



**HAL**  
open science

## Fierce Selection and Interference in B-Cell Repertoire Response to Chronic HIV-1

Armita Nourmohammad, Jakub Otwinowski, Marta Łuksza, Thierry Mora, Aleksandra M Walczak

► **To cite this version:**

Armita Nourmohammad, Jakub Otwinowski, Marta Łuksza, Thierry Mora, Aleksandra M Walczak. Fierce Selection and Interference in B-Cell Repertoire Response to Chronic HIV-1. *Molecular Biology and Evolution*, 2019, 36 (10), pp.2184-2194. 10.1093/molbev/msz143 . hal-02364221

**HAL Id: hal-02364221**

**<https://hal.science/hal-02364221v1>**

Submitted on 20 Sep 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Fierce Selection and Interference in B-Cell Repertoire Response to Chronic HIV-1

Armita Nourmohammad,<sup>\*,1,2</sup> Jakub Otwinowski,<sup>1</sup> Marta Łuksza,<sup>3</sup> Thierry Mora,<sup>†,4</sup> and Aleksandra M. Walczak<sup>†,5</sup>

<sup>1</sup>Max Planck Institute for Dynamics and Self-organization, Göttingen, Germany

<sup>2</sup>Department of Physics, University of Washington, Seattle, WA

<sup>3</sup>Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY

<sup>4</sup>Laboratoire de Physique Statistique, CNRS, Sorbonne University, Paris-Diderot University, École Normale Supérieure (PSL), Paris, France

<sup>5</sup>Laboratoire de Physique Théorique, CNRS, Sorbonne University, École Normale Supérieure (PSL), Paris, France

<sup>†</sup>These authors contributed equally to this work.

\*Corresponding author: E-mail: armita@ds.mpg.de.

Associate editor: Thomas Leitner

## Abstract

During chronic infection, HIV-1 engages in a rapid coevolutionary arms race with the host's adaptive immune system. While it is clear that HIV exerts strong selection on the adaptive immune system, the characteristics of the somatic evolution that shape the immune response are still unknown. Traditional population genetics methods fail to distinguish chronic immune response from healthy repertoire evolution. Here, we infer the evolutionary modes of B-cell repertoires and identify complex dynamics with a constant production of better B-cell receptor (BCR) mutants that compete, maintaining large clonal diversity and potentially slowing down adaptation. A substantial fraction of mutations that rise to high frequencies in pathogen-engaging CDRs of BCRs are beneficial, in contrast to many such changes in structurally relevant frameworks that are deleterious and circulate by hitchhiking. We identify a pattern where BCRs in patients who experience larger viral expansions undergo stronger selection with a rapid turnover of beneficial mutations due to clonal interference in their CDR3 regions. Using population genetics modeling, we show that the extinction of these beneficial mutations can be attributed to the rise of competing beneficial alleles and clonal interference. The picture is of a dynamic repertoire, where better clones may be outcompeted by new mutants before they fix.

**Key words:** B-cell somatic evolution, adaptive immunity, population genetics, fluctuating selection, clonal interference.

## Introduction

HIV-1 evolves and proliferates quickly within the human body (Richman et al. 2003; Moore et al. 2009; Liao et al. 2013), rapidly mutating and often recombining its genetic material among different viral genomes. These factors make it very hard for the host immune system to maintain a sustained control of an infection, leading to a long-term chronic condition. While it is clear that the virus exerts strong selective pressure on the host immune system, the quantitative nature of the evolutionary dynamics of the adaptive immune system during chronic infections remains unknown.

The immune system has a diverse set of B- and T-cells with specialized surface receptors that recognize foreign antigens, such as viral epitopes, to protect the organism. We focus on the chronic phase of HIV infection, where the immune response is dominated by antibody-mediated mechanisms, following the strong response of cytotoxic T-lymphocytes (i.e., CD8+ killers T-cells), ~50 days after infection (McMichael et al. 2010). During the chronic phase, the symptoms are minor and the viral load is relatively stable, but its genetic

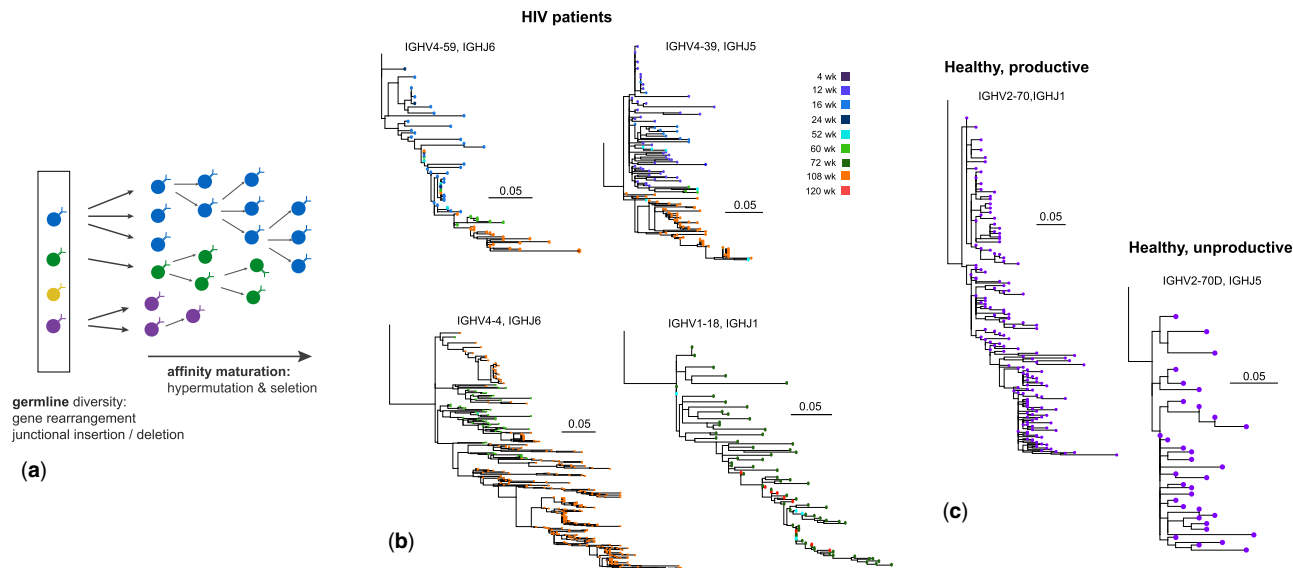
composition undergoes rapid turnover. After an infection, B-cells undergo a rapid somatic hypermutation in lymph node germinal centers, with a rate that is approximately four to five orders of magnitude larger than an average germline mutation rate in humans (Campbell and Eichler 2013). Mutated B-cells compete for survival and proliferation signals from helper T-cells, based on the B-cell receptor's (BCR) binding to antigens. This process of *affinity maturation* is Darwinian evolution within the host and can increase binding affinities of BCRs up to 10- to 100-fold (Victora and Nussenzweig 2012). It generates memory and plasma B-cells with distinct receptors, forming lineages that reflect their coevolution with viruses (Nourmohammad et al. 2016), (see schematic in fig. 1a). A B-cell repertoire consists of many such lineages forming a forest of coexisting genealogies. The outcome of an affinity maturation process shifts the overall repertoire response against the pathogen (Berek and Milstein 1987).

Immune repertoire high-throughput sequencing has been instrumental in quantifying the diversity of B-cell repertoires (Weinstein et al. 2009; Elhanati et al. 2015). Statistical methods have been developed to characterize the processes

© The Author(s) 2019. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access



**FIG. 1.** Affinity maturation forms B-cell lineages. (a) Schematic of B-cell affinity maturation and lineage formation. The naive immune repertoire consists of a diverse set of B-cell receptors, generated by gene rearrangement (VDJ) recombination and junctional sequence insertion and deletion (distinct colored cells in the box). Affinity maturation with somatic hypermutations and selection for strong binding of BCRs to antigens forms lineages of BCRs stemmed from a germline progenitor, shown by three growing lineages in this figure. (b) Examples of B-cell lineages reconstructed from the heavy chain sequences of BCR repertoires in HIV patients (see Materials and Methods and [Supplementary Material](#) online). The distance between the nodes along the horizontal axis indicates their sequence hamming distance. The nodes are colored according to the time they were sampled from a patient over the period of  $\sim 2.5$  years. (c) Examples of a productive (left) and unproductive (right) B-cell lineage reconstructed from the heavy chain repertoire of a healthy individual sampled at a single time point (see Materials and Methods and [Supplementary Material](#) online).

involved in the generation of diversity in repertoires and to infer the underlying heterogeneous hypermutation preferences in BCRs (Yaari et al. 2013; Elhanati et al. 2015; McCoy et al. 2015). Deviation of the observed mutations in BCRs from the expected hypermutation patterns are used to infer selection effects of mutations from repertoire snapshots in order to identify functional changes that contribute to the response against pathogens (Yaari et al. 2013; Uduman et al. 2014). Recently, longitudinal data, with repertoires sampled over multiple time points from the same individuals, have brought insight into the dynamics of affinity maturation in response to antigens (Vollmers et al. 2013; Laserson et al. 2014; Hoehn et al. 2015; Horns et al. 2019). The dynamics of affinity maturation and selection in response to HIV have also been characterized for chosen monoclonal broadly neutralizing antibody lineages (Liao et al. 2013; Vieira et al. 2018). Yet, the effect of a chronic infection on the dynamics of the whole BCR repertoire remains unknown.

Here, we analyze the history and structure of BCR lineages in the full repertoire of HIV-1-infected patients. We uncover distinct modes of immune response, including selection and competitive clonal interference among BCRs, a fraction of which may be HIV-specific. We identify a pattern, where BCR repertoires in patients who experience a larger viral expansions undergo stronger selection and clonal interference in their pathogen-engaging CDR3 regions. We show that clonal interference in CDR3 regions reflects a macro-evolutionary drive of the repertoire, either caused by the virus or the overall reorganization of the BCRs, even those that do not directly target HIV-1. Our results are based on advanced

statistical measures informed by population genetics theory that capture the differences between baseline affinity maturation and long-term selection in response to HIV-1 infection.

## Results

We compare the structure and dynamics of BCR repertoires sampled over 2.5 years in HIV patients (data from Hoehn et al. 2015; collected through the SPARTAC study [SPARTAC Trial Investigators et al. 2013]). Among these individuals are two untreated patients and four patients who had interrupted ART after a year of treatment. We have also analyzed the BCR repertoire structure in three healthy individuals (data from DeWitt et al. 2016). The sequencing depth of the two data sets differ, with on an average  $\sim 172,000$  unique BCR sequences per HIV patients, and  $\sim 880,000$  unique BCRs in healthy individuals and an average of about 3,500 lineages with size  $>20$  per HIV patient and 17,700 per healthy individuals; see Materials and Methods and [supplementary fig. S1](#), and [table S1](#), [Supplementary Material](#) online, for details on BCR data and processing. Additionally, due to the differences in the sequencing protocols (Hoehn et al. 2015; DeWitt et al. 2016), the read length of the receptors in healthy individuals ( $\sim 130$  bp) is much smaller than in HIV patients ( $\sim 300$  bp with  $\sim 35$  bp gap), making a direct comparison between the two data sets difficult. We have performed our statistical analysis both on the complete BCR repertoire data in healthy individuals and on the subsampled data with a depth comparable to the BCR repertoires in HIV patients; see [Supplementary Material](#) online. However, the healthy

repertoires serve as a guideline in our analysis, rather than a null model for selection in chronically challenged BCR repertoires, due to the differences in the structure of the data sets and the underlying sequencing protocols. Our primary conclusions rely on the analysis of selection in BCR repertoires of HIV patients and relating the differences among patients to the state of their viral load over time.

### Statistics of BCR Lineage Genealogies Indicate Positive Selection

We reconstruct genealogical trees for BCR lineages inferred from BCR repertoires in each individual (see Materials and Methods and [Supplementary Material](#) online). B-cell lineages of HIV patients, a few examples of which are shown in [figure 1b](#), can persist from months to years after the initial infection, which is much longer than the lifetime of a germinal center (weeks), indicating the recruitment of memory cells for further cycles of affinity maturation in response to the evolving virus. Reconstructed lineage trees show a skewed and asymmetric structure, consistent with rapid evolution under positive selection (see [supplementary fig. S2, Supplementary Material](#) online) (Neher and Hallatschek 2013). To quantify these asymmetries, we estimated two indices of tree imbalance and terminal branch length anomaly. In both HIV patients and healthy individuals, we observe a significant branching imbalance at the root of the BCR lineage trees, indicated by the U-shaped distribution of the sublineage weight ratios (see [Supplementary Material](#) online), in contrast to the flat prediction of neutral evolution, calculated from Kingman's coalescent ([fig. 2a](#)). Moreover, we observe elongated terminal branches (i.e., larger coalescence time) in BCR trees compared with their internal branches, with the strongest effect seen in trees from HIV patients, again in violation of neutrality ([fig. 2b](#) and [supplementary fig. S2, Supplementary Material](#) online); see [Supplementary Material](#) online for inference of coalescence time. These asymmetric features of BCR trees are clear signs of intralinearage positive selection. They break the assumptions of neutral models that are based on nonbiased growth of all terminal branches, which results in all branches and sublineages growing at equal rates. However, the considered statistics only reflect the history of lineage replication and give limited insight into the mechanisms and dynamics of selection. For instance, tree asymmetry is also observed in unproductive BCR lineages, which lack any immunological function but are carried along with the productive version of the recombined gene expressed on the other chromosome ([fig. 2a and b](#)).

### Site Frequency Spectra Indicate Rapid Adaptation in CDR3 Regions

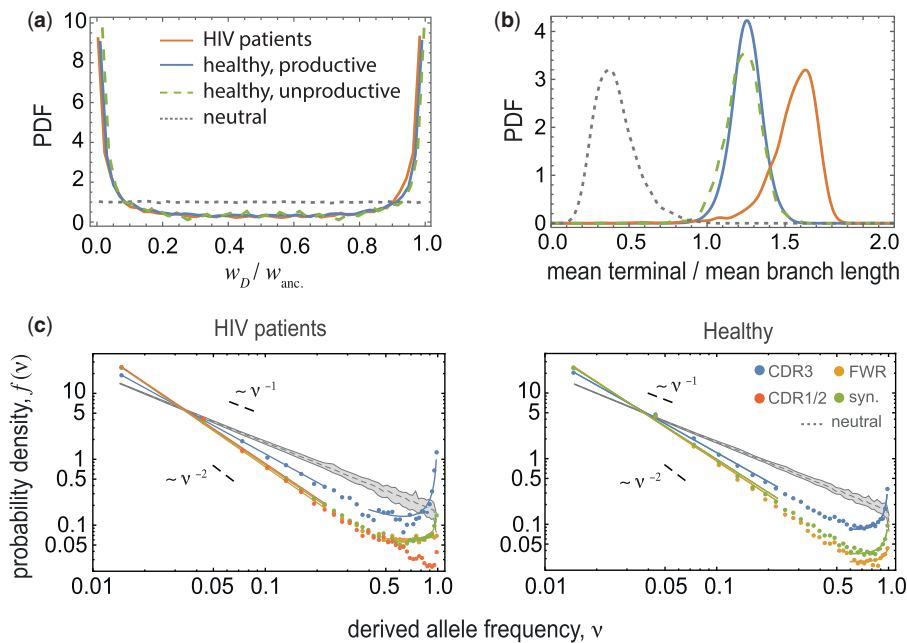
To characterize the selection effect of mutations in more detail, we evaluate the spectrum of mutation frequencies in a lineage, known as the site frequency spectrum (SFS). The SFS is the probability density  $f(\nu)$  of observing a derived mutation (allele) with a given frequency  $\nu$  in a lineage. A mutation that occurs along the phylogeny of a lineage forms a clade and is present in all the descendent nodes (leaves) of

its clade (see [supplementary fig. S2, Supplementary Material](#) online). Therefore, SFS carries information about the shape of the phylogeny, including both the topology and the branch lengths. In neutrality, mutations rarely reach high frequency, and hence, the SFS decays monotonically with allele frequency as,  $f(\nu) \sim \nu^{-1}$  (Kingman 1982). In phylogenies with skewed branching, many mutations reside on the larger subclade following a branching event, and hence, are present in the majority of the descendent leaves on the tree. The SFS of such lineages is often nonmonotonic with an upturn in the high frequency part of the spectrum (Neher and Hallatschek 2013). We evaluate the SFS separately for synonymous and nonsynonymous mutations in different regions of BCRs ([fig. 2c](#) and [supplementary fig. S3, Supplementary Material](#) online). In HIV patients, we see a significant upturn of SFS polarized on nonsynonymous mutations in pathogen-engaging CDR3 regions, consistent with rapid adaptive evolution (Neher and Hallatschek 2013), and in contrast to monotonically decaying SFS in neutrality ([fig. 2c](#) and [Supplementary Material](#) online). In addition, we observe significant overrepresentation of high-frequency synonymous mutations in productive lineages of HIV patients and healthy individuals, which indicates hitchhiking of neutral mutations with positively selected alleles. We evaluate the significance of the signal by comparing to a bootstrapped distribution of an ensemble of neutrally generated trees with otherwise similar statistics to experimentally observed BCR lineages ([fig. 2c](#) and [supplementary fig. S3, Supplementary Material](#) online). The signal of positive selection is strongest in HIV patients with an order of magnitude increase in the high end of the spectrum, suggesting that the BCR population rapidly adapts in HIV patients. In addition, this signal is not an artifact of heterogeneous hypermutation patterns in BCRs, as shown by simulations in [supplementary figure S4, Supplementary Material](#) online.

A similar signal of adaptation based on the upturn in the SFS has been observed among BCR lineages in response to influenza vaccine in healthy individuals (Horns et al. 2019). Although the upturn of SFS is often used as a standard signal for selection in population genetics, it has low power in distinguishing between hitchhiking under selection or out-of-equilibrium effects due to population structure in neutrality (Jensen et al. 2005) (see [Supplementary Material](#) online). In particular, the signal may be confounded in expanding populations of B-cells during transient response to acute infections or vaccination.

### Inferring Intralinearage Selection and Interference from Longitudinal Data

To understand the dynamics and fate of adaptive mutations during chronic infection, we use the longitudinal nature of the data to analyze the temporal structure of the lineages. We estimate the likelihood that a new mutation appearing in a certain region of a BCR reaches frequency  $x$  at some later time within the lineage ([fig. 3a](#)), and evaluate a measure of selection  $g(x)$  as the ratio of this likelihood between nonsynonymous and synonymous mutations (Strelkova and Lässig 2012) (see Materials and Methods and [Supplementary](#)



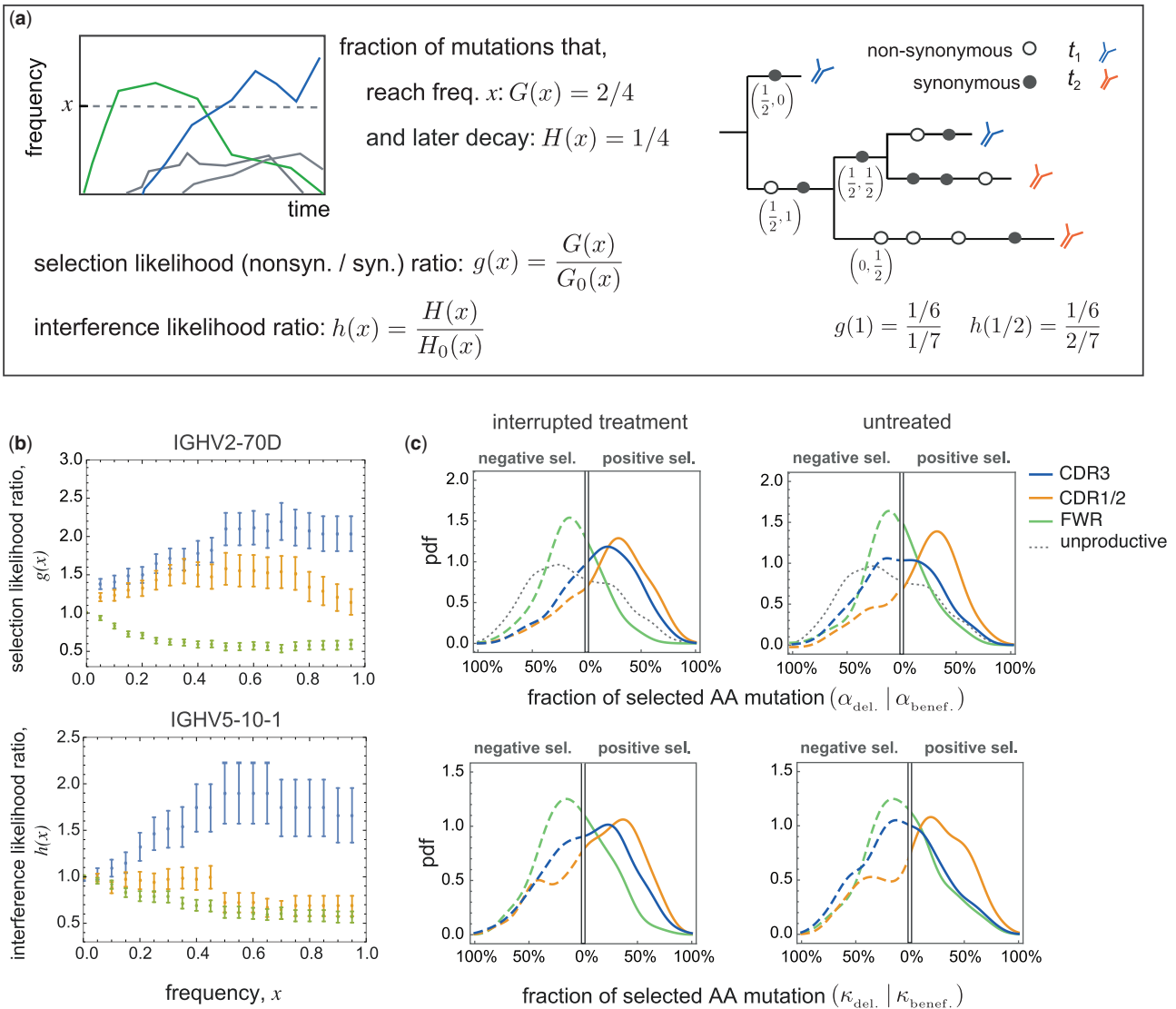
**Fig. 2.** Statistics of BCR lineage genealogies indicate positive selection. (a) The U-shaped distribution of sublineage weight ratios at the root of lineage trees (see [Supplementary Material](#) online)  $w_D/w_{anc.}$  and (b) the distribution of elongated mean terminal branch length (in units of divergence time) relative to the mean length of all branches in BCR lineages indicate positive selection in HIV patients and in healthy individuals (colors), in contrast to the neutral expectation (dotted lines); see [supplementary figure S2, Supplementary Material](#) online, for comparison of tree statistics under different evolutionary scenarios. (c) The site frequency spectrum (SFS)  $f(v)$  is shown for mutations in different regions of BCRs (distinct colors) in HIV patients (left) and in healthy individuals (right); see [supplementary figure S3, Supplementary Material](#) online, for SFS of unproductive BCR lineages. The frequencies are estimated within each lineage and the distributions are aggregates across lineages of size  $> 100$ , amounting to a total of 1,524 lineages in HIV patients and 2,795 lineages in healthy individuals; see [supplementary table S1, Supplementary Material](#) online for details. The gray area shows the span of SFS across 100 realizations of simulated neutral trees (Kingman’s coalescent) with sizes equal to the BCR lineages in HIV patients (left) and in healthy individuals (right). The significant upturn of the SFS for nonsynonymous mutations in the CDR3 region is indicative of rapid evolution under positive selection. The upturn for synonymous mutations indicates hitchhiking of neutral mutations along with the positively selected alleles (see [Supplementary Material](#) online).

[Material](#) online). The frequency of a mutation  $x_t$  is estimated as the relative size of its descendent clade at time  $t$  (number of leaves in its subclade) to the total number of leaves in the lineage at that time ([fig. 3a](#)). At frequency  $x = 1$  (i.e., substitution), the likelihood ratio  $g(x)$  is equivalent to the McDonald–Kreitman test for selection ([McDonald and Kreitman 1991](#)). Generalizing it to  $x < 1$  makes it a more flexible measure applicable to the majority of mutations that only reach intermediate frequencies. Similar to McDonald–Kreitman test, the likelihood ratio  $g(x)$  is relatively robust to effects due to demography in comparison to the SFS, as both synonymous and nonsynonymous mutations experience similar demographic biases.

A major reason why many beneficial mutations never fix in a lineage is clonal interference, whereby BCR mutants within and across lineages compete with each other ([Nourmohammad et al. 2016](#)). Clonal interference in population genetics refers to a specific regime of evolution by natural selection, where multiple beneficial mutations simultaneously and independently arise on different genetic backgrounds and form competing clones. Here, we use the population genetics definition of a “clone,” which refers to the descendants (i.e., subclade) of a given mutation in a lineage phylogeny, and although related, it should not be

confused by the immunological analogue in “clonal selection theory” ([Burnet 1976](#)). In the absence of clonal interference, beneficial mutations can readily fix after they rise to intermediate frequencies, beyond which stochastic effects cannot impact their fate ([Desai and Fisher 2007](#)) (see [Materials and Methods](#) and [Supplementary Material](#) online). Clonal competition among beneficial mutations is common in large adaptive asexual populations and reduces the rate of evolution by slowing down the successive fixation of beneficial mutations ([Schiffels et al. 2011](#)). In this evolutionary regime, the dynamics of beneficial mutations becomes more neutral ([Schiffels et al. 2011](#)), resulting in a reduced efficacy of selection that hinders the emergence of very fit strains (e.g., a high-affinity BCR). Moreover, the nonlinearity due to competition among clones reduces the predictability of the fate of beneficial mutations during evolution ([Lässig et al. 2017](#)).

To quantify the prevalence of clonal interference, we evaluate the nonsynonymous-to-synonymous ratio  $h(x)$  as the likelihood for a mutation to reach frequency  $x$  and later to go extinct ([Strelkova and Lässig 2012](#)) (see [fig. 3a](#), [Materials and Methods](#), and [Supplementary Material](#) online). In short, the selection likelihood  $g(x)$  identifies “surges” and interference likelihood  $h(x)$  “bumps” in frequency trajectories of clones. These likelihood ratios have intuitive interpretations:



**Fig. 3.** Inference of selection and clonal interference in BCR lineages. (a) Schematic shows time-dependent frequencies of four distinct mutations that rise within a population (left). We denote the fraction of mutations that reach frequency  $x$  (at any time point) within a population (or lineage) by  $G(x)$  (blue and green) and the subset that later goes extinct due to clonal interference by  $H(x)$  (green). Classifying the mutations into nonsynonymous and synonymous groups, we quantify the strength of selection using the likelihood ratios between the nonsynonymous and synonymous mutations,  $g(x) = G(x)/G_0(x)$ , and  $h(x) = H(x)/H_0(x)$ . A schematic B-cell genealogy (right) is shown for a lineage sampled at two time points (colors). Nonsynonymous and synonymous mutations are shown by empty and filled circles and their time-dependent frequencies ( $x_{t_1}, x_{t_2}$ ), as observed in the sampled tree leaves, are indicated below a number of branches. The corresponding likelihood ratios are given below. (b) Selection likelihood ratio  $g(x)$  in the V-gene class IGHV2-70D (top; pooled from 35 lineages) and the interference likelihood ratio  $h(x)$  for the V-gene class IGHV5-10-1 (bottom; pooled from 18 lineages) in patient 5 are plotted against frequency  $x$  for mutations in different BCR regions (colors). The likelihood ratios indicate positive selection and strong clonal interference in the CDR3 region, negative selection on the FWR region and positive selection on mutations that rise to intermediate frequencies in the joint CDR1/CDR2 regions. We do not observe interference in the FWR and the joint CDR1/CDR2 regions. The error bars are estimated assuming a binomial sampling of the mutations (see [Supplementary Material](#) online). (c) Each panel shows the probability density across distinct VJ-gene classes in HIV patients with interrupted treatment (left) and without treatment (right), of the fractions of beneficial and deleterious mutations  $\alpha_{benef.}$  and  $\alpha_{del.}$  on the right  $x$ -axis and the left inverted  $x$ -axis, respectively, that reach frequency  $x = 80\%$  within a lineage (top), and similarly, the fractions of beneficial and deleterious mutations ( $\kappa_{benef.}$  and  $\kappa_{del.}$ ) that reach frequency  $x = 60\%$  within a lineage and later go extinct (bottom). The fraction of selected mutations  $\alpha, \kappa$  are estimated based on the deviation of the likelihood ratios,  $g(x)$  and  $h(x)$ , from 1 in VJ-gene classes within each patient separately (see Materials and Methods and [Supplementary Material](#) online). These aggregate statistics are then pooled together to form the histograms, without averaging over VJ-genes across patients. The dotted gray line indicates the null distribution from unproductive lineages of healthy individuals ([supplementary fig. S8, Supplementary Material](#) online). The probability densities are evaluated from 13,601 lineages and aggregated over 661 VJ-gene classes pooled from the four patients with interrupted treatment (left), from 7,043 lineages with 373 VJ-gene classes pooled from the two untreated patients (right) and from 2,903 unproductive lineages with 417 VJ-gene classes pooled from three healthy individuals (dotted gray). The color code for distinct BCR regions in all panels is consistent with the legend; see [Supplementary Material](#) online for statistical details, [supplementary table S1, Supplementary Material](#) online, for details on lineages and [figure 4 and supplementary figs. S5, S6, S8, and S10, Supplementary Material](#) online, for further analysis of likelihood ratios and selection statistics.

$g(x) > 1$  shows overrepresentation of nonsynonymous to synonymous mutations at frequency  $x$  and indicates evolution under positive selection, with a fraction of at least  $\alpha_{\text{benef.}}(x) = 1 - 1/g(x)$  strongly beneficial amino acid mutations in a given region (Smith and Eyre-Walker 2002). On the other hand, the likelihood ratio  $g(x) < 1$  is indicative of negative selection, where nonsynonymous mutations are suppressed, with a fraction of at least  $\alpha_{\text{del.}}(x) = 1 - g(x)$  strongly deleterious mutations (see [Supplementary Material](#) online for a derivation of these bounds). Likewise,  $\kappa_{\text{benef.}}(x) = 1 - 1/h(x)$  or  $\kappa_{\text{del.}}(x) = 1 - h(x)$  define a lower bound on the fraction of either beneficial or deleterious mutations that go extinct.

### Region-Specific Patterns of Intralineage Selection and Interference

To demonstrate the structure of the signal, [figure 3b](#) shows the selection likelihood ratio  $g(x)$  in an HIV patient (patient 5) for lineages belonging to a typical V-gene class IGHV2-70D (see Materials and Methods); see [supplementary fig. S5, Supplementary Material](#) online, for repertoire averaged statistics in all individuals. In this gene family, we detect positive selection ( $g > 1$ ) in the CDR3 region. We observe  $\sim 10\%$  of the 854 nonsynonymous mutations in this gene-family reach frequency  $x = 0.6$ , in comparison to only 5% of the 884 synonymous mutations; mutations are pooled across 35 lineages with an average CDR3 length of 45 bp. Therefore, the selection likelihood ratio  $g(x)$  has around a 2-fold larger fraction of nonsynonymous compared with synonymous mutations in the CDR3 region, which indicates that at least  $\alpha_{\text{benef.}} = 46\%$  of mutations that reach frequency  $x = 0.6$  are strongly beneficial. On the other hand, the likelihood ratio in FWR signals strong negative selection ( $g < 1$ ), where nonsynonymous mutations reaching frequencies  $x = 0.6$  are two times less frequent than the synonymous mutations, which indicates at least  $\alpha_{\text{del.}} = 37\%$  of these mutations are strongly deleterious. In FWR, we identify 3,987 nonsynonymous mutations, 2% of which reach frequency  $x = 0.6$ , in comparison to 4% of the 3,114 synonymous mutations; the average FWR length among the 35 pooled lineages is 213 bp. Similarly, the interference likelihood ratio  $h(x)$  for a V-gene class IGHV5-10-1 in patient 5 indicates that at least  $\kappa_{\text{benef.}} = 37\%$  of CDR3 mutations in this gene family that go extinct due to clonal interference are strongly beneficial ([fig. 3b](#)). This likelihood ratio is estimated based on the observed 16% of the 231 nonsynonymous mutations that reach frequency  $x = 0.6$  and later go extinct, in comparison to 8% of the 190 synonymous mutations, pooled from 18 lineages that span over multiple time points, with an average CDR3 length of 45 bp. We should emphasize that the mutation frequencies  $x$  used for statistics of a gene-family are evaluated within their respective lineages but the likelihood ratios  $g(x)$ ,  $h(x)$  and their uncertainty estimates are aggregate measures in the given gene family (see [Supplementary Material](#) online).

To see how these observations generalize at the repertoire level, we quantify the region-specific fraction of beneficial and deleterious mutations within BCR lineages of distinct gene classes and also the fraction of selected mutations that are

impeded by clonal interference ([fig. 3c](#) and [table 1](#)). Overall, we observe that a substantial fraction of lineages (aggregated into VJ-gene classes) carry positively selected amino acid mutations in their CDR regions and negatively selected amino acid mutations in FWRs. We infer that at least  $\bar{\alpha}_{\text{benef.}} = 12 - 30\%$  of CDR mutations that reach frequency  $x = 0.8$  are strongly beneficial and  $\bar{\alpha}_{\text{del.}} = 16 - 20\%$  of FWR mutations that reach frequency  $x = 0.8$  are strongly deleterious ([table 1](#)); overbars indicate averages over VJ-gene classes. [Figure 4](#) shows the detailed statistics of selected mutations in each patient and [supplementary figure S7, Supplementary Material](#) online, shows the inferred effective selection strengths  $\bar{s}$  for different V-gene classes (see Materials and Methods and [Supplementary Material](#) online). The inferred effective selection strengths within the repertoire indicate a significantly larger fraction of V-gene classes to carry positively selected alleles in their CDRs as opposed to the overrepresented negatively selected alleles in FWRs ([supplementary fig. S7, Supplementary Material](#) online). A similar region-specific selection pattern is evident in healthy individuals ([supplementary fig. S8, Supplementary Material](#) online).

### Macro- and Microevolutionary Selection Fluctuations Shape the Structure of BCR Lineages

The frequency-dependent behavior of the selection likelihood ratio in [figure 3](#) is a strong indicator for the underlying evolutionary mode. In the absence of any competition and clonal interference (i.e., independent site model), the likelihood for a beneficial (deleterious) mutation to reach high frequencies should deviate strongly from the neutral expectation, leading to a rapidly increasing (decreasing) likelihood ratio  $g(x)$  as a function of the frequency  $x$ ; see Materials and Methods and [Supplementary Material](#) online, for theoretical expectation in this regime. As shown in [supplementary figure S6, Supplementary Material](#) online, the data significantly deviates from the expected behavior of the selection likelihood ratio  $g(x)$  for independent site evolution under selection. Competition among beneficial mutations reduces the rate of BCR adaptation by slowing down the successive fixation of beneficial mutations and can ultimately hinder the evolution of high-affinity BCRs. This clonal interference effectively reduces the efficacy of selection (Schiffels et al. 2011) and can lead to flattening of the selection likelihood ratio at high frequencies, consistent with [figure 3](#). Until a secondary competing allele takes up a substantial fraction of a lineage, the dynamics of the focal allele may be characterized by selection with uncorrelated fluctuations, for example, due to spontaneous environmental noise such as access to T-cell help or signaling molecules. We describe a theoretical model for this process in Materials and Methods and show that the flattening of the likelihood ratio, up to intermediate frequencies, can be explained by an effective selection strength  $\bar{s}$  subject to rapid *microevolutionary* fluctuations with an amplitude  $v$ ; see Materials and Methods and [Supplementary Material](#) online, and the fitted model to the likelihood ratios in [supplementary figure S6, Supplementary Material](#) online.

The interference likelihood ratio  $h(x)$  captures the long-term turnover of circulating alleles, which we characterize by

using the extinction probability of a rising allele under various evolutionary scenarios (see [Supplementary Material](#) online). New beneficial mutations overcome the risk of stochastic extinction by genetic drift once they reach the establishment frequency, which is inversely proportional to their selection coefficient (Desai and Fisher 2007). Therefore, for the majority of strongly beneficial mutations at high frequencies, extinction by genetic drift is an unlikely scenario. As we discussed earlier, rapid microevolutionary fitness fluctuations due to the environmental noise can slow down the rise of beneficial mutations in a lineage. An allele which is on an average deleterious can have a positive selection coefficient at some instances due to fitness fluctuations and intermittently hamper the rise of the dominant beneficial allele. However, these fluctuations do not persist long enough for a deleterious mutant to fully replace an established beneficial allele. Therefore, neither microevolutionary fluctuations nor drift can explain the mutation turnovers observed in [figure 3](#) and [supplementary figure S6, Supplementary Material](#) online.

We explain the rise and fall of beneficial mutations with selection strength  $s_0$  by clonal interference where a new beneficial mutation with selection strength  $s_1$  arises on a distinct (formerly deleterious or neutral) genetic background and outcompetes the circulating allele. We model this process as evolution with *macroevolutionary* selection fluctuations that occur at rates lower than the lifetime of a polymorphism in a lineage (Mustonen and Lässig 2009). In this picture, a rising new beneficial allele with selection strength  $s_1$  makes a shift in selection coefficient of the dominant allele  $s_0 \rightarrow s_0 - s_1$ . The persistence of such fitness shifts over macroevolutionary time scales can lead to successive turnover of new beneficial mutations in a lineage, consistent with the observation in [supplementary figure S6, Supplementary Material](#) online. The comparison between the interference likelihood ratios  $h(x)$  predicted by different evolutionary scenarios strongly indicates the prevalence of clonal interference and macroevolutionary selection fluctuations in shaping the structure of BCR lineages for the V-gene classes in [supplementary figure S6, Supplementary Material](#) online.

Overall, we observe that the positively selected mutations in CDR3 and the pooled CDR1/CDR2 regions are strongly impacted by clonal interference, in contrast to mutations in FWR ([fig. 3c](#) and [table 1](#)). In particular, using the interference likelihood ratio  $h(x)$ , we infer a significant macroevolutionary turnover in the preference for the selected alleles (see Materials and Methods and [Supplementary Material](#) online). [Supplementary figure S9, Supplementary Material](#) online, shows that strongly selected alleles in a lineage are often replaced by a competing allele, with a selection strength larger than expected. These observations indicate the abundance of beneficial mutations, leading to pervasive clonal interference and a long-term selection turnover in the regions of the BCR with the most important functional role, at the repertoire level. Importantly, our simulations of affinity maturation show that the inference of region-specific selection and clonal interference in BCRs is insensitive to the heterogenous hypermutation statistics and the presence of mutational hotspots

**Table 1.** Beneficial and Deleterious Mutations in BCRs.

	HIV Infected Untreated			HIV Infected Interrupted Treatment			Healthy Productive			Healthy Unproductive		
	$\bar{\alpha}_{\text{benef.}}$	$\bar{\alpha}_{\text{del.}}$	$\bar{\kappa}_{\text{benef.}}$	$\bar{\alpha}_{\text{benef.}}$	$\bar{\alpha}_{\text{del.}}$	$\bar{\kappa}_{\text{benef.}}$	$\bar{\alpha}_{\text{benef.}}$	$\bar{\alpha}_{\text{del.}}$	$\bar{\kappa}_{\text{del.}}$	$\bar{\alpha}_{\text{benef.}}$	$\bar{\alpha}_{\text{del.}}$	$\bar{\kappa}_{\text{del.}}$
CDR3	0.12 ± 0.01	0.14 ± 0.01	0.11 ± 0.01	0.20 ± 0.01	0.08 ± 0.01	0.17 ± 0.01	0.31 ± 0.01	0.03 ± 0.004	0.12 ± 0.01	0.09 ± 0.01	0.21 ± 0.01	0.24 ± 0.01
CDR1/2	0.24 ± 0.01	0.07 ± 0.01	0.22 ± 0.02	0.27 ± 0.01	0.07 ± 0.01	0.23 ± 0.01	NA	NA	0.10 ± 0.01	NA	NA	NA
FWR	0.08 ± 0.01	0.13 ± 0.01	0.09 ± 0.01	0.05 ± 0.01	0.18 ± 0.01	0.08 ± 0.01	0.07 ± 0.01	0.22 ± 0.01	0.17 ± 0.01	0.11 ± 0.01	0.24 ± 0.01	0.24 ± 0.01

NOTE.—The average fraction of beneficial  $\bar{\alpha}_{\text{benef.}}$  and deleterious  $\bar{\alpha}_{\text{del.}}$  mutations that reach frequency  $x = 80\%$  (based on selection likelihood ratio  $g(0.8)$ ) in different regions of BCRs are reported for HIV patients (with interrupted and without treatment) and for healthy individuals (productive and unproductive lineages); averages are estimated over VJ-gene classes. Similarly, the average fraction of beneficial  $\bar{\kappa}_{\text{benef.}}$  and deleterious  $\bar{\kappa}_{\text{del.}}$  mutations that reach frequency  $x = 60\%$  followed by extinction (based on interference likelihood ratio  $h(0.6)$ ) are reported for HIV patients with interrupted and without treatment; we cannot estimate the interference likelihood ratio in healthy individuals due to the lack of time-resolved data. The errors indicate the SEM. The corresponding distributions are presented in [figure 3](#) and [supplementary figure S8, Supplementary Material](#) online. In addition, see Materials and Methods and [supplementary figure S5, Supplementary Material](#) online; and [figure 4](#) and [supplementary figure S7, Supplementary Material](#) online, for detailed comparison of selection and clonal interference statistics among the patient groups. The sequence reads in healthy individuals (DeWitt et al. 2016) are shorter than in HIV patients (Hoehn et al. 2015) and do not extend to CDR1/2.



in CDRs (see Materials and Methods and [supplementary fig. S10, Supplementary Material](#) online).

In short, we observe a large fraction of adaptive mutations, and also a substantial amount of clonal interference among them which prevents some of the mutations from dominating within lineages.

### Viral Expansion Drives the BCR Repertoire Response with Strong Selection and Clonal Interference

In patients with interrupted ART, we infer a substantially larger fraction of beneficial mutations to rise with strong clonal interference in pathogen-engaging CDR3 regions following the interruption of treatment, compared with the ART-naive patients with a stable chronic infection ([table 1](#) and [fig. 3](#)); The CDR3 statistics of the two patient groups are significantly distinct based on the two-sample KS test for the selection statistics,  $P$  value =  $5 \times 10^{-8}$  and for the clonal interference statistics,  $P$  value =  $5 \times 10^{-4}$ . Such a shift is not present for mutations in CDR1, CDR2, and FWR ( $P$  values  $>0.1$ , two-sample KS test); see [figure 3](#), [table 1](#), and [Supplementary Material](#) online. Moreover, we observe that the expansion of the HIV population is met with strong positive selection and clonal interference of beneficial mutations in BCRs. Specifically, selection and clonal interference in the CDR3 region strongly correlate with changes in viral load during the 2.5 years of study ([fig. 4](#) and [supplementary fig. S7](#) and [table S2, Supplementary Material](#) online). No such correlation is observed in CDR1, CDR2, and FWR ([fig. 4](#)). This result is consistent with our inference of strong positive selection and clonal interference in CDR3 of patients who had terminated ART after the first year of treatment, and hence, have the largest change in their viral load.

This evolutionary pattern is consistent with the rate of HIV-1 evolution in patients with different states of therapy. Genome-wide analysis of HIV-1 has revealed that evolution of the virus within ART-naive patients slows down during chronic infections with the majority of mutations happening as reversions rather than immune escape and with limited clonal interference in viral populations ([Zanini et al. 2015](#)). In a separate study by SPARTAC ([Roberts et al. 2015](#)), ART-naive patients show a slow and steady viral escape from the CTL immune response over the first 2 years of infection. Our analysis suggests that the response at the repertoire level traces the slow evolution of the virus during the chronic phase. On the other hand, rapid expansion of HIV-1 following the interruption of ART drives a strong immune response. We hypothesize that evolution of the HIV-1 population during expansion introduces a time-dependent target for the adaptive immune system and opens up room for many beneficial mutations in the pathogen-engaging CDR3 regions. The emergence of beneficial mutations on separate backgrounds results in evolution with clonal interference between clones in the repertoire ([figs. 3 and 4](#)).

## Discussion

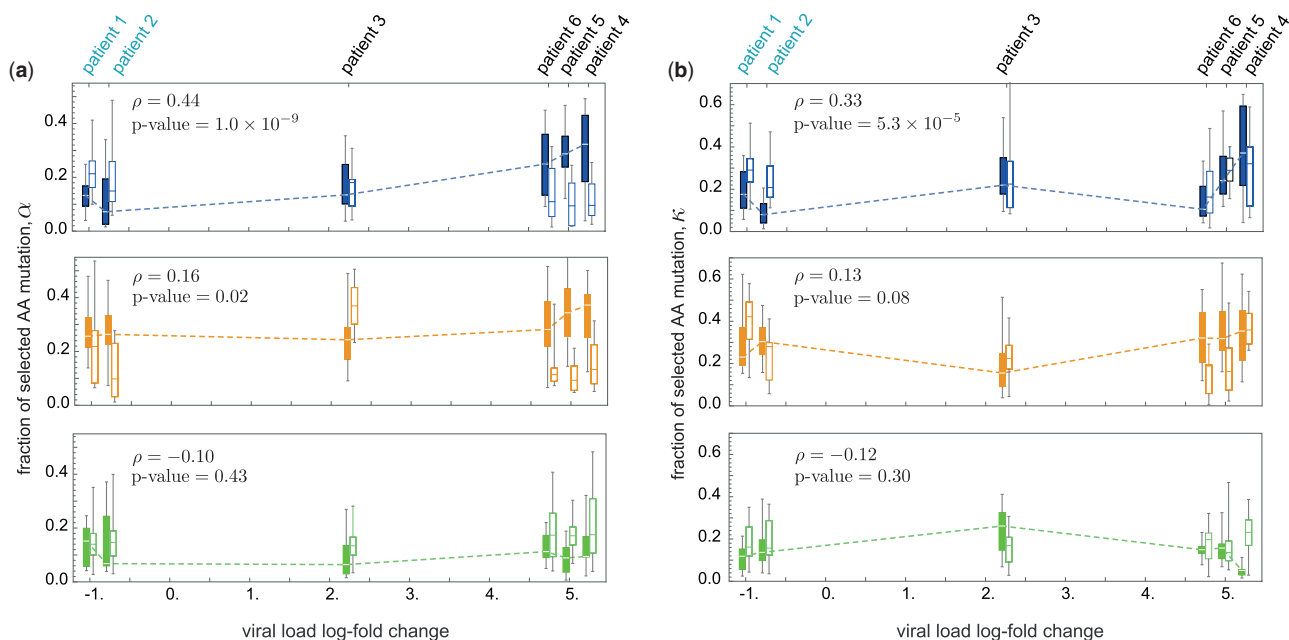
Somatic evolution during affinity maturation is complex: there is no one winner of the race for the best antibody.

We show that the B-cell repertoire mounts a relatively slow response to the stable chronic HIV during the early stages of infection in ART-naive patients. On the other hand, following the interruption of ART in a number of patients, the expanding HIV population drives strong affinity maturation in B-cells with rapid dynamics and clonal interference. Overall, the change in viral load correlates with the strength of selection and clonal interference in BCR repertoires ([fig. 4](#)). Expansion and growth of the HIV-1 population is often accompanied with rapid evolution of the virus, which can exert a strong selection on the adaptive immune system. We hypothesize that the observed strong selection on BCRs is to counter the viral evolution during its expansion. The extent of such co-evolution can be tested in future experiments that trace the inpatient dynamics of BCR and HIV-1 populations over time.

The lack of sequence fixation in a repertoire has been previously observed at the level of monoclonal antibody lineages in response to vaccination in mouse models ([Berek and Milstein 1987](#)) and among many rising BCR clones over short time scales ( $\sim$  weeks), during a transient response of human immune repertoires to the influenza vaccine ([Horns et al. 2019](#)). Many factors, including idiotypic interactions, resulting in frequency-dependent selection, or spatial structure of a population have been hypothesized to contribute to the large scale shifts in the repertoire structure, leading to a constant rise of beneficial clones within a repertoire. Here, we provide a principled approach to characterize *macroevolutionary* shifts of immune repertoires ([figs. 3 and 4](#); [supplementary fig. S9, Supplementary Material](#) online) and show that the somatic evolution of BCRs is not limited by beneficial mutations, the supply of which can last over many years of a chronic infection.

The dynamics of an adaptive immune response resembles rapid evolution in asexual populations where many beneficial mutations rise to intermediate frequencies leading to complex clonal interference and genetic hitchhiking. Such evolutionary dynamics is prominent in microbial populations ([Lässig et al. 2017](#)), in viruses including HIV within a patient ([Pandit and De Boer 2014](#); [Zanini et al. 2015](#)) and global influenza ([Strelkova and Lässig 2012](#); [Łuksza and Lässig 2014](#); [Neher et al. 2014](#)). In this evolutionary regime different beneficial mutations arise at nearly the same time and compete with each other, reducing the rate at which beneficial mutations can accumulate. This is distinct from selection, which is merely a difference in growth rate or survival of different cells. Clonal interference can result from competition of BCRs for the same antigen or other stimulatory or activation factors. In HIV patients, we expect that as long as the CD4<sup>+</sup> T-cell levels stay at its normal range (500–1,500 per ml) to activate a large population of B-cells, as it is the case in this study ([supplementary table S2, Supplementary Material](#) online), clonal interference among many positively selected mutations which chase the viral evolution should remain the prominent mode of somatic evolution in BCRs on long time-scales.

On one hand, the nonlinear clonal competition among BCRs reduces the rate of adaptation during affinity



**Fig. 4.** Fraction of beneficial and deleterious mutations in BCRs of all patients. (a) The box plots show the distributions across V-gene classes for the fraction of beneficial  $\alpha_{\text{benef.}}$  (filled boxes) and deleterious  $\alpha_{\text{del.}}$  (empty boxes) mutations that reach frequency  $x = 80\%$  within a lineage in each BCR region: CDR3 (top, blue), CDR1/2 (middle, orange), and FWR (bottom, green); midline: median; box: 40% around median; bars: 80% around median. (b) shows similar statistics for the fraction of beneficial and deleterious mutations ( $k_{\text{benef.}}$ ,  $k_{\text{del.}}$ ) that reach frequency  $x = 60\%$  within a lineage and later go extinct, measuring the extent of clonal interference in B-cells. The patients are positioned along the x-axis in accordance to their viral load change  $\Delta V$  over the course of infection, estimated as the difference of the averaged log-viral load between the second year ( $t > 48$  weeks) and the first year ( $t \leq 48$  weeks) of the study,  $\Delta V = \langle \log V \rangle_{\text{yr.2}} - \langle \log V \rangle_{\text{yr.1}}$ ; see [supplementary table S2, Supplementary Material](#) online. Patients 1–2 are ART-naïve and the rest had interrupted their treatments after week 48. The dashed line in each figure traces the median for the distribution of the beneficial mutations (filled boxes) as a function of the change in viral load. In the CDR3 region, positive selection and clonal interference among beneficial mutations are significantly stronger in patients with a larger change in viral load, with Spearman rank correlation and significance indicated in the corresponding panels. The correlations in other regions are small  $\sim 0.1$  and insignificant with  $P$  values  $> 0.01$ . Estimated correlations for the fraction of negatively selected mutations are small and insignificant (the values are not shown). Correlations are estimated from the whole data and not only the median of the distributions, indicated by the dashed lines.

maturation, leading to a less predictable fate for good clones, as they can be outcompeted by new mutants before they dominate the immune response. On the other hand, in this evolutionary regime, the fate of a lineage is not strongly impacted by stochastic uncertainties due to waiting for new beneficial mutations to arise, as a large supply of such mutations is available in response to a pathogen. Thus, we hypothesize that it should be feasible to infer *fitness models* that primarily rely on the selection differences among circulating BCRs to forecast the outcome of an immune response. Similar approaches have been previously successful in forecasting the fate of a selection-dominated evolving process to predict the annually dominant strain of the influenza virus (Łuksza and Lässig 2014; Neher et al. 2014) and the response of evolving tumors to cancer immunotherapy (Łuksza et al. 2017). Predicting the outcome and efficacy of B-cell response is of significant consequence for designing targeted immune-based therapies. Currently, the central challenge in HIV vaccine research is to devise a means to stimulate a lineage producing highly potent broadly neutralizing antibodies (BnAbs). A combination of successive immunization and ART has been suggested as an approach to elicit a stable and effective BnAb response (Caskey et al. 2016). An optimal treatment strategy should leverage the information on the selected

clones among BCRs during a rapid immune response to antigen stimulation, to overcome the nonlinear impact of clonal interference and derive the immune response toward a desired BnAb within the repertoire.

## Materials and Methods

### B-Cell Repertoire Data, Annotation, and Genealogies

We analyze B-cell repertoire data from six HIV patients from Hoehn et al. (2015) with raw sequence reads accessible from the European Nucleotide Archive under study accession numbers, ERP009671 and ERP000572. The data covers  $\sim 2.5$  years of study with six to eight sampled time points per patient; see [supplementary table S1, Supplementary Material](#) online, for details. The B-cell repertoire sequences consist of 150-bp nonoverlapping paired-end reads (Illumina MiSeq), with one read covering much of the V gene and the other read covering the area around the CDR3 region and the J gene. We analyze memory B-cell repertoire data of three individuals published in DeWitt et al. (2016). We annotate the BCR repertoire sequences of each individual (pooled time points) by Partis (version 0.11.0) (Ralph and Matsen 2016) and further process by MiXCR (Bolotin et al. 2015, 2017). To identify BCR lineages, we first group sequences by the

assigned V gene, J gene, and CDR3 length, and then used single linkage clustering with a threshold of 90% Hamming distance. We reconstruct a maximum-likelihood genealogical tree for sequences in each lineage. We use FastTree (Price et al. 2010) to construct the initial tree by maximum parsimony. We use this tree as seed for the maximum likelihood construction of the phylogeny with RAxML (Stamatakis 2014), using the GTRCAT substitution model. Details of data processing and error corrections and genealogy reconstruction are discussed in [supplementary figure S1, Supplementary Material](#) online.

### Selection and Interference Likelihood Ratio

Hypermutations during affinity maturation create new clades within a lineage. The frequency  $x$  of these clades changes over time (fig. 3a). A mutation under positive (or negative) selection should reach a higher (lower) frequency than a neutral mutation. We quantify the likelihood of a mutation reaching a frequency  $x$  in its lifetime by  $G(x) \equiv n(x)/N$ , where  $n(x)$  is the number of mutations that reach frequency  $x$  and  $N$  is the total number of mutations. We determine the selection likelihood ratio between nonsynonymous  $G(x)$  and synonymous  $G_0(x)$  mutations:

$$g(x) = \frac{G(x)}{G_0(x)}$$

We characterize clonal interference by the likelihood that a mutation reaches frequency  $x$  and later goes extinct,  $H(x) = G(x) \times G(0|x)$  (Strelkowa and Lässig 2012);  $G(0|x)$  is the conditional probability that a mutation starting at frequency  $x$  goes extinct (fig. 3a). We estimate the interference likelihood ratio by comparing the interference likelihood between nonsynonymous  $H(x)$  and synonymous mutations,  $H_0(x)$  (fig. 3a):

$$h(x) = \frac{H(x)}{H_0(x)} \equiv \frac{G(x) \times G(0|x)}{G_0(x) \times G_0(0|x)}.$$

### Affinity Maturation with Fluctuating Selection

For independently occurring mutations, the conditional probability  $G(x|x_i, t)$  that a mutation with a starting frequency  $x_i$  reaches a frequency  $x$  by time  $t$  satisfies the backward Kimura's equation (Kimura 1964):

$$\frac{\partial}{\partial t} G(x|x_i, t) = \frac{1}{2N} \left[ x(1-x) \frac{\partial^2}{\partial x^2} + s(t)x(1-x) \frac{\partial}{\partial x} \right] G(x|x_i, t),$$

with the boundary conditions  $G(x|0) = 0$  and  $G(x|x_i) = 1$ . Here,  $s(t) = \bar{s} + \eta(t)$  is a fluctuating selection coefficient with average  $\bar{s}$  and uncorrelated Gaussian fluctuations  $\eta(t)$ , with amplitude  $\nu$ , which we interpret as *microevolutionary* fluctuations, as opposed to *macroevolutionary* (i.e., long term) correlations in environmental fluctuations. The solution to the stationary state follows (Takahata et al. 1975):

$$G(x|x_i; \bar{s}, \nu) = \frac{1 - \left| \frac{1-x_i/a_+}{1-x_i/a_-} \right|^{\lambda(\bar{s}, \nu)}}{1 - \left| \frac{1-x/a_+}{1-x/a_-} \right|^{\lambda(\bar{s}, \nu)}},$$

with  $a_{\pm} = \frac{1}{2} \left[ 1 \pm \sqrt{1 + 4/\nu} \right]$  and  $\lambda(\bar{s}, \nu) = \bar{s}/(\nu \sqrt{1 + 4/\nu})$ . Expected selection likelihood ratio with microevolutionary fluctuations  $g(x; \bar{s}, \nu) = G(x|x_i; \bar{s}, \nu)/G(x|x_i; 0, \nu)$  fits the data up to intermediate frequencies  $x < 0.5$  ([supplementary fig. S6, Supplementary Material](#) online), indicating the prevalence of long-term fluctuations in the system. We assume a simple scenario with macroevolutionary shift in selection preference, where the competing allele can become more beneficial over time, resulting in interference likelihood:

$$H(x|x_i; \bar{s}_0, \bar{s}_1, \nu) = G(x|x_i; \bar{s}_0, \nu) G(1|1-x; \bar{s}_1, \nu)$$

where  $s_1 \neq -s_0$ . The expected likelihood ratio  $h(x; \bar{s}_0, \bar{s}_1, \nu) = H(x|x_i; \bar{s}_0, \bar{s}_1, \nu)/H(x|x_i; 0, 0, \nu)$  fits the data ([supplementary fig. S6, Supplementary Material](#) online). [Supplementary figure S9, Supplementary Material](#) online, shows the prevalence of such macroevolutionary fluctuations throughout the repertoire.

### Supplementary Material

[Supplementary data](#) are available at *Molecular Biology and Evolution* online.

### Acknowledgments

We are thankful to Duncan Ralph and Eric Matsen for help with implementation of the BCR analysis software Partis, and Oliver Pybus and Kenneth Hoehn for assisting us with access to the BCR data from HIV patients. A.N. acknowledges the funding from the Lewis-Sigler Institute for Integrative Genomics at Princeton University, where this work was initiated. We acknowledge the support by Max Planck Society (A.N.), SFB1310 (A.N. and A.M.W.), National Institute of Health (Grant No. T32AI055400 to J.O.), the Janssen Research & Development LLC (M.L.), and ERCCoG (Grant No. 724208 to T.M. and A.M.W.). This work was performed in part at the Aspen Center for Theoretical Physics, which is supported by the National Science Foundation (Grant No. PHY-1066293). The authors declare no competing interests.

### References

- Berek C, Milstein C. 1987. Mutation drift and repertoire shift in the maturation of the immune response. *Immunol Rev.* 96(1):23–41.
- Bolotin DA, Poslavsky S, Davydov AN, Frenkel FE, Fanchi L, Zolotareva OI, Hemmers S, Putintseva EV, Obraztsova AS, Shugay M, et al. 2017. Antigen receptor repertoire profiling from RNA-seq data. *Nat Biotechnol.* 35(10):908–911.
- Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, Chudakov DM. 2015. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods.* 12(5):380–381.
- Burnet FM. 1976. A modification of Jerne's theory of antibody production using the concept of clonal selection. *CA Cancer J Clin.* 26(2):119–121.
- Campbell CD, Eichler EE. 2013. Properties and rates of germline mutations in humans. *Trends Genet.* 29(10):575–584.

- Caskey M, Klein F, Nussenzweig MC. 2016. Broadly neutralizing antibodies for HIV-1 prevention or immunotherapy. *N Engl J Med*. 375(21):2019–2021.
- Desai MM, Fisher DS. 2007. Beneficial mutation–selection balance and the effect of linkage on positive selection. *Genetics* 17(5):385–394.
- DeWitt WS, Lindau P, Snyder TM, Sherwood AM, Vignali M, Carlson CS, Greenberg PD, Duerkopp N, Emerson RO, Robins HS. 2016. A public database of memory and naive B-cell receptor sequences. *PLoS One* 11(8):e0160853.
- Elhanati Y, Sethna Z, Marcou Q, Callan CG, Mora T, Walczak AM. 2015. Inferring processes underlying B-cell repertoire diversity. *Philos Trans R Soc B*. 370(1676):20140243.
- Hoehn KB, Gall Bashford-Rogers AR, Fidler SJ, Kaye S, Weber JN, McClure MO, SPARTAC Trial Investigators, Kellam P, Pybus OG. 2015. Dynamics of immunoglobulin sequence diversity in HIV-1 infected individuals. *Philos Trans R Soc B*. 370(1676):pii: 20140241.
- Horns F, Vollmers C, Dekker CL, Quake SR. 2019. Signatures of selection in the human antibody repertoire: selective sweeps, competing subclones, and neutral drift. *Proc Natl Acad Sci U S A*. 116(4):1261–1266.
- Jensen JD, Kim Y, DuMont VB, Aquadro CF, Bustamante CD. 2005. Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics* 170(3):1401–1410.
- Kimura M. 1964. Diffusion models in population genetics. *J Appl Probab*. 1(2):177–232.
- Kingman JFC. 1982. On the genealogy of large populations. *J Appl Probab*. 19(A):27–43.
- Laserson U, Vigneault F, Gadala-Maria D, Yaari G, Uduman M, Vander Heiden JA, Kelton W, Jung ST, Liu Y, Laserson J, et al. 2014. High-resolution antibody dynamics of vaccine-induced immune responses. *Proc Natl Acad Sci U S A*. 111(13):4928–4933.
- Lässig M, Mustonen V, Walczak AM. 2017. Predicting evolution. *Nat Ecol Evol*. 1(3):77.
- Liao H-X, Lynch R, Zhou T, Gao F, Alam SM, Boyd SD, Fire AZ, Roskin KM, Schramm CA, Zhang Z, et al. 2013. Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* 496(7446):469–476.
- Łuksza M, Lässig M. 2014. A predictive fitness model for influenza. *Nature* 507(7490):57–61.
- Łuksza M, Riaz N, Makarov V, Balachandran VP, Hellmann MD, Soloviyov A, Rizvi NA, Merghoub T, Levine AJ, Chan TA, et al. 2017. A neo-antigen fitness model predicts tumour response to checkpoint blockade immunotherapy. *Nature* 551(7681):517–520.
- McCoy CO, Bedford T, Minin VN, Bradley P, Robins H, Matsen FA. 2015. Quantifying evolutionary constraints on B-cell affinity maturation. *Philos Trans R Soc B*. 370(1676):20140244.
- McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351(6328):652–654.
- 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(1):11–23.
- Moore PL, Ranchohe N, Lambson BE, Gray ES, Cave E, Abrahams M-R, Bandawe G, Mlisana K, Abdool Karim SS, Williamson C, et al. 2009. Limited neutralizing antibody specificities drive neutralization escape in early HIV-1 subtype C infection. *PLoS Pathog*. 5(9):e1000598.
- Mustonen V, Lässig M. 2009. From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation. *Trends Genet*. 25(3):111–119.
- Neher RA, Hallatschek O. 2013. Genealogies of rapidly adapting populations. *Proc Natl Acad Sci U S A*. 110(2):437–442.
- Neher RA, Russell CA, Shraiman BI. 2014. Predicting evolution from the shape of genealogical trees. *eLife* 3:e03568.
- Nourmohammad A, Otwinowski J, Plotkin JB. 2016. Host-pathogen co-evolution and the emergence of broadly neutralizing antibodies in chronic infections. *PLoS Genet*. 12(7):e1006171.
- 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):56.
- Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* 5(3):e9490.
- Ralph DK, Matsen FA. 2016. Consistency of VDJ rearrangement and substitution parameters enables accurate B-cell receptor sequence annotation. *PLoS Comput Biol*. 12(1):e1004409.
- Richman DD, Wrinn T, Little SJ, Petropoulos CJ. 2003. Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc Natl Acad Sci U S A*. 100(7):4144–4149.
- Roberts HE, Hurst J, Robinson N, Brown H, Flanagan P, Vass L, Fidler S, Weber J, Babiker A, Phillips RE, et al. 2015. Structured observations reveal slow HIV-1 CTL escape. *PLoS Genet*. 11(2):e1004914.
- Schiffels S, Szöllösi GJ, Mustonen V, Lässig M. 2011. Emergent neutrality in adaptive asexual evolution. *Genetics* 189(4):1361–1375.
- Smith NGC, Eyre-Walker A. 2002. Adaptive protein evolution in *Drosophila*. *Nature* 415(6875):1022–1024.
- SPARTAC Trial Investigators, Fidler S, Porter K, Ewings F, Frater J, Ramjee G, Cooper D, Rees H, Fisher M, Schechter M, et al. 2013. Short-course antiretroviral therapy in primary HIV infection. *N Engl J Med*. 368(3):207–217.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
- Strelkowa N, Lässig M. 2012. Clonal interference in the evolution of influenza. *Genetics* 192(2):671–682.
- Takahata N, Ishii K, Matsuda H. 1975. Effect of temporal fluctuation of selection coefficient on gene frequency in a population. *Proc Natl Acad Sci U S A*. 72(11):4541–4545.
- Uduman M, Shlomchik MJ, Vigneault F, Church GM, Kleinstein SH. 2014. Integrating B cell lineage information into statistical tests for detecting selection in Ig sequences. *J Immunol*. 192(3):867–874.
- Victoria GD, Nussenzweig MC. 2012. Germinal centers. *Annu Rev Immunol*. 30:429–457.
- Vieira MC, Zinder D, Cobey S. 2018. Selection and neutral mutations drive pervasive mutability losses in long-lived B cell lineages. *Mol Biol Evol*. 35(5):1135–1146.
- Vollmers C, Sit RV, Weinstein JA, Dekker CL, Quake SR. 2013. Genetic measurement of memory B-cell recall using antibody repertoire sequencing. *Proc Natl Acad Sci U S A*. 110(33):13463–13468.
- Weinstein JA, Jiang N, White RA, Fisher DS, Quake SR. 2009. High-throughput sequencing of the zebrafish antibody repertoire. *Science* 324(5928):807–810.
- Yaari G, Vander Heiden JA, Uduman M, Gadala-Maria D, Gupta N, Stern JNH, O'Connor KC, Hafler DA, Laserson U, Vigneault F, et al. 2013. Models of somatic hypermutation targeting and substitution based on synonymous mutations from high-throughput immunoglobulin sequencing data. *Front Immunol*. 4:358.
- Zanini F, Brodin J, Thebo L, Lanz C, Bratt G, Albert J, Neher RA. 2015. Population genomics of inpatient HIV-1 evolution. *eLife* 4:e11282.