

Perspective

Fifteen Years Later: Hard and Soft Selection Sweeps Confirm a Large Population Number for HIV In Vivo

Igor M. Rouzine^{1*}, John M. Coffin^{2,3}, Leor S. Weinberger¹

1 The Gladstone Institutes, Gladstone Institute of Virology and Immunology, University of California San Francisco, San Francisco, California, United States of America, **2** Tufts University, Sackler School of Biomedical Sciences, Boston, Massachusetts, United States of America, **3** HIV Drug Resistance Program, Center for Cancer Research, National Cancer Institute, Frederick, Maryland, United States of America

Even among RNA viruses, which generally exhibit high evolutionary plasticity due to low fidelity of their RNA polymerases, HIV-1 is second only to HCV for its ability to generate within-host genetic diversity [1]. HIV's rapid generation time leads to this high genetic diversity. The unfortunate consequences of HIV's rapid evolution are resistance to antiretroviral drugs [1], partial escape from immune responses [2–4], the ability to switch tropism for target cells [5], and potential threats to new therapeutic strategies [6,7]. The forces driving and influencing HIV evolution include Darwinian selection, limited population size, linkage, recombination, epistasis, spatial aspects, and dynamic factors (particularly due to the immune response). These factors, and the parameters that define them, can be difficult to discern. One of the most elusive parameters critically important for the rate of evolution in every medically relevant scenario is the “effective population number” (N_{eff}) (Figure 1). By definition, the census population size of HIV is the total number of infectious proviruses integrated into the cellular DNA of an individual at a given time. However, the genetically relevant N_{eff} may differ substantially from the census population size. In this volume of *PLOS Genetics*, Pennings and colleagues [8] use new insights into “hard” and “soft” selective sweeps to estimate the effective population size of HIV.

The search for N_{eff} (and other HIV evolutionary parameters) has gone on for almost two decades, following every turn and hitting each pothole on the eventful road of HIV modeling [9]. The rapidity of resistance to monotherapy (in 1–2 weeks) was explained by the deterministic selection of alleles that preexist therapy in minute quantities [1]. The large numbers of virus-producing cells ($\sim 10^8$) in the lymphoid tissue of experimentally infected macaques seemed to confirm this simple Darwinian selection model [10]. However, the Darwinian view has faced challenges. Tajima's “neutrality test” applied to HIV sequences in untreated patients assumed that selection was neutral and predicted

much smaller “effective” populations, of $N_{\text{eff}} \sim 10^3$ [11]. Since Tajima's approach was designed to detect isolated selective sweeps at one or a few mutant sites—while HIV exhibits hundreds of diverse sites in vivo—two groups re-tested the result. A linkage disequilibrium (LD) test [12] and analysis of the variation in the time to drug resistance [13] arrived at the same value, $N_{\text{eff}} = (5-10) \times 10^5$, for an average patient (with the mutation rate $\sim 10^{-5}$ per base). Such populations are sufficiently large for deterministic selection to dominate, yet not large enough to neglect stochastic effects altogether. The LD test [12] is affected by recombination, and HIV's recombination rate had not been well measured at that time. The recent measurement of 5×10^{-6} crossovers per base per HIV replication cycle in an average untreated individual [14–16] updates N_{eff} to $(1-2) \times 10^5$, not far from the original value. A recent study of the pattern of diversity accumulation in early and late HIV infection confirms the range of N_{eff} [17]. However, all these estimates of N_{eff} are lower bounds.

Pennings et al. [8] continue this quest for an effective population size of HIV using a new method based on a theoretical calculation of the probability of multiple introductions of a beneficial allele at a site before it is fixed in a population [18]. The prediction does not depend on whether mutations are new or result from standing variation prior to therapy. The authors use sequence data obtained from 30 patients who failed suboptimal antiretroviral regi-

mens, including efavirenz [19]—a non-nucleoside reverse transcriptase (RT) inhibitor (NNRTI)—and who exhibited a rise of drug-resistant alleles in RT. The sequence data reveal fixation of two alleles, both corresponding to an amino-acid replacement K103N. Pennings et al.'s analysis focuses on the genetic composition at RT codon 103 and the adjacent 500 nucleotides. Based on the changes in the genetic diversity in this region, 30 fixations are classified into “hard” selective sweeps with a single parental sequence, or “soft” sweeps with multiple parental sequences. Observing that both types of sweep occurred at similar frequencies (also confirmed by observations in other resistance codons), the authors predict $N_{\text{eff}} = 1.5 \times 10^5$, in agreement with the LD test.

Pennings et al. also discuss why “selectively neutral” methods based on synonymous diversity underestimate the population size. It is well known that a selection sweep lowers the diversity at linked sites (hence the term “sweep”) and any method assuming selective neutrality translates lower diversity to smaller N_{eff} . The interesting part is the dynamic component of this effect. Pennings et al. demonstrate that rapid sweeps are followed by long periods when the diversity recovers at the linked sites (for synonymous sites, these periods are very long). From another angle, we can add that selection shortens the time to the common ancestor, which decreases the sequence divergence. The ancestral-tree argument is rather general and also applies

Citation: Rouzine IM, Coffin JM, Weinberger LS (2014) Fifteen Years Later: Hard and Soft Selection Sweeps Confirm a Large Population Number for HIV In Vivo. *PLoS Genet* 10(2): e1004179. doi:10.1371/journal.pgen.1004179

Editor: Christophe Fraser, Imperial College London, United Kingdom

Published: February 20, 2014

Copyright: © 2014 Rouzine et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported through an Alfred P. Sloan Research Fellowship (to LSW). The funder had no role in the preparation of the article.

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

* E-mail: igor.rouzine@gladstone.ucsf.edu

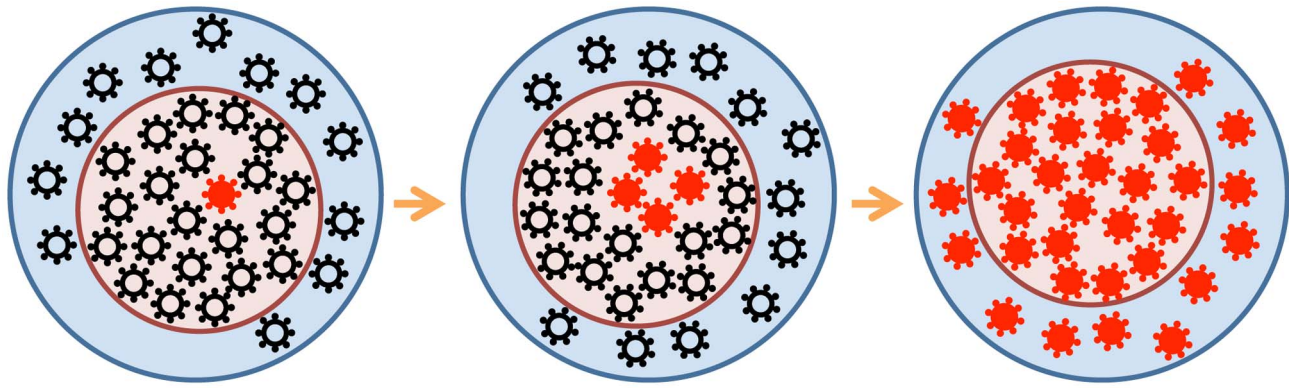


Figure 1. Beneficial viral mutants (red) arise in the “effective” virus subpopulation (N_{eff} , pink circle) and spread gradually to the entire “census” population (blue circle). For a number of reasons (see the text), the effective population may be much smaller than the census population.

doi:10.1371/journal.pgen.1004179.g001

to a large number of linked sites evolving under selection [20–23].

The previous estimates [12,13,17] were lower bounds on N_{eff} . In contrast, the Pennings et al. study puts a number on N_{eff} . However, this number ($N_{\text{eff}} = 1.5 \times 10^5$) raises a question: why is N_{eff} so far below the census population size of 10^8 or more? Pennings et al. offer an elegant explanation of this relatively small N_{eff} in the spirit of the “traveling wave” approach [24–27]. They note that resistant alleles at different sites emerge against different fitness backgrounds. To be fixed, alleles conferring a small benefit must emerge in the most-fit genomes [28,29]; hence, the effective N_{eff} for these alleles is small. Alleles with a larger beneficial effect can explore a larger fraction of population (larger N_{eff}). Conceptually, this idea is quite correct; quantitatively, in the context of drug resistance, some problems arise. For example, the fitness benefit from a resistance mutation (under drug) is almost 100%, while the difference between the fittest and the average genome (in untreated patients) is a modest $\sim 10\%$ [14]. Indeed, the average selection coefficient is quite small, $\sim 0.5\%$ [14,15].

There may be several other reasons for $N_{\text{eff}} < 10^8$, as follows.

- (i) By considering only 500 bases ($\sim 5\%$) of the HIV genome, the study may underestimate the number of genetic backgrounds in which the resistant allele can be observed.
- (ii) N_{eff} is likely to vary in time—similar to viremia, which decays

strongly after the onset of therapy and rebounds after its failure—and the placement of the inferred population size within the therapy time frame is unclear. Specifically, it is unclear from the empirical source [19] whether K103N mutations are generated before therapy (which is likely, considering that the mutation of interest decays very slowly in vivo in untreated patients and therefore has a low mutation cost [30]) or after therapy fails for another reason (see Figure 1 in [19]). In the first scenario, inferred $N_{\text{eff}} = 10^5$ is the pretreatment number. In the second scenario, the pretreatment number must be much higher than 10^5 , since the replicating census population is reduced by a large factor (~ 100) following initiation of therapy.

- (iii) Other factors, such as variation of the population number among patients and the spatial organization of the infected tissue [31] (both neglected in the test), may be relevant. Furthermore, the authors’ calculations rely on the assumption of equal mutation rates for the two resistance mutations analyzed (both transversions). If the underlying rate of AAA to AAC is much greater than that of to AAT, the cited analysis would have underestimated the frequency of soft sweeps, yielding an underestimate of N_{eff} .
- (iv) A significant complicating factor is the presence, in the parent study

[19], of other drugs, particularly the nucleoside RT inhibitors (NRTIs) AZT and 3TC. In some cases, mutations conferring resistance to these drugs may have also contributed to failure (e.g., during the precursor monotherapy; see Figure 1 in [19]), and the requirement for these additional changes would have made the frequency of resistant strains much less than the estimate. For virus that escaped the combination treatment in the absence of NRTI mutations, replication was most likely occurring only in a fraction, or “sanctuary,” of cells that did not receive an inhibitory dose of these drugs. Either or both of these effects would have led to a potentially large underestimate of N_{eff} . Indeed, a recent study of rapid NNRTI resistance, in SIV-infected monkeys treated with efavirenz monotherapy, used an ultrasensitive PCR assay to estimate the pre-therapy level of either K103N mutation as less than 0.0001% [32], implying a total replicating population of $> 10^6$.

For these reasons, the value $N_{\text{eff}} = 1.5 \times 10^5$ obtained in the study of Pennings et al. should probably still be regarded as a lower bound. At the same time, the study solidifies our understanding of HIV evolution as a Darwinian process and leads to important questions regarding the structure of HIV population, which are still waiting for new insights.

References

- Coffin JM (1995) HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science* 267: 483–488.
- Ganusov VV, Goonetilleke N, Liu MK, Ferrari G, Shaw GM, et al. (2011) Fitness costs and diversity of the cytotoxic T lymphocyte (CTL) response determine the rate of CTL escape during acute and chronic phases of HIV infection. *J Virol* 85: 10518–10528.
- Liu Y, McNevin JP, Holte S, McElrath MJ, Mullins JI (2011) Dynamics of viral evolution and

- CTL responses in HIV-1 infection. *PLOS One* 6: e15639. doi:10.1371/journal.pone.0015639
4. Goonetilleke N, Liu MK, Salazar-Gonzalez JF, Ferrari G, Giorgi E, et al. (2009) The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. *J Exp Med* 206: 1253–1272.
 5. Coakley E, Petropoulos CJ, Whitcomb JM (2005) Assessing chemokine co-receptor usage in HIV. *Curr Opin Infect Dis* 18: 9–15.
 6. Rouzine IM, Weinberger LS (2013) Design requirements for interfering particles to maintain co-adaptive stability with HIV-1. *J Virol* 87: 2081–2093.
 7. Metzger VT, Lloyd-Smith JO, Weinberger LS (2011) Autonomous targeting of infectious super-spreaders using engineered transmissible therapies. *PLOS Comput Biol* 7: e1002015. doi:10.1371/journal.pcbi.1002015
 8. Pennings PS, Kryazhimsky S, Wakeley J (2014) Loss and recovery of genetic diversity in adapting populations of HIV. *PLOS Genet* 10: e1004000. doi:10.1371/journal.pgen.1004000
 9. Rouzine IM, Weinberger L (2013) The quantitative theory of within-host viral evolution [review]. *J Stat Mech*: P01009.
 10. Haase AT (1999) Population biology of HIV-1 infection: viral and CD4+ T cell demographics and dynamics in lymphatic tissues. *Annu Rev Immunol* 17: 625–656.
 11. Leigh-Brown AJ (1997) Analysis of HIV-1 env gene sequences reveals evidence for a low effective number in the viral population. *Proc Natl Acad Sci U S A* 94: 1862–1865.
 12. Rouzine IM, Coffin JM (1999) Linkage disequilibrium test implies a large effective population number for HIV in vivo. *Proc Natl Acad Sci U S A* 96: 10758–10763.
 13. Frost SD, Nijhuis M, Schuurman R, Boucher CA, Brown AJ (2000) Evolution of lamivudine resistance in human immunodeficiency virus type 1-infected individuals: the relative roles of drift and selection. *J Virol* 74: 6262–6268.
 14. Batorsky R, Kearney MF, Palmer SE, Maldarelli F, Rouzine IM, et al. (2011) Estimate of effective recombination rate and average selection coefficient for HIV in chronic infection. *Proc Natl Acad Sci U S A* 108: 5661–5666.
 15. Neher RA, Leitner T (2010) Recombination rate and selection strength in HIV intra-patient evolution. *PLOS Comput Biol* 6: e1000660. doi:10.1371/journal.pcbi.1000660
 16. Josefsson L, King MS, Makitalo B, Brannstrom J, Shao W, et al. (2011) Majority of CD4+ T cells from peripheral blood of HIV-1-infected individuals contain only one HIV DNA molecule. *Proc Natl Acad Sci U S A* 108: 11199–11204.
 17. Maldarelli F, Kearney M, Palmer S, Stephens R, Mican J, et al. (2013) HIV populations are large and accumulate high genetic diversity in a nonlinear fashion. *J Virol* 87: 10313–10323.
 18. Pennings PS, Hermisson J (2006) Soft sweeps II—molecular population genetics of adaptation from recurrent mutation or migration. *Mol Biol Evol* 23: 1076–1084.
 19. Bacheiler LT, Anton ED, Kudish P, Baker D, Bunville J, et al. (2000) Human immunodeficiency virus type 1 mutations selected in patients failing efavirenz combination therapy. *Antimicrob Agents Chemother* 44: 2475–2484.
 20. Brunet E, Derrida B, Mueller AH, Munier S (2007) Effect of selection on ancestry: An exactly soluble case and its phenomenological generalization. *Phys Rev E Stat Nonlin Soft Matter Phys* 76: 041104–041101.
 21. Seger J, Smith WA, Perry JJ, Hunn J, Kaliszewska ZA, et al. (2010) Gene genealogies strongly distorted by weakly interfering mutations in constant environments. *Genetics* 184: 529–545.
 22. Rouzine IM, Coffin JM (2010) Multi-site adaptation in the presence of infrequent recombination. *Theor Popul Biol* 77: 189–204.
 23. Neher RA, Hallatschek O (2013) Genealogies of rapidly adapting populations. *Proc Natl Acad Sci U S A* 110: 437–442.
 24. Tsimring LS, Levine H, Kessler D (1996) RNA virus evolution via a fitness-space model. *Phys Rev Lett* 76: 4440–4443.
 25. Rouzine I, Wakeley J, Coffin J (2003) The solitary wave of asexual evolution. *Proc Natl Acad Sci U S A* 100: 587–592.
 26. Desai MM, Fisher DS (2007) Beneficial mutation selection balance and the effect of linkage on positive selection. *Genetics* 176: 1759–1798.
 27. Hallatschek O (2010) The noisy edge of traveling waves. *Proc Natl Acad Sci U S A* 108: 1783–1787.
 28. Neher RA, Shraiman BI, Fisher DS (2010) Rate of adaptation in large sexual populations. *Genetics* 184: 467–481.
 29. Good BH, Rouzine IM, Balick DJ, Hallatschek O, Desai MM (2012) Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proc Natl Acad Sci U S A* 109: 4950–4955.
 30. Palmer S, Boltz V, Martinson N, Maldarelli F, Gray G, et al. (2006) Persistence of nevirapine-resistant HIV-1 in women after single-dose nevirapine therapy for prevention of maternal-to-fetal HIV-1 transmission. *Proc Natl Acad Sci U S A* 103: 7094–7099.
 31. Frost SD, Dumaurier MJ, Wain-Hobson S, Brown AJ (2001) Genetic drift and within-host metapopulation dynamics of HIV-1 infection. *Proc Natl Acad Sci U S A* 98: 6975–6980.
 32. Boltz VF, Ambrose Z, Kearney MF, Shao W, Kewalramani VN, et al. (2012) Ultrasensitive allele-specific PCR reveals rare preexisting drug-resistant variants and a large replicating virus population in macaques infected with a simian immunodeficiency virus containing human immunodeficiency virus reverse transcriptase. *J Virol* 86: 12525–12530.