

RESEARCH ARTICLE

Recombination Enhances HIV-1 Envelope Diversity by Facilitating the Survival of Latent Genomic Fragments in the Plasma Virus Population

Taina T. Immonen^{1*}, Jessica M. Conway², Ethan O. Romero-Severson¹, Alan S. Perelson¹, Thomas Leitner¹

1 Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, New Mexico, United States of America, **2** Department of Mathematics, The Pennsylvania State University, University Park, Pennsylvania, United States of America

* tti@lanl.gov



OPEN ACCESS

Citation: Immonen TT, Conway JM, Romero-Severson EO, Perelson AS, 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(12): e1004625. doi:10.1371/journal.pcbi.1004625

Editor: Roger Dimitri Kouyos, University of Zurich, SWITZERLAND

Received: February 21, 2015

Accepted: October 25, 2015

Published: December 22, 2015

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This research was funded by the National Institutes of Health grants R01-AI028433 and R01-OD011095 (awarded to ASP), and the National Institutes of Health grant AI106437 (awarded to TL). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

HIV-1 is subject to immune pressure exerted by the host, giving variants that escape the immune response an advantage. Virus released from activated latent cells competes against variants that have continually evolved and adapted to host immune pressure. Nevertheless, there is increasing evidence that virus displaying a signal of latency survives in patient plasma despite having reduced fitness due to long-term immune memory. We investigated the survival of virus with latent envelope genomic fragments by simulating within-host HIV-1 sequence evolution and the cycling of viral lineages in and out of the latent reservoir. Our model incorporates a detailed mutation process including nucleotide substitution, recombination, latent reservoir dynamics, diversifying selection pressure driven by the immune response, and purifying selection pressure asserted by deleterious mutations. We evaluated the ability of our model to capture sequence evolution *in vivo* by comparing our simulated sequences to HIV-1 envelope sequence data from 16 HIV-infected untreated patients. Empirical sequence divergence and diversity measures were qualitatively and quantitatively similar to those of our simulated HIV-1 populations, suggesting that our model invokes realistic trends of HIV-1 genetic evolution. Moreover, reconstructed phylogenies of simulated and patient HIV-1 populations showed similar topological structures. Our simulation results suggest that recombination is a key mechanism facilitating the persistence of virus with latent envelope genomic fragments in the productively infected cell population. Recombination increased the survival probability of latent virus forms approximately 13-fold. Prevalence of virus with latent fragments in productively infected cells was observed in only 2% of simulations when we ignored recombination, while the proportion increased to 27% of simulations when we allowed recombination. We also found that the selection pressures exerted by different fitness landscapes influenced the shape of phylogenies, diversity trends, and survival of virus with latent genomic fragments. Our model predicts that the

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

persistence of latent genomic fragments from multiple different ancestral origins increases sequence diversity in plasma for reasonable fitness landscapes.

Author Summary

Increasing evidence suggests that HIV-1 released from activated latent cells survives in productively infected cells in patient plasma despite competition against better adapted virus variants that have evolved in response to the host immune pressure. Long-term survival requires that latent virus forms adapt to the host immune response so that they are not outcompeted. We simulated the dynamics of HIV-1 envelope sequence evolution in response to host immune pressure to investigate how virus from activated latent cells can survive despite having reduced fitness compared to the more evolved virus variants in patient plasma. The evolutionary trends of our simulated virus populations followed closely those observed in HIV-1 sequence data from 16 patients. Our simulation results suggest that recombination facilitates the survival of genomic fragments originating from virus activated from latent cells. Our model further predicts that sequence diversity increases with the number of latent genomic fragments from different origins that persist in plasma.

Introduction

Patients infected with HIV-1 require lifelong highly active antiretroviral therapy (HAART) to suppress infection. Treatment cessation typically leads to HIV viral rebound to pre-therapy levels; the resurgence is thought to be associated with activation of long-lived, latently HIV-infected cells. A cure for HIV therefore requires either clearance of all cells harboring latent virus, or prevention of virus release from the reservoirs after discontinuation of treatment. Increasing evidence suggests that latency plays an integral role throughout the life cycle of the virus. We recently observed that the majority of HIV-1 plasma sequences in two untreated chronically infected patients had accumulated significantly less mutations than expected, suggesting a period of latency during which no replication occurred in the history of these lineages [1]. Moreover, viral variants with reduced evolution consistent with periods of latency are frequently involved in transmission events [2–4]. While recent advances have shed light on the mechanisms leading to the establishment and maintenance of latent reservoirs [5, 6], the prevalence of viral sequences displaying a signal of latency in the replicating population remains enigmatic, especially in the absence of antiviral treatment.

HIV-1 is subject to selection pressure exerted by the immune system; strains that can avoid the immune response have an advantage within a host. The neutralizing antibody response in patient sera is much stronger against virus circulating in infection earlier than contemporaneous variants, with immunological memory persisting for years [7–9]. Virus from activated latent cells is therefore less fit due to long-term immunological memory than variants that have continually evolved in response to the host immune pressure. Yet, we detect a signal of latency in the replicating population, suggesting that virus from activated latent cells persists despite competing against better-adapted contemporaneous variants.

In order to survive, activated latent forms must either have a replicative advantage outweighing immune selection, or quickly acquire escape mutations to evade the immune response. While mutations associated with immune escape decrease the replicative capacity of

the virus between 0–24% [10], compensatory mutations can completely restore fitness in some variants [11]. It is thus unknown whether virus from activated latent cells is able to compete against better adapted contemporaneous virus due to higher replicative fitness. To persist in plasma after activation, the virus must adapt to immune pressure before being outcompeted. The virus may adapt by sequentially accumulating escape mutations at recognized epitopes, or by recombining with contemporaneous virus and simultaneously gaining multiple epitopes not recognized by the immune system.

Recombination provides a mechanism to preserve genomic fragments originating from activated latent cells within contemporaneous virus backbones adapted to the immune response. Further recombination events may propagate latent genomic fragments through the replicating plasma population, enhancing the adaptability potential of the virus. In this paper, we investigate the persistence of latent envelope genomic fragments in the replicating plasma population by developing a detailed model of HIV-1 sequence evolution and the cycling of lineages in and out of the latent reservoir.

There has been an increasing focus on modeling the dynamics of the latent reservoir, particularly in the context of delineating the effect of latency on viral blips during treatment, and the re-emergence of infection after treatment failure/interruption [12–16]. A common feature of these models is the cycling of infected cells through a latent compartment, in which they may divide or die before becoming activated and re-joining the productively infected cell population. Our agent-based model integrates the dynamics of the latent reservoir proposed in [12–14], with a sequence evolution framework first introduced in [17, 18], which follows how individual viruses mutate and recombine.

Vijay et al. developed an HIV sequence evolution model of infected patients that incorporates mutation, infection of cells by multiple virions, recombination, fitness selection, and epistatic interactions between multiple loci [17]. Their model, which describes quantitatively the evolution of HIV-1 diversity and divergence in patients, has been applied to a wide variety of questions, including the effect of recombination on sequence diversification [17], the effective population size of HIV-1 [18], the genetic structure of HIV-1 quasispecies and its error threshold [19], and the fraction of progeny viruses that must incorporate a drug treatment target for suppression of productive infection [20].

We extend the sequence evolution model by incorporating a latent compartment in which no evolution takes place, and by explicitly simulating how the virus population interacts with and adapts to the immune response. To evaluate whether our model captures sequence evolution *in vivo*, we compare our simulated sequences to sequences from 16 untreated or unsuccessfully treated patients followed longitudinally from seroconversion [9, 21, 22] in terms of sequence divergence (i.e. average distance of a population of sequences from a founder sequence), diversity (i.e. average pairwise distance between all sequences in a population), and phylogenetic tree shape measures.

Models

Model overview

We developed an agent-based model to simulate within-host HIV-1 sequence evolution after primary HIV-1 infection (PHI), tracking how long each virus lineage has spent in the latent reservoir throughout its history (Fig 1). The model incorporates a detailed mutation process including recombination, latent reservoir dynamics, diversifying selection pressure driven by the immune response, and purifying selection pressure asserted by deleterious mutations. During each generation, a population of virus infects a population of uninfected cells, with infecting viruses chosen according to their relative fitness. Genetic variability is introduced through

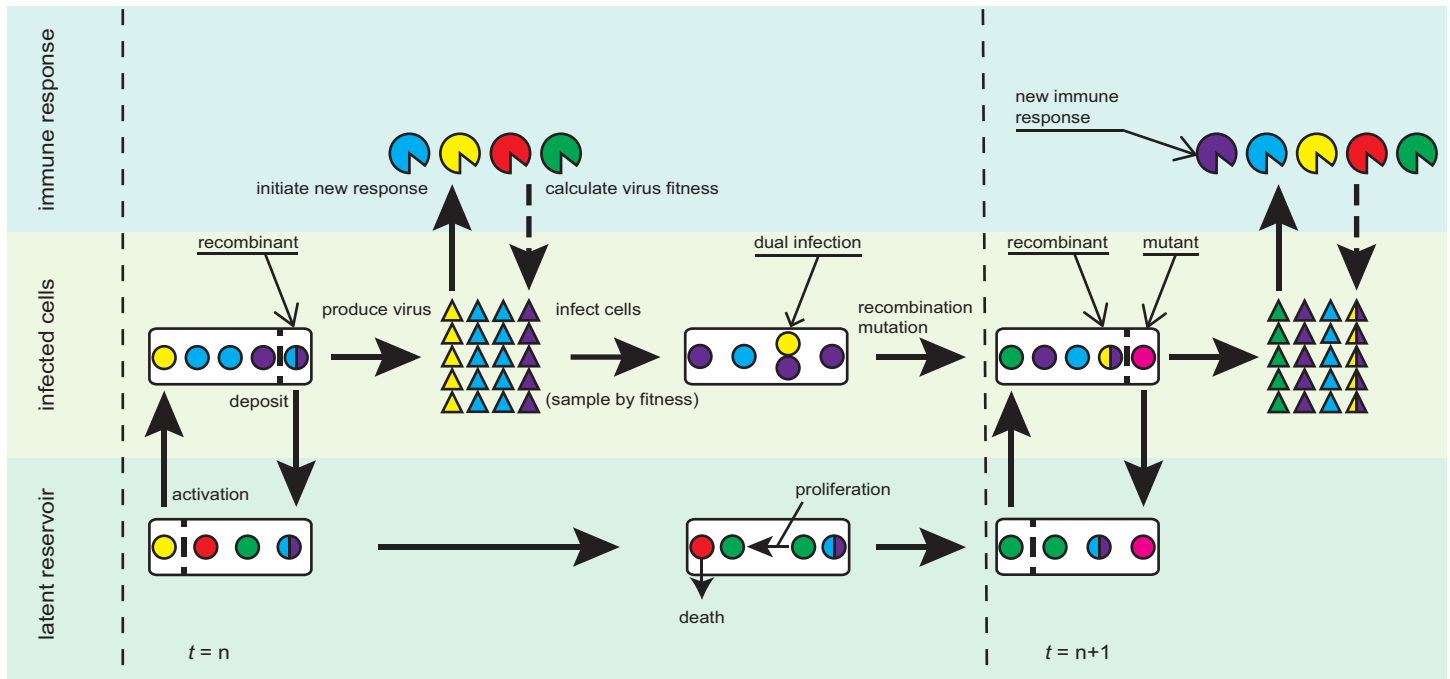


Fig 1. Model schematic. We model interactions between and within three compartments: the latent reservoir, productively infected cells, and the immune response during generation n . The circles represent infected cells (both productively infected and latent), the triangles progeny virus, and the pies (circular sectors) the immune response. The colors of the circles and triangles represent different viral epitopes, and the colors in the pies indicate which viral epitopes are recognized by the immune response. Each productively infected cell produces the same number of infectious virus particles. The population of virus is sampled based on fitness to form the next set of productively infected cells, where the fitness of each virus depends on whether it is recognized by the immune response. If a new viral antigen reaches high enough numbers in the plasma, it triggers an immune response. Upon infection a small fraction of cells becomes latent. To mimic this, we assign a small probability to an infected cell moving to the latent reservoir. Also, cells in the latent reservoir have some probability of being activated and joining the replicating population. The viral sequences in productively infected cells are mutated, mimicking events that occur during reverse transcription, and the two parental strains in dually infected cells have some probability of recombining. Cells in the latent reservoir have some probability of dying, and homeostatically proliferate such that the size of the reservoir is maintained.

doi:10.1371/journal.pcbi.1004625.g001

recombination and substitution during reverse transcription of viral RNAs to proviral DNA in infected cells. Infected cells have a nonzero probability of moving into the latent reservoir, while latent cells have a nonzero probability of dying, proliferating, or becoming activated. Infected cells and activated latent cells produce progeny virus, forming the population of virus infecting the next generation of cells.

The relative fitness of each virus depends on whether it has acquired mutations at invariant sites, which are considered deleterious, and how well it is recognized by the immune system. All virus sequences have multiple epitopes at pre-defined sites, which are presented to the immune system. If an epitope variant reaches a sufficient frequency in the plasma, it elicits an immune response, which remains in memory for the duration of the simulation. The breadth of the immune response is updated each generation by adding any newly recognized epitope variants to immunological memory.

While we simulate infections established by a single founder strain, the model can also accommodate multiple founder strains, which are observed in 20% of heterosexual cases [23]. The simulations were implemented using a computer program written in R [24], and the phylogenies were created using the R package ape (Analyses of Phylogenetics and Evolution) [25]. The algorithm is outlined in the supplement (S1 Text).

Dynamics of productively infected cells

We initialized the simulation by generating a set of N_V identical virions, where each virion consists of a string of L nucleotides (A, U, G, C). While HIV contains two RNA molecules, we considered only a single RNA molecule for each virus to simplify computations. Because recombination between two identical parental strains does not lead to new genetic variants, restricting our simulations to one RNA molecule per virus should have a negligible effect on our simulation results. We chose as the starting sequence the consensus of the first time point sequences, sampled close to seroconversion, of Patient A from the Amsterdam cohort [9]. The sequence is from *env*, and consists of approximately 700 nucleotides.

At each generation, N_C viruses are chosen to infect N_C uninfected cells such that each cell is infected with a single virus. Multiple infection of cells is described below. The probability of selecting a given virus is proportional to its relative fitness, which is a function of the strength of the immune response against it, and the number of deleterious mutations it has acquired. Fitness selection is described in detail below. Upon infection each cell has a small probability η of becoming latent and moving to the latent reservoir. The virus in the infected cells that remain productive undergo reverse transcription, where the viral RNA is copied into proviral DNA. This process includes both recombination and substitution.

To simplify computations, we only considered dual infections that lead to recombination. Based on estimates of Josefsson et al. [26], we assumed that 10% of cells are infected with a second virion, ignoring multiple infection with three or more virions. On average half of the progeny virus produced by a cell infected with two virions acquires one RNA molecule from each parent. To simplify computations, we allowed at most one cross-over event per sequence. This simplification is justified because while there are on average 2.8 cross-overs per each HIV-1 genome of 10,000 nucleotides every life cycle [27], our simulated *env* sequences were over 10 times shorter.

For a template switching rate of 3×10^{-4} per site per generation [27, 28], the expected number of cells accommodating a recombination event between two different parental strains is $N_R = 0.1 \times 0.5 \times 3 \times 10^{-4} \times N_C L$. We incorporated dual infection into the model by sampling N_R additional viruses from the virus population with probability proportional to their relative fitness and adding them to randomly chosen infected cells. The dually infected cells then undergo recombination.

We assumed that cross-over events are distributed uniformly across the genome, selecting the starting strand and the cross-over position $\in \{2, L - 1\}$ randomly. The sequence of the starting strand is copied until the cross-over position, the strand is then switched, and the second strand is copied to the end of the sequence. The recombinant sequence replaces the two parent strands, so that all infected cells contain a single viral sequence. The sequence is then mutated according to a general-time-reversible substitution model [29]. We used the nucleotide-specific substitution rates estimated from the sequences of Patient A via maximum likelihood (*PAUP**, [30]), which were representative of the rates found for all 16 patients. The proviral DNA of each infected cell is transcribed to viral RNAs, which are released as new virions. Each infected cell produces P identical virions, which form the current population of virus from which viruses that infect the next generation of cells are selected.

Dynamics of the latent reservoir

As the latent reservoir is established early in infection [31, 32], we initialized the reservoir by infecting each of the N_L cells with the same founder virus used to infect the replicating population. During effective antiretroviral treatment, the reservoir's half-life has been estimated to range from 4 to 44 months [33–36]. Assuming that the size of the latent reservoir, N_L , remains constant during untreated chronic infection, we set the probability of an infected cell becoming

latent equal to the decay rate of the latent reservoir. Each generation, we randomly select which (if any) infected cells become latent.

For simplicity, we assumed homogeneous activation and death rates, randomly selecting which cells in the latent compartment die or become activated each generation. Activated latent cells join the population of productively infected cells and produce progeny virus. We assumed that the latent reservoir is maintained through homeostatic proliferation [12, 13, 37], which we simulated by randomly duplicating cells to replace those lost to death and activation to keep the size of the latent pool constant. Specifically, if the size of latent reservoir N_L after the addition of any newly latent cells and the removal of any activated or dead cells is less than the target size N_L^* , we randomly duplicate $N_L^* - N_L$ of the remaining cells in the latent compartment. No proliferation takes place if the size of the latent reservoir exceeds N_L^* .

The model tracks how many generations each cell spends in the latent reservoir before being activated, with progeny viruses retaining the information for the duration of the lineage. Because viruses that have been latent for different amounts of time can recombine, we track latency separately at each site in the genome. If a virus that has never been latent recombines with a virus that has been latent for n generations, we consider the recombinant progeny a latent virus form, with the portion of its genome acquired from the non-latent parent recorded as having been latent for zero generations, and the portion acquired from the latent parent as having been latent n generations. At the end of the simulation, we know the position and age of every latent genomic fragment for each virus sequence. We are interested in the persistence of sequences that contain at least one latent genomic fragment consisting of at least one site originally from a latent ancestor.

Immune selection and fitness

The immune response against antigens with recognized epitopes consists of two major arms; neutralization of a pathogen via antibodies, and killing of infected cells by cell-mediated immune responses. Because we simulated the viral envelope, we focused primarily on the antibody response, assuming that 1) variants with a recognized epitope are less fit than those able to evade the immune response, and 2) once an epitope is recognized, it is retained in memory for all time. The immune response imposes diversifying selection pressure on the virus population, resulting in sequential escape at each epitope. A virus variant not recognized by the immune system has a fitness advantage, and starts to take over the virus population, eventually eliciting an immune response, and thus selection for new escape variants.

We chose the location of N_E epitopes randomly in the envelope fragment that we are simulating. Epitopes are approximately 30 nucleotides in length, and may be non-contiguous and overlapping [38]. As synonymous mutations at the nucleotide level have no effect at the amino acid level, we ignored third-codon positions, and assumed that the epitopes are 20 nucleotides long and non-overlapping.

Serum from HIV-1 infected individuals does not neutralize contemporaneous virus, but rather virus that dominated the population 3 to 6 months earlier [7]. To model this behavior, we imposed an immune response against a new escape variant not when it first appeared but after it had risen in frequency in the plasma, and increased the strength of the subsequent response gradually over time.

We began the immune response against the virus after 30 generations. Every generation thereafter, we determined which epitope variants had reached a sufficient frequency, f , in the plasma, and initiated an immune response against them. The epitopes were stored in immunological memory for the duration of the simulation, and all virus variants containing them had a fitness disadvantage. For simplicity, we assumed that to be recognized the nucleotide sequence

of a virus epitope must match perfectly with the sequence of a stored epitope. Note that virus may be recognized at multiple epitopes simultaneously.

Because antibody responses mature over time, we assumed that the fitness cost to the virus imposed by the immune response to epitope i of variant j increases linearly until it reaches some maximum value c_i^* . Thus, the fitness cost of epitope i of variant j is given by

$$c_{ij}(t) = \min\{c_i^*, c_i^*(t - t_{ij}^0)/d\}\delta_{ij}, t \geq t_{ij}^0,$$

where t_{ij}^0 is the time when the immune response against epitope i of variant j is introduced, d is the time it takes to reach full potency, and δ_{ij} is the Dirac delta function, where $\delta_{ij} = 1$ when epitope i of variant j is recognized, and $\delta_{ij} = 0$ when it is not recognized.

The strength of the immune response against a particular epitope depends in part on how accessible it is to antibodies/killer T-cells. While escape mutations appear shortly after seroconversion in HIV-1/SIV infection at some epitopes [10, 39], it may take years to see evidence of selection at others [40]. We initialized the maximum fitness cost of each epitope in the beginning of the simulation by drawing it from a uniform distribution, $c_i^* \sim U[0, c_{max}]$. We assumed that neutralization via antibodies is the primary driver of selection in chronic infection, and that virus is saturated with antibodies. In this scenario, the fitness loss of a virus variant upon antibody recognition is driven by the most potent antibody that binds it, $c_j^{imm} = \max(c_{ij}\delta_{ij})$. The distribution of fitness costs associated with the epitopes defines the fitness landscape, which influences the evolutionary trajectory of the virus population.

Purifying selection conserves sites in the HIV-1 genome where mutations would be deleterious to the virus. We assume that mutations at invariant sites (where no variability is observed across sequences from different time points) are inherently deleterious, and incur a multiplicative fitness cost, $c_j^{inv} = 1 - (1 - \psi)^k$, where ψ is the reduction in fitness per mutation, and k is the number of mutations. At the beginning of each simulation, we randomly distribute Lp_{inv} invariant sites across the sequence, where p_{inv} is the proportion of invariant sites. Note that the positions of the invariant sites vary between simulations, and may occur in both non-epitope and epitope regions. Following Ganusov et al. [41], we define the relative fitness of variant j as $f_j = (1 - c_j^{imm})(1 - c_j^{inv}) = (1 - \max(c_{ij}\delta_{ij}))(1 - \psi)^k$.

Selection of parameter values

We considered a pool of $N_C = 15000$ infected cells, which is in line with the mean effective population size of HIV-1 in chronic infection [18]. HIV-1 has a large burst size of approximately 50,000 [42], but only a small fraction of one in 1000 to 10,000 virions appear to be infectious [43–45]. Each productively infected cell therefore produces between 5 and 50 infectious virions. For consistency with the sequence evolution models developed by Vijay et al. and Balagam et al., we set the number of progeny virions produced by each productively infected cell to $P = 5$ [17, 18]. Following again Balagam et al., we set the generation time to 1.2 days [18], and the substitution rate to the mean of the mutation rates estimated by Mansky et al., $\mu = 3.5 \times 10^{-5}$ [46]. We set the proportion of invariant sites to 50%, corresponding to the mean proportion of invariant sites estimated from the HIV-1 sequence alignments of patients described in Bunnik et al. and Karlsson et al. [9, 22]. We ran the simulations for 3000 generations, corresponding to 10 years.

To maintain the ratio of productively infected to latent cells predicted by the homeostatic proliferation model of latency introduced by Kim et al. and Rong et al., we set the size of the latent reservoir to $N_L = 100$ [12, 13]. Assuming a total body load of 2×10^8 productively infected cells [47] and between 2.2×10^5 to 1.6×10^7 latent cells with replication competent DNA [48], there are between 0.0011 to 0.08 latent cells for every productively infected cell. In our

simulations with 15,000 productively infected cells, this corresponds to a latent reservoir size N_L of 16 to 1200 cells; our chosen latent reservoir size is close to the geometric mean of this range.

We chose a conservative estimate of 44 months for the half life of the latent reservoir [33, 36], which we used to calibrate the probability η of an infected cell becoming latent upon infection such that the size of the latent reservoir was maintained. Archin et al. recently estimated that the composite parameter $\beta\eta$ is on the order of 10^{-14} [15], where β is the mass-action infection rate constant in $ml^{-1} day^{-1}$. The latter has been estimated to be approximately $\beta = 1.5 \times 10^{-8}$ [49, 50]. Our value of $\eta = 3.5 \times 10^{-6}$ is therefore in line with these estimates. We set the death rate of latent cells to $d_L = 0.004$ per generation, corresponding to the estimated death rate of memory cells, which make up the bulk of the latent reservoir [13, 14]. Following Rong et al., we varied the activation rate from $a_L \in (0.001, 0.01)$ per generation so that between 0.1 and 1 latent cells were activated each generation [13, 14].

Barr et al. estimated that the minimum efficacy of three early neutralizing antibodies at blocking *de novo* infections ranged from 19.6% to 35.2% [10]. To account for the observation of Richman et al. [7] that some antibodies isolated from patient serum show no neutralization activity, we allowed the maximum fitness cost, c_i^* , due to recognition of a viral epitope to range from 0 to 0.40, with a mean of 0.2.

The remaining parameters associated with the immune response are not well known. For our default parameter setting, we assumed that there are ten epitopes in the approximately 700 nucleotide region of envelope that we simulate. In subsequent sensitivity analyses, we varied the number of epitopes between 0 and 15. While it is not known how many epitopes there are on the envelope, analysis of escape mutations to SIV in rhesus monkeys suggests that there are several [51]. We assumed that an antibody is created against an epitope variant once it reaches 1% frequency in the plasma. In sensitivity analyses, we varied the antigen frequency required for stimulating a new antibody response between 0.1% and 10%. Because patient serum does not neutralize contemporaneous virus but rather virus that circulated in the plasma at least 3 months earlier [7], we assumed that it takes 90 generations for a newly introduced immune response to reach its full potency (i.e. impose maximal fitness cost to the virus variant it recognizes, with fitness cost increasing linearly from zero over the 90 generations). The default values of the model parameters are summarized in Table 1.

To investigate the robustness of our results, we also considered alternative parameter values proposed in the literature. Recent estimates of the basic reproductive number of HIV-1 during primary infection suggest that an infectious cell generates at least eight new infected cells at the start of infection when target cells are not limiting. Thus, each infected cell produces at least $P = 8$ infectious progeny virions [52]. Following Pearson et al., we set the number of progeny virions to $P = 10$ to account for viral clearance [53]. We set the generation time of HIV-1 *in vivo* to two days, estimated from the decay dynamics of productively infected cells [54], and the substitution rate to the reverse transcriptase nucleotide substitution rate of approximately 2.2×10^{-5} [46, 55]. We further assumed that 10% of sites in the envelope region of interest are invariant, corresponding to the minimum proportion of invariant sites in the patient data. Because the mechanism of recombination is central to our investigation, we also varied the number of recombination events per generation between $0.1N_R$, and $10N_R$.

Sequence divergence and diversity

Estimating sequence divergence and diversity allows us to quantify viral evolution. Divergence is a measure of how far viral genomes have evolved from the founder strain, whereas diversity is a measure of the genomic variation in the viral population at any given time. Every 30 generations, we calculated the divergence and diversity of the HIV-1 sequences in the entire

Table 1. Parameters used in our model.

Parameter	Value	Description
N_C	15000	Number of cells in the replicating population [18]
N_L	100	Number of cells in the latent reservoir [13]
τ	1.2	Duration of each generation [days] [18]
L	669	Length of simulated DNA sequences [nucleotides] [9, 21, 22]
μ	3.5×10^{-5}	Mutation rate per site per generation [46]
N_R	150	Mean number of recombination events per generation [26–28]
P	5	Number of progeny virus produced by each infected cell [17, 42]
η	3.5×10^{-6}	Probability of infected cell becoming latent, calibrated to maintain size of latent reservoir, assuming 44 month half-life [35, 36]
a_L	0.001 – 0.01	Activation rate of latent cells [13, 14]
δ	0.004	Death rate of latent cells [13, 14]
N_E	10 (0–15)	Number of epitopes [51]
L_E	20	Length of epitopes, based on non-synonymous sites [nucleotides] [38]
c_i^*	$U[0,0.4]$	Maximum fitness cost of epitope i upon recognition [10]
ρ_{inv}	0.5	Proportion of invariant sites [9, 21, 22]
ψ	0.9	Fitness cost per mutation on invariant site
f	0.01 (0.001–0.1)	Proportion new viral antigen must reach to initiate immune response against corresponding epitope
d	90	Number of generations for a new immune response to reach full potency

doi:10.1371/journal.pcbi.1004625.t001

populations of productively infected and latent cells, recording the mean, median, and 5% and 95% quantiles. Furthermore, we randomly sampled 100 productively infected cells and 100 cells from the latent reservoir every 300 generations, storing the HIV-1 sequences for later phylogenetic analysis. Using $d(i, j)$ to denote the mean number of differences per position between sequences i and j , divergence and diversity are defined as follows for a collection of k sequences:

$$divergence = \frac{1}{k} \sum_{i=1}^k d_{i,founder}$$

$$diversity = \frac{\sum_{i=1}^{k-1} \sum_{j=i+1}^k d(i, j)}{\frac{1}{2} k(k-1)}$$

The effect of simulated population size on survival of latent genomic fragments

Stochastic effects dominate at small population sizes—the smaller the population, the larger the probability that stochastic fluctuations lead to extinction. We therefore investigated how the size of the simulated population affects the ability of latent genomic fragments to survive in patient plasma, assuming that recombination is necessary for survival. To this end we determined the expected number of dual infections, a precursor for recombination, involving at least one latent form when the sizes of the replicating and latent populations were increased ten-fold. We ran 100 simulations, tracking the progeny of the activated latent cells for ten generations, introduced into a replicating population of size n_r at generation zero. The relative fitness of each latent virus was set to 0.5, while the relative fitness of each non-latent virus was drawn from a uniform distribution between 0.5 and 1. As in our full model, each generation of

the simulation consisted of every infected cell producing five progeny virions, and downsampling the virus population to infect the next generation of n_r uninfected cells. The expected number of dual infections involving at least one latent virus was then estimated for each simulation run.

Patient data

Simulation results were compared to HIV-1 DNA sequence data from 16 untreated or unsuccessfully treated patients [9, 21, 22]. All patients were infected with HIV-1 subtype B. The patients were followed longitudinally from seroconversion with 2–22 sequences sampled at intervals of 1–67 months (mean 13 months). The sequences were derived from the envelope, and were between 532–948 nucleotides long (mean 683 nucleotides). For each patient, we calculated the mean divergence and diversity at each time point the same way as for the simulated data. Note that in [21], sequence divergence and diversity were estimated from mean pair-wise distances determined using either a two-parameter Kimura model or a general time-reversible model with site-to-site variation in substitution rates, while here only simple pairwise Hamming distances were used. We considered the unsuccessfully treated patients as effectively untreated, because their viremia was not suppressed, and their divergence and diversity trends were similar to those of the untreated patients. We also estimated the proportion of invariant sites across all time points for each patient using PhyML [56]. The mean proportion of invariant sites across all patients was 39%. The lowest proportion of invariant sites was 13% for Patient 9 in the Shankarappa cohort, while the mean was greater than 50% for both the Amsterdam and Karlsson patients.

Results

Simulated HIV-1 genetic trends emulate real patient data

To be sure that our model invokes realistic trends of HIV-1 genetic evolution, we compared our simulated patient HIV-1 populations generated by simulations incorporating recombination to HIV-1 DNA sequence data from 16 untreated or unsuccessfully treated patients [9, 21, 22]. Reassuringly, the empirical HIV-1 diversity and divergence trends over time were qualitatively and quantitatively similar to our simulated patient HIV-1 populations (Fig 2). Divergence increased linearly over time to a mean of 0.06 substitutions per site at 10 years post-primary HIV-1 infection (post-PHI), while diversity initially increased rapidly, and then saturated to approximately 0.05 substitutions per site at 10 years post-PHI. As seen in S1 Table in the supplement, there was no significant trend between the activation rate and mean sequence diversity ($p = 0.67$, $cor = 0.15$; Pearson's product-moment correlation). Increasing the activation rate slightly decreased sequence divergence at the end of the simulations ($p = 0.04$, $cor = -0.65$; Pearson's product-moment correlation). Unsurprisingly, both mean sequence divergence and diversity increased in the latent reservoir at a much slower initial rate than in productively infected cells (Fig 3). After approximately 6 years post-PHI, divergence in the latent reservoir proceeded at nearly the same rate as in productively infected cells. Because diversity in the latent reservoir grew linearly but started to saturate in plasma, by 10 years post-PHI diversity in the latent reservoir almost reached that in productively infected cells.

Fitness landscape influences tree shape and genetic trends

Reconstructed phylogenies of the simulated patient HIV-1 populations, with 20 sequences sampled every 12 months, showed similar topological structures as patient HIV-1 populations. While the serially sampled HIV-1 phylogenies ranged from star- to ladder-like structures, they

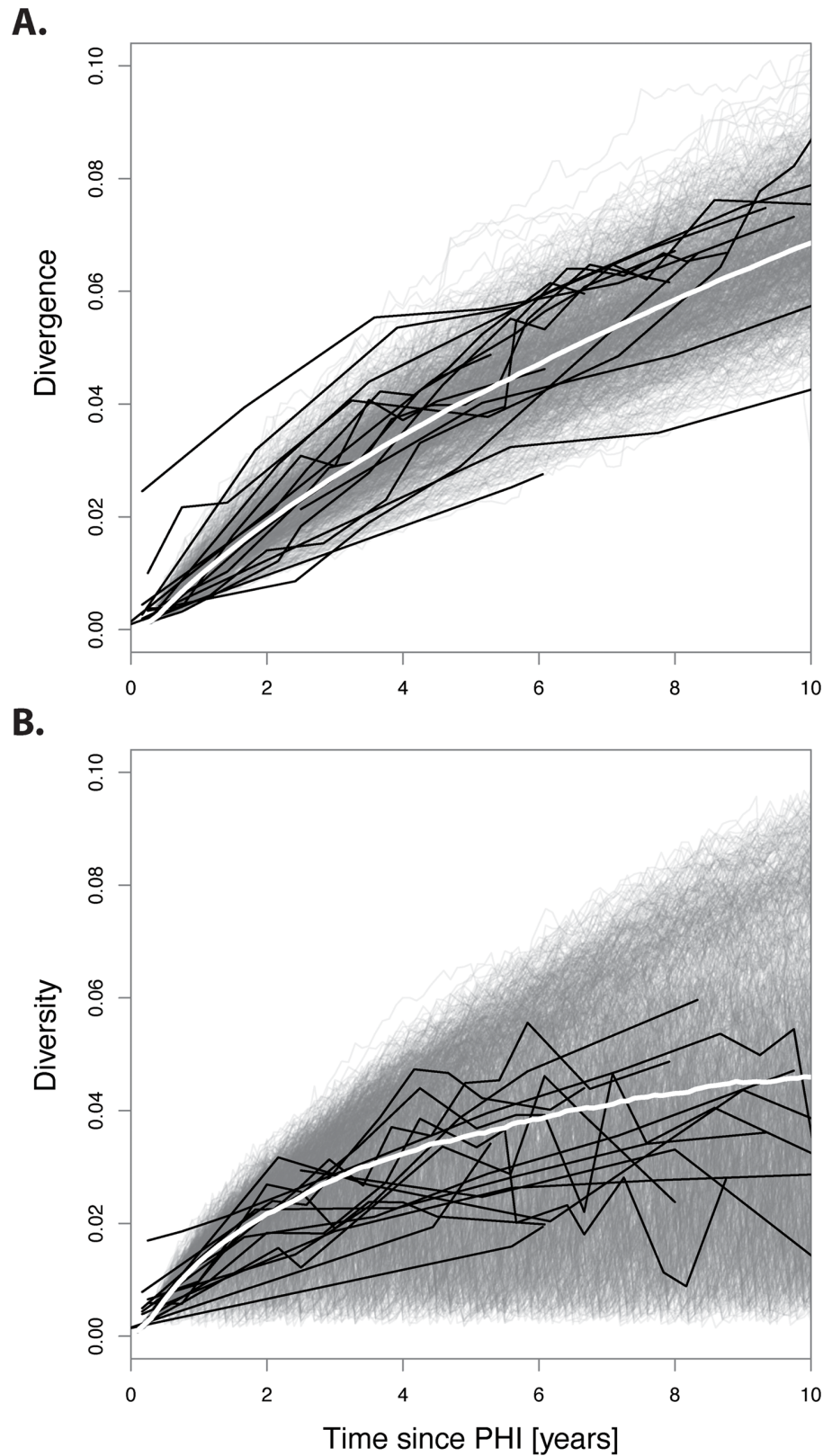


Fig 2. The divergence and diversity of simulated sequences capture evolutionary trends in clinical data. Grey lines correspond to the divergence (A) or diversity (B) of the sequences in each simulation, and

white lines show mean of simulations. Black lines show real patient data. Note that divergence and diversity were calculated every 30 generations for 15,000 sequences in each of the 1000 simulations.

doi:10.1371/journal.pcbi.1004625.g002

generally displayed a clear time trend (Fig 4). Both in the simulated and clinical data, the fraction of surviving phylogenetic lineages between samplings ranged from 0.1 to 1.0 (S1 Fig). Low survival of lineages corresponds to a ladder-like phylogeny while high survival corresponds to a star-like phylogeny.

The tree shapes were largely explained by the fitness landscape, where convex profiles lead to more ladder-like trees, and concave profiles lead to bushier, star-like, trees. Intermediate fitness landscapes (nearly linear profiles) generated trees more typically observed in HIV-1 infected patients. Importantly, these intermediate landscapes also showed typical patient diversity trends (Fig 4). Interestingly, in intermediate and concave profiles we sometimes observed dramatic selective sweeps where the diversity dropped drastically and then recovered. The fitness landscape also influenced the number of latent origins (latent virus of different ages reaching at least 1% frequency in plasma virus population). The mean (s.d.) number of different

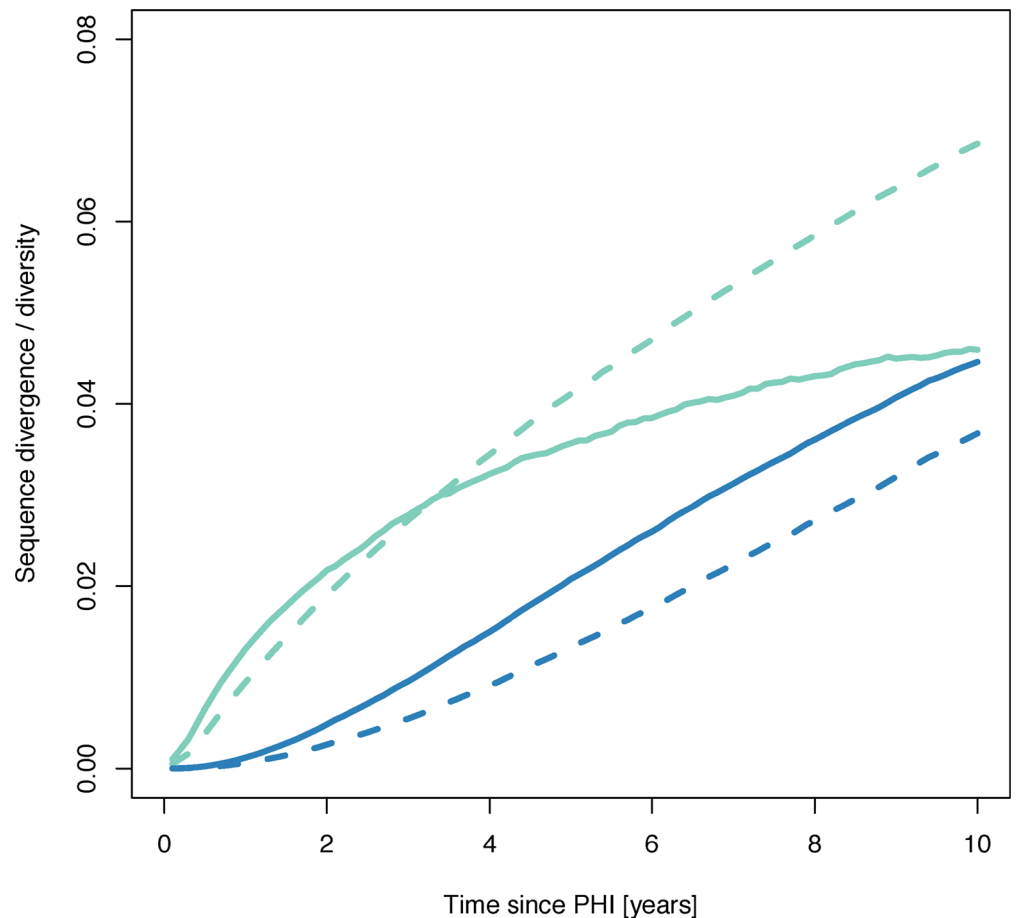


Fig 3. The divergence and diversity of simulated sequences in the latent reservoir initially increase much more slowly than in productively infected cells. After approximately 2 years post-PHI, diversity grows linearly in the latent reservoir (blue solid line) while it starts to saturate in plasma (green solid line). Divergence in the latent reservoir (blue dashed line) grows at approximately the same rate as in the plasma (green dashed line) 6 years post-PHI.

doi:10.1371/journal.pcbi.1004625.g003

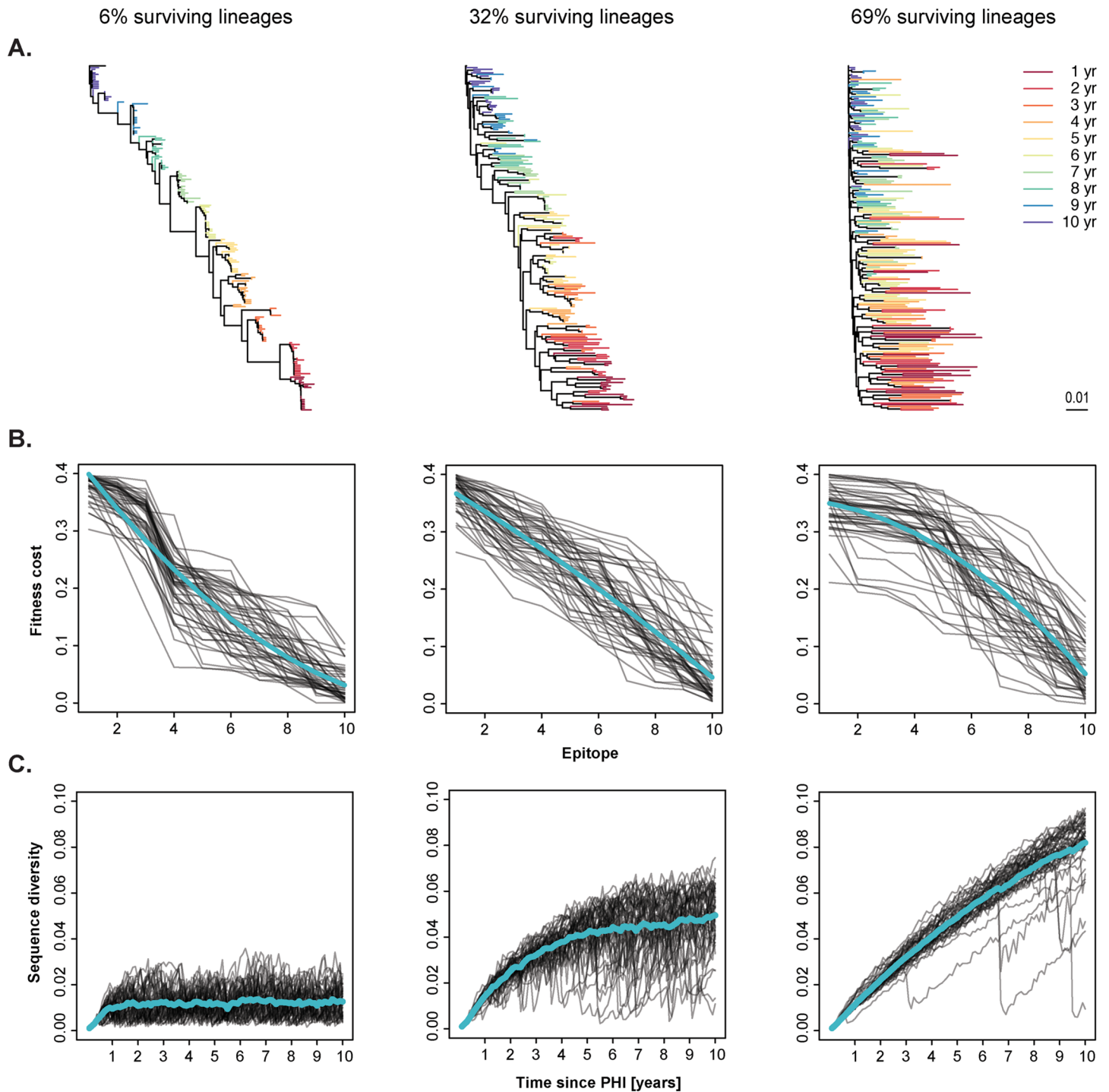


Fig 4. Tree shape and diversity are influenced by the fitness landscape. We ordered the simulations based on the average proportion of phylogenetic lineages surviving between samplings, and formed three groups consisting of 50 simulations each, with low (mean 6%), intermediate (mean 32%), and high (mean 69%) survival of lineages, respectively. (A) Typical trees of each group. Branch lengths are according to the indicated scale. Color indicates sampling time. (B) Individual fitness landscapes (grey lines) and average profile (quadratic fit, turquoise lines) of each group. (C) Individual diversity curves (grey lines) and average trends (turquoise lines).

doi:10.1371/journal.pcbi.1004625.g004

latent origins was 0.16 (0.42) for convex landscapes, 0.42 (0.76) for nearly linear landscapes, and 7.6 (7.5) for concave ones. Thus, our simulations suggest that the fitness landscape influences general tree shape, diversity trends, and the survival of latent forms.

The shape of the landscape is defined by the maximum fitness costs exerted by immune responses against different viral epitopes. The concave landscapes have a higher number of different immune responses exerting similar high fitness costs on the virus than convex or nearly linear landscapes. Escaping the immune response exerting the maximum fitness cost only marginally increases fitness if there are simultaneous immune responses against other epitopes that also impose high fitness costs. Therefore, a more concave fitness landscape exerts a stronger selection pressure on the virus, reducing both the mean and variance of the relative fitness in the virus population. The survival of less adapted latent forms in plasma therefore increases, while the reduction in fitness differences between virus variants in productively infected cells increases diversity.

Latency increases sequence diversity

[Fig 5](#) shows that the more genomic fragments with different latent origins persist in the plasma, the higher the sequence diversity (panel A). Compared to simulations with no surviving latent origins, sequence diversity became significantly higher after 9.5 years post-PHI in simulations where there were between 1 and 5 latent origins ($p < 0.05$, Wilcoxon rank-sum test), while in simulations where there were more than 5 latent origins, sequence diversity became significantly higher much earlier, after approximately 5 years post-PHI ($p < 0.05$, Wilcoxon rank-sum test). At 10 years post-PHI, mean sequence diversity was 0.038 substitutions/site when no latent genomic fragments survived, increasing by 16%, when there were between 1 and 5 latent origins, and by 95% when there were more than 15 latent origins. The propagation of genomic fragments that have accumulated fewer mutations than contemporaneous variants should reduce average sequence divergence in the plasma. As expected, when the number of surviving genomic fragments of different latent origins increased, mean sequence divergence decreased ([Fig 5](#), panel B).

Our model predicts that multiple introductions of latent forms into plasma are necessary for increased diversity; simulations with only one surviving latent origin did not have significantly higher diversity than those with no latent forms in plasma. Because sequence diversity increased almost linearly for concave fitness landscapes, the introduction of older virus forms into plasma did not further increase diversity ([S2 Fig](#)). Furthermore, because the survival of multiple latent origins was rare for convex fitness landscapes, it was not possible to determine whether latency could have had an effect on diversity under such conditions. However, the phylogenies and diversity trends generated by convex or concave fitness landscapes did not resemble those observed in typical patients.

Recombination facilitates survival of latent HIV-1 genomic fragments

[Fig 6](#) shows the effect of recombination on the proportion of virus with latent genomic fragments in productively infected cells in 2000 simulations of untreated patients. Our simulation results suggest that recombination facilitates the survival of latent genomic fragments. In simulations with recombination, 27% of the HIV-1 replicating plasma populations reached a mean of $\geq 10\%$ virus with latent genomic fragments between 5–10 years post-PHI, while only 2.0% reached $\geq 10\%$ latent forms without recombination. Thus, recombination increased the survival of latent genomic fragments in productively infected cells approximately 13-fold.

The proportion of virus with latent genomic fragments was higher when we allowed for recombination than when we ignored it. This distinction is clear from even the first sample

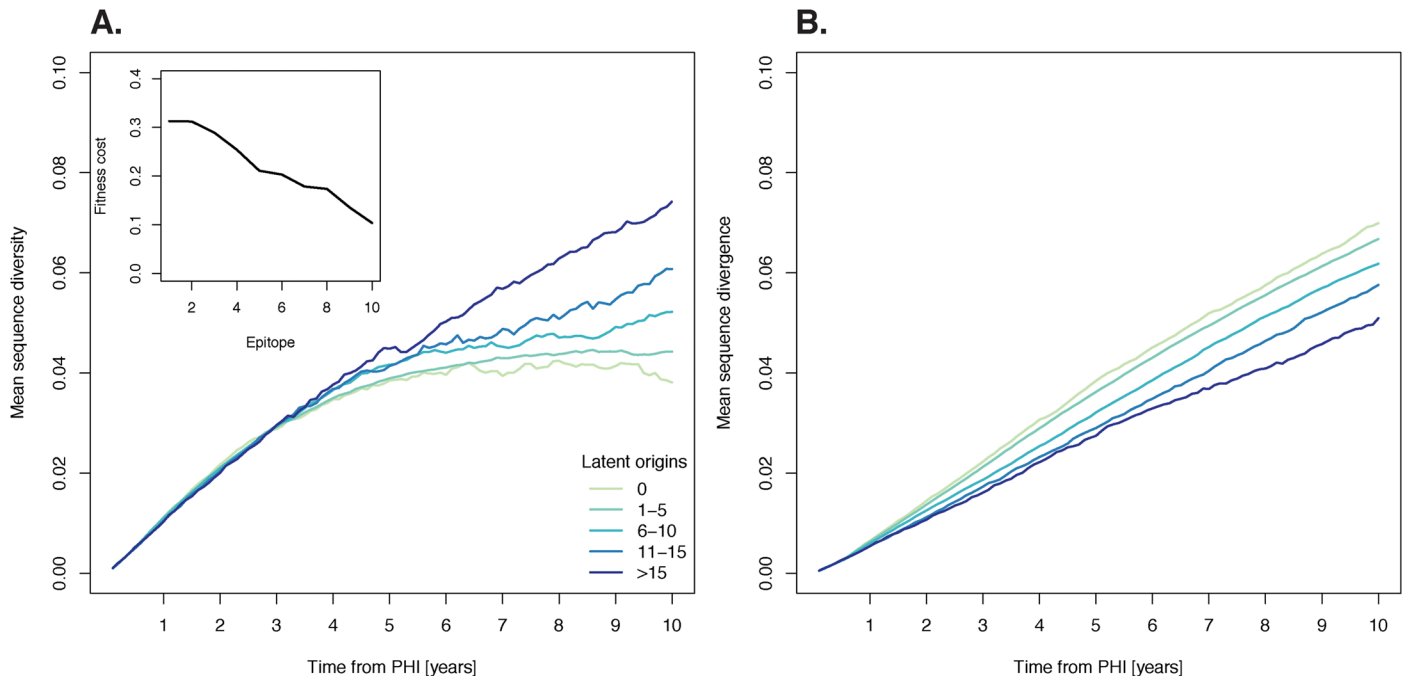


Fig 5. Survival of different latent genomic fragments increases sequence diversity. We ran 1000 simulations with the intermediate fitness landscape shown in the insert, and categorized the results based on the number of latent genomic fragments from different origins at 1% or greater frequency at 10 years post-PHI. (A) As the number of latent origins increases, so does mean sequence diversity. (B) As the number of latent origins increases, mean sequence divergence decreases.

doi:10.1371/journal.pcbi.1004625.g005

time, after 30 generations ($p < 0.05$, Wilcoxon rank-sum test). Furthermore, the rate of increase of virus with latent genomic fragments over time in the replicating population was approximately 9 times faster when recombination was involved (Fig 6). In most simulations without recombination, virus with latent fragments immediately disappeared from productively infected cells upon introduction, while in some simulations, sharp peaks were observed where the proportion of latent forms increased rapidly to a high fraction but then fell quickly and disappeared completely. In simulations with recombination, the proportion of virus with latent genomic fragments often increased gradually, and stabilized both at intermediate values and the extremes.

Latent HIV display many different recombination patterns

Most recombinant HIV-1 that include a fragment from a latent virus underwent relatively few recombination events; 96% had fewer than 5 latent fragments (Fig 7A). However, only 0.7% of sequences with latent sites at 10 years post-PHI had not acquired genomic fragments from non-latent contemporaneous virus through recombination. The most common latent recombinant observed in our simulations had one latent fragment, with a mean (s.d.) of 85 (123) sites (Fig 7A). When such a recombinant proliferates further, additional recombination with contemporaneous virus removes sites from the latent fragment. Overall, the mean (s.d.) number of latent sites at 10 years post-PHI was 99 (121), while the mean (s.d.) fragment length was 66 sites (Fig 7B).

When the number of latent fragments in a viral sequence increased, the fragment length decreased (Fig 7B). Interestingly, viruses that had more latent fragments also had more latent origins (Fig 7C). Hence, the most common forms of latent recombinants had few breakpoints,

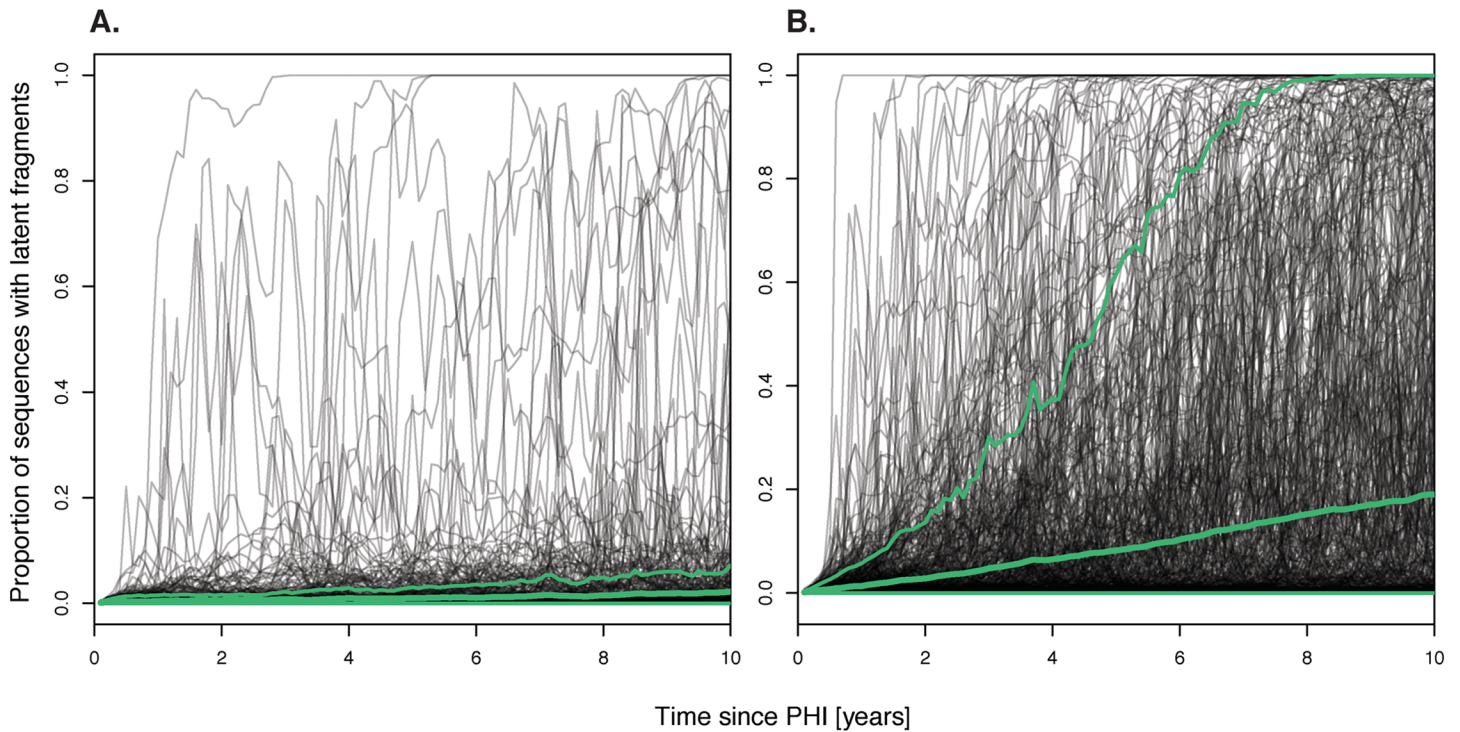


Fig 6. The effect of recombination on survival of activated latent HIV in the plasma population. A) Simulations without recombination. B) Simulations with recombination. Grey lines show the proportion of virus with latent genomic fragments in the productively infected cell population of individual simulations, where the bold green line is the mean proportion and the thin green lines outline the 95% confidence envelope. Comparing panels A and B, clearly shows that recombination facilitates survival of latent forms.

doi:10.1371/journal.pcbi.1004625.g006

and multiple breakpoints frequently involved latent virus with different origins in time. Increasing the recombination rate increased the survival of virus with latent genomic fragments (S2 Table). As expected, the number of latent fragments per sequence also increased, while the fragment length decreased.

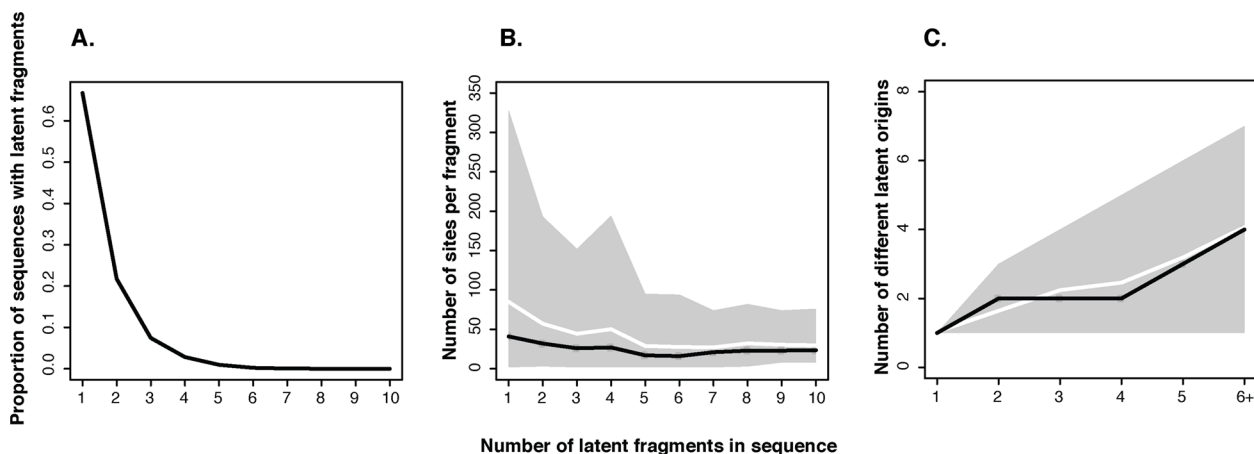


Fig 7. Latent genomic fragment patterns. (A) The proportion of sequences with latent fragments declines rapidly as the number of latent fragments per sequence increases. (B) The number of latent sites in recombinants decreases as more latent fragments are introduced. (C) As more latent fragments exist in a sequence, they have different origins in time. The grey envelopes indicate 95% of simulation results, white lines indicate the mean, and the black lines the median trends.

doi:10.1371/journal.pcbi.1004625.g007

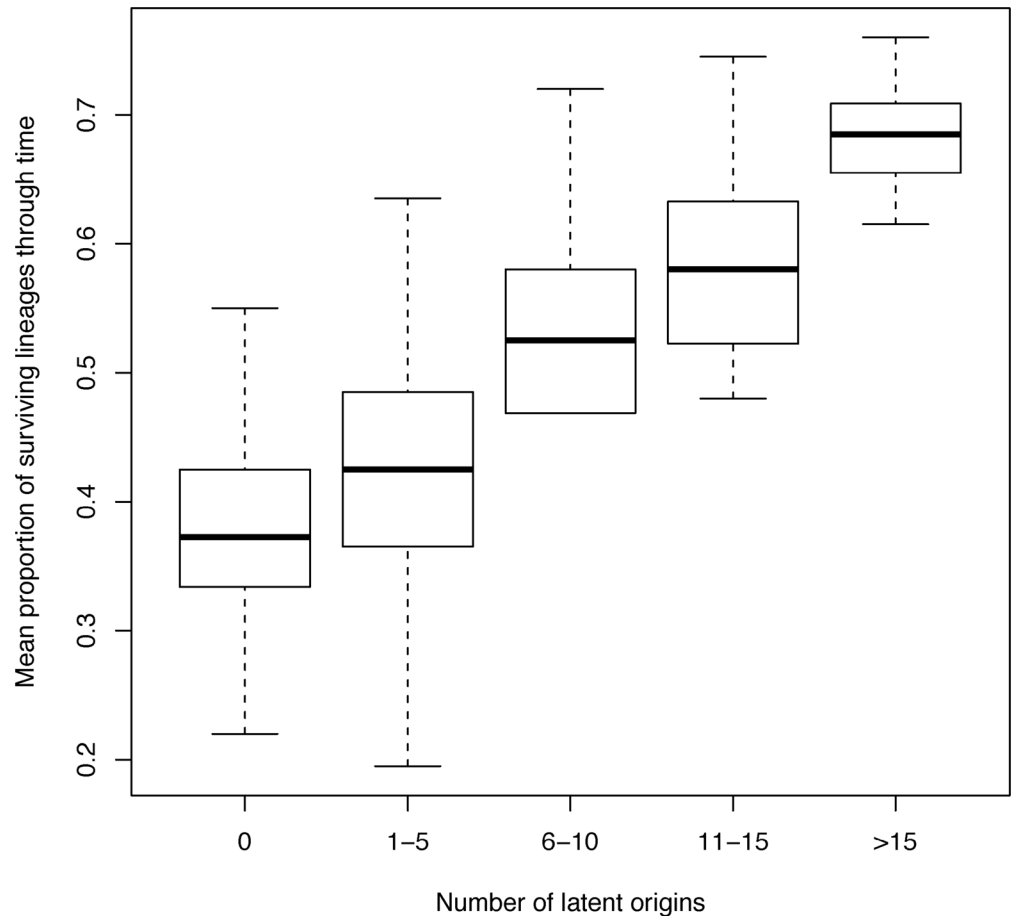


Fig 8. Survival of lineages through time increases with number of latent origins. The boxplots show the distributions of the mean proportion of lineages through time (averaged from 0 to 10 years post-PHI for the phylogeny generated from each simulation) when the number of latent genomic fragments from different origins increases. The survival of lineages increases almost linearly with the number of latent origins.

doi:10.1371/journal.pcbi.1004625.g008

Many phylogenetic lineages over time implies many different latent origins

Higher proportions of surviving lineages through time imply higher numbers of virus with genomic fragments of different latent origins that have long branches reaching far back into the phylogeny (Fig 8). When the prevalence of latent genomic fragments in productively infected cells was high, they were typically of different latency ages, i.e., deposited and activated at different times (S3 Fig). These observations are supported by the noted increase in both the proportion of virus with latent genomic fragments and the observed number of different latent origins in productively infected cells when the activation rate was increased (S4 Fig).

Recent latent forms are more common in plasma even when old latent forms are overrepresented in the reservoir

When a person becomes infected with HIV-1 the initial spike in viral load may deposit many similar copies of the initial virus in the reservoir. We modeled this by initially filling the reservoir with the infecting strain. Thereafter, a relatively slow deposit rate and proliferation rate refill and maintain a diversifying latent virus population. Interestingly, at 10 years post-PHI,

our simulations predict that about 27% of the reservoir still consists of virus deposited in the first year of infection (Fig 9A). Conversely, the latent genomic fragments in the plasma population originated mostly from more recently deposited and activated virus (Fig 9B). Overall, the age structure among latent genomic HIV-1 fragments in the plasma of an untreated patient followed that of the latent reservoir with the exception of the initially deposited virus, which is strongly selected against upon activation by the refined immune response.

Survival of latent genomic fragments under recombination is robust to perturbations of the model parameters

The evolutionary trends of the virus populations simulated under the alternative model parameters ($P = 10$, $\mu = 2.2 \times 10^{-5}$, $inv = 0.1$, $\tau = 2$ days) were qualitatively and quantitatively similar to those seen previously under the default parameters. The mean sequence divergence and diversity of 100 simulations with the alternative model parameters are shown in (S5 Fig). We found that the prevalence of virus with latent genomic fragments in productively infected cells was much higher in simulations with recombination than without (S6 Fig). Our simulation results are therefore robust to the choice of parameters defining the infection and replication dynamics.

The sensitivity of the model to the immune system parameters, i.e., the antigen frequency necessary to elicit a new immune response, and the number of epitopes accessible to the immune system, are described in (S3 and S4 Tables), respectively. Fifty simulation runs were performed for each parameter value. When the antigen frequency required for stimulating a new immune response was increased from 0.1% to 10%, the survival of virus with latent fragments and sequence diversity decreased ($p = 0.025$, $cor = -0.67$; $p = 0.013$, $cor = -0.71$, respectively; Pearson's product-moment correlation), while sequence divergence increased ($p = 0.003$, $cor = 0.79$, Pearson's product-moment correlation).

When the immune response to a new escape mutant is delayed, the frequency of virus variants carrying the advantageous escape mutation increases. The increase in the number of virus variants that have evolved away from the founder sequence increases divergence but the simultaneous convergence towards the new escape variant reduces diversity. Because the mean relative fitness of virus in productively infected cells increases, the survival probability of virus from activated latent cells decreases. The survival of virus with latent genomic fragments was low in all simulations without recombination.

On the other hand, when the number of epitopes was increased, sequence diversity at 10 years post-PHI initially decreased from one epitope to 4 epitopes and then increased from 4 to 15 epitopes (S4 Table). Sequence divergence initially increased from one to 4 epitopes, and then slightly decreased as the number of epitopes was increased. The survival of virus with latent genomic fragments was more stochastic, but generally displayed a similar trend. Virus with latent genomic fragments rarely survived in simulations without recombination. Therefore, our fundamental model result suggesting that recombination is an essential mechanism facilitating the survival of latent genomic fragments is robust to large perturbations in the largely unknown parameters describing the immune response.

When the number of epitopes was zero, no immune pressure was exerted on the virus. This resulted in high diversity (0.081 subst/site) because all virus variants without mutations at invariant sites had equal probability of surviving, and low divergence (0.047 subst/site), because rapid accumulation of escape mutations was not selected for. Virus with latent genomic fragments survived in all simulations, because earlier virus forms were not at a fitness disadvantage compared to more recent variants.

When the number of epitopes was increased from 1 to 4, it became increasingly more difficult but not impossible for virus to escape the immune response. When two epitopes had high

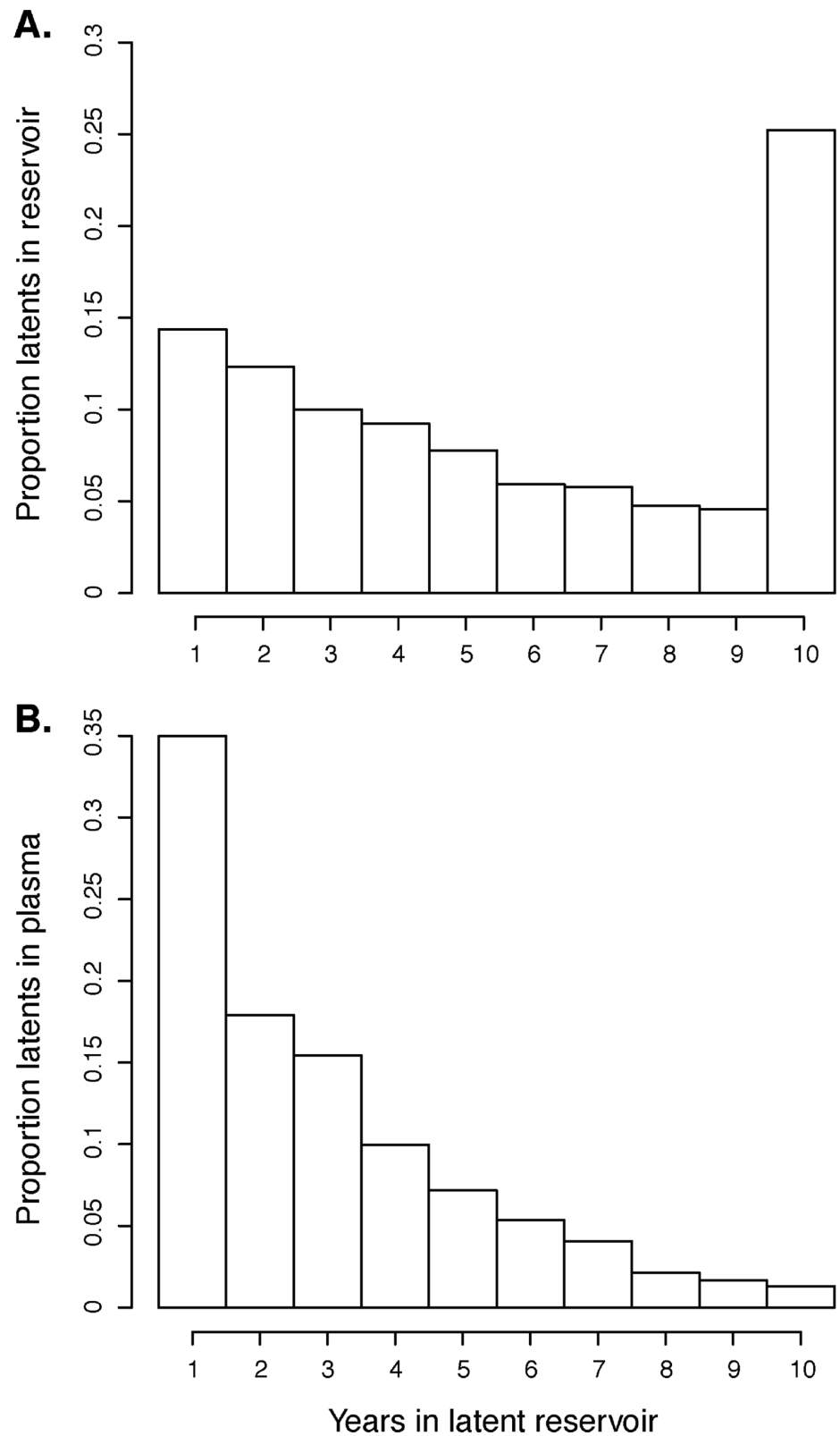


Fig 9. Distribution of latency periods in the latent reservoir and productively infected cells. (A) Distribution of time that cells have spent in the latent reservoir. (B) Distribution of the time that latent genomic fragments in plasma have spent in the latent reservoir.

doi:10.1371/journal.pcbi.1004625.g009

fitness costs associated with them, it took longer for any virus variant in the population to simultaneously escape both responses, and the significant fitness gain and proliferation due to escape resulted in decreased diversity (S7 Fig). Such genetic sweeps were more dramatic when three epitopes had high fitness costs. When four epitopes had high fitness costs, escape at all epitopes to gain fitness was nearly impossible, and happened in only a small fraction of the simulations. Therefore, more concave fitness landscapes with four or more epitopes with high fitness costs imposed selection pressure resulting in high diversity (S7 Fig). The probability that a randomly drawn fitness landscapes had 2 or 3 epitopes with high fitness costs was highest when the total number of epitopes was 4 or 5. When the number of epitopes was increased from 6 to 15, the probability that each fitness landscape had 4 or more epitopes of high fitness costs also increased.

Small sample size may underestimate the survival of latent genomic fragments

Surviving in the replicating population long enough to recombine with a replicating virus forms the bottleneck that largely determines the long-term fate of a latent lineage. When one latent form was introduced into a population of 15,000 cells at generation $t = t_0$, corresponding to the expected number of cells activated each generation from the latent reservoir of 100 cells at the highest activation rate $\alpha = 0.01$, the expected number of dual infections involving at least one latent virus was 0.1 over 10 generations. However, when 10 latent forms were introduced to a population of 150,000 cells, i.e., at the same proportion but in a ten-fold population, the total expected number of dual infections was 0.94. Thus, when the simulated population size was increased ten-fold, the probability that at least one latent form ended up in a dually infected cell before disappearing from the replicating population increased nearly ten-fold. The long-term behavior in simulations where virus with latent genomic fragments proliferates is expected to be robust to stochastic effects due to the increased number of such viruses, however, the variance observed in the survival of latent virus forms between simulations is overestimated in a smaller population due to early stochastic extinction. The proportion of simulations with persistence of virus with latent genomic fragments is therefore likely an underestimate.

Discussion

We developed a within-host HIV-1 evolution model that includes point mutation, recombination, immune (positive) selection, negative selection, and latency. This model quantitatively captures previously observed viral sequence divergence and diversity trends and mechanistically explains these patterns, allowing us to model realistic within-host natural evolution of HIV-1. We used this model to investigate when and how latent forms of HIV-1 can survive in productively infected cells despite not being well adapted to the immune response, therefore having reduced relative fitness. Our simulation results suggest that recombination is a key mechanism facilitating the survival of virus forms with latent genomic fragments. We have previously shown that the majority of phylogenetic lineages in HIV-1 populations of untreated patients display a statistically defined signal of latency [1]. The results in this paper suggest that such lineages likely survived because of recombination. Our model further predicts that the survival of latent genomic fragments in plasma depends on the fitness landscape induced by the immune response.

By comparing simulations where the surviving genomic fragments originated from different numbers of latent ancestors, we found that latency reduced mean sequence divergence, but increased mean sequence diversity for reasonable fitness landscapes. This effect may explain how HIV-1 can keep a high adaptation potential without mutating too far away from the

infecting strain. High adaptation potential is useful for escaping immune and antiviral drug pressures, while less divergent sequences closer to the infecting form are arguably more fit to infect new hosts [2].

The dynamics of our simulated immune response depend on the number of epitopes and the antigen frequency triggering a new immune response, neither of which are well known. To test the sensitivity of our model, we varied these parameters to the extremes where measures of sequence divergence and diversity became unrealistic. Importantly, the trend of much higher survival of virus with latent genomic fragments in simulations with recombination than in simulations without recombination was robust for the full range of immune system parameters. This suggests that our main result is not sensitive to the exact details of how the immune response was implemented in our model.

While *in vivo* HIV-1 viral populations are typically very large, computational considerations limited us to follow instead the relatively small population of 15,000 productively infected cells, which is in line with recent estimates of the effective population size in real HIV-1 populations [18]. The variability in the survival of virus with latent genomic fragments observed in our simulations is likely a result of stochastic fluctuations when latent virus first enters the replicating population. The proportion of simulations in which latent virus forms survive is therefore likely to be an underestimate. However, we expect the mean dynamics of virus with latent genomic fragments to be robust to scaling issues after the initial bottleneck.

Previous studies have suggested that escape mutations have a fitness cost [10], and thus virus closer to the infecting strain may have higher replicative fitness due to having accumulated fewer escape mutations to evade immune surveillance. One of the limitations of our model is that we do not consider fitness costs associated with immune escape. However, compensatory mutations may arise in other parts of the genome, and at least partially restore fitness [11]. Furthermore, mutations can hitchhike and eventually affect population fitness. We do not attempt to incorporate epistatic effects here but rather investigate the theoretical mechanisms that may allow the persistence of latent genomic fragments in the plasma population despite virus activated from latent cells having a fitness disadvantage. Assuming that latent virus does not have a replicative advantage may lead us to underestimate the survival of latent genomic fragments in the plasma virus population.

Furthermore, if activated virus from latent cells has similar fitness as contemporaneous variants, recombination may not be as important for survival. However, recombination does not merely facilitate survival, but rather allows latent genomic fragments to propagate through the plasma virus population. In the most extreme case where no selection pressure was exerted by the immune response (0 epitopes, S4 Table), approximately 60% of viral sequences contained latent genomic fragments during the last five years of infection in simulations with recombination, while only 6% of sequences were latent in simulations without recombination. While continual recombination with virus circulating in the plasma reduces the size of the genomic fragment inherited from a latent ancestor, even a few latent sites of high replicative fitness may allow the virus to better adapt to different evolutionary pressures. Latency may therefore expand the adaptive potential of HIV-1 besides enabling the virus to hide from immune surveillance and antiretroviral treatment.

Because the patient sequences we chose to study were in envelope, we focused on antibody responses and escape from them. While we did not explicitly account for CTL responses, the epitopes in our model are approximately the same length as CTL epitopes, so our model could be interpreted as one in which CTL responses instead of or in addition to antibodies are providing immune pressure. However, CTL responses exert selection pressure on the whole proteome. Since we only simulated sequence evolution in *env* and ignored the interplay between different genes, our model predictions may not be generalizable to the whole virus. Our future

directions include investigating further the balance between replicative fitness and immune escape on the level of the whole virus genome, and adapting our model to simulate HIV-1 evolution and latency under different treatment scenarios.

Supporting Information

S1 Text. The model algorithm. The pseudo-code of our computer simulations, implemented in R.

(DOCX)

S1 Fig. The proportion of surviving lineages between samples is similar in phylogenies derived from simulated sequences and from clinical data. (A) Typical phylogeny from our simulations. The proportion of surviving lineages from earlier samples is representative of the tree shape (star- to ladder-like). The legend shows samples through time and the surviving proportion of lineages in parentheses. (B) Green dots represent the proportion of surviving lineages between samples at adjacent time points in clinical data, and black dots represent our simulated data. The stronger green color indicates overlapping data points. The phylogenetic trees were generated from 20 sequences sampled every year per simulation.

(TIF)

S2 Fig. Survival of different latent genomic fragments has no effect on diversity for concave fitness maps. We ran 1000 simulations with the concave fitness landscape shown in the insert, and categorized the results based on the number of latent genomic fragments from different origins at 1% or greater frequency at 10 years post-PHI. Increasing the number of latent origins does not increase sequence diversity, which grows linearly regardless of the latent survival.

(TIF)

S3 Fig. Number of different latent origins as a function of the proportion of latent lineages in productively infected cells. Envelope indicates the span of 95% of observations, with the mean given by the white line, and the median given by the green line.

(TIF)

S4 Fig. The effect of activation rate on the persistence and composition of latent genomic fragments. A) The distribution of the number of different latent origins in productively infected cells as a function of activation rate. B) The distribution of the proportion of virus with latent genomic fragments in productively infected cells as a function of activation rate. 95% envelopes with white lines as the mean and green lines as the median.

(TIF)

S5 Fig. The mean sequence divergence and diversity of HIV-1 populations simulated under the alternative model parameters. After approximately 2 years post-PHI, diversity starts to grow linearly in the latent reservoir (blue solid line) while it starts to saturate in plasma (green solid line). Divergence in the latent reservoir (blue dashed line) grows at a slightly slower rate than in plasma (green dashed line).

(TIF)

S6 Fig. The effect of recombination on survival of latent lineages under the alternative model parameters. A) Simulations without recombination. B) Simulations with recombination. Grey lines show the proportion of latent lineages in the productively infected cell population of individual simulations, where the bold tan line is the mean proportion of latent lineages and the thin tan lines outline the 95% confidence envelope. Comparing panels A and B clearly

shows that recombination facilitates survival of latent forms.
(TIF)

S7 Fig. The number of epitopes associated with high fitness cost influences sequence diversity. We compiled the simulation results of varying the number of epitopes between 1 and 15 (S4 Table), and re-categorized them based on the number of epitopes associated with high fitness costs (defined here as fitness costs within 15% of the maximum fitness cost observed in the fitness landscape for a given simulation). Sequence diversity dramatically decreases when the number of epitopes with high fitness costs is either 2 or 3, making escape difficult but not impossible, and is the highest when the number of epitopes is 5 or greater, fully preventing escape from all simultaneous immune responses.
(TIF)

S1 Table. Simulation results when the activation rate is varied, with and without recombination. Increasing the activation rate in simulations with recombination increases the proportion of simulations with persistence ($\geq 10\%$) of virus with latent genomic fragments, and the mean proportion of virus sequences with latent genomic fragments. 100 simulations were performed for each activation rate. In simulations without recombination, latent persistence is rare.
(DOC)

S2 Table. Simulation results when the recombination rate is varied. Increasing the recombination rate increases the proportion of simulations with persistence ($\geq 10\%$) of virus with latent genomic fragments, and the mean proportion of virus sequences with latent genomic fragments. As the recombination rate increases, the mean number of latent fragments per sequence increases, but the mean size of the fragments decreases. 50 simulations were performed for each recombination rate.
(DOC)

S3 Table. Simulation results when the sensitivity of the immune response to new viral antigens is varied, with and without recombination. Increasing the antigen frequency required for eliciting a new immune response in simulations with recombination decreases the survival of virus with latent genomic fragments and sequence diversity while sequence divergence increases. In simulations without recombination, latent persistency is rare. 50 simulations were performed for each value of antigen frequency.
(DOC)

S4 Table. Simulation results when the number of epitopes recognized by the immune system is varied, with and without recombination. When the number of epitopes is increased in simulations with recombination, sequence diversity initially decreases, reaching lowest values at 3 and 4 epitopes, and increases thereafter. Sequence divergence follows the reverse trend. Virus with latent genomic fragments rarely survived in simulations without recombination. 50 simulations were performed for each number of epitopes.
(DOC)

Author Contributions

Conceived and designed the experiments: TTI JMC EORS ASP TL. Performed the experiments: TTI. Analyzed the data: TTI TL. Contributed reagents/materials/analysis tools: ASP TL. Wrote the paper: TTI JMC EORS ASP TL.

References

1. Immonen TT, Leitner T. Reduced evolutionary rates in HIV-1 reveal extensive latency periods among replicating lineages. *Retrovirology*. 2014; 11(1):81. doi: [10.1186/s12977-014-0081-0](https://doi.org/10.1186/s12977-014-0081-0) PMID: [25318357](https://pubmed.ncbi.nlm.nih.gov/25318357/)
2. Lythgoe KA, Fraser C. New insights into the evolutionary rate of HIV-1 at the within-host and epidemiological levels. *Proc. Biol. Sci.* 2012; 279(1741):3367–3375. doi: [10.1098/rspb.2012.0595](https://doi.org/10.1098/rspb.2012.0595) PMID: [22593106](https://pubmed.ncbi.nlm.nih.gov/22593106/)
3. Alizon S, Fraser C. Within-host and between-host evolutionary rates across the HIV-1 genome. *Retrovirology*. 2013; 10(1):49.
4. Vrancken B, Rambaut A, Suchard MA, Drummond A, Baele G, Derdelinckx I, et al. The genealogical population dynamics of HIV-1 in a large transmission chain: Bridging within and among host evolutionary rates. *PLoS Comput Biol*. 2014; 10(4):e1003505. doi: [10.1371/journal.pcbi.1003505](https://doi.org/10.1371/journal.pcbi.1003505) PMID: [24699231](https://pubmed.ncbi.nlm.nih.gov/24699231/)
5. Van Lint C, Bouchat S, Marcello A. HIV-1 transcription and latency: an update. *Retrovirology*. 2013; 10(1):67. doi: [10.1186/1742-4690-10-67](https://doi.org/10.1186/1742-4690-10-67) PMID: [23803414](https://pubmed.ncbi.nlm.nih.gov/23803414/)
6. Siliciano RF, Greene WC. HIV latency. *Cold Spring Harb Perspect Med*. 2011; 1(1):a007096. doi: [10.1101/cshperspect.a007096](https://doi.org/10.1101/cshperspect.a007096) PMID: [22229121](https://pubmed.ncbi.nlm.nih.gov/22229121/)
7. Richman DD, Wrin T, Little SJ, Petropoulos CJ. Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc Natl Acad Sci USA*. 2003; 100(7):4144–4149. doi: [10.1073/pnas.0630530100](https://doi.org/10.1073/pnas.0630530100) PMID: [12644702](https://pubmed.ncbi.nlm.nih.gov/12644702/)
8. Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X, et al. Antibody neutralization and escape by HIV-1. *Nature*. 2003; 422(6929):307–312. doi: [10.1038/nature01470](https://doi.org/10.1038/nature01470) PMID: [12646921](https://pubmed.ncbi.nlm.nih.gov/12646921/)
9. Bunnik EM, Pisas L, van Nuenen AC, Schuitemaker H. Autologous neutralizing humoral immunity and evolution of the viral envelope in the course of subtype B human immunodeficiency virus type 1 infection. *J Virol*. 2008; 82(16):7932–7941. doi: [10.1128/JVI.00757-08](https://doi.org/10.1128/JVI.00757-08) PMID: [18524815](https://pubmed.ncbi.nlm.nih.gov/18524815/)
10. Bar KJ, Tsao Cy, Iyer SS, Decker JM, Yang Y, Bonsignori M, et al. Early low-titer neutralizing antibodies impede HIV-1 replication and select for virus escape. *PLoS Pathog*. 2012; 8(5):e1002721. doi: [10.1371/journal.ppat.1002721](https://doi.org/10.1371/journal.ppat.1002721) PMID: [22693447](https://pubmed.ncbi.nlm.nih.gov/22693447/)
11. Song H, Pavlicek JW, Cai F, Bhattacharya T, Li H, Iyer SS, et al. Impact of immune escape mutations on HIV-1 fitness in the context of the cognate transmitted/founder genome. *Retrovirology*. 2012; 9(1):1–14. doi: [10.1186/1742-4690-9-89](https://doi.org/10.1186/1742-4690-9-89)
12. Kim H, Perelson AS. Viral and latent reservoir persistence in HIV-1-infected patients on therapy. *PLoS Comput Biol*. 2006; 2(10):e135. doi: [10.1371/journal.pcbi.0020135](https://doi.org/10.1371/journal.pcbi.0020135) PMID: [17040122](https://pubmed.ncbi.nlm.nih.gov/17040122/)
13. Rong L, Perelson AS. Modeling latently infected cell activation: viral and latent reservoir persistence, and viral blips in HIV-infected patients on potent therapy. *PLoS Comput Biol*. 2009; 5(10):e1000533. doi: [10.1371/journal.pcbi.1000533](https://doi.org/10.1371/journal.pcbi.1000533) PMID: [19834532](https://pubmed.ncbi.nlm.nih.gov/19834532/)
14. Rong L, Perelson AS. Modeling HIV persistence, the latent reservoir, and viral blips. *J Theor Biol*. 2009; 260(2):308–331. doi: [10.1016/j.jtbi.2009.06.011](https://doi.org/10.1016/j.jtbi.2009.06.011) PMID: [19539630](https://pubmed.ncbi.nlm.nih.gov/19539630/)
15. 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. *Proc Natl Acad Sci USA*. 2012; 109(24):9523–9528. doi: [10.1073/pnas.1120248109](https://doi.org/10.1073/pnas.1120248109) PMID: [22645358](https://pubmed.ncbi.nlm.nih.gov/22645358/)
16. Hill AL, Rosenbloom DI, Fu F, Nowak MA, Siliciano RF. Predicting the outcomes of treatment to eradicate the latent reservoir for HIV-1. *Proc Natl Acad Sci USA*. 2014; 111(37):13475–13480. doi: [10.1073/pnas.1406663111](https://doi.org/10.1073/pnas.1406663111) PMID: [25097264](https://pubmed.ncbi.nlm.nih.gov/25097264/)
17. Vijay N, Ajmani R, Perelson AS, Dixit NM, et al. Recombination increases human immunodeficiency virus fitness, but not necessarily diversity. *J Gen Virol*. 2008; 89(6):1467–1477. doi: [10.1099/vir.0.83668-0](https://doi.org/10.1099/vir.0.83668-0) PMID: [18474563](https://pubmed.ncbi.nlm.nih.gov/18474563/)
18. Balagam R, Singh V, Sagi AR, Dixit NM. Taking multiple infections of cells and recombination into account leads to small within-host effective-population-size estimates of HIV-1. *PloS One*. 2011; 6(1):e14531. doi: [10.1371/journal.pone.0014531](https://doi.org/10.1371/journal.pone.0014531) PMID: [21249189](https://pubmed.ncbi.nlm.nih.gov/21249189/)
19. Tripathi K, Balagam R, Vishnoi NK, Dixit NM. Stochastic simulations suggest that HIV-1 survives close to its error threshold. *PLoS Comput Biol*. 2012; 8(9):e1002684. doi: [10.1371/journal.pcbi.1002684](https://doi.org/10.1371/journal.pcbi.1002684) PMID: [23028282](https://pubmed.ncbi.nlm.nih.gov/23028282/)
20. Thangavelu PU, Gupta V, Dixit NM. Estimating the fraction of progeny virions that must incorporate APOBEC3G for suppression of productive HIV-1 infection. *Virology*. 2014; 449:224–228. doi: [10.1016/j.virol.2013.11.026](https://doi.org/10.1016/j.virol.2013.11.026) PMID: [24418556](https://pubmed.ncbi.nlm.nih.gov/24418556/)

21. Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, Farzadegan H, et al. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol.* 1999; 73(12):10489–10502. PMID: [10559367](#)
22. Karlsson AC, Lindkvist A, Lindbäck S, Gaines H, Sönnernborg A. Recent origin of human immunodeficiency virus type 1 variants in resting CD4+ T lymphocytes in untreated and suboptimally treated subjects. *J Infect Dis.* 2001; 184(11):1392–1401. doi: [10.1086/324405](#) PMID: [11709781](#)
23. Novitsky V, Wang R, Margolin L, Baca J, Rossenkhan R, Moyo S, et al. Transmission of single and multiple viral variants in primary HIV-1 subtype C infection. *PLoS One.* 2011; 6(2):e16714. doi: [10.1371/journal.pone.0016714](#) PMID: [21415914](#)
24. R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>
25. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics.* 2004; 20(2): 289–290. doi: [10.1093/bioinformatics/btg412](#) PMID: [14734327](#)
26. Josefsson L, Palmer S, Faria NR, Lemey P, Casazza J, Ambrozak D, et al. Single cell analysis of lymph node tissue from HIV-1 infected patients reveals that the majority of CD4+ T-cells contain one HIV-1 DNA molecule. *PLoS Pathog.* 2013; 9(6):e1003432. doi: [10.1371/journal.ppat.1003432](#) PMID: [23818847](#)
27. Zhuang J, Jetzt AE, Sun G, Yu H, Klarmann G, Ron Y, et al. Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. *J Virol.* 2002; 76(22):11273–11282. doi: [10.1128/JVI.76.22.11273-11282.2002](#) PMID: [12388687](#)
28. Jetzt AE, Yu H, Klarmann GJ, Ron Y, Preston BD, Dougherty JP. High rate of recombination throughout the human immunodeficiency virus type 1 genome. *J Virol.* 2000; 74(3):1234–1240. doi: [10.1128/JVI.74.3.1234-1240.2000](#) PMID: [10627533](#)
29. Leitner T, Kumar S, Albert J. Tempo and mode of nucleotide substitutions in gag and env gene fragments in human immunodeficiency virus type 1 populations with a known transmission history. *J Virol.* 1997; 71(6):4761–4770. PMID: [9151870](#)
30. Swofford D. PAUP*: phylogenetic analysis using parsimony, version 4.0 b10. 2003;.
31. Whitney JB, Hill AL, Sanisetty S, Penalzoza-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):74–77. doi: [10.1038/nature13594](#) PMID: [25042999](#)
32. Ananworanich J, Schuetz A, Vandergeeten C, Sereti I, De Souza M, Rerknimitr R, et al. Impact of multi-targeted antiretroviral treatment on gut T cell depletion and HIV reservoir seeding during acute HIV infection. *PLoS One.* 2012; 7(3):e33948. doi: [10.1371/journal.pone.0033948](#) PMID: [22479485](#)
33. 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. *Nat Med.* 2003; 9(6):727–728. doi: [10.1038/nm880](#) PMID: [12754504](#)
34. Zhang L, Ramratnam B, Tenner-Racz K, He Y, Vesanen M, Lewin S, et al. Quantifying residual HIV-1 replication in patients receiving combination antiretroviral therapy. *N Engl J Med.* 1999; 340(21):1605–1613. doi: [10.1056/NEJM199905273402101](#) PMID: [10341272](#)
35. 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. *Nat Med.* 1999; 5(5):512–517. doi: [10.1038/8394](#) PMID: [10229227](#)
36. Crooks AM, Bateson R, Cope AB, Dahl NP, Griggs MK, Kuruc JD, et al. Precise quantification of the latent HIV-1 reservoir: implications for eradication strategies. *J Infect Dis.* 2015; jiv218.
37. 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. *Nat Med.* 2009; 15(8):893–900. doi: [10.1038/nm.1972](#) PMID: [19543283](#)
38. Llano A, Williams A, Olvera A, Silva-Arrieta S, Brander C. Best-characterized HIV-1 CTL epitopes: The 2013 Update. *HIV Molecular Immunology* 2013. 2013;p. 3–25.
39. Cao J, McNeven J, Malhotra U, McElrath MJ. Evolution of CD8+ T cell immunity and viral escape following acute HIV-1 infection. *J Immunology.* 2003; 171(7):3837–3846. doi: [10.4049/jimmunol.171.7.3837](#)
40. Goulder PJ, Phillips RE, Colbert RA, McAdam S, Ogg G, Giangrande P, et al. Late escape from an immunodominant cytotoxic T lymphocyte response associated with progression to AIDS. *Immunol Lett.* 1997; 56:25. doi: [10.1016/S0165-2478\(97\)85097-3](#)
41. Ganusov VV, Goonetilleke N, Liu MK, Ferrari G, Shaw GM, McMichael AJ, et al. Fitness costs and diversity of the cytotoxic T lymphocyte (CTL) response determine the rate of CTL escape during acute and chronic phases of HIV infection. *J Virol.* 2011; 85(20):10518–10528. doi: [10.1128/JVI.00655-11](#) PMID: [21835793](#)

42. Chen HY, Di Mascio M, Perelson AS, Ho DD, Zhang L. Determination of virus burst size in vivo using a single-cycle SIV in rhesus macaques. *Proc Natl Acad Sci USA*. 2007; 104(48):19079–19084. doi: [10.1073/pnas.0707449104](https://doi.org/10.1073/pnas.0707449104) PMID: [18025463](https://pubmed.ncbi.nlm.nih.gov/18025463/)
43. Kwon YJ, Hung G, Anderson WF, Peng CA, Yu H. Determination of infectious retrovirus concentration from colony-forming assay with quantitative analysis. *J Virol*. 2003; 77(10):5712–5720. doi: [10.1128/JVI.77.10.5712-5720.2003](https://doi.org/10.1128/JVI.77.10.5712-5720.2003) PMID: [12719564](https://pubmed.ncbi.nlm.nih.gov/12719564/)
44. Marozsan AJ, Fraundorf E, Abraha A, Baird H, Moore D, Troyer R, et al. Relationships between infectious titer, capsid protein levels, and reverse transcriptase activities of diverse human immunodeficiency virus type 1 isolates. *J Virol*. 2004; 78(20):11130–11141. doi: [10.1128/JVI.78.20.11130-11141.2004](https://doi.org/10.1128/JVI.78.20.11130-11141.2004) PMID: [15452233](https://pubmed.ncbi.nlm.nih.gov/15452233/)
45. Rusert P, Fischer M, Joos B, Leemann C, Kuster H, Flepp M, et al. Quantification of infectious HIV-1 plasma viral load using a boosted in vitro infection protocol. *Virology*. 2004; 326(1):113–129. doi: [10.1016/j.virol.2004.05.022](https://doi.org/10.1016/j.virol.2004.05.022) PMID: [15262500](https://pubmed.ncbi.nlm.nih.gov/15262500/)
46. Mansky LM. Forward mutation rate of human immunodeficiency virus Ttpe 1 in a T lymphoid cell line*. *AIDS Res Hum Retroviruses*. 1996; 12(4):307–314. doi: [10.1089/aid.1996.12.307](https://doi.org/10.1089/aid.1996.12.307) PMID: [8906991](https://pubmed.ncbi.nlm.nih.gov/8906991/)
47. Haase AT, Henry K, Zupancic M, Sedgewick G, Faust RA, Melroe H, et al. Quantitative image analysis of HIV-1 infection in lymphoid tissue. *Science*. 1996; 274(5289):985–989. doi: [10.1126/science.274.5289.985](https://doi.org/10.1126/science.274.5289.985) PMID: [8875941](https://pubmed.ncbi.nlm.nih.gov/8875941/)
48. Chun TW, Carruth L, Finzi D, Shen X, DiGiuseppe JA, Taylor H, et al. Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature*. 1997; 387(6629):183–188. doi: [10.1038/387183a0](https://doi.org/10.1038/387183a0) PMID: [9144289](https://pubmed.ncbi.nlm.nih.gov/9144289/)
49. Stafford MA, Corey L, Cao Y, Daar ES, Ho DD, Perelson AS. Modeling plasma virus concentration during primary HIV infection. *J Theor Biol*. 2000; 203(3):285–301. doi: [10.1006/jtbi.2000.1076](https://doi.org/10.1006/jtbi.2000.1076) PMID: [10716909](https://pubmed.ncbi.nlm.nih.gov/10716909/)
50. Vaidya NK, Rong L, Marconi VC, Kuritzkes DR, Deeks SG, Perelson AS. Treatment-mediated alterations in HIV fitness preserve CD4+ T cell counts but have minimal effects on viral load. *PLoS Comput Biol*. 2010; 6(11):e1001012. doi: [10.1371/journal.pcbi.1001012](https://doi.org/10.1371/journal.pcbi.1001012) PMID: [21124866](https://pubmed.ncbi.nlm.nih.gov/21124866/)
51. Yeh WW, Rahman I, Hraber P, Coffey RT, Nevidomskyye D, Giri A, et al. Autologous neutralizing antibodies to the transmitted/founder viruses emerge late after simian immunodeficiency virus SIVmac251 infection of rhesus monkeys. *J Virol*. 2010; 84(12):6018–6032. doi: [10.1128/JVI.02741-09](https://doi.org/10.1128/JVI.02741-09) PMID: [20357097](https://pubmed.ncbi.nlm.nih.gov/20357097/)
52. Ribeiro RM, Qin L, Chavez LL, Li D, Self SG, Perelson AS. Estimation of the initial viral growth rate and basic reproductive number during acute HIV-1 infection. *J Virol*. 2010; 84(12):6096–6102. doi: [10.1128/JVI.00127-10](https://doi.org/10.1128/JVI.00127-10) PMID: [20357090](https://pubmed.ncbi.nlm.nih.gov/20357090/)
53. Pearson JE, Krapivsky P, Perelson AS. Stochastic theory of early viral infection: continuous versus burst production of virions. *PLoS Comput Biol*. 2011; 7(2):e1001058. doi: [10.1371/journal.pcbi.1001058](https://doi.org/10.1371/journal.pcbi.1001058) PMID: [21304934](https://pubmed.ncbi.nlm.nih.gov/21304934/)
54. Markowitz M, Louie M, Hurley A, Sun E, Di Mascio M, Perelson A, et al. A novel antiviral intervention results in more accurate assessment of human immunodeficiency virus type 1 replication dynamics and T-cell decay in vivo. *J Virol*. 2003; 77(8):5037–5038. doi: [10.1128/JVI.77.8.5037-5038.2003](https://doi.org/10.1128/JVI.77.8.5037-5038.2003) PMID: [12663814](https://pubmed.ncbi.nlm.nih.gov/12663814/)
55. 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. *Proc Natl Acad Sci USA*. 2008; 105(21):7552–7557. doi: [10.1073/pnas.0802203105](https://doi.org/10.1073/pnas.0802203105) PMID: [18490657](https://pubmed.ncbi.nlm.nih.gov/18490657/)
56. 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. *Systematic biology*. 2010; 59(3):307–321. doi: [10.1093/sysbio/syq010](https://doi.org/10.1093/sysbio/syq010) PMID: [20525638](https://pubmed.ncbi.nlm.nih.gov/20525638/)