PLOS GENETICS

VIEWPOINTS

The clarifying role of time series data in the population genetics of HIV

Alison F. Feder [1] *, **Pleuni S. Pennings** [2], **Dmitri A. Petrov** [3]

1 Department of Integrative Biology, University of California, Berkeley, Berkeley, California, United States of America, **2** Department of Biology, San Francisco State University, San Francisco, California, United States of America, **3** Department of Biology, Stanford University, Stanford, California, United States of America

* affeder@berkeley.edu

Abstract

HIV can evolve remarkably quickly in response to antiretroviral therapies and the immune system. This evolution stymies treatment effectiveness and prevents the development of an HIV vaccine. Consequently, there has been a great interest in using population genetics to disentangle the forces that govern the HIV adaptive landscape (selection, drift, mutation, and recombination). Traditional population genetics approaches look at the current state of genetic variation and infer the processes that can generate it. However, because HIV evolves rapidly, we can also sample populations repeatedly over time and watch evolution in action. In this paper, we demonstrate how time series data can bound evolutionary parameters in a way that complements and informs traditional population genetic approaches. Specifically, we focus on our recent paper (Feder et al., 2016, *eLife*), in which we show that, as improved HIV drugs have led to fewer patients failing therapy due to resistance evolution, less genetic diversity has been maintained following the fixation of drug resistance mutations. Because soft sweeps of multiple drug resistance mutations spreading simultaneously have been previously documented in response to the less effective HIV therapies used early in the epidemic, we interpret the maintenance of post-sweep diversity in response to poor therapies as further evidence of soft sweeps and therefore a high population mutation rate (θ) in these intra-patient HIV populations. Because improved drugs resulted in rarer resistance evolution accompanied by lower post-sweep diversity, we suggest that both observations can be explained by decreased population mutation rates and a resultant transition to hard selective sweeps. A recent paper (Harris et al., 2018, *PLOS Genetics*) proposed an alternative interpretation: Diversity maintenance following drug resistance evolution in response to poor therapies may have been driven by recombination during slow, hard selective sweeps of single mutations. Then, if better drugs have led to faster hard selective sweeps of resistance, recombination will have less time to rescue diversity during the sweep, recapitulating the decrease in post-sweep diversity as drugs have improved. In this paper, we use time series data to show that drug resistance evolution during ineffective treatment is very fast, providing new evidence that soft sweeps drove early HIV treatment failure.

Introduction

Diversity signatures from a single time point provide information about evolution

In population genetics, we often sample data at a single time point and build models to explain how the observed patterns of genetic diversity might have been generated. Many species evolve so slowly that researchers cannot directly observe the evolutionary process, and working with snapshot data is the only possibility.

One particularly active field of study uses patterns of genetic diversity to understand the historical and ongoing role of selection. Selection can lead to many characteristic signatures, including the depression of diversity surrounding a recently fixed beneficial allele. As a single adaptive allele rises to high frequency, hitchhiking genetic neighbors also fix in the population in an event known as a **hard selective sweep**. The ratio of selection strength and recombination rate governs the distance on the chromosome from the adaptive site with depressed diversity following a sweep. The most widely used methods for detecting positive selection scan the genome for such sweep signatures [1,2].

More recently, there has been increased attention to cases in which beneficial mutations enter the population rapidly and repeatedly. If a beneficial mutation lands on multiple genetic backgrounds before any single background can sweep, the backgrounds carrying the beneficial mutation will spread concurrently. In such a case, substantially, more genetic diversity will be retained following the fixation of the beneficial mutation, because diverse genetic backgrounds can be brought to intermediate frequency. This is known as a **soft selective sweep**. Because hard and soft sweeps leave different diversity signatures, traditional selective scans do not detect them sensitively, and new approaches have been developed to identify soft sweeps [3,4].

Whether sweeps will be hard or soft depends largely on the rate at which beneficial mutations enter the population each generation, which can be computed by the effective population size (N_e) times the beneficial mutation rate (μ_b). Note that the effective population size must be computed over the timescale relevant to adaptation [5–7]. The product of these terms (called θ) controls how long the population must wait to produce a beneficial mutation with fitness $1 + s$ relative to the nonmutant ($\approx 1/\theta s$ generations). If θ is small ($\theta \ll 1$), adaptation is mutation limited: In this regime, the population must wait between beneficial mutations, and when one does appear and establish, it will spread in the population to the exclusion of the ancestral type, in a hard sweep. However, if θ is large ($\theta \sim O(1)$), beneficial mutations enter the population rapidly, and adaptation is not mutation limited. If 1 or more additional beneficial mutations enter the population before the first one fixes, the result is a soft sweep. It is therefore possible under certain circumstances to determine θ by looking for characteristic signatures of hard and soft sweeps [7–10].

Because of an interdependence through θ, whether a sweep will be hard or soft is also related to the probability of adaptation in a fixed period of time. Assuming that beneficial mutations sweep rapidly, when many mutations are expected to enter a population in quick succession (i.e., θ large), adaptation will be likely in a short period of time, and sweeps will be soft. When the population must wait for a beneficial mutation of large effect (θ small), adaptation is less likely to happen in a fixed period of time, and if it does happen, the sweep is more likely to be hard. Later in this paper, we will demonstrate this principle via simulations. However, the reasoning is much more general: If an independent event is likely to occur in a fixed time, it is likely to happen multiple times.
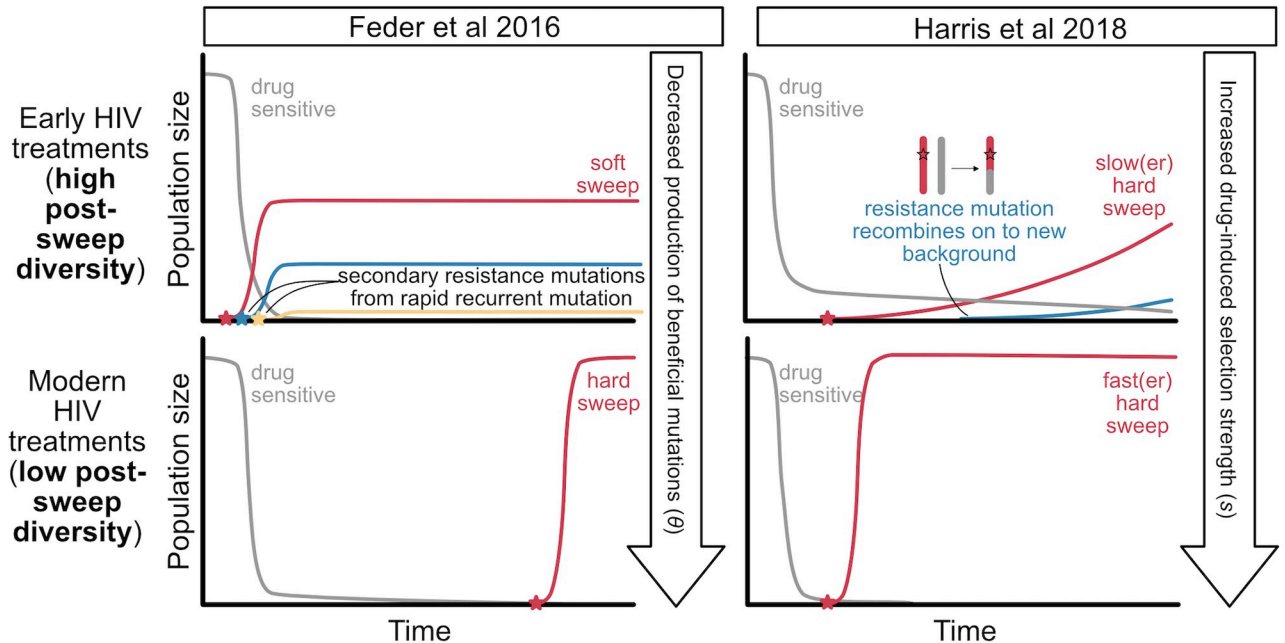
Fig 1. Cartoon of explanations for post-sweep diversity loss patterns observed in Feder et al. [11]. We seek to explain why more diversity is retained following the fixation of drug resistance to early, ineffective therapies than modern, effective ones. **The soft sweeps of recurrent mutations hypothesis** [11] suggest that the primary difference is that θ is high for ineffective drugs, leading to short waiting times between resistance mutations and subsequent soft sweeps (top left), whereas θ is lower for effective drugs, leading to long waiting times between resistance mutations subsequent and hard sweeps (bottom left). **The recombination-rescued diversity hypothesis** [12] suggests that the primary difference is the opportunity for recombination to rescue diversity following a beneficial mutation: Sweeps in response to ineffective drugs are slow and occur after wider population bottlenecks, which gives recombination time to rescue diversity (top right). Sweeps in response to effective drugs are fast and occur after narrow population bottlenecks, so recombination has less opportunity to rescue genetic diversity (bottom right). Stars represent drug resistance mutations entering the population de novo and drug pressure starts at the x-intercept.

https://doi.org/10.1371/journal.pgen.1009050.g001

Diversity signatures in single time point HIV data suggest 2 hypotheses characterizing early HIV evolution

As described above, we expect the probability of evolution in a fixed period of time to be dependent on θ and s. What might this imply about HIV evolution? Better HIV drugs have resulted in the probability of a patient evolving drug resistance in a year on therapy falling sharply from over 80% in the early 1990s to < 10% in the early 2010s. We might therefore expect that as the probability of resistance evolution has decreased, hard sweeps have become more prevalent. In [11], we examined differences in the diversity signatures of intra-patient HIV populations to understand if this decreasing prevalence of drug resistance evolution was indeed accompanied by a transition from soft sweeps to hard sweeps, as might be expected if these better drugs had effectively increased the waiting time for a beneficial mutation. This is illustrated in the left-hand panels of Fig 1.

Consistent with our hypothesis, we found that in response to early, ineffective therapies with a high probability of failure via drug resistance evolution, genetic diversity was largely unaffected by the number of sweeps of resistance mutations within a patient. This is consistent with soft sweep signatures, whereby populations can fix mutations without losing genetic diversity. However, in response to more modern therapies with a low probability of resistance evolution, fixing a drug resistance mutation was associated with a significant loss in genetic diversity, consistent with the signature of a hard selective sweep.

Harris and colleagues [12] have recently proposed that different levels of recombination during a selective sweep can generate the patterns we observed in Feder and colleagues [11] without invoking soft sweeps. They argue that first, if ineffective drugs induce less severe population bottlenecks than those from effective drugs, more diversity will be present that can be rescued via recombination during a selective sweep. Second, if ineffective drugs have longer sweep times than effective drugs, recombination will have more time to recombine genetic diversity on to the sweeping haplotype. This hypothesis is illustrated in the right-hand panels of Fig 1.

Fig 1 shows that under both models, when resistance is acquired in response to ineffective drugs, little population diversity is lost. We discuss 2 explanatory hypotheses: soft sweeps in which recurrent mutation brings multiple genetic backgrounds to intermediate frequency or slow sweeps that allow for the rescue of diversity via recombination (upper row, Fig 1). These hypotheses cannot be readily distinguished using the snapshot data analyzed in Feder and colleagues [11]. However, because HIV is a well-studied system, we can draw upon external sources of information about drug resistance evolution in HIV to understand constraints on the timescales at play, specifically the speed and the probability of adaptation. We argue here that because time series data show that drug resistance evolution in HIV is rapid and predictable, it requires that HIV evolution in response to ineffective drugs is driven by a very rapid input of strongly selected mutations. This means post-sweep diversity is unlikely to be maintained through recombination-driven diversity rescue during hard, slow sweeps of rare weakly adaptive mutations.

HIV time series data suggest rapid adaptation and soft sweeps likely in early HIV evolution

Time series data show just how rapidly HIV can evolve. For example, Fig 2 shows the frequencies of drug resistance mutation M184V, which arises rapidly in 19 of 20 patients who were treated with 3TC monotherapy in the early 1990s [13]. Among the 20 patients, 16 had drug
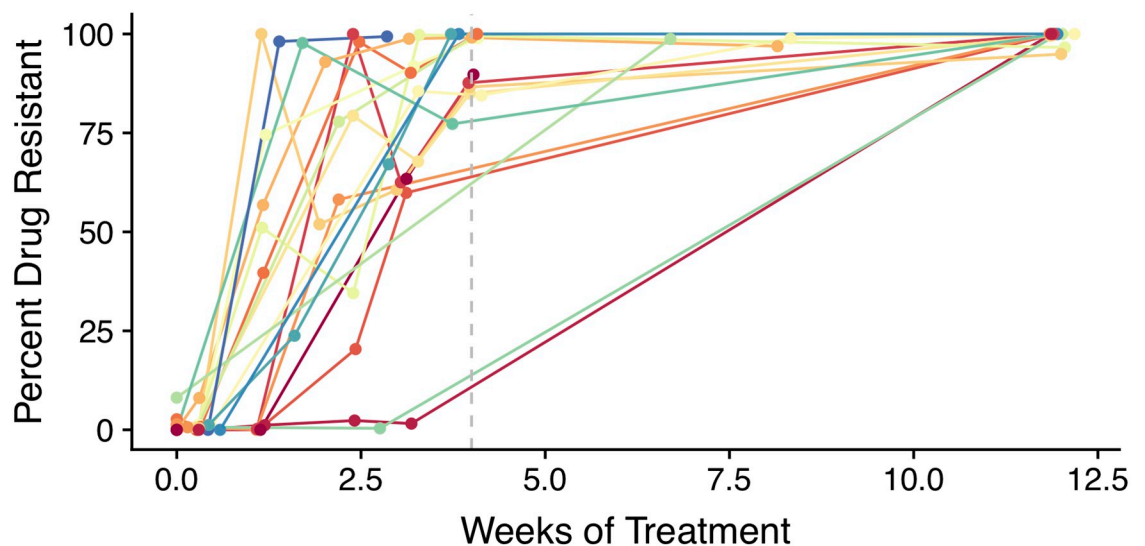

Fig 2. Frequencies of drug resistance mutation M184V in response to 3TC monotherapy in 20 patients. In total, 16/20 have drug resistance observed above frequency 50% in 4 weeks (approximately **28** generations, shown in dashed line). Moreover, 19/20 have drug resistance above 50% in 12 weeks. The patient with no drug resistance is not plotted. Data replotted from Schuurman et al. [13] via WebPlotDigitizer [18].

https://doi.org/10.1371/journal.pgen.1009050.g002

resistance over 50% frequency by week 4 (approximately 28 generations, at 1 HIV generation/day). Note, M184V is costly and unlikely to be present at high frequency as standing genetic variation before treatment [14]. Also note that 4 of the 20 patients had prior exposure to drugs for which M184V rendered partial resistance to therapy, but the percentage of these patients with M184V emerging within 4 weeks was very similar to the cohort as a whole (3/4 patients). This represents just 1 of many examples showing that HIV evolves drug resistance **rapidly** and **predictably** in response to early drug therapies [15–17]. These examples provide information about multiple population genetic parameters.

HIV's ability to adapt **rapidly** depends strongly on the strength of selection, s, for the drug resistant variant under monotherapy. Understanding the sweep time of a beneficial variant ($\approx \log(NS)/s$ generations when $s \ll 1$) allows us to place bounds on s. The **predictability** of HIV's evolution in a short period of time tells us about the number of beneficial mutations entering the population each generation, θ. That mutations appear promptly and consistently across populations suggests that θ must be high.

To quantify what the rapidity and predictably of drug resistance evolution can tell us about HIV evolution in response to monotherapy, we simulated a single locus model of drug resistance during evolution of resistance in HIV for a wide variety of values for θ and s (see Materials and methods for full details). The value of θ is set by the mutation rate, which we set at 10^{-5} per site [19] and on the short-term effective population size relevant for such rapid adaptation. This short-term effective population size is bounded by the census population size of HIV, which can be on the order of 10^9 and possibly much larger [20] (giving us $\theta < 10^4$). The effective population size is smaller than the census size, but how much smaller depends on many biological specifics of the HIV population in question, such as the proportion of infectious particles. Here, we varied θ from 10^{-3} to 10. For each parameter combination, we recorded the probability that beneficial mutations reached frequency $\geq 50\%$ in 30 generations (Fig 3A), as an analog to the data presented in Fig 2. For sweeps to be very predictable on this timescale (80% of cases), both s and θ must be large: s needs to be on the order of 1 and θ must be greater than 0.1. For higher, potentially more realistic values of N (or, more correctly, the short-term N_e relevant to rapid adaptation on the timescales of tens of generation) [5,8], even higher s values are needed for sweeps as fast as observed.

We then asked, conditional on a sweep occurring at each combination of parameters, what is the probability of that sweep being soft (Fig 3B). The parameter combinations in which sweeps are likely (i.e., $> 80\%$ probability of a sweep in 30 generations) are exactly those in which soft sweeps are most likely to occur. Specifically, as θ increases, the probability of both a sweep and a soft sweep increases (Fig 3C).

Note, the probability of a soft sweep depends on both θ and s, and if $s >> \theta$, hard sweeps can occur because the sweep time is less than the waiting time for a new beneficial mutation. For example, when $s = 10$, $\theta = 0.1$, sweeps are likely to occur in 30 generations (93%), but they are hard more often than not (87% of sweeps hard). However, even when s is so large, there is a relatively constrained part of the parameter space with respect to θ that makes both sweeps and hard sweeps very likely. Specifically, θ must be on order 0.1: Indeed, if $\theta < 0.025$, sweeps do not happen with necessity ($< 50\%$ frequency), but if $\theta > 0.25$, sweeps are soft more than 25% of the time. Moreover, if $\theta \approx 0.1$ and s is large, these hard sweeps will retain little diversity anyway and thus are unlikely to be the dominant mode of early adaptation given the observation of diversity retention in Feder and colleagues [11].

Note that standing genetic variation may also contribute to these sweeps. However, for these mutations to be reliably present before the start of treatment, θ must also be high. Although this does depend on the fitness cost of the mutation in the absence of the drug, much more important for predictability of presence is the population mutation rate [21]. In
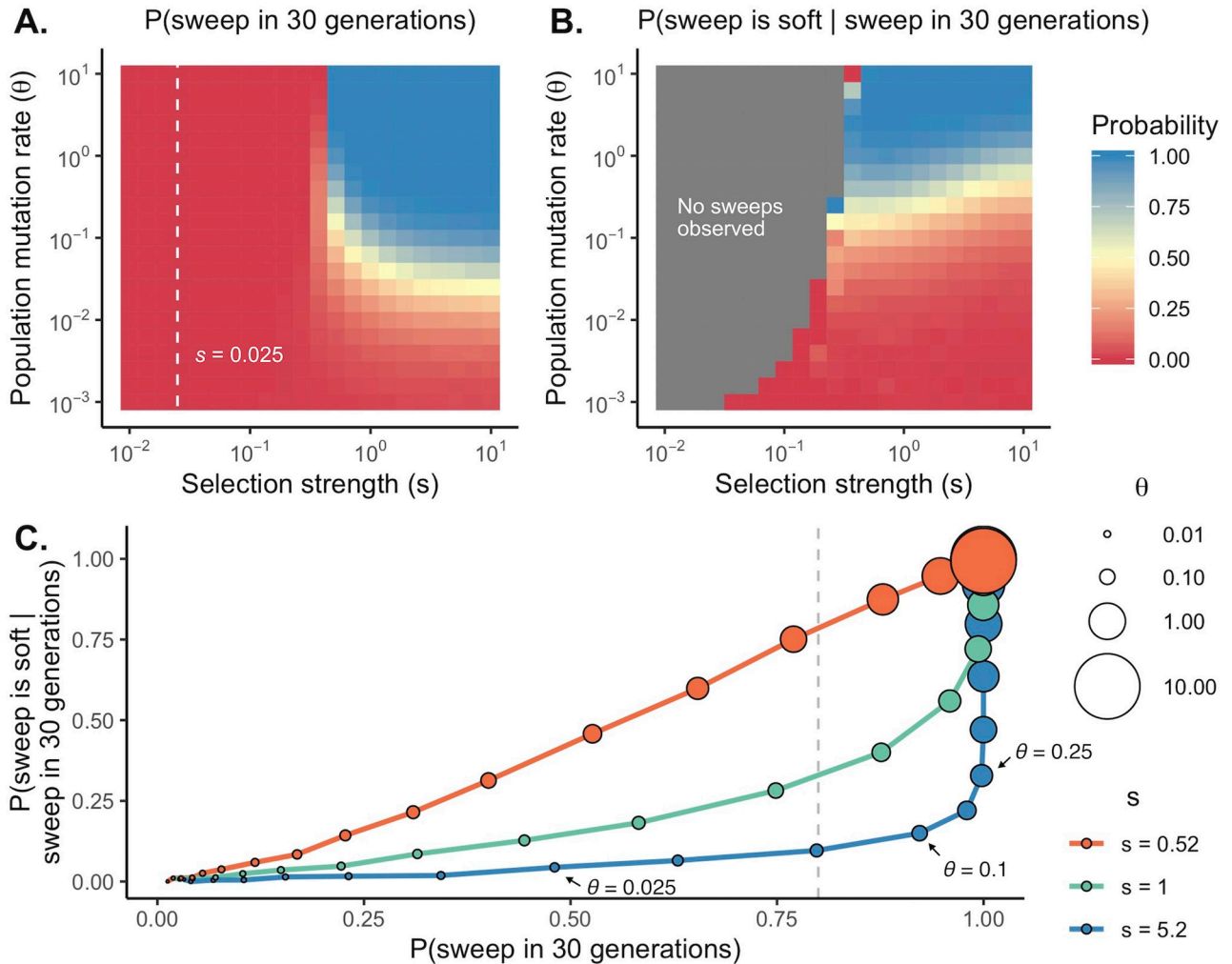
Fig 3. The probability of a sweep and a soft sweep are closely connected. For a variety of combinations of *θ* and *s*, we plot (**A**) the probability of the 1 or more beneficial alleles reaching 50% frequency in 30 generations, and (**B**) we plot the proportion of sweeps that are soft conditional on occurring. Gray squares indicate the beneficial allele(s) did not reach 50% frequency in any simulations, and the white dashed line indicates the selection strength of mutations conferring resistance to inefficient drugs simulated by Harris et al. (*s* = **0.025**). In (**C**), we plot the relationship between how likely a sweep is and its probability of being soft conditional on occurring and find that the 2 are strongly, positively correlated *P*(*sweep*) = **80**% (i.e., [13]) is shown in a dashed line. Point size shows increasing values of *θ* for fixed values of *s* (*s* = (**0.52, 1, 5.2**), shown in red, green, and blue). Other simulation parameters: **10⁴** replicates, *μ* = **10⁻⁵**.

the *θ* regimes in which the beneficial mutation is likely to arise via standing genetic variation, it is also likely to arise multiply via recurrent mutation after the onset of treatment [22].

The simulations above suggest that when sweeps happen on the timescales observed in early HIV drug resistance evolution data, we might expect them to be driven at least in part via soft sweeps of recurrent mutations. Indeed, we have clear, corroborating evidence that soft sweeps happen in HIV treated with ineffective therapies. Pennings, Kryazhimskiy, and Wakeley studied how a strongly beneficial drug resistance mutation to protease inhibitors, K103N, emerges in viruses from treated patient sampled over time. K103N can be created by either an $A \rightarrow T$ or $A \rightarrow C$ mutation at the third nucleotide of the codon. Among a dataset of 17 patients fixing K103N in response to therapy, both $A \rightarrow T$ or $A \rightarrow C$ mutations were simultaneously observed in $> 40\%$ (7/17) of patients. One such patient is shown in Fig 4. At day 0, no drug
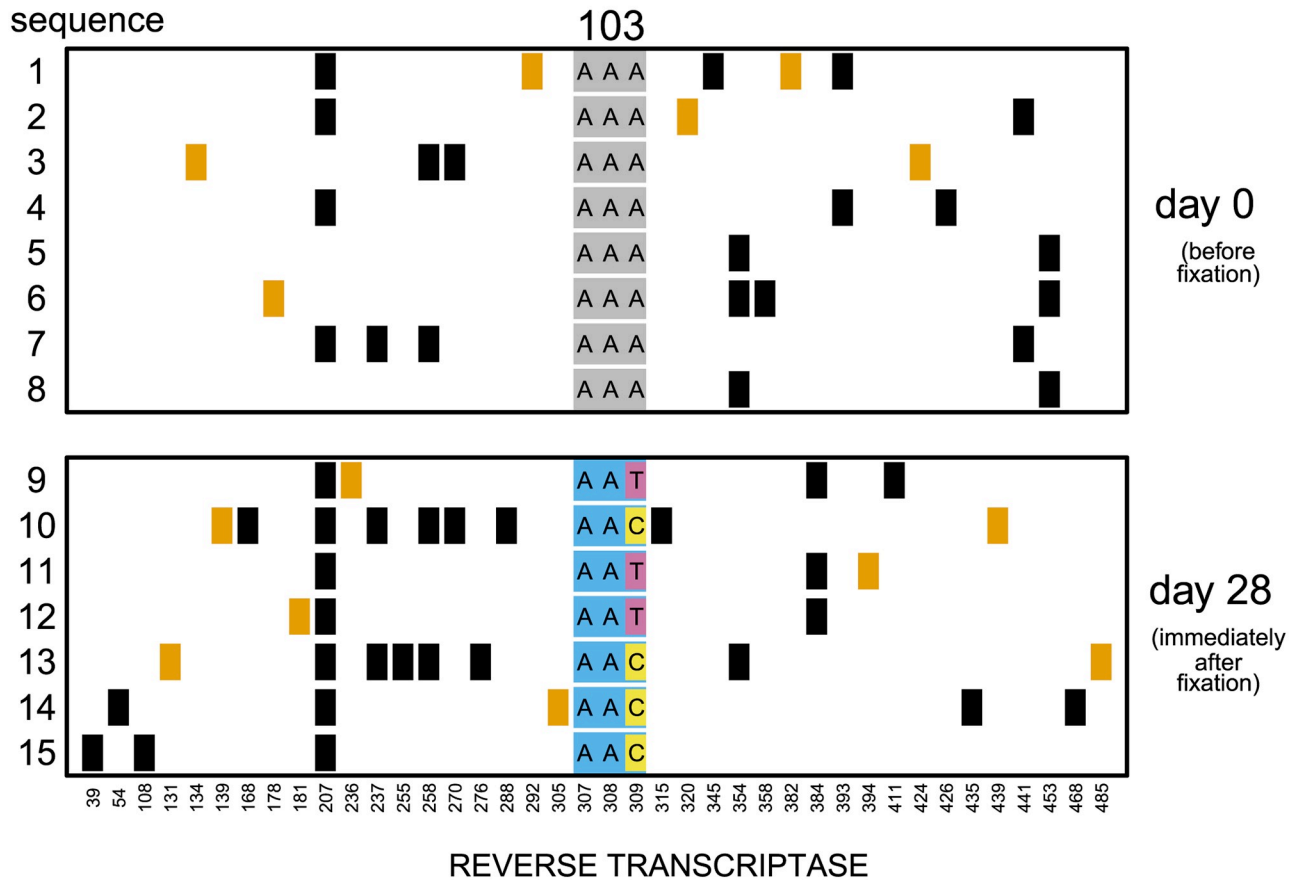
Selective sweep in patient 058

Fig 4. Soft selective sweep from Pennings, Kryazhimskiy, and Wakeley [8]. Viruses were sampled from a patient before (day 0) and after (day 28) the fixation of drug resistance mutation K103N. At site 103, both $A \rightarrow T$ and $A \rightarrow C$ mutations spread at the third codon position. All genetically variable sites are plotted, with black squares representing synonymous substitutions and orange squares representing nonsynonymous substitutions.

https://doi.org/10.1371/journal.pgen.1009050.g004

resistance is present in 8 sequenced viral samples. At day 28, drug resistance is visible in 2 different encodings. Drug resistance must therefore have arisen from at least 2 independent mutations within these patients [8] and could not have been generated via recombination of a single mutation on to different backgrounds. This is a conservative estimate of the number of soft sweeps, because when 2 or more mutations of the same type happen (e.g., $A \rightarrow T$ occurs twice), the soft sweep appears to us as hard. There are extensive other examples of soft sweeps in response to ineffective therapies from studies where multiple sequences are available per individual in HIV-infected humans [13,15,23–29] and Simian-HIV-infected macaques [30–33].

Rapid adaptation allows recombination little opportunity to rescue diversity

Given the parameter constraints on HIV's evolutionary process in response to monotherapy demonstrated above, we return to the hypothesis that slow hard sweeps in response to ineffective drugs allow the rescue of genetic diversity. Recombination during a selective sweep rescues

genetic diversity proportionally to the ratio of the selection strength and the recombination rate (s/r) [34]. The strength of selection therefore determines whether recombination will result in rescue of genetic diversity (assuming r fixed).

Harris and colleagues suggest that if sweeps in response to ineffective drugs are slow ($s = 0.025$), they can fix mutations without a loss of diversity. They compare these sweeps to those in response to simulated modern drugs ($s = 0.1$) which do lose genetic diversity with the sequential fixations of drug resistance mutations [12]. However, time series data suggest that selection in response to monotherapy is stronger than either of these values (Figs 2 and 3). The selection strength simulating monotherapy ($s = 0.025$) is more than an an order of magnitude too low to recapitulate the speed of drug resistance evolution expected from Fig 3. Even the strongest selection strength considered ($s = 0.1$) is not strong enough to lead to any sweeps within 4 weeks. Since even stronger selection observed in data will lead to even more drastic reductions in diversity, the recombination-rescued diversity hypothesis cannot explain how diversity is maintained for HIV responding to ineffective drugs.

We also note that the recombination rate put forward in support of the recombination-rescued diversity hypothesis may be too high. It was estimated from untreated patients where coinfection (and therefore recombination) is more likely than in the simulated bottlenecked HIV populations [35]. Harris and colleagues therefore likely overestimate how much diversity will be maintained due to recombination during a sweep. Because HIV populations treated with monotherapies experience both much stronger selection and potentially weaker recombination than what was assumed by Harris and colleagues [12], abundant recombination during slow hard sweeps does not represent a viable alternative to create the patterns observed in Feder and colleagues [11].

Conclusions

Population genetics is a powerful tool to translate modern day genetic variation into an understanding of the forces acting over evolutionary history. Integrating time series data can provide important calibration on our understanding of the evolutionary process. For example, even small amounts of ancient DNA found in specific places at specific times have transformed the understanding our history as a species (reviewed in [36]). In the case of HIV, detailed time series data help provide strong bounds on the population genetic forces at play. Through tracking changing populations over time, we can see that HIV has high θ and large s for drug resistance mutations in response to ineffective therapies.

Bounding the timescales of these population genetic forces can help us understand whether soft sweeps of multiple recurrent mutations or hard, slow sweeps with recombination can explain how HIV populations became resistant to early therapies without losing genetic diversity. When populations fix beneficial mutations rapidly and predictably, soft sweeps are not only possible, but probable. While slow, hard sweeps can retain diversity after a sweep in general, parameter constraints on s and θ suggest that this is not an appropriate model for HIV.

Materials and methods

We simulated a 1 locus model of drug resistance with a fixed mutation rate of $\mu = 10^{-5}$ [19]. We simulated a range of parameter values with $N = 10^2 - 10^6$ (resulting in $N\mu = \theta = 0.001 - 10$) and $s = 0.01 - 10$. For each parameter combination, we ran 10^4 forward simulation replicates and recorded the percentage of runs in which the sum of all drug resistance mutations reached 50% frequency by generation 30, which we called a sweep. If a sweep occurred, we also recorded whether more than 1 mutation was at frequency above 5%, which we recorded as a soft selective sweep.

Acknowledgments

References

1. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. PLoS Biol. 2006; 4(3):e72. https://doi.org/10.1371/journal.pbio.0040072 PMID: 16494531

2. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, et al. Genome-wide detection and characterization of positive selection in human populations. Nature. 2007; 449(7164):913. https://doi.org/10.1038/nature06250 PMID: 17943131

3. Garud NR, Messer PW, Buzbas EO, Petrov DA. Recent selective sweeps in North American Drosophila melanogaster show signatures of soft sweeps. PLoS Genet. 2015; 11(2):e1005004. https://doi.org/10.1371/journal.pgen.1005004 PMID: 25706129

4. Schrider DR, Kern AD. S/HIC: robust identification of soft and hard sweeps using machine learning. PLoS Genet. 2016; 12(3):e1005928. https://doi.org/10.1371/journal.pgen.1005928 PMID: 26977894

5. Karasov T, Messer PW, Petrov DA. Evidence that adaptation in Drosophila is not limited by mutation at single sites. PLoS Genet. 2010; 6(6):e1000924. https://doi.org/10.1371/journal.pgen.1000924 PMID: 20585551

6. Wilson BA, Petrov DA, Messer PW. Soft selective sweeps in complex demographic scenarios. Genetics. 2014; 198:669–84. https://doi.org/10.1534/genetics.114.165571 PMID: 25060100

7. Khatri BS, Burt A. Robust estimation of recent effective population size from number of independent origins in soft sweeps. bioRxiv. 2018; 472266.

8. Pennings PS, Kryazhimskiy S, Wakeley J. Loss and recovery of genetic diversity in adapting populations of HIV. PLoS Genet. 2014; 10(1):e1004000. https://doi.org/10.1371/journal.pgen.1004000 PMID: 24465214

9. Anderson TJ, Nair S, McDew-White M, Cheeseman IH, Nkhoma S, Bilgic F, et al. Population parameters underlying an ongoing soft sweep in southeast Asian malaria parasites. Mol Biol Evol. 2016; 34 (1):131–44. https://doi.org/10.1093/molbev/msw228 PMID: 28025270

10. Garud NR, Messer PW, Petrov DA. Detection of hard and soft selective sweeps from Drosophila melanogaster population genomic data. bioRxiv. 2020;.

11. Feder AF, Rhee SY, Holmes SP, Shafer RW, Petrov DA, Pennings PS. More effective drugs lead to harder selective sweeps in the evolution of drug resistance in HIV-1. eLife. 2016; 5:e10670. https://doi.org/10.7554/eLife.10670 PMID: 26882502

12. Harris RB, Sackman A, Jensen JD. On the unfounded enthusiasm for soft selective sweeps II: Examining recent evidence from humans, flies, and viruses. PLoS Genet. 2018; 14(12):1–21. https://doi.org/10.1371/journal.pgen.1007859 PMID: 30592709

13. Schuurman R, Nijhuis M, van Leeuwen R, Schipper P, de Jong D, Collis P, et al. Rapid changes in human immunodeficiency virus type 1 RNA load and appearance of drug-resistant virus populations in persons treated with lamivudine (3TC). J Infect Dis. 1995; 171(6):1411–9. https://doi.org/10.1093/infdis/171.6.1411 PMID: 7539472

14. Jain V, Sucupira MC, Bacchetti P, Hartogensis W, Diaz RS, Kallas EG, et al. Differential persistence of transmitted HIV-1 drug resistance mutation classes. J Infect Dis. 2011; 203(8):1174–81. https://doi.org/10.1093/infdis/jiq167 PMID: 21451005

15. Kellam P, Boucher CA, Tijnagel JM, Larder BA. Zidovudine treatment results in the selection of human immunodeficiency virus type 1 variants whose genotypes confer increasing levels of drug resistance. J Gen Virol. 1994; 75(2):341–51. https://doi.org/10.1099/0022-1317-75-2-341 PMID: 7509370

16. Larder BA, Darby G, Richman DD. HIV with reduced sensitivity to zidovudine (AZT) isolated during prolonged therapy. Science. 1989; 243(4899):1731–4. https://doi.org/10.1126/science.2467383 PMID: 2467383

17. Boucher CA, O'Sullivan E, Mulder JW, Ramautarsing C, Kellam P, Darby G, et al. Ordered appearance of zidovudine resistance mutations during treatment of 18 human immunodeficiency virus-positive subjects. J Infect Dis. 1992; 165(1):105–10. https://doi.org/10.1093/infdis/165.1.105 PMID: 1370174

18. Rohatgi A. WebPlotDigitizer. 2011.

19. 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(19):9864–78. https://doi.org/10.1128/JVI.00915-10 PMID: 20660205

20. German Advisory Committee Blood (Arbeitskreis Blut), Subgroup 'Assessment of Pathogens Transmissible by Blood'. Human Immunodeficiency Virus (HIV). Transfus Med Hemother. 2016; 43(3):203–22. https://doi.org/10.1159/000445852 PMID: 27403093

21. Hermisson J, Pennings PS. Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. Methods Ecol Evol. 2017; 8(6):700–16.

22. Barton N. Understanding adaptation in large populations. PLoS Genet. 2010; 6(6):e1000987. https://doi.org/10.1371/journal.pgen.1000987 PMID: 20585547

23. Wainberg MA, Salomon H, Gu Z, Montaner J, Cooley TP, McCaffrey R, et al. Development of HIV-1 resistance to (-) 2'-deoxy-3'-thiacytidine in patients with AIDS or advanced AIDS-related complex. AIDS (London, England). 1995; 9(4):351–7. PMID: 7540846

24. Frost SD, Nijhuis M, Schuurman R, Boucher CA, Brown AJL. Evolution of lamivudine resistance in human immunodeficiency virus type 1-infected individuals: the relative roles of drift and selection. J Virol. 2000; 74(14):6262–8. https://doi.org/10.1128/jvi.74.14.6262-6268.2000 PMID: 10864635

25. Flys T, Nissley DV, Claasen CW, Jones D, Shi C, Guay LA, et al. Sensitive drug-resistance assays reveal long-term persistence of HIV-1 variants with the K103N nevirapine (NVP) resistance mutation in some women and infants after the administration of single-dose NVP: HIVNET 012. J Infect Dis. 2005; 192(1):24–9. https://doi.org/10.1086/430742 PMID: 15942890

26. Lecossier D, Shulman NS, Morand-Joubert L, Shafer RW, Joly V, Zolopa AR, et al. Detection of minority populations of HIV-1 expressing the K103N resistance mutation in patients failing nevirapine. J Acquir Immune Defic Syndr. 2005; 38(1):37–42. https://doi.org/10.1097/00126334-200501010-00007 PMID: 15608522

27. Palmer S, Boltz V, Maldarelli F, Kearney M, Halvas EK, Rock D, et al. Selection and persistence of non-nucleoside reverse transcriptase inhibitor-resistant HIV-1 in patients starting and stopping non-nucleoside therapy. AIDS. 2006; 20(5):701–10. https://doi.org/10.1097/01.aids.0000216370.69066.7f PMID: 16514300

28. Hauser A, Mugenyi K, Kabasinguzi R, Bluethgen K, Kuecherer C, Harms G, et al. Detection and quantification of minor human immunodeficiency virus type 1 variants harboring K103N and Y181C resistance mutations in subtype A and D isolates by allele-specific real-time PCR. Antimicrob Agents Chemother. 2009; 53(7):2965–73. https://doi.org/10.1128/AAC.01672-08 PMID: 19433556

29. Rowley CF, Boutwell CL, Lee EJ, MacLeod IJ, Ribaudo HJ, Essex M, et al. Ultrasensitive detection of minor drug-resistant variants for HIV after nevirapine exposure using allele-specific PCR: clinical significance. AIDS Res Hum Retrovir. 2010; 26(3):293–300. https://doi.org/10.1089/aid.2009.0082 PMID: 20334564

30. Ambrose Z, Palmer S, Boltz VF, Kearney M, Larsen K, Polacino P, et al. Suppression of viremia and evolution of human immunodeficiency virus type 1 drug resistance in a macaque model for antiretroviral therapy. J Virol. 2007; 81(22):12145–55. https://doi.org/10.1128/JVI.01301-07 PMID: 17855539

31. Shao W, Kearney M, Maldarelli F, Mellors JW, Stephens RM, Lifson JD, et al. RT-SHIV subpopulation dynamics in infected macaques during anti-HIV therapy. Retrovirology. 2009; 6(1):101. https://doi.org/10.1186/1742-4690-6-101 PMID: 19889213

32. Boltz VF, Ambrose Z, Kearney MF, Shao W, KewalRamani VN, Maldarelli F, et al. 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. 2012; 86(23):12525–30. https://doi.org/10.1128/JVI.01963-12 PMID: 22933296

33. Feder AF, Kline C, Polacino P, Cottrell M, Kashuba AD, Keele BF, et al. A spatio-temporal assessment of simian/human immunodeficiency virus (SHIV) evolution reveals a highly dynamic process within the host. PLoS Pathog. 2017; 13(5):e1006358. https://doi.org/10.1371/journal.ppat.1006358 PMID: 28542550

34. Gillespie JH. Population genetics: a concise guide. JHU Press; 2010.

35. Neher RA, Leitner T. Recombination Rate and Selection Strength in HIV Intra-patient Evolution. PLoS Comput Biol. 2010; 6(1):1–7. https://doi.org/10.1371/journal.pcbi.1000660 PMID: 20126527

36. Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E. Tracing the peopling of the world through genomics. Nature. 2017; 541(7637):302–10. https://doi.org/10.1038/nature21347 PMID: 28102248