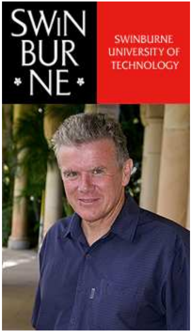
Within-host viral infection dynamics and evolution

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

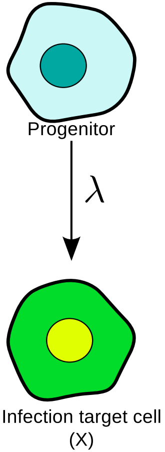
- Model effects of mutation due to different microscopic mechanisms
- Examine correlations arising between genetically distinct viral subpopulations
- Determine extent to which mutation mechanism affects these correlations

Part I

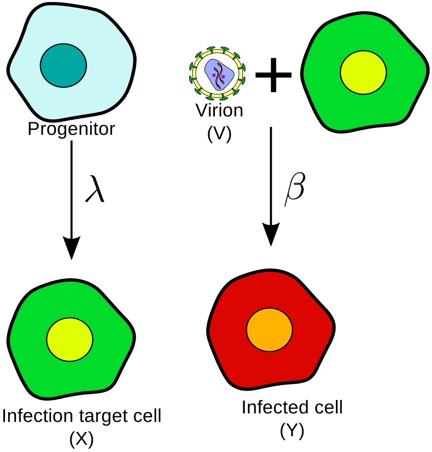
Demographic fluctuations

Primitive infection processes

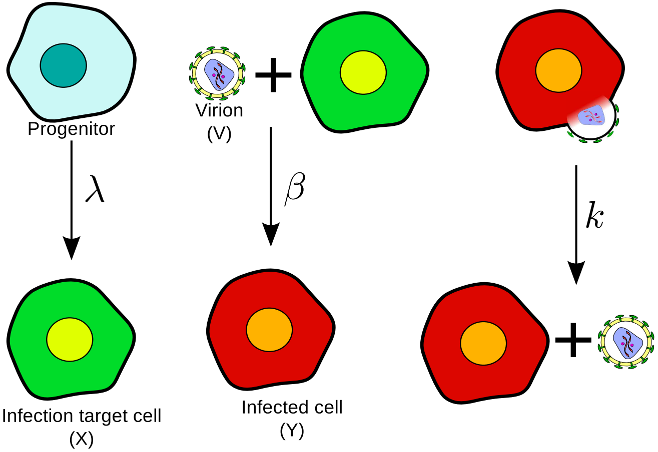
Primitive infection processes



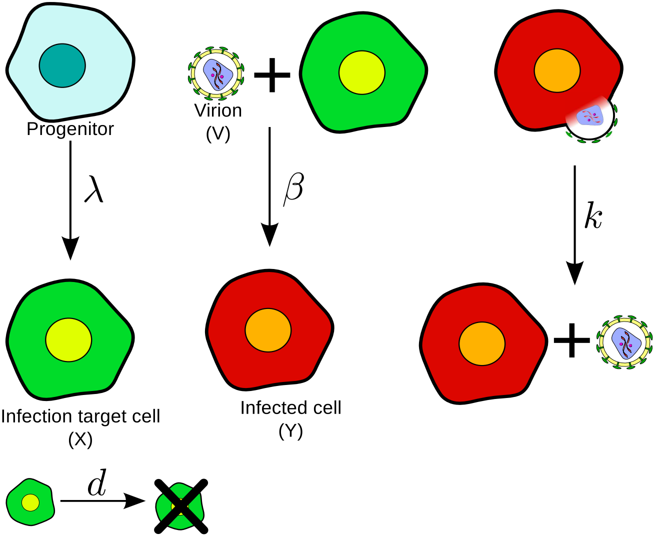
Primitive infection processes



Primitive infection processes



Primitive infection processes



Progenitor

λ

Virion
(V)

β

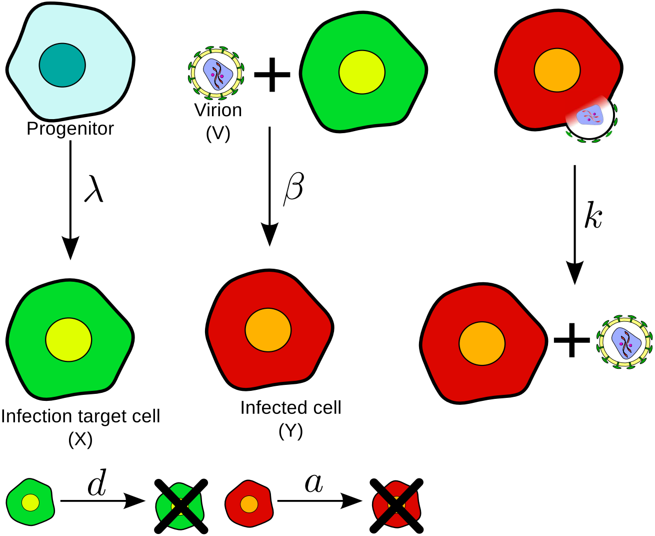
k

Infection target cell
(X)

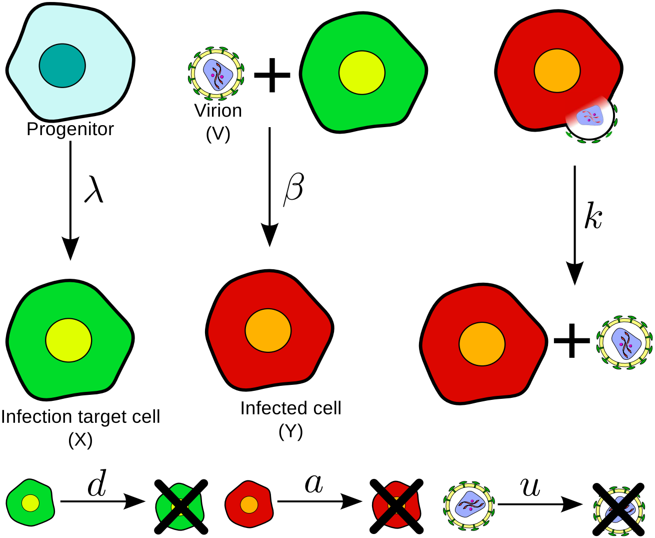
Infected cell
(Y)

d

Primitive infection processes



Primitive infection processes



Values of rate constants

Parameter	Description	Value
λ^*	T cell production rate	2.5×10^8/day
β^*	T cell infection rate	5×10^{-13}/T cell/virion/day
k	virion production rate	10^3/infected T cell/day
d	T cell death rate	10^{-3}/T cell/day
a	infected T cell death rate	1/infected T cell/day
u	virion clearance rate	3/virion/day

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

Values of rate constants

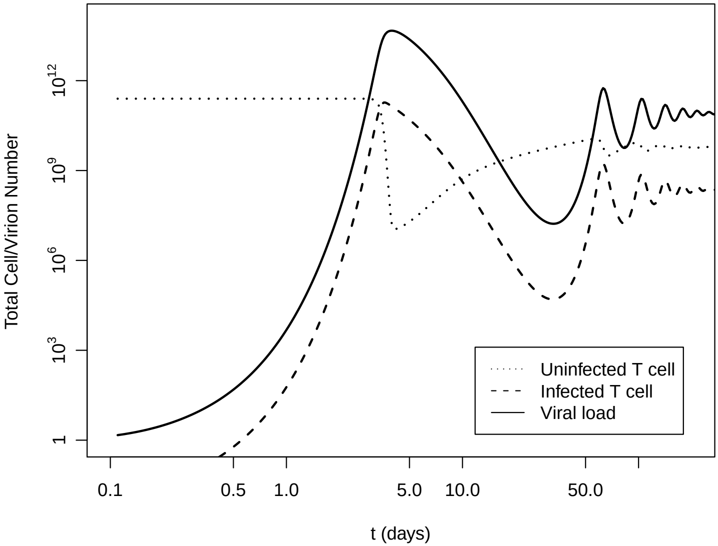
Parameter	Description	Value
λ^*	T cell production rate	2.5×10^8/day
β^*	T cell infection rate	5×10^{-13}/T cell/virion/day
k	virion production rate	10^3/infected T cell/day
d	T cell death rate	10^{-3}/T cell/day
a	infected T cell death rate	1/infected T cell/day
u	virion clearance rate	3/virion/day

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

Deterministic Model

$$\begin{aligned}\dot{x} &= \lambda - \beta x v - d x\\ \dot{y} &= \beta x v - a y\\ \dot{v} &= k y - \beta x v - u v\end{aligned}$$

Deterministic predictions



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)

$$
\begin{aligned}
\frac{\partial}{\partial t} P(N_x, N_y, N_v) \quad = \quad & \lambda \big[P(N_x - 1, N_y, N_v) - P(N_x, N_y, N_v) \big] \\
& + \beta \big[(N_x - 1)(N_v - 1) P(N_x - 1, N_y + 1, N_v - 1) \\
& \quad - N_x N_y P(N_x, N_y, N_v) \big] \\
& + k N_y \big[P(N_x, N_y + 1, N_v - 1) - P(N_x, N_y, N_v) \big] \\
& + d \big[(N_x + 1) P(N_x + 1, N_y, N_v) - N_x P(N_x, N_y, N_v) \big] \\
& + a \big[(N_y + 1) P(N_x, N_y + 1, N_v) - N_y P(N_x, N_y, N_v) \big] \\
& + u \big[(N_v + 1) P(N_x, N_y, N_v + 1) - N_v P(N_x, N_y, N_v) \big]
\end{aligned}
$$

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

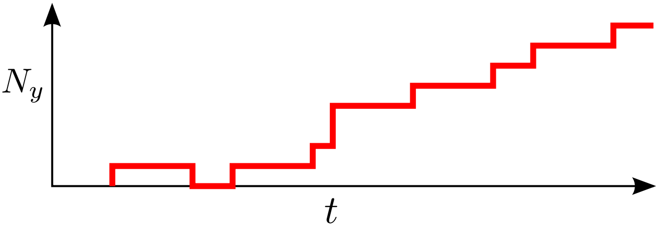
- Given Poissonian fluctuations in the variables N_x, N_y and N_v, expect the 'steady-state' volume of occupied state space to be $\gtrsim 10^{15}$: too large for direct numerical integration of CME.
- Instead, use Monte Carlo approaches to draw samples from probability distribution over possible dynamical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

Traditional Monte Carlo approach

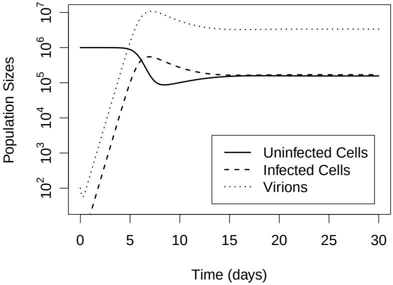
- Given Poissonian fluctuations in the variables N_x, N_y and N_v, expect the 'steady-state' volume of occupied state space to be $\gtrsim 10^{15}$: too large for direct numerical integration of CME.
- Instead, use Monte Carlo approaches to draw samples from probability distribution over possible dynamical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

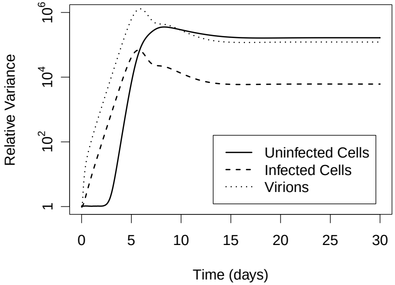
Traditional Monte Carlo approach

- 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):

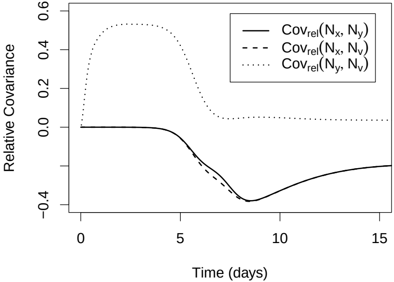


SSA results for small system

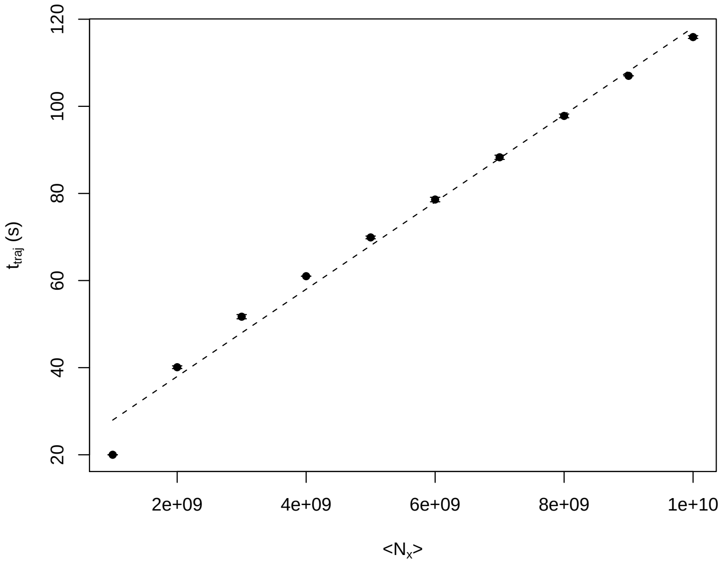




$$\mathrm{Cov}_{\mathrm{rel}}(N_i, N_j) \equiv \frac{\langle N_i N_j \rangle}{\langle N_i \rangle \langle N_j \rangle} - 1$$



Computational burden of the SSA



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.

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:

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)

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

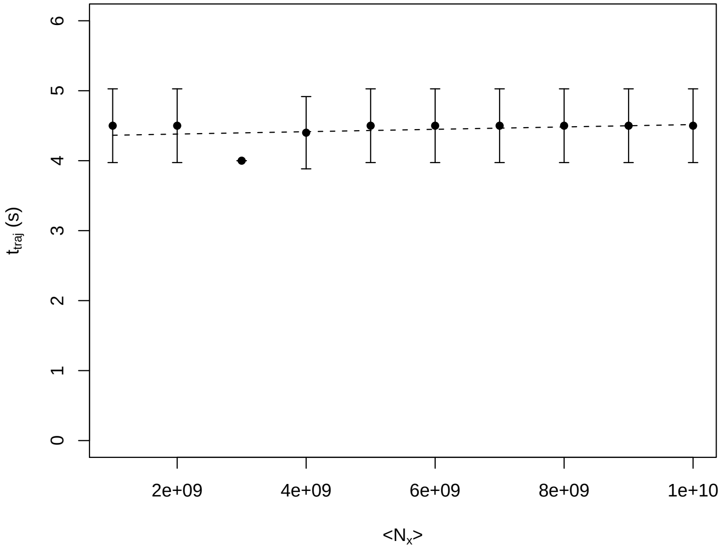
Difficulties with Poisson representation approach

- 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 $|\vec{\alpha}| \rightarrow \infty$. 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 originiating from a Poisson process using a Poissonian basis can result in a transformed distribution with tigher 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 originiating from a Poisson process using a Poissonian basis can result in a transformed distribution with tigher support.

   $\Longrightarrow$ Fewer trajectories needed to sample the distribution.

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 originiating from a Poisson process using a Poissonian basis can result in a transformed distribution with tigher support.

   $\Longrightarrow$ 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, *the size of which does not explicitly depend on the magnitude of the variables involved.*

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.

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:

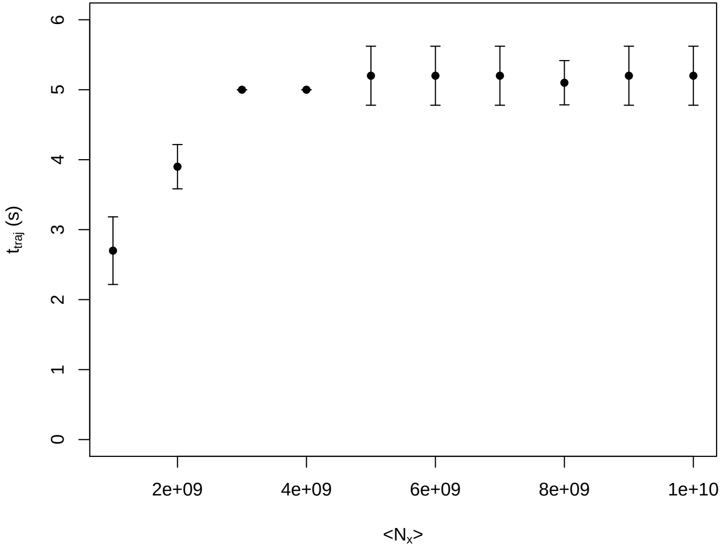
$$P(N'_x, t+\tau | N_x, t) = \sum_{m=0}^{\infty} \delta_{N'_x - N_x, m} e^{-\tau\lambda} \frac{(\tau\lambda)^m}{m!}$$

- Finite time step approximation to stochastic trajectory for N_x can be generated by iterating:

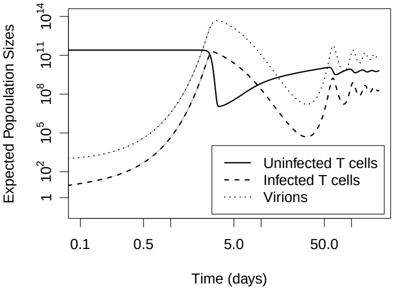
$$N_x(t+(q+1)\tau) = N_x(t+q\tau) + m_q$$

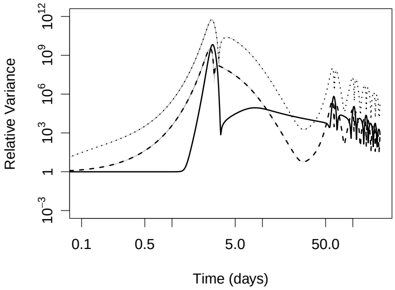
with each m_q chosen from a Poisson distrib. with mean $\tau\lambda$. (Gillespie, 2001)

Finite time step integration for discrete processes

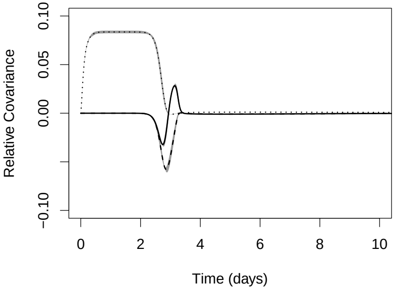


τ-leaping results for full-sized model

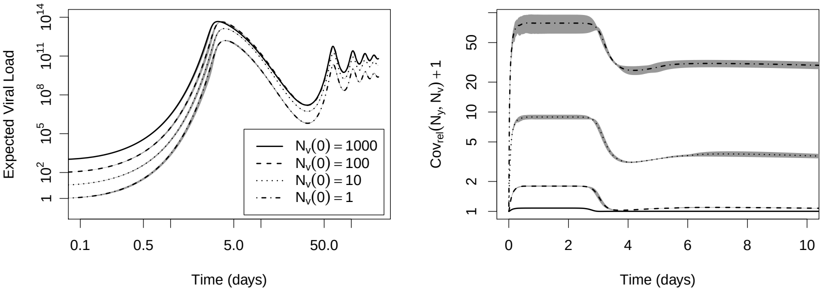




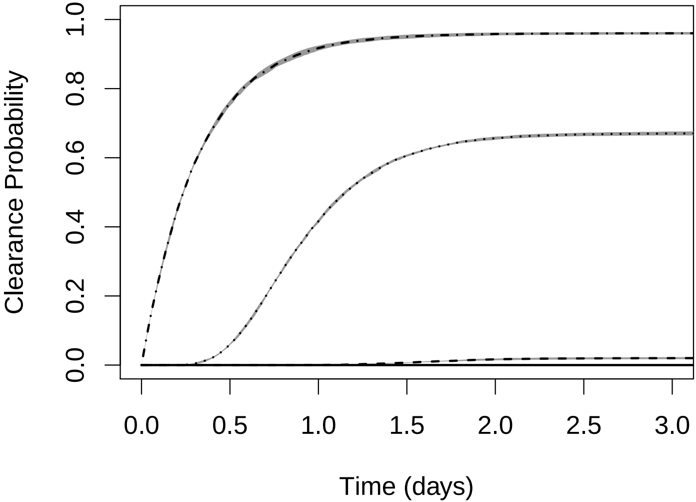
$$\mathrm{Cov}_{\mathrm{rel}}(N_i, N_j) \equiv \frac{\langle N_i N_j \rangle}{\langle N_i \rangle \langle N_j \rangle} - 1$$



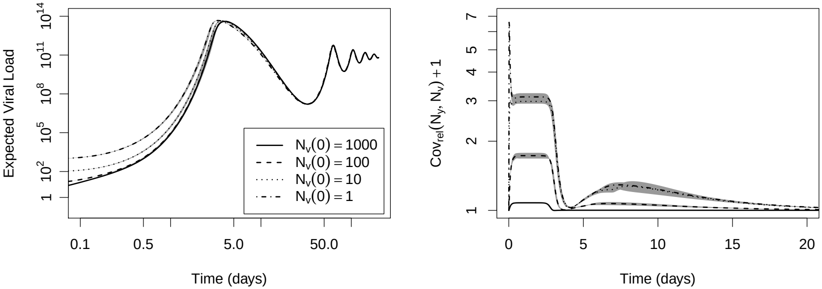
Impact of initial population size



Impact of initial population size



Impact of initial population size



Part II

Genetic correlations

Mutation in HIV replication

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

Mutation in HIV replication

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

1. recombination (two viruses infect a single cell), and

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 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.

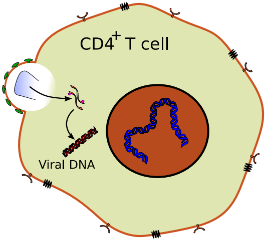
Mutation occurs during:

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:



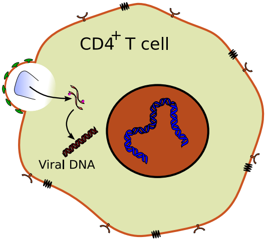
Reverse Transcription

Mutation in HIV replication

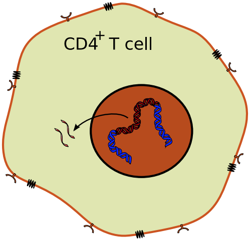
There are two primary generators of within-host genetic diversity:

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

Mutation occurs during:



Reverse Transcription



Transcription

Scale of the problem

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.

Scale of the problem

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, $L \simeq 10^4$—sequence space impossibly large.

Scale of the problem

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, $L \simeq 10^4$—sequence space impossibly large.
- Sub-sequence of length $L = 105$ corresponds to protein sequence which does not directly affect viral fecundity. $4^{105} \sim 10^{63}$

Scale of the problem

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, $L \simeq 10^4$—sequence space impossibly large.
- Sub-sequence of length $L = 105$ corresponds to protein sequence which does not directly affect viral fecundity. $4^{105} \sim 10^{63}$
- Mutation rate $\mu \sim 2 \times 10^{-5}$/character/replication: small, but viral population reaches $\sim 10^{14}$ at its peak.

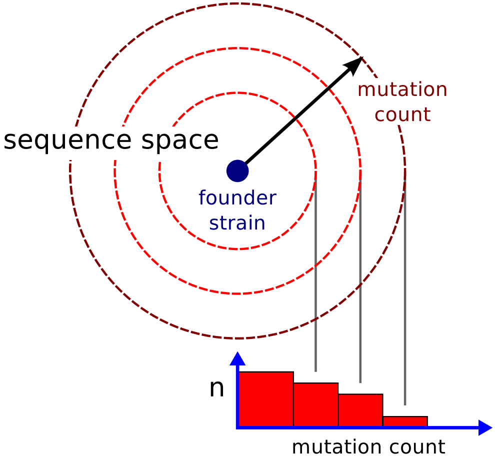
Scale of the problem

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.

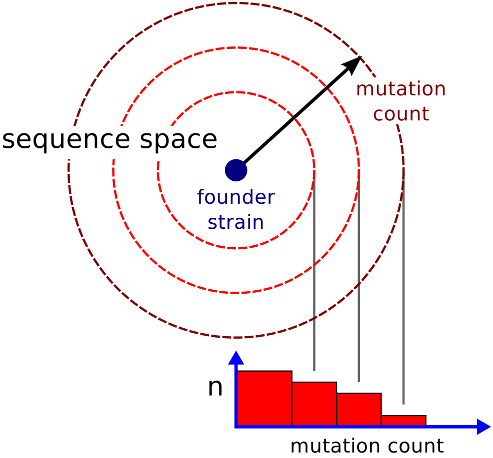
- For HIV, $L \simeq 10^4$—sequence space impossibly large.
- Sub-sequence of length $L = 105$ corresponds to protein sequence which does not directly affect viral fecundity. $4^{105} \sim 10^{63}$
- Mutation rate $\mu \sim 2 \times 10^{-5}$/character/replication: small, but viral population reaches $\sim 10^{14}$ at its peak.

$\Longrightarrow$ stochastic simulation (even τ-leaping) without further simplification is unwieldy at best.

Sequence space projection

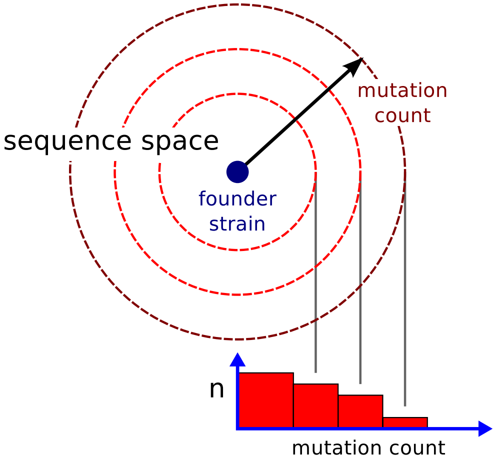


Sequence space projection



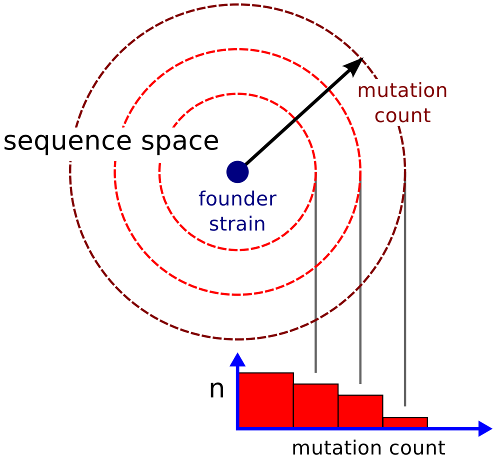
- Partition seq. space into ‘hyperspheres’.

Sequence space projection



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

Sequence space projection



- 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.

Summary

Summary

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

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.

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.

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.

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!