2.86	EFV, TDF, FTC	0.91
N133M	16-May-2007	16-Aug-2007	01-Jul-2007	46	2.87	EFV, TDF-FTC	0.56, 1.51
Z1788F	10-Nov-2006	10-Feb-2007	26-Dec-2006	46	2.98	EFV, TDF, FTC	0.72, 2.09
Z634F	22-Jun-2007	24-Jul-2007	08-Jul-2007	16	3.11	EFV, TDF-FTC	0.75, 1.40
Z2006M	09-Apr-2009	13-May-2009	26-Apr-2009	17	3.32	NVP, AZT, 3TC/EFV, TDF, FTC	0.32
Z1047M	24-May-2007	14-Aug-2007	31-Jul-2007	24	3.66	EFV, TDF, FTC	0.50
Z1165M	08-Dec-2005	16-Mar-2006	26-Jan-2006	49	3.82	ABC, 3TC, EFV	1.36, 2.14
Z1658F	11-Mar-2006	08-Jun-2006	24-Apr-2006	45	4.63	ABC, 3TC, NVP	0.89
Z1808F	18-Dec-2007	24-Jan-2008	10-Jan-2008	14	4.68	EFV, TDF, FTC	0.31
Z1124F	15-Feb-2006	18-May-2006	02-Apr-2006	46	5.22	ABC, 3TC, NVP	0.54
Z1044M	29-Nov-2005	25-Mar-2006	26-Jan-2006	58	5.23	AZT, 3TC, EFV	0.33
Z326M	21-Nov-2006	17-Feb-2007	04-Jan-2007	44	6.22	EFV, TDF, FTC	0.48
MEDIAN	NA	NA	NA	44	3.66	NA	0.54, 1.80

N = Ndola, Zambia; Z = Lusaka, Zambia; M = Male; F = Female; SC = seroconversion

^aSee materials and methods for calculation of EDI

<https://doi.org/10.1371/journal.ppat.1008378.t001>

of the infection date within these time frames is calculated as the midpoint of the antibody test dates. To corroborate our clinically-based EDIs from HIV testing in the single-variant infections, we used Bayesian approaches to infer the mean root date, a proxy for infection date, directly from sequence data. We used all pre-therapy *env* sequences without evidence of hypermutation or recombination for the analysis, annotated with the sample collection dates. Though Bayesian-estimated mean root dates predated the clinical EDI in all cases, the 95% confidence intervals of the Bayesian root dates overlapped the clinical EDIs in all but four infections (Fig 2). These results confirm the early nature of the infections when sampled at the seroconversion time point.

Reservoir variants are distributed throughout the phylogenetic trees of viral sequences from each individual

For all 13 participants, a maximum-likelihood phylogeny was inferred from within-host alignments of plasma RNA and proviral DNA *vpu-env-nef* sequences from all time points, where the tree was rooted on the inferred TF virus (Fig 3 and S2 Fig). This permits insight into the divergence of all descendant variants in each population from the TF virus, as well as the inferred evolutionary relationships between them. Divergence of plasma sequences from the

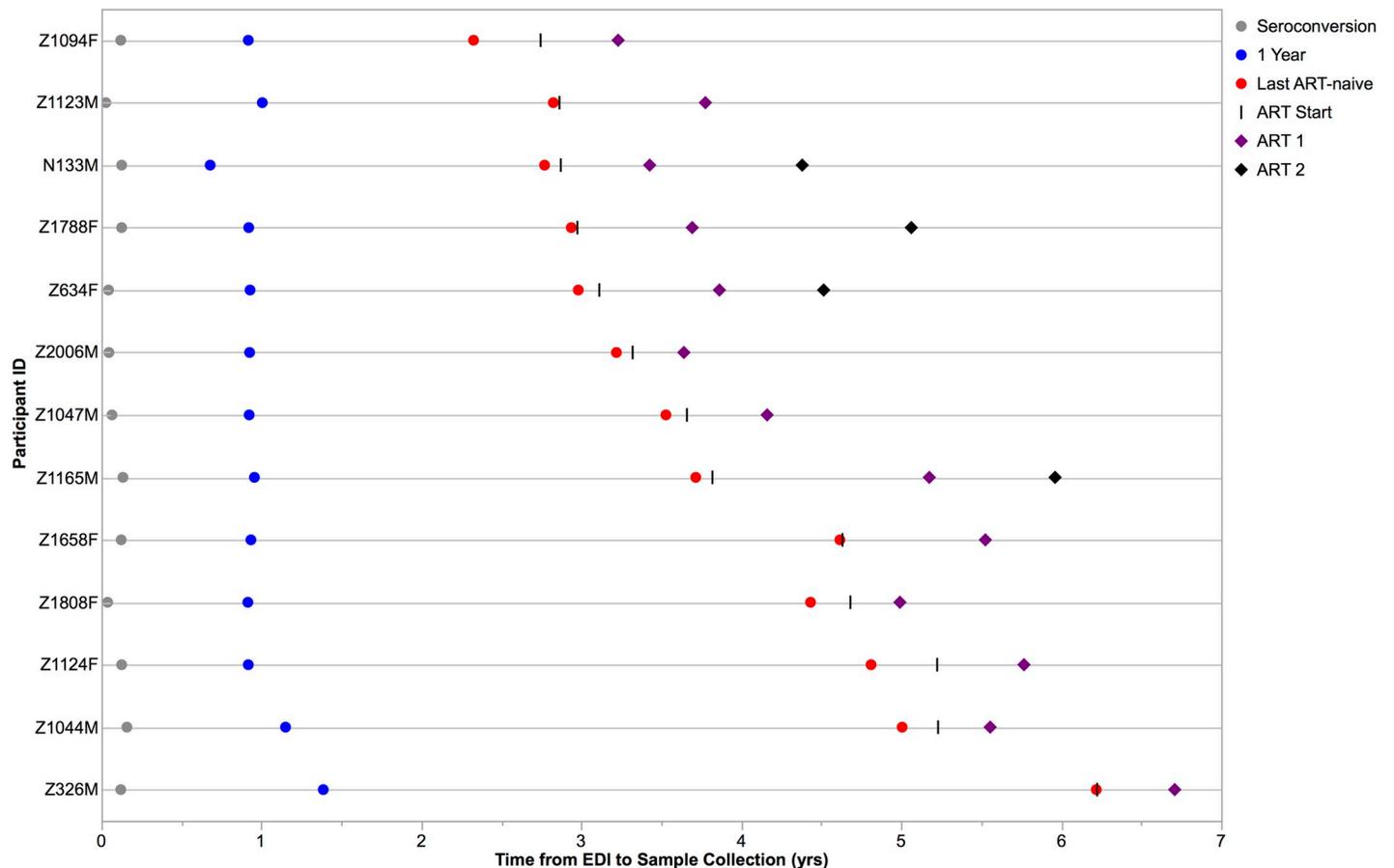


Fig 1. Sampling strategy for study participants. Viral RNA was reverse transcribed, amplified, and sequenced from pre-therapy plasma samples, while proviral DNA was amplified from cells collected during treatment. For six individuals (IDs Z1094F, Z1123M, Z1788F, Z1047M, Z1165M, and Z1658F), we additionally amplified and sequenced proviral DNA from cells at the last pre-therapy time point (red) to assess divergence from the transmitted/founder virus of contemporaneous sequences from plasma vs. cells. EDI = Estimated Date of Infection.

<https://doi.org/10.1371/journal.ppat.1008378.g001>

TF virus by patristic (root-to-tip) distances increases significantly from one year post-infection to the last ART-naïve time point, yet there is a significant decrease in the patristic distances of proviral sequences sampled during treatment compared to those sampled at the last ART-naïve time point (Fig 4B). This decrease does not appear to be an artifact of analyzing sequences from cells versus plasma alone, as ART sequences are significantly closer to the TF virus than sequences from cells at the last ART-naïve time point (S3 Fig). Sequences with minimal divergence from the TF virus are persisting in the reservoir, as can be seen for participants Z1123M and Z326M (Fig 3A and 3C), where sequences identical to the TF were recovered in the reservoir, having persisted for over two years and six years of ART-naïve infection, respectively.

Although proviral variants closely related to and including the TF virus persist in the reservoir, reservoir variants are distributed throughout the viral phylogenies for each individual. Plasma variants from one year post-EDI and the last pre-therapy time point exhibited a "ladder-like" topology characteristic of within-host phylogenies [34], where plasma sequences from a given time point formed distinct clades in all inpatient trees, and reservoir variants fell within or between these clades. In an initial analysis, we classified individual reservoir variants as being most closely related to variants of the clade within which they fall, or as

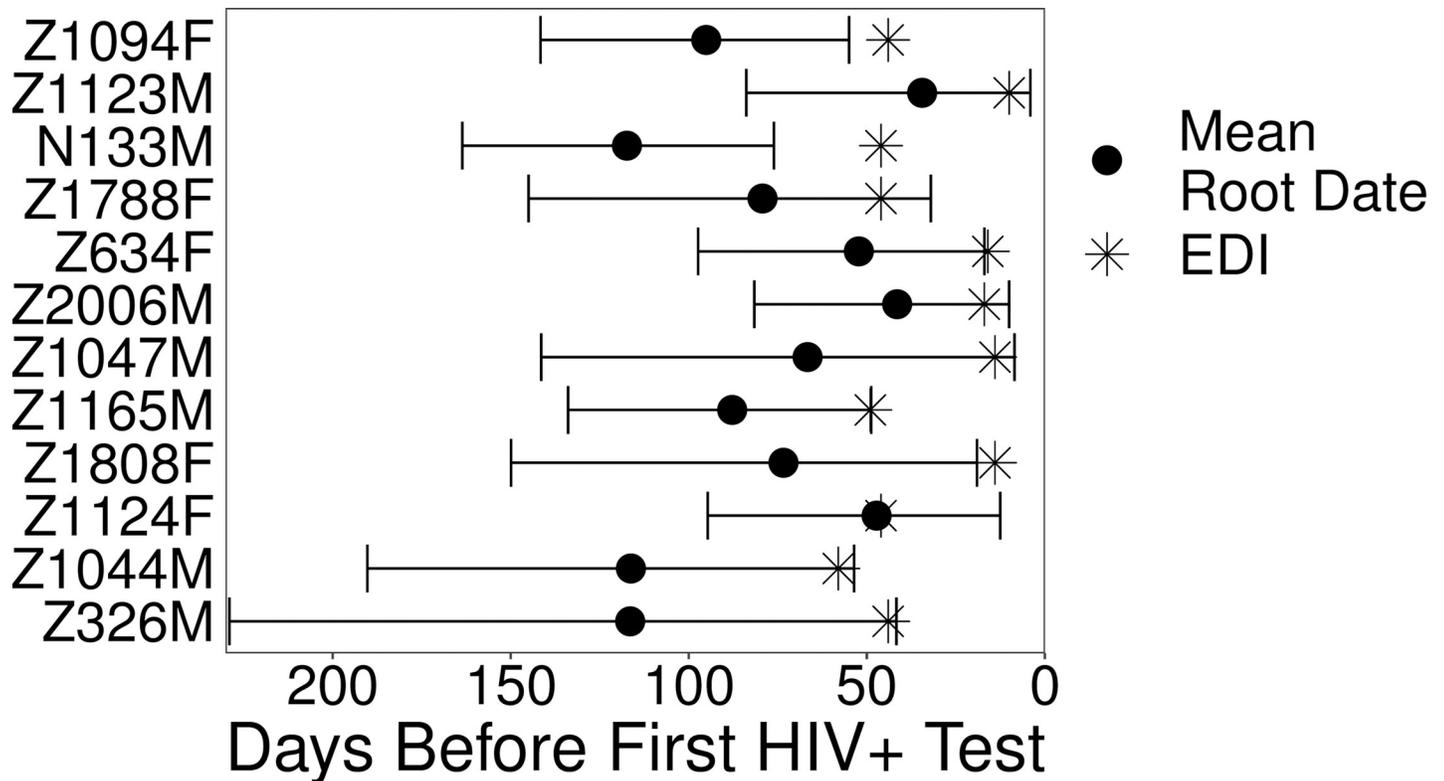


Fig 2. Bayesian root date estimation. Bayesian inference was used to estimate the root date, or time of infection, from pre-therapy plasma *env* sequences. The 95% highest posterior density intervals surround the mean estimated root dates of the Bayesian inference (circles). The clinically estimated infection dates (EDIs) are depicted as stars.

<https://doi.org/10.1371/journal.ppat.1008378.g002>

intermediate in cases where they do not fall within the one-year or last ART-naïve clades (Fig 4C, S4 Fig). Additionally, variants that did not fall within either the one year post-infection or last ART-naïve clade and were within six nucleotide changes from the TF virus were classified as seroconversion variants, as they exhibit divergence from the TF virus equivalent to the level of diversity among sequences from the seroconversion time point. In 10 of 13 individuals, we observed at least one proviral variant classified as most closely related to seroconversion or one year post-EDI plasma variants, while in all participants we observed proviral variants most closely related to the last ART-naïve plasma variants, with these sequences making up the greatest portion of the total proviral populations overall (Fig 4C). Taken together, we consider the presence of TF or seroconversion variants, as well as variants classified as intermediate and chronic, to indicate archiving of viral sequences throughout infection.

Dating of provirus integration indicates variant archiving throughout infection

While visualization of within-host phylogenies inferred from pre-therapy plasma sequences and proviruses persisting during ART allowed us to estimate the era in which the latter integrated into the reservoir, to more precisely estimate the age of proviral variants with respect to the ART-naïve infection, we applied the method developed by Jones et al. [32]. This method utilizes pre-therapy plasma sequences to develop a model of within-host evolution relative to sampling time, and places reservoir variants at a distinct date along the infection history, rather than into a broad category based on relatedness to discrete pre-therapy sampling. For this

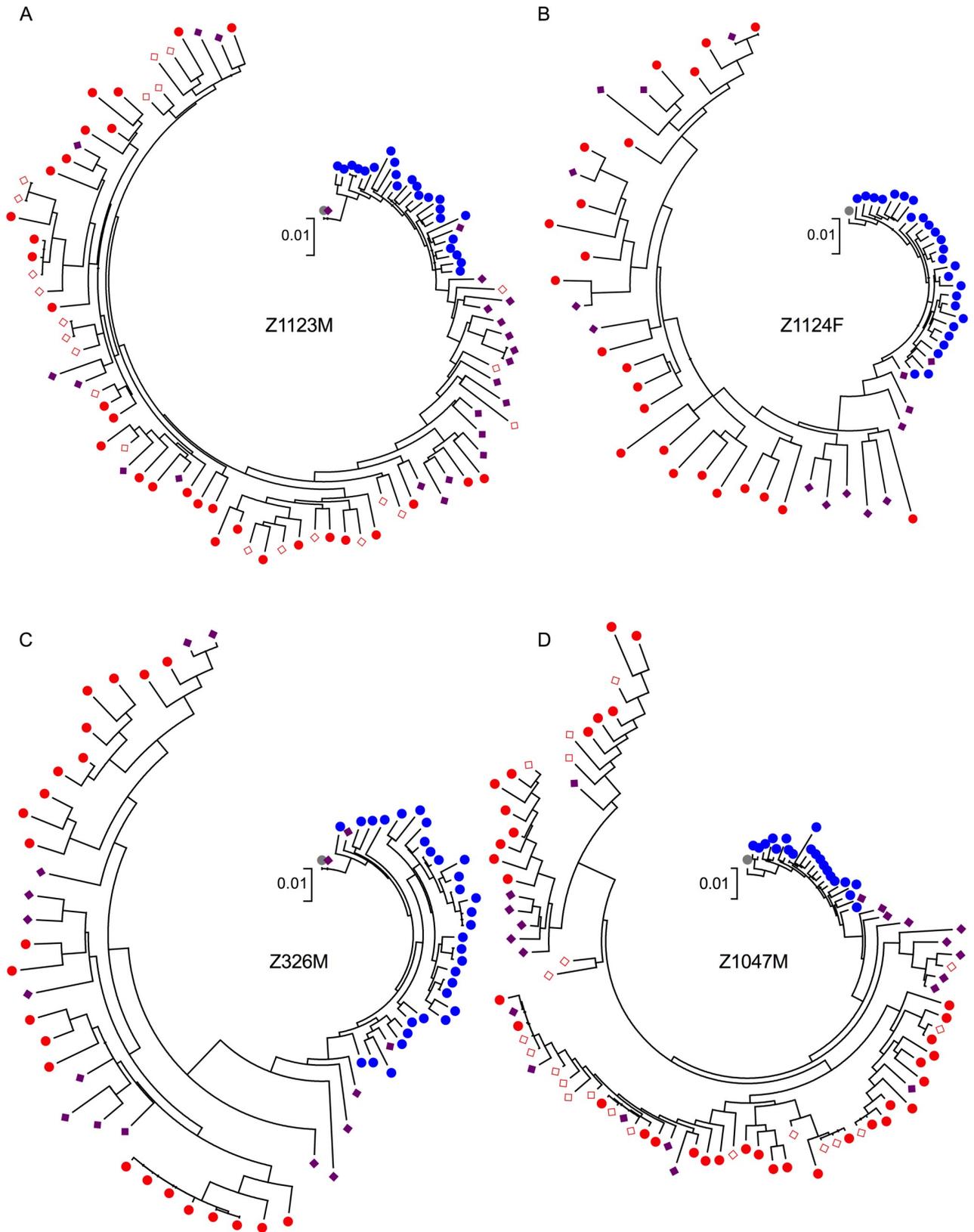


Fig 3. Maximum-likelihood (ML) trees for viral and proviral variants in four individuals. Representative ML phylogenetic trees for participants Z1123M (A), Z1124F (B), Z326M (C), and Z1047M (D) rooted on the respective transmitted/founder (TF) virus (grey) identified from the seroconversion sample and depicting all viral variants from one year post-infection (blue), the last ART-naïve sample (red), and during treatment (purple diamonds). Variants from cells collected at the last ART-naïve time point are shown in open red diamonds, while all plasma variants are in filled circles.

<https://doi.org/10.1371/journal.ppat.1008378.g003>

analysis, we inferred maximum-likelihood phylogenies from pre-therapy plasma and reservoir proviral *env* sequences, where the root was placed at the location that maximized the correlation between the divergence from the root and sample collection date of the pre-therapy plasma sequences (Fig 5A and 5D, S5 Fig). The pre-ART plasma variants were then used to train a linear model that related their divergence from the root and their sample collection dates. The linear model was used to infer the integration date and 95% confidence interval of the age of each proviral variant based upon its divergence from the root (Fig 5B, 5C, 5E and 5F, S5 Fig).

Consistent with reservoir proviral variants being seeded at various times spanning infection to treatment initiation, there is considerable discrepancy between sample collection dates and inferred integration dates for proviral sequences sampled during ART, with some variants estimated to have been integrated near the time of seroconversion. In a representative case, the point estimates of integration dates for Z1165M indicate a variant was archived within three months of the root date of Feb 21, 2006. In contrast, several variants displaying considerably higher divergence from the root were present as well, including those with estimated integration dates consistent with the last ART-naïve plasma variant date estimates. In participant Z634F, there are two variants dating to approximately one year post-infection, but no earlier variants, and only provirus dating to shortly prior to the initiation of therapy was detected in participants Z1044M and Z1808F (S5 Fig). Interestingly, where cells were sampled at the last ART-naïve time point, such as for participant Z1165M, the integration date estimates for provirus of these cells typically fell slightly after that of the proviral variants persisting during treatment. Across the group of participants, however, and consistent with our initial analysis in Fig 4C, integration date estimates for proviral sequences supported periodic seeding of variants in the reservoir throughout the infection.

Repeated sampling during ART and persistence of early infection variants

The dynamics of proviral decay during short-term ART influence the results of this investigation, as proviral DNA decays most rapidly within approximately the first one to two years of therapy [35–38]. By sampling during this time frame, we therefore may be sampling provirus that persists only transiently rather than comprising the more stable population of latently infected cells with a slower decay rate. To determine if within-host proviral composition was influenced by the relatively short time on treatment, we sampled an additional time point six months to a year later in four participants. All proviral variants without APOBEC hypermutation were included in phylogenetic trees along with all pre-therapy sequences, and phylogenies were again rooted on the TF virus (Fig 6). Early variants were observed throughout the repeated sampling during ART, as proviral variants classified as seroconversion variants were in both the first and second time point during treatment for participant N133M (Fig 6C). For participant Z1165M, a single proviral sequence most closely related to seroconversion sequences was observed in the second time point during treatment (Fig 6D), and in participant Z1788F, seroconversion and early infection variants differing from the TF virus by up to approximately 30 nucleotides were found during the first and second time points following treatment (Fig 6A). Overall, however, proviral sequences from both time points during treatment were intermingled with each other and the sequences sampled prior to treatment.

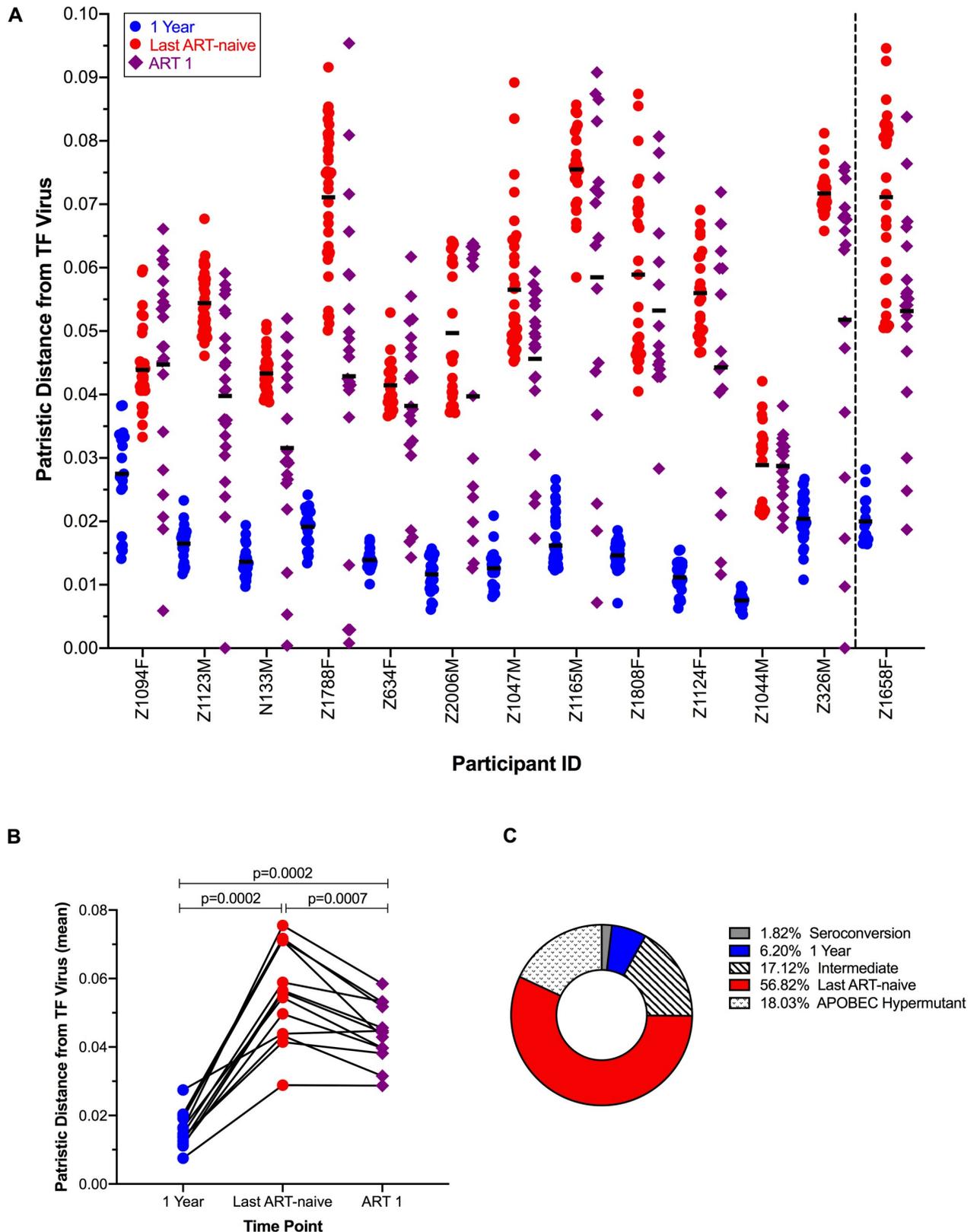


Fig 4. Divergence from transmitted/founder (TF) virus increases during ART-naïve infection and decreases on treatment. (A) Patristic distances from the TF virus, or root-to-tip values, from the maximum-likelihood (ML) trees are shown for all 13 participants, with the single instance of multivariant infection in participant Z1658F shown at the far right. Where two samples were assessed during treatment, only the first is shown here. Mean values are indicated by horizontal black bars. (B) Summary of the mean inpatient patristic distances, where the mean distance is significantly different between each time point assessed (Wilcoxon matched-pairs signed rank tests). (C) Proportion of proviruses seeded into the reservoir, by era,

as estimated from the placement of reservoir sequence in the phylogeny. The mean values of proportions for all 13 volunteers are shown, while proportions for each participant are given in S4 Fig. A majority of variants are most closely related to variants present at the last ART-naïve time point. The percentage of variants demonstrating APOBEC hypermutation is also indicated, though these sequences were excluded from analysis.

<https://doi.org/10.1371/journal.ppat.1008378.g004>

To extend this analysis, we formally compared proviruses sampled at both time points during ART with respect to their genetic divergence from the TF virus. To facilitate combining data across participants, root-to-tip or patristic distances of each reservoir sequence were normalized to the participant's total tree height. Comparison of these scaled root-to-tip distances by sampling time point during ART revealed shorter mean distances for the later reservoir samples compared to the earlier ones (unpaired t test, $p = 0.1054$, Fig 7). This suggests that, with ongoing treatment, viral variants are not continuing to evolve, as this would bring about an increase in patristic distance. Furthermore, it suggests that early viral variants may be enriched in the reservoir during ART.

Discussion

We observed that proviral sequences from 13 individuals who had undergone short-term ART were distributed among pre-therapy sequences in phylogenetic trees, with the majority of proviral sequences most closely related to variants from the last ART-naïve time point. However, as analysis of the estimated time of integration for proviral sequences indicates, there is archiving of variants throughout ART-naïve infection, from the earliest time of infection to treatment initiation. This finding is consistent with previous work by Jones et al. [32], but extends the stages of pre-therapy infection explored to acute infection. We identified TF viruses from acute infections with longitudinal follow-up through chronic ART-naïve infection and treatment initiation, while the pre-therapy samples of the two HIV-1 infections investigated in Jones et al. are from chronic infections [32]. Archiving of variants throughout ART-naïve infection is complementary to the observations that the reservoir is smaller and less diverse in individuals beginning treatment early in infection versus during chronic infection [37, 39, 40], since preventing replication with ART ensures a halt in viral evolution and concomitant latent infection with progressively more diverse variants.

Within the diverse populations of proviral sequences we observed, we identified variants that were identical to or contemporaneous with the TF virus after as many as six years of ART-naïve infection and following six to 24 months of ART. These very early, TF-related sequences were observed in five of the 13 individuals sampled and represented from 2.6–7.5% of all reservoir variants in those individuals. It is clear that these very early viral sequences can persist for several years in the absence of therapy, consistent with their integration in long-lived CD4+ T cells. Persistence of ancestral variants is not unprecedented, as several studies assessing drug resistance in patients receiving virologically suppressive ART after a history of non-suppressive therapy found that both ancestral, drug susceptible virus and variants with resistance mutations persist during years of effective treatment [41–44].

Recent studies have shown that a majority of proviruses persisting during ART exhibit large internal deletions or other defects, such as nonsense mutations resulting from APOBEC-induced hypermutation, which render the provirus defective [22, 24]. Due to sample limitations, we assessed approximately one-third of the genome encompassing the *vpu*, *env*, and *nef* genes, and thus cannot exclude the possibility that sequences we have observed as exact matches to the TF virus in this amplicon might contain differences elsewhere in the genome, including mutations and/or deletions that would prevent viral replication. Nevertheless, all of the sequences used for analysis do represent biologically functional gene regions, since sequences with frameshifting INDELS or nonsense mutations were excluded. Unlike

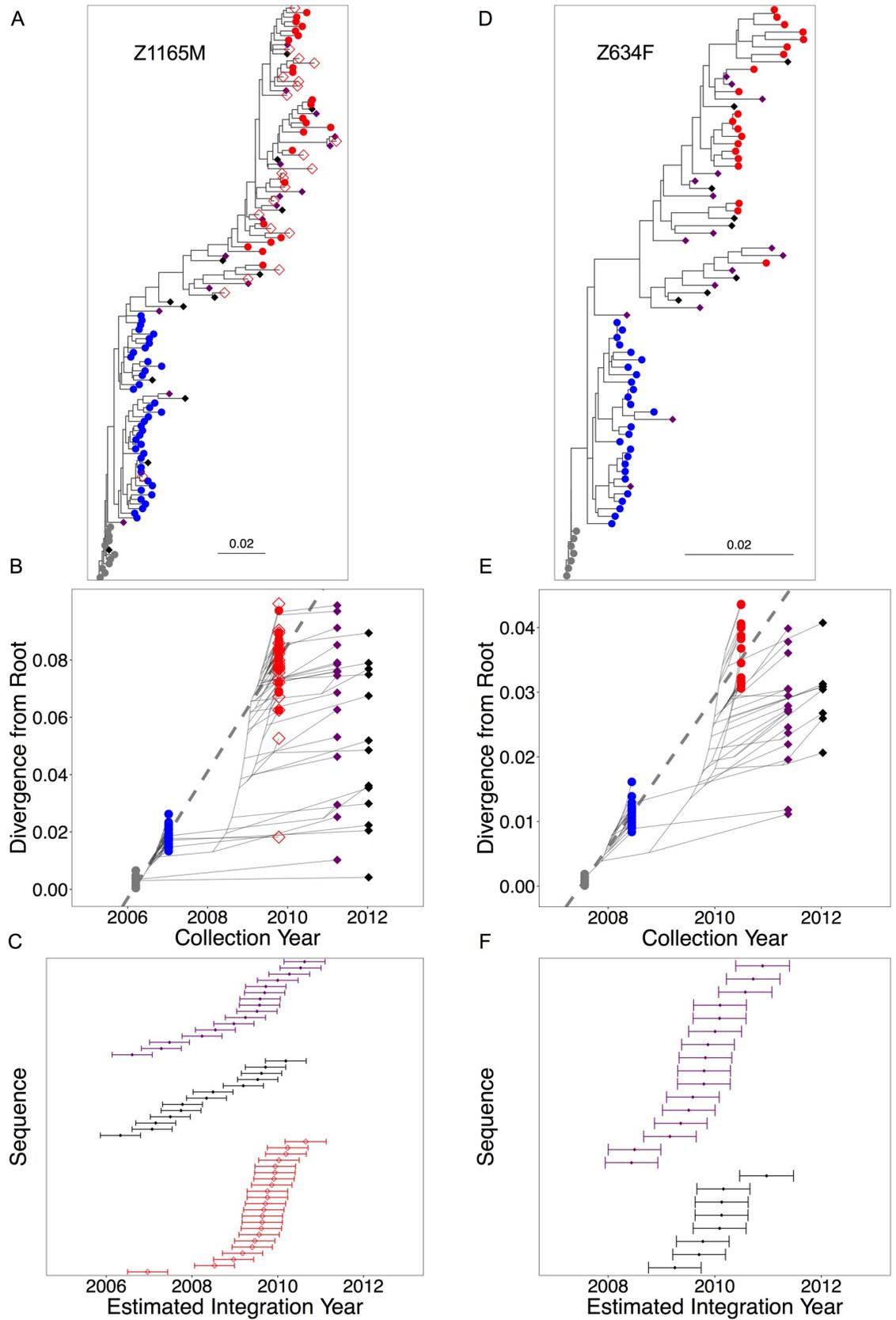


Fig 5. Regression-based inference of time of provirus integration. Representative figures for participant Z1165M (A-C) and participant Z634F (D-F). Maximum-likelihood trees of the *env* gene for pre-therapy variants (circles), including individual seroconversion variants (grey), one year sequences (blue), last pre-therapy sequences (red circles, plasma and open red diamonds, cells), and proviral variants (filled diamonds) in A and D. Two samples during treatment were assessed for both participants, with the first in purple, and subsequent in black. Trees were rooted to optimize the correlation between root-to-tip distance and sampling time for all pre-therapy plasma variants. The linear model relating root-to-tip distances to sampling time is shown in the dashed lines of figures B and E, with the pre-therapy variants denoted as colored dots, and the phylogenetic relationships between them denoted as faint grey lines. Proviral variants from samples collected during treatment are shown in filled diamonds in the same manner. The estimated integration dates of the proviral variants and 95% confidence intervals are shown in the plots C and F. Figures for additional individuals are in S5 Fig.

<https://doi.org/10.1371/journal.ppat.1008378.g005>

Abrahams et al. [31], who used QVOA to characterize sequences reactivated *in vitro*, we are not exclusively addressing the replication-competent reservoir. However, QVOA are known to underestimate the size of the reservoir, as the bulk of replication-competent proviruses are not induced with single or successive rounds of stimulation [22]. Phylogenetic assessment of HIV-1 DNA during virologically suppressive ART serves to address the broad population of persistent provirus within which the replication-competent reservoir is contained, and address its relationship to pre-therapy virus.

In addition to containing TF virus or very early infection variants in some individuals, reservoir proviral populations were overall less evolved from the TF virus than the sequences at the last ART-naïve time point (Fig 4). This finding may be influenced by the short duration of treatment, as all individuals studied here received ART for less than three years at the time of sample collection during treatment, and three participants were sampled within six months of treatment initiation while the reservoir is less stable. However, we did find that early infection variants persisted with continued time on treatment in individuals sampled twice while receiving ART. Furthermore, sequential sampling indicated that with continued time on treatment, the distance of reservoir variants from the TF virus decreased (Fig 7), indicating a potential enrichment for variants dating to earlier in the course of the infection. As viremia rapidly declines in the first phase of viral decay following treatment initiation, followed by a second, slower decay phase [45], latently infected cells decay in stages [35–38], perhaps with those infected most recently by variants circulating in the plasma just prior to treatment initiation decaying first. This mechanism would be consistent with the observation that CD4+ central memory T cells from four to eight years of ART harbor HIV-1 DNA more closely related to early infection sequences than HIV-1 DNA of shorter-lived CD4+ effector memory T cells, in which there is a more prominent decline of HIV-1 DNA with continued time on treatment [23]. However, the relationship between CD4+ differentiation status and the ages of proviruses persisting during ART is by no means clear [46]. Further studies must address the phylogenetic influence of latently-infected cell decay.

As HIV-1 prevention and treatment efforts are scaled up globally, research efforts to reduce and/or eliminate the reservoir in pursuit of an HIV-1 cure are expanding as well. Towards this goal, it is critical to characterize the genetic diversity of the reservoir to assess the variants that HIV-1 eradication strategies must target. Our findings indicate that virus is archived throughout infection, and cure strategies should therefore address the genetic diversity of reservoir proviral quasispecies with many unique variants, including those dating back to the time of infection.

Materials and methods

Human subjects

Zambian volunteers were enrolled as heterosexual couples in Couples Voluntary Counseling and Testing (CVCT), with HIV testing and counseling of both partners conducted upon

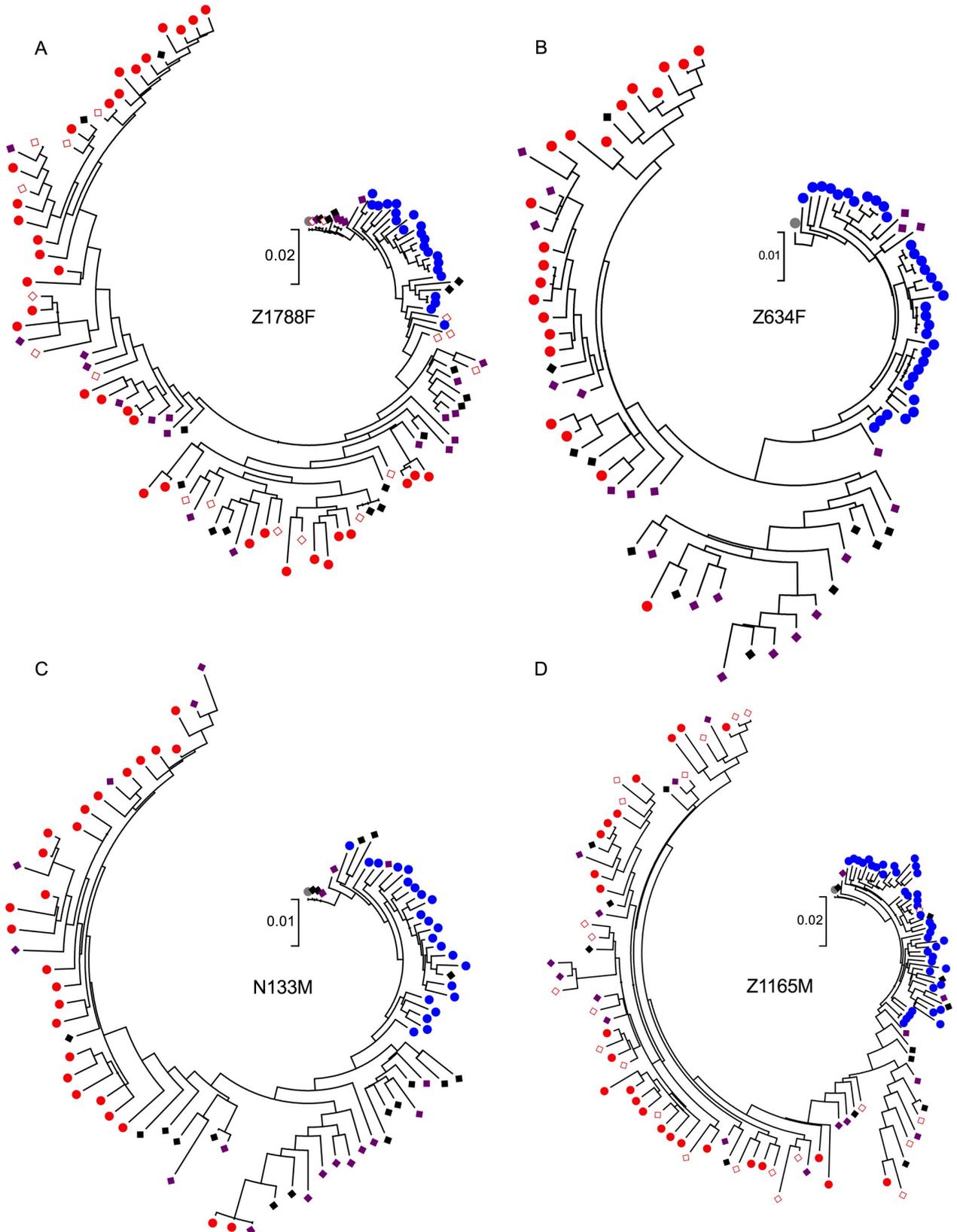


Fig 6. Maximum-likelihood (ML) trees of all variants for participants with two samples during treatment. ML phylogenetic trees for all four participants: Z1788F (A), Z634F (B), N133M (C), and Z1165M (D) rooted on the respective transmitted/founder virus (grey) and depicting all viral variants from one year post-infection (blue), the last ART-naïve sample (red), and during treatment (purple and black diamonds, with second sample in black). Sequences from cells collected at the last ART-naïve time point are in open red diamonds, while plasma sequences are depicted in circles.

<https://doi.org/10.1371/journal.ppat.1008378.g006>

enrollment. Follow-up HIV testing was conducted approximately every three months for the negative partners of serodiscordant couples, and blood samples were collected from both partners in the event of a positive test as a component of the Zambia-Emory HIV Research Project (ZEHRP). Dates for ART initiation were self-reported, and clinically-based estimated dates of infection (EDIs) were calculated with the appropriate formula of the three following: 1. mid-point of dates for the last antibody negative and first antibody positive test; 2. Fourteen days prior to the first p24 antigen positive, antibody negative test; 3. Ten days prior to the first viral load >1600 copies/mL, antibody negative test. All participants had antibody positive tests for their first HIV+ test, with the exception of participants Z1123M, Z1047M, and Z1808F. EDIs for participants Z1047M and Z1808F were estimated as in formula 2 above, and Z1123M EDI was calculated according to formula 3. For participant Z1047M, plasma from the first HIV + test was not available for HIV sequence amplification, and therefore a sample from 10 days

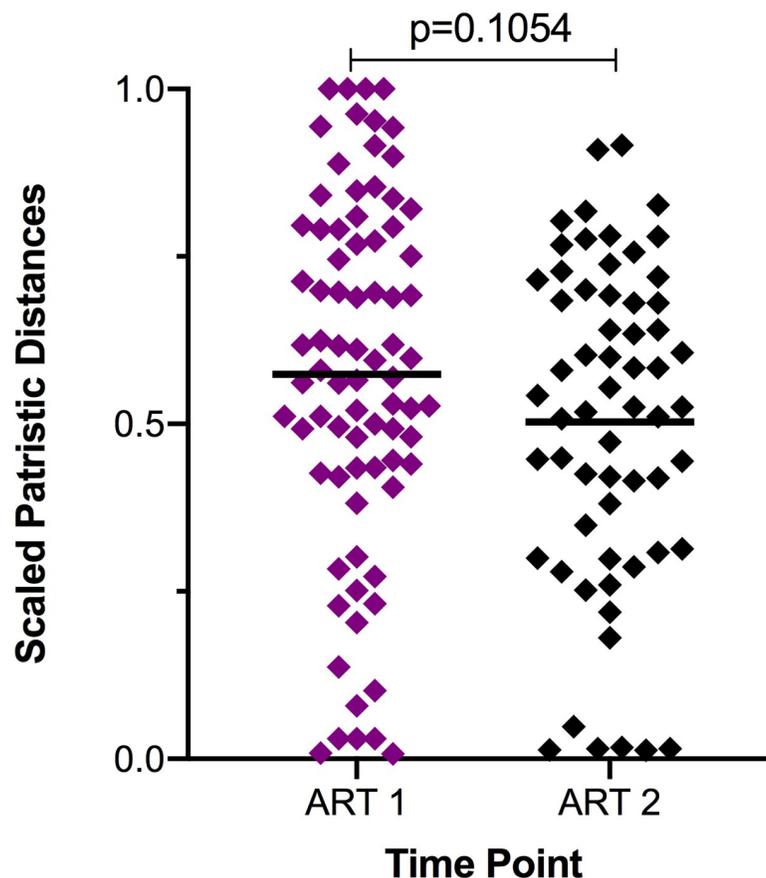


Fig 7. Distance from transmitted/founder (TF) virus decreases with subsequent sampling during treatment. Patristic distances from the TF virus, or root, for reservoir variants as a proportion of the greatest intrapatient patristic distance (tree height) from the maximum-likelihood phylogenetic tree; means shown in horizontal black bars. Distances are lower for variants sampled at the second time point during treatment compared to the first (unpaired t test).

<https://doi.org/10.1371/journal.ppat.1008378.g007>

later on 24 Aug 2007 was used as the seroconversion sample. All other seroconversion samples were collected the day of the first HIV+ test.

Ethics statement

Human subjects protocols for ZEHRP were approved by the University Teaching Hospital Ethics Committee in Lusaka, Zambia, while additional approval for sample or data use was granted by the Institutional Review Boards of Emory University, Simon Fraser University, and Providence Health Care/University of British Columbia. Written informed consent for sample collection was obtained for each volunteer upon enrollment in CVCT.

Nucleic acid extraction and cDNA synthesis

Viral RNA in plasma samples was extracted using the QIAamp Viral RNA Mini Kit (QIAGEN) or E.Z.N.A Viral RNA Kit (Omega Bio-Tek) according to the manufacturer's instructions. Briefly, 150 μ L plasma was lysed with buffer and centrifuged through a silica column, which was then washed with appropriate buffers. RNA was eluted in >60 μ L Buffer AVE (QIAGEN) or DEPC H₂O (Omega). RNA served as template in cDNA synthesis reactions described below.

Eleven microliters of viral RNA were used in each 20 μ L reverse-transcriptase reaction for cDNA synthesis utilizing SuperScript III or IV Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions, but with an extension time of up to one hour. SuperScript III protocols additionally included a 4°C pause following the one hour extension time for addition of 200 Units RT enzyme proceeding a second extension for two hours at 55°C. Both SuperScript III and IV protocols included RNase H digestion of RNA-DNA heteroduplexes with 20 min incubations at 37°C. Oligo dT (5'-TTTTTTTTTTTTTTTTTT-3') or 1.3'3'PICb (5'-ACTACTTAGAGCACTCAAGGCAAGCTTTATTG-3') primers were used as anchors in the reactions, and cDNA was directly used in PCR or frozen for subsequent use.

Nucleic acids were extracted from cells using the QIAamp DNA Blood Mini or Midi Kit (QIAGEN) according to the manufacturer's instructions. Briefly, white cell pellets of total white blood cells in RNAlater were processed for lysis in QIAGEN Protease or Proteinase K and lysis buffer. Following addition of 100% ethanol, lysate was applied to a silica column and centrifuged, following by washing of the column. Samples were eluted in QIAGEN buffer AVE and used directly in PCR or frozen for subsequent use.

PCR and amplicon purification

PCR for single genome amplification (SGA) of near full-length genomes (NFLGs) consisted of two rounds of PCR utilizing appropriate template for $\leq 40\%$ positive reactions of approximately nine kilobases as visualized by gel electrophoresis. Each round of PCR consisted of 25 μ L reactions with 0.5 Units Q5 Hot Start High-Fidelity Enzyme (NEB), 1x Q5 Reaction Buffer, 1x Q5 High GC Enhancer, 350 μ M each dNTP, 500 nM each primer, plus template and nuclease-free H₂O to reach 25 μ L. PCR primers are described in Rousseau 2006 [47] for both first and second rounds, and first round primers are as follows: 1.U5Cc (5'-CCTTGAGTGCTCTAAGTAGTGTGTGCCCGTCTGT-3', forward primer) and 1.3'3'PICb (5'-ACTACTTAGAGCACTCAAGGCAAGCTTTATTG-3', reverse primer). Second round PCR primers are as follows, with 1 μ L of first round PCR product used as template in the second round PCR: 2.U5Cd (5'-AGTAGTGTGTGCCCGTCTGTTGTGTGACTC-3', forward primer) and 2.3'3'PICb (5'-TAGAGCACTCAAGGCAAGCTTTATTGAGGCTTA-3', reverse primer). Both first and second round PCR utilized the following program: 98°C for 30 sec, 35 cycles of 98°C for 10 sec and 72°C for 7:30 sec, 72°C for 10:00 min, and 4°C forever (end). Amplicons were purified

with the Wizard SV Gel and PCR Clean-Up System (Promega) according to manufacturer's instructions, eluting in H₂O.

For amplification of *vpu*, *env*, and *nef* gene amplicons, two rounds of PCR were used as above for NFLG amplification. Reactions were 20 uL with first round primers: Vif1 KB (5'-GGTTTATTACAGRGACAGCAGAG-3', forward primer) and Ofm19 (5'-GCACTCAAGGC AAGCTTTATTGAGGCTTA-3', reverse primer). First round PCR product (0.8 uL) was used as a template for second round PCR with the following primers: EA1F KB (5'-GCTTAGGCATY TCMTATGGCAGGAAGAAG-3', forward primer) and O1R (5'-AAAGCAGCTGCTTATA TGCAGCWTC-3', reverse primer). First round PCR program was as follows: 98°C for 45 sec, 30 cycles of 98°C for 15 sec, 60°C for 30 sec, and 72°C for 4:00 min, then 10 min at 72°C, and 4°C forever (end). Second round program was the same but for a 62°C annealing temperature and 3:00 min extension step. Amplicons were purified with the NucleoSpin Gel and PCR Clean-Up (Takara), eluting in Elution Buffer NE or H₂O.

Next-generation sequencing

All sequencing was performed with Pacific Biosciences SMRTbell sequencing on the RS II, with individual DNA libraries run on a single SMRT cell. Libraries were generated with 30–60 NFLG amplicons combined at equimolar concentrations and identified by nucleotide barcode following reamplification of first round PCR products with barcoded primers, or by barcoded adapter from the SMRTbell Barcoded Adapter Complete Prep Kit-96 (Pacific Biosciences). Libraries for *vpu*, *env*, and *nef* amplicons were generated with 80–100 amplicons per library and identified with barcoded adapters. Libraries were made according to manufacturer's instruction, followed by appropriate size selection with the BluePippin (Sage Science). We are greatly appreciative of library size selection and quality control, as well as sample run of libraries on the RS II performed at the University of Delaware Sequencing and Genotyping Center.

Amplicon reads generated from the SMRT sequencing were analyzed with a unique algorithm to perform read phasing and error correction in generation of final sequences [33]. Libraries generated using Pacific Biosciences barcoded adapters were first analyzed with PacBio SMRT analysis software PB Barcode to separate reads by barcoded adapter prior to additional read phasing and error correction with the algorithm described in Dileria et al. [33].

Phylogenetic trees and reservoir variant dating

Maximum-likelihood phylogenetic trees for complete amplicons lacking frameshifting INDELS, APOBEC hypermutation, or other deleterious mutations were made with the PhyML plugin [48] of Geneious software v9.0.4 [49] using a general time reversible model with six nucleotide substitution categories and gamma distribution parameter with 100 bootstraps. APOBEC hypermutants were first removed from analysis with the LANL Hypermut v2.0 tool [50] with the appropriate transmitted/founder (TF) virus as the reference sequence, and all sequences of $p < 0.05$ considered hypermutated. Trees were rooted on the appropriate TF virus sequence trimmed to the *vpu*, *env*, and *nef* gene amplicon. Patristic distances from the TF virus were extracted from the distance matrix. Trees were edited for visualization with MEGA v7.0.26 [51]. Statistics for patristic distance from phylogenetic trees were performed using Prism v8.3.0. Figures were made using Prism or, for Fig 1, JMP Pro 14 v14.2.

For all 12 participants with single-variant infections, we estimated the root date of their within-host plasma HIV-1 RNA sequences using established Bayesian methods. Briefly, within-host pre-therapy plasma HIV-1 *env* sequences were first screened for hypermutation (using Hypermut v2.0) and recombination (RDP v4.95 [52]) and any hypermutated or within-host recombinant sequences were removed. Sequences with ambiguous bases were also

excluded, and identical sequences discarded but for one sequence from the earliest time point at which it was sampled. We ran two parallel 100,000,000 length chains sampling every 10,000 states in the software package BEAST v1.10.4 [53] for each participant. Posterior distributions for the root date were estimated using the unlinked SRD06 substitution model [54], the uncorrelated relaxed lognormal clock models [55], and the coalescent GMRF Bayesian skyride tree model [56, 57]. After discarding 10–30% of the initial run as burn-in, the chains from parallel runs were combined with LogCombiner v2.5.2 [58] and analyzed in Tracer v1.7.1 [59] to ensure convergence and verify that effective sample size values were >200 for all parameters.

Proviral variant integration dates were estimated as previously described [32] for the 12 participants infected by a single transmitted/founder virus. Briefly, *env* genes were trimmed from seroconversion and all other pre-therapy variants, as well as proviral variants from samples collected during treatment. Any sequences demonstrating hypermutation, ambiguous bases, or recombination were excluded from analysis, and only the earliest variant of duplicate sequences was kept for analysis. Maximum-likelihood trees were generated with RAxML v8.2.12 [60] and trees were rooted with root-to-tip regression (RTT) using the R package ape v5.3 [61] to maximize the correlation between the divergence from root and the sample collection date of the pre-therapy sequences. The pre-therapy variants were used to train a linear model of the divergence from root and the sample collection date. Finally, the integration dates and confidence intervals of the proviral variants were estimated from this model.

Supporting information

S1 Table. Extended sampling details.

(PDF)

S1 Fig. Seroconversion sequences. (A) Maximum-likelihood phylogenetic tree of the near full-length HIV-1 genomes from all 13 study individuals at the seroconversion time point with color-coding of sequences from each participant. A single distinct clade and very short branch lengths within each participant viral population are indicative of the low sequence diversity, except in the case of participant Z1658F, where two clades are present. Sequences within each clade for Z1658F are low-diversity, consistent with infection being established by two transmitted/founder (TF) viruses. (B) Highlighter plot of the two viral populations in seroconversion sequences for Z1658F, with each TF virus as a master sequence. Polymorphisms matching TF virus A are shown in red, those matching TF virus B are in blue, and unique polymorphisms are in grey. Highlighter plot made with LANL Highlighter tool [62].

(PDF)

S2 Fig. Maximum-likelihood (ML) trees for five participants with one sample available during treatment. (A) Z1094F (B) Z2006M (C) Z1808F (D) Z1044M and (E) Z1658F. Trees for participants with one sample during treatment not shown in Fig 3 are shown here. Participant Z1658F was infected with two transmitted/founder (TF) viruses, both included in grey in the ML tree, which is rooted on a Zambian subtype C consensus sequence (black square). All other trees are rooted on the respective TF virus (grey) identified from the seroconversion sample and depict all viral variants from one year post-infection (blue), the last ART-naïve sample (red), and during treatment (purple diamonds). Sequences from cells collected at the last ART-naïve time point are shown in open red diamonds, while all plasma variants are in filled circles.

(PDF)

S3 Fig. Sequences during treatment are closer to transmitted/founder (TF) virus than last ART-naïve sequences. To compare distances across participants, each variant's patristic

distance from the TF virus or root is expressed as a proportion of the greatest patristic distance or branch length in a given participant's maximum-likelihood tree. Means are shown in horizontal black bars. The proportional or scaled distances of sequences during treatment are significantly lower than sequences from either the cells or plasma at the last ART-naïve time point (Mann-Whitney tests).

(PDF)

S4 Fig. Classifications of reservoir variants for each participant. Where sequences of the given era were not present and the percentage of the reservoir proviral population was therefore zero, the classification is omitted from the pie chart.

(PDF)

S5 Fig. Proviral variant integration date estimates for each participant. All trees, linear models, and variant integration date estimates not shown in [Fig 5](#) are provided here.

(PDF)

Acknowledgments

We would like to thank all the ZEHRP study participants, as this work would not have been possible without them. Additionally, we are grateful for the support of Jon Allen and Charlott Morel Sanchez at the Emory Vaccine Center for sample management, and Dr. Paul Farmer at the Emory Vaccine Center for technical assistance regarding the ZEHRP cohort. Furthermore, we are indebted to the efforts of Olga Shevchenko at the University of Delaware Sequencing and Genotyping Center for all Pacific Biosciences sequencing.

Author Contributions

Conceptualization: Kelsie Brooks, Eric Hunter.

Data curation: Kelsie Brooks, Dario A. Dilernia, Daniel J. Wilkins, William Kilembe.

Formal analysis: Kelsie Brooks, Bradley R. Jones.

Funding acquisition: Jill Gilmour, William Kilembe, Jeffrey B. Joy, Susan A. Allen, Zabrina L. Brumme, Eric Hunter.

Investigation: Kelsie Brooks, Bradley R. Jones, Daniel T. Claiborne, Samantha McNally.

Methodology: Bradley R. Jones, Jeffrey B. Joy, Zabrina L. Brumme, Eric Hunter.

Project administration: Eric Hunter.

Resources: Jill Gilmour, William Kilembe, Susan A. Allen, Eric Hunter.

Software: Dario A. Dilernia, Daniel J. Wilkins.

Supervision: Jeffrey B. Joy, Zabrina L. Brumme, Eric Hunter.

Writing – original draft: Kelsie Brooks, Eric Hunter.

Writing – review & editing: Kelsie Brooks, Bradley R. Jones, Daniel J. Wilkins, Daniel T. Claiborne, Samantha McNally, Jeffrey B. Joy, Zabrina L. Brumme, Eric Hunter.

References

1. UNAIDS. Fact sheet: global AIDS update 2019. 2019.

2. Hutter G, Nowak D, Mossner M, Ganepola S, Mussig A, Allers K, et al. Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. *N Engl J Med*. 2009; 360(7):692–8. Epub 2009/02/14. <https://doi.org/10.1056/NEJMoa0802905> PMID: 19213682.
3. Allers K, Hutter G, Hofmann J, Loddenkemper C, Rieger K, Thiel E, et al. Evidence for the cure of HIV infection by CCR5Delta32/Delta32 stem cell transplantation. *Blood*. 2011; 117(10):2791–9. Epub 2010/12/15. <https://doi.org/10.1182/blood-2010-09-309591> PMID: 21148083.
4. Gupta RK, Abdul-Jawad S, McCoy LE, Mok HP, Peppas D, Salgado M, et al. HIV-1 remission following CCR5Delta32/Delta32 haematopoietic stem-cell transplantation. *Nature*. 2019; 568(7751):244–8. Epub 2019/03/06. <https://doi.org/10.1038/s41586-019-1027-4> PMID: 30836379.
5. Finzi D, Hermankova M, Pierson T, Carruth LM, Buck C, Chaisson RE, et al. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science (New York, NY)*. 1997; 278(5341):1295–300. Epub 1997/11/21. <https://doi.org/10.1126/science.278.5341.1295> PMID: 9360927.
6. Wong JK, Hezareh M, Gunthard HF, Havlir DV, Ignacio CC, Spina CA, et al. Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. *Science (New York, NY)*. 1997; 278(5341):1291–5. Epub 1997/11/21. <https://doi.org/10.1126/science.278.5341.1291> PMID: 9360926.
7. Chun TW, Stuyver L, Mizell SB, Ehler LA, Mican JA, Baseler M, et al. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proceedings of the National Academy of Sciences of the United States of America*. 1997; 94(24):13193–7. Epub 1997/12/16. <https://doi.org/10.1073/pnas.94.24.13193> PMID: 9371822; PubMed Central PMCID: PMC24285.
8. Chomont N, El-Far M, Ancuta P, Trautmann L, Procopio FA, Yassine-Diab B, et al. HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nature medicine*. 2009; 15(8):893–900. Epub 2009/06/23. <https://doi.org/10.1038/nm.1972> PMID: 19543283; PubMed Central PMCID: PMC2859814.
9. Maldarelli F, Wu X, Su L, Simonetti FR, Shao W, Hill S, et al. HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science (New York, NY)*. 2014; 345(6193):179–83. Epub 2014/06/28. <https://doi.org/10.1126/science.1254194> PMID: 24968937; PubMed Central PMCID: PMC4262401.
10. Wagner TA, McLaughlin S, Garg K, Cheung CY, Larsen BB, Styrchak S, et al. HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science (New York, NY)*. 2014; 345(6196):570–3. Epub 2014/07/12. <https://doi.org/10.1126/science.1256304> PMID: 25011556; PubMed Central PMCID: PMC4230336.
11. Wang Z, Gurule EE, Brennan TP, Gerold JM, Kwon KJ, Hosmane NN, et al. Expanded cellular clones carrying replication-competent HIV-1 persist, wax, and wane. *Proceedings of the National Academy of Sciences of the United States of America*. 2018; 115(11):E2575–E84. Epub 2018/02/28. <https://doi.org/10.1073/pnas.1720665115> PMID: 29483265; PubMed Central PMCID: PMC5856552.
12. Finzi D, Blankson J, Siliciano JD, Margolick JB, Chadwick K, Pierson T, et al. Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nature medicine*. 1999; 5(5):512–7. Epub 1999/05/06. <https://doi.org/10.1038/8394> PMID: 10229227.
13. Siliciano JD, Kajdas J, Finzi D, Quinn TC, Chadwick K, Margolick JB, et al. Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nature medicine*. 2003; 9(6):727–8. Epub 2003/05/20. <https://doi.org/10.1038/nm880> PMID: 12754504.
14. Zhang L, Chung C, Hu BS, He T, Guo Y, Kim AJ, et al. Genetic characterization of rebounding HIV-1 after cessation of highly active antiretroviral therapy. *J Clin Invest*. 2000; 106(7):839–45. Epub 2000/10/06. <https://doi.org/10.1172/JCI10565> PMID: 11018071; PubMed Central PMCID: PMC517816.
15. Joos B, Fischer M, Kuster H, Pillai SK, Wong JK, Boni J, et al. HIV rebounds from latently infected cells, rather than from continuing low-level replication. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105(43):16725–30. Epub 2008/10/22. <https://doi.org/10.1073/pnas.0804192105> PMID: 18936487; PubMed Central PMCID: PMC2575487.
16. Kearney MF, Spindler J, Shao W, Yu S, Anderson EM, O'Shea A, et al. Lack of detectable HIV-1 molecular evolution during suppressive antiretroviral therapy. *PLoS pathogens*. 2014; 10(3):e1004010. Epub 2014/03/22. <https://doi.org/10.1371/journal.ppat.1004010> PMID: 24651464; PubMed Central PMCID: PMC3961343.
17. Kearney MF, Wiegand A, Shao W, Coffin JM, Mellors JW, Lederman M, et al. Origin of Rebound Plasma HIV Includes Cells with Identical Proviruses That Are Transcriptionally Active before Stopping of Antiretroviral Therapy. *Journal of virology*. 2016; 90(3):1369–76. Epub 2015/11/20. <https://doi.org/10.1128/JVI.02139-15> PMID: 26581989; PubMed Central PMCID: PMC4719635.
18. Cohen YZ, Lorenzi JCC, Krassnig L, Barton JP, Burke L, Pai J, et al. Relationship between latent and rebound viruses in a clinical trial of anti-HIV-1 antibody 3BNC117. *The Journal of experimental*

- medicine. 2018; 215(9):2311–24. Epub 2018/08/04. <https://doi.org/10.1084/jem.20180936> PMID: 30072495; PubMed Central PMCID: PMC6122972.
19. Palich R, Ghosn J, Chaillon A, Boilet V, Nere ML, Chaix ML, et al. Viral rebound in semen after antiretroviral treatment interruption in an HIV therapeutic vaccine double-blind trial. *AIDS (London, England)*. 2019; 33(2):279–84. Epub 2018/10/17. <https://doi.org/10.1097/QAD.0000000000002058> PMID: 30325777.
 20. Colby DJ, Trautmann L, Pinyakorn S, Leyre L, Pagliuzza A, Kroon E, et al. Rapid HIV RNA rebound after antiretroviral treatment interruption in persons durably suppressed in Fiebig I acute HIV infection. *Nature medicine*. 2018; 24(7):923–6. Epub 2018/06/13. <https://doi.org/10.1038/s41591-018-0026-6> PMID: 29892063; PubMed Central PMCID: PMC6092240.
 21. Vibholm LK, Lorenzi JCC, Pai JA, Cohen YZ, Oliveira TY, Barton JP, et al. Characterization of Intact Proviruses in Blood and Lymph Node from HIV-Infected Individuals Undergoing Analytical Treatment Interruption. *Journal of virology*. 2019; 93(8):e01920–18. Epub 2019/02/01. <https://doi.org/10.1128/JVI.01920-18> PMID: 30700598; PubMed Central PMCID: PMC6450127.
 22. Ho YC, Shan L, Hosmane NN, Wang J, Laskey SB, Rosenbloom DI, et al. Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell*. 2013; 155(3):540–51. Epub 2013/11/19. <https://doi.org/10.1016/j.cell.2013.09.020> PMID: 24243014; PubMed Central PMCID: PMC3896327.
 23. Buzon MJ, Sun H, Li C, Shaw A, Seiss K, Ouyang Z, et al. HIV-1 persistence in CD4+ T cells with stem cell-like properties. *Nature medicine*. 2014; 20(2):139–42. Epub 2014/01/15. <https://doi.org/10.1038/nm.3445> PMID: 24412925; PubMed Central PMCID: PMC3959167.
 24. Bruner KM, Murray AJ, Pollack RA, Soliman MG, Laskey SB, Capoferri AA, et al. Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nature medicine*. 2016; 22(9):1043–9. Epub 2016/08/09. <https://doi.org/10.1038/nm.4156> PMID: 27500724; PubMed Central PMCID: PMC5014606.
 25. Hiener B, Horsburgh BA, Eden JS, Barton K, Schlub TE, Lee E, et al. Identification of Genetically Intact HIV-1 Proviruses in Specific CD4(+) T Cells from Effectively Treated Participants. *Cell reports*. 2017; 21(3):813–22. Epub 2017/10/19. <https://doi.org/10.1016/j.celrep.2017.09.081> PMID: 29045846; PubMed Central PMCID: PMC5960642.
 26. Bruner KM, Wang Z, Simonetti FR, Bender AM, Kwon KJ, Sengupta S, et al. A quantitative approach for measuring the reservoir of latent HIV-1 proviruses. *Nature*. 2019; 566(7742):120–5. Epub 2019/02/01. <https://doi.org/10.1038/s41586-019-0898-8> PMID: 30700913; PubMed Central PMCID: PMC6447073.
 27. Daar ES, Bai J, Hausner MA, Majchrowicz M, Tamaddon M, Giorgi JV. Acute HIV syndrome after discontinuation of antiretroviral therapy in a patient treated before seroconversion. *Ann Intern Med*. 1998; 128(10):827–9. Epub 1998/05/23. <https://doi.org/10.7326/0003-4819-128-10-199805150-00005> PMID: 9599194.
 28. Whitney JB, Hill AL, Sanisetty S, Penaloza-MacMaster P, Liu J, Shetty M, et al. Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. *Nature*. 2014; 512(7512):74–7. Epub 2014/07/22. <https://doi.org/10.1038/nature13594> PMID: 25042999; PubMed Central PMCID: PMC4126858.
 29. Henrich TJ, Hatano H, Bacon O, Hogan LE, Rutishauser R, Hill A, et al. HIV-1 persistence following extremely early initiation of antiretroviral therapy (ART) during acute HIV-1 infection: An observational study. *PLoS Med*. 2017; 14(11):e1002417. Epub 2017/11/08. <https://doi.org/10.1371/journal.pmed.1002417> PMID: 29112956; PubMed Central PMCID: PMC5675377.
 30. Brodin J, Zanini F, Thebo L, Lanz C, Bratt G, Neher RA, et al. Establishment and stability of the latent HIV-1 DNA reservoir. *eLife*. 2016; 5(2050-084X (Electronic)). Epub 2016/11/18. <https://doi.org/10.7554/eLife.18889> PMID: 27855060; PubMed Central PMCID: PMC5201419.
 31. Abrahams MR, Joseph SB, Garrett N, Tyers L, Moeser M, Archin N, et al. The replication-competent HIV-1 latent reservoir is primarily established near the time of therapy initiation. *Sci Transl Med*. 2019; 11(513):eaaw5589. Epub 2019/10/11. <https://doi.org/10.1126/scitranslmed.aaw5589> PMID: 31597754.
 32. Jones BR, Kinloch NN, Horacek J, Ganase B, Harris M, Harrigan PR, et al. Phylogenetic approach to recover integration dates of latent HIV sequences within-host. *Proceedings of the National Academy of Sciences of the United States of America*. 2018; 115(38):E8958–E67. Epub 2018/09/07. <https://doi.org/10.1073/pnas.1802028115> PMID: 30185556; PubMed Central PMCID: PMC6156657.
 33. Dilerma DA, Chien JT, Monaco DC, Brown MP, Ende Z, Deymier MJ, et al. Multiplexed highly-accurate DNA sequencing of closely-related HIV-1 variants using continuous long reads from single molecule, real-time sequencing. *Nucleic acids research*. 2015; 43(20):e129. Epub 2015/06/24. <https://doi.org/10.1093/nar/gkv630> PMID: 26101252; PubMed Central PMCID: PMC4787755.
 34. Theys K, Libin P, Pineda-Pena AC, Nowe A, Vandamme AM, Abecasis AB. The impact of HIV-1 within-host evolution on transmission dynamics. *Curr Opin Virol*. 2018; 28:92–101. Epub 2017/12/25. <https://doi.org/10.1016/j.coviro.2017.12.001> PMID: 29275182.

35. Yerly S, Pernerger TV, Vora S, Hirschel B, Perrin L. Decay of cell-associated HIV-1 DNA correlates with residual replication in patients treated during acute HIV-1 infection. *AIDS (London, England)*. 2000; 14(18):2805–12. Epub 2001/01/12. <https://doi.org/10.1097/00002030-200012220-00001> PMID: 11153661.
36. Blankson JN, Finzi D, Pierson TC, Sabundayo BP, Chadwick K, Margolick JB, et al. Biphasic decay of latently infected CD4+ T cells in acute human immunodeficiency virus type 1 infection. *The Journal of infectious diseases*. 2000; 182(6):1636–42. Epub 2000/11/09. <https://doi.org/10.1086/317615> PMID: 11069234.
37. Hocqueloux L, Avettand-Fenoel V, Jacquot S, Prazuck T, Legac E, Melard A, et al. Long-term antiretroviral therapy initiated during primary HIV-1 infection is key to achieving both low HIV reservoirs and normal T cell counts. *J Antimicrob Chemother*. 2013; 68(5):1169–78. Epub 2013/01/22. <https://doi.org/10.1093/jac/dks533> PMID: 23335199.
38. Besson GJ, Lalama CM, Bosch RJ, Gandhi RT, Bedison MA, Aga E, et al. HIV-1 DNA decay dynamics in blood during more than a decade of suppressive antiretroviral therapy. *Clin Infect Dis*. 2014; 59(9):1312–21. Epub 2014/07/31. <https://doi.org/10.1093/cid/ciu585> PMID: 25073894; PubMed Central PMCID: PMC4200019.
39. Archin NM, Vaidya NK, Kuruc JD, Liberty AL, Wiegand A, Kearney MF, et al. Immediate antiviral therapy appears to restrict resting CD4+ cell HIV-1 infection without accelerating the decay of latent infection. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109(24):9523–8. Epub 2012/05/31. <https://doi.org/10.1073/pnas.1120248109> PMID: 22645358; PubMed Central PMCID: PMC3386138.
40. Buzon MJ, Martin-Gayo E, Pereyra F, Ouyang Z, Sun H, Li JZ, et al. Long-term antiretroviral treatment initiated at primary HIV-1 infection affects the size, composition, and decay kinetics of the reservoir of HIV-1-infected CD4 T cells. *Journal of virology*. 2014; 88(17):10056–65. Epub 2014/06/27. <https://doi.org/10.1128/JVI.01046-14> PMID: 24965451; PubMed Central PMCID: PMC4136362.
41. Ruff CT, Ray SC, Kwon P, Zinn R, Pendleton A, Hutton N, et al. Persistence of wild-type virus and lack of temporal structure in the latent reservoir for human immunodeficiency virus type 1 in pediatric patients with extensive antiretroviral exposure. *Journal of virology*. 2002; 76(18):9481–92. Epub 2002/08/21. <https://doi.org/10.1128/jvi.76.18.9481-9492.2002> PMID: 12186930; PubMed Central PMCID: PMC136462.
42. Verhofstede C, Noe A, Demecheleer E, De Cabooter N, Van Wanseele F, Van Der Gucht B, et al. Drug-resistant variants that evolve during nonsuppressive therapy persist in HIV-1-infected peripheral blood mononuclear cells after long-term highly active antiretroviral therapy. *Journal of acquired immune deficiency syndromes*. 2004; 35(5):473–83. Epub 2004/03/17. <https://doi.org/10.1097/00126334-200404150-00005> PMID: 15021312.
43. Kieffer TL, Finucane MM, Nettles RE, Quinn TC, Broman KW, Ray SC, et al. Genotypic analysis of HIV-1 drug resistance at the limit of detection: virus production without evolution in treated adults with undetectable HIV loads. *The Journal of infectious diseases*. 2004; 189(8):1452–65. Epub 2004/04/10. <https://doi.org/10.1086/382488> PMID: 15073683.
44. Bailey JR, Sedaghat AR, Kieffer T, Brennan T, Lee PK, Wind-Rotolo M, et al. Residual human immunodeficiency virus type 1 viremia in some patients on antiretroviral therapy is dominated by a small number of invariant clones rarely found in circulating CD4+ T cells. *Journal of virology*. 2006; 80(13):6441–57. Epub 2006/06/16. <https://doi.org/10.1128/JVI.00591-06> PMID: 16775332; PubMed Central PMCID: PMC1488985.
45. Perelson AS, Essunger P, Cao Y, Vesanen M, Hurley A, Saksela K, et al. Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature*. 1997; 387(6629):188–91. Epub 1997/05/08. <https://doi.org/10.1038/387188a0> PMID: 9144290.
46. Jones BR, Miller RL, Kinloch NN, Tsai O, Rigsby H, Sudderuddin H, et al. Genetic diversity, compartmentalization and age of HIV proviruses persisting in CD4+ T cell subsets during long-term combination antiretroviral therapy. *Journal of virology*. 2019; JVI.01786-19. Epub 2019/11/30. <https://doi.org/10.1128/JVI.01786-19> PMID: 31776273.
47. Rousseau CM, Birditt BA, McKay AR, Stoddard JN, Lee TC, McLaughlin S, et al. Large-scale amplification, cloning and sequencing of near full-length HIV-1 subtype C genomes. *Journal of virological methods*. 2006; 136(1–2):118–25. Epub 2006/05/17. <https://doi.org/10.1016/j.jviromet.2006.04.009> PMID: 16701907.
48. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010; 59(3):307–21. Epub 2010/06/09. <https://doi.org/10.1093/sysbio/syq010> PMID: 20525638.
49. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data.

- Bioinformatics. 2012; 28(12):1647–9. Epub 2012/05/01. <https://doi.org/10.1093/bioinformatics/bts199> PMID: 22543367; PubMed Central PMCID: PMC3371832.
50. Rose PP, Korber BT. Detecting hypermutations in viral sequences with an emphasis on G→A hypermutation. *Bioinformatics*. 2000; 16(4):400–1. Epub 2000/06/27. <https://doi.org/10.1093/bioinformatics/16.4.400> PMID: 10869039.
 51. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol*. 2016; 33(7):1870–4. Epub 2016/03/24. <https://doi.org/10.1093/molbev/msw054> PMID: 27004904.
 52. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol*. 2015; 1(1):vev003. Epub 2015/05/26. <https://doi.org/10.1093/vev/vev003> PMID: 27774277; PubMed Central PMCID: PMC5014473.
 53. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol*. 2018; 4(1):vey016. Epub 2018/06/27. <https://doi.org/10.1093/ve/vey016> PMID: 29942656; PubMed Central PMCID: PMC6007674.
 54. Shapiro B, Rambaut A, Drummond AJ. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol Biol Evol*. 2006; 23(1):7–9. Epub 2005/09/24. <https://doi.org/10.1093/molbev/msj021> PMID: 16177232.
 55. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 2006; 4(5):e88. Epub 2006/05/11. <https://doi.org/10.1371/journal.pbio.0040088> PMID: 16683862; PubMed Central PMCID: PMC1395354.
 56. Minin VN, Bloomquist EW, Suchard MA. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol Biol Evol*. 2008; 25(7):1459–71. Epub 2008/04/15. <https://doi.org/10.1093/molbev/msn090> PMID: 18408232; PubMed Central PMCID: PMC3302198.
 57. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics*. 2002; 161(3):1307–20. Epub 2002/07/24. PMID: 12136032; PubMed Central PMCID: PMC1462188.
 58. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchene S, Fourment M, Gavryushkina A, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol*. 2019; 15(4):e1006650. Epub 2019/04/09. <https://doi.org/10.1371/journal.pcbi.1006650> PMID: 30958812; PubMed Central PMCID: PMC6472827.
 59. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst Biol*. 2018; 67(5):901–4. Epub 2018/05/03. <https://doi.org/10.1093/sysbio/syy032> PMID: 29718447; PubMed Central PMCID: PMC6101584.
 60. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014; 30(9):1312–3. Epub 2014/01/24. <https://doi.org/10.1093/bioinformatics/btu033> PMID: 24451623; PubMed Central PMCID: PMC3998144.
 61. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*. 2019; 35(3):526–8. Epub 2018/07/18. <https://doi.org/10.1093/bioinformatics/bty633> PMID: 30016406.
 62. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, et al. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105(21):7552–7. Epub 2008/05/21. <https://doi.org/10.1073/pnas.0802203105> PMID: 18490657; PubMed Central PMCID: PMC2387184.