PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

Cite this article: Mazzolini A, Mora T, Walczak AM. 2023 Inspecting the interaction between human immunodeficiency virus and the immune system through genetic turnover. *Phil. Trans. R. Soc. B* **378**: 20220056. https://doi.org/10.1098/rstb.2022.0056

Received: 25 July 2022
Accepted: 15 November 2022

One contribution of 13 to a theme issue 'Interdisciplinary approaches to predicting evolutionary biology'.

Subject Areas:
evolution, theoretical biology, genetics, immunology

Keywords:
viral–immune coevolution, repertoire sequencing, viral sequencing, population genetics, statistical analysis

Authors for correspondence:
Thierry Mora
e-mail: thierry.mora@phys.ens.fr
Aleksandra M. Walczak
e-mail: aleksandra.walczak@phys.ens.fr

†These authors contributed equally to the study.

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.6476964.

THE ROYAL SOCIETY PUBLISHING

Inspecting the interaction between human immunodeficiency virus and the immune system through genetic turnover

Andrea Mazzolini, Thierry Mora† and Aleksandra M. Walczak†

Laboratoire de physique de l'École normale supérieure, PSL Université, CNRS, Sorbonne Université and Université Paris Cité, 75005 Paris, France

AM, 0000-0003-3194-2052; AMW, 0000-0002-2686-5702

Chronic infections of the human immunodeficiency virus (HIV) create a very complex coevolutionary process, where the virus tries to escape the continuously adapting host immune system. Quantitative details of this process are largely unknown and could help in disease treatment and vaccine development. Here we study a longitudinal dataset of ten HIV-infected people, where both the B-cell receptors and the virus are deeply sequenced. We focus on simple measures of turnover, which quantify how much the composition of the viral strains and the immune repertoire change between time points. At the single-patient level, the viral–host turnover rates do not show any statistically significant correlation, however, they correlate if one increases the amount of statistics by aggregating the information across patients. We identify an anti-correlation: large changes in the viral pool composition come with small changes in the B-cell receptor repertoire. This result seems to contradict the naïve expectation that when the virus mutates quickly, the immune repertoire needs to change to keep up. However, a simple model of antagonistically evolving populations can explain this signal. If it is sampled at intervals comparable with the sweep time, one population has had time to sweep while the second cannot start a counter-sweep, leading to the observed anti-correlation.

This article is part of the theme issue 'Interdisciplinary approaches to predicting evolutionary biology'.

1. Introduction

The adaptive immune system has been shaped by evolution to provide an effective response against a practically infinite reservoir of pathogens. During an infection, B-cells undergo affinity maturation in lymph-node germinal centres [1,2]. This mechanism is a Darwinian evolutionary process, where B-cell receptors are subject to somatic hypermutations [3] and are selected depending on their ability to recognize an external pathogen. This increases the affinity of naïve B-cells against the pathogen up to 10–100 factors [4–6], generating memory and plasma B-cells.

During chronic infections of the human immunodeficiency virus (HIV), the immune response is dominated by the action of antibody-secreting plasma B-cells [7]. However, most of the time, this machinery is not enough to control or clear the virus. The reason can be identified in the extremely rapid evolution of the virus escaping immune adaptation [8–10] and the fact that regions of the viral structure sensitive to B-cell targeting are made inaccessible [11,12]. Nevertheless, an effective immune response can instead occur naturally in 10–20% of the patients, and it is related to the emergence of broadly neutralizing antibodies (bNAbs) [13–15]. This promising discovery is the basis of the search for an HIV vaccine [16–19].

The current picture is that of the two populations of HIV and B-cell repertoire undergoing rapid and complex antagonistic coevolution. A lot of effort has been put into quantitatively understanding evolutionary properties of these two populations, which are usually considered separately. For example,

previous work has studied the dynamics of HIV variants escaping the immune system [20–22], or diversity patterns and linkage equilibrium properties of the virus [23]. On the immune system side, a large body of work has been dedicated to studying the immune response to HIV and the emergence of bNAbs [24–26], as well as to characterize lineage evolution during the affinity maturation process, using high-throughput sequencing of B-cell repertoire [27,28]. Much less work focuses on the coupled evolutionary dynamics of the two populations. Coevolutionary work has typically been theoretical, with a general focus on bNAb generation [29–31]. The datasets used for the data-based studies contain sequences of either HIV or immune repertoires. Given the interacting nature of this coevolutionary process, genetic data of both the populations evolving in time would provide valuable information.

To our knowledge, only one dataset of this kind has been made public, where a portion of the HIV envelope gene and B-cell receptors have both been deeply sequenced in time for different patients [32]. We base our analysis on this dataset and define simple macroscopical observables for the evolution of the two populations, which quantify genetic turnover and selective pressure. We find that a few of those measures display temporal correlations between the viral population composition and the immune system, showing that the coevolutionary 'arms-race' leaves traces at the whole-population level. To make sense of these correlations, the second part of the article introduces population–genetics models with different levels of complexity. Through numerical and analytical analysis, these models show that the observed statistical patterns can emerge for a biologically reasonable set of parameters. These theoretical models can help to build intuition about how and under which conditions these correlations arise.

2. Methods

(a) Longitudinal data for human immunodeficiency virus and the immune repertoire

Our study is based on the dataset described in [32]. The samples originate from 10 HIV-infected male participants. For each individual we have 10–20 longitudinal samples, taken before administration of antiretroviral therapy. For most of the time points the viral genetic composition and the immune repertoire are tracked in parallel, see electronic supplementary material, figure S1 and section S1 for more information. Specifically, on the viral side, the C2–V3 region of the *env* gene is deeply sequenced. This gene is known to be a potential target of the antibody repertoire [33,34]. On the immune repertoire side, the samples correspond to the deep sequencing of the variable region of the immunoglobulin (Ig) heavy chain locus. We omitted patient 10 because of lack of HIV samples, as shown in electronic supplementary material, section S1.

We wrote a pipeline to download the dataset, assemble the HIV sequences and the immunoglobulin heavy chain clonotypes. The description of our pipeline is in electronic supplementary material, section S2. The public repository https://github.com/statbiophys/HIV_coevo.git contains all the scripts and the instructions necessary for running the pipeline and reproducing our results. The obtained dataset is composed of HIV samples having an average number of 1000 unique sequences and 250 000 total sequences. For the B-cell receptors there are 130 000 unique clonotypes and 210 000 total clonotypes. The Ig clonotypes have been clustered into lineages as described in electronic supplementary material, §S2D.

(b) Simple measures for quantifying evolutionary properties of human immunodeficiency virus and the immune repertoire

We ask whether the evolutionary dynamics of HIV are temporally correlated with those of the immune repertoire. To this end, we define and explore a few simple measures characterizing the evolution of the two populations.

On the viral side, most of these measures are based on how single nucleotide polymorphisms (SNPs) of the viral sequences change over time. Details about how the SNPs are tracked in data are discussed in electronic supplementary material, §S2E. As sketched in figure 1*a*, for a given position in the sequences and a given nucleotide, we can count its frequency and its abundance. The frequency is the number of unique sequences in which the nucleotide appears divided by the total number of sequences. The abundance is the sum of the sequence counts of all the sequences in which the letter appears. Therefore, the abundance contains the information related to the sequence counts, while the frequency does not. Since we do not know *a priori* if this information is important for the later analysis, we keep track of both quantities. These numbers can be computed for a given SNP, i (e.g. $i = (1, \mathrm{A})$ for an A at position 1) at different time points to create a trajectory $x_i(t)$ (figure 1*b*).

To quantify how much the genetic composition of a set of sequences varies between two timepoints t_1 and t_2, we add together the absolute difference of all the single SNP trajectories, as in figure 1*c*. We call this quantity *absolute change*:

$$\text{absolute change}\ (t_1, t_2) = \sum_i |x_i(t_2) - x_i(t_2)|. \tag{2.1}$$

The choice of the absolute value makes no distinction between an increase or a decrease in the trajectory of the same amount. This is because we are only interested in the magnitude of the change and not in its sign.

The absolute change can be applied to the HIV sequences of a patient across time points, both for SNP frequencies and SNP abundances. This defines the first two entries of table 1: 'HIV turnover fr/ab'. Since these measures compute how much new mutations spread in the population from one time point to the other, they can be interpreted as the genetic turnover.

In a similar way, these measures can also be applied to the SNPs of Ig repertoire (entries 'Ig turnover fr/ab' of table 2). However, in that case, SNPs have to be computed at the level of lineages, obtaining trajectories for SNP i in lineage l, $x_{i,l}(t)$. Note that for 'Ig turnover fr', the frequency is normalized by the number of sequences in the lineage, so that $\sum_{i\in l} x_{i,l}(t) = 1$. The absolute turnover then sums across all SNPs and lineages.

To quantify the fact that B-cell lineages themselves are subject to turnover, and rise and fall in time, we introduce a measure that computes the absolute change of lineage sizes. The size of a lineage is the sum of all its sequence abundances or frequencies (here normalized by their total counts in the sample, so that $\sum_i x_{i,l}(t) = 1$). Applying the formula for absolute change, equation (2.1), to these quantities defines the *Ig lineage turnover fr/ab* measures in table 2. We define two additional measures derived from the Ig lineage turnover abundances: one that includes only the top 10% changes in abundances in the computation of the absolute change (*Ig lineage large turnover*), and one with only the bottom 50% (*Ig lineage small turnover*). In general we expect that smaller absolute changes are more subject to noise and, in turn, carry less signal, which is verified by our correlation analysis.

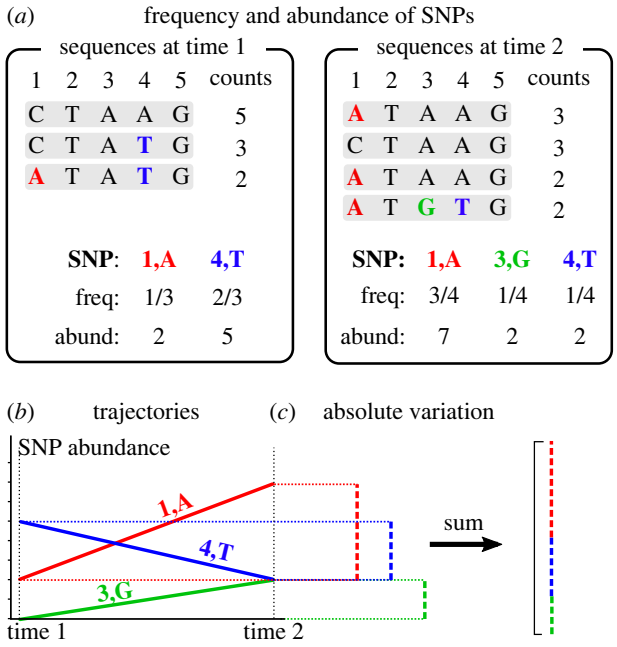

Figure 1. (*a*) Definition of SNP frequency and SNP abundance in a toy example. (*b*) Trajectories of SNP abundances. The values correspond to the example in (*a*). (*c*) Absolute-change computation of the trajectories. The dashed vertical lines are the different contributions to the variation given by each trajectory. They are summed together leading to the final absolute change. (Online version in colour.)

Finally, we define an estimate of the selective pressure on a population, closely related with the well-known measure of adaptation dN/dS [35,36]. To this end, we compute separately the absolute change of the non-synonymous SNPs and synonymous ones and we take their ratio:

$$\frac{\mathrm{dN}}{\mathrm{dS}}(t_1, t_2) = k\frac{\sum_{i\in\text{non-syn}} |x_i(t_2) - x_i(t_2)|}{\sum_{j\in\text{syn}} |x_j(t_2) - x_j(t_2)|}. \tag{2.2}$$

The coefficient k is the ratio of the probabilities of randomly generating synonymous and non-synonymous mutations from the reference sequence. This fixes the ratio to 1 in the case of uniform random mutations and a small mutation rate (see electronic supplementary material, §S2E). We computed this measure only for HIV and not for the immune system. The reason for this is that most of the lineages are small and it can happen that their synonymous change is zero, dS = 0, leading to several undefined values.

3. Results

(a) A single-patient analysis does not show any interaction

We will show that the coevolutionary interaction between HIV and the immune system leaves a trace in the dynamics of two populations. More precisely, some of the evolutionary measures defined for HIV, table 1, significantly temporally correlate with properties of the immune repertoire, table 2.

However, this signal is almost not visible at the single-patient level. For example, in figure 2, we compute the Spearman correlation between the *HIV turnover ab* and the *Ig lineage turnover ab* trajectories in the nine patients. Eight out of nine patients show an anti-correlation, but the signal is weak, with only three out of eight showing a *p*-value below 0.1, and only 1 (patient 4) showing a significant correlation, with $p = 0.004$ (which comes very close to the significance threshold of 0.05 after the Bonferroni correction). Nevertheless, this kind of analysis is not catching the fact that all these weak correlations point towards the same direction. Below we propose a statistical measure which shows that the correlation between these pairs of measures deviates significantly from a null model where no correlations exist.

Table 1. List of HIV evolutionary measures. ab, abundances; fr, frequencies; non-syn, non-synonymous; syn, synonymous.

short name	definition
HIV turnover fr	absolute change of SNP frequencies
HIV turnover ab	absolute change of SNP abundances
HIV dN/dS fr	ratio of the non-syn and syn absolute changes for SNP frequencies
HIV dN/dS ab	ratio of the non-syn and syn absolute changes for SNP abundances

(b) Statistical procedure for combining temporal correlations across patients

As we saw in §3(a) and figure 2, the correlations between the considered measures were negative for almost all the patients, but with a non-significant *p*-value, which led us to conclude that there is no significant correlation. However, it is very unlikely that, if the two measures are really independent, they generate a coherent positive (or negative) correlation across all patients. In the following we describe a procedure whose aim is to quantify this observation of co-change and 'integrate' the information contained in coherent correlations.

We graphically illustrate the statistical integration procedure in the cartoon of figure 3 using an artificial dataset composed of three patients. For each patient we consider a given pair of trajectories: one for a measure of the HIV turnover, taken from table 1 and one for the Ig turnover, table 2. We then calculate the Spearman correlation, ρ, between the two trajectories in each patient, resulting in a set of three coefficients. As discussed in the previous section, these coefficients are not significant but they all point towards the same sign (figure 2). These three coefficients, in general, cannot be compared with each other: performing random shuffles of the trajectories leads to null distributions that are specific to each patient. These null distributions approximately follow centred Gaussian laws, but with standard deviations that depend on the number of points n, σ_n. For instance, if two patients have the same Spearman coefficient but with different numbers of points n and thus a different width of the null distribution σ_n, the patient with the smaller σ_n will be more significant than the other. To make these coefficients comparable across patients, we re-scale them as $\hat{\rho} = \rho/\sigma_n$, so that when the $\hat{\rho}$ are equal, they also have the same significance. The next step of the procedure is to test if this set of re-scaled coefficients deviates *as a whole* significantly from the null uncorrelated scenario. To do so, we perform a one-sample Kolmogorov–Smirnov test against a normal distribution, i.e. the null distribution of all the $\hat{\rho}$. The obtained Kolmogorov–Smirnov *p*-value quantifies this significance.

Table 2. List of Ig-repertoire evolutionary measures. ab, abundances; fr, frequencies; non-syn, non-synonymous; syn, synonymous.

short name	definition
Ig turnover fr	absolute change of SNPs frequencies
Ig turnover ab	absolute change of SNPs abundances
Ig lineage turnover fr	absolute change of lineage frequencies
Ig lineage turnover ab	absolute change of lineage abundances
Ig lineage large	the largest 10% absolute change turnover of lineage abundances
Ig lineage small	the smallest 50% absolute change turnover of lineage abundances

(c) Human immunodeficiency virus turnover and Ig lineages turnover are significantly anti-correlated

To combine information across patients, we employ the statistical procedure explained in §3(b). For each pair of the HIV–Ig absolute-change measures, we compute a Kolmogorov–Smirnov *p*-value on the distribution of re-scaled correlation coefficients. The 24 *p*-values are shown in electronic supplementary material, figure S3*a*. We correct for multiple testing through a Benjamini–Hochberg test at a false discovery rate of 0.05. This selects three significant pairs of measures, which are shown in figure 4. This figure displays a scatter plot for each pair, which contains all the points of all the patients together (normalized as explained in the caption, and colour-coded by patient). For example the second panel displays all the points of the nine plots of figure 2. The histogram of Kolmogorov–Smirnov *p*-values is shown above each scatter plot. The first two significant pairs are the Ig lineage turnover and the HIV turnover. We find a correlation when all the lineage turnovers are considered (middle plot, *Ig lineage turnover ab*), or when only the largest lineage changes are taken (left plot, *Ig lineage large turnover*). The signal disappears if one considers only small changes, *Ig lineage small turnover*, electronic supplementary material, figure S3*a*, a measure that is probably more susceptible to noise. Interestingly, the correlation is negative, as can be seen from the top histograms of figure 4, and from the grey density of points of the scatter plots. The third significant pair correlates lineage turnover versus the dN/dS measure of HIV. This group of re-scaled correlations points again towards an anti-correlation.

The fact that the correlation between turnovers is negative seems counterintuitive. It means that if the composition of the viral population is changing quickly, the immune repertoire abundances are not changing much, while the immune repertoire changes its composition faster when the viral population is varying more slowly.

We performed additional tests to verify that the observed signals are not caused by spurious effects of the data. A possible confounding factor is sequencing depth: if the sizes of the HIV and Ig samples are correlated for some reason, and one of the considered measures is, in turn, correlated with size, a spurious correlation can appear. However, as shown by electronic supplementary material, figure S3*a*, the number of HIV sequences does not correlate with the number of Ig clonotypes as well as

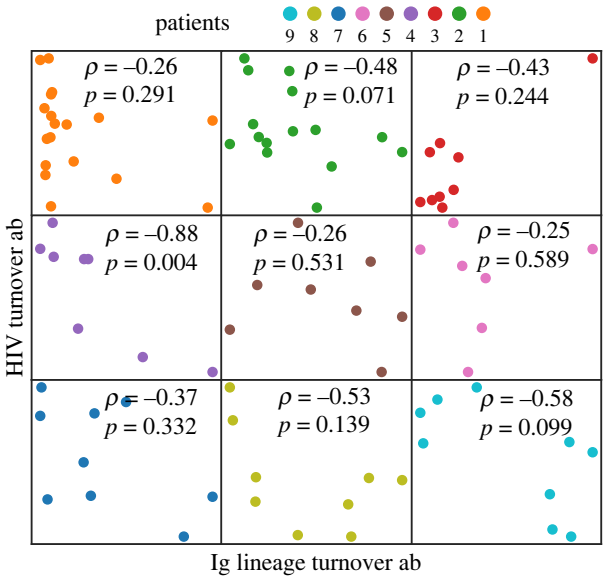

Figure 2. Scatter plots of the trajectories of *Ig lineage turnover ab* and *HIV turnover ab* for the nine patients of the dataset. The Spearman correlation coefficient ρ and the correlation *p*-value p are reported within each plot. (Online version in colour.)

with any of the Ig measures (and similarly for the Ig number of sequences). Another spurious correlation can be generated by the fact that the time points are not homogeneously distributed (electronic supplementary material, figure S1), and the considered change measures are dependent on the length of the time windows. However, we find that the size of time windows (*Delta time* in electronic supplementary material, figure S3*a*) does not show a correlation with any of the other measures.

Conversely, the Benjamini–Hochberg procedure that we use to account for multiple testing makes the conservative assumption that the different tests we tried are independent of each other. But it is likely that many of the defined quantities are strongly correlated, e.g. *Ig lineage turnover ab* and *Ig lineage large turnover*, making the effective number of tests smaller than actually used in the procedure and leading us to overestimate the corrected *p*-values. Electronic supplementary material, §S3 discusses this problem using a different null model generates trajectories that reproduce the internal correlations present within the HIV measures and within the Ig measures. This test confirms that the observed patterns are strongly unlikely to be generated by the refined null model.

We then asked if the sign of the correlation depended on time-interval size at which the populations were sampled. The procedure of selecting only long time intervals decreases the number of points of the trajectories, leading to fewer statistics. The *p*-values do not show significant correlations for longer sampling times, as shown in electronic supplementary material, figure S3*c*. However, we cannot conclude whether this lack of significance is due to the larger sampling time, or to reduced statistical power.

(d) The HIV affects the future state of the immune system

We repeated the analysis of §3(c) but with a temporal shift to the trajectories. In particular, we computed the correlations between the points of the Ig trajectories against the points of the HIV ones, but one step forward in time (Ig time shift +1, blue panel of figure 5) or one step back (Ig time

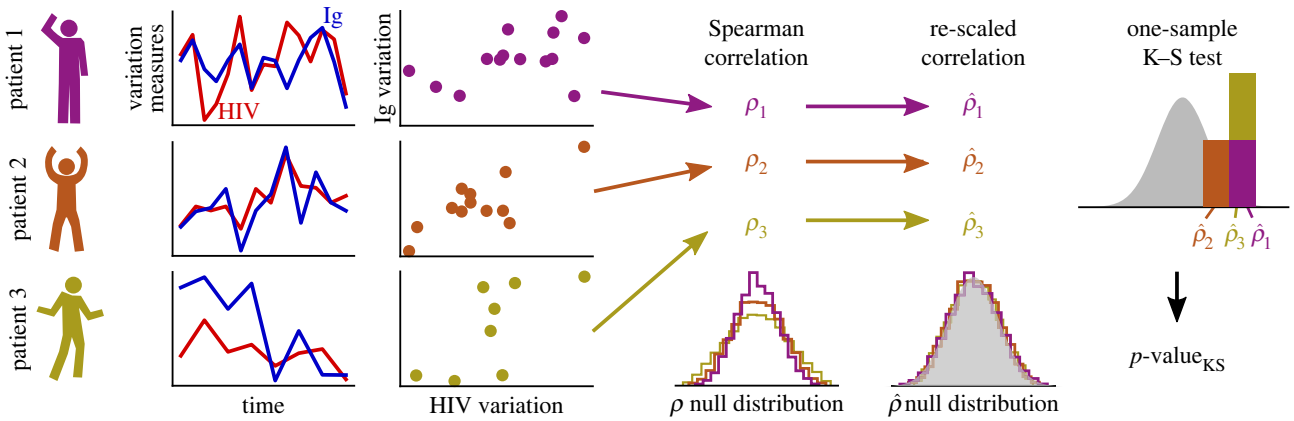

Figure 3. Cartoon of the statistical procedure to integrate correlations across different patients (not real data). Given two measures of absolute change (one for the HIV and one for the Ig repertoire) computed in all the patients, the Spearman correlations between the trajectories are computed and then re-scaled in such a way that their null distributions collapse across patients. Finally, a one-sample Kolmogorov–Smirnov test verifies if these re-scaled coefficients deviate significantly from the null distribution of correlations. (Online version in colour.)

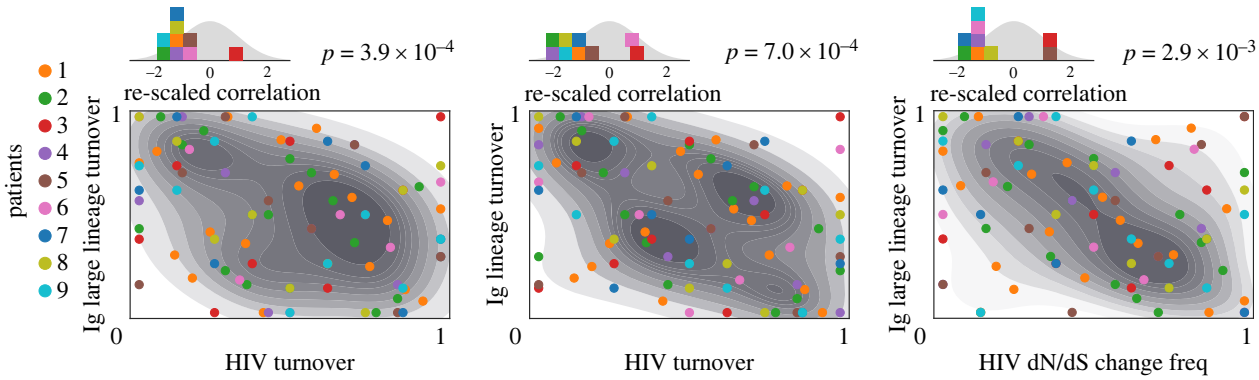

Figure 4. The three significant pairs of (anti)correlated measures. The scatter plots show the rank (normalized between 0 and 1) of all the values of the two absolute-change measures in all the patients (differentiated by the colour of the dots). The shaded areas below are the densities of these points computed with a kernel density estimate. Above each scatter plot, the histogram of the re-scaled correlations is shown, compared with the null distribution (shaded grey area). This is the same plot as the left-most one of figure 3. They all show an average negative re-scaled correlation. The Kolmogorov–Smirnov p-values are displayed on the right of the histograms. (Online version in colour.)

shift −1, red panel of figure 5). The central green panel corresponds to the case considered in the §3(c) (no time shift).

The statistical procedure discussed in §3(b) were then applied in all these cases, leading to a distribution of re-scaled correlations whose average is plotted in the lower panels of figure 5. The significant pairs (Benjamini–Hochberg test with 5% false discovery rate) are highlighted with a black border. The p-values for all the pairs are displayed in electronic supplementary material, figure S4. In addition to the previously observed significant correlations with no time shift, significance is achieved by six pairs, in which changes in HIV precedes changes of the immune system at the future time step (red dots). In other words, a larger turnover of HIV is followed by a smaller turnover of the Ig lineages at a later time. The average time of this delay, i.e. the average time distance between consecutive points of the dataset, is around 225 days. No pair shows a significant correlation in the opposite case, suggesting that the immune system does not immediately affect the future dynamics of the HIV population.

(e) Turnover correlations can be reproduced by a population–genetics model

Assuming a coevolutionary arms-race, one could have expected a positive correlation between HIV and Ig repertoire turnovers, in contradiction with the anti-correlation observed in figures 4 and 5. When one population is changing a lot, the other population is expected to change as well to keep up, leading to a positive correlation. In the following we show using population-genetics models that this intuitive argument is not generally correct, and that anti-correlations can emerge for a reasonable range of parameters.

We investigated the emergence of correlation patterns in genetic turnovers within a model used before in the context of HIV and immune system coevolution [30], with a few minor modifications. The two populations are characterized by binary strings of length L. The genotype of a virus v is denoted by $v = (\sigma_1^v, \sigma_2^v, \ldots, \sigma_L^v)$, where $\sigma_i = 1$ or -1. Similarly, a clonotype of a B-cell receptor is characterized by a binary string $r = (\sigma_1^r, \sigma_2^r, \ldots, \sigma_L^r)$. We considered populations of fixed size, N_V and N_R. The fraction of a given virus strain or Ig clone are denoted by x_v, x_r. The ability of a B-cell to recognize a viral strain and expand depends on how much its Ig string is similar to that of the virus. At the same time, a virus strain proliferates more easily if it is different from the receptor sequences. We defined the affinity between v and r as follows:

$$E_{v,r} = \sum_{i=1}^{L} \sigma_i^v \sigma_i^r. \tag{3.1}$$

Affinity enters the definition of fitness for the two populations, which is proportional to its average over the whole

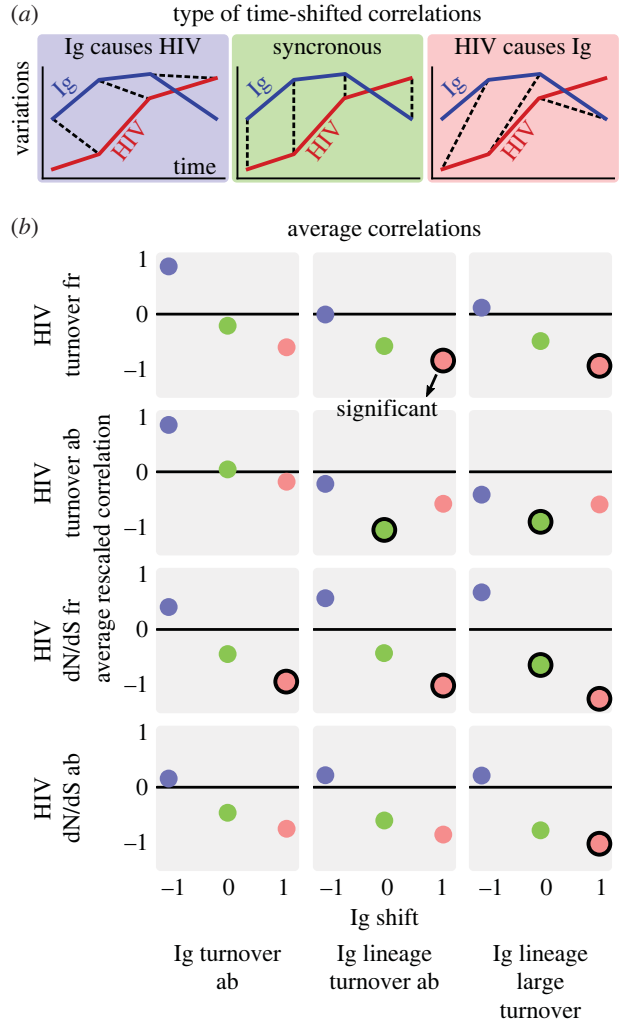

Figure 5. Testing correlations across time, using temporal shifts of the trajectories. (*a*) Cartoon of the three considered scenarios: HIV population change follows repertoire change (Ig shift −1), synchronous change (Ig shift 0), repertoire change follows HIV population change (Ig shift +1). (*b*) For each pair of HIV–Ig measures, we tested the average re-scaled correlation for each Ig time shift, whose values are shown on the *x*-axis. The dots with a black border are significant according to the Benjamini–Hochberg test with a false discovery rate of 0.05. The pairs that do not show any significant correlation are shown in electronic supplementary material, figure S4. See table 1 and table 2 for abbreviation definitions. (Online version in colour.)

adversary population:

$$f_R = \frac{s_R}{2}\sum_v x_v E_{r,v}, \quad f_V = -\frac{s_V}{2}\sum_r x_r E_{r,v}, \tag{3.2}$$

where the two selection coefficients s_R, $s_V > 0$ control the strength of selection.

This antagonistic coevolutionary dynamics is simulated using a Wright–Fisher model, whose details are described in electronic supplementary material, section S4. Briefly, in one generation, each genotype i generates a binomially distributed number of offspring, with a probability proportional to the exponential fitness (given by equation (3.2) for Ig, or its negative value for HIV) times the genotype frequency, $\exp(f_i)x_i$. The total number of individuals in the population is fixed and imposed via multinomial sampling of the population from one generation to the next. After each reproduction step, there is a probability that a given site switches sign, leading to a mutated offspring. We call the mutation rates per site per generation μ_V and μ_R.

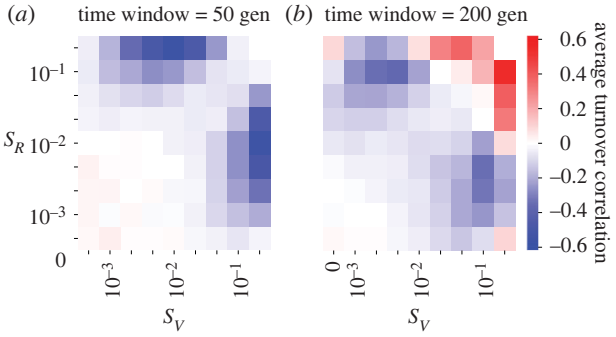

Figure 6. Colour map of the average re-scaled correlation of the genotype absolute change in the coevolving populations of binary strings. The parameters are $N_R = N_V = 10^3$, $\mu_R = \mu_V = 10^{-3}/L$, $L = 50$. For each set of parameters an ensemble of 10 000 realizations is generated. The absolute change of genotype frequencies is computed by considering 20 bins of times separated by 50 (*a*) or 200 (*b*) generations. Given the two trajectories of absolute changes, the Spearman correlation is computed, re-scaled and averaged across realizations. Details of the simulations are given in electronic supplementary material, §S4. (Online version in colour.)

We performed simulations using values for the population size and the mutation rate estimated from data. The effective population size of viral population and number of B-cells in germinal centres is approximately $N_R \sim 10^3 - 10^4$, $N_V \sim 10^2 - 10^3$ [31,37,38]. The mutation rates are estimated from neutral sequence diversity as $\mu_R \sim \mu_V \sim 10^{-5} - 10^{-4}$ over a length of $L_R \sim 200$ for the receptor variable region, and $L_V \sim 1000$ for the *env* protein of HIV. This leads to 1–10 mutations per generation in both populations [23,30,39]

As done in [30], in our model we needed to set the same L for both of the populations, choosing $L = 50$ amino acids, which means a gene of a similar length of the B cell receptor variable region. The other parameters were chosen in a way to keep the total number of mutations per generation equal to 1 in both the populations: $N_R = N_V = 10^3$, $\mu_R = \mu_V = 10^{-3}/L$.

Figure 6 shows the correlation between turnovers generated by the simulation. The genotype frequencies of the two populations were sampled every 50 or 200 generations. The real generation time of B-cells or HIV can be of order of 1 day [31], meaning that we sampled our simulations every few months, similarly to what was done in the experimental data. We compute the absolute change by considering 20 consecutive samples, using equation (2.1), where x corresponds to the genotype frequency. We then correlate the two obtained trajectories of absolute changes for the two populations and plot the average of those re-scaled correlations over repeated realizations of the simulation. The behaviour is complex, depending on the selection coefficients and the interval of time chosen for the sampling. However, figure 6 shows that the region of parameter for which the correlation is negative is wide, consistent with empirical observations.

(f) Minimal population–genetics models can reproduce the observed signal

To gain further intuition about the observed turnover correlations, we make a strong simplification of neglecting the mutational background [40] and consider the antagonistic coevolution of two populations, each with only one locus. This corresponds to the previous model with $L = 1$.

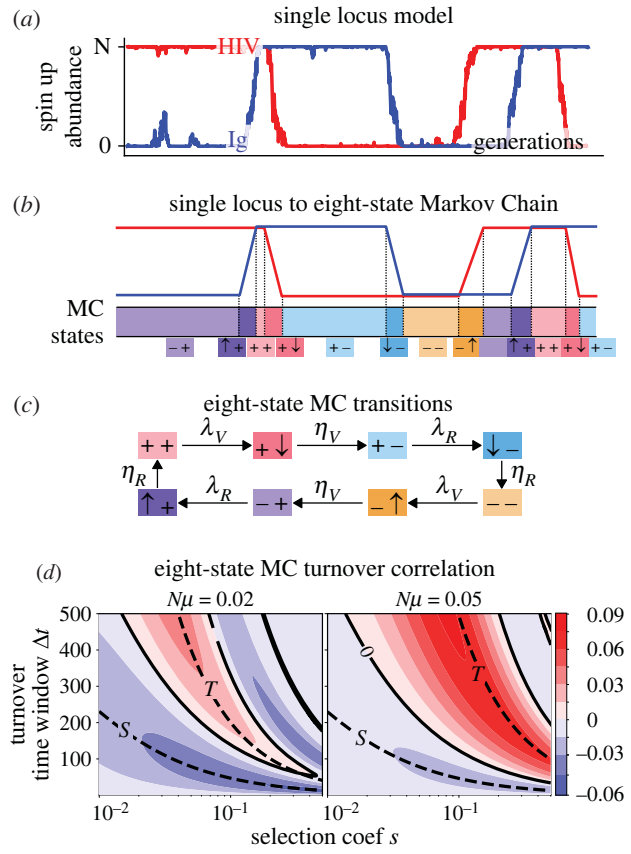

Figure 7. (*a*) Number of up states of two populations of antagonistically interacting loci (model of §3(e) with $L = 1$). The parameters are $N = N_R = N_V = 10^3$, $\mu = \mu_R = \mu_V = 10^{-3}/50$, $s = s_R = s_V = 0.005$. (*b*) Schematized trajectories where the dynamics is composed of periods of fixed populations, interleaved with periods of switches to the next fixation. (*c*) Equivalent eight-state Markov chain whose states are represented with different colours shown in (*b*). (*d*) Analytical solution for the turnover correlation (electronic supplementary material, equation E3) for two populations with identical size, mutation and selection parameters, as a function of the selection coefficient s and of the time window for turnover computation Δt. Black lines indicate zero correlation. The dashed lines are the establishment time, T, and the switch time S as a function of s. The two different plots refer to two choices of μN, with $N = 10^3$. MC, Markov chain. Rate $\eta_i = 1/S_i$, where $i = \{R, V\}$. (Online version in colour.)

Using parameters corresponding to the full model, within the simple model each population alternates between states of almost complete fixation and switching (figure 7*a*), characteristic of the successional-mutation or selective sweep regime [41,42]. Within this regime, the simple model estimates the time for finding and establishing a beneficial mutation as $T \sim 1/(s\ N\mu)$. This time can be applied to HIV when its population is fixed and the spin sign is concordant with that of the immune system. For the immune system, it applies when its sign is opposite to HIV. In both cases mutations are beneficial with a selective advantage s. The time for a beneficial mutation to expand in the whole-population scales as $S \sim \log(s\ N)/s$ [41]. The selective sweep regime, $S \ll T$, means that each sweep finishes before the sweep of the other population starts, which corresponds to the limit of weak mutations $N\mu \ll 1$.

This simple model maps the trajectories in figure 7*a* onto an effective eight-state discrete Markov chain (figure 7*b*,*c*). Each of the two populations can either be at a fixed or sweep state that, when combined with the four possible all up (+) and down (–) configurations of the two populations, lead to eight possible states. Transitions to the sweep states happen with rate $\lambda_i = 1/T_i$ and to the transitional states with rate $\eta_i = 1/S_i$, where $i = \{R, V\}$. For example, in the state where both populations are in the + fixation state (the pink state ++ in figure 7*b*,*c*), a beneficial mutation occurs in the virus with rate λ_V moving it to a sweep state +↓. The system moves to the discordant configuration +– with effective rate η_V, until a beneficial mutation in the repertoire occurs with rate, λ_R and so on. In this setting, the turnover of a population in a time window Δt is 1 if the state at t has a different fixation sign from the state at $t + \Delta t$, and 0 if the fixation sign is the same between the two time points (see electronic supplementary material, §S5 for more details).

This simplified Markov chain qualitatively captures the correlations between turnovers of the full model in §3(e) (electronic supplementary material, figure S7*a* and section S5). This case can be also solved analytically (electronic supplementary material, equation S3), allowing us to understand how the correlation regime depends on the population parameters and the timescales involved. In general, we see that we need selection to observe any correlation. Negative correlations, such as those observed in the data, require small or intermediate selection coefficients s, and small turnover time windows Δt (figure 7*d* and electronic supplementary material, section S5C). The intuition behind the negative correlation is that a sweep in the second population is unlikely to occur before the sweep of the first population is almost complete because the selective advantage for the second population is weaker during the sweep than after fixation. As a consequence, two sweeps are unlikely to be synchronous, implying an anti-correlation in the turnovers of the two populations when Δt is of the order of the sweep time S. Consistent with this reasoning, figure 7*d* shows that the first maximum of negative correlation scales with the sweep time $S \sim \log(s\ N)/s$, which sets the timescale for negative correlations.

In contrast, a longer Δt will not be affected by this interference effect and can lead to positive correlations when $\Delta t \sim T$, capturing the intuition that a sweep in one population favours a subsequent sweep in the other. Consistently, the maximum of positive correlation scales with the establishment time $T \sim 1/(s\ N\mu)$, determining the timescale of positive correlations (figure 7*d* and electronic supplementary material, section S5C). Increasing the time window Δt after this scale, the first population has enough time to sweep back, generating a second region of negative correlation at larger Δt. This alternating behaviour of positive and negative correlations is damped as Δt is further increased, eventually approaching zero correlation for $\Delta t \gg T$ (electronic supplementary material, section S5C).

Together, the simple model shows that the timescales of turnover determine the negative correlations observed in data. The results also suggest the limits of relatively weak selection coefficients and small turnover windows compared to mutation rates.

4. Discussion

HIV and the immune system likely interact antagonistically in a coevolutionary process, whose quantitative details are still not well understood. Previous works provide evidence that the virus [43] and the immune system [44] taken separately are measurably evolving, but an analysis that identifies signals

of the coevolutionary interaction between the two populations was still lacking. Taking advantage of a dataset that tracks in time HIV sequences and B-cell repertoires, we show that this interaction exists and leaves traces at the population scale. In particular, we find that HIV genetic turnover and lineage turnover are significantly negatively correlated. Since lineage turnover is a measure that considers the immune system as a whole—we do not select for specific Ig lineages—the fact that the signal is significant makes us speculate that the viral population interacts with a large number of distinct immune lineages, as suggested by previous studies [27,30]. A second surprising finding is that this correlation is negative, meaning that when the viral population slows down, the B-cell clones increase their rate of change (and *vice versa*).

Using a simple model of coevolution, we were able to show that negative correlations appear when the time delay for computing the turnover is comparable with the sweep time of mutations, while positive correlations emerge for delays of the order of the establishment time of the new mutant. The simplest one-locus model that we considered to derive these timescales clearly lacks biological realism, in particular in its neglecting mutational background and competition between lineages. Nevertheless, it shows that the observed negative correlations are driven by a specific interaction between the timescales of the problem and constrains the evolutionary regimes of the viral population and the repertoire. Specifically, it suggests that the viral population sweeps before the immune repertoire can respond.

While we cannot offer a detailed mechanism for how this evolutionary regime is obtained, we may speculate that the viral population within a sweep rapidly mutates away from the regions covered by the repertoire, so that the immune system cannot immediately adapt. Eventually the immune system catches up, but this happens on timescales where new beneficial mutations in the viral population may occur. While the simple model proved very useful to gain intuition how different turnover correlations can occur, its mapping between the correlation sign and the population parameters (selection coefficients, mutation rates) has to be taken with caution. For example, both the HIV [45] and the B-cell repertoire [28] show evidence of clonal interference. A consequence of clonal interference is a decrease in the sweep times of beneficial mutations [46], and the dependence between the distribution of selection coefficients and observed sweep dynamics is much more difficult to estimate.

By investigating the correlations between the two populations with a timeshift, we found that a few measures of viral turnover and dN/dS negatively correlate with Ig measures one step forward in the future, while the opposite was never true. This suggests that HIV evolution impacts the future state of the immune system. One possible explanation for the absence of reciprocity could be that there is a difference in the timescales of response to changes in the other population. Assume that the HIV virus responds fast to repertoire changes, while repertoires respond more slowly to changes in HIV composition. Then the immune system would hold a longer memory of the past states of the viral population, leading to the observed delay in the anti-correlation. Characterizing the timescales involved in these correlations could help us unveil crucial properties of the HIV–immune system interaction. However, our attempts in this direction have been hampered by the limited size of the dataset.

Both the full model of HIV–Ig coevolution inspired by Nourmohammad *et al.* [30] and its simplified Markov chain description reproduce the negative correlations in a wide range of biologically plausible parameters. The simplified model does not include the possibility of clonal interference, indicating that these properties are not necessary for the appearance of the correlation patterns. Both models consider only the evolution of clonotypes within a single lineage against a single epitope, rather than a the multi-lineage dynamics with multiple targets. Generalizing to more realistic models would require making choices about the kind of dynamics and parameters for ways in which lineages interact. A future interesting line of research would be to better characterize this competitive dynamics, since it is known that different lineages act together against the same pathogen [4], for example targeting different viral epitopes. One could also introduce multiple viral epitopes in the model [47]. Different receptor lineages would compete to increase the affinity against a given epitope, but at the same time cooperate in targeting different epitopes. The emerging overall dynamics generalizes previous models [30] and it would be interesting to study how this more realistic setting influences the statistics of the generated turnover correlations.

To conclude, in this work we have identified an unexpected pattern of anti-correlation in the longitudinal tracking of coevolving HIV and Ig repertoires. The approach we developed can help to analyse genetic data from other coevolving populations with antagonist interactions, and to understand better the general rules that govern them.

Data accessibility. The open access database of Strauli *et al.* [32] is available from from the Sequence Read Archive (SRA), BioProject ID PRJNA543982 at: https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA543982.

The data are provided in electronic supplementary material [48].

Authors' contributions. A.M.: conceptualization, data curation, formal analysis, investigation, methodology, resources, software, validation, visualization, writing—original draft, writing—review and editing; T.M.: conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, visualization, writing—original draft, writing—review and editing; A.M.W.: conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, visualization, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. This work was partially supported by the European Research Council Consolidator grant no. 724208 and ANR-19-CE45-0018 'RESP-REP' from the Agence Nationale de la Recherche.

References

1. MacLennan ICM. 1994 Germinal centers. *Annu. Rev. Immunol.* **12**, 117–139. (doi:10.1146/annurev.iy.12.040194.001001)
2. Allen CDC, Okada T, Cyster JG. 2007 Germinal-center organization and cellular dynamics. *Immunity* **27**, 190–202. (doi:10.1016/j.immuni.2007.07.009)
3. Campbell CD, Eichler EE. 2013 Properties and rates of germline mutations in humans. *Trends Genet.* **29**, 575–584. (doi:10.1016/j.tig.2013.04.005)

4. Victora GD, Nussenzweig MC. 2012 Germinal centers. *Annu. Rev. Immunol.* **30**, 429–457. (doi:10.1146/annurev-immunol-020711-075032)
5. Shlomchik MJ, Weisel F. 2012 Germinal center selection and the development of memory B and plasma cells. *Immunol. Rev.* **247**, 52–63. (doi:10.1111/j.1600-065X.2012.01124.x)
6. Mesin L, Ersching J, Victora GD. 2016 Germinal center B cell dynamics. *Immunity* **45**, 471–482. (doi:10.1016/j.immuni.2016.09.001)
7. McMichael AJ, Borrow P, Tomaras GD, Goonetilleke N, Haynes BF. 2010 The immune response during acute HIV-1 infection: clues for vaccine development. *Nat. Rev. Immunol.* **10**, 11–23. (doi:10.1038/nri2674)
8. Fauci AS. 2003 HIV and AIDS: 20 years of science. *Nat. Med.* **9**, 839–843. (doi:10.1038/nm0703-839)
9. Richman DD, Wrin T, Little SJ, Petropoulos CJ. 2003 Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc. Natl Acad. Sci. USA* **100**, 4144–4149. (doi:10.1073/pnas.0630530100)
10. Moore PL *et al.* 2009 Limited neutralizing antibody specificities drive neutralization escape in early HIV-1 subtype C infection. *PLoS Pathog.* **5**, e1000598. (doi:10.1371/journal.ppat.1000598)
11. Kwong PD *et al.* 2002 HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature* **420**, 678–682. (doi:10.1038/nature01188)
12. Lyumkis D *et al.* 2013 Cryo-em structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science* **342**, 1484–1490. (doi:10.1126/science.1245627)
13. Simek MD *et al.* 2009 Human immunodeficiency virus type 1 elite neutralizers: individuals with broad and potent neutralizing activity identified by using a high-throughput neutralization assay together with an analytical selection algorithm. *J. Virol.* **83**, 7337–7348. (doi:10.1128/JVI.00110-09)
14. Liao H-X *et al.* 2013 Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* **496**, 469–476. (doi:10.1038/nature12053)
15. McCoy LE, Burton DR. 2017 Identification and specificity of broadly neutralizing antibodies against HIV. *Immunol. Rev.* **275**, 11–20. (doi:10.1111/imr.12484)
16. Walker LM *et al.* 2009 Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* **326**, 285–289. (doi:10.1126/science.1178746)
17. Walker LM *et al.* 2011 Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* **477**, 466–470. (doi:10.1038/nature10373)
18. Kwong PD, Mascola JR, Nabel GJ. 2013 Broadly neutralizing antibodies and the search for an HIV-1 vaccine: the end of the beginning. *Nat. Rev. Immunol.* **13**, 693–701. (doi:10.1038/nri3516)
19. Klein F, Mouquet H, Dosenovic P, Scheid JF, Scharf L, Nussenzweig MC. 2013 Antibodies in HIV-1 vaccine development and therapy. *Science* **341**, 1199–1204. (doi:10.1126/science.1241144)
20. Fischer W *et al.* 2010 Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. *PLoS ONE* **5**, e12303. (doi:10.1371/journal.pone.0012303)
21. Henn MR *et al.* 2012 Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. *PLoS Pathog.* **8**, e1002529. (doi:10.1371/journal.ppat.1002529)
22. Barton JP, Goonetilleke N, Butler TC, Walker BD, McMichael AJ, Chakraborty AK. 2016 Relative rate and location of intra-host HIV evolution to evade cellular immunity are predictable. *Nat. Commun.* **7**, 1–10. (doi:10.1038/ncomms11660)
23. Zanini F, Brodin J, Thebo L, Lanz C, Bratt G, Albert J, Neher RA. 2015 Population genomics of intrapatient HIV-1 evolution. *Elife* **4**, e11282. (doi:10.7554/eLife.11282)
24. Mouquet H. 2014 Antibody B cell responses in HIV-1 infection. *Trends Immunol.* **35**, 549–561. (doi:10.1016/j.it.2014.08.007)
25. Victora GD, Mouquet H. 2018 What are the primary limitations in B-cell affinity maturation, and how much affinity maturation can we drive with vaccination? Lessons from the antibody response to HIV-1. *Cold Spring Harbor perspect. Biol.* **10**, a029389. (doi:10.1101/cshperspect.a029389)
26. Kreer C *et al.* 2022 Determining probabilities of HIV-1 bnab development in healthy and chronically infected individuals. *bioRxiv* 2022.07.11.499584. (doi:10.1101/2022.07.11.499584)
27. Hoehn KB *et al.* 2015 Dynamics of immunoglobulin sequence diversity in HIV-1 infected individuals. *Phil. Trans. R. Soc. B* **370**, 20140241. (doi:10.1098/rstb.2014.0241)
28. Nourmohammad A, Otwinowski J, Łuksza M, Mora T, Walczak AM. 2019 Fierce selection and interference in B-cell repertoire response to chronic HIV-1. *Mol. Biol. Evol.* **36**, 2184–2194. (doi:10.1093/molbev/msz143)
29. Wang S *et al.* 2015 Manipulating the selection forces during affinity maturation to generate cross-reactive HIV antibodies. *Cell* **160**, 785–797. (doi:10.1016/j.cell.2015.01.027)
30. Nourmohammad A, Otwinowski J, Plotkin JB. 2016 Host-pathogen coevolution and the emergence of broadly neutralizing antibodies in chronic infections. *PLoS Genet.* **12**, e1006171. (doi:10.1371/journal.pgen.1006171)
31. Molari M, Eyer K, Baudry J, Cocco S, Monasson R. 2020 Quantitative modeling of the effect of antigen dosage on B-cell affinity distributions in maturating germinal centers. *Elife* **9**, e55678. (doi:10.7554/eLife.55678)
32. Strauli N, Fryer EK, Pham O, Abdel-Mohsen M, Facente SN, Pilcher C, Pennings P, Pillai S, Hernandez RD. 2019 The genetic interaction between hiv and the antibody repertoire. *BioRxiv* 646968. (doi:10.1101/646968)
33. Hatada M, Yoshimura K, Harada S, Kawanami Y, Shibata J, Matsushita S. 2010 Human immunodeficiency virus type 1 evasion of a neutralizing anti-V3 antibody involves acquisition of a potential glycosylation site in V2. *J. Gen. Virol.* **91**, 1335–1345. (doi:10.1099/vir.0.017426-0)
34. Ringe R, Das L, Choudhary I, Sharma D, Paranjape R, Chauhan VS, Bhattacharya J. 2012 Unique C2V3 sequence in HIV-1 envelope obtained from broadly neutralizing plasma of a slow progressing patient conferred enhanced virus neutralization. *PLoS ONE* **7**, e46713. (doi:10.1371/journal.pone.0046713)
35. Nei M, Gojobori T. 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**, 418–426. (doi:10.1093/oxfordjournals.molbev.a040410)
36. Yang Z, Bielawski JP. 2000 Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**, 496–503. (doi:10.1016/S0169-5347(00)01994-7)
37. Tas JMJ *et al.* 2016 Visualizing antibody affinity maturation in germinal centers. *Science* **351**, 1048–1054. (doi:10.1126/science.aad3439)
38. Lemey P, Rambaut A, Pybus OG. 2006 HIV evolutionary dynamics within and among hosts. *AIDS Rev.* **8**, 125–140.
39. Elhanati Y, Sethna Z, Marcou Q, Callan Jr CG, Mora T, Walczak AM. 2015 Inferring processes underlying B-cell repertoire diversity. *Phil. Trans. R. Soc. B* **370**, 20140243. (doi:10.1098/rstb.2014.0243)
40. Neher RA. 2013 Genetic draft, selective interference, and population genetics of rapid adaptation. *Ann. Rev. Ecol. Evol. Syst.* **44**, 195–215. (doi:10.1146/annurev-ecolsys-110512-135920)
41. Desai MM, Fisher DS. 2007 Beneficial mutation–selection balance and the effect of linkage on positive selection. *Genetics* **176**, 1759–1798. (doi:10.1534/genetics.106.067678)
42. Ebert D, Fields PD. 2020 Host–parasite co-evolution and its genomic signature. *Nat. Rev. Genet.* **21**, 754–768. (doi:10.1038/s41576-020-0269-1)
43. Rambaut A, Posada D, Crandall KA, Holmes EC. 2004 The causes and consequences of HIV evolution. *Nat. Rev. Genet.* **5**, 52–61. (doi:10.1038/nrg1246)
44. Hoehn KB, Turner JS, Miller FI, Jiang R, Pybus OG, Ellebedy AH, Kleinstein SH. 2021 Human B cell lineages associated with germinal centers following influenza vaccination are measurably evolving. *Elife* **10**, e70873. (doi:10.7554/eLife.70873)
45. Pandit A, de Boer RJ. 2014 Reliable reconstruction of HIV-1 whole genome haplotypes reveals clonal interference and genetic hitchhiking among immune escape variants. *Retrovirology* **11**, 1–15. (doi:10.1186/1742-4690-11-56)
46. Schiffels S, Szöllősi GJ, Mustonen V, Lässig M. 2011 Emergent neutrality in adaptive asexual evolution. *Genetics* **189**, 1361–1375. (doi:10.1534/genetics.111.132027)
47. Zolla-Pazner S. 2004 Identifying epitopes of HIV-1 that induce protective antibodies. *Nat. Rev. Immunol.* **4**, 199–210. (doi:10.1038/nri1307)
48. Mazzolini A, Mora T, Walczak AM. 2023 Inspecting the interaction between human immunodeficiency virus and the immune system through genetic turnover. Figshare. (doi:10.6084/m9.figshare.c.6476964)