Spin models inferred from patient-derived viral sequence data faithfully describe HIV fitness landscapes

Karthik Shekhar,[1,2] Claire F. Ruberman,[3] Andrew L. Ferguson,[4] John P. Barton,[1,2]
Mehran Kardar,[5,*] and Arup K. Chakraborty[1,2,5,6,7,8,†]
[1]*Department of Chemical Engineering, MIT, Cambridge, Massachusetts 02139, USA*
[2]*Ragon Institute of MGH, MIT and Harvard, Boston, Massachusetts 02129, USA*
[3]*Department of Mathematics, Pomona College, Claremont, California 91711, USA*
[4]*Department of Materials Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*
[5]*Department of Physics, MIT, Cambridge, Massachusetts 02139, USA*
[6]*Department of Chemistry, MIT, Cambridge, Massachusetts 02139, USA*
[7]*Department of Biological Engineering, MIT, Cambridge, Massachusetts 02139, USA*
[8]*Institute for Medical Engineering and Science, MIT, Cambridge, Massachusetts 02139, USA*
(Received 10 June 2013; published 4 December 2013)

Mutational escape from vaccine-induced immune responses has thwarted the development of a successful vaccine against AIDS, whose causative agent is HIV, a highly mutable virus. Knowing the virus' fitness as a function of its proteomic sequence can enable rational design of potent vaccines, as this information can focus vaccine-induced immune responses to target mutational vulnerabilities of the virus. Spin models have been proposed as a means to infer intrinsic fitness landscapes of HIV proteins from patient-derived viral protein sequences. These sequences are the product of nonequilibrium viral evolution driven by patient-specific immune responses and are subject to phylogenetic constraints. How can such sequence data allow inference of intrinsic fitness landscapes? We combined computer simulations and variational theory á la Feynman to show that, in most circumstances, spin models inferred from patient-derived viral sequences reflect the correct rank order of the fitness of mutant viral strains. Our findings are relevant for diverse viruses.

DOI: 10.1103/PhysRevE.88.062705 PACS number(s): 87.10.−e, 87.19.xd, 87.18.Vf, 87.23.Cc

I. INTRODUCTION

The staggering sequence diversity of HIV [1] and its ability to evade most natural or vaccine-induced immune responses by mutational escape [2] have precluded the development of a successful vaccine against this global epidemic [3]. It has been proposed that a vaccine-induced immune response should target regions in the viral proteome, where escape mutations are most likely to damage replicative fitness. Single residues that appear highly conserved in proteins derived from virus samples extracted from diverse patients have been suggested as vaccine targets [4], but the fitness cost [5] of making mutations at such sites can be restored by additional compensatory mutations [6]. Groups of sites in HIV proteins that are collectively constrained such that multiple simultaneous mutations within such groups impose a high fitness penalty have been identified and shown to be targeted by patients whose immune systems naturally control HIV [7]. But these models cannot identify which specific sites in these collectively coevolving groups should be targeted to maximally compromise viral fitness and how mutational escape pathways that exist even within these regions may be blocked by additional immune responses. Answering these questions requires knowledge of the complex, multidimensional structure of HIV's fitness landscape [8,9]—a measure of the virus' replicative capacity as a function of the amino acid sequence of its constituent proteins. Knowledge of the fitness landscape can guide systematic identification of the mutational vulnerabilities of viruses (not just HIV), and the rational design of vaccines that can target these weaknesses.

Using inference principles rooted in entropy maximization [10,11], Ferguson *et al.* [9] recently employed publicly available multiple sequence alignments (MSA) [12] of four HIV proteins to obtain the prevalence of HIV strains bearing multiple mutations in these proteins. In this approach, a viral protein of N sites is described by a coarse-grained binary code $\vec{s} = \{0,1\}^N$, wherein the "wild-type" (most frequent) amino acid at site i is denoted by $s_i = 0$ and any mutant is denoted by $s_i = 1$ (irrespective of its identity). Given the protein MSA, the maximum entropy framework seeks a minimally biased probability distribution $P[\vec{s}]$ over the space of all possible mutant strains $\{\vec{s}\}$ that reproduces the marginal one-site and two-site mutational probabilities $\langle s_i \rangle_{\mathrm{MSA}}$ and $\langle s_i s_j \rangle_{\mathrm{MSA}}$ measured from sequence data. The resulting inference leads to a model where the probability of a particular strain $\vec{s}$, the *prevalence landscape*, is described by a Boltzmann distribution, $P[\vec{s}] \propto e^{-H_0[\vec{s}]}$, where the Hamiltonian takes the form of an infinite-range Ising spin glass [13],

$$H_0[\vec{s}] = \sum_{i<j=1}^{N} J_{ij} s_i s_j + \sum_{i=1}^{N} h_i s_i. \qquad (1)$$

This can be generalized to account for the identities of mutant amino acids using Potts models rather than an Ising model [9].

Only under certain restricted circumstances [14] can it be shown that this prevalence landscape is the replicative fitness landscape, with the "energy" $H_0[\vec{s}]$ of a strain $\vec{s}$ being inversely correlated with its replicative capacity. Exact solutions to the mathematical model of evolution originally

*kardar@mit.edu
†arupc@mit.edu

proposed by M. Eigen [15] have been found under certain conditions, in particular for the class of fitness landscapes where fitness is a function of the mutational distance from the wild-type strain [16]. But, these conditions are likely inapplicable for a complex problem like HIV evolution in a population of human patients, and the connection between fitness and prevalence is not obvious. Ferguson *et al.* [9] report a statistically significant negative correlation between values of $H_0[\vec{s}]$ predicted by the inferred model and *in vitro* fitness measurements of several engineered strains (including both published and *de novo* experiments). Their predictions also tested positively against clinical data. These results provide evidence that the prevalence landscape of the virus described by $H_0[\vec{s}]$ is a good proxy for the intrinsic fitness landscape.

This pleasing result is surprising. However, the quantities used to parametrize the prevalence landscape ($\langle s_i \rangle_{\mathrm{MSA}}$ and $\langle s_i s_j \rangle_{\mathrm{MSA}}$) were obtained from "consensus" (most common) protein strains in genetically diverse patients. The sequence evolution of HIV within a particular host is a nonequilibrium process driven by a genetically determined pattern of *immune pressure* acting on the viral proteome [17], which determines the effectively fittest viral strains in each patient. Thus, the HIV sequences used to infer the prevalence landscape were not sampled from an equilibrium ensemble of sequences distributed according to their "intrinsic" fitness. In contrast, the *in vitro* measurements are not subject to human immune responses and therefore assay the intrinsic fitness of the virus. The robust correlation between model predictions and *in vitro* fitness measurements observed by Ferguson *et al.* [9], and their ability to describe clinical data in humans with diverse genotypes when the immune responses were known, therefore pose an important question: How does a prevalence landscape inferred from the statistics of mutations in a nonequilibrium ensemble of sequences evolving under diverse adaptive immune responses faithfully reflect the intrinsic fitness of mutant viral strains?

Here we combine computer simulations and analytical theory to address this question. We find that the presence of genetically diverse immune responses imposed by patients across the population is necessary for comprehensive sampling of the sequence space of viral proteins. We then show that the prevalence landscape inferred from mutational correlations observed in the sequence databases correctly reflects the rank order of the intrinsic replicative fitness of mutant viral strains in most circumstances. We provide mechanistic insights into why this is so and circumstances wherein this may not be true.

II. SIMULATIONS

For the computer simulations, as an example, we study the 132-residue HIV matrix protein p17 [18]. We consider a growing population of infected hosts and model the network of viral transmission between hosts as illustrated in Fig. 1. In all simulations, infection in the first host is seeded with N_v copies of the wild-type (WT) strain ($\vec{s^0} = \vec{0}$). New hosts added to this network are infected with N_v identical copies of a strain randomly selected from the quasispecies within a host chosen randomly from the existing pool of hosts. For simplicity, the number of viral strains in a host is chosen to be a constant N_v, making our intrahost evolutionary model similar in spirit to the Wright-Fisher model [19]. N_v serves as an "effective population size" as in conventional population genetics, and

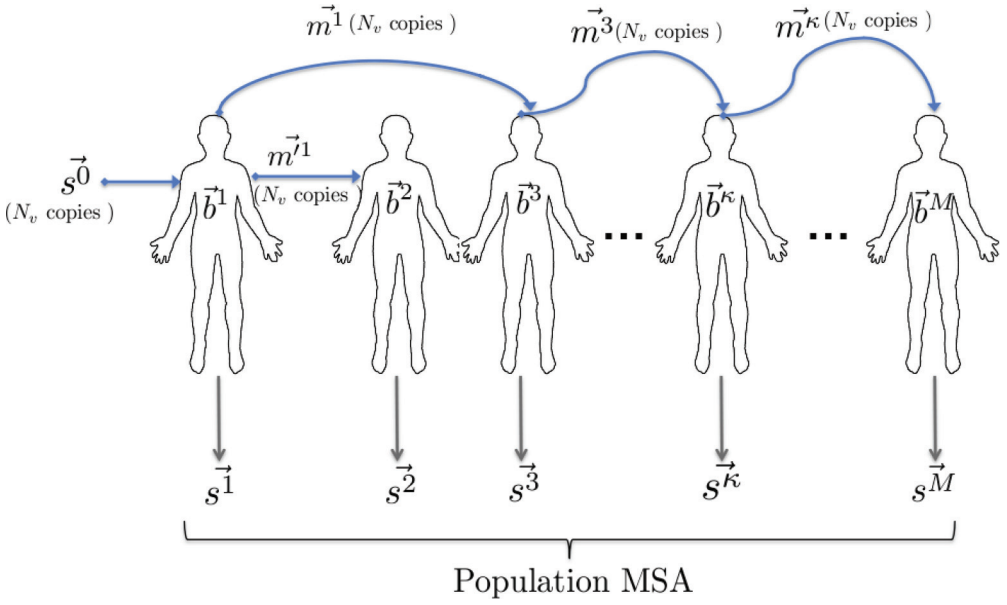

FIG. 1. (Color online) Graphical description of simulation model. We consider an expanding network of infected hosts. The first host in the network is infected with N_v copies of $\vec{s^0} = \vec{0}$, corresponding to the wild-type (WT) strain. Each new host added to the network is infected with N_v copies of a single viral strain derived from the quasispecies within an existing host chosen at random (based on evidence that most infections are initiated by a single virus strain). $\vec{b^\kappa}$ is a "field" that acts on the proteomic sites within host κ and represents the genetically determined immune response within that host. The "consensus" viral strain is extracted from every productively infected host, $i \in 1, 2, \ldots, M$, and added to the *in silico* population ensemble, mimicking the way these sequences were sampled from a real population. For example, in this figure, infection in host κ is seeded with N_v copies of strain $\vec{m^3}$, which is randomly chosen from the quasispecies within host 3. At a randomly chosen generation of viral quasispecies evolution within host κ, the consensus strain $\vec{s^\kappa}$ is derived from the quasispecies and added to the population ensemble.

not the actual number of strains in a host, which is usually much larger than the range of N_v ($2 \times 10^3 - 5 \times 10^5$) we have considered [20].

We assume that the fitness of a particular strain $\vec{s}$ of the N-site protein in a given host is described by an *effective* Hamiltonian $H[\vec{s}] = H_{\rm int}[\vec{s}] + I[\vec{s}]$. Here, $H_{\rm int}[\vec{s}] = \sum_{i<j=1}^{N} J_{ij} s_i s_j + \sum_{i=1}^{N} h_i s_i$ constitutes intrinsic fitness that is independent of host, while $I[\vec{s}] = -\sum_{i=1}^{N} b_i s_i$ is a "host-specific" immune pressure that applies only to some sites: the sites s_i for which b_i are nonzero are distinct for different hosts. Within each host, the fields $\{b_i\}$ are chosen in a manner consistent with known clinical information (see Appendix A and Supplemental Material [21]). Post infection, the viral quasispecies in a host evolve through a nonequilibrium mutation-selection process in discrete generations described in Appendix B. In our model, the evolutionary timescale is coarse-grained such that each generation corresponds to multiple replication cycles of the viral quasispecies. We evolve the quasispecies in a host for a random number of generations τ_S, chosen uniformly between 25 and 500. Empirically, we find that our results do not change qualitatively as long $\langle \tau_S \rangle > 150$ (cf. Supplementary Note 5), which might tentatively correspond to a time scale in which the quasispecies is able to sense the immune pressure and respond through adaptive mutations. The consensus strain within each host is extracted at a randomly chosen generation, and an ensemble of such strains is recorded, which we refer to as the *population ensemble*. This mimics how actual sequences in public databases were collected from patients.

For $\{J_{ij}, h_i\}$, the parameters of $H_{\rm int}[\vec{s}]$, we have used numerical values of the maximum entropy model $H_0[\vec{s}]$ of Ferguson *et al.* [9] inferred from available p17 sequences [12]. In other words, we assume that the intrinsic fitness landscape is correctly given by these parameters and carry out our numerical simulations of the nonequilibrium dynamics with the effects of immune pressure included. Recall that $H_0[\vec{s}]$ (and therefore $H_{\rm int}[\vec{s}]$) reproduces the one- and two-point mutational probabilities within the real p17 sequences (see Supplementary Fig. S4). From the simulations we obtain a "virus sample" from each *in silico* "patient," and from this ensemble, we compute the one- and two-point mutational probabilities. We then ask whether the mutational probabilities in this *in silico* population ensemble are the same as those obtained from real p17 sequences, which are reproduced by $H_{\rm int}[\vec{s}]$. If they are the same, then our assumption that the intrinsic fitness landscape is accurately described by the maximum entropy Hamiltonian inferred from patient sequences is exactly correct. If they differ, we can evaluate what the differences are and determine how the inferred prevalence landscape relates to an intrinsic fitness landscape.

III. RESULTS

To aid visualization, we computed a 2D embedding of the intrinsic fitness landscape associated with $H_{\rm int}[\vec{s}]$, which charts the peaks and valleys of fitness in sequence space (see Appendix C). We first simulated our model in the absence of immune pressure ($\{b_i\} = 0$) for values of μ in the range 10^{-5}–10^{-2}/site/generation. In these simulations, the quasispecies within every host stays localized around the WT strain $\vec{s} = \vec{0}$ [Fig. 2(a) and Supplemental Fig. S8]. The population ensemble in this case is entirely composed of WT strains and mutations are never selected at the population level. The frequencies of single and double mutations in the population ensemble, $\langle s_i \rangle_{\rm dyn}$ and $\langle s_i s_j \rangle_{\rm dyn}$ are zero (unlike in the real MSA) and reveal no information about the correlation structure of the fitness landscape.

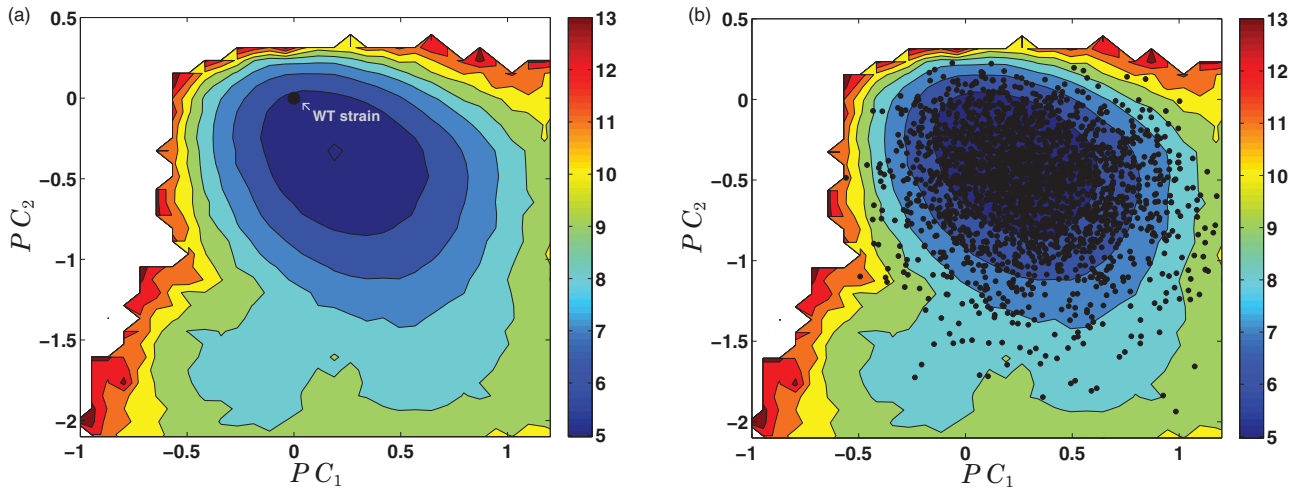

FIG. 2. (Color) Immune pressure facilitates exploration of the virus in sequence space. Viral exploration of sequence space in our simulations is depicted using a lower dimensional representation of the intrinsic fitness landscape of the protein p17, computed by applying principal component analysis (PCA) to sequences resulting from an equilibrium sampling of $H_{\rm int}[\vec{s}]$ (cf. Supplemental Notes 1 and 2). We only focus on the primary basin relevant to our simulations (cf. Supplemental Fig. S2). Different colors represent contours of the free energy computed as $A(x,y) = -\log P(x,y)$, where $P(x,y)$ is the normalized density of sequences at point (x,y) on the $[PC_1, PC_2]$ plane. Low values correspond to regions of high fitness. (a) In the absence of immune pressure ($\{b_i^\kappa\} = 0$), the population ensemble extracted from our simulations consists of only WT sequences, which are represented by a single bullet (•), located at (0,0). (b) In the presence of immune pressure, the population ensemble consists of sequences that explore different parts of the landscape. Each bullet (•) represents the most frequent strain in a particular host.

This is because in the absence of any immune pressure, for reasonably large values of N_v typical of the chronic phase of infection during which virus samples are collected, selective forces dominate genetic drift and suppress the fixation of mutations that are deleterious to intrinsic fitness. This behavior persists until values of μ beyond which selective adaptation is ineffective and the quasispecies collapse [15] because of the rapid accumulation of deleterious mutations (cf. Supplemental Note 5).

The presence of an immune response changes the "effective" fitness landscape by favoring mutations that enable the virus to escape immune pressure despite lowering intrinsic fitness. This causes the quasispecies to shift away from the WT strain, and the viral quasispecies sample different parts of sequence space in different hosts because of the great diversity of human genes associated with T cell immune responses (see Fig. 2(b) and Supplemental Fig. S9). Primary mutations that enable escape also influence the emergence of secondary mutations at sites which are not directly targeted by the immune pressure but are coupled to the primary mutations to compensate the incurred fitness cost [7,9,22]. At the population level, Fig. 2(b) shows that due to immune pressure, the consensus sequences from different hosts explore diverse regions of the fitness landscape. Thus, the immune pressure imposed by different hosts acts like a "higher temperature" that facilitates sampling of sequence space.

The exploration of sequence space is sensitive to the value of the mutation rate μ (cf. Fig. 3). There is an intermediate

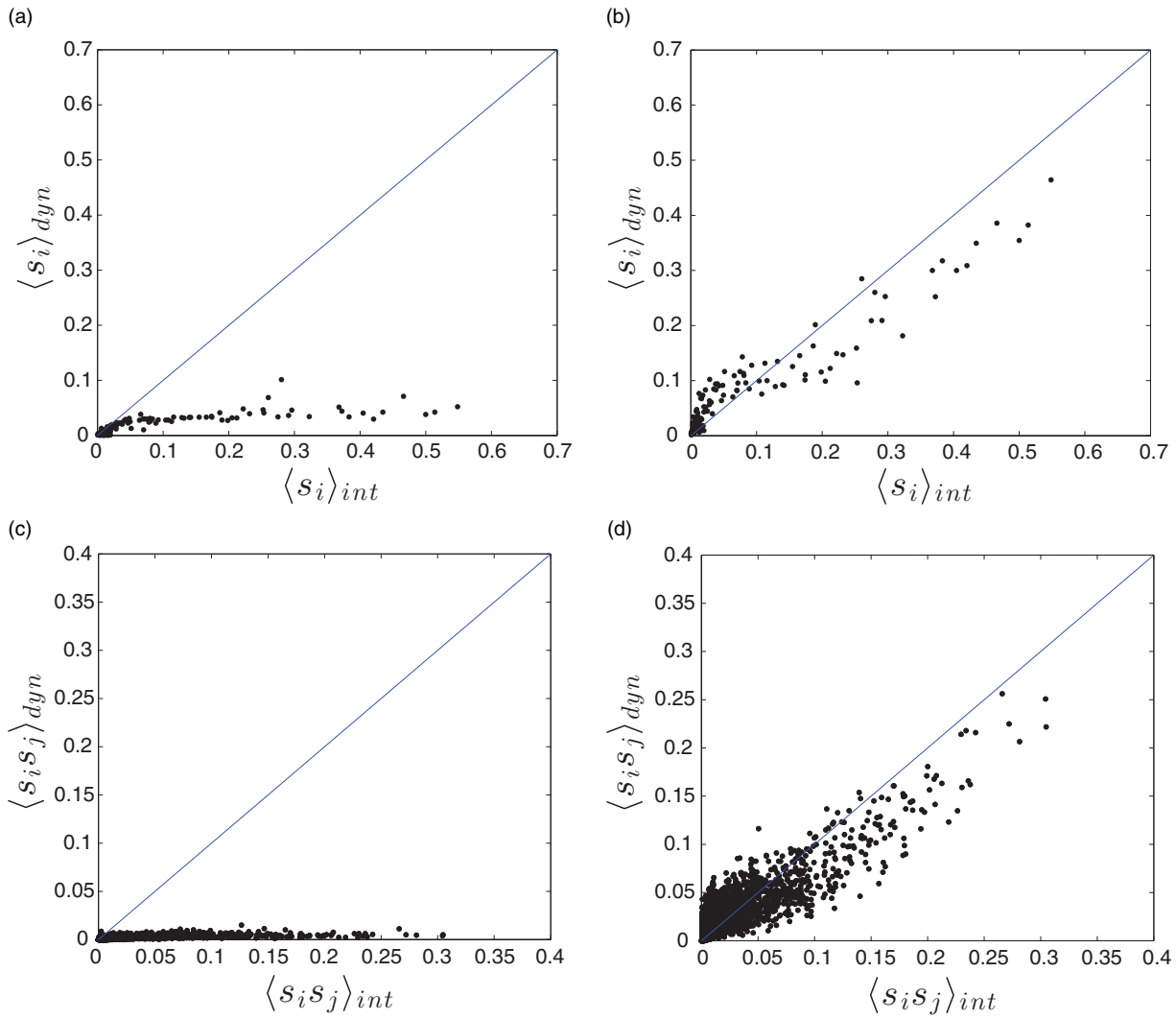

FIG. 3. (Color online) Comparison of one- and two-body mutational probabilities at low and intermediate mutation rates.(a), (c) $\mu = 5 \times 10^{-5}$/site/generation, $N_v = 15\,000$. The one- and two-body mutational probabilities (average per sequence) computed from the population ensemble resulting from our simulations, $\langle s_i \rangle_{\rm dyn}$ and $\langle s_i s_j \rangle_{\rm dyn}$, are compared with their counterparts, $\langle s_i \rangle_{\rm int}$ and $\langle s_i s_j \rangle_{\rm int}$, both computed from an equilibrium sampling of the intrinsic fitness Hamiltonian $H_{\rm int}[\vec{s}]$, which agree with values computed from sequences used to infer the maximum entropy model. At small mutation rates, escape mutations are rarely sampled. The viral quasispecies within each host stays frozen near the ground state [cf. Fig. 2(a)] and mutations are not selected at the population level, resulting in $\langle s_i \rangle_{\rm dyn}$, $\langle s_i s_j \rangle_{\rm dyn} \approx 0$. (b), (d) For intermediate mutation rates $\mu \in (10^{-4}, 10^{-2})$ (here, $\mu = 5 \times 10^{-3}$/site/generation and $N_v = 15\,000$), we find that immune selection leads to the accumulation of mutations across the viral proteome, and at the population level the one- and two-body mutational probabilities $\langle s_i \rangle_{\rm dyn}$ and $\langle s_i s_j \rangle_{\rm dyn}$ correlate monotonically with their counterparts $\langle s_i \rangle_{\rm int}$ and $\langle s_i s_j \rangle_{\rm int}$. At higher mutation rates ($\mu > 10^{-2}$/site/generation), the quasispecies within most hosts become unable to survive selection as deleterious mutations are rapidly accumulated.

range $\mu \in (10^{-4}, 10^{-2})$ where immune selection is stable and favors viral adaptation. This range is higher than the mutation rate of HIV ($\sim 10^{-4}$) when measured in units of per amino acid site per replication cycle. But, this is reasonable as each generation in our simulations corresponds to a number of replication cycles for reasons described in Appendix B.

In this intermediate range of μ, the marginal single-, two-, and three-site probabilities $\langle s_i \rangle_{\rm dyn}$, $\langle s_i s_j \rangle_{\rm dyn}$, and $\langle s_i s_j s_k \rangle_{\rm dyn}$ are monotonically correlated with their intrinsic fitness counterparts $\langle s_i \rangle_{\rm int}$, $\langle s_i s_j \rangle_{\rm int}$, and $\langle s_i s_j s_k \rangle_{\rm int}$ [Figs. 3(b), 3(d), and Supplemental Fig. S10). Taken together, our results suggest that immune pressure plays a necessary role in facilitating exploration of sequence space so that the viral quasispecies sample the fitness landscape, and furthermore, the correlation structure of the prevailing consensus strains in the population ensemble is monotonically related to the correlations that characterize mutant strains selected according to intrinsic replicative fitness.

IV. VARIATIONAL THEORY

Our ultimate interest, however, is not just in correlation structure but in characterizing the relationship between the intrinsic fitness landscape of the virus and the prevalence landscape inferred from patient-derived sequences. Toward this goal we exploit a mapping by Leuthäusser [23] to describe nonequilibrium quasispecies evolution according to Eigen's equation [15]: Each evolutionary path in sequence space is denoted by $\Sigma = \{\vec{s^0}, \vec{s^1}, \vec{s^2}, \ldots \vec{s^n}\}$, where $\vec{s^\alpha}$ denotes a strain in generation α, and can be regarded as a configuration of an inhomogenous Ising model. Different generations $\vec{s^\alpha}$ realized in a particular evolutionary path correspond to sequentially arranged rows of this Ising system. The probability of a particular evolutionary path, Σ, is $p(\Sigma) \propto e^{-\mathcal{H}(\Sigma)}$, with the Hamiltonian

$$\mathcal{H}(\Sigma) = -J \sum_{\alpha=0}^{n-1} (\vec{1} - 2\vec{s^\alpha})(\vec{1} - 2\vec{s^{\alpha+1}}) + \sum_{\alpha=0}^{n} H^\alpha[\vec{s^\alpha}]. \quad (2)$$

The first term in Eq. (2) describes a coupling between the same site in successive generations, with $J = \frac{1}{2}\log(\frac{1-\mu}{\mu})$ (since μ is small, J is positive, preferring sites in successive generations to be the same [23]). This longitudinal coupling describes the phylogenetic relationship between strains in a population. Note that $\vec{s^0}$ is coupled only to its progeny and $\vec{s^n}$ is coupled only to its parent, while an intermediate strain $\vec{s^\alpha}$ is phylogenetically coupled to both its parent and progeny. The second term in $\mathcal{H}(\Sigma)$ describes the aggregate effective fitness of the strains in the trajectory, where each term in the sum is decomposed as

$$\begin{aligned} H^\alpha[\vec{s^\alpha}] &= H_{\rm int}[\vec{s^\alpha}] + I^\alpha[\vec{s^\alpha}] \\ &= \sum_{i<j=1}^{N} J_{ij} s_i^\alpha s_j^\alpha + \sum_{i=1}^{N} h_i s_i^\alpha - \sum_i b_i^\alpha s_i^\alpha. \end{aligned} \quad (3)$$

$H_{\rm int}$ (parameterized by $\{J_{ij}, h_i\}$) is α independent, while the immune pressure in generation α is described by the fields $\{b_i^\alpha\}$.

The nonequilibrium dynamics captured by Eqs. (2) and (3), while not identical to our simulations, contain the important elements of phylogeny and immune pressure in different hosts. Instead of numerically sampling configurations with normalized probability $p(\Sigma) = e^{-\mathcal{H}(\Sigma)}/Z$ (which we expect will lead to results consistent with our simulations), we developed analytical approximations. The prevalence landscape in Eq. (1), inferred from sequences extracted from patients at a multitude of different times, makes no reference to phylogeny or immune pressure. Thus, we asked how well a "phylogeny and immune pressure independent" probability of the form $p_T(\Sigma) = (\prod_\alpha e^{-H_T^\alpha})/Z_T$ can approximate $p(\Sigma)$.

In particular, we chose a trial Hamiltonian of the same form as the inferred Hamiltonian in Eq. (1), $H_T^\alpha(\{K_{ij}^\alpha, a_i^\alpha\}, \vec{s^\alpha}) = \sum_{i<j=1}^{N} K_{ij}^\alpha s_i^\alpha s_j^\alpha + \sum_i a_i^\alpha s_i^\alpha$, and approximate $\mathcal{H}(\Sigma)$ in Eq. (2) as $\mathcal{H}_T = \sum_\alpha H_T^\alpha(\{K_{ij}^\alpha, a_i^\alpha\}, \vec{s^\alpha})$. We then variationally estimated the α-dependent parameters $\{K_{ij}^\alpha, a_i^\alpha\}$, which best approximate Eq. (2). These parameters can be estimated through the Gibbs-Feynman-Bogoliobov variational bound [24]

$$\ln Z \geqslant \ln Z_T - \langle \mathcal{H} - \mathcal{H}_T \rangle_T. \quad (4)$$

Extremizing the bound through variations of the parameters $\{K_{ij}^\alpha\}$ and $\{a_i^\alpha\}$ leads to the self-consistent mean-field relations for the parameters of the variational Hamiltonian (cf. Appendix D). The optimal field and coupling constants governing the Hamiltonian for a strain in the "bulk" ($0 < \alpha < n$) are

$$a_i^\alpha = h_i + 4J\left(1 - \left\langle s_i^{\alpha-1}\right\rangle_T - \left\langle s_i^{\alpha+1}\right\rangle_T\right) - b_i^\alpha, \quad (5)$$

$$K_{ij}^\alpha = J_{ij}. \quad (6)$$

Within this mapping, however, the bulk layers are not "observable" quantities, and the structure of the quasispecies after n generations are given by the statistics of the surface layer at $\alpha = n$ [25]. The parameters of the variational Hamiltonian for this layer are as follows (cf. Appendix D):

$$a_i^n = h_i + 2J\left(1 - 2\left\langle s_i^n \right\rangle_T\right) - b_i^n, \quad (7)$$

$$K_{ij}^n = J_{ij}, \quad (8)$$

where we have additionally assumed that $\langle s_i^{n-1} \rangle_T$ is well-approximated by $\langle s_i^n \rangle_T$ the mutational probability at the surface layer. This is reasonable if the number of generations is large and the replicative fidelity is high (see Appendix D). Note that for both the bulk and the surface, the effects of phylogeny and immune pressure appear only through the onsite fields.

Within the variational approximation then, the prevalence of a strain $\vec{s}$ after n generations of quasispecies evolution is completely determined by the parameters of the surface layer Hamiltonian $H_T^n[\vec{s}]$. The prevalence of $\vec{s}$ as encoded by the Hamiltonian $H_0[\vec{s}]$ in Eq. (1), however, reflects surface samples obtained from evolutionary trajectories of varying number of generations. These viral sequences have been sampled from genetically diverse hosts in whom the virus has evolved for different times. To mimic this, we average the Hamiltonian $H_T^n(\vec{s})$ over a large number (M) of realizations, leading to

$$H_T[\vec{s}] = \sum_{i<j=1}^{N} J_{ij} s_i s_j + \sum_{i=1}^{N} [h_i + 2J(1 - 2\langle s_i \rangle_T) - \bar{b_i}] s_i, \quad (9)$$

where $\bar{b_i} = \frac{1}{M}\sum_\alpha b_i^\alpha$ is the average immune pressure, and $\langle s_i\rangle_T = \overline{\langle s_i^n\rangle_T}$, the mutational probability at the surface layer of the Ising system, averaged over evolutionary trajectories containing different number of generations.

Comparing the parameters in $H_T[\vec{s}]$ to that in $H_{\rm int}[\vec{s}]$, we see that while the coupling constant between sites remains unmodified, the effective field at a particular site is changed from h_i by two competing contributions. The first term is self-consistently related to $\langle s_i\rangle_T$, the average single-site mutation probability at site i under Hamiltonian $H_T[\vec{s}]$. For $\langle s_i\rangle_T < 1/2$, which is true for $J > 0$, the first term represents an increase in the field at site i that disfavors mutation. This is because replicative fidelity disfavors sampling of sequence space since one mutant strain must be the progeny of another one.

In the absence of immune pressure, this "phylogenetic coupling" term favors freezing into the ground state ($\vec{s} = \vec{0}$), accounting for the localization of quasispecies in the vicinity of the intrinsically fittest sequence. The immune response counters this effect, bringing the effective field a_i closer to the intrinsic field h_i and drives the statistics of mutations closer to those governed by intrinsic fitness.

Armed with this variational approximation, we can now ask if $H_T[\vec{s}]$ preserves the fitness ranks of different viral strains as encoded in $H_{\rm int}[\vec{s}]$. The values of $\{\bar{b_i}\}$ are easily obtained from the simulations; and $\langle s_i\rangle_T$ are approximated by $\langle s_i\rangle_{\rm dyn}$ from our simulations (calculating $\langle s_i\rangle_T$ self-consistently is not practical given the complexity of H_T). This assumes that the variational Hamiltonian $\mathcal{H}_T$ is a reasonable approximation of true quasispecies evolution according to Eigen's equation and that the latter (where the number of strains is unbounded) is a reasonable facsimile of our simulations (where the quasispecies population is bounded). It must be recognized that fluctuations play an important role in finite populations [26,27]. In our simulations, however, the product of the population size and mutation rate is large ($N_v\mu > 1$), and in this regime theoretical and computational studies of finite populations (relevant for RNA viruses) have shown consistency with the predictions of quasispecies theory (reviewed in Ref. [28]). Also, Dixit *et al.* [29] recently proved formally for a class of finite population evolution models (similar to the one we have considered in this work) that as the population size increases, the stationary distribution of genotypes converges to the distribution predicted by Eigen's quasispecies model [30].

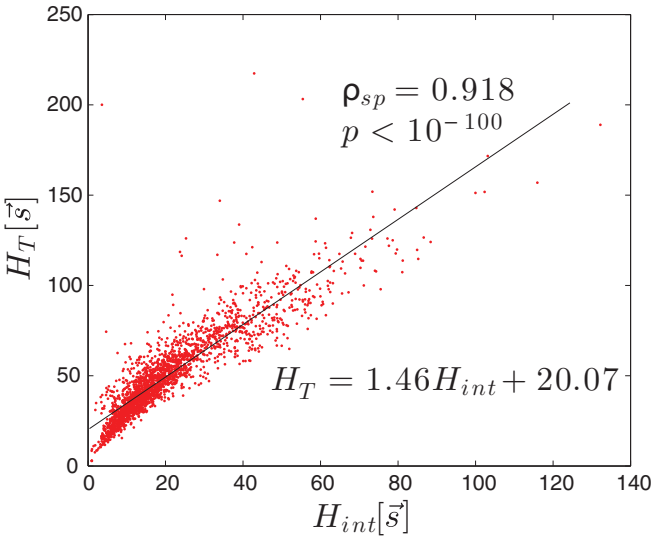

FIG. 4. (Color) Variational analysis using the Feynman bound predicts that the prevalence landscape is correlated with the fitness landscape. A numerical comparison between the variational estimate of the prevalence Hamiltonian $H_T[\vec{s}]$ [Eq. (9)] to the intrinsic Hamiltonian $H_{\rm int}[\vec{s}]$ for the case for 2500 subtype B p17 sequences. Both $H_T[\vec{s}]$ and $H_{\rm int}[\vec{s}]$ have units of dimensionless energy. $H_T[\vec{s}]$ is a variational estimate that is related to the logarithm of the prevalence of the strain $\vec{s}$, whose relation to $H_{\rm int}[\vec{s}]$ is corrupted by phylogeny and immune pressure. The parameters $\mu = 5\times10^{-3}$/site/generation, $n_{\rm max} = 6$, $N_v = 15\,000$ were used in the simulations. Each point on the plot corresponds to one p17 sequence. The sequences were downloaded from the Los Alamos sequence database [12] and converted to the binary code. $\rho_{\rm sp}$ is the standard rank correlation computed from Spearman's test [31] and p is the corresponding significance value. The line is computed from fitting an ordinary least squares regression model.

In Fig. 4, we plot $H_T[\vec{s}]$ versus $H_{\rm int}[\vec{s}]$ for 2474 subtype B HIV-1 strains extracted from a public database [12], after converting them to the binary code. A Spearman rank test [31] shows that the order of ranking is preserved with very high statistical accuracy for most strains ($\rho = 0.918$, $p < 10^{-100}$). Thus, at least within a mean-field approximation, prevalence landscapes inferred from patient-derived virus protein sequences preserve the rank order of intrinsic replicative fitnesses of mutant virus strains.

V. DISCUSSION

The underlying reason for this result may be simple. Because of the great diversity of genes that determine the immune response, individual sites in viral proteins are targeted by a small fraction of infected patients (see Supplemental Note 4). Furthermore, clinical data show that when the immune pressure in a particular host results in escape mutations, and the mutated virus is transmitted to another host who does not target the mutated sites, the virus rapidly reverts to WT in these regions [32]. Within a given host, the magnitude of immune pressure at particular sites (b_i^α) is large enough to drive exploration of sequence space. But this effect is present at any site only in a small fraction of hosts and acts as a perturbation ($\bar{b_i}$) when averaged over many consensus sequences. Therefore, although the immune pressure imposed by genetically diverse patients enables exploration of sequence space by modifying the fitness landscape [Fig. 2(b)], using a sufficient number of sequences ensures that the inferred prevalence model preserves the rank order of the intrinsic fitnesses of mutant viral strains.

This should, however, only be true if we compare sequences that are not phylogenetically distant. The effects of phylogeny, immune pressure, and intrinsic fitness are concatenated in the parameters that define our inferred fitness landscape [Eq. (9)]. For reasons discussed above, the immune pressure is a critical but perturbative field. The effect of phylogeny can, however, be quite strong when comparing phylogenetically distal strains. Because the infection in a population is initiated by a "founder strain," the relatively high fidelity of replication ($\mu \ll 0.5$)

in each cycle limits the rate at which the sequence space of the virus is explored in the following generations, favoring appearance of mutants that are proximal in sequence space to the parent strain. Therefore, although a strain with a large number of mutations might have a high intrinsic fitness, it is less likely to appear because of the mutational distance that separates it from the parent strain. For such strains the "phylogenetic term" in Eq. (9) increases the disparity between their prevalence in a population and their intrinsic fitness. In other words, because of replicative fidelity, a phylogenetically distant strain is less likely to be prevalent than one that is phylogenetically closer, even if they are of comparable fitness. The excellent agreement between experimental measurements of replicative fitness and the inferred prevalence landscape described by Ferguson *et al*. [9] may reflect the fact that different strains were phylogenetically proximal. This is also likely the case for strains used to construct Fig. 4. From a practical standpoint of using the fitness landscape for immunogen design, this issue presents little difficulty as a vaccine-induced immune response is unlikely to generate mutants that are phylogenetically distal. But, care needs to be taken in using the inferred prevalence landscape when comparing *in vitro* fitness measurements of phylogenetically distant strains. This is because our estimate of the correction due to replicative fidelity and phylogeny in Eq. (9) is not expected to be quantitatively correct.

Spin-glass models have been employed previously to explore the consequences of original antigenic sin [33] and measures of antigenic distance on influenza vaccination [34]. In our context, the possibility of using sequence data to infer such fitness landscapes of diverse viruses can accelerate the rational design of immunogens from sequence data, which may be able to induce potent and protective immune responses in humans against infectious diseases. Taken together, our results, and those in Ferguson *et al*. [9] show that maximum entropy models inferred from viral protein sequences sampled from patients can faithfully represent the intrinsic fitness landscape for phylogenetically related strains. Further work needs to be done to develop general methods for deconvoluting the effects of phylogeny and intrinsic fitness in inferred landscapes in order to reliably predict the fitness of strains regardless of phylogenetic distance. With the rapid expansion of available genomic data a promising and efficient route to rational immunogen design is thus suggested.

ACKNOWLEDGMENTS

This research was supported by the Ragon Institute of MIT, MGH, and Harvard, a NIH Director's Pioneer Award (A.K.C.), and a Poitras pre-doctoral fellowship (K.S.). Fruitful discussions with Dr. T. Butler are gratefully acknowledged.

APPENDIX A: MODELING IMMUNE PRESSURE

The immune response in a particular host is chosen to randomly target k sites in the protein where k is a random integer ranging from 0 to $n_{\max} \ll N$, based on clinical evidence within Caucasian Americans that the p17 protein is targeted predominantly by T-cells, and that a given protein site is expected to be targeted by a very small fraction of individuals in a population (Supplemental Note 4 and Supplemental Fig. S5). For the simulations reported in the main text, we employ $n_{\max} = 6$. The targeted sites $\alpha_1, \alpha_2, \ldots, \alpha_k$ are chosen randomly from protein sites $1, 2, \ldots, N$ in each host without bias, thereby mimicking the highly polymorphic nature of genes encoding the cell machinery that presents viral protein fragments, which are recognized by T cells. Thus, each individual is likely to target different sites compared to other persons. We assume that the parameters $\{b_{\alpha_i}\}_{i=1}^{k}$ are independent random variables, drawn from the same Gaussian distribution of mean $\bar{h}$ and variance σ_h^2 as determined by the intrinsic fitness parameters $\{h_i\}$. Changing the parameters of this distribution to increase the magnitude of the immune fields does not change the main conclusions of our work (see Supplemental Note 5.2) as long as the immune pressure in any given host targets only a small fraction of sites ($<10\%$) and that the typical magnitude of $\{b_i\}$ do not greatly exceed $\bar{h}$.

Incorporating host-specific effects as a linear term $-\sum_{i=1}^{N} b_i s_i$ constitutes the simplest model of T-cell-mediated immune pressure, in contrast to models where T-cell dynamics are treated explicitly [35,36], which, however, do not consider viral strains at proteomic resolution. T cells impose immune pressure on short viral peptides ($\sim$10 amino acids long). In our simulations, individual amino acids in the protein, corresponding to residues in these peptides, are under immune pressure. The scenario is analogous to an Ising system in an external magnetic field. Including higher-order terms would imply the presence of correlations within the immune-mediated targeting of different residues. Indeed, T-cell-mediated immune pressure can be abrogated by mutation in a single residue within the peptide or in the flanking residues, thus "releasing" the immune pressure from all residues. As a consequence, short-range correlations over the length of the targeted peptide are present, but two well-separated peptides are targeted largely independently of each other because of the large diversity in the genes that determine which peptides are targeted in a host. While it would be straightforward to make peptides rather than individual residues the loci of immune pressure in our simulations, we do not expect our qualitative results to change. So long as the T-cell response in different hosts targets diverse regions within viral proteins (a consequence of the highly polymorphic nature of the underlying genes), the resulting selective pressure will drive exploration of the virus to different regions of sequence space.

APPENDIX B: QUASISPECIES SIMULATIONS

Within each host (parameterized by immune pressure $\{b_i\}$) the viral quasispecies evolve for τ_S generations following infection with N_v copies of a "founder" strain. Each generation is composed of the following steps,

(1) *Mutation:* For each viral strain $\vec{s}$ within the quasispecies, every site i is mutated with probability μ. In the binary representation, this amounts to the operation $s_i \to 1 - s_i$.

(2) *Prescreening:* Eliminate sequences that escape the region in the reduced free energy landscape defined by subtype B sequences (see Supplemental Fig. S2). (This step is rarely necessary unless the mutation rate μ exceeds 10^{-2}).

(3) *Selection:* Strain $\vec{s}$ is selected to survive with probability $p_s(\vec{s}) = \frac{e^{-H[\vec{s}]}}{1+e^{-H[\vec{s}]}}$. Here, $H[\vec{s}] = H_{\rm int}[\vec{s}] - \sum_{i=1}^{N} b_i s_i$ and parameters $\{b_i\}$ are host-specific.

(4) *Replenishment:* The numbers of surviving strains is randomly resampled with replacement to replenish the viral population to size N_v.

The survival probability in step 3 above has a functional form consistent to the "death probability" employed in Amitrano *et al.* [37]. Assuming $f_{\vec{s}} \sim e^{-H[\vec{s}]}$ is the fitness of strain $\vec{s}$, the survival probability has a simple interpretation. In each generation a strain $\vec{s}$ is compared with a WT strain and is elected to survive with probability $\frac{f_{\vec{s}}}{f_{\rm WT}+f_{\vec{s}}}$, where $f_{\rm WT} \sim e^{-H_{\rm WT}} = 1$. Alternatively, one can envision a selection rule where a strain $\vec{s}$ is compared with the average strain in the current quasispecies and is elected to survive with probability $\frac{f_{\vec{s}}}{\bar{f}_{\vec{s}}+f_{\vec{s}}}$, where $\bar{f}_{\vec{s}} \sim \langle e^{-H[\vec{s}]}\rangle$ is the average fitness of the quasispecies. The latter selection rule produces results that are in qualitative agreement with the former selection rule within the range of parameters we have explored (data not shown).

1. A note on time scales and the mutation rate

In these idealized simulations that mirror the Wright-Fisher process [19], each "generation" combines the processes of mutation of extant viruses and selection of viruses according to their effective fitness values in the presence of immune pressure. Biologically, however, these two processes are disparate and act on different entities and time scales—while mutation occurs on the genome of an infecting virus during reverse transcription within the cytoplasm of the infected cell, selection occurs at the level of infected cells (not considered explicitly in our simulations), which are recognized and killed by cytotoxic T cells.

Mathematical models fitted to viral kinetic data have suggested that the mean lifetime of individual virions is much smaller (∼0.3 days) than those of infected cells (∼2.2 days) [38]. Therefore, if infection is to be sustained, then at least one of the virions produced from an infected cell during the latter's lifetime must lead to a new infection in a healthy cell. Infection of a healthy cell results in the commencement of a replication cycle, wherein the genome of the virus is reverse-transcribed, resulting in mutations. Since a large fraction of mutations are deleterious even for a virus like HIV, every "attempted" infection does not necessarily lead to "productive" infection. Combining this with the fact that viral lifetimes are smaller than those of infected cells, it follows that multiple progeny virions from an infected cell must make "infection attempts" within the latter's lifetime to ensure successful propagation of the infection. At steady state, the rate at which new productively infected cells are produced must equal the rate at which they are cleared by the immune system.

In summary, if we assume that each generation in the Wright-Fisher introduced in Appendix B process corresponds to a timescale governed by selection, which is set by the lifetime of infected cells, we must correct for the fact that multiple infection and viral replication events occur in this period. We account for this in the simplest possible manner by increasing the mutation rate in the *Mutation* step above, since this effectively enables viruses in our simulations to explore larger combinations of mutations in a single "generation."

APPENDIX C: LANDSCAPE VISUALIZATION

Mutational states of the protein were sampled according to $H_{\rm int}[\vec{s}]$ by the Metropolis Monte-Carlo (MC) algorithm [39] to generate an "equilibrium" ensemble of approximately 10^6 sequences. We applied principal component analysis (PCA) [40] to the covariance matrix corresponding to double mutations in this ensemble of sequences (see Supplemental Note 1). We computed the projection of each sequence in the equilibrium ensemble onto this space along the top two principal components (PCs). Using an appropriately sized square mesh, the density of sequences at different locations in this 2D embedding was converted to the analog of free-energy contours in statistical mechanics using the relation

$$A(x,y) = -\log P(x,y),$$

where (x,y) is the center of a cell in the $[PC_1, PC_2]$ plane and $P(x, y)$ is the sample probability of a sequence in the equilibrium ensemble occupying this point.

In the projection along PC_1-PC_2, the landscape exhibits three high fitness (or low "free-energy") peaks (see Supplemental Fig. S2). But two of these peaks are unexplored by the subtype B sequences in the MSA that were used to parametrize $H_{\rm int}[\vec{s}]$ (Supplemental Fig. S3) and represent extrapolations of the fitted nonlinear model. To focus on the question of how the inferred prevalence of subtype B HIV strains in the population relates to the intrinsic fitness landscape, we restricted quasispecies sequences in our computer simulations to lie in the region corresponding to observed sequences by placing reflecting boundaries on the PC_1-PC_2 space as described in Supplemental Fig. S2.

APPENDIX D: VARIATIONAL CALCULATIONS

In variational mean-field theory, the parameters of the approximate Hamiltonian are obtained by maximizing the right-hand side of the Gibbs-Feynman-Bogoliobov bound [24] (see Eq. (4)) with respect to the variational parameters $\{K_{ij}^{\alpha}, a_i^{\alpha}\}$. The stationarity conditions are

$$\begin{aligned} \frac{\partial \ln Z_T}{\partial K_{ij}^{\alpha}} - \frac{\partial}{\partial K_{ij}^{\alpha}} \langle \mathcal{H} - \mathcal{H}_T \rangle_T = 0, \\ \frac{\partial \ln Z_T}{\partial a_i^{\alpha}} - \frac{\partial}{\partial a_i^{\alpha}} \langle \mathcal{H} - \mathcal{H}_T \rangle_T = 0. \end{aligned} \tag{D1}$$

Here, $\mathcal{H}$ is the original Hamiltonian while $\mathcal{H}_T = \sum_{\alpha=0}^{n} H_T^{\alpha}(\{K_{ij}^{\alpha}, a_i^{\alpha}\}, \vec{s^{\alpha}})$ is the trial Hamiltonian with the form, $H_T^{\alpha}(\{K_{ij}^{\alpha}, a_i^{\alpha}\}, \vec{s^{\alpha}}) = \sum_{i<j=1}^{N} K_{ij}^{\alpha} s_i^{\alpha} s_j^{\alpha} + \sum_i a_i^{\alpha} s_i^{\alpha}$. The trial partition function is defined as $Z_T = \sum_{\{\vec{s^{\alpha}}\}} \prod_{\alpha=0}^{n} \exp(-H_T^{\alpha}[\{K_{ij}^{\alpha}, a_i^{\alpha}\}, \vec{s^{\alpha}}])$. As different generations are uncoupled from each other the sum and the product can be interchanged. Taking its logarithm,

$$\begin{aligned} \ln Z_T &= \sum_{\alpha} \ln \left(\sum_{\vec{s^{\alpha}}} \exp\left(-H_T\left[\left\{K_{ij}^{\alpha}, a_i^{\alpha}\right\}, \vec{s^{\alpha}}\right]\right) \right) \\ &= \sum_{\alpha} \ln Z_T^{\alpha}, \end{aligned} \tag{D2}$$

where Z_T^{α} is the partition function for a single generation in the evolutionary trajectory. Thus, Eq. (D1) can be further

simplified as

$$\frac{\partial \ln Z_T^\alpha}{\partial K_{ij}^\alpha} - \frac{\partial}{\partial K_{ij}^\alpha}\langle \mathcal{H} - \mathcal{H}_T\rangle_T = 0, \quad \frac{\partial \ln Z_T^\alpha}{\partial a_i^\alpha} - \frac{\partial}{\partial a_i^\alpha}\langle \mathcal{H} - \mathcal{H}_T\rangle_T = 0. \tag{D3}$$

As is well-known in equlibrium statistical mechanics, $\ln Z_T^\alpha$ is the equivalent of a scaled free energy and its derivatives with regards to the coupling constants and fields yield thermal averages of different quantities of interest. Thus, it can be shown that $\frac{\partial \ln Z_T^\alpha}{\partial K_{ij}^\alpha} = -\langle s_i^\alpha s_j^\alpha\rangle_T$ and $\frac{\partial \ln Z_T^\alpha}{\partial a_i^\alpha} = -\langle s_i^\alpha\rangle_T$. Substituting in Eq. (D3),

$$-\langle s_i^\alpha s_j^\alpha\rangle_T - \frac{\partial\langle\mathcal{H}\rangle_T}{\partial K_{ij}^\alpha} + \frac{\partial\langle\mathcal{H}_T\rangle_T}{\partial K_{ij}^\alpha} = 0, \quad -\langle s_i^\alpha\rangle_T - \frac{\partial\langle\mathcal{H}\rangle_T}{\partial a_i^\alpha} + \frac{\partial\langle\mathcal{H}_T\rangle_T}{\partial a_i^\alpha} = 0. \tag{D4}$$

Substituting expressions for $\langle\mathcal{H}\rangle_T$ (cf. main text) and $\langle\mathcal{H}_T\rangle_T$,

$$\begin{aligned}
&\cancel{-\langle s_i^\alpha s_j^\alpha\rangle_T} - \sum_{\substack{k<m\\ \beta=0,\ldots,n}} J_{km}\frac{\partial\langle s_k^\beta s_m^\beta\rangle_T}{\partial K_{ij}^\alpha} - \sum_{\substack{k\\ \beta=0,\ldots,n}} \left(h_k - b_k^\beta\right)\frac{\partial\langle s_k^\beta\rangle_T}{K_{ij}^\alpha} - 2J \sum_{\substack{k\\ \beta=0,\ldots,n-1}} \left\{\frac{\partial}{\partial K_{ij}^\alpha}\left(\langle s_k^\beta\rangle_T + \langle s_k^{\beta+1}\rangle_T\right)\right.\\
&\left. - 2\frac{\partial\langle s_k^\beta s_k^{\beta+1}\rangle_T}{\partial K_{ij}^\alpha}\right\} + \cancel{\langle s_i^\alpha s_j^\alpha\rangle_T} + \sum_{\substack{k<m\\ \beta=0,\ldots,n}} K_{km}^\beta\frac{\partial\langle s_k^\beta s_m^\beta\rangle_T}{\partial K_{ij}^\alpha} + \sum_{\substack{k\\ \beta=0,\ldots,n}} a_k^\beta\frac{\partial\langle s_k^\beta\rangle_T}{\partial K_{ij}^\alpha} = 0.\\
&\cancel{-\langle s_i^\alpha\rangle_T} - \sum_{\substack{k<m\\ \beta=0,\ldots,n}} J_{km}\frac{\partial\langle s_k^\beta s_m^\beta\rangle_T}{\partial a_i^\alpha} - \sum_{\substack{k\\ \beta=0,\ldots,n}} \left(h_k - b_k^\beta\right)\frac{\partial\langle s_k^\beta\rangle_T}{\partial a_i^\alpha} - 2J \sum_{\substack{k\\ \beta=0,\ldots,n-1}} \left\{\frac{\partial}{\partial a_i^\alpha}\left(\langle s_k^\beta\rangle_T + \langle s_k^{\beta+1}\rangle_T\right)\right.\\
&\left. - 2\frac{\partial\langle s_k^\beta s_k^{\beta+1}\rangle_T}{\partial a_i^\alpha}\right\} + \cancel{\langle s_i^\alpha\rangle_T} + \sum_{\substack{k\\ \beta=0,\ldots,n}} K_{km}^\beta\frac{\partial\langle s_k^\beta s_m^\beta\rangle_T}{\partial a_i^\alpha} + \sum_{\substack{k\\ \beta=0,\ldots,n}} a_k^\beta\frac{\partial\langle s_k^\beta\rangle_T}{a_i^\alpha} = 0.
\end{aligned} \tag{D5}$$

The derivatives encoding the phylogenetic (intergenerational) coupling can be simplified as follows:

$$\begin{aligned}
\sum_{\substack{k\\ \beta=0,\ldots,n-1}} \frac{\partial\langle s_k^\beta s_k^{\beta+1}\rangle_T}{K_{ij}^\alpha} &= \sum_{\substack{k\\ \beta=0,\ldots,n}} \left(\langle s_k^\beta\rangle_T\frac{\partial\langle s_k^{\beta+1}\rangle_T}{\partial K_{ij}^\alpha} + \langle s_k^{\beta+1}\rangle_T\frac{\partial\langle s_k^\beta\rangle_T}{\partial K_{ij}^\alpha}\right)\\
&= \sum_{\substack{k\\ \beta=1,\ldots,n}} \langle s_k^{\beta-1}\rangle_T\frac{\partial\langle s_k^\beta\rangle_T}{\partial K_{ij}^\alpha} + \sum_{\substack{k\\ \beta=0,\ldots,n-1}} \langle s_k^{\beta+1}\rangle_T\frac{\partial\langle s_k^\beta\rangle_T}{\partial K_{ij}^\alpha}\\
&= \sum_k \langle s_k^1\rangle_T\frac{\partial\langle s_k^0\rangle_T}{\partial K_{ij}^\alpha} + \sum_{\substack{k\\ \beta=1,\ldots,n-1}} \left(\langle s_k^{\beta-1}\rangle_T + \langle s_k^{\beta+1}\rangle_T\right)\frac{\partial\langle s_k^\beta\rangle_T}{\partial K_{ij}^\alpha} + \sum_k \langle s_k^{n-1}\rangle_T\frac{\partial\langle s_k^n\rangle_T}{\partial K_{ij}^\alpha},
\end{aligned} \tag{D6}$$

and similarly for the derivatives with respect to a_i^α's. Note that we have written separately the terms that correspond to the bulk ($0 < \alpha < n$) and the two surfaces ($\alpha = 0, n$). Equations (D6) and the corresponding equation for a_i^α can be substituted into equation (D5), while carefully separating the bulk and surface terms,

$$\begin{aligned}
&\sum_{k<m}\left\{K_{km}^0 - J_{km}\right\}\frac{\partial\langle s_k^0 s_m^0\rangle_T}{\partial K_{ij}^\alpha} + \sum_k\left\{a_k^0 - h_k + b_k^0 - 4J\left(\frac{1}{2} - \langle s_k^1\rangle_T\right)\right\}\frac{\partial\langle s_k^0\rangle_T}{\partial K_{ij}^\alpha}\\
&+ \sum_{\substack{k<m\\ \beta=1,\ldots,n-1}}\left\{K_{km}^\beta - J_{km}\right\}\frac{\partial\langle s_k^\beta s_m^\beta\rangle_T}{\partial K_{ij}^\alpha} + \sum_{\substack{k\\ \beta=1,\ldots,n-1}}\left\{a_k^\beta - h_k + b_k^\beta - 4J\left(1 - \langle s_k^{\beta-1}\rangle_T - \langle s_k^{\beta+1}\rangle_T\right)\right\}\frac{\partial\langle s_k^\beta\rangle_T}{\partial K_{ij}^\alpha}\\
&+ \sum_{k<m}\left\{K_{km}^n - J_{km}\right\}\frac{\partial\langle s_k^n s_m^n\rangle_T}{\partial K_{ij}^\alpha} + \sum_k\left\{a_k^n - h_k + b_k^n - 4J\left(\frac{1}{2} - \langle s_k^{n-1}\rangle_T\right)\right\}\frac{\partial\langle s_k^n\rangle_T}{\partial K_{ij}^\alpha} = 0.
\end{aligned} \tag{D7}$$

$$\sum_{k<m}\left\{K^0_{km}-J_{km}\right\}\frac{\partial\langle s^0_k s^0_m\rangle_T}{\partial a^\alpha_i}+\sum_k\left\{a^0_k-h_k+b^0_k-4J\left(\frac{1}{2}-\langle s^1_k\rangle_T\right)\right\}\frac{\partial\langle s^0_k\rangle_T}{\partial a^\alpha_i}$$
$$+\sum_{\substack{k<m\\ \beta=1,\ldots,n-1}}\left\{K^\beta_{km}-J_{km}\right\}\frac{\partial\langle s^\beta_k s^\beta_m\rangle_T}{\partial a^\alpha_i}+\sum_{\substack{k\\ \beta=1,\ldots,n-1}}\left\{a^\beta_k-h_k+b^\beta_k-4J\left(1-\langle s^{\beta-1}_k\rangle_T-\langle s^{\beta+1}_k\rangle_T\right)\right\}\frac{\partial\langle s^\beta_k\rangle_T}{\partial a^\alpha_i}$$
$$+\sum_{k<m}\left\{K^n_{km}-J_{km}\right\}\frac{\partial\langle s^n_k s^n_m\rangle_T}{\partial a^\alpha_i}+\sum_k\left\{a^n_k-h_k+b^n_k-4J\left(\frac{1}{2}-\langle s^{n-1}_k\rangle_T\right)\right\}\frac{\partial\langle s^n_k\rangle_T}{\partial a^\alpha_i}=0. \quad \text{(D8)}$$

Equations (D7) and (D8) are satisfied if the coefficients of the derivatives are set identically to zero and we obtain Eqs. (5)–(8). We expect the profile of the expectation values $\langle s^\beta_i\rangle_T$ to vary gradually from a surface value to the bulk value over a surface correlation length ξ. Due to the strength of the phylogenetic coupling we expect a variation over the span of several generations, such that $\langle s^{n-1}_i\rangle_T \approx \langle s^n_i\rangle_T$ (see, e.g., Ref. [41] for mean-field calculations relevant to a uniform spin model close to a surface).

[1] B. Gaschen *et al.*, Science **296**, 2354 (2002).
[2] P. J. R. Goulder and D. I. Watkins, Nature Rev. Immunol. **4**, 630 (2004).
[3] B. D. Walker and D. R. Burton, Science **320**, 760 (2008).
[4] S. Létourneau *et al.*, PLoS One **2**, e984 (2007).
[5] A. Schneidewind *et al.*, J. Virol. **81**, 12382 (2007).
[6] R. M. Troyer *et al.*, PLoS Pathogens **5**, e1000365 (2009).
[7] V. Dahirel *et al.*, Proc. Natl. Acad. Sci. USA **108**, 11530 (2011).
[8] R. D. Kouyos, G. E. Leventhal, T. Hinkley, M. Haddad, J. M. Whitcomb, C. J. Petropoulos, and S. Bonhoeffer, PLoS Genetics **8**, e1002551 (2012).
[9] A. L. Ferguson, J. K. Mann, S. Omarjee, T. Ndungu, B. D. Walker, and A. K. Chakraborty, Immunity **38**, 606 (2013).
[10] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).
[11] G. Tkacik, E. Schneidman, M. J. Berry II, and W. Bialek, arXiv:0912.5409.
[12] http://www.hiv.lanl.gov/.
[13] K. Binder and A. P. Young, Rev. Mod. Phys. **58**, 801 (1986).
[14] G. Sella and A. E. Hirsh, Proc. Natl. Acad. Sci. USA **102**, 9541 (2005).
[15] M. Eigen, Naturwissenschaften **58**, 465 (1971).
[16] D. B. Saakian and C.-K. Hu, Proc. Natl. Acad. Sci. USA **103**, 4935 (2006).
[17] Z. L. Brumme *et al.*, PLoS One **4**, e6687 (2009).
[18] B. G. Turner and M. F. Summers., J. Mol. Biol. **285**, 1 (1999).
[19] I. M. Rouzine and J. M. Coffin, Proc. Natl. Acad. Sci. USA **96**, 10758 (1999).
[20] R. D. Kouyos, C. L. Althaus, and S. Bonhoeffer, Trends Microbiol. **14**, 507 (2006).
[21] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.88.062705 for Supplementary Notes and Supplementary Figures related to landscape visualization using PCA, parameter sensitivity analyses, comparison of mutational probabilities, and statistics of T-cell targeting.
[22] S. Bonhoeffer, C. Chappey, N. T. Parkin, J. M. Whitcomb, and C. J. Petropoulos, Science **306**, 1547 (2004).
[23] I. Leuthäusser, J. Stat. Phys. **48**, 343 (1987).
[24] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (MacGraw Hill, New York, 1965).
[25] P. Tarazona, Phys. Rev. A **45**, 6038 (1992).
[26] Y.-C. Zhang, Phys. Rev. E **55**, R3817 (1997).
[27] J.-M. Park, E. Munoz, and M. W. Deem, Phys. Rev. E **81**, 011902 (2010).
[28] C. O. Wilke, BMC Evolution. Biol. **5**, 44 (2005).
[29] N. M. Dixit, P. Srivastava, and N. K. Vishnoi, J. Comput. Biol. **19**, 1176 (2012).
[30] In Ref. [29], the authors simulate a finite population model called the "RSM" model, which is structurally similar to the model of intrahost evolution we have considered in this paper.
[31] G. W. Corder and D. I. Foreman, *Nonparametric Statistics for Non-statisticians: A Step-by-step Approach* (Wiley, New York, 2009).
[32] M. R. Henn *et al.*, PLoS Pathogens **8**, e1002529 (2012).
[33] M. W. Deem and H. Y. Lee, Phys. Rev. Lett. **91**, 068101 (2003).
[34] V. Gupta, D. J. Earl, and M. W. Deem, Vaccine **24**, 3881 (2006).
[35] C. L. Althaus and R. J. De Boer, PLoS Comput. Biol. **4**, e1000103 (2008).
[36] G. Wang and M. W. Deem, Phys. Rev. Lett. **97**, 188106 (2006).
[37] C. Amitrano, L. Peliti, and M. Saber, J. Mol. Evol. **29**, 513 (1989).
[38] A. S. Perelson, A. U. Neumann, M. Markowitz, J. M. Leonard, and D. D. Ho, Science **271**, 1582 (1996).
[39] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[40] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning* (Springer, Berlin, 2001).
[41] H. Li, M. Paczuski, M. Kardar, and K. Huang, Phys. Rev. B **44**, 8274 (1991).