A binary trait model reveals the fitness effects of HIV-1 escape from T cell responses

Yirui Gao[1] **and John P. Barton**[1,2,3,†]

[1]Department of Physics and Astronomy, University of California, Riverside, USA. [2]Department of Computational and Systems Biology, University of Pittsburgh School of Medicine, USA. [3]Department of Physics and Astronomy, University of Pittsburgh, USA. †Address correspondence to: jpbarton@pitt.edu.

Natural selection often acts on multiple traits simultaneously. For example, the virus HIV-1 faces pressure to evade host immunity while also preserving replicative fitness. While past work has studied selection during HIV-1 evolution, it is challenging to quantitatively separate different contributions to fitness. This task is made more difficult because a single mutation can affect both immune escape and replication. Here, we develop an evolutionary model that disentangles the effects of escaping CD8$^+$ T cell-mediated immunity, which we model as a binary trait, from other contributions to fitness. After validation in simulations, we applied this model to study within-host HIV-1 evolution in a clinical data set. We observed strong selection for immune escape, sometimes greatly exceeding past estimates, especially early in infection. Conservative estimates suggest that roughly half of HIV-1 fitness gains during the first months to years of infection can be attributed to T cell escape. Our approach is not limited to HIV-1 or viruses, and could be adapted to study the evolution of quantitative traits in other contexts.

Introduction

Natural selection acts on phenotypic traits, as individuals with traits that are well-adapted to their environment enjoy relative reproductive success. Understanding how selective pressures shape evolution is a major goal of evolutionary biology. Applications include studying the drug resistance in bacteria[1] and patterns of evolution in cancer[2] or viruses[3].

In population genetics, methods have been developed to detect natural selection by examining the genetic diversity of a population and how it changes over time[4–11]. However, inferred selection for a particular allele is not always easy to interpret. A single mutation can affect multiple traits, each of which may be subject to selection. Many phenotypic traits are also polygenic, meaning that they are affected by genetic variation at multiple sites and/or genes[12,13]. How can selection for a particular trait be disentangled from other evolutionary forces?

To approach this question, we developed a model to quantify selection on a binary trait, distinct from other contributions to fitness, using sequence data from an evolving population. Our model is motivated in particular by human immunodeficiency virus (HIV-1) infection. HIV-1 evolves rapidly during infection, and specific mutations allow the virus to escape immune control[14–17]. Past work has shown that the pressure to escape from CD8$^+$ T cell responses is particularly strong[11,18–20], though this is difficult to quantify precisely[20–22]. While CD8$^+$ T cell escape mutations can rapidly sweep through viral populations, the same mutations also often revert when the virus is transmitted to a new individual with a different immune response[18,20,23–25]. This suggests that some escape mutations may harm viral replication in the absence of immune pressure.

After testing our approach in simulations, we applied it to study viral evolution in a data set derived from 13 HIV-1-infected individuals. In this data, we observed strong selection for CD8$^+$ T cell escape independent of other contributions to fitness. This was balanced by a decrease in the estimated fitness effects of specific escape mutations, compared to estimates in a model without the escape trait. Remarkably, we found that fitness gains due to CD8$^+$ T cell escape were responsible for roughly half of the gains in viral fitness estimated from data, especially early in infection. Reversions to clade consensus were also strongly selected, comprising roughly a quarter of viral fitness gains after several years of evolution. Thus, in this data set we found that the vast majority of HIV-1 adaptation was driven by CD8$^+$ T cell escape and reversions.

While the current work focuses on HIV-1 evolution, our framework for estimating the strength of selection on a binary trait is generic. This approach could be applied to study trait evolution in other contexts using temporal sequence data.

Results

CD8$^+$ T cell escape as a binary trait

T cells are a vital part of the adaptive immune system. A subset of T cells that express CD8 on the cell surface specializes in eliminating intracellular pathogens and cancer. Each T cell is equipped with a T cell receptor (TCR) that binds to a specific peptide, roughly 10 amino acids in length, which can be displayed by major histocompatibility complex type I (MHC-I) molecules on the target cell surface. When an activated CD8$^+$ T cell binds to its cognate antigen, this can trigger the release of cytotoxins that kill the target cell. CD8$^+$ T cells that recognize peptides derived from viral proteins can thus play an important role in controlling infection by killing infected cells before they have the chance to release viral particles. The importance of cytotoxic T cells in the context of HIV-1 infection has been appreciated since the early days of HIV-1 research[26].

Here we model the susceptibility of HIV-1 viruses to killing by CD8$^+$ T cell clones as a binary trait. That is, for each HIV-1-specific T cell clone within an individual, each virus is either vulnerable (0) or not (1). This is determined by whether or not viral proteins contain the peptide that a T

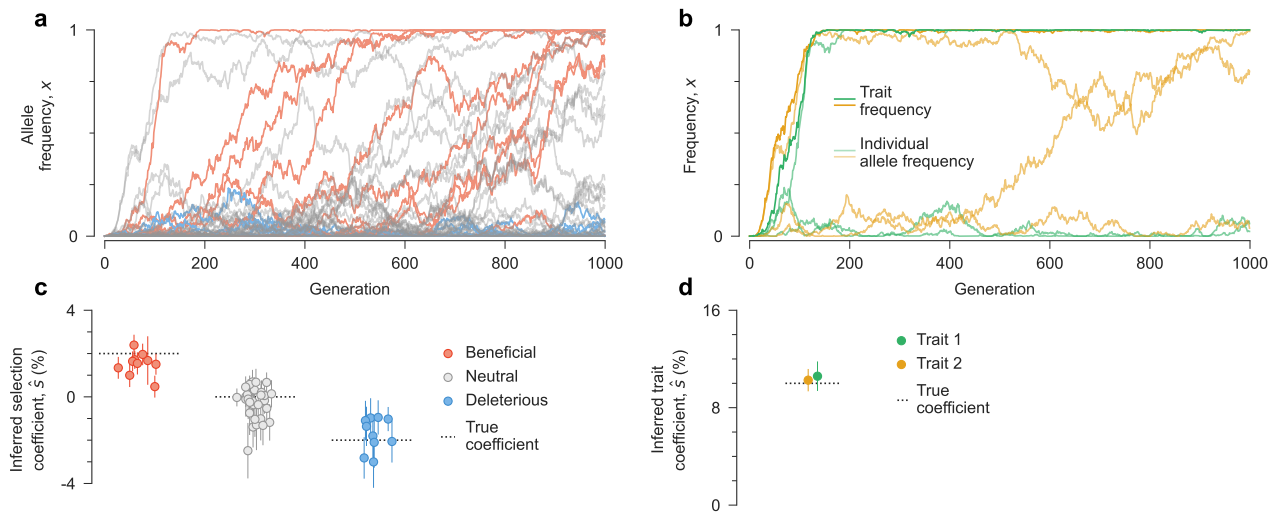


Fig. 1. Recovering selection on individual mutations and traits from temporal genetic data. a, Simulated mutant allele frequency trajectories. **b**, Trait frequencies and their contributing individual mutant allele frequencies in the same simulation. Trait frequencies can approach one even when the frequencies of contributing alleles remain modest. The fitness contributions of individual mutations (**c**) and traits (**d**) that we infer are close to their true values. Simulation parameters: $L = 50$ loci with two alleles at each locus (mutant and WT), ten beneficial mutants with $s = 0.02$, 30 neutral mutants with $s = 0$ and ten deleterious mutants with $s = -0.02$. We consider two binary traits, each with three contributing alleles and trait coefficients $s = 0.1$. Mutation probability per site per generation $\mu = 2 \times 10^{-4}$, recombination probability per site per generation $r = 2 \times 10^{-4}$, population size $N = 10^3$. The initial population contains all WT sequences, evolved over $T = 1000$ generations.

cell binds, which is also referred to as an epitope.

In principle, one may imagine that states of partial immunity are possible: a virus may not express precisely the same epitope that a CD8$^+$ T cell recognizes, but it may express one that is similar enough that it can also be detected by T cells. However, the binding of TCRs to peptide-MHC-I complexes is highly specific. While experiments have shown that some TCRs are cross-reactive to highly similar peptides, most mutations within the epitope abrogate T cell recognition[27–30]. Thus, as we describe below, we modeled HIV-1 susceptibility to CD8$^+$ T cells as "all or nothing," where any nonsynonymous mutation within a T cell epitope results in a loss of immune recognition.

Evolutionary model including selection on binary traits

Our starting point is a Wright-Fisher model, which describes the evolution of a population of N individuals represented by genetic sequences of length L. For simplicity, in the main text we will assume that genetic variants are binary, with each locus consisting of either a wild-type or mutant allele (see Methods for a more detailed model). We write the fitness f_a of an individual with sequence a as

$$f_a = 1 + \sum_i^L s_i g_i^a + \sum_n^\Lambda s_n g_n^a \,. \qquad \textbf{(1)}$$

Here, the s_i are selection coefficients that quantify the fitness effects of the mutant allele at each site i. The g_i^a are indicator variables, with g_i^a equal to one if sequence a has a mutant allele at site i, and zero otherwise. The last term in (1) gives the fitness contribution of the binary trait (CD8$^+$ T cell escape, in our case). Here the n index labels different CD8$^+$ T cell epitopes, with s_n quantifying the fitness effect of escape. We set g_n^a to one if sequence a contains nonsynonymous mutations within epitope n, and zero otherwise. During each generation of replication, the probability that an individual replicates is proportional to its fitness.

In addition to natural selection, our model includes spontaneous mutations, recombination, and genetic drift. We assume a simple probability μ per site per generation for the allele at each site i to change from wild-type (WT) to mutant (and vice versa). This can easily be extended to model different, asymmetric mutation rates between nucleotides in a realistic sequence model (Methods). Similarly, we use a probability r per site per generation for a recombination breakpoint to occur between each pair of neighboring sites in the sequence. Recombination occurs frequently during natural HIV-1 infection, with an estimated probability $r = 1.4 \times 10^{-5}$ (ref.[31]). This is comparable to estimates of HIV-1 mutation rates[32]. Finally, genetic drift due to the finite population size of N individuals adds randomness to the evolutionary dynamics.

Inference of natural selection from temporal genetic data

We developed a method to infer selection, including both selection coefficients s_i for individual mutations and coefficients s_n for binary traits, from sequence data sampled over time. Our approach works by computing the probabilities of different evolutionary histories (i.e., frequencies of individual sequences in the population over time) as a function of the selection coefficients. We then find the selection coefficients that maximize the posterior probability of the data using Bayes' theorem.

To make this inference problem tractable, we used a number of approximations and analytical techniques. First, we considered our evolutionary model in the so-called diffusion limit[33]. In this limit, we assume that the population size N is very large, and that the selection coefficients s_i, trait coefficients s_n, mutation probabilities μ, and recombination probability r are small. We can then derive effective equations

for the dynamics of the allele frequencies x_i and the escape trait frequencies x_n, which are described by a Fokker-Planck equation (Methods). Allele frequencies x_i are defined as the fraction of individuals in the population with a mutant allele at site i, and trait frequencies x_n represent the fraction of individuals with a mutant allele at least one of the epitope sites for epitope n.

While the effective equations are mathematically complicated, their meaning can be clearly interpreted. Natural selection, mutation, and recombination drive deterministic changes in allele and/or trait frequencies, while the finite population size contributes to noise. Alleles and traits are likely to increase in frequency if they appear on sequences that have higher fitness than the average fitness of the population. Mutation introduces new genetic variation, driving allele and trait frequencies away from zero or one. The effects of recombination are more subtle, with no net effect on expected changes in allele frequencies. However, recombination can drive changes in trait frequencies in a way that depends on the arrangement of mutations within the epitope (see Methods).

Finally, we used methods from statistical physics to convert the Fokker-Planck equation into a path integral[11,34–37] (Methods). The path integral quantifies the likelihood of different evolutionary histories or "paths" as a function of the selection coefficients. We can then derive an analytical expression for the vector of selection coefficients $\hat{\boldsymbol{s}}$, including both the s_i and the s_n, that best fits the data (see Methods),

$$\hat{\boldsymbol{s}} = \left(C_{\text{int}} + \gamma I\right)^{-1} \left(\Delta \boldsymbol{x} - \boldsymbol{\mu}_{\text{fl}} - \boldsymbol{R}_{\text{fl}}\right) . \qquad \textbf{(2)}$$

Here C_{int} is the allele/trait frequency covariance matrix integrated over time, which accounts for the speed of evolution and the correlations between mutations/traits. The parameter γ specifies the width of a Gaussian prior distribution for the selection coefficients, and I is the identity matrix. The net change in allele/trait frequencies over the trajectory is $\Delta \boldsymbol{x}$. Finally, $\boldsymbol{\mu}_{\text{fl}}$ and $\boldsymbol{R}_{\text{fl}}$ quantify the expected cumulative change (or flux) in allele/trait frequencies due to mutation and recombination, respectively. Intuitively, (2) attributes net changes in allele/trait frequencies that are not already explained by mutation or recombination to natural selection. The speed of evolution and the genetic background then control how beneficial or deleterious an allele or trait is inferred to be.

Performance on simulated data

To benchmark the performance of our method, we simulated population evolution in the Wright-Fisher model using the fitness function defined in (1) (**Fig. 1**). We considered sequences with $L = 50$ loci (10 beneficial with selection coefficients $s = 2\%$, 30 neutral, and 10 deleterious with selection coefficients $s = -2\%$), with a population size of $N = 1000$ individuals. We also included strong selection ($s = 10\%$) on $\Lambda = 2$ binary traits, each with three contributing alleles that were randomly chosen across the sequence. Individuals thus receive a substantial increase in fitness for mutations at at least one of the trait sites, but multiple mutations confer no additional benefit. **Figure 1** shows that our method is able to distinguish between beneficial, neutral, and deleterious alleles, and to estimate selection on these binary traits.

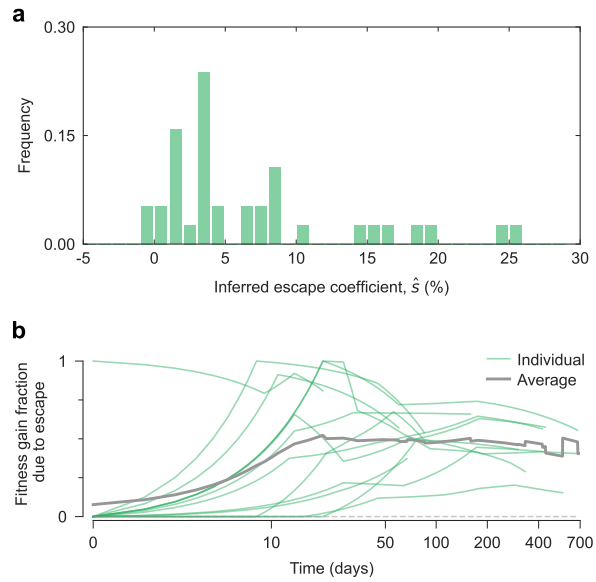


Fig. 2. CD8$^+$ T cell escape is strongly selected and contributes substantially to intrahost HIV-1 fitness gains. a, Distribution of inferred fitness effects of CD8$^+$ T cell escape in HIV-1 patient data. Mutations within T cell epitopes allow infected cells to escape killing by T cells, thus allowing the virus to continue productive replication. **b**, The fraction of the total increase in within-host HIV-1 fitness inferred for each of 13 individuals that is due to CD8$^+$ T cell escape.

Robustness to limited sampling

Real data sets often face significant limitations in both the number of sequences that can be obtained and the frequency of sampling. To test the robustness of our approach to finite sampling, we ran 100 simulations across a wide range of sampling conditions. We assessed our ability to accurately identify selection by computing the average area under the receiver operating characteristic (AUROC) for classifying beneficial/deleterious mutations. We also computed the normalized root mean square error (NRMSE) for the inferred trait coefficients. While inference becomes more difficult with limited sampling, declines in performance occur gradually (**Supplementary Fig. 1**). Our method is particularly robust to limited numbers of sequences, with the ability to identify beneficial or deleterious mutations using as little as 10 sequences per time point.

Selection for immune escape in intrahost HIV-1 evolution

We applied our approach to study HIV-1 evolution using 13 patient data sets[38] (Methods). For each individual, longitudinal HIV-1 half-genome sequences were collected from the time of infection up to several months or years afterward. Early-phase CD8$^+$ T cell epitopes were also carefully verified[30,38], which allowed us to identify putative CD8$^+$ T cell escape mutations. We then estimated the selection coefficients for individual mutations and T cell escape across all 13 individuals.

In order to confidently separate the fitness contribution of escape as a trait from the individual escape mutations, we focused on a well-sampled subset of all T cell epitopes. Specif-

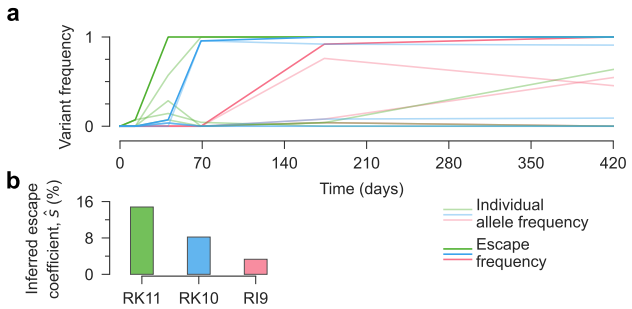


Fig. 3. Estimates of selection for CD8$^+$ T cell escape in a simple example. **a**, Frequency of individual escape mutations and the escape trait for three epitopes in the 3′ half-genome for individual CH470. Escape in each epitope is mediated by multiple mutations. **b**, We infer strong positive selection for escape for all three epitopes.

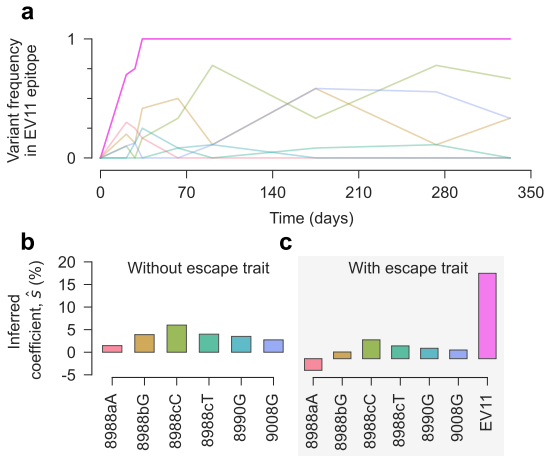


Fig. 4. Inclusion of escape as a trait increases the inferred selective advantage of escape and decreases the apparent advantage of individual escape mutations. **a**, Allele frequencies and escape frequency for epitope EV11 in individual CH131. Escape at this epitope is complex, with contributions from six distinct mutations. **b**, Without treating escape as a trait, all EV11 escape mutations are inferred to be positively selected. **c**, When the selective advantage of escape is considered separately, the inferred selection coefficients for individual mutations become more deleterious. Conversely, the estimated advantage of EV11 escape is substantial, much larger than the inferred selection coefficient for any individual escape mutation without the escape trait.

ically, we estimated escape coefficients for epitopes where at least three putative escape mutations were observed. Out of 71 CD8$^+$ T cell epitopes in this data set, 38 had at least three escape mutations, thus allowing us to estimate their escape coefficients. For the remaining 33 epitopes, we still estimated selection coefficients for individual escape mutations, but we did not attempt to disentangle the contributions of T cell escape versus intrinsic fitness.

Overall, we found very strong selection for CD8$^+$ T cell escape in most epitopes (**Fig. 2a**). Inferred escape coefficients ranged from nearly neutral to highly beneficial ($s \sim 26\%$), with a mean value of 7%. **Figure 3** shows a typical example of HIV-1 evolution to escape from T cell responses across three epitopes in individual CH470. Here, multiple escape mutations are observed for each epitope, making it challenging to infer the strength of selection for escape by looking at single mutations alone.

Next, we quantified how much CD8$^+$ T cell escape contributes to observed changes in HIV-1 fitness *in vivo*. To do this, we computed the fitness of each sequence in each individual and measured the fraction of fitness increase (relative to the transmitted/founder (TF) virus) due to escape coefficients. Here we also included the contribution from escape mutations in epitopes where escape coefficients were not separately estimated. We observed a consistent pattern where escape mutations were rapidly selected within the first weeks of infection, followed by a long, slow decline (**Fig. 2b**). In only 10 days, T cell escape comprises roughly half of HIV-1 fitness gains. This is especially remarkable given that the identified T cell epitopes cover less than 2% of the HIV-1 proteome in these individuals.

Reversions and the intrinsic fitness effects of CD8$^+$ T cell escape mutations

CD8$^+$ T cell escape mutations have been observed to revert after the virus is transmitted to a new host[18,20,23–25], suggesting that escape mutations may be deleterious in the absence of immune pressure. Before explicitly modeling T cell escape as a trait, a large fraction (35%) of escape mutations are inferred to be substantially beneficial ($\hat{s} > 1\%$; **Supplementary Fig. 2**). When escape coefficients are included, the distribution of inferred selection coefficients for escape mutations shifts substantially towards more deleterious values, with a peak near zero. In this case, only 15% of escape mutations are inferred to be substantially beneficial. HIV-1 evolutionary dynamics within each host are therefore best explained by a model where CD8$^+$ T cell escape provides a substantial fitness benefit, which is achieved through escape mutations that may be nearly neutral or even moderately deleterious.

Figure 4 shows an example of the shift in inferred selection coefficients for escape mutations after escape is modeled as a trait. Escape at the EV11 epitope targeted by individual CH131 occurs rapidly. However, there is no single, primary escape mutation. Instead, a set of six escape mutations compete for dominance in the viral population. Without the inclusion of an escape coefficient, the inferred selection coefficients for all of these mutations are positive, with a maximum of around 6%. When we estimate an escape coefficient for EV11, the inferred selection coefficients for individual escape mutations shift to appear more deleterious. In addition, the inferred fitness benefit of escape for the EV11 epitope is around 15%, far higher than the selection coefficient for any individual escape mutation that we estimated in the absence of the escape coefficient.

The intrinsic fitness cost of escape mutations is also highlighted by the benefit of reversions toward the clade consensus sequence. Similar to our previous analysis for CD8$^+$ T cell escape mutations, we computed the fraction of HIV-1 fitness gains in each individual that can be attributed to reversions over time. We observed that reversions contributed substantially to HIV-1 fitness *in vivo*, in a way that grows steadily over time (**Supplementary Fig. 3**). After several years of infection, an average of around 25% of HIV-1 fitness gains relative to the TF virus are due to reversions.

For both T cell escape and reversions, we note that fitness

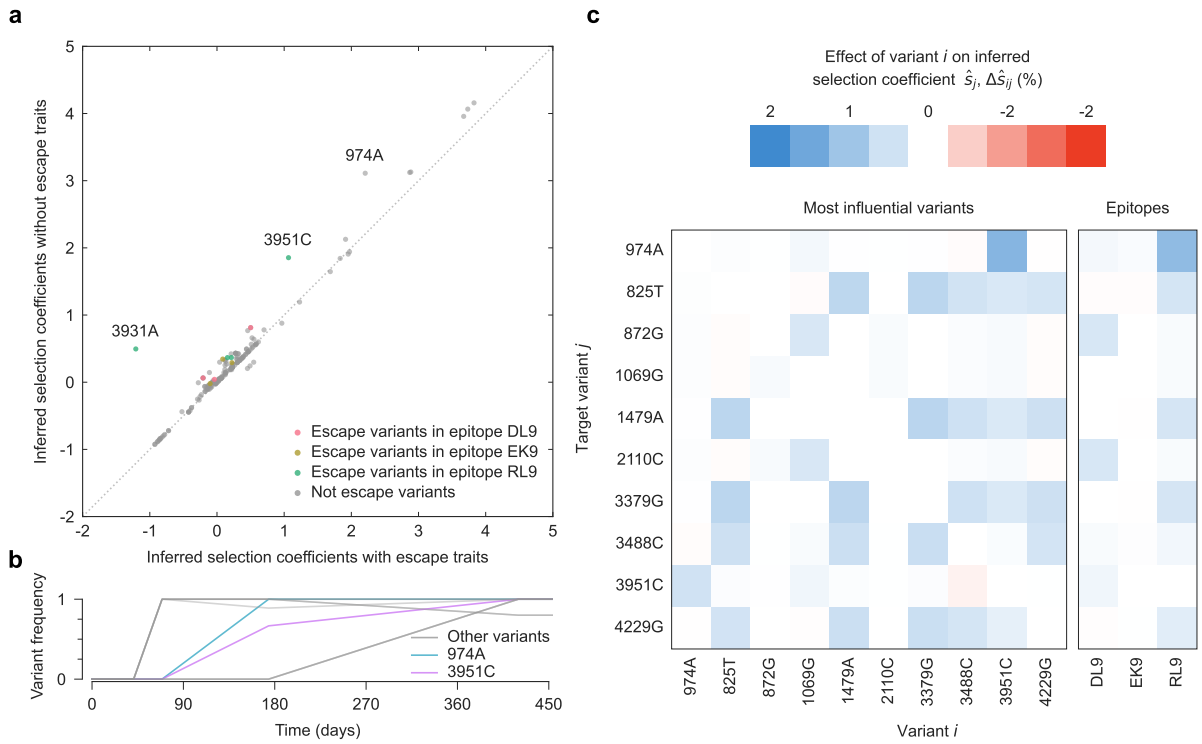


Fig. 5. Inclusion of escape as a trait only strongly affects inferred selection coefficients for escape mutations and alleles that are strongly linked to them. a, Inferred selection coefficients with and without escape terms inferred from 5′ HIV-1 half-genomes from individual CH470. There are three alleles with substantial shifts in inferred selection. Two of these (3931A and 3951C) are escape variants in RL9. **b**, Allele frequencies for variants that appear frequently with 974A, demonstrating strong linkage with 3951C. **c**, Linkage effects on inferred selection coefficients for mutations linked to 974A, and escape coefficients for the three CD8$^+$ T cell epitopes. Effects shown the strong interaction between variant 974A and variant 3951C, with RL9 escape decreasing the inferred selection coefficient for 974A.

contributions are heterogeneously distributed across individual mutations. Inferred selection coefficients for both types of mutations have long tails (**Supplementary Fig. 4**). A few, large selection coefficients thus make large contributions to fitness.

Stability of inferred selection outside T cell epitopes

Changing the fitness model that we use to describe HIV-1 evolution can affect the fitness effects of mutations that we infer. For mutations within CD8$^+$ T cell epitopes, these effects can be quite large, as described above (see also **Supplementary Fig. 2**). Does the inclusion of escape coefficients also affect mutations outside of T cell epitopes? Large changes in these selection coefficients could indicate that the inference process is highly sensitive to specific modeling choices, casting doubt on its reliability.

Overall, the inclusion of escape coefficients shifts the inferred effects of escape mutations, which are typically inferred to be highly beneficial without considering the escape trait, toward much more deleterious values, but without substantially reshaping the overall distribution of selection coefficients (**Supplementary Fig. 4**). **Figure 5** shows a representative example, demonstrating that inferred selection coefficients for individual mutations with and without the inclusion of escape coefficients are very strongly correlated. Most selection coefficients that change significantly are located within CD8$^+$ T cell epitopes, as expected. However, the selection coefficient for one non-epitope mutation, 974A, undergoes a substantial change ($\hat{s} \sim 3\%$ to around 2% before and after the inclusion of escape coefficients, respectively). We found that 974A was strongly linked with 3951C, an escape mutation in the RL9 epitope, which potentially explains this difference.

In this example, we comprehensively explored the effects of linkage disequilibrium (i.e., correlations) between mutations/escape on inferred selection coefficients. Estimated selection coefficients for all mutations and traits are connected due to the integrated allele/trait frequency covariance matrix C_{int} in (2). For each escape mutation and trait j, we computed the selection coefficients that would be inferred if each other mutation/trait i was ignored, denoted $\hat{s}_j^{\setminus i}$ (see Methods). The difference between the coefficient inferred for j using the full data and the coefficient inferred when mutation/trait i is held out yields the effect of linkage with i on j (ref. [11]),

$$\Delta\hat{s}_{ij} = \hat{s}_j - \hat{s}_j^{\setminus i}.$$

Figure 5c shows $\Delta\hat{s}_{ij}$ values for 974A, the escape coefficients for epitopes DL9, EK9, and RL9, and the individual escape mutations that occur within these epitopes. Here we find that linkage disequilibrium with 3951C and RL9 escape mutants is indeed primarily responsible for the shift in selection inferred for 974A. The fitness benefit of escape partly explains the rise to fixation of 974A, reducing its estimated selection coefficient.

Discussion

In this paper, we developed a new method to estimate the fitness effect of a binary trait, jointly with selection coefficients for individual mutations, from temporal genetic data. After validation in simulations, we applied this method to study intrahost HIV-1 evolution in a clinical data set. We modeled $CD8^+$ T cell escape as a binary trait to disentangle the contributions of immune evasion and intrinsic replication to HIV-1 fitness. In this data, HIV-1 evolution was best explained by models with strong selection for T cell escape, but with individual escape mutations that typically did not substantially increase in intrinsic fitness. Reversions to clade consensus were also inferred to be beneficial. The contributions of reversions to HIV-1 fitness gains grew steadily over time.

Our findings are consistent with past observations that have reported strong selection for T cell escape[11,18,20,38]. However, here we found an even stronger benefit for $CD8^+$ T cell escape compared to past work[11]. There are also reasons to believe that the selection for T cell escape may be even stronger than what we have estimated. The Bayesian regularization that we have used controls the inference of large selection coefficients that are not well-supported by data, but it favors spreading the effects of a single large parameter (e.g., a large escape coefficient) into multiple smaller ones (e.g., selection coefficients for individual escape mutations) to maximize the posterior distribution. This is a well-known effect of L_2-norm regularization. This effect could be alleviated in future work through the choice of a different prior distribution, though this could make the analytical derivation of the estimator for selection/trait coefficients more challenging.

The apparent selective advantage of reversions has also been noted in prior work[11,20,25]. However, to our knowledge, the relative contributions of escape and reversions to HIV-1 fitness over time have not previously been quantified. Remarkably, after several years of infection, roughly three quarters of within-host HIV-1 fitness gains can be attributed to $CD8^+$ T cell escape or reversions (**Supplementary Fig. 5**). As discussed by Zanini and collaborators[20], the apparent consistency of selective pressures on HIV-1 across many infected individuals, reflected by strong selection for reversions, could help to explain the success of HIV-1 fitness models that are based on sequence conservation across hosts[30,39–41].

Our study has some limitations due to the constraints on the data available. While our approach is very robust to limited numbers of sequences, long gaps in time between samples limit our ability to fully resolve evolutionary dynamics, leading to less accurate inferences (**Supplementary Fig. 1**). In T cell epitopes, when there are few observed escape mutations it is also statistically challenging to distinguish between the fitness effects of individual mutations and the fitness benefit of escape. This is why we have not attempted to estimate escape coefficients for epitopes with only one or two escape mutations.

The fitness model that we have used could also be extended. While the form of our model was motivated by the study of HIV-1 escape from $CD8^+$ T cell responses, it could apply binary traits in other evolutionary contexts as well. In future work, we aim to extend this approach to estimate selection on more general quantitative traits. Future studies should also consider time-varying selection. In principle, we may expect to observe time-varying selection for T cell escape as immune responses rise and fall. With our current approach, the selection coefficients that we estimate are roughly equivalent to the average selection coefficient when selection changes in time[11,36]. Ultimately, the development of new methods to fit more realistic models to data will improve our quantitative understanding of evolution.

ACKNOWLEDGEMENTS

The work of Y.G. and J.P.B. reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R35GM138233.

AUTHOR CONTRIBUTIONS

All authors contributed to methods development, data analysis, interpretation of results, and writing the paper. J.P.B. supervised the project.

References

1. Davies, J. & Davies, D. Origins and evolution of antibiotic resistance. *Microbiology and molecular biology reviews* **74**, 417–433 (2010).
2. Vendramin, R., Litchfield, K. & Swanton, C. Cancer evolution: Darwin and beyond. *The EMBO Journal* **40**, e108389 (2021).
3. Lauring, A. S. Within-host viral diversity: a window into viral evolution. *Annual Review of Virology* **7**, 63–81 (2020).
4. Illingworth, C. J. & Mustonen, V. Distinguishing driver and passenger mutations in an evolutionary history categorized by interference. *Genetics* **189**, 989–1000 (2011).
5. Feder, A. F., Kryazhimskiy, S. & Plotkin, J. B. Identifying signatures of selection in genetic time series. *Genetics* **196**, 509–522 (2014).
6. Lacerda, M. & Seoighe, C. Population genetics inference for longitudinally-sampled mutants under strong selection. *Genetics* **198**, 1237–1250 (2014).
7. Foll, M., Shim, H. & Jensen, J. D. Wfabc: a w right–f isher abc-based approach for inferring effective population sizes and selection coefficients from time-sampled data. *Molecular ecology resources* **15**, 87–98 (2015).
8. Terhorst, J., Schlötterer, C. & Song, Y. S. Multi-locus analysis of genomic time series data from experimental evolution. *PLoS genetics* **11**, e1005069 (2015).
9. Schraiber, J. G., Evans, S. N. & Slatkin, M. Bayesian inference of natural selection from allele frequency time series. *Genetics* **203**, 493–511 (2016).
10. Taus, T., Futschik, A. & Schlötterer, C. Quantifying selection with pool-seq time series data. *Molecular biology and evolution* **34**, 3023–3034 (2017).
11. Sohail, M. S., Louie, R. H., McKay, M. R. & Barton, J. P. Mpl resolves genetic linkage in fitness inference from complex evolutionary histories. *Nature biotechnology* **39**, 472–479 (2021).
12. Pritchard, J. K., Pickrell, J. K. & Coop, G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current biology* **20**, R208–R215 (2010).
13. Barghi, N., Hermisson, J. & Schlötterer, C. Polygenic adaptation: a unifying framework to understand positive selection. *Nature Reviews Genetics* **21**, 769–781 (2020).
14. Borrow, P. *et al.* Antiviral pressure exerted by hiv-l-specific cytotoxic t lymphocytes (ctls) during primary infection demonstrated by rapid selection of ctl escape virus. *Nature medicine* **3**, 205–211 (1997).
15. Goulder, P. J. *et al.* Late escape from an immunodominant cytotoxic t-lymphocyte response associated with progression to aids. *Nature medicine* **3**, 212–217 (1997).
16. Price, D. A. *et al.* Positive selection of hiv-1 cytotoxic t lymphocyte escape variants during primary infection. *Proceedings of the National Academy of Sciences* **94**, 1890–1895 (1997).
17. McMichael, A. J., Borrow, P., Tomaras, G. D., Goonetilleke, N. & Haynes, B. F. The immune response during acute hiv-1 infection: clues for vaccine development. *Nature Reviews Immunology* **10**, 11–23 (2010).
18. Allen, T. M. *et al.* Selective escape from cd8+ t-cell responses represents a major driving force of human immunodeficiency virus type 1 (hiv-1) sequence diversity and reveals constraints on hiv-1 evolution. *Journal of virology* **79**, 13239–13249 (2005).
19. Liu, Y. *et al.* Selection on the human immunodeficiency virus type 1 proteome following primary infection. *Journal of virology* **80**, 9519–9529 (2006).
20. Zanini, F. *et al.* Population genomics of intrapatient hiv-1 evolution. *Elife* **4**, e11282 (2015).
21. Schneidewind, A. *et al.* Transmission and long-term stability of compensated cd8 escape mutations. *Journal of virology* **83**, 3993–3997 (2009).
22. Lythgoe, K. A. & Fraser, C. New insights into the evolutionary rate of hiv-1 at the within-host and epidemiological levels. *Proceedings of the Royal Society B: Biological Sciences* **279**, 3367–3375 (2012).

23. Leslie, A. *et al.* Hiv evolution: Ctl escape mutation and reversion after transmission. *Nature medicine* **10**, 282–289 (2004).
24. Friedrich, T. C. *et al.* Reversion of ctl escape–variant immunodeficiency viruses in vivo. *Nature medicine* **10**, 275–281 (2004).
25. Li, B. *et al.* Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution. *Journal of virology* **81**, 193–201 (2007).
26. Walker, B. & McMichael, A. The t-cell response to hiv. *Cold Spring Harbor perspectives in medicine* a007054 (2012).
27. Lee, J. K. *et al.* T cell cross-reactivity and conformational changes during tcr engagement. *The Journal of experimental medicine* **200**, 1455–1466 (2004).
28. Huseby, E. S. *et al.* How the t cell repertoire becomes peptide and mhc specific. *Cell* **122**, 247–260 (2005).
29. Huseby, E. S., Crawford, F., White, J., Marrack, P. & Kappler, J. W. Interface-disrupting amino acids establish specificity between t cell receptors and complexes of major histocompatibility complex and peptide. *Nature immunology* **7**, 1191–1199 (2006).
30. Barton, J. P. *et al.* Relative rate and location of intra-host hiv evolution to evade cellular immunity are predictable. *Nature communications* **7**, 11660 (2016).
31. Neher, R. A. & Leitner, T. Recombination rate and selection strength in hiv intra-patient evolution. *PLoS computational biology* **6**, e1000660 (2010).
32. Zanini, F., Puller, V., Brodin, J., Albert, J. & Neher, R. A. In vivo mutation rates and the landscape of fitness costs of hiv-1. *Virus evolution* **3**, vex003 (2017).
33. Ewens, W. J. *Mathematical population genetics: theoretical introduction*, vol. 27 (Springer, 2004).
34. Mustonen, V. & Lässig, M. Fitness flux and ubiquity of adaptive evolution. *Proceedings of the National Academy of Sciences* **107**, 4248–4253 (2010).
35. Sohail, M. S., Louie, R. H., Hong, Z., Barton, J. P. & McKay, M. R. Inferring epistasis from genetic time-series data. *Molecular biology and evolution* **39**, msac199 (2022).
36. Lee, B. *et al.* Inferring effects of mutations on sars-cov-2 transmission from genomic surveillance data. *medRxiv* 2021–12 (2022).
37. Shimagaki, K. & Barton, J. P. Bézier interpolation improves the inference of dynamical models from data. *Physical Review E* **107**, 024116 (2023).
38. Liu, M. K. *et al.* Vertical t cell immunodominance and epitope entropy determine hiv-1 escape. *The Journal of clinical investigation* **123** (2012).
39. Ferguson, A. L. *et al.* Translating hiv sequences into quantitative fitness landscapes predicts viral vulnerabilities for rational immunogen design. *Immunity* **38**, 606–617 (2013).
40. Mann, J. K. *et al.* The fitness landscape of hiv-1 gag: advanced modeling approaches and validation of model predictions by in vitro testing. *PLoS computational biology* **10**, e1003776 (2014).
41. Louie, R. H., Kaczorowski, K. J., Barton, J. P., Chakraborty, A. K. & McKay, M. R. Fitness landscape of the human immunodeficiency virus envelope protein that is targeted by antibodies. *Proceedings of the National Academy of Sciences* **115**, E564–E573 (2018).

Methods

Evolutionary model with selection on binary traits

In our study, model the evolution of a population of N individuals subject to recombination, mutation, and natural selection, following the Wright-Fisher (WF) model. For simplicity, we first describe a model with only two alleles per site, wild-type (WT) or mutant (MT). Individuals possess a genetic sequence length of L, resulting in a total of $M = 2^L$ genotypes. We further assume that there exist Λ binary traits, which depend on the presence or absence of mutant alleles at specific sites, and which are also subject to selection. For clarity, we will use $i, j, \ldots$ indices to denote the different loci or sites in the sequence, $n, m, \ldots$ are used for trait indices, and $a, b, \ldots$ represent genotype indices. Finally, $t_1, t_2, \ldots, t_K$ are used as generation (time) indices. The first two sets of indices are presented as subscripts, the genotype indices as superscripts, and the generation indices are displayed within parentheses to indicate quantities that vary with time.

Let $n_a(t_k)$ represent the number of individuals with genotype a at generation t_k, with $z_a(t_k) = n_a(t_k)/N$ the frequency of genotype a at generation t_k. The vector $z(t_k) = (z_a(t_k), z_b(t_k), \ldots, z_M(t_k))$ describes the state of the population at generation t_k. In our model, the probability of recombination occurring per site per generation is denoted by r. After recombination, the mean frequency of genotype a at generation t_{k+1}, is

$$y_a(t_{k+1}) = (1-r)^{L-1} z_a(t_k) + (1-(1-r)^{L-1})\psi_a(t_k)\,. \tag{3}$$

The term $(1-r)^{L-1}$ gives the likelihood of an individual not experiencing recombination, and $\psi_a(t_k)$ denotes the probability that the random recombination of any two individuals within the population results in an offspring of genotype a. This includes scenarios where both parent and offspring share the same genotype a.

After recombination, the mean frequency of each genotype in the next generation $p_a(t_k)$ is shaped by selection and mutation,

$$p_a(z(t_k)) = \frac{y_a(t_k) f_a + \mu \sum\limits_{b|d(a,b)=1} [y_b(t_k) f_b - y_a(t_k) f_a]}{\sum_{b=1}^{M} y_b(t_k) f_b}\,. \tag{4}$$

In this case, all alleles and loci share an equal mutation rate μ. For simplicity, we assume that each sequence undergoes at most one mutation per generation, given the exceedingly low mutation rate μ. The notation $b|d(a,b) = 1$ indicates that genotypes a and b differ by just a single mutation.

The fitness of genotype a, denoted as f_a, is given by

$$f_a = 1 + \sum_{i}^{L} s_i g_i^a + \sum_{n}^{\Lambda} s_n g_n^a\,.$$

Contributions to fitness come from the effects of individual alleles (quantified by *selection coefficients* s_i) and traits (*trait coefficients* s_n). For the former, the impact of mutations is cumulative. In other words, the effects of mutant alleles at different sites add together. In contrast, the number of mutant alleles within a trait group does not affect fitness, so long as it is one or more. Thus, multiple mutations within the trait sites do not lead to higher or lower fitness. However, the fitness effects of multiple traits are considered additive.

Path integral likelihood

Under WF dynamics, the probability of observing genotype frequencies $z(t_{k+1})$ at generation t_{k+1}, given genotype frequencies of $z(t_k)$= $(z_a(t_k), z_b(t_k), \ldots, z_M(t_k))$ at generation t_k, is

$$P(z(t_{k+1})|z(t_k)) = N! \prod_{a=1}^{M} \frac{(p_a(t_{k+1}))^{N z_a(t_{k+1})}}{(N z_a(t_{k+1}))!}\,. \tag{5}$$

Consequently, the likelihood that the genotype frequency vector follows a specific evolutionary trajectory, or "path", $\boldsymbol{z} = (z(t_1), z(t_2), \cdots, z(t_K)$, is

$$P(\boldsymbol{z}|z(t_0)) = \prod_{k=0}^{K-1} P(z(t_{k+1})|z(t_k))\,,$$

conditioned upon the initial state $z(t_0)$. This expression is difficult to work with directly, so we use several approximations to make our analysis more tractable.

To simplify this expression, we will project dynamics to the level of individual mutant loci i and trait groups n, instead of genotypes a. We denote the frequency of mutant alleles at site i in the population as x_i, and the frequency of individuals with one or more mutant alleles in trait group n as x_n. Additionally, the frequency of paired mutant alleles x_{ij} is used to describe the correlation between different mutations. The formulas for these are as follows:

$$\begin{aligned} x_i &= \sum_{a}^{M} g_i^a z_a\,, \\ x_{ij} &= \sum_{a}^{M} g_i^a g_j^a z_a\,, \\ x_n &= \sum_{a}^{M} g_n^a z_a = \sum_{a}^{M} \left[1 - \prod_{i \in n} (1 - g_i^a)\right] z_a \end{aligned} \tag{6}$$

Here the term g_i^a indicates whether genotype a contains a mutant allele at locus i, with the wild type (WT) having $g_i^a = 0$ and the mutant type (MT) having $g_i^a = 1$. Similarly, g_n^a specifies whether trait group n contains a mutant allele. If all loci in trait group n of genotype a are WT, then $g_n^a = 0$; however, if there is at least one mutant allele in the trait group, meaning that at least one locus within the trait group has $g_i^a = 1$, then $g_n^a = 1$.

It is worth noting that, in most cases, trait groups can be regarded as a special type of locus. Consequently, we utilize the new subscripts $\boldsymbol{i}, \boldsymbol{j}$ to represent generic loci — encompassing both individual loci and trait groups — to distinguish them from individual loci i, j. Taking the pair allele frequency as an example, $x_{\boldsymbol{ij}}$ will represent not only the correlation between different individual loci but also between an individual locus and a trait group, as well as between different trait groups.

Next we consider the dynamics of the mutant allele frequencies (and trait frequencies) in the diffusion limit[42]. We assume that the population size $N \to \infty$ and that the selection coefficients, mutation rate, and recombination rate are all small ($\mathcal{O}(1/N)$). In this limit, applying methods from statistical physics, the probability of an evolutionary trajectory $\boldsymbol{x} = (x(t_1), x(t_2), \ldots, x(t_K))$ can be quantified using a path integral (see refs.[43–45] for more details on this approach)

$$P(\boldsymbol{x}|x(t_0)) \propto \exp\left[-\frac{N}{2}\boldsymbol{S}(\boldsymbol{x})\right], \tag{7}$$

$$\boldsymbol{S}(\boldsymbol{x}) = \sum_{k=0}^{K-1}\frac{1}{\Delta t_k}\sum_{\boldsymbol{i},\boldsymbol{j}=1}^{L+\Lambda}[\Delta x_{\boldsymbol{i}}(t_k) - \Delta t_k D_{\boldsymbol{i}}(t_k)] \times \left[C_{\boldsymbol{ij}}(x(t_k))\right]^{-1}\left[\Delta x_{\boldsymbol{j}}(t_k) - \Delta t_k D_{\boldsymbol{j}}(t_k)\right].$$

Here $\Delta t_k = t_{k+1} - t_k$ and $\Delta x_{\boldsymbol{i}}(t_k) = x_{\boldsymbol{i}}(t_{k+1}) - x_{\boldsymbol{i}}(t_k)$. $\boldsymbol{S}(\boldsymbol{x}(t_k))$ is referred to as the action in physics. In this expression, trait terms are considered as special individual loci, thus the total length of frequencies is the length of generic loci, which is $L + \Lambda$ (binary case). In statistical physics, $D_{\boldsymbol{i}}(t_k)$ and $C_{\boldsymbol{ij}}(t_k)/N$ are commonly referred to as the drift vector and the diffusion matrix respectively, which will be discussed in the following section. To prevent confusion, here we note that the drift vector quantifies the effects of natural selection, mutation, and recombination, which affect the expected change in allele frequencies, and *not* genetic drift.

The drift vector and diffusion matrix

In this section, unless otherwise specified, the time parameter for all physical quantities is assumed to be t_k, and the corresponding time indices will be omitted. We first begin with the diffusion matrix $C_{\boldsymbol{ij}}/N$, which is the same for both allele and trait frequencies

$$\frac{C_{\boldsymbol{ij}}}{N} = \langle (x_{\boldsymbol{i}}(t_{k+1}) - x_{\boldsymbol{i}}(t_k))(x_{\boldsymbol{j}}(t_{k+1}) - x_{\boldsymbol{j}}(t_k))\rangle \approx \frac{1}{N}\begin{cases} x_{\boldsymbol{i}}(1-x_{\boldsymbol{i}}) & \boldsymbol{i} = \boldsymbol{j}, \\ x_{\boldsymbol{ij}} - x_{\boldsymbol{i}}x_{\boldsymbol{j}} & \boldsymbol{i} \neq \boldsymbol{j}. \end{cases}$$

The expectation in the first line is taken over the WF model dynamics, which yields the second line in the limit that N is large and terms of $\mathcal{O}(1/N^2)$ are omitted. Here $\boldsymbol{i}, \boldsymbol{j}$ can be used in both individual and trait terms. $x_{\boldsymbol{i}}$ is the mutant frequencies at generic loci $\boldsymbol{i}$, while $x_{\boldsymbol{ij}}$ is the frequency of individuals with mutations at both generic loci $\boldsymbol{i}$ and $\boldsymbol{j}$ (see (6)). The diffusion matrix quantifies the amount of "noise" in changes in allele/trait frequencies due to finite population size N, i.e., genetic drift.

The drift vector describes the expected change in mutant allele frequencies in time due to selection, mutation, and recombination. For the trait frequencies, this term is especially complex.

Using (3) and (4) and dropping infinitesimals smaller than $\mathcal{O}(1/N)$, we find

$$\begin{aligned} D_{\boldsymbol{i}} = \langle x_{\boldsymbol{i}}(t_{k+1}) - x_{\boldsymbol{i}}(t_k)\rangle &= \sum_a g_{\boldsymbol{i}}^a (p_a(z) - z_a) \\ &\approx \sum_a^M g_{\boldsymbol{i}}^a \Bigg\{ \Big[\sum_j^L g_j^a s_j z_a - \sum_b^M \sum_j^L g_j^b s_j z_a z_b\Big] \\ &+ \Big[\sum_n^\Lambda g_n^a s_n z_a - \sum_b^M \sum_n^\Lambda g_n^b s_n z_a z_b\Big] \\ &+ \Big[\mu \sum_{b|d(a,b)=1}(z_b - z_a)\Big] + \Big[r(L-1)(\psi_a - z_a)\Big]\Bigg\}. \end{aligned}$$

Here we have four terms, which describe the evolutionary forces of selection for individual mutant alleles, selection on binary traits, mutation, and recombination, respectively.

Contributions due to natural selection can be written in the same way for both mutant alleles i and traits n (which, as a reminder, we group under a general bold-faced index $\boldsymbol{i}$). Using (6) and the expression for pair allele/trait frequencies $x_{\boldsymbol{ij}} = \sum_a^M g_{\boldsymbol{i}}^a g_{\boldsymbol{j}}^a z_a$, these terms become

$$\sum_j^L (x_{\boldsymbol{i}j} - x_{\boldsymbol{i}}x_j)s_j + \sum_n^\Lambda (x_{\boldsymbol{i}n} - x_{\boldsymbol{i}}x_n)s_n = \sum_{\boldsymbol{j}}^{L+\Lambda} C_{\boldsymbol{ij}}s_{\boldsymbol{j}}.$$

The mutation and recombination terms are different for mutant alleles and for traits. Here, we start with the mutation term for mutant alleles, which can be simplified as

$$\begin{aligned} &\mu \sum_a^M g_i^a \sum_{b|d(a,b)=1}(z_b - z_a) \\ =&\mu \sum_a^M \Big[\sum_{\substack{b|d(a,b)=1 \\ g_i^a = g_i^b}}(g_i^b z_b - g_i^a z_a) + \sum_{\substack{b|d(a,b)=1 \\ g_i^a \neq g_i^b}}((1-g_i^b)z_b - g_i^a z_a)\Big] \\ =& \quad 0 \quad + \mu \sum_a^M ((1-g_i^a z_a) - g_i^a z_a) \\ =&\mu(1-2x_i). \end{aligned} \tag{8}$$

For the first term, $b|d(a,b) = 1$, $g_i^a = g_i^b$ means that a and b have the same allele at locus i, but they differ by one mutation at some other locus. If we sum over every genotype, these terms make no contribution to the change in mutant allele frequency i because every pair (a,b) has one conjugate pair (b,a). The second term is not symmetrical, but every a only has one b that can satisfy the conditions $b|d(a,b) = 1$, $g_i^a \neq g_i^b$ (i.e., the genotype with a MT allele at site i flipped to WT, or vice versa). Thus, we can use a to replace b and its summation.

Physically, (8) expresses the total flux due to mutation, which is the probability of mutation that can change the allele at locus i from WT to MT (flux in) minus the probability from MT to WT (flux out). The probability of MT at locus i is the mutant allele frequency. Since the state for the allele

is binary, the sum of MT frequency and WT frequency is 1. Thus, we can easily simplify the expression as

$$\mu \sum_{a}^{M} g_i^a \sum_{b|d(a,b)=1} (z_b - z_a) = \mu(1-x_i) - \mu x_i = \mu(1-2x_i)\,.$$

Next, we will consider the recombination term for mutant alleles. Similar to the mutation term, this represents the recombination flux for locus i. However, the expected mutant allele frequency change due to recombination alone is always zero. This is because, for every case in which a sequence without a mutant allele recombines with a sequence that has a mutant allele such that the recombinant sequence has the mutant allele (thus increasing the mutant allele frequency), there is another case with the MT and WT sequences switched (thus decreasing mutant allele frequency by the same amount) which occurs with the same probability. Thus, this term vanishes by symmetry.

In total, then, the drift vector for mutant frequencies i is

$$D_i = \sum_{\boldsymbol{j}}^{L+\Lambda} C_{i\boldsymbol{j}} s_{\boldsymbol{j}} + \mu\,(1-2x_i)\,.$$

Now we consider the mutation and recombination terms for traits. These terms are more complex because the conditions that can change the trait are different from the ones for a single locus, and generally depend on the state of all alleles that contribute to the trait. For example, mutation at one site within a trait group n will change the allele from WT to MT (or vice versa), but it may not change the state for the trait as a whole if there are other mutations among the trait group.

Let us consider the mutation term for traits, which we can expand as

$$\begin{aligned}
&\mu \sum_{a}^{M} g_n^a \sum_{b|d(a,b)=1} (z_b - z_a) \\
=&\mu \sum_{a}^{M} \Big[\sum_{\substack{b|d(a,b)=1 \\ g_n^a = g_n^b}} (g_n^b z_b - g_n^a z_a) + \sum_{\substack{b|d(a,b)=1 \\ g_n^a \neq g_n^b}} ((1-g_n^b) z_b - g_n^a z_a) \Big]\,.
\end{aligned}$$

As in (8), the first term is also zero. However, the second mutation term is different. In this case, genotypes a and b that satisfy $b|d(a,b)=1$, $g_n^a \neq g_n^b$ do not exhibit a one-to-one correspondence: if all alleles in the trait group for a are WT, for example, then there are many b can satisfy these conditions and the number of b is the length of the trait group n. Alternately, if genotype a contains more than one mutant allele in the trait group, then it cannot be changed to WT within a single mutation.

To address this issue, we introduce a new variable, denoted as y_n^i, which is the frequency of genotypes that contain only one mutation in trait group n, and the mutation is in locus i. This can be written as

$$y_n^i = \sum_{a}^{M} g_n^a g_i^a \prod_{j \neq i, j \in n} (1 - g_j^a) z_a\,.$$

With this new variable, we can rewrite the mutation term for traits. Physically, to change the trait value (starting from WT), every locus among the trait group needs to be considered, as any mutation in these loci can change the state of the trait group from WT to MT. Conversely, mutation can only change the trait value from MT to WT if it affects genotypes that have one mutation among the trait group (i.e., by converting the single mutant allele to WT). Thus,

$$\mu \sum_{a}^{M} g_n^a \sum_{b|d(a,b)=1} (z_b - z_a) = \mu \sum_{i \in n} ((1-x_n) - y_n^i)\,.$$

Finally, we turn to the recombination term for traits. Here we start by building physical intuition for the result. In each generation of the evolutionary process, recombination occurs numerous times. For this issue, we need to focus solely on the recombination that can alter the state of a trait group. In particular, we must enumerate the probabilities of recombination events that change traits *and* for which the reverse process (i.e., switching the order of the recombination partner sequences, which has the same probability) does not change the trait in the opposite way, thus yielding no net expected change in the trait frequency.

We use the index k to represent a recombination breakpoint. To change a trait's value, k must be located after the first site in the trait group and before the last. Since we have assumed that the recombination probability per site per generation r is small, we will only consider cases where a single recombination breakpoint falls within the trait sites, but the analysis below could be generalized to allow for two or more breakpoints. If recombination occurs between two sequences, where one is WT for trait group n and the other is MT before and after a breakpoint k for the same trait, then all the recombinants will be MT for trait group n. For such sequences, recombination can change WT to MT for the trait, but not MT to WT. This is the only scenario in which recombination leads to a net *increase* in trait frequency.

In the opposite direction, consider recombination between two sequences, where one is WT for trait n before a breakpoint k and MT after, and the second sequence is MT before k and WT after. Thus, both sequences are MT for trait n, but through recombination a WT sequence can be produced. This is the only recombination process that leads to a net *decrease* in trait frequency.

Collecting these two terms together, we arrive at the following expression for the net expected change in trait frequency due to recombination,

$$\begin{aligned}
&\sum_{a}^{M} g_n^a r(L-1)(\psi_a - z_a) \\
&= \psi_{a_in}^n - \psi_{a_out}^n \\
&= r \sum_{k \in n} P_{W,W}^{k,n} P_{M,M}^{k,n} - r \sum_{k \in n} P_{W,M}^{k,n} P_{M,W}^{k,n}\,.
\end{aligned}$$

Here $P_{M,W}^{k,n}$ represents the frequency of sequences that have at least one mutant allele in the trait group n on or before site

k, and all WT alleles in the same trait group after k. In total, then, the drift vector for traits is

$$\begin{aligned} D_n &= \langle x_n(t_{k+1}) - x_n(t_k) \rangle \\ &= \sum_{j}^{L+\Lambda} C_{n\boldsymbol{j}} s_{\boldsymbol{j}} + \mu \sum_{i \in n} ((1-x_n) - y_n^i) \\ &+ r \left(\sum_{k \in n} P_{W,W}^{k,n} P_{M,M}^{k,n} - \sum_{k \in n} P_{W,M}^{k,n} P_{M,W}^{k,n} \right). \end{aligned}$$

Marginal path likelihood (MPL) inference

To find the fitness effects of alleles/traits that best fit the data, we will attempt to find parameters $s_{\boldsymbol{i}}$ that maximize the likelihood of the data. Since the likelihood P is proportional to action $\boldsymbol{S}$ (see (7)), the unknown selection coefficients s_i and trait coefficients s_n that can maximize action $\boldsymbol{S}$ will maximize the path probability P.

To control our estimates, we also introduce a Gaussian prior distribution for the selection coefficients with mean zero and precision γN,

$$\begin{aligned} P_{prior}(\boldsymbol{s}(t)) &= \frac{1}{(2\pi\sigma^2)^{(L+\Lambda)/2}} \exp\left(-\frac{1}{2\sigma^2} \boldsymbol{s}^T \boldsymbol{s} \right) \\ &\propto \exp\left(-\frac{N}{2} \gamma \boldsymbol{s}^2 \right). \end{aligned}$$

Here we have absorbed the population size N into the width of the prior for convenience (see below). Including the prior distribution, the action becomes

$$\begin{aligned} \boldsymbol{S}(\boldsymbol{x}(t_k)) = \sum_{k=0}^{K-1} \frac{1}{\Delta t_k} \sum_{\boldsymbol{i},\boldsymbol{j}=1}^{L+\Lambda} \Big\{ &[\Delta x_{\boldsymbol{i}}(t_k) - \Delta t_k D_{\boldsymbol{i}}(t_k)] \\ \times \left[C_{\boldsymbol{ij}}(x(t_k)) \right]^{-1} &\left[\Delta x_{\boldsymbol{j}}(t_k) - \Delta t_k D_{\boldsymbol{j}}(t_k) \right] \Big\} + \sum_{\boldsymbol{i}}^{L+\Lambda} \gamma s_{\boldsymbol{i}}^2 \end{aligned}$$

Finally, we can apply Bayes' theorem to infer selection coefficients $\hat{s}_{\boldsymbol{i}}$ that maximize the posterior distribution, providing the best compromise between the prior and the likelihood. While the posterior distribution is a complicated function of allele/trait frequencies, it is simply a Gaussian function of the selection coefficients. This allows us to write an analytical expression for the maximum a posteriori parameters, as shown in equation (2) in the main text, with

$$\begin{aligned} \boldsymbol{C}_{int} &= \sum_{k=0}^{K-1} \Delta t_k \boldsymbol{C}(t_k), \\ \Delta \boldsymbol{x} &= \boldsymbol{x}(t_K) - \boldsymbol{x}(t_0), \end{aligned}$$

$$\boldsymbol{\mu}_{\text{fl}} = \begin{cases} \sum_{k=0}^{K-1} \Delta t_k \mu (1 - 2\boldsymbol{x}(t_k)), \\ \sum_{k=0}^{K-1} \Delta t_k \mu \sum_{i \in N} (1 - \boldsymbol{x}_n(t_k) - \boldsymbol{y}_n^i(t_k)), \end{cases}$$

$$\boldsymbol{R}_{\text{fl}} = \begin{cases} 0, \\ \sum_{k=0}^{K-1} \Delta t_k r \left(\sum_{k \in n} P_{W,W}^{k,n} P_{M,M}^{k,n} - \sum_{k \in n} P_{M,W}^{k,n} P_{W,M}^{k,n} \right). \end{cases}$$

Here, the top expressions for $\boldsymbol{\mu}_{\text{fl}}$ and $\boldsymbol{R}_{\text{fl}}$ are for mutant alleles, and the bottom ones are for traits. In simulations, we used $\gamma = 1$ for selection coefficients and 0.1 for traits. In noisier HIV-1 data, we used $\gamma = 10$ for selection coefficients and 1 for traits.

Extension to multiple alleles per locus and asymmetric mutation probabilities

To study real sequences, we can extend the simple model presented in the previous sections to allow for multiple alleles per locus and asymmetric mutation probabilities. We use $\alpha, \beta, \ldots$ indices to represent different alleles (ranging, for example, over nucleotides or amino acids), where we write the total number of possible alleles at a locus as l. Our fitness model is then:

$$f_a = 1 + \sum_{i}^{L} \sum_{\alpha}^{l} s_{i,\alpha} g_{i,\alpha}^a + \sum_{n}^{\Lambda} s_n g_n^a.$$

Similarly, $s_{i,\alpha}$ represents the selection coefficient for allele α at locus i, and $g_{i,\alpha}^a$ equals 1 if genotype a has allele α at locus i. The trait term is the same as the binary case since it does not have a natural counterpart (only 2 states, WT and MT). We use $x_{i,\alpha}(t_k)$ to represent the frequency of allele α at locus i at generation t_k, and $\mu_{\alpha\beta}$ to denote the probability per locus per generation of mutation from allele α to β. Following parallel arguments to before, the MPL estimate of the selection coefficient $s_{i,\alpha}$ for each allele α at each locus i and the trait coefficients s_n can be obtained.

First, we write the diffusion matrix $C_{i\alpha,j\beta}(t_k)/N$,

$$C_{\boldsymbol{i}\alpha,\boldsymbol{j}\beta}(t_k) = x_{\boldsymbol{i}\alpha,\boldsymbol{j}\beta}(t_k) - x_{\boldsymbol{i}\alpha}(t_k) \cdot x_{\boldsymbol{j}\beta}(t_k).$$

where $x_{\boldsymbol{i}\alpha,\boldsymbol{j}\beta}(t_k)$ is the frequency of sequences with alleles α and β at loci $\boldsymbol{i}$ and $\boldsymbol{j}$ at generation t_k. When one of the indices corresponds to a trait, allele subscripts are not needed. For example, the covariance between trait group n and allele α at locus i can be written as $C_{n,i\alpha}(t_k) = x_{n,i\alpha}(t_k) - x_n(t_k) \cdot x_{i\alpha}(t_k)$.

The estimates selection coefficients for alleles are then

$$s_{i,\alpha} = \sum_{\boldsymbol{j}}^{L+\Lambda} \sum_{\beta}^{l} \Big[\sum_{k=0}^{K-1} \Delta t_k C_{i\alpha,\boldsymbol{j}\beta}(t_k) + \gamma \hat{I} \Big]_{i\alpha\boldsymbol{j}\beta}^{-1} \cdot \Big[x_{\boldsymbol{j}\beta}(t_K) - x_{\boldsymbol{j}\beta}(t_0) - \sum_{k=0}^{K-1} \Delta t_k \sum_{\delta \neq \alpha} \left(\mu_{\delta\alpha} x_{\boldsymbol{j}\delta}(t_k) - \mu_{\alpha\delta} x_{\boldsymbol{j}\alpha}(t_k) \right) \Big], \quad \textbf{(9)}$$

And the selection coefficients for traits are given by

$$s_n = \sum_{\boldsymbol{j}}^{L+\Lambda} \sum_{\beta}^{l} \Big[\sum_{k=0}^{K-1} \Delta t_k C_{n,\boldsymbol{j}\beta}(t_k) + \gamma \hat{I} \Big]^{-1}_{n\boldsymbol{j}\beta} \cdot \Big[x_{\boldsymbol{j}\beta}(t_K) - x_{\boldsymbol{j}\beta}(t_0) - \sum_{k=0}^{K-1} \Delta t_k \sum_{i\in n} \sum_{\delta\neq\epsilon} \sum_{\epsilon} \Big(\mu_{\epsilon\delta} y_n^{i\epsilon} - \mu_{\delta\epsilon} y_n^{i\delta}(t_k) \Big) \quad \textbf{(10)}$$
$$- \sum_{k=0}^{K-1} \Delta t_k r \Big(\sum_{k\in n} P^{k,n}_{W,W} P^{k,n}_{M,M} - \sum_{k\in n} P^{k,n}_{M,W} P^{k,n}_{W,M} \Big) \Big] .$$

Here the index ϵ runs over WT alleles and synonymous mutations for each binary trait n. This is because only nonsynonymous mutations contribute to changes in the trait in our model.

In this version of the model, we compute selection coefficients for both mutant and WT alleles. However, what is important is not the selection coefficient for an individual allele in isolation, but rather the change in fitness upon mutation. Thus, the selection coefficients we report are normalized by subtracting the WT selection coefficient from the selection coefficient of each allele at each locus ($s_{i\alpha} - s_{iW}$). In this way, the selection coefficients are normalized such that the WT selection coefficient is 0, and the selection coefficients for other alleles quantify the advantage or disadvantage relative to WT. In the limit that the selection coefficients are small, one can show that the WF dynamics are invariant under such shifts in the selection coefficients. In physics, this phenomenon is referred to as gauge invariance.

Simulation data

We simulated the WF model with discrete generations and binary (mutant/WT) states in Python. Briefly, we evolved populations of sequences according to (5) over multiple generations, starting with an initial population of all WT sequences and recording the entire evolutionary history. After simulation, we computed the single x_i, double x_{ij} mutant frequencies, and trait frequencies x_n from the sampled sequences trajectories and used them to infer individual selection coefficients (9) and trait coefficients (10). Parameter values are detailed in **Fig. 1**. The simulation and analysis code with original simulation data are contained in the GitHub repository.

Since real data typically contains only a small portion of population, and is not sampled at every generation, we also studied how different sampling depths and sampling time intervals affect the performance of our method. We chose part of the sequences from the population and time points to estimate individual selection and trait coefficients. n_s denotes the number of sequences we randomly selected from the population and Δt is the time interval, which means we choose the data every Δt generations. The initial population and simulation parameters are described in **Supplementary Fig. 1**.

HIV-1 sequence data

We obtained HIV-1 sequence data from 13 individuals of the CHAVI 001 and CAPRISA 002 studies in the United States, Malawi, and South Africa from the Los Alamos National Laboratory (LANL) HIV Sequence Database. We applied several selection criteria[44] to minimize the influence of noise in the data, including removing the sequences with large numbers of gaps, sites with high gap frequencies, and time points with very small numbers of sequences or large gaps in time from the last sample. We also imputed ambiguous nucleotides with the most common nucleotides observed at the same site within the same individual.

For these 13 individuals, sequence data consisted of 3′ and 5′ half-genome sequences, which were approximately 4,500 bp in length. Our analysis focused only on polymorphic sites, where more than one nucleotide (including gaps/deletions) was observed in an individual (approximately 100-900 bp in length). To infer selection, we used a recombination rate $r = 1.4 \times 10^{-5}$ (ref.[46]) and mutation rate matrix estimated in ref.[47] as input. We treated the transmitted/founder (TF) sequence as the "wild-type" sequence for each individual.

The locations of $\mathrm{CD8}^+$ T cell epitopes for these sequences were experimentally[48] or computationally[49] determined. *Escape sites* refer to polymorphic sites where nonsynonymous mutations were observed in the reading frame of a $\mathrm{CD8}^+$ T cell epitope. In this way, we anticipate that nonsynonymous mutations in escape sites will affect T cell recognition. We consider the escape sites that can change the same epitope are among one trait group.

Calculation of effects of linkage on inferred selection

$\Delta\hat{s}_{ij}$ can tell us the effects of linkage from variant i to j. To compute it, we calculate the coefficients for mutant variant j and eliminate the influence from mutant variant i by artificially reverting variant i to WT[44].

For mutant alleles, we generated a modified version of the sequence data where all mutant variants i are replaced by the corresponding TF nucleotide for all sequences at all time points. For traits, we treat all mutant variants within one epitope as WT. With these modified data, we can infer the coefficients again for all variants $\boldsymbol{j}$, denoted as $\hat{s}_j^{\setminus i}$. Then we define

$$\Delta\hat{s}_{ij} = \hat{s}_j - \hat{s}_j^{\setminus i}.$$

Positive values of $\Delta\hat{s}_{ij}$ indicate that linkage with variant i can increase the selection coefficient inferred for variant j and vice versa. By computing the $\Delta\hat{s}_{ij}$, we can quantify the effect of linkage on inferred selection.

The effects of recombination on the inference

Unlike selection on individual alleles, which has been studied previously using similar approaches[44,45], the selection that we infer for binary traits is affected by recombination. How large is the contribution of recombination in this analysis?

To answer this question, we inferred selection on mutant alleles and traits in simulations and in the HIV-1 data sets, with and without the inclusion of recombination. In general, we find that the effect of recombination on the inferred coefficients is small. This is reasonable, as recombination only affects the traits, which constitute a small portion of a sequence. For simulation data, there are three traits in a sequence length of 50; however, for experimental data, it is often a few epitopes contained within sequences that are thousands of base pairs long.

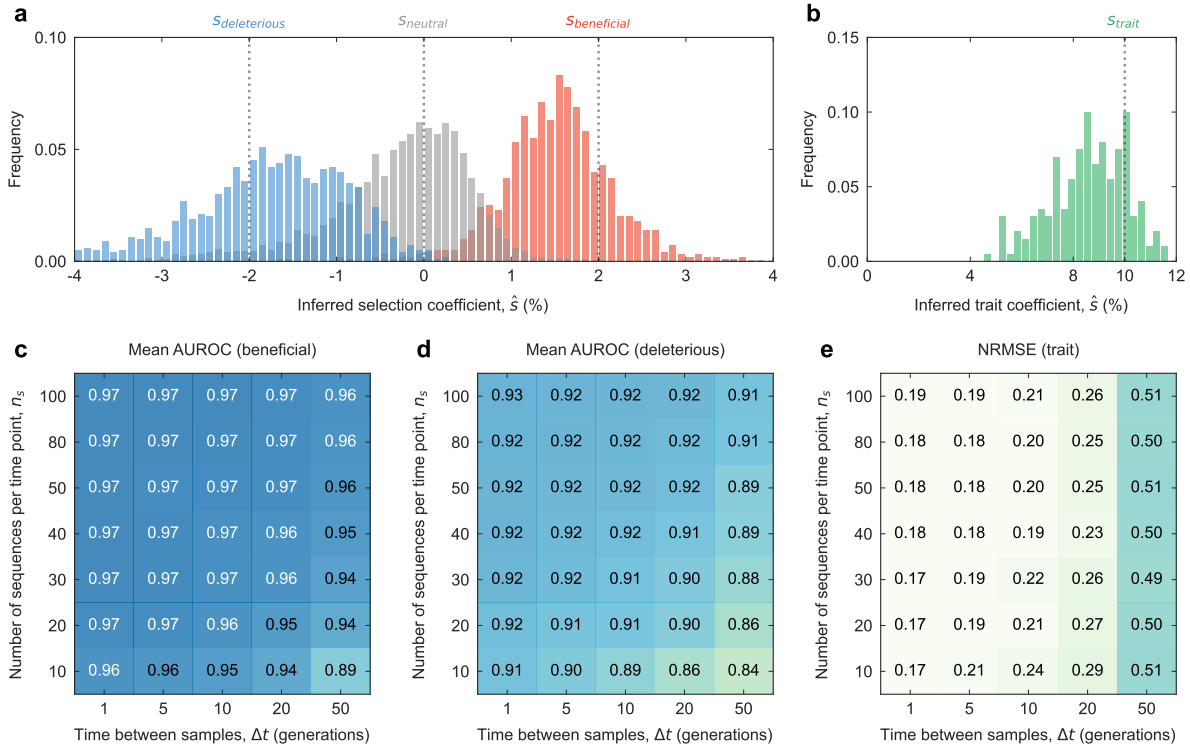
Simulation results (**Supplementary Fig. 6**) illustrate that the recombination term has negligible effects on the inference of individual selection coefficients. For trait terms, incorporating recombination does lead to small but noticeable improvements in inference. In HIV-1 data, the influence of the recombination term is small (**Supplementary Fig. 7**). Although the recombination rate in HIV-1 is relatively high, the scenarios under which recombination will lead to net change in trait frequencies are rare.

Data and code

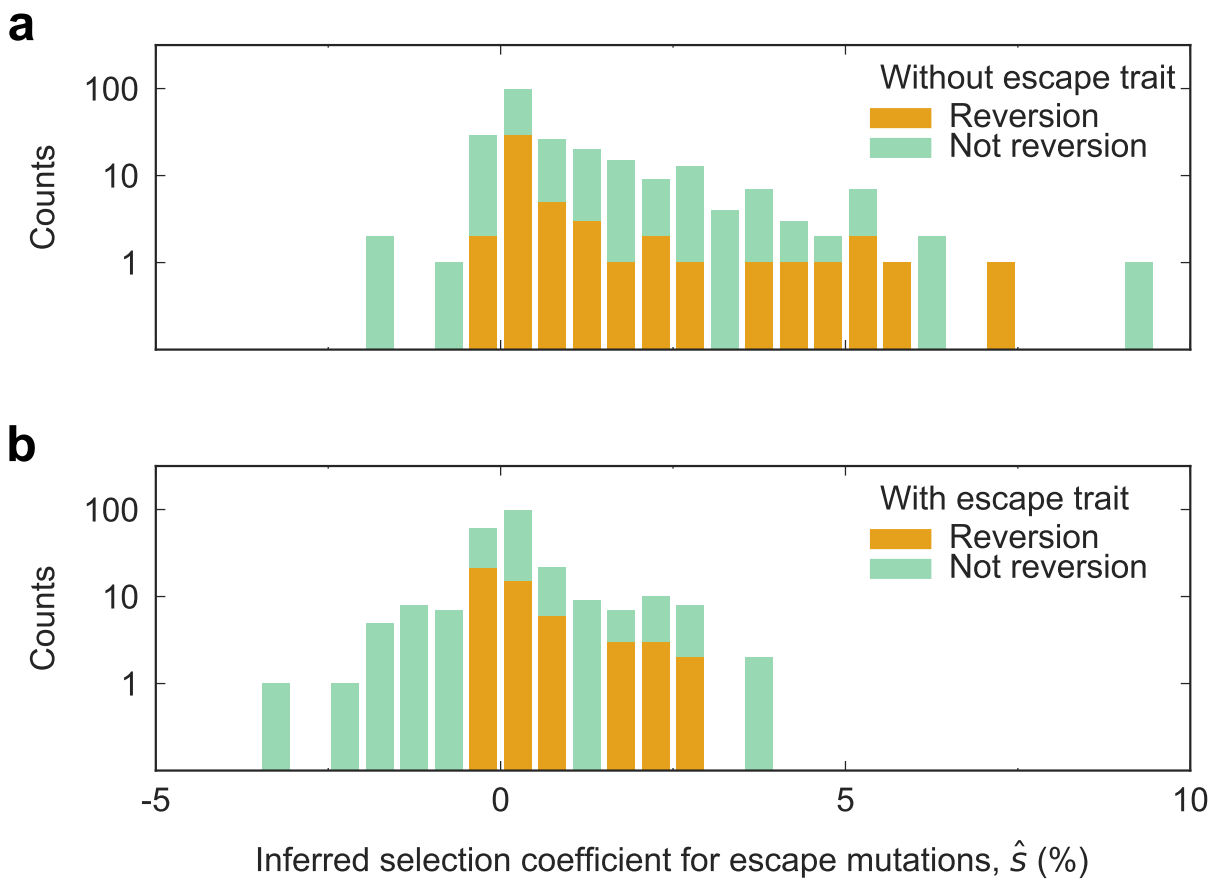
Data and code used in our analysis is available in the GitHub repository located at `https://github.com/bartonlab/paper-binary-trait-inference`. This repository also contains Jupyter notebooks that can be run to reproduce our figures and analysis.

References

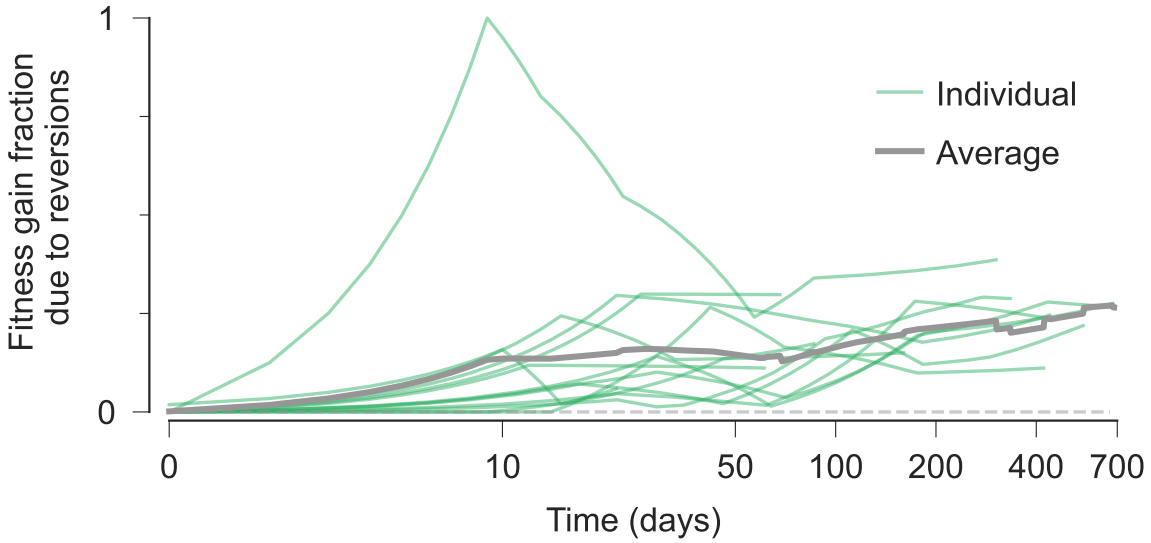
42. Ewens, W. J. *Mathematical population genetics: theoretical introduction*, vol. 27 (Springer, 2004).
43. Mustonen, V. & Lässig, M. Fitness flux and ubiquity of adaptive evolution. *Proceedings of the National Academy of Sciences* **107**, 4248–4253 (2010).
44. Sohail, M. S., Louie, R. H., McKay, M. R. & Barton, J. P. Mpl resolves genetic linkage in fitness inference from complex evolutionary histories. *Nature biotechnology* **39**, 472–479 (2021).
45. Lee, B. *et al.* Inferring effects of mutations on sars-cov-2 transmission from genomic surveillance data. *medRxiv* 2021–12 (2022).
46. Neher, R. A. & Leitner, T. Recombination rate and selection strength in hiv intra-patient evolution. *PLoS computational biology* **6**, e1000660 (2010).
47. Zanini, F., Puller, V., Brodin, J., Albert, J. & Neher, R. A. In vivo mutation rates and the landscape of fitness costs of hiv-1. *Virus evolution* **3**, vex003 (2017).
48. Liu, M. K. *et al.* Vertical t cell immunodominance and epitope entropy determine hiv-1 escape. *The Journal of clinical investigation* **123** (2012).
49. Barton, J. P. *et al.* Relative rate and location of intra-host hiv evolution to evade cellular immunity are predictable. *Nature communications* **7**, 11660 (2016).



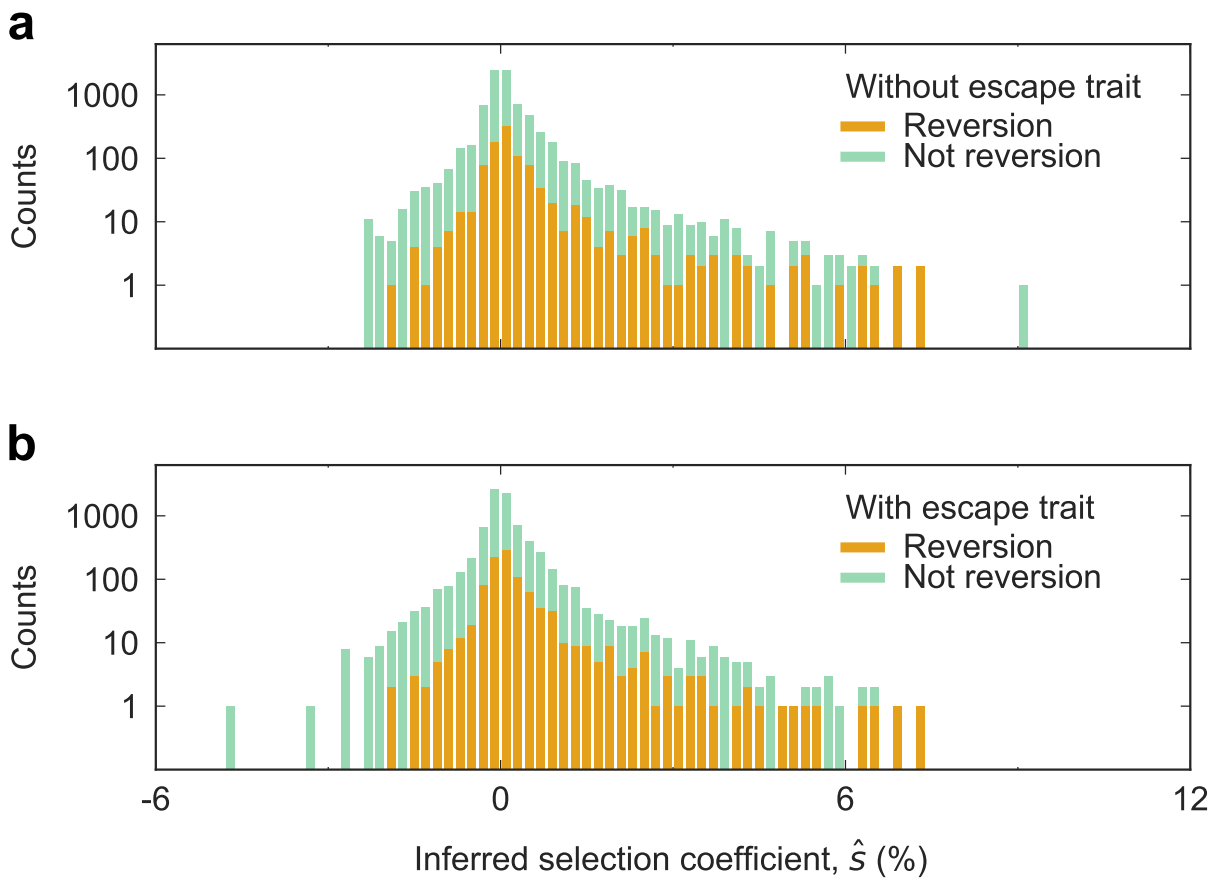
Supplementary Fig. 1. MPL recovers selection from complex dynamics even from limited data. Distribution of (**a**) individual selection coefficient and (**b**) trait coefficient estimates across 100 replicate simulations, using the same parameters as in **Fig. 1**. Our approach is robust to finite sampling constraints, as measured by the accurate classification of (**c**) beneficial and (**d**) deleterious mutants and (**e**) inference of trait coefficients, even when the number of sequences sampled per time point n_s is low and the spacing between time samples Δt is large. AUROC, area under the receiver operating characteristic; NRMSE, normalized root mean square error. Simulation parameters: $L = 50$ loci with two alleles at each locus (mutant and WT), ten beneficial mutants with $s = 0.02$, 30 neutral mutants with $s = 0$, and ten deleterious mutants with $s = -0.02$. We consider $\Lambda = 2$ trait groups, each with three contributing alleles and trait coefficients $s = 0.1$. Mutation probability per locus per generation $\mu = 2 \times 10^{-4}$, population size $N = 10^3$. The initial population is all wild type, evolved over $T = 1000$ generations.



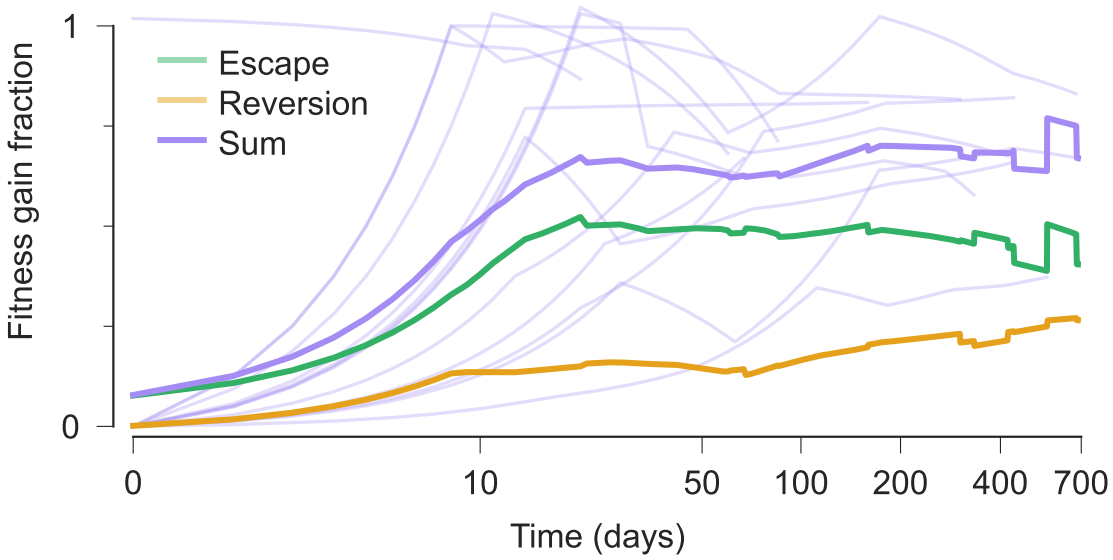
Supplementary Fig. 2. Inclusion of escape coefficients substantially shifts inferred selection coefficients for escape mutations toward more deleterious values. **a**, Distribution of inferred selection coefficients without escape traits. **b**, Distribution of inferred selection coefficients with escape traits. Reversions are slightly more likely to be inferred to be beneficial/less like to be deleterious than non-reversions.



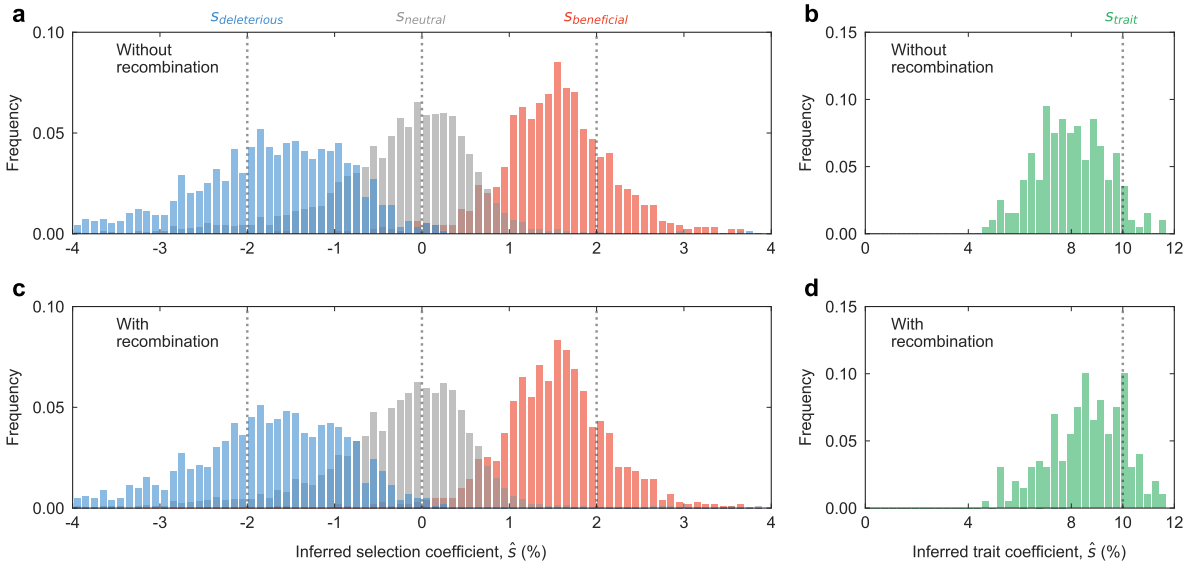
Supplementary Fig. 3. The contribution of reversions to intrahost HIV-1 fitness gains is significant and grows over time. Reversions refer to mutations that revert from the transmitted/founder (TF) variant to the nucleotide of the HIV-1 clade consensus sequence.



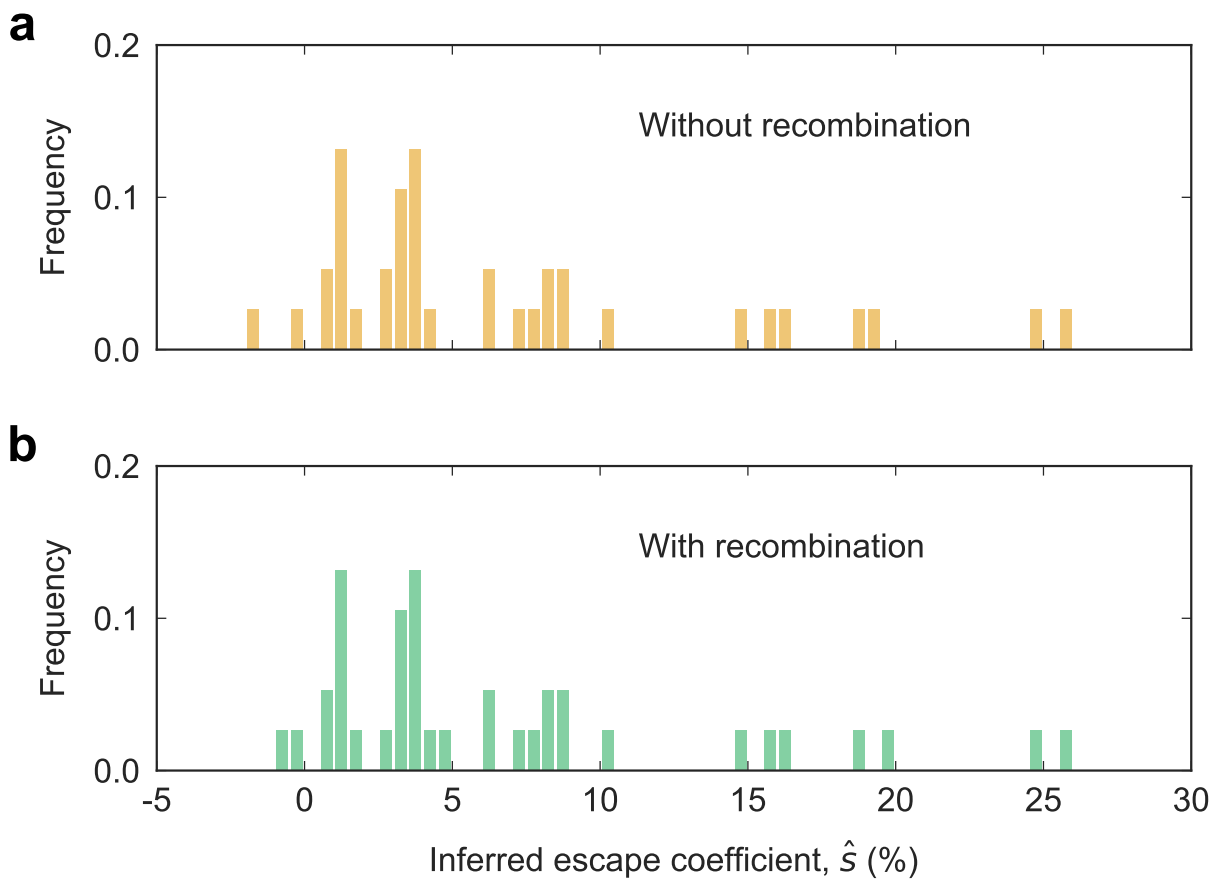
Supplementary Fig. 4. The distribution of inferred selection coefficients with and without the inclusion of escape traits. a, Distribution of inferred selection coefficients without escape traits. **b**, Distribution of inferred selection coefficients with escape traits. We note an enrichment in substantially beneficial mutations that are also reversions to the HIV-1 clade consensus sequence. While some large-effect mutations shift substantially with the inclusion of escape traits (i.e., escape mutations), the effect on the bulk of the inferred selection coefficients is minimal.



Supplementary Fig. 5. Intrahost HIV-1 fitness gains due to CD8$^+$ T cell escape mutations and reversions. In total, CD8+ T cell escape (see **Fig. 2b**) and reversions (see **Supplementary Fig. 3**) make dominant contributions to HIV-1 fitness gains *in vivo*. Across the 13 patient data sets that we studied, the fraction of fitness gains due to escape and reversions is around 75% on average. At earlier times, T cell escape plays a larger role, while the contribution of reversions grows steadily over time.



Supplementary Fig. 6. Effects of recombination on inferred selection coefficients and trait coefficients in 100 replicate simulations. The distribution of inferred (**a**) selection coefficients and (**b**) trait coefficients in simulations, using the same parameters as in **Fig. 1**, without including the effects of recombination in the estimator. **c**, **d**, Analogous distributions of inferred coefficients when recombination is included in the estimator. The effects of recombination on the inferred parameters are subtle, with the most prominent feature being a shift in the inferred trait coefficients toward the true value and away from zero.



Supplementary Fig. 7. Effects of recombination on inferred escape coefficients in HIV-1 data. Distribution of inferred escape coefficients (**a**) without and (**b**) with the inclusion of recombination in the estimator. Even though the recombination rate for HIV-1 is relatively high ($r \sim 1.4 \times 10^{-5}$), recombination events capable of altering trait frequencies are rare. The contribution of $\boldsymbol{R}_{\mathrm{fl}}$ to the inferred escape coefficients is thus small.