



Constructing lower-bounds for CTL escape rates in early SIV infection



Sivan Leviyang

Georgetown University, Department of Mathematics and Statistics, United States

HIGHLIGHTS

- In HIV and SIV infections, the viral population repeatedly escapes from CTL response.
- We develop inference methods to estimate the rate of the first CTL escape.
- Using frequency data, we construct estimators which serve as lower bounds for the escape rate.
- The early part of the first CTL escape proceeds quicker than later parts.
- The rate of the first CTL escape differs across different infected compartments.

ARTICLE INFO

Article history:

Received 30 July 2013

Received in revised form

15 January 2014

Accepted 17 February 2014

Available online 3 March 2014

Keywords:

CTL escape

Escape rate

SIV

HIV

ABSTRACT

Intrahost human and simian immunodeficiency virus (HIV and SIV) evolution is marked by repeated viral escape from cytotoxic T-lymphocyte (CTL) response. Typically, the first such CTL escape starts around the time of peak viral load and completes within one or two weeks. Many authors have developed methods to quantify CTL escape rates, but existing methods depend on sampling at two or more timepoints. Since many datasets capture the dynamics of the first CTL escape at a single timepoint, we develop inference methods applicable to single timepoint datasets. To account for model uncertainty, we construct estimators which serve as lower bounds for the escape rate. These lower-bound estimators allow for statistically meaningful comparison of escape rates across different times and different compartments. We apply our methods to two SIV datasets, showing that escape rates are relatively high during the initial days of the first CTL escape and drop to lower levels as the escape proceeds.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

During HIV and SIV infections, the viral population repeatedly escapes from selective pressure exerted by cytotoxic T-lymphocytes (CTLs), a type of immune system cell. Each CTL targets a specific peptide, referred to as an epitope, associated with a locus on the viral genome. Mutation at the locus may change the epitope, making it partially or completely unrecognisable by existing CTLs. Viruses possessing such mutations are at a selective advantage, leading to a selective sweep referred to as a CTL escape. See [Goulder and Watkins \(2004\)](#) for a review of CTL escape in both HIV and SIV infections.

In this work, we consider the first CTL escape to occur during an infection. In SIV and HIV infections, CTL response initiates roughly at 14 and 21 days after infection, respectively, just prior to peak viral load ([Borrow et al., 1994](#); [Cohen et al., 2011](#); [Goulder and Watkins, 2004](#); [McMichael et al., 2010](#)). In the week or two following the initiation of CTL response, CTL escape often occurs at a single targeted epitope ([Boutwell et al., 2010](#); [McMichael et al., 2010](#); [Goonetilleke et al., 2009](#); [Henn et al., 2012](#); [Allen et al., 2000](#)). T-cell

tetramer studies suggest that this escape is driven by an especially focused CTL response in comparison to subsequent responses and escapes ([Turnbull et al., 2009](#); [Yasutomi et al., 1994](#); [Veazey et al., 2003](#)).

Many authors have attempted to quantify the strength of CTL response by measuring the rate at which CTL escape occurs. A commonly used method (e.g. [Goonetilleke et al., 2009](#); [Love et al., 2008](#); [Loh et al., 2008](#); [Asquith and McLean, 2007](#); [Ganusov et al., 2011](#)), introduced in [Fernandez et al. \(2005\)](#), [Asquith et al. \(2006\)](#), fits escape mutation frequencies at two timepoints to a differential equation model. The model fit is determined by a parameter, known as the escape rate, which is used to quantify the strength of CTL response at a given epitope. Since this approach requires frequency data at two timepoints, we call it the two-point method.

Using the two-point method to analyze the first CTL escape is difficult because rarely do both sampled timepoints capture the escape. For example, the first two timepoints available in HIV studies of acute infection are typically in the range of days 30 and 50, e.g. [Fisher et al. \(2010\)](#) and [Goonetilleke et al. \(2009\)](#). Using the two-point method on such data estimates escape rates between

days 30 and 50, while CTL response is likely strongest prior to day 30. The situation is different for SIV studies. Since the time of infection can be controlled, sampling timepoints can be chosen that straddle day 14, the approximate time of CTL response; for example sampling can occur at days 7 and 21. But usually the CTL escape has not started at day 7, so the two-point method must be applied using data collected at day 21 and a later timepoint, leading to the same difficulties seen in HIV datasets.

Other authors (e.g. Mandl et al., 2007; Althaus and de Boer, 2008; Petravic et al., 2008; Monteiro et al., 2000) have developed methods based on the standard model of viral dynamics (Perelson, 2002; Nowak and May, 2000). These methods depend on models with many parameters, in contrast the two-point method depends only on the escape rate and the mutation frequencies at the two timepoints. Further, fitting the standard model and its variants requires multiple timepoints, so the time period to which such escape rate estimates apply is often unclear. Recently, haplotype data has been used to estimate escape rates, but this method is more applicable to later timepoints in infection, when the viral population possesses significant genetic diversity (Messer and Neher, 2012).

The rate of CTL escape can be defined in different ways. For example, some authors measure the timespan from initiation of CTL response to the time when mutant frequencies reach a prescribed level (Liu et al., 2006; Palmer et al., 2013). In the two-point method, using the underlying model, the escape rate is the difference between the average CTL kill rate and the fitness cost of mutation (Asquith et al., 2006; Fernandez et al., 2005). We take this as our definition of the escape rate.

In this work, we develop inference methods for estimating the rate of the first CTL escape using frequency data from a single timepoint. We apply these methods to SIV datasets, a setting in which inference is slightly easier because infection time is typically known, but our methods extend to HIV escape as well. For SIV infection, we have in mind frequency data collected somewhere between days 14 and 28, times that capture the first CTL escape when the mutation frequency is substantial, but before escape at other epitopes has developed. Single timepoint methods have been used to infer early growth rates for cytomegalovirus, a situation in which immune response and viral mutation are less important (Cromer et al., 2013).

The price we pay for using a single timepoint is the need for an underlying model describing viral dynamics and evolution in the early stages of infection, prior to peak viral load. Specifying such a model is difficult because the dynamics of early SIV and HIV infections are poorly understood. We solve this difficulty by introducing a model that allows for a range of assumptions on early viral dynamics depending on the parameters chosen. Then, in order to cope with the resulting large parameter space, we derive estimators which serve as lower bounds for the escape rate over a large range of possible models.

We apply our methods to the two SIV datasets presented in Bimber et al. (2009) and Vanderford et al. (2011). By combining lower-bound estimators and the two-point method, we are able to compare escape rates between early and late time periods during the first CTL escape, as well as across different compartments. Our results also clarify the role of different modeling assumptions on escape rate inference.

2. Results

2.1. Model

Our model distinguishes between two types of infected cells: wild type and mutant. Wild types contain the epitope at which the

Table 1
Model parameters.

Parameter	Description	Units
t_A	CTL response time	day
t_F	Sampling time	day
μ	Epitope mutation rate	$\frac{\text{mutant infected cell}}{\text{day} \cdot \text{wild type infected cell}}$
$r(t)$	Wild type growth rate	day^{-1}
$k(t)$	CTL kill rate	day^{-1}
$X(t_A; s)$	Offspring distribution	infected cells
c	Fitness cost	day^{-1}

first CTL escape occurs; mutants contain a nucleotide mutation at that epitope. $w(t)$ and $m(t)$ represent the number of wild type and mutant type, respectively, at t days post infection. $f(t)$ represents the mutant frequency at t , i.e. $f(t) = m(t)/(w(t) + m(t))$.

The model depends on the seven parameters listed in Table 1. To start, we present the model assuming no mutation-associated fitness costs; $c=0$. The parameter t_A specifies the time, in units of days post infection, when CTL response initiates. Specification of the model splits according to whether times are before or after t_A .

For times prior to t_A , wild type dynamics are specified through the parameter $r(t)$ and the equation

$$\frac{dw}{dt}(t) = r(t)w(t). \quad (1)$$

$r(t)$ is the wild type growth rate in units of day^{-1} . By choosing $r(t)$ appropriately, arbitrary profiles for $w(t)$ are possible, reflecting the flexibility of the model.

Given $w(t)$, the parameters μ and $X(t_A; s)$ specify the distribution of $m(t_A)$ through the equation

$$m(t_A) = \sum_{i=1}^N X(t_A; s_i). \quad (2)$$

The s_i for $i = 1, 2, \dots, N$ are the times prior to t_A at which a wild type mutated. The s_i are stochastic and are generated at a non-constant rate $\mu w(s)$. $X(t_A; s)$, termed the offspring distribution, gives the number of mutants at time t_A that descend from a wild type infected cell which mutated at time s . Importantly, $X(t_A; s)$ is a random variable. When no fitness costs exist, we assume

$$E[X(t_A; s)] = \exp \left[\int_s^{t_A} r(s') ds' \right], \quad (3)$$

so that mutants have the same average growth rate as wild types. In (3), $\exp[x]$ represents the exponential taken to the power of x , i.e. e^x , and $E[\cdot]$ denotes the expected value.

To explain $X(t_A; s)$ more concretely, we provide three examples:

$$X_1(t_A; s) = \exp \left[\int_s^{t_A} r(s') ds' \right], \quad (4)$$

$$X_2(t_A; s) = 2 * \text{Bernoulli}(.5) * \exp \left[\int_s^{t_A} r(s) ds \right], \quad (5)$$

$$X_3(t_A; s) = H / 2 * \exp \left[\int_s^{t_A} r(s) ds \right]. \quad (6)$$

Above, $\text{Bernoulli}(.5)$ is a Bernoulli random variable with success probability .5 and H is a continuous distribution on $[1, \infty)$ with density $2/y^3$, a heavy tailed distribution. All three $X_i(t_A; s)$ satisfy $E[X_i(t_A; s)] = \exp[\int_s^{t_A} r(s') ds']$, however, the variance increases from zero for $X_1(t_A; s)$ to infinity for $X_3(t_A; s)$. Little is known about the form of $X(t_A; s)$ in SIV and HIV infections, but experimental results suggesting that HIV has an effective population size much smaller than its census size could correspond to offspring distributions

such as $X_2(t_A; s)$ or $X_3(t_A; s)$ (Kouyos et al., 2006; Leigh-Brown, 1997).

For times after t_A , we switch to a deterministic model. In the datasets considered below, CTL response arises one or two days prior to or just at peak viral load. As discussed below, $f(t_A) \approx \mu t_A$, leading to $m(t)$ in the 100 s or greater. As several authors have noted, when mutant population size reaches such values, averaging effects reduce the impact of stochasticity and the dynamics become deterministic, see Rouzine and Coffin (2005), Kessinger et al. (2013), Leviyang (2013) and Desai and Fisher (2007). The stochastic model could be extended past t_A in cases for which $m(t_A)$ is modest, see Section 3.

After t_A , we model $w(t)$ and $m(t)$ through the equations

$$\begin{aligned} \frac{dw}{dt}(t) &= (r(t) - k(t))w(t), \\ \frac{dm}{dt}(t) &= r(t)m(t) + \mu w(t), \end{aligned} \quad (7)$$

where $k(t)$ is the CTL-mediated killing rate of wild type infected cells between times t_A and t_F in units of day⁻¹, see Ganusov et al. (2013) for a similar model. (7) can be recast in terms of the mutant frequency $f(t)$ to give

$$\frac{df}{dt}(t) = \mu(1 - f(t)) + (k(t) - \mu)f(t)(1 - f(t)). \quad (8)$$

To model a fitness cost c , we change (8) to

$$\frac{df}{dt}(t) = \mu(1 - f(t)) + (k(t) - c - \mu)f(t)(1 - f(t)), \quad (9)$$

and (3) to

$$E[X(t_A, s)] = \exp \left[\int_s^{t_A} (r(s) - c) ds \right]. \quad (10)$$

2.2. Inference methods

Let \hat{f}_{data} be the estimate of mutant frequency obtained by sampling viral sequences at time t_F . Using \hat{f}_{data} , our goal is to infer the escape rate $\bar{k} - c$, where \bar{k} is the average kill rate between t_A and t_F

$$\bar{k} = \frac{1}{t_F - t_A} \int_{t_A}^{t_F} k(s) ds,$$

and c is the mutation-associated fitness cost. To start, we assume no fitness costs, i.e. $c=0$, and present three estimators of \bar{k} , \bar{k}_D , \bar{k}_G and \bar{k}_R , referred to as the deterministic, general, and restricted estimator, respectively. \bar{k}_G and \bar{k}_R serve as lower bounds for \bar{k} . Towards the end of the section we consider fitness costs, showing that \bar{k}_G and \bar{k}_R are lower bounds for $\bar{k} - c$.

Regardless of the estimator, our approach involves the same steps. We assume that the parameters μ, t_A, t_F are known and based on these parameters, we select a value for $f(t_A)$, labeled \hat{f}_{silico} , and a family of profiles $k(t; \bar{k})$. For every possible value of \bar{k} , $k(t; \bar{k})$ is a specific CTL kill rate profile with average \bar{k} . Then, starting (8) at time t_A with $f(t_A) = \hat{f}_{\text{silico}}$, we fit \bar{k} by integrating (8) to time t_F and selecting the \bar{k} satisfying $f(t_F) = \hat{f}_{\text{data}}$. The distinction between the three estimators lies in the choice for \hat{f}_{silico} and the family $k(t; \bar{k})$.

To construct \bar{k}_D , we take a deterministic approach, using (8) to compute $f(t_A)$. Setting $f(0)=0$ and integrating (8) to t_A with $k(t)=0$ gives $f(t_A) = \mu t_A$ and so we set $\hat{f}_{\text{silico}} = \mu t_A$. Equivalently, $\hat{f}_{\text{silico}} = E[f(t_A)]$ since (8) gives the mean dynamics of $f(t)$, as can be seen by taking the expected value of (2). To build $k(t; \bar{k})$, we assume that CTL kill rates are constant once CTL response begins, making $k(t; \bar{k}) = \bar{k}$ for $t \in [t_A, t_F]$.

Given these choices for \hat{f}_{silico} and $k(t; \bar{k})$, \bar{k}_D satisfies the relation:

$$\hat{f}_{\text{data}} = \frac{1}{1 + \frac{\exp[-\bar{k}_D(t_F - t_A)]}{\mu t_A + \frac{\mu}{\bar{k}_D}(1 - \exp[-\bar{k}_D(t_F - t_A)])}}. \quad (11)$$

If we ignore mutations occurring after t_A , the two-point method with $f(t_A) = \mu t_A$ and $f(t_F) = \hat{f}_{\text{data}}$ can be applied, leading to the following approximation:

$$\bar{k}_D \approx \frac{1}{t_F - t_A} \log \left(\frac{\hat{f}_{\text{data}}}{(1 - \hat{f}_{\text{data}})\mu t_A} \right). \quad (12)$$

As we show below through numerical experiments, \bar{k}_D is a useful estimator. However, \bar{k}_D often overestimates \bar{k} and we would like estimators which serve as lower bounds for \bar{k} in order to compare escape rates across different times and compartments. To develop an estimator which is less than \bar{k} with confidence $1 - \sigma$, we choose \hat{f}_{silico} so that $f(t_A) < \hat{f}_{\text{silico}}$ with probability at least $1 - \sigma$ across a range of parametrizations. To choose $k(t; \bar{k})$, we select profiles maximizing the number of mutations occurring after t_A given fixed values of $f(t_A)$ and \bar{k} . Overestimating $f(t_A)$ through \hat{f}_{silico} and maximizing the number of mutations after t_A , leads to underestimates of \bar{k} because less CTL-mediated killing is necessary to achieve $f(t_F) = \hat{f}_{\text{data}}$. See Section 3 for more details. As a result, under a null model in which \hat{f}_{data} is generated according to the stochastic model with \bar{k} as the average for the $k(t)$ parameter chosen, estimators constructed in this manner will be less than \bar{k} with probability at least $1 - \sigma$.

To construct \bar{k}_G we set $\hat{f}_{\text{silico}} = \mu t_A / \sigma$. For the $k(t; \bar{k})$ profiles see Section 3, but roughly we choose profiles which delay most of the killing until time t_F . Intuitively, delaying killing allows more mutations to occur after time t_A , thereby raising $f(t_F)$. \bar{k}_G satisfies the relation

$$\hat{f}_{\text{data}} = \frac{1}{1 + \frac{\exp[-\bar{k}_G(t_F - t_A)]}{\frac{\mu t_A}{\sigma} + \mu(t_F - t_A)}}, \quad (13)$$

which can be solved to find

$$\bar{k}_G = \frac{1}{t_F - t_A} \log \left(\frac{\hat{f}_{\text{data}}}{\frac{\mu t_A}{\sigma} + \mu(t_F - t_A)} \right). \quad (14)$$

\bar{k}_G is often a poor lower bound; in many cases, it significantly underestimates \bar{k} . The poor behaviour derives from the large parameter space, namely all choices for $r(t), k(t), X(t_A; s)$. To produce a better lower-bound estimator, we consider smaller parameter spaces by requiring $r(t) \geq r_{\min}$ for $t < t_A$ and $k''(t) < 0$, where $r_{\min} = .8$. $X(t_A; s)$ is allowed to take any value. The restriction $r(t) \geq r_{\min}$ assumes a minimum expansion rate for the number of infected cells prior to CTL response time t_A . $k''(t) < 0$ assumes CTL kill rates rise quickest at the beginning of the CTL response. Other restrictions are possible, reflecting different biological assumptions. Assuming these restrictions, \bar{k}_R is constructed using $\hat{f}_{\text{silico}} = \mu t_A(1 + 2/(r_{\min} t_A \sigma))$ and $k(t; \bar{k}) = 2\bar{k}(t - t_A)/(t_F - t_A)$, a linearly increasing profile. \bar{k}_R satisfies the relation

$$\hat{f}_{\text{data}} = \frac{1}{1 + \frac{\exp[-\bar{k}_R(t_F - t_A)]}{\mu t_A \left(1 + \frac{2}{\sigma t_A r_{\min}} \right) + \mu \sqrt{\frac{\pi(t_F - t_A)}{4\bar{k}_R}}}}. \quad (15)$$

Ignoring mutations after t_A allows the two-point method to be applied giving the approximation,

$$\bar{k}_R \approx \frac{1}{t_F - t_A} \log \left(\frac{\hat{f}_{\text{data}}}{(1 - \hat{f}_{\text{data}})\mu t_A \left(1 + \frac{2}{\sigma t_A r_{\min}}\right)} \right). \quad (16)$$

Now we consider the presence of a fitness cost c . \bar{k}_G and \bar{k}_R are constructed assuming no fitness costs, but the two estimators are lower-bounds for the escape rate when fitness costs exist. Between t_A and t_F , introducing a fitness cost can be seen as a shift of $k(t)$ to $k(t) - c$. Correspondingly, \bar{k}_G and \bar{k}_R shift from lower bounds for \bar{k} to lower bounds for $\bar{k} - c$. \hat{f}_{silico} is still an upper bound for $f(t_A)$ because fitness costs reduce $f(t_A)$, meaning that the probability \hat{f}_{silico} is greater than $f(t_A)$ will increase. As a result, \bar{k}_G and \bar{k}_R are lower-bounds for the escape rate in the absence or presence of fitness costs.

In practice, μ and t_A are unknown, but are needed for all three estimation methods. Using the approximate formulas for the three estimators and letting $t_F - t_A = 7$, we calculate that mistaking μ by a factor of 10 shifts all estimators by roughly .3. The estimators are shifted down as μ is increased, so to maintain the lower-bounds μ should be overestimated. As discussed in Mandl et al. (2007), assuming an epitope composed of roughly 30 nucleotides, a mutation rate of 3×10^{-5} per base pairing (Mansky, 1996), and about 2/3 of mutations being non-synonymous, leads to an epitope mutation rate of $\mu = 6 \times 10^{-4}$ which is likely greater than the true rate.

If the parameter t_A is less than the true t_A value, then the true $k(t)$ will be zero for times greater than the parameter t_A but less than the true t_A . There is nothing in the model and estimators that prohibit this, except that the restriction $k''(t) > 0$ for \bar{k}_R will not hold. In contrast, using a t_A value greater than the true t_A value will bias the estimators up because more mutants will exist at t_A than predicted under the model. To preserve lower-bounds, tetramer data should be used to estimate t_A , with underestimation preferred to overestimation.

2.3. Numerical experiments

We conducted numerical experiments to assess the effect of different $r(t)$, $k(t)$ and $X(t_A; s)$ on the inference methods. We considered three different choices for each of $r(t)$, $X(t_A; s)$ and $k(t)$. We set $\mu = 10^{-4}$, $t_A = 14$, $t_F = 21$ and assumed no fitness costs. The three choices for $r(t)$; referred to as constant, logistic, and slow; have the log $w(t)$ profiles shown in Fig. 1. All three profiles satisfy $w(t_A) = 10^8$. For $X(t_A; s)$, the three choices are given by the $X_1(t_A; s)$, $X_2(t_A; s)$, $X_3(t_A; s)$ defined above and are correspondingly

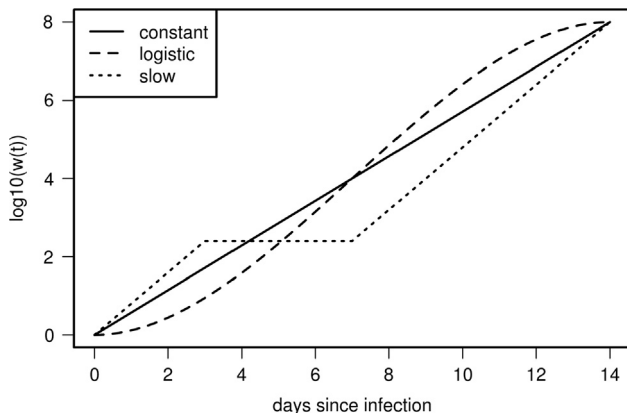


Fig. 1. $w(t)$ profiles for numerical experiments.

labeled no-variance, Bernoulli, and heavy-tail. To define $k(t)$ on the interval $[t_A, t_F]$, we chose $\bar{k} = .8$ and then considered a constant profile, $k(t) = \bar{k}$, a linear increasing profile, $k(t) = 2\bar{k}(t - t_A)/(t_F - t_A)$, and a linear decreasing profile, $k(t) = 2\bar{k}(t_F - t)/(t_F - t_A)$. All the $k(t)$ profiles have average kill rate $\bar{k} = .8$. The actual $r(t)$ profiles seen in HIV and SIV infections are unknown and current understanding of $k(t)$ profiles depends on tetramer and ELISPOT data which may not translate simply to kill rates (Turnbull et al., 2009). We chose our $r(t)$ and $k(t)$ profiles as special cases through which the effect of general profiles can be understood. For example, the slow $r(t)$ profile could be biologically explained as an initial focus of approximately 100 infected cells formed in the first two days of infection, followed by a waiting time until infection spreads to the lymph nodes and gut, but here we present the slow $r(t)$ as a simple profile through which to understand the role of early expansion rates in shaping the stochasticity of early escape.

Tables 2–4 show results for different combinations of $r(t)$, $X(t_A; s)$, $k(t)$. To produce the tables, we ran 1000 simulations of the stochastic model for each $r(t)$, $X(t_A; s)$, $k(t)$ combination. Each simulation returned a value of $f(t_F)$ (mutant frequency at sample time) which was used by the inference methods to estimate \bar{k} . Importantly, to implement the inference methods, we assumed that μ , t_A , t_F were known, but no further information other than $f(t_F)$ was used. Since we are interested in lower bounds, the table gives one-sided 95% CIs, i.e. the range of values seen over the 1000 simulations with the top 50 ignored. The tables use the exact formulas for \bar{k}_D , \bar{k}_G , and \bar{k}_R ; approximate formulas yielded similar patterns.

Table 2 examines the effect of the three $r(t)$ profiles on $f(t_F)$ and the \bar{k} estimators. $X(t_A; s)$ and $k(t)$ were fixed as noted. Across the three parametrizations, $f(t_F)$ has significant variance, with the endpoints of the 95% CIs varying by roughly 50% from their averaged value. Variance in $f(t_F)$ translates into error in the \bar{k} estimates. \bar{k}_D is slightly biased down from the true $\bar{k} = .8$ value, but also has a significant probability of overestimating \bar{k} . For example, with constant $r(t)$, the 95% CI reaches to .9. Overestimation is reduced under the logistic $r(t)$ and increases, to a right endpoint value of .97, under the slow $r(t)$. Intuitively, when the population is large, averaging reduces variance and when the population is small, the probability of mutations occurring is small, also leading to reduced variance. As a result, variance is influenced by the time the wild type population spends at levels of order $1/\mu$, a population size at which mutations are likely but not numerous. Under the logistic $r(t)$, this time period is short. On the other hand, we constructed the slow $r(t)$ to make this time period long. The constant $r(t)$ represents a middle ground. General profiles can be understood within this context. Notice that across all three $r(t)$ profiles, the restricted and general CIs are below \bar{k} .

Table 2
Simulation results: effect of $r(t)$.

$r(t)$	$X(t_A; s)$	$k(t)$	$f(t_F)$	\bar{k}_D	\bar{k}_R	\bar{k}_G
Constant	No-variance	Constant	[.14,.45]	[.67,.9]	[.46,.69]	[.25,.48]
Logistic	No-variance	Constant	[.15,.37]	[.68,.85]	[.47,.64]	[.26,.43]
Slow	No-variance	Constant	[.11,.57]	[.63,.97]	[.42,.76]	[.21,.55]

Table 3
Simulation results: effect of $X(t_A; s)$.

$r(t)$	$X(t_A; s)$	$k(t)$	$f(t_F)$	\bar{k}_D	\bar{k}_R	\bar{k}_G
Slow	No-variance	Constant	[.11,.57]	[.63,.97]	[.42,.76]	[.21,.55]
Slow	Bernoulli	Constant	[.10,.69]	[.61,.104]	[.40,.83]	[.19,.62]
Slow	Heavy	Constant	[.10,.55]	[.61,.96]	[.40,.75]	[.20,.54]

Table 3 examines the effect of $X(t_A; s)$. The Bernoulli $X(t_A; s)$ increases the variance of $f(t_F)$ leading to wider CIs across all three inference methods. Somewhat surprisingly, the heavy distribution leads to slightly less $f(t_F)$ variance. While the heavy distribution allows for samples resulting in extremely high values of $f(t_F)$, such samples have small probability and their occurrence falls outside the 95% CI. Notice that, under a Bernoulli $X(t_A; s)$ and slow $r(t)$, the restricted method's CI exceeds $\bar{k} = .8$, reflecting the erroneous assumption of $r(t) \geq r_{\min}$. In contrast, the general estimator's CI stays below $\bar{k} = .8$.

To understand the effect of $k(t)$ on $f(t_F)$ and the three estimators, consider \bar{k}_D as defined through (11); the right hand side of (11) is $f(t_F)$ under the assumptions $f(t_A) = \hat{f}_{\text{silico}} = \mu t_A$ and $k(t; \bar{k}) = \bar{k}_D$. The expressions μt_A and $\mu / \bar{k}_D (1 - \exp[-\bar{k}_D(t_F - t_A)])$ represent the contributions of mutations before and after t_A , respectively, to $f(t_F)$. When $\bar{k}_D t_A$ is large, the μt_A term is dominant and $f(t_F)$ is mostly influenced by mutations occurring prior to t_A . Intuitively, a large kill rate pushes the frequency of wild types down, reducing the number of mutations occurring after t_A , and a large t_A value increases the number of mutations arising prior to t_A . More generally, for arbitrary parametrizations of the stochastic model, as $\bar{k} t_A$ increases, the profile of $k(t)$ matters less to the value of $f(t_F)$. Consequently, numerical experiments with $\bar{k} = .8$, $t_A = 14$, $t_F = 21$ show minor dependence on the $k(t)$ profile (results not shown). In contrast, Table 4 shows simulation results under $\bar{k} = .2$, $t_A = 9$, $t_F = 35$. The increasing and decreasing $k(t)$ profiles produce more and less mutations after t_A , respectively, with the constant $k(t)$ profile occupying a middle ground. As the results show, producing more mutants after t_A shifts all the CIs to the right. In particular, under the increasing profile the \bar{k}_D CI is largely to the right of the true $\bar{k} = .2$ value. Intuitively, the \bar{k}_D estimator underestimates

the number of mutations after t_A , causing it to overestimate \bar{k} . Notice that \bar{k}_G and \bar{k}_R still serve as lower bounds.

2.4. Datasets

Most existing analyses of CTL escape infer a single rate for the entire escape. For the datasets below, we consider three times: t_A , t_F and t_S . t_A and t_F are as previously defined, the CTL response time and sampling time, but now we add a second sampling time t_S subsequent to t_F . To t_A and t_F , we apply the discussed methods, estimating the escape rate between t_A and t_F . To t_F and t_S , we apply the two-point method, estimating the escape rate between t_F and t_S . Using the lower-bound estimators \bar{k}_G and \bar{k}_R , we investigate whether escape rates from t_A to t_F are greater than escape rates from t_F to t_S . In the case of the Vanderford et al. data, we also compare escape rates across compartments.

Given frequency data, our estimators and the two-point method estimator provide point estimates for the escape rate. When sampling variance associated with the data is included, the point estimates generalize to CIs in a standard way. Conservatively, we assume that the frequency data is accurate to roughly 10%, corresponding to 100 viral sequences at each timepoint. The CIs shown below are constructed accordingly. In some of the following figures, sampling variance causes the two-point method CIs to include negative escape rates. As an example of how this might occur, suppose that the mutation frequencies provided by the data at times t_F and t_S are .4 and .45, respectively. Sampling variance allows the true mutation frequencies from which the data was sampled to be, say, .43 and .42, reflecting a drop in mutation frequency and a negative escape rate.

2.4.1. Bimber et al. dataset

The data from Bimber et al. involves four Mauritian cynomolgus macaques (MCMs) and four Rhesus macaques (RMs). (The full dataset included eight RMs, we considered the four unvaccinated RMs.) We refer the reader to the article and references therein for full details (Bimber et al., 2009). Briefly, all animals were intrarectally infected with SIVmac239. The first CTL escape in the MCMs was at the epitope NEF-RM9, while CTL escape first occurred at TAT-SL8 in the RMs. Pyrosequencing of the epitopes was performed at various timepoints. At day 14 after infection, for both

Table 4
Simulation results: effect of $k(t)$.

$r(t)$	$X(t_A; s)$	$k(t)$	$f(t_F)$	\bar{k}_D	\bar{k}_R	\bar{k}_G
Constant	Constant	Constant	[.09,.29]	[.16,.22]	[.10,.16]	[.06,.12]
Constant	Constant	Decreasing	[.05,.27]	[.14,.22]	[.08,.15]	[.04,.11]
Constant	Constant	Increasing	[.16,.34]	[.19,.23]	[.13,.17]	[.08,.12]

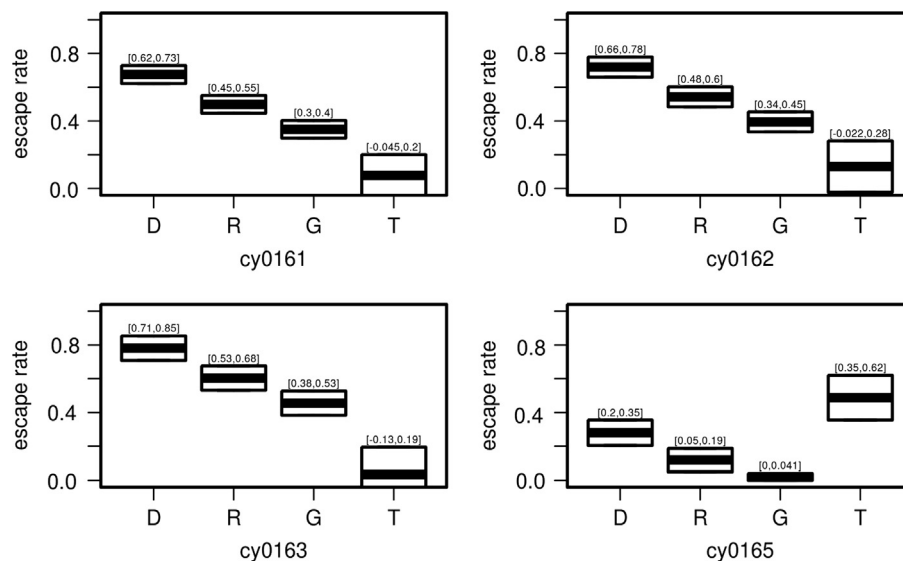


Fig. 2. Escape rates CIs for MCMs in Bimber et al. Each subfigure represents a single animal. Within each subfigure, the ticks labeled D, R and G give escape rate CIs for days 12–21 using, respectively, the deterministic, restricted, and general methods. The right tick gives the escape rate CI for days 21–28 as computed by the two-point method. All CIs are at 95% significance and include sampling variance.

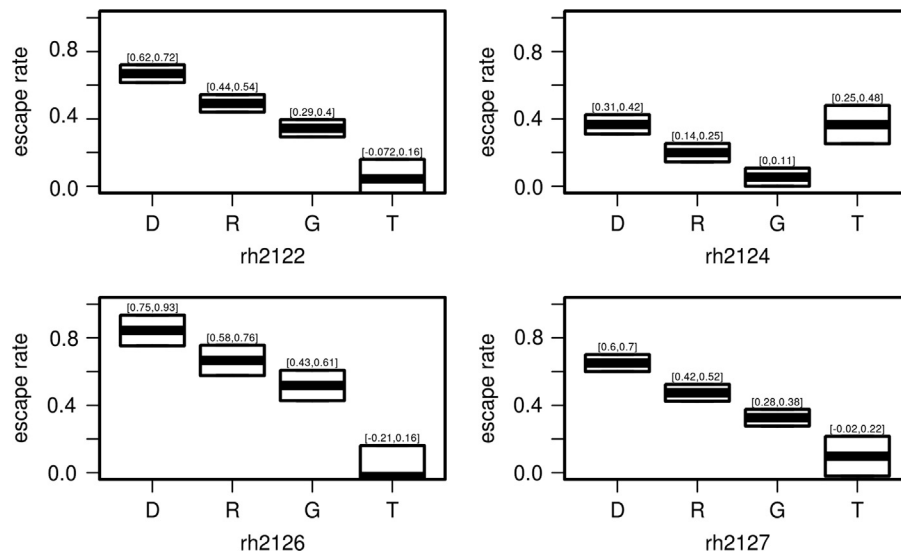


Fig. 3. Escape rates CIs for RMs in Bimber et al. See Fig. 2 for details.

MCMs and RMs, the sampled sequences were roughly homogeneous; by days 21 and 28, MCMs and RMs had a significant frequency of escape mutants at NEF-RM9 and TAT-SL8, respectively. Tetramer data showed no CTL response at day 10, but a strong CTL response by day 14. We set $t_A=12$ with the aim of underestimating t_A but still accounting for no response at day 10. The results reported below are essentially unchanged given $t_A=10$ or $t_A=11$. We set $t_F=21$ and $t_S=28$ since these were the first sampled timepoints after day 14. Finally, we set $\mu = 6 \times 10^{-4}$ with the aim of overestimating μ .

Fig. 2 shows results for the four MCMs. In each subfigure, the three tics to the left give CIs for the deterministic, restricted, and general estimate of the escape rate during days 12–21. The right-most tic gives the CI for the escape rate during days 21–28 according to the two-point method. As the figure shows, in three of the four animals, the escape rate estimates under $\bar{k}_D, \bar{k}_R, \bar{k}_G$ for days 12–21 are higher than the two-point method estimate for days 21–28. Interestingly, animal cy0165, which is the only animal with overlapping two-point and lower-bound CIs, had a weak CD8+ response from days 14 to 21 and a response that was increasing shortly after day 21 (see Fig. 3 in Bimber et al.).

Fig. 3 shows analogous results for the four RMs. The pattern is similar to the MCMs, but in this case no tetramer data was available.

2.4.2. Vanderford et al. dataset

The Vanderford et al. dataset includes 15 Rhesus macaques (RMs) infected intravenously with SIVmac239. As in Bimber et al., the RMs experienced initial escape at TAT-SL8 and escape dynamics were sampled using pyrosequencing at different timepoints in four compartments: viral RNA in the plasma (PL) and genomic DNA from peripheral blood mononuclear cells (PBMC), lymph node biopsies (LN), and rectal mucosa biopsies (RB). Using tetramer data, Vanderford et al. estimate the frequency of CD8+ T-lymphocytes specific for TAT-SL8 in the different compartments at days 7, 14 and 28.

Vanderford et al. show that lymph nodes and rectal mucosa are the primary source of TAT-SL8 escape mutants, with escape mutants often first arising in the lymph nodes. Given this result, we focused on comparing rates of escape between the LN and RB compartments. In order to consider escapes starting in lymph nodes, we restricted our attention to animals in which PL, PBMC, and RB epitope frequencies were above 90% at day 14 and for which LN escape data was available at days 14 and 28 (not all animals were sampled at all timepoints). Six RMs fulfilled these requirements.

For the LN, tetramer data showed a very weak CD8+ response at day 7 but a strong response by day 14. Given this data, we set $t_A=9$ as an underestimate of CTL response time, results were similar for $t_A=7$ and $t_A=8$. t_F and t_S were set at 14 and 28, respectively, reflecting the first sampled timepoints at which mutant frequency was significant in the LN. We set $\mu = 6 \times 10^{-4}$.

For RB we set $t_A=9$, reflecting tetramer data showing a response similar to the LN. We set $t_F=28$; this was the first timepoint at which mutant frequency was significant in the RB. Subsequent to day 28, frequency data was available at day 56, by which time the escape at TAT-SL8 had already ended. As a result, for RB we estimate the escape rate during the single time interval of days 9–28 and solely through our methods. We set $\mu = 6 \times 10^{-4}$.

Fig. 4 gives confidence intervals for the escape rates in the six RMs. Each subfigure corresponds to a single animal. The first tic from the left gives the escape rates in the LN during days 9–14 using the deterministic and general estimators. The second tic gives the escape rate in the LN during days 14–28, constructed using the two-point method. The third tic gives the escape rate in RB during days 9–28, constructed using the deterministic and general estimators. We have not included results for the restricted estimator to make the figure more readable; roughly, the restricted estimator CIs fall midway between the deterministic and general estimator CIs.

Across all animals the deterministic estimator CI predicts a significantly higher LN escape rate during days 9–14 than escape rates inferred for days 14–28 using the two-point method. This pattern is supported by the general estimator for the RMs represented in the top row, but not in the bottom row. However, if instead of assuming $\mu = 6 \times 10^{-4}$ we set $\mu = 10^{-4}$, all general estimator CIs are raised above the two-point method CIs.

Similarly, across all animals, the deterministic CIs predict higher RB escape rates during days 9–28 than those seen in the LN during days 14–28. In this case, the general CIs for the RB escape rate are either slightly above or below the LN escape rates in four of the six animals. If μ is raised to 10^{-4} , the deterministic and general CIs are shifted up by roughly .13, so all general CIs are above or even with the LN escape rates.

3. Methods

3.1. \bar{k}_G and \bar{k}_R are lower bounds for \bar{k}

Suppose that $f(t_F)$ is generated through a single simulation of our stochastic model with no fitness costs and using parameters

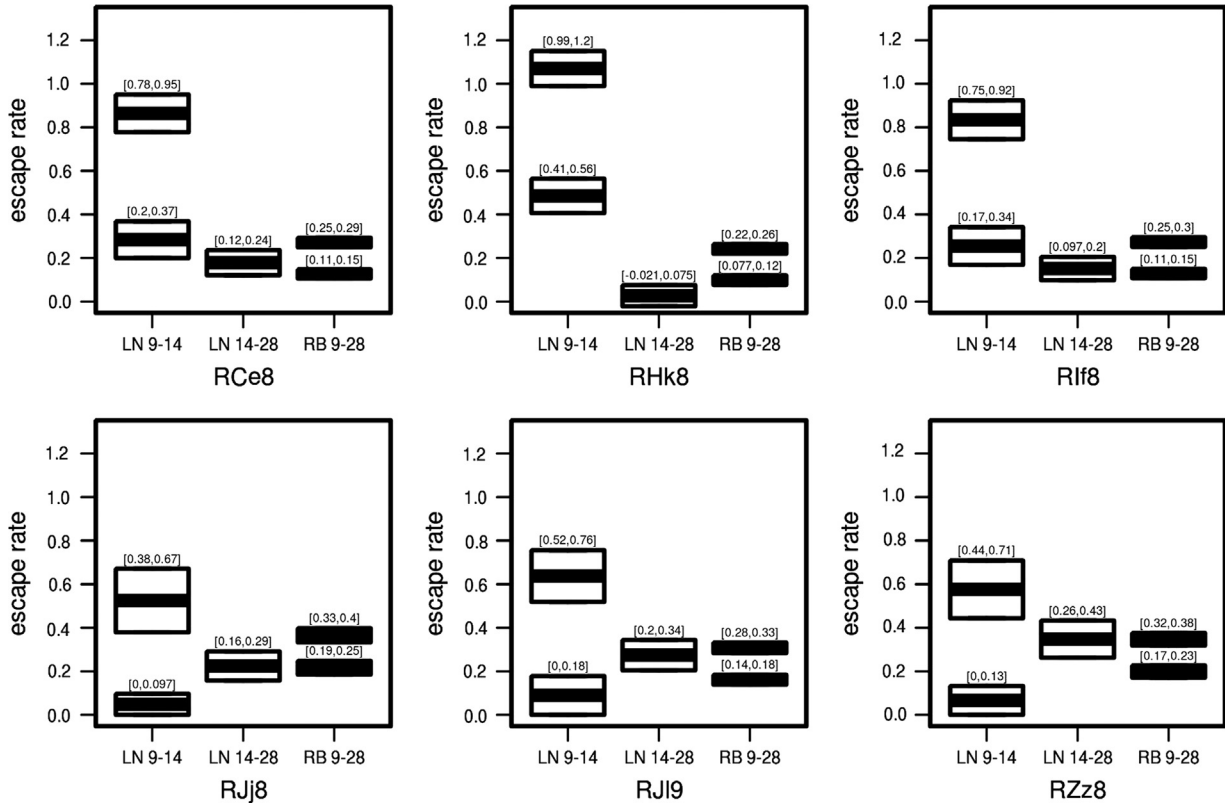


Fig. 4. Escape rates CIs for Vanderford et al. Each subfigure represents a single animal. Tics, from left to right, represent escape rates in the lymph node during days 9–14, lymph nodes during days 14–28, and rectal biopsies during days 9–28. Within each subfigure, the left-most and right-most tics show the deterministic CI (upper box) and the general lower-bound CI (lower box). The center tic shows the two-point method CI. All CIs are at 95% significance.

$r(t), k(t), X(t_A; s)$ and μ, t_A, t_F . As previously defined, let \bar{k} be the average of $k(t)$ from t_A to t_F and let $f(t_A)$ be the mutant frequency seen during the simulation at time t_A . Here, we would like to show that $\bar{k}_G < \bar{k}$ and $\bar{k}_R < \bar{k}$ with probability $1 - \sigma$ when \bar{k}_G and \bar{k}_R are constructed using $f(t_F)$ and knowledge of μ, t_A, t_F . $\bar{k}_R < \bar{k}$ further requires $r(t)$ and $k(t)$ to fulfill the requirements stated in the Results. We will show $\bar{k}_R < \bar{k}$, the arguments for \bar{k}_G are similar.

Between t_A and t_F , $f(t)$ satisfies (8). Integrating (8) relates the simulation values $f(t_A), f(t_F), k(t), \bar{k}$ as follows:

$$f(t_F) = \frac{1}{1 + \frac{\exp[-\bar{k}(t_F - t_A)]}{f(t_A) + \mu\Gamma}} \quad (17)$$

where

$$\Gamma = \int_{t_A}^{t_F} \exp\left[-\int_{t_A}^s ds' k(s')\right] ds \quad (18)$$

\bar{k}_R is chosen as the $k^\#$ value which satisfies the following expression:

$$f(t_F) = \frac{1}{1 + \frac{\exp[-k^\#(t_F - t_A)]}{\hat{f}_{\text{silico}} + \mu\Gamma_{\text{silico}}(k^\#)}} \quad (19)$$

where

$$\Gamma_{\text{silico}}(k^\#) = \int_{t_A}^{t_F} \exp\left[-\int_{t_A}^s ds' k(s'; k^\#)\right] ds \quad (20)$$

In (20), $k(s; k^\#) = 2k^\#(s - t_A)/(t_F - t_A)$, which is the family of kill rate profiles used to construct \bar{k}_R . $\Gamma_{\text{silico}}(k^\#)$ essentially reduces to $\sqrt{\pi(t_F - t_A)/4k^\#}$ seen in (15), but the integral form written above makes the connection to Γ clear.

If in (19) we choose $k^\# = \bar{k}$, then the right side of (19) will be less than the right side of (17) with probability $1 - \sigma$ because we construct \hat{f}_{silico} to be greater than $f(t_A)$ with probability $1 - \sigma$ and we construct $k(t; k^\#)$ so that $\Gamma_{\text{silico}}(\bar{k}) > \Gamma$. Therefore, to make the equality in (19) true, $k^\#$ needs to be chosen to make the right side greater. The derivative of the right side in $k^\#$ is always negative, so we must lower $k^\#$, meaning $\bar{k}_R < \bar{k}$.

3.2. Constructing \hat{f}_{silico}

Define f_{max} by $\hat{f}_{\text{silico}} = f_{\text{max}} \mu t_A$, working with f_{max} instead of \hat{f}_{silico} makes for cleaner formulas. For a confidence level σ , we need to construct f_{max} satisfying

$$P(f(t_A) > f_{\text{max}} \mu t_A) \leq \sigma. \quad (21)$$

To simplify the arguments below, let $Z(t_A; s)$ be a normalization of $X(t_A; s)$:

$$Z(t_A; s) = X(t_A; s) / E[X(t_A; s)] = X(t_A; s) / \left(\frac{w(t_A)}{w(s)} \right). \quad (22)$$

The rightmost equality follows from $E[X(t_A; s)] = w(t_A)/w(s)$. Intuitively, wild types at time s collectively produce $w(t_A)$ offspring at time t_A ; so on average, each wild type produces $w(t_A)/w(s)$ offspring. (A rigorous demonstration follows from rewriting $w(t_A)$ and $w(s)$ in terms of $r(t)$.) Assuming no mutation-associated fitness costs, on average a mutant should produce the same number of offspring prior to CTL response as a wild type.

Using $Z(t; t_0)$, $f(t_A)$ can be written as

$$f(t_A) = \int_0^{t_A} P(\mu w(s) ds) \left[\frac{w(t_A)}{w(t_A) + m(t_A)} \right] \frac{Z(t; t_0)}{w(s)}, \quad (23)$$

where $P(\mu w(s) ds)$ is a Poisson process which jumps one unit during the time interval $[s, s + \Delta s]$ with probability $\mu w(s) \Delta s$. The

integral above always reduces to the sum (2), with s_i as the jump times, but the integral form is easier to analyze. Specifically, for an arbitrary integral of such form, $I = \int_0^{t_A} P(\rho(s) ds)h(s)$, where $\rho(s)$ and $h(s)$ play the role of $\mu w(s)$ and $Z(t; s)/w(s)$, respectively, the mean and variance of I are given by

$$E[I] = \int_0^{t_A} ds \frac{\rho(s)}{h(s)}, \quad (24)$$

$$V[I] = \int_0^{t_A} ds \frac{\rho(s)}{h^2(s)}, \quad (25)$$

and the probability of no jump during a time interval $[0, t_1]$ is given by

$$\exp\left[-\int_0^{t_1} ds \rho(s)\right]. \quad (26)$$

See Kyprianou (2006) for a nice introduction to such computations.

Returning to (23), the expression in the brackets, $w(t_A)/w(t_A) + m(t_A)$, is of order $1 - O(\mu t_A)$. Ignoring small second order effects, we keep the 1 and drop the $O(\mu t_A)$, leading to the approximation

$$f(t_A) \approx \int_0^{t_A} P(\mu w(s) ds) \frac{Z(t; s)}{w(s)}. \quad (27)$$

The value of f_{\max} for the general estimator arises from applying a Chebyshev inequality to $E[f(t_A)]$. Since $E[Z(t; s)] = 1$, using (24) we have $E[f(t_A)] = \mu t_A$. Applying a Chebyshev bound gives

$$P(f(t_A) > f_{\max} \mu t_A) \leq \frac{E[f(t_A)]}{f_{\max} \mu t_A} = \frac{1}{f_{\max}}, \quad (28)$$

and we set $f_{\max} = 1/\sigma$.

To construct f_{\max} for the restricted estimator we apply a Chebyshev bound using the variance, i.e. $V[f(t_A)]$. However, directly applying a Chebyshev bound using the variance does not work. To see this, consider the special case of $Z(t; t_0) = 1$, i.e. no offspring stochasticity, and $r(t) = 1$, which leads to $V[f(t_A)] \approx \mu$. A Chebyshev bound using the second moment gives

$$P(|f(t_A) - \mu t_A| > f_{\max} \mu t_A) \leq \frac{V[f(t_A)]}{f_{\max}^2 \mu^2 t_A^2} = \frac{1}{f_{\max}^2 \mu t_A^2}. \quad (29)$$

Bounding the probability by σ requires $f_{\max} = O(1/\sqrt{\mu})$, a value greater than for the general estimator. The trouble derives from the heavy tails of $f(t_A)$; intuitively, heavy tails arise from the small probability that a mutation will occur soon after infection.

We handle the heavy tails by lopping them off and computing the variance of what remains, an example will demonstrate the approach. Consider again the case $Z(t; t_0) = 1$ and $r(t) = 1$, and split $f(t_A)$ into $f_{\text{tail}}(t_A)$ and $f_{\text{center}}(t_A)$ according to whether a mutation occurs before or after a time t_1

$$f_{\text{tail}}(t_A) = \int_0^{t_1} P(\mu w(s) ds) \frac{1}{w(s)},$$

$$f_{\text{center}}(t_A) = \int_{t_1}^{t_A} P(\mu w(s) ds) \frac{1}{w(s)}. \quad (30)$$

$f_{\text{tail}}(t_A)$ is the tail; we choose t_1 so the probability of a mutation occurring in this tail is $\sigma/2$, from (26):

$$\exp\left[-\int_0^{t_1} \mu w(s) ds\right] = 1 - \frac{\sigma}{2}. \quad (31)$$

Since $r(t) = 1$, we can calculate $t_1 \approx \log(\sigma/2\mu)$. Turning to $f_{\text{center}}(t_A)$, $V[f_{\text{center}}(t_A)] = 2\mu^2/\sigma$, which gives the Chebyshev bound

$$P(|f_{\text{center}}(t_A) - E[f_{\text{center}}(t_A)]| > a\mu t_A) \leq \frac{2\mu^2/\sigma}{a^2\mu^2 t_A^2}. \quad (32)$$

To bound $P(f_{\text{center}}(t_A) > f_{\max} \mu t_A)$ by $\sigma/2$ requires $f_{\max} = (1 + 2/\sigma t_A)$, which for $t_A = 14$ is about 1/5th of the $f_{\max} = 1/\sigma$ provided by the

general bound. Combining the two $\sigma/2$ bounds above shows $f_{\text{center}}(t_A) > f_{\max} \mu t_A$ with probability less than σ .

When offspring stochasticity is present, the tail is not identified with early mutation times because $Z(t_A; s)$ may itself have a heavy tail, allowing late mutations to produce large numbers of offspring. Instead, the tail corresponds to jumps of large size, meaning jumps of the Poisson process $P(\mu w(s) ds)$ for which $Z(t; s)/w(s)$ is large. Once the jumps are re-ordered according to size, the same arguments given in the example can be applied. Technical details provided in the appendix demonstrate $P(f(t_A) > f_{\max} \mu t_A) \leq \sigma$ when $f_{\max} = 1 + 2/\sigma t_A r_{\min}$, a result similar to the example except for the appearance of r_{\min} .

3.3. Construction of $k(t; \bar{k})$

Before constructing $k(t; \bar{k})$ for the general and restricted estimators, we explain the dependence of $f(t_F)$ on the profile of $k(t)$. Integrating the $w(t)$ equation in (7) from time t_A to t_F makes the dependence of $w(t_F)$ on $k(t)$ explicit:

$$w(t_F) = w(t_A) \exp\left[\int_{t_A}^{t_F} r(s) ds\right] \exp\left[-\int_{t_A}^{t_F} k(s) ds\right]. \quad (33)$$

The second exponential, $\exp[-\int_{t_A}^{t_F} k(s) ds]$, represents the contribution of $k(t)$ to $w(t_F)$; notice that the exponential can be rewritten as $\exp[-\bar{k}(t_F - t_A)]$, showing that $k(t)$ affects $w(t_F)$ only through \bar{k} .

In contrast, $m(t_F)$ depends on \bar{k} and the profile of $k(t)$. Consider the ratio $m(t_F)/w(t_F)$, its dependence on $m(t_A)/w(t_A)$ and $k(t)$ is given by

$$\frac{m(t_F)}{w(t_F)} = \exp[\bar{k}(t_F - t_A)] \left[\frac{m(t_A)}{w(t_A)} + \mu \Gamma \right], \quad (34)$$

where Γ is defined above in (18).

To obtain the general estimator, we replace Γ by an upper bound $t_F - t_A$, thereby achieving the maximum possible value of $m(t_F)/w(t_F)$. Intuitively, once \bar{k} is chosen, the $k(t)$ profile that maximizes the mutant frequency at t_A delays all of the killing until time t_F . Since \bar{k} is fixed, the number of wild types at t_F is fixed but delaying killing until time t_F raises the number of wild types extant during the interval t_A to t_F , thereby raising the number of wild type mutations during that interval and, in turn, the mutant frequency at t_F . As a specific example, consider the following $k(t; \bar{k})$ family of profiles:

$$k(t; \bar{k}) = \begin{cases} 0 & \text{for } t < t_F - \epsilon \\ \frac{2\bar{k}}{\epsilon}(t_F - t) & \text{for } t \geq t_F - \epsilon, \end{cases} \quad (35)$$

where ϵ is the length of some small time period prior to t_F during which all the killing occurs. The general bound is achieved by taking $\epsilon \rightarrow 0$.

Under the restriction $k''(t) > 0$, an upper bound on Γ is achieved using the profile family $k(t; \bar{k}) = 2\bar{k}(t - t_A)/(t_F - t_A)$. Intuitively, we again delay killing, but the restriction $k''(t) > 0$ prevents the biologically unrealistic case of all killing occurring at t_F . For $\bar{k}(t_F - t_A) > 3$, which is the regime of all our datasets, $k(t; \bar{k})$ is well approximated by $\sqrt{\pi(t_F - t_A)/4\bar{k}_R}$.

3.4. Model extension

The switch in the model from stochastic to deterministic dynamics need not occur at t_A . In particular, when $m(t_A)$ is small, perhaps due to an early immune response, the stochastic dynamics can be extended through a parameter t_{switch} which specifies the switch time. Wild type dynamics prior to t_{switch}

become

$$\frac{dw}{dt}(t) = (r(t) - k(t))w(t), \quad (36)$$

where $k(t) = 0$ prior to t_A . As before, $r(t)$ and $k(t)$ can be specified arbitrarily. Evaluation of $m(t_A)$ through (2) is replaced by evaluation of $m(t_{\text{switch}})$ through

$$m(t_A) = \sum_{i=1}^N X(t_{\text{switch}}; s_i) \quad (37)$$

and

$$E[X(t_{\text{switch}}, s)] = \exp \left[\int_s^{t_{\text{switch}}} r(s') ds' \right], \quad (38)$$

Importantly, (36) assumes that $w(t)$ dynamics are independent of $m(t)$ dynamics, a plausible assumption when mutant frequencies are small. t_{switch} should be chosen to satisfy $f(t_{\text{switch}}) \ll 1$, to ensure small frequencies, but also to satisfy $m(t_{\text{switch}}) \gg 1$, to ensure deterministic dynamics after t_{switch} . For our datasets $t_{\text{switch}} = t_A$ satisfies this requirement.

4. Discussion

Inferring the rate of the first CTL escape involves a trade-off. On one hand, the existing two-point method is largely model independent but can only be applied using two sampled time-points, meaning that the early part of the escape is often missed and inference implicitly focuses on later parts of the escape. On the other hand, if a more parametrized method is used, the early part of the escape can be considered but inference results depend on model structure and the parameter values chosen.

In this work, we have developed escape rate inference methods applicable to single timepoint datasets with an effort to minimize model dependence. To do this, we developed a general model and constructed lower-bound estimators valid across large portions of the model's parameter space.

Lower-bound estimators allow us to compare escape rates in a statistically meaningful way which accounts for model and parameter uncertainty. Through the Bimber et al. dataset, lower-bound estimates combined with the two-point method reveal faster rates for the first CTL escape during days 12–21 than days 21–28. The Vanderford et al. data shows roughly the same pattern, with faster rates of escape in the lymph nodes during days 9–14 than during days 14–28.

In this work, we have not distinguished between different types of epitope mutations, although CTL escape typically involves multiple mutation variants (Boutwell et al., 2010). As CTL binding affinity may differ between mutation variants, we are really inferring an average escape rate over all mutations at the epitope. For TAT-SL8, most mutations arising in escape have low binding affinity, so our estimates may apply without much modification (Allen et al., 2000; O'Conner et al., 2002). Nevertheless, more work is required to address this limitation.

Besides developing estimators, our model and accompanying analysis provides a basis through which to understand early infection stochasticity and its impact on escape rate inference. Soon after peak viral load, in both HIV and SIV, multiple CTL escapes occur, often overlapping in time (Bimber et al., 2010; Henn et al., 2012; Goonetilleke et al., 2009; Boutwell et al., 2010). The interaction of viral variants involved in such sweeps, both through inter-variant competition for target cells and possible recombination events, makes modeling and inference complex (Leviyang, 2013; Neher and Leitner, 2010; Kessinger et al., 2013; Batorsky et al., 2011; Gantsov et al., 2011). The parameter space becomes much larger, multiple escapes and multiple variants within each escape lead to

potentially dozens of parameters. Inferring escape rates in such a high dimensional space through deterministic models will likely lead to overfitting. Extending the current work to multiple escape settings may be helpful in avoiding such difficulties.

The lower-bound estimators as currently constructed are overly conservative. Often, as demonstrated by numerical experiments in Section 3, the lower-bound CIs significantly underestimate the escape rate. Improved lower-bound estimators require better quantitative understanding of acute infection. For example, the number of offspring infected cells descendant from a single HIV or SIV infected cell is not well understood. Future work leading to improved lower-bound estimators would expand our ability to analyze individual escapes as well as compare escapes against each other.

Acknowledgements

I would like to thank Guido Silvestri and Thomas Vanderford for providing the Vanderford et al. dataset analyzed in this paper. I thank Shelby O'Connor for providing several datasets that helped me understand early SIV infection and also for answering several questions relating to the Bimber et al. dataset and SIV infection in general. I thank Roland Regoes for several helpful suggestions. I thank Hai Zhou for many profitable discussions and for finding several errors in an earlier version of this manuscript.

I am deeply indebted to two anonymous reviewers whose comments and suggestions greatly improved this work.

This work was supported by NSF grant DMS-1225601.

Appendix A

Here we derive the bound $P(f(t_A) > f_{\max} \mu t_A) < \sigma$ for $f_{\max} = 1 + (2/\sigma t_A r_{\min})$ under the restriction $r(t) \geq r_{\min}$. For simplicity we assume that $Z(t_A, s)$ has a density, written as $g(z, s)$ to reflect possible dependence on the mutation time s , although the arguments below work for discrete distributions as well. The first few computations below use some basic ideas from the theory of Levy processes, see Kyprianou (2006) for an introduction.

To start, we order the jump sizes of $f(t_A)$ so the tail can be identified; this corresponds to writing the Laplace exponent of $f(t_A)$ in standard form. Starting with (27), we can compute the Laplace transform of $f(t_A)$ and the associated Laplace exponent $\Psi(\lambda)$ (i.e. $E[\exp[-\lambda f(t_A)]] = \exp[-\Psi(\lambda)]$):

$$\Psi(\lambda) = \int_0^{t_A} ds \mu w(s) \int_0^\infty dz g(z, s) \left(1 - \exp \left[-\lambda \frac{z}{w(s)} \right] \right) \quad (39)$$

Changing variables through $v = w(s)$ and flipping the order of integration give

$$\Psi(\lambda) = \int_0^\infty dx \rho(x) (1 - \exp[-\lambda x]) \quad (40)$$

with

$$\rho(x) = \int_1^{w(t_A)} dv \frac{\mu v^2}{w'(s)} f(xv, s). \quad (41)$$

By well known results in the theory of Levy distributions, $\rho(x)$ is the rate at which jumps of size x occur, allowing us to re-express $f(t_A)$ as

$$f(t_A) = \int_0^\infty P(\rho(x) dx) x. \quad (42)$$

Beyond this point, the computations only depend on calculus and (24)–(26).

Through (42), the tail of $f(t_A)$ can be identified, and we split $f(t_A)$ into two pieces according to whether x is greater or less than

a value x_0 .

$$f_{\text{tail}}(t_A) = \int_{x_0}^{\infty} P(\rho(x)) dx, \\ f_{\text{center}}(t_A) = \int_0^{x_0} P(\rho(x)) dx \quad (43)$$

We choose x_0 so the probability of no jump in the tail, i.e. $f_{\text{tail}}(t_A) = 0$, is $\sigma/2$, which requires

$$\int_{x_0}^{\infty} dx \rho(x) = \sigma/2. \quad (44)$$

To identify x_0 , we execute the following arguments:

1. Using the inequality $w'(s) \geq r_{\min} w(s)$, a consequence of the assumption $r(t) \geq r_{\min}$ gives the bound

$$\int_{x_0}^{\infty} dx \rho(x) \leq \frac{\mu}{r_{\min}} \int_{x_0}^{\infty} dx \int_1^{w(t_A)} dv v f(xv, s) \quad (45)$$

2. Apply the transform $v' = xv$ to the integral on the right in step 1 to find,

$$\int_{x_0}^{\infty} dx \rho(x) \leq \frac{\mu}{r_{\min}} \int_{x_0}^{\infty} dx \frac{1}{x^2} \int_{x_0}^{x_0 w(t_A)} dv v f(v, s) \quad (46)$$

3. Since $E[Z(t_A; s)] = 1$, the dv integral on the right of step 2 is less than 1, leading to

$$\int_{x_0}^{\infty} dx \rho(x) \leq \frac{\mu}{x_0 r_{\min}} \quad (47)$$

At the end of these steps we conclude

$$x_0 \leq \frac{2\mu}{\sigma r_{\min}}. \quad (48)$$

Next, as in the example, we bound the variance of $f_{\text{center}}(t_A)$.

$$V[f_{\text{center}}(t_A)] = \int_0^{x_0} dx \rho(x) x^2. \quad (49)$$

Using the same transforms as steps 1 and 3 above

$$V[f_{\text{center}}(t_A)] \leq \frac{\mu}{r_{\min}} \int_0^{x_0} dx = \frac{\mu x_0}{r_{\min}} \quad (50)$$

The rest is exactly as in the example.

References

- Allen, T.M., O'Conner, D.H., Jing, P., Dzuris, J.L., et al., 2000. Tat-specific cytotoxic T lymphocytes select for SIV escape variants during resolution of primary viraemia. *Nature* 407, 386–390.
- Althaus, C.L., de Boer, R.J., 2008. Dynamics of immune escape during HIV/SIV infection. *PLOS Comput. Biol.* 4 (7), 1–10.
- Asquith, B., Edwards, C.T.T., Lipsitch, M., McLean, A.R., 2006. Inefficient CTL mediated killing of HIV-1 infected cells in-vivo. *PLOS Biol.* 4 (4), 583–592.
- Asquith, B., McLean, A.R., 2007. In vivo CD8+ T cell control of immunodeficiency virus infection in humans and macaques. *Proc. Natl. Acad. Sci.* 104 (15), 6365–6370.
- Batorsky, R., Kearney, M.F., Palmer, S.E., Maldarelli, F., Rouzine, I.M., et al., 2011. Estimate of effective recombination rate and selection coefficient for HIV chronic infection. *Proc. Natl. Acad. Sci.* 108 (14).
- Bimber, B.N., Dudley, D.M., Lauck, M., Becker, E.A., et al., 2010. Whole-genome characterization of human and simian immunodeficiency virus intrahost diversity by ultradeep pyrosequencing. *J. Virol.* 84 (22), 12087–12092.
- Bimber, B.N., et al., 2009. Ultradeep pyrosequencing detects complex patterns of CD8 T-lymphocyte escape in simian immunodeficiency virus-infected macaques. *J. Virol.* 83 (16), 8247–8253.
- Borrow, P.H., et al., 1994. Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary HIV-1 infection. *J. Virol.* 68, 6103–6110.
- Boutwell, C.L., et al., 2010. Viral evolution and escape during acute HIV-1 infection. *J. Infect. Dis.* 202 (2), 309–314.
- Cohen, M.S., Shaw, G.M., McMichael, A.J., Haynes, B.F., et al., 2011. Acute HIV-1 infection. *New Engl. J. Med.* 364 (20), 1943–1954.
- Cromer, D.C., Khanna R. Tey, S.-K., Davenport, M.P., 2013. Estimating cytomegalovirus growth rates by using only a single timepoint. *J. Virol.* 87 (6), 3376–3381.
- Desai, M.M., Fisher, D.S., 2007. Beneficial mutation-selection balance and the effect of linkage on positive selection. *Genetics* 176, 1759–1798.
- Fernandez, C.S., Stratov, I., De Rose, R., Walsh, K., Dale, C.J., et al., 2005. Rapid viral escape at an immunodominant simian-human immunodeficiency CTL epitope exacts a dramatic fitness cost. *J. Virol.* 79 (9), 5721–5731.
- Fisher, W., Ganusov, V.V., Giorgi, E.E., Hrabner, P.T., Keele, B.F., et al., 2010. Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. *PLOS One* 5 (8).
- Ganusov, V.V., Goonetilleke, N., Liu, M.K.P., Ferrari, G., Shaw, G.M., et al., 2011. Fitness costs and diversity of the CTL response determine the rate of CTL escape during acute and chronic phases of HIV infection. *J. Virol.* 85 (50), 10518–10528.
- Ganusov, V.V., Neher, R.A., Perelson, A.S., 2013. Mathematical modeling of escape of HIV from cytotoxic T lymphocyte responses. *J. Stat. Mech.: Theor. Exp.*
- Goonetilleke, N., Liu, M.K.P., Salazar-Gonzalez, J.F., Ferrari, G., Giorgi, E., et al., 2009. The first t cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. *J. Exp. Med.* 206 (6), 1253–1272.
- Goulder, P., Watkins, D.I., 2004. HIV and SIV CTL escape: implications for vaccine design. *Nat. Rev. Immunol.* 4, 630–640.
- Henn, M.R., et al., 2012. Whole genome sequencing of HIV-1 reveals impact of early minor immune variants on immune recognition during acute infection. *PLOS Pathog.* 8 (3).
- Kessinger, T.A., Perelson, A.S., Neher, R.A., 2013. Inferring HIV escape rates from multi-locus genotype data. *Frontiers in Immunology* 4, 252.
- Kouyos, R.D., et al., 2006. Stochastic or deterministic: what is the effective population size of HIV-1. *Trends Microbiol.* 14 (12), 507–511.
- Kyprianou, A.E., 2006. Fluctuations of Lévy Processes with Applications: Introductory Lectures. Universitext.
- Leigh-Brown, A.J., 1997. Analysis of HIV-1 env gene sequences reveals evidence for a low effective number in the viral population. *Proc. Natl. Acad. Sci.* 94, 1862–1865.
- Leviyang, S., 2013. Computational inference methods for selective sweeps arising in acute HIV infection. *Genetics* 194, 737–752.
- Liu, Y., et al., 2006. Selection on the human immunodeficiency virus type 1 in a T lymphoid cell line. *Aids Res. Human Retrovir.* 12 (4), 9519–9529.
- Loh, L., Petravic, J., Batten, C.J., Davenport, M.P., Kent, S.J., 2008. Vaccination and timing influence SIV immune escape viral dynamics in vivo. *PLOS Pathog.* 4 (1), 27–37.
- Love, T., Thurston, S.W., Keefer, M.C., Dewhurst, S., Lee, H.Y., 2008. Mathematical modeling of ultradeep sequencing data reveals that acute CD8+ T-lymphocyte responses exert strong selective pressure in simian immunodeficiency virus-infected macaques but still fail to clear founder epitope sequences. *J. Virol.* 84 (11), 5802–5814.
- Mandl, J., et al., 2007. Estimating the effectiveness of SIV-specific CD8+ T cells from the dynamics of viral immune escape. *J. Virol.* 81, 11982–11991.
- Mansky, L.M., 1996. Forward mutation rate of human immunodeficiency virus type 1 in a T lymphoid cell line. *Aids Res. Human Retrovir.* 12 (4), 307–314.
- McMichael, A.J., Borrow, P., Tomaras, G.D., Goonetilleke, N., Haynes, B.F., 2010. The immune response during acute HIV-1 infection: clues for vaccine development. *Nat. Rev. Immunol.* 10, 11–23.
- Messer, P.W., Neher, R.A., 2012. Estimating the strength of selective sweeps from deep population diversity data. *Genetics* 191, 563–605.
- Monteiro, L.H.A., Gonhalves, C.H.O., Piqueira, J.R.C., 2000. A condition for successful escape of a mutant after primary HIV infection. *J. Theor. Biol.* 203 (1), 399–406.
- Neher, R.A., Leitner, T., 2010. Recombination rate and selection strength in HIV intra-patient evolution. *PLOS Comput. Biol.* 6 (1).
- Nowak, M.A., May, R.M., 2000. *Virus Dynamics: Mathematical Principles of Immunology and Virology*. Oxford University Press.
- O'Conner, D.H., Allen, T.M., Vogel, T.U., Jing, P., et al., 2002. Acute phase cytotoxic T lymphocyte escape is a hallmark of simian immunodeficiency virus infection. *Nat. Med.* 8 (5), 493–499.
- Palmer, D., Frater, J., Phillips, R., McLean, A.R., McVean, G., 2013. Integrating genealogical and dynamical modelling to infer escape and reversion rates in HIV epitopes. *Proc. R. Soc. B* 280.
- Perelson, A.S., 2002. Modeling viral and immune system dynamics. *Nat. Rev.* 2, 28–36.
- Petravic, J., Loh, L., Kent, S.J., Davenport, M.P., 2008. CD4+ target cell availability determines the dynamics of immune escape and reversion in vivo. *J. Virol.* 82 (8), 4091–4101.
- Rouzine, I.M., Coffin, J.M., 2005. Evolution of HIV under selection and weak recombination. *Genetics* 170, 7–18.
- Turnbull, E.L., Wong, M., Wang, S., Wei, X., et al., 2009. Kinetics of expansion of epitope-specific T cell responses during primary HIV-1 infection. *J. Immunol.* 182, 7131–7145.
- Vanderford, T.H., Bleckwehl, C., Engram, J.C., Dunham, R.M., et al., 2011. Viral CTL escape mutants are generated in lymph nodes and subsequently become fixed in plasma and rectal mucosa during acute SIV infection of macaques. *PLOS Pathog.* 7 (5), 1–10.
- Veazey, R.S., Lifson, J.D., Schmitz, M., Kuroda, J.E., et al., 2003. Dynamics of simian immunodeficiency virus-specific cytotoxic t-cell responses in tissues. *J. Med. Primatol.* 32, 194–200.
- Yasutomi, Y., Reimann, K.A., Lord, C.I., Miller, M.D., Letvin, N.L., et al., 1994. Simian immunodeficiency virus-specific CD8+ lymphocyte response in acutely infected rhesus monkeys. *J. Virol.* 67, 1707–1711.