


RESEARCH ARTICLE

Extremely High Mutation Rate of HIV-1 In Vivo

José M. Cuevas¹ , Ron Geller¹ , Raquel Garijo¹, José López-Aldeguer^{2,3}, Rafael Sanjuán^{1,4*}

1 Instituto Cavanilles de Biodiversidad y Biología Evolutiva, Valencia, Spain, **2** Hospital Universitario La Fe, Valencia, Spain, **3** CoRIS and HIV Biobank, Spanish AIDS Research Network, Spain, **4** Departament de Genètica, Universitat de València, Valencia, Spain

 These authors contributed equally to this work.

* rafael.sanjuan@uv.es



CrossMark
click for updates

 OPEN ACCESS

Citation: Cuevas JM, Geller R, Garijo R, López-Aldeguer J, Sanjuán R (2015) Extremely High Mutation Rate of HIV-1 In Vivo. *PLoS Biol* 13(9): e1002251. doi:10.1371/journal.pbio.1002251

Academic Editor: Sarah L. Rowland-Jones, Weatherall Institute of Molecular Medicine, UNITED KINGDOM

Received: April 3, 2015

Accepted: August 10, 2015

Published: September 16, 2015

Copyright: © 2015 Cuevas 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 paper, its Supporting Information files ([S1–S5 Data](#)), and from the GenBank database (accessions KT200348–KT200358 and KT205403–KT205555).

Funding: This work was funded by the Instituto de Salud Carlos III (RD12/0017 -RIS), a Starting Grant from the European Research Council (erc.europa.eu) (ERC-2011-StG- 281191-VIRMUT), and a grant from the Spanish Ministerio de Economía y Competitividad (www.mineco.gob.es) (BFU2013-41329-P) to RS. The HIV BioBank, integrated in the Spanish AIDS Research Network, is supported by

Abstract

Rates of spontaneous mutation critically determine the genetic diversity and evolution of RNA viruses. Although these rates have been characterized in vitro and in cell culture models, they have seldom been determined in vivo for human viruses. Here, we use the intrapatient frequency of premature stop codons to quantify the HIV-1 genome-wide rate of spontaneous mutation in DNA sequences from peripheral blood mononuclear cells. This reveals an extremely high mutation rate of $(4.1 \pm 1.7) \times 10^{-3}$ per base per cell, the highest reported for any biological entity. Sequencing of plasma-derived sequences yielded a mutation frequency 44 times lower, indicating that a large fraction of viral genomes are lethally mutated and fail to reach plasma. We show that the HIV-1 reverse transcriptase contributes only 2% of mutations, whereas 98% result from editing by host cytidine deaminases of the A3 family. Hypermutated viral sequences are less abundant in patients showing rapid disease progression compared to normal progressors, highlighting the antiviral role of A3 proteins. However, the amount of A3-mediated editing varies broadly, and we find that low-edited sequences are more abundant among rapid progressors, suggesting that suboptimal A3 activity might enhance HIV-1 genetic diversity and pathogenesis.

Author Summary

The high levels of genetic diversity of the HIV-1 virus grant it the ability to escape the immune system, to rapidly evolve drug resistance, and to circumvent vaccination strategies. However, our knowledge of HIV-1 mutation rates has been largely restricted to in vitro and cell culture studies because of the inherent complexity of measuring these rates in vivo. Here, by analyzing the frequency of premature stop codons, we show that the HIV-1 mutation rate in vivo is two orders of magnitude higher than that predicted by in vitro studies, making it the highest reported mutation rate for any biological system. A large component of this rate is from host cellular cytidine deaminases, which induce mutations in the viral DNA as a defense mechanism. While the HIV-1 genome is hypermutated in blood cells, only a very small fraction of these mutations reach the plasma, indicating

Instituto de Salud Carlos III, Spanish Health Ministry (grant RD06/0006/0035 and RD12/0017/0037) as part of the Plan Nacional R + D + I and cofinanced by ISCIII- Subdirección General de Evaluación y el Fondo Europeo de Desarrollo Regional (FEDER) and Fundación para la investigación y prevención del SIDA en España (FIPSE). The RIS Cohort (CoRIS) is funded by the Instituto de Salud Carlos III through the Red Temática de Investigación Cooperativa en SIDA (RIS C03/173 and RD12/0017/0018) as part of the Plan Nacional R+D+I and cofinanced by ISCIII- Subdirección General de Evaluación y el Fondo Europeo de Desarrollo Regional (FEDER). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

Abbreviations: A3, apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3; ADAR, double-stranded RNA-specific adenosine deaminase; HCV, hepatitis C virus; HSV, herpes simplex virus 1; MHV, murine hepatitis virus; NSMT, nonsense mutational targets; PBMC, peripheral blood mononuclear cells; RT, reverse transcriptase; TEV, tobacco etch virus; TMV, tobacco mosaic virus; Vif, viral infectivity factor; VSV, vesicular stomatitis virus.

that many viruses are defective as a result of the extremely high mutation load. In addition, we find that the HIV-1 mutation rate tends to be higher in patients showing normal disease progression than in those undergoing rapid progression, emphasizing the negative impact on viral fitness of hypermutation by host cytidine deaminases. However, we also observe subpopulations of weakly-mutated viral genomes whose sequence diversity may influence viral pathogenesis. Our work highlights the fine balance for HIV-1 between enough mutation to evade host responses and too much mutation that can inactivate the virus.

Introduction

RNA viruses exist as extremely diverse populations, with every possible spontaneous mutation along the genome appearing within each patient every day [1]. This diversity plays a fundamental role in HIV-1 biology, enabling the virus to successfully evade the immune system, rapidly modify cell tropism, evolve drug resistances, and thwart vaccination strategies [2,3]. Similar to other RNA viruses, the great diversity of HIV-1 stems from its high rate of spontaneous mutation, defined as the probability that a new mutation appears per base in each infected cell [4,5]. The HIV-1 reverse transcriptase (RT) lacks proofreading activity and has an estimated error rate on the order of 3×10^{-5} per base per round of copying as determined in cell culture studies [4,6–9]. However, these estimates may not truly reflect the mutational process of HIV-1 in patients, because cellular factors such as dNTP levels or sequence context can affect the frequency and type of mutations produced [10–12]. Furthermore, HIV-1 is subject in vivo to editing by cellular enzymes of the apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3 (A3) family, which are packaged into the virion and, upon infection of a new cell, mediate the edition of cytidine to uracil in the negative-strand viral cDNA, resulting in G→A substitutions in the viral genomic RNA [13–15]. Numerous A3 enzymes are expressed in humans, but A3D, A3F, A3G and A3H are believed to be the most relevant to HIV-1 pathogenesis [14,15]. Their antiviral role is underscored by the function of the HIV-1 Viral infectivity factor (Vif), which promotes A3 proteasomal degradation and is required for infectivity in CD4 cells [13]. Multiple studies have demonstrated the existence of high levels of A3-induced substitutions in patient-derived HIV-1 sequences (hypermutation), but the contribution of A3 to the total mutation rate of the virus has not been quantified. Also, conflicting results have been reported regarding the role played by A3 in HIV-1 diversity and evolution. While cell culture experiments have suggested that hypermutation is invariably lethal for the virus [16], others have reported that A3 can promote immune escape and drug resistance [17,18]. Interestingly, A3 expression differs among patients, and recent work has shown that at least seven A3H haplotypes with different geographic distributions exist, of which only three represent stable enzymes capable of editing HIV-1 but can be counteracted by some Vif variants [14,19–22]. However, the role played by A3 edition in disease progression still remains debated, with some studies suggesting a positive association between A3 activity and clinical outcome [23–25], but not others [26]. Hence, the HIV-1 mutation rate in vivo, the contribution of the HIV-1 RT, and host A3 proteins to this rate, as well as its relevance to disease progression, remain to be elucidated.

In this study, we estimate the mutation rate of HIV-1 using sequences derived from both intracellular viral DNA and plasma viral RNA. Although many factors determine the genetic diversity of a virus, including, among others, natural selection, transmission bottlenecks, and cell turnover rates, the effect of mutation rate on genetic diversity can be disentangled from these other factors by focusing on mutations that abrogate viral infectivity (lethal mutations)

[27]. We thus performed massive parallel sequencing of the entire HIV-1 protein coding region from 11 infected donors, obtained a high confidence set of 3,069 likely lethal mutations in viral DNA from cells, and contrasted these results with those obtained from previously published HIV-1 sequences derived from plasma RNA. Whereas the observed mutation rate in plasma was compatible with the known fidelity of the HIV-1 RT, our results reveal a strong inflation of the mutation rate in DNA sequences, with an estimated $(4.1 \pm 1.7) \times 10^{-3}$ mutations per base per cell. By examining the relative contribution of the HIV-1 RT, A3G, and A3D/F/H to the overall mutation rate of the virus, we show that this extremely high rate is essentially driven by the action of A3 enzymes. Supporting the antiviral role of A3-mediated HIV-1 genome editing, we find a significantly lower viral mutation rate in patients showing rapid disease progression compared to normal progressors and a negative correlation between the viral mutation rate and the set-point viral load. However, the extent of A3 editing varies broadly among sequences, with some sequences showing only few A3-driven mutations. Interestingly, we find that low-level A3 editing is more abundant in rapid progressors than in normal progressors, suggesting that failure of A3 to inactivate the virus by hypermutation may promote HIV-1 intrapatient diversity and pathogenesis.

Results

Inference of the HIV-1 Mutation Rate In Vivo

To infer the HIV-1 mutation rate from patient-derived sequences, we used the lethal mutation method [27], which builds on the principle that the frequency of lethal mutations in a population equals their rate of production, as these cannot be inherited. We used premature stop codons as a surrogate for lethal mutations, and the mutation rate was calculated by dividing the observed number of stops by the total number of possible single-base substitutions leading to stop codons (nonsense mutational targets, NSMTs; see [methods](#)). NSMTs are distributed throughout protein-coding regions, thus allowing for a genome-wide assessment of the mutation rate. When we applied the lethal mutation method to three large available datasets of HIV-1 sequences from plasma RNA (2689 subtype B *env*, 1450 subtype C *env*, and 1310 *gag* sequences from either subtype), we obtained a mutation rate of $(9.3 \pm 2.3) \times 10^{-5}$ per base per cell, which is slightly higher than those reported in cell culture studies [4]. However, A3-driven mutations may be largely absent from plasma because selection against highly mutated sequences may impede viral assembly and budding [28,29]. We therefore reasoned that focusing on plasma RNA may underestimate the mutation rate in vivo. Hence, we amplified the entire protein-coding region of the viral DNA directly from peripheral blood mononuclear cells (PBMCs) of 11 treatment-naive patients with known infection, viral load, and CD4 count histories ([Table 1](#), [S1 Fig](#), [S2 Fig](#)). High-fidelity limiting-dilution PCR was carried out in three overlapping fragments. Since in this method each positive PCR is effectively initiated from a single template molecule, cloning takes place before PCR, thus preventing the accumulation of PCR errors and amplification biases [30–34]. Using DNA as template, rather than RNA, has the additional advantage of avoiding the RT step prior to PCR, which is a major source of errors resulting from the low in vitro fidelity of these enzymes [35,36]. For each patient, 50 limiting-dilution clonal PCR products were obtained per fragment, thereby enabling us to analyze $11 \times 50 = 550$ full-length protein-coding sequences.

For practical reasons, PCRs were arranged in 110 libraries each containing an equimolar pool of five PCRs per fragment representing five full-length protein-coding regions, and libraries were uniquely tagged and subjected to Illumina paired-end sequencing. As a control, we similarly amplified 50 HIV-1 protein-coding regions from the reference plasmid pNL4-3. An average coverage of $19,150 \pm 5,801$ reads per site was achieved (minimum of 4,000), and

Table 1. Patient clinical data and mutation rate summary.

| Patient | Sex | Age | Infection time (years) | CD4count (cell/ μ L) | Per-year CD4 count decay rate (cell/ μ L) ^a | log ₁₀ set-point viral load (copies/mL) ^b | Progression rate ^c | HIV-1 mutation rate per base per cell |
|---------|-----|-----|------------------------|--------------------------|--|---|-------------------------------|---------------------------------------|
| R3 | M | 22 | 1.31 | 380 | 321 ± 82 | 5.4 ± 0.2 | Rapid | 1.9 × 10 ⁻³ |
| R5 | M | 37 | 1.82 | 432 | 183 ± 41 | 5.1 ± 0.1 | Rapid | 1.6 × 10 ⁻³ |
| R6 | M | 42 | 3.10 | 218 | 243 ± 39 | 4.6 ± 0.3 | Rapid | 3.7 × 10 ⁻³ |
| R7 | M | 54 | 1.63 | 291 | 459 ± 33 | 5.1 ± 0.1 | Rapid | 2.8 × 10 ⁻³ |
| R8 | M | 25 | 1.00 | 338 | 160 | 4.4 | Rapid | 2.4 × 10 ⁻³ |
| R11 | M | 52 | 0.96 | 689 | 87 ± 25 | 4.0 ± 0.2 | Normal | 3.4 × 10 ⁻³ |
| R14 | M | 31 | NA | 637 | 76 ± 22 | 4.1 ± 0.2 | Normal | 3.7 × 10 ⁻³ |
| R15 | M | 25 | 1.24 | 446 | 68 ± 25 | 4.2 ± 0.1 | Normal | 12.6 × 10 ⁻³ |
| R4 | M | 26 | 2.16 | 439 | 48 ± 57 | 4.3 ± 0.1 | Normal | 3.2 × 10 ⁻³ |
| R9 | M | 35 | 0.65 | 1146 | 124 ± 111 | 3.9 ± 0.1 | Normal | 4.8 × 10 ⁻³ |
| P6 | F | 49 | 20.94 | 431 | 37 ± 4 | 4.3 ± 0.1 | Normal | 4.5 × 10 ⁻³ |

^a Estimated by linear regression as the slope of CD4 counts against infection time. For R8, although there were few data points for reliably estimating the CD4 decay rate, rapid progression was supported by the observation that the CD4 count dropped to 338 cells/ μ L within the first year of infection. Full data are shown in [S1 Fig](#), the [S1 Data](#) file, and the [S2 Data](#) file.

^b Average log viral load obtained from all available patient samples taken at least one year postinfection but before the onset of treatment and/or symptoms. R8 had only one available viral load determination. For R9, a viral load determination taken 329 d after infection was included in the calculation to have at least three data points. Full data are shown in [S2 Fig](#), the [S1 Data](#) file, and the [S2 Data](#) file.

^c Rapid progressors were defined as patients showing a CD4 count decay rate greater than 150 cell/ μ L per y.

doi:10.1371/journal.pbio.1002251.t001

mutations relative to the patient consensus sequence were then called. Application of the lethal mutation method to the control pNL4-derived sequences yielded no mutations, as expected. In contrast, in samples from the 11 patients, we found 3,069 total stop-codon mutations in the 732,350 NSMTs examined, giving a per-patient average mutation rate of $(4.1 \pm 1.7) \times 10^{-3}$ per base per cell or, equivalently, one mutation every 250 bases. This value is 44 times greater than the rate observed in plasma samples ([Fig 1A](#)) and approximately two orders of magnitude higher than the rates reported previously in cell culture studies.

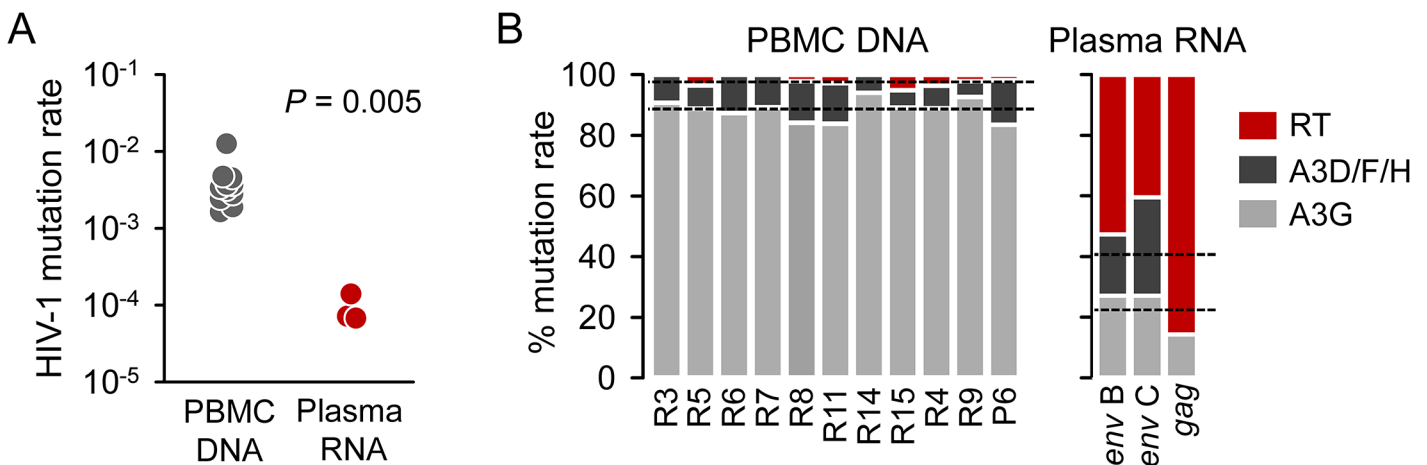


Fig 1. HIV-1 mutation rate and source of mutations in DNA from PBMCs and in RNA from plasma. **A.** The mutation rate was inferred from PBMC DNA and from plasma RNA using the lethal mutation method. Each data point represents the average mutation rate obtained for one patient (PBMC DNA) or the average rate calculated from one large publicly available dataset of *gag* or *env* sequences (plasma RNA). **B.** Contribution of A3G, A3D/F/H, and the viral RT to the observed mutation rate in PBMC DNA and plasma RNA sequences. Dotted lines indicate the average contribution of each enzyme. Individual numerical values shown in this Figure are available in [Table 1](#) and in the [S1 Data](#) file.

doi:10.1371/journal.pbio.1002251.g001

Contribution of A3 and RT to the HIV-1 Mutation Rate

The sequence contexts for cDNA editing by different A3 enzymes have been previously defined in detail [13,14,37,38]. A3G shows a strong preference for CC sequences on minus-strand cDNA leading to GG → AG mutations in the positive strand. In contrast, GA → AA editing is largely mediated by A3D/F/H. Hence, A3 activity leads to stop codons by edition of tryptophan TGG codons, with A3G and A3D/F/H leading to different stop codons depending on sequence context. A3G edition mutates TGG codons to TAG stops, and to either TGA or TAA stops if the TGG codon is followed by a G (TGGG), whereas A3D/F/H lead to TGA stops if the TGG codon is followed by an A (TGGA). In contrast, the HIV-1 RT can mediate all possible mutations at both TGG codons as well as numerous other codons. Using these sequence context preferences allowed us to assign mutations to A3G, A3D/F/H, or the HIV-1 RT (Table 2).

Analysis of sequences from viral RNA suggested a relatively balanced contribution of A3G, A3D/F/H, and HIV-1 RT to the total mutation rate (22.8 ± 4.2%, 17.6 ± 9.5%, and 59.7 ± 13.5%, respectively, assuming that all G→A changes falling at canonical A3 targets were produced by A3 and not by the HIV-1 RT; Fig 1B). However, these percentages may be inaccurate due to the low fidelity of the RT-PCR step needed for sequencing of viral RNA [35]. In contrast, in viral DNA sequences from PBMCs, A3G contributed 88.4 ± 1.1% of the total mutation rate, A3D/F/H contributed 9.7 ± 1.1%, and the HIV-1 RT only 2.0 ± 0.54% (Fig 1B). Among the total 56,800 NSMTs available for A3G editing in the viral DNA sequences from the 11 patients, 2,752 contained stop codon mutations, the fraction of A3G targets edited being thus 4.8%. This shows that the extremely high mutation rate observed in viral DNA was essentially driven by A3 edition, and that sequences edited by A3 were comparatively much more difficult to sample in RNA sequences from plasma than nonmutated or RT-mutated sequences. Indeed, the observed RT mutation rate was nearly identical in DNA (6.3 × 10⁻⁵) and plasma RNA sequences (6.2 × 10⁻⁵). In theory, G→A changes falling at canonical A3 targets could also be produced by the RT. However, since mutations were clearly concentrated in these targets and the RT does not show such sequence-context preferences [36], the vast majority of mutations should be correctly assigned to A3. Further confirming our assignment of mutations, the observed edits were consistent with known A3 sequence-context preferences beyond the GG and GA dinucleotides [37,38]. Specifically, A3G mutations were enhanced by the presence of a T at position -2 relative to the mutated G (TNGG; Binomial test: *p* < 0.01 in 6 of 11 patients), a T at position -1 (TGG; *p* < 0.01 in 11 of 11 patients), and a G at position +2 (GGG; *p* < 0.01 in 10 of 11 patients), whereas they were negatively influenced by an A or a G at position -1 (A/GGG; *p* < 0.01 in 7 and 9 of 11 patients, respectively) and a C at position +2 (GGC; *p* < 0.01 in 10 of 11 patients). In turn, A3D/F/H mutations were also favored by a T at position -1 (TGA; *p* < 0.01 in 7 of 11 patients) and negatively affected by a C at position -2 (CNGA; *p* < 0.01 in 7 of 11 patients) or a C at position +2 (GAC; *P* < 0.01 in 9 of 11 patients).

Table 2. Assignment of mutations in NSMTs to A3G, A3D/F/H, or RT.

| NSMT | Resulting stop codon ^a | | |
|-------------------|-----------------------------------|---------|---------------|
| | TAG | TGA | TAA |
| TGG A | A3G | A3D/F/H | A3G + A3D/F/H |
| TGG C | A3G | RT | A3G + RT |
| TGG G | A3G | A3G | A3G + A3G |
| TGG T | A3G | RT | A3G + RT |
| Other (17 codons) | RT | RT | RT |

^aIn all cases the indicated mutations can also be produced by the RT in addition to A3 enzymes.

doi:10.1371/journal.pbio.1002251.t002

Variation in the HIV-1 Mutation Rate

In our dataset, stop codons were observed throughout the viral genome, and their abundance tended to vary accordingly with the number of available NSMTs, as expected (S3 Fig). Calculation of the mutation rate (stops/NSMTs) using a 50-codon sliding window revealed several regions with a higher-than-average rate, which mapped mainly to gp41, the N-terminal region of gp120, and the central regions of *gag* and *pol* (Fig 2). Comparison of mutation rates for the major five genes (*gag*, *pol*, *vif*, *env*, and *nef*) also yielded significant differences (Kruskal-Wallis test: $p < 0.001$), the rate being highest in gp41 and lowest in *vif* (Tukey HSD posthoc test: $p < 0.05$). A3 editing depends on the amount of time the viral DNA is found as single-stranded DNA, the preferred A3 substrate, and this produces twin-peaks gradients of A3 edition, which are largely consistent with our observed distribution of mutation rates [37,38].

In addition to the across-genome mutation rate variation, we were interested in characterizing how the number of mutations varied among sequencing libraries. The effects of A3 edition on viral population fitness may be relatively minor if the vast majority of mutations were clustered in a small subset of sequences, whereas they should be substantial otherwise. While an all-or-nothing model has been suggested wherein any HIV-1 genome subjected to editing is so extensively mutated that it has an exceedingly low probability of being viable [16], other works have shown varying levels of A3 editing [17,18]. To address this, we first examined the number of total, A3G-, and A3D/F/H-driven lethal mutations in each of the 110 sequencing libraries. We found that only 17 libraries (15.4%) were free of stop codon mutations and that, among the rest of the libraries, mutation counts varied gradually across two orders of magnitude (from 1 to 164), arguing against an all-or-nothing pattern (Fig 3A and 3B). Examination of each patient separately also indicated a large variation in the number of mutations and that mutation-free libraries were rare (0% to 30% depending on the patient; Fig 3C).

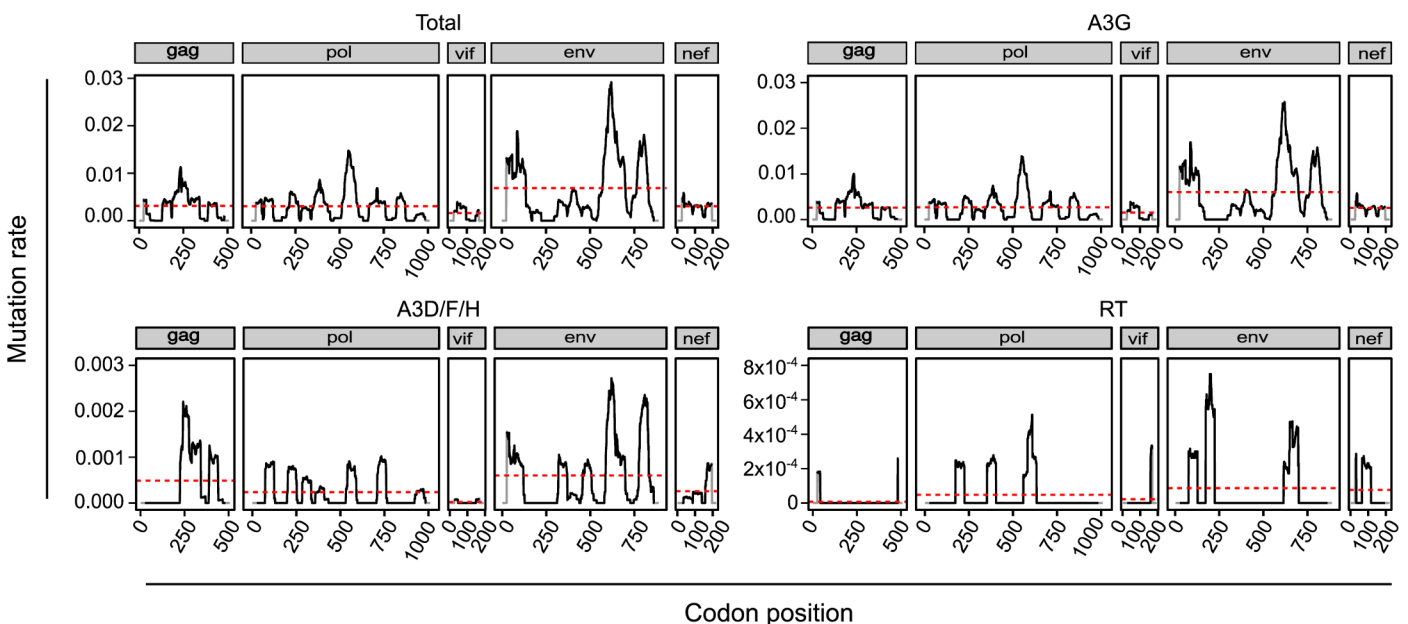


Fig 2. Distribution of mutation rates across HIV-1 genes. The average total, A3G, A3D/F/H, and RT mutation rate within a sliding window of 50 codons across HIV1 genes (black skyline), and the average mutation rate for each gene (red dashed line) are shown. Stop codon and NSMT counts are represented in S3 Fig, and numerical values are available from the S3 Data file.

doi:10.1371/journal.pbio.1002251.g002

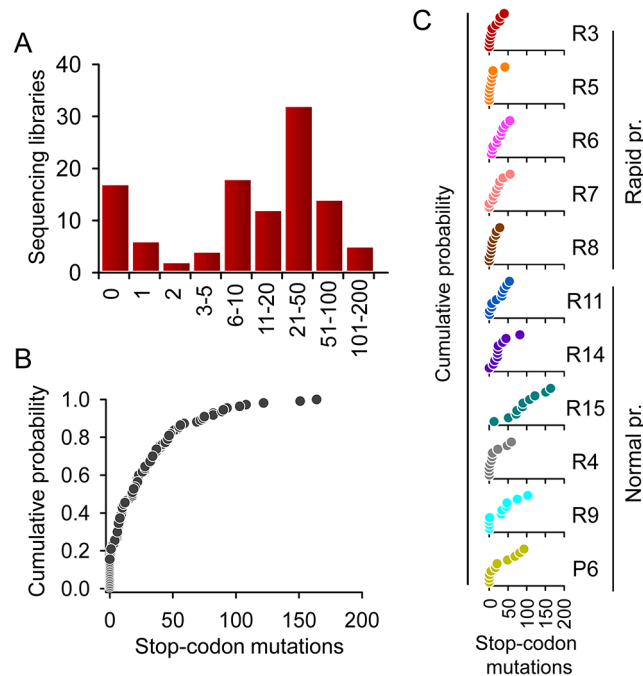


Fig 3. Variation in mutation rate among HIV-1 sequences. **A.** Histogram showing the distribution of the number of stop codon mutations per sequencing library. **B and C.** Cumulative probability distribution of the number of stop codon mutations per sequencing library for the entire dataset (**B**) and for each patient (**C**). Notice that the number of mutations tends to be higher in normal progressors (bottom) than in rapid progressors (top). Individual numerical values shown in this Figure are available in the [S4 Data](#) file.

doi:10.1371/journal.pbio.1002251.g003

However, since each library was made of five limiting-dilution PCRs, these data did not have enough resolution to assess the number of mutations per individual sequence. To address this, we performed a more detailed, intralibrary analysis of mutations for the *nef* gene. In total, there were 147 *nef* premature stop codon mutations in 45,400 NSMTs, giving a mutation rate of 3.2×10^{-3} per base per cell. Of the 110 libraries, 31 contained two or more stop codon mutations in this single gene. For each of these 31 libraries, we sequenced the five limiting-dilution PCR clones of the library by the Sanger method to ascertain whether the multiple stops occurred in one or several clones. We found that in 13 libraries (42%), stop codon mutations were significantly clustered in only one or two of the clones (test against Poisson expectation: $p < 0.05$), whereas in the remaining 18 libraries (58%), we could not reject the possibility that mutations were randomly spread across clones ([S4 Fig](#); [S5 Data](#)). These data confirm that there are hyper-mutated sequences but that A3 edition does not follow an all-or-nothing pattern in vivo.

Correlation between the HIV-1 Mutation Rate and Disease Progression

Five of the patients were defined as rapid progressors based on CD4 count decay rates (>150 cells/ μ L per year), whereas the other six patients were normal progressors ([Table 1](#)). This grouping was supported by differences in set-point viral load, defined as the approximately stable viral load reached after acute infection but before the onset of symptoms and/or treatment. The viral load set-point has been shown to be a good predictor of disease progression [[39,40](#)]. Here, the set-point load averaged $(1.1 \pm 0.4) \times 10^5$ copies/mL for rapid progressors, versus $(1.4 \pm 0.2) \times 10^4$ copies/mL for normal progressors (Mann-Whitney test: $p = 0.004$). Despite the low number of patients in each group, we found several lines of evidence linking the HIV-1 mutation rate with disease progression. Firstly, the distribution of the number of stop-codon

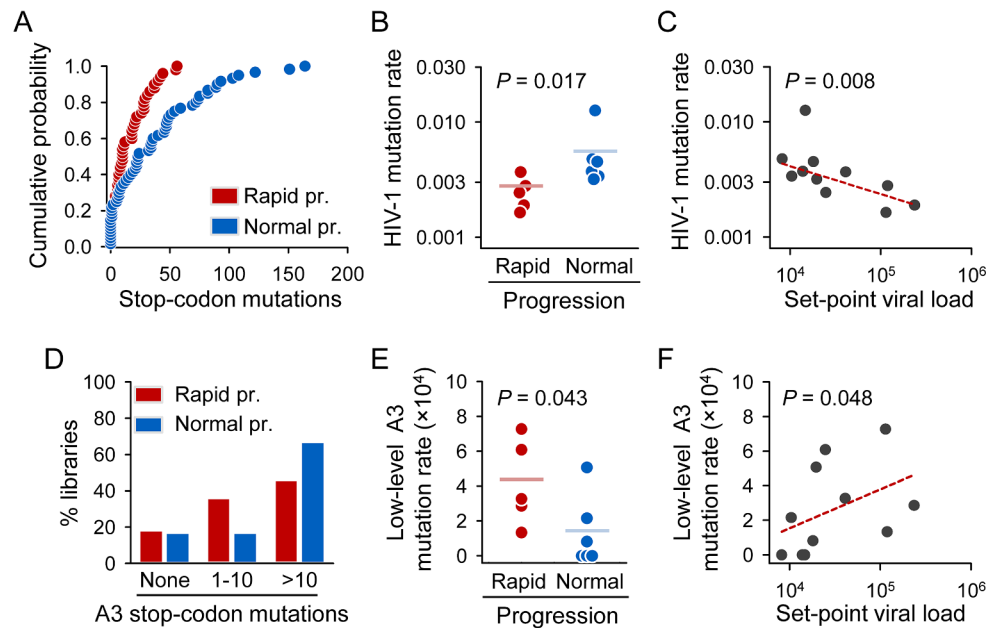


Fig 4. Association between HIV-1 mutation rate and disease progression markers. **A.** Cumulative distribution of the number of stop-codon mutations per sequencing library for rapid and normal progressors. **B.** Differences in mutation rate between these two patient groups. **C.** Inverse correlation between the HIV-1 mutation rate and the set-point viral load. The linear regression line obtained after removing the high-mutation outlier patient R15 is shown in red. **D.** Percent of libraries showing no (0), low-level (1–10), or high-level A3 editing (>10) in rapid and normal progressors. **E.** Differences in low-level A3 mutation rate (using libraries with 1–10 A3-driven mutations) between normal and rapid progressors. **F.** Positive correlation between the low-level A3 mutation rate and the set-point viral load. The linear regression line is shown in red. Individual numerical values shown in this Figure are available on [Table 1](#) and in the [S1 Data](#) file.

doi:10.1371/journal.pbio.1002251.g004

mutations per sequencing library differed between rapid and normal progressors (Fig 4A; Kolmogorov-Smirnov test: $p = 0.003$). Secondly, there was a statistically significant, 2.2-fold decrease in the HIV-1 mutation rate in rapid progressors compared with normal progressors, from $(5.4 \pm 3.6) \times 10^{-3}$ to $(2.5 \pm 0.8) \times 10^{-3}$ (Mann-Whitney test: $p = 0.017$; Fig 4B). Thirdly, we found that the mutation rate correlated negatively with the set-point viral load (Spearman $\rho = -0.755$; $p = 0.007$; Fig 4C). Examination of the distribution of mutations across the viral genome revealed that the difference between rapid and normal progressors was consistent in all genes except *gag*, in agreement with findings showing that A3G expression in infected individuals correlates with hypermutation levels in *vif* and *env* but not in *gag* [41] (S5 Fig).

Interestingly, the consensus viral sequence from patient R15 had a nonconservative amino acid replacement in the histidine(H)/cysteine(C)-containing (HCCH) motif of Vif (E134T), which has been previously implicated in the interaction of Vif with CBF- β , a requirement for A3 degradation [42]. As predicted from the anti-A3 effect of Vif, the HIV-1 mutation rate for this patient was four times higher than the average for the other patients (Table 1). Importantly, the association between disease progression and the HIV-1 mutation rate was not driven exclusively by this single data point, since significant differences were still observed after removal of this patient from the dataset (Mann-Whitney test: $p = 0.030$; correlation with set-point viral load: $\rho = -0.782$; $p = 0.008$). Finally, since A3H is the most genetically diverse of the seven human A3 genes, with some alleles encoding stable A3H forms while others encoding short-lived forms with little or no antiviral activity [22], we determined the A3H genotype of each patient. All patients were homozygous for an inactive A3H allele with the exception of

R15. This patient was heterozygous for an active A3H allele (see [Methods](#)), but R15 Vif sequences contained the A3H susceptibility residue V39 [19–21], thus suggesting an additional mechanism for the extremely high HIV-1 mutation rate shown by R15. Although our results support the antiviral effect of A3-mediated hypermutation, previous work has suggested that low-level editing may also contribute to viral diversity and, potentially, to pathogenesis [17,18]. We found that the number of A3 mutations per library did not follow the same distribution in rapid and normal progressors. In rapid progressors, 36% of the libraries had between one and ten mutations (low-level editing), whereas this fraction dropped to 17% in normal progressors (Fisher test: $p = 0.021$; [Fig 4D](#)). Whereas highly-edited libraries (>10 stop-codon mutations) recapitulated the results obtained with the full dataset (mutation rates: 2.0×10^{-3} and 5.2×10^{-3} for rapid and normal progressors, respectively; Mann-Whitney test: $p = 0.018$), the situation was reversed for low-level editing, with rapid progressors showing a 3.2-fold increase in mutation rate compared to normal progressors (4.2×10^{-4} and 1.3×10^{-4} , respectively; $p = 0.043$; [Fig 4E](#)). Similarly, in contrast to the total mutation rate, the low-level A3 editing rate correlated positively with set-point viral load (Spearman $\rho = 0.606$; $P = 0.048$; [Fig 4F](#)). Notice that the low-level A3 editing rate was still at least twice as high as the RT mutation rate. Our results thus support a dual role for A3 in disease progression. On one hand, A3 exert an antiviral effect by introducing a large number of mutations in the HIV-1 genome, which usually abolish infectivity. On the other hand, though, a less extensive edition does not ensure the loss of infectivity and may promote pathogenicity.

Discussion

The rate of spontaneous mutation is a major determinant of viral diversity and evolution, plays a role in the success of vaccination strategies [43,44], determines the likelihood that live attenuated vaccines revert to virulence [45], and influences the risk of disease emergence at the epidemiological level [46,47]. In the last years, direct experimental evidence has established that the viral mutation rate is also a virulence factor. For instance, a poliovirus with a high-fidelity polymerase showed lower ability to evolve drug resistance and to escape antibody neutralization in cell culture and was significantly attenuated in mice as a result of its impaired ability to evade the immune response or to adapt to different microenvironments in vivo [48,49]. Similar results have been obtained with a high-fidelity variant of chikungunya virus [50] and, recently, with other fidelity variants of poliovirus [51] and enterovirus 71 [52], suggesting that RNA viruses have optimized their mutation rates for maximal adaptability and underscoring the importance of quantifying viral mutation rates in vivo. To date, this has been largely done under cell culture conditions for different RNA viruses, including HIV-1 and other retroviruses, influenza virus, measles virus, poliovirus, plant viruses, and bacteriophages and has indicated that RNA virus mutation rates range from 10^{-6} to 10^{-4} per base per cell [4]. In contrast, quantitation of these rates in vivo has been more problematic. The only available estimates for a human virus in vivo correspond to hepatitis C virus (HCV) and are consistent with the values obtained for other RNA viruses in cell culture [53,54].

Inference of the rate of spontaneous mutations in vivo is complicated by the unknown number of viral generations (i.e., infection cycles), the removal of a large number of deleterious mutations by selection, and genetic drift, among others. By focusing on a single cell infection cycle, the lethal mutation method bypasses most of these uncontrolled factors. The assumption that sequences carrying stop codons will not undergo subsequent infection cycles is substantiated by the essential or quasiessential nature of all HIV-1 genes in vivo. However, some stop codon mutants may be able to complete multiple infection cycles if they are genetically complemented by other genomes coinfecting the same cells, as has been shown for some specific stop codon mutations in other RNA viruses [55]. However, if this process was widespread in HIV-1,

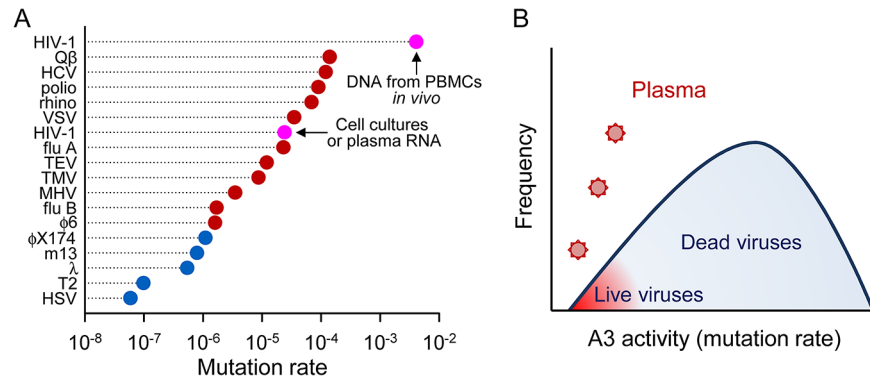


Fig 5. Comparison of viral mutation rates and effect of A3-mediated editing. **A.** Mutation rates per base per cell are shown for HIV-1, bacteriophage Q β , HCV, poliovirus, human rhinovirus 14, vesicular stomatitis virus (VSV), influenza A virus, tobacco etch virus (TEV), tobacco mosaic virus (TMV), murine hepatitis virus (MHV), influenza B virus, bacteriophages ϕ 6, ϕ X174, m13, λ and T2, and herpes simplex virus 1 (HSV). RNA viruses are shown in red and DNA viruses in blue. Two HIV-1 data points are shown in pink, one obtained from cell culture studies [4] (which is similar to the estimate obtained here using plasma RNA), and the estimate obtained here from PBMC DNA. All other mutation rates were taken from a review [4] except for Q β [63]. Numerical values can be retrieved from these references. Other reverse-transcribing viruses are believed to exhibit mutation rates similar to those of RNA viruses, but these are not shown because previous work did not address the potentially strong contribution of A3 in vivo. Notice that the mutation rate axis is in log-scale, such that the HIV-1 mutation rate is 37 times higher than the second highest rate, and also substantially higher than the rate obtained from plasma RNA or under cell culture conditions. **B.** According to our interpretation, this discrepancy occurs because most A3-edited sequences are lethally mutated and are thus unable to reach the plasma. Therefore, analysis of plasma RNA sequences would lead to a gross underestimation of the actual HIV-1 mutation rate.

doi:10.1371/journal.pbio.1002251.g005

the abundance of stop codons in sequences from plasma would approach that of viral DNA sequences, but our results clearly show that this does not occur. Another potential bias may come from the fact that CD4 cells infected with stop codon-containing, defective proviruses may have a longer lifespan than those infected with nondefective viruses, thus becoming overrepresented in the PBMC population. However, HIV-1 genome integration typically requires that the host cell is activated, because otherwise nucleotide pools are too low to support efficient reverse transcription and nuclear transport of the preintegration complex [56], but activated lymphocytes have short lifespans even if they are not infected [57]. Additionally, it has been estimated that only 1/5 to 1/10 of the HIV-1 DNA is comprised of integrated proviral genomes and thus that most viral DNA is short-lived [58,59]. Therefore, usage of stop codons for inferring the HIV-1 mutation rate should not be a major source of bias. Supporting this view, application of this same method to HCV yielded estimates that are within the accepted range for RNA viruses [53,54], and our inference from plasma RNA sequences is also consistent with this range [4,6–9].

In contrast, our results using DNA from PBMCs provide a much higher mutation rate for HIV-1, which clearly deviates from the accepted range of rates for RNA viruses (Fig 5A). This extremely high mutation rate is essentially driven by A3 editing, the relative contributions of A3G, A3D/F/H, and RT being approximately 100:10:1, respectively. Hypermutation should be largely lethal for the virus, particularly considering the low robustness exhibited by some HIV-1 proteins [60,61]. This is consistent with the observation that 88% of latently integrated proviruses are genetically defective [62]. However, low-edited viruses may be able to form infectious particles and reach the plasma (Fig 5B).

As we have shown, A3 enzymes still contribute approximately 50% of the spontaneous mutations in plasma. Furthermore, since these sequences were obtained by RT-PCR and given that RTs tend to exhibit considerably lower fidelity in vitro than in vivo [35,36], it is likely that

the actual contribution of A3 to sequence diversity in plasma is greater than estimated here. Previous analyses of inpatient viral diversity have suggested a role for A3 in promoting immune escape and drug resistance [17,18], and antiviral therapy failure was found to be more frequent among patients infected with defective *vif* alleles [64]. Our results showing that A3 introduces enormous numbers of mutations in HIV-1 but that the amount of A3 editing varies support such as dual role for A3 as an antiviral factor and a diversity-generating agent. The correlations found with disease progression markers underscore the importance of viral mutation rates for pathogenesis. However, noncausal relationships between viral mutation rates and disease progression may be envisaged. For instance, variants showing different cell tropism may be subject to varying editing rates depending on A3 expression levels in the host cell type. Since cell tropism and disease progression are associated [65,66], this might produce a correlation between the HIV-1 mutation rate and disease progression. Computational analysis of consensus V3-loop sequences suggested CCR-5 coreceptor usage in all patients except R5 and, therefore, we found no evidence that systematic differences in cell tropism among patients may drive the observed correlations.

In conclusion, we have inferred an extremely high mutation rate in HIV-1 patients, which is mainly caused by A3-driven editing of the viral genome. We argue that analysis of sequences from plasma grossly underestimates the HIV-1 mutation rate due to the abundance of lethal mutations that are incompatible with the release of viral particles. A3 proteins have also been shown to mutate hepatitis B virus [67] and nonreverse transcribing DNA viruses such as papillomaviruses [68] or herpesviruses [69]. In addition, the double-stranded RNA-specific adenosine deaminase (ADAR) can edit the genomes of many RNA viruses including measles virus [70], human parainfluenza virus [71], respiratory syncytial virus [72], lymphocytic choriomeningitis virus [73], and Rift Valley fever virus [74]. Future work may elucidate whether intracellular sequences also harbor higher-than-average mutation rates in these viruses, similar to A3-edited HIV-1 sequences. As shown for HIV-1, host-mediated hypermutation of viral genomes can be regarded as an antiviral mechanism, but a downside of this process is that it may contribute to viral genetic diversity and pathogenesis. In our study, we have focused only on normal and rapid progressors, but it remains unclear whether similar results will be observed in other disease progression categories such as long-term nonprogressors or elite controllers.

Methods

Patients

Ten samples from patients were kindly provided by the HIV BioBank integrated in the Spanish AIDS Research Network (RIS) [75], and one sample (P6) was provided by Hospital La Fe (Valencia, Spain). Samples were processed following standard procedures and frozen immediately upon reception. For sample P6, 10 mL of blood was provided by Hospital La Fe (Valencia, Spain), and DNA extraction from PBMC was performed following buffy coat purification. For all samples, approximately 10 million PBMCs were used for DNA extraction using QIAamp DNA Blood Mini Kit (Qiagen). All patients participating in the study gave their informed consent, and protocols were approved by institutional ethics committees. The clinical and epidemiological data provided for patients were included in the cohort of adults with HIV infection of the AIDS Research Network (CoRIS). The program was approved by the Institutional Review Boards of the participating hospitals and centers. The cohort of adults with HIV infection of the AIDS Research Network (CoRIS) is an open, multicenter cohort of patients newly diagnosed with HIV infection at the hospital or treatment center, over 13 years of age, and naïve to antiretroviral treatment. The information is subject to internal quality controls; once every two years, information on 10% of the cohort is audited by an externally contracted agency.

Amplification of Viral DNA by Limiting Dilution PCR

Nearly complete HIV genomes were amplified in three overlapping PCR fragments (named as regions 1, 2, and 3) using different sets of primer described previously [76], and which are provided in S1 Table. High-fidelity limiting dilution PCR was performed using Phusion polymerase (Thermo Scientific) to amplify clonal sequences, setting the dilution of the sample such that the percentage of positive PCRs was about 10%. Primary PCRs were performed at 2 min at 98°C, 35 cycles of 5 s at 98°C, 30 s at 62°C, and 2 min at 72°C, and a final extension of 10 min at 72°C. Secondary amplification was done by nested PCR under the same conditions but with an annealing temperature of 65°C. Positive reactions were picked from the 96-well plates and visually checked in agarose gels for equal concentration. For each patient, fifty clonal PCR products were obtained for each region. PCR products were pooled as indicated, purified using High Pure PCR Product Purification Kit (Roche), and sequenced by Illumina HiSeq2000 using paired-end libraries (Genoscreen, France).

NGS Sequence Mapping and Mutation Calling

Fastq files were cleaned and trimmed using FASTX toolkit version 0.0.14 and dereplicated using Prinseq-lite version 0.20.3 [77]. To obtain the consensus for each patient, 50,000 paired reads were subsampled from each library, pooled, and mapped using Bowtie 2 version 2.2.4 [78] to reference libraries generated from overlapping half-genomes alignments (position 1–5,000 and 4,000–end) of 6 subtype B reference sequences (HXB2, pNL4.3, K03455, AY423387, AY173951, AY331295). Reads mapping to each region were then split using a custom script and the consensus sequence obtained using VICUNA [79] with default settings. Aligning reads to each half genome region was necessary to properly assemble the 5' and 3' UTR regions. Contigs from both regions were then merged using the contigMerger.pl script of V-FAT (Broad institute) to generate the overall consensus for each patient, which were deposited in GenBank (accession numbers KT200348–KT200358). Using these references, reads were mapped using MOSAIK-2.2.3 aligner [80]. V-profiler [81] was then used to call mutations at each codon position, excluding mutations occurring only at the last 10 bases of reads. The codon details output file from V-phaser 2 was then parsed using a custom R script to keep only codons that have <30% low quality reads, and a mutation frequency >6%. Positions with stop codons were then extracted and their occurrence in each mix of five clones estimated using the formula 6%–30% = 1 genome, 31%–50% = 2 genomes, 51%–70% = 3 genomes, and 71%–90% = 4 genomes. In addition, a consensus sequence for each mix (mix consensus) was obtained from the nucleotide frequency output file of V-phaser 2 and an overall consensus for each patient derived from these.

Sanger Sequencing of *nef* Clones

Positive primary PCRs of fragment 3 from libraries showing at least two stop codons per pool as determined by NGS were reamplified, column-purified, and subjected to Sanger sequencing using the reverse PCR primer. Chromatograms were analyzed using the Staden package (version 2.0.0b10). Sequences were converted to Fasta using Egglib software (version 2.1.7), and the number and nature of stop codons determined by visual inspection in Aliview (version 1.17.1). Sequences were deposited in GenBank (accession numbers KT205403–KT205555).

Sequences from Plasma RNA

For *env* gene, two available datasets of full length, codon alignments were downloaded, one of subtype B (http://www.hiv.lanl.gov/content/sequence/HIV/USER_ALIGNMENTS/keele.html) and another of subtype C (http://www.hiv.lanl.gov/content/sequence/HIV/SI_alignments/set5).

[html](#)). These were separated into individual alignments for each patient including the precalculated consensus using a custom script. For *gag*, sequences were downloaded from the HIV database that encompassed the entire gene, included at least four sequences per patient, and were derived by single genome amplification, cloning, or limiting dilution. The *gag* region was then obtained using the GeneCutter tool available from this database, and sequences were codon-aligned using HIV Align with HMM-align option. Sequences from individual patients were then separated, and a consensus calculated using Biostrings R package. For each patient in the three datasets, sequences with frameshifts were removed.

Calculation of A3G, A3D/F/H, and RT Mutation Rates

NSMTs are codons that can mutate to a stop codon via a single nucleotide substitution [53,54]. The presence of these codons was identified in each reference sequence and their abundance calculated by multiplying by the number of sequences for each patient. Based on sequence preference of A3G ($\underline{GG} \rightarrow \underline{AG}$) or A3D/F/H ($\underline{GA} \rightarrow \underline{AA}$), NSMTs were assigned as potential targets for A3G, A3D/F/H, or RT (see [Table 1](#)), with all mutations which do not occur at A3G or A3D/F/H sequence motifs being assigned to the RT category. Subsequently, all stop codons were identified and assigned as an A3G, A3D/F/H, or RT mutation (e.g., TAG codon in TGG is A3G; CAG to TAG is RT). Mutation rates for A3 enzymes were calculated as the number of A3 mutations per A3 NSMT. Mutation rates for RT were calculated similarly but were multiplied by three to account for the fact that only one out of three bases at each NSMT produces an observable stop codon. In all cases, the last 5% of each gene was not considered in order to avoid inclusion of stop codons with reduced effect on protein function. Finally, for three patients, a single PCR fragment had a mutation rate that was <3% of the other 2 PCR fragments. These are likely due to selective amplification of nonhypermutated genomes due to primer mismatches and were therefore not included in the analysis.

A3H Haplotype Determination

PCR was performed on patient DNA to amplify exon 3 of A3H using primers 5-CATGGGA CTGGACGAAACGCA (A3H105F) and 5-TGGGATCCACACAGAAGCCGCA (A3H105R), Phusion high-fidelity polymerase, and 35 cycles. The resulting PCR was directly sequenced to ascertain the presence of a glycine or arginine at amino acid position 105 (reference SNP 139297). For patients homozygous (P6, R7, and R8) or heterozygous (R15) for an arginine at position 105, indicating a potentially active haplotype, PCR was performed to amplify A3H exon 2 using primers 5-GTGGCTTGAGCTGGGGTGA (A3H15F) and 5-CAGAGA GCCCGTGTGGCACC (A3H15R). The PCR product was then cloned, and 5 clones per patient were analyzed for the presence of a deletion at amino acid position 15 (reference SNP 79323350). With the exception of R15, all patients were homozygous for a deletion at position 15, indicating an unstable A3H genotype. R15 was homozygous for the presence of an asparagine at position 15 and hence carried one stable and one unstable allele.

Coreceptor Usage Prediction

V3 loop sequences from the consensus sequence from each patient were analyzed using WebPSSM tool (<http://indra.mullins.microbiol.washington.edu/webpssm/>), and only R5 was predicted to use CXCR4.

Supporting Information

S1 Data. Average mutation rates and clinical information for each patient. Total, A3G, A3D/F/H, low-level A3, and high-level A3 mutation rates are provided. Mutation rates inferred from plasma RNA are also provided. Sequences from PBMC DNA were obtained by Illumina sequencing in this study. Consensus sequences for each patient are available from Genbank accessions KT200348–KT200358. Sequences from plasma RNA are publicly available at http://www.hiv.lanl.gov/content/sequence/HIV/USER_ALIGNMENTS/keele.html and http://www.hiv.lanl.gov/content/sequence/HIV/SI_alignments/set5.html.

(XLSX)

S2 Data. Detailed clinical data. The infection time, viral load, and CD4 count are provided for each patient sample.

(XLSX)

S3 Data. List and location of stop codons in each patient. For each patient (columns), genes, NSMT-containing codons, and the observed number of stop codons are shown.

(XLSX)

S4 Data. List of stop codons found in each sequencing library. For each patient and sequencing library, the total, A3G, A3D/F/H, and RT stops are indicated, and libraries are classified depending on the A3 mutation level (high/low).

(XLSX)

S5 Data. Stop codons in individual *nef* sequences. Each library containing at least two stops was analyzed by sequencing the *nef* gene of each of the five clones (i.e., limiting-dilution PCRs) constituting the library by the Sanger method. The patient, library, and number of stop codons found in each clone are shown. The observed number of clones with zero stops is compared with the number expected under a Poisson model, and a *p*-value is provided for each library. Sequences are available from Genbank accessions KT205403–KT205555.

(XLSX)

S1 Fig. CD4 counts of the 11 patients included in this study. The per-year CD4 count decay rate was obtained by linear regression (dashed lines). The numerical values shown in this Figure are provided in the [S2 Data](#) file.

(TIF)

S2 Fig. Viral load determinations for the 11 patients included in this study. The set-point viral load was obtained by averaging the log load values obtained at least one year postinfection (dashed lines). The numerical values shown in this Figure are provided in the [S2 Data](#) file.

(TIF)

S3 Fig. Distribution of stop codons and NSMTs across HIV-1 genes. Total, A3G, A3D/F/H, and RT stop codons, NSMTs within a sliding window of 50 codons (black skyline), and the average for each gene (red dashed line) are shown. Numerical values can be obtained from the [S3 Data](#) file.

(TIF)

S4 Fig. Distribution of the number of stop codons per library in *nef*. Each column corresponds to a sequencing library showing at least two total *nef* stop-codons from the indicated patient. Each stacked bar shows the number of stop codons found in each clone of that library (clones with no stops are not represented). Asterisks indicate libraries in which mutations were

significantly clustered in a subset of clones. Numerical values are provided in the [S5 Data](#) file. (TIF)

S5 Fig. Distribution of mutation rates across HIV-1 genes in rapid and normal progressors.

The mutation rate within a sliding window of 50 codons (skylines) and the average for each gene (dashed lines) are shown for rapid (red) and normal (blue) progressors. Numerical values can be obtained from the [S3 Data](#) file.

(TIF)

S1 Table. List of primers used for limiting-dilution PCR.

(XLSX)

Acknowledgments

We thank Dr. Cecilio López-Galíndez for useful comments and for primers and PCR conditions, and Alejandro Manzano-Marín for help with sequence analysis. We want to particularly acknowledge the patients in this study for their participation and the HIV BioBank integrated in the Spanish AIDS Research Network and collaborating centers for the generous gifts of clinical samples used in this work. The HIV BioBank, integrated in the Spanish AIDS Research Network, is supported by the Spanish Instituto de Salud Carlos III.

Author Contributions

Conceived and designed the experiments: RS. Performed the experiments: JMC RGe. Analyzed the data: RGe RS. Contributed reagents/materials/analysis tools: JLA. Wrote the paper: RGe RS.

References

1. Perelson AS. Modelling viral and immune system dynamics. *Nat Rev Immunol.* 2002; 2:28–36. PMID: [11905835](#)
2. Fraser C, Lythgoe K, Leventhal GE, Shirreff G, Hollingsworth TD, Alizon S, et al. Virulence and pathogenesis of HIV-1 infection: an evolutionary perspective. *Science.* 2014; 343:1243727. doi: [10.1126/science.1243727](#) PMID: [24653038](#)
3. Smyth RP, Davenport MP, Mak J. The origin of genetic diversity in HIV-1. *Virus Res.* 2012; 169:415–29. doi: [10.1016/j.virusres.2012.06.015](#) PMID: [22728444](#)
4. Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral mutation rates. *J Virol.* 2010; 84:9733–48. doi: [10.1128/JVI.00694-10](#) PMID: [20660197](#)
5. Lauring AS, Frydman J, Andino R. The role of mutational robustness in RNA virus evolution. *Nat Rev Microbiol.* 2013; 11:327–36. doi: [10.1038/nrmicro3003](#) PMID: [23524517](#)
6. Abram ME, Ferris AL, Shao W, Alvord WG, Hughes SH. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *J Virol.* 2010; 84:9864–78. doi: [10.1128/JVI.00915-10](#) PMID: [20660205](#)
7. Mansky LM. Retrovirus mutation rates and their role in genetic variation. *J Gen Virol.* 1998; 79:1337–45. PMID: [9634073](#)
8. Roberts JD, Bebenek K, Kunkel TA. The accuracy of reverse transcriptase from HIV-1. *Science.* 1988; 242:1171–3. PMID: [2460925](#)
9. Ji JP, Loeb LA. Fidelity of HIV-1 reverse transcriptase copying RNA in vitro. *Biochemistry.* 1992; 31:954–8. PMID: [1370910](#)
10. Julias JG, Pathak VK. Deoxyribonucleoside triphosphate pool imbalances in vivo are associated with an increased retroviral mutation rate. *J Virol.* 1998; 72:7941–9. PMID: [9733832](#)
11. Mansky LM, Le Rouzic E, Benichou S, Gajary LC. Influence of reverse transcriptase variants, drugs, and Vpr on human immunodeficiency virus type 1 mutant frequencies. *J Virol.* 2003; 77:2071–80. PMID: [12525642](#)
12. Holtz CM, Mansky LM. Variation of HIV-1 mutation spectra among cell types. *J Virol.* 2013; 87:5296–9. doi: [10.1128/JVI.03576-12](#) PMID: [23449788](#)

13. Desimmie BA, Delviks-Frankenberry KA, Burdick RC, Qi D, Izumi T, Pathak VK. Multiple APOBEC3 restriction factors for HIV-1 and one Vif to rule them all. *J Mol Biol.* 2014; 426:1220–45. doi: [10.1016/j.jmb.2013.10.033](https://doi.org/10.1016/j.jmb.2013.10.033) PMID: [24189052](https://pubmed.ncbi.nlm.nih.gov/24189052/)
14. Moris A, Murray S, Cardinaud S. AID and APOBECs span the gap between innate and adaptive immunity. *Front Microbiol.* 2014; 5:534. doi: [10.3389/fmicb.2014.00534](https://doi.org/10.3389/fmicb.2014.00534) PMID: [25352838](https://pubmed.ncbi.nlm.nih.gov/25352838/)
15. Santa-Marta M, de Brito PM, Godinho-Santos A, Goncalves J. Host Factors and HIV-1 Replication: Clinical Evidence and Potential Therapeutic Approaches. *Front Immunol.* 2013; 4:343. doi: [10.3389/fimmu.2013.00343](https://doi.org/10.3389/fimmu.2013.00343) PMID: [24167505](https://pubmed.ncbi.nlm.nih.gov/24167505/)
16. Armitage AE, Deforche K, Chang CH, Wee E, Kramer B, Welch JJ, et al. APOBEC3G-induced hypermutation of human immunodeficiency virus type-1 is typically a discrete "all or nothing" phenomenon. *PLoS Genet.* 2012; 8:e1002550. doi: [10.1371/journal.pgen.1002550](https://doi.org/10.1371/journal.pgen.1002550) PMID: [22457633](https://pubmed.ncbi.nlm.nih.gov/22457633/)
17. Monajemi M, Woodworth CF, Zipperlen K, Gallant M, Grant MD, Larijani M. Positioning of APOBEC3G/F mutational hotspots in the human immunodeficiency virus genome favors reduced recognition by CD8+ T cells. *PLoS ONE.* 2014; 9:e93428. doi: [10.1371/journal.pone.0093428](https://doi.org/10.1371/journal.pone.0093428) PMID: [24722422](https://pubmed.ncbi.nlm.nih.gov/24722422/)
18. Sadler HA, Stenglein MD, Harris RS, Mansky LM. APOBEC3G contributes to HIV-1 variation through sublethal mutagenesis. *J Virol.* 2010; 84:7396–404. doi: [10.1128/JVI.00056-10](https://doi.org/10.1128/JVI.00056-10) PMID: [20463080](https://pubmed.ncbi.nlm.nih.gov/20463080/)
19. Binka M, Ooms M, Steward M, Simon V. The activity spectrum of Vif from multiple HIV-1 subtypes against APOBEC3G, APOBEC3F, and APOBEC3H. *J Virol.* 2012; 86:49–59. doi: [10.1128/JVI.06082-11](https://doi.org/10.1128/JVI.06082-11) PMID: [22013041](https://pubmed.ncbi.nlm.nih.gov/22013041/)
20. Ooms M, Brayton B, Letko M, Maio SM, Pilcher CD, Hecht FM, et al. HIV-1 Vif adaptation to human APOBEC3H haplotypes. *Cell Host Microbe.* 2013; 14:411–21. doi: [10.1016/j.chom.2013.09.006](https://doi.org/10.1016/j.chom.2013.09.006) PMID: [24139399](https://pubmed.ncbi.nlm.nih.gov/24139399/)
21. Refsland EW, Hultquist JF, Luengas EM, Ikeda T, Shaban NM, Law EK, et al. Natural polymorphisms in human APOBEC3H and HIV-1 Vif combine in primary T lymphocytes to affect viral G-to-A mutation levels and infectivity. *PLoS Genet.* 2014; 10:e1004761. doi: [10.1371/journal.pgen.1004761](https://doi.org/10.1371/journal.pgen.1004761) PMID: [25411794](https://pubmed.ncbi.nlm.nih.gov/25411794/)
22. Wang X, Abudu A, Son S, Dang Y, Venta PJ, Zheng YH. Analysis of human APOBEC3H haplotypes and anti-human immunodeficiency virus type 1 activity. *J Virol.* 2011; 85:3142–52. doi: [10.1128/JVI.02049-10](https://doi.org/10.1128/JVI.02049-10) PMID: [21270145](https://pubmed.ncbi.nlm.nih.gov/21270145/)
23. Amoedo ND, Afonso AO, Cunha SM, Oliveira RH, Machado ES, Soares MA. Expression of APOBEC3G/3F and G-to-A hypermutation levels in HIV-1-infected children with different profiles of disease progression. *PLoS ONE.* 2011; 6:e24118. doi: [10.1371/journal.pone.0024118](https://doi.org/10.1371/journal.pone.0024118) PMID: [21897871](https://pubmed.ncbi.nlm.nih.gov/21897871/)
24. Eyzaguirre LM, Charurat M, Redfield RR, Blattner WA, Carr JK, Sajadi MM. Elevated hypermutation levels in HIV-1 natural viral suppressors. *Virology.* 2013; 443:306–12. doi: [10.1016/j.virol.2013.05.019](https://doi.org/10.1016/j.virol.2013.05.019) PMID: [23791226](https://pubmed.ncbi.nlm.nih.gov/23791226/)
25. Kourteva Y, De PM, Allos T, McMunn C, D'Aquila RT. APOBEC3G expression and hypermutation are inversely associated with human immunodeficiency virus type 1 (HIV-1) burden in vivo. *Virology.* 2012; 430:1–9. doi: [10.1016/j.virol.2012.03.018](https://doi.org/10.1016/j.virol.2012.03.018) PMID: [22579353](https://pubmed.ncbi.nlm.nih.gov/22579353/)
26. Piantadosi A, Humes D, Chohan B, McClelland RS, Overbaugh J. Analysis of the percentage of human immunodeficiency virus type 1 sequences that are hypermutated and markers of disease progression in a longitudinal cohort, including one individual with a partially defective Vif. *J Virol.* 2009; 83:7805–14. doi: [10.1128/JVI.00280-09](https://doi.org/10.1128/JVI.00280-09) PMID: [19494014](https://pubmed.ncbi.nlm.nih.gov/19494014/)
27. Gago S, Elena SF, Flores R, Sanjuán R. Extremely high mutation rate of a hammerhead viroid. *Science.* 2009; 323:1308. doi: [10.1126/science.1169202](https://doi.org/10.1126/science.1169202) PMID: [19265013](https://pubmed.ncbi.nlm.nih.gov/19265013/)
28. Kieffer TL, Kwon P, Nettles RE, Han Y, Ray SC, Siliciano RF. G→A hypermutation in protease and reverse transcriptase regions of human immunodeficiency virus type 1 residing in resting CD4+ T cells in vivo. *J Virol.* 2005; 79:1975–80. PMID: [15650227](https://pubmed.ncbi.nlm.nih.gov/15650227/)
29. Russell RA, Moore MD, Hu WS, Pathak VK. APOBEC3G induces a hypermutation gradient: purifying selection at multiple steps during HIV-1 replication results in levels of G-to-A mutations that are high in DNA, intermediate in cellular viral RNA, and low in virion RNA. *Retrovirology.* 2009; 6:16. doi: [10.1186/1742-4690-6-16](https://doi.org/10.1186/1742-4690-6-16) PMID: [19216784](https://pubmed.ncbi.nlm.nih.gov/19216784/)
30. 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; 105:7552–7. doi: [10.1073/pnas.0802203105](https://doi.org/10.1073/pnas.0802203105) PMID: [18490657](https://pubmed.ncbi.nlm.nih.gov/18490657/)
31. Kravtsov Y, Khrapko K. Single-molecule PCR: an artifact-free PCR approach for the analysis of somatic mutations. *Expert Rev Mol Diagn.* 2005; 5:809–15. PMID: [16149882](https://pubmed.ncbi.nlm.nih.gov/16149882/)
32. Palmer S, Kearney M, Maldarelli F, Halvas EK, Bixby CJ, Bazmi H, et al. Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *J Clin Microbiol.* 2005; 43:406–13. PMID: [15635002](https://pubmed.ncbi.nlm.nih.gov/15635002/)

33. Ramachandran S, Xia GL, Ganova-Raeva LM, Nainan OV, Khudyakov Y. End-point limiting-dilution real-time PCR assay for evaluation of hepatitis C virus quasispecies in serum: performance under optimal and suboptimal conditions. *J Virol Methods*. 2008; 151:217–24. doi: [10.1016/j.jviromet.2008.05.005](https://doi.org/10.1016/j.jviromet.2008.05.005) PMID: [18571738](https://pubmed.ncbi.nlm.nih.gov/18571738/)
34. Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, Keele BF, et al. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. *J Virol*. 2008; 82:3952–70. doi: [10.1128/JVI.02660-07](https://doi.org/10.1128/JVI.02660-07) PMID: [18256145](https://pubmed.ncbi.nlm.nih.gov/18256145/)
35. 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; 69:5087–94. PMID: [7541846](https://pubmed.ncbi.nlm.nih.gov/7541846/)
36. Menéndez-Arias L. Mutation rates and intrinsic fidelity of retroviral reverse transcriptases. *Viruses*. 2009; 1:1137–65. doi: [10.3390/v1031137](https://doi.org/10.3390/v1031137) PMID: [21994586](https://pubmed.ncbi.nlm.nih.gov/21994586/)
37. Armitage AE, Katzourakis A, de OT, Welch JJ, Belshaw R, Bishop KN, et al. Conserved footprints of APOBEC3G on Hypermutated human immunodeficiency virus type 1 and human endogenous retrovirus HERV-K(HML2) sequences. *J Virol*. 2008; 82:8743–61. doi: [10.1128/JVI.00584-08](https://doi.org/10.1128/JVI.00584-08) PMID: [18562517](https://pubmed.ncbi.nlm.nih.gov/18562517/)
38. Kijak GH, Janini LM, Tovanabutra S, Sanders-Buell E, Arroyo MA, Robb ML, et al. Variable contexts and levels of hypermutation in HIV-1 proviral genomes recovered from primary peripheral blood mononuclear cells. *Virology*. 2008; 376:101–11. doi: [10.1016/j.virol.2008.03.017](https://doi.org/10.1016/j.virol.2008.03.017) PMID: [18436274](https://pubmed.ncbi.nlm.nih.gov/18436274/)
39. Langford SE, Ananworanich J, Cooper DA. Predictors of disease progression in HIV infection: a review. *AIDS Res Ther*. 2007; 4:11. PMID: [17502001](https://pubmed.ncbi.nlm.nih.gov/17502001/)
40. Mellors JW, Rinaldo CR Jr., Gupta P, White RM, Todd JA, Kingsley LA. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science*. 1996; 272:1167–70. PMID: [8638160](https://pubmed.ncbi.nlm.nih.gov/8638160/)
41. Ulenga NK, Sarr AD, Hamel D, Sankale JL, Mboup S, Kanki PJ. The level of APOBEC3G (hA3G)-related G-to-A mutations does not correlate with viral load in HIV type 1-infected individuals. *AIDS Res Hum Retroviruses*. 2008; 24:1285–90. doi: [10.1089/aid.2008.0072](https://doi.org/10.1089/aid.2008.0072) PMID: [18851679](https://pubmed.ncbi.nlm.nih.gov/18851679/)
42. Wang H, Lv G, Zhou X, Li Z, Liu X, Yu XF, et al. Requirement of HIV-1 Vif C-terminus for Vif-CBF-ss interaction and assembly of CUL5-containing E3 ligase. *BMC Microbiol*. 2014; 14:290. doi: [10.1186/s12866-014-0290-7](https://doi.org/10.1186/s12866-014-0290-7) PMID: [25424878](https://pubmed.ncbi.nlm.nih.gov/25424878/)
43. Davenport MP, Loh L, Petravic J, Kent SJ. Rates of HIV immune escape and reversion: implications for vaccination. *Trends Microbiol*. 2008; 16:561–6. doi: [10.1016/j.tim.2008.09.001](https://doi.org/10.1016/j.tim.2008.09.001) PMID: [18964018](https://pubmed.ncbi.nlm.nih.gov/18964018/)
44. Korber BT, Letvin NL, Haynes BF. T-cell vaccine strategies for human immunodeficiency virus, the virus with a thousand faces. *J Virol*. 2009; 83:8300–14. doi: [10.1128/JVI.00114-09](https://doi.org/10.1128/JVI.00114-09) PMID: [19439471](https://pubmed.ncbi.nlm.nih.gov/19439471/)
45. Vignuzzi M, Wendt E, Andino R. Engineering attenuated virus vaccines by controlling replication fidelity. *Nat Med*. 2008; 14:154–61. doi: [10.1038/nm1726](https://doi.org/10.1038/nm1726) PMID: [18246077](https://pubmed.ncbi.nlm.nih.gov/18246077/)
46. Holmes EC. *The evolution and emergence of RNA viruses*. Oxford University Press; 2009.
47. Pepin KM, Lass S, Pulliam JR, Read AF, Lloyd-Smith JO. Identifying genetic markers of adaptation for surveillance of viral host jumps. *Nat Rev Microbiol*. 2010; 8:802–13. doi: [10.1038/nrmicro2440](https://doi.org/10.1038/nrmicro2440) PMID: [20938453](https://pubmed.ncbi.nlm.nih.gov/20938453/)
48. Pfeiffer JK, Kirkegaard K. Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. *PLoS Pathog*. 2005; 1:e11. PMID: [16220146](https://pubmed.ncbi.nlm.nih.gov/16220146/)
49. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature*. 2006; 439:344–8. PMID: [16327776](https://pubmed.ncbi.nlm.nih.gov/16327776/)
50. Coffey LL, Beeharry Y, Borderia AV, Blanc H, Vignuzzi M. Arbovirus high fidelity variant loses fitness in mosquitoes and mice. *Proc Natl Acad Sci USA*. 2011; 108:16038–43. doi: [10.1073/pnas.1111650108](https://doi.org/10.1073/pnas.1111650108) PMID: [21896755](https://pubmed.ncbi.nlm.nih.gov/21896755/)
51. Korboukh VK, Lee CA, Acevedo A, Vignuzzi M, Xiao Y, Arnold JJ, et al. RNA virus population diversity, an optimum for maximal fitness and virulence. *J Biol Chem*. 2014; 289:29531–44. doi: [10.1074/jbc.M114.592303](https://doi.org/10.1074/jbc.M114.592303) PMID: [25213864](https://pubmed.ncbi.nlm.nih.gov/25213864/)
52. Meng T, Kwang J. Attenuation of human enterovirus 71 high-replication-fidelity variants in AG129 mice. *J Virol*. 2014; 88:5803–15. doi: [10.1128/JVI.00289-14](https://doi.org/10.1128/JVI.00289-14) PMID: [24623423](https://pubmed.ncbi.nlm.nih.gov/24623423/)
53. Cuevas JM, Gonzalez-Candelas F, Moya A, Sanjuán R. The effect of ribavirin on the mutation rate and spectrum of Hepatitis C virus in vivo. *J Virol*. 2009; 83:5760–4. doi: [10.1128/JVI.00201-09](https://doi.org/10.1128/JVI.00201-09) PMID: [19321623](https://pubmed.ncbi.nlm.nih.gov/19321623/)
54. Ribeiro RM, Li H, Wang S, Stoddard MB, Learn GH, Korber BT, et al. Quantifying the diversification of hepatitis C virus (HCV) during primary infection: estimates of the in vivo mutation rate. *PLoS Pathog*. 2012; 8:e1002881. doi: [10.1371/journal.ppat.1002881](https://doi.org/10.1371/journal.ppat.1002881) PMID: [22927817](https://pubmed.ncbi.nlm.nih.gov/22927817/)

55. Aaskov J, Buzacott K, Thu HM, Lowry K, Holmes EC. Long-term transmission of defective RNA viruses in humans and *Aedes* mosquitoes. *Science*. 2006; 311:236–8. PMID: [16410525](#)
56. Coiras M, López-Huertas MR, Pérez-Olmeda M, Alcamí J. Understanding HIV-1 latency provides clues for the eradication of long-term reservoirs. *Nat Rev Microbiol*. 2009; 7:798–812. doi: [10.1038/nrmicro2223](#) PMID: [19834480](#)
57. Macallan DC, Asquith B, Irvine AJ, Wallace DL, Worth A, Ghattas H, et al. Measurement and modeling of human T cell kinetics. *Eur J Immunol*. 2003; 33:2316–26. PMID: [12884307](#)
58. Koelsch KK, Liu L, Haubrich R, May S, Havlir D, Gunthard HF, et al. Dynamics of total, linear nonintegrated, and integrated HIV-1 DNA in vivo and in vitro. *J Infect Dis*. 2008; 197:411–9. doi: [10.1086/525283](#) PMID: [18248304](#)
59. Murray JM, McBride K, Boesecke C, Bailey M, Amin J, Suzuki K, et al. Integrated HIV DNA accumulates prior to treatment while episomal HIV DNA records ongoing transmission afterwards. *AIDS*. 2012; 26:543–50. doi: [10.1097/QAD.0b013e328350fb3c](#) PMID: [22410637](#)
60. Rihn SJ, Wilson SJ, Loman NJ, Alim M, Bakker SE, Bhella D, et al. Extreme genetic fragility of the HIV-1 capsid. *PLoS Pathog*. 2013; 9:e1003461. doi: [10.1371/journal.ppat.1003461](#) PMID: [23818857](#)
61. Rihn SJ, Hughes J, Wilson SJ, Bieniasz PD. Uneven genetic robustness of HIV-1 integrase. *J Virol*. 2015; 89:552–67. doi: [10.1128/JVI.02451-14](#) PMID: [25339768](#)
62. Ho YC, Shan L, Hosmane NN, Wang J, Laskey SB, Rosenbloom DI, et al. Replication-competent non-induced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell*. 2013; 155:540–51. doi: [10.1016/j.cell.2013.09.020](#) PMID: [24243014](#)
63. Bradwell K, Combe M, Domingo-Calap P, Sanjuán R. Correlation between mutation rate and genome size in riboviruses: mutation rate of bacteriophage Qbeta. *Genetics*. 2013; 195:243–51. doi: [10.1534/genetics.113.154963](#) PMID: [23852383](#)
64. Fourati S, Malet I, Lambert S, Soulie C, Wirlden M, Flandre P, et al. E138K and M184I mutations in HIV-1 reverse transcriptase coemerge as a result of APOBEC3 editing in the absence of drug exposure. *AIDS*. 2012; 26:1619–24. doi: [10.1097/QAD.0b013e3283560703](#) PMID: [22695298](#)
65. Goetz MB, Leduc R, Kostman JR, Labriola AM, Lie Y, Weidler J, et al. Relationship between HIV coreceptor tropism and disease progression in persons with untreated chronic HIV infection. *J Acquir Immune Defic Syndr*. 2009; 50:259–66. doi: [10.1097/QAI.0b013e3181989a8b](#) PMID: [19194318](#)
66. Poveda E, Briz V, Quinones-Mateu M, Soriano V. HIV tropism: diagnostic tools and implications for disease progression and treatment with entry inhibitors. *AIDS*. 2006; 20:1359–67. PMID: [16791010](#)
67. Suspène R, Guetard D, Henry M, Sommer P, Wain-Hobson S, Vartanian JP. Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases in vitro and in vivo. *Proc Natl Acad Sci USA*. 2005; 102:8321–6. PMID: [15919829](#)
68. Vartanian JP, Guetard D, Henry M, Wain-Hobson S. Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. *Science*. 2008; 320:230–3. doi: [10.1126/science.1153201](#) PMID: [18403710](#)
69. Suspène R, Aynaud MM, Koch S, Padeloup D, Labetoulle M, Gaertner B, et al. Genetic editing of herpes simplex virus 1 and Epstein-Barr herpesvirus genomes by human APOBEC3 cytidine deaminases in culture and in vivo. *J Virol*. 2011; 85:7594–602. doi: [10.1128/JVI.00290-11](#) PMID: [21632763](#)
70. Cattaneo R, Schmid A, Eschle D, Baczko K, ter M, V, Billeter MA. Biased hypermutation and other genetic changes in defective measles viruses in human brain infections. *Cell*. 1988; 55:255–65. PMID: [3167982](#)
71. Murphy DG, Dimock K, Kang CY. Numerous transitions in human parainfluenza virus 3 RNA recovered from persistently infected cells. *Virology*. 1991; 181:760–3. PMID: [1849685](#)
72. Martínez I, Melero JA. A model for the generation of multiple A to G transitions in the human respiratory syncytial virus genome: predicted RNA secondary structures as substrates for adenosine deaminases that act on RNA. *J Gen Virol*. 2002; 83:1445–55. PMID: [12029160](#)
73. Zahn RC, Schelp I, Utermohlen O, von LD. A-to-G hypermutation in the genome of lymphocytic choriomeningitis virus. *J Virol*. 2007; 81:457–64. PMID: [17020943](#)
74. Suspène R, Renard M, Henry M, Guetard D, Puyraimond-Zemmour D, Billecocq A, et al. Inverting the natural hydrogen bonding rule to selectively amplify GC-rich ADAR-edited RNAs. *Nucleic Acids Res*. 2008; 36:e72. doi: [10.1093/nar/gkn295](#) PMID: [18515351](#)
75. García-Merino I, de Las Cuevas N, Jiménez JL, Gallego J, Gómez C, Prieto C, et al. The Spanish HIV BioBank: a model of cooperative HIV research. *Retrovirology*. 2009; 6:27–6. doi: [10.1186/1742-4690-6-27](#) PMID: [19272145](#)
76. Sandoñis V, Casado C, Alvaro T, Pernas M, Olivares I, García S, et al. A combination of defective DNA and protective host factors are found in a set of HIV-1 ancestral LTNP. *Virology*. 2009; 391:73–82. doi: [10.1016/j.virol.2009.05.022](#) PMID: [19559455](#)

77. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*. 2011; 27:863–4. doi: [10.1093/bioinformatics/btr026](https://doi.org/10.1093/bioinformatics/btr026) PMID: [21278185](https://pubmed.ncbi.nlm.nih.gov/21278185/)
78. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012; 9:357–9. doi: [10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923) PMID: [22388286](https://pubmed.ncbi.nlm.nih.gov/22388286/)
79. Yang X, Charlebois P, Gnerre S, Coole MG, Lennon NJ, Levin JZ, et al. De novo assembly of highly diverse viral populations. *BMC Genomics*. 2012; 13:475–13. doi: [10.1186/1471-2164-13-475](https://doi.org/10.1186/1471-2164-13-475) PMID: [22974120](https://pubmed.ncbi.nlm.nih.gov/22974120/)
80. Lee WP, Stromberg MP, Ward A, Stewart C, Garrison EP, Marth GT. MOSAIK: a hash-based algorithm for accurate next-generation sequencing short-read mapping. *PLoS ONE*. 2014; 9:e90581. doi: [10.1371/journal.pone.0090581](https://doi.org/10.1371/journal.pone.0090581) PMID: [24599324](https://pubmed.ncbi.nlm.nih.gov/24599324/)
81. Henn MR, Boutwell CL, Charlebois P, Lennon NJ, Power KA, Macalalad AR, et al. Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. *PLoS Pathog*. 2012; 8:e1002529. doi: [10.1371/journal.ppat.1002529](https://doi.org/10.1371/journal.ppat.1002529) PMID: [22412369](https://pubmed.ncbi.nlm.nih.gov/22412369/)