Within-host viral infection dynamics and evolution

The surprising complexity of the classical world

Tim Vaughan, Alexei Drummond, Peter Drummond

September 13, 2011
Goals of study
Goals of study

Part 1
Goals of study

Part 1

- Study short-term within-host HIV infection dynamics
Goals of study

Part 1

- Study short-term within-host HIV infection dynamics
- Investigate effects of stochasticity on these dynamics
Goals of study

Part 1

- Study short-term within-host HIV infection dynamics
- Investigate effects of stochasticity on these dynamics
- Hunt for non-trivial correlation dynamics in realistically-sized systems
Goals of study

Part 1

- Study short-term within-host HIV infection dynamics
- Investigate effects of stochasticity on these dynamics
- Hunt for non-trivial correlation dynamics in realistically-sized systems

Part 2
Goals of study

Part 1

- Study short-term within-host HIV infection dynamics
- Investigate effects of stochasticity on these dynamics
- Hunt for non-trivial correlation dynamics in realistically-sized systems

Part 2

- Model effects of mutation due to different microscopic mechanisms
Goals of study

Part 1

- Study short-term within-host HIV infection dynamics
- Investigate effects of stochasticity on these dynamics
- Hunt for non-trivial correlation dynamics in realistically-sized systems

Part 2

- Model effects of mutation due to different microscopic mechanisms
- Examine correlations arising between genetically distinct viral subpopulations
Goals of study

Part 1

- Study short-term within-host HIV infection dynamics
- Investigate effects of stochasticity on these dynamics
- Hunt for non-trivial correlation dynamics in realistically-sized systems

Part 2

- Model effects of mutation due to different microscopic mechanisms
- Examine correlations arising between genetically distinct viral subpopulations
- Determine extent to which mutation mechanism affects these correlations
Part I

Demographic fluctuations
Primitive infection processes
Primitive infection processes

Progenitor \rightarrow \lambda \rightarrow \text{Infection target cell (X)}
Primitive infection processes

- Progenitor
- Virion (V)
- Infection target cell (X)
- Infected cell (Y)

\[ \lambda \] and \[ \beta \] represent the infection processes.
Primitive infection processes

Progenitor → \( \lambda \) → Infection target cell \((X)\)

Virion \((V)\) + Infected cell \((Y)\) → \( \beta \) → Infected cell \((Y)\)

Infected cell \((Y)\) + +
Primitive infection processes

\[ \lambda \]

\[ \beta \]

\[ k \]

Progenitor

Virion (V)

Infected target cell (X)

Infected cell (Y)

\[ d \]
Primitive infection processes

- Progenitor
- Virion (V)
- Infection target cell (X)
- Infected cell (Y)

Symbols:
- $\lambda$
- $\beta$
- $k$
- $d$
- $\alpha$
Primitive infection processes
### Values of rate constants

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda^*$</td>
<td>T cell production rate</td>
<td>$2.5 \times 10^8$/day</td>
</tr>
<tr>
<td>$\beta^*$</td>
<td>T cell infection rate</td>
<td>$5 \times 10^{-13}$/T cell/virion/day</td>
</tr>
<tr>
<td>$k$</td>
<td>virion production rate</td>
<td>$10^3$/infected T cell/day</td>
</tr>
<tr>
<td>$d$</td>
<td>T cell death rate</td>
<td>$10^{-3}$/T cell/day</td>
</tr>
<tr>
<td>$a$</td>
<td>infected T cell death rate</td>
<td>$1$/infected T cell/day</td>
</tr>
<tr>
<td>$u$</td>
<td>virion clearance rate</td>
<td>$3$/virion/day</td>
</tr>
</tbody>
</table>

*These parameter values obtained by aligning predictions of deterministic model with observed infection dynamics:*
## Values of rate constants

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda^*$</td>
<td>T cell production rate</td>
<td>$2.5 \times 10^8$/day</td>
</tr>
<tr>
<td>$\beta^*$</td>
<td>T cell infection rate</td>
<td>$5 \times 10^{-13}$/T cell/virion/day</td>
</tr>
<tr>
<td>$k$</td>
<td>virion production rate</td>
<td>$10^3$/infected T cell/day</td>
</tr>
<tr>
<td>$d$</td>
<td>T cell death rate</td>
<td>$10^{-3}$/T cell/day</td>
</tr>
<tr>
<td>$a$</td>
<td>infected T cell death rate</td>
<td>$1$/infected T cell/day</td>
</tr>
<tr>
<td>$u$</td>
<td>virion clearance rate</td>
<td>$3$/virion/day</td>
</tr>
</tbody>
</table>

*These parameter values obtained by aligning predictions of deterministic model with observed infection dynamics:

### Deterministic Model

\[
\begin{align*}
\dot{x} &= \lambda - \beta xv - dx \\
\dot{y} &= \beta xv - ay \\
\dot{v} &= ky - \beta xv - uv
\end{align*}
\]
Deterministic predictions

<table>
<thead>
<tr>
<th>t (days)</th>
<th>Total Cell/Virion Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>10</td>
</tr>
<tr>
<td>1.0</td>
<td>100</td>
</tr>
<tr>
<td>5.0</td>
<td>10^9</td>
</tr>
<tr>
<td>10.0</td>
<td>10^12</td>
</tr>
<tr>
<td>50.0</td>
<td>10^12</td>
</tr>
</tbody>
</table>

- **Uninfected T cell**
- **Infected T cell**
- **Viral load**
Why use a stochastic model?
Why use a stochastic model?

- Earliest models of infection dynamics employed Lotka-Volterra style ODEs.
Why use a stochastic model?

- Earliest models of infection dynamics employed Lotka-Volterra style ODEs.
- Used to demonstrate the importance of T cell depletion in fixing location and height of peak viral load.
Why use a stochastic model?

- Earliest models of infection dynamics employed Lotka-Volterra style ODEs.
- Used to demonstrate the importance of T cell depletion in fixing location and height of peak viral load.
- Implicitly assume continuous population sizes—often sensible in the large population limit.
Why use a stochastic model?

- Earliest models of infection dynamics employed Lotka-Volterra style ODEs.
- Used to demonstrate the importance of T cell depletion in fixing location and height of peak viral load.
- Implicitly assume continuous population sizes—often sensible in the large population limit.
- We need to relax this assumption to consider the effect of integer numbers of cells and virions.
Why use a stochastic model?

- Earliest models of infection dynamics employed Lotka-Volterra style ODEs.
- Used to demonstrate the importance of T cell depletion in fixing location and height of peak viral load.
- Implicitly assume continuous population sizes—often sensible in the large population limit.
- We need to relax this assumption to consider the effect of integer numbers of cells and virions.
- This leads naturally to the question of when microscopic interactions occur.
Why use a stochastic model?

- Earliest models of infection dynamics employed Lotka-Volterra style ODEs.
- Used to demonstrate the importance of T cell depletion in fixing location and height of peak viral load.
- Implicitly assume continuous population sizes—often sensible in the large population limit.
- We need to relax this assumption to consider the effect of integer numbers of cells and virions.
- This leads naturally to the question of *when* microscopic interactions occur.
- Assume reactions occur at *known rates* (dependent on population sizes) but at completely *unknown times*: i.e. Poisson stochastic processes.
A stochastic description

With these goals in mind, we assemble the

Chemical Master Equation (CME)

\[
\frac{\partial}{\partial t} P(N_x, N_y, N_v) = \lambda \left[ P(N_x - 1, N_y, N_v) - P(N_x, N_y, N_v) \right] \\
+ \beta \left[ (N_x - 1)(N_v - 1)P(N_x - 1, N_y + 1, N_v - 1) \\
- N_x N_y P(N_x, N_y, N_v) \right] \\
+ k N_y \left[ P(N_x, N_y + 1, N_v - 1) - P(N_x, N_y, N_v) \right] \\
+ d \left[ (N_x + 1)P(N_x + 1, N_y, N_v) - N_x P(N_x, N_y, N_v) \right] \\
+ a \left[ (N_y + 1)P(N_x, N_y + 1, N_v) - N_y P(N_x, N_y, N_v) \right] \\
+ u \left[ (N_v + 1)P(N_x, N_y, N_v + 1) - N_v P(N_x, N_y, N_v) \right]
\]
Traditional Monte Carlo approach

- Given Poissonian fluctuations in the variables $N_x$, $N_y$ and $N_v$, expect the ‘steady-state’ volume of occupied state space to be $\gtrsim 10^{15}$: too large for direct numerical integration of CME.
Traditional Monte Carlo approach

- Given Poissonian fluctuations in the variables $N_x$, $N_y$ and $N_v$, expect the ‘steady-state’ volume of occupied state space to be $\gtrsim 10^{15}$: too large for direct numerical integration of CME.
- Instead, use Monte Carlo approaches to draw samples from probability distribution over possible dynamical trajectories.
Given Poissonian fluctuations in the variables $N_x$, $N_y$ and $N_v$, expect the ‘steady-state’ volume of occupied state space to be $\gg 10^{15}$: too large for direct numerical integration of CME.

Instead, use Monte Carlo approaches to draw samples from probability distribution over possible dynamical trajectories.

Simplest and most robust approach is to directly simulate each Poisson process via Stochastic Simulation Algorithm (Gillespie, 1976):
Traditional Monte Carlo approach

- Given Poissonian fluctuations in the variables $N_x$, $N_y$ and $N_v$, expect the ‘steady-state’ volume of occupied state space to be $\gtrsim 10^{15}$: too large for direct numerical integration of CME.
- Instead, use Monte Carlo approaches to draw samples from probability distribution over possible dynamical trajectories.
- Simplest and most robust approach is to directly simulate each Poisson process via Stochastic Simulation Algorithm (Gillespie, 1976):

![Graph showing the simulation of $N_y$ over time $t$.]
SSA results for small system

\[
\text{Cov}_{\text{rel}}(N_i, N_j) \equiv \frac{\langle N_i N_j \rangle}{\langle N_i \rangle \langle N_j \rangle} - 1
\]
Computational burden of the SSA

The diagram shows the computational burden of the SSA, with the x-axis representing the average number of states, $\langle N_x \rangle$, and the y-axis representing the trajectory time, $t_{\text{traj}}$ (s). The graph plots a linear relationship between the two variables.
Stochastic simulation of realistically-sized model
Stochastic simulation of realistically-sized model

- Calculation of SSA results for small system took 23 hrs on 100-node cluster: use of this method to perform full-sized calculation clearly impractical.
Calculation of SSA results for small system took 23 hrs on 100-node cluster: use of this method to perform full-sized calculation clearly impractical.

Tried to employ the Poisson representation phase space approach:
Calculation of SSA results for small system took 23 hrs on 100-node cluster: use of this method to perform full-sized calculation clearly impractical.

Tried to employ the Poisson representation phase space approach:

Poisson Representation (Gardner, Chaturveidi, 1976)
Stochastic simulation of realistically-sized model

- Calculation of SSA results for small system took 23 hrs on 100-node cluster: use of this method to perform full-sized calculation clearly impractical.

- Tried to employ the Poisson representation phase space approach:

Poisson Representation (Gardner, Chaturveidi, 1976)

1. Expand $P(N_x, N_y, N_v)$ in terms of multivariate Poisson distributions of means distributed according to $f(\alpha_x, \alpha_y, \alpha_v)$. 
Stochastic simulation of realistically-sized model

- Calculation of SSA results for small system took 23 hrs on 100-node cluster: use of this method to perform full-sized calculation clearly impractical.
- Tried to employ the Poisson representation phase space approach:

Poisson Representation *(Gardner, Chaturveidi, 1976)*

1. Expand $P(N_x, N_y, N_v)$ in terms of multivariate Poisson distributions of means distributed according to $f(\alpha_x, \alpha_y, \alpha_v)$.
2. Equation of motion of positive definite distribution $f$ takes the form of a Fokker-Planck equation (FPE).
Stochastic simulation of realistically-sized model

- Calculation of SSA results for small system took 23 hrs on 100-node cluster: use of this method to perform full-sized calculation clearly impractical.
- Tried to employ the Poisson representation phase space approach:

### Poisson Representation (Gardner, Chaturveidi, 1976)

1. Expand $P(N_x, N_y, N_v)$ in terms of multivariate Poisson distributions of means distributed according to $f(\alpha_x, \alpha_y, \alpha_v)$.
2. Equation of motion of positive definite distribution $f$ takes the form of a Fokker-Planck equation (FPE).
3. Map the FPE to stochastic differential equations (SDEs) and solve numerically.
Difficulties with Poisson representation approach
Difficulties with Poisson representation approach

- Derivation of FPE involves an “integration by parts” step which assumes tails of \( f(\alpha_x, \alpha_y, \alpha_v) \) decay rapidly as \( |\vec{\alpha}| \to \infty \). Violation of this assumption can lead to large sampling errors and/or systematic errors in results.
Difficulties with Poisson representation approach

- Derivation of FPE involves an “integration by parts” step which assumes tails of \( f(\alpha_x, \alpha_y, \alpha_v) \) decay rapidly as \( |\vec{\alpha}| \to \infty \). Violation of this assumption can lead to large sampling errors and/or systematic errors in results.

- Can in principle address with stochastic gauges (Drummond, 2004), but there are an infinite number of different gauges and no a priori way of assessing performance of a particular gauge is currently known.
Why does Poisson rep. offer scaling improvement?
Why does Poisson rep. offer scaling improvement?

There are two primary reasons the Poisson Representation approach, when it can be made to work, ‘wins’ computationally against the SSA:
Why does Poisson rep. offer scaling improvement?

There are two primary reasons the Poisson Representation approach, when it can be made to work, ‘wins’ computationally against the SSA:

1. Expanding a probability distribution originating from a Poisson process using a Poissonian basis can result in a transformed distribution with tighter support.
Why does Poisson rep. offer scaling improvement?

There are two primary reasons the Poisson Representation approach, when it can be made to work, ‘wins’ computationally against the SSA:

1. Expanding a probability distribution originating from a Poisson process using a Poissonian basis can result in a transformed distribution with tighter support.

   $\Rightarrow$ Fewer trajectories needed to sample the distribution.
There are two primary reasons the Poisson Representation approach, when it can be made to work, ‘wins’ computationally against the SSA:

1. Expanding a probability distribution originating from a Poisson process using a Poissonian basis can result in a transformed distribution with tighter support.
   \[ \Rightarrow \] Fewer trajectories needed to sample the distribution.

2. Integration algorithms for continuous variable SDEs involve approximating the integral using a series of finite time steps, \textit{the size of which does not explicitly depend on the magnitude of the variables involved.}
Basic integration scheme for SDEs can be derived by solving a short-time approximation to the FPE and using this to generate an approximate form of the path integral.
Finite time step integration for discrete processes

- Basic integration scheme for SDEs can be derived by solving a short-time approximation to the FPE and using this to generate an approximate form of the path integral.

- The short-time approximate solution to the T cell birth component of the CME is:

\[ P(N'_x, t + \tau | N_x, t) = \sum_{m=0}^{\infty} \delta_{N'_x - N_x, m} e^{-\tau \lambda} \frac{(\tau \lambda)^m}{m!} \]
Finite time step integration for discrete processes

- Basic integration scheme for SDEs can be derived by solving a short-time approximation to the FPE and using this to generate an approximate form of the path integral.

- The short-time approximate solution to the T cell birth component of the CME is:
  \[
P(N'_x, t + \tau | N_x, t) = \sum_{m=0}^{\infty} \delta_{N'_x-N_x,m} e^{-\tau \lambda} \frac{(\tau \lambda)^m}{m!}
\]

- Finite time step approximation to stochastic trajectory for \( N_x \) can be generated by iterating:
  \[
  N_x(t + (q+1)\tau) = N_x(t + q\tau) + m_q
  \]
  with each \( m_q \) chosen from a Poisson distrib. with mean \( \tau \lambda \).
  (Gillespie, 2001)
Finite time step integration for discrete processes
\(\tau\)-leaping results for full-sized model

\[
\text{Expected Population Sizes}
\]

\[
\begin{array}{c|c|c|c|c|c}
\text{Time (days)} & 0.1 & 0.5 & 5.0 & 50.0 \\
\hline
\text{Uninfected T cells} & 1 \times 10^2 & 1 \times 10^8 & 1 \times 10^{11} & 1 \times 10^{14} \\
\text{Infected T cells} & 1 \times 10^5 & 1 \times 10^{11} & 1 \times 10^{14} & 1 \times 10^{17} \\
\text{Virions} & 1 \times 10^3 & 1 \times 10^6 & 1 \times 10^9 & 1 \times 10^{12} \\
\end{array}
\]

\[
\text{Relative Variance}
\]

\[
\text{Relative Covariance}
\]

\[
\text{Cov}_{\text{rel}}(N_i, N_j) \equiv \frac{\langle N_i N_j \rangle}{\langle N_i \rangle \langle N_j \rangle} - 1
\]
Impact of initial population size

Expected Viral Load

\[ N_v(0) = 1000 \]
\[ N_v(0) = 100 \]
\[ N_v(0) = 10 \]
\[ N_v(0) = 1 \]
Impact of initial population size

![Graph showing impact of initial population size on clearance probability over time. The x-axis represents time in days, ranging from 0.0 to 3.0, and the y-axis represents clearance probability ranging from 0.0 to 1.0. The graph includes multiple curves, each representing different initial population sizes.](image-url)
Impact of initial population size

![Graph showing expected viral load and covrel over time for different initial population sizes.](image)

- $N_v(0) = 1000$
- $N_v(0) = 100$
- $N_v(0) = 10$
- $N_v(0) = 1$
Part II

Genetic correlations
Mutation in HIV replication

There are two primary generators of within-host genetic diversity:
There are two primary generators of within-host genetic diversity:

1. recombination (two viruses infect a single cell), and
There are two primary generators of within-host genetic diversity:

1. recombination (two viruses infect a single cell), and
2. mutation of single characters.
There are two primary generators of within-host genetic diversity:

1. recombination (two viruses infect a single cell), and
2. mutation of single characters.

Mutation occurs during:
Mutation in HIV replication

There are two primary generators of within-host genetic diversity:
1. recombination (two viruses infect a single cell), and
2. mutation of single characters.

Mutation occurs during:

Reverse Transcription
There are two primary generators of within-host genetic diversity:

1. recombination (two viruses infect a single cell), and
2. mutation of single characters.

Mutation occurs during:

- **Reverse Transcription**
- **Transcription**
The evolutionary dynamics of a mutating virus genome of length $L$ takes place in an $L$-dimensional sequence space containing $4^L$ possible genetic sequences.
The evolutionary dynamics of a mutating virus genome of length $L$ takes place in an $L$-dimensional sequence space containing $4^L$ possible genetic sequences.

- For HIV, $L \approx 10^4$—sequence space impossibly large.
The evolutionary dynamics of a mutating virus genome of length $L$ takes place in an $L$-dimensional sequence space containing $4^L$ possible genetic sequences.

- For HIV, $L \approx 10^4$—sequence space impossibly large.
- Sub-sequence of length $L = 105$ corresponds to protein sequence which does not directly affect viral fecundity. $4^{105} \sim 10^{63}$
The evolutionary dynamics of a mutating virus genome of length \( L \) takes place in an \( L \)-dimensional sequence space containing \( 4^L \) possible genetic sequences.

- For HIV, \( L \approx 10^4 \)—sequence space impossibly large.
- Sub-sequence of length \( L = 105 \) corresponds to protein sequence which does not directly affect viral fecundity. \( 4^{105} \approx 10^{63} \)
- Mutation rate \( \mu \approx 2 \times 10^{-5} / \text{character/replication} \): small, but viral population reaches \( \sim 10^{14} \) at its peak.
The evolutionary dynamics of a mutating virus genome of length $L$ takes place in an $L$-dimensional sequence space containing $4^L$ possible genetic sequences.

- For HIV, $L \approx 10^4$—sequence space impossibly large.
- Sub-sequence of length $L = 10^5$ corresponds to protein sequence which does not directly affect viral fecundity. $4^{10^5} \approx 10^{63}$
- Mutation rate $\mu \approx 2 \times 10^{-5}$/character/replication: small, but viral population reaches $\sim 10^{14}$ at its peak.  
  \[ \Rightarrow \text{stochastic simulation (even } \tau\text{-leaping) without further simplification is unwieldy at best.} \]
Sequence space projection

sequence space

founder strain

mutation count

n

mutation count
Sequence space projection

Partition seq. space into ‘hyperspheres’.

Sequence space

founder strain

mutation count

n

mutation count
Sequence space projection

- Partition seq. space into ‘hyperspheres’.
- Master equation for marginal probability distribution exists.
Partition seq. space into ‘hyperspheres’.

Master equation for marginal probability distribution exists.

Can obtain exact results using only $L + 1$ effective sites instead of $4^L$. 

Sequence space projection
Summary
Populations interacting within hosts during viral infection are very large—viral population approaching $10^{14}$ particles.
Populations interacting within hosts during viral infection are very large—viral population approaching $10^{14}$ particles.

Stochastic simulation impossible, Poisson representation promising but difficult to manage, direct finite time-step integration of birth-death process is the way to go.
Summary

- Populations interacting within hosts during viral infection are very large—viral population approaching $10^{14}$ particles.

- Stochastic simulation impossible, Poisson representation promising but difficult to manage, direct finite time-step integration of birth-death process is the way to go.

- Have predicted correlation between infected cell and viral population size fluctuations which is unique to the acute phase of the infection. Correlation remains even when extinction is accounted for.
Populations interacting within hosts during viral infection are very large—viral population approaching $10^{14}$ particles.

Stochastic simulation impossible, Poisson representation promising but difficult to manage, direct finite time-step integration of birth-death process is the way to go.

Have predicted correlation between infected cell and viral population size fluctuations which is unique to the acute phase of the infection. Correlation remains even when extinction is accounted for.

Sequence space projection promises to allow full stochastic simulation of viral evolution.
Summary

- Populations interacting within hosts during viral infection are very large—viral population approaching $10^{14}$ particles.
- Stochastic simulation impossible, Poisson representation promising but difficult to manage, direct finite time-step integration of birth-death process is the way to go.
- Have predicted correlation between infected cell and viral population size fluctuations which is unique to the acute phase of the infection. Correlation remains even when extinction is accounted for.
- Sequence space projection promises to allow full stochastic simulation of viral evolution.

Thank-you!