Host-pathogen co-evolution and the emergence of broadly neutralizing antibodies in chronic infections

Armita Nourmohammad[1]*†, Jakub Otwinowski[2]†, Joshua B Plotkin[2]

[1] *Joseph-Henri Laboratories of Physics and Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA*

[2] *Department of Biology, University of Pennsylvania, Philadelphia, PA 19104*

The vertebrate adaptive immune system provides a flexible and diverse set of molecules to neutralize pathogens. Yet, viruses that cause chronic infections, such as HIV, can survive by evolving as quickly as the adaptive immune system, forming an evolutionary arms race within a host. Here we introduce a mathematical framework to study the co-evolutionary dynamics of antibodies with antigens within a patient. We focus on changes in the binding interactions between the antibody and antigen populations, which result from the underlying stochastic evolution of genotype frequencies driven by mutation, selection, and drift. We identify the critical viral and immune parameters that determine the distribution of antibody-antigen binding affinities. We also identify definitive signatures of co-evolution that measure the reciprocal response between the antibody and viruses, and we introduce experimentally measurable quantities that quantify the extent of adaptation during continual co-evolution of the two opposing populations. Finally, we analyze competition between clonal lineages of antibodies and characterize the fate of a given lineage dependent on the state of the antibody and viral populations. In particular, we derive the conditions that favor the emergence of broadly neutralizing antibodies, which may be used in designing a vaccine against HIV.

Introduction

It takes decades for humans to reproduce, but our pathogens can reproduce in less than a day. How can we coexist with pathogens whose potential to evolve is 10^4-times faster than our own? The answer no doubt lies in the vertebrate adaptive immune system, which uses recombination, mutation, and selection to evolve a response on the same time-scale at which pathogens themselves evolve.

One of the central actors in the adaptive immune system are B-cells, which recognize pathogens using highly diverse membrane-bound receptors. Naive B-cells are created by processes which generate extensive genetic diversity in their receptors via recombination, insertions and deletions, and hypermutations [1] which can potentially produce $\sim 10^{18}$ variants in a human repertoire [2]. This estimate of potential lymphocyte diversity outnumbers the total population size of B-cells in humans, i.e., $\sim 10^{10}$ [3, 4]. During an infection, B-cells aggregate to form *germinal centers*, where they hypermutate at a rate of about $\sim 10^{-3}$ per base pair per cell division over a region of 1-2 kilo base pairs [5]. The B-cell hypermutation rate is approximately $4-5$ orders of magnitude larger than an average germline mutation rate per cell division in humans [6]. Mutated B-cells compete for survival and proliferation signals from helper T-cells, based on the B-cell receptor's binding to antigens. This form of natural selection is known as *affinity maturation*, and it can increase binding affinities up to 10-100 fold [7–9], see Fig. 1. B-cells with high binding affinity may leave germinal centers to become antibody secreting plasma cells, or dormant memory cells that can be reactivated quickly upon future infections [1]. Secreted antibodies, which are the soluble form of B-cell receptors, can bind directly to pathogens to mark them for neutralization by other parts of the immune system. Plasma B-cells may recirculate to other germinal centers and undergo further hypermutation [8].

Some viruses, such as seasonal influenza viruses, evolve quickly at the population level, but the adaptive immune system can nonetheless remove them from any given host within a week or two. By contrast, chronic infections can last for decades within an individual, either by pathogen dormancy or by pathogens avoiding neutralization by evolving as rapidly as B-cell populations. HIV mutation rates, for example, can be as high as $0.1-0.2$ per generation per genome [10]. Neutralizing assays and phylogenetic analyses suggest an evolutionary arms race between B-cells and HIV populations during infection in a single patient [11–14]. Viruses such as HIV have evolved to keep the sensitive regions of their structure inaccessible by the immune system e.g., through glycan restriction or immuno-dominant variable loops [15, 16]. As a result, the majority of selected antibodies bind to the most easily accessible regions of the virus, where viruses can tolerate mutations and thereby escape immune challenge. Nonetheless, a remarkably large proportion of HIV patients ($\sim 20\%$) eventually produce antibodies that neutralize a broad panel of virions [17, 18] by attacking structurally conserved regions, such as the CD4 binding site of HIV *env* protein [13, 19–22]. These broadly neutralizing antibodies (BnAbs), can even neutralize HIV viruses

*Correspondence should be addressed to: Armita Nourmohammad (armitan@princeton.edu).

†Authors with equal contribution

from other clades, suggesting it may be possible to design an effective HIV vaccine if we can understand the conditions under which BnAbs arise [13, 19, 22–26].

Recent studies have focused on mechanistic modeling of germinal centers in response to one or several antigens [7, 27], and elicitation of BnAbs [26, 28]. However, these studies did not model the co-evolution of the virus and B-cell repertoire, which is important to understand how BnAbs arise *in vivo*. Modeling of such co-evolution is difficult because the mechanistic details of germinal center activity are largely unknown [14], and the multitude of parameters make it difficult to identify generalizable aspects of a model. While evidence of viral escape mutations and B-cell adaptation has been observed experimentally [11–14] and modeled mechanistically [26, 28], it is not clear what are the generic features and relevant parameters in an evolutionary arms race that permit the development, or, especially, the early development of BnAbs. Phenomenological models ignore many details of affinity maturation and heterogeneity in the structure of germinal centers and yet produce useful qualitative predictions [14, 29]. Past models typically described only a few viral types [26, 27, 30], and did not account for the vast genetic diversity and turnover seen in infecting populations.

In this paper, we take a phenomenological approach to model the within-host co-evolution of *diverse* populations of B-cells and chronic viruses. We focus on the latency phase of an infection, during which the population sizes of viruses and lymphocytes are relatively constant but their genetic compositions undergo rapid turnover [31]. We characterize the interacting sites of B-cell receptors and viruses as mutable binary strings, with binding affinity, and therefore selection, defined by matching bits. We keep track of both variable regions in the viral genome and conserved regions, asking specifically when B-cell receptors will evolve to bind to the conserved region, i.e., to develop broad neutralization capacity. The main simplification that makes our analysis tractable is that we focus on the evolution of a shared interaction phenotype, namely the distribution of binding affinities between viral and receptor populations. Specifically, we model the effects of mutations, selection and reproductive stochasticity on the distribution of binding affinities between the two populations. Projecting from the high-dimensional space of genotypes to lower dimension of binding phenotypes allows for a predictive and analytical description of the co-evolutionary process [32], whilst retaining the salient information about the quantities of greatest biological and therapeutic interest.

Using this modeling approach we show that the evolution of the binding affinity does not depend on details of any single-locus contribution, but is an emerging property of all constitutive loci. Even though the co-evolution of antibodies and viruses is perpetually out of equilibrium, we develop a framework to quantify the amount of adaptation in each of the two populations by defining fitness and transfer flux, which partition changes in mean fitness. We discuss how to measure the fitness and transfer flux from time-shifted experiments, where antibodies are competed against past and future viruses; and we show how such measurements provide a signature of co-evolution. We discuss the consequences of competition between clonal B-cell lineages within and between germinal centers. In particular, we derive analytic expressions for the fixation probability of a newly arisen, broadly neutralizing antibody lineage. We find that BnAbs have an elevated chance of fixation in the presence of a diverse viral population, whereas specific neutralizing antibody lineages do not. We discuss the implications of these results for the design of preventive vaccines that elicit BnAbs against HIV.

Model

Interaction phenotype between antibodies and viruses. B-cell receptors undergo mutation and selection in germinal centers, whereas viruses are primarily affected by the receptors secreted into the blood, known as antibodies. Our model does not distinguish between antibodies and B-cells, so we will use the terms interchangeably. To represent genetically diverse populations we define genotypes for antibodies and viruses as binary sequences of ±1, where mutations change the sign of individual loci. Mutations in some regions of a viral genome are highly deleterious, e.g. at sites that allow the virus to bind target cell receptors, including CD4-binding sites for HIV. To capture this property we explicitly model a conserved region of the viral genome that does not tolerate mutations, so that its bits are always set to $+1$. We let viruses have variable bits at positions $i = 1 \ldots \ell$, and conserved bits at positions $i = \ell+1, \ldots, \ell+\hat{\ell}$; while antibodies have variable bits at positions $i = 1 \ldots \ell+\hat{\ell}$; see Fig. 1B.

Naive B-cells generate diversity by recombination events (VDJ recombination), which differentiates their ability to bind to different epitopes of the virus; and then B-cells diversify further by somatic hypermutation and selection during affinity maturation. We call the set of B-cells that originate from a common germline sequence a clonal lineage. A lineage with access to conserved regions of the virus can effectively neutralize more viral genotypes, since no escape mutation can counteract this kind of neutralization.

The binding affinity between antibody and virus determines the likelihood of a given antigen neutralization by an antibody, and therefore it is the key molecular phenotype that determines selection on both immune and viral populations. We model the binding affinity as a weighted dot product over all loci, which for antibody A^α chosen from the genotype space $\alpha \in 1 \ldots 2^{\ell+\hat{\ell}}$ and virus V^γ with

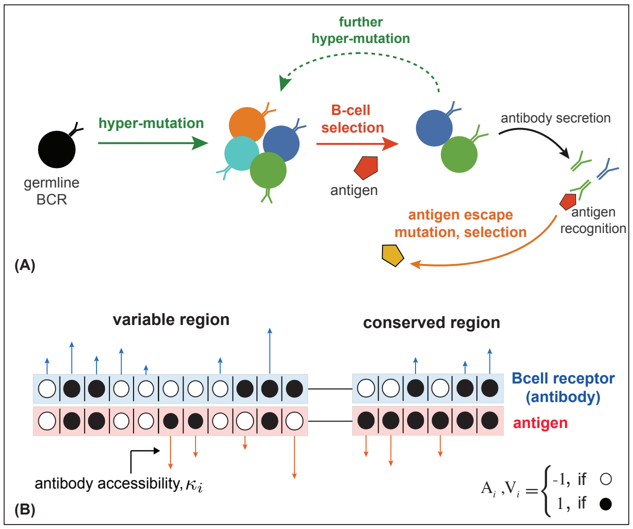

FIG. 1: **Co-evolution of antibodies and viruses. (A)** Schematic of affinity maturation in a germinal center. A naive, germline B-cell receptor (black) with marginal binding affinity for the circulating antigen (red pentagon) enters the process of affinity maturation in a germinal center. Hypermutations produce a diverse set of B-cell receptors (colors), the majority of which do not increase the neutralization efficacy of B-cells, except for some beneficial mutations that increase binding affinity (dark blue and green) to the presented antigen. The selected B-cells may enter the blood and secrete antibodies, or enter further rounds of hypermutations to enhance their neutralization ability. Antigens mutate and are selected (yellow pentagon) based on their ability to escape the current immune challenge. **(B)** We model the interaction between the genotype of a B-cell receptor and its secreted antibody (blue) with a viral genotype (red) in both variable and conserved regions of the viral genome. The black and white circles indicate the state of the interacting loci with values ±1. Loci in the conserved region of the virus are fixed at +1. The length of the arrows indicate the contribution of each locus to the binding affinity, κ_i, which is a measure of the accessibility of an antibody lineage to viral epitopes. The blue arrows indicate the interactions that increase binding affinity (i.e., loci with same signs in antibody and viral genotype), whereas red arrows indicate interaction that decrease the affinity (i.e., loci with opposite signs in antibody and viral genotype.)

$\gamma \in 1 \ldots 2^\ell$ has binding affinity

$$\begin{aligned} E^{\mathcal{C}}_{\text{tot}}(A^\alpha, V^\gamma) &= \underbrace{\sum_{i=1}^{\ell} \kappa_i^{\mathcal{C}} A_i^\alpha V_i^\gamma}_{\text{variable viral region}} + \underbrace{\sum_{i=\ell+1}^{\ell+\hat\ell} \hat\kappa_i^{\mathcal{C}} A_i^\alpha}_{\text{conserved viral region}} \\ &\equiv E^{\mathcal{C}}_{\alpha\,\gamma} + \hat E^{\mathcal{C}}_\alpha \end{aligned} \tag{1}$$

where, $A_i^\alpha = \pm 1$ denotes the i^{th} locus of the α antibody genotype, and V_i^γ the i^{th} locus of the γ viral genotype. Matching bits at interacting positions enhance binding affinity between an antibody and a virus; see Fig. 1B. Similar models have been used to describe B-cell maturation in germinal centers [26], and T-cell selection based on the capability to bind external antigens and avoid self proteins [33, 34]. The conserved region of the virus with $V_i = 1$ is located at positions $i = \ell+1, \ldots, \ell+\hat\ell$ for all viral sequences. Consequently, the total binding affinity is decomposed into the interaction with the variable region of the virus, $E^{\mathcal{C}}_{\alpha\,\gamma}$ and with the conserved region of the virus, $\hat E^{\mathcal{C}}_\alpha$. We call the lineage-specific binding constants $\{\kappa_i^{\mathcal{C}} \geq 0\}$ and $\{\hat\kappa_i^{\mathcal{C}} \geq 0\}$ the *accessibilities*, because they characterize the intrinsic sensitivity of an antibody lineage to individual sites in viral epitopes. We begin by analyzing the evolution of a single antibody lineage, and suppress the $\mathcal{C}$ notation for brevity. Co-evolution with multiple antibody lineages is discussed in a later section.

Both antibody and viral populations are highly polymorphic, and therefore contain many genotypes simultaneously. While the binding affinity between a virus V^γ and an antibody A^α is constant, given by eq. (1), the frequencies of the antibody and viral genotypes, x^α and y^γ, and all quantities derived from them, change over time as the two populations co-evolve. To characterize the distribution of binding affinities we define the genotype-specific binding affinities in each population, which are marginalized quantities over the opposing population: $E_{\alpha\,\cdot} = \sum_\gamma E_{\alpha\,\gamma} y^\gamma$ for the antibody A^α, and $E_{\cdot\,\gamma} = \sum_\alpha E_{\alpha\,\gamma} x^\alpha$ for the virus V^γ. We will describe the time evolution of the joint distribution of $E_{\alpha\,\cdot}$, $\hat E_\alpha$, and $E_{\cdot\,\gamma}$, by considering three of its moments: (i) the mean binding affinity, which is the same for both populations $\mathcal{E} = \sum_\alpha E_{\alpha\,\cdot} x^\alpha = \sum_\gamma E_{\cdot\,\gamma} y^\gamma$, (ii) the diversity of binding affinity in the antibodies, $M_{A,2} = \sum_\alpha (E_{\alpha\,\cdot} - \mathcal{E})^2 x^\alpha$ and (iii) the diversity of binding affinities in the viruses, $M_{V,2} = \sum_\gamma (E_{\cdot\,\gamma} - \mathcal{E})^2 y^\gamma$. Analogous statistics of binding affinities can be defined for the conserved region of the virus, which we denote by $\hat{\mathcal{E}}$ for the mean interaction, and $\hat M_{A,2}$ for the diversity across antibodies. The diversity of viral interactions in the conserved region must always equal zero, $\hat M_{V,2} = 0$.

The model outlined above is similar to models for the evolution of other molecular traits developed in the context of quantitative genetics [35–37]. Our analysis neglects the correlation between the variable and the conserved regions of the virus, which is due to physical linkage of the segments. In Methods and in Fig. S3 we show that there is a difference in evolutionary time-scales of these two regions, which reduces the magnitude of this correlation.

Co-evolution of a single antibody lineage and viral population. We first characterize the affinity maturation process of a single clonal antibody lineage coevolving with a viral population, which includes hypermutation, selection, and stochasticity due to population size in germinal centers, i.e., genetic drift.

In the bi-allelic model outlined in Fig. 1B, a hypermutation changes the sign of an antibody site, i.e., $A_i^\alpha \to -A_i^\alpha$, affecting binding affinity in proportion to the lineage's intrinsic accessibility at that site, κ_i. Therefore, a mutation in an antibody at position i changes $E_{\alpha\cdot}$ by $\delta_i E_{\alpha\cdot} = -2\,\kappa_i A_i^\alpha \sum_\gamma V_i^\gamma y^\gamma$. Likewise, a mutation at position j of a virus $V_j^\gamma \to -V_j^\gamma$ affects binding affinity in proportion to κ_j. We assume constant mutation rates in the variable regions of the viruses and antibodies: μ_v and μ_a per site per generation. Summing the effects of mutations in both antibody and viral populations yields the per-generation change in mean binding affinity, $\Delta\mathcal{E} = -2(\mu_a + \mu_v)\mathcal{E}$ in the variable region and $\Delta\hat{\mathcal{E}} = -2\mu_a\hat{\mathcal{E}}$ in the conserved region of the virus. In the absence of selection, $\mathcal{E}$ and $\hat{\mathcal{E}}$ approach zero over time, which is the state of highest sequence entropy. In Methods and in SI we discuss the evolution of the higher central moments in detail.

Frequencies of genotypes change according to their relative growth rate, or fitness. The change in the frequency of antibody A^α with fitness f_{A^α} is $\Delta\, x^\alpha = (f_{A^\alpha} - F_A)x^\alpha$ per generation, where $F_A = \sum_\alpha f_{A^\alpha} x^\alpha$ denotes the mean fitness of the antibody population. Likewise, the change in frequency of virus V^γ due to selection per generation is, $\Delta\, y^\gamma = (f_{V^\gamma} - F_V)y^\gamma$, where F_V denotes the mean fitness in the viral population.

The most important choice in formulating an evolutionary model is to specify the form of fitness for each genotype. During affinity maturation in a germinal center B-cell growth rate depends on their ability to bind to the limited amounts of antigen, and to solicit survival signals from helper T-cells [8]. The simplest functional form that approximates this process, and for which we can provide analytical insight, is linear with respect to the binding affinity,

$$f_{A^\alpha} = S_a(E_{\alpha\cdot} + \hat{E}_\alpha) \tag{2}$$

$$f_{V^\gamma} = -S_v(E_{\cdot\gamma} + \hat{\mathcal{E}}) \tag{3}$$

for antibody A^α and virus V^γ. The selection coefficient $S_a > 0$ quantifies the strength of selection on the binding affinity of antibodies. The value of S_a may in fact decrease in late stages of a long-term HIV infection, as the host's T-cell count decays [29], but we do not model this behavior. The viral selection coefficient $S_v > 0$ represents immune pressure impeding the growth of the virus.

Changes of the genotype frequencies Δx^α and Δy^γ in the antibody and viral populations due to selection affect the mean binding affinity between the two populations. In the linear fitness formulation in eqs. (2,3), the change in the mean binding affinity per generation is $\Delta\mathcal{E} = S_a M_{A,2} - S_v M_{V,2}$ and $\Delta\hat{\mathcal{E}} = S_a \hat{M}_{A,2}$, in the variable and conserved regions respectively. As these equations reflect, mean binding affinity increases in proportion to the diversity of antibody binding affinities, and it decreases in proportion to the diversity of viral binding affinities, in accordance with the Price equation [38].

Many aspects of affinity maturation are not well known, and so it is worth considering other forms of selection. Therefore, in the Methods and in Section 2.5 of SI we describe fitness as a non-linear function of the binding affinity. In particular, we consider fitness that depends on the antibody activation probability, which is a sigmoid function of the *strongest* binding affinity among a finite number of interactions with antigens. The linear fitness function in eq. (2) is a limiting case of this more general fitness model. While most of our analytical results are based on the assumption of linear fitness function, we also discuss how to quantify adaptation for arbitrary fitness models, and we numerically study the effect of nonlinearity on the rate of antibody adaptation during affinity maturation.

Although the population of B-cells can reach large numbers within an individual host, significant bottlenecks occur in germinal centers, where there may be on the order of $\sim 10^3 - 10^4$ B-cells [7]. While our model does not describe heterogeneity between germinal centers, we do model the effects of finite population size as stochasticity in reproductive success, known as genetic drift. Similarly, for HIV, estimates for intra-patient viral divergence suggests an effective population size of about $\sim 10^2 - 10^3$, which is much smaller than the number of infected cells within a patient $\sim 10^7 - 10^9$ [39]. We use diffusion equations in which the strength of genetic drift is described by the effective population sizes N_a and N_v of antibodies and viruses (see Section 2 of SI for details).

Without loss of generality, we assume that generation times in antibodies and viruses are equal, but we define distinct characteristic time-scales for the two populations. The relevant time-scale for evolution of polymorphic populations is the neutral coalescence time, N generations – namely, the characteristic time that two randomly chosen neutral alleles in the population coalesce to their most recent common ancestor. The neutral coalescence time is estimated by reconstructing phylogenetic trees from sequences and is often interpreted as an effective population size, which may be different from the census population size. Coalescence time can be mapped onto real units of time (e.g., days) if data are collected with sufficient time resolution. To distinguish between the neutral coalescence time of antibodies and viruses, we use distinct values for their population sizes, i.e., N_a in antibodies and N_v in viruses.

Combining genetic drift with mutation and selection, and assuming a continuous-time and continuous-frequency process, results in a stochastic dynamical equation for the evolution of mean binding affinity in the variable region,

$$\frac{d}{dt}\mathcal{E} = -2(\mu_a+\mu_v)\mathcal{E}+S_a M_{A,2}-S_v M_{V,2}+\sqrt{\frac{M_{A,2}}{N_a}+\frac{M_{V,2}}{N_v}}\chi_{\mathcal{E}} \tag{4}$$

and in the conserved region,

$$\frac{d}{dt}\hat{\mathcal{E}} = -2\mu_a\hat{\mathcal{E}} + S_a\hat{M}_{A,2} + \sqrt{\frac{\hat{M}_{A,2}}{N_a}}\chi_{\hat{\mathcal{E}}} \qquad (5)$$

where $\chi_{\mathcal{E}}$ and $\chi_{\hat{\mathcal{E}}}$ are standard Gaussian noise terms.

The number of sites and their accessibilities, which are implicit in eqs. (4, 5), affect the overall strength of selection on binding affinity. Therefore, it is useful to absorb the intrinsic effects of the trait magnitude into the selection strength, and keep the binding affinities comparable across lineages of antibodies, and across experiments. We therefore rescale quantities related to the binding affinity by the total scale of the phenotypes $E_0^2 = \sum_i \kappa_i^2$ and $\hat{E}_0^2 = \sum_i \hat{\kappa}_i^2$, such that $\varepsilon = \mathcal{E}/E_0$, $\hat{\varepsilon} = \hat{\mathcal{E}}/\hat{E}_0$, $m_{A,2} = M_{A,2}/E_0^2$ and $m_{V,2} = M_{V,2}/E_0^2$, and $\hat{m}_{A,2} = \hat{M}_{A,2}/\hat{E}_0^2$. Accordingly, we define rescaled selection coefficients $s_a = N_aS_aE_0$, $\hat{s}_a = N_aS_a\hat{E}_0$ and $s_v = N_vS_vE_0$, which describe the total strength of selection on binding affinity.

Empirical estimates of per-generation mutation rates for viruses μ_v or hypermutation rates of BCR sequences μ_a are extremely imprecise, and so we rescale mutation rates by neutral coalescence times. To do this, we use measurements of standing neutral sequence diversity, estimated from genetic variation in, e.g., the four-fold synonymous sites of the protein sequences at each position. Neutral sequence diversity for the antibody variable region, which spans a couple of hundred base pairs, is about $\theta_a = N_a\mu_a = 0.05 - 0.1$ [2]. Nucleotide diversity of HIV increases over time within a patient, and ranges between $\theta_v = N_v\mu_v = 10^{-3} - 10^{-2}$ in the *env* protein of HIV-1 patients, with a length of about a thousand base pairs [40]. Interestingly, the total diversity of the variable region in BCRs is comparable to the diversity of its main target, the *env* protein, in HIV. Therefore, both populations have on the order of 1-10 mutations per genotype per generation, which we use as a guideline for parameterizing simulations of our model.

Results

Dynamics of the mean binding affinity. The model defined above is analytically tractable, which allows us to study the dynamics of antibody-viral co-evolution in terms of their basic underlying parameters. We focus initially on understanding the (rescaled) mean binding affinity ε, $\hat{\varepsilon}$ between a clonal antibody lineage and the viral population, since this is a proxy for the overall neutralization ability that is commonly monitored during an infection. Appropriate rescaling of eqns. [4,5] shows that the efficacy of selection on binding affinity from the antibody or the viral populations depends on the rescaled diversity $m_{A,2}$, $m_{V,2}$ in each of the populations, as shown in eqns. [4,5].

If a population harbors a large diversity of binding affinities then it has more potential for adaptation from the favorable tail of the distribution, which contains the most fit individuals in each generation [38, 41]. It follows that selection on viruses does not affect the evolution of their conserved region, where the viral diversity of binding is always zero, $\hat{m}_{V,2} = 0$. In the following we describe the stochastic dynamics of the mean binding affinity and diversities, as well as their ensemble-averaged stationary solutions.

The dynamics described by eqns. [4,5] present a common difficulty: the change in mean binding affinity depends on the binding diversities, and the diversities in turn depend on higher moments, forming an infinite hierarchy (see Methods and Section 2.3 of SI). However, this moment hierarchy can be simplified to produce accurate analytical approximations in the regime where selection on individual loci is weak but the additive effects of selection on the total binding affinity are strong, i.e., $N_aS_a\kappa_i < 1$ and $s_a \gtrsim 1$ for antibodies, and likewise for viruses and the conserved region. This parameter regime, which has been observed in chronic HIV infections [40], makes it possible to truncate the moment hierarchy and produce reliable predictions for the mean and the diversity of binding affinities, when the rescaled coefficients satisfy $s_a\theta_a < 1$ and $s_v\theta_v < 1$. As shown in detail in Section 2 of SI, truncation at the 4^{th} central moment is a suitable choice for our model.

As shown in Fig. S3 and discussed in Methods, the binding diversities are fast variables compared to the mean affinity, and therefore, we can describe the dynamics of the mean in terms of the stationary binding diversities. The ensemble-averaged binding diversities depend only weakly on the strength of selection and can be approximated by $\langle m_{A,2}\rangle \simeq 4\theta_a$ and $\langle m_{V,2}\rangle \simeq 4\theta_v$, even for substantial selection $s \sim 1$. Here $\langle\cdot\rangle$ indicates ensemble averaging. Higher-order corrections to the diversity of the binding affinity are given in Section 2 of SI and shown in Fig. S2. In this regime, the ensemble-averaged mean binding affinities relax exponentially towards their stationary values,

$$\langle\varepsilon\rangle \simeq \frac{2(s_a\theta_a - s_v\theta_v(N_a/N_v))}{\theta_a + \theta_v(N_a/N_v)} \equiv 2\,\Delta s_{av} \qquad (6)$$

$$\langle\hat{\varepsilon}\rangle \simeq 2\,\hat{s}_a \qquad (7)$$

where Δs_{av} is an effective selection coefficient for binding affinity in the variable region, combining the effect of selection from both populations and accounting for their distinct genetic diversities. The stationary mean binding affinity quantifies the balance of mutation and selection acting on both populations. Mutations drive the mean affinity towards the neutral value, zero, whereas selection pushes it towards positive or negative values. Positive values indicate that binding is more strongly influenced by the antibodies, whereas negative values indicate more influence from the viruses. Strong selection difference between two populations $\Delta s_{av} \gtrsim 1$, results in selective sweeps for genotypes with extreme values of binding affinity in each population, and hence, reduces the diversity of interactions as shown in Fig. S2. In Section 2.3 of SI we discuss in detail the effect of selection on the diversity of binding affinities. We validate our analytical solutions for stationary mean binding and diversities by comparison with Wright-Fisher simulations across a broad range of selection strengths.

In the Methods and in Section 2.5 of SI we numerically study the non-linear fitness landscapes described in the Model section, and their effect on the stationary mean binding and rate of adaptation (Fig. S4A). While the results differ quantitatively, we can qualitatively understand how the stationary mean binding affinity depends on the form of non-linearity.

The rescaled binding affinities $\varepsilon \equiv \mathcal{E}/E_0$, $\hat{\varepsilon} \equiv \hat{\mathcal{E}}/\hat{E}_0$ are independent of the total scale of the phenotype, and can therefore be used for comparisons across experiments. Measuring the total scale of the phenotype E_0, $\hat{E}_0$ directly would require a library of all single point mutations and measurements of their presumably very small fitness differences, κ_i, $\hat{\kappa}_i$. Nonetheless the rescaled mean binding affinities can be approximated simply by measurements of the binding affinity distribution and neutral sequence diversities, which are experimentally accessible. The ratio of the binding diversity $M_{A,2}$ and the neutral sequence diversity θ_a provides a reasonable approximation for the overall scale of the phenotype: $\varepsilon \approx \langle \mathcal{E} \rangle / \sqrt{\langle M_{A,2} \rangle / 4\theta_a}$ and $\hat{\varepsilon} \approx \langle \hat{\mathcal{E}} \rangle / \sqrt{\langle \hat{M}_{A,2} \rangle / 4\theta_a}$. Fig. S1 demonstrates the utility of this approximation and shows, moreover, that heterogeneous binding accessibilities in an antibody lineage, κ_i, drawn from several different distributions, do not affect predictions for stationary mean binding. Even though we have formulated a high-dimensional stochastic model of antibody-antigen co-evolution in polymorphic populations, Fig. S1 demonstrates that we can nonetheless understand the long-term binding affinities, which are commonly measured in patients, in terms of only a few key parameters.

Fitness and transfer flux. The antagonistic co-evolution of antibodies and viruses is a non-equilibrium process, with each population constantly adapting to a dynamic environment, namely, the state of the opposing population. As a result, any time-independent quantity, such as the stationary mean binding affinity studied above, is itself not informative for the extent of co-evolution that is occurring. For example, a stationary mean binding affinity of zero (equivalently $\Delta s_{av} = 0$ in eq. (6)) can indicate either neutral evolution or rapid co-evolution induced by equally strong selection in antibody and viral populations.

To quantify the amount of adaptation and extent of interaction in two co-evolving populations we will partition the change in mean fitness of each population into two components. We measure adaptation by the *fitness flux* [42, 43], which quantifies the change in mean fitness of a population in response to the state of the environment (that is, in this case, the opposing population). More specifically, the fitness flux of the antibody population quantifies the effect of changing genotype frequencies on mean fitness, and it is defined as $\phi_A(t) = \sum_\alpha \partial_{x^\alpha} F_A(t)\, dx^\alpha(t)/dt$, where F_A denotes the mean fitness of antibodies, and the derivative $dx^\alpha(t)/dt$ measures the change in frequency of the antibody A^α. The forces of mutation, drift, and selection all contribute to fitness flux, however the portion of fitness flux due to selection always equals the population variance of fitness, in accordance with Fisher's theorem [41]. The second quantity we study, which we term the *transfer flux*, measures the amount of interaction between the two populations by quantifying the change in mean fitness due to the response of the opposing population. The transfer flux from viruses to antibodies is defined as $\mathcal{T}_{V\to A}(t) = \sum_\gamma \partial_{y^\gamma} F_A(t)\, dy^\gamma(t)/dt$. Analogous measures of adaptation and interaction can be defined for the

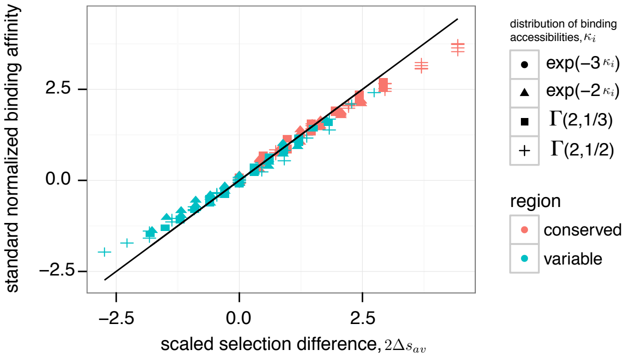

FIG. 2: **Effect of selection on immune-virus binding affinity.** Our mathematical model predicts the stationary mean binding affinity between viruses and antibodies in terms of small number of key parameters. The figure plots the stationary mean binding affinity, rescaled by the diversity of binding in the antibody population ($\mathcal{E}/\sqrt{M_{A,2}/4\theta_a}$), as predicted by our analysis (line) compared with Wright-Fisher simulations (dots). Remarkably, the stationary mean binding is a simple function of the selection difference between antibody and viral populations, Δs_{av} (eq. (6)), which is insensitive to the details of the heterogeneous binding accessibilities, κ_i, associated with an antibody lineage (drawn from several different distributions, shown in legend). Small deviations from this collapse onto a universal predicted mean binding are due to higher moments of binding affinity, which can also be understood analytically (Fig. S1). Stationary mean binding affinities are averaged from simulations with selection coefficients, s_a, $\hat{s}_a$, s_v ranging from 0 to 6, and with $N_a = N_v = 1000$, $\ell = \hat{\ell} = 25$, $\theta_a = \theta_v = 1/50$.

viral population (see Section 3 of SI).

The fitness flux and transfer flux represent rates of adaptation and interaction, and they are typically time dependent, except in stationary state. The total amount of adaptation and interaction during non-stationary evolution, where the fluxes change over time, can be measured by the cumulative fluxes over a period of time: $\Phi_A(\tau_a) = N_a \int_{t'=0}^{t} \phi_A(t')\,dt'$ and $\mathbf{T}_{V\to A}(\tau_a) = N_a \int_{t'=0}^{t} \mathcal{T}_{V\to A}(t')\,dt'$, where time $\tau_a = t/N_a$ is measured in units of neutral coalescence time of antibodies N_a. In the stationary state, the ensemble-averaged cumulative fluxes grow linearly with time. For co-evolution on the fitness landscapes given by equations [2,3], the ensemble-averaged, stationary cumulative fitness flux and transfer flux in antibodies are

$$\langle \Phi_A(\tau_a)\rangle = \big[-2\theta_a s_a \langle\varepsilon\rangle + s_a^2 \langle m_{A,2}\rangle\big]\tau_a \tag{8}$$

$$\langle \mathbf{T}_{V\to A}(\tau_a)\rangle = \big[-2\theta_v s_a \langle\varepsilon\rangle - s_a s_v \langle m_{V,2}\rangle\big](N_a/N_v)\tau_a \tag{9}$$

Note that the factor $(N_a/N_v)\tau_a$ in eq. (9), which is a rescaling of time in units of viral neutral coalescence time $\tau_v = t/N_v$, emphasizes the distinction between the evolutionary time scales of antibodies and viruses. The first terms on the right hand side of eqs. (8,9) represent the fitness changes due to mutation, second terms are due to selection, and the effects of genetic drift are zero in the ensemble average for our linear fitness landscape. Notably, the flux due to the conserved region of the virus is zero in stationarity, as is the case for evolution in a static fitness landscape (i.e., under equilibrium conditions). In the stationary state, the cumulative fitness and transfer fluxes sum up to zero, $\langle \Phi_A(\tau_a)\rangle + \langle \mathbf{T}_{V\to A}(\tau_a)\rangle = 0$.

Fitness flux and transfer flux are generic quantities that are independent of the details of our model, and so they provide a natural way to compare the rate of adaptation in different evolutionary models or in different experiments. In the regime of strong selection $s_a, s_v \gtrsim 1$, non-linearity of the fitness function results in a more narrow distribution of fitness values in the antibody population, and hence, reduces the rate of adaptation and fitness flux; see Fig. S4. In the following section we show how to use fitness and transfer flux to detect signatures of significant antibody-antigen co-evolution.

Signature of co-evolution. Measuring interactions between antibody and viral populations sampled from different time points provides a powerful way to identify signatures of immune-pathogen co-evolution. These types of "time-shifted" experiments are informative both in theoretical models and in empirical assays. In general, antibodies sampled at the present perform best against viruses sampled from the past, to which they have been selected to bind, whereas they perform worst against viruses from the future, due to viral escape (Fig. 3). Under our model, and neglecting the conserved region of the virus, the rescaled time-shifted binding affinity between antibodies at time t and viruses at time $t+\tau$ is $\varepsilon_\tau(t) = \sum_{\alpha,\gamma} E_{\alpha\,\gamma} x^\alpha(t)\, y^\gamma(t{+}\tau)/E_0$, and the corresponding antibody and viral mean fitnesses are $N_a F_{A;\tau}(t) = s_a \varepsilon_\tau(t)$ and $N_v F_{V;\tau}(t) = -s_v \varepsilon_{-\tau}(t)$. The slope of the time-shifted fitness at the time where the two populations co-occur (i.e., $\tau = 0$), approaching from the negative τ, i.e., the past, measures the amount of adaptation of the focal population in response to the state of the viral population, and is precisely equal to the fitness flux: $\partial_\tau F_{A;\tau}(t-\tau)|_{\tau=0^-} = \phi_A(t)$. The slope approaching from positive time-shifts, i.e., the future, measures the change in the mean fitness of the focal population due to adaptation of the viral population, and is precisely equal to the

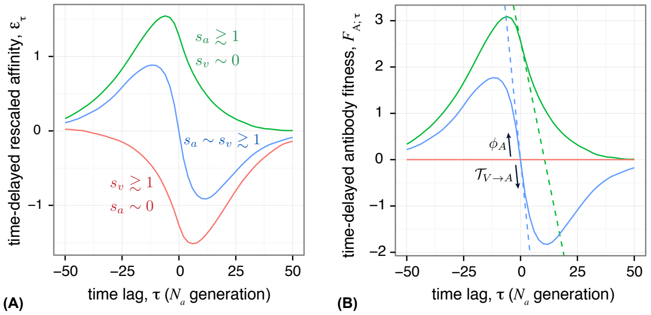

FIG. 3: **Time-shifted binding assays between antibodies and antigens provide a definitive signature of viral-immune co-evolution.** **(A)** Stationary binding affinity between the antibody population sampled at time t, and the viral population at time $t+\tau$, averaged over all t: $\varepsilon_\tau = \langle \sum_{\alpha,\gamma} E_{\alpha\gamma} x^\alpha(t) y^\gamma(t+\tau)\rangle_t / E_0$, and **(B)** the time-shifted mean fitness of the antibody population $N_a F_{A;\tau} = s_a \varepsilon_\tau$ are shown for three distinct regimes of co-evolutionary dynamics: (i) strong and comparable selection strengths on both antibodies and antigens $s_a = s_v = 2$ (blue), (ii) strong selection in antibodies $s_a = 2$ and no selection on viruses $s_v = 0$ (green), and (iii) strong selection on viruses and $s_v = 2$ and no selection on antibodies $s_a = 0$ (red). The "S" shape curve in blue is a signature of co-evolution between the two populations, $s_a \sim s_v$: Antibodies perform best against viruses in the past, which they have been selected to bind, and perform worst against viruses in the future due to viral escape by selection. For large time-shifts the binding strength relaxes to its neutral value, zero, as mutations randomize the populations with respect to each other. In the absence of selection in one of the populations, the time-shifted binding affinity shows the selective response of one population against stochastic variation in the other population due to mutation and genetic drift. The slope of time-shifted fitness at time-lag $\tau = 0$ is a measure of the antibody population's fitness flux (towards the past) and the transfer flux from the opposing population (towards the future), which are equal in stationary state as depicted in (B). The dashed lines indicated the predicted fitness flux and transfer flux given by eqs. (8,9). Other simulation parameters are: $N_a = N_v = 1000$, $\ell = 50$, $\hat{\ell} = 0$, $\theta_a = \theta_v = 1/25$, and $\kappa_i = 1$ for all loci.

transfer flux $\partial_\tau F_{A;\tau}(t)|_{\tau=0^+} = \mathcal{T}_{V\to A}(t)$. In stationarity, the sum of fitness flux and transfer flux is zero on average, and so the slopes from either side of $\tau = 0$ are equal, as in Fig. 3. Note that the relationships between time-shifted fitness and the flux variables hold beyond the specific case of linear landscape. In a non-stationary state, the fitness flux and transfer flux are not balanced, and so $\langle F_{A;\tau}(t)\rangle$ has a discontinuous derivative at $\tau = 0$ (Fig. S5). Therefore, such a discontinuity provides a mechanism to identify stationarity versus transient dynamics.

Whether in stationarity or not, the signature of out-of-equilibrium evolution is a positive fitness flux and negative transfer flux. For time-shifted fitness, this means that for short time shifts, where dynamics are dominated by selection, antibodies have a higher fitness against viruses from the past, and worse binding to viruses from the future. This is true even when one population is evolving neutrally and the other has substantial selection, as shown in Fig. 3. For long time shifts, the sequences are randomized by mutations and the fitness decays exponentially to the neutral value. When selection and mutation are substantial on both sides the time-shifted fitness curve has a characteristic "S" shape – a signature of co-evolution, whose inflection form can be understood in terms of the fitness and transfer fluxes.

Time-shifted measurements have already been studied empirically. Measurements of neutralization in HIV patients have found that antibodies perform better at neutralizing viruses from the past, and worse at neutralizing viruses from the future [11, 12]. The signature of co-evolution can be obscured in in clinical studies because fitnesses of antibodies and viruses also depend on intrinsic and environmental factors, such as a time-dependent environment from drug treatments. However, linear regression of time-shifted antibody-HIV neutralization measurements was able to decompose antibody fitnesses into components due to interaction with the viruses and other factors, resulting in a characteristic S shape in the time-shifted fitness due to interaction [44].

Competition between multiple antibody lineages. B-cells in the adaptive immune system are associated with clonal lineages that originate from distinct ancestral naive cells, generated by germline (VDJ) recombination and junctional diversification [1]. Multiple lineages may be stimulated within a germinal center, and also circulate to other germinal centers [8]. Lineages compete for activation agents (e.g., helper T-cells) and interaction with a finite number of presented antigens [8]. We extend our theoretical framework to study how multiple lineages compete with each other and co-evolve with viruses. This generalization allows us to show that lineages with higher overall binding ability, higher fitness flux, and lower (absolute) transfer flux have a better chance of surviving. In particular, we show that an antibody repertoire fighting against a highly diversified viral population, e.g., during late stages of HIV infection, favors elicitation of broadly neutralizing antibodies compared to normal antibodies.

We define an antibody lineage $\mathcal{C}$ based on its site-specific accessibilities to the viral sequence $\{\kappa_i^{\mathcal{C}}, \hat{\kappa}_j^{\mathcal{C}}\}$, defined in Fig. 1A. The distribution of site-specific accessibilities over different antibody lineages $P_{\mathcal{C}}\big(\{\kappa_i^{\mathcal{C}}, \hat{\kappa}_i^{\mathcal{C}}\}\big)$ characterizes the ability of an antibody repertoire to respond to a specific virus. Without continual introduction of new lineages, one lineage will ultimately dominate and the rest will go extinct within the coalescence time-scale of antibodies, N_a (Fig. 4A). In reality, constant turn-over of lineages results in a highly diverse B-cell response, with multiple lineages acting simultaneously against an infection [45].

Stochastic effects are significant when the size of a lineage is small, so an important question is to find the probability that a low-frequency antibody lineage reaches an appreciable size and fixes in the population. We denote the frequency of an antibody lineage with size $N_a^{\mathcal{C}}$ by $\rho^{\mathcal{C}} = N_a^{\mathcal{C}}/N_a$. The growth rate of a given lineage $\mathcal{C}$ depends on its relative fitness $F_{A^{\mathcal{C}}}$ compared to the rest of the population,

$$\frac{d}{dt}\rho^{\mathcal{C}} = (F_{A^{\mathcal{C}}} - F_A)\rho^{\mathcal{C}} + \sqrt{\frac{\rho^{\mathcal{C}}(1-\rho^{\mathcal{C}})}{N_a}}\chi_C \tag{10}$$

where $F_A = \sum_{\mathcal{C}} F_{A^{\mathcal{C}}}\rho^{\mathcal{C}}$ is the average fitness of the entire antibody population. For the linear fitness landscape from eq. (2), the mean fitness of lineage $\mathcal{C}$ is $F_{A^{\mathcal{C}}} = S_a(\mathcal{E}^{\mathcal{C}} + \hat{\mathcal{E}}^{\mathcal{C}})$. The probability of fixation of lineage $\mathcal{C}$ equals the asymptotic (i.e., long time) value of the ensemble-averaged lineage frequency, $P_{\text{fix}}(\mathcal{C}) = \lim_{t\to\infty}\langle\rho^{\mathcal{C}}(t)\rangle$.

Similar to evolution of a single lineage, the dynamics of a focal lineage are defined by an infinite hierarchy of moment equations for the fitness distribution. In the regime of $s_a \sim 1$, where terms due to mutation can be neglected, a suitable truncation of the moment hierarchy allows us to estimate the long-time limit of the lineage frequency, and hence, its fixation probability (see Section 4 of SI). The result can be expressed in terms of the ensemble-averaged relative mean fitness, fitness flux and transfer flux at the time of introduction of the focal lineage,

$$\begin{aligned} P_{\text{fix}}(\mathcal{C})/P_{0_{\text{fix}}} &\simeq 1 + \big\langle N_a(F_{A^{\mathcal{C}}}(0) - F_{A(0)})\big\rangle \\ &+ \frac{N_a^2}{3}\big\langle\phi_{A^{\mathcal{C}}}(0) - \phi_A(0)\big\rangle - N_aN_v\big\langle\big|\mathcal{T}_{V\to A^{\mathcal{C}}}(0)\big| - \big|\mathcal{T}_{V\to A}(0)\big|]\big\rangle \end{aligned} \tag{11}$$

where $P_{0_{\text{fix}}}$ is the fixation probability of the lineage in neutrality, which equals its initial frequency at the time of introduction, $P_{0_{\text{fix}}} = \rho^{\mathcal{C}}(0)$. The first order term that determines the excess probability for fixation of a lineage is the difference between its mean fitness and the average fitness of the whole population. Thus, a lineage with higher relative mean fitness at the time of introduction, e.g., due to its better accessibility to either the variable or conserved region, will have a higher chance of fixation. Moreover, lineages with higher rate of adaptation, i.e., fitness flux $\phi_{\mathcal{A}^{\mathcal{C}}}(t=0)$, and lower (absolute) transfer flux from viruses $\left|\mathcal{T}_{\mathcal{V}\to\mathcal{A}^{\mathcal{C}}}(t=0)\right|$ tend to dominate the population.

For evolution in the linear fitness landscape, we can calculate a more explicit expansion of the fixation probability that includes mutation effects. In this case, the fixation probability of a focal lineage can be expressed in terms of the experimentally observable lineage-specific moments of the binding affinity distribution, instead of the moments of the fitness distribution (see Methods).

Emergence of broadly neutralizing antibodies. With our multi-lineage model, we can understand the conditions for emergence of broadly neutralizing antibodies (BnAbs) in an antibody repertoire. Similar to any other lineage, the progenitor of a BnAb faces competition with other resident antibody lineages that may be dominating the population. The dominant term in the fixation probability is the relative fitness difference of the focal lineage to the total population at the time of introduction. Lineages may reach different fitnesses because they differ in their scale of interaction with the viruses, $E_0^{\mathcal{C}}$ in the variable region and $\hat{E}_0^{\mathcal{C}}$ in the conserved region. Lineages which bind primarily to the conserved region, i.e., $\hat{E}_0^{\mathcal{C}} \gg E_0^{\mathcal{C}}$, do not have an opposing viral population that reduce their binding affinity. Such BnAbs may be able to reach higher fitnesses compared to normal antibodies which bind to the variable region with a comparable scale of interaction. The difference in the mean fitness of the two lineages becomes even stronger, when viruses are more diverse (i.e., high $M_{V,2}$), so that they can strongly compromise the affinity of the lineage that binds to the variable region; see eq. (11).

If the invading lineage has the same fitness as the resident lineage, then the second order terms in eq. (11) proportional to the fitness and transfer flux may be relevant. A BnAb lineage that binds to the conserved region has a reduced transfer flux than a normal antibody lineage, all else being equal. The difference in transfer flux of the two lineages depends on the viral diversity $M_{V,2}$, and becomes more favorable for BnAbs when the viral diversity is high. Overall, a BnAb generating lineage has a higher advantage for fixation compared to normal antibodies, when the repertoire is co-evolving against a highly diversified viral population, e.g., during late stages of HIV infection. As previously suggested by Luo & Perelson [30], this effect might be the reason for BnAbs to be detected late in infection, and to contain a relatively large amount of mutations compared to the germline sequence ($\sim 30\%$ of residues).

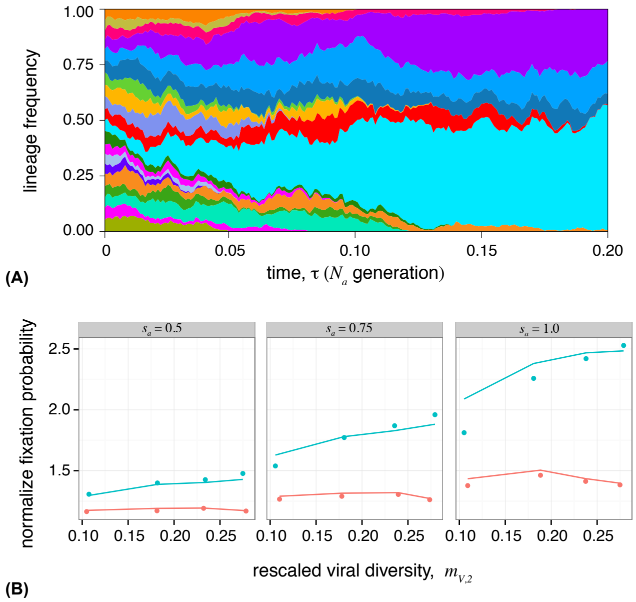

FIG. 4: **Competition between antibody lineages, and fixation of broadly neutralizing antibodies.** **(A)** Simulation of competition between 20 clonal antibody lineages introduced at time 0 against a common viral population. Lineages with higher mean fitness, higher fitness flux, and lower transfer flux tend to dominate the antibody repertoire. Each color represents a distinct antibody lineage, however there is also diversity within each lineage from somatic mutations. The reduction in the number of circulating lineages resembles the reduction in the number of active B-cell clones within the life-time of a germinal center [8]. **(B)** Analytical estimates (lines, eq. S2) of the fixation probability of a new antibody lineage, based on the state of the populations at the time of its introduction, compared to Wright-Fisher simulations (points) with two competing antibody lineages. A non-BnAb resident population is invaded by a BnAb (blue) or non-BnAb (red), (simulation procedures are detailed in Methods). Emergence of a broadly neutralizing antibody lineage is facilitated when the viral population is diverse. Normal antibody lineages have $\ell = 15$, $\hat{\ell} = 0$, whereas broadly neutralizing lineages have $\ell = 0$, $\hat{\ell} = 15$. Panels show results for different values of antibody selection $s_a = 0.5, 0.75, 1$ against a common viral selection strength $s_v = 0.75$. Viral diversity is influenced mostly by the viral mutation rate θ_v, which ranges from 0.05 to 0.2. Other simulation parameters are: $N_a = N_v = 1000$, $\theta_a = 1/10$, and $\kappa_i = \hat{\kappa}_i = 1$ for all loci.

In Fig. 4B we compare the fixation probability of a BnAb lineage, that binds only to the conserved region, with a normal antibody lineage that binds only to the variable region. In both cases we assume that the emerging lineage competes against a resident population of normal antibodies. We compare our analytical predictions for fixation probability as a function of the initial state of the antibody and viral populations in eq. (11), with Wright-Fisher simulations of co-evolving populations (numerical procedures are detailed in Methods). Increasing viral diversity $M_{2,V}$ increases the fixation of BnAbs, but does not influence fixation of normal lineages. For low viral diversity, fixation of BnAbs is closer to normal Abs and therefore, they might arise and be outcompeted by other antibody lineages.

Discussion

We have presented an analytical framework to describe the co-evolutionary dynamics between two antagonistic populations, based on molecular interactions between them. We have focused our analysis on antibody-secreting B-cells and chronic infections, such as HIV. We identified effective parameters for selection on B-cells during hypermutation that enhance their binding and neutralization efficacy, and conversely parameters for selection on viruses to escape the antibody binding. The resulting "red-queen" dynamic between antibodies and viruses produces an optimum response of antibodies against viral genotypes sampled from the past (to which they are adapted) and a deficient response to viruses sampled from the future, which are the successful escape mutants. Finally, we have shown that emergence and fixation of a given clonal antibody lineage is determined by competition between circulating antibody lineages [30], and that broadly neutralizing antibody lineages, in particular, are more likely to dominate in the context of a diverse viral population.

Our model is simple enough to clarify some fundamental concepts of antibody-antigen dynamics. However, understanding more refined aspects of B-cell-virus co-evolution will require adding details specific to affinity maturation and viral reproduction, such as non-neutralizing binding between antibodies and antigens [14, 46], epitope masking by antibodies [47] and spatial structure of germinal centers [8]. Importantly, viral recombination [39, 40, 48] and latent viral reservoirs [49] are also known to influence the evolution of HIV within a patient. Similarly, the repertoire of the memory B-cells and T-cells, which effectively keep a record of prior viral interactions, influence the response of the adaptive immune system against viruses with antigenic similarity.

While our analysis has focused on co-evolution of chronic viruses with the immune system, our framework is general enough to apply to other systems, such as bacteria-phage co-evolution. Likewise, the notions of fitness and transfer flux as measures of adaptation during non-stationary evolution are independent of the underlying model. Bacteria-phage interactions have been studied by evolution experiments [50, 51], and by time-shifted assays of fitness [52, 53], but established models of co-evolution describe only a small number of alleles with large selection effects [54]. In contrast, our model offers a formalism for bacteria-phage co-evolution where new genotypes are constantly produced by mutation, consistent with experimental observations [52]. Similarly, our formalism may be applied to study the evolution of seasonal influenza virus in response to the "global" immune challenge, imposed by a collective immune landscape of all recently infected or vaccinated individuals [55]. Time-shifted binding assays of antibodies to influenza surface proteins are already used to gauge the virulence and cross-reactivity of viruses. Quantifying the fitness flux and transfer flux, based on these assays, is therefore a principled way to measure rates of immunologically important adaptation in these systems.

One central challenge in HIV vaccine research is to devise a means to stimulate a lineage producing broadly neutralizing antibodies. Common characteristics of BnAbs, such as high levels of somatic mutation or large insertions, often lead to their depletion by mechanisms of immune tolerance control [56]. Therefore, one strategy to elicit these antibodies is to stimulate the progenitors of their clonal lineage, which may be inferred by phylogenetic methods [57], and to guide their affinity maturation process towards a broadly neutralizing state. Understanding the underlying evolutionary process is necessary to make principled progress towards such strategies, and this study represents a step in that direction. For example, our results suggest that a vaccine based on a genetically diverse set of viral antigens is more likely to stimulate BnAbs. Similar preference for stimulation of BnAbs was shown by a numerical study of co-evolution between B-cell receptors and HIV within a patient [30].

Materials and Methods

Method summary

Alternative fitness models. We assume that the malthusian (log-) fitness of an antibody is proportional to the logarithm of the activation probability, $f_A(A^\alpha; \{V\}) \propto c_a \log[\rho(A^\alpha; \{V\})]$. The activation probability $\rho(A^\alpha; \{V\})$ is a nonlinear sigmoid function of the binding affinity with a binding threshold E^* and a nonlinearity β_0. We assume that the affinity of an antibody is determined by its strongest binding among R interactions, which we estimate by means of extreme value statistics, $E^\alpha_{\max} = E_{\alpha\,.} + \hat{E}_{\alpha\,.} + \sqrt{2 I_{\alpha\,.} \ln R}$. Here, $I_{\alpha\,.}$ is the variance of binding affinities of a given antibody A^α across the viral population. The *linear-averaged* fitness function in eq. (2) is an approximation to this sigmoidal function, when the nonlinearity and the logarithm of the number of interactions are small. The selection coefficient of antibodies in eq. (2) relates to the biophysical conditions for the binding interactions, $S_a = c_a \beta_0 / (1 + \exp[-\beta_0 E^*])$. Similarly, we relate viral fitness to the probability that it avoids binding to the circulating antibodies. The effect of nonlinearities of the fitness function on the stationary binding statistics and the rate of adaptation are shown in Fig. S4.

Evolution of the binding affinity diversity. Similar to the mean binding affinity in eqs. (4,5), we can derive evolution equations for the diversity of binding in the antibody and viral populations, with mutations, selection and genetic drift. The diffusion equation for the binding diversity in the variable regions of antibodies follows,

$$\frac{d}{dt} M_{A,2} = -4\mu_a (M_{A,2} - \ell \mathcal{K}_2) - 4\mu_v M_{A,2} - \frac{M_{A,2}}{N_a} + S_a M_{A,3} + \chi_{M_{A,2}} \tag{S1}$$

where, $\chi_{M_{A,2}}$ is a Gaussian correlated noise with mean 0 and variance, $(M_{A,4} - (M_{A,2})^2)/N_a$. Similar equations can be derived for antibody diversity in the conserved interaction regions, $\hat{M}_{A,2}$, and diversity in viruses $M_{V,2}$. The stationary solutions for the binding diversity and the higher central moments of the binding affinity distributions are given by eqs. (S78-S81).

Separation of time-scales. The binding diversities fluctuate with an auto-correlation time of the order of the neutral coalescence times: $\tau_{M_{A,2}} \sim N_a$, and $\tau_{M_{V,2}} \sim N_v$ in generations, respectively for antibodies and viruses (see SI and Fig. S3). By contrast, the mean binding affinity in the variable region exhibits a slower dynamics, with an auto-correlation time proportional to the inverse of the neutral sequence diversities, $\tau_{\mathcal{E}} \sim N_a/(\theta_a + \theta_v(N_a/N_v))$. The separation of time scales between mean and diversity of binding affinity allows us to describe the dynamics of the mean in eqs. (4,5) in terms of the stationary binding diversities. Moreover, the mean binding affinity in the conserved region has an autocorrelation time longer than in the variable region, $\tau_{\hat{\mathcal{E}}} \sim N_a/\theta_a$. This difference in time-scales allows us to neglect the correlations between the means and diversities, as well as between the conserved and variable regions.

Competition between multiple lineages. In the linear fitness landscape of eq. (2), the relative mean fitness and fitness flux of a lineage, and hence, its fixation probability in eq. (11), can be expressed in terms of the binding affinity deviations of the lineage's constituent genotypes from the population mean. For example, the r^{th} lineage-specific moment in the variable region is defined as $\overset{\mathcal{C}}{L}_{A_r} = \left\langle \sum_\alpha (E^{\mathcal{C}}_{\alpha\,.} - \mathcal{E})^r \underset{\mathcal{C}}{x^\alpha} \right\rangle$, where $\underset{\mathcal{C}}{x^\alpha}$ is the frequency of genotype A^α from lineage $\mathcal{C}$ in the population, and $E^{\mathcal{C}}_{\alpha\,.}$ is its mean binding affinity. The transfer flux from viruses to the lineage $\mathcal{C}$ can be expressed in terms of the cross-statistics between the binding affinity of the lineage and the viral population: $\overset{\mathcal{C}}{L}_{A_1,V_1} = \langle \sum_{\alpha\gamma} (E_{.\,\gamma} - \mathcal{E})\,(E_{\alpha\gamma} - \mathcal{E})\, y^\gamma \underset{\mathcal{C}}{x^\alpha} \rangle$. Analogous lineage-specific moments can be defined for the conserved viral region.

The hierarchy of evolution equations for the lineage-specific moments are given by eqs. (S115-S122). The fixation probability is the asymptotic value of the ensemble-averaged lineage frequency $P_{\text{fix}}(\mathcal{C}) = \lim_{t\to\infty} \overset{\mathcal{C}}{L}_0(t)$, which we compute by using the Laplace transform of the lineage-specific moments, $\mathcal{L}_{A_r}(z) = \sum_t L_{A_r}(t) \exp[-zt]$. It can be separated up to orders of $\mathcal{O}((NS)^2, NS\theta)$ to contributions from (i) the relative mean fitness of the lineage which is proportional to $N_a S_a$, (ii) the relative fitness flux of the lineage which is proportional to $(N_a S_a)^2$ and (iii) the relative transfer flux from viruses to the lineage which is proportional to $(N_a S_a) \times (N_v S_v)$,

$$\begin{aligned} P_{\text{fix}}(\mathcal{C}) - P_{0_{\text{fix}}}(\mathcal{C}) &= N_a\, S_a \big(L^{\mathcal{C}}_{A_1}(0) + \hat{L}^{\mathcal{C}}_{A_1}(0)\big) \\ &+ (N_a\, S_a)^2 \big[L^{\mathcal{C}}_{A_2}(0) + \hat{L}^{\mathcal{C}}_{A_2}(0) - L^{\mathcal{C}}_{A_{(0;2)}}(0) - \hat{L}^{\mathcal{C}}_{A_{(0;2)}}(0)\big] \\ &- (N_v\, S_v) \times (N_a S_a) \big[L^{\mathcal{C}}_{A_1,V_1}(0) - L^{\mathcal{C}}_{A_0,V_2}(0)\big] \end{aligned} \tag{S2}$$

where, $P_{0_{\text{fix}}}(\mathcal{C}) = L^{\mathcal{C}}_{A_0}(0)$ is the expected fixation probability in neutrality. $L^{\mathcal{C}}_{A_{(0;2)}}$ and $L_{A_0,V_2} = \langle \rho^{\mathcal{C}} M_{V,2} \rangle$ are respectively the total diversity of binding in the antibody and in the viral population, scaled by the frequency of the lineage $\mathcal{C}$; see Section 4 of SI for detailed derivation of eq. (S2). The diversity of binding affinity in viruses is a population observable which affects the lineage fixation probability, as shown in Fig. 4.

Simulations. Simulations specified two populations with genotypes, selection, and mutation processes as defined in Model section. For each generation, the populations were replaced by their offspring, which inherit genotypes from their parents, and the number of individuals of each genotype is drawn from a multinomial distribution, with probabilities proportional to the exponential of the fitnesses. The number of mutations was determined by a poisson random number with mean equal to the expected total number of mutations. Populations were initialized as monomorphic, i.e., a uniformly random genotype. Programs were written in luajit, utilizing the gnu scientific library and gnu parallel [58]. To measure quantities in the stationary state (Figs. S1, 3) simulations were run for $10^4 N_a$ generations, and quantities were averaged from samples every $2N_a$ generations, omitting the first $2\mu_a^{-1}$ generations.

To produce the simulations shown in Fig. 4, the newly emerging antibody lineages compete with the resident population as follows. First, the resident lineage is evolved with the virus for $20N_a$ generations. Then the resident and invading lineage are evolved with the virus for $10N_a$ generations, with fixed sizes 10^3, and viral fitness determined only by the resident lineage. This ensures that competing lineages can marginally bind to the viral population, and are functional lineage progenitors; a process that happens prior to affinity maturation in germinal centers. The pre-adaptation of the invading lineage can also be interpreted as initial rounds of affinity maturation in germinal centers isolated from competition with adapted antibody lineages. Then the two lineages are combined with resident at 95% and invader at 5%, with a total size of 10^3, and the state of the system is recorded. The two lineages are evolved until one is extinct, repeated over 100 replicates to estimate the fixation probability. The whole procedure is repeated 3×10^3 times for ensemble averaging. The invader, is either a normal lineage or a BnAb that binds only to the conserved region.

1. Antibody-viral co-evolution in the genotype space

We represent the antibody population as a set of k genotypes consisting of vectors, $\mathbf{A}^\alpha$ $(\alpha = 1 \dots k)$, and corresponding genotype frequencies $\mathbf{x}$, with elements x^α satisfying $\sum_{\alpha=1}^{k} x^\alpha = 1$. Similarly, we consider a viral population with k' possible genotypes $\mathbf{V}^a$, and frequencies $\mathbf{y}$ with elements y^γ $(\gamma = 1, \dots, k')$ with $\sum_{\gamma=1}^{k'} y^\gamma = 1$. Note that superscripts are indices, not exponentiation, unless next to parentheses, e.g. $(a)^b$. The frequencies change over time, although we omit explicit notation for brevity, and hence every quantity that depends on the frequencies is itself time-dependent.

In the following, we describe the co-evolution of an antibody and a viral population in terms of three evolutionary forces: mutation, selection and genetic drift. We assume that population sizes are large enough, and changes in frequencies are small enough to accommodate a continuous time and continuous frequency stochastic process [59, 60].

Mutations. The change of the genotype frequencies due to mutations follow the Master equation,

$$\begin{aligned} \frac{dx^\alpha}{dt} &= m^\alpha_{A}(\mathbf{x}) \equiv \sum_{\beta=1}^{k} \mu_{\beta\to\alpha} x^\beta - \left(\sum_{\beta=1}^{k} \mu_{A^\alpha \to A^\beta}\right) x^\alpha \\ \frac{dy^\gamma}{dt} &= m^\gamma_{V}(\mathbf{y}) \equiv \sum_{\lambda=1}^{k'} \mu_{V^\lambda \to V^\gamma} y^\lambda - \left(\sum_{\lambda=1}^{k'} \mu_{V^\gamma \to V^\lambda}\right) y^\gamma \end{aligned} \tag{S3}$$

where we define m_A and m_V as mutational fields in the antibodies and viruses, and $\mu_{A^\beta \to A^\alpha}$ is the antibody mutation rate from genotype $\mathbf{A}^\beta$ to $\mathbf{A}^\alpha$, and similarly, $\mu_{V^\lambda \to V^\gamma}$ is the viral mutation rate from the genotype $\mathbf{V}^\lambda$ to $\mathbf{V}^\gamma$. We assume constant mutation rates μ_a, μ_v, per generation per site for antibodies and viruses, with the exception of $\mu_v = 0$ for the viral constant region, which implies that mutations in that region are lethal for the virus.

Selection and interacting fitness functions. The fitness of a genotype determines its growth rate at each point in time. We define fitness of genotypes in one population as a function of the genotypes in the other population. The general form of change in genotype frequencies due to selection follows,

$$\frac{1}{x^\alpha}\frac{dx^\alpha}{dt} = f_{A^\alpha}(\mathbf{x};\mathbf{y}) - \sum_\alpha x^\alpha f_{\mathbf{A}^\alpha}(\mathbf{x};\mathbf{y}) \tag{S4}$$

$$\frac{1}{y^\gamma}\frac{dy^\gamma}{dt} = f_{V^\gamma}(\mathbf{y};\mathbf{x}) - \sum_\gamma y^\gamma f_{V^\gamma}(\mathbf{y};\mathbf{x}) \tag{S5}$$

The subscript for the antibody and viral fitness functions, $f_{A^\alpha}(\mathbf{x};\mathbf{y})$ and $f_{V^\gamma}(\mathbf{y};\mathbf{x})$, refer to the genotypes in the corresponding population. The explicit dependence of the antibody fitness function on the viruses $\{\mathbf{V}\}$, denoted by the conditional dependence of the fitness function $f_{A^\alpha}(\mathbf{x};\mathbf{y})$ on the frequency vector $\mathbf{y}$, emphasizes that fitness of an antibody depends on the interacting viral population. Similar notation is used for the fitness function of the viruses. The subtraction of the population's mean fitness, $F_A = \sum_\alpha x^\alpha f_{A^\alpha}(\mathbf{x};\mathbf{y})$ and $F_V = \sum_\gamma y^\gamma f_{V^\gamma}(\mathbf{y};\mathbf{x})$, ensures that the genotype frequencies remain normalized in each population. In terms of linearly independent frequencies $\mathbf{x} = (x^1, \ldots, x^{k-1})$ and $\mathbf{y} = (y^1, \ldots, y^{k'-1})$, these evolution equations take the forms,

$$\frac{dx^\alpha}{dt} = {\sigma_A}^\alpha(\mathbf{x};\mathbf{y}) \equiv \sum_\beta g^{\alpha\beta}\,\sigma_{A^\beta}(\mathbf{x};\mathbf{y}) \tag{S6}$$

$$\frac{dy^\gamma}{dt} = {\sigma_V}^\gamma(\mathbf{y};\mathbf{x}) \equiv \sum_\lambda h^{\gamma\lambda}\,\sigma_{V^\lambda}(\mathbf{y};\mathbf{x}) \tag{S7}$$

with selection coefficients,

$$\sigma_{A^\alpha}(\mathbf{x},\mathbf{y}) = f_{A^\alpha}(\mathbf{x};\mathbf{y}) - f_{A^k}(\mathbf{x};\mathbf{y}) \tag{S8}$$

$$\sigma_{V^\gamma}(\mathbf{x},\mathbf{y}) = f_{V^\gamma}(\mathbf{y};\mathbf{x}) - f_{V^{k'}}(\mathbf{y};\mathbf{x}) \tag{S9}$$

$\sigma_{A^\alpha}(\mathbf{x};\mathbf{y})$ and $\sigma_{V^\gamma}(\mathbf{y};\mathbf{x})$ are the respective time dependent selection coefficients of the antibody A^α and the viral strain V^γ, which depend on the state of the both populations at that moment in time. The inverse of the response matrices, $g_{\alpha\beta} = (g^{\alpha\beta})^{-1}$ and $h_{\gamma\lambda} = (h^{\gamma\lambda})^{-1}$, play the role of metric in the genotype space (see below and e.g., [61]). The change in the mean fitness due to selection after an infinitesimal amount of time follows,

$$F_A(\mathbf{x}+\delta\mathbf{x};\mathbf{y}+\delta\mathbf{y}) = \sum_\alpha \sigma_{A^\alpha}(\mathbf{x};\mathbf{y})\delta x^\alpha + \sum_{\gamma,\alpha} x^\alpha\,\sigma_{V^\gamma \to A^\alpha}(\mathbf{x};\mathbf{y})\,\delta y^\gamma \tag{S10}$$

$$F_V(\mathbf{y}+\delta\mathbf{y};\mathbf{x}+\delta\mathbf{x}) = \sum_\gamma \sigma_{V^\gamma}(\mathbf{y};\mathbf{x})\delta y^\gamma + \sum_{\gamma,\alpha} y^\gamma\,\sigma_{A^\alpha \to V^\gamma}(\mathbf{y};\mathbf{x})\,\delta x^\alpha \tag{S11}$$

where δx^α and δy^γ are the infinitesimal changes in the genotype frequencies, and $\sigma_{V^\gamma \to A^\alpha} = \partial\sigma_{A^\alpha}/\partial y^\gamma$ and, $\sigma_{A^\alpha \to V^\gamma} = \partial\sigma_{V^\gamma}/\partial x^\alpha$ are respectively transfer selection from the viral population to antibodies and vice versa. The transfer of fitness is a useful concept for interacting populations. Intuitively, it can be understood as the change of fitness in one population only due to the changes of allele or genotype frequencies in the opposing population.

Genetic drift and stochasticity. The stochasticity of reproduction and survival, commonly known as genetic drift, is represented as binomial sampling of genotypes in the next generation, with the constraint of a total population size, in discrete processes such as the Wright-Fisher process. The magnitude of this sampling noise is proportional to inverse population size. In the continuous time, continuous frequency limit, genetic drift is represented as noise terms with magnitude proportional to inverse population size [60]. N_a and N_v are the effective population sizes of the antibody and the viral populations, which represent the number of interacting partners in a germinal center. The diffusion coefficients are characteristics of the Fisher metric [61, 62],

$$g^{\alpha\beta} = \begin{cases} -x^\alpha x^\beta & \text{if } \alpha \neq \beta \\ x^\alpha(1-x^\alpha) & \text{if } \alpha = \beta \end{cases}, \qquad h^{\gamma\lambda} = \begin{cases} -y^\gamma y^\lambda & \text{if } \gamma \neq \lambda \\ y^\gamma(1-y^\gamma) & \text{if } \gamma = \lambda \end{cases} \tag{S12}$$

The generalized Kimura diffusion equation [63] for the joint distribution of genotype frequencies $P(\mathbf{x}, \mathbf{y}, t)$ in both populations reads,

$$\begin{aligned} \frac{\partial}{\partial t} P(\mathbf{x}, \mathbf{y}, t) = \sum_{\alpha,\beta,a,b} \Big[& \frac{1}{2N_a} \frac{\partial^2}{\partial x^\alpha x^\beta} g^{\alpha\beta}(\mathbf{x}) + \frac{1}{2N_v} \frac{\partial^2}{\partial y^\gamma \partial y^\lambda} h^{\gamma\lambda}(\mathbf{y}) \\ & + \frac{\partial}{\partial x^\alpha} \left({m_{_A}}^\alpha + {\sigma_{_A}}^\alpha(\mathbf{x}; \mathbf{y}) \right) + \frac{\partial}{\partial y^\gamma} \left({m_{_V}}^\gamma + {\sigma_{_V}}^\gamma(\mathbf{y}; \mathbf{x}) \right) \Big] P(\mathbf{x}, \mathbf{y}, t) \end{aligned} \tag{S13}$$

This Fokker-Planck equation acts on the high dimensional genotype space of antibodies and viruses, which are likely to be under sampled in any biological setting. In the following, we introduce a projection from genotype space onto a lower dimensional space of molecular traits (phenotypes) to make the problem more tractable.

2. Antibody-viral co-evolution in the phenotype space

2.1 Molecular traits for antibody-viral interaction

We define the interaction between an antibody and viral genotype, the binding affinity, which serves a molecular traits for which we will describe the evolutionary dynamics. Binding affinity is related to the neutralization efficacy of antibodies and the immune pressure on the virus. Therefore, we will define selection on antibodies and viruses as a function of binding affinity. Antibody and viral genotypes are represented by binary sequences of ±1. Antibody sequences are of length $\ell + \hat{\ell}$, while viral sequences consist of a mutable region of length ℓ, and a conserved (i.e. sensitive) region of length $\hat{\ell}$, where each site is always $+1$, as was similarly done in [26]. We model the binding affinity between antibody $\mathbf{A}^\alpha$ and virus $\mathbf{V}^\gamma$ as a weighted dot product over all sites,

$$\begin{aligned} E_{\text{tot}}(\mathbf{A}^\alpha, \mathbf{V}^\gamma) &= \sum_{i=1}^{\ell} \varepsilon_i A_i^\alpha V_i^\gamma + \sum_{i=\ell+1}^{\ell+\hat{\ell}} \hat{\varepsilon}_i \, A_i^\alpha \\ &\equiv E_{\alpha\,\gamma} + \hat{E}_\alpha \end{aligned} \tag{S14}$$

where A_i^α, and V_i^γ denote the i^{th} site for antibodies and viruses. The set of coupling constants for the mutable and conserved region, $\{\varepsilon_i, \hat{\varepsilon}_i \geq 0\}$ represent the accessibility of a clonal antibody lineage to regions of the viral sequence. Matching sites between an antibody and a viral string has been used as a model for binding affinity in the context of T-cell selection in the adaptive immune system, which is based on the capability to bind well to the external antigens and to avoid the self proteins [33, 34]. In Section 4, we extend our model to multiple lineages, where each lineage has its own set of accessibilities. Antibody lineages with access to the conserved regions of the virus can potentially fix as broadly neutralizing antibodies. In the following, quantities related to the conserved sites of the virus are note with a hat: $\hat{\cdot}$.

We project the evolutionary forces from the genotype level onto the binding phenotype (trait) E, and quantify the changes of the binding phenotype distribution in each population over time. For a single antibody genotype $\mathbf{A}^\alpha$ we characterize its interactions with the viral population by the *genotype-specific moments*,

Statistics of the binding affinity distribution for antibody $\mathbf{A}^\alpha$:

(i) average in the variable region:

$$E_{\alpha\,\cdot} = \sum_{\gamma\in\text{ viruses}} E_{\alpha\,\gamma} y^\gamma \tag{S15}$$

(ii) average in the conserved region:

$$\hat{E}_{\alpha\,\cdot} = \hat{E}_\alpha \tag{S16}$$

(iii) r^{th} central moment in the variable region:

$$I^{(r)}_{\alpha\,\cdot} = \sum_{\gamma\in\text{ viruses}} (E_{\alpha\,\gamma} - E_{\alpha\,\cdot})^r y^\gamma \tag{S17}$$

Since the viral population is monomorphic in the conserved region, the average mean binding affinity of an antibody is independent of the state of the viral population, $\hat{E}_{\alpha\,\cdot} = \hat{E}_\alpha$, and the higher central moments are zero, $\hat{I}^{(r)}_{\alpha\,\cdot} = 0$. Similarly, for a viral genotype $\mathbf{V}^\gamma$, we characterize its interactions with the antibody population by the genotype specific moments,

Statistics of the binding affinity distribution for virus $\mathbf{V}^\gamma$:

(i) average in the variable region:

$$E_{\cdot\,\gamma} = \sum_{\alpha\in\text{antibodies}} E_{\alpha\,\gamma} x^\alpha \tag{S18}$$

(ii) average in the conserved region:

$$\hat{E}_{\cdot\,\gamma} = \sum_{\alpha\in\text{antibodies}} \hat{E}_\alpha x^\alpha \equiv \hat{E}_\cdot \tag{S19}$$

(iii) r^{th} central moment in the variable region:

$$I^{(r)}_{\cdot\,\gamma} = \sum_{\alpha\in\text{antibodies}} (E_{\alpha\,\gamma} - E_{\cdot\,\gamma})^r x^\alpha \tag{S20}$$

(iii) r^{th} central moment in the conserved region:

$$\hat{I}^{(r)}_{\cdot\,\gamma} = \sum_{\alpha\in\text{antibodies}} (\hat{E}_\alpha - \hat{\mathcal{E}})^r x^\alpha \tag{S21}$$

One of the most informative statistics that we characterize is the distribution of averaged antibody and viral interactions, respectively denoted by $P_A(E_{\alpha\,\cdot}, \hat{E}_\alpha)$ and $P_V(E_{\cdot\,\gamma}, \hat{E}_\cdot)$. The mean of these distributions are equal to each other, but the higher moments differ. We denote the *population-specific moments* of the average interaction affinity by,

Average binding affinity in,

(i) the variable region:

$$\mathcal{E} = \sum_{\alpha,\gamma} E_{\alpha\,\gamma}\, x^\alpha y^\gamma \tag{S22}$$

(ii) the conserved region:

$$\hat{\mathcal{E}} = \sum_{\alpha} \hat{E}_\alpha\, x^\alpha \tag{S23}$$

r^{th} central moment of distributions of mean affinities in,

(i) the variable region of antibody population:

$$M_{A,r} = \sum_{\alpha} (E_{\alpha\,.} - \mathcal{E})^r x^{\alpha} \tag{S24}$$

(ii) the conserved region of antibody population:

$$\hat{M}_{A,r} = \sum_{\alpha} (\hat{E}_{\alpha\,.} - \hat{\mathcal{E}})^r x^{\alpha} \tag{S25}$$

(iii) the variable region of viral population:

$$M_{V,r} = \sum_{\gamma} (E_{.\,\gamma} - \mathcal{E})^r y^{\gamma} \tag{S26}$$

Note that the population central moments $M_{A,r}$ and $M_{V,r}$ are distinct from the genotype-specific moments, $I^{(r)}_{\alpha\,.}$ and $I^{(r)}_{.\,\gamma}$. The central moments of the viral population in the conserved region of the virus are equal to zero, $\hat{M}_{V,r} = 0$.

Trait scale and dimensionless quantities. It is useful to measure traits in natural units, which avoids the arbitrariness of the physical units of ε_i, and the total number of sites. As previously shown in [32, 36], there exist summary statistics of the site specific effects, (here ε_i), which define a natural scale of the molecular trait. We define the moments of ε_i along the genome,

$$\mathcal{K}_r = \frac{1}{\ell} \sum_{i=1}^{\ell} (\varepsilon_i)^r, \qquad \hat{\mathcal{K}}_r = \frac{1}{\hat{\ell}} \sum_{i=\ell+1}^{\ell+\hat{\ell}} (\hat{\varepsilon}_i)^r \tag{S27}$$

We express the trait statistics in units of the squared sum of site specific effects, $E_0^2 = \mathcal{K}_2 \ell$ for the variable region, and $\hat{E}_0^2 = \hat{\mathcal{K}}_2 \hat{\ell}$, for the conserved region.

$$\varepsilon \equiv \frac{\mathcal{E}}{E_0}, \qquad \hat{\varepsilon} \equiv \frac{\mathcal{E}}{\hat{E}_0} \qquad \text{and,} \qquad m_{Z,r} \equiv \frac{M_{Z,r}}{E_0^r}, \qquad \hat{m}_{Z,r} \equiv \frac{\hat{M}_{Z,r}}{\hat{E}_0^r} \qquad (\text{for } Z = A, V) \tag{S28}$$

These scaled values are pure numbers (we distinguish them by use of lower case letters from the raw data). E_0^2 and $\hat{E}_0^2$ are a natural way to standardize the relevant quantities because they are the stationary ensemble variances of the population mean binding affinity in an ensemble of genotypes undergoing neutral evolution in the weak-mutation regime (see Section 2.3 for derivation of the stationary statistics),

$$E_0^2 = \lim_{\mu_v, \mu_a \to 0} \langle (\mathcal{E} - \langle \mathcal{E} \rangle)^2 \rangle, \qquad \hat{E}_0^2 = \lim_{\mu_a \to 0} \langle (\hat{\mathcal{E}} - \langle \hat{\mathcal{E}} \rangle)^2 \rangle \tag{S29}$$

where $\langle \cdot \rangle$ indicates averages over an ensemble of independent populations.

Binding probability. The probability that an antibody is bound by an antigen determines its chance of proliferation and survival during the process of affinity maturation, which will help us define the fitness of genotypes. We describe two distinct models for antibody activation. The simplest model assumes that the binding probability of a given antibody $\mathbf{A}^{\alpha}$ is a sigmoid function of its *mean binding affinity* against the viral population,

$$\rho_{A}(\mathbf{A}^{\alpha}) = \frac{1}{1 + \exp[-\beta_0 (E_{\alpha\,.} + \hat{E}_{\alpha\,.} - E^*)]} \tag{S30}$$

where E^* is the threshold for the binding affinity and β_0 determines the amount of nonlinearity, and is related to the inverse of temperature in thermodynamics. Following the rescaling introduced in eq. (S28), the binding threshold and the nonlinearity in eq. (S30) rescale as $e^* \equiv E^* / \sqrt{\hat{E}_0^2 + E_0^2}$ and $\beta = \beta_0 \sqrt{\hat{E}_0^2 + E_0^2}$. In the following, we will use eq. (S30) to characterize a biophysically grounded fitness function for an antibody strain.

For the virus, binding to an antibody reduces the chances of its survival. Therefore, we use the averaged binding affinity approach to characterize its binding probability to the antibodies. Similar to eq. (S30), the binding probability of a given virus $\mathbf{V}^{\gamma}$ is,

$$\rho_V(\mathbf{V}^\gamma) = \frac{1}{1+\exp[-\beta_0(E_{\cdot\,\gamma}+\hat{E}_{\cdot\,\gamma}-E^*)]}, \tag{S31}$$

where E^* and β_0 are similar to eq. (S30).

In Section 2.5, we will discuss an alternative model for activation of an antibody which is based on its *strongest binding affinity* with a subset of viruses.

2.2 Co-evolutionary forces on the binding affinity

Similar to genotype evolution, stochastic evolution of a quantitative trait generates a probability distribution, $Q(\mathcal{E}, M_{A,r}, M_{V,r})$, which describes an ensemble of independently evolving populations, each having a trait distribution with mean of average affinity $\mathcal{E}$ and central moments of the averaged affinity in the antibody population, $M_{A,r}$, $\hat{M}_{A,r}$, and in the viral population, $M_{V,r}$ (see also [36]). The probability $Q(\mathcal{E}, M_{A,r}, \hat{M}_{A,r}, M_{V,r})$ is a sum of probabilities of genotype frequencies,

$$Q(\mathcal{E}, M_{A,r}, \hat{M}_{A,r}, M_{V,r}) = \\ \int \delta(\mathcal{E}(\mathbf{x},\mathbf{y})-\mathcal{E}) \prod_r \delta(M_{A,r}(\mathbf{x},\mathbf{y})-M_{A,r})\delta(\hat{M}_{A,r}(\mathbf{x})-\hat{M}_{A,r})\delta(M_{V,r}(\mathbf{x},\mathbf{y})-M_{V,r})P(\mathbf{x},\mathbf{y},t)d\mathbf{x}d\mathbf{y} \tag{S32}$$

where $\delta(\cdot)$ is the Dirac delta function. Furthermore, we assume that fitnesses of genotypes (defined below) only depend on the trait distribution. Below, we characterize the effect of mutations, selection and genetic drift on the evolution of the trait moments $\mathcal{E}$, $M_{A,r}$, $\hat{M}_{A,r}$ and $M_{V,r}$.

Mutation. A mutation at site i changes the sign of the site, and its effect on the binding affinity is proportional to κ_i. To compute the effect of mutations on the traits moments, we classify pairs of genotypes $(\mathbf{A}^\alpha, \mathbf{V}^\gamma)$, in mutational classes, defined by the number of $+1$ positions, of their product vector, $(A_1^\alpha \cdot V_1^\gamma, \ldots, A_{\ell+\hat{\ell}}^\alpha \cdot V_{\ell+\hat{\ell}}^\gamma)$ in the variable n_+ and in the conserved interaction region $\hat{n}_+$

$$n_+(\mathbf{A}^\alpha, \mathbf{V}^\gamma) = \sum_{i=1}^{\ell} \delta(1-A_i^\alpha \cdot V_i^\gamma), \qquad \hat{n}_+(\mathbf{A}^\alpha, \mathbf{V}^\gamma) = \sum_{i=\ell+1}^{\ell+\hat{\ell}} \delta(1-A_i^\alpha \cdot V_i^\gamma) \tag{S33}$$

The frequency of a each mutational class $\rho(n_+)$ is estimated from interactions between all pairs of antibody and viral genotypes in the population for both variable and conserved regions,

$$\rho^{(1)}(n_+) = \frac{1}{N_a N_V}\sum_{\alpha,\gamma}\delta(n_+(\mathbf{A}^\alpha, \mathbf{V}^\gamma)-n_+), \qquad \rho^{(2)}(n_+) = \frac{1}{N_a N_V}\sum_{\alpha,\gamma}\delta(\hat{n}_+(\mathbf{A}^\alpha, \mathbf{V}^\gamma)-n_+) \tag{S34}$$

We find the change in frequency $\rho^{(\lambda)}(n_+)$ (with $\lambda = 1, 2$) by mutations occurring in both populations in one generation. The superscript $\lambda = 1, 2$ indicates the interacting region of the virus, i.e. $\lambda = 1$ refers to the variable region of the virus with $\mu_v^{(1)} = \mu_v$ and the length $\ell^{(1)} = \ell$, and $\lambda = 2$ refers to the conserved region of the viral genome with $\mu_v^{(2)} = 0$ and the sequence length $\ell^{(2)} = \hat{\ell}$. As shown in [36, 64, 65], for a large number of sites in a trait, the $\mathcal{K}_r$ sufficiently determines the effect of mutations on the trait moments $M_{A,r}$, $\hat{M}_{A,r}$ and $M_{V,r}$. If the mutational effects of all sites were equal to ε, we have: $\mathcal{E} = (2[n_+]_{A,V} - \ell)\varepsilon$, where $[\cdot]_{A,V}$ indicates averaging of a quantity in the subscript populations, which in this case are both the viral and the antibody populations. If the interaction contributions κ_i differ between sites, the mean and central moments of the binding affinity can be approximated with an *annealed average* of site contributions. For the mean affinity $\mathcal{E}$ it results in $\mathcal{E} = (2[n_+]_{A,V} - \ell)\mathcal{K}_1$, and similarly for the higher central moments, $M_{V,r} = 2^r \mathcal{K}_r \left[\left([n_+]_A - [n_+]_{A,V}\right)^r\right]_V$ and $M_{A,r} = 2^r \mathcal{K}_r \left[\left([n_+]_V - [n_+]_{A,V}\right)^r\right]_A$. The Master equation for the evolution of the mutational classes under neutrality (mutation and genetic drift) follows,

$$\begin{aligned} d\,\rho^{(\lambda)}(n_+) \;&=\; \Big[(\mu_a+\mu_v^{(\lambda)})(\ell^{(\lambda)}-(n_+-1))\rho^{(\lambda)}(n_+-1)+\mu^{(\lambda)}(n_++1)\rho^{(\lambda)}(n_++1)-\mu^{(\lambda)}\ell^{(\lambda)}\,\rho^{(\lambda)}(n_+)\Big]\,dt \\ &\quad +\big(\delta_{n'_+,n_+}-\rho^{(\lambda)}(n_+)\big)\left[\sqrt{\frac{\rho^{(\lambda)}(n_+)}{N_a}}dW_{{}_A}(t)+\sqrt{\frac{\rho^{(\lambda)}(n_+)}{N_v}}dW_{{}_V}(t)\right] \end{aligned} \tag{S35}$$

$W_{{}_A}(t)$ and $W_{{}_V}(t)$ are delta-correlated Gaussian noise (Wiener process) with an ensemble mean $\langle W_i\rangle = 0$ and variance, $\langle W_i(t)W_j(t')\rangle = \delta_{i,j}\,\delta(t-t')$ where $i,j\in\{A,V\}$ indicate antibodies and viruses. The stochasticity (genetic drift) is due to finite population size of the interacting genotypes N_a and N_v.

In neutrality, the ensemble mean for the averaged number of upwards spins $\big\langle [{n_+}^{(\lambda)}]_{{}_{A,V}}\big\rangle$ and the central moments of the upward spins, $\langle Y^{(\lambda)}_{A,r}\rangle \equiv \Big\langle \big[\big([n_+^{(\lambda)}]_{{}_V}-[n_+^{(\lambda)}]_{{}_{A,V}}\big)^r\big]_{{}_A}\Big\rangle$ and $\langle Y^{(\lambda)}_{V,r}\rangle \equiv \Big\langle \big[\big([n_+^{(\lambda)}]_{{}_A}-[n_+^{(\lambda)}]_{{}_{A,V}}\big)^r\big]_{{}_V}\Big\rangle$ in both variable $(\lambda=1)$ and conserved $(\lambda=2)$ interaction regions follow [64, 65],

$$\begin{aligned} \frac{\partial\langle[{n_+}^{(\lambda)}]_{{}_{A,V}}\rangle}{\partial t} \;&=\; \left\langle (\mu_a+\mu_v^{(\lambda)})\sum_{n_+} n_+\Big[(\ell^{(\lambda)}-n_++1)\rho^{(\lambda)}(n_+-1)+(n_++1)\rho^{(\lambda)}(n_++1)-\ell^{(\lambda)}\rho^{(\lambda)}(n_+)\Big]\right\rangle \\ &=\; \begin{cases} -2(\mu_a+\mu_v)\Big[(\big\langle[n_+]_{{}_{A,V}}\big\rangle-\ell/2\Big) & \text{variable region, } \lambda=1 \\ \\ -2\mu_a\Big(\big\langle[n_+]_{{}_{A,V}}\big\rangle-\hat{\ell}/2\Big) & \text{constant region, } \lambda=2 \end{cases} \end{aligned} \tag{S36}$$

$$\begin{aligned} \frac{\partial}{\partial t}\left\langle Y^{(\lambda)}_{A,r}\right\rangle \;&=\; \mu_a\ell^{(\lambda)}\sum_{i=0}^{r-2}\binom{r}{i}\left\langle Y^{(\lambda)}_{A,i}\right\rangle+\frac{\binom{r}{2}\left\langle Y^{(\lambda)}_{A,2}Y^{(\lambda)}_{A,r-2}\right\rangle-r\left\langle Y^{(\lambda)}_{A,r}\right\rangle}{N_a}-2r(\mu_a+\mu_v^{(\lambda)})\left\langle Y^{(\lambda)}_{A,r}\right\rangle \\ &\quad -\mu_a\sum_{i=0}^{r-2}\binom{r}{i}\left(\left\langle Y^{(\lambda)}_{A,i+1}\right\rangle+\left\langle[n_+]_{{}_{A,V}}Y^{(\lambda)}_{A,i}\right\rangle\right)\,[1+(-1)^{r-i+1}] \end{aligned} \tag{S37}$$

$$\begin{aligned} \frac{\partial}{\partial t}\left\langle Y^{(\lambda)}_{V,r}\right\rangle \;&=\; \mu_v^{(\lambda)}\ell^{(\lambda)}\sum_{i=0}^{r-2}\binom{r}{i}\left\langle Y^{(\lambda)}_{V,i}\right\rangle+\frac{\binom{r}{2}\left\langle Y^{(\lambda)}_{V,2}Y^{(\lambda)}_{V,r-2}\right\rangle-r\left\langle Y^{(\lambda)}_{V,r}\right\rangle}{N_v}-2r(\mu_a+\mu_v^{(\lambda)})\langle Y^{(\lambda)}_{V,r}\rangle \\ &\quad -\mu_v^{(\lambda)}\sum_{i=0}^{r-2}\binom{r}{i}\left(\left\langle Y^{(\lambda)}_{V,i+1}\right\rangle+\left\langle[n_+]_{{}_{A,V}}Y^{(\lambda)}_{V,i}\right\rangle\right)\,[1+(-1)^{r-i+1}] \end{aligned} \tag{S38}$$

where $\langle\cdot\rangle$ denotes averages over independent ensembles of populations. The second term in the right hand side of equations (S37) and (S38) is a consequence of the Itô calculus in stochastic processes [59]. The transformations from $[n_+^{(1)}]_{{}_{A,V}}$ to $\mathcal{E}$ in the variable region, and from $[n_+^{(2)}]_{{}_{A,V}}$ to $\hat{\mathcal{E}}$ in the conserved region result in

$$\frac{\partial\langle\mathcal{E}\rangle}{\partial t} = -2(\mu_a+\mu_v)\langle\mathcal{E}\rangle \tag{S39}$$

$$\frac{\partial\langle\hat{\mathcal{E}}\rangle}{\partial t} = -2\mu_a\langle\hat{\mathcal{E}}\rangle \tag{S40}$$

The transformations from $Y^{(1)}_{A,r}$ to $M_{A,r}$, from $Y^{(2)}_{A,r}$ to $\hat{M}_{A,r}$ and from $Y^{(1)}_{V,r}$ to $M_{V,r}$ result in,

$$\begin{aligned}\frac{\partial\langle M_{A,r}\rangle}{\partial t} &= \mu_a \ell \sum_{i=0}^{r-2} 2^{r-i}\mathcal{K}_{r-i}\binom{r}{i}\langle M_{A,i}\rangle + \frac{\binom{r}{2}\mathcal{K}_2\,\mathcal{K}_{r-2}\langle M_{A,2}M_{A,r-2}\rangle - r\langle M_{A,r}\rangle}{N_a} - 2r(\mu_a+\mu_v)\langle M_{A,r}\rangle \\ &-\mu_a \sum_{i=0}^{r-2} 2^{r-i-1}\binom{r}{i}\left[\mathcal{K}_{r-i-1}\langle M_{A,i+1}\rangle + \frac{\mathcal{K}_{r-i}}{\mathcal{K}_1}\langle\mathcal{E}M_{A,i}\rangle\right]\left[1+(-1)^{r-i+1}\right]\end{aligned} \tag{S41}$$

$$\begin{aligned}\frac{\partial\langle \hat{M}_{A,r}\rangle}{\partial t} &= \mu_a \hat{\ell} \sum_{i=0}^{r-2} 2^{r-i}\mathcal{K}_{r-i}\binom{r}{i}\langle \hat{M}_{A,i}\rangle + \frac{\binom{r}{2}\mathcal{K}_2\,\mathcal{K}_{r-2}\langle \hat{M}_{A,2}\hat{M}_{A,r-2}\rangle - r\langle \hat{M}_{A,r}\rangle}{N_a} - 2r\mu_a\langle \hat{M}_{A,r}\rangle \\ &-\mu_a \sum_{i=0}^{r-2} 2^{r-i-1}\binom{r}{i}\left[\mathcal{K}_{r-i-1}\langle \hat{M}_{A,i+1}\rangle + \frac{\mathcal{K}_{r-i}}{\mathcal{K}_1}\langle\hat{\mathcal{E}}\hat{M}_{A,i}\rangle\right]\left[1+(-1)^{r-i+1}\right]\end{aligned} \tag{S42}$$

$$\begin{aligned}\frac{\partial\langle M_{V,r}\rangle}{\partial t} &= \mu_v \ell \sum_{i=0}^{r-2} 2^{r-i}\mathcal{K}_{r-i}\binom{r}{i}\langle M_{V,i}\rangle + \frac{\binom{r}{2}\mathcal{K}_2\,\mathcal{K}_{r-2}\langle M_{V,2}M_{V,r-2}\rangle - r\langle M_{V,r}\rangle}{N_v} - 2r(\mu_v+\mu_a)\langle M_{V,r}\rangle \\ &-\mu_v \sum_{i=0}^{r-2} 2^{r-i-1}\binom{r}{i}\left[\mathcal{K}_{r-i-1}\langle M_{V,i+1}\rangle + \frac{\mathcal{K}_{r-i}}{\mathcal{K}_1}\langle\mathcal{E}M_{V,i}\rangle\right]\left[1+(-1)^{r-i+1}\right]\end{aligned} \tag{S43}$$

Selection. We assume that fitness of an antibody is proportional to the logarithm of its activation probability given by equation (S30) based on its average interaction strength,

$$f_{A^\alpha} \equiv f_A(\mathbf{A}^\alpha;\{V\}) = c_a \log[\rho_A(\mathbf{A}^\alpha)] = -c_a\log(1+\exp[-\beta_0(E_{\alpha\,.}+\hat{E}_{\alpha\,.}-E^*)]) \tag{S44}$$

$$\simeq f_A^* + S_a(E_{\alpha\,.}+\hat{E}_{\alpha\,.}) \tag{S45}$$

with $f_A^* = -c_a\log\left(1+\exp[\beta_0 E^*]\right)$ and the selection coefficient $S_a = c_a\beta_0/(1+\exp[-\beta_0 E^*])$. The approximation in (S45) is by expansion of the nonlinear fitness function around the neutral binding affinity, $\mathcal{E}=0$. The antibody selection coefficient S_a can be thought as the amount of stimulation that a bound antibody experiences, e.g. due to helper T-cells. If the chronic infection is HIV, where the virus attacks the helper T-cells, S_a may decrease as HIV progresses and the T-cell count decays. Furthermore, f_A^* affects the absolute growth rate, but does not affect the relative growth rate between genotypes. We call the fitness models based on the averaged binding affinity in eq. (S44) as "*nonlinear-averaged*" and in eq. (S45) as "*linear-averaged*". In Section 2.5 we introduce an alternative model of antibody activation, which assumes that proliferation of an antibody is related to its *best binding affinity* against $R\le N_v$ antigens, that are presented to the antibody during its life time. We should note that the analytical results in this paper are all based on the antibody evolution in "*linear-averaged*" fitness landscapes (S45), and the other fitness models are only studied numerically.

The viral fitness is related to the probability that it escapes the binding interactions with antibodies. We define the fitness of an antigen (virus) as the negative logarithm of its binding probability to the average antibodies that it interacts with in eq. (S31),

$$f_{V^\gamma} \equiv f_V(\mathbf{V}^\gamma;\{A\}) = -c_v \log[\rho_V(\mathbf{V}^\gamma)] = c_v\log(1+\exp[-\beta_0(E_{.\,\gamma}+\hat{E}_{.\,\gamma}-E^*)]) \tag{S46}$$

$$\simeq f_V^* - S_v(E_{.\,\gamma}+\hat{E}_{.\,\gamma}) \tag{S47}$$

with $f_V^* = c_v\log\left(1+\exp[\beta_0 E^*]\right)$ and the selection coefficient $S_v = c_v\beta_0/(1+\exp[-\beta_0 E^*])$.

As shown in equations (S6, S7) the change in the frequency of an antibody or a virus is proportional to its fitness, which is related to its average binding affinity. Therefore, the change of a given phenotype statistic $U(\mathbf{x},\mathbf{y})$ due to selection follows,

$$\frac{d}{dt}U(\mathbf{x},\mathbf{y}) = \sum_{\alpha,\gamma}\left[\frac{\partial U}{\partial x^{\alpha}}(f_{A^{\alpha}} - F_{A})\,x^{\alpha} + \frac{\partial U}{\partial y^{\gamma}}(f_{V^{\gamma}} - F_{V})\,y^{\gamma}\right] \tag{S48}$$

where F_A and F_V are respectively the mean fitness in the antibody and in the viral population. With this formulation we can compute the effect of selection on the statistics of the binding affinity distribution, i.e., the mean affinity $\mathcal{E}$, $\hat{\mathcal{E}}$, and the central moments, $M_{A,r}$, $\hat{M}_{A,r}$ and $M_{V,r}$, which we present in the following section.

Similar to the rescaling procedure in eq. (S28), we use the total scale of the traits to define the rescaled strength of selection,

$$s_a = N_a S_a E_0, \qquad \hat{s}_a = N_a S_a \hat{E}_0, \qquad s_v = N_a S_v E_0, \qquad \hat{s}_v = 0 \tag{S49}$$

Genetic drift. We can project the stochasticity of the genotype space onto the phenotype space. The projected diffusion coefficients show correlation between the noise levels of the phenotypic statistics A and B.

$$\mathcal{G}^{AB} = \frac{1}{N_a}\sum_{\alpha,\beta}\frac{\partial A}{\partial x^{\alpha}}\,\frac{\partial B}{\partial x^{\beta}}\,g^{\alpha\beta} + \frac{1}{N_v}\sum_{\gamma,\lambda}\frac{\partial A}{\partial y^{\gamma}}\,\frac{\partial B}{\partial y^{\lambda}}\,h^{\gamma\lambda} \tag{S50}$$

and the genotypic diffusion constants $g^{\alpha\beta}$ and $h^{\gamma\lambda}$ are given by eq. (S12). As an example, we compute the diffusion term for $\mathcal{E}$,

$$\begin{aligned}
\mathcal{G}^{\mathcal{E}\mathcal{E}} &= \frac{1}{N_a}\sum_{\alpha,\beta}\frac{\partial \mathcal{E}}{\partial x^{\alpha}}\,\frac{\partial \mathcal{E}}{\partial x^{\beta}}\,g^{\alpha\beta} + \frac{1}{N_v}\sum_{\gamma,\lambda}\frac{\partial \mathcal{E}}{\partial y^{\gamma}}\,\frac{\partial \mathcal{E}}{\partial y^{\lambda}}\,h^{\gamma\lambda} \\
&= \frac{1}{N_a}\sum_{\alpha,\beta}E_{\alpha\,.}E_{\beta\,.}\Big[-x^{\alpha}x^{\beta}(1-\delta_{\alpha}^{\beta}) + x^{\alpha}(1-x^{\alpha})\delta_{\alpha}^{\beta}\Big] \\
&\quad + \frac{1}{N_v}\sum_{\gamma,\lambda}E_{.\,\gamma}E_{.\,\lambda}\Big[-y^{\gamma}y^{\lambda}(1-\delta_{\gamma}^{\lambda}) + y^{\gamma}(1-y^{\gamma})\delta_{\gamma}^{\lambda}\Big] \\
&= \frac{1}{N_a}\Big[\sum_{\alpha}(E_{\alpha\,.}-\mathcal{E})^2 x^{\alpha}\Big] + \frac{1}{N_v}\Big[\sum_{\gamma}(E_{.\,\gamma}-\mathcal{E})^2 y^{\gamma}\Big] \\
&= \frac{1}{N_a}M_{A,2} + \frac{1}{N_v}M_{V,2}
\end{aligned} \tag{S51}$$

A similar approach finds the diffusion term for the second moments and the cross-correlation terms between the first and the second moments (see e.g., [36] for further details),

$$\mathcal{G}^{M_{A,2},M_{A,2}} = \frac{1}{N_a}(M_{A,4} - M_{A,2}^2), \qquad \mathcal{G}^{M_{V,2},M_{V,2}} = \frac{1}{N_v}(M_{V,4} - M_{V,2}^2) \tag{S52}$$

$$\mathcal{G}^{\mathcal{E},M_{A,2}} = \frac{1}{N_a}M_{A,3}, \qquad \mathcal{G}^{\mathcal{E},M_{V,2}} = \frac{1}{N_v}\langle M_{V,3}\rangle \tag{S53}$$

*2.3 Stochastic evolution of molecular traits (*linear-averaged *fitness)*

Putting all the evolutionary forces together, we can write down evolution equations for the statistics of binding affinities in a linear fitness landscape introduced in equations (S45, S47),

$$\frac{d}{dt}\mathcal{E} = -2(\mu_v + \mu_a)\mathcal{E} + S_a M_{A,2} - S_v\,M_{V,2} + \chi_{\mathcal{E}} \tag{S54}$$

$$\frac{d}{dt}\hat{\mathcal{E}} = S_a\,\hat{M}_{A,2} - 2\mu_a\,\hat{\mathcal{E}} + \chi_{\hat{\mathcal{E}}} \tag{S55}$$

with the Gaussian correlated noise statistics due to the genetic drift,

$$\langle \chi_{\mathcal{E}} \rangle = 0, \qquad \langle \chi_{\mathcal{E}}(t)\chi_{\mathcal{E}}(t')\rangle = \left[\frac{M_{A,2}}{N_a} + \frac{M_{V,2}}{N_v}\right]\delta(t-t') \tag{S56}$$

$$\langle \chi_{\hat{\mathcal{E}}} \rangle = 0, \qquad \langle \chi_{\hat{\mathcal{E}}}(t)\chi_{\hat{\mathcal{E}}}(t')\rangle = \left[\frac{\hat{M}_{A,2}}{N_a}\right]\delta(t-t') \tag{S57}$$

It should be noted that we ignore the linkage correlations between the binding affinity of the variable region $\mathcal{E}$ and conserved region $\hat{\mathcal{E}}$ of the virus. From the numerical analysis we see that the covariance between the linked variable and conserved regions, $\langle [(\mathcal{E}_{\alpha\,.} - \mathcal{E})\,(\hat{\mathcal{E}}_{\alpha\,.} - \hat{\mathcal{E}})]_A \rangle$ is small compared to the diversity of the binding affinity $\langle M_{A,2}\rangle$ and $\langle \hat{M}_{A,2}\rangle$ in each part; Fig. S2D. Lineages with access to the conserved region of the virus adapt by aligning their sites to the conserved sequence, and hence remain relatively conserved with variations arising only from the stochastic forces of mutation and genetic drift. In Section 2.4 we explicitly show that the auto-correlation time for the binding affinity in the conserved region is longer than in the variable interaction region; see equations (S89, S88). Therefore, the correlation between the binding affinity of the variable and the conserved regions remains small, throughout the evolutionary process.

We can write down the stochastic evolution equations for the second moments $M_{A,2}$, $\hat{M}_{A,2}$ and $M_{V,2}$,

$$\frac{d}{dt}M_{A,2} = -4\mu_a(M_{A,2} - \ell\mathcal{K}_2) - 4\mu_v M_{A,2} - \frac{M_{A,2}}{N_a} + S_a M_{A,3} + \chi_{M_{A,2}} \tag{S58}$$

$$\frac{d}{dt}\hat{M}_{A,2} = -4\mu_a(\hat{M}_{A,2} - \hat{\ell}\hat{\mathcal{K}}_2) - \frac{\hat{M}_{A,2}}{N_a} + S_a \hat{M}_{A,3} + \chi_{\hat{M}_{A,2}} \tag{S59}$$

$$\frac{d}{dt}M_{V,2} = -4\mu_v(M_{V,2} - \ell\mathcal{K}_2) - 4\mu_a M_{V,2} - \frac{M_{V,2}}{N_v} - S_v M_{V,3} + \chi_{M_{V,2}} \tag{S60}$$

with Gaussian correlated noise statistics,

$$\langle \chi_{M_{A,2}} \rangle = 0, \qquad \langle \chi_{M_{A,2}}(t)\chi_{M_{A,2}}(t')\rangle = \left[\frac{M_{A,4} - (M_{A,2})^2}{N_a}\right]\delta(t-t') \tag{S61}$$

$$\langle \chi_{\hat{M}_{A,2}} \rangle = 0, \qquad \langle \chi_{\hat{M}_{A,2}}(t)\chi_{\hat{M}_{A,2}}(t')\rangle = \left[\frac{\hat{M}_{A,4} - (\hat{M}_{A,2})^2}{N_a}\right]\delta(t-t') \tag{S62}$$

$$\langle \chi_{M_{V,2}} \rangle 0, \qquad \langle \chi_{M_{V,2}}(t)\chi_{M_{V,2}}(t')\rangle = \left[\frac{M_{V,4} - (M_{V,2})^2}{N_v}\right]\delta(t-t') \tag{S63}$$

$$\langle \chi_{M_{A,2}}(t)\chi_{\mathcal{E}}(t')\rangle = \frac{M_{A,3}}{N_a}\,\delta(t-t'), \qquad \langle \chi_{\hat{M}_{A,2}}(t)\chi_{\hat{\mathcal{E}}}(t')\rangle = \frac{\hat{M}_{A,3}}{N_a}\,\delta(t-t') \tag{S64}$$

$$\langle \chi_{M_{V,2}}(t)\chi_{\mathcal{E}}(t')\rangle = \frac{\langle M_{V,3}\rangle}{N_v}\,\delta(t-t') \tag{S65}$$

Stationary solutions for trait mean and diversity. From equations above we can solve for the stationary moments of the mean binding affinity and its diversity in both populations, and the cross-correlations between them as a function of the higher moments,

$$\langle \mathcal{E} \rangle = \frac{1}{2\tilde{\theta}_a} N_a S_a \langle M_{A,2}\rangle - \frac{1}{2\tilde{\theta}_v} N_v S_v \langle M_{V,2}\rangle \tag{S66}$$

$$\langle \mathcal{E}, \mathcal{E} \rangle = \frac{1}{4\tilde{\theta}_a}\left[\langle M_{A,2}\rangle + 2N_a S_a \langle \mathcal{E}, M_{A,2}\rangle\right] + \frac{1}{4\tilde{\theta}_v}\left[\langle M_{V,2}\rangle - 2N_v S_v \langle \mathcal{E}, M_{V,2}\rangle\right] \tag{S67}$$

$$\langle \hat{\mathcal{E}} \rangle = N_a S_a \langle \hat{M}_{A,2}\rangle / 2\theta_a \tag{S68}$$

$$\langle \hat{\mathcal{E}}, \hat{\mathcal{E}} \rangle = \frac{1}{4\theta_a}\left[\langle \hat{M}_{A,2}\rangle + 2N_a S_a \langle \hat{\mathcal{E}}, \hat{M}_{A,2}\rangle\right] \tag{S69}$$

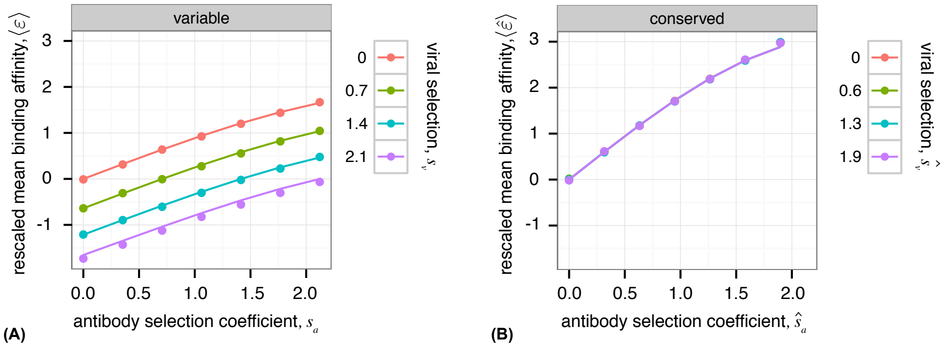

FIG. S1: **Effect of selection on the mean binding affinity.** The mean binding affinity for **(A)** the variable interaction region $\varepsilon = \mathcal{E}/E_0$, and **(B)** the conserved region $\hat{\varepsilon} = \hat{\mathcal{E}}/E_0$, as a function of selection coefficients. Stationary mean binding affinity is sensitive to selection on antibodies in both variable and conserved regions. The conserved region is not sensitive to viral selection strength. Points indicate simulation results, and solid lines indicate the stationary solution in eqs. (S66, S68). Parameters are: $\kappa_i = \hat{\kappa}_i = 1$ for all sites, $\ell = 50$, $\hat{\ell} = 40$, $N_a = N_v = 1000$, $\theta_a = \theta_v = 1/90$. Points are time averaged values from simulations run for 10^7 generations, with values sampled every N_a generations, and data from first $100N_a$ generations discarded, and then ensemble-averaged over 100 replicate simulations.

$$\left\langle M_{A,2}\right\rangle = \frac{1}{1+4\tilde{\theta}_a}\left[4\ell\mathcal{K}_2\theta_a + \left(N_a S_a\right)\left\langle M_{A,3}\right\rangle\right] \tag{S70}$$

$$\left\langle M_{V,2}\right\rangle = \frac{1}{1+4\tilde{\theta}_v}\left[4\ell\mathcal{K}_2\theta_v - \left(N_v S_v\right)\left\langle M_{V,3}\right\rangle\right] \tag{S71}$$

$$\langle\mathcal{E}, M_{A,2}\rangle = \frac{1}{1+6\tilde{\theta}_a}\left[\langle M_{A,3}\rangle + N_a S_a\left[\langle\mathcal{E}, M_{A,3}\rangle + \langle(M_{A,2})^2\rangle\right]\right) \tag{S72}$$

$$\langle\mathcal{E}, M_{V,2}\rangle = \frac{1}{1+6\tilde{\theta}_v}\left(\langle M_{V,3}\rangle - N_v S_v\left[\langle\mathcal{E}, M_{V,3}\rangle + \langle(M_{V,2})^2\rangle\right]\right) \tag{S73}$$

$$\langle\mathcal{E}, M_{A,3}\rangle = \frac{\langle M_{A,4}\rangle/3 - \langle(M_{A,2})^2\rangle}{1+8/3\tilde{\theta}_a}, \qquad \langle\mathcal{E}, M_{V,3}\rangle = \frac{\langle M_{V,4}\rangle/3 - \langle(M_{V,2})^2\rangle}{1+8/3\tilde{\theta}_v} \tag{S74}$$

where $\tilde{\theta}_a = \theta_a + (N_a/N_v)\,\theta_v$ and $\tilde{\theta}_v = \theta_v + (N_v/N_a)\,\theta_a$. We denote the ensemble covariance of two stochastic observables x and y by,

$$\left\langle x, y\right\rangle \equiv \left\langle\left(x - \langle x\rangle\right)\left(y - \langle y\rangle\right)\right\rangle \tag{S75}$$

and hence, $\langle x, x\rangle$ indicates the ensemble variance of the variable x. Similar forms of the stationary solutions apply for the statistics of the binding affinity in the conserved interaction region, $\hat{M}_{A,2}$, $\langle\hat{\mathcal{E}}, \hat{M}_{A,2}\rangle$, and can be found by setting the viral mutation μ_v rate and selection coefficient S_v equal to zero. For brevity we do not present the solutions of the central moments in the conserved region.

In equations (S54-S65), the evolution of each moment depends on the higher moments in the presence of selection, which leads to an infinite moment hierarchy. However, in the regime where rescaled coefficients satisfy $s_a\theta_a < 1$ and $s_v\theta_v < 1$, we can truncate the moment hierarchy. From the comparisons of the Wright-Fisher simulations with our theoretical results we choose to truncate the hierarchy after the 4^{th} moment. Furthermore, higher central moments are fast stochastic variables (see e.g., [36] and the discussion in Section 2.4 and Fig. S3), and their ensemble averages can sufficiently characterize the evolution of the trait mean $\mathcal{E}$ and the trait diversity $M_{A,2}$, $\hat{M}_{A,2}$ and $M_{V,2}$. Therefore, we will only present ensemble-averaged equations for the 3^{rd} and 4^{th} moments of the trait distributions. In order to clarify the truncation of the moment hierarchy, we explicitly show the evolution equations and stationary solutions of the rescaled moments, which are defined in eq. (S28).

$$\frac{d}{d\tau_a}\langle m_{A,3}\rangle = -6\theta_a\langle m_{A,3}\rangle - 8\theta_a\left(\frac{\mathcal{K}_3}{E_0^2\mathcal{K}_1}\langle\varepsilon\rangle\right) - 6\theta_v(N_a/N_v)\,\langle m_{A,3}\rangle - 3\langle m_{A,3}\rangle + s_a\left[\langle m_{A,4}\rangle - 3\langle(m_{A,2})^2\rangle\right] \tag{S76}$$

$$\frac{d}{d\tau_v}\langle m_{V,3}\rangle = -6\theta_v\langle m_{V,3}\rangle - 8\theta_v\left(\frac{\mathcal{K}_3}{E_0^2\mathcal{K}_1}\langle\varepsilon\rangle\right) - 6\theta_a(N_a/N_v)\,\langle m_{V,3}\rangle - 3\langle m_{V,3}\rangle - s_v\left[\langle m_{V,4}\rangle - 3\langle(m_{V,2})^2\rangle\right] \tag{S77}$$

$$\frac{d}{d\tau_a}\langle(m_{A,2})^2\rangle = -8\theta_a\left[\langle(m_{A,2})^2\rangle - \langle m_{A,2}\rangle\right] - 8\theta_v\,(N_v/N_a)\langle(m_{A,2})^2\rangle + \langle m_{A,4}\rangle - 3\langle(m_{A,2})^2\rangle \tag{S78}$$

$$\frac{d}{d\tau_v}\langle(m_{V,2})^2\rangle = -8\theta_v\left[\langle(m_{V,2})^2\rangle - \langle m_{V,2}\rangle\right] - 8\theta_a(N_v/N_a)\,\langle(m_{V,2})^2\rangle + \langle m_{V,4}\rangle - 3\langle(m_{V,2})^2\rangle \tag{S79}$$

$$\frac{d}{d\tau_a}\langle m_{A,4}\rangle = -8\theta_a\left[\langle m_{A,4}\rangle - 2\frac{\mathcal{K}_4}{\ell\,\mathcal{K}_2^2} - (3-4/\ell)\,\langle m_{A,2}\rangle\right] - 8\theta_v(N_a/N_v)\,\langle m_{A,4}\rangle + 6\langle(m_{A,2})^2\rangle - 4\langle m_{A,4}\rangle \tag{S80}$$

$$\frac{d}{d\tau_v}\langle m_{V,4}\rangle = -8\theta_v\left[\langle m_{V,4}\rangle - 2\frac{\mathcal{K}_4}{\ell\,\mathcal{K}_2^2} - (3-4/\ell)\,\langle m_{V,2}\rangle\right] - 8\theta_a(N_v/N_a)\,\langle m_{V,4}\rangle + 6\langle(m_{V,2})^2\rangle - 4\langle m_{V,4}\rangle \tag{S81}$$

where τ_a and τ_v are respectively time measured in units of the neutral coalescence in the antibodies, N_a and in the viral population, N_v. The term $\langle\varepsilon\rangle = 4(s_a\theta_a - s_v\theta_v\,(N_a/N_v))\big/(\theta_a + \theta_v\,(N_a/N_v))$ in equations (S76, S77) is the stationary solution for the rescaled mean binding affinity up to orders of $\mathcal{O}(\theta_a^2, \theta_v^2)$. The stationary solutions for the rescaled central moments of the antibody population follow,

$$\big\langle m_{A,2}\big\rangle = \frac{4\theta_a}{1+4\tilde{\theta}_a} - \frac{8\theta_a}{3+18\tilde{\theta}_a}s_a\left[\frac{\mathcal{K}_3}{E_0^2\mathcal{K}_1}\langle\varepsilon\rangle - 4s_a\theta_a^2 + \mathcal{O}\big(\theta_a^3\big)\right] + \mathcal{O}(s_a^2\theta^2) \tag{S82}$$

$$\big\langle m_{A,3}\big\rangle = -\frac{8}{3}\,\frac{\theta_a}{1+2\tilde{\theta}}\left(\frac{\mathcal{K}_3}{E_0^2\mathcal{K}_1}\langle\varepsilon\rangle\right) + \frac{32}{3}s_a\left[\theta_a^2 + \mathcal{O}(\theta_a^3)\right] + \mathcal{O}(s_a^2\theta^3) \tag{S83}$$

$$\big\langle(m_{A,2})^2\big\rangle = \frac{8\theta_a}{3+28\,\tilde{\theta}_a}\left[\frac{1}{\ell}\,\frac{\mathcal{K}_4}{\mathcal{K}_2^2} + 2\theta_a(7-4/\ell)\right] + \mathcal{O}(s_a\theta_a^3) \tag{S84}$$

$$\langle m_{A,4}\rangle = \frac{24\theta_a}{3+28\,\tilde{\theta}_a}\left[\frac{1}{\ell}\,\frac{\mathcal{K}_4}{\mathcal{K}_2^2} + 2\theta_a(5-4/\ell)\right] + \mathcal{O}(s_a\theta_a^3) \tag{S85}$$

with $\tilde{\theta}_a = \theta_a + \theta_v(N_a/N_v)$. Similar solutions can be found for the moments of the variable sequence region in the viral population $m_{V,r}$ by replacing the subscripts a and v in the equations above. The stationary solutions for the central moments of the antibody population in the conserved interaction region, $\hat{m}_{A,r}$ can be found by setting the viral mutation rate and selection coefficient equal to zero, i.e., $\theta_v = 0$ and $s_v = 0$, and by using the characteristics of the conserved region i.e., genetic length $\hat{\ell}$ and sites contributions $\hat{\mathcal{K}}_r$ in the equation above. Fig. S1 shows a good agreement between the numerical results for the rescaled stationary mean binding affinity $\langle\varepsilon\rangle = \langle\mathcal{E}\rangle/E_0$, $\langle\hat{\varepsilon}\rangle = \langle\hat{\mathcal{E}}\rangle/\hat{E}_0$ from the Wright-Fisher simulations and the analytical solutions (S66, S68), by using the stationary ensemble averages for the diversity of the binding affinity $\langle m_{A,2}\rangle$, $\langle\hat{m}_{A,2}\rangle$ and $\langle m_{V,2}\rangle$ in eq. (S82). Fig. S2 compares the analytical solution for the second central moments $m_{A,2}$ and $m_{V,2}$ with numerical results from the Wright-Fisher simulations, by inserting the empirical estimates of the higher moments from the simulations in equations (S70) and (S71), (dashed lines), and by using the analytical solutions for the higher moments to estimate the stationary value for the trait diversity given by eq. (S82), (solid lines).

2.4 Time-dependent statistics

Statistics of the trait mean. As we show below, the higher central moments $M_{V,r}$ and $M_{A,r}$ for $(r > 1)$ are fast stochastic variables. Therefore, it is sufficient to use their stationary ensemble averages to compute the finite time correlation of the mean interaction variables, $\mathcal{E}(\tau)$ and $\hat{\mathcal{E}}(\tau)$.

The time dependent solution of the trait mean $\mathcal{E}(\tau)$ and $\hat{\mathcal{E}}(\tau)$ and the covariance between two time points $\tau_2 \geq$

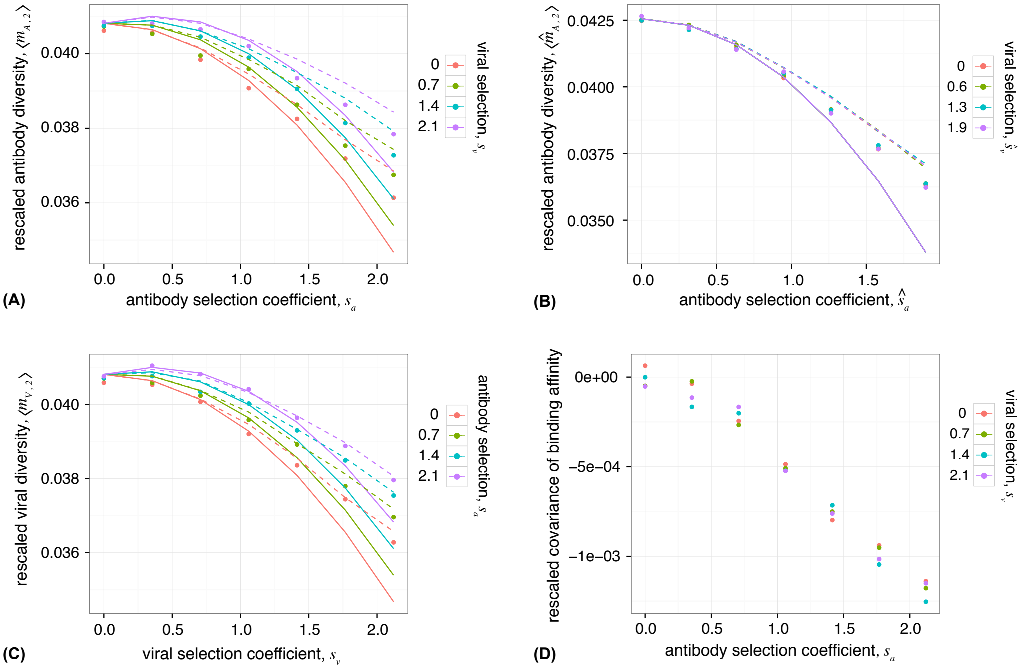

FIG. S2: **Effect of selection on the diversity of binding affinity in antibodies and viruses**. Stationary diversity of the binding affinity for **(A)** the variable interaction region $m_{A,2} = M_{A,2}/E_0^2$, **(B)** the conserved interaction region $\hat{m}_{A,2} = \hat{M}_{A,2}/\hat{E}_0^2$ in the antibody population, and **(C)** the variable region in the viral population $m_{V,2} = M_{V,2}/E_0^2$ plotted as a function of viral and antibody selection coefficients. The diversity of binding across the antibodies in the conserved region, **(B)** $\hat{m}_{A,2}$ is not sensitive to viral selection strength. **(D)** The magnitude of the rescaled covariance due to genetic linkage between binding of the antibody to the conserved and the variable regions, $\langle [(E_{\alpha\,.} - \mathcal{E})(\hat{E}_{\alpha\,.} - \hat{\mathcal{E}})]_A\rangle / E_0\hat{E}_0$, is much smaller than the diversity of binding in each region shown in **(A)** and **(B)**, supporting our decision to neglect the effects of genetic linkage on the evolution of binding affinity. Points indicate simulation results, described in Fig. S1, dashed lines indicate the stationary solution which depends on measured higher moments (eq. (S70), (S71)), and solid lines indicate the stationary solution (eq. (S82), and similarly for viruses). Theory lines begin to deviate from simulation results for large selection strengths $s_a, s_v > 1$. The deviations are larger in antibodies due to neglecting the linkage correlation between the variable and the conserved regions.

τ_1, starting from initial conditions at $\tau_0 = 0$ with the ensemble averages $\langle\mathcal{E}(0)\rangle$, $\langle\hat{\mathcal{E}}(0)\rangle$ and variance $\langle\mathcal{E}(0), \mathcal{E}(0)\rangle$, $\langle\hat{\mathcal{E}}(0), \hat{\mathcal{E}}(0)\rangle$ follows,

$$\langle\mathcal{E}(\tau)\rangle = (1 - \mathrm{e}^{-2\tilde{\theta}_a\tau})\,\langle\mathcal{E}\rangle + \mathrm{e}^{-2\tilde{\theta}_a\tau}\,\langle\mathcal{E}(0)\rangle \tag{S86}$$

$$\langle\hat{\mathcal{E}}(\tau)\rangle = (1 - \mathrm{e}^{-2\theta_a\tau})\,\langle\hat{\mathcal{E}}\rangle + \mathrm{e}^{-2\theta_a\tau}\,\langle\hat{\mathcal{E}}(0)\rangle \tag{S87}$$

$$\begin{aligned}\langle\mathcal{E}(\tau_1), \mathcal{E}(\tau_2)\rangle &= \mathrm{e}^{-2\tilde{\theta}_a\tau_2}\langle\mathcal{E}(0), \mathcal{E}(0)\rangle + \left[\frac{\langle M_{A,2}\rangle}{N_a} + \frac{\langle M_{V,2}\rangle}{N_v}\right]\int_0^{\tau_1} \mathrm{e}^{-2\tilde{\theta}_a(\tau_1-\tau')}\mathrm{e}^{-2\tilde{\theta}_a(\tau_2-\tau')}d\tau' \\ &= \mathrm{e}^{-2\tilde{\theta}_a\tau_2}\,\langle\mathcal{E}(0), \mathcal{E}(0)\rangle + \left[\frac{\langle M_{A,2}\rangle}{4\tilde{\theta}_a} + \frac{\langle M_{V,2}\rangle}{4\tilde{\theta}_v}\right]\left[\mathrm{e}^{-2\tilde{\theta}_a(\tau_2-\tau_1)} - \mathrm{e}^{-2\tilde{\theta}_a(\tau_1+\tau_2)}\right]\end{aligned} \tag{S88}$$

$$\langle\hat{\mathcal{E}}(\tau_1), \hat{\mathcal{E}}(\tau_2)\rangle = \mathrm{e}^{-2\theta_a\tau_2}\,\langle\hat{\mathcal{E}}(0), \hat{\mathcal{E}}(0)\rangle + \frac{\langle\hat{M}_{A,2}\rangle}{4\theta_a}\,\left[\mathrm{e}^{-2\theta_a(\tau_2-\tau_1)} - \mathrm{e}^{-2\theta_a(\tau_1+\tau_2)}\right] \tag{S89}$$

where $\langle\mathcal{E}\rangle$ and $\langle\hat{\mathcal{E}}\rangle$ are the stationary values of the trait mean in the variable and the conserved interaction regions, given by equations (S66, S68). Time τ is measured in units of the neutral coalescence time for antibodies, N_a. The characteristic time-scale for the decay of the mean binding affinity in the variable interaction region of the virus is $1/2\tilde{\theta}_a = 1/(2(\theta_a + (N_a/N_v)\theta_v))$ in units of N_a, which is shorter than the time-scale for the conserved region, $1/2\theta_a$. Therefore, binding affinity in the conserved region is correlated over a longer period of time compared to the variable

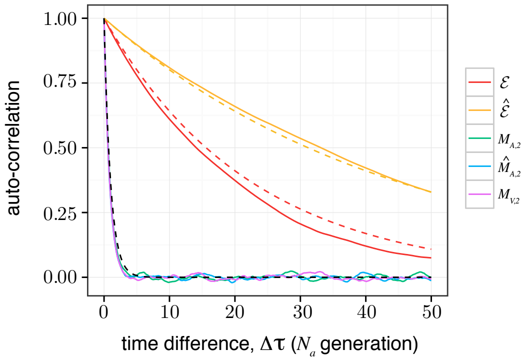

FIG. S3: **Time-dependent statistics.** Auto-correlation of the stationary mean binding affinity in the variable region (red), eq. (S88), has a shorter decay time than the conserved interaction region (orange), eq. (S89). The decay time for the auto-correlation of the trait mean in both variable and conserved regions, which is of order of the inverse mutation rate, is much longer than the second moments (green, blue, purple), which decay on a timescale of N generations. Solid lines are from stationary simulations, and dashed lines are eq. (S88) (orange), eq. S89 (red) and eq. S90 (black), normalized to be correlations. Parameters are: all $\kappa_i = \hat{\kappa}_i = 1$, $\ell = 50$, $\hat{\ell} = 40$, $N_a = N_v = 1000$, $\theta_a = \theta_v = 1/90$, $s_a = s_v = 0.7$, $\hat{s}_a = 0.6$. Simulation results were time-averaged over $10^4 N_a$ generations, with values sampled every 100 generations, and first N_a/θ_a generations omitted.

region (i.e., about twice as long if $\theta_a \sim \theta_v$). The difference in time-scale explains the small covariance due to the genetic linkage between the conserved and the variable region of the virus shown in Fig. S3.

Statistics of the trait diversity. As shown in [36], the fluctuations in the trait diversity are scale invariant, which is a consequence of coherent, genome-wide linkage disequilibrium fluctuations in the absence of recombination. It is generated by sampling from a set of genotypes with binding affinities $E_{\alpha\,\cdot}$ in antibodies and $E_{\cdot\,\gamma}$ in viruses from the underlying distributions with variance $M_{A,2}$ and $M_{V,2}$, which scale like the genome length ℓ. These large fluctuations result in a relatively short correlation time for the trait diversity, shown in Fig. S3. Similar to the mean binding affinity, we can estimate the typical lifetime of these fluctuations from the stationary auto-correlation function,

$$\langle M_{A,2}(\tau_a), M_{A,2}(\tau'_a)\rangle \sim \mathrm{e}^{-(\tau_a-\tau'_a)}, \qquad \langle M_{V,2}(\tau_v), M_{V,2}(\tau'_v)\rangle \sim \mathrm{e}^{-(\tau_v-\tau'_v)} \tag{S90}$$

where τ_a, τ'_a are measured in units of the antibody neutral coalescence time N_a, and τ_v, τ'_v are measured in units of the viral neutral coalescence time N_v. Fig. S3 shows the decay of the stationary auto-correlation for the diversity of the binding affinity in antibodies $M_{A,2}$ as a function of the evolutionary time τ. It is evident that the characteristic decay time for the trait diversity (S90) is much shorter than that of the trait mean, given by the auto-correlation function in equations (S88), (S89).

2.5 Alternative fitness models

Nonlinear activation probability based on average binding (nonlinear-averaged). We assume that the growth rate (fitness) of an antibody is proportional to the logarithm of its activation probability (eq. (S44)), which may be approximated by a linear function if the non-linearity is small (S45). Here, we numerically study the effect of nonlinear fitness by comparing the evolutionary dynamics of populations in fitness landscapes with different values of nonlinearity $\beta = \beta_0 E_0$ and binding threshold $e^* = E^*/E_0$, while keeping the overall strength of (rescaled) selection, $s_a = c_a\beta/(1+\exp[-\beta e^*])$ constant. The strength of selection corresponds to the slope of the approximate *linear-averaged* fitness function in eq. (S45).

As the rescaled nonlinearity $\beta = \beta_0 E_0$ of the fitness landscape (S44) increases, the mean binding affinity $\mathcal{E}$ becomes closer to the neutral value; see Fig. S4A. This is a result of the sigmoid form of the fitness function, which reduces fitness differences between genotypes at extreme values of binding affinity. Since mutations push the mean binding affinity towards zero, the reduced advantage of binding at the extremes moves the stationary binding affinity towards zero.

Similar arguments suggest that the rate of adaptation in the antibody population should decrease as the fitness landscapes become more non-linear. The rate of adaptation is determined by fitness flux [42, 43], and is approximately

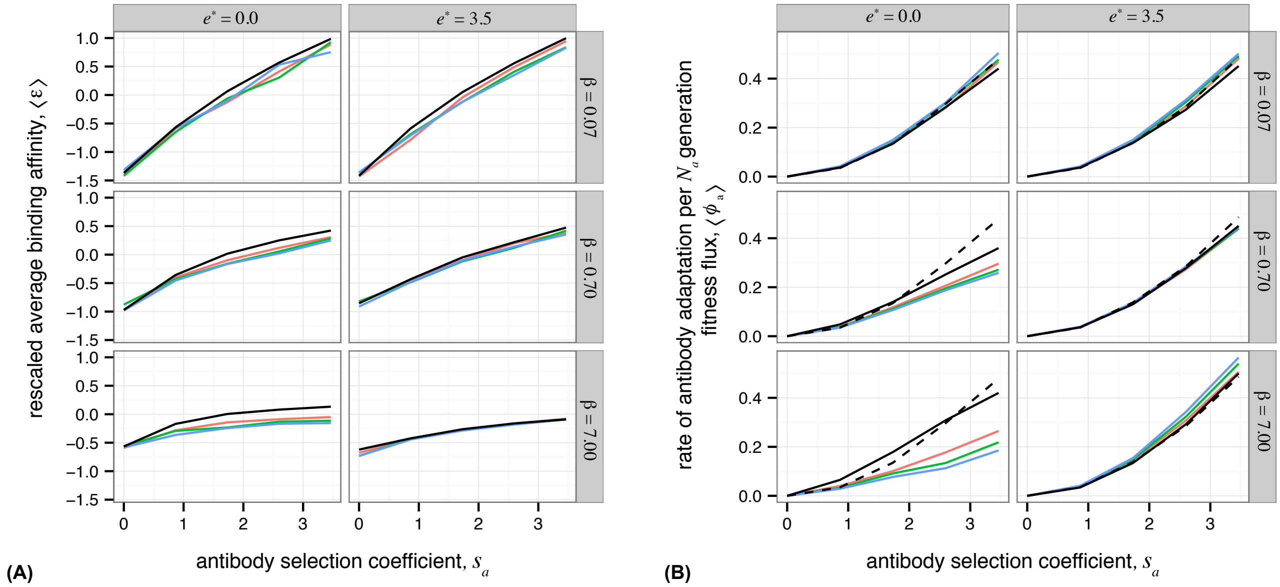

FIG. S4: **Alternative fitness models. (A)** Stationary mean binding affinity and **(B)** rate of antibody adaptation (fitness flux) due to selection, estimated by population fitness variance, for nonlinear- averaged fitness model (black) and the nonlinear-EVD fitness model with the number of interactions, $R = 10$ (red), $R = 100$ (green), and $R = 1000$ (blue). The mean binding affinity is sensitive to the degree of non-linearity β, and binding threshold e^*, but it is not very sensitive to the number of interactions R. The selection coefficient s_a is defined as in eq. S45. Dashed line in **(B)** indicates the expected fitness variance for a linear-averaged fitness model, $\langle \phi_A \rangle \simeq s_a^2 \langle m_{A,2} \rangle$, which is the selection component of the fitness flux in eq. (S94). Parameters are: $\kappa_i = \hat{\kappa}_i = 1$ for all sites, $\ell = 75$, $\hat{\ell} = 0$, $N_a = N_v = 1000$, $\theta_a = \theta_v = 1/75$. The unscaled selection coefficients are, $N_v S_v = s_v/E_0 = 0.2$, and $N_a S_a = s_a/E_0$ from 0 to 0.4 on the $x-$axis, with $\beta_0 = .01, .1, 1$ from top to bottom panels, and $E^* = 0, 25$ in the left and in the right panels. Points are time averaged values from simulations run for $10^4 N_a$ generations, with values sampled every $2N_a$ generations, and data from first $100 N_a$ generations discarded.

equal to the variance of fitness in the population [41]; see Section 3 for detailed discussion. Due to the sigmoidal shape of the fitness function, fitness differences become small at large values of binding affinity (i.e., the functional antibodies), which reduces the population fitness variance, and hence, the rate of adaptation. However, this effect is less pronounced when the threshold for specific interaction is very large, $e^* \gg 1/\beta$. In this case, the fitness function is nearly linear for most antibodies, since their binding affinity fall below the binding threshold e^*. Therefore, fitness variance and the rate of adaptation are only sensitive to the selection strength s_a (i.e., slope of fitness at $e = 0$), but not the nonlinearity of the fitness landscape. Evidently, the fitness variance (Fig. S4B) is less sensitive to the non-linearity, than the mean binding affinity (Fig. S4A).

Nonlinear activation probability based on the strongest binding (nonlinear-EVD). We also study a model for activation of antibodies which is based on their *strongest binding affinity* with a subset of viruses. The basic assumption is that an antibody attempts to bind to a set of viruses (which may be smaller than the viral population size), and once binding occurs due to a high affinity, it begins to proliferate. Similar treatments have been introduced in the context of T-cell activation [66, 67]. The probability distribution function, $\Pi(E^*_{\alpha\,.})$ of the strongest of R independent binding interactions between the antibody $\mathbf{A}^\alpha$ and the viral population $\{\mathbf{V}\}$ can be obtained by extreme value statistics. According to extreme value theory, if the distribution of binding affinities for a given antibody has an exponential tail, then the corresponding distribution for its strongest binding affinity belongs to the class of Gumbel distributions [68]. In the evolutionary regime that we study here, the amount of genetic polymorphism in the population of antibodies results in a Gaussian-like distribution for the binding affinities, with mean $E_{\alpha\,.} + \hat{E}_{\alpha\,.}$, and variance $I^{(2)}_{\alpha\,.}$ given by eq. (S17). Therefore, the corresponding probability distribution for the strongest binding affinity out of R independent trials, is a Gumbel distribution [68] with a peak at,

$$E^\alpha_{\max} = E_{\alpha\,.} + \hat{E}_{\alpha\,.} + \sqrt{2 I^{(2)}_{\alpha\,.} \ln R} \qquad \text{(S91)}$$

and a width $\Sigma^\alpha = \sqrt{\pi I^{(2)}_{\alpha\cdot}/(12\ln R)}$. If we assume that $\ln R \gg 1$, the distribution is sharply peaked, and $E^\alpha_{\max}$ is sufficient to describe it. In addition, we assume the activation probability is a sigmoid function of $E^\alpha_{\max}$,

$$\rho_{A,\max}(\mathbf{A}^\alpha) = \frac{1}{1+\exp[-\beta_0(E^\alpha_{\max} - E^*)]}. \tag{S92}$$

The fitness function $f_{A,\max}(\mathbf{A}^\alpha; \{V\})$ for the *nonlinear-EVD* model is related to the logarithm of the activation probability,

$$f_{A,\max}(\mathbf{A}^\alpha; \{V\}) \;=\; c_a \log[\rho_{\max}(\mathbf{A}^\alpha)] = -c_a \log(1+\exp[-\beta_0(E^\alpha_{\max} - E^*)]) \tag{S93}$$

where the coefficients are similarly defined as in eq. (S44). Fig. S4A shows the stationary binding affinity for the *nonlinear-EVD* fitness model. While the mean binding affinity is sensitive to the nonlinearity parameter β, it is relatively insensitive to the number of interactions R, and is similar to the *nonlinear-averaged* model. This is not surprising given the logarithmic dependence of binding affinity on the number of interactions R. As in the nonlinear-averaged model, the fitness variance may be reduced due to the smaller fitness differences at high binding. In the non-linear-EVD model (Fig. S4B, colors), the fitness variance is further reduced as the number of interactions increases, because there is a higher chance of binding with a large affinity when there are more interactions. However, when the threshold for specific interaction is very large, $e^* \gg 1/\beta$, the binding affinity of most antibodies fall below the threshold, where the fitness function is nearly linear.

3. Fitness flux and the co-evolutionary transfer flux.

The fitness flux $\phi(t)$ characterizes the adaptive response of a population by genotypic or phenotypic changes in a population [37, 42, 43, 69, 70]. The cumulative fitness flux, $\Phi(\tau) = \int_t^{t+\tau} \phi(t')dt'$, measures the total amount of adaptation over an evolutionary period τ [43, 69]. The evolutionary statistics of this quantity is specified by the fitness flux theorem [43]. The fitness flux for the antibodies $\phi_A(t)$ and the viruses $\phi_V(t)$ follow,

$$\phi_A(t) \;=\; \sum_{\alpha\in\text{antibodies}} \frac{\partial F_A(t)}{\partial x^\alpha} \times \frac{dx^\alpha(t)}{dt} \tag{S94}$$

$$\phi_V(t) \;=\; \sum_{\gamma\in\text{viruses}} \frac{\partial F_V(t)}{\partial y^\gamma} \times \frac{dy^\gamma(t)}{dt} \tag{S95}$$

where, F_A and F_V are the stationary mean fitness of the antibody and the viral populations, and t is measured in units of generations.

We introduce a new measure of interaction between co-evolving populations "*transfer flux*", which is the change in the mean fitness of a population due to the evolution of the opposing population. The transfer flux from antibodies to viruses $\mathcal{T}_{A\to V}$ and from viruses to antibodies $\mathcal{T}_{V\to A}$ follow,

$$\mathcal{T}_{A\to V}(t) \;=\; \sum_{\alpha\in\text{antibodies}} \frac{\partial F_A(t)}{\partial x^\alpha} \times \frac{dx^\alpha(t)}{dt} \tag{S96}$$

$$\mathcal{T}_{V\to A}(t) \;=\; \sum_{\gamma\in\text{viruses}} \frac{\partial F_A(t)}{\partial y^\gamma} \times \frac{dy^\gamma(t)}{dt} \tag{S97}$$

In the regime of substantial selection $s_a, s_v \gtrsim 1$, the transfer flux in antagonistically interacting populations of antibodies and viruses is always negative, implying that adaptation of one population reduces the fitness of the opposing population.

The fitness flux and transfer flux are rates of adaptation and interaction that are time-independent only in the stationary state. The total amount of adaptation for non-stationary evolution, where the fluxes change in time, can be generally measured by the cumulative fitness and transfer flux. In the stationary state, the cumulative flux values grow linearly with the evolutionary time. For co-evolution in the *linear-averaged* fitness landscape of equations (S45, S47) the stationary cumulative fitness flux over an evolutionary time for antibodies and viruses follow from a simple genotype-to-phenotype projection,

$$\begin{aligned}\langle \Phi_a(\tau_a)\rangle &= N_a \int_{t'=0}^{t} \phi_{_A}(t')dt' = \left[\left\langle \frac{\partial F_{_A}}{\partial \mathcal{E}} \frac{\partial \mathcal{E}}{\partial t}\Big|_{\{\mathbf{V}\}}\right\rangle + \left\langle \frac{\partial F_{_A}}{\partial \hat{\mathcal{E}}} \frac{\partial \hat{\mathcal{E}}}{\partial t}\Big|_{\{\mathbf{V}\}}\right\rangle\right]\tau_a \\ &= \Big[-2\theta_a s_a\big(\langle\varepsilon\rangle + \langle\hat{\varepsilon}\rangle\big) + s_a^2\left(\langle m_{A,2}\rangle + \langle \hat{m}_{A,2}\rangle\right)\Big]\tau_a = \big[-2\theta_a s_a\langle\varepsilon\rangle + s_a^2\langle m_{A,2}\rangle\big]\tau_a \qquad \text{(S98)}\end{aligned}$$

$$\begin{aligned}\langle \Phi_v(\tau_v)\rangle &= N_v \int_{t'=0}^{t} \phi_{_V}(t')dt' = \left\langle \frac{\partial F_{_V}}{\partial \mathcal{E}} \frac{\partial \mathcal{E}}{\partial t}\Big|_{\{\mathbf{A}\}}\right\rangle \tau_v \\ &= \big[2\theta_v s_v\langle\varepsilon\rangle + s_v^2\,\langle m_{V,2}\rangle\big]\tau_v \qquad \text{(S99)}\end{aligned}$$

where ε is the rescaled mean binding affinity, and $m_{A,2}\,(m_{V,2})$ and $s_a\,(s_v)$ are the rescaled diversity of the binding affinity and the selection coefficient in the antibody (viral) population, according to the rescaling procedures in equations (S28) and (S49). $\tau_a = t/N_a$ and $\tau_v = t/N_v$ are respectively the evolutionary time in natural units of the neutral coalescence time in the antibody population N_a and in the viral population N_v. The first terms in eqs. (S98, S99) are the fitness changes due to mutation, the second terms are due to selection, and the changes due to genetic drift are zero in the ensemble average for our linear model. In the regime of substantial selection $s_a, s_v \gtrsim 1$, the fitness flux in a polymorphic population asymptotically converges to the variance of the stationary fitness distribution in the population [43], which is in accordance with the rate of adaptation given by Fisher's fundamental theorem and Price's equation [38, 41]. In this regime, fitness flux is the change in the mean fitness of the population due to selection.

Similarly, the cumulative stationary transfer fluxes for co-evolution in the *linear-averaged* fitness landscape eqs. (S45, S47) follow,

$$\begin{aligned}\langle \mathbf{T}_{V\to A}(\tau_a)\rangle &= N_a \int_{t'=0}^{t} \mathcal{T}_{V\to A}(t')dt' = \left\langle \frac{\partial F_{_A}}{\partial \mathcal{E}} \frac{\partial \mathcal{E}}{\partial t}\Big|_{\{\mathbf{A}\}}\right\rangle \tau_a(N_a/N_v) \\ &= \big[-2\theta_v s_a\langle\varepsilon\rangle - s_a s_v\,\langle m_{V,2}\rangle\big]\,\tau_a(N_a/N_v) \qquad \text{(S100)}\end{aligned}$$

$$\begin{aligned}\langle \mathbf{T}_{A\to V}(\tau_v)\rangle &= N_v \int_{t'=0}^{t} \mathcal{T}_{A\to V}(t')dt' = \left\langle \frac{\partial F_{_V}}{\partial \mathcal{E}} \frac{\partial \mathcal{E}}{\partial t}\Big|_{\{\mathbf{V}\}}\right\rangle \tau_v(N_v/N_a) \\ &= \big[2\theta_a s_v\langle\varepsilon\rangle - s_v s_a\,\langle m_{A,2}\rangle\big]\tau_v(N_v/N_a) \qquad \text{(S101)}\end{aligned}$$

The first terms in equations (S100, S101) are the fitness changes due to mutation, the second terms are due to selection. Note that the rescaling of the time $(N_a/N_v)\tau_a$ and $(N_v/N_a)\tau_v$ in eqs. (S100, S101) are respectively equivalent to measurements in units of neutral coalescence time in antibodies $\tau_v = t/N_v$ and in viruses $\tau_a = t/N_a$, which are the natural characteristic times for adaptation in each of these populations. In the stationary state, the fitness flux in each population and the transfer flux from the opposing population sum up to 0, e.g., $\langle \Phi_A + \mathbf{T}_{V\to A}\rangle = 0$, keeping the mean fitness of both populations constant. Non-stationary states occur during transient evolutionary dynamics of the whole population, or when considering a subset of the population, such as a clonal lineage, whose size fluctuates to fixation or extinction. In particular, the imbalance between the fitness flux and the transfer flux may determine the evolutionary fate of a clonal lineage which we discuss in Section 4. A convenient way to measure fitness and transfer flux is from time-shifted fitness measurements, from stationary (Fig. 3) and non-stationary (Fig. S5) co-evolving populations.

4. Evolution of multiple antibody lineages

Fixation probability in a general fitness landscape. We extend our results to multiple clonal antibody lineages evolving with a viral population. The lineages are distinguished by their coupling constants $\{\varepsilon_i^{\mathcal{C}}, \hat{\varepsilon}_i^{\mathcal{C}} \geq 0\}$ for lineage $\mathcal{C}$, which characterizes the lineages' accessibility to regions of the viral sequence. The fraction of all antibodies in lineage $\mathcal{C}$, $\rho_{_\mathcal{C}}(t)$, and its long time behavior determines whether a lineage goes extinct or expands to an appreciable size. Assuming that mutations cannot change one lineage to another, the growth of a given lineage $\mathcal{C}$ depends on the relative mean fitness of the lineage $F_{_{A^\mathcal{C}}}$ compared to the mean fitness of the whole population $F_{_A}(t) = \sum_{\mathcal{C}} F_{_{A^\mathcal{C}}}(t)\rho_{_\mathcal{C}}(t)$,

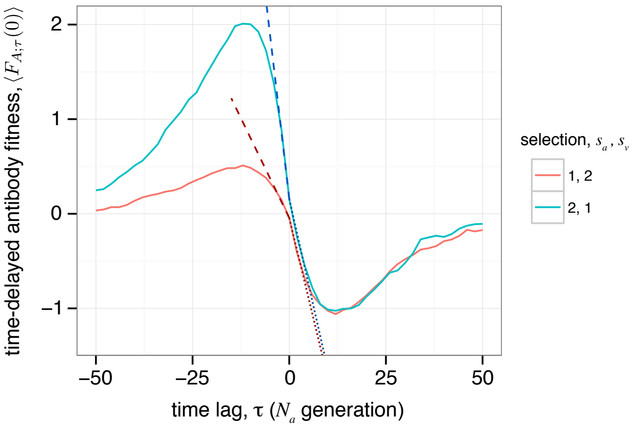

FIG. S5: **Non-stationary signature of co-evolution from time-shifted interactions.** Transient (non-stationary) co-evolutionary dynamics are quantified by the ensemble-averaged time-shifted mean fitness of the antibody population $\langle F_{A;\tau}(t)\rangle = s_a\langle\varepsilon_\tau\rangle = s_a\langle\sum_{\alpha,\gamma} E_{\alpha\gamma}x^\alpha(t)y^\gamma(t+\tau)\rangle/E_0$ for $\tau > 0$ and $\langle F_{A;\tau}(t)\rangle = s_a\langle\sum_{\alpha,\gamma} E_{\alpha\gamma}x^\alpha(t-\tau)y^\gamma(t)\rangle/E_0$ for $\tau < 0$, estimated at a reference time point that is N_a generations after the initial state. The fitness function is shown for two evolutionary regimes, (i) stronger antibody selection, $s_a = 2$, $s_v = 1$ in blue and (ii) weaker antibody selection, $s_a = 1$, $s_v = 2$ in red. The slope of time-shifted fitness at $\tau = 0$ measures the population's fitness flux (towards the past) shown by dashed lines, and the transfer flux from the opposing population (towards the future) shown by dotted lines . Fitness flux and transfer flux do not have equal values in a non-stationary state, leading to the discontinuity in the slope of the time-shifted fitness function at $\tau = 0$. The dashed and dotted lines show the estimated fitness flux and transfer flux which are the slopes of the cumulative flux values in eqs. (S98, S100). The non-stationary Fitness flux in antibodies is larger than the transfer flux from the viruses to the antibody population, when selection on antibodies is stronger (blue). Parameters $\ell = 50$, $\hat{\ell} = 0$, $N_a = N_v = 1000$, $\theta_a = \theta_v = 1/25$, and results were ensemble-averaged over 10^3 simulations.

and on the strength of stochasticity due to genetic drift,

$$\frac{d}{dt}\rho_{\mathcal{C}}(t) = \sum_\alpha \left(f_{\mathcal{C}^\alpha}(t) - F_{\scriptscriptstyle A}(t)\right) x^\alpha_{\mathcal{C}}(t) + \sqrt{\frac{\rho_{\mathcal{C}}(1-\rho_{\mathcal{C}})}{N_a}} \tag{S102}$$

where $f_{\mathcal{C}^\alpha}(t)$ is the fitness of the genotype A^α in the lineage $\mathcal{C}$. Similar to the evolution of single lineage, the growth of multiple lineages also follows an infinite hierarchy of moment equations for the fitness distribution. Here, we truncate these equations at the second central moment of fitness, i.e., the lineage-specific fitness flux $\phi_{\scriptscriptstyle A^{\mathcal{C}}}$ and the transfer flux $\mathcal{T}_{V\to A^{\mathcal{C}}}$. The changes of the ensemble-averaged mean fitness of a lineage $F_{\scriptscriptstyle A^{\mathcal{C}}}(t)$ and the mean fitness of the whole population $F_{\scriptscriptstyle A}(t)$, weighted by the frequency $\rho_{\mathcal{C}}(t)$ follow,

$$\begin{aligned}\left\langle \frac{d}{dt}\sum_\alpha f_{\mathcal{C}^\alpha}(t)\,x^\alpha_{\mathcal{C}}(t)\right\rangle &= \left\langle \rho_{\mathcal{C}}(t)\sum_{\alpha\in\mathcal{C}} \frac{\partial F_{\scriptscriptstyle A^{\mathcal{C}}}}{\partial x^\alpha_{\mathcal{C}}} \times \frac{dx^\alpha_{\mathcal{C}}}{dt}\right\rangle + \left\langle \rho_{\mathcal{C}}(t)\sum_{\gamma} \frac{\partial F_{\scriptscriptstyle A^{\mathcal{C}}}}{\partial y^\gamma} \times \frac{dy^\gamma}{dt}\right\rangle - \frac{1}{N_a}\left\langle F_{\scriptscriptstyle A^{\mathcal{C}}}(t)\,\rho_{\mathcal{C}}(t)\right\rangle \\ &\equiv \left\langle \rho_{\mathcal{C}}(t)\,\phi_{\scriptscriptstyle A^{\mathcal{C}}}(t)\right\rangle + \left\langle \rho_{\mathcal{C}}(t)\,\mathcal{T}_{V\to A^{\mathcal{C}}}(t)\right\rangle - \frac{1}{N_a}\left\langle F_{\scriptscriptstyle A^{\mathcal{C}}}(t)\,\rho_{\mathcal{C}}(t)\right\rangle \end{aligned} \tag{S103}$$

Similarly,

$$\left\langle \frac{d}{dt}\sum_\alpha F_{\scriptscriptstyle A}(t)\,x^\alpha_{\mathcal{C}}(t)\right\rangle = \left\langle \rho_{\mathcal{C}}(t)\,\phi_{\scriptscriptstyle A}(t)\right\rangle + \left\langle \rho_{\mathcal{C}}(t)\,\mathcal{T}_{V\to A}(t)\right\rangle - \frac{1}{N_a}\left\langle F_{\scriptscriptstyle A}(t)\,\rho_{\mathcal{C}}(t)\right\rangle \tag{S104}$$

where $x^\alpha_{\mathcal{C}} \equiv x_{\mathcal{C}}(\mathbf{A}^\alpha)$ is the frequency of the antibody genotype A^α from lineage $\mathcal{C}$ in the total population. Here, we assume that the mean fitness of a lineage only depends on the genotypes within the lineage, as is the case for the fitness functions given in eqs. (S45, S47). The ensemble-averaged changes of the fitness flux and the transfer flux due to selection depend on higher central moments of the fitness distribution, which we neglect in our analysis. The effects of mutation and genetic drift (using Itô calculus) on the flux quantities follow,

$$\frac{d}{dt}\Big\langle \rho_{\mathcal{C}}(t)\,\phi_{_{A^{\mathcal{C}}}}(t)\Big\rangle \simeq \Big\langle \rho_{\mathcal{C}}(t)\, m^{\alpha}_{_A}\frac{\partial}{\partial x^{\alpha}}\phi_{_{A^{\mathcal{C}}}}(t)\Big\rangle + \frac{1}{N_a}\Big[\big\langle \rho_{\mathcal{C}}(t)\,\phi_{_A}(t)\big\rangle - 2\big\langle \rho_{\mathcal{C}}(t)\,\phi_{_{A^{\mathcal{C}}}}(t)\big\rangle\Big] \tag{S105}$$

$$\frac{d}{dt}\Big\langle \rho_{\mathcal{C}}(t)\,\phi_{_A}(t)\Big\rangle \simeq \Big\langle \rho_{\mathcal{C}}(t)\, m^{\alpha}_{_A}\frac{\partial}{\partial x^{\alpha}}\phi_{_A}(t)\Big\rangle + \frac{1}{N_a}\Big[\big\langle \rho_{\mathcal{C}}(t)\,\phi_{_{A^{\mathcal{C}}}}(t)\big\rangle - 2\big\langle \rho_{\mathcal{C}}(t)\,\phi_{_A}(t)\big\rangle\Big] \tag{S106}$$

$$\frac{d}{dt}\Big\langle \rho_{\mathcal{C}}(t)\,\mathcal{T}_{V\to A^{\mathcal{C}}}(t)\Big\rangle \simeq \Big\langle \rho_{\mathcal{C}}(t)\,\Big[m^{\alpha}_{_A}\,\frac{\partial}{\partial x^{\alpha}}\mathcal{T}_{V\to A^{\mathcal{C}}}(t) + m^{\gamma}_{_V}\,\frac{\partial}{\partial y^{\gamma}}\mathcal{T}_{V\to A^{\mathcal{C}}}(t)\Big]\Big\rangle - \frac{1}{N_v}\Big\langle \rho_{\mathcal{C}}(t)\,\mathcal{T}_{V\to A^{\mathcal{C}}}(t)\Big\rangle \tag{S107}$$

$$\frac{d}{dt}\Big\langle \rho_{\mathcal{C}}(t)\,\mathcal{T}_{V\to A}(t)\Big\rangle \simeq \Big\langle \rho_{\mathcal{C}}(t)\,\Big[m^{\alpha}_{_A}\,\frac{\partial}{\partial x^{\alpha}}\mathcal{T}_{V\to}(t) + m^{\gamma}_{_V}\,\frac{\partial}{\partial y^{\gamma}}\mathcal{T}_{V\to A}(t)\Big]\Big\rangle - \frac{1}{N_v}\Big\langle \rho_{\mathcal{C}}(t)\,\mathcal{T}_{V\to A}(t)\Big\rangle \tag{S108}$$

where $m^{\alpha}_{_A}$ and $m^{\gamma}_{_V}$ are the mutational fields associated with frequency changes due to mutations in the antibody $\mathbf{A}^{\alpha}$ and in the virus $\mathbf{V}^{\gamma}$, as defined by eq. (S3).

In order to compute the fixation probability $P_{\text{fix}} = \lim_{t\to\infty}\langle\rho_{\mathcal{C}}(t)\rangle$, it is convenient to use the Laplace transform of the lineage frequency, and compute its asymptotic behavior at large time (see e.g., [71]). The Laplace transform of a given function $A(t)$ can be computed as, $\mathcal{A}(z) = \sum_t A(t)\exp[-zt]$ with the inverse transform: $A(t) = \lim_{T\to\infty}\frac{1}{2\pi i}\int_{\gamma-iT}^{\gamma+iT}\exp[zt]\mathcal{A}(z)$. Following this procedure for the hierarchy of equations (S102-S108) entails a general form for the fixation probability of a lineage, depending on the initial states of the antibody and the viral populations,

$$\begin{aligned} P_{\text{fix}}(\mathcal{C}) &= \lim_{t\to\infty}\langle\rho_{\mathcal{C}}(t)\rangle \\ &= \langle\rho_{\mathcal{C}}(0)\rangle + \Big\langle N_a\big(F_{_{A^{\mathcal{C}}}}(0) - F(0)\big)\,\rho_{\mathcal{C}}(0)\Big\rangle + \frac{1}{3}\Big\langle N_a^2\big(\phi_{_{A^{\mathcal{C}}}}(0) - \phi_{_A}(0)\big)\,\rho_{\mathcal{C}}(0)\Big\rangle \\ &\quad - \Big\langle N_a N_v\Big(\big|\mathcal{T}_{V\to A^{\mathcal{C}}}(0)\big| - \big|\mathcal{T}_{V\to A}(0)\big|\Big)\,\rho_{\mathcal{C}}(0)\Big\rangle + \mathcal{O}(\theta\langle(N\delta f)^2\rangle, \langle(N\delta f)^3\rangle) \end{aligned} \tag{S109}$$

where $\langle(\delta f)^r\rangle$ denotes the r^{th} central moment of the fitness distribution. Here, we have neglected the change in fitness and transfer flux due to mutations, which is of order $\mathcal{O}(\theta\langle(N\delta f)^2\rangle)$. Below, we will explicitly study the mutational terms for the specific case of the linear fitness model in eqs. (S45, S47). The first term in eq. (S109) is the ensemble-averaged initial frequency of the lineage at time $t = 0$, and equals its fixation probability in neutrality. In the presence of selection, lineages of antibodies with higher relative mean fitness, $F_{_{A^{\mathcal{C}}}}(0) - F(0)$, higher rate of adaptation, $\phi_{_{A^{\mathcal{C}}}}(0) - \phi_{_A}(0)$, and lower (absolute) transfer flux from viruses, $\big|\mathcal{T}_{V\to A^{\mathcal{C}}}(0)\big| - \big|\mathcal{T}_{V\to A}(0)\big|$, tend to dominate the population. In the following, we will derive in detail the exact form of the fixation probability for evolution in linear fitness landscapes given by eqs. (S45, S47).

Fixation probability in the linear fitness landscape. For evolution in a linear fitness landscape, the growth of a lineage depends on its relative binding affinity compared to the rest of the population. In order to quantify the competition between the lineages, we define the lineage-specific moments,

$$L^{\mathcal{C}}_{A_m} = \langle \sum_{\alpha}(E_{\alpha\,.} - \mathcal{E})^m\, x^{\alpha}_{\mathcal{C}}\rangle, \qquad \hat{L}^{\mathcal{C}}_{A_m} = \langle \sum_{\alpha}(\hat{E}_{\alpha\,.} - \hat{\mathcal{E}})^m\, x^{\alpha}_{\mathcal{C}}\rangle \tag{S110}$$

$$L^{\mathcal{C}}_{A_{m;n}} = \Big\langle \sum_{\alpha}(E_{\alpha\,.} - \mathcal{E})^m\, x^{\alpha}_{\mathcal{C}} \sum_{\beta,\mathcal{C}'}(E_{\beta\,.} - \mathcal{E})^n\, x^{\beta}_{\mathcal{C}'}\Big\rangle \tag{S111}$$

$$\hat{L}^{\mathcal{C}}_{A_{m;n}} = \Big\langle \sum_{\alpha}(\hat{E}_{\alpha\,.} - \hat{\mathcal{E}})^m\, x^{\alpha}_{\mathcal{C}} \sum_{\beta,\mathcal{C}'}(\hat{E}_{\beta\,.} - \hat{\mathcal{E}})^n\, x^{\beta}_{\mathcal{C}'}\Big\rangle \tag{S112}$$

$$L^{\mathcal{C}}_{A_m,V_k} = \Big\langle \sum_{\gamma}(E_{.\,\gamma} - \mathcal{E})^k y^{\gamma} \sum_{\alpha}(E_{\alpha\,\gamma} - \mathcal{E})^m\, x^{\alpha}_{\mathcal{C}}\Big\rangle \tag{S113}$$

In this notation, $L^{\mathcal{C}}_{A_0} \equiv \langle\rho_{\mathcal{C}}\rangle$. As given by eq. (S105), the change in the frequency of the lineage $\mathcal{C}$, follows from the evolution equation,

$$\frac{d}{dt}L^{\mathcal{C}}_{A_0} = S_a(L^{\mathcal{C}}_{A_1} + \hat{L}^{\mathcal{C}}_{A_1}) \tag{S114}$$

As discussed above, the dynamics of multiple lineages follows from an infinite hierarchy of moment equations, which we truncate at the second moment to estimate the fixation probability of an antibody lineage up to the order of $\mathcal{O}(s^2)$. The hierarchy of evolution equations for the lineage specific moments $L^{\mathcal{C}}_{A_m}$ and the cross-statistics $L^{\mathcal{C}}_{A_m,V_k}$ follow,

variable region:

$$\frac{d}{dt}L^{\mathcal{C}}_{A_1} = S_a(L^{\mathcal{C}}_{A_2} - L^{\mathcal{C}}_{A_{(0;2)}}) - S_v(L^{\mathcal{C}}_{A_1,V_1} - L^{\mathcal{C}}_{A_0,V_2}) - 2(\mu_a + \mu_v)L^{\mathcal{C}}_{A_1} - \frac{L^{\mathcal{C}}_{A_1}}{N_a} \tag{S115}$$

$$\frac{d}{dt}L^{\mathcal{C}}_{A_2} = -4\mu_a(L^{\mathcal{C}}_{A_2} - \ell\,\mathcal{K}_{2,\mathcal{C}}L^{\mathcal{C}}_{A_0}) - 4\mu_v L^{\mathcal{C}}_{A_2} + \frac{L^{\mathcal{C}}_{A_{(0;2)}} - 2L^{\mathcal{C}}_{A_2}}{N_a} + \mathcal{O}(S_a, S_v) \tag{S116}$$

$$\frac{d}{dt}L^{\mathcal{C}}_{A_{(0;2)}} = -4\mu_a(L^{\mathcal{C}}_{A_{(0;2)}} - \ell\,[\mathcal{K}_{2,\mathcal{C}}]_{\mathcal{C}}L^{\mathcal{C}}_{A_0}) - 4\mu_v L^{\mathcal{C}}_{A_{(0;2)}} + \frac{L^{\mathcal{C}}_{A_2} - 2L^{\mathcal{C}}_{A_{(0;2)}}}{N_a} + \mathcal{O}(S_a) \tag{S117}$$

$$\frac{d}{dt}L^{\mathcal{C}}_{A_1,V_1} = -4\mu_a L^{\mathcal{C}}_{A_1,V_1} - 4\mu_v(L^{\mathcal{C}}_{A_1,V_1} - \ell\sqrt{\mathcal{K}_{2,\mathcal{C}}\,[\mathcal{K}_{2,\mathcal{C}}]_{\mathcal{C}}}\;L^{\mathcal{C}}_{A_0}) - \frac{L^{\mathcal{C}}_{A_1,V_1}}{N_v} + \mathcal{O}(S_a) \tag{S118}$$

$$\frac{d}{dt}L^{\mathcal{C}}_{A_0,V_2} = -4\mu_a L^{\mathcal{C}}_{A_0,V_2} - 4\mu_v(L^{\mathcal{C}}_{A_0,V_2} - \ell\,[\mathcal{K}_{2,\mathcal{C}}]_{\mathcal{C}}L^{\mathcal{C}}_{A_0}) - \frac{L^{\mathcal{C}}_{A_0,V_2}}{N_v} + \mathcal{O}(S_a) \tag{S119}$$

conserved region:

$$\frac{d}{dt}\hat{L}^{\mathcal{C}}_{A_1} = S_a(L^{\mathcal{C}}_{A_2} - L^{\mathcal{C}}_{A_{(0;2)}}) - 2\mu_a\hat{L}^{\mathcal{C}}_{A_1} - \frac{\hat{L}^{\mathcal{C}}_{A_1}}{N_a} \tag{S120}$$

$$\frac{d}{dt}\hat{L}^{\mathcal{C}}_{A_2} = -4\mu_a(\hat{L}^{\mathcal{C}}_{A_2} - \hat{\ell}\,\hat{\mathcal{K}}_{2,\mathcal{C}}L^{\mathcal{C}}_{A_0}) + \frac{\hat{L}^{\mathcal{C}}_{A_{(0;2)}} - 2\hat{L}^{\mathcal{C}}_{A_2}}{N_a} + \mathcal{O}(S_a) \tag{S121}$$

$$\frac{d}{dt}\hat{L}^{\mathcal{C}}_{A_{(0;2)}} = -4\mu_a(\hat{L}^{\mathcal{C}}_{A_{(0;2)}} - \hat{\ell}\,[\hat{\mathcal{K}}_{2,\mathcal{C}}]_{\mathcal{C}}L^{\mathcal{C}}_{A_0}) + \frac{\hat{L}^{\mathcal{C}}_{A_2} - 2\hat{L}^{\mathcal{C}}_{A_{(0;2)}}}{N_a} + \mathcal{O}(S_a) \tag{S122}$$

with the lineage averaged statistics,

$$[\mathcal{K}_{2,\mathcal{C}}]_{\mathcal{C}} = \sum_{\text{lineages } \mathcal{C}} \mathcal{K}_{2,\mathcal{C}}\,\rho_{\mathcal{C}}, \qquad [\hat{\mathcal{K}}^{\mathcal{C}}_2]_{\mathcal{C}} = \sum_{\text{lineages } \mathcal{C}} \hat{\mathcal{K}}_{2,\mathcal{C}}\,\rho_{\mathcal{C}} \tag{S123}$$

In order to compute the fixation probability, we use the Laplace transform of the lineage specific moments $\mathcal{L}_{A_m,V_k}$ and compute the asymptotic behavior of the 0^{th} moment, $L^{\mathcal{C}}_0$ after the inverse transform (see e.g., [65, 71]). The Laplace transform for the hierarchy of moment equations (S115-S122) up to order of $\mathcal{O}(S^2)$ in $\mathcal{L}^{\mathcal{C}}_{A_0}$ entails,

$$z\mathcal{L}^{\mathcal{C}}_{A_0}(z) - L^{\mathcal{C}}_{A,0}(0) = S_a(\mathcal{L}^{\mathcal{C}}_{A_1}(z) + \hat{\mathcal{L}}^{\mathcal{C}}_1(z)) \tag{S124}$$

variable region:

$$z\mathcal{L}^{\mathcal{C}}_{A_1}(z) - L^{\mathcal{C}}_{A_1}(0) = S_a(\mathcal{L}^{\mathcal{C}}_{A_2}(z) - \mathcal{L}^{\mathcal{C}}_{A_{(0;2)}}(z)) - S_v(\mathcal{L}^{\mathcal{C}}_{A_1,V_1} - \mathcal{L}^{\mathcal{C}}_{A_0,V_2}) - 2(\mu_a + \mu_v)\mathcal{L}^{\mathcal{C}}_{A_1}(z) - \frac{\mathcal{L}^{\mathcal{C}}_{A_1}(z)}{N_a} \tag{S125}$$

$$z\mathcal{L}^{\mathcal{C}}_{A_2}(z) - L^{\mathcal{C}}_{A_2}(0) = -4\mu_a(\mathcal{L}^{\mathcal{C}}_{A_2}(z) - \ell\,\mathcal{K}^{\mathcal{C}}_2\mathcal{L}^{\mathcal{C}}_{A_0}(z)) - 4\mu_v\mathcal{L}^{\mathcal{C}}_{A_2}(z) + \frac{\mathcal{L}^{\mathcal{C}}_{A_{(0;2)}}(z) - 2\mathcal{L}^{\mathcal{C}}_{A_2}(z)}{N_a} \tag{S126}$$

$$z\mathcal{L}^{\mathcal{C}}_{A_{(0;2)}} - L^{\mathcal{C}}_{A_{(0;2)}}(0) = -4\mu_a(\mathcal{L}^{\mathcal{C}}_{A_{(0;2)}} - \ell\,[\mathcal{K}^{\mathcal{C}}_2]_{\mathcal{C}}\,\mathcal{L}^{\mathcal{C}}_{A_0}) - 4\mu_v\mathcal{L}^{\mathcal{C}}_{A_{(0;2)}} + \frac{\mathcal{L}^{\mathcal{C}}_{A_2} - 2\mathcal{L}^{\mathcal{C}}_{A_{(0;2)}}}{N_a} \tag{S127}$$

$$z\mathcal{L}^{\mathcal{C}}_{A_1,V_1} - L^{\mathcal{C}}_{A_1,V_1}(0) = -4\mu_a\mathcal{L}^{\mathcal{C}}_{A_1,V_1} - 4\mu_v(\mathcal{L}^{\mathcal{C}}_{A_1,V_1} - \ell\sqrt{\mathcal{K}_{2,\mathcal{C}}\,[\mathcal{K}_{2,\mathcal{C}}]_{\mathcal{C}}}\;\mathcal{L}^{\mathcal{C}}_{A_0}) - \frac{\mathcal{L}^{\mathcal{C}}_{A_1,V_1}}{N_v} \tag{S128}$$

$$z\mathcal{L}^{\mathcal{C}}_{A_0,V_2} - L^{\mathcal{C}}_{A_0,V_2}(0) = -4\mu_a\mathcal{L}^{\mathcal{C}}_{A_0,V_2} - 4\mu_v(\mathcal{L}^{\mathcal{C}}_{A_0,V_2} - \ell\,[\mathcal{K}_{2,\mathcal{C}}]_{\mathcal{C}}\mathcal{L}^{\mathcal{C}}_{A_0}) - \frac{\mathcal{L}^{\mathcal{C}}_{A_0,V_2}}{N_v} \tag{S129}$$

conserved region:

$$z\hat{\mathcal{L}}^{\mathcal{C}}_{A_1}(z) - \hat{L}^{\mathcal{C}}_{A_1}(0) = S_a(\hat{\mathcal{L}}^{\mathcal{C}}_{A_2}(z) - \hat{\mathcal{L}}^{\mathcal{C}}_{A_{(0;2)}}(z)) - 2\mu_a\hat{\mathcal{L}}^{\mathcal{C}}_{A_1}(z) - \frac{\hat{\mathcal{L}}^{\mathcal{C}}_{A_1}(z)}{N_a} \tag{S130}$$

$$z\hat{\mathcal{L}}^{\mathcal{C}}_{A_2}(z) - \hat{L}^{\mathcal{C}}_{A_2}(0) = \frac{\hat{\mathcal{L}}^{\mathcal{C}}_{A_{(0;2)}}(z) - 2\hat{\mathcal{L}}^{\mathcal{C}}_{A_2}(z)}{N_a} - 4\mu_a(\hat{\mathcal{L}}^{\mathcal{C}}_{A_2}(z) - \hat{\ell}\hat{\mathcal{K}}^{\mathcal{C}}_2\hat{\mathcal{L}}^{\mathcal{C}}_{A_0}(z)) \tag{S131}$$

$$z\hat{\mathcal{L}}^{\mathcal{C}}_{A_{(0;2)}} - \hat{L}^{\mathcal{C}}_{A_{(0;2)}}(0) = -4\mu_a(\hat{\mathcal{L}}^{\mathcal{C}}_{A_{(0;2)}} - \hat{\ell}\,[\hat{\mathcal{K}}^{\mathcal{C}}_2]_{\mathcal{C}}\hat{\mathcal{L}}^{\mathcal{C}}_0) + \frac{\hat{\mathcal{L}}^{\mathcal{C}}_{A_2} - 2\hat{\mathcal{L}}^{\mathcal{C}}_{A_{(0;2)}}}{N_a} \tag{S132}$$

The inverse transform of $\mathcal{L}^{\mathcal{C}}_{A_0}(z)$ in the limit of $z \to 0$ results in the asymptotic behavior of the ensemble-averaged frequency of the lineage $\mathcal{C}$, $\lim_{t\to\infty} L^{\mathcal{C}}_{A_0}$, which corresponds to the fixation probability P_{fix} of the lineage,

$$\begin{aligned}P_{\text{fix}}(\mathcal{C}) &= \lim_{t\to\infty} L^{\mathcal{C}}_{A_0}(t)\\ &= L^{\mathcal{C}}_{A_0}(0) + \frac{N_a\,S_a}{1+2\tilde{\theta}_a}L^{\mathcal{C}}_{A_1}(0) + \frac{N_a\,S_a}{(1+2\tilde{\theta}_a)}\left[\frac{N_aS_a(L^{\mathcal{C}}_{A_2}(0) - L^{\mathcal{C}}_{A_{(0;2)}}(0))}{3+4\tilde{\theta}_a} - \frac{N_v\,S_v\left(L^{\mathcal{C}}_{A_1,V_1}(0) - L^{\mathcal{C}}_{A_0,V_2}(0)\right)}{1+4\tilde{\theta}_v}\right]\\ &\quad + \frac{N_a\,S_a}{1+2\theta_a}\hat{L}^{\mathcal{C}}_{A_1}(0) + \frac{N_a\,S_a}{(1+2\theta_a)}\left[\frac{N_aS_a(\hat{L}^{\mathcal{C}}_{A_2}(0) - \hat{L}^{\mathcal{C}}_{A_{(0;2)}}(0))}{3+4\theta_a}\right]\end{aligned} \tag{S133}$$

The fixation probability of a lineage can be characterized by the state of the antibody population and viral population upon its introduction. The first term in eq. (S133) is the frequency of the antibody lineage at the time of introduction, and is equal to the neutral fixation probability. The second term, which is favored by the antibody selection coefficient, measures the relative fitness of the lineage $\mathcal{C}$ to the mean fitness of the population. The terms proportional to the $(N_aS_a)^2$ measure the relative fitness flux of the lineage $\mathcal{C}$ to the fitness flux of the whole population. The terms proportional to $(N_aS_a) \times (N_vS_v)$ measure the transfer flux from the viral population to the antibody lineage $\mathcal{C}$ relative to the total transfer flux from viruses to the antibody population.

As mentioned in the main text, the higher viral diversity favors the fixation of broadly neutralizing antibodies for two reasons. First, the larger viral diversity compromises the mean fitness of the resident non-broad antibody population, and makes it easier for the potential BnAb lineage to take over the existing antibody lineages. This effect is captured by terms proportional to N_aS_a in eq. (S133). Second, the transfer flux from the viral population to the lineage with access to the conserved interaction regions (i.e, a lineage with $\hat{E}^2_0/E^2_0 \gg 1$) is small. Therefore, the viral escape from binding to a potential BnAb lineage is less efficient than from the resident non-broad antibody population, which increases the chances of fixation for a potential BnAb lineage. This effect is captured by terms proportional to $(N_aS_a) \times (N_vS_v)$ in eq. (S133). Similar conclusions regarding the elicitation of BnAbs have been drawn from numerical analysis by Luo & Pelerson [30].

[1] C. A. Janeway, P. Travers, M. Walport, and M. Shlomchik, *Immunobiology: The Immune System in Health and Disease (Garland Science, New York)* (2005).

[2] Y. Elhanati and et al., Phil. Trans. R. Soc. B **370** (2015).

[3] F. Trepel, Klin. Wochenschrift **52**, 511 (1974).

[4] J. Glanville and et al., Proc. Natl. Acad. Sci. U.S.A. **106**, 20216 (2009).

[5] V. H. Odegard and D. G. Schatz, Nature Rev. Immunol. **6**, 573 (2006).

[6] C. D. Campbell and E. E. Eichler, Trends. Genet. **29**, 575 (2013).

[7] M. Meyer-Hermann and et al., Cell Rep. **2**, 162 (2012).

[8] G. D. Victora and M. C. Nussenzweig, Annu. Rev. Immunol. **30**, 429 (2012).

[9] S. Cobey, P. Wilson, and F. A. Matsen, Phil. Trans. R. Soc. B **370** (2015).

[10] S. Duffy, L. A. Shackelton, and E. C. Holmes, Nature Rev. Genet. **9**, 267 (2008).

[11] D. D. Richman, T. Wrin, S. J. Little, and C. J. Petropoulos, Proc. Natl. Acad. Sci. U.S.A. **100**, 4144 (2003).
[12] P. L. Moore and et al., PLoS Pathog. **5**, e1000598 (2009).
[13] H.-X. Liao and et al., Nature **496**, 469 (2013).
[14] S. Luo and A. S. Perelson, Phil. Trans. R. Soc. B **370** (2015).
[15] P. D. Kwong and et al., Nature **420**, 678 (2002).
[16] D. Lyumkis and et al., Science **342**, 1484 (2013).
[17] M. D. Simek and et al., J. Virol. **83**, 7337 (2009).
[18] N. A. Doria-Rose and et al., J. Virol. **84**, 1631 (2010).
[19] L. M. Walker and et al., Science **326**, 285 (2009).
[20] L. Chen and et al., Science **326**, 1123 (2009).
[21] T. Zhou and et al., Science **329**, 811 (2010).
[22] L. M. Walker and et al., Nature **477**, 466 (2011).
[23] H. Mouquet and M. C. Nussenzweig, Nature **496**, 441 (2013).
[24] P. D. Kwong, J. R. Mascola, and G. J. Nabel, Nature Rev. Immunol. **13**, 693 (2013).
[25] F. Klein and et al., Science **341**, 1199 (2013).
[26] S. Wang and et al., Cell **160**, 785 (2015).
[27] L. M. Childs, E. B. Baskerville, and S. Cobey, Phil. Trans. R. Soc. B **370** (2015).
[28] S. Chaudhury, J. Reifman, and A. Wallqvist, J. Immunol. **193**, 2073 (2014).
[29] A. S. Perelson, Nature Rev. Immunol. **2**, 28 (2002).
[30] S. Luo and A. S. Perelson, Proc Natl Acad Sci USA **112**, 11654 (2015).
[31] R. Shankarappa and et al., J. Virol. **73**, 10489 (1999).
[32] A. Nourmohammad, T. Held, and M. Lässig, Curr. Opin. Genet. Dev. **23**, 684 (2013).
[33] V. Detours and A. S. Perelson, Proc. Natl. Acad. Sci. U.S.A. **96**, 5153 (1999).
[34] V. Detours and A. S. Perelson, Proc. Natl. Acad. Sci. U.S.A. **97**, 8479 (2000).
[35] M. Lynch and B. Walsh, *Genetics and analysis of quantitative traits* (Sinauer Associates Inc, 1998).
[36] A. Nourmohammad, S. Schiffels, and M. Lässig, J. Stat. Mech. Theor. Exp. **2013**, P01012 (2013).
[37] T. Held, A. Nourmohammad, and M. Lässig, J. Stat. Mech. Theor. Exp. **2014**, P09029 (2014).
[38] G. R. Price, Nature **227**, 520 (1970).
[39] P. Lemey, A. Rambaut, and O. G. Pybus, AIDS Rev. **8**, 125 (2006).
[40] F. Zanini and et al., eLife **10.7554/eLife.11282** (2015).
[41] R. A. Fisher, *The genetical theory of natural selection* (Oxford University Press, USA, 1930), 1st ed.
[42] V. Mustonen and M. Lässig, Trends Genet. **25**, 111 (2009).
[43] V. Mustonen and M. Lässig, Proc. Natl. Acad. Sci. U.S.A. **107**, 4248 (2010).
[44] F. Blanquart and S. Gandon, Ecol. Lett. **16**, 31 (2013).
[45] K. B. Hoehn and et al., Phil. Trans. R. Soc. B **370** (2015).
[46] R. Wyatt and J. Sodroski, Science **280**, 1884 (1998).
[47] Y. Zhang and et al., The Journal of experimental medicine **210**, 457 (2013).
[48] R. A. Neher and T. Leitner, PLoS Comput. Biol. **6**, e1000660 (2010).
[49] T. W. Chun and et al., Nature **387**, 183 (1997).
[50] M. A. Brockhurst and B. Koskella, Trends Ecol. Evol. **28**, 367 (2013).
[51] B. Koskella and M. A. Brockhurst, FEMS Microbiol. Rev. **38**, 916 (2014).
[52] A. R. Hall, P. D. Scanlan, A. D. Morgan, and A. Buckling, Ecol. Lett. **14**, 635 (2011).
[53] A. Betts, O. Kaltz, and M. E. Hochberg, Proc. Natl. Acad. Sci. U.S.A. **111**, 11109 (2014).
[54] A. Agrawal and C. M. Lively, Evol. Ecol. Res. **4**, 91 (2002).
[55] J. M. Fonville and et al., Science **346**, 996 (2014).
[56] L. Verkoczy, G. Kelsoe, M. Moody, and B. Haynes, Curr. Opin. Immunol. **23**, 383 (2011).
[57] T. Kepler and et. al., Front. Immunol. **5**, 170 (2014).
[58] O. Tange, ;login: The USENIX Magazine **36**, 42 (2011).
[59] C. Gardiner, *Handbook of Stochastic methods: for physics, chemistry and the natural sciences* (Springer, 2004), 3rd ed.
[60] M. Kimura, J. Appl. Probab. **1**, 177 (1964).
[61] P. L. Antonelli and C. Strobeck, Adv. Appl. Probab. **9**, 238 (1977).
[62] R. A. Fisher, Proc. R. Soc. Edinb. **50**, 205 (1930).
[63] M. Kimura, *The neutral allele theory of molecular evolution* (Cambridge University Press, Cambridge, UK, 1983).
[64] P. G. Higgs and G. Woodcock, J. Math. Biol. **33**, 677 (1995).
[65] B. H. Good and M. M. Desai, Theor. Popul. Biol. **85**, 86 (2013).
[66] A. Kosmrlj, A. K. Jha, E. S. Huseby, M. Kardar, and A. K. Chakraborty, Proc. Natl. Acad. Sci. U.S.A. **105**, 16671 (2008).
[67] A. Kosmrlj, A. K. Chakraborty, M. Kardar, and E. I. Shakhnovich, Phys. Rev. Lett. **103**, 068103 (2009).
[68] L. de Haan and A. Ferreira, *Extreme value theory: an introduction* (Springer US, New York, 2006).
[69] V. Mustonen and M. Lässig, Proc. Natl. Acad. Sci. U.S.A. **104**, 2277 (2007).
[70] A. Nourmohammad, J. Rambeau, T. Held, J. Berg, and M. Lässig, arXiv pp. q–bio/1502.06406v2 (2015).
[71] M. M. Desai and D. S. Fisher, Genetics **17**, 385 (2007).