

5-19-2016


HIV-1 Genetic Variation Resulting in the Development of New Quasispecies Continues to Be Encountered in the Peripheral Blood of Well-Suppressed Patients.

Will Dampier
Drexel University College of Medicine

Michael R Nonnemacher
Drexel University College of Medicine

Joshua Mell
Drexel University College of Medicine

Joshua Earl
Drexel University College of Medicine
Follow this and additional works at: <https://jdc.jefferson.edu/kimmelccfp>

 Garth D Ehrlich
Part of the Immune System Diseases Commons, and the Medical Immunology Commons
Drexel University College of Medicine

[Let us know how access to this document benefits you](#)

See next page for additional authors
Recommended Citation

Dampier, Will; Nonnemacher, Michael R; Mell, Joshua; Earl, Joshua; Ehrlich, Garth D; Pirrone, Vanessa; Aiamkitsumrit, Benjamas; Zhong, Wen; Kercher, Katherine; Passic, Shendra; Williams, Jean W; Jacobson, Jeffrey M; and Wigdahl, Brian, "HIV-1 Genetic Variation Resulting in the Development of New Quasispecies Continues to Be Encountered in the Peripheral Blood of Well-Suppressed Patients." (2016). *Kimmel Cancer Center Faculty Papers*. Paper 64.
<https://jdc.jefferson.edu/kimmelccfp/64>

This Article is brought to you for free and open access by the Jefferson Digital Commons. The Jefferson Digital Commons is a service of Thomas Jefferson University's [Center for Teaching and Learning \(CTL\)](#). The Commons is a showcase for Jefferson books and journals, peer-reviewed scholarly publications, unique historical collections from the University archives, and teaching tools. The Jefferson Digital Commons allows researchers and interested readers anywhere in the world to learn about and keep up to date with Jefferson scholarship. This article has been accepted for inclusion in Kimmel Cancer Center Faculty Papers by an authorized administrator of the Jefferson Digital Commons. For more information, please contact: JeffersonDigitalCommons@jefferson.edu.

Authors

Will Dampier, Michael R Nonnemacher, Joshua Mell, Joshua Earl, Garth D Ehrlich, Vanessa Pirrone, Benjamas Aiamkitsumrit, Wen Zhong, Katherine Kercher, Shendra Passic, Jean W Williams, Jeffrey M Jacobson, and Brian Wigdahl

RESEARCH ARTICLE

HIV-1 Genetic Variation Resulting in the Development of New Quasispecies Continues to Be Encountered in the Peripheral Blood of Well-Suppressed Patients

Will Dampier^{1,2,3}, Michael R. Nonnemacher^{1,2}, Joshua Mell^{1,4}, Joshua Earl^{1,4}, Garth D. Ehrlich^{1,4,5}, Vanessa Pirrone^{1,2}, Benjamas Aiamkitsumrit^{1,2}, Wen Zhong^{1,2}, Katherine Kercher^{1,2}, Shendra Passic^{1,2}, Jean W. Williams^{1,2}, Jeffrey M. Jacobson^{1,6,7}, Brian Wigdahl^{1,2,8*}



CrossMark
click for updates

1 Department of Microbiology and Immunology, Drexel University College of Medicine, Philadelphia, Pennsylvania, United States of America, **2** Center for Molecular Virology and Translational Neuroscience, Institute for Molecular Medicine and Infectious Disease, Drexel University College of Medicine, Philadelphia, Pennsylvania, United States of America, **3** School of Biomedical Engineering, Science, and Health Systems, Drexel University, Philadelphia, Pennsylvania, United States of America, **4** Center for Genomic Sciences, Institute for Molecular Medicine and Infectious Disease, Drexel University College of Medicine, Philadelphia, Pennsylvania, United States of America, **5** Center for Advanced Microbial Processing, Institute for Molecular Medicine and Infectious Disease, Drexel University College of Medicine, Philadelphia, Pennsylvania, United States of America, **6** Department of Medicine, Division of Infectious Diseases and HIV Medicine, Drexel University College of Medicine, Philadelphia, Pennsylvania, United States of America, **7** Center for Clinical and Translational Medicine, Institute for Molecular Medicine and Infectious Disease, Drexel University College of Medicine, Philadelphia, Pennsylvania, United States of America, **8** Sidney Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, PA, United States of America

* bwigdahl@drexelmed.edu

OPEN ACCESS

Citation: Dampier W, Nonnemacher MR, Mell J, Earl J, Ehrlich GD, Pirrone V, et al. (2016) HIV-1 Genetic Variation Resulting in the Development of New Quasispecies Continues to Be Encountered in the Peripheral Blood of Well-Suppressed Patients. PLoS ONE 11(5): e0155382. doi:10.1371/journal.pone.0155382

Editor: Fatah Kashanchi, George Mason University, UNITED STATES

Received: February 11, 2016

Accepted: April 27, 2016

Published: May 19, 2016

Copyright: © 2016 Dampier et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript except for the sequence files which have been submitted to Genbank under Bioproject ID SUB1322660.

Funding: These studies were funded in part by the Public Health Service, National Institutes of Health through grants from the National Institute of Neurological Disorders and Stroke, NS32092 and NS46263, the National Institute of Drug Abuse, DA19807 (Dr. Brian Wigdahl, Principal Investigator), NS089435 (Dr. Michael Nonnemacher, Principal

Abstract

As a result of antiretroviral therapeutic strategies, human immunodeficiency virus type 1 (HIV-1) infection has become a long-term clinically manageable chronic disease for many infected individuals. However, despite this progress in therapeutic control, including undetectable viral loads and CD4⁺ T-cell counts in the normal range, viral mutations continue to accumulate in the peripheral blood compartment over time, indicating either low level reactivation and/or replication. Using patients from the Drexel Medicine CNS AIDS Research and Eradication Study (CARES) Cohort, whom have been sampled longitudinally for more than 7 years, genetic change was modeled against to the dominant integrated proviral quasispecies with respect to selection pressures such as therapeutic interventions, AIDS defining illnesses, and other factors. Phylogenetic methods based on the sequences of the LTR and *tat* exon 1 of the HIV-1 proviral DNA quasispecies were used to obtain an estimate of an average mutation rate of 5.3 nucleotides (nt)/kilobasepair (kb)/year (yr) prior to initiation of antiretroviral therapy (ART). Following ART the baseline mutation rate was reduced to an average of 1.02 nt/kb/yr. The post-ART baseline rate of genetic change, however, appears to be unique for each patient. These studies represent our initial steps in quantifying rates of genetic change among HIV-1 quasispecies using longitudinally sampled sequences from patients at different stages of disease both before and after initiation of combination ART.

Investigator), National Institute of Mental Health, P30 MH092177, and under the Ruth L. Kirschstein National Research Service Award T32 MH079785. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

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

Notably, while long-term ART reduced the estimated mutation rates in the vast majority of patients studied, there was still measurable HIV-1 mutation even in patients with no detectable virus by standard quantitative assays. Determining the factors that affect HIV-1 mutation rates in the peripheral blood may lead to elucidation of the mechanisms associated with changes in HIV-1 disease severity.

Introduction

The introduction of combination therapeutics have extended the life expectancy of those infected with the human immunodeficiency virus type 1 (HIV-1) for many years past what would have been observed even a decade ago. Due to the mutable nature of the HIV-1 genome, longer life expectancy and associated selective pressures has led to increased opportunities for the emergence of viral genetic variants that escape therapy [1, 2], contribute to neurocognitive decline [3–5], or result in more or less frequent activation/reactivation of persistent/latent infection [6], among many other potential complications resulting from a large number of well-known comorbidities [7]. As such, it is important to understand the nature of viral genetic variation within individual hosts both prior to and after the initiation of combination antiretroviral therapy (cART) and perhaps just as importantly, among the various combination therapeutic regimens.

Over the past several years, a great deal of attention has focused on the fate of the infecting viral inoculum and the cells that are initially targeted in the genital tract, regional lymph nodes, and peripheral blood during the early acute stages of infection. The general conclusions from studies of heterosexual transmission have indicated that infection is typically acquired from a single viral genotype (80% of cases), and the transmitted virion typically utilizes the CCR5 coreceptor (designated an R5 virus), which replicates well in CD4+ T-cell cultures but not monocyte-derived macrophage cultures [8–11]. Although intra-host genetic bottlenecks have been identified by a number of mechanisms and at different anatomic sites following the initial encounter, these results are consistent with a small founder population and host tropisms that have previously been observed in the simian immunodeficiency virus (SIV) rhesus macaque model [12, 13].

Viral genetic variation over time in a patient is likely the result of at least three processes. The first process involves the introduction of novel mutations between the entry and integration steps of the viral life cycle, due to errors in the proof-reading mechanism of HIV-1 reverse transcriptase [14]. These errors have been estimated to occur at an average rate of 0.1 mutations per genome per cell generation as measured by single-round integration experiments [14]. Proofreading errors results in an average mutation rate of ~2 nucleotides per kilobase per year per host, as measured in a large-scale phylogenetic analysis of the Los Alamos database [15]. The high mutation rates of reverse transcription indicate that 1 in 10 cellular infection events results in a novel, replication competent, viral genome being integrated into the host cell. This results in “clouds” of similar genomes that collectively span a large portion of the replication competent sequence space [16], each of which is referred to as a “quasispecies”. Through natural selection, viral genomes harboring beneficial variants will increase in frequency. However, individual patients present a complex and variable fitness landscape, and thus HIV’s high mutation rate allows it to maintain a high degree of diversity to evade immune pressure, antiretroviral therapy, and other host-specific selective pressures [17].

The second major modulator of viral genetic variation within a host centers on how variant sequences are affected by selective pressures resulting from immune responses, therapeutic interventions, drug abuse, and other environmental factors, which all affect the frequency of preexisting viral variants. In most cases, this leads to a predominant sequence comprising between 20% to 80% of the total number of viral quasispecies sequences along with potentially dozens of lower frequency variant sequences [18]. Viral quasispecies that contain new mutations or preexisting low frequency viral quasispecies that acquire drug resistance alleles or immune escape variants can increase in frequency to predominate when the selective pressures change over time [1, 19, 20]. Due to changes in patient lifestyles (e.g. lapses in therapeutic compliance, changing substance abuse patterns, or general health), selective pressure can also change, shifting the frequency of individual viral genetic variants within individual HIV-1-infected patients.

The third general source of viral genetic variation in the peripheral blood is based on the periodic escape or release of HIV-1 quasispecies from cellular and anatomical reservoirs harboring virus. Subsequent or continued replication of such previously compartmentalized viruses after transition to the peripheral blood allows for the seemingly hidden evolution of new quasispecies in the peripheral blood. During the acute infectious process when the virus achieves very high viral loads it seeds several anatomic reservoirs including the brain [21], lung [22], gut, as well as other locations including the bone marrow [23]. Integration into the host cell's chromosome of the HIV-1 genome as a provirus allows it to remain latent in resting memory CD4+ T cells where it can evade most selective pressures of the immune system [24, 25]. New viral quasispecies and their cognate viral particles can emerge from these reservoirs through either activation of the latent memory CD4+ T-cell population, or from resident cell populations within the anatomic reservoirs through unknown mechanisms. Understanding the clinical factors that influence activation and viral release from more therapeutically refractile reservoirs will be crucial to the success of the next generation of therapeutics such as shock-and-kill and related strategies [26], therapeutic vaccine development [27], and proviral excision techniques that hold great promise to eradicate HIV infection [28].

The Drexel Medicine CNS AIDS Research and Eradication Study (CARES) Cohort in Philadelphia, PA is a unique resource to examine the interplay between these three biologic processes that impact the course of viral genetic variation in individual infected patients. More than 550 HIV-1-infected patients within the CARES Cohort have been longitudinally sampled with return visits scheduled approximately every six months. PCR-based Sanger sequencing has been utilized to sample the viral sequence from multiple genomic regions of integrated HIV-1 provirus populations, and this has provided for a detailed inspection of the sequential variation of the dominant proviral quasispecies across many HIV-1-infected patients [29–32].

Modern computational methods are able to integrate multiple time-stamped sequences into a phylogenetic reconstruction of current and past divergence histories. Bayesian Estimation of Ancestors by Sampling Trees (BEAST) is a well-established tool used to estimate the divergence time of historic sequences, as well as lineage-specific evolutionary rates using a relaxed molecular clock assumption [33]. It has also been used to estimate HIV-1 transmission networks in Kenya [34] and the emergence of novel HIV-1 variants in China [35]. Josefsson and coworkers examined the HIV-1 mutation rates in the resting memory CD4+ T-cell and gut reservoirs of several patients that were well-suppressed on long-term ART and found a rate of <3 mutations per gene over a 4–12 year time-period, depending on the gene sampled and time between measurements [36].

In this study, the LTR and *tat* exon 1 regions of 36 patients in the CARES Cohort were sampled over a six-year period and BEAST phylogenetic analyses were performed to determine how the mutation rates changed over time, and how clinical factors may affect mutation rates.

Materials and Methods

Patients enrolled in the Drexel Medicine CARES Cohort were recruited from the Partnership Comprehensive Care Practice of the Division of Infectious Disease and HIV Medicine in the Department of Medicine at Drexel University College of Medicine (Philadelphia, Pennsylvania, USA) and the Center for Clinical and Translational Medicine in the Drexel Institute for Molecular Medicine and Infectious Disease. Patients from this Cohort encompass a broad range of clinical severities; however, in general, the CARES Cohort consists primarily of African American males that have been infected for a median of 15.9 [15.6–16.7] years and who have been on therapy for a median of 13.5 [12.9–14.3] years (Table 1). For the archival analysis presented herein, 36 patients were selected from the CARES Cohort using the following inclusion criteria with each patient having: (1) eight or more visits, (2) at least one visit naïve to therapy, and (3) at least 2 years on ART. Patients were excluded if they switched ART therapy while in the Cohort or if they have ever admitted to IV drug use (rationale discussed below).

Whole blood was collected from each of the 36 patients, processed, and assessed as previously described [29]. This procedure was performed at the initial visit and at each subsequent return visit. Peripheral blood samples were used for drug screening, plasma analysis, and PBMC isolation as previously described [32].

The LTR and *tat* exon 1 regions were amplified from genomic DNA isolated from PBMCs with three different methods: (1) The “PCR Read” method which consisted of direct Sanger sequencing of a PCR product generated from genomic DNA; approximately 90% of the patient samples have a PCR Read. (2) The “Clone Read” consisted of subcloning to isolate individual fragments from PCR products with subsequent sequencing of the clones; 31 patients had at least 10 clones and an average of 16 clones per sample (Table 1, column 2). (3) The “4.4 kb Fragment Read” consisted of amplifying a 4.4 kb fragment from the 3’-end of the viral genome followed by isolation of individual fragments by subcloning and Sanger sequencing; approximately 36% of patients have a “4.4 kb Fragment Read”. Greater detail on the PCR procedure has been described [29, 32]. The approximately 2,300 individual sequences used in this analysis were deposited in Genbank under Bioproject ID PRJNA318511.

Phylogenetic reconstruction

Phylogenetic reconstruction of the evolutionary history of the HIV-1 sequences derived from the peripheral blood of each patient was performed using version 1.8.1 of the BEAST package [33]. This method uses a Bayesian Markov-Chain Monte-Carlo estimation technique to find the most likely tree given a set of sequences along with the provision of sampling dates and assumptions about the population size, root height, nucleotide distribution, as well other parameters as defined below. The assumptions behind the BEAST model are critical to fitting reasonable estimates to the data. The model was fit with a Bayesian skyline assumption [37] in which the effective population size was allowed to vary over 10 equally spaced points between the root and tip of the tree. The effective sample size was estimated as an exponentially distributed variable with a mean calculated from the viral loads of each patient. The root height was estimated from medical records of HIV infection and modeled as a truncated normally distributed variable with a standard deviation of 5 years and limited to their age on the left end and their first HIV-related medical event on the right. This assumption implies a sexual transmission route, and as such, patients who admitted to IV drug use or who had likely been infected by transfusion based on clinical history were excluded from the analysis because of a higher probability of having experienced a blood stream transmission event rather than transmission by sexual intercourse. The average mutation rate was modeled as a normally distributed variable, truncated at 0, with a mean of 2 mutations/kb/year with a standard deviation of +/- 1,

Table 1. Demographics of the entire CARES Cohort as well as the subset of patients that have been included in the CARES longitudinal arm and those that were used for confirmation by subcloning and sequencing. The values indicate the median and 95% confidence intervals. *Due to the longitudinal and archival nature of this dataset, the limit of detection with respect to viral load has changed but has ranged from <100 to <20.

Category	Note	Entire Cohort	Clone Sampling	Longitudinal Sampling
Patient number		506	31	36
Age		49 [47.8–49.1]	44 ± [41.8–46.0]	51.5 [49.8–54.2]
Gender	Male	62%	62%	59%
	Female	37%	37%	41%
CD4 (cells/mL)	Latest	522 [536–585]	494 [447–592]	675 [645–830]
	Nadir	223 [249–281]	199 [179–249]	212 [179–267]
Viral Load (copies/uL)	Latest	48 [48–18,200]	235 [48–442,000]	Undetectable*
	Peak	45,500 [151,000–313,000]	90,860 [122,000–786,000]	75,012 [84,700–221,000]
Years Seropositive		15.9 [15.6–16.7]	10.9 [10.5–14.2]	19.1 [18.2–21.8]
Years on ART		14.1 [12.9–14.3]	5.6 [3.22–724]	14.7 [12.5–16.8]

doi:10.1371/journal.pone.0155382.t001

based on previous experimental estimates [15]. Nucleotide substitution models were selected independently for each alignment using the Bayesian information selection criteria as implemented in the jModelTest 2 package [38]. All other parameters were left in the default mode.

Sequences were first aligned with the MUSCLE alignment tool using default parameters [39]. Simulations were run for 1,000,000 iterations with samplings every 1,000 chains. The maximum likelihood phylogenetic trees were reconstructed using the TreeAnnotator tool packaged with BEAST to estimate the tree topology and 95% confidence intervals on divergence times and mutation rates. The multiple alignments were also processed using the RAxML tool version 8.1.17 for bootstrapping [40]. Only splits that occurred in 80% of the trees were maintained. The maximum likelihood tree was compared to the tree generated by BEAST and those with less than 90% agreement in the complete phylogenetic topology were excluded from further analysis.

Statistical Analysis

Trees were processed with the Dendropy [41] python package to measure mutation rate estimates for each branch. The mutation rates across all branches at a particular time point were aggregated by averaging the estimates across each branch that existed at that time point taking into account the confidence intervals provided by the BEAST analysis. Due to the longitudinal nature of this study, accounting for non-independence between measurements was non-trivial. Mixed effects models were used to account for the fixed effects of the measured variables in the presence of an infinite array of patient-specific random effects. All statistical analyses were performed with the MixedML package of the Statsmodels python toolset [<http://statsmodels.sourceforge.net/>].

Results

Use of PCR to directly amplify members of the HIV-1 quasispecies from genomic DNA

In an effort to simplify the process of longitudinal sampling of HIV-1 quasispecies in the Drexel Medicine CARES Cohort, PCR amplification of specific regions of the HIV-1 genome was performed on genomic DNA and the resulting PCR products were directly sequenced without the traditional step of sub-cloning to isolate individual sequence fragments prior to sequencing the cloned product. To validate this direct PCR-based sequencing protocol, it was

compared to the longer and more expensive sequencing strategy, in which PCR products were cloned, and several individual clones were sequenced. Towards this end, a set of 31 patient samples were subjected to the traditional sub-cloning PCR strategy [29, 32]. From these 31 samples, 507 total clones were obtained with an average of 16.3 Clone Reads per sample. By counting the number of unique sequences it was estimated that there were 2.32 QS per sample (Fig 1A) each representing 4% to 100% (mean 43%) of the observed QS (Fig 1B).

To further determine whether the PCR read generated a single member of the quasispecies or an amalgamation of the entire quasispecies, the PCR Read was compared to all Clone Reads. In all 31 samples (each with at least 10 Clone Reads) the PCR Read exactly matched at least one of the Clone Reads. Also of particular interest is whether the PCR Read matched the most predominant unique Clone Read. In 23/31 cases the PCR Read matched the predominant Clone Read (Fig 1B). While the PCR Read does not always provide a measurement of the Predominant Clone Read, it does provide a member of the quasispecies; as such it is possible to use these sequences for subsequent phylogenetic reconstructions.

Longitudinal analysis of mutation rate

Using a subset of patients from the CARES Cohort, phylogenetic reconstruction using the BEAST analysis tool was performed to estimate the change in evolutionary rate over time. In brief, this process involved fitting a date-tipped maximum likelihood tree, in which the date at the root of the tree was fit to a probability model using the patient-reported infection date as a prior. A relaxed molecular clock assumption allowed for a variable mutation rate across each branch. Trees were generated for the LTR and *tat* exon 1. Due to the current sequencing progress in the Drexel Medicine CARES Cohort and the exclusion criteria discussed in Materials and Methods, this resulted in the construction of 84 trees for the LTR, 12 trees for *tat* exon 1, and 6 trees for *tat* exon 2. BEAST trees that disagreed with standard phylogenetic trees

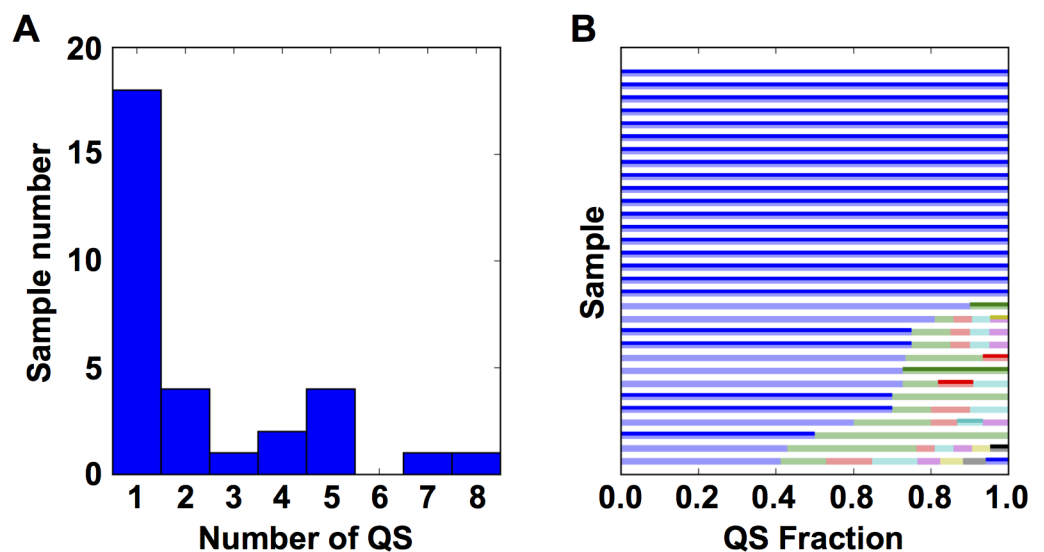


Fig 1. Direct sequencing of PCR products generated directly from genomic DNA very often returns the predominant genotype. (A) (Left) A histogram of the number of QS observed in each of the 31 samples with at least 10 Clone Reads. (B) A bar plot showing the fraction of the total observed QS for each unique Clone Read. Each horizontal bar represents one of the 31 samples with the width of the bar indicating the QS fraction. Each color in the bar indicates unique QS. Darker bars denote the QS that match the PCR Read.

doi:10.1371/journal.pone.0155382.g001

generated using RAxML were then excluded. This resulted in a high confidence set of 36 LTR trees and 8 *tat* exon 1 trees (Table 1, column 3).

In order to estimate the time-varying rate of HIV-1 mutation in these patients, a summary method was developed to obtain a mutation rate. Using the date-tipped BEAST trees, the mutation rate was averaged across all existing branches at a particular time-point, taking into account the Bayesian credible intervals provided by the BEAST algorithm. The BEAST-constructed trees for the LTR and *tat* exon 1 regions of three representative patients are shown (Fig 2). Aggregated estimates of the mutation rate over time are shown for the same three patients (Fig 2).

The BEAST trees shown in Fig 2 are representative of all the patients and illustrate several known features of HIV-1 clinical disease progression. These particular three patients were chosen as those with the most sequences of both LTR and *tat* exon 1. All six trees show a major split within the first few years of infection. The *tat* exon 1 tree of patient A0095 shows sequential mutation in the predominant variant as evidenced by the branching pattern of the bottom arm of the tree. Importantly, in all six trees there is a marked decrease in the mutation rate

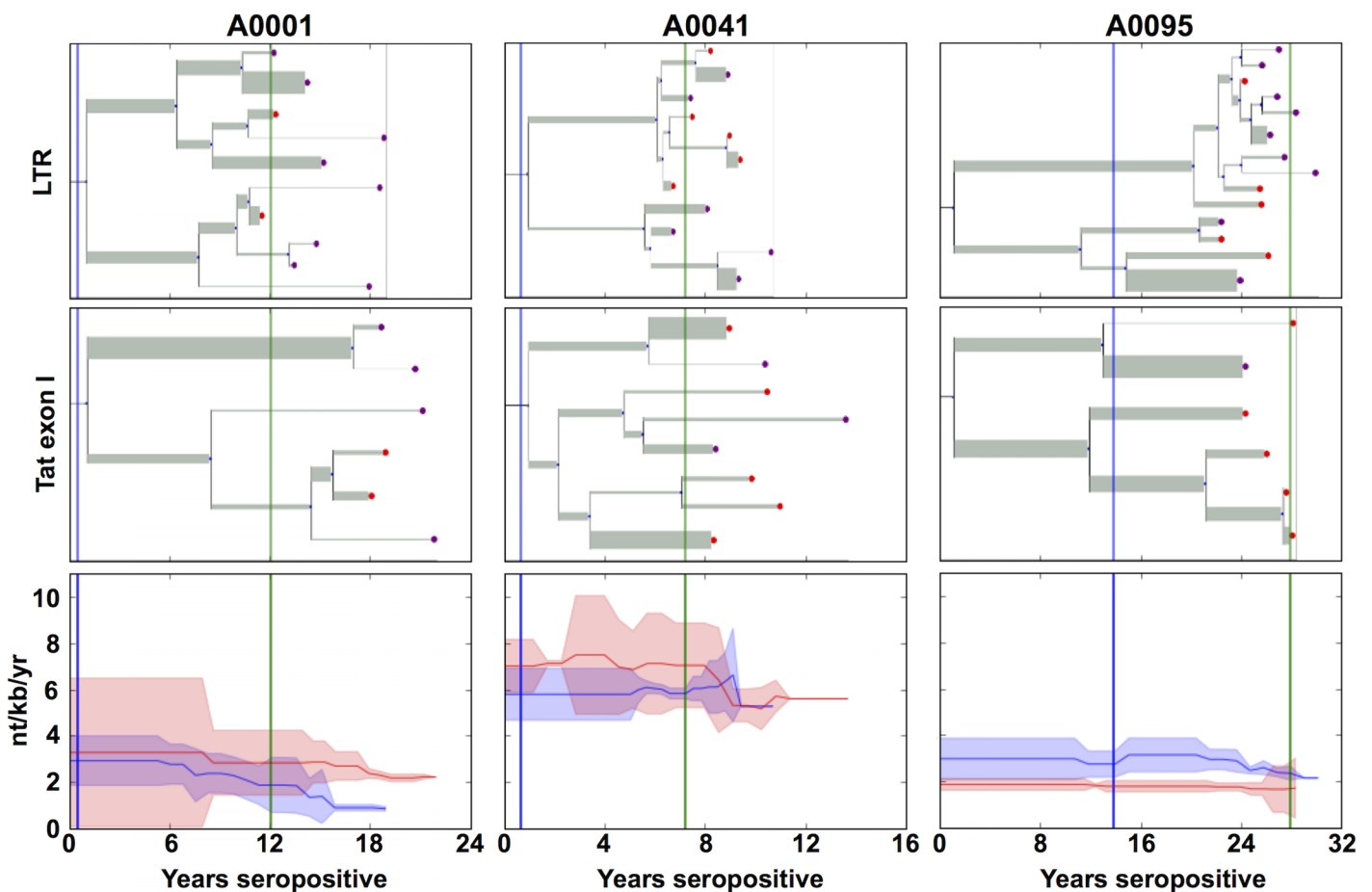


Fig 2. The mutation rate of HIV varies over time as determined by BEAST analysis. BEAST trees from three representative patients (A0001, A0041 and A0095) are shown for both the LTR (Top) and *tat* exon 1 (Middle) regions. The width of the branch indicates the rate of mutation across the branch and the branch-length represents the time since divergence. The purple nodes indicate PCR Reads, the red nodes 4.4 Kb Fragment Reads, and the green nodes indicate Clone Reads. (Bottom) The time-series of each patient was synchronized such that the estimated time of infection for all samples have been shifted to 0 years. The red line shows the average mutation rate of *tat* exon 1 and the red shadow shows the 95% confidence interval of the estimate. The blue line shows the average mutation rate of the LTR and the blue shadow shows the 95% confidence interval of the estimate.

doi:10.1371/journal.pone.0155382.g002

once ART therapy has lowered viral loads to undetectable levels (green line), as indicated by narrower branches and a lower aggregated mutation rate.

Interestingly, [Fig 2](#) also shows instances of sequences emerging from long-armed phylogenetic branches. These imply that either a variant has emerged from an unsampled reservoir, the reactivation of previously latent genomes, or a member of the quasispecies moving from low frequency to a higher frequency. The last two could be resolved by an increase in the number of samples with multiple reads while the first requires sampling all potential reservoirs. The complications surrounding these topics are reviewed in greater detail in the discussion.

To quantify the effect of ART therapy on the HIV-1 mutation rate, 36 patients were further analyzed. This subset of patients was selected such that more than one sample had been collected from the patient prior to the initiation of combination ART (PreART) and at least one sample had been collected, in which they had been on combination ART with undetectable viral loads. The samples from each patient were subdivided into the two groups. Specifically; the PreART Group refers to samples collected before the introduction of ART. The Suppressed ART Group refers to samples collected after the virus has been reduced to undetectable levels until therapy was changed or the latest sampled time point. As expected, the 36 patients have spent different lengths of time on combination ART and have been well-suppressed for different lengths of time. In order to synchronize the time series across multiple patients, the time points were synchronized using clinical markers. The PreART Group time points were measured from the reported HIV infection date ([Fig 3A](#)) while the Suppressed ART Group time points were measured from the first well-suppressed viral load measurement. In order to estimate the effect of ART, the mean mutation rate for each patient was calculated for the PreART Group and Suppressed ART Group and plotted as normalized histograms ([Fig 3B](#)). The mean mutation rates were calculated as 5.3 +/- 4.7, and 3.1 +/- 2.4 nt/kb/year for PreART and controlled ART, respectively. This suggested that, while viral control by combination ART is correlated with a decrease in mutation rate, the magnitude of this decrease was small.

To directly test the hypothesis that combination ART has a significant effect on the mutation rate, a mixed effects statistical model was built such that the mutation rate was due to three fixed effects: years seropositive, ART status, and genetic region sampled (LTR or *tat* exon 1). The random effects were assumed to be an independent intercept unique to each patient. This assumes that the virus in each patient has a unique basal mutation rate that is an aggregate of all unmeasured variables affecting viral mutation in that patient, thus accounting for repeated measure issues inherent with longitudinal data [42]. This model was compared to a null model, in which the only fixed effect was years seropositive. When compared to the null model, the alternative model has a likelihood ratio of 16.2 with an extra 2 degrees of freedom; which results in a chi-squared p-value = 7.38E-5. The estimated effect sizes of the parameters in this model are shown in [Table 2](#). Based on this model, HIV-1 has a basal mutation rate of 5.21 mutations/kb/year and combination ART decreases this by 0.97 mutations/kb/year.

Numerous other clinical parameters contained within the Drexel Medicine CARES Cohort database were examined with respect to their correlation with time-varying mutation rate but no associations with a statistically significant effect were identified. Peak viral load and nadir CD4⁺ T-cell count showed no improvement in model fit ($p > 0.1$). Age showed a mild correlation with mutation rate ($p = 0.074$); however, this may be due to its covariance with Years Seropositive; more data is needed to resolve this difference. Current viral loads (those within the detectable range), CD4⁺ T-cell, and CD8⁺ T-cell counts also did not correlate with the mutation rate ($p > 0.1$).

The results, taken as a whole, indicate that the mutation rate was lowered by ART therapy in this dataset but that the overall effect is small and mutations remain ongoing even in the absence of detectable viral load for prolonged periods of time. Using this population of 36

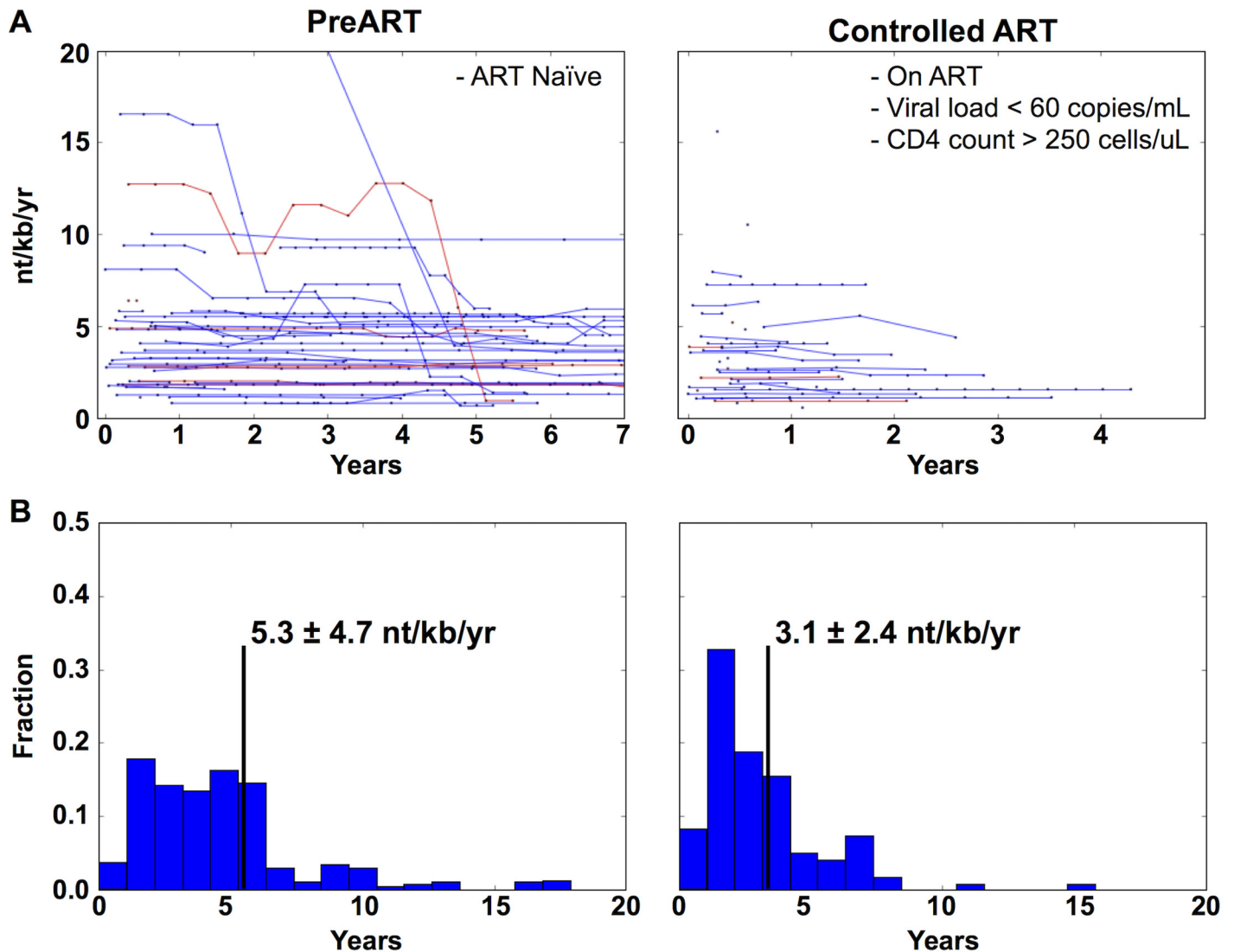


Fig 3. The mutation rate of HIV-1 is lowered by the introduction of ART but not to undetectable levels. (A) The time series of mutation rates for all patients (N = 36) that fall into the PreART Group and Suppressed ART Group. Red lines represent the mutation rate of *tat* exon 1, while blue lines represent the LTR mutation rate. In order to synchronize each patient, their time points were aligned such that all patients in the PreART Group had their t = 0 set to their estimated time of infection. Patients in the Suppressed ART Group were synchronized such that t = 0 was set to the first clinical visit in which their viral loads dropped to undetectable levels and continued until either they switched therapy, lost viral control, or the current time was reached. (B) A normalized histogram of the mean mutation rate for each patient in either the PreART or Suppressed ART Group categories is shown. The black vertical line denotes the average rate.

doi:10.1371/journal.pone.0155382.g003

Table 2. The calculated effect sizes of fixed effects for the patients shown in Fig 3.

Parameter	nt/kb/yr	95% CI	p-value
Baseline (Intercept)	5.21	[4.17, 6.25]	9.68E-23
Patient Baseline (random effect)	0.0012	[-0.85, 0.85]	N/A
<i>Tat</i> exon 1 versus LTR	0.636	[0.067, 1.20]	2.85E-02
Suppressed ART	-0.972	[-1.35, -0.592]	5.47E-07
Years Seropositive	-0.064	[-0.088, -0.041]	6.31E-08

doi:10.1371/journal.pone.0155382.t002

patients sampled longitudinally for six years it is estimated that the effect of therapy on the mutation rate of HIV represents a 20% decrease. Notably, despite undetectable levels of viral loads the mutation rate has never been estimated to be zero. These results show that, even in the absence of detectable viral load, new genetic variation continues to arise in the proviral DNA population. Importantly, the analysis performed treated distinct lineages, or quasispecies, of provirus individually since lineage-specific mutation rates were estimated, therefore reactivation of latent reservoirs and patients infected by multiple distinct genotypes are not confounding the results.

Discussion and Conclusions

We have hypothesized that the genetic variation over time within the integrated HIV-1 proviral genome in the infected cells derived from the peripheral blood represents (1) a combination of sequential mutations within quasispecies, (2) changes in the relative frequencies of distinct quasispecies, and (3) the reemergence of compartmentalized genomes leaking from viral reservoirs including the brain, gastrointestinal tract, and perhaps other tissues that may include the bone marrow. Longitudinal analysis of HIV-1 proviral sequences derived from patients in the Drexel Medicine CARES Cohort is beginning to quantify the relative contribution of each of these forces. In the present analysis, we have demonstrated that it is possible to construct date-tipped phylogenies and estimate changes in mutation rates in response to clinical factors/events.

Of the three potential sources of variation, this study focused on lineage-specific mutation rates and how these changed over time to ART. The sampling of Clone Reads performed on the 31 samples was insufficient to capture low frequency variants, usually defined as those below 1% [1, 20]. This makes it difficult to distinguish between the resurgence of a low-frequency variant and the reemergence of a historical compartmentalized differentially dormant or latent provirus. Future research will focus on disentangling the effect of all three modes of variation during the course of HIV-1 disease. This will be accomplished through the examination of similar longitudinal samples but utilizing next generation sequencing platforms including the Illumina NextSeq and Pacific Bioscience (PacBio) platforms to explore low frequency variants and their change in relative abundance over time.

Examining the release of virus from a “leaky” viral reservoir will require additional experimental approaches in addition to a greater depth of sequence coverage. For example, fluorescence-activated cell sorting can be used to separate cells to specifically examine the CD4 + memory T-cell compartment, a likely cellular reservoir [24, 25]. Examining a number of anatomical reservoirs using cerebrospinal fluid [21], GI tract biopsies [23], and bronchiolar lavage [22] sampling have been utilized to gain an overall picture of the corresponding viral reservoir. A dataset that can combine deep sequencing from multiple cellular and anatomical reservoirs will be an invaluable tool with respect to defining the exact nature of the longitudinal genetic change or variation observed in the peripheral blood and the critical cellular reservoirs during the course of HIV-1 disease.

The variable rate of sequence evolution between the differing regions of the HIV-1 genome is something that has been explored previously [31, 32]. This study confirms that *tat* exon 1 has a higher observed mutation rate (0.636 nt/kb/yr) when compared to the LTR (0.527 nt/kb/yr). The LTR and *tat* differ in that *tat* is expressed as a protein where the LTR is not. Because of this Tat is processed as an antigen which is recognized by the immune system. This leads to an additional selective pressure on Tat with respect to immune evasion [43] that does not exist for the LTR. Future studies will examine other HIV-1 genes in this patient population. We hypothesize that changes in mutation rate with the introduction of therapy will be more pronounced

in genes encoding proteins that are directly targeted, such as the reverse transcriptase, protease, and integrase.

The mutation rates calculated above conform to previous estimates of HIV-1 mutation rates as observed by Josefsson and colleagues [36]. They found that the mutation rate of HIV-1 in untreated individuals to be 5–7 nt/kb/yr, which compares favorably to our estimate of 4–6 nt/kb/yr. However, the estimates of mutation rates in patients well-suppressed on combination ART are not directly comparable between this study and their study. Josefsson and colleagues examined the gag-pol region of the genome from the memory T-cell compartment. As such, they estimated a mutation rate of 0.01–0.15 nt/kb/yr, 50 times lower than the 3.5–4.6 nt/kb/yr that our study estimates. Their study provides an estimate of the mutation rate in the latent/persistent resting memory CD4⁺ T-cell population as opposed to the total cellular population in the PBMC compartment of well-suppressed HIV-1-infected patients examined in our investigations; both of which are important cellular populations to understand within the context of HIV-1 disease. Furthermore, their study specifically targeted the polymerase gene and the selective pressure imposed by ART confounds direct comparison between the two studies with respect to comparing mutation rates.

Attempts to account for these data using other clinical variables (including CD4⁺/CD8⁺ cell counts, current and peak viral loads, and age) did not have a significant impact on the result. The viral load was weakly correlated with the mutation rate but only when using a threshold at the level of detection and combined with ART status. The PreART Group was examined with respect to direct correlations between viral load and mutation rate but with no success. This could be due to numerous factors that include the fact that (1) increased viral loads may occur from a clonal expansion of a small number of infected cells leading to the development of a population of identical viral genomes leading to less overall variation [44, 45]; (2) the spike in viral load may be due to replication- or entry-incompetent virions [46]; (3) there was insufficient resolution at current time-points to measure the branch that spawned the spike in viral load; and (4) there was a general lack of statistical power when examining this relatively small number of patients. Future studies focusing on PreART time points may find additional quantitative correlations with respect to mutation rates in selected regions of the HIV-1 genome.

After initiation of ART, the mutation rate converges to an apparently patient-specific rate that stays constant over numerous years, after therapy has reduced viral load to undetectable levels. This rate was estimated to be 3.69 +/- 2.88 mutations/kb/year. It is currently unclear what parameters affect this measure, but it is likely that a combination of the specific ART regimen of each patient, patient adherence to therapy, and viral factors such as the initial viral set point all play a role. In this dataset there is no correlation between the baseline rate of genetic change and typical clinical markers such as Nadir CD4, peak viral load, or years on therapy ($p > 0.1$). It was not possible to test the effect of specific ART regimens due to the breadth of the therapeutic approaches used in this relatively small cohort. As such, a larger dataset will be required to associate the patient-specific baseline rate of genetic change with clinical parameters such as therapeutic protocol, viral set point, as well as others.

This study is limited by the assumptions inherent in both the phylogenetic framework as well as the linear modeling. The BEAST analysis assumes that the tree is rooted at a specific time and that all observed variation results from mutations to that common founder virus. While this is likely to be true in most cases [10, 11], individuals in this cohort may have been infected numerous times. It is currently unclear how to find these instances beyond the trivial cases of superinfection with multiple HIV-1 subtypes. Additional effort needs to be expended to answer these questions to increase the likelihood of resolving these issues.

Assumptions surrounding the linear modeling may also lead to potential misinterpretations of the results. Both ART suppression and increased length of infection are correlated with a

decreased mutation rate. These results may lead one to hypothesize that after enough time on therapy the mutation rate would eventually reach zero. This mirrors previous hypotheses that sufficient therapeutic control would eventually lead to viral clearance [47, 48]. However, this is not what was observed; instead the rate of genetic change appears to plateau to a patient-specific baseline mutation rate. Instead, the results presented showed that reservoirs may persist despite the most effective therapeutic control available. The results shown also confirm that the viral quasispecies in the PBMC compartment can continue to change and/or mutate with only an average of a 20% reduction of the PreART mutation rate.

With the improvement of therapeutic interventions, HIV-1 patients are living longer. As such it is important to understand the factors that affect the viral mutation rate including (1) relative frequencies abundances of quasispecies, (2) reemergence of historical sequences, and (3) the nature of the reservoirs that harbor defective and fully functional proviruses as this will be critical to treating patients in the long-term. The results reported herein indicate that even under complete therapeutic elimination of detectable viral replication, mutations within the integrated proviral DNA still occur, albeit at a lower rate. It is likely that this is a combination of active replication in areas of poor therapeutic penetrance, such as the brain, gut, or other tissues, and undetectable levels of replication in the PBMC compartment. Deep sequencing experiments on longitudinally collected samples will allow researchers to sample all quasispecies at each time-point along with relative frequencies. When combined with rich clinical information, the factors influencing each mode of variation will be examined in even greater detail.

Ethics statement

The Drexel University College of Medicine Institutional Review Board (IRB) has approved this work under protocol 1201000748 (Brian Wigdahl, PI). All patient samples were collected under the auspices of protocol 1201000748 through written consent.

Acknowledgments

We would like to thank all patients who are part of the Drexel Medicine CNS AIDS Research and Eradication Study (CARES) Cohort. We would also like to thank the clinical staff within the Division of Infectious Diseases and HIV Medicine at the Drexel University College of Medicine who were involved in the recruitment, enrollment, obtaining consent, obtaining clinical history, venipuncture, and delivery of peripheral blood to the research laboratories in the Center for Molecular Virology and Translational Neuroscience in the Institute for Molecular Medicine and Infectious Disease.

Author Contributions

Conceived and designed the experiments: WD MRN VP BW. Performed the experiments: BA WZ KK SP JW. Analyzed the data: WD JM. Contributed reagents/materials/analysis tools: WD MRN BW. Wrote the paper: WD MRN JM JE GE VP JMJ BW.

References

1. Pennings PS. HIV Drug Resistance: Problems and Perspectives. *Infectious disease reports*. 2013; 5 (Suppl 1):e5. doi: [10.4081/idr.2013.s1.e5](https://doi.org/10.4081/idr.2013.s1.e5) PMID: [24470969](https://pubmed.ncbi.nlm.nih.gov/24470969/); PubMed Central PMCID: [PMC3892620](https://pubmed.ncbi.nlm.nih.gov/PMC3892620/).
2. Quashie PK, Mesplede T, Wainberg MA. Evolution of HIV integrase resistance mutations. *Curr Opin Infect Dis*. 2013; 26(1):43–9. doi: [10.1097/QCO.0b013e32835ba81c](https://doi.org/10.1097/QCO.0b013e32835ba81c) PMID: [23242340](https://pubmed.ncbi.nlm.nih.gov/23242340/).
3. Samikkannu T, Atluri VS, Arias AY, Rao KV, Mulet CT, Jayant RD, et al. HIV-1 subtypes B and C Tat differentially impact synaptic plasticity expression and implicates HIV-associated neurocognitive disorders. *Curr HIV Res*. 2014; 12(6):397–405. PMID: [25613138](https://pubmed.ncbi.nlm.nih.gov/25613138/).

4. Fields J, Dumaop W, Elueteri S, Campos S, Serger E, Trejo M, et al. HIV-1 Tat alters neuronal autophagy by modulating autophagosome fusion to the lysosome: implications for HIV-associated neurocognitive disorders. *J Neurosci*. 2015; 35(5):1921–38. doi: [10.1523/JNEUROSCI.3207-14.2015](https://doi.org/10.1523/JNEUROSCI.3207-14.2015) PMID: [25653352](https://pubmed.ncbi.nlm.nih.gov/25653352/); PubMed Central PMCID: PMC4315828.
5. Krebs FC, Ross H, McAllister J, Wigdahl B. HIV-associated neurologic dysfunction. In HIV: Molecular mechanisms and Clinical Applications. In: Jeang KT, editor. *Advances in Pharmacology*. 492000. p. 315–85.
6. Wilhelm E, Bell B. Selective recognition of viral promoters by host cell transcription complexes: challenges and opportunities to control latency. *Current opinion in virology*. 2013; 3(4):380–6. doi: [10.1016/j.coviro.2013.06.006](https://doi.org/10.1016/j.coviro.2013.06.006) PMID: [23827503](https://pubmed.ncbi.nlm.nih.gov/23827503/).
7. Tedaldi EM, Minniti NL, Fischer T. HIV-associated neurocognitive disorders: the relationship of HIV infection with physical and social comorbidities. *BioMed research international*. 2015; 2015:641913. doi: [10.1155/2015/641913](https://doi.org/10.1155/2015/641913) PMID: [25815329](https://pubmed.ncbi.nlm.nih.gov/25815329/); PubMed Central PMCID: PMC4359826.
8. Wolfs TF, Zwart G, Bakker M, Goudsmit J. HIV-1 genomic RNA diversification following sexual and parenteral virus transmission. *Virology*. 1992; 189(1):103–10. PMID: [1376536](https://pubmed.ncbi.nlm.nih.gov/1376536/).
9. Zhang LQ, MacKenzie P, Cleland A, Holmes EC, Brown AJ, Simmonds P. Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. *J Virol*. 1993; 67(6):3345–56. PMID: [8497055](https://pubmed.ncbi.nlm.nih.gov/8497055/); PubMed Central PMCID: PMC237678.
10. 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 U S A*. 2008; 105(21):7552–7. doi: [10.1073/pnas.0802203105](https://doi.org/10.1073/pnas.0802203105) PMID: [18490657](https://pubmed.ncbi.nlm.nih.gov/18490657/); PubMed Central PMCID: PMC2387184.
11. Salazar-Gonzalez JF, Salazar MG, Keele BF, Learn GH, Giorgi EE, Li H, et al. Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *J Exp Med*. 2009; 206(6):1273–89. doi: [10.1084/jem.20090378](https://doi.org/10.1084/jem.20090378) PMID: [19487424](https://pubmed.ncbi.nlm.nih.gov/19487424/); PubMed Central PMCID: PMC2715054.
12. Tsai L, Tasovski I, Leda AR, Chin MP, Cheng-Mayer C. The number and genetic relatedness of transmitted/founder virus impact clinical outcome in vaginal R5 SHIVSF162P3N infection. *Retrovirology*. 2014; 11:22. doi: [10.1186/1742-4690-11-22](https://doi.org/10.1186/1742-4690-11-22) PMID: [24612462](https://pubmed.ncbi.nlm.nih.gov/24612462/); PubMed Central PMCID: PMC3975242.
13. Gonzalez MW, DeVico AL, Lewis GK, Spouge JL. Conserved molecular signatures in gp120 are associated with the genetic bottleneck during simian immunodeficiency virus (SIV), SIV-human immunodeficiency virus (SHIV), and HIV type 1 (HIV-1) transmission. *J Virol*. 2015; 89(7):3619–29. doi: [10.1128/JVI.03235-14](https://doi.org/10.1128/JVI.03235-14) PMID: [25589663](https://pubmed.ncbi.nlm.nih.gov/25589663/); PubMed Central PMCID: PMC4403421.
14. Mansky LM, Temin HM. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol*. 1995; 69(8):5087–94. PMID: [7541846](https://pubmed.ncbi.nlm.nih.gov/7541846/); PubMed Central PMCID: PMC189326.
15. Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A, et al. Timing the ancestor of the HIV-1 pandemic strains. *Science*. 2000; 288(5472):1789–96. PMID: [10846155](https://pubmed.ncbi.nlm.nih.gov/10846155/).
16. Domingo E, Sheldon J, Perales C. Viral quasispecies evolution. *Microbiol Mol Biol Rev*. 2012; 76(2):159–216. doi: [10.1128/MMBR.05023-11](https://doi.org/10.1128/MMBR.05023-11) PMID: [22688811](https://pubmed.ncbi.nlm.nih.gov/22688811/); PubMed Central PMCID: PMC3372249.
17. Wargo AR, Kurath G. Viral fitness: definitions, measurement, and current insights. *Current opinion in virology*. 2012; 2(5):538–45. doi: [10.1016/j.coviro.2012.07.007](https://doi.org/10.1016/j.coviro.2012.07.007) PMID: [22986085](https://pubmed.ncbi.nlm.nih.gov/22986085/).
18. Nowak MA. Variability of HIV infections. *J Theor Biol*. 1992; 155(1):1–20. PMID: [1619947](https://pubmed.ncbi.nlm.nih.gov/1619947/).
19. Shankarappa R, Gupta P, Learn GH Jr., Rodrigo AG, Rinaldo CR Jr., Gorry MC, et al. Evolution of human immunodeficiency virus type 1 envelope sequences in infected individuals with differing disease progression profiles. *Virology*. 1998; 241(2):251–9. doi: [10.1006/viro.1997.8996](https://doi.org/10.1006/viro.1997.8996) PMID: [9499799](https://pubmed.ncbi.nlm.nih.gov/9499799/).
20. Briones C, Domingo E. Minority report: hidden memory genomes in HIV-1 quasispecies and possible clinical implications. *AIDS Rev*. 2008; 10(2):93–109. PMID: [18615120](https://pubmed.ncbi.nlm.nih.gov/18615120/).
21. Churchill M, Nath A. Where does HIV hide? A focus on the central nervous system. *Curr Opin HIV AIDS*. 2013; 8(3):165–9. doi: [10.1097/COH.0b013e32835fc601](https://doi.org/10.1097/COH.0b013e32835fc601) PMID: [23429501](https://pubmed.ncbi.nlm.nih.gov/23429501/).
22. Kanmogne GD, Kennedy RC, Grammas P. Analysis of human lung endothelial cells for susceptibility to HIV type 1 infection, coreceptor expression, and cytotoxicity of gp120 protein. *AIDS Res Hum Retroviruses*. 2001; 17(1):45–53. doi: [10.1089/088922201750056771](https://doi.org/10.1089/088922201750056771) PMID: [11177382](https://pubmed.ncbi.nlm.nih.gov/11177382/).
23. Nannini EC, Okhuysen PC. HIV1 and the gut in the era of highly active antiretroviral therapy. *Current gastroenterology reports*. 2002; 4(5):392–8. PMID: [12228041](https://pubmed.ncbi.nlm.nih.gov/12228041/).
24. van der Sluis RM, Jeeninga RE, Berkhout B. Establishment and molecular mechanisms of HIV-1 latency in T cells. *Current opinion in virology*. 2013; 3(6):700–6. doi: [10.1016/j.coviro.2013.07.006](https://doi.org/10.1016/j.coviro.2013.07.006) PMID: [23953324](https://pubmed.ncbi.nlm.nih.gov/23953324/).

25. Alexaki A, Liu Y, Wigdahl B. Cellular reservoirs of HIV-1 and their role in viral persistence. *Curr HIV Res.* 2008; 6(5):388–400. Epub 2008/10/16. PMID: [18855649](#); PubMed Central PMCID: [PMC2683678](#).
26. Shirakawa K, Chavez L, Hakre S, Calvanese V, Verdin E. Reactivation of latent HIV by histone deacetylase inhibitors. *Trends in microbiology.* 2013; 21(6):277–85. doi: [10.1016/j.tim.2013.02.005](#) PMID: [23517573](#); PubMed Central PMCID: [PMC3685471](#).
27. Ensoli B, Cafaro A, Monini P, Marcotullio S, Ensoli F. Challenges in HIV Vaccine Research for Treatment and Prevention. *Frontiers in immunology.* 2014; 5:417. doi: [10.3389/fimmu.2014.00417](#) PMID: [25250026](#); PubMed Central PMCID: [PMC4157563](#).
28. Gu WG. Genome editing-based HIV therapies. *Trends in biotechnology.* 2015; 33(3):172–9. doi: [10.1016/j.tibtech.2014.12.006](#) PMID: [25600622](#).
29. Li L, Aiamkitsumrit B, Pirrone V, Nonnemacher MR, Wojno A, Passic S, et al. Development of co-selected single nucleotide polymorphisms in the viral promoter precedes the onset of human immunodeficiency virus type 1-associated neurocognitive impairment. *J Neurovirol.* 2011; 17(1):92–109. Epub 2011/01/13. doi: [10.1007/s13365-010-0014-1](#) PMID: [21225391](#); PubMed Central PMCID: [PMC3057211](#).
30. Aiamkitsumrit B, Sullivan NT, Nonnemacher MR, Pirrone V, Wigdahl B. Human Immunodeficiency Virus Type 1 Cellular Entry and Exit in the T Lymphocytic and Monocytic Compartments: Mechanisms and Target Opportunities During Viral Disease. *Adv Virus Res.* 2015; 93:257–311. doi: [10.1016/bs.aivir.2015.04.001](#) PMID: [26111588](#).
31. Aiamkitsumrit B, Dampier W, Antell G, Rivera N, Martin-Garcia J, Pirrone V, et al. Bioinformatic analysis of HIV-1 entry and pathogenesis. *Curr HIV Res.* 2014; 12(2):132–61. PMID: [24862329](#); PubMed Central PMCID: [PMC4382797](#).
32. Aiamkitsumrit B, Dampier W, Martin-Garcia J, Nonnemacher MR, Pirrone V, Ivanova T, et al. Defining differential genetic signatures in CXCR4- and the CCR5-utilizing HIV-1 co-linear sequences. *PLoS One.* 2014; 9(9):e107389. doi: [10.1371/journal.pone.0107389](#) PMID: [25265194](#); PubMed Central PMCID: [PMC4180074](#).
33. Bouckaert R, Heled J, Kuhnert D, Vaughan T, Wu CH, Xie D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol.* 2014; 10(4):e1003537. doi: [10.1371/journal.pcbi.1003537](#) PMID: [24722319](#); PubMed Central PMCID: [PMC3985171](#).
34. Bezemer D, Faria NR, Hassan A, Hamers RL, Mutua G, Anzala O, et al. HIV Type 1 transmission networks among men having sex with men and heterosexuals in Kenya. *AIDS Res Hum Retroviruses.* 2014; 30(2):118–26. doi: [10.1089/AID.2013.0171](#) PMID: [23947948](#); PubMed Central PMCID: [PMC3910480](#).
35. Kondo M, Lemey P, Sano T, Itoda I, Yoshimura Y, Sagara H, et al. Emergence in Japan of an HIV-1 variant associated with transmission among men who have sex with men (MSM) in China: first indication of the International Dissemination of the Chinese MSM lineage. *J Virol.* 2013; 87(10):5351–61. doi: [10.1128/JVI.02370-12](#) PMID: [23365432](#); PubMed Central PMCID: [PMC3648132](#).
36. Josefsson L, von Stockenstrom S, Faria NR, Sinclair E, Bacchetti P, Killian M, et al. The HIV-1 reservoir in eight patients on long-term suppressive antiretroviral therapy is stable with few genetic changes over time. *Proc Natl Acad Sci U S A.* 2013; 110(51):E4987–96. doi: [10.1073/pnas.1308313110](#) PMID: [24277811](#); PubMed Central PMCID: [PMC3870728](#).
37. Ho SY, Shapiro B. Skyline-plot methods for estimating demographic history from nucleotide sequences. *Molecular ecology resources.* 2011; 11(3):423–34. doi: [10.1111/j.1755-0998.2011.02988.x](#) PMID: [21481200](#).
38. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nature methods.* 2012; 9(8):772. doi: [10.1038/nmeth.2109](#) PMID: [22847109](#).
39. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004; 32(5):1792–7. Epub 2004/03/23. doi: [10.1093/nar/gkh340](#) PMID: [15034147](#); PubMed Central PMCID: [PMC390337](#).
40. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014; 30(9):1312–3. doi: [10.1093/bioinformatics/btu033](#) PMID: [24451623](#); PubMed Central PMCID: [PMC3998144](#).
41. Sukumaran J, Holder MT. DendroPy: a Python library for phylogenetic computing. *Bioinformatics.* 2010; 26(12):1569–71. doi: [10.1093/bioinformatics/btq228](#) PMID: [20421198](#).
42. Verbeke G, Fieuws S, Molenberghs G, Davidian M. The analysis of multivariate longitudinal data: a review. *Statistical methods in medical research.* 2014; 23(1):42–59. doi: [10.1177/0962280212445834](#) PMID: [22523185](#); PubMed Central PMCID: [PMC3404254](#).

43. Vider-Shalit T, Almani M, Sarid R, Louzoun Y. The HIV hide and seek game: an immunogenomic analysis of the HIV epitope repertoire. *AIDS*. 2009; 23(11):1311–8. doi: [10.1097/QAD.0b013e32832c492a](https://doi.org/10.1097/QAD.0b013e32832c492a) PMID: [19550286](https://pubmed.ncbi.nlm.nih.gov/19550286/).
44. Krakauer DC, Sasaki A. The greater than twofold cost of integration for retroviruses. *Proc Biol Sci*. 2006; 273(1590):1159–63. doi: [10.1098/rspb.2005.3322](https://doi.org/10.1098/rspb.2005.3322) PMID: [16600896](https://pubmed.ncbi.nlm.nih.gov/16600896/); PubMed Central PMCID: PMC1560259.
45. Joos B, Fischer M, Kuster H, Pillai SK, Wong JK, Boni J, et al. HIV rebounds from latently infected cells, rather than from continuing low-level replication. *Proc Natl Acad Sci U S A*. 2008; 105(43):16725–30. doi: [10.1073/pnas.0804192105](https://doi.org/10.1073/pnas.0804192105) PMID: [18936487](https://pubmed.ncbi.nlm.nih.gov/18936487/); PubMed Central PMCID: PMC2575487.
46. Chun TW. Tracking replication-competent HIV reservoirs in infected individuals. *Curr Opin HIV AIDS*. 2013; 8(2):111–6. doi: [10.1097/COH.0b013e32835d6e1c](https://doi.org/10.1097/COH.0b013e32835d6e1c) PMID: [23380652](https://pubmed.ncbi.nlm.nih.gov/23380652/).
47. Richman DD, Margolis DM, Delaney M, Greene WC, Hazuda D, Pomerantz RJ. The challenge of finding a cure for HIV infection. *Science*. 2009; 323(5919):1304–7. doi: [10.1126/science.1165706](https://doi.org/10.1126/science.1165706) PMID: [19265012](https://pubmed.ncbi.nlm.nih.gov/19265012/).
48. Richman DD. Introduction: challenges to finding a cure for HIV infection. *Curr Opin HIV AIDS*. 2011; 6(1):1–3. doi: [10.1097/COH.0b013e328340ffa6](https://doi.org/10.1097/COH.0b013e328340ffa6) PMID: [21242886](https://pubmed.ncbi.nlm.nih.gov/21242886/); PubMed Central PMCID: PMC3077900.