Parallel HIV-1 evolutionary dynamics in humans and rhesus macaques who develop broadly neutralizing antibodies

Kai S. Shimagaki[1,2], **Rebecca M. Lynch**[3], **and John P. Barton**[1,2,†]

[1]Department of Computational and Systems Biology, University of Pittsburgh School of Medicine, USA. [2]Department of Physics and Astronomy, University of Pittsburgh, USA. [3]Department of Microbiology, Immunology and Tropical Medicine, School of Medicine and Health Sciences, George Washington University, USA. †Address correspondence to: jpbarton@pitt.edu.

Human immunodeficiency virus (HIV)-1 exhibits remarkable genetic diversity. For this reason, an effective HIV-1 vaccine must elicit antibodies that can neutralize many variants of the virus. While broadly neutralizing antibodies (bnAbs) have been isolated from HIV-1 infected individuals, a general understanding of the virus-antibody coevolutionary processes that lead to their development remains incomplete. We performed a quantitative study of HIV-1 evolution in two individuals who developed bnAbs. We observed strong selection early in infection for mutations affecting HIV-1 envelope glycosylation and escape from autologous strain-specific antibodies, followed by weaker selection for bnAb resistance later in infection. To confirm our findings, we analyzed data from rhesus macaques infected with viruses derived from the same two individuals. We inferred remarkably similar fitness effects of HIV-1 mutations in humans and macaques. Moreover, we observed a striking pattern of rapid HIV-1 evolution, consistent in both humans and macaques, that precedes the development of bnAbs. Our work highlights strong parallels between infection in rhesus macaques and humans, and it reveals a quantitative evolutionary signature of bnAb development.

Introduction

Human immunodeficiency virus (HIV-1) rapidly mutates and proliferates in infected individuals. The immune system is a major driver of HIV-1 evolution, as the virus accumulates mutations to escape from host T cells and antibodies[1–3]. Due to the chronic nature of HIV-1 infection, coupled with high rates of mutation and replication, HIV-1 genetic diversity within and between infected individuals is incredibly high. Genetic diversity challenges vaccine development, as vaccine-elicited antibodies must be able to neutralize many strains of the virus to protect against infection[4].

However, there exist rare antibodies that are capable of neutralizing a broad range of HIV-1 viruses. These broadly neutralizing antibodies (bnAbs) have therefore been the subject of intense research[5–8]. Eliciting bnAbs through vaccination remains a major goal of HIV-1 vaccine design. However, the development of bnAbs in infected individuals is rare. Only around 10-20% of HIV patients are observed to develop bnAbs, and typically only after several years of infection[9,10].

Recent years have yielded important insights into the coevolutionary process between HIV-1 and antibodies that sometimes leads to the development of bnAbs. Clinical studies have collected serial samples of HIV-1 sequences from a few individuals who developed bnAbs and characterized the resulting antibodies, their developmental stages, and binding sites[11–13]. The contributions of HIV-1 and its coevolution to bnAb development are complex[14,15]. High viral loads and viral diversity have been positively associated with bnAb development[14–16]. However, superinfection, which can vastly increase HIV-1 diversity, is not always associated with bnAb development[17], and it does not appear to broaden antibody responses in the absence of other factors[15].

Here, we sought to characterize the evolutionary dynamics of HIV-1 that accompany the development of bnAbs in clinical data. In particular, we inferred the landscape of selective pressures that shape the evolution of HIV-1 within hosts, reflecting the effects of the immune environment. We first analyzed data from two individuals who developed bnAbs within a few years after HIV-1 infection[11,12]. In both individuals, HIV-1 mutations inferred to be the most beneficial were observed early in infection. In general, mutations that provided resistance to autologous strain-specific antibodies were inferred to be more strongly selected than ones that escaped from bnAbs. We also observed clusters of beneficial mutations along the HIV-1 genome, which were associated with envelope protein (Env) structure.

To confirm the generality of these patterns in a broader sample, we studied recent data from rhesus macaques (RMs) infected with simian-human immunodeficiency viruses (SHIV) that incorporated HIV-1 Env proteins derived from the two individuals above[18]. We observed striking parallels between the inferred fitness effects of HIV-1 mutations in RMs and humans, suggesting highly similar selective pressures on the virus despite different host species and differences in individual immune responses. Furthermore, we found that RMs that developed broad, potent antibody responses could clearly be distinguished from those with narrowly focused responses using the evolutionary dynamics of the virus. Specifically, the virus population in individuals who developed greater breadth was distinguished by larger and more rapid gains in fitness than in other individuals. Collectively, these results show high similarity between SHIV evolutionary dynamics in RMs and HIV-1 in humans, and that viral fitness gain is associated with antibody breadth.

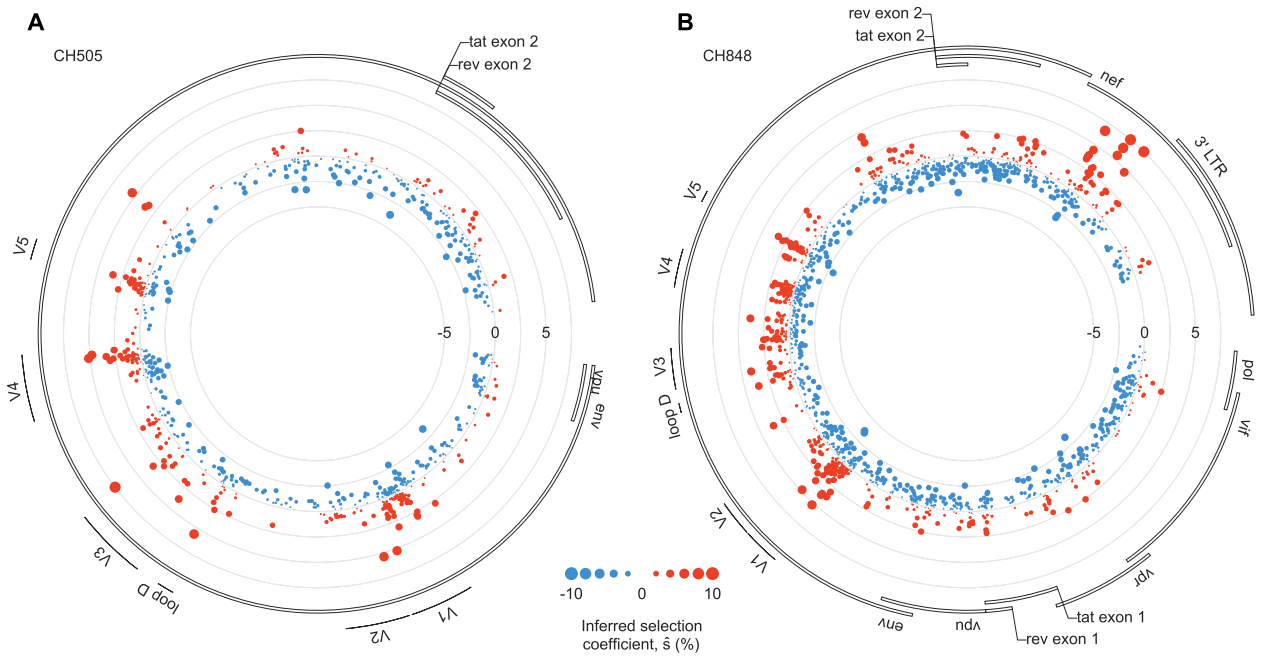


Fig. 1. Beneficial mutations occur in clusters along the genome. Inferred fitness effects of HIV-1 mutations in CH505 (**A**) and CH848 (**B**). The position along the radius of each circle specifies the strength of selection: mutations plotted closer to the center are more deleterious, while those closer to the edge are more beneficial. For both individuals, clusters of beneficial mutations are observed in the variable loops of Env, some of which are associated with antibody escape. For CH848, a group of strongly beneficial mutations also appears in Nef.

Results

Quantifying HIV-1 evolutionary dynamics

We studied HIV-1 evolution accompanying the development of bnAbs in two donors, CH505 and CH848, enrolled in the Center for HIV/AIDS Vaccine Immunology 001 acute infection cohort[19]. CH505 developed the CD4 binding site-targeting bnAb CH103, which was first detectable 14 weeks after HIV-1 infection[11]. CH103 maturation was found to be associated with viral escape from another antibody lineage, CH235, that ultimately developed significant breadth[20,21]. CH848 developed a bnAb, DH270, targeting a glycosylated site near the third variable loop (V3) of Env[12]. Similar to the bnAb development process in CH505, escape from "cooperating" DH272 and DH475 lineage antibodies was observed to contribute to the maturation of DH270 (ref.[12]).

To quantify HIV-1 evolutionary dynamics, we sought to infer a fitness model that best explained the changes in the genetic composition of the viral population observed in each individual over time. We assume that the effect on viral fitness of each individual mutation a at each site i is quantified by a selection coefficient $s_i(a)$, with positive coefficients $s_i(a) > 0$ denoting mutations that are beneficial for the virus and $s_i(a) < 0$ denoting deleterious ones. We further assume that the cumulative fitness effects of mutations are additive, such that the overall fitness F of a viral sequence is given by the sum of the selection coefficients for all the mutations that it bears. Following this model, we can then quantify the probability of any evolutionary history of the viral population (i.e., the distribution of viral genotypes over time) as a function of the selection coefficients (Methods). We then used MPL[22,23] to infer the selection coefficients that best fit the viral dynamics observed in data from CH505 and CH848 (Methods). This approach has been successfully applied to study the evolution of HIV-1 (ref.[22]), as well as SARS-CoV-2 (ref.[24]) and experimental evolution in bacteria[25,26].

Broad patterns of HIV-1 selection

While the great majority of HIV-1 mutations were inferred to be neutral ($s_i(a) \sim 0$), a few mutations substantially increase viral fitness (**Fig. 1**). Strongly beneficial mutations occurred in clusters along the genome and preferentially appeared in specific regions of Env (**Fig. 2**).

To quantify patterns of selection in Env, we examined the top 2% of mutations inferred to be the most beneficial in CH505 and CH848. For CH505, we found 10.9-fold more strongly beneficial mutations in the first variable loop (V1) than expected by chance ($p = 2.5 \times 10^{-3}$, Fisher's exact test; Methods). This is consistent with the presence of V1 mutations conferring resistance to autologous strain-specific antibodies[20]. Mutations in V4, a region targeted by $CD8^+$ T cells[20], were also 10.0-fold enriched in this subset ($p = 9.0 \times 10^{-5}$). For CH848, mutations in V1, V3, and V5 were enriched by factors of 14.2, 6.3, and 19.1 among the top 2% most beneficial mutations ($p = 5.9 \times 10^{-10}$, 6.6×10^{-3}, and 4.8×10^{-6}). Mutations in these regions were shown to play a role in resistance to DH270 and DH475 lineage antibodies[12].

Reversions and mutations affecting N-linked glycosylation motifs were also likely to be beneficial. We define reversions as mutations where the transmitted/founder (TF) amino acid changes to match the subtype consensus sequence at the same site. Among the top 2% most beneficial mutations, reversions were enriched by factors of 19.9 and 17.8 for CH505 and CH848 viruses, respectively ($p = 2.1 \times 10^{-8}$

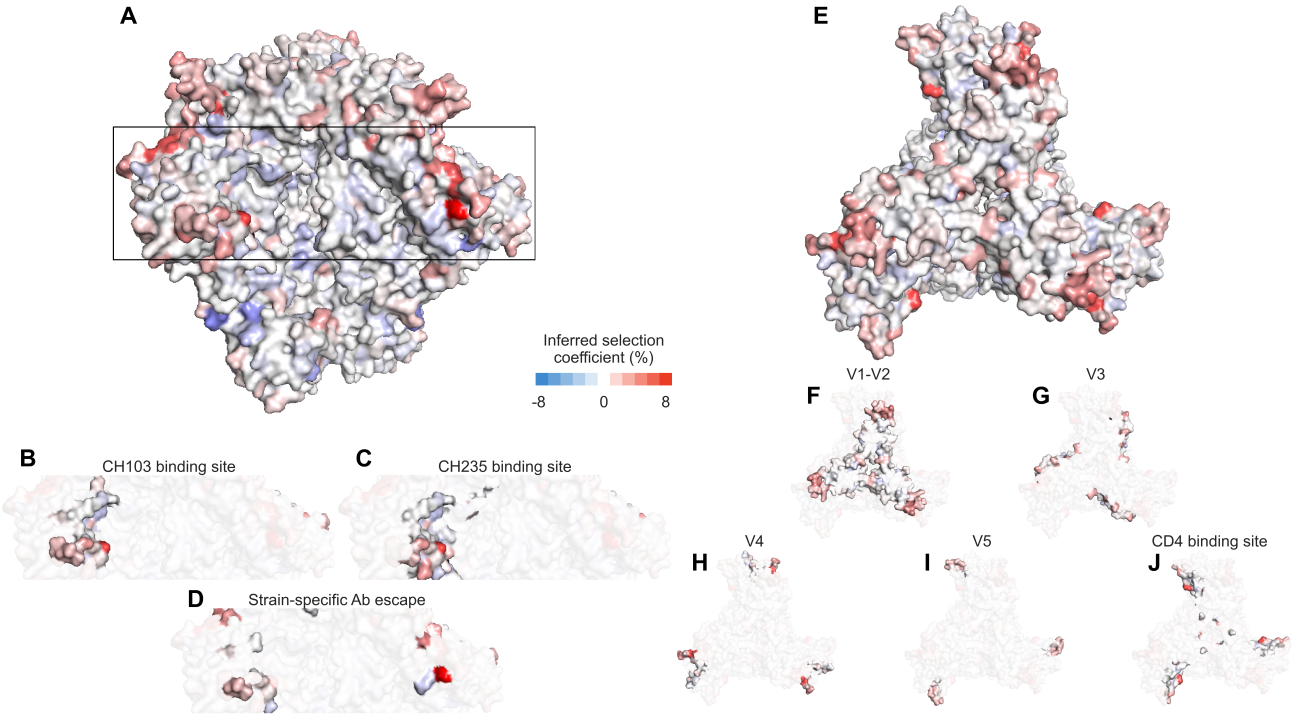


Fig. 2. Visualization of the effects of HIV-1 mutations in CH505 on the Env trimer. A, Side view of the Env trimer[27], with detail views of selection for mutations in the CH103 binding site (**B**), CH235 binding site (**C**), and sites associated with escape from autologous strain-specific antibodies (**D**). **E**, Top view of the trimer, with detail views of variable loops (**F-I**) and the CD4 binding site (**J**). Generally, beneficial mutations appear more frequently near exposed regions at the top of the Env trimer, and deleterious mutations appear in more protected regions. Mutations near the V1/V2 apex can affect the binding and neutralization of antibodies targeting the CD4 binding site[20].

and 8.5×10^{-13}), consistent with past work finding strong selection for reversions[22,28]. For CH848, this group also includes several strongly selected mutations observed in Nef (**Fig. 1B**), a protein that plays multiple roles during HIV-1 infection[29,30]. Mutations affecting N-linked glycosylation motifs (i.e., by adding, removing, or shifting a glycosylation motif) were enriched by factors of 4.6 ($p = 7.0 \times 10^{-3}$) and 8.7 ($p = 1.4 \times 10^{-10}$). Changes in glycosylation patterns contributed to antibody escape for both CH505 and CH848 (refs. [12,20]).

Selection for antibody escape

To quantify levels of selection for antibody escape, we computed selection coefficients for mutations that were observed to contribute to resistance to bnAbs as well as bnAb precursors and autologous strain-specific antibodies[20,31–34]. We mapped the inferred selection coefficients to the Env protein structure[11], highlighting the binding sites for bnAbs and resistance mutations for strain-specific antibodies, as well as important parts of Env (**Fig. 2**, **Supplementary Fig. 1**). We also analyzed when these resistance mutations were first observed in each individual.

Overall, we observed stronger selection for escape from autologous strain-specific antibodies and/or changes in glycosylation during the first six months of infection (**Supplementary Tables 3-4**). This was then followed by more modest selection for escape from bnAb lineages (**Fig. 3**). For CH505, mutations that conferred resistance to the intermediate-breadth CH235 lineage[20] were less beneficial than top mutations escaping from strain-specific antibodies (**Fig. 3A**). In turn, resistance mutations for the broader CH103 lineage were less beneficial than CH235 resistance mutations (**Fig. 3A**). CH848 is similar, with some highly beneficial mutations affecting glycosylation observed early in infection (**Fig. 3B**). While a few mutations affecting DH270 appear strongly selected, these mutations appeared long before the DH270 lineage was detected (around 3.5 years after infection[12]). Thus, these mutations may have initially been selected for other reasons.

Consistent patterns of selection in SHIV evolution

Simian-human immunodeficiency viruses (SHIVs) have numerous applications in HIV/AIDS research[35,36]. Recently, Roark and collaborators studied SHIV-antibody coevolution in rhesus macaques (RMs)[18]. Two of the SHIV constructs in this study included envelope sequences derived from CH505 and CH848 transmitted/founder (TF) viruses. There, it was found that 2 out of 10 RMs inoculated with SHIV.CH505 and 2 out of 6 RMs inoculated with SHIV.CH848 developed antibodies with substantial breadth.

To understand whether the patterns of HIV-1 selection observed in CH505 and CH848 are repeatable, and to search for viral factors that distinguish between individuals who develop bnAbs and those who do not, we analyzed SHIV.CH505 and SHIV.CH848 evolution in RMs[18]. To prevent spurious inferences, we first omitted data from RMs with <3 sampling times or <4 sequences in total (Methods). After processing, we examined evolutionary data from 7 RMs

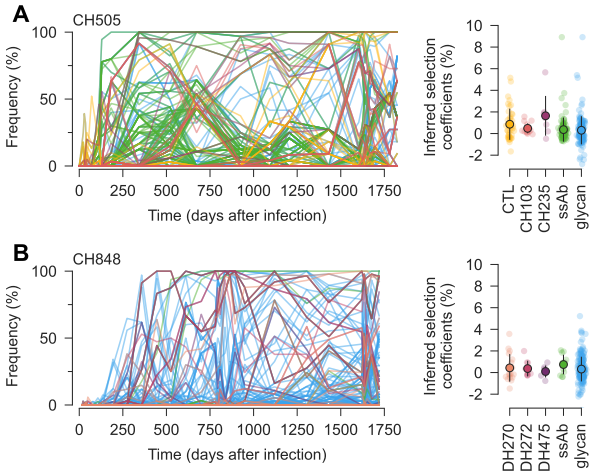


Fig. 3. Trajectories and inferred selection coefficients for immune escape mutations and mutations affecting Env glycosylation. A, For CH505, early, strongly selected mutations include ones that escape from CD8$^+$ T cells and autologous strain-specific antibodies (ssAbs). More moderately selected bnAb resistance mutations tend to arise later. Note that mutations that affect bnAb resistance can appear in the viral population before bnAbs are generated. Open circles and error bars reflect the mean and standard deviation of inferred selection coefficients in each category. **B**, For CH848, mutations affecting glycosylation dominate the early phase of evolution, followed later by mutations affecting bnAb resistance.

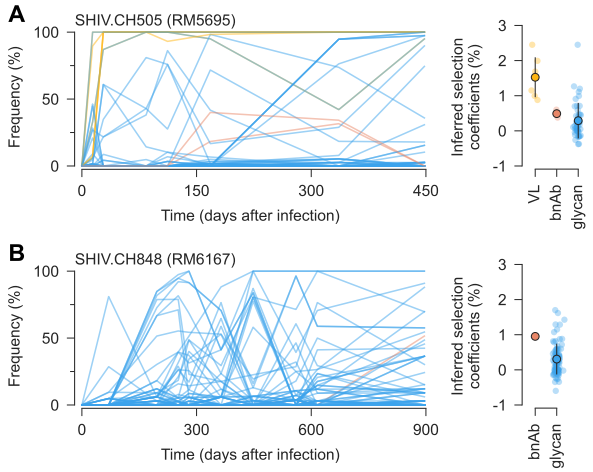


Fig. 4. Example trajectories and selection coefficients for SHIV mutations that affect viral load, bnAb recognition, or glycosylation. A, In RM5695, infected with SHIV.CH505, mutations known to increase viral load[34] and ones affecting glycosylation were rapidly selected. Mutations affecting resistance to broad antibodies[18] arose later under moderate selection. Open circles and error bars reflect the mean and standard deviation of inferred selection coefficients in each category. **B**, Slower but qualitatively similar evolutionary patterns were observed in RM6163, infected with SHIV.CH848.

inoculated with SHIV.CH505 and 6 RMs inoculated with SHIV.CH848 (**Supplementary Tables 1-2**). We then computed selection coefficients for SHIV mutations within each RM. Reasoning that selective pressures across SHIV.CH505 and SHIV.CH848 viruses are likely to be similar, we also inferred two sets of joint selection coefficients that best describe SHIV evolution in SHIV.CH505- and SHIV.CH848-inoculated RMs, respectively (Methods).

As before, we examined the top 2% of SHIV.CH505 and SHIV.CH848 mutations that we inferred to be the most beneficial for the virus. Overall, we found consistent selection for reversions (17.9- and 14.2-fold enrichment, $p = 4.3 \times 10^{-11}$ and 1.2×10^{-11} for SHIV.CH505 and SHIV.CH848, respectively), with slightly attenuated enrichment in mutations that affect N-linked glycosylation (2.7- and 4.2-fold enrichment, $p = 5.4 \times 10^{-2}$ and 3.4×10^{-6}). However, there is a small subset of mutations that *shift* glycosylation sites by simultaneously disrupting one N-linked glycosylation motif and completing another, where highly beneficial mutations occur far more often than expected by chance (158.3- and 191.1-fold enrichment, $p = 1.7 \times 10^{-4}$ and 7.9×10^{-11} for CH505 and SHIV.CH505, respectively; 118.7-fold and 90.4-fold enrichment, $p = 2.0 \times 10^{-4}$ and 2.3×10^{-7} for CH848 and SHIV.CH848).

We observed some differences between HIV-1 and SHIV in the precise locations of the most beneficial Env mutations. For example, mutations in V4 are highly enriched in CH505 due to a CD8$^+$ T cell epitope in this region, but not in SHIV.CH505 (2.7-fold enrichment, $p = 0.13$). For SHIV.CH848, beneficial mutations are modestly enriched in V1 and V5 (4.7- and 4.2-fold, $p = 4.0 \times 10^{-3}$ and 6.7×10^{-2}), as in CH848, but not for V3 (0.9-fold enrichment, $p = 0.22$).

Despite some differences in the top mutations, patterns of selection over time in SHIV were very similar to those found for HIV-1. As before, highly beneficial mutations, including ones affecting glycosylation, tended to appear earlier in infection. This was followed by modestly beneficial mutations at later times, including ones involved in resistance to bnAbs in the RMs who developed antibodies with significant breadth (examples in **Fig. 4**).

Detection of SHIV mutations that increase viral load

A major goal of nonhuman primate studies with SHIV is to faithfully recover important aspects of HIV-1 infection in humans. However, due to the divergence of simian immunodeficiency viruses and HIV-1, SHIVs are not always well-adapted to replication in RMs[34]. To combat this problem, a recent study identified six SHIV.CH505 mutations that increase viral load (VL) in RMs[34]. These mutations result in viral kinetics that better mimic HIV-1 infection in humans.

Our analysis readily identifies the SHIV.CH505 mutations shown to increase VL. Five out of the top six SHIV.CH505 mutations with the largest average selection coefficients are associated with increased VL (**Supplementary Table 5**). The final mutation identified by Bauer et al., N130D, is ranked tenth. We also find highly beneficial mutations in SHIV.CH848 that are distinct from those in SHIV.CH505 (**Supplementary Table 6**). Highly-ranked mutations identified here may be good experimental targets for future studies aimed at increasing SHIV.CH848 replication *in vivo*.

Fitness agreement between HIV-1 and SHIV

Next, we explored the similarity in the overall viral fitness landscapes inferred for HIV-1 and SHIV, beyond just the top mutations. First, we computed the fitness of each SHIV sequence using the joint SHIV.CH505 and SHIV.CH848 selection coefficients inferred from RM data. Then, we computed

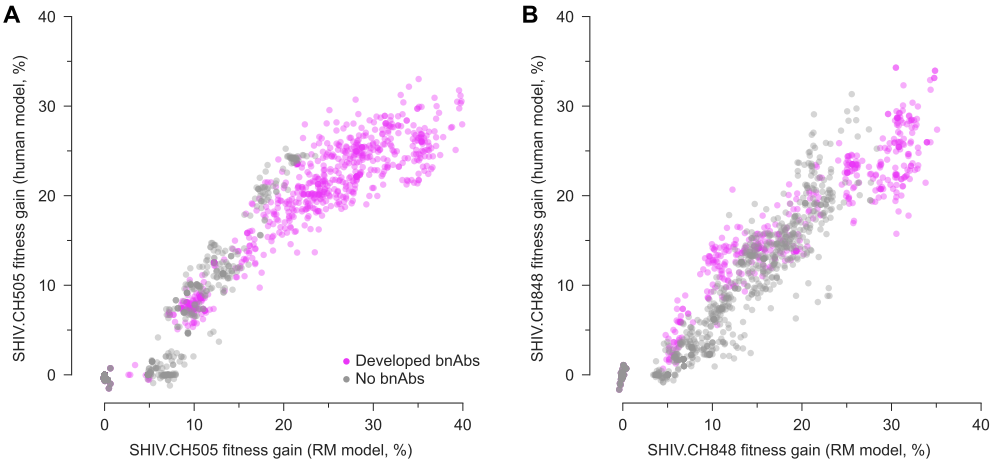


Fig. 5. Inferred fitness landscapes in HIV-1 and SHIV are highly similar. A, Fitness of SHIV.CH505 sequences relative to the TF sequence across 7 RMs, including 3 that developed bnAbs and 4 that did not, evaluated using fitness effects of mutations using data from CH505 and using RM data. The fitness values are strongly correlated, indicating the similarity of Env fitness landscapes inferred using HIV-1 or SHIV data. Values are normalized such that the fitness gain of the TF sequence is zero. **B**, Fitness values for SHIV.CH848 sequences also show strong agreement between CH848 and SHIV.CH848 landscapes.

fitness values for SHIV sequences using selection coefficients inferred from HIV-1 evolution in CH505 and CH848.

We observed a remarkable agreement between SHIV fitness values computed from these two sources (**Fig. 5**). For both SHIV.CH505 and SHIV.CH848, the correlation between viral fitness estimated using data from humans (i.e., CH505 and CH848) and RMs is strongly and linearly correlated (Pearson's $r = 0.96$ and 0.95, $p < 10^{-20}$). This implies that evolutionary pressures on the envelope protein SHIV-infected RMs are highly similar to those on HIV-1 in humans with the same TF Env. In fact, this relationship holds even beyond the SHIV sequences observed during infection. Fitness estimates for sequences with randomly shuffled sets of mutations are also strongly correlated (**Supplementary Fig. 2**).

Evolutionary dynamics forecast antibody breadth

Given the similarity of HIV-1 and SHIV evolution, we sought to identify evolutionary features that distinguish between hosts who develop broad antibody responses and those who do not. **Figure 5** shows that SHIV sequences from hosts with bnAbs often reach higher fitness values than those in hosts with only narrow-spectrum antibodies. We hypothesized that stronger selective pressures on the virus might drive viral diversification, stimulating the development of antibody breadth. Past studies have associated higher viral loads with bnAb development and observed viral diversification around the time of bnAb emergence[14,15]. Computational studies and experiments have also shown that sequential exposure to diverse antigens can induce cross-reactive antibodies[37–39].

To further quantify SHIV evolutionary dynamics, we computed the average fitness gain of viral populations in each RM over time. We observed a striking difference in SHIV fitness gains between RMs that developed broad antibody responses and those that did not (**Fig. 6**). In particular, SHIV fitness increased rapidly before the development of antibody breadth. SHIV fitness gains in RM5695, which developed exceptionally broad and potent antibodies[18], were especially rapid and dramatic. These fitness differences were not attributable to bnAb resistance mutations, which were only moderately selected and generally appeared after bnAbs developed.

One outlier in this pattern is RM6072, infected with SHIV.CH505. Antibody development in RM6072 followed a path that was remarkably similar to CH505, including a lineage of antibodies, DH650, directed toward the CD4 binding site of Env[18]. However, resistance to the DH650 lineage is conferred by a strongly selected mutation that adds a glycan at site 234 (T234N, with an inferred selection coefficient of 4.5%). Broadly neutralizing antibodies similar to DH650 are able to accommodate this glycan due to shorter and/or more flexible light chains[18,40], but DH650 cannot. Antibody evolution in RM6072 thus proceeded along a clear pathway toward bnAb development, but lacked critical mutations to achieve breadth.

Discussion

The induction of bnAbs is a major goal of HIV-1 vaccine design[8]. Both computational[37,39,41,42] and experimental[38,43,44] studies, as well as observations from individuals who developed bnAbs[11,12,20], suggest that the co-evolution of antibodies and HIV-1 is important to stimulate broad antibody responses. However, a general understanding of the coevolutionary pathways by which bnAbs develop remains incomplete. Here, we quantitatively studied the evolution of HIV-1 and SHIV (featuring HIV-1-derived Env sequences) that accompanies the development of bnAbs.

In both HIV-1 and SHIV, we found strong selection for reversions to clade consensus and some mutations that affected N-linked glycosylation motifs or resistance to autologous strain-specific antibodies. Few $CD8^+$ T cell epitopes were identified in this data set, but the T cell escape mutations that we did observe were highly beneficial for the virus. Consistent with past work studying VRC26 escape in CAP256, we observed more modest selection for bnAb resistance mutations[22].

Overall, we found striking similarities between Env evolu-

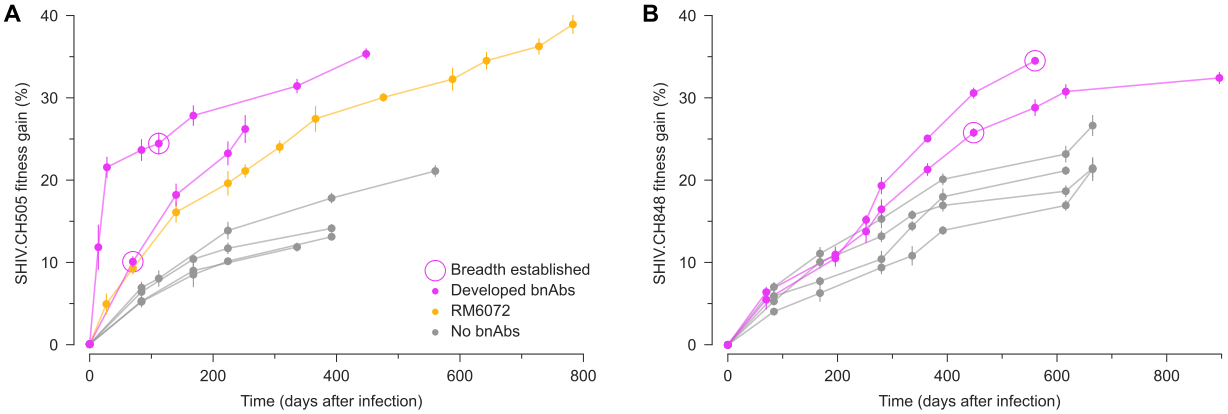


Fig. 6. Rapid SHIV fitness gains precede the development of broadly neutralizing antibodies. For both SHIV.CH505 (**A**) and SHIV.CH848 (**B**), viral fitness gains over time display distinct patterns in RM hosts that developed bnAbs versus those that did not. Notably, the differences in SHIV fitness gains between hosts with and without broad antibody responses appear *before* the development of antibody breadth, and cannot be attributed to selection for bnAb resistance mutations. RM6072 is an unusual case, exhibiting antibody development that was highly similar to CH505. Although RM6072 developed tier 2 nAbs, they lacked key mutations critical for breadth[18]. Points and error bars reflect the mean and standard deviation of fitness gains across SHIV samples in each RM at each time.

tion in humans and RMs. Importantly, these parallels extend beyond the observation of repetitive mutations. Our inferred Env fitness values in humans and RMs were highly correlated, even for artificial sequences with randomly permuted mutations. This indicates that the evolutionary pressures on Env in HIV-1 and SHIV infection are very similar. Our findings reinforce SHIV as a model system that closely mirrors HIV-1 infection.

We discovered that the speed of SHIV fitness gains was clearly higher in RMs that developed broad antibody responses than in those with narrow-spectrum antibodies. Fitness gains in the viral population preceded the development of bnAbs, and they were not driven by bnAb resistance mutations. This suggests that rapid changes in the viral population are a cause rather than a consequence of antibody breadth. While our sample is limited to 13 RMs and two founder Env sequences, we find a clear separation between RMs that did or did not develop antibody breadth. Thus, the dynamics of viral fitness could serve as a quantitative signal associated with bnAb development.

Beyond the specific context of HIV-1 and bnAb development, these studies also provide insight into viral evolution. Despite differences in host biology, the inferred fitness effects of Env mutations in HIV-1 and SHIV were surprisingly similar. This is consistent with methods that aim to use sequence statistics across multiple individuals and hosts to predict the fitness effects of mutations[45–50]. Our study also helps highlight how different classes of mutations (e.g., mutations affecting T cell escape or Env glycosylation) affect fitness *in vivo*.

ACKNOWLEDGEMENTS

The work of K.S.S. 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 research designs, methods development, interpretation of results, and writing the paper. K.S.S. performed simulations and computational analyses. J.P.B. supervised the project.

References

1. Wei, X. *et al.* Antibody neutralization and escape by hiv-1. *Nature* **422**, 307–312 (2003).
2. 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).
3. Li, B. *et al.* Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution. *Journal of virology* **81**, 193–201 (2007).
4. Altfeld, M. & Allen, T. M. Hitting hiv where it hurts: an alternative approach to hiv vaccine design. *TRENDS in Immunology* **27**, 504–510 (2006).
5. Kwong, P. D., Mascola, J. R. & Nabel, G. J. Broadly neutralizing antibodies and the search for an hiv-1 vaccine: the end of the beginning. *Nature Reviews Immunology* **13**, 693–701 (2013).
6. Burton, D. R. & Hangartner, L. Broadly neutralizing antibodies to hiv and their role in vaccine design. *Annual review of immunology* **34**, 635–659 (2016).
7. Sok, D. & Burton, D. R. Recent progress in broadly neutralizing antibodies to hiv. *Nature immunology* **19**, 1179–1188 (2018).
8. Haynes, B. F. *et al.* Strategies for hiv-1 vaccines that induce broadly neutralizing antibodies. *Nature Reviews Immunology* **23**, 142–158 (2023).
9. Haynes, B. F. *et al.* Hiv-host interactions: implications for vaccine design. *Cell host & microbe* **19**, 292–303 (2016).
10. Hraber, P. *et al.* Prevalence of broadly neutralizing antibody responses during chronic hiv-1 infection. *AIDS (London, England)* **28**, 163 (2014).
11. Liao, H.-X. *et al.* Co-evolution of a broadly neutralizing hiv-1 antibody and founder virus. *Nature* **496**, 469–476 (2013).
12. Bonsignori, M. *et al.* Staged induction of hiv-1 glycan–dependent broadly neutralizing antibodies. *Science translational medicine* **9**, eaai7514 (2017).
13. Doria-Rose, N. A. *et al.* Developmental pathway for potent v1v2-directed hiv-neutralizing antibodies. *Nature* **509**, 55–62 (2014).
14. Moore, P. L., Williamson, C. & Morris, L. Virological features associated with the development of broadly neutralizing antibodies to hiv-1. *Trends in microbiology* **23**, 204–211 (2015).
15. Landais, E. & Moore, P. L. Development of broadly neutralizing antibodies in hiv-1 infected elite neutralizers. *Retrovirology* **15**, 1–14 (2018).
16. Gray, E. S. *et al.* The neutralization breadth of hiv-1 develops incrementally over four years and is associated with cd4+ t cell decline and high viral load during acute infection. *Journal of virology* **85**, 4828–4840 (2011).
17. Cornelissen, M. *et al.* The neutralizing antibody response in an individual with triple hiv-1 infection remains directed at the first infecting subtype. *AIDS research and human retroviruses* **32**, 1135–1142 (2016).
18. Roark, R. S. *et al.* Recapitulation of hiv-1 env-antibody coevolution in macaques leading to neutralization breadth. *Science* **371**, eabd2638 (2021).
19. Tomaras, G. D. *et al.* Initial b-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin m (igm) and igg antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. *Journal of virology* **82**, 12449–12463 (2008).
20. Gao, F. *et al.* Cooperation of b cell lineages in induction of hiv-1-broadly neutralizing antibodies. *Cell* **158**, 481–491 (2014).
21. Bonsignori, M. *et al.* Maturation pathway from germline to broad hiv-1 neutralizer of a cd4-mimic antibody. *Cell* **165**, 449–463 (2016).
22. 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).
23. 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).
24. Lee, B. *et al.* Inferring effects of mutations on sars-cov-2 transmission from genomic surveillance data. *medRxiv* 2021–12 (2022).
25. Li, Y. & Barton, J. P. Estimating linkage disequilibrium and selection from allele frequency trajectories. *Genetics* **223**, iyac189 (2023).
26. Li, Y. & Barton, J. P. Correlated allele frequency changes reveal clonal structure and selection in temporal genetic data. *Molecular Biology and Evolution* **41**, msae060 (2024).
27. Jardine, J. G. *et al.* Minimally mutated hiv-1 broadly neutralizing antibodies to guide reductionist vaccine design. *PLoS pathogens* **12**, e1005815 (2016).
28. Zanini, F. *et al.* Population genomics of intrapatient hiv-1 evolution. *Elife* **4**, e11282 (2015).
29. Das, S. R. & Jameel, S. Biology of the hiv nef protein. *Indian J Med Res* **121**, 315–32 (2005).
30. Barton, J. P. *et al.* Modelling and in vitro testing of the hiv-1 nef fitness landscape. *Virus evolution* **5**, vez029 (2019).
31. McCurley, N. P. *et al.* Hiv transmitted/founder vaccines elicit autologous tier 2 neutralizing antibodies for the cd4 binding site. *PLoS One* **12**, e0177863 (2017).
32. Kong, R. *et al.* Antibody lineages with vaccine-induced antigen-binding hotspots develop broad hiv neutralization. *Cell* **178**, 567–584 (2019).
33. Saunders, K. O. *et al.* Stabilized hiv-1 envelope immunization induces neutralizing antibodies to the cd4bs and protects macaques against mucosal infection. *Science translational medicine* **14**, eabo5598 (2022).
34. Bauer, A. *et al.* Adaptation of a transmitted/founder simian-human immunodeficiency virus for enhanced replication in rhesus macaques. *PLoS Pathogens* **19**, e1011059 (2023).
35. Hatziioannou, T. & Evans, D. T. Animal models for hiv/aids research. *Nature Reviews Microbiology* **10**, 852–867 (2012).
36. Li, H. *et al.* Envelope residue 375 substitutions in simian–human immunodeficiency viruses enhance cd4 binding and replication in rhesus macaques. *Proceedings of the National Academy of Sciences* **113**, E3413–E3422 (2016).
37. Wang, S. *et al.* Manipulating the selection forces during affinity maturation to generate cross-reactive hiv antibodies. *Cell* **160**, 785–797 (2015).
38. Escolano, A. *et al.* Sequential immunization elicits broadly neutralizing anti-hiv-1 antibodies in ig knockin mice. *Cell* **166**, 1445–1458 (2016).
39. Sprenger, K. G., Louveau, J. E., Murugan, P. M. & Chakraborty, A. K. Optimizing immunization protocols to elicit broadly neutralizing antibodies. *Proceedings of the National Academy of Sciences* **117**, 20077–20087 (2020).
40. Zhou, T. *et al.* Multidonor analysis reveals structural elements, genetic determinants, and maturation pathway for hiv-1 neutralization by vrc01-class antibodies. *Immunity* **39**, 245–258 (2013).
41. Shaffer, J. S., Moore, P. L., Kardar, M. & Chakraborty, A. K. Optimal immunization cocktails can promote induction of broadly neutralizing abs against highly mutable pathogens. *Proceedings of the National Academy of Sciences* **113**, E7039–E7048 (2016).
42. Nourmohammad, A., Otwinowski, J. & Plotkin, J. B. Host-pathogen coevolution and the emergence of broadly neutralizing antibodies in chronic infections. *PLoS genetics* **12**, e1006171 (2016).
43. Dosenovic, P. *et al.* Immunization for hiv-1 broadly neutralizing antibodies in human ig knockin mice. *Cell* **161**, 1505–1515 (2015).
44. Williams, W. B. *et al.* Vaccine induction of heterologous hiv-1 neutralizing antibody b cell lineages in humans. *medRxiv* 2023–03 (2023).
45. Ferguson, A. L. *et al.* Translating hiv sequences into quantitative fitness landscapes predicts viral vulnerabilities for rational immunogen design. *Immunity* **38**, 606–617 (2013).
46. 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).
47. Łuksza, M. & Lässig, M. A predictive fitness model for influenza. *Nature* **507**, 57–61 (2014).
48. 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).
49. 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).
50. Hie, B., Zhong, E. D., Berger, B. & Bryson, B. Learning the language of viral evolution and escape. *Science* **371**, 284–288 (2021).

Methods

Data

We retrieved HIV-1 sequences from CH505 (703010505) and CH848 (703010848) from the HIV sequence database at Los Alamos National Laboratory (LANL)[51]. The rhesus macaque (RM) SHIV sequences[52] were obtained from GenBank[53]. We then co-aligned SHIV.CH505 and SHIV.CH848 sequences with CH505 and CH848 HIV-1 sequences, respectively, using HIValign[54].

CH505. CH505 developed two distinct lineages of CD4 binding site (CD4bs) bnAbs, CH103 and CH235 (refs.[55,56]). CH103 antibodies were detectable by 14 weeks after infection and further developed neutralization breadth between 41-92 weeks[57]. IC50 values of CH235 against the TF virus were 6.5-fold lower than those of CH103 (ref.[55]). CH235 lineages could neutralize autologous viruses at week 30. However, viruses that acquired mutations at loop D from 53-100 weeks escaped CH235 neutralization[55]. Although the neutralization breadth of CH235 was not as broad as that of CH103, this lineage played a critical role; escaping mutations from the CH235 lineage stimulated the development of another lineage with broader neutralization depth[55]. Mutations in loop D enabled the virus to escape from CH235, but these sequential mutations in loop D, such as E275K, N279D, and V281S, favorably bound to the mature CH103 and continuously increased the binding affinity between mature CH103 and loop D[55]. Gradually, CH103 matured, developing a broader neutralization breadth.

CH848. CH848 developed DH270, a bnAb that targets the glycosylated site adjacent to the third variable loop (V3). DH270 was detectable three and a half years after infection[58]. Similar to the CH505 case, the CH848 case exhibited cooperative virus and antibody coevolution. The earlier antibody lineages, DH272 and DH475, could neutralize autologous viruses until week 51 and weeks 15-39, respectively. The virus escaped from DH272 and DH475 afterward, with escape mutations including a longer V1V2 loop. DH270 then developed, with potent and broad neutralization breadth[58].

Rhesus macaques. Chimeric viruses, SHIVs, were constructed by bearing the transmitted/founder (TF) Env from three HIV-1 patients, including CH505 and CH848 (ref.[52]). In some RMs, SHIV developed similar patterns of mutations to those observed in human donors. In our analysis, we considered RMs with SHIV sequences sampled at at least three points in time. This yielded a set of 7 RMs and 6 RMs for SHIV.CH505 and SHIV.CH848, respectively. The **Tables 1-2** summarize the number of sequences, time points, and the development of bnAbs for each individual in the SHIV cases as well as HIV-1 cases.

Sequence data processing

Data quality control. To focus our analysis on functional sequences, we removed sequences with more than 200 gaps. To eliminate rare insertions or possible alignment errors, we also masked sites where gaps occurred in more than 95% of sequences within each individual host. To limit errors in virus frequencies, we only considered data from time points with four or more sequences.

identifier	# of sequences	# of time points	# of variable sites	bnAbs detected	bnAb type
703010505	578	23	1131	**Yes**	CD4bs
RM5695	330	8	467	**Yes**	V2 Apex
RM6070	215	5	334	**Yes**	V2 Apex
RM6072	355	13	710	No	
RM6697	228	6	386	No	
RM6699	118	5	207	No	
RM6701	98	5	148	No	
RM6703	131	4	275	No	

Table 1. CH505 and SHIV.CH505 sequence statistics. The number of sequences and time points listed here are given after filtering. Variable sites are defined as sites where at least two different nucleotides (including deletions) were observed.

identifier	# of sequences	# of time points	# of variable sites	bnAbs detected	bnAb type
703010848	1205	31	2694	**Yes**	V3-glycan
RM6163	261	8	467	**Yes**	V3-glycan
RM6167	349	10	657	**Yes**	V3-glycan
RM6700	260	7	568	No	
RM6713	281	7	567	No	
RM6714	284	8	516	No	
RM6720	193	8	340	No	

Table 2. CH848 and SHIV.CH848 sequence statistics. The content of this table is the same as in **Table 1**.

Identifying reversions. A mutation is classified as a reversion if the new (mutant) nucleotide matches with the nucleotide at the same site in the HIV-1 consensus sequence from the same subtype. Here, all viruses were subtype C, so we compared with the subtype C consensus sequence as defined by LANL.

Identifying mutations that affect N-linked glycosylation. To identify mutations that affect glycosylation, we search for Env mutations that modify the N-linked glycosylation motif Asn-X-Ser/Thr, where X can be any amino acid except proline. We identified three types of mutations affecting glycosylation: "shields," which complete a previously incomplete glycosylation motif, "holes," which disrupt an existing glycosylation motif, and "shifts," which simultaneously complete one N-linked glycosylation motif and disrupt another.

Enrichment analysis. We used fold enrichment values and Fisher's exact test to quantify the excess or lack of mutations. For a particular subset of mutations (for example, the top $x\%$ beneficial mutations), we first computed the number of mutations in that subset that do (n_{sel}) and do not (N_{sel}) have a particular property (for example, nonsynonymous mutations in the CD4 binding site). We then computed the total number of mutations that do and do not have the property (n_{null} and N_{null}, respectively) across the entire data set. The fold enrichment value is then $\frac{n_{\text{sel}}/N_{\text{sel}}}{n_{\text{null}}/N_{\text{null}}}$. The term $\frac{n_{\text{sel}}}{N_{\text{sel}}}$ quantifies the fraction of mutations having specific properties across the selected mutations, while the denominator $\frac{n_{\text{null}}}{N_{\text{null}}}$ is the fraction of all mutations that have the property. Fisher's exact p values are computed from the 2×2 table with n_{sel}, $n_{\text{null}} - n_{\text{sel}}$

in the first row, and $N_{\text{sel}} - n_{\text{sel}}$, $N_{\text{null}} - N_{\text{sel}} - (n_{\text{null}} - n_{\text{sel}})$ in the second row[59].

Inferring fitness effects of mutations

In this section, we describe the inference framework used to obtain selection coefficients, referred to as the fitness effects of mutations, from temporal genetic data.

Evolutionary model. We model viral evolution with the Wright-Fisher (WF) model, a fundamental model in population genetics[60]. In this model, a population of N individuals (viruses or infected cells, in our case) undergo discrete rounds of selection, mutation, and replication. Each individual a has a genetic sequence $g^a = (g_1^a, g_2^a, \ldots, g_L^a)$. Here L is the length of the genetic sequence (number of loci). We write the number of states at each locus (i.e., nucleotides or amino acids) as q. For simplicity, we will use $q = 2$ and $g_i \in \{0,1\}$ to describe the model below, though we use $q = 5$ and $q = 21$ for DNA and amino acid sequences, respectively, in real HIV-1 and SHIV data. In this binary representation, 0 represents a "wild-type" (WT) or TF allele, and 1 is a mutant allele.

We define the fitness of an individual with genetic sequence g by

$$F(g|s) = 1 + \sum_{i=1}^{L} s_i g_i \,. \tag{1}$$

Here s_i is a selection coefficient, quantifying the fitness effect of a mutant allele at locus i. If $s_i > 0$, the mutant allele is beneficial (enhancing replication), and if $s_i < 0$ it is deleterious (impairing replication). By convention, we set the selection coefficient for TF alleles to zero. Individuals with higher fitness values are more likely to replicate than those with lower fitness.

Mutations introduce new genotypes and drive the evolution of the population. Let us define μ^{ab} as the probability of mutation from genotype a to genotype b per replication cycle. Below, we will express this probability in terms of a mutation rate per site per round of replication. In the analysis of real data, we use asymmetric mutation rates estimated from intra-host HIV-1 data[61].

Given these parameters, the WF model describes the dynamics of the frequencies of different genotypes in the population over time. We write the frequency of genotype a at time t as $z_a(t)$. Given that the frequency of genotypes in the population at time t is $z(t) = (z_1(t), z_2(t), \ldots, z_M(t))$, where M is the total number of genotypes, the probability distribution of the frequency of genotypes in the next generation $z(t+1)$ is

$$P(z(t+1)|z(t); s, \mu, N) = N! \prod_{a=1}^{M} \left(\frac{p_a(z(t))^{N z_a(t+1)}}{[N z_a(t+1)]!} \right) . \tag{2}$$

Here p_a is

$$p_a(z(t)) \propto F^a z_a + \sum_{b(\neq a)} [\mu^{ba} z_b(t) - \mu^{ab} z_a(t)] \,. \tag{3}$$

where F^a is the fitness value of genotype a, based on Eq. (1). Across K generations, the probability of an entire evolutionary trajectory, defined by the vector of genotype frequencies at each time, is then

$$P((z(t_k))_{k=0}^{K}|s, \mu, N) = \prod_{k=0}^{K-1} P(z(t_{k+1})|z(t_k); s, \mu, N) \,. \tag{4}$$

Diffusion limit. When the population size is sufficiently large, the evolution of the population defined in Eq. (2) can be reasonably well approximated by a Gaussian process, which is a solution to the Fokker-Planck (forward Kolmogorov) equation[62,63].

$$P(z(t+\Delta t)|z(t); s, \mu, N) \sim \mathcal{N}\Big(z(t) + \Delta t d(t), C(z(t))/N\Big), \tag{5}$$

with the drift vector $d(t)$ and the diffusion matrix $C(z(t))/N$ such that

$$C_{ab}(z(t)) = \begin{cases} -z_a(t) z_b(t) & \text{for } a \neq b \\ z_a(t)(1 - z_a(t)) & \text{for } a = b \,, \end{cases} \tag{6}$$

and

$$d(t) = C(z(t)) s + \mu^{\text{flux}} \,. \tag{7}$$

Dimensional reduction. While the WF process in genotype space provides valuable insights into genotype dynamics, the mathematical expressions are sometimes challenging to interpret. To obtain more intuitive expressions, we can project the dynamics onto the space of allele frequencies,

$$x_i = \sum_a g_i^a z_a \,. \tag{8}$$

One can then find the drift vector d and diffusion matrix C/N in allele frequency space,

$$d_i = C_{ii} s_i + \sum_{j(\neq i)} C_{ij} s_j + \mu_i^{\text{flux}} \,, \tag{9}$$

and

$$C_{ij} = \begin{cases} x_{ij} - x_i x_j & \text{for } i \neq j \\ x_i(1 - x_i) & \text{for } i = j \,. \end{cases} \tag{10}$$

Here, x_{ij} is the frequency of individuals with alleles i and j, and μ_i^{flux} is net expected change in frequency of allele i due to mutations, which is given explicitly in Eq. (14) below. The first term in Eq. (9) gives the expected change in the frequency x_i due to the direct fitness effect s_i, while the second term represents the contributions due to indirect or genetic linkage effects with other alleles j.

Maximum path likelihood. Following recent work[64], we employed Bayes' rule to find the selection coefficients that best explain the data. These are the coefficients that maximize the posterior distribution

$$P_{\text{posterior}}(s|(z(t_k))_k) \propto P((z(t_k))_k|s) \; P_{\text{prior}}(s) \,, \tag{11}$$

which is a product of the likelihood of the evolutionary trajectory observed in the data Eq. (4) (under the diffusion limit Eq. (5)) and a prior distribution for the selection coefficients. We chose a Gaussian prior distribution with zero mean and a covariance of $I/(N\gamma)$, where I is the identity matrix. This prior distribution penalizes the inferences of large selection coefficients when they are not well-supported by the data. The maximum *a posteriori* selection coefficients are then given by[64]

$$\hat{s} = (C_{\rm int} + \gamma I)^{-1} \left[\Delta x_{\rm int} - \Delta\mu_{\rm int}\right]. \tag{12}$$

Here $C_{\rm int}$, $\Delta x_{\rm int}$, and $\Delta\mu_{\rm int}$ represent the covariance matrix, vector of frequency changes, and mutational flux integrated over the evolution

$$\begin{aligned} C_{\rm int} &= \sum_{k=0}^{K} \Delta t_k C(x(t_k)) \\ \Delta x_{\rm int} &= x(t_K) - x(t_0) \\ \Delta\mu_{\rm int} &= \sum_{k=0}^{K-1} \Delta t_k \mu^{\rm flux}(t_k). \end{aligned} \tag{13}$$

The mutational flux $\mu^{\rm flux}$ is characterized by the rates of mutations from nucleotides β to α, denoted by $\mu_{\alpha\beta}$, which are determined from longitudinal HIV-1 populations in untreated patients[61]. The change of the α nucleotide frequency at locus i due to mutation is given by:

$$\mu^{\rm flux}_{i,\alpha} = \sum_{\beta} (\mu_{\alpha\beta} x_{i,\beta}(t_k) - \mu_{\beta\alpha} x_{i,\alpha}(t_k)). \tag{14}$$

Inverting the integrated covariance matrix effectively reveals the underlying direct allele interactions and resolves the genetic linkage effects.

The shift in the covariance diagonal in Eq. (12), arising from the selection coefficients' posterior distribution, reflects the uncertainty in the selection distribution. We used $\gamma = 10$ for all simulations, but the model is robust to variation in the strength of regularization[64]. For the mutation rates, we incorporated the transition probabilities among arbitrary DNA nucleotides, estimated from whole-genome deep sequencing of multiple untreated HIV-1 patients followed for 5 to 8 years post-infection[61].

Integration of covariance. When the time interval of the observation Δt is sufficiently short, the trajectory of the allele frequency would be continuous and ideally it would be a smooth curve[65]. To accurately estimate the covariance matrix, we employ piecewise linear interpolation for frequencies. Let $\tau \in [0,1]$, then the linear interpolation for a frequency vector can be expressed as:

$$\begin{aligned} x_i^{[k,k+1]}(\tau) &= (1-\tau)x_i(t_k) + \tau x_i(t_{k+1}), \\ x_{ij}^{[k,k+1]}(\tau) &= (1-\tau)x_{ij}(t_k) + \tau x_{ij}(t_{k+1}), \end{aligned} \tag{15}$$

which yields ,

$$\begin{aligned} x_i^{\rm int} &= \sum_{k=0}^{K-1} \Delta t_k \int_0^1 d\tau x_i^{[k,k+1]}(\tau) \\ C_{ij}^{\rm int} &= \sum_{k=0}^{K-1} \Delta t_k \int_0^1 d\tau \left(x_{ij}^{[k,k+1]}(\tau) - x_i^{[k,k+1]}(\tau) x_j^{[k,k+1]}(\tau) \right). \end{aligned}$$

The explicit expression of the integrated covariance is given in refs.[64,65].

Joint RM model

In addition to fitness models derived from SHIV data for individual RMs, we inferred a joint model under the assumption that virus evolution within each individual RM with the same TF virus is governed by a similar fitness landscape. The joint path likelihood for allele frequency trajectories across RMs with the same TF virus is then

$$\begin{aligned} &p(((x^\alpha(t_k))_{k=0}^{K_\alpha})_{\alpha=1}^{R} | s, \mu, \gamma) \\ &= \left(\prod_{\alpha=1}^{R} \prod_{k=0}^{K-1} p_{\rm M}(x^\alpha(t_{k+1}) | x^\alpha(t_k); s, \mu, N) \right) \\ &\quad \times p(x(t_0))\, p(s|\gamma). \end{aligned} \tag{16}$$

Here, x^α is the allele frequency of the α-th individual and R is the number of replicate individuals (i.e., the number of RMs sharing the same TF virus). The initial state is $p(x^\alpha(t_0)) = p(x(t_0))$ for all α for individuals with the same TF virus. The solution of the joint path likelihood is given by

$$\hat{s} = (\overline{C}_{\rm int} + \gamma I)^{-1} [\overline{\Delta x}_{\rm int} - \overline{\Delta\mu}_{\rm int}]. \tag{17}$$

Here, the overbar denotes the sum over the replicate RMs.

We emphasize that the joint selection coefficients in Eq. (17) are not the same as selection coefficients that are simply averaged across RMs with the same TF virus. The joint selection coefficients are more robust, as they are guided by the level of evidence within each individual rather than naive averaging.

Geometrical interpretation of the fitness comparison

The Pearson values we utilized to compare the fitness landscapes denoted as $F_s(g) := F(g|s)$ and $F_h(g) := F(g|h)$ can be expressed by the following simple relation:

$$r = \frac{{\rm Cov}(F_s, F_h)}{\sqrt{{\rm Var}(F_s)}\sqrt{{\rm Var}(F_h)}} = \frac{s^\top C h}{\sqrt{s^\top C s}\sqrt{h^\top C h}}. \tag{18}$$

Here, ${\rm Cov}(F_s, F_h)$ and ${\rm Var}(F_s)$ represent the covariance and variance values estimated from the samples being compared, $(F_s(g^n), F_h(g^n))_n$. C is the covariance matrix defined between arbitrary loci. The last equation can be interpreted as an angle between two vectors, s and h, with a metric matrix C; if $s = h$, the Pearson value clearly becomes 1. However, the "similarity" also depends on how these vectors are projected by C; eigenmodes associated with larger variance of statistics will be more emphasized. The last expression

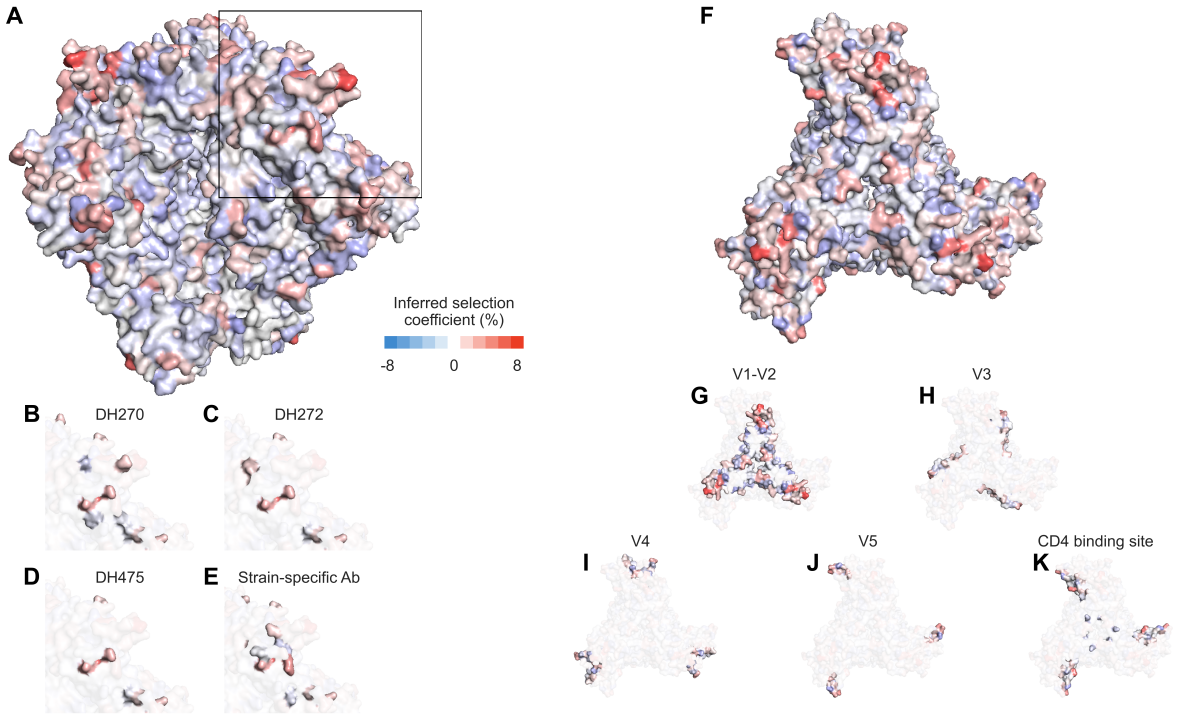
readily implies an interpretation for the case of shuffled sequences; shuffling the sequences equates to diluting the covariance between loci, resulting in the metric matrix becoming a diagonal matrix. Removing the off-diagonal elements corresponds to lifting the constraints on the fitness landscape.

Code availability

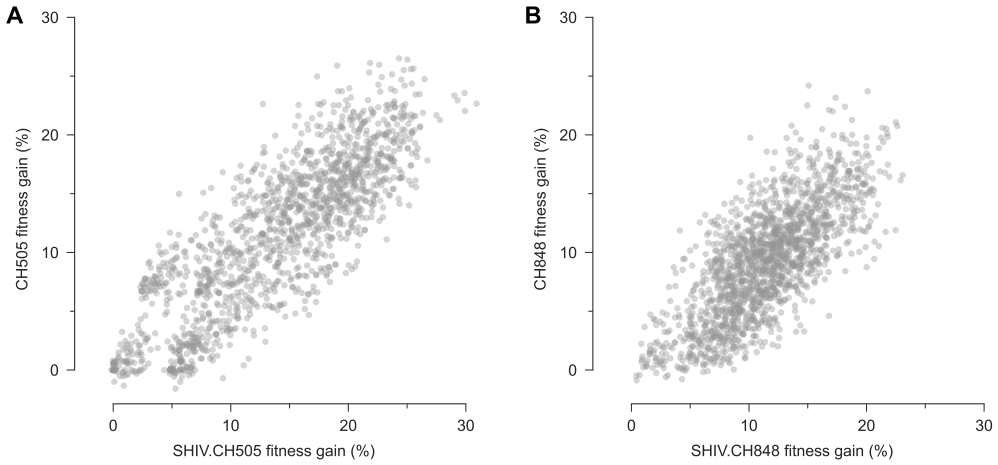
Sets of processed data and computer code used in this study are available in the GitHub repository of https://github.com/bartonlab/paper-HIV-coevolution. This repository contains source files that process HIV-1 and SHIV sequences, infer selection coefficients, and identify and characterize mutations. The included Jupyter Notebooks can be run to reproduce the figures presented here. The original HIV-1 sequences can be retrieved from the LANL database (https://www.hiv.lanl.gov/content/index), and SHIV sequences can be found at GenBank (https://www.ncbi.nlm.nih.gov/).

References

51. Los Alamos National Laboratory. Hiv sequence database (2023). URL https://www.hiv.lanl.gov. Accessed: 703010505 and 703010848 inpatient code.
52. Roark, R. S. *et al.* Recapitulation of hiv-1 env-antibody coevolution in macaques leading to neutralization breadth. *Science* **371**, eabd2638 (2021).
53. Benson, D. A. *et al.* Genbank. *Nucleic acids research* **41**, D36–D42 (2012).
54. Los Alamos National Laboratory. Hivalign: Hiv sequence alignment tool (2023). URL https://www.hiv.lanl.gov/content/sequence/VIRALIGN/viralign.html.
55. Gao, F. *et al.* Cooperation of b cell lineages in induction of hiv-1-broadly neutralizing antibodies. *Cell* **158**, 481–491 (2014).
56. Kreer, C. *et al.* Probabilities of developing hiv-1 bnab sequence features in uninfected and chronically infected individuals. *Nature Communications* **14**, 7137 (2023).
57. Liao, H.-X. *et al.* Co-evolution of a broadly neutralizing hiv-1 antibody and founder virus. *Nature* **496**, 469–476 (2013).
58. Bonsignori, M. *et al.* Staged induction of hiv-1 glycan–dependent broadly neutralizing antibodies. *Science translational medicine* **9**, eaai7514 (2017).
59. Ruxton, G. D. & Neuhäuser, M. Good practice in testing for an association in contingency tables. *Behavioral Ecology and Sociobiology* **64**, 1505–1513 (2010).
60. Ewens, W. J. *Mathematical population genetics: theoretical introduction*, vol. 1 (Springer, 2004).
61. Zanini, F. *et al.* Population genomics of intrapatient hiv-1 evolution. *Elife* **4**, e11282 (2015).
62. Kimura, M. Diffusion models in population genetics. *Journal of Applied Probability* **1**, 177–232 (1964).
63. Crow, J. F. *An introduction to population genetics theory* (Scientific Publishers, 2017).
64. 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).
65. Shimagaki, K. & Barton, J. P. Bézier interpolation improves the inference of dynamical models from data. *Physical Review E* **107**, 024116 (2023).
66. Saunders, K. O. *et al.* Stabilized hiv-1 envelope immunization induces neutralizing antibodies to the cd4bs and protects macaques against mucosal infection. *Science translational medicine* **14**, eabo5598 (2022).
67. Bricault, C. A. *et al.* Hiv-1 neutralizing antibody signatures and application to epitope-targeted vaccine design. *Cell host & microbe* **25**, 59–72 (2019).
68. Bauer, A. *et al.* Adaptation of a transmitted/founder simian-human immunodeficiency virus for enhanced replication in rhesus macaques. *PLoS Pathogens* **19**, e1011059 (2023).



Supplementary Fig. 1. CH848 exhibits similar spatial patterns of selection coefficients; significantly strongly selected mutations are located at the apex, particularly on the edge of the Env protein. **A** and **B** Side and top views of the Env protein with inferred selection coefficient values. The apex region is enriched with moderately and strongly beneficial mutations. **C-F** Mutations that are presumably resistant to bnAbs and autologous nAbs are highlighted. The selection values in the autologous nAbs binding regions were slightly larger than those in the bnAbs binding regions. **G-K** The top views highlight the individual variable regions.



Supplementary Fig. 2. Broad similarity between HIV-1 and SHIV fitness landscapes with the same TF Env sequence. A, Fitness estimates for a sample of artificial Env sequences, obtained by independently shuffling observed amino acids in SHIV.CH505 sequences at each residue. This random sequence ensemble conserves single-residue frequencies, but not correlations between mutations. Even on these artificial sequences, the fitness estimates using a model trained on HIV-1 data strongly agree with the SHIV model (Pearson's $r = 0.84$, $p < 10^{-20}$). This implies that the similarity of the fitness landscapes is not confined to the specific genotypes observed in HIV-1 or SHIV evolution, but also extends to more distant sequences. **B** Similar results also hold for fitness landscapes based on CH848 and SHIV.CH848 data (Pearson's $r = 0.74$, $p < 10^{-20}$).

Rank	Mutation (DNA)	Mutation (AA)	Date	Region	Selection coefficient (%)	Note
1	A7225G	N334S	113	Near V3 Loop	8.92	Reversion, N-linked glycan shift, resistance to DH950 antibodies[66]
2	T7066A	V281D	183	CD4bs/Loop D	5.66	Resistant to intermediate CH235 lineage antibodies[55], combination with N280S or N280T gives resistance to mature CH235 (ref.[55])
3	G7776A	V518M	1093	Fusion Peptide	5.42	
4	C6709T	T162I	129	V2 Loop	5.40	N-linked glycan hole
5	G6682A	G153E	71	V1 Loop	5.24	Reversion
6	C7462A	T413N	129	V4 Loop	5.16	N-linked glycan shift
7	C7468A	T415K	71	V4 Loop	4.82	Putative $CD8^+$ T cell escape mutation[55]
8	A7130C	K302N	183	V3 Loop	3.74	Reduces neutralization of DH950-like Abs[66]
9	G7769A	M515I	1603	Fusion Peptide	3.57	Reversion
10	A7614G	N464D	183	V5 Loop	3.46	Reversion
11	C7462T	T413I	23	V4 Loop	3.42	Putative $CD8^+$ T cell escape mutation[55]
12	G7776T	V518L	1640	Fusion Peptide	3.37	
13	T7212C	Y330H	129	V3 Loop	3.28	Reversion, reduces neutralization of DH950-like Abs[66]
14	G6661A	S146N	129	V1 Loop	3.14	Reversion, N-linked glycan shield
15	G7453A	S410N	23	V4 Loop	3.08	Putative $CD8^+$ T cell escape mutation[57]
16	G6634T	S137I	183	V1 Loop	3.00	N-linked glycan hole
17	G6634C	S137A	129	V1 Loop	2.82	Reversion, N-linked glycan hole
18	T6588A	L122M	183	Near CD4bs	2.81	
19	C7456A	T411K	29	V4 Loop	2.75	Reversion, N-linked glycan hole, and putative $CD8^+$ T cell escape mutation[57]
20	A6596C	P124P	183	CD4bs	2.73	
21	A7474G	H417R	23	V4 Loop	2.63	Putative $CD8^+$ T cell escape mutation[57]
22	A7601dG	N459D	1603	CD4bs	2.61	N-linked glycan shield, located in the CH235 antibody binding region[57]
23	A7056T	T278S	1603	Loop D	2.59	Located in the CH235 antibody binding region[57]
24	G6630A	A136T	113	V1 Loop	2.59	Reversion
25	A8143G	N640S	1780	Near Glycosite	2.55	

Table 3. Biological effects of the HIV-1 mutations inferred to be the most beneficial in CH505. The table lists the 25 mutations in CH505 with the highest (i.e., most beneficial) selection coefficients, among over 2500 mutations in total. All of the top mutations are nonsynonymous. Mutations associated with immune evasion, changes in glycosylation, and reversions to the subtype C consensus sequence are frequently observed among these top mutations.

Rank	Mutation (DNA)	Mutation (AA)	Date	Region	Selection coefficient (%)	Note
1	T9043G	S83A	49	Nef	12.55	Reversion
2	A8988cT	E65D	38	Nef	8.28	
3	G8911A	E39K	38	Nef	6.68	Reversion
4	A8988bG	E65G	38	Nef	6.42	
5	A8988cC	E65D	43	Nef	5.48	
6	G6640A	S139N	358	V1 Loop	5.22	Reversion, N-linked glycan shift, presence of N-linked glycan at site 138 decreases neutralization by the earlier DH270 UCA and removes glycan at site 137 (ref.[66])
7	G6679A	G152E	38	V1 Loop	4.83	
8	C9044G	A84G	78	Nef	4.70	Reversion
9	A8915G	H41R	38	Nef	4.48	Reversion
10	A8083T	K620M	808	gp41	4.44	
11	A8988cG	E65E	38	Nef	4.35	
12	G7123A	S300N	700	V3 Loop	4.06	Reversion, confers resistance to DH272 and DH475 (ref.[58])
13	G6640T	S139V	445	V1 Loop	3.92	N-linked glycan hole
14	A7310C	K362N	780	V1 Loop	3.87	N-linked glycan shield
15	A8083C	K620T	780	gp41	3.84	
16	A8912C	E40A	32	Nef	3.83	
17	A8895G	A34A	49	Nef	3.77	Reversion
18	A8912T	E40V	43	Nef	3.74	
19	G6640C	S139T	1432	V1 Loop	3.73	N-linked glycan shield
20	A6621C	N133H	1120	V1 Loop	3.66	N-linked glycan hole
21	G8911C	E39Q	49	Nef	3.64	
22	C7066T	A281V	1651	CD4BS	3.62	Reversion
23	A6912G	N230D	194	Glycosite	3.57	N-linked glycan hole, amino acid N is sensitive to DH270.5 (ref.[67])
24	Δ8911G	Δ39E	78	Nef	3.57	
25	A7614G	T464E	445	V5 Loop	3.39	Reversion, N-linked glycan hole

Table 4. Biological effects of the HIV-1 mutations inferred to be the most beneficial in CH848. This table is analogous to **Table 3**, but for CH848. Unlike for CH505, for CH848 longer sequences were available, including genes beyond Env. We observed multiple beneficial mutations in Nef and Env.

Rank	Mutation (AA)	Region	Selection coefficient (%)	Note
1	N334S	Near V3 Loop	3.17	N-linked glycan shift, enhanced viral replication[68], reduced early plasma neutralization[52]
2	Y330H	V3 Loop	2.55	Improved virus entry and replication[68]
3	K302N	V3 Loop	2.20	Enhanced replication and affects virus entry[68]
4	N279D	CD4BS	1.92	Enhanced replication[68], reduced early plasma neutralization[52]
5	H356N	Glycosite	1.69	
6	H417R	V4 Loop	1.66	Enhanced replication[68], putative $CD8^+$ T cell escape mutation[52,68]
7	K462E	V5 Loop	1.61	
8	T234N	Glycosite	1.58	N-linked glycan shield, reduced early plasma neutralization[52]
9	G153E	V1 Loop	1.27	
10	N130D	Near V1 Loop	1.08	Reduced early plasma neutralization[52]

Table 5. Biological effects of strongly selected SHIV.CH505 mutations. Five out of the top six mutations have been shown to increase viral load in RMs[68]. Selection coefficients and ranks are based on the joint evolutionary model across SHIV.CH505-infected RMs.

Rank	Mutation (AA)	Region	Selection coefficient (%)	Note
1	P179L	V2 Loop	2.03	
2	N413T	V4 Loop	1.64	N-linked glycan hole, loss of N413 alters sensitively to early nAbs in RM6167 (ref.[52])
3	N462S	V4 Loop	1.59	N-linked glycan shift
4	R336E	Near V3 Loop	1.48	Mutations at site 336 observed in all RMs, R336G reduced sensitivity to plasma in RM6167 (ref.[52])
5	N413S	V4 Loop	1.45	N-linked glycan hole, aids in antibody escape[52]
6	T459aI	V4 Loop	1.44	
7	N462D	V4 Loop	1.39	N-linked glycan hole
8	Δ459aT	V4 Loop	1.35	
9	S407N	V4 Loop	1.28	
10	R336K	Near V3 Loop	1.25	Mutations at site 336 observed in all RMs, R336G reduced sensitivity to plasma in RM6167 (ref.[52])

Table 6. Biological effects of strongly selected SHIV.CH848 mutations. Selection coefficients and ranks are based on the joint evolutionary model across SHIV.CH848-infected RMs.