



Published in final edited form as:

Cell Rep. 2025 May 27; 44(5): 115663. doi:10.1016/j.celrep.2025.115663.

SIV proviruses seeded later in infection are harbored in short-lived CD4⁺ T cells

Narmada Sambaturu^{1,2}, Emily J. Fray³, Vivek Hariharan³, Fengting Wu³, Carolin Zitzmann¹, Francesco R. Simonetti³, Dan H. Barouch⁴, Janet D. Siliciano³, Robert F. Siliciano^{3,5}, Ruy M. Ribeiro¹, Alan S. Perelson¹, Carmen Molina-París¹, Thomas Leitner^{1,6,*}

¹Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

²School of Systems Science and Industrial Engineering, State University of New York at Binghamton, Binghamton, NY, USA

³Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

⁴Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, Boston, MA 02215, USA

⁵Howard Hughes Medical Institute, Baltimore, MD 21205, USA

⁶Lead contact

SUMMARY

The human immunodeficiency virus (HIV) can persist in a latent form as integrated DNA (provirus) in resting CD4⁺ T cells unaffected by antiretroviral therapy. Despite being a major obstacle for eradication efforts, it remains unclear which infected cells survive, persist, and ultimately enter the long-lived reservoir. Here, we determine the genetic divergence and integration times of simian immunodeficiency virus (SIV) envelope sequences collected from infected macaques. We show that the proviral divergence and the phylogenetically estimated integration times display a biphasic decline over time. Investigating the dynamics of the mutational distributions, we show that SIV genomes in short-lived cells are, on average, more diverged, while long-lived cells contain less diverged virus. The change in the mutational distributions over time explains the observed biphasic decline in the divergence of the proviruses.

This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

*Correspondence: tkl@lanl.gov.

AUTHOR CONTRIBUTIONS

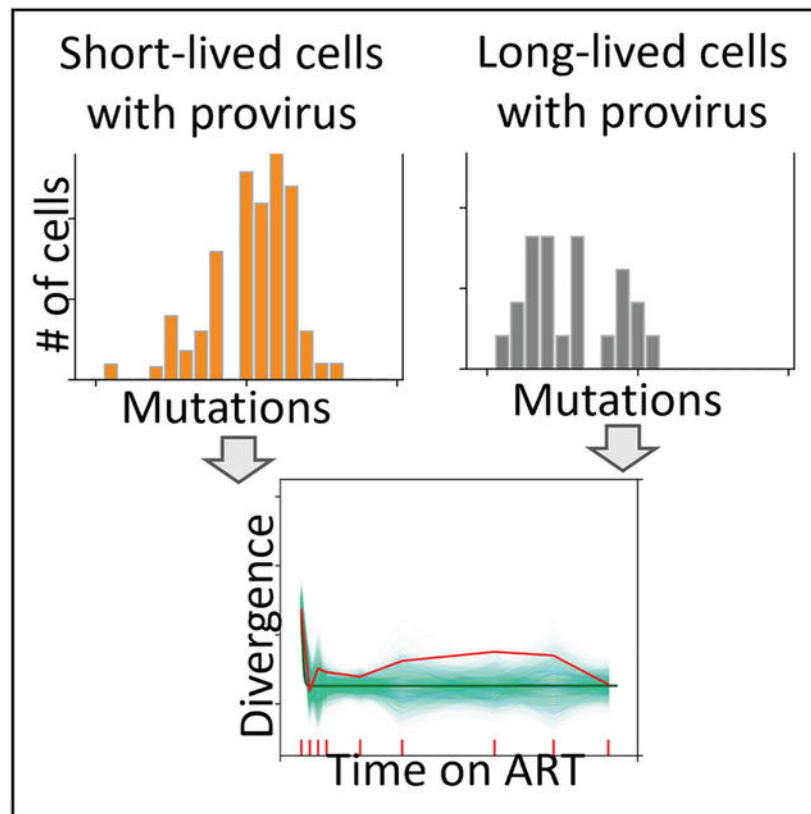
T.L., C.M.-P., A.S.P., R.M.R., R.F.S., J.D.S., and D.H.B. designed the study. N. S., E.J.F., V.H., F.W., C.Z., F.R.S., D.H.B., J.D.S., R.F.S., R.M.R., A.S.P., C. M.-P., and T.L. participated in the discussion and interpretation of the results. E.J.F., V.H., and F.W. performed the experiments. D.H.B. provided samples, information, and drugs related to the animal work. N.S., C.Z., C.M.-P., T.L., R.M.R., A.S.P., E.J.F., V.H., F.W., F.R.S., and R.F.S. performed the data analysis and mathematical modeling. N.S. wrote the manuscript. E.J.F., F.R.S., R. F.S., J.D.S., R.M.R., A.S.P., C.M.-P., and T.L. revised the manuscript. All authors approved of the final version of this report.

DECLARATION OF INTERESTS

Aspects of HIV-1 IPDA are the subject of patent application PCT/US16/28822 filed by Johns Hopkins University. R.F.S. is an inventor on this application. Acelevir Diagnostics holds an exclusive license for this patent application. R.F. S. holds no equity interest in Acelevir Diagnostics.

This suggests that long-lived cells harbor viruses deposited earlier in infection, while short-lived cells predominantly harbor more recent viruses.

Graphical abstract



In brief

Sambaturu et al. show that SIV proviral DNA exists in both short- and long-lived CD4⁺ T cells, which harbor different genetically diverged virus populations. Thus, when CD4⁺ T cells decay under effective drug treatment, which prevents virus replication, the resulting proviral divergence decreases predictably over time.

INTRODUCTION

Human immunodeficiency virus (HIV) has infected over 84 million people and caused an estimated 40.1 million deaths worldwide since the beginning of the epidemic in 1981.^{1,2} Antiretroviral therapy (ART) suppresses HIV replication, reduces viral load, and prolongs life.¹ However, even during suppressive ART, a reservoir of transcriptionally silent proviruses persists, integrated into the genomes of CD4⁺ T cells of the host.^{3–7} This proviral reservoir is capable of re-seeding infection if therapy is interrupted, making HIV infection, as of yet, incurable with ART alone.^{8,9} Despite the importance of this population of cells, it is largely unknown which infected cells ultimately persist and become part of the stable reservoir after ART is initiated.

Attempts to identify when the latent reservoir is established have yielded seemingly contradictory results. Several studies have observed that even when ART is initiated within days of infection, viral rebound occurs upon treatment interruption,^{10–12} suggesting rapid and continuous seeding of the latent reservoir. Other studies discovered that viral rebound sequences obtained during treatment interruption were genetically closest to the founder virus, also suggesting that the latent reservoir is established primarily near the start of infection.^{13–16} Still others have found the latent reservoir to be genetically heterogeneous, recapitulating the within-host HIV evolutionary history, suggesting that the latent reservoir is seeded throughout the course of infection.^{17–19} Comprehensive simulations of within-host sequence evolution that accounted for a latent reservoir, immune selection, point mutations, and recombination demonstrated that empirical diversity and divergence trends were consistent with continuous deposition of HIV variants throughout natural infection.²⁰ In contradiction, more recent studies have found that a majority of proviral DNA sequences obtained several years after the initiation of ART clustered closer on a phylogenetic tree to plasma viral sequences sampled from ≈ 1 to 2 years preceding initiation of ART than to sequences from earlier in infection.^{21–26} This finding suggests that the replication-competent latent reservoir is mainly established near the time of therapy initiation. Interestingly, Pankau et al.²⁵ tested whether a “last time point” model (HIV DNA reservoir mimics HIV RNA at the last time point prior to ART initiation) or a “cumulative” model (constant reservoir seeding and decay) better reflects observed fractions of lineage variants in the reservoir and found that the difference between the two models was not statistically significant.

Furthermore, contradictory findings regarding the fate of T cells in the latent reservoir have also been reported. Lorenzo-Redondo et al.²⁷ and others²⁸ found that HIV-1 replication and evolution can occur in some tissue compartments during ART. However, studies proposing ongoing replication during ART have been criticized for being inconsistent with clinical data.^{29,30} Several studies have found that persistence and clonal expansion of long-lived, latently infected T cells, and not ongoing viral replication, were predominantly responsible for the maintenance of the latent reservoir.^{22,31–35} Cho et al.³⁵ sequenced intact proviruses from CD4⁺ T cells and found that the diversity decreased over time, suggesting that some sequences within the latent reservoir may be replaced as the clones of infected cells expand.

Using the intact proviral DNA assay (IPDA),³⁶ White et al.³⁷ showed that after the initiation of ART, the number of intact HIV genomes declined with biphasic kinetics with an initial rapid decline ($t_{1/2} = 12.9$ days) and a second slower decline ($t_{1/2} = 19$ months). We recently reported³⁸ a triphasic decline in the number of intact simian immunodeficiency virus (SIV) genomes after the initiation of ART in SIV-infected macaques, with an initial rapid phase ($t_{1/2} = 3.3$ days), a second slower phase ($t_{1/2} = 8.1$ months), and a stable third phase with no observable decay reached after 1.6–2.9 years. We also found that the number of non-synonymous mutations in these proviral sequences displayed a biphasic decline. These observations point to the existence of at least two sub-populations (short and long lived) of CD4⁺ T cells harboring intact proviruses. However, it remains unclear whether there is a difference in the genetic composition of the proviruses harbored in these sub-populations

of CD4⁺ T cells and how such a difference may impact experimental measurements of the latent reservoir.

Taking these observations together, we postulate that observations on reservoir composition and dynamics may be confounded by the heterogeneous nature of the pool of infected CD4⁺ T cells comprising the reservoir. To test this hypothesis, we further studied the same cohort of SIV-infected macaques,³⁸ focusing our analysis on the first ≈ 3 years on ART, corresponding to the conservative estimate of the time to reach the stable third phase. We gathered 2,591 additional non-defective proviral *env* sequences from circulating CD4⁺ T cells during ART and found a biphasic decline in divergence (the evolutionary distance from the founder sequence[s]), corresponding to the biphasic decline observed in the number of non-synonymous mutations (Figure 7 in Fray et al.³⁸). We also estimated the integration time of on-ART sequences in 4 macaques by comparing them with pre-ART plasma RNA sequences through a phylogenetic method and found that the estimated integration times follow a trend nearly identical to the divergence dynamics, with estimated integration times approaching the start of infection as ART progressed. We developed a new method to estimate the distributions of mutations in the proviruses harbored in the short- and long-lived sub-populations of CD4⁺ T cells and concluded that the short-lived cells harbored more diverged proviruses, suggesting they were integrated later in infection, while long-lived cells harbored less diverged proviruses, presumably integrated earlier in infection. Furthermore, we developed a mathematical model to predict divergence kinetics during ART, given these initial distributions and decay rates. The model captures the observed biphasic decline of divergence, thus providing a mechanistic explanation for the observed dynamics in divergence and estimated integration times. This difference in the genetic composition of proviruses harbored in short- and long-lived CD4⁺ T cells circulating during ART may explain the seemingly contradictory observations in previous experiments studying the latent reservoir.

RESULTS

Biphasic decline in divergence and integration times of proviral SIV during early ART

In this paper, we further studied the cohort of SIV-infected macaques we reported in Fray et al.³⁸ The cohort consisted of 10 Indian-origin rhesus macaques infected with SIV (SIVmac₂₅₁ stock) and initiated ART at ≈ 48 weeks post-infection. Intact proviruses in CD4⁺ T cells were quantified using the IPDA³⁶ periodically from the day ART was initiated. Envelope gene (*env*) sequences from SIV DNA in circulating CD4⁺ T cells were then obtained using single-genome sequencing and filtered to remove any sequences with defects (deletions, premature stop codons, hypermutations, etc.), resulting in non-defective proviral *env* sequences (see STAR Methods for details). We focused our analysis on the first 2.8 years ($t = 146$ weeks) on ART and augmented the data by gathering 2,591 additional sequences, corresponding to plasma viral RNA sequences at ART initiation for 9 of 10 macaques and for 6 time points pre-ART for 4 macaques (see Methods S1 and Tables S1 and S2 for details).

We determined the divergence of non-defective proviral *env* sequences from the consensus sequence of the infecting virus stock. We found that the divergence showed a biphasic

decline over time, with an approximately 230 times faster first phase compared to the second phase (Figure 1) ($p = 0.0019$, $n = 10$, paired two-sided Wilcoxon signed rank exact test; Table S3). This is comparable to the biphasic decline in the mean number of non-synonymous mutations reported previously.³⁸ Macaque T624 is an exception and was found to have spontaneously controlled viral replication to low levels before ART. The elbow point at which the slope of the divergence curve changes (time when $|\text{slope 1}/\text{slope 2}|$ was largest; see STAR Methods) was found to have a wide variation among the macaques, with an average of 11.6 weeks and a range of 4–28 weeks (Table S3). To test whether this heterogeneity in elbow points corresponded to a variation in half-lives of the CD4⁺ T cells harboring these proviruses, we re-estimated the half-lives using IPDA data on the number of intact proviral sequence over time in two ways: (1) by using pooled data for all 10 macaques and then adjusting for individual variations (referred to here as mixed-effects parameters) and (2) by estimating the half-lives separately for each macaque using only its own data. We found that method 1 (mixed effects) confirmed the half-lives of 3.3 days and 8 months estimated earlier.³⁸ Method 2, newly carried out here, revealed that the estimated half-lives did indeed display a wide variation across the macaques, from 0.9 to 19.8 days for short-lived cells and from 3.9 to 12.2 months for long-lived cells (Table S4). Both methods estimated the fraction of short-lived cells at the start of ART to be 0.6 (Table S4).

To test how the non-defective *env* DNA sequences compared with non-defective plasma *env* RNA sequences at the start of ART, we computed the divergence of plasma *env* RNA sequences from the same reference sequence (consensus of infecting stock). We found that the plasma *env* sequences were more diverged (blue circles in Figure 1) than the proviral *env* sequences (gray circles in Figure 1) at the start of ART. In fact, when the divergence dynamics were computed using the divergence of plasma *env* RNA sequences at the start of ART and proviral *env* DNA at subsequent time points, we found a much faster first-phase divergence decline, approximately 390 times faster than the second phase ($p = 0.0039$, $n = 9$, paired two-sided Wilcoxon signed rank exact test) (Figure 1; Table S5). This likely reflects the fact that the plasma virus sequences sampled at the start of ART represent actively replicating viruses, while proviral sequences obtained at the same time point include actively replicating viruses as well as archived sequences.³⁸ We computed divergence using the following reference sequences: (1) consensus sequence of the sequences in the SIVmac₂₅₁ stock (Figures S1 and S2), (2) each of the sequences in the SIVmac₂₅₁ stock (Figures S1 and S2), and (3) each of the sequences obtained from the first time point after confirmed infection (Figures S3–S6). We found that the divergence displayed a biphasic decline irrespective of the reference sequence used. We, therefore, used the consensus of the sequences in the SIVmac₂₅₁ stock as the reference for all further analyses.

During infection, actively replicating SIV accumulates mutations and diverges from the founder virus as infection progresses.^{20,40–42} Therefore, the divergence of proviral sequences can serve as a proxy for the time when the viruses were integrated into the CD4⁺ T cells. In order to verify that the observed divergence trends indeed corresponded to integration times, we gathered pre-ART sequences from 4 macaques (T530, T623, T625, and T627) and estimated the integration times for the sequences collected after the start of ART using a phylogenetic method similar to that used in Kinloch et al.²²

In brief, a linear regression was used to relate the sampling times of pre-ART sequences to their root-to-tip distances on a phylogenetic tree (rooted to the most recent common ancestor of the sequences from the first time point after confirmed infection). The root-to-tip distances of on-ART sequences were then mapped to their estimated integration time using this regression (see STAR Methods for more details). We found that the mean estimated integration time followed a nearly identical biphasic decline to that observed in the divergence profiles (Figure 2). This further supports the observation that the pool of SIV-infected CD4⁺ T cells surviving after the start of ART is dynamic, initially dominated by proviruses deposited near the start of ART, and older proviruses becoming relatively more dominant over time.

Short- and long-lived CD4⁺ T cell sub-populations contain provirus with different divergences

Sequencing proviral DNA at each time point provides the total distribution of mutations across infected CD4⁺ T cells without distinguishing between the short- and long-lived sub-populations. To the best of our knowledge, we are not aware of a method to experimentally or computationally distinguish CD4⁺ T cells based on their half-life. Thus, we developed a new computational method to distinguish the distributions of mutations of proviruses in short- and long-lived cells. The method is based on the assumption that at the last sampled time point of ≈ 3 years on ART, the short-lived sub-population of CD4⁺ T cells has died out, an assumption we believe to be reasonable given the half-life of 3.3 days or even the upper bound of 19.8 days estimated for short-lived cells. Using this upper bound, one can calculate that after 3 years, the population size of the short-lived cells would have decreased by a factor of at least 10^{16} . Thus, we take the distribution of mutations at the last time point to correspond to the proviruses in long-lived cells. By scaling this distribution up to account for the decline in T cell numbers during the ≈ 3 years, we obtain the distribution of mutations of proviruses in long-lived cells at the start of ART. We then subtract this distribution from the total distribution of mutations at the start of ART to give the distribution for proviruses in short-lived T cells. Further details are presented in STAR Methods.

This method revealed a difference in the genetic composition of proviruses harbored in short- and long-lived CD4⁺ T cells at the start of ART (Figure 3). Short-lived cells, on average, had more diverged proviruses than long-lived ones ($p = 0.002$ and $n = 10$ when non-defective proviral DNA sequences were used at all time points and $p = 0.004$ and $n = 9$ when plasma viral RNA sequences were used at the start of ART; paired Wilcoxon signed rank test; Tables S6 and S7; Figure S7). During untreated infection, SIV undergoes continuous replication, with mutations accumulating as infection progresses.^{20,40,43} Thus, on average, the actively replicating virus diverges further and further away from the founder virus as infection progresses,^{20,41,42} allowing the use of divergence as an indicator of the last time in infection when the virus was actively replicating. Therefore, the higher divergence in the provirus integrated into short-lived cells suggests that these viruses were integrated later in infection, closer to the start of ART. Conversely, the lower divergence found in the provirus integrated into long-lived cells indicates that these viruses were integrated earlier in infection.

Interestingly, the distributions of the number of mutations in the short- and long-lived cells were shifted relative to the overall divergence median: shifted right (dominated by sequences with a larger number of mutations) in the short-lived cells and shifted left (dominated by sequences with fewer mutations) in the long-lived cells. This lends further support to the finding that the CD4⁺ T cell sub-populations represent different eras of virus deposition ($p = 2.46 \times 10^{-4}$, $n = 10$, Mann Whitney U test) (Figure 3). Again, comparing the subsequently sampled proviral sequences to the plasma RNA population at the start of ART showed that the RNA population was even more diverged and, therefore, corresponds to the most recent virus variants ($p = 4.12 \times 10^{-4}$, $n = 9$, Mann Whitney U test; Tables S6 and S7; Figure S7).

Decay of short- and long-lived CD4⁺ T cells with different divergences explains the biphasic divergence decline

To test whether these differences in the proviruses harbored in short- and long-lived CD4⁺ T cells can explain the observed biphasic decline in divergence, we developed a model to simulate the divergence dynamics of proviruses integrated into two CD4⁺ T cell sub-populations (see STAR Methods for details). Simulations start with the estimated distributions of mutations for the proviruses in short- and long-lived T cell sub-populations based on the experimental sequence data from each macaque (Figure 4). CD4⁺ T cells then decay stochastically over time at the estimated rate (Table S4). We sample each simulated macaque at the same times and with the same number of sequences as in the empirical data.

We found that the overall divergence dynamics predicted by the model matched the empirically observed divergence decline trends (Figure 5). Simulations were carried out with the half-lives estimated using both methods described in biphasic decline in divergence and integration times of proviral SIV during early ART (also see STAR Methods for details): (1) pooling data for all macaques and then adjusting for variations (blue lines in Figure 5, with the bold blue line showing the average of 10^3 simulations) and (2) estimating for each macaque separately using only its own data (green lines in Figure 5, with the bold green line showing the average of 10^3 simulations). We found that the half-lives estimated using either method 1 (e.g., T523 and T625) or method 2 (e.g., T628, T627, T545, and T544) captured the empirically observed heterogeneity across macaques. The variability of the empirical data was based on about 20 sequences per time point (Figure 1; Tables S1 and S2). Encouragingly, with the same sampling schedule (times and number of sequences), our simulations showed similar variability in individual runs, matching the observed dynamics (e.g., T623, T544, and T627). Agreement is seen between the experimental observations and model simulations when the proviral *env* DNA sequences are replaced with plasma viral RNA sequences at the start of ART, retaining non-defective proviral DNA sequences for subsequent time points (Figure S8).

Divergence decline results were robust to perturbations in parameters and mutational profiles

We investigated the relative contribution of the model parameters and sensitivity to perturbations in parameters to the divergence over time. Here, the model parameters were the initial frequency of short-lived cells ($f(0)$), the half-lives of short- and long-lived cells

(τ_s and τ_L , respectively), and the mean divergence in short- and long-lived cells (d_s and d_L , respectively) (Methods S1; Figures S9 and S10). At the start of ART, the mean proviral divergences, d_s and d_L , both contribute about 50%, and $f(0)$ contributes 10%–20% to the divergence. As time on ART proceeds, $f(0)$ and d_s become less important, while d_L becomes dominant and eventually the sole determinant of the divergence.

To investigate whether the shape of the mutation distributions of diverged proviral sequences in short- and long-lived cell sub-populations has any impact on the biphasic decline, we simulated different combinations of mutation profiles (Figures S11–S15). Overall, as long as the mean divergence in the short-lived cells was greater than that in the long-lived ones, qualitatively, we found that biphasic divergence declined. Thus, the mutation distribution had no qualitative impact on the overall dynamics of proviral divergence under ART. However, we observed that the choice of distribution did impact the quantitative behavior of the biphasic divergence decline.

Taken together, these observations show that our results are robust to perturbations in parameters and mutational profiles.

DISCUSSION

HIV persists as an integrated provirus in a reservoir of CD4⁺ T cells, even during suppressive ART.^{3–5} This reservoir can serve as a source of viral rebound if treatment is interrupted, making this a major hurdle to curing HIV infection.^{6,7,44–46} Contradictory results exist on when the latent reservoir is formed and what role it may have in maintaining infection. These results have led to different authors recommending different strategies for its elimination.^{10–13,17,18,20–26} Previous studies have suggested the presence of at least two sub-populations of CD4⁺ T cells (short- and long-lived) harboring intact proviruses.^{37,38,47} However, it remained unclear whether there was a difference in the proviruses harbored in these two sub-populations and how such a difference might impact experimental observations of the latent reservoir. In this study, we observed a biphasic decline in the divergence and the estimated integration times of proviruses in CD4⁺ T cells from SIV-infected rhesus macaques during the first ≈ 3 years on ART. We developed a method to estimate from data the distribution of mutations of proviruses harbored in short- and long-lived CD4⁺ T cells at the start of ART. We found that more mutated proviruses (compared to the infecting stock) were harbored in short-lived cells, while relatively less mutated ones were present in long-lived cells. During untreated infection, mutations accumulate in actively replicating SIV, resulting in increasing divergence from the founder virus.^{20,40–42} As a result of this, the divergence of proviral sequences can be used as a proxy for the deposition time,^{48,49} also supported by our estimated integration time analysis. Thus, our results indicated that SIV proviruses from near the start of therapy, which occurred ≈ 48 weeks after infection, are harbored in short-lived CD4⁺ T cells, while proviruses from relatively earlier in infection are harbored in long-lived T cells. We further found that in the SIV-infected macaques studied, the mutational profile in the short-lived cells was shifted right, toward a higher number of mutations, suggesting that they are dominated by proviruses from near the start of therapy. Conversely, and in relative terms, the mutational profile in long-lived cells was shifted left, toward a lower number of mutations,

consistent with being dominated by proviruses from earlier in infection. The forces driving this shift in the mutational profiles in short- and long-lived cells are as of yet unknown and require further investigation. A possible explanation is that the short-lived cells are mostly comprised of productively infected CD4⁺ T cells, measurable for a short while after the initiation of ART.

Using the estimated distribution of mutations and half-lives as inputs, we developed a mathematical model to characterize the divergence dynamics during suppressive ART (assuming no viral replication or clonal expansion during the modeled time of suppressive ART). The model predicts that when these cell populations decay during ART, the overall virus divergence declines in a biphasic manner, similar to that observed in our datasets. By computationally exploring different combinations of mutational profiles at the start of ART, we observed that a biphasic decline occurred whenever the short-lived population had, on average, more diverged viral sequences than the long-lived one.

Interestingly, since short-lived cells form a greater fraction of the total CD4⁺ T cell population before ART,^{37,38} the overall distribution of mutations of the two sub-populations combined would mostly reflect the highly diverged sequences present in the short-lived cells initially after the start of ART. This agrees with recent studies showing that the majority of replication-competent proviruses are genetically similar to the viruses circulating near the time of therapy initiation.^{21–23,25,26} As these short-lived cells die out, the divergence in proviruses harbored in CD4⁺ T cells increasingly reflects the viral genomes harbored in the long-lived sub-population, which we found to be relatively less diverged on average. This explains the observed biphasic divergence decline of integrated proviruses. This means that the choice of sampling times (with respect to the start of ART) can lead to rather different conclusions of how diverged the proviruses in an individual are with respect to a founder virus and, as a consequence, the time that cells were deposited into the reservoir. Thus, the presence of short- and long-lived sub-populations of CD4⁺ T cells harboring different genetically diverged proviruses may be one of the confounding factors contributing to the seemingly contradictory observations in the literature regarding reservoir deposition times.

The model proposed here has the characteristic of “last in, first out”—the viruses replicating later in infection, which tend to be highly diverged, will be predominantly harbored in short-lived CD4⁺ T cells and will be lost earlier than the long-lived cells harboring less diverged proviruses, at least in the early phases of therapy, and when considering both intact and defective proviruses. This suggests that recent clinical trials that focus almost exclusively on viruses circulating near the start of ART⁵⁰ may miss the less diverged proviruses in long-lived cells. The good news is that people who have been on suppressive ART long enough for their short-lived cells to have died out would tend to have less evolved proviruses, which, upon activation, may be more susceptible to immune pressure.

Limitations of the study

This study builds on our previous work,³⁸ where we reported a biphasic decline in the number of non-synonymous mutations in SIV proviral *env* sequences obtained from circulating CD4⁺ T cells in 10 SIV-infected macaques during 4 years on ART. In this study, we further investigated the same cohort of SIV-infected macaques, focusing our

analysis on the first ≈ 3 years of ART corresponding to the conservative estimate of the time to reach a stable phase with essentially no decline in the number of SIV genomes. We build on our previous findings³⁸ by (1) computing divergence dynamics (as opposed to non-synonymous mutations) from an average of 92 different reference sequences and by gathering 2,591 additional sequences, showing a biphasic decline; (2) phylogenetically estimating the integration time of sequences sampled during ART, showing a corresponding biphasic decline; (3) identifying the wide variation in half-lives of CD4⁺ T cells across the macaques (ranges: 0.9–19.8 days for short-lived and 3.9–12.2 months for long-lived sub-populations); (4) developing a new method to estimate the distribution of mutations in proviruses harbored in short- and long-lived cells, showing differences in their genetic composition; (5) developing a model to predict the divergence dynamics given these estimated distributions and half-lives, whose predictions track the observed divergence trends, thus providing a mechanistic explanation; and (6) rigorously studying the effect of perturbations, showing that the presence of a larger number of mutations in the proviruses harbored in short-lived cells compared to long-lived ones can result in a biphasic decline in divergence.

Our model does not incorporate mechanisms for reservoir maintenance, such as clonal expansion. However, the transition from the first rapid decline in divergence to the second slower one takes place in a matter of months (≈ 1 to ≈ 7 months) after the start of ART in this cohort of SIV-infected macaques, well before the potential impact of clonal expansion is significant (≈ 5 –10 years in people living with HIV [PLWH]^{35,51}; similar or shorter in SIV-infected macaques on ART^{38,52,53}). We restricted our analysis to the first ≈ 3 years (146 weeks), corresponding to the conservative estimate of the time to reach the stable third phase,³⁸ possibly maintained through clonal expansion. Importantly, given the distributions of mutations at the start of ART, the predictions of a simple decay model can reproduce the observed divergence dynamics. This suggests that the composition of proviruses in the CD4⁺ T cell sub-populations at the start of therapy and the corresponding decay rates play a greater role in determining the dynamics of divergence than clonal expansion over the timescales considered here.

Our experimental data, and therefore our model, investigated CD4⁺ T cells with integrated proviruses having non-defective *env* sequences. This may include both productively infected and resting cells and may also include proviruses with defects in other parts of their genome. Recent studies into viral genetic features have found that the latent reservoir is mostly comprised of genetically defective proviruses,^{22–24} although SIV genomes tend to be more intact than full-length genomes sampled from PLWH.^{52,54} In fact, some studies have found a marked difference between the estimated integration times of intact and defective proviruses sampled after several years on ART,^{22,24} while other studies found no strong influence of replication competency and cellular tropism on the timing of deposition of the proviruses that persist during ART.²³ Despite the possible differences between replication-competent and defective proviruses, it is important to understand both, as defective proviruses, which have different kinetics,^{37,38} have been found to be capable of producing transcripts and proteins,^{55,56} triggering immune responses,^{56,57} and even generating virions⁵⁸ and may have consequences for the design of therapeutic strategies.

Differences between HIV/SIV and human/monkey biology should always be considered and cannot be ruled out. An analogous dataset (in terms of sampling frequency relative to ART initiation, sample type: CD4s vs. peripheral blood mononuclear cells [PBMCs], sequencing target: Env vs. full length, and knowledge regarding the transmitted/founder sequence) from a cohort of PLWH does not exist at this time to our knowledge. The field would benefit greatly from such a comparison, however, and although this is beyond the scope of this paper, we are interested in further exploring this issue with an appropriate human dataset in the future.

This work is based on sequences collected from PBMCs and does not contain samples from other tissue compartments, including lymph nodes or gut-associated lymphoid tissue (GALT). A recent study in rhesus macaques found no significant changes in the SIV DNA population during ART in samples collected from PBMCs, lymph nodes, and spleen.⁵⁹ Previous studies using samples from PLWH have suggested that in participants sampled during ART, there is relative mixing between lymph node and peripheral blood populations in terms of sequence diversity.⁶⁰ Another study found no evidence of compartmentalization of HIV in blood, lymphoid, and other infected tissues obtained at colonoscopy or autopsy in individuals who were on ART for 8–16 years.⁶¹ Based on this evidence, we believe that the divergence, dynamics, and decay results we have obtained from PBMC-derived proviruses may be representative of those in other tissue compartments as well.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Dr. Thomas Leitner (tkl@lanl.gov).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Sequences generated in this study have been deposited in GenBank and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- All original code has been deposited in GitHub (https://github.com/MolEvolEpid/SIV_divergence_dynamics) and is publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

STAR★METHODS

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals: Animals were housed at Bioqual Inc., Rockville, MD. Approval for all animal work was granted by the Institutional Care and Use Committees of Bioqual and the National Institutes of Health, and has been determined to be in accordance with the guidelines outlined by the Animal Welfare Act and Regulation (USDA) and The Guide for the Care & Use of Laboratory Animals, 8th Edition (NIH). The regulatory standards of the indicated committees have also been followed in all the experiments which involved laboratory animals.

Ten outbred Indian-origin rhesus macaques (*Macaca mulatta*) were infected intrarectally (IR) via repetitive challenge with the SIVmac₂₅₁ swarm, used earlier in Del Prete et al.³⁹ until SIV RNA was detected in the plasma using qPCR. The macaques were all challenged with the same SIVmac₂₅₁ swarm via the same route (IR). Starting at \approx 48 weeks post-infection, all the animals were treated with a daily combination antiretroviral drug regimen of tenofovir disoproxil fumarate, emtricitabine and dolutegravir (TDF/FTC/DTG, Gilead Sciences, Inc).^{38,62} The sample size was 10, and all 10 macaques were challenged and treated in the same way (no experimental groups).

Peripheral blood mononuclear cells (PBMCs) were isolated, from which CD4⁺ T cells were purified using the Miltenyi NHP CD4 Isolation Kit according to the manufacturer's protocol. DNA was extracted using the QiAmp DNA Mini Kit according to the manufacturer's directions. The intact proviral DNA assay (IPDA) and single genome sequencing of the envelope gene were carried out as described in Fray et al.^{36,38,52} Sequences with clear defects such as deletion, APOBEC3G/F-induced stop codons or hypermutation were excluded. At the start of ART (week 0), plasma viral RNA was also sequenced for 9 of 10 animals. Plasma sequences could not be recovered from animal T624 due to the low viral load at this time point. During untreated infection, plasma viral RNA was sequenced for 4 of 10 macaques; see Tables S1 and S2 for sampling details.

The animals were followed for up to 216 weeks (\approx 4 years) on ART, with longitudinal sampling of peripheral blood at weeks 0, 4, 8, 12, 28, 48, 92, 104, 120, 146 and 208 or 216 post initiation of ART. When the decay kinetics of CD4⁺ T cells with intact proviruses were studied in this cohort,³⁸ the two sub-populations modeled here were identified in the first < 3 years. However, between 2 and 3 years, a third stable sub-population with no observable decay was observed, possibly maintained by clonal expansion. In this work, we restrict our analysis to the first 146 weeks to reduce the impact of this third CD4⁺ T cell sub-population and clonal expansion. This dataset contains more data than the one reported in Fray et al.,³⁸ with additional sequences collected from samples corresponding to weeks 0, 4, 8 and 92 (Table S1). With these additional sequences, an average of 22 sequences were obtained per animal at each time point (minimum 0, maximum 118 across animals). A complete list of the sampling times and number of sequences at each time is provided in Table S1. Further assay details are described in Fray et al.³⁸

METHOD DETAILS

Divergence dynamics from data: The non-defective proviral *env* sequences were aligned to a reference sequence using MAFFT (v7.490⁶³). Various possibilities were explored for the reference sequence - consensus or each sequence of the SIVmac₂₅₁ swarm,³⁹ and where possible, the consensus or each sequence from the first time point after confirmed infection. Divergence was computed as follows

$$d(t) = \frac{\sum_{m=0}^{\max} m \times N_m(t)}{\sum_{m=0}^{\max} N_m(t)},$$

(Equation 1)

Where $d(t)$ is the divergence at time t , m is the raw distance (normalized by aligned sequence length) between each sequence and the reference, and $N_m(t)$ is the number of sequences at time t having distance m from the reference sequence. Considering the start of ART as time 0, $N_m(0)$ is estimated using experimental data as described in section 7.3 below. $N_m(t)$ is then computed for each time t by modeling the decay of cells based on their half-life. The R *lm* function⁶⁴ was used to fit a linear model to identify the slopes in the divergence curve given by the mean divergence at each time point. The elbow point at which the slope of the divergence curve changes was identified by the time point at which the absolute value of the ratio of the resulting slopes ($|\text{slope 1}/\text{slope 2}|$) was the largest. The slopes were estimated using at least two time points, and the time point at the potential elbow was used for both the first and second slope. These computations were carried out in R (v4.2.0⁶⁴), using packages *ape* (v5.6.2⁶⁵) and *ips* (v0.0.11⁶⁶).

Estimation of CD4⁺ T cell half-lives: We fitted a biphasic exponential decline model to intact proviral SIV DNA measured using IPDA in the 10 macaques under ART. This model is given by $y = y_0(fe^{-b_1t} + (1-f)e^{-b_2t})$, where y is intact provirus and y_0 is its baseline value. The fraction f decays in the first (short-lived) phase with decay rate b_1 , while the fraction $(1-f)$ decays in the second (long-lived) phase with decay rate b_2 . The model was fitted in two ways, (i) using a mixed effects approach, where all macaques were fitted simultaneously, and (ii) using the data for each macaque separately. The two methods of estimation were carried out to account for the high degree of variability across macaques. Fitting was carried out in Monolix 2021R1.⁶⁷ The estimated half-lives are provided in Table S4.

Phylogenetic estimation of integration time: Phylogenetic trees were built using IQ-Tree (v2.3.2^{68,69}), using all the sequences (pre- and on-ART) for each macaque. Tree visualization was done in R (v4.2.0⁶⁴), using the package *ape* (v5.6.2⁶⁵). Integration times of on-ART sequences were estimated as follows. First, the trees were rooted to the most recent common ancestor of the sequences obtained during the first timepoint after confirmed infection. Then, linear regression was carried out using the *lm* function in R, to fit a line relating the sampling times of the pre-ART plasma RNA *env* sequences to their root-to-tip distances. This fitted line was then used to map the root-to-tip distance of on-ART sequences to their estimated integration times. The estimated integration times were used as-is even if

they happened to be after the known time when ART was initiated. For the timepoint when ART was initiated, pre-ART plasma sequences were used to build the phylogeny.

Estimation of initial divergence distributions: To estimate the distribution of mutations in non-defective *env* sequences of infected cells at the beginning of ART in the short- and long-lived CD4⁺ T cells, the following assumptions were made: (i) the sequences at the start of ART come from a mixture of short- and long-lived cells, and (ii) the sequences at the last time point come only from long-lived cells, i.e., when all the short-lived cells have died out. Thus, to obtain the distribution of mutations in the long-lived cells at the start of ART ($D_L(m, 0)$), we first scale the distribution of mutations in the total set of sequences at the last measured time point ($D_T(m, t_{max})$) to account for decay using the scaling factor $\frac{N_T(0)}{N_T(t_{max})}$. Here, $N_T(0)$ and $N_T(t_{max})$ refer to the total number of sequences at the start of ART and at the last sampled time point, respectively. We then multiply by $(1 - f)$ to ensure that the fraction of long-lived cells at the start of ART matches the value estimated from data. The distribution of mutations in the short-lived cells at the start of ART $D_S(m, 0)$, is then a direct subtraction of the estimated $D_L(m, 0)$ from the total distribution of mutations at the start of ART $D_T(m, 0)$. This can be written as

$$D_L(m, 0) = D_T(m, t_{max}) \frac{N_T(0)}{N_T(t_{max})} (1 - f),$$

(Equation 2)

$$D_S(m, 0) = D_T(m, 0) - D_L(m, 0).$$

(Equation 3)

We obtained a histogram of mutations by defining each bin to represent 1 mutation. The same procedure was followed whether the data for the start of ART comprises non-defective proviral *env* DNA or plasma viral *env* RNA.

Model of divergence kinetics: We developed a model of the dynamics of sequence divergence based on different initial distributions of mutations in short- and long-lived cells, and the corresponding dynamics of these cells once treatment was started, assuming no new cell infections and no cell proliferation; Figure 4. Each of the short- and long-lived cell populations had a distribution of mutations reflecting the divergence of the sequences at the initiation of ART. The initial distributions were either the experimentally estimated distributions or idealized test profiles for theoretical simulations (linear, exponential, or uniform distribution profiles). The full set of parameter values for the model are provided in Table S4. In each time step (=1 day), the populations decayed according to their corresponding decay rate, irrespective of which mutational bin they belong to. The divergence of the surviving cells was computed according to Equation 1. The model was run for the same amount of time as the experiments, or for the theoretical simulations for 4 years, with 10^3 runs for each pair of initial distributions.

We computed $N_m(t)$, the number of cells with provirus having m mutations at any time t , analytically as

$$N_m(t) = N_m(0)e^{-t/\tau},$$

(Equation 4)

where $N_m(0)$ is the number of sequences with m mutations at time 0, and τ is the half-life (divided by $\ln(2)$). This was used to analyze the relative contribution of parameters.

All decay simulations were carried out in Python 3.8.12 using packages pandas (v1.3.5)^{70,71} and numpy (1.20.3).⁷²

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical testing was performed using the paired two-sided Wilcoxon signed rank exact test (Results section 2.1, data in Tables S3 and S5; Results section 2.2, data in Tables S6 and S7) and the Mann Whitney U test (Results section 2.2, data in Tables S6 and S7). N numbers are provided in each result section where the p -value is reported. p values <0.05 are considered statistically significant.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

Support was provided by NIH/NIAID grants R01-AI087520 (to T.L.), Martin Delany Collaboratory UM1-AI164561 (to R.M.R.), PAVE-UM1AI164566 (to F.R. S.), P01-AI169615 (to R.F.S.), and R01AI028433 and R01-OD011095 (to A.S. P.). N.S. was supported by LANL LDRD Fellowship 20210959PRD3. F.R.S. was also supported by the Office of the NIH Director and National Institute of Dental & Craniofacial Research (DP5OD031834) and the Johns Hopkins University CFAR (P30AI094189). Support was also provided by the NIH Martin Delaney Collaboratory grant UM1AI164556 to D.H.B. and by the Howard Hughes Medical Institute to R.F.S. Animal studies were supported by AI124377, AI128751, AI149670, AI164556, and AI169615 for D.H.B.

REFERENCES

1. CDC centers for disease control, HIV and AIDS timeline. <https://npin.cdc.gov/pages/hiv-and-aids-timeline>, 2022.
2. WHO world health organization. <https://www.who.int/data/gho/data/themes/hiv-aids>, 2022.
3. Chun TW, Stuyver L, Mizell SB, Ehler LA, Mican JA, Baseler M, Lloyd AL, Nowak MA, and Fauci AS (1997). Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc. Natl. Acad. Sci. USA* 94, 13193–13197. [PubMed: 9371822]
4. Finzi D, Blankson J, Siliciano JD, Margolick JB, Chadwick K, Pierson T, Smith K, Lisiewicz J, Lori F, Flexner C, et al. (1999). Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nat. Med* 5, 512–517. [PubMed: 10229227]
5. Wong JK, Hezareh M, Günthard HF, Havlir DV, Ignacio CC, Spina CA, and Richman DD (1997). Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. *Science* 278, 1291–1295. [PubMed: 9360926]
6. Crooks AM, Bateson R, Cope AB, Dahl NP, Griggs MK, Kuruc JAD, Gay CL, Eron JJ, Margolis DM, Bosch RJ, et al. (2015). Precise quantitation of the latent HIV-1 reservoir: implications for eradication strategies. *J. Infect. Dis* 212, 1361–1365. [PubMed: 25877550]

7. Siliciano JD, Kajdas J, Finzi D, Quinn TC, Chadwick K, Margolick JB, Kovacs C, Gange SJ, and Siliciano RF (2003). Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nat. Med* 9, 727–728. [PubMed: 12754504]
8. Davey RT Jr., Bhat N, Yoder C, Chun TW, Metcalf JA, Dewar R, Natarajan V, Lempicki RA, Adelsberger JW, Miller KD, et al. (1999). HIV-1 and T cell dynamics after interruption of highly active antiretroviral therapy (HAART) in patients with a history of sustained viral suppression. *Proc. Natl. Acad. Sci. USA* 96, 15109–15114. [PubMed: 10611346]
9. Rothenberger MK, Keele BF, Wietgreffe SW, Fletcher CV, Beilman GJ, Chipman JG, Khoruts A, Estes JD, Anderson J, Callisto SP, et al. (2015). Large number of rebounding/founder HIV variants emerge from multifocal infection in lymphatic tissues after treatment interruption. *Proc. Natl. Acad. Sci. USA* 112, E1126–E1134. [PubMed: 25713386]
10. Whitney JB, Hill AL, Sanisetty S, Penaloza-MacMaster P, Liu J, Shetty M, Parenteau L, Cabral C, Shields J, Blackmore S, et al. (2014). Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. *Nature* 512, 74–77. [PubMed: 25042999]
11. Henrich TJ, Hatano H, Bacon O, Hogan LE, Rutishauser R, Hill A, Kearney MF, Anderson EM, Buchbinder SP, Cohen SE, et al. (2017). HIV-1 persistence following extremely early initiation of antiretroviral therapy (ART) during acute HIV-1 infection: an observational study. *PLoS Med* 14, e1002417. [PubMed: 29112956]
12. Colby DJ, Trautmann L, Pinyakorn S, Leyre L, Pagliuzza A, Kroon E, Rolland M, Takata H, Buranapraditkun S, Intasan J, et al. (2018). Rapid HIV RNA rebound after antiretroviral treatment interruption in persons durably suppressed in Fiebig I acute HIV infection. *Nat. Med* 24, 923–926. [PubMed: 29892063]
13. Joos B, Fischer M, Kuster H, Pillai SK, Wong JK, Böni J, Hirschel B, Weber R, Trkola A, and Günthard HF; Swiss HIV Cohort Study (2008). HIV rebounds from latently infected cells, rather than from continuing low-level replication. *Proc. Natl. Acad. Sci. USA* 105, 16725–16730. [PubMed: 18936487]
14. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, Sun C, Grayson T, Wang S, Li H, et al. (2008). Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. USA* 105, 7552–7557. [PubMed: 18490657]
15. Gondim MVP, Sherrill-Mix S, Bibollet-Ruche F, Russell RM, Trimboli S, Smith AG, Li Y, Liu W, Avitto AN, DeVito JC, et al. (2021). Heightened resistance to host type 1 interferons characterizes HIV-1 at transmission and after antiretroviral therapy interruption. *Sci. Transl. Med* 13, eabd8179. [PubMed: 33441429]
16. Keele BF, Okoye AA, Fennessey CM, Varco-Merth B, Immonen TT, Kose E, Conchas A, Pinkevych M, Lipkey L, Newman L, et al. (2024). Early antiretroviral therapy in siv-infected rhesus macaques reveals a multiphasic, saturable dynamic accumulation of the rebound competent viral reservoir. *PLoS Pathog* 20, e1012135. [PubMed: 38593120]
17. Jones BR, Kinloch NN, Horacek J, Ganase B, Harris M, Harrigan PR, Jones RB, Brockman MA, Joy JB, Poon AFY, and Brumme ZL (2018). Phylogenetic approach to recover integration dates of latent HIV sequences within-host. *Proc. Natl. Acad. Sci. USA* 115, E8958–E8967. [PubMed: 30185556]
18. Brooks K, Jones BR, Diletnia DA, Wilkins DJ, Claiborne DT, McInally S, Gilmour J, Kilembe W, Joy JB, Allen SA, et al. (2020). HIV-1 variants are archived throughout infection and persist in the reservoir. *PLoS Pathog* 16, e1008378. [PubMed: 32492044]
19. Jones BR, Miller RL, Kinloch NN, Tsai O, Rigsby H, Sudderuddin H, Shahid A, Ganase B, Brumme CJ, Harris M, et al. (2020). Genetic diversity, compartmentalization, and age of HIV proviruses persisting in CD4+ T cell subsets during long-term combination antiretroviral therapy. *J. Virol* 94, e01786. [PubMed: 31776273]
20. Immonen TT, Conway JM, Romero-Severson EO, Perelson AS, and Leitner T (2015). Recombination enhances HIV-1 envelope diversity by facilitating the survival of latent genomic fragments in the plasma virus population. *PLoS Comput. Biol* 11, e1004625. [PubMed: 26693708]
21. Abrahams M-R, Joseph SB, Garrett N, Tyers L, Moeser M, Archin N, Council OD, Matten D, Zhou S, Doolabh D, et al. (2019). The replication-competent HIV-1 latent reservoir is primarily

- established near the time of therapy initiation. *Sci. Transl. Med* 11, eaaw5589. [PubMed: 31597754]
22. Natalie NK, Shahid A, Dong W, Kirkby D, Jones BR, Beelen CJ, MacMillan D, Lee GQ, Mota TM, Sudderuddin H, et al. (2023). HIV reservoirs are dominated by genetically younger and clonally enriched proviruses. *mBio* 14, e02417–e02423. [PubMed: 37971267]
 23. Joseph SB, Abrahams M-R, Moeser M, Tyers L, Archin NM, Council OD, Sondgeroth A, Spielvogel E, Emery A, Zhou S, et al. (2024). The timing of HIV-1 infection of cells that persist on therapy is not strongly influenced by replication competency or cellular tropism of the provirus. *PLoS Pathog* 20, e1011974. [PubMed: 38422171]
 24. Shahid A, MacLennan S, Jones BR, Sudderuddin H, Dang Z, Co-barrubias K, Duncan MC, Kinloch NN, Dapp MJ, Archin NM, et al. (2024). The replication-competent HIV reservoir is a genetically restricted, younger subset of the overall pool of hiv proviruses persisting during therapy, which is highly genetically stable over time. *J. Virol* 98. e01655–23, 2024.
 25. Pankau MD, Reeves DB, Harkins E, Ronen K, Jaoko W, Mandaliya K, Graham SM, McClelland RS, Matsen IV FA, Schiffer JT, et al. (2020). Dynamics of HIV DNA reservoir seeding in a cohort of superinfected Kenyan women. *PLoS Pathog* 16, e1008286. [PubMed: 32023326]
 26. Brodin J, Zanini F, Thebo L, Lanz C, Bratt G, Neher RA, and Albert J (2016). Establishment and stability of the latent HIV-1 DNA reservoir. *Elife* 5, e18889. [PubMed: 27855060]
 27. Lorenzo-Redondo R, Fryer HR, Bedford T, Kim E-Y, Archer J, Pond SLK, Chung Y-S, Penugonda S, Chipman J, Fletcher CV, et al. (2016). Persistent HIV-1 replication maintains the tissue reservoir during therapy. *Nature* 530, 51–56. [PubMed: 26814962]
 28. Rose R, Lamers SL, Nolan DJ, Maidji E, Faria NR, Pybus OG, Dollar JJ, Maruniak SA, McAvoy AC, Salemi M, et al. (2016). HIV maintains an evolving and dispersed population in multiple tissues during suppressive combined antiretroviral therapy in individuals with cancer. *J. Virol* 90, 8984–8993. [PubMed: 27466425]
 29. Rosenbloom DIS, Hill AL, Laskey SB, and Siliciano RF (2017). Reevaluating evolution in the HIV reservoir. *Nature* 551, E6–E9. [PubMed: 29168805]
 30. Kearney MF, Wiegand A, Shao W, McManus WR, Bale MJ, Luke B, Maldarelli F, Mellors JW, and Coffin JM (2017). Ongoing HIV replication during ART reconsidered. *Open forum infectious diseases* 4 (Oxford University Press), pp. ofx173.
 31. Bozzi G, Simonetti FR, Watters SA, Anderson EM, Gouzoulis M, Kearney MF, Rote P, Lange C, Shao W, Gorelick R, et al. (2019). No evidence of ongoing HIV replication or compartmentalization in tissues during combination antiretroviral therapy: Implications for HIV eradication. *Sci. Adv* 5, eaav2045. [PubMed: 31579817]
 32. Kearney MF, Spindler J, Shao W, Yu S, Anderson EM, O'Shea A, Rehm C, Poethke C, Kovacs N, Mellors JW, et al. (2014). Lack of detectable HIV-1 molecular evolution during suppressive antiretroviral therapy. *PLoS Pathog* 10, e1004010. [PubMed: 24651464]
 33. Hosmane NN, Kwon KJ, Bruner KM, Capoferri AA, Beg S, Rosenbloom DIS, Keele BF, Ho Y-C, Siliciano JD, and Siliciano RF (2017). Proliferation of latently infected CD4+ T cells carrying replication-competent HIV-1: Potential role in latent reservoir dynamics. *J. Exp. Med* 214, 959–972. [PubMed: 28341641]
 34. Wang Z, Gurule EE, Brennan TP, Gerold JM, Kwon KJ, Hosmane NN, Kumar MR, Beg SA, Capoferri AA, Ray SC, et al. (2018). Expanded cellular clones carrying replication-competent HIV-1 persist, wax, and wane. *Proc. Natl. Acad. Sci. USA* 115, E2575–E2584. [PubMed: 29483265]
 35. Cho A, Gaebler C, Oliveira T, Ramos V, Saad M, Lorenzi JCC, Gazumyan A, Moir S, Caskey M, Chun T-W, and Nussenzweig MC (2022). Longitudinal clonal dynamics of HIV-1 latent reservoirs measured by combination quadruplex polymerase chain reaction and sequencing. *Proc. Natl. Acad. Sci. USA* 119, e2117630119. [PubMed: 35042816]
 36. Bruner KM, Wang Z, Simonetti FR, Bender AM, Kwon KJ, Sengupta S, Fray EJ, Beg SA, Antar AAR, Jenike KM, et al. (2019). A quantitative approach for measuring the reservoir of latent HIV-1 proviruses. *Nature* 566, 120–125. [PubMed: 30700913]
 37. White JA, Simonetti FR, Beg S, McMyn NF, Dai W, Bachmann N, Lai J, Ford WC, Bunch C, Jones JL, et al. (2022). Complex decay dynamics of HIV virions, intact and defective proviruses,

- and 2LTR circles following initiation of antiretroviral therapy. *Proc. Natl. Acad. Sci. USA* 119, e2120326119. [PubMed: 35110411]
38. Fray EJ, Wu F, Simonetti FR, Zitzmann C, Sambaturu N, Molina-Paris C, Bender AM, Liu P-T, Ventura JD, Wiseman RW, et al. (2023). Antiretroviral therapy reveals triphasic decay of intact siv genomes and persistence of ancestral variants. *Cell Host Microbe* 31, 356–372. [PubMed: 36809762]
 39. Del Prete GQ, Scarlotta M, Newman L, Reid C, Parodi LM, Roser JD, Oswald K, Marx PA, Miller CJ, Desrosiers RC, et al. (2013). Comparative characterization of transfection-and infection-derived simian immunodeficiency virus challenge stocks for in vivo nonhuman primate studies. *J. Virol* 87, 4584–4595. [PubMed: 23408608]
 40. Coffin J, and Swanstrom R (2013). HIV pathogenesis: dynamics and genetics of viral populations and infected cells. *Cold Spring Harb. Perspect. Med* 3, a012526. [PubMed: 23284080]
 41. Siliciano JD, and Siliciano RF (2013). Recent trends in HIV-1 drug resistance. *Curr. Opin. Virol* 3, 487–494. [PubMed: 24021560]
 42. Sung YP, Love TMT, Perelson AS, Mack WJ, and Lee HY (2016). Molecular clock of HIV-1 envelope genes under early immune selection. *Retrovirology* 13, 1–12. [PubMed: 26728316]
 43. Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, Farzadegan H, Gupta P, Rinaldo CR, Learn GH, He X, et al. (1999). Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J. Virol* 73, 10489–10502. [PubMed: 10559367]
 44. Saha A, and Dixit NM (2020). Pre-existing resistance in the latent reservoir can compromise vrc01 therapy during chronic hiv-1 infection. *PLoS Comput. Biol* 16, e1008434. [PubMed: 33253162]
 45. Schröter J, Anelone AJN, Yates AJ, and De Boer RJ (1999). Time to viral suppression in perinatally hiv-infected infants depends on the viral load and cd4 t-cell percentage at the start of treatment. *Journal of acquired immune deficiency syndromes* 83, 522–529.
 46. Bachmann N, Von Siebenthal C, Vongrad V, Turk T, Neumann K, Beerenwinkel N, Bogojeska J, Fellay J, Roth V, Kok YL, et al. (2019). Determinants of hiv-1 reservoir size and long-term dynamics during suppressive art. *Nat. Commun* 10, 3193. [PubMed: 31324762]
 47. Morris SE, Dziobek-Garrett L, Strehlau R, Schröter J, Shiao S, Anelone AJN, Paximadis M, de Boer RJ, Abrams EJ, Tiemessen CT, et al. (2020). Quantifying the dynamics of hiv decline in perinatally-infected neonates on antiretroviral therapy. *J. Acquir. Immune Defic. Syndr* 85, 209–218. [PubMed: 32576731]
 48. Giardina F, Romero-Severson EO, Axelsson M, Svedhem V, Leitner T, Britton T, and Albert J (2019). Getting more from heterogeneous hiv-1 surveillance data in a high immigration country: estimation of incidence and undiagnosed population size using multiple biomarkers. *Int. J. Epidemiol* 48, 1795–1803. [PubMed: 31074780]
 49. Lundgren E, Romero-Severson E, Albert J, and Leitner T (2022). Combining biomarker and virus phylogenetic models improves hiv-1 epidemiological source identification. *PLoS Comput. Biol* 18, e1009741. [PubMed: 36026480]
 50. Gunst JD, Pahus MH, Rosás-Umbert M, Lu IN, Benfield T, Nielsen H, Johansen IS, Mohey R, Østergaard L, Kastrup V, et al. (2022). Early intervention with 3BNC117 and romidepsin at antiretroviral treatment initiation in people with HIV-1: a phase 1b/2a, randomized trial. *Nat. Med* 28, 2424–2435. [PubMed: 36253609]
 51. Wagner TA, McKernan JL, Tobin NH, Tapia KA, Mullins JI, and Frenkel LM (2013). An increasing proportion of monotypic HIV-1 DNA sequences during antiretroviral treatment suggests proliferation of HIV-infected cells. *J. Virol* 87, 1770–1778. [PubMed: 23175380]
 52. Bender AM, Simonetti FR, Kumar MR, Fray EJ, Bruner KM, Timmons AE, Tai KY, Jenike KM, Antar AAR, Liu PT, et al. (2019). The landscape of persistent viral genomes in ART-treated SIV, SHIV, and HIV-2 infections. *Cell Host Microbe* 26, 73–85. [PubMed: 31295427]
 53. Ferris AL, Wells DW, Guo S, Del Prete GQ, Swanstrom AE, Coffin JM, Wu X, Lifson JD, and Hughes SH (2019). Clonal expansion of SIVinfected cells in macaques on antiretroviral therapy is similar to that of HIV-infected cells in humans. *PLoS Pathog* 15, e1007869. [PubMed: 31291371]
 54. Long S, Fennessey CM, Newman L, Reid C, O'Brien SP, Li Y, Del Prete GQ, Lifson JD, Gorelick RJ, and Keele BF (2019). Evaluating the intactness of persistent viral genomes in simian

- immunodeficiency virus-infected rhesus macaques after initiating antiretroviral therapy within one year of infection. *J. Virol* 94, e01308–19. [PubMed: 31597776]
55. Imamichi H, Dewar RL, Adelsberger JW, Rehm CA, O'doherty U, Paxinos EE, Fauci AS, and Lane HC (2016). Defective HIV-1 proviruses produce novel protein-coding RNA species in HIV-infected patients on combination antiretroviral therapy. *Proc. Natl. Acad. Sci. USA* 113, 8783–8788. [PubMed: 27432972]
 56. Pollack RA, Jones RB, Perteu M, Bruner KM, Martin AR, Thomas AS, Capoferri AA, Beg SA, Huang S-H, Karandish S, et al. (2017). Defective HIV-1 proviruses are expressed and can be recognized by cytotoxic T lymphocytes, which shape the proviral landscape. *Cell Host Microbe* 21, 494–506. [PubMed: 28407485]
 57. Stevenson EM, Ward AR, Truong R, Thomas AS, Huang SH, Dilling TR, Terry S, Bui JK, Mota TM, Danesh A, et al. (2021). HIV-specific T cell responses reflect substantive in vivo interactions with antigen despite long-term therapy. *JCI insight* 6, e142640. [PubMed: 33400687]
 58. White JA, Wu F, Yasin S, Moskovljevic M, Varriale J, Dragoni F, Camilo-Contreras A, Duan J, Zheng MY, Tadzong NF, et al. (2023). Clonally expanded HIV-1 proviruses with 5'-leader defects can give rise to nonsuppressible residual viremia. *J. Clin. Investig* 133, e165245. [PubMed: 36602866]
 59. Immonen TT, Fennessey CM, Lipkey L, Newman L, Macairan A, Bosche M, Waltz N, Del Prete GQ, Lifson JD, and Keele BF (2024). No evidence for ongoing replication on ART in SIV-infected macaques. *Nat. Commun* 15, 5093. [PubMed: 38877003]
 60. Kuo H-H, Banga R, Lee GQ, Gao C, Cavassini M, Corpataux J-M, Blackmer JE, Zur Wiesch S, Yu XG, Pantaleo G, et al. (2020). Blood and lymph node dissemination of clonal genome-intact human immunodeficiency virus 1 DNA sequences during suppressive antiretroviral therapy. *J. Infect. Dis* 222, 655–660. [PubMed: 32236405]
 61. Bozzi G, Simonetti FR, Watters SA, Anderson EM, Gouzoulis M, Kearney MF, Rote P, Lange C, Shao W, Gorelick R, et al. (2019). No evidence of ongoing HIV replication or compartmentalization in tissues during combination antiretroviral therapy: Implications for HIV eradication. *Sci. Adv* 5, eaav2045. [PubMed: 31579817]
 62. Del Prete GQ, Smedley J, Macallister R, Jones GS, Li B, Hattersley J, Zheng J, Piatak M Jr., Keele BF, Hesselgesser J, et al. (2016). Comparative evaluation of coformulated injectable combination antiretroviral therapy regimens in simian immunodeficiency virus-infected rhesus macaques. *AIDS Res. Hum. Retroviruses* 32, 163–168. [PubMed: 26150024]
 63. Katoh K, and Standley DM (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol* 30, 772–780. [PubMed: 23329690]
 64. R Core Team (2021). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).
 65. Paradis E, and Schliep K (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. [PubMed: 30016406]
 66. Heibl C (2008). PHYLOCH: R language tree plotting tools and interfaces to diverse phylogenetic software packages. <http://www.christopheibl.de/rpackages.html>.
 67. Lavielle M (2014). Mixed Effects Models for the Population Approach: Models, Tasks, Methods and Tools (CRC press).
 68. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, and Lanfear R (2020). Iq-tree 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol* 37, 1530–1534. [PubMed: 32011700]
 69. Kalyanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, and Jeremiin LS (2017). Modelfinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. [PubMed: 28481363]
 70. (2020). The pandas development team. pandas-dev/pandas: Pandas. <https://github.com/pandas-dev/pandas>.
 71. McKinney W (2010). Data Structures for Statistical Computing in Python. In van der Walt Stéfan, Millman J, ed., pp. 56–61.

72. Harris CR, Millman KJ, Van der Walt van SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, et al. (2020). Array programming with NumPy. *Nature* 585, 357–362. [PubMed: 32939066]

Highlights

- SIV proviral divergence declines in two phases over time
- Short-lived CD4⁺ cells harbor more diverged viruses than long-lived cells
- Long-lived CD4⁺ cells contain SIVs deposited earlier in infection than short-lived cells

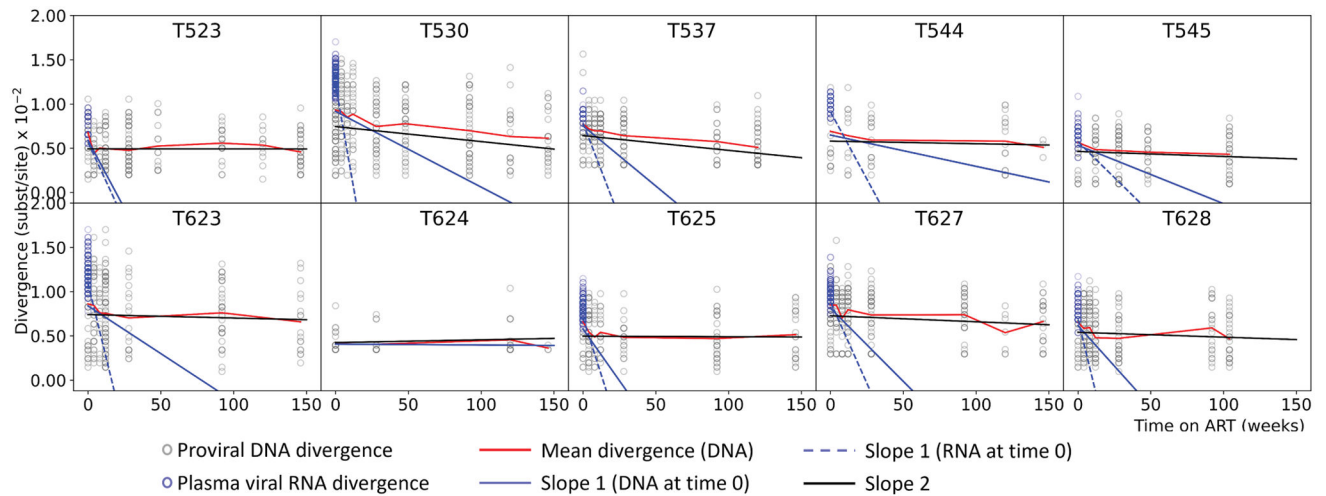


Figure 1. Biphasic divergence decline

Biphasic divergence decline of non-defective proviral *env* sequences while on suppressive ART (mean divergence curve shown in red), with an initial rapid decline followed by a second slower decline. Divergence in individual sequences is shown as circles (non-defective proviral *env* divergence as gray circles and plasma viral *env* RNA divergence as blue circles). The first slope is shown by blue lines, using non-defective proviral *env* DNA at the start of ART as a solid line and using plasma viral *env* RNA at the start of ART as a dashed line. The second slope is shown by a black solid line. Divergence was computed from the consensus sequence of the SIVmac₂₅₁ swarm.³⁹ Plasma viral RNA could not be collected for macaque T624.

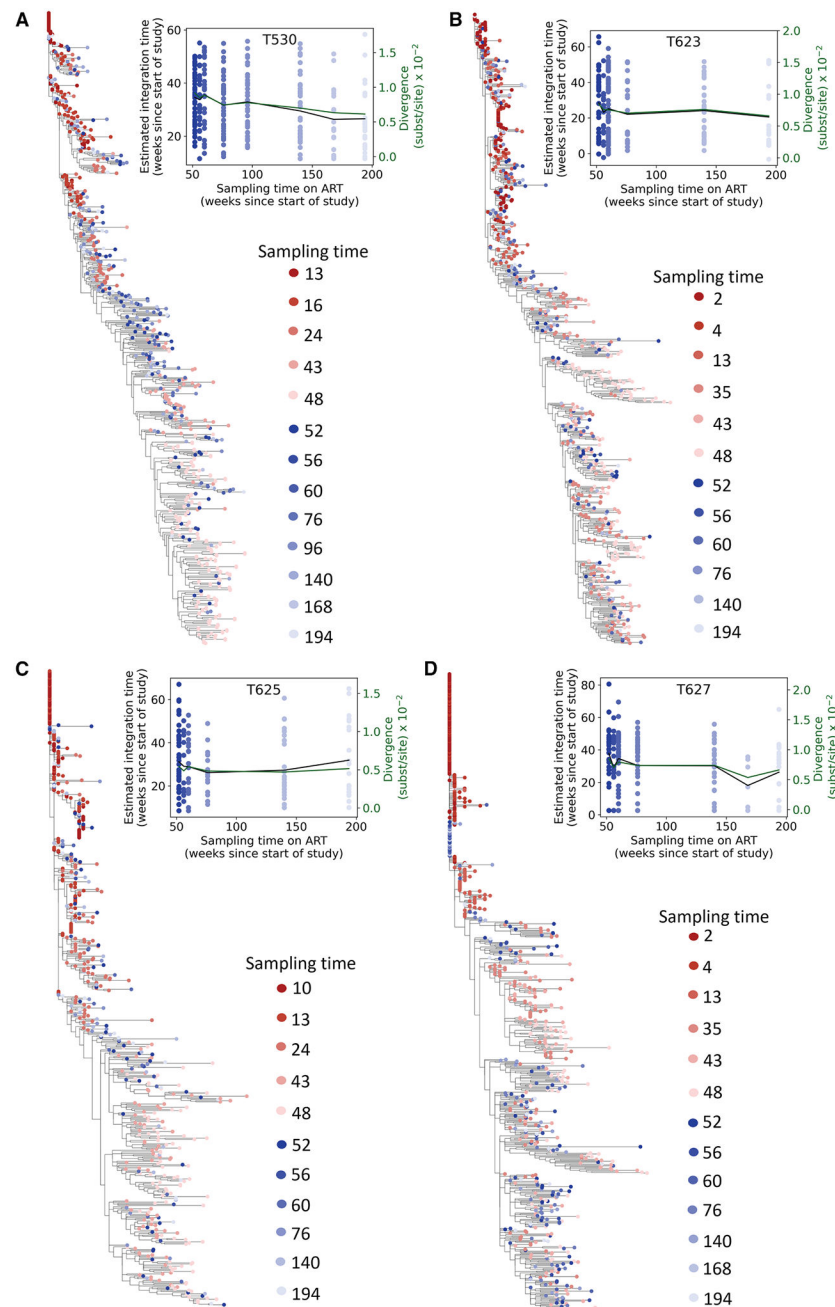


Figure 2. Integration time

Phylogenetic analyses for macaques (A) T530, (B) T623, (C) T625, and (D) T627, for whom pre-ART sequences were available. The tips in the phylogenetic trees corresponding to samples obtained pre-ART are colored red, while those obtained on-ART are colored blue, with a gradient such that samples obtained later are lighter. The inset plot for each macaque shows the estimated integration time for each proviral non-defective *env* DNA sequence collected on-ART (circles, colored in the same way as the tips in the phylogenetic tree) and the mean integration time for each such time point (black line) on the left y axis.

The right y axis shows the divergence dynamics of the on-ART sequences (dark green line).
All times are in terms of weeks since the start of the study.

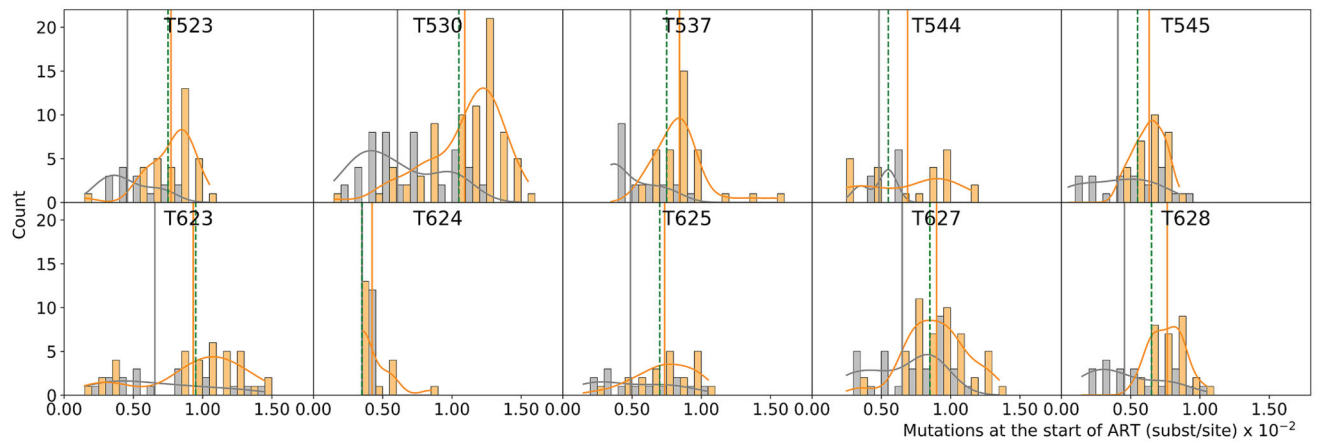


Figure 3. Mutational distributions in short- and long-lived cells

Mutational distributions of proviruses in short-lived (orange) and long-lived (gray) populations at the start of ART. Orange and gray curves show the smoothed-out histograms. Solid vertical lines correspond to the mean divergence of short-lived (orange) and long-lived (gray) populations. Vertical green dashed lines show the median divergence of the total population

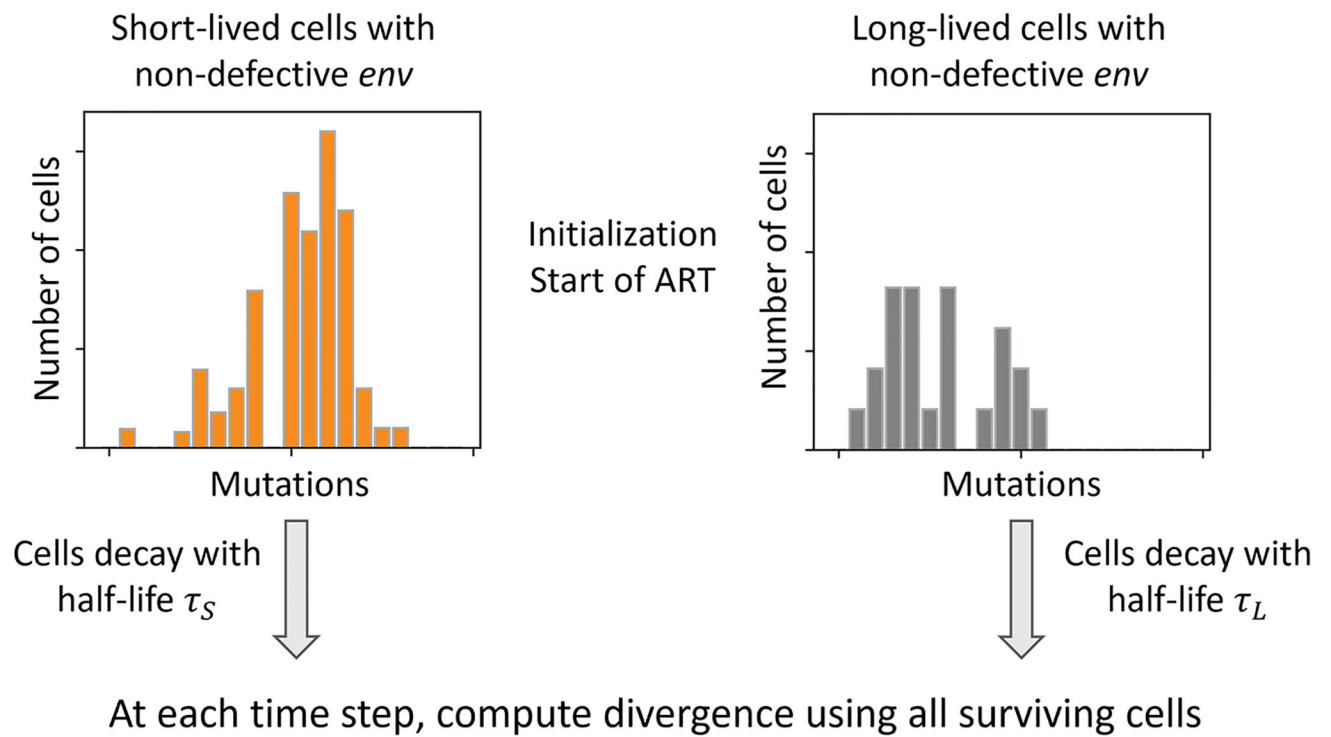


Figure 4. Model schematic

Schematic figure of the decay model used in this work. Divergence is computed using Equation 1.

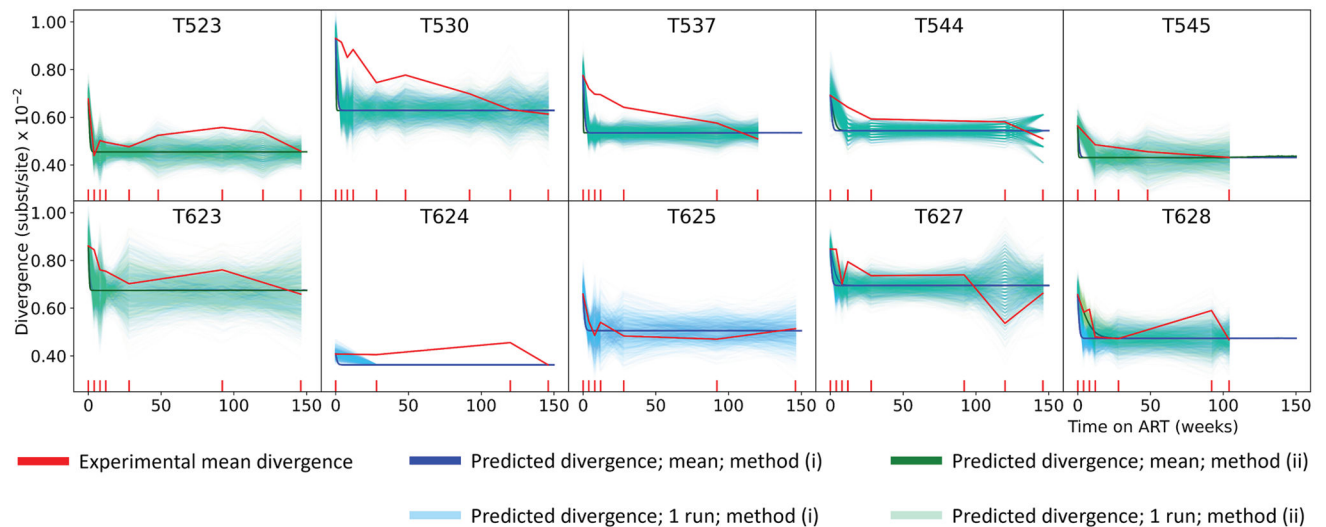


Figure 5. Experimental and simulated divergence dynamics

Empirically observed mean divergence dynamics (red lines) compared to model simulated divergence. Divergence dynamics using CD4⁺ T cell half-lives estimated by first pooling the data of all 10 macaques and then adjusting for individual variations are shown by blue lines. Divergence dynamics using half-lives learned separately for each individual macaque are shown by green lines. Bold blue/green lines show the means of 10^3 simulations. Thin blue/green lines show individual stochastic simulations of divergence obtained by sampling according to experimental times and number of sequences. Turquoise is the result of the thin blue and green lines overlapping. For macaque T523, the blue and green lines overlap exactly, causing only the green lines to be visible. For macaques T624 and T625, green lines are absent, as the data were insufficient to estimate half-lives individually. In T624, the thin blue lines overlap exactly with the thick blue line after ≈ 20 weeks. The red tick marks on the x axis show the times when experimental samples were obtained

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
SIVmac ₂₅₁ swarm	Dr. Dan Barouch, BIDMC	N/A
SIVmac ₂₅₁ -infected Rhesus macaque PBMCs	Dr. Dan Barouch, BIDMC	N/A
Experimental models: Organisms/strains		
Rhesus macaque (<i>Macaca mulatta</i>) infected with SIVmac ₂₅₁	Indian origin	Animals T523, T530, T537, T544, T545, T623, T624, T625, T627 & T628
Deposited data		
SIV – plasma RNA sequences	Fray et al. ³⁷	GenBank: OQ168641-OQ168979
SIV – non-defective proviral DNA sequences	Fray et al. ³⁷	GenBank: OQ168980-OQ170751
SIV – plasma RNA sequences	This paper	GenBank: PV293187 - PV294728
SIV – non-defective proviral DNA sequences	This paper	GenBank: PV292351 - PV293186
Software and algorithms		
Monolix	https://lixoft.com/	2020R1
MAFFT	https://mafft.cbrc.jp/	7.490
R	https://www.r-project.org/	4.2.0
R package ape	https://cran.r-project.org/package=ape	5.6.2
R package ips	https://cran.r-project.org/package=ips	0.0.11
IQ-Tree	http://www.iqtree.org/	2.3.2
Python	https://www.anaconda.com/	3.8.12
Python package pandas	https://pypi.org/project/pandas/	1.3.5
Python package numpy	https://pypi.org/project/numpy/	1.20.3