bioRxiv preprint doi: https://doi.org/10.1101/271130; this version posted June 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fierce selection and interference in B-cell repertoire response to chronic HIV-1

Armita Nourmohammad[1,2*], Jakub Otwinowski[1], Marta Łuksza[3], Thierry Mora[4†], Aleksandra M Walczak[5†]

[1] Max Planck Institute for Dynamics and Self-organization, Am Faßberg 17, 37077 Göttingen, Germany
[2]Department of Physics, University of Washington, 3910 15th Avenue Northeast, Seattle, WA 98195, USA
[3]Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, 1470 Madison Ave, New York, NY 10029, USA
[4]Laboratoire de Physique Statistique, CNRS, Sorbonne University, Paris-Diderot University, École Normale Supérieure (PSL), 24, rue Lhomond, 75005 Paris, France
[5]Laboratoire de Physique Théorique, CNRS, Sorbonne University, École Normale Supérieure (PSL), 24, rue Lhomond, 75005 Paris, France

* correspondence should be addressed to Armita Nourmohammad: armita@ds.mpg.de
† equal contribution

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 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 B-cell receptors (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.

bioRxiv preprint doi: https://doi.org/10.1101/271130; this version posted June 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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), around 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 $4 - 5$ 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 binding to antigens. This process of *affinity maturation* is Darwinian evolution within the host and can increase binding affinities of B-cell receptors (BCRs) up to 10-100 fold [Victora and Nussenzweig, 2012]. It generates memory and plasma B-cells with distinct receptors, forming lineages that reflect their co-evolution with viruses [Nourmohammad et al., 2016], (see schematic in Fig. 1a). A B-cell repertoire consists of many such lineages forming a forest of co-existing 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 involved in the generation of diversity in repertoires and to infer the underlying heterogenous hypermutation preferences in B-cell receptors [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 ref. [Hoehn et al., 2015], collected through the SPARTAC study [SPARTAC Trial Investigators et al., 2013]). Among these individuals are 2 untreated patients and 4 patients who had interrupted ART after a year of treat-

bioRxiv preprint doi: https://doi.org/10.1101/271130; this version posted June 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

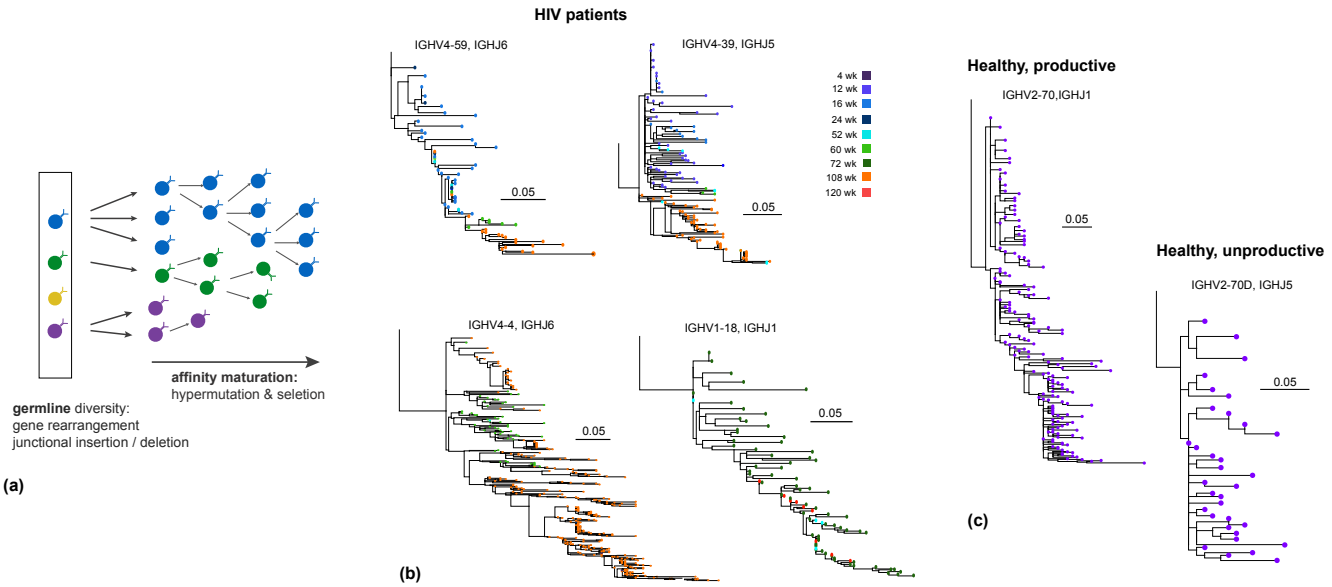

Figure 1: Affinity maturation forms B-cell lineages. **(a)** Schematic of B-cell affinity maturation and lineage formation. The naive immune repertoire consists of a diverse set of B-cell receptors, generated by gene rearrangement (VDJ recombination) and junctional sequence insertion and deletion (distinct 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 Methods and SI). 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 ~ 2.5 yrs. **(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 (Methods, SI).

ment. We have also analyzed the BCR repertoire structure in 3 healthy individuals (data from ref. [DeWitt et al., 2016]). The sequencing depth of the two datasets differ, with on 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 Methods, SI, Fig. S1 and Table S1 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 (~ 130 bp) is much smaller than in HIV patients (~ 300 bp with ~ 35 bp gap), making a direct comparison between the two datasets difficult. We have performed our statistical analysis both on the complete BCR repertoire data in healthy individuals and on the sub-sampled data with a depth comparable to the BCR repertoires in HIV patients; see SI. 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 datasets and the underlying sequencing protocols. Our primary conclusions rely on the analysis of selection in BCR repertories 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 B-cell receptor lineages inferred from BCR repertoires in each individual (Methods, SI). B-cell lineages of HIV patients, a few examples of which are shown in Fig. 1b, can persist from months to years after the initial infection, which is much longer than the lifetime of a germinal centre (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 Fig. S2) [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 distribu-

bioRxiv preprint doi: https://doi.org/10.1101/271130; this version posted June 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

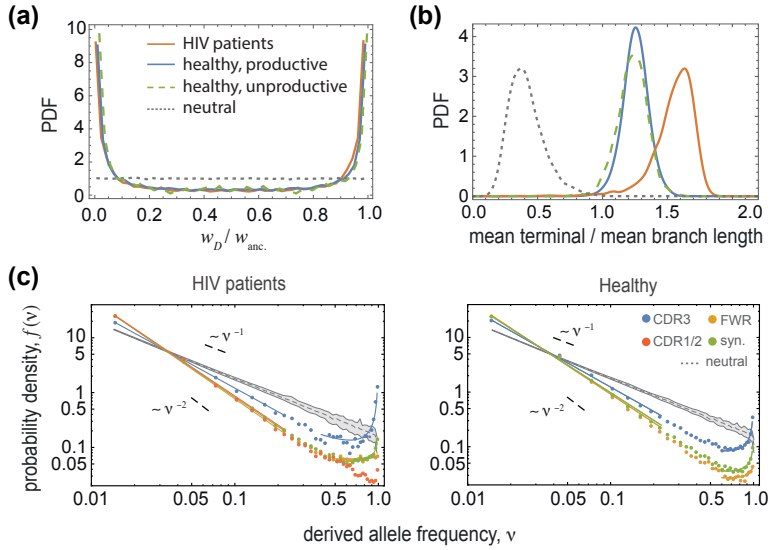

Figure 2: Statistics of BCR lineage genealogies indicate positive selection. (a) The U-shaped distribution of sub-lineage weight ratios at the root of lineage trees (SI) $w_D/w_{\text{anc.}}$ and **(b)** the distribution of elongated mean terminal branch length (in units of divergence time) relative 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 Fig. S2 for comparison of tree statistics under different evolutionary scenarios. **(c)** The Site Frequency Spectrum (SFS) $f(\nu)$ is shown for mutations in different regions of BCRs (distinct colors) in HIV patients (left) and in healthy individuals (right); see Fig. S3 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 1524 lineages in HIV patients and 2795 lineages in healthy individuals; see Table S1 for details. The grey 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 non-synonymous 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 (SI).

tion of the sub-lineage weight ratios (see SI), 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 to their internal branches, with the strongest effect seen in trees from HIV patients, again in violation of neutrality (Fig. 2b, Fig. S2); see SI for inference of coalescence time. These asymmetric features of BCR trees are clear signs of intra-lineage positive selection. They break the assumptions of neutral models that are based on non-biased growth of all terminal branches, which results in all branches and sub-lineages 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,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 ν 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 Fig. S2). 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

bioRxiv preprint doi: https://doi.org/10.1101/271130; this version posted June 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SFS of such lineages is often non-monotonic with an upturn in the high frequency part of the spectrum [Neher and Hallatschek, 2013]. We evaluate the SFS separately for synonymous and non-synonymous mutations in different regions of BCRs (Fig. 2c, Fig. S3 and SI). In HIV patients, we see a signifiant upturn of SFS polarized on non-synonymous 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, SI). In addition, we observe significant over-representation 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, Fig. S3 and SI). 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 heterogenous hypermutation patterns in BCRs, as shown by simulations in SI and in Fig. S4.

A similar signal of adaptation based on the upturn in the site frequency spectrum 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] (SI). In particular, the signal may be confounded in expanding populations of B-cells during transient response to acute infections or vaccination.

Inferring intra-lineage 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 non-synonymous and synonymous mutations [Strelkowa and Lässig, 2012] (Methods, SI). 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 sub-clade) 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 non-synonymous 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., sub-clade) 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] (Methods, SI). 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 non-linearity 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 non-synonymous-to-synonymous ratio $h(x)$

bioRxiv preprint doi: https://doi.org/10.1101/271130; this version posted June 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

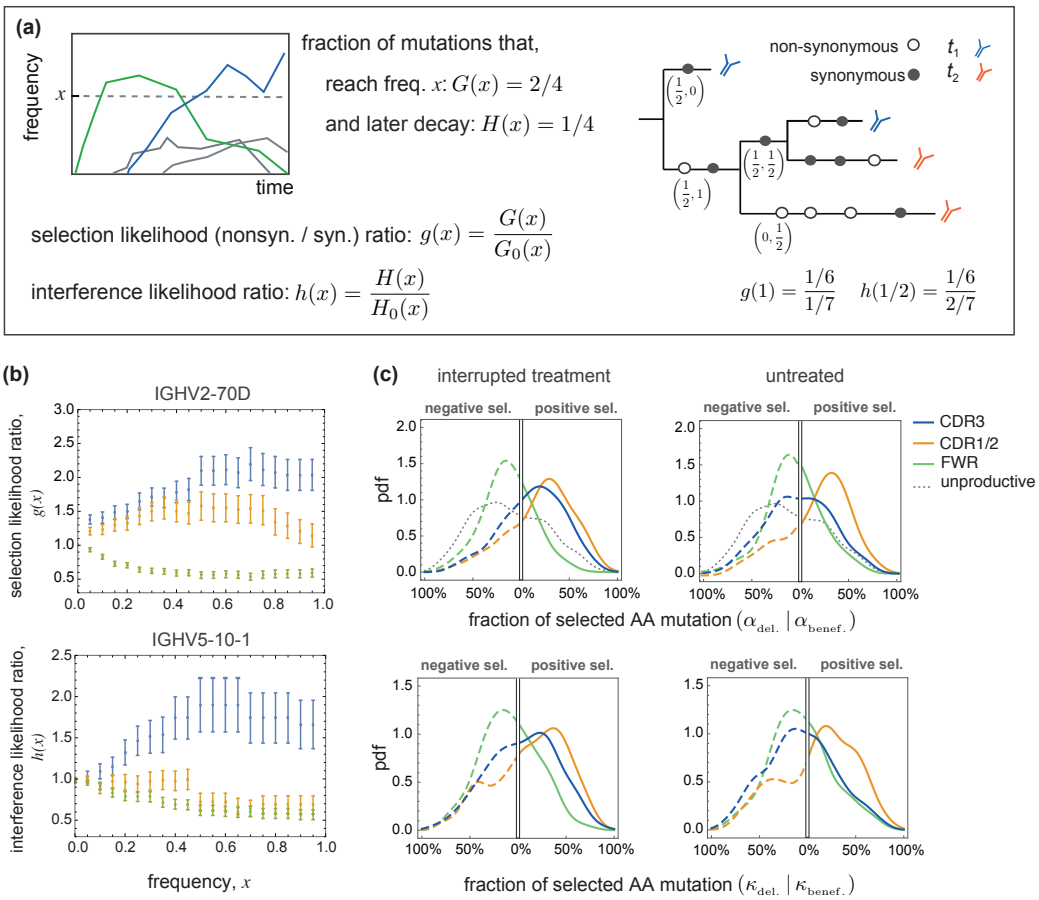

Figure 3: Inference of selection and clonal interference in BCR lineages. (a) Schematic shows time-dependent frequencies of four distinct mutations that rise within a population (left). We denote the fraction of mutations that reach frequency x (at any time point) within a population (or lineage) by $G(x)$ (blue and green) and the subset that later goes extinct due to clonal interference by $H(x)$ (green). Classifying the mutations into non-synonymous and synonymous groups, we quantify the strength of selection using the likelihood ratios between the non-synonymous 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). Non-synonymous 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 region. The error bars are estimated assuming a binomial sampling of the mutations (see SI). **(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 $\alpha_{\text{benef.}}$ and $\alpha_{\text{del.}}$ 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 ($\kappa_{\text{benef.}}$ and $\kappa_{\text{del.}}$) that reach frequency $x = 60\%$ within a lineage and later go extinct (bottom). The fraction of selected mutations α, κ 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 (Methods, SI). These aggregate statistics are then pooled together to form the histograms, without averaging over VJ-genes across patients. The dotted grey line indicates the null distribution from unproductive lineages of healthy individuals (Fig. S8). The probability densities are evaluated from 13,601 lineages and aggregated over 661 VJ-gene classes pooled from the 4 patients with interrupted treatment (left), from 7,043 lineages with 373 VJ-gene classes pooled from the 2 untreated patients (right) and from 2,903 unproductive lineages with 417 VJ-gene classes pooled from 3 healthy individuals (dotted grey). The color code for distinct BCR regions in all panels is consistent with the legend; see SI for statistical details, Table S1 for details on lineages and Fig. 4, Figs. S5, S7, S8, S10 for further analysis of likelihood ratios and selection statistics.

bioRxiv preprint doi: https://doi.org/10.1101/271130; this version posted June 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

as the likelihood for a mutation to reach frequency x and later to go extinct [Strelkowa and Lässig, 2012] (Fig. 3a, Methods and SI). 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: $g(x) > 1$ shows over-representation of non-synonymous 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 $g(x)$ smaller than 1 is indicative of negative selection, where non-synonymous mutations are suppressed, with a fraction of at least $\alpha_{\text{del.}}(x) = 1 - g(x)$ strongly deleterious mutations (see SI 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 intra-lineage selection and interference

To demonstrate the structure of the signal, Fig. 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 (Methods); see Fig. S5 for repertoire averaged statistics in all individuals. In this gene family, we detect positive selection ($g > 1$) in the CDR3 region. We observe about 10% of the 854 non-synonymous 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 two fold larger fraction of non-synonymous compared to 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 non-synonymous 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 3987 non-synonymous mutations, 2% of which reach frequency $x = 0.6$, in comparison to 4% of the 3114 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 non-synonymous 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 (SI).

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 $\overline{\alpha}_{\text{benef.}} = 12\% - 30\%$ of CDR mutations that reach frequency $x = 0.8$ are strongly beneficial and $\overline{\alpha}_{\text{del.}} = 16\% - 20\%$ of FWR mutations that reach frequency $x = 0.8$ are strongly deleterious (Table 1); overbars indicate averages over VJ-gene classes. Fig. 4 shows the detailed statistics of selected mutations in each patient and Fig. S7 shows the inferred effective selection strengths $\overline{s}$ for different V-gene classes (Methods, SI). 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 over-represented negatively selected alleles in FWRs (Fig. S7). A similar region-specific selection pattern is evident in healthy individuals (Fig. S8).

bioRxiv preprint doi: https://doi.org/10.1101/271130; this version posted June 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

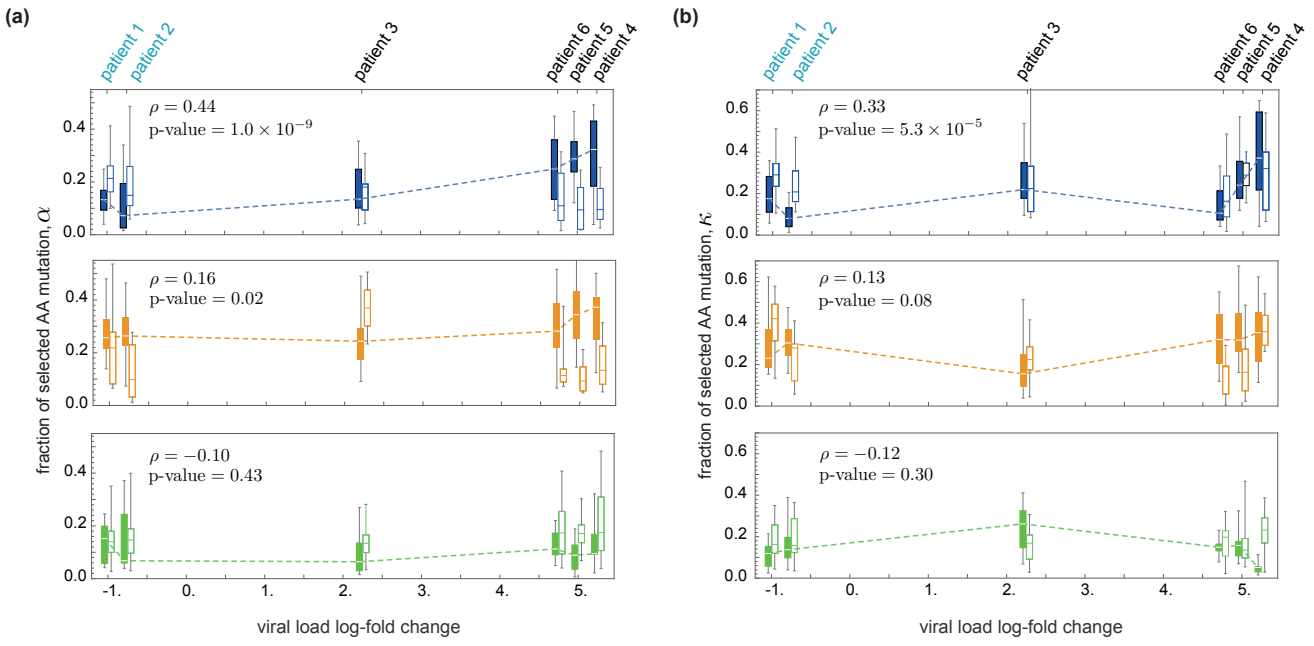

Figure 4: Fraction of beneficial and deleterious mutations in BCRs of all patients. **(a)** The box plots show the distributions across V-gene classes for the fraction of beneficial $\alpha_{\text{benef.}}$ (filled boxes) and deleterious $\alpha_{\text{del.}}$ (empty boxes) mutations that reach frequency $x = 80\%$ within a lineage in each BCR region: CDR3 (top, blue), CDR1/2 (middle, orange) and FWR (bottom, green); mid-line: median; box: 40% around median; bars: 80% around median. **(b)** shows similar statistics for the fraction of beneficial and deleterious mutations ($\kappa_{\text{benef.}}, \kappa_{\text{del.}}$) 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 ΔV over the course of infection, estimated as the difference of the averaged log-viral load between the second year ($t > 48$ wk) and the first year ($t \leq 48$ wk) of the study, $\Delta V = \langle \log V \rangle_{yr.2} - \langle \log V \rangle_{yr.1}$; see Table S2. Patients 1-2 are ART-naive 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 ~ 0.1 and insignificant with $\text{p} - \text{values} > 0.01$. 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.

Macro- and micro-evolutionary selection fluctuations shape the structure of BCR lineages

The frequency-dependent behaviour of the selection likelihood ratio in Fig. 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 Methods and SI for theoretical expectation in this regime. As shown in Fig. S6, the data significantly deviates from the expected behaviour 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 Fig. 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, e.g., due to spontaneous environmental noise such as access to T-cell help or signalling molecules. We describe a theoretical model for this process in the Methods and show that the flattening of the likelihood ratio, up to intermediate frequencies, can be explained by an effective selection strength $\overline{s}$ subject to rapid *micro-evolutionary* fluctuations with an amplitude v; see Methods, SI and the fitted model to the likelihood ratios in Fig. S6.

bioRxiv preprint doi: https://doi.org/10.1101/271130; this version posted June 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 (SI). 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 above, rapid micro-evolutionary fitness fluctuations due to the environmental noise can slow down the rise of beneficial mutations in a lineage. An allele which is on 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 micro-evolutionary fluctuations nor drift can explain the mutation turnovers observed in (Figs. 3, S6).

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 *macro-evolutionary* selection fluctuations that occur at rates lower than the life-time 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 $s_0 \to s_0 - s_1$. The persistence of such fitness shifts over macro-evolutionary time scales can lead to successive turnover of new beneficial mutations in a lineage, consistent with the observation in Fig. S6. The comparison between the interference likelihood ratios $h(x)$ predicted by different evolutionary scenarios strongly indicates the prevalence of clonal interference and macro-evolutionary selection fluctuations in shaping the structure of BCR lineages for the V-gene classes in Fig. S6.

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, Table 1). In particular, using the interference likelihood ratio $h(x)$, we infer a significant macro-evolutionary turnover in the preference for the selected alleles (see Methods, SI). Fig. S9 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 in CDRs (Methods and Fig. S10).

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 to the ART-naive patients with a stable chronic infection (Table 1, 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 Fig. 3, Table 1, SI. 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, Fig. S7 and Table S2). 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.

bioRxiv preprint doi: https://doi.org/10.1101/271130; this version posted June 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

	HIV infected untreated				HIV infected interrupted treatment				Healthy productive		Healthy unproductive	
	$\overline{\alpha}_{\text{benef.}}$	$\overline{\alpha}_{\text{del.}}$	$\overline{\kappa}_{\text{benef.}}$	$\overline{\kappa}_{\text{del.}}$	$\overline{\alpha}_{\text{benef.}}$	$\overline{\alpha}_{\text{del.}}$	$\overline{\kappa}_{\text{benef.}}$	$\overline{\kappa}_{\text{del.}}$	$\overline{\alpha}_{\text{benef.}}$	$\overline{\alpha}_{\text{del.}}$	$\overline{\alpha}_{\text{benef.}}$	$\overline{\alpha}_{\text{del.}}$
CDR3	0.12 ±0.01	0.14 ±0.01	0.11 ±0.01	0.18 ±0.01	0.20 ±0.01	0.08 ±0.01	0.17 ±0.01	0.12 ±0.01	0.31 ±0.01	0.03 ±0.004	0.09 ±0.01	0.21 ±0.01
CDR1/2	0.24 ±0.01	0.07 ±0.01	0.22 ±0.02	0.11 ±0.01	0.27 ±0.01	0.07 ±0.01	0.23 ±0.01	0.10 ±0.01	NA	NA	NA	NA
FWR	0.08 ±0.01	0.13 ±0.01	0.09 ±0.01	0.17 ±0.01	0.05 ±0.01	0.18 ±0.01	0.08 ±0.01	0.17 ±0.01	0.07 ±0.01	0.22 ±0.01	0.11 ±0.01	0.24 ±0.01

Table 1: Beneficial and deleterious mutations in BCRs. The average fraction of beneficial $\overline{\alpha}_{\text{benef.}}$ and deleterious $\overline{\alpha}_{\text{del.}}$ 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 $\overline{\kappa}_{\text{benef.}}$ and deleterious $\overline{\kappa}_{\text{del.}}$ 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 standard error of the mean. The corresponding distributions are presented in Fig. 3 and Fig. S8; see Methods, SI and Fig S5, Fig. 4 and Fig S7 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.

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 result in evolution with clonal interference between clones in the repertoire (Figs. 3, 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 coevolution can be tested in future experiments that trace the intra-patient 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 idiotypical 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

bioRxiv preprint doi: https://doi.org/10.1101/271130; this version posted June 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

approach to characterize *macro-evolutionary* shifts of immune repertoires (Figs. 3, 4 and Fig S9) 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 [Strelkowa 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+ T-cell levels stay at its normal range (500 -1500 per ml) to activate a large population of B-cells, as it is the case in this study (Table S2), 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 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 [Neher et al., 2014, Łuksza and Lässig, 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; see e.g. ref. [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 towards a desired BnAb within the repertoire.

Methods

B-cell repertoire data, annotation and genealogies. We analyze B-cell repertoire data from 6 HIV patients from ref. [Hoehn et al., 2015] with raw sequence reads accessible from the European Nucleotide Archive under study accession numbers, ERP009671 and ERP000572. The data covers ~ 2.5 years of study with 6-8 sampled time points per patient; see Table S1 for details. The B-cell repertoire sequences consist of 150bp non-overlapping 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 3 individuals published in ref. [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 the SI and Fig. S1.

bioRxiv preprint doi: https://doi.org/10.1101/271130; this version posted June 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Selection and interference likelihood ratio. Hypermutations during affinity maturation create new clades within a lineage. The frequency x of these clades change 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 non-synonymous $G(x)$ an 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 non-synonymous $H(x)$ 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 $G(x_i|x_i) = 1$. Here, $s(t) = \overline{s} + \eta(t)$ is a fluctuating selection coefficient with average $\overline{s}$ and uncorrelated Gaussian fluctuations $\eta(t)$, with amplitude v, which we interpret as *micro-evolutionary* fluctuations, as opposed to *macro-evolutionary* (i.e., long term) correlations in environmental fluctuations. The solution to the stationary state follows [Takahata et al., 1975],

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

with $a_\pm = \frac{1}{2}[1 \pm \sqrt{1 + 4/v}]$ and $\lambda(\overline{s}, v) = \overline{s}/(v\sqrt{1 + 4/v})$. Expected selection likelihood ratio with micro-evolutionary fluctuations $g(x; \overline{s}, v) = G(x|\overline{x_i}; \overline{s}, v) \Big/ G(x|\overline{x_i}; 0, v)$ fits the data up to intermediate frequencies $x < 0.5$ (Fig. S6), indicating the prevalence of long-term fluctuations in the system. We assume a simple scenario with macro-evolutionary shift in selection preference, where the competing allele can become more beneficial over time, resulting in interference likelihood,

$$H(x|\overline{x_i}; \overline{s_0}, \overline{s_1}, v) = G(x|\overline{x}_i; \overline{s_0}, v)\, G(1|1 - x; \overline{s_1}, v)$$

where $s_1 \neq -s_0$. The expected likelihood ratio $h(x; \overline{s_0}, \overline{s_1}, v) = H(x|\overline{x_i}; \overline{s_0}, \overline{s_1}, v) \Big/ G(x|\overline{x_i}; 0, v)$ fits the data (Fig. S6). Fig. S9 shows the prevalence of such macro-evolutionary fluctuations throughout the repertoire.

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. AN acknowledges the funding from the Lewis-Sigler Institute for Integrative Genomics at Princeton University, where

bioRxiv preprint doi: https://doi.org/10.1101/271130; this version posted June 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

this work was initiated. We acknowledge the support by Max Planck Society (AN), SFB1310 (AN and AMW), National Institute of Health grant T32AI055400 (JO), the Janssen Research & Development LLC (MŁ), ERCCoG grant no. 724208 (TM and AMW). This work was performed in part at the Aspen Center for Theoretical Physics, which is supported by the National Science Foundation grant PHY-1066293.

Competing interests

The authors declare no competing interests.

References

C. Berek and C. Milstein. Mutation drift and repertoire shift in the maturation of the immune response. *Immunol. Rev.*, 96:23–41, Apr. 1987.

D. A. Bolotin, S. Poslavsky, I. Mitrophanov, M. Shugay, I. Z. Mamedov, E. V. Putintseva, and D. M. Chudakov. MiXCR: software for comprehensive adaptive immunity profiling. *Nat. Methods*, 12(5):380–381, May 2015.

D. A. Bolotin, S. Poslavsky, A. N. Davydov, F. E. Frenkel, L. Fanchi, O. I. Zolotareva, S. Hemmers, E. V. Putintseva, A. S. Obraztsova, M. Shugay, R. I. Ataullakhanov, A. Y. Rudensky, T. N. Schumacher, and D. M. Chudakov. Antigen receptor repertoire profiling from RNA-seq data. *Nat. Biotechnol.*, 35(10):908–911, Oct. 2017.

F. M. Burnet. A modification of Jerne's theory of antibody production using the concept of clonal selection. *CA Cancer J Clin.*, 26(2):119–121, Mar. 1976.

C. D. Campbell and E. E. Eichler. Properties and rates of germline mutations in humans. *Trends Genet*, 29(10): 575–584, Oct. 2013.

M. Caskey, F. Klein, and M. C. Nussenzweig. Broadly neutralizing antibodies for HIV-1 prevention or immunotherapy. *N. Engl. J. Med.*, 375(21):2019–2021, 2016.

M. M. Desai and D. S. Fisher. Beneficial mutation–selection balance and the effect of linkage on positive selection. *Genetics*, 17(5):385–394, 2007.

W. S. DeWitt, P. Lindau, T. M. Snyder, A. M. Sherwood, M. Vignali, C. S. Carlson, P. D. Greenberg, N. Duerkopp, R. O. Emerson, and H. S. Robins. A Public Database of Memory and Naive B-Cell Receptor Sequences. *PLoS ONE*, 11(8):e0160853, 2016.

Y. Elhanati, Z. Sethna, Q. Marcou, C. G. Callan, T. Mora, and A. M. Walczak. Inferring processes underlying B-cell repertoire diversity. *Phil. Trans. R. Soc. B*, 370(1676):20140243, Sept. 2015.

K. B. Hoehn, A. Gall, R. Bashford-Rogers, S. J. Fidler, S. Kaye, J. N. Weber, M. O. McClure, SPARTAC Trial Investigators, P. Kellam, and O. G. Pybus. Dynamics of immunoglobulin sequence diversity in HIV-1 infected individuals. *Phil. Trans. R. Soc. B*, 370(1676), Sept. 2015.

F. Horns, C. Vollmers, C. L. Dekker, and S. R. Quake. Signatures of selection in the human antibody repertoire: selective sweeps, competing subclones, and neutral drift. *Proc Natl Acad Sci USA*, 116(4):1261–1266, 2019.

J. D. Jensen, Y. Kim, V. B. DuMont, C. F. Aquadro, and C. D. Bustamante. Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics*, 170:1401–1410, 2005.

M. Kimura. Diffusion models in population genetics. *J. Appl. Probab.*, 1(2):177–232, 1964.

bioRxiv preprint doi: https://doi.org/10.1101/271130; this version posted June 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

J. Kingman. On the genealogy of large populations. *Journal of Applied Probability*, 19:27–43, 1982.

U. Laserson, F. Vigneault, D. Gadala-Maria, G. Yaari, M. Uduman, J. A. Vander Heiden, W. Kelton, S. Taek Jung, Y. Liu, J. Laserson, R. Chari, J.-H. Lee, I. Bachelet, B. Hickey, E. Lieberman-Aiden, B. Hanczaruk, B. B. Simen, M. Egholm, D. Koller, G. Georgiou, S. H. Kleinstein, and G. M. Church. High-resolution antibody dynamics of vaccine-induced immune responses. *Proc. Natl. Acad. Sci. U.S.A.*, 111(13):4928–4933, Apr. 2014.

M. Lässig, V. Mustonen, and A. M. Walczak. Predicting evolution. *Nat Ecol Evol*, 1(3):77, Feb. 2017.

H.-X. Liao, R. Lynch, T. Zhou, F. Gao, S. M. Alam, S. D. Boyd, A. Z. Fire, K. M. Roskin, C. A. Schramm, Z. Zhang, J. Zhu, L. Shapiro, NISC Comparative Sequencing Program, J. C. Mullikin, S. Gnanakaran, P. Hraber, K. Wiehe, G. Kelsoe, G. Yang, S.-M. Xia, D. C. Montefiori, R. Parks, K. E. Lloyd, R. M. Scearce, K. A. Soderberg, M. Cohen, G. Kamanga, M. K. Louder, L. M. Tran, Y. Chen, F. Cai, S. Chen, S. Moquin, X. Du, M. G. Joyce, S. Srivatsan, B. Zhang, A. Zheng, G. M. Shaw, B. H. Hahn, T. B. Kepler, B. T. M. Korber, P. D. Kwong, J. R. Mascola, and B. F. Haynes. Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature*, 496(7446):469–476, Apr. 2013.

M. Łuksza and M. Lässig. A predictive fitness model for influenza. *Nature*, 507(7490):57–61, Mar. 2014.

M. Łuksza, N. Riaz, V. Makarov, V. P. Balachandran, M. D. Hellmann, A. Solovyov, N. A. Rizvi, T. Merghoub, A. J. Levine, T. A. Chan, J. D. Wolchok, and B. D. Greenbaum. A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. *Nature*, 551(7681):517–520, Nov. 2017.

C. O. McCoy, T. Bedford, V. N. Minin, P. Bradley, H. Robins, and F. A. Matsen. Quantifying evolutionary constraints on B-cell affinity maturation. *Phil. Trans. R. Soc. B*, 370(1676):20140244, Sept. 2015.

J. H. McDonald and M. Kreitman. Adaptive protein evolution at the Adh locus in Drosophila. *Nature*, 351(6328): 652–654, 1991.

A. J. McMichael, P. Borrow, G. D. Tomaras, N. Goonetilleke, and B. F. Haynes. The immune response during acute HIV-1 infection: clues for vaccine development. *Nature Rev. Immunol.*, 10(1):11–23, Jan. 2010.

P. L. Moore, N. Ranchobe, B. E. Lambson, E. S. Gray, E. Cave, M.-R. Abrahams, G. Bandawe, K. Mlisana, S. S. Abdool Karim, C. Williamson, L. Morris, CAPRISA 002 Study, and NIAID Center for HIV/AIDS Vaccine Immunology (CHAVI). Limited neutralizing antibody specificities drive neutralization escape in early HIV-1 subtype C infection. *PLoS Pathog.*, 5(9):e1000598, Sept. 2009.

V. Mustonen and M. Lässig. From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation. *Trends Genet*, 25(3):111–119, 2009.

R. A. Neher and O. Hallatschek. Genealogies of rapidly adapting populations. *Proc. Natl. Acad. Sci. U.S.A.*, 110 (2):437–442, Jan. 2013.

R. A. Neher, C. A. Russell, and B. I. Shraiman. Predicting evolution from the shape of genealogical trees. *eLife*, 3: e03568, 2014.

A. Nourmohammad, J. Otwinowski, and J. B. Plotkin. Host-pathogen coevolution and the emergence of broadly neutralizing antibodies in chronic infections. *PLoS Genet*, 12(7):e1006171, July 2016.

A. Pandit and R. J. De Boer. Reliable reconstruction of HIV-1 whole genome haplotypes reveals clonal interference and genetic hitchhiking among immune escape variants. *Retrovirology*, 11(1):56, July 2014.

M. N. Price, P. S. Dehal, and A. P. Arkin. FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS ONE*, 5(3):e9490, Mar. 2010.

bioRxiv preprint doi: https://doi.org/10.1101/271130; this version posted June 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

D. K. Ralph and F. A. Matsen. Consistency of VDJ rearrangement and substitution parameters enables accurate B-cell receptor sequence annotation. *PLoS Comput. Biol.*, 12(1):e1004409, Jan. 2016.

D. D. Richman, T. Wrin, S. J. Little, and C. J. Petropoulos. Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc. Natl. Acad. Sci. U.S.A.*, 100(7):4144–4149, Apr. 2003.

H. E. Roberts, J. Hurst, N. Robinson, H. Brown, P. Flanagan, L. Vass, S. Fidler, J. Weber, A. Babiker, R. E. Phillips, A. R. McLean, J. Frater, and S. t. investigators. Structured observations reveal slow HIV-1 CTL escape. *PLoS Genet*, 11:e1004914, 2015.

S. Schiffels, G. J. Szöllősi, V. Mustonen, and M. Lässig. Emergent Neutrality in Adaptive Asexual Evolution. *Genetics*, 189(4):1361–1375, Dec. 2011.

N. G. C. Smith and A. Eyre-Walker. Adaptive protein evolution in Drosophila. *Nature*, 415(6875):1022–1024, Feb. 2002.

SPARTAC Trial Investigators, S. Fidler, K. Porter, F. Ewings, J. Frater, G. Ramjee, D. Cooper, H. Rees, M. Fisher, M. Schechter, P. Kaleebu, G. Tambussi, S. Kinloch, J. M. Miro, A. Kelleher, M. McClure, S. Kaye, M. Gabriel, R. Phillips, J. Weber, and A. Babiker. Short-course antiretroviral therapy in primary HIV infection. *N. Engl. J. Med.*, 368(3):207–217, Jan. 2013.

A. Stamatakis. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9):1312–1313, May 2014.

N. Strelkowa and M. Lässig. Clonal interference in the evolution of influenza. *Genetics*, 192(2):671–682, Oct. 2012.

N. Takahata, K. Ishii, and H. Matsuda. Effect of temporal fluctuation of selection coefficient on gene frequency in a population. *Proc. Natl. Acad. Sci. U.S.A.*, 72(11):4541–4545, Nov. 1975.

M. Uduman, M. J. Shlomchik, F. Vigneault, G. M. Church, and S. H. Kleinstein. Integrating B cell lineage information into statistical tests for detecting selection in Ig sequences. *J. Immunol.*, 192(3):867–874, Feb. 2014.

G. D. Victora and M. C. Nussenzweig. Germinal centers. *Annu. Rev. Immunol.*, 30:429–457, 2012.

M. C. Vieira, D. Zinder, and S. Cobey. Selection and neutral mutations drive pervasive mutability losses in long-lived B cell lineages. *Mol. Biol. Evol.*, 35:1135–1146, 2018.

C. Vollmers, R. V. Sit, J. A. Weinstein, C. L. Dekker, and S. R. Quake. Genetic measurement of memory B-cell recall using antibody repertoire sequencing. *Proc. Natl. Acad. Sci. U.S.A.*, 110(33):13463–13468, 2013.

J. A. Weinstein, N. Jiang, R. A. White, D. S. Fisher, and S. R. Quake. High-throughput sequencing of the zebrafish antibody repertoire. *Science*, 324(5928):807–810, May 2009.

G. Yaari, J. A. Vander Heiden, M. Uduman, D. Gadala-Maria, N. Gupta, J. N. H. Stern, K. C. O'Connor, D. A. Hafler, U. Laserson, F. Vigneault, and S. H. Kleinstein. Models of somatic hypermutation targeting and substitution based on synonymous mutations from high-throughput immunoglobulin sequencing data. *Front Immunol*, 4:358, 2013.

F. Zanini, J. Brodin, L. Thebo, C. Lanz, G. Bratt, J. Albert, and R. A. Neher. Population genomics of intrapatient HIV-1 evolution. *eLife*, 4:e11282, Dec. 2015.