Linking Dynamical and Population Genetic Models of Persistent Viral Infection

John K. Kelly,[1,*] Scott Williamson,[1] Maria E. Orive,[1] Marilyn S. Smith,[2] and Robert D. Holt[3]

1. Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas 66045-7534;
2. Department of Microbiology, University of Kansas Medical Center, Kansas City, Kansas 66160-7420;
3. Department of Zoology, P.O. Box 118525, University of Florida, Gainsville, Florida 32611-8525

Submitted July 22, 2002; Accepted December 18, 2002; Electronically published June 12, 2003

abstract: This article develops a theoretical framework to link dynamical and population genetic models of persistent viral infection. This linkage is useful because, while the dynamical and population genetic theories have developed independently, the biological processes they describe are completely interrelated. Parameters of the dynamical models are important determinants of evolutionary processes such as natural selection and genetic drift. We develop analytical methods, based on coupled differential equations and Markov chain theory, to predict the accumulation of genetic diversity within the viral population as a function of dynamical parameters. These methods are first applied to the standard model of viral dynamics and then generalized to consider the infection of multiple host cell types by the viral population. Each cell type is characterized by specific parameter values. Inclusion of multiple cell types increases the likelihood of persistent infection and can increase the amount of genetic diversity within the viral population. However, the overall rate of gene sequence evolution may actually be reduced.

Keywords: HIV, parasites, rapid evolution, viral dynamic.

In recent years, biologists have become increasingly aware that evolution frequently occurs over ecological timescales (Thompson et al. 2001). This is particularly apparent for infectious diseases, where significant evolution can occur even within a single infected host. Because of their public health importance, viral pathogens that establish persistent infections have been the focus of intensive theoretical study. This category includes lentiviruses such as the human immunodeficiency virus (HIV), hepatitis B and C viruses, simian immunodeficiency virus (SIV), and equine infectious anemia virus, among others. These pathogens undergo many replication cycles and experience rapid evolutionary changes within a single infected host. The statistical nature of these processes has prompted the development of mathematical models. Quantitative models can generate nonintuitive predictions and suggest new avenues for experimental investigation (Nowak and Bangham 1996; Levin et al. 1999).

Most of the theory can be classified into two distinct categories: dynamical models, which characterize the population dynamic interaction between virus and host cells, and population genetic models, which predict gene sequence evolution of the viral population. The dynamical theory is an adaptation of classical epidemiological models in which the dynamical variables are the numbers of virions (free viral particles) and the numbers of cells, both infected and uninfected, of various types. These models have been used to predict the conditions for the establishment of infection, the quantitative responses to antiviral drug therapy, and the causes of disease progression (Perelson 1989, 2002; Nowak et al. 1991; Perelson et al. 1993, 1996; Wodarz and Nowak 1999; Nowak and May 2000). Complementary models have been developed for bacterial infections (Lipsitch and Levin 1997; Kirschner 2001).

The population genetic theory has developed in response to the observation that viral populations undergo extensive gene sequence evolution within a single infected individual (Hahn et al. 1986; Burns and Desrosiers 1991; Simmonds et al. 1991; Shankarappa et al. 1999). The high mutation rate of lentiviruses in concert with high numbers of replication cycles in vivo generates a great deal of genetic variability. For example, the mutation rate of HIV-1 is on the order of 10^{-5} to 10^{-4} per base pair (Preston et al. 1988; Roberts et al. 1988; Mansky 1996), and an estimated 10^{9} replication cycles occur per day within a single patient (Ho et al. 1995; Wei et al. 1995; Perelson et al. 1996). Population genetic models have been developed to investigate the forces acting on this variability and to estimate a diversity

* Corresponding author; e-mail: jkk@ku.edu.

Am. Nat. 2003. Vol. 162, pp. 14–28.

of evolutionary parameters (Kelly 1994, 1996; Crandall 1996, 1999; Wolinsky et al. 1996; Leigh Brown 1997; Muse 1999; Rodrigo and Felsenstein 1999; Rodrigo et al. 1999; Zanotto et al. 1999; Moya et al. 2000; Drummond et al. 2001; Frost et al. 2001*a*, 2001*b*).

While the dynamical and population genetic theories have developed independently to a large extent, the biological processes they describe are completely interrelated. Parameters of the dynamical models are important determinants of evolutionary processes such as natural selection and genetic drift. In particular, gene sequence evolution is typically modeled as a mutation-limited process. The rate of evolution along a lineage is directly proportional to the number of replication cycles per unit time (Kimura 1983; Muse 1999; Rodrigo et al. 1999). For viral infections, the number of replication cycles per unit time is determined by dynamical parameters such as the availability of target cells, the half-life of free virus in the host, the half-life of infected cells, and the rate of cell-to-cell transmission of virus (Herz et al. 1996; Perelson et al. 1996).

Our first purpose with this article is to develop an appropriate theoretical framework to link dynamical and population genetic models. We introduce an analytical method, based on coupled differential equations, to predict the accumulation of replication cycles (and hence genetic diversity) as a function of dynamical parameters. The elements of this framework are developed with a simple dynamical model and then extended to more complicated and realistic situations. Here, we consider one important complication, cellular population structure (CPS). We use this term to refer to viral infection of multiple cell types with differing biological properties. Other generalizations will follow in subsequent articles. We also suggest that the general approach may be more widely applicable than host-viral systems because there are many parallels between infectious disease processes and the dynamics of other biological systems (Holt 2000).

Infection of multiple distinct cell types in vivo has been documented in a number of viruses. The primary targets of HIV are activated CD4+ T cells and monocyte/macrophages (Levy 1993), but other related cells in the peripheral blood and in tissues are also infected. These include long-lived CD4+ memory T cells and dendritic cells of the monocyte lineage (Chun et al. 1997; Granelli-Piperno et al. 1998). Simian immunodeficiency virus infects the same range of cell types. Hepatitis B virus infects hepatocytes in addition to T and B lymphocytes. The importance of CPS lies in the fact that the virus behaves differently in each cell type.

Basic dynamical parameters vary among cell types. For example, the death rate of different cell types when infected with HIV-1 has been estimated using patient responses to highly active antiretroviral therapy (HAART). The estimated death rate is 0.99/d for activated T cells, 0.049/d for monocyte/macrophages, and 0.0005/d for "latently infected" memory CD4 T cells (Finzi et al. 1997, 1999). Latently infected cells harbor the provirus genome but are not currently expressing viral transcripts. Other important parameters, such as infectivity or the rate at which infected cells produce virions, will also generally differ among cell types. There is accumulating evidence to suggest that CPS has a range of important consequences for viral dynamics and evolution (see "Discussion").

In this article, we develop a model to address a specific set of questions. How does CPS affect conditions for persistent infection? In other words, does the ability to infect multiple cell types allow the viral population to persist under conditions in which infection fails with only one cell type? When infection is maintained, how does CPS impact equilibrium levels of virus load and the numbers of infected cells? Finally, we explore the effect of CPS on the rate and pattern of viral gene sequence evolution.

Theory

Single Cell Type Model

We first consider the dynamical model with one type of target cell. The dynamical variables are the number of virions (q), the number of uninfected cells (n), and the number of infected cells (n^*). We assume that changes in these quantities are governed by the system of differential equations

$$\frac{dq}{dt} = \nu\mu^* n^* - \mu' q - \beta q n, \tag{1}$$

$$\frac{dn}{dt} = \lambda - \mu n - \beta q n, \tag{2}$$

$$\frac{dn^*}{dt} = \beta q n - \mu^* n^*, \tag{3}$$

where ν is the "burst size," the number of virions released when an infected cell dies, μ' is the death rate of virions (clearance rate), β is the infection constant, λ is the input rate of uninfected cells, μ is the death rate of uninfected cells, and μ^* is the death rate of infected cells. We will introduce a substantial number of variables in this paper (a summary is provided in table 1). Throughout, we will use Roman letters to represent dynamical variables and Greek letters to represent constants.

Equations (1)–(3) are similar to the basic viral dynamics model of Nowak and Bangham (1996; see also Nowak and

Table 1: Index of variables

Constants:
- ν = burst size, the number of virions produced from an infected cell
- λ = input rate of uninfected cells
- μ = death rate of uninfected cells
- μ^* = death rate of infected cells
- μ' = death rate of virions (clearance rate by host immune system)
- β = infection constant
- c = decline rate of virions
- τ = mean recurrence time to the virion state
- ω = variance in recurrence time to the virion state
- ε_i = proportion of new infections into cell type i

Dynamical variables:
- q = number of virions
- n = number of uninfected cells
- n^* = number of infected cells
- $U_i(t)$ = proportion of viral genomes that are virions with i replication cycles in their ancestry
- $A_i(t)$ = proportion of viral genomes that are provirus and have i replication cycles in their ancestry
- $M_U(t)$ = mean number of ancestral replication cycles among virions
- $M_A(t)$ = mean number of ancestral replication cycles among provirus
- $V_U(t)$ = variance in number of ancestral replication cycles among virions
- $V_A(t)$ = variance in number of ancestral replication cycles among provirus

Note: The constants ν, λ, μ, μ^*, and β are subscripted in the multiple cell type model. Each of the dynamical variables, except q, is subscripted in the multiple cell type model.

May 2000 and references therein). There are two equilibrium solutions, and the outcome depends on R_0:

$$R_0 = \frac{\beta\lambda(\nu - 1)}{\mu\mu'}. \tag{4}$$

If $R_0 < 1$, the infection fails, leaving only uninfected cells ($q = 0$, $n^* = 0$, $n = \lambda/\mu$). If $R_0 > 1$, the system rapidly converges to the following equilibrium:

$$q = \frac{\lambda(\nu - 1)}{\mu'} - \frac{\mu}{\beta} = \frac{\mu}{\beta}(R_0 - 1), \tag{5}$$

$$n = \frac{\mu'}{\beta(\nu - 1)} = \frac{\lambda}{\mu R_0}, \tag{6}$$

$$n^* = \frac{\lambda}{\mu^*} - \frac{\mu\mu'}{\beta\mu^*(\nu - 1)} = \frac{\lambda}{\mu^* R_0}(R_0 - 1). \tag{7}$$

Nowak and May (2000, chap. 3) describe the basic dependencies of this equilibrium on the parameters. Briefly, the proportion of cells that are infected is positively related to β but inversely related to μ^*. The total number of viral genomes ($q + n^*$) increases with β, ν, and λ but decreases with μ, μ^*, and μ'.

Equations (4)–(7) indicate the conditions necessary for infection and the abundance of virus during infection but not the rate that it will evolve. In order to derive predictions relevant to molecular evolution, we need to track the accumulation of replication cycles within the viral population through time. A new replication cycle is added to a lineage with each successful cellular infection. This method of accounting is appropriate for retroviruses such as HIV because most mutations will occur at this stage of the life cycle. The error-prone process of reverse transcription is responsible for the conversion of the viral genome from virion RNA to proviral DNA.

Among virions and provirus at time t, what proportion have i replication cycles in their ancestry (since the initiation of infection)? To address this question, we assume that the viral population is at the dynamical equilibrium defined by equations (5)–(7). This approximation is justified by the rapid approach of the system to equilibrium. We further assume that each infected cell harbors a single viral genome. While this may not be the case for many viruses (Potash and Volsky 1998), the theory can be generalized to allow superinfection of cells. Let f_U and f_A represent the proportions of viral genomes in virion and proviral form,

$$f_U = \frac{q}{n^* + q},$$

$$f_A = \frac{n^*}{n^* + q} = 1 - f_U, \tag{8}$$

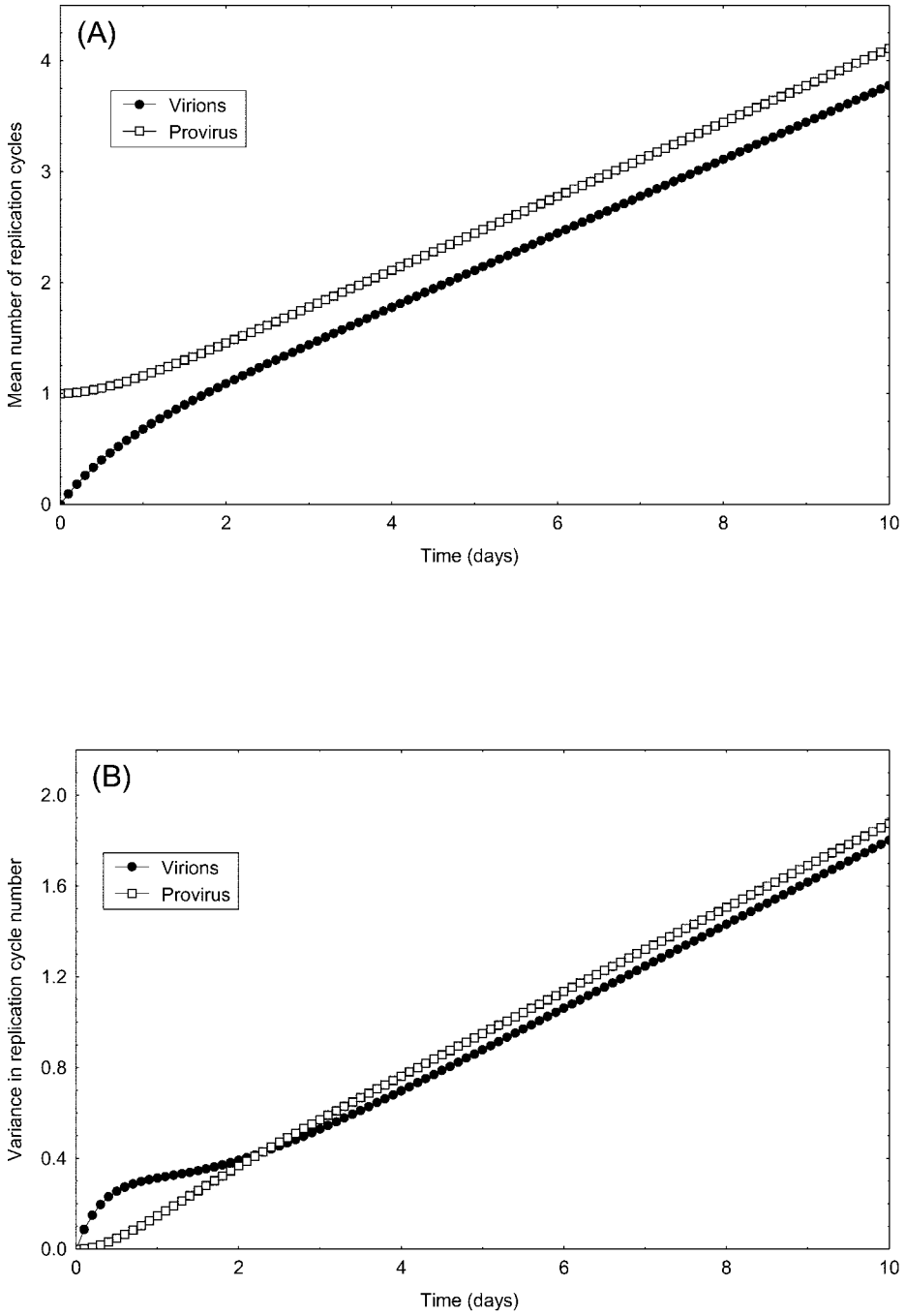

Figure 1: Mean (*A*) and variance (*B*) of replication cycle number as a function of time; in this example, $\mu' = 1$ and $\mu^* = 0.5$

respectively, where q and n^* are given by equations (5) and (7).

Let $U_i(t)$ denote the proportion of viral genomes that are virions with i replication cycles in their ancestry. Let $A_i(t)$ denote the proportion of all viral genomes that exist as proviral DNA (incorporated in infected cells) and have i replication cycles in their ancestry. Both series are specific to a given time t and satisfy the following equation:

$$U_0(t) + \sum_{i=1}^{\infty} [U_i(t) + A_i(t)] = 1. \quad (9)$$

There is no zero class for provirus ($A_0[t] = 0$) because all infected cells have at least one replication cycle in their ancestry. The initial virion inoculum declines exponentially due to both successful infection of cells and clearance by the host immune system:

$$U_0(t) = f_U e^{-ct}, \quad (10)$$

where

$$c = \mu' + \beta n = \mu'\left(1 + \frac{1}{\nu - 1}\right). \quad (11)$$

The right-hand side of equation (11) is obtained by substituting equation (6) in the central term. The following differential equations describe changes in $U_i(t)$ and $A_i(t)$ for $i > 0$:

$$\frac{\partial U_i(t)}{\partial t} = \nu\mu^* A_i(t) - cU_i(t), \quad (12)$$

$$\frac{\partial A_i(t)}{\partial t} = \beta n U_{i-1}(t) - \mu^* A_i(t)$$

$$= \frac{\mu'}{\nu - 1} U_{i-1}(t) - \mu^* A_i(t). \quad (13)$$

The two terms in each differential equation represent the flow into and out of a given category. The change in $A_i(t)$ is determined by new infections from virions with $i - 1$ replication cycles (the inflow) minus the death of infected cells with i replication cycles (the outflow). The change in $U_i(t)$ is determined by the influx of new virions from the death of infected cells in the ith class minus the clearance of virions from this category.

Equations (12) and (13) are linear differential equations, and closed form solutions can be found. Unfortunately, these solutions become increasingly cumbersome and uninformative as i increases (except in the special case discussed below). For this reason, it is useful to summarize the distribution of replication cycles in terms of the mean and the variance. Let $M_U(t)$ denote the mean number of replication cycles among virions and $M_A(t)$ denote the corresponding mean among provirus at time t:

$$M_U(t) = \frac{1}{f_U} \sum_{i=0}^{\infty} iU_i(t),$$

$$M_A(t) = \frac{1}{f_A} \sum_{i=0}^{\infty} iA_i(t). \quad (14)$$

Taking derivatives of equations (14), we obtain the following:

$$\frac{\partial M_U(t)}{\partial t} = \frac{f_A}{f_U} \nu\mu^* M_A(t) - cM_U(t), \quad (15a)$$

$$\frac{\partial M_A(t)}{\partial t} = -\mu^* M_A(t) + \frac{f_U}{f_A} \frac{\mu'}{\nu - 1} [M_U(t) + 1]. \quad (15b)$$

This is a simple 2×2 linear system with initial conditions $M_U(0) = 0$ and $M_A(0) = 1$. These equations can be further simplified by noting that ν will generally be much greater than 1. If $\nu \gg 1$, then $c \cong \mu'$, $f_A \cong \mu'/(\mu^*\nu + \mu')$, and $f_U \cong \mu^*\nu/(\mu^*\nu + \mu')$. Using this approximation, we find that

$$\frac{\partial M_U(t)}{\partial t} = \mu'[M_A(t) - M_U(t)], \quad (16a)$$

$$\frac{\partial M_A(t)}{\partial t} = \mu^*[1 + M_U(t) - M_A(t)]. \quad (16b)$$

Incorporating the initial conditions, we obtain the solution

$$M_A(t) = 1 + \frac{\mu'\mu^*}{(\mu' + \mu^*)^2} [e^{-t(\mu' + \mu^*)} - 1 + t(\mu' + \mu^*)], \quad (17)$$

$$M_U(t) = \frac{\mu'\mu^*}{(\mu' + \mu^*)^2} \left[\frac{\mu'}{\mu^*}(1 - e^{-t(\mu' + \mu^*)}) + t(\mu' + \mu^*)\right]. \quad (18)$$

Equations (17) and (18) are asymptotically linear functions of t with a common slope $\mu'\mu^*/(\mu' + \mu^*)$. The reciprocal of this slope, $(\mu' + \mu^*)/(\mu'\mu^*)$, is perhaps the most meaningful measure of generation time from an evolutionary perspective.

Let $V_U(t)$ denote the variance in replication cycle number among virions and $V_A(t)$ denote the corresponding variance among provirus:

$$V_U(t) = \left[\frac{1}{f_U} \sum_{i=0}^{\infty} i^2 U_i(t)\right] - M_U(t)^2,$$

$$V_A(t) = \left[\frac{1}{f_A} \sum_{i=0}^{\infty} i^2 A_i(t)\right] - M_A(t)^2. \quad (19)$$

Using an analysis similar to that leading to equations (17) and (18), we find that

$$V_A(t) = \frac{\mu^*\mu'}{(\mu^* + \mu')^4} \{3\mu^*\mu' - (\mu^*)^2 - (\mu')^2 + t[(\mu')^2 + (\mu^*)^2](\mu' + \mu^*) + K_1(t)\}, \quad (20)$$

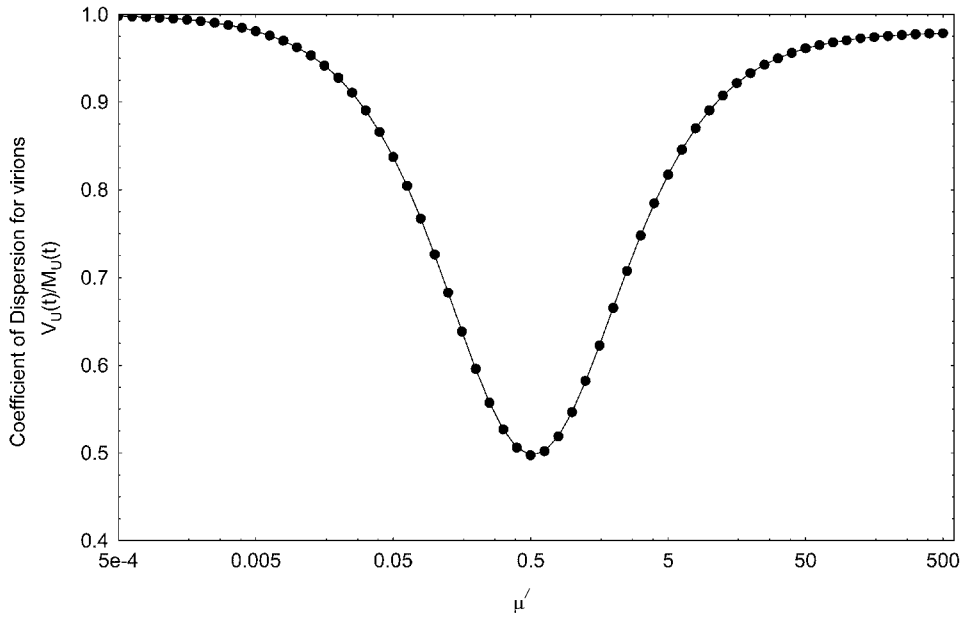

Figure 2: Ratio of $V_U(t)$ to $M_U(t)$, also known as the coefficient of dispersion, at $t = 100$ as a function of μ'; in this example, $\mu^* = 0.5$

where

$$K_1(t) = e^{-2t(\mu^*+\mu')}\{-\mu'\mu^* + e^{t(\mu'+\mu^*)} \times [(\mu^*)^2 - \mu'\mu^*(2 + 4t\mu^*) - (\mu')^2(4t\mu^* - 1)]\}, \quad (21)$$

and

$$V_U(t) = \frac{\mu^*\mu'}{(\mu^* + \mu')^4}\{3\mu^*\mu' - 2(\mu')^2 + t[(\mu')^2 + (\mu^*)^2](\mu' + \mu^*) + K_2(t)\}, \quad (22)$$

where

$$K_2(t) = \frac{\mu' e^{-2t(\mu^*+\mu')}}{\mu^*}\{-(\mu')^2 + e^{t(\mu'+\mu^*)}[-3(\mu^*)^2 + 2\mu'\mu^*(1 + 2t\mu^*) + (\mu')^2(1 + 4t\mu^*)]\}. \quad (23)$$

$K_1(t)$ and $K_2(t)$ are transitory terms that disappear rapidly as t increases.

Both the mean and variance of replication cycle counts increase in an approximately linear way with time (eqq. [17]–[23]; fig. 1). A nonlinear increase in $M_A(t)$ and $M_U(t)$ occurs immediately following initial infection, but both functions rapidly converge to parallel lines with $M_A(t) - M_U(t) = \mu^*/(\mu' + \mu^*)$. The increase in $V_A(t)$ and $V_U(t)$ is also approximately linear with t. In the example of figure 1*B*, $V_U(t)$ initially exceeds $V_A(t)$ but then rapidly settles to a lower value. After this transitory period, $V_A(t)$ and $V_U(t)$ are parallel lines with $V_A(t) - V_U(t) = \mu^*\mu'(\mu' - \mu^*)/(\mu' + \mu^*)^3$. Thus, $V_A(t) > V_U(t)$ if $\mu' > \mu^*$ and $V_A(t) < V_U(t)$ if $\mu' < \mu^*$.

The linearity of these relationships implies that we can assess the effects of the parameters through their effects on the slopes. The slope for $M_A(t)$ and $M_U(t)$ increases with both μ' and μ^*, but its value is less than the smaller value. The slope for $V_A(t)$ and $V_U(t)$ is smaller than the slope for $M_A(t)$ and $M_U(t)$. The ratio of variance to mean, known as the coefficient of dispersion, is illustrated for virion replication cycle statistics in figure 2. Note that the coefficient of dispersion approaches 1 if $\mu' \gg \mu^*$ or if $\mu^* \gg \mu'$. Under this condition, it is possible to predict the full distribution of replication cycles and not just the first two moments of this distribution (eqq. [17]–[23]). Consider the case in which $\mu' \gg \mu^*$. Virion dynamics are essentially instantaneous: virions either infect new cells or are cleared by the immune system immediately after they are released. Under this condition, proviral DNA accounts for essentially all viral genomes at any one point in time ($f_U \to 0$), and the distribution of replication cycles among provirus is Poisson:

$$A_{i+1}(t) = \frac{e^{-\mu^* t}(\mu^* t)^i}{i!}. \quad (24)$$

The asymptotic linearity of the mean and variance with t also suggests that a simpler analysis might suffice to derive

the slopes associated with $M_A(t)$ and $V_A(t)$. It is useful to take a retrospective approach and consider the ancestral lineage of a virion sampled at time t. This lineage is a series of intervals with the virus persisting as either provirus or virion. The duration of each interval is a random variable. Under the assumptions of our model, the amount of time spent in a state (provirus or virion) before jumping to the alternate state is exponentially distributed. The mean time spent as virion is $1/\mu'$ (actually $1/c$), and the mean as provirus is $1/\mu^*$.

Viewed in this way, lineage history is a simple Markovian stochastic process. In the language of stochastic processes, the recurrence time is the amount of time between successive passages into the same state and is essentially the length of a viral generation. Let τ denote the mean recurrence time and ω denote the variance in recurrence time. For this model, $\tau = 1/\mu' + 1/\mu^*$ and $\omega = (1/\mu')^2 + (1/\mu^*)^2$. The mean and variance in replication cycle number, $M_U(t)$ and $V_U(t)$, are derived by noting that replication cycles can be equated to the number of passages through the virion state in a viral lineage. Asymptotically, the expected number of passages is t/τ, and the variance in number of passages is $t\omega/\tau^3$ (Feller 1968, pp. 320–321). Direct substitution of τ and ω into these terms yields the slopes for the mean and variance (eqq. [18]–[22]).

This retrospective approach is not necessary for the single cell model. The forward equations are explicit, tractable, and yield the slope terms directly. They also provide information about the transitory period, which the retrospective analysis does not. However, the retrospective approach is very useful in generalizations of the model. It allows us to extract analytical results as the forward equations become increasingly unmanageable. This is illustrated below for the multiple cell type model, but the retrospective approach will also be used in subsequent papers.

Infection of Multiple Cell Types

It is straightforward to generalize the dynamical model (eqq. [1]–[3]) to allow infection of multiple cell types. We allow the parameters ν, β, λ, μ, and μ^* to vary among target cells. The number of cell types is arbitrary, and a subscript is added to each variable to distinguish types. Thus,

$$\frac{dn_i}{dt} = \lambda_i - \mu_i n_i - \beta_i n_i q, \tag{25}$$

$$\frac{dn_i^*}{dt} = \beta_i n_i q - \mu_i^* n_i^*. \tag{26}$$

The differential equation for virion abundance needs to include contributions from, and losses to, each class of infected cells:

$$\frac{dq}{dt} = \left(\sum_i \mu_i^* \nu_i n_i^*\right) - \mu' q - \left(\sum_i \beta_i n_i q\right). \tag{27}$$

Like the single cell type model (eqq. [1]–[3]), the infection will fail if $R_0 < 1$, where

$$R_0 = \frac{1}{\mu'} \sum_i \frac{\beta_i \lambda_i (\nu_i - 1)}{\mu_i}. \tag{28}$$

The leading term in this sum is the R_0 for the single cell type model (eq. [4]). Since all of the terms are positive, we can conclude that the ability of the virus to infect multiple cell types increases the likelihood of maintaining a persistent infection (all else being equal).

If $R_0 > 1$, the dynamical variables approach equilibrium values. The equilibrium solutions for q, n_i, and n_i^* become increasingly cumbersome as the number of cell types increases. The solution for two cell types is a complicated quadratic (not shown but available on request), and we have been unable to extract closed form solutions for more than three cell types. An important simplification is possible if $\nu_i \gg 1$ and $\beta_i q \gg \mu_i$ for each cell type. Under these conditions, most cells are infected before natural death, and the system approaches the equilibrium

$$n_i^* \approx \frac{\lambda_i}{\mu_i^*}, \tag{29a}$$

$$n_i \approx \frac{\lambda_i \mu'}{\beta_i \sum_j \nu_j \lambda_j}, \tag{29b}$$

$$q \approx \frac{\sum_j \nu_j \lambda_j}{\mu'}, \tag{29c}$$

where the summations are taken over all cell types. The approximate equilibria are surprisingly close to the exact values under a wide range of parameter values that allow persistent infection. This may reflect the fact that a high rate of cellular infection is typical of this class of dynamical models when $R_0 \gg 1$ (Nowak and May 2000). For simplicity, we use this approximation for the development of the population genetic theory (eqq. [30]–[32]). However, it is straightforward to derive predictions using the exact dynamical equilibria in cases where the approximation of equations (29) is not accurate.

A simple way to distill this equilibrium is to consider the effects of parameter variation on the total number of viral genomes, $q + \sum n_i^*$. Equations (29) indicate that the number of genomes increases linearly with the supply rates

of uninfected cells (λ_i). The number of virions increases linearly with the burst sizes (ν_i) of infected cells. Numbers for both virions and provirus in each cell type are inversely related to their respective death rates (μ' and μ_i^*). The primary effect of the transmission parameters (β_i) is on the number of uninfected cells. As the value of β_i increases for a particular cell type, uninfected cells become infected more rapidly. Parameter differences between cell types can substantially change the equilibrium values for the severity of infection, as measured by the total number of virus genomes. In the multiple cell type model, the number of viral genomes depends primarily on the average value of parameters across cell types, but variability in cell death rates (either infected or uninfected) can increase the equilibrium number.

What is the effect of allowing infection of multiple cell types on the evolution of the virus? More specifically, how does cellular population structure affect the mean and variance in replication cycles per viral genome through time? With regard to this question, the dynamical model provides critical information concerning the proportion of viral genomes residing in each cell type and the proportion of virions produced by each cell type. Let ε_i denote the latter proportion

$$\varepsilon_i = \frac{\mu_i^* n_i^* \nu_i}{\sum_j \mu_j^* n_j^* \nu_j} = \frac{\lambda_i \nu_i}{\sum_j \lambda_j \nu_j}, \tag{30}$$

where the rightmost term relies on the approximation of equations (29).

It is straightforward to develop differential equations that are structurally similar to those from the single cell type model (see appendix). The analysis of these equations indicates that, after a transitory period, the mean and variance of the replication cycle number increase as a linear function of *t*. Virions and provirus from all cell types share the same slope, although the slope value associated with the mean will generally differ from that of the variance. These values can be derived most easily by using the retrospective approach. Figure 3 depicts the Markov process associated with the multiple cell type model. The average recurrence time for a virion is

$$\tau = \frac{1}{\mu'} + \sum_i \varepsilon_i \frac{1}{\mu_i^*}, \tag{31}$$

and the variance in recurrence time is

$$\omega = \left(\frac{1}{\mu'}\right)^2 + 2\sum_i \varepsilon_i \left(\frac{1}{\mu_i^*}\right)^2 - \left[\sum_i \varepsilon_i \left(\frac{1}{\mu_i^*}\right)\right]^2. \tag{32}$$

Numerical evaluation of the forward equations (see appendix) confirms that the slope associated with the mean is equal to $1/\tau$ and that the variance slope is equal to ω/τ^3 (with τ and ω calculated from eqq. [29]–[32]).

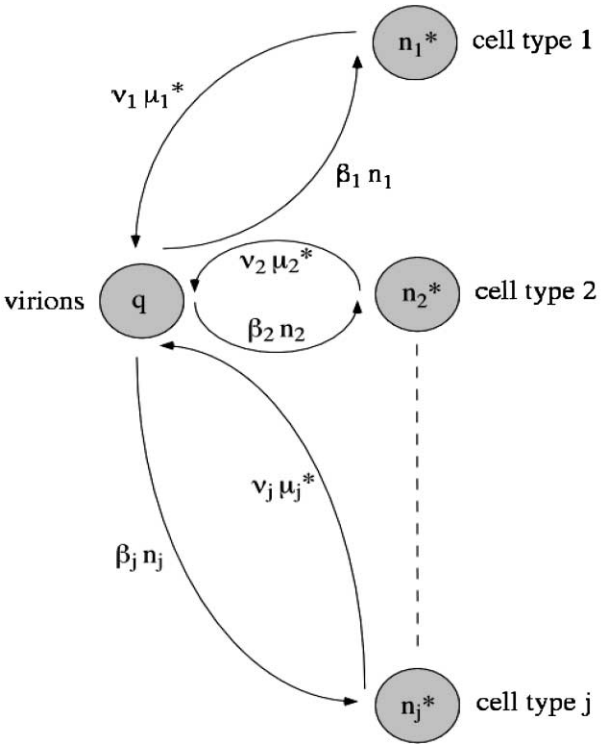

Figure 3: Markov loop diagram for the multiple cell type model. The values beside the arrows give the transition probabilities. The values within the shaded circles are the densities within each category.

How does CPS effect the distribution of replication cycles? Equations (31) and (32) indicate that the key variables are the death rates of infected cells (μ_i^*) and the relative contributions of each cell type to the virion pool (ε_i). Figure 4 illustrates the special case of two cell types. We consider the effect of variation in μ_i^* among cell types by allowing the difference between μ_1^* and μ_2^* to increase while the average is held constant at 0.05. With $\varepsilon_1 = 0.5$, the slope determining the average accumulation of replication cycles declines as $\mu_1^* - \mu_2^*$ increases (fig. 4*A*). With $\varepsilon_1 = 0.9$, this slope is actually highest for intermediate values of $\mu_1^* - \mu_2^*$, but its value declines precipitously as the difference becomes large (fig. 4*B*). In contrast to the single cell type model, the variance slope is often greater than the mean slope. The coefficient of dispersion is greatest when there is a substantial difference in contributions of the two cell types to the virion pool ($\varepsilon_1 > \varepsilon_2$ in fig. 4*B*).

Discussion

We present a theoretical framework composed of two components. The dynamic component of the theory predicts

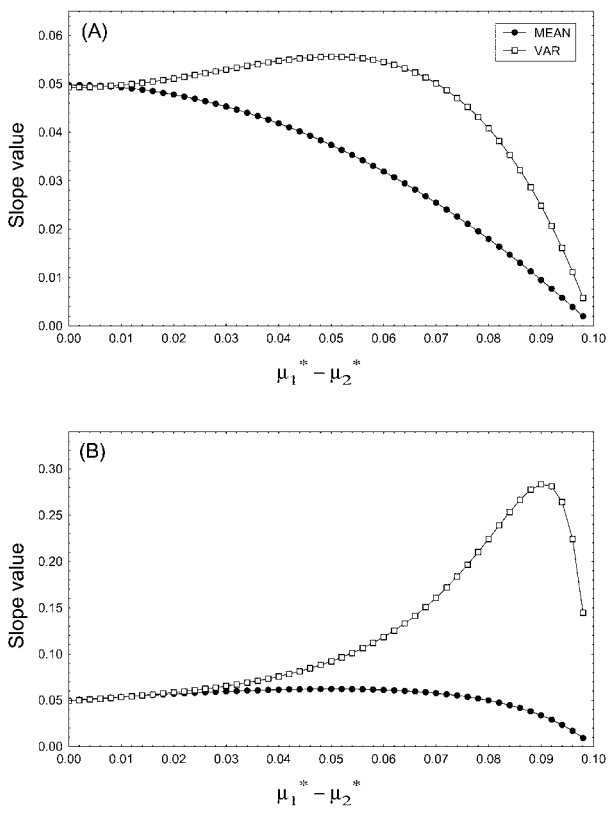

Figure 4: Linear slope associated with the mean (*circles*) and variance (*squares*) of replication cycle number as a function of the difference in death rates of infected cells ($\mu_1^* - \mu_2^*$). The average death rate, $(\mu_1^* + \mu_2^*)/2$, is held constant at 0.05. The value of μ' is 10 in both parts of the figure, while $\varepsilon_1 = 0.5$ in *A* and $\varepsilon_1 = 0.9$ in *B*. Note the difference in scale between *A* and *B*.

the numbers of infected and uninfected cells of various types and the abundance of virions (free viral particles). Equilibria derived from the dynamical models are essential inputs to the subsequent population genetic component of the theory. The latter predicts the distribution of ancestral replication cycles within the viral population as a function of time and the dynamical model parameters. This linkage between dynamical and population genetic models provides a natural conceptual bridge between virological and gene sequence studies. From an empirical point of view, the value of such a bridge is that it may suggest ways to bring genetic data to bear on virological hypotheses and vice versa.

The dynamical variables approach equilibrium values rapidly in both our single and multiple cell type models. An important difference between the models is that allowing infection of multiple cell types increases the likelihood that the viral population will persist. Consider a virus whose primary target is cell type 1 with the following parameter values: $\nu = 900$, $\mu = 0.1$, $\mu^* = 0.5$, $\beta = 10^{-4}$, $\lambda = 10$, and $\mu' = 10$. If type 1 is the only target cell, the viral population will not persist because R_0 is less than 1 (eq. [4]). However, the viral population will persist if it can also infect a second cell type (type 2) with the following parameter values: $\nu = 100$, $\mu = 0.01$, $\mu^* = 0.05$, $\beta = 10^{-4}$, and $\lambda = 2$. Cell type 2 is essential despite the fact that it is only responsible for approximately 2% of virion production. This example illustrates the subtle but potentially critical role of latently infected cells in the persistence of infection.

Latently infected cells may also provide the viral population with a refuge from attack, either from natural immune response or from medical intervention (Zack et al. 1990; Chun et al. 1995; Wolinsky and Learn 1999). Antiviral drugs can rapidly reduce HIV plasma virus load by orders of magnitude (Ho et al. 1995; Wei et al. 1995). Initial calculations suggested such treatments could eradicate HIV-1 from a patient's system within a few years. However, it is now clear that the virus can "hide" in latently infected memory T cells that can persist for many years (Chun et al. 1995, 1997). These cells, programmed to recognize rarely encountered antigens, become activated and permissive to virus replication when they eventually come into contact with their specific antigen (Blankson et al. 2002).

Evolutionary Predictions

Mutation is the ultimate source of genetic variation, and replication cycles provide the input of mutations into the viral population. We thus expect that dynamical parameters that influence the evolution of the distribution of replication cycles will consequently influence the rate and pattern of gene sequence evolution. To derive numerical predictions for gene sequence evolution, it is necessary to combine information regarding the distribution of replication cycles with a substitution model (see Muse 1999). However, it is possible to extract some qualitative predictions without imposing a particular substitution model. This allows a preliminary comparison of our models with the kind of data that virologists actually collect.

Shankarappa et al. (1999) describe one of the most intensive studies of viral gene sequence evolution within HIV-infected patients. They tracked viral evolution within nine HIV-positive men over a 6–12-yr period starting at the time of seroconversion, which typically occurs within a few months of initial infection. Minimal gene sequence variation was observed within the envelope gene in the seroconversion samples. This allowed Shankarappa et al. (1999) to define a progenitor sequence for each patient and determine the extent of divergence from this progenitor (number of substitutions) for viruses sampled later

during each infection. Figure 5*A* describes the mean genetic divergence within three of these patients, where each point denotes a sample of gene sequences. The coefficient of dispersion, the variance in divergence divided by the mean, was also calculated for each sample (fig. 5*B*). However, because the number of sequences per sample was limited (about 12 on average), the variance estimates are subject to large statistical uncertainty.

There is a roughly linear increase in divergence with time in these patients (fig. 5*A*), at least over the asymptomatic period of infection (see fig. 3*C* in Shankarappa et al. 1999) for the comparable graph for all nine patients). This pattern has been documented in numerous longitudinal studies (e.g., Balfe et al. 1990; Zhang et al. 1992) and is consistent with the theory in that both the single and multiple cell type models predict that the mean number of replication cycles should increase in an approximately linear way with time. However, there is some indication that the rate of gene sequence divergence actually declines in the terminal stages of infection as AIDS develops (Wolfs et al. 1991; Shankarappa et al. 1999). In population genetic terms, changes in rate of divergence can be due to either a change in the average number of replication cycles per unit time (and hence the number of mutations introduced per unit time) or a change in the probability that mutations will become fixed in the population (or at least reach high frequencies). A rigorous comparison of synonymous and nonsynonymous substitution rates over the course of infection may allow us to distinguish these alternatives.

In the single cell type model, the rate at which replication cycles accumulate increases with both μ' and μ^* and is limited primarily by the smaller value (eqq. [17], [18]). In the multiple cell type model, the rate depends on a weighted harmonic average of the death rates of infected cells (eq. [31]), with the weightings determined by the relative contributions of each cell type to the virion pool (ε_i). This kind of averaging implies that the inclusion of multiple cell types can actually decrease the rate of evolution. This can be illustrated by considering the case of two cell types (e.g., fig. 4). For HIV infection, type 1 might denote activated T cells, while type 2 might denote memory T cells. Inclusion of type 2 cells reduces the mean numbers of replication cycles because the majority of viral genomes will have at least one sojourn through cell type 2 in their ancestry. Gene sequence data from HIV does provide some support for the idea that latent infection does reduce the overall rate of gene sequence evolution (Kelly 1996).

The predicted reduction in evolutionary rate may be counteracted, to some extent, if the infection of multiple cell types opens new avenues for adaptive evolution of the viral population. There is clear evidence of genetic differentiation between viral populations inhabiting different cell types within a single host in both SIV and HIV (Smith et al. 1993; Poss et al. 1998). Gene sequence evolution will be accelerated, at least at specific sites, if novel mutations are required for the virus to infect different cell types (e.g., Hwang et al. 1991; Cann et al. 1992). Human immunodeficiency virus infections are typically initiated by macrophage-tropic genotypes that use the CCR5 molecule as a coreceptor on the cell surface (Zhu et al. 1993; Wolinsky et al. 1996; Veazey et al. 2000; Meng et al. 2002). However, T cell tropic genotypes that use a different molecule as coreceptor (CXCR4), or dual tropic genotypes that can use either CCR5 or CXCR4, often predominate later in infection. Changes in the frequencies of these variants may have important implications for disease progression (Tersmette et al. 1988; Wolinsky and Learn 1999). More generally, this example illustrates that the relationship between viral evolution and CPS is likely to be both

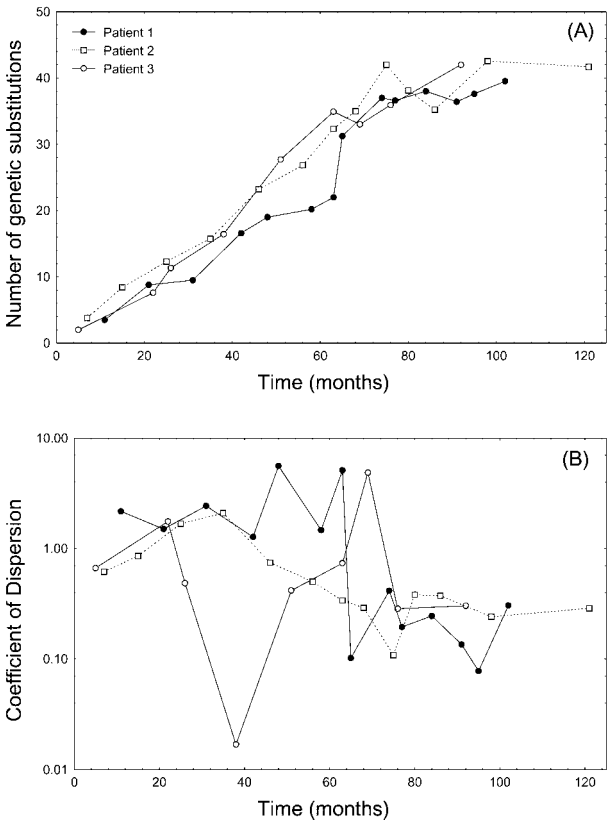

Figure 5: Mean divergence (*A*) and coefficient of dispersion (*B*) for each sample of sequences are given as a function of the amount of time since seroconversion for three patients. The statistics were calculated from the original sequence data downloaded from GenBank (accession numbers AF137629–AF138163, AF138166–AF138263, AF138305–AF138703; Shankarappa et al. 1999). Note that a logarithmic scale is used in *B*.

complicated and reciprocal. Cellular population structure should directly impact the rate and pattern of genetic changes within the viral population, but as these changes occur, the nature of the interaction between virus and host cells will also change.

All else equal, the variance among sequences in divergence (e.g., fig. 5*B*) will increase with the variance in replication cycle number. In our single cell type model, the variance is generally smaller than the mean, whereas with multiple cell types, the variance is typically greater than the mean (cf. figs. 2, 4). Higher values of $V_U(t)$ might serve to increase genetic variability because the longer-lived cell type will act to preserve relatively ancient sequences in much the same way that a seed bank preserves variation in a plant population (Kelly 1996; Blankson et al. 2002).

The preceding argument is subject to the caveat that the relationship between genetic variability and variance statistics such as $V_U(t)$ may be rather complicated. There is more to gene sequence evolution than the mean and variance of divergence. For example, the number of nucleotide differences between two viruses is more directly related to the number of replication cycles in the ancestry of each sequence since their most recent common ancestor than to $V_U(t)$. Genealogical and coalescent approaches have proven a useful tool for investigating ancestral relations and predicting patterns of extant variation (Rodrigo and Felsenstein 1999; Zanotto et al. 1999). Here, we employed a retrospective approach to derive the slopes associated with $M_U(t)$ and $V_U(t)$, but the same methods can be extended to determine patterns of common ancestry. Important variables, such as the viral population sizes, can be deduced from the dynamical portion of the model.

Generalizations

The dynamical models described by equations (1)–(3) and (25)–(27) are highly abstracted from a virological point of view. By design, we have neglected a number of complications that are undoubtedly important to the interaction between host and pathogen. The development of simple models is an essential precursor to more complicated and realistic models. In this article, we used simple models to derive the basic machinery for integrating dynamical and population genetic theories. The predictions derived from these simple models also provide an important baseline for comparison. We can determine the dynamical and evolutionary effects of a particular complicating factor only by comparison to a simpler model that lacks this complication.

One important complication is tissue level compartmentalization of viral population. Our multiple cell type model considers only one sort of structure. Cells of varying types will exist within a variety of tissues and tissues within numerous interconnected systems. As a consequence, the viral population will be structured at a hierarchy of levels, and this may impact both viral dynamics and evolution (Kepler and Perelson 1998; Kirschner et al. 1998; Holt 2000). Genetic differentiation between HIV populations inhabiting different organs within the same infected patient has been documented (Epstein et al. 1991), and this kind of tissue-level compartmentalization may also be relevant to the transmission of the pathogen among host organisms (Poss et al. 1998; Wolinsky and Learn 1999, pp. 294–295).

A second complication, particularly important to HIV and other viruses that attack the immune system, concerns the dynamics of individual cell types. Here, we consider the simplest possible situation: cells of each type are recruited at a constant rate and removed by either natural death or viral lysis. In reality, recruitment is likely to be density dependent for most cell types. In other words, the number of new cells produced in a given time interval will depend on the number of cells (infected and uninfected of various types) currently present. Recruitment may also depend on other factors, such as the degree of thymus dysfunction and immune hyperactivation (Carcelain et al. 2001; Douek et al. 2001). Moreover, the cells of the immune system may change from one type to another. For example, T cells may transform from a resting to an activated state and vice versa. Each state would be classified as a different cell type in this theory because they are likely to have very different dynamical properties.

Our focus in this article has been within-host viral evolution and population dynamics. These are certainly important systems, and they allow detailed study of the interplay of ecological and evolutionary dynamics (e.g., Fenner and Ratcliffe 1965). However, our basic approach of linking explicit population dynamics with the distribution of replication events may prove broadly applicable. Consider the example of leaf-mining insects that feed on oak trees. An individual oak of the species *Quercus geminata* can harbor an entire population of the herbivore *Stilbosis quadricustatella.* Like the viral pathogens considered here, the herbivore population undergoes many generations over the lifespan of an infested oak (the host). Over the course of these generations, *S. quadricustatella* genetically adapts to the characteristics of the host plant (Mopper et al. 2000) in a process analogous to intrahost viral evolution. While details of the model would certainly change when considering leaf miners (or other systems), we suggest that the approach and its associated analytical tools may prove useful for a broad range of applications.

Acknowledgments

We would like to thank M. Barfield, A. Rodrigo, and one anonymous reader for careful review of the manuscript. The work was supported by National Institutes of Health grant 1 R01 GM60792-01A1, the Murphy Scholarship fund (University of Kansas), and the University of Florida Foundation.

APPENDIX

For the multiple cell type model, we require two subscripts. Let j denote replication cycle and i denote cell type. The multiple cell extension of equation (9) is

$$U_0(t) + \sum_{j=1}^{\infty}\left[U_j(t) + \sum_i A_{ij}(t)\right] = 1, \tag{A1}$$

where $A_{ij}(t)$ is the fraction of viral genomes that are provirus in cell type i that have j replication cycles in their ancestry. We assume the dynamical variables are at the approximate equilibrium of equations (29) and that $\nu_i \gg 1$. The analogs of equations (12) and (13) for the multiple cell type model are

$$\frac{\partial U_j(t)}{\partial t} = -\mu' U_j(t) + \sum_i \mu_i^* \nu_i A_{ij}(t), \tag{A2}$$

$$\frac{\partial A_{ij}(t)}{\partial t} = \left(\frac{\lambda_i \mu'}{\sum_k \nu_k \lambda_k}\right) U_{j-1}(t) - \mu_i^* A_{ij}(t). \tag{A3}$$

The differential equations for mean replication cycle number are

$$\frac{\partial M_{\mathrm{U}}(t)}{\partial t} = -\mu' M_{\mathrm{U}}(t) + \frac{1}{f_{\mathrm{U}}}\sum_i \mu_i^* \nu_i f_{\mathrm{A}i} M_{\mathrm{A}i}(t), \tag{A4}$$

$$\frac{\partial M_{\mathrm{A}i}(t)}{\partial t} = -\mu_i^* M_{\mathrm{A}i}(t) + \frac{f_{\mathrm{U}}}{f_{\mathrm{A}i}}\left(\frac{\lambda_i \mu'}{\sum_k \nu_k \lambda_k}\right)[M_{\mathrm{U}}(t) + 1], \tag{A5}$$

where

$$f_{\mathrm{U}} = \frac{q}{q + \sum_i n_i^*},$$

$$f_{\mathrm{A}i} = \frac{n_i^*}{q + \sum_k n_k^*}, \tag{A6}$$

with q and n_i^* given by equations (29). This is a linear system that can be analyzed by standard methods.

The variance of replication cycle numbers can be derived from the equations for second noncentral moments: $V_{\mathrm{U}}(t) = M_{2\mathrm{U}}(t) - [M_{\mathrm{U}}(t)]^2$ and $V_{\mathrm{A}i}(t) = M_{2\mathrm{A}i}(t) - [M_{\mathrm{A}i}(t)]^2$. The new terms are defined as

$$M_{2\mathrm{U}}(t) = \frac{1}{f_{\mathrm{U}}}\sum_{j=0}^{\infty} j^2 U_j(t), \tag{A7}$$

$$M_{2\mathrm{A}i}(t) = \frac{1}{f_{\mathrm{A}i}}\sum_{j=0}^{\infty} j^2 A_{ij}(t). \tag{A8}$$

The differential equations for these quantities are

$$\frac{\partial M_{2\mathrm{U}}(t)}{\partial t} = -\mu' M_{2\mathrm{U}}(t) + \frac{1}{f_{\mathrm{U}}}\sum_{i=1}^{\infty} f_{\mathrm{A}i}\mu_i^* \nu_i M_{2\mathrm{A}i}(t), \tag{A9}$$

$$\frac{\partial M_{2\mathrm{A}i}(t)}{\partial t} = -\mu_i^* M_{2\mathrm{A}i}(t) + \frac{f_{\mathrm{U}}}{f_{\mathrm{A}i}}\left(\frac{\lambda_i \mu'}{\sum_k \nu_k \lambda_k}\right) \times [M_{2\mathrm{U}}(t) + 2M_{\mathrm{U}}(t) + 1]. \tag{A10}$$

Numerical analysis of these equations was used to verify results from the retrospective analysis (eqq. [31], [32]).

Literature Cited

Balfe, P., P. Simmonds, C. A. Ludlam, J. O. Bishop, and A. J. L. Brown. 1990. Concurrent evolution of human immunodeficiency type 1 in patients infected from the same source: rate of sequence change and low frequency of inactivating mutations. Journal of Virology 64: 6221–6233.

Blankson, J. N., D. Persaud, and R. F. Siliciano. 2002. The challenge of viral reservoirs in HIV-1 infection. Annual Review of Medicine 53:557–593.

Burns, D. P., and R. C. Desrosiers. 1991. Selection of genetic variants of simian immunodeficiency virus in persistently infected rhesus monkeys. Journal of Virology 65:1843–1854.

Cann, A. J., M. J. Churcher, M. Boyd, W. O'Brien, J. Q. Zhao, J. A. Zack, and I. S. Chen. 1992. The region of the envelope gene of human immunodeficiency virus type 1 responsible for cell tropism. Journal of Virology 66:305–309.

Carcelain, G., P. Debre, and B. Autran. 2001. Reconsti-

tution of CD4+ T lymphocytes in HIV-infected individuals following antiretroviral therapy. Current Opinion in Immunology 13:483–488.

Chun, T. W., D. Finzi, J. B. Margolick, K. Chadwick, D. Schwartz, and R. F. Silicano. 1995. In vivo fate of HIV-1 infected T cells: quantitative analysis of the transition to stable latency. Nature Medicine 1:1284–1290.

Chun, T. W., L. Stuyver, S. B. Mizell, L. A. Ehler, J. A. Mican, M. Baseler, A. L. Lloyd, M. A. Nowak, and A. S. Fauci. 1997. Presence of an inducible HIV-1 latent reservoir during highly active antiretrovial therapy. Proceedings of the National Academy of Sciences of the USA 94:13193–13197.

Crandall, K. 1996. Multiple interspecies transmissions of human and simian T-cell leukemia/lymphoma virus type I sequences. Molecular Biology and Evolution 13: 115–131.

———. 1999. The evolution of HIV. Johns Hopkins University Press, Baltimore.

Douek, D. C., M. Betts, B. Hill, S. Little, R. Lempicki, J. Metcalf, J. Casazza, et al. 2001. Evidence for increased T cell turnover and decreased thymic output in HIV infection. Journal of Immunology 167:6663–6668.

Drummond, A., R. Forsberg, and A. G. Rodrigo. 2001. The inference of stepwise changes in substitution rates using serial sequence samples. Molecular Biology and Evolution 18:1365–1371.

Epstein, L. G., C. Kuiken, M. M. Blumberg, S. Hartman, L. R. Sharer, M. Clement, and J. Goudsmit. 1991. HIV-1 V3 domain variation in brain and spleen of children with AIDS: tissue-specific evolution within host-determined quasispecies. Virology 180:583–590.

Feller, W. 1968. An introduction to probability theory and its applications. Vol. 1. Wiley, New York.

Fenner, F., and F. N. Ratcliffe. 1965. Myxomatosis. Cambridge University Press, London.

Finzi, D., M. Hermankova, T. Pierson, L. M. Carruth, C. Buck, R. E. Chaisson, T. C. Quinn, et al. 1997. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. Science (Washington, D.C.) 278:1295–1300.

Finzi, D., J. Blankson, J. D. Silicano, J. B. Margolick, K. Chadwick, T. Pierson, K. Smith, et al. 1999. Latent infection of CD4 T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. Nature Medicine 5:512–517.

Frost, S. D. W., H. F. Gunthard, J. K. Wong, D. Havlir, D. D. Richman, and A. J. Leigh Brown. 2001*a*. Evidence for positive selection driving the evolution of HIV-1 env under potent antiviral therapy. Virology 284:250–258.

Frost, S. D. W., M.-J. Dumaurier, S. Wain-Hobson, and A. J. Leigh Brown. 2001*b*. Genetic drift and within-host metapopulation dynamics of HIV-1 infection. Proceedings of the National Academy of Sciences of the USA 98:6975–6980.

Granelli-Piperno, A., E. Delgado, V. Finkel, W. Paxton, and R. M. Steinman. 1998. Immature dendritic cells selectively replicate macrophagetropic (M-tropic) human immunodeficiency virus type 1, while mature cells efficiently transmit both M- and T-tropic virus to T cells. Journal of Virology 72:2733–2737.

Hahn, B. H., G. M. Shaw, M. E. Taylor, R. R. Redfield, P. D. Markham, S. Z. Salahuddin, F. Wong-Stall, R. C. Gallo, E. S. Parks, and W. Parks. 1986. Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. Science (Washington, D.C.) 232: 1548–1553.

Herz, A. V., S. Bonhoeffer, R. M. Anderson, R. M. May, and M. A. Nowak. 1996. Viral dynamics in vivo: limitations on estimates of intracellular delay and virus decay. Proceedings of the National Academy of Sciences of the USA 93:7247–7251.

Ho, D. D., A. U. Neumann, A. S. Perelson, W. Chen, J. M. Leonard, and M. Markowitz. 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature 373:123–126.

Holt, R. D. 2000. A biogeographical and landscape perspective on within-host infection dynamics. Pages 583–588 *in* C. R. Bell, M. Brylinsky, and P. Johnson-Green, eds. Proceedings of the Eighth International Symposium of Microbial Ecology: Ecology of Infectious Diseases. Atlantic Canada Society for Microbial Ecology, Halifax.

Hwang, S. S., T. J. Boyle, H. K. Lyerly, and B. R. Cullen. 1991. Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. Science (Washington, D.C.) 253:71–74.

Kelly, J. K. 1994. An application of population genetic theory to synonymous gene sequence evolution in the human immunodeficiency virus (HIV). Genetical Research 64:1–9.

———. 1996. Replication rate and evolution in the human immunodeficiency virus. Journal of Theoretical Biology 180:359–364.

Kepler, T. B., and A. S. Perelson. 1998. Drug concentration heterogeneity facilitates the evolution of drug resistance. Proceedings of the National Academy of Sciences of the USA 95:11514–11519.

Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, New York.

Kirschner, D. 2001. Reconstructing microbial pathogenesis. ASM News 67:566–573.

Kirschner, D. E., R. Mehr, and A. S. Perelson. 1998. Role of the thymus in pediatric HIV-1 infection. Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology 18:95–109.

Leigh Brown, A. J. 1997. Analysis of HIV-1 env gene sequences reveals evidence for a low effective number in the viral population. Proceedings of the National Academy of Sciences of the USA 94:1862–1865.

Levin, B. R., M. Lipsitch, and S. Bonhoeffer. 1999. Population biology, evolution, and infectious disease: convergence and synthesis. Science (Washington, D.C.) 283: 806–809.

Levy, J. A. 1993. Pathogenesis of human immunodeficiency virus infection. Microbiological Reviews 57:183–289.

Lipsitch, M., and B. R. Levin. 1997. The population dynamics of antimicrobial chemotherapy. Antimicrobial Agents and Chemotherapy 41:363–373.

Mansky, L. M. 1996. Forward mutation rate of human immunodeficiency virus type 1 in a T-lymphoid cell line. AIDS Research Human Retroviruses 12:307–314.

Meng, G., X. Wei, X. Wu, M. T. Sellers, J. M. Decker, Z. Moldoveanu, J. M. Orenstein, et al. 2002. Primary intestinal epithelial cells selectively transfer R5 HIV-1 to CCR5+ cells. Nature Medicine 8:150–156.

Mopper, S., P. Stirling, K. Landau, D. Simberloff, and P. Van Zandt. 2000. Spatiotemporal variation in leafminer population structure and adaptation to individual oak trees. Ecology 81:1577–1587.

Moya, A., S. F. Elena, A. Bracho, R. Miralles, and E. Barrio. 2000. The evolution of RNA viruses: a population genetics view. Proceedings of the National Academy of Sciences of the USA 97:6967–6973.

Muse, M. V. 1999. Modelling the molecular evolution of HIV. Pages 122–152 *in* K. Crandall, ed. The evolution of HIV. Johns Hopkins University Press, Baltimore.

Nowak, M., and C. R. M. Bangham. 1996. Population dynamics of immune responses to persistent viruses. Science (Washington, D.C.) 272:74–79.

Nowak, M., and R. M. May. 2000. Virus dynamics: mathematical principles of immunology and virology. Oxford University Press, New York.

Nowak, M., R. M. Anderson, A. R. McLean, T. F. Wolfs, J. Goudsmit, and R. M. May. 1991. Antigenic diversity thresholds and the development of AIDS. Science (Washington, D.C.) 254:963–969.

Perelson, A. S. 1989. Modeling the interaction of the immune system with HIV. Pages 350–370 *in* C. Castillo-Chávez, ed. Mathematical and statistical approaches to AIDS epidemiology. Springer, Berlin.

———. 2002. Modelling viral and immune system dynamics. Nature Reviews 2:28–36.

Perelson, A. S., D. E. Kirschner, and R. De-Boer. 1993. Dynamics of HIV infection of CD4+ T cells. Mathematical Biosciences 114:81–125.

Perelson, A. S., A. U. Neumann, M. Markowitz, J. M. Leonard, and D. D. Ho. 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science (Washington, D.C.) 271: 1582–1586.

Poss, M., A. G. Rodrigo, J. J. Gosink, G. H. Learn, D. de Vange Panteleeff, H. L. Martin, J. Bwayo, J. K. Kreiss, and J. Overbaugh. 1998. Evolution of envelope sequences from the genital tract and peripheral blood of women infected with clade A human immunodeficiency virus type 1. Journal of Virology 72:8240–8251.

Potash, M. J., and D. J. Volsky. 1998. Viral interference in HIV-1 infected cells. Reviews in Medical Virology 8: 203–211.

Preston, B. D., B. J. Poiesz, and L. A. Loeb. 1988. Fidelity of HIV reverse transcriptase. Science (Washington, D.C.) 242:1168–1171.

Roberts, J. D., K. Bebenek, and T. A. Kunkel. 1988. The accuracy of reverse transcriptase from HIV-1. Science (Washington, D.C.) 242:1171–1173.

Rodrigo, A. G., and J. Felsenstein. 1999. Coalescent approaches to HIV population genetics. Pages 233–272 *in* K. Crandall, ed. The evolution of HIV. Johns Hopkins University Press, Baltimore.

Rodrigo, A. G., E. G. Shpaer, E. L. Delwart, A. K. N. Iversen, M. V. Gallo, J. Brojatsch, M. S. Hirsch, B. D. Walker, and J. I. Mullins. 1999. Coalescent estimates of HIV-1 generation time in vivo. Proceedings of the National Academy of Sciences of the USA 96:2187–2191.

Shankarappa, R., J. B. Margolick, S. J. Gange, A. G. Rodrigo, D. Upchurch, H. Farzadegan, P. Gupta, et al. 1999. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. Journal of Virology 73:10489–10502.

Simmonds, P., L. Q. Zhang, F. McOmish, P. Balfe, C. A. Ludlam, and A. J. L. Brown. 1991. Discontinuous sequence change of human immunodeficiency virus (HIV) type 1 env sequences in plasma viral and lymphocyte-associate proviral populations in vivo: implications for models of HIV pathogenesis. Journal of Virology 65:6266–6276.

Smith, M. S., K. L. Koerber, and J. S. Pagano. 1993. Zidovudine-resistant human immunodeficiency virus type 1 genomes detected in plasma distinct from viral genomes in peripheral blood mononuclear cells. Journal of Infectious Diseases 167:445–448.

Tersmette, M., R. E. de Goede, B. J. Al, I. N. Winkel, R. A. Gruters, H. T. Cuypers, H. G. Huisman, and F. Miedema. 1988. Differential syncytium-inducing capacity of human immunodeficiency virus isolates: frequent detection of syncytium-inducing isolates in patients with acquired immunodeficiency syndrome (AIDS) and AIDS-related complex. Journal of Virology 62: 2026–2032.

Thompson, J. N., O. J. Reichman, P. J. Morin, G. A. Polis,

M. E. Power, R. W. Sterner, C. A. Couch, et al. 2001. Frontiers of ecology. BioScience 51:15–24.

Veazey, R. S., K. Mansfield, I. Tham, A. Carville, D. Shvetz, A. Forand, and A. Lackner. 2000. Dynamics of CCR5 expression by CD4(+) T cells in lymphoid tissues during simian immunodeficiency virus infection. Journal of Virology 74:11001–11007.

Wei, X., S. K. Ghosh, E. Taylor, V. A. Johnson, E. A. Emini, P. Deutsch, J. D. Lifson, et al. 1995. Viral dynamics in human immunodeficiency virus type 1 infection. Nature 373:117–122.

Wodarz, D., and M. A. Nowak. 1999. Specific therapy regimes could lead to long-term immunological control of HIV. Proceedings of the National Academy of Sciences of the USA 96:14464–14469.

Wolfs, T. F. W., G. Zwart, M. Bakker, M. Valk, C. L. Kuiken, et al. 1991. Naturally occurring mutations within HIV-1 V3 genomic RNA lead to antigenic variation dependent on a single amino acid substitution. Virology 185: 195–205.

Wolinsky, S. M., and G. H. Learn. 1999. Levels of diversity within and among host individuals. Pages 275–314 *in* K. Crandall, ed. The evolution of HIV. Johns Hopkins University Press, Baltimore.

Wolinsky, S. M., B. T. Korber, and A. U. Neumann. 1996. Adaptive evolution of human immunodeficiency virus–type 1 during the natural course of infection. Science (Washington, D.C.) 272:537–542.

Zack, J. A., S. J. Arrigo, S. R. Weitsman, A. S. Go, A. Haislip, and I. S. Chen. 1990. HIV-1 entry into quiescent primary lymphocytes: molecular analysis reveals a labile, latent viral structure. Cell 61:213–222.

Zanotto, P. M. D. A., E. G. Kallas, R. F. de Souza, and E. C. Holmes. 1999. Genealogical evidence for positive selection in the nef gene of HIV-1. Genetics 153: 1077–1089.

Zhang, L. Q., P. MacKenzie, A. Cleland, E. Holmes, A. Brown, and P. Simmonds. 1992. Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. Journal of Virology 67:3345–3356.

Zhu, T., H. Mo, N. Wang, D. Nam, Y. Cao, R. Koup, and D. Ho. 1993. Genotypic and phenotypic characterization of HIV-1 patients with primary infection. Science (Washington, D.C.) 261:1179–1181.

Associate Editor: Alan R. Templeton