

# Stochastic Simulations and Phylogeography

Tim Vaughan

Centre for Computational Evolution  
The University of Auckland, New Zealand

Engelberg, Switzerland, 2016

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# Part I

## Simulation

# What is simulation?

## Wikipedia

Simulation is the imitation of the operation of a real-world process or system over time.

[Introduction](#)

[Background theory](#)

[MASTER](#)

[MASTER population simulation examples](#)

[Linear birth/death](#)

[Stochastic logistic model](#)

[Stochastic epidemic model](#)

[Within-host virus dynamics](#)

[Simulating trees](#)

[Birth/death trees](#)

[SEIR trees](#)

[Moran trees](#)

[Conditioning and Sampling](#)

[Serial sampling](#)

[Contemporaneous sampling](#)

[Simulating coalescent trees](#)

# What is simulation?

## Wikipedia

Simulation is the imitation of the operation of a real-world process or system over time.

- ▶ Implies the existence of a mathematical model.
- ▶ Imitation of the real world: some physical/biological grounding.
- ▶ Time-dependence of central importance.

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# What is simulation?

## Wikipedia

Simulation is the imitation of the operation of a real-world process or system over time.

- ▶ Implies the existence of a mathematical model.
- ▶ Imitation of the real world: some physical/biological grounding.
- ▶ Time-dependence of central importance.

*When we say “simulation” we are exclusively thinking of “computer simulation”.*

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Why simulate?

Generally:

- ▶ In order to compare reality to modelling predictions we need to know what those predictions are.

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Why simulate?

Generally:

- ▶ In order to compare reality to modelling predictions we need to know what those predictions are.
- ▶ It is very easy to specify models that are too realistic that predictions cannot be made analytically.

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Why simulate?

Generally:

- ▶ In order to compare reality to modelling predictions we need to know what those predictions are.
- ▶ It is very easy to specify models that are too realistic that predictions cannot be made analytically.
- ▶ Therefore, simulate!

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Why simulate?

Generally:

- ▶ In order to compare reality to modelling predictions we need to know what those predictions are.
- ▶ It is very easy to specify models that are too realistic that predictions cannot be made analytically.
- ▶ Therefore, simulate!

In the context of phylogenetic inference:

- ▶ Bayesian phylogenetic inference seeks to draw conclusions from sequence data using models of:

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Why simulate?

Generally:

- ▶ In order to compare reality to modelling predictions we need to know what those predictions are.
- ▶ It is very easy to specify models that are too realistic that predictions cannot be made analytically.
- ▶ Therefore, simulate!

In the context of phylogenetic inference:

- ▶ Bayesian phylogenetic inference seeks to draw conclusions from sequence data using models of:
  - ▶ the substitution process,

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Why simulate?

Generally:

- ▶ In order to compare reality to modelling predictions we need to know what those predictions are.
- ▶ It is very easy to specify models that are too realistic that predictions cannot be made analytically.
- ▶ Therefore, simulate!

In the context of phylogenetic inference:

- ▶ Bayesian phylogenetic inference seeks to draw conclusions from sequence data using models of:
  - ▶ the substitution process,
  - ▶ the genealogical process,

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Why simulate?

Generally:

- ▶ In order to compare reality to modelling predictions we need to know what those predictions are.
- ▶ It is very easy to specify models that are too realistic that predictions cannot be made analytically.
- ▶ Therefore, simulate!

In the context of phylogenetic inference:

- ▶ Bayesian phylogenetic inference seeks to draw conclusions from sequence data using models of:
  - ▶ the substitution process,
  - ▶ the genealogical process,
  - ▶ population dynamics and or species richness.

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Why simulate?

Generally:

- ▶ In order to compare reality to modelling predictions we need to know what those predictions are.
- ▶ It is very easy to specify models that are too realistic that predictions cannot be made analytically.
- ▶ Therefore, simulate!

In the context of phylogenetic inference:

- ▶ Bayesian phylogenetic inference seeks to draw conclusions from sequence data using models of:
  - ▶ the substitution process,
  - ▶ the genealogical process,
  - ▶ population dynamics and or species richness.
- ▶ Simulation helps us better understand our models (and thus our priors).

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Why simulate?

Generally:

- ▶ In order to compare reality to modelling predictions we need to know what those predictions are.
- ▶ It is very easy to specify models that are too realistic that predictions cannot be made analytically.
- ▶ Therefore, simulate!

In the context of phylogenetic inference:

- ▶ Bayesian phylogenetic inference seeks to draw conclusions from sequence data using models of:
  - ▶ the substitution process,
  - ▶ the genealogical process,
  - ▶ population dynamics and or species richness.
- ▶ Simulation helps us better understand our models (and thus our priors).
- ▶ It is also possible to directly apply simulation techniques to inference. (e.g. ABC.)

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Deterministic Modelling

- ▶ The predictions of deterministic models are fully specified by their parameters and initial conditions.

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

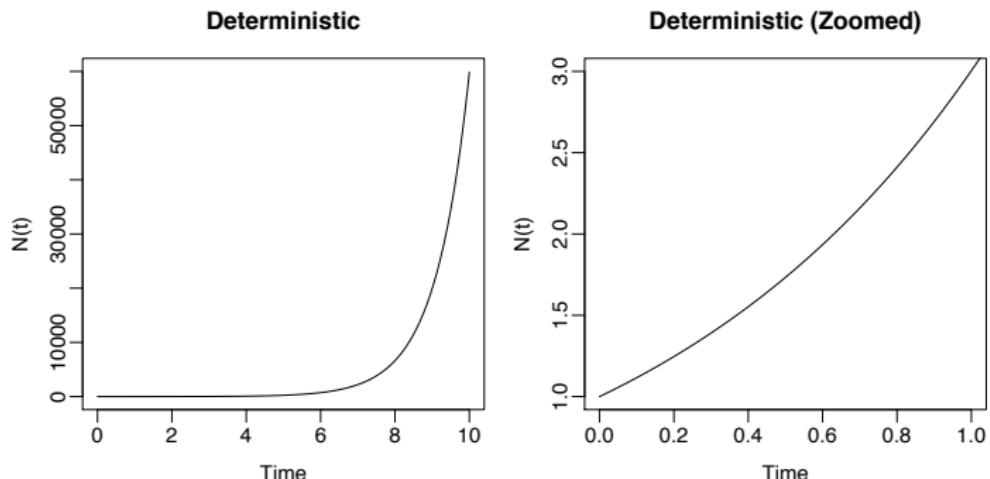
Contemporaneous sampling

Simulating coalescent  
trees

# Deterministic Modelling

- ▶ The predictions of deterministic models are fully specified by their parameters and initial conditions.

Consider the following exponential model of population growth:



Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

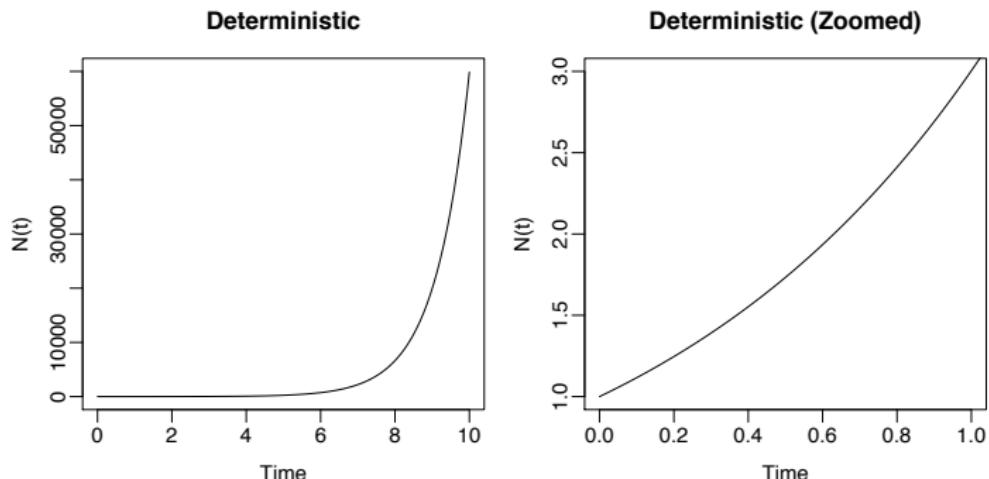
Serial sampling

Contemporaneous sampling

Simulating coalescent trees

- ▶ The predictions of deterministic models are fully specified by their parameters and initial conditions.

Consider the following exponential model of population growth:



- ▶ Model becomes unrealistic for short times.

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Stochastic Modelling

- ▶ Stochastic models allow portions of their dynamics to be described probabilistically.

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

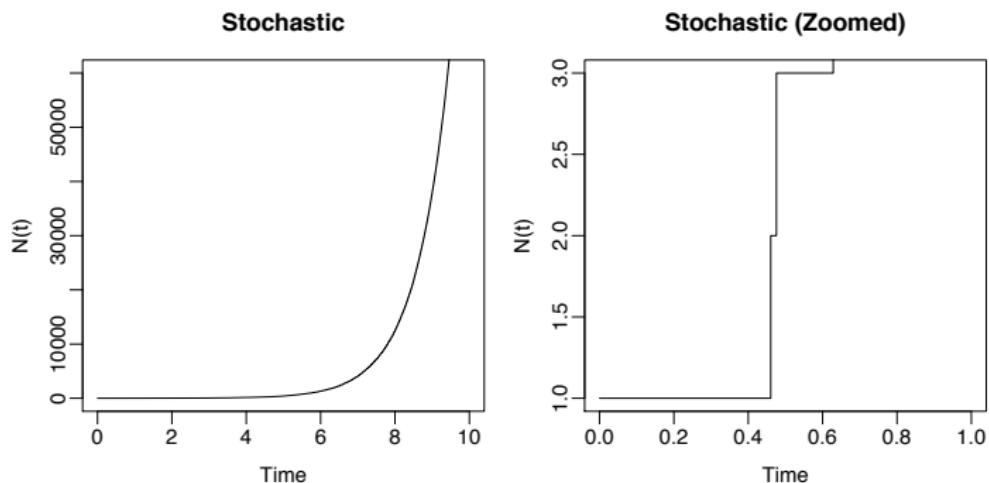
Contemporaneous sampling

Simulating coalescent  
trees

# Stochastic Modelling

- ▶ Stochastic models allow portions of their dynamics to be described probabilistically.

Consider the following stochastic birth/death model of population growth:

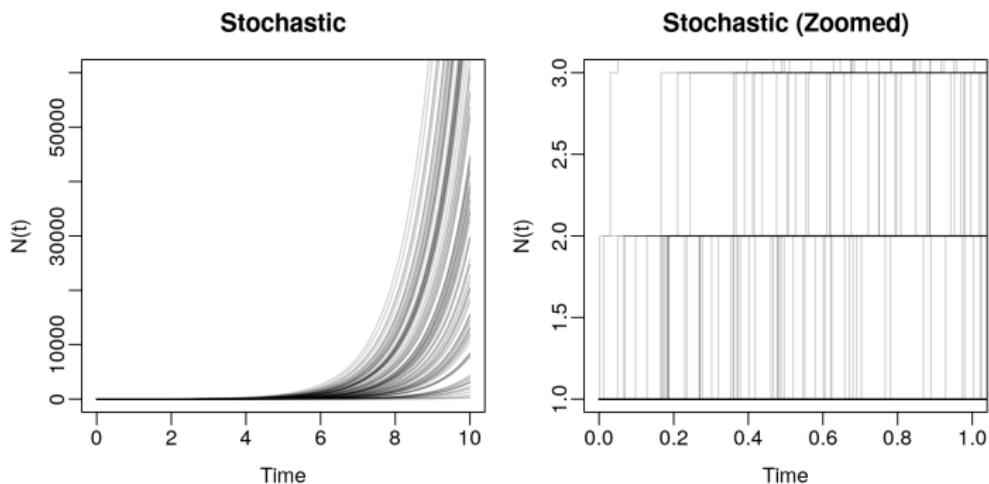


- Introduction
- Background theory
- MASTER
  - MASTER population simulation examples
  - Linear birth/death
  - Stochastic logistic model
  - Stochastic epidemic model
  - Within-host virus dynamics
- Simulating trees
  - Birth/death trees
  - SEIR trees
  - Moran trees
- Conditioning and Sampling
  - Serial sampling
  - Contemporaneous sampling
- Simulating coalescent trees

# Stochastic Modelling

- ▶ Stochastic models allow portions of their dynamics to be described probabilistically.

Consider the following stochastic birth/death model of population growth:

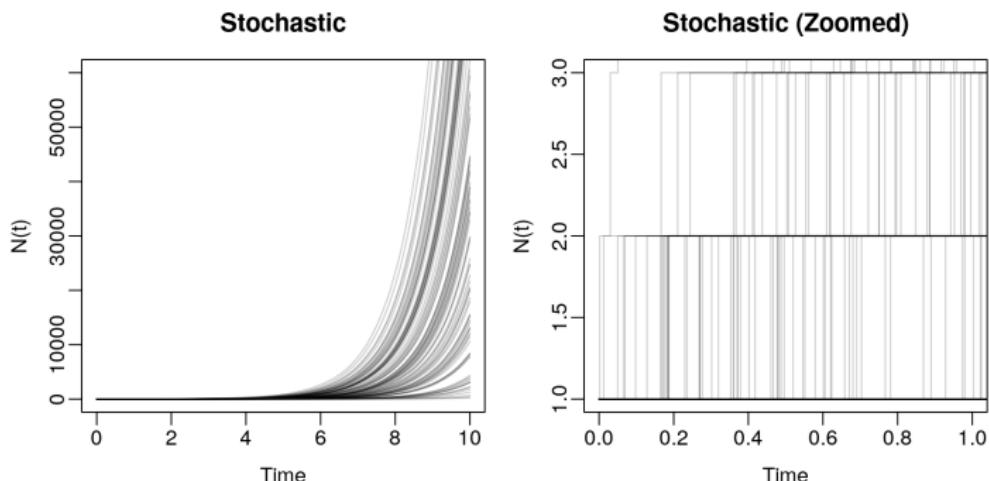


- Introduction
- Background theory
- MASTER
- MASTER population simulation examples
  - Linear birth/death
  - Stochastic logistic model
  - Stochastic epidemic model
  - Within-host virus dynamics
- Simulating trees
  - Birth/death trees
  - SEIR trees
  - Moran trees
- Conditioning and Sampling
  - Serial sampling
  - Contemporaneous sampling
- Simulating coalescent trees

# Stochastic Modelling

- ▶ Stochastic models allow portions of their dynamics to be described probabilistically.

Consider the following stochastic birth/death model of population growth:

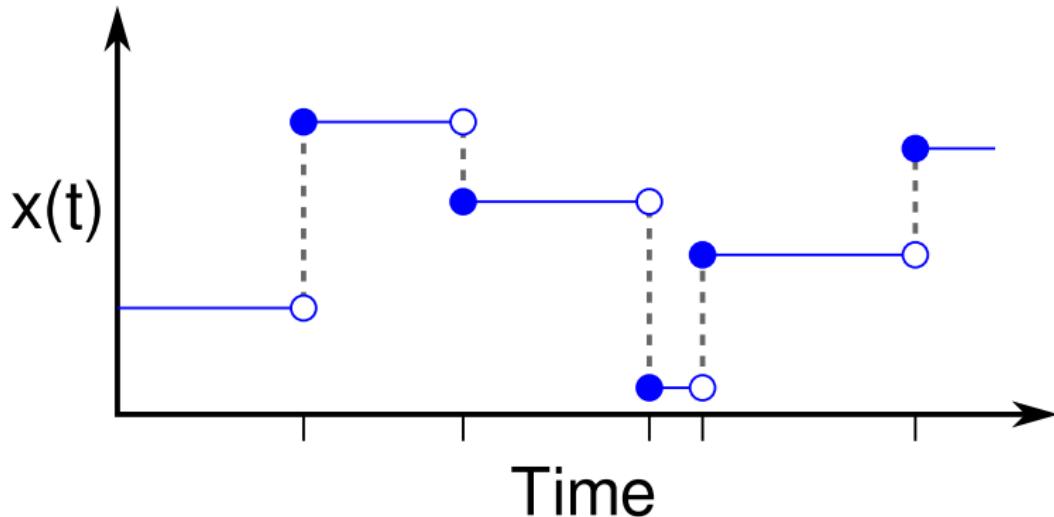


- ▶ Probability enters once we admit that we do not know everything about the dynamics.

Introduction  
Background theory  
MASTER  
MASTER population simulation examples  
Linear birth/death  
Stochastic logistic model  
Stochastic epidemic model  
Within-host virus dynamics  
Simulating trees  
Birth/death trees  
SEIR trees  
Moran trees  
Conditioning and Sampling  
Serial sampling  
Contemporaneous sampling  
Simulating coalescent trees

# Markovian jump processes

- ▶ Characterized by discrete system states affected by instantaneous events that can occur at any time.
- ▶ Rate of transition depends only on current state.



Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

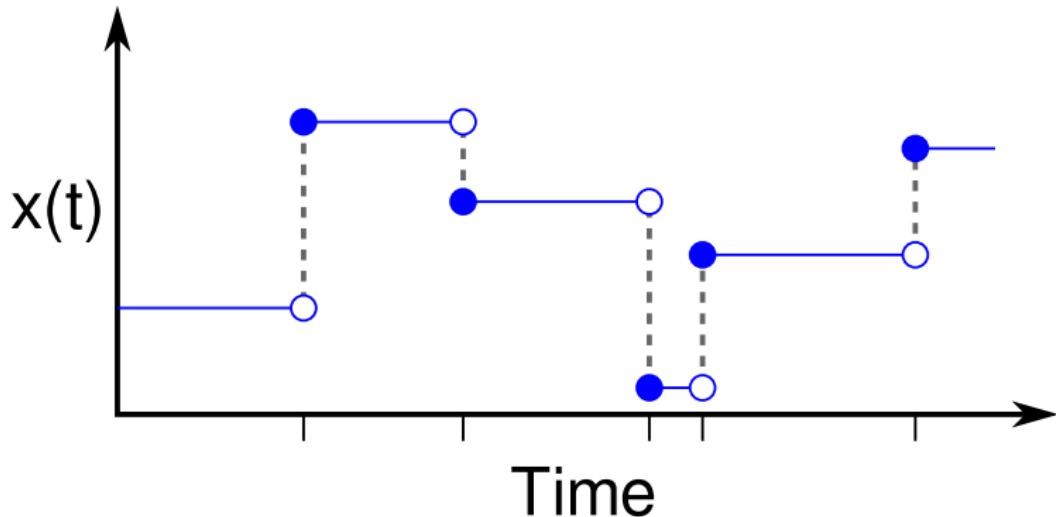
Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Markovian jump processes

- ▶ Characterized by discrete system states affected by instantaneous events that can occur at any time.
- ▶ Rate of transition depends only on current state.

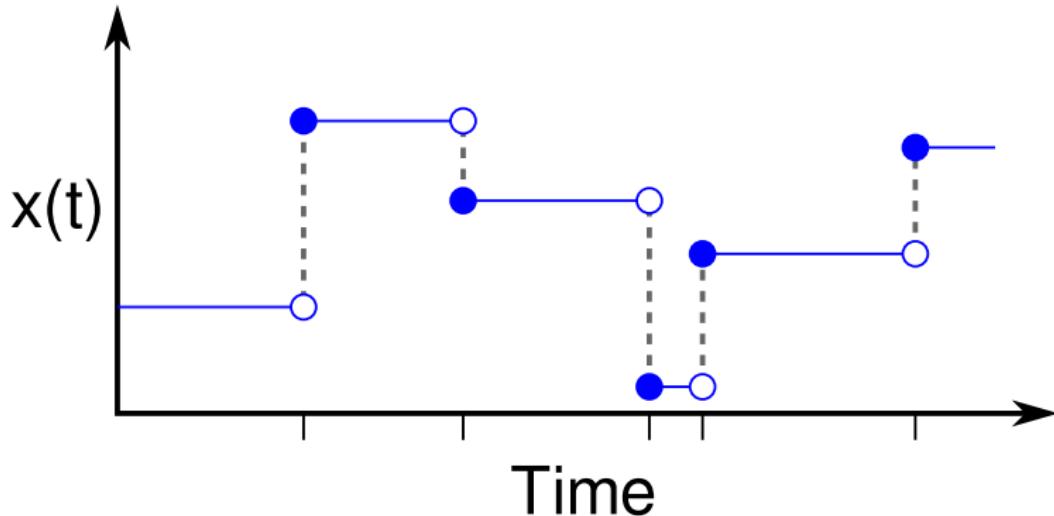


- ▶ Used to model stochastic "population" dynamics.

Introduction  
Background theory  
MASTER  
MASTER population simulation examples  
Linear birth/death  
Stochastic logistic model  
Stochastic epidemic model  
Within-host virus dynamics  
Simulating trees  
Birth/death trees  
SEIR trees  
Moran trees  
Conditioning and Sampling  
Serial sampling  
Contemporaneous sampling  
Simulating coalescent trees

# Markovian jump processes

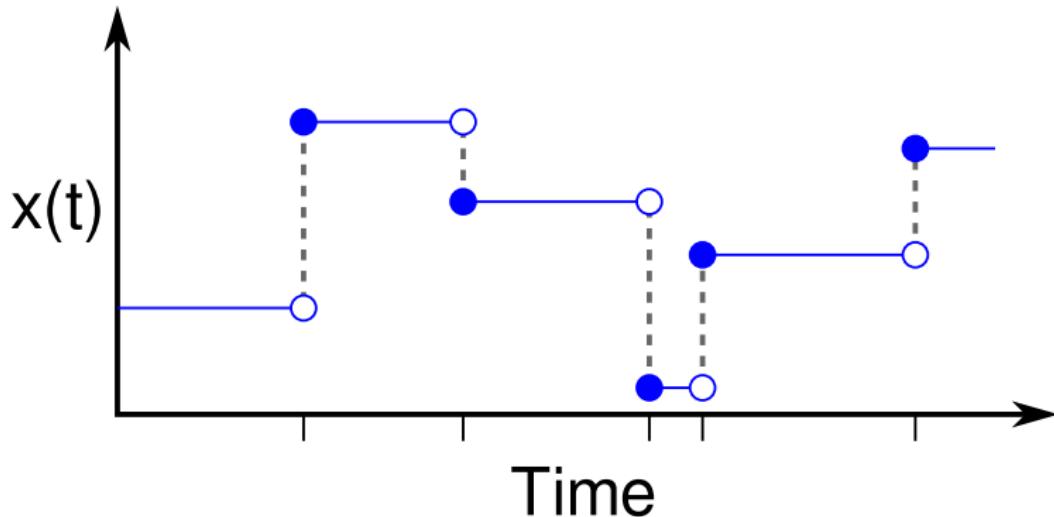
- ▶ Characterized by discrete system states affected by instantaneous events that can occur at any time.
- ▶ Rate of transition depends only on current state.



- ▶ Used to model stochastic "population" dynamics.
- ▶ Continuous time Markov chains (CTMCs) are a special case where the number of states is finite.

Introduction
Background theory
MASTER
MASTER population simulation examples
Linear birth/death
Stochastic logistic model
Stochastic epidemic model
Within-host virus dynamics
Simulating trees
Birth/death trees
SEIR trees
Moran trees
Conditioning and Sampling
Serial sampling
Contemporaneous sampling
Simulating coalescent trees

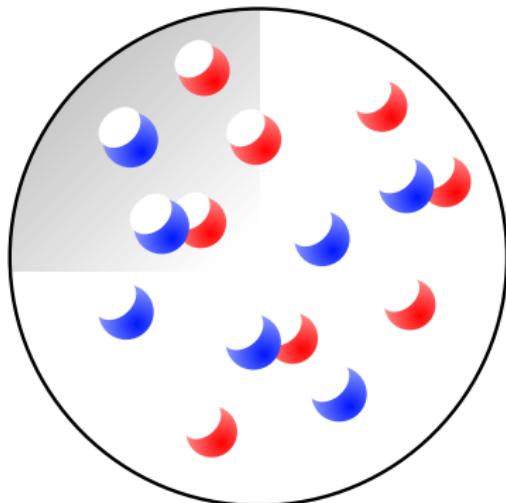
- ▶ Characterized by discrete system states affected by instantaneous events that can occur at any time.
- ▶ Rate of transition depends only on current state.



- ▶ Used to model stochastic "population" dynamics.
- ▶ Continuous time Markov chains (CTMCs) are a special case where the number of states is finite.
- ▶ CTMCs used to model substitution process.

Introduction  
Background theory  
MASTER  
MASTER population simulation examples  
Linear birth/death  
Stochastic logistic model  
Stochastic epidemic model  
Within-host virus dynamics  
Simulating trees  
Birth/death trees  
SEIR trees  
Moran trees  
Conditioning and Sampling  
Serial sampling  
Contemporaneous sampling  
Simulating coalescent trees

# Mass-action chemical kinetics



Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

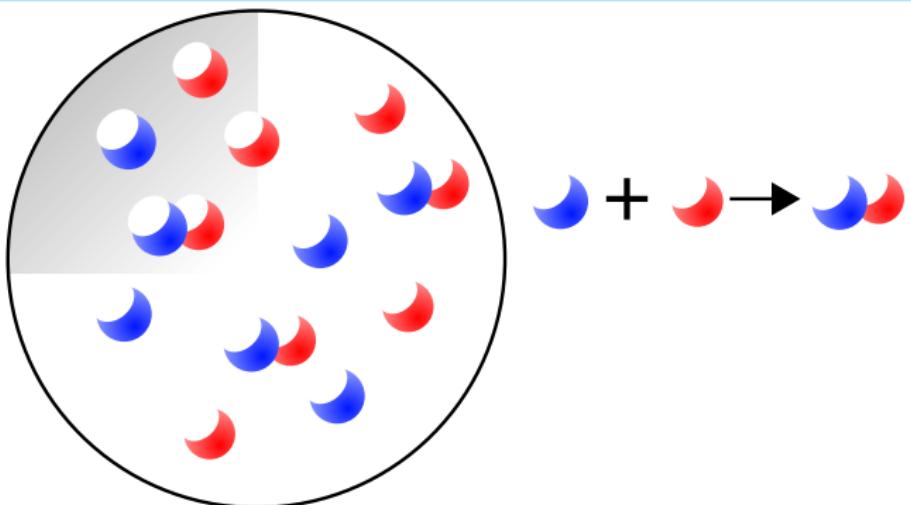
Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Mass-action chemical kinetics



Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

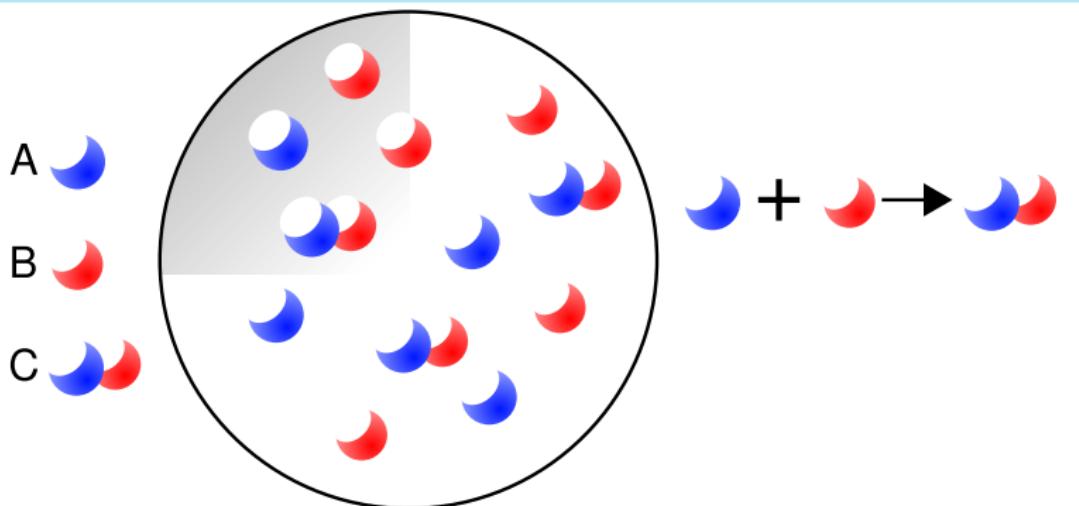
Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Mass-action chemical kinetics



Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

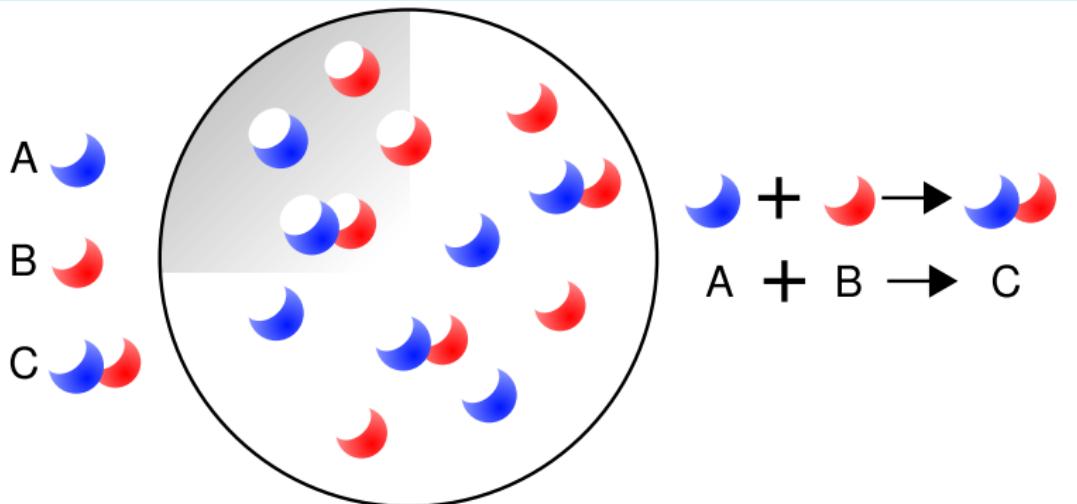
Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Mass-action chemical kinetics



Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

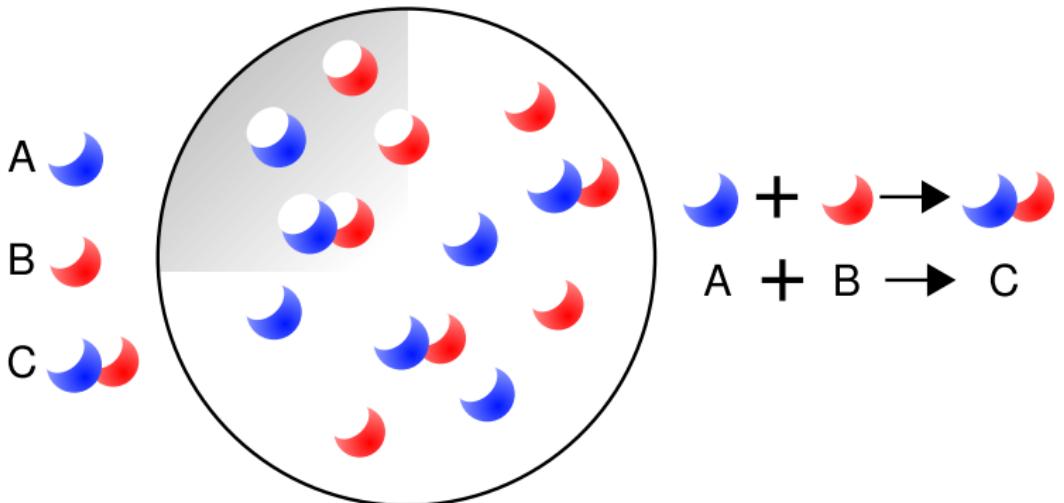
Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

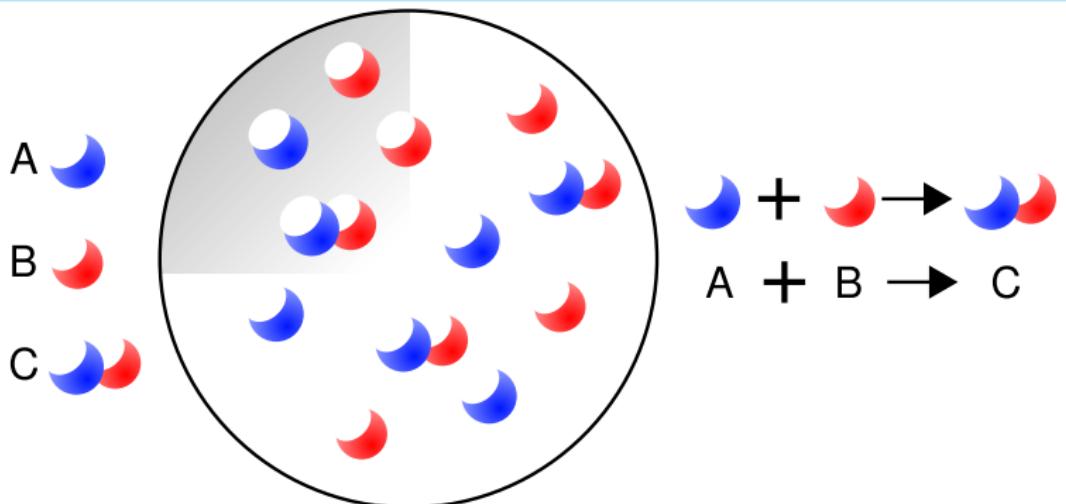
# Mass-action chemical kinetics



- ▶ Probability of reaction in some  $\delta t \ll 1$ :

Introduction  
Background theory  
MASTER  
MASTER population simulation examples  
Linear birth/death  
Stochastic logistic model  
Stochastic epidemic model  
Within-host virus dynamics  
Simulating trees  
Birth/death trees  
SEIR trees  
Moran trees  
Conditioning and Sampling  
Serial sampling  
Contemporaneous sampling  
Simulating coalescent trees

# Mass-action chemical kinetics

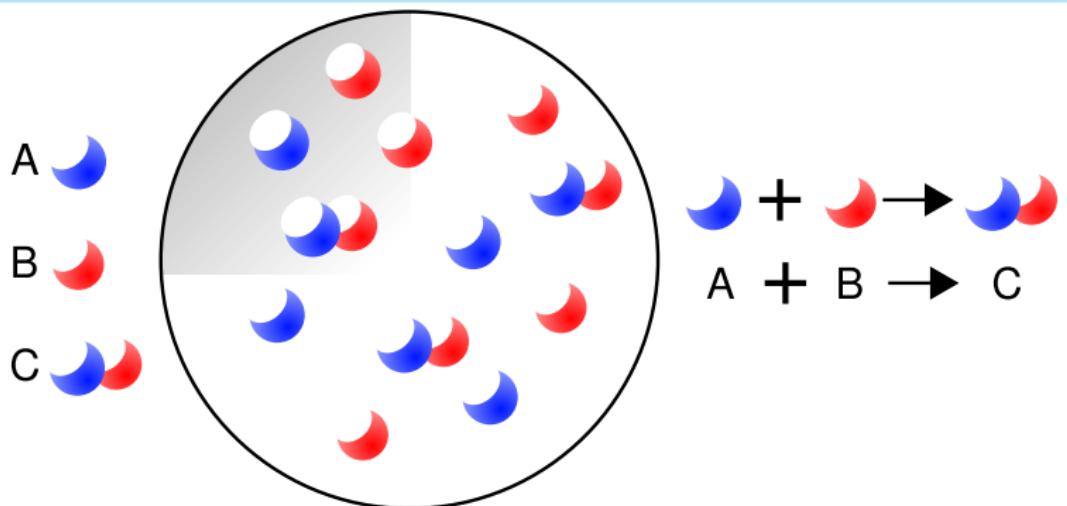


- ▶ Probability of reaction in some  $\delta t \ll 1$ :

$$k \times N_A \times N_B \times \delta t \equiv T(N_A, N_B) \delta t$$

Introduction  
Background theory  
MASTER  
MASTER population simulation examples  
Linear birth/death  
Stochastic logistic model  
Stochastic epidemic model  
Within-host virus dynamics  
Simulating trees  
Birth/death trees  
SEIR trees  
Moran trees  
Conditioning and Sampling  
Serial sampling  
Contemporaneous sampling  
Simulating coalescent trees

# Mass-action chemical kinetics



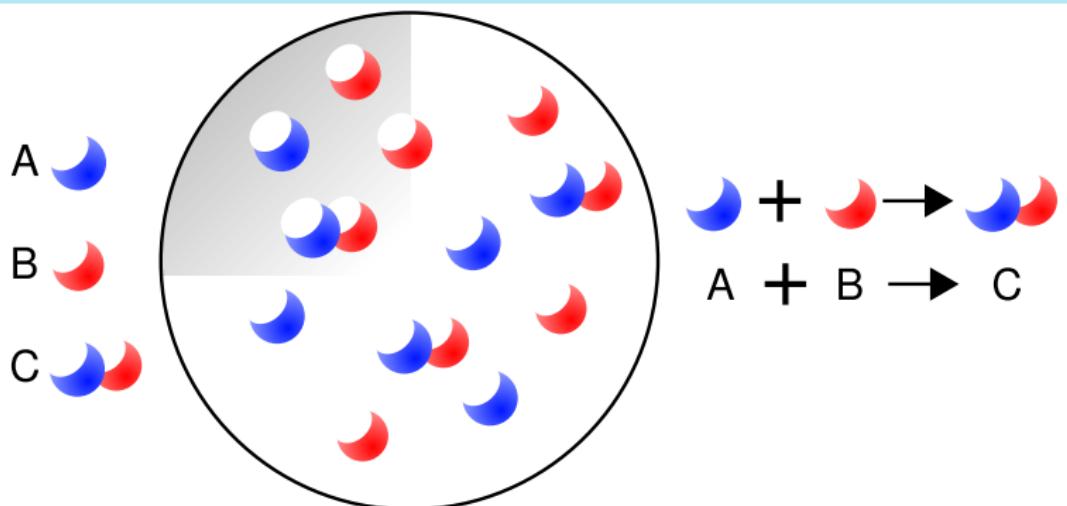
- ▶ Probability of reaction in some  $\delta t \ll 1$ :

$$k \times N_A \times N_B \times \delta t \equiv T(N_A, N_B) \delta t$$

- ▶ With the combinatoric term understood, write

Introduction  
Background theory  
MASTER  
MASTER population simulation examples  
Linear birth/death  
Stochastic logistic model  
Stochastic epidemic model  
Within-host virus dynamics  
Simulating trees  
Birth/death trees  
SEIR trees  
Moran trees  
Conditioning and Sampling  
Serial sampling  
Contemporaneous sampling  
Simulating coalescent trees

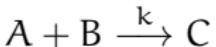
# Mass-action chemical kinetics



- ▶ Probability of reaction in some  $\delta t \ll 1$ :

$$k \times N_A \times N_B \times \delta t \equiv T(N_A, N_B) \delta t$$

- ▶ With the combinatoric term understood, write



Introduction  
Background theory  
MASTER  
MASTER population simulation examples  
Linear birth/death  
Stochastic logistic model  
Stochastic epidemic model  
Within-host virus dynamics  
Simulating trees  
Birth/death trees  
SEIR trees  
Moran trees  
Conditioning and Sampling  
Serial sampling  
Contemporaneous sampling  
Simulating coalescent trees

# Chemical Master Equations

- ▶ For simulation, only thing that matters are population sizes  $N_A$ ,  $N_B$  and  $N_C$ .

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# Chemical Master Equations

- ▶ For simulation, only thing that matters are population sizes  $N_A$ ,  $N_B$  and  $N_C$ .
- ▶ These vary stochastically, so really want  $P(N_A(t) = n_A, N_B(t) = n_B, N_C(t) = n_C) = p(\vec{n}, t)$ .

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# Chemical Master Equations

- ▶ For simulation, only thing that matters are population sizes  $N_A$ ,  $N_B$  and  $N_C$ .
- ▶ These vary stochastically, so really want  $P(N_A(t) = n_A, N_B(t) = n_B, N_C(t) = n_C) = p(\vec{n}, t)$ .
- ▶ Dynamics of  $p(\vec{n}, t)$  governed by Chemical Master Equation:

$$\frac{d}{dt}p(\vec{n}, t) = T(N_A + 1, N_B + 1)p(\vec{n}_{+A+B-C}, t) - T(N_A, N_B)p(\vec{n}, t)$$

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# Chemical Master Equations

- ▶ For simulation, only thing that matters are population sizes  $N_A$ ,  $N_B$  and  $N_C$ .
- ▶ These vary stochastically, so really want  $P(N_A(t) = n_A, N_B(t) = n_B, N_C(t) = n_C) = p(\vec{n}, t)$ .
- ▶ Dynamics of  $p(\vec{n}, t)$  governed by Chemical Master Equation:

$$\frac{d}{dt}p(\vec{n}, t) = T(N_A + 1, N_B + 1)p(\vec{n}_{+A+B-C}, t) - T(N_A, N_B)p(\vec{n}, t)$$

- ▶ Can interpret  $A + B \xrightarrow{k} C$  as exactly equivalent.

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Chemical Master Equations

- ▶ For simulation, only thing that matters are population sizes  $N_A$ ,  $N_B$  and  $N_C$ .
- ▶ These vary stochastically, so really want  $P(N_A(t) = n_A, N_B(t) = n_B, N_C(t) = n_C) = p(\vec{n}, t)$ .
- ▶ Dynamics of  $p(\vec{n}, t)$  governed by Chemical Master Equation:

$$\frac{d}{dt}p(\vec{n}, t) = T(N_A + 1, N_B + 1)p(\vec{n}_{+A+B-C}, t) - T(N_A, N_B)p(\vec{n}, t)$$

- ▶ Can interpret  $A + B \xrightarrow{k} C$  as exactly equivalent.
- ▶ This is generally HUGE linear system of differential equations: one for each value of  $n_A$ ,  $n_B$  and  $n_C$ .

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Chemical Master Equations

- ▶ For simulation, only thing that matters are population sizes  $N_A$ ,  $N_B$  and  $N_C$ .
- ▶ These vary stochastically, so really want  $P(N_A(t) = n_A, N_B(t) = n_B, N_C(t) = n_C) = p(\vec{n}, t)$ .
- ▶ Dynamics of  $p(\vec{n}, t)$  governed by Chemical Master Equation:

$$\frac{d}{dt}p(\vec{n}, t) = T(N_A + 1, N_B + 1)p(\vec{n}_{+A+B-C}, t) - T(N_A, N_B)p(\vec{n}, t)$$

- ▶ Can interpret  $A + B \xrightarrow{k} C$  as exactly equivalent.
- ▶ This is generally HUGE linear system of differential equations: one for each value of  $n_A$ ,  $n_B$  and  $n_C$ .

*Solve using Monte Carlo methods.*

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

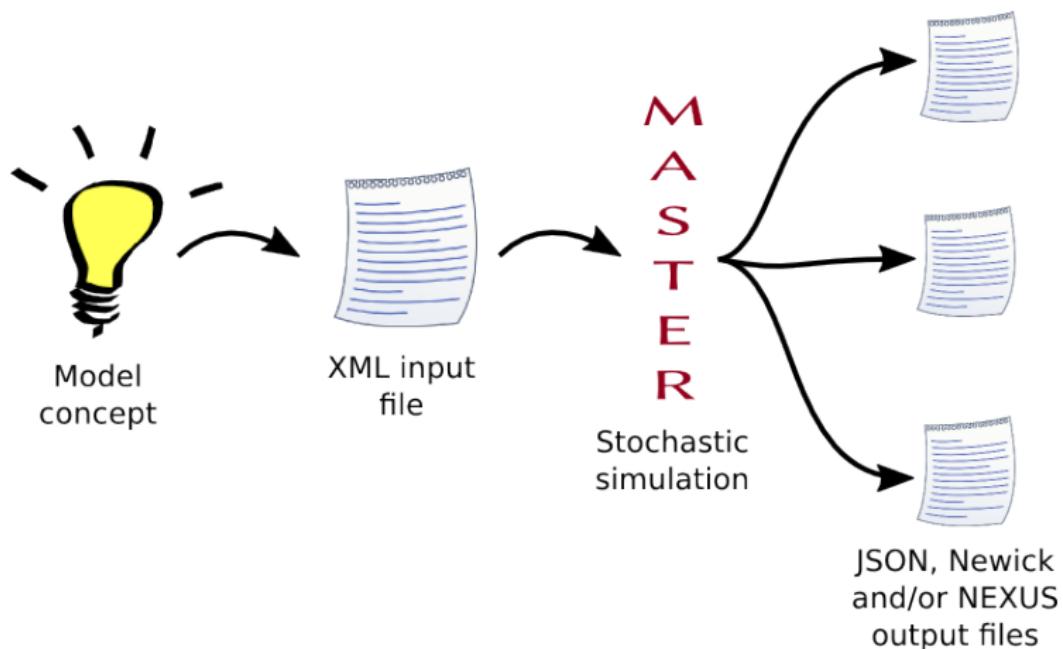
Contemporaneous sampling

Simulating coalescent  
trees

# Simulation using MASTER

## What is MASTER?

- ▶ BEAST 2 package for Monte Carlo treatment of CMEs.
- ▶ i.e. Simulation of systems that behave like chemical systems.



Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

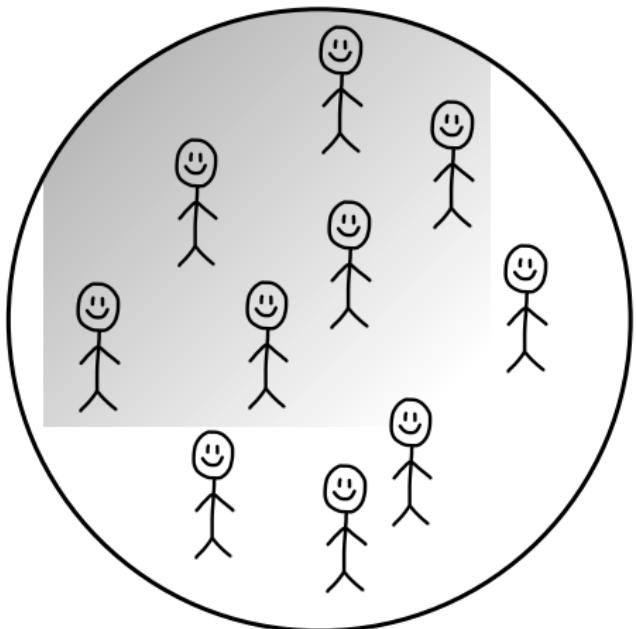
Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Example: Linear birth/death process



Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

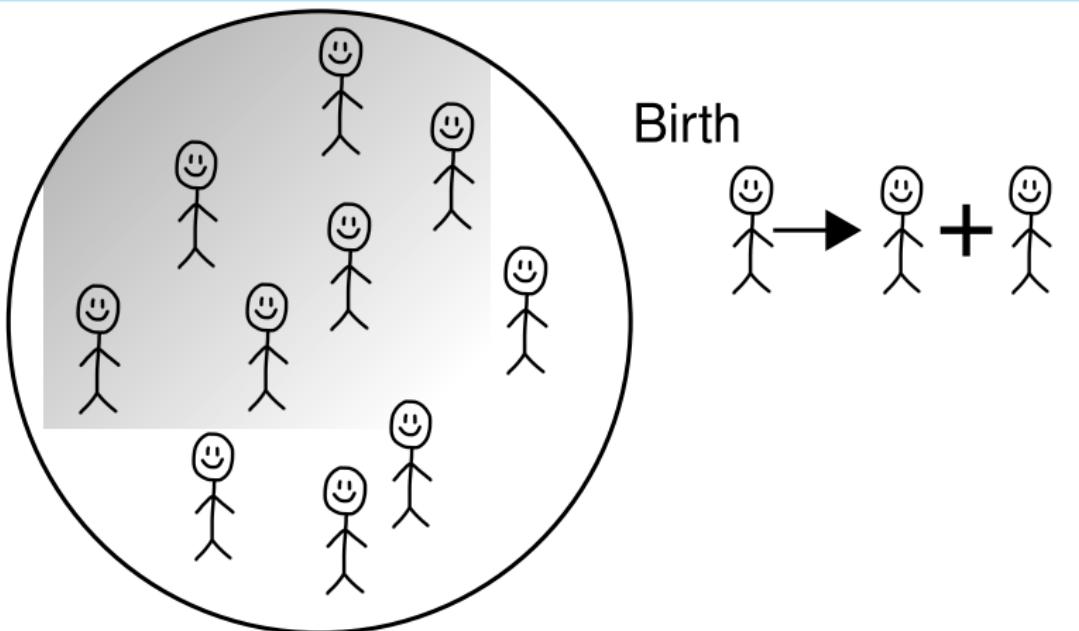
Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# Example: Linear birth/death process



[Introduction](#)

[Background theory](#)

[MASTER](#)

[MASTER population simulation examples](#)

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

[Simulating trees](#)

Birth/death trees

SEIR trees

Moran trees

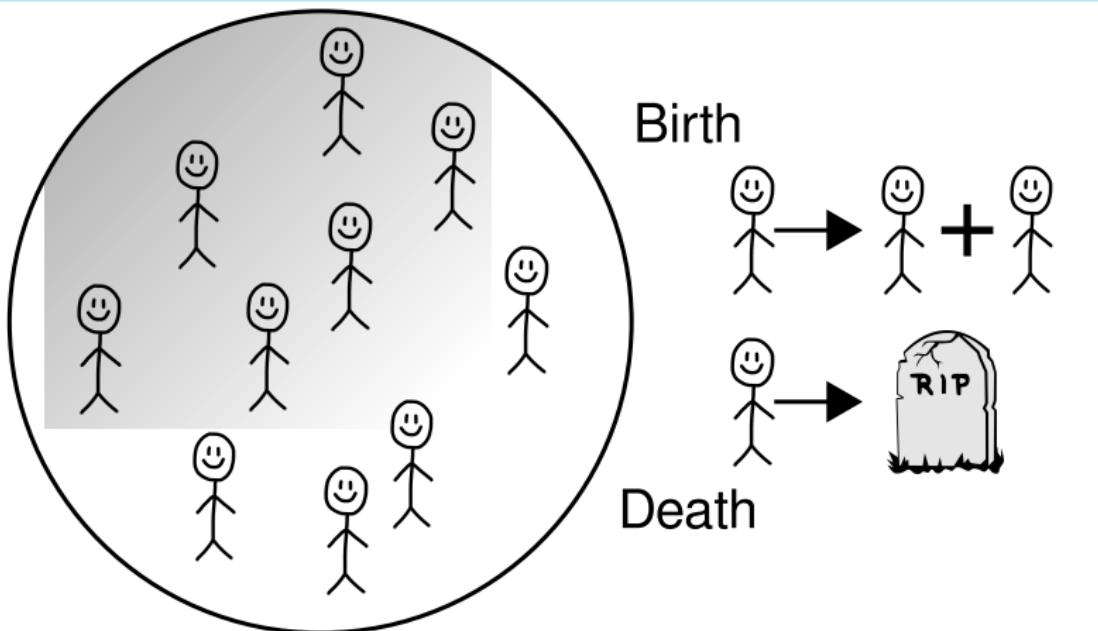
[Conditioning and Sampling](#)

Serial sampling

Contemporaneous sampling

[Simulating coalescent trees](#)

# Example: Linear birth/death process



Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

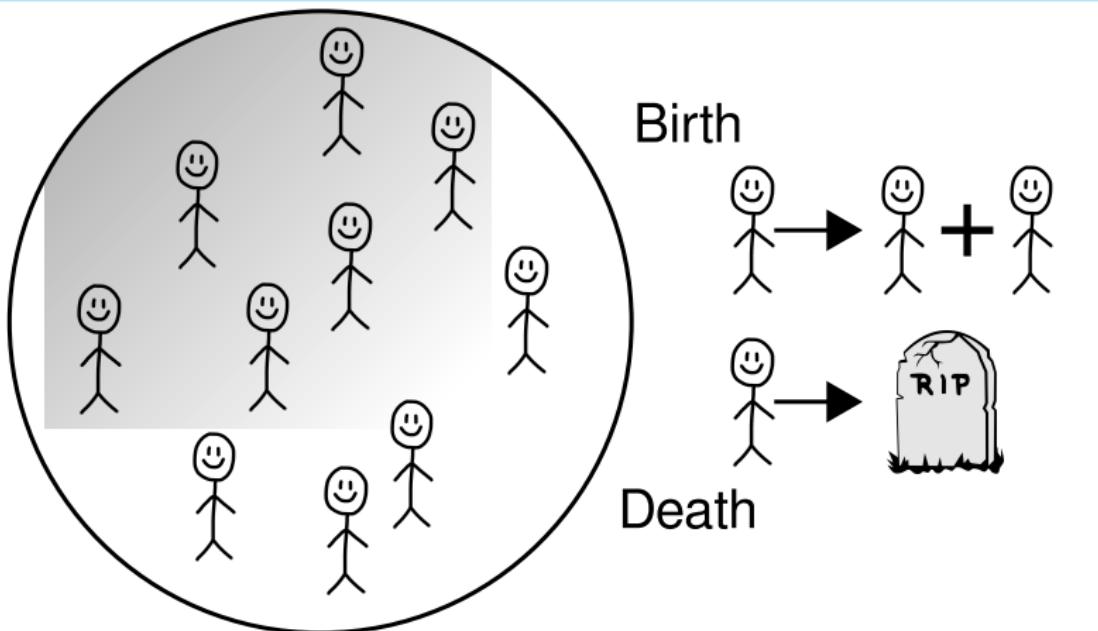
Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

## Example: Linear birth/death process

[Introduction](#)[Background theory](#)[MASTER](#)[MASTER population simulation examples](#)[Linear birth/death](#)[Stochastic logistic model](#)[Stochastic epidemic model](#)[Within-host virus dynamics](#)[Simulating trees](#)[Birth/death trees](#)[SEIR trees](#)[Moran trees](#)[Conditioning and Sampling](#)[Serial sampling](#)[Contemporaneous sampling](#)[Simulating coalescent trees](#)

# Example: Linear birth/death process

MASTER input file:

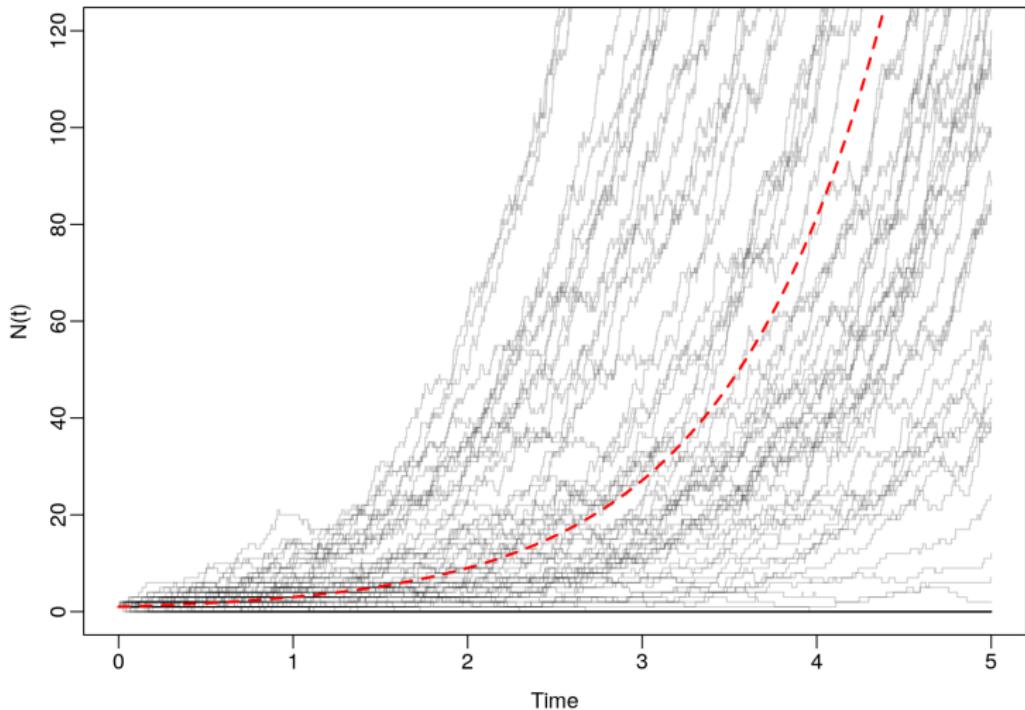
```
<beast version='2.0' namespace='...'>
  <run spec='Trajectory' simulationTime="5">
    <model spec='Model'>
      <population spec='Population' populationName='X' id='X' />
      <reaction spec='Reaction' reactionName='Birth' rate="2">
        X -> 2X
      </reaction>
      <reaction spec='Reaction' reactionName='Death' rate="0.9">
        X -> 0
      </reaction>
    </model>

    <initialState spec='InitState'>
      <populationSize spec='PopulationSize' population="@X" size='1' />
    </initialState>

    <output spec='JsonOutput' fileName='${filebase}_output.json' />
  </run>
</beast>
```

Introduction
Background theory
MASTER
MASTER population simulation examples
Linear birth/death
Stochastic logistic model
Stochastic epidemic model
Within-host virus dynamics
Simulating trees
Birth/death trees
SEIR trees
Moran trees
Conditioning and Sampling
Serial sampling
Contemporaneous sampling
Simulating coalescent trees

# Example: Linear birth/death process



Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

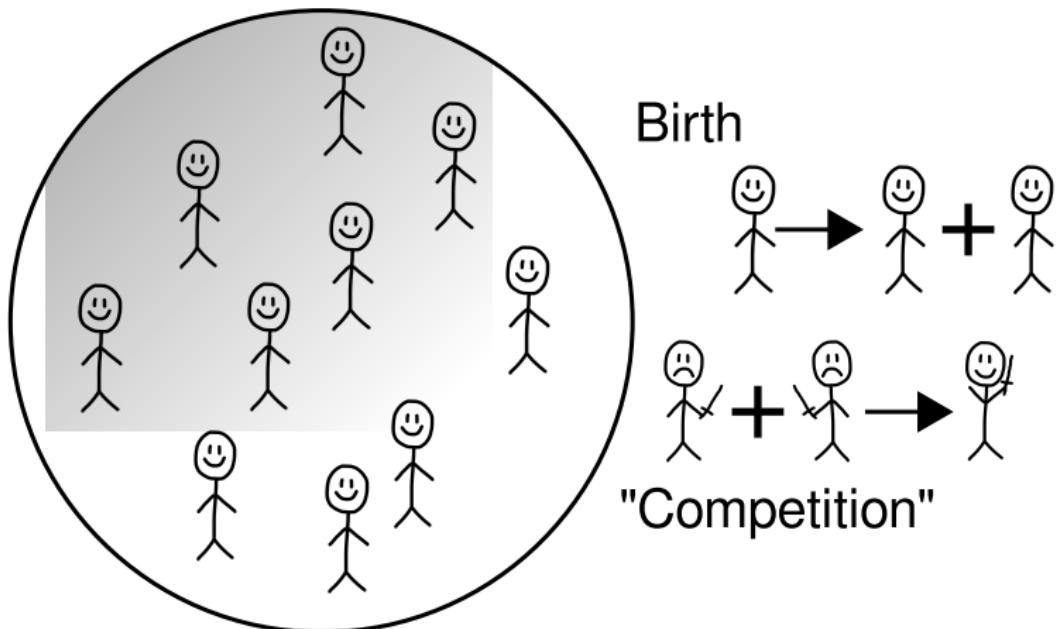
Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# Example: Stochastic logistic process



Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

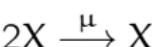
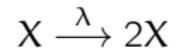
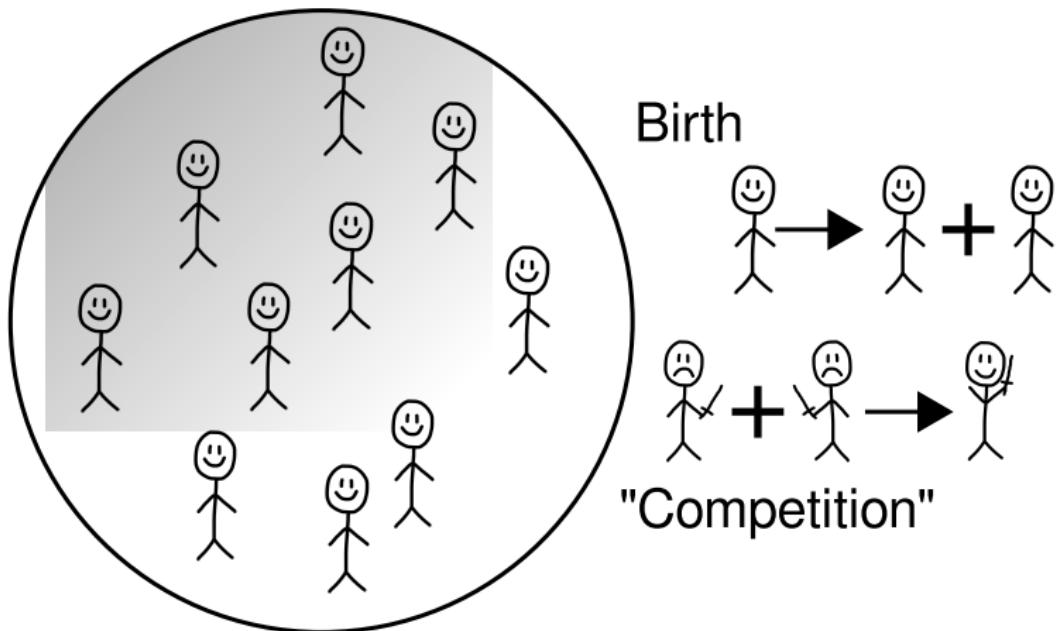
Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

## Example: Stochastic logistic process

[Introduction](#)[Background theory](#)[MASTER](#)[MASTER population simulation examples](#)[Linear birth/death](#)[Stochastic logistic model](#)[Stochastic epidemic model](#)[Within-host virus dynamics](#)[Simulating trees](#)[Birth/death trees](#)[SEIR trees](#)[Moran trees](#)[Conditioning and Sampling](#)[Serial sampling](#)[Contemporaneous sampling](#)[Simulating coalescent trees](#)

# Example: Stochastic logistic process

MASTER input file:

```
<beast version='2.0' namespace='...'>
  <run spec='Trajectory' simulationTime="5">
    <model spec='Model'>
      <population spec='Population' populationName='X' id='X' />
      <reaction spec='Reaction' reactionName='Birth' rate="5">
        X -> 2X
      </reaction>
      <reaction spec='Reaction' reactionName='Death' rate="0.1">
        2X -> X
      </reaction>
    </model>

    <initialState spec='InitState'>
      <populationSize spec='PopulationSize' population="@X" size='1' />
    </initialState>

    <output spec='JsonOutput' fileName='${filebase}_output.json' />
  </run>
</beast>
```

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

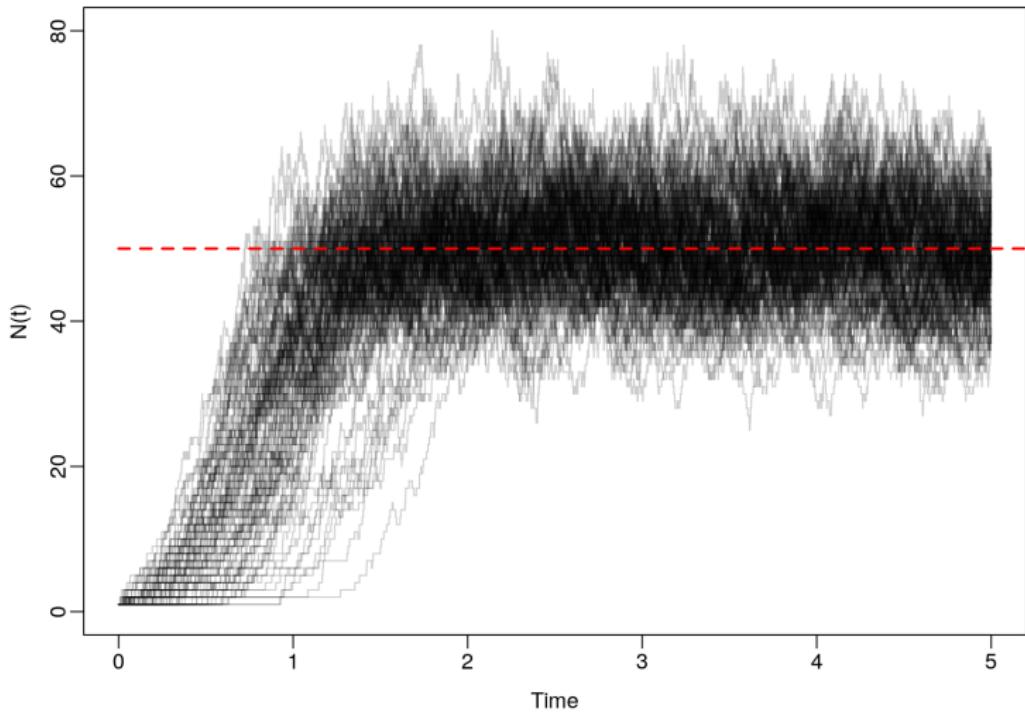
Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# Example: Stochastic logistic process



Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

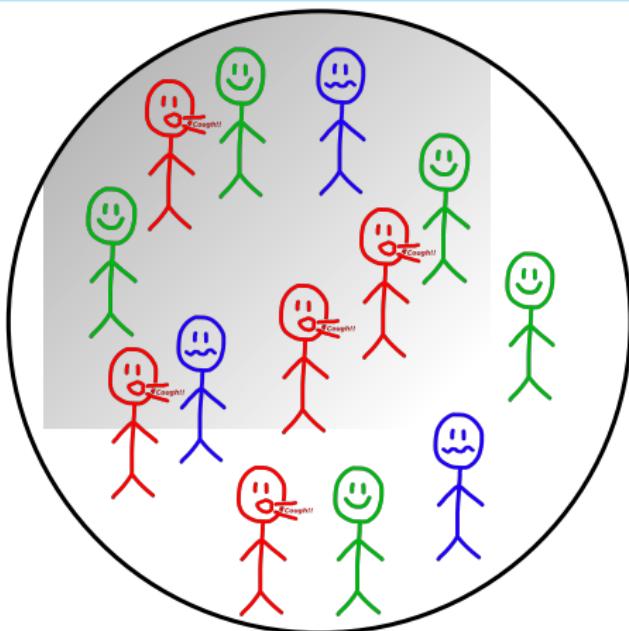
Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Example: SEIR model for epidemics



Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

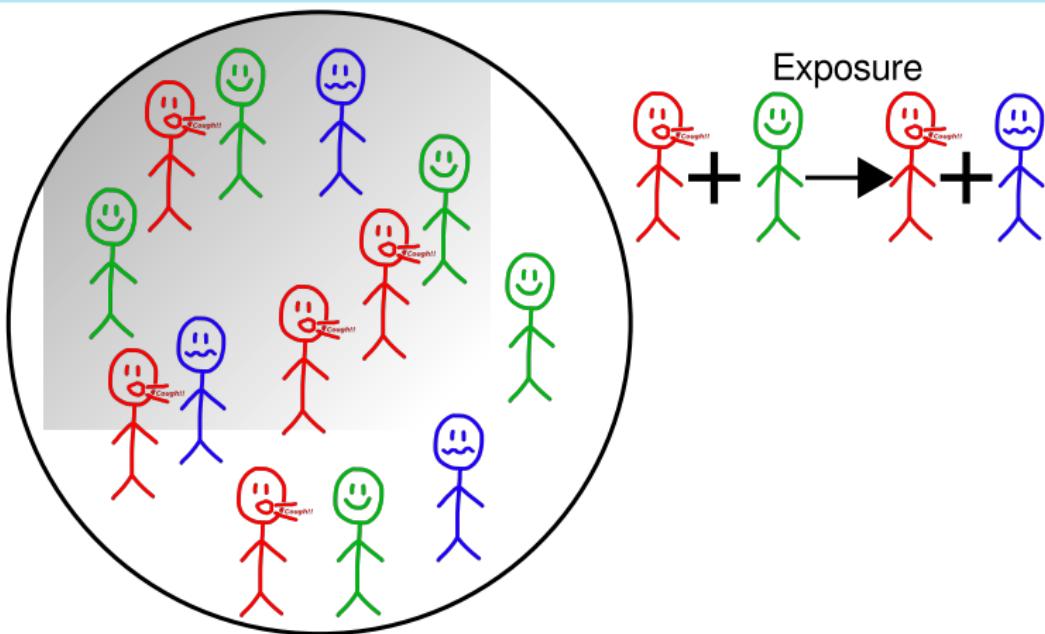
Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Example: SEIR model for epidemics



Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

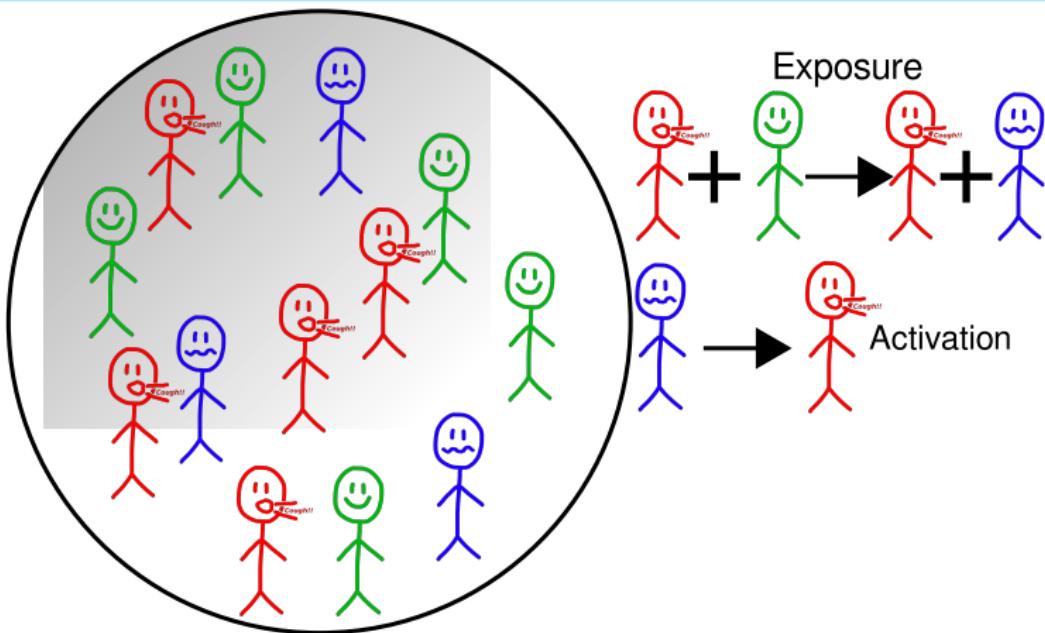
Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Example: SEIR model for epidemics



Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

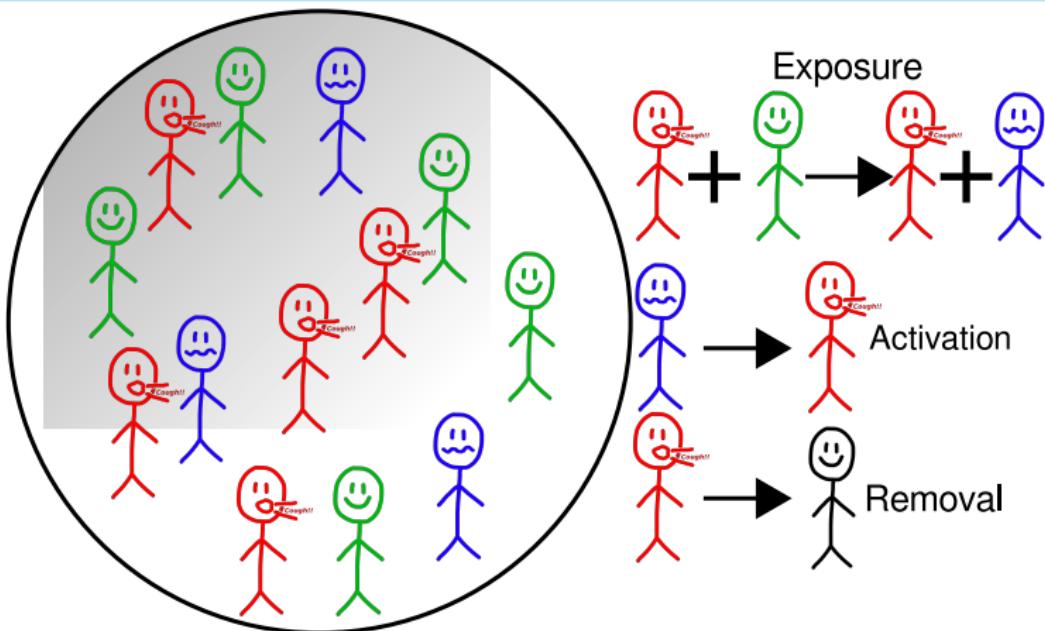
Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Example: SEIR model for epidemics



Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

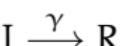
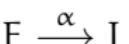
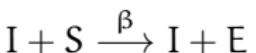
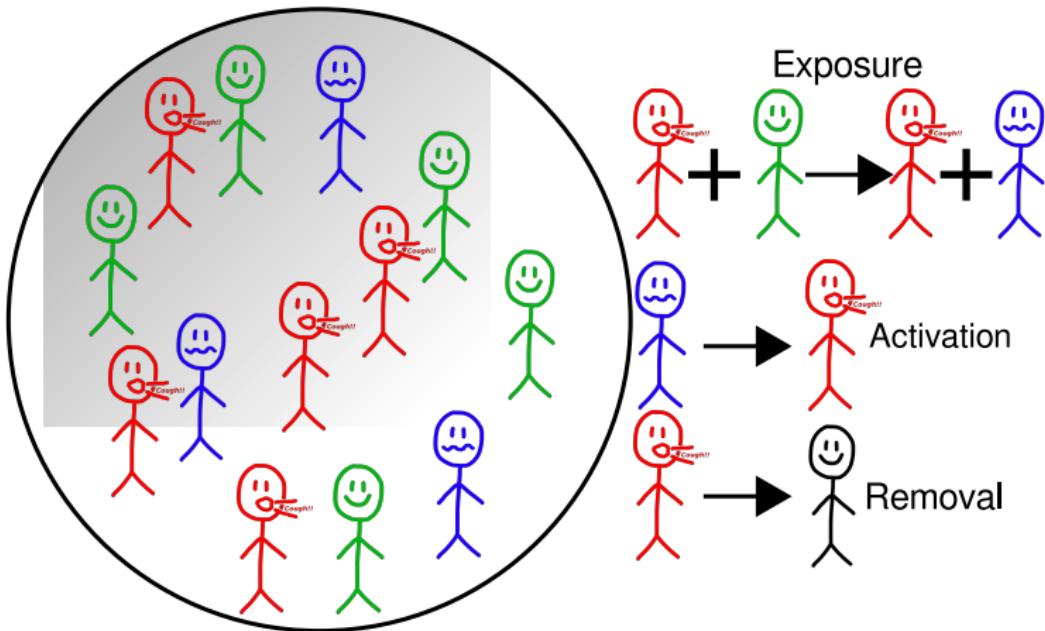
Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

## Example: SEIR model for epidemics



- Introduction
- Background theory
- MASTER
  - MASTER population simulation examples
    - Linear birth/death
    - Stochastic logistic model
    - Stochastic epidemic model
    - Within-host virus dynamics
  - Simulating trees
    - Birth/death trees
    - SEIR trees
    - Moran trees
  - Conditioning and Sampling
    - Serial sampling
    - Contemporaneous sampling
  - Simulating coalescent trees

# Example: SEIR model for epidemics

MASTER input file:

```
<model spec='Model'>
    <population spec='Population' populationName='S' id='S' />
    <population spec='Population' populationName='E' id='E' />
    <population spec='Population' populationName='I' id='I' />
    <population spec='Population' populationName='R' id='R' />
    <reaction spec='Reaction' reactionName='Exposure' rate="0.1">
        I + S -> I + E
    </reaction>
    <reaction spec='Reaction' reactionName='Activation' rate="1">
        E -> I
    </reaction>
    <reaction spec='Reaction' reactionName='Removal' rate="1">
        I -> R
    </reaction>
</model>
<initialState spec='InitState'>
    <populationSize spec='PopulationSize' population="@S" size='99' />
    <populationSize spec='PopulationSize' population="@I" size='1' />
</initialState>
```

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

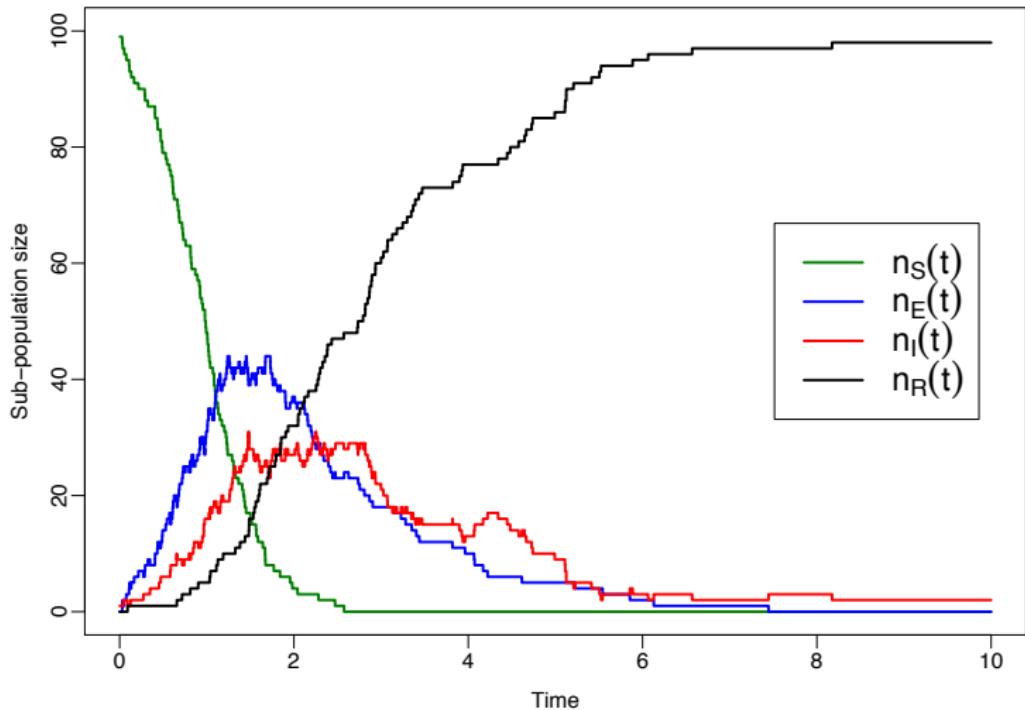
Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

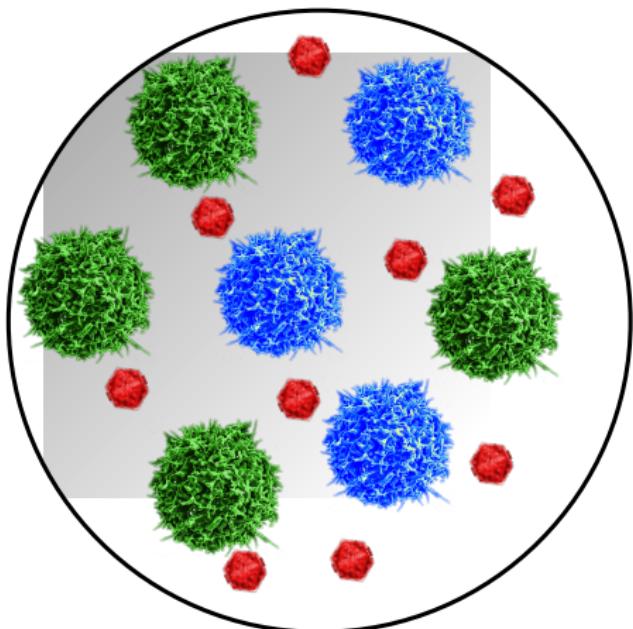
Simulating coalescent  
trees

# Example: SEIR model for epidemics

[Introduction](#)[Background theory](#)[MASTER](#)[MASTER population simulation examples](#)[Linear birth/death](#)[Stochastic logistic model](#)[Stochastic epidemic model](#)[Within-host virus dynamics](#)[Simulating trees](#)[Birth/death trees](#)[SEIR trees](#)[Moran trees](#)[Conditioning and Sampling](#)[Serial sampling](#)[Contemporaneous sampling](#)[Simulating coalescent trees](#)

# Example: Within-host HIV dynamics

Simulations and  
Phylogeography



Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

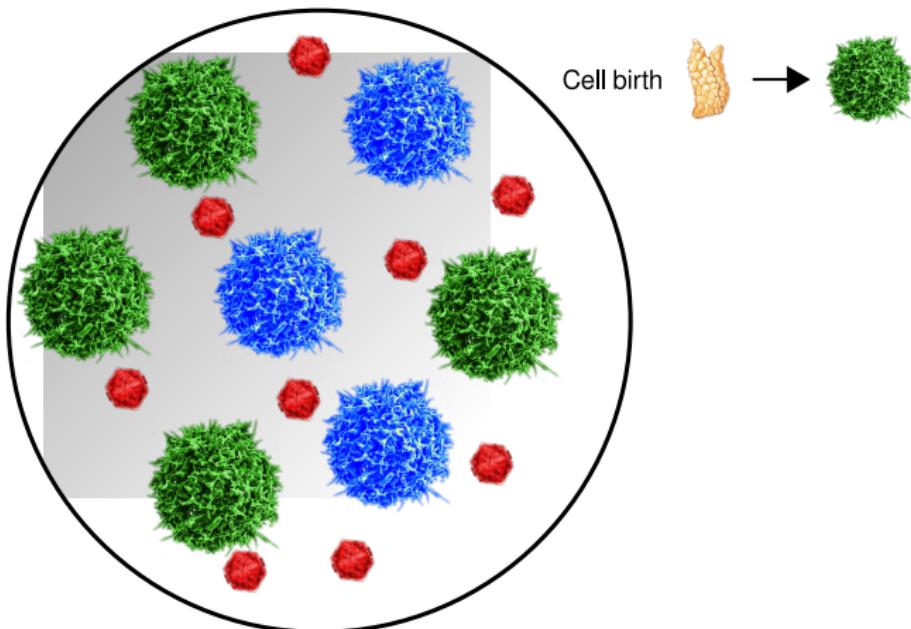
Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# Example: Within-host HIV dynamics

Simulations and  
Phylogeography



Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

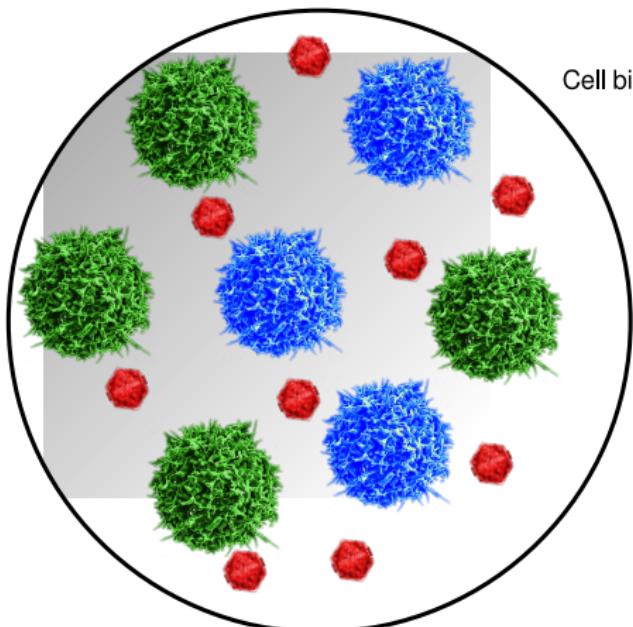
Conditioning and Sampling

Serial sampling

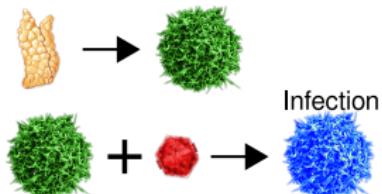
Contemporaneous sampling

Simulating coalescent trees

# Example: Within-host HIV dynamics



Cell birth



Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

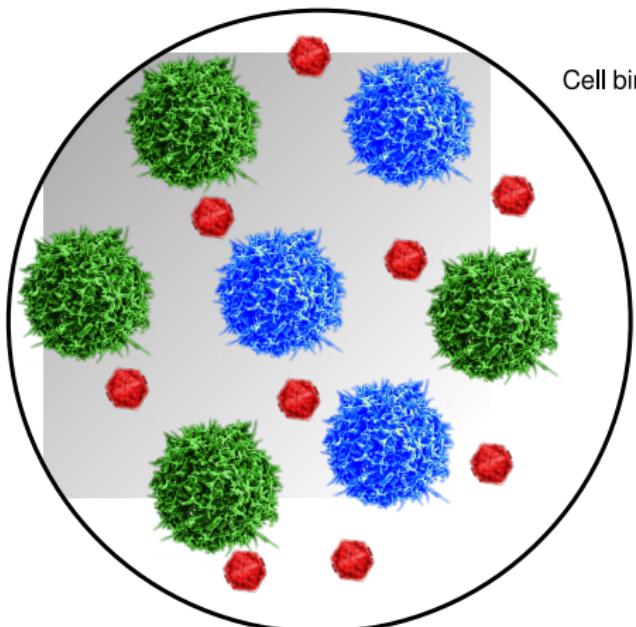
Conditioning and Sampling

Serial sampling

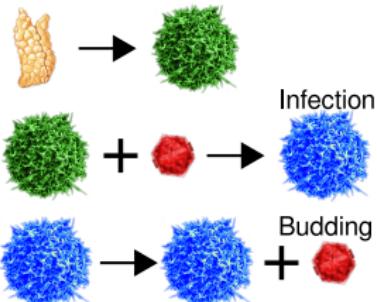
Contemporaneous sampling

Simulating coalescent trees

# Example: Within-host HIV dynamics



Cell birth



Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

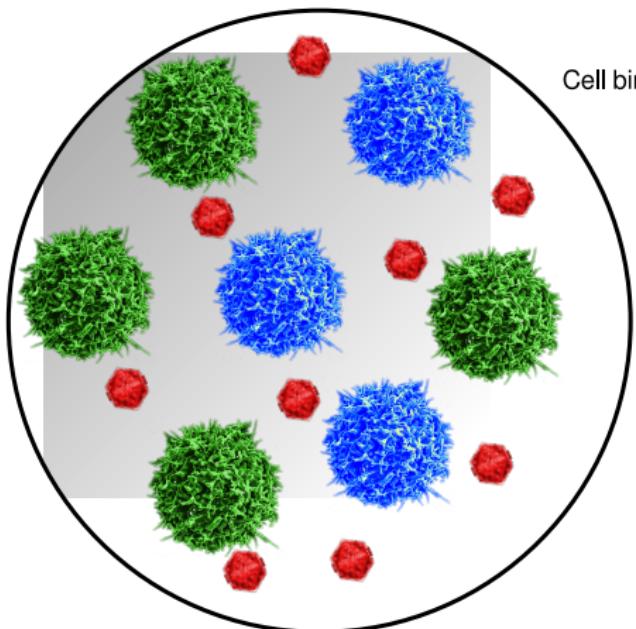
Conditioning and Sampling

Serial sampling

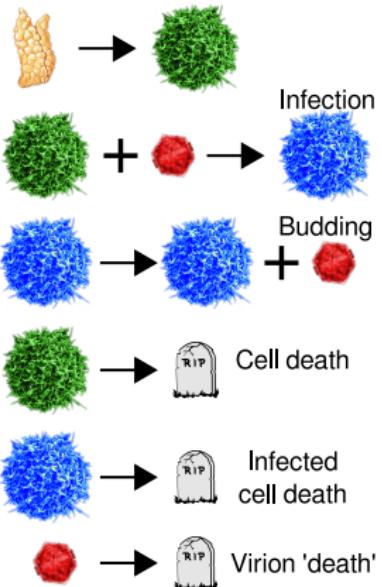
Contemporaneous sampling

Simulating coalescent trees

# Example: Within-host HIV dynamics



Cell birth



Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Example: Within-host HIV dynamics

MASTER input file:

```
<stepper spec='TauLeapingStepper' stepSize='0.01'>

<model spec='Model' id='model'>
    <population spec='Population' id='X' populationName='X' />
    ...
    <reaction spec='Reaction' reactionName="CellBirth" rate="2.5e8">
        0 -> X
    </reaction>
    <reaction spec='Reaction' reactionName="Infection" rate="5e-13">
        X + V -> Y
    </reaction>
    ...
</model>

<initialState spec='InitState'>
    <populationSize spec='PopulationSize' population='@X' size='2.5e11' />
    <populationSize spec='PopulationSize' population='@V' size='10' />
</initialState>
```

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

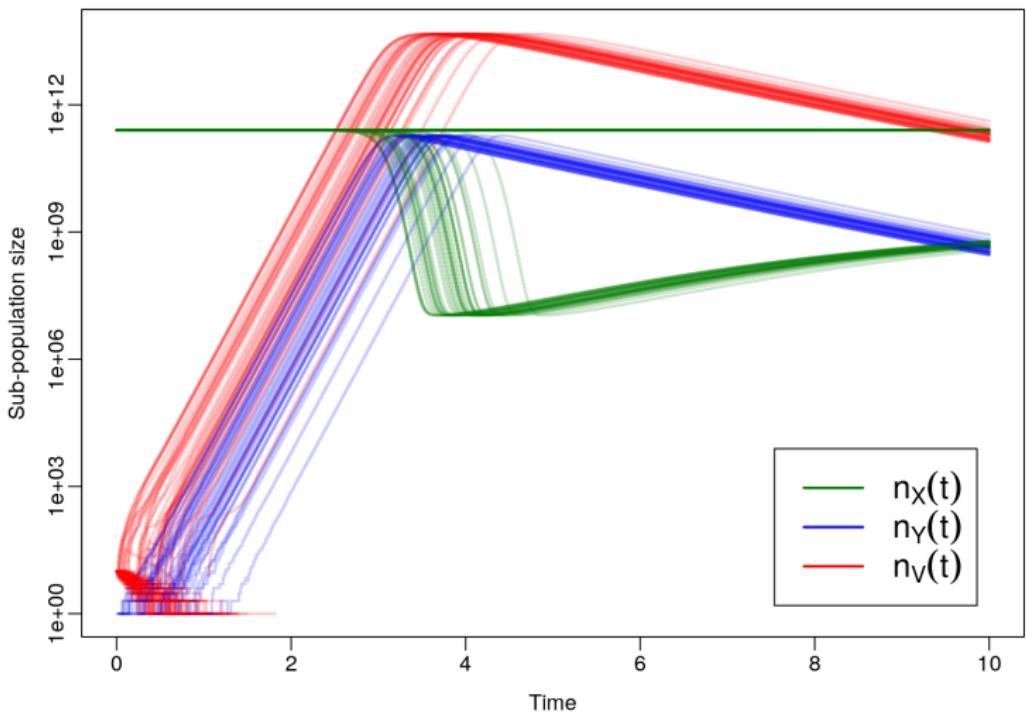
Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# Example: Within-host HIV dynamics



- Introduction
- Background theory
- MASTER
- MASTER population simulation examples
  - Linear birth/death
  - Stochastic logistic model
  - Stochastic epidemic model
  - Within-host virus dynamics
- Simulating trees
  - Birth/death trees
  - SEIR trees
  - Moran trees
- Conditioning and Sampling
  - Serial sampling
  - Contemporaneous sampling
- Simulating coalescent trees

# Continuous time branching processes: Trees!

- ▶ Continuous branching processes describe Markov stochastic population models where the parent of each individual is uniquely identified.

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

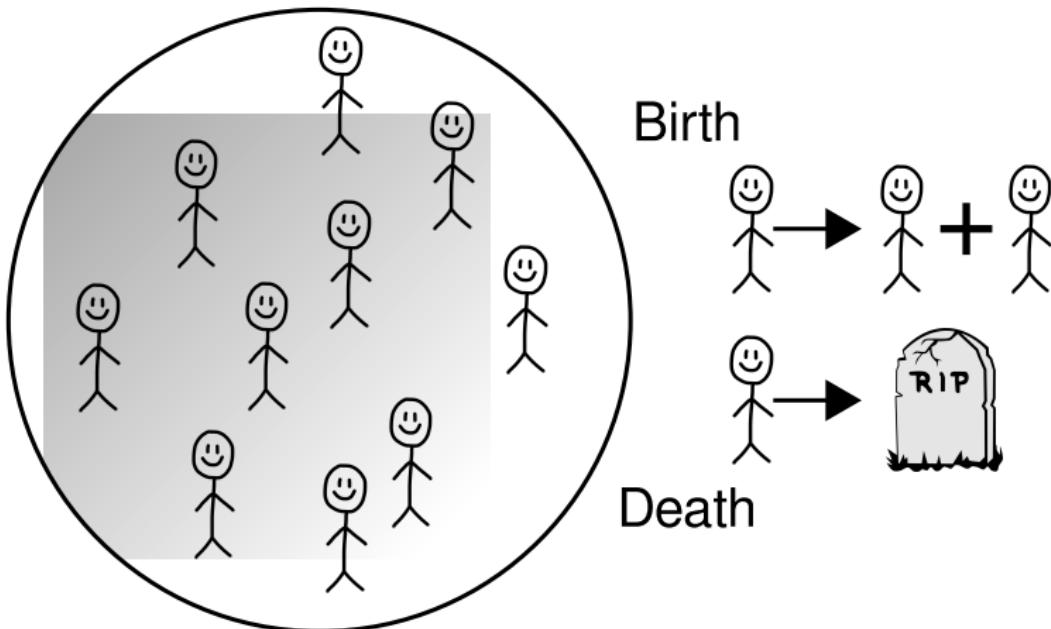
Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Continuous time branching processes: Trees!

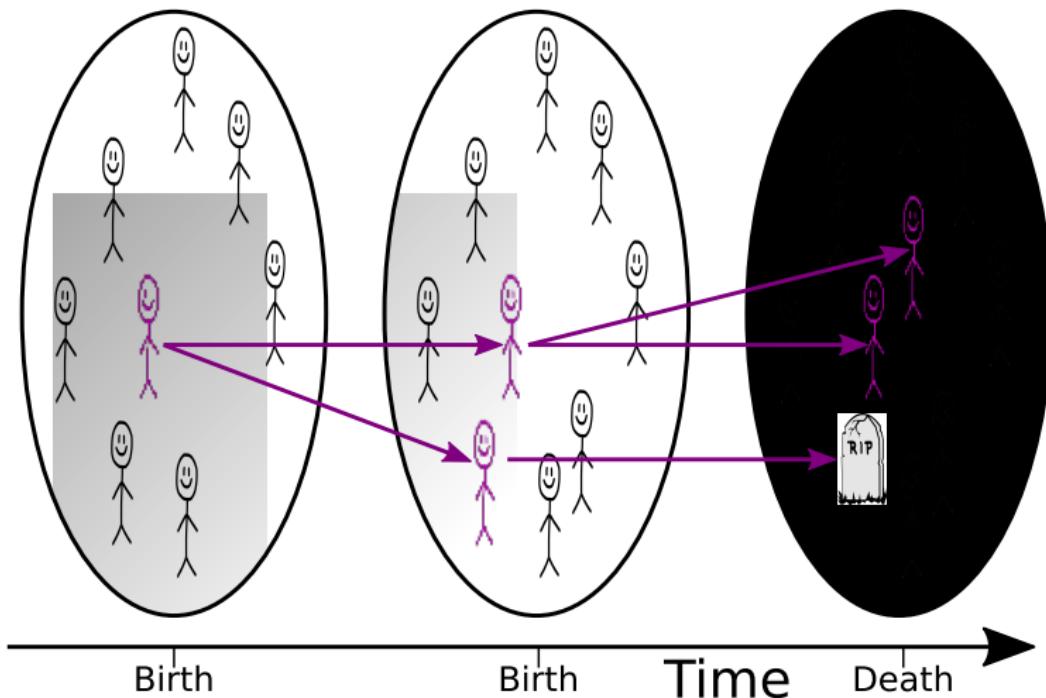
- ▶ Continuous branching processes describe Markov stochastic population models where the parent of each individual is uniquely identified.



Introduction  
Background theory  
MASTER  
MASTER population simulation examples  
Linear birth/death  
Stochastic logistic model  
Stochastic epidemic model  
Within-host virus dynamics  
Simulating trees  
Birth/death trees  
SEIR trees  
Moran trees  
Conditioning and Sampling  
Serial sampling  
Contemporaneous sampling  
Simulating coalescent trees

# Continuous time branching processes: Trees!

- ▶ Continuous branching processes describe Markov stochastic population models where the parent of each individual is uniquely identified.



Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

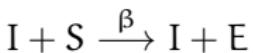
Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Inheritance reaction notation

Simulations and  
Phylogeography



Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

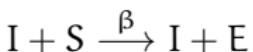
Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees



## Problem

Chemical reaction notation is ambiguous with respect to parent-child relationships. (Not really useful in chemistry so no surprise!)

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

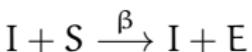
Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees



## Problem

Chemical reaction notation is ambiguous with respect to parent-child relationships. (Not really useful in chemistry so no surprise!)

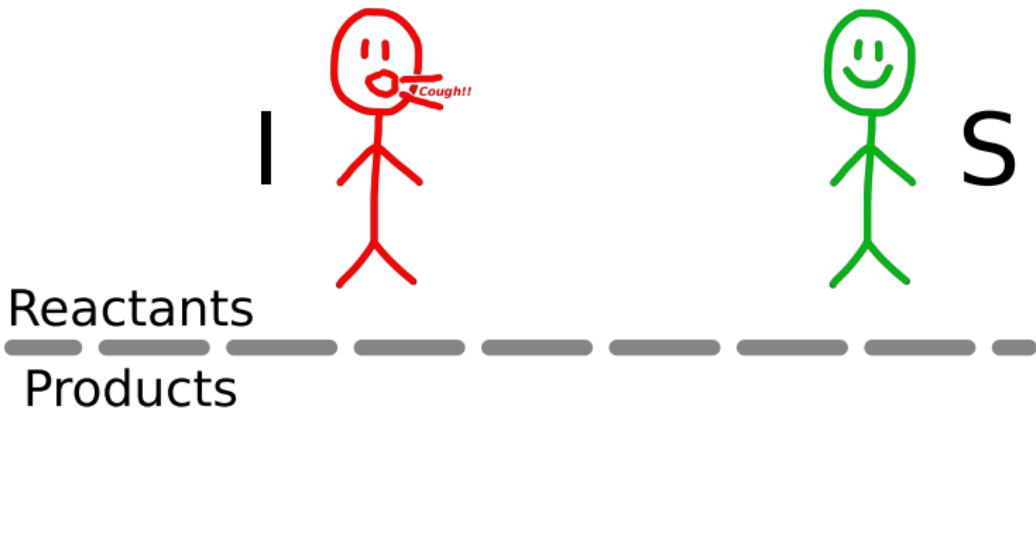
## Question

How can we express parent-child (inheritance) relationships in chemical reaction notation?

Reactants  
  
Products

- Introduction
- Background theory
- MASTER
  - MASTER population simulation examples
    - Linear birth/death
    - Stochastic logistic model
    - Stochastic epidemic model
    - Within-host virus dynamics
  - Simulating trees
    - Birth/death trees
    - SEIR trees
    - Moran trees
  - Conditioning and Sampling
    - Serial sampling
    - Contemporaneous sampling
  - Simulating coalescent trees

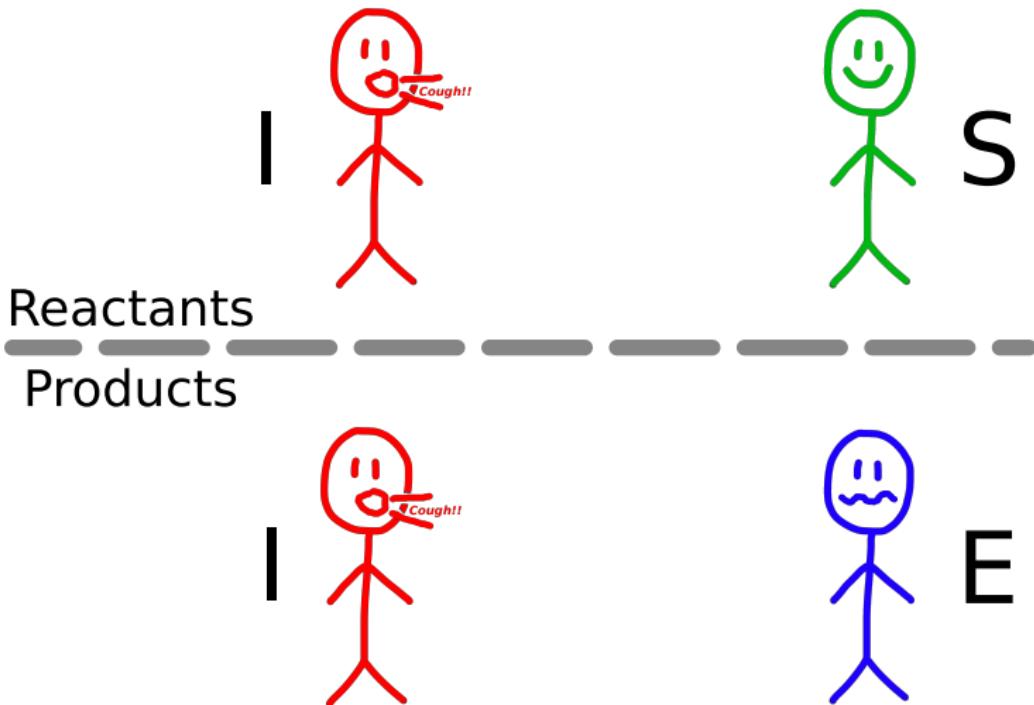
# Inheritance reaction notation



- Introduction
- Background theory
- MASTER
- MASTER population simulation examples
  - Linear birth/death
  - Stochastic logistic model
  - Stochastic epidemic model
  - Within-host virus dynamics
- Simulating trees
  - Birth/death trees
  - SEIR trees
  - Moran trees
- Conditioning and Sampling
  - Serial sampling
  - Contemporaneous sampling
- Simulating coalescent trees

# Inheritance reaction notation

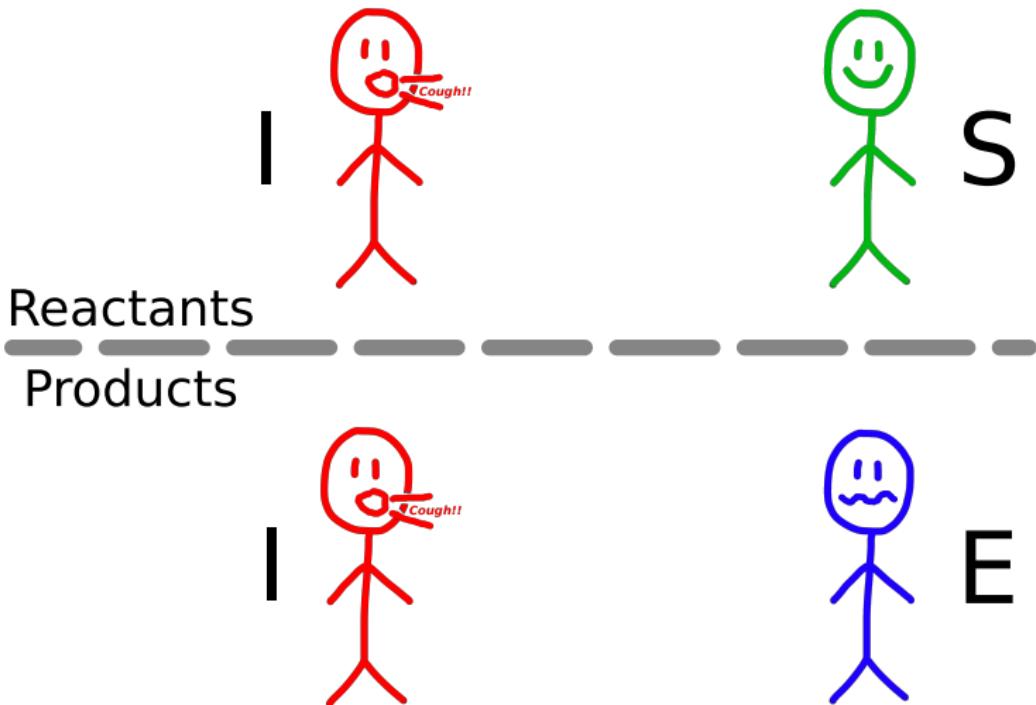
Simulations and  
Phylogeography



- Introduction
- Background theory
- MASTER
- MASTER population simulation examples
  - Linear birth/death
  - Stochastic logistic model
  - Stochastic epidemic model
  - Within-host virus dynamics
- Simulating trees
  - Birth/death trees
  - SEIR trees
  - Moran trees
- Conditioning and Sampling
  - Serial sampling
  - Contemporaneous sampling
- Simulating coalescent trees

# Inheritance reaction notation

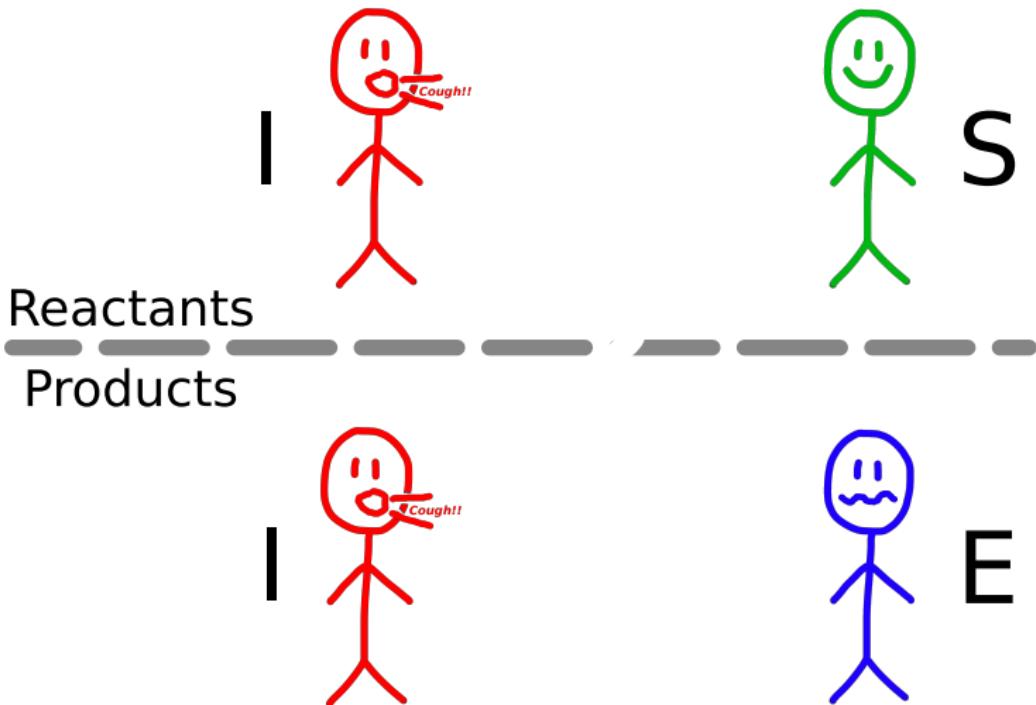
Simulations and  
Phylogeography



- Introduction
- Background theory
- MASTER
- MASTER population simulation examples
  - Linear birth/death
  - Stochastic logistic model
  - Stochastic epidemic model
  - Within-host virus dynamics
- Simulating trees
  - Birth/death trees
  - SEIR trees
  - Moran trees
- Conditioning and Sampling
  - Serial sampling
  - Contemporaneous sampling
- Simulating coalescent trees

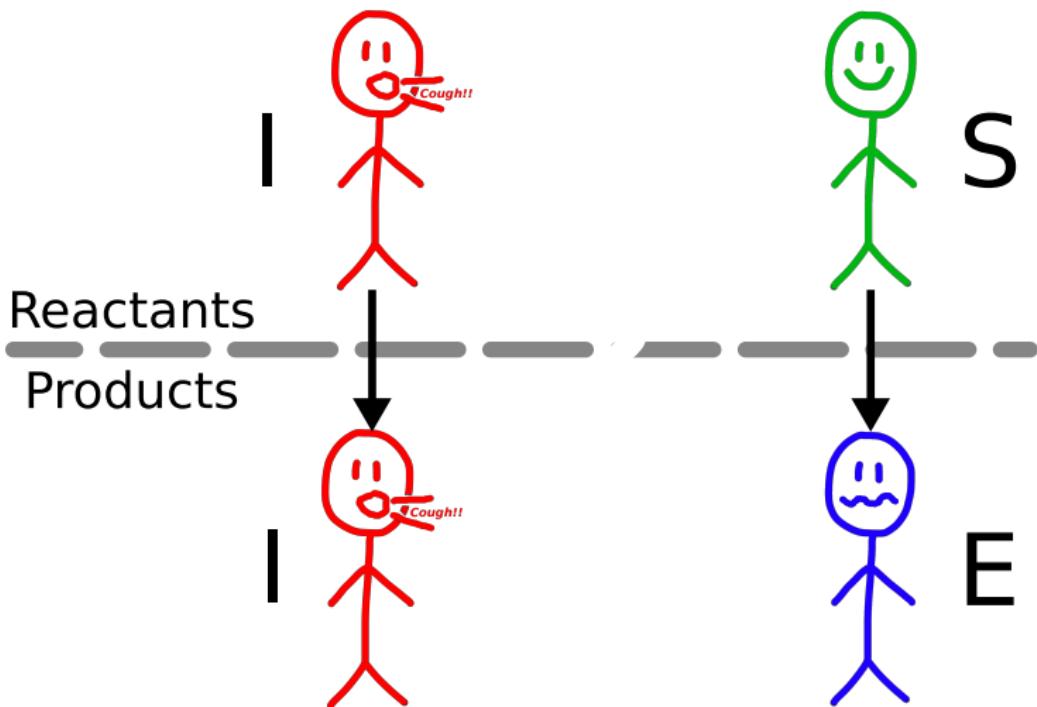
# Inheritance reaction notation

Simulations and  
Phylogeography



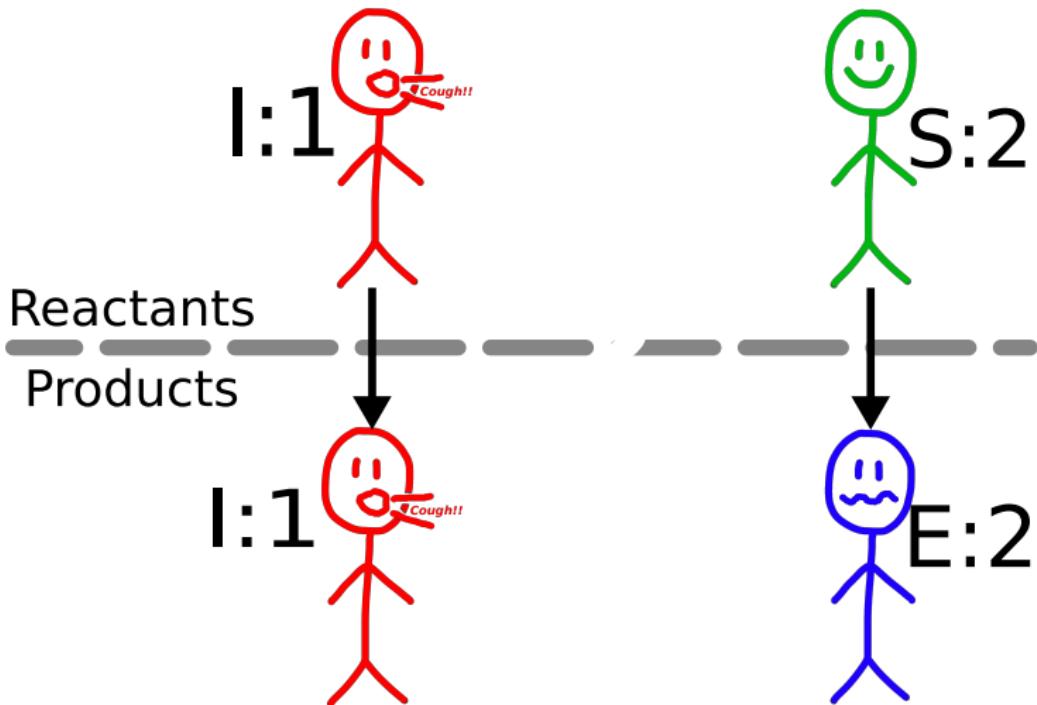
- Introduction
- Background theory
- MASTER
- MASTER population simulation examples
  - Linear birth/death
  - Stochastic logistic model
  - Stochastic epidemic model
  - Within-host virus dynamics
- Simulating trees
  - Birth/death trees
  - SEIR trees
  - Moran trees
- Conditioning and Sampling
  - Serial sampling
  - Contemporaneous sampling
- Simulating coalescent trees

# Inheritance reaction notation



- Introduction
- Background theory
- MASTER
- MASTER population simulation examples
  - Linear birth/death
  - Stochastic logistic model
  - Stochastic epidemic model
  - Within-host virus dynamics
- Simulating trees
  - Birth/death trees
  - SEIR trees
  - Moran trees
- Conditioning and Sampling
  - Serial sampling
  - Contemporaneous sampling
- Simulating coalescent trees

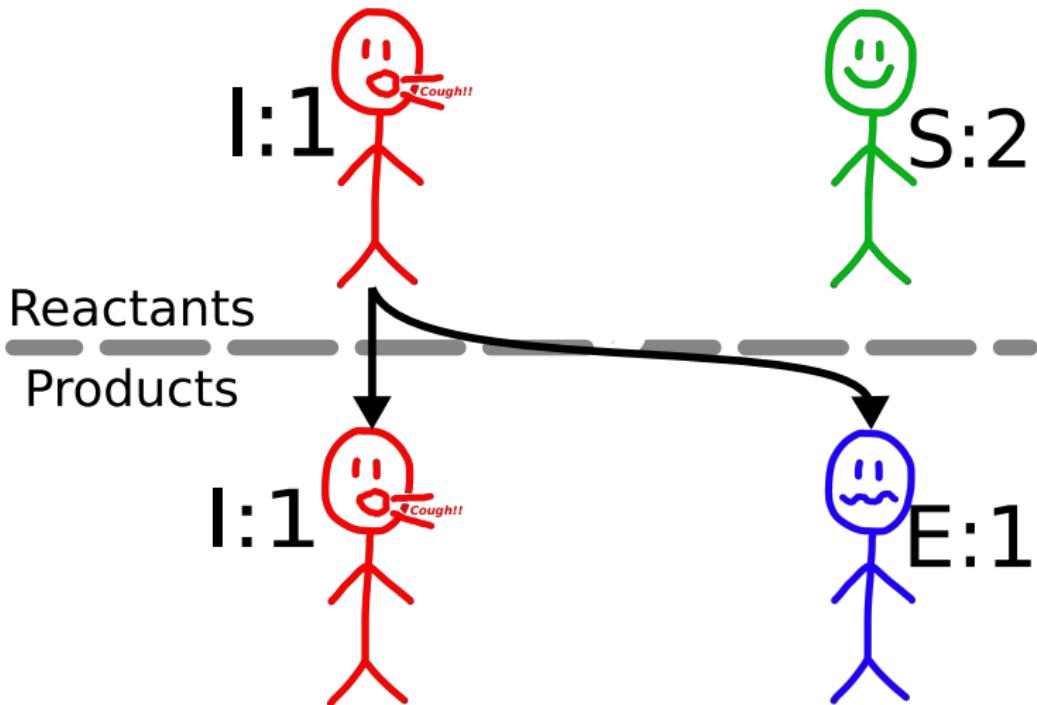
# Inheritance reaction notation



- Introduction
- Background theory
- MASTER
  - MASTER population simulation examples
    - Linear birth/death
    - Stochastic logistic model
    - Stochastic epidemic model
    - Within-host virus dynamics
  - Simulating trees
    - Birth/death trees
    - SEIR trees
    - Moran trees
  - Conditioning and Sampling
    - Serial sampling
    - Contemporaneous sampling
  - Simulating coalescent trees

# Inheritance reaction notation

Simulations and  
Phylogeography



- Introduction
- Background theory
- MASTER
- MASTER population simulation examples
  - Linear birth/death
  - Stochastic logistic model
  - Stochastic epidemic model
  - Within-host virus dynamics
- Simulating trees
  - Birth/death trees
  - SEIR trees
  - Moran trees
- Conditioning and Sampling
  - Serial sampling
  - Contemporaneous sampling
- Simulating coalescent trees

# Example: birth/death trees

MASTER input file:

```
<beast version='2.0' namespace='...'>
<run spec='InheritanceTrajectory' simulationTime="5.0"
      samplePopulationSizes="true">
  <model spec='Model'>
    <population spec='Population' populationName='X' id='X' />
    <reaction spec='Reaction' reactionName='Birth' rate="2">
      X -> 2X
    </reaction>
    <reaction spec='Reaction' reactionName='Death' rate="0.9">
      X -> 0
    </reaction>
  </model>

  <initialState spec='InitState'>
    <lineageSeed spec='Individual' population="@X"/>
  </initialState>

  <output spec='JsonOutput' fileName='${filebase}_output.json' />
  <output spec='NexusOutput' fileName='${filebase}_output.nexus' />
</run>
```

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

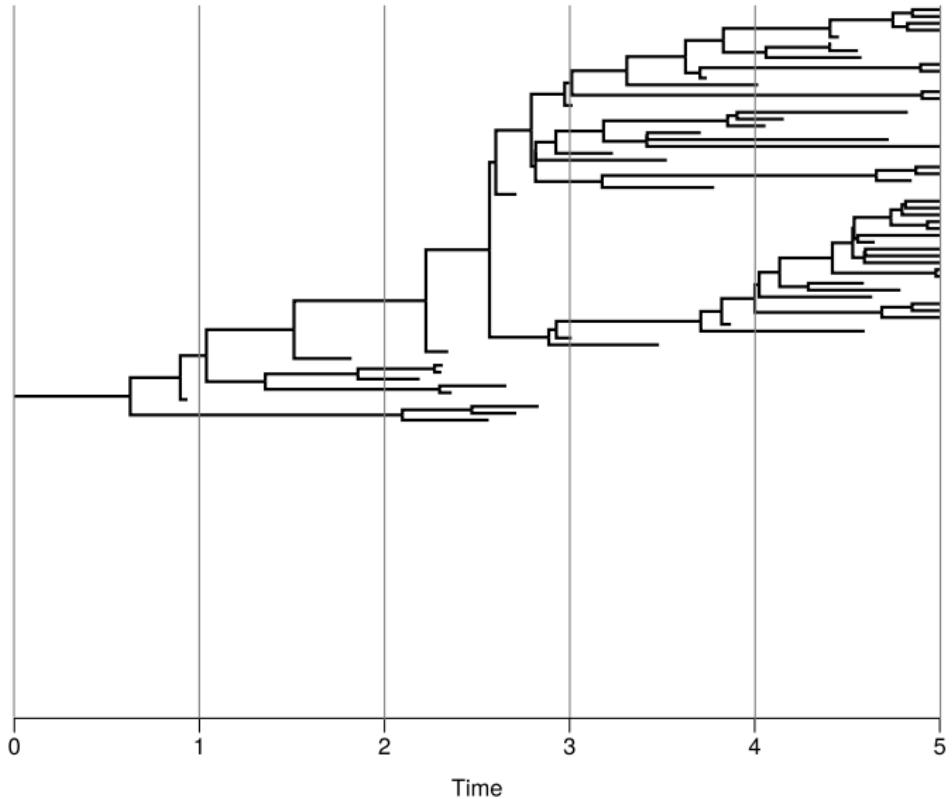
Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

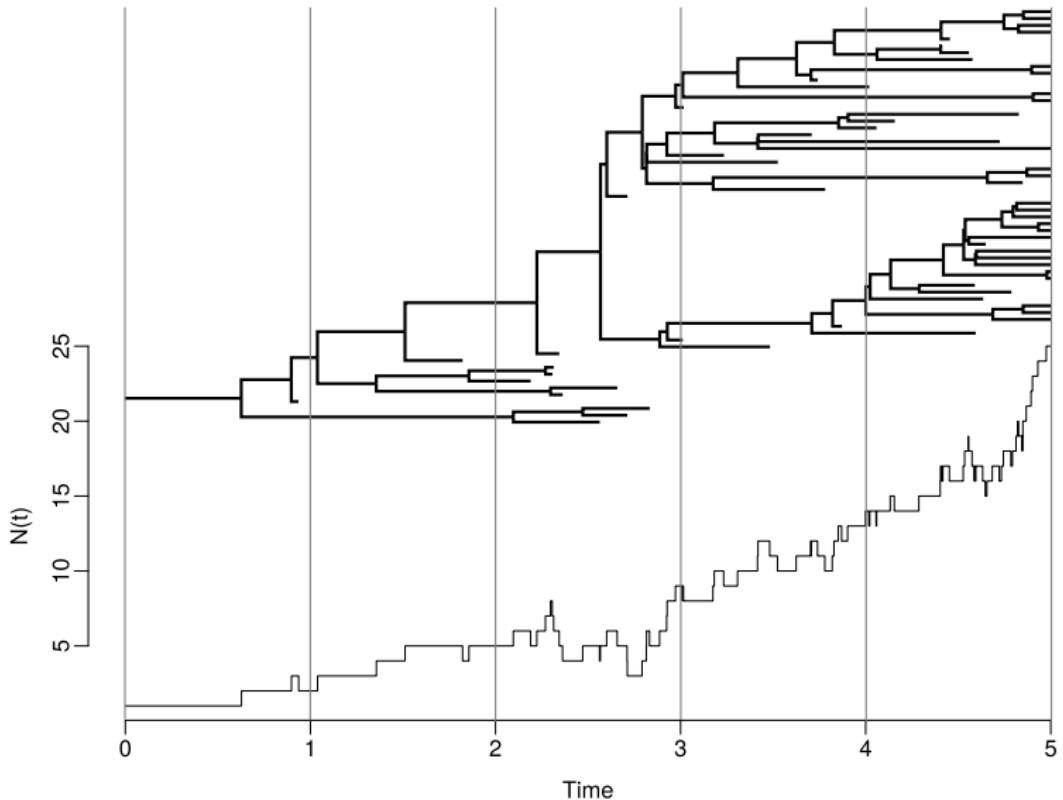
Simulating coalescent  
trees

# Example: birth/death trees



- Introduction
- Background theory
- MASTER
- MASTER population simulation examples
  - Linear birth/death
  - Stochastic logistic model
  - Stochastic epidemic model
  - Within-host virus dynamics
- Simulating trees
  - Birth/death trees
  - SEIR trees
  - Moran trees
- Conditioning and Sampling
  - Serial sampling
  - Contemporaneous sampling
- Simulating coalescent trees

# Example: birth/death trees



- Introduction
- Background theory
- MASTER
- MASTER population simulation examples
  - Linear birth/death
  - Stochastic logistic model
  - Stochastic epidemic model
  - Within-host virus dynamics
- Simulating trees
  - Birth/death trees
  - SEIR trees
  - Moran trees
- Conditioning and Sampling
  - Serial sampling
  - Contemporaneous sampling
- Simulating coalescent trees

# Example: SEIR trees

MASTER input file:

```
<model spec='Model'>
    <population spec='Population' populationName='S' id='S' />
    <population spec='Population' populationName='E' id='E' />
    <population spec='Population' populationName='I' id='I' />
    <population spec='Population' populationName='R' id='R' />
    <reaction spec='Reaction' reactionName='Exposure' rate="0.1">
        I:1 + S:2 -> I:1 + E:1
    </reaction>
    <reaction spec='Reaction' reactionName='Activation' rate="1">
        E:1 -> I:1
    </reaction>
    <reaction spec='Reaction' reactionName='Removal' rate="1">
        I:1 -> R:2
    </reaction>
</model>

<initialState spec='InitState'>
    <populationSize spec='PopulationSize' population="@S" size='99' />
    <lineageSeed spec='Individual' population="@I" />
</initialState>
```

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

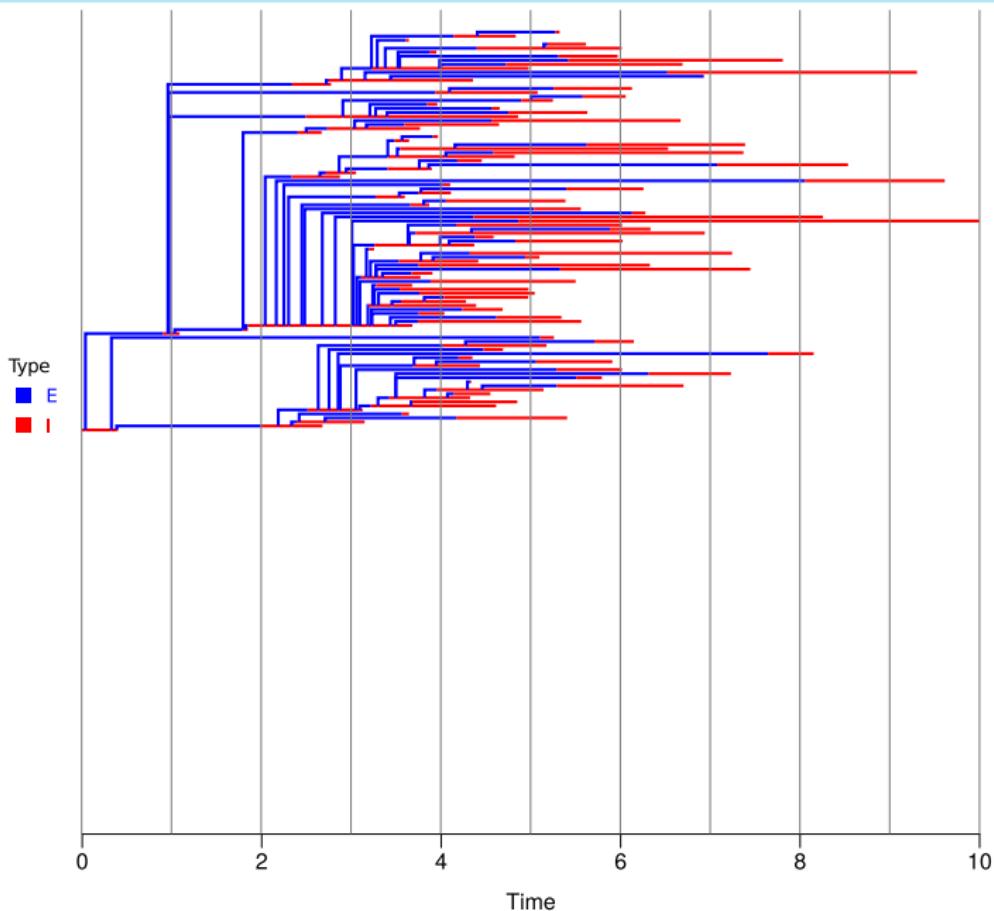
Conditioning and  
Sampling

Serial sampling

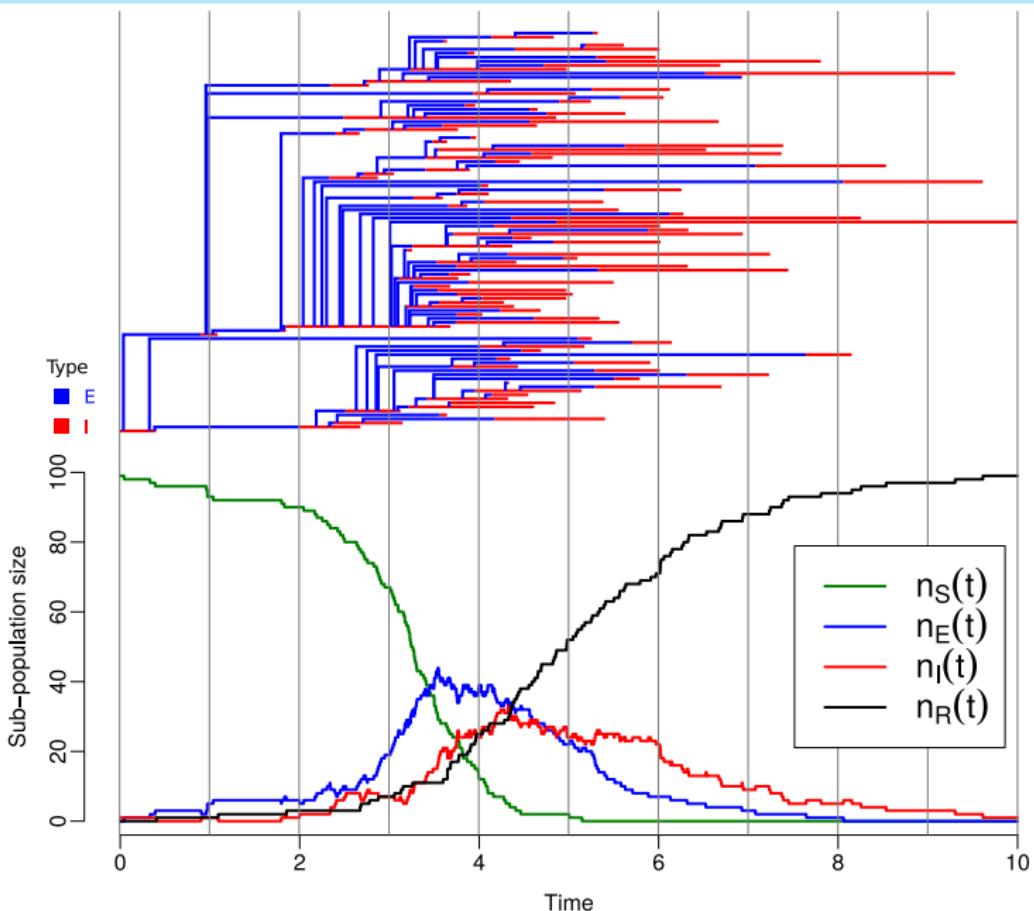
Contemporaneous sampling

Simulating coalescent  
trees

# Example: SEIR trees

[Introduction](#)[Background theory](#)[MASTER](#)[MASTER population simulation examples](#)[Linear birth/death](#)[Stochastic logistic model](#)[Stochastic epidemic model](#)[Within-host virus dynamics](#)[Simulating trees](#)[Birth/death trees](#)[SEIR trees](#)[Moran trees](#)[Conditioning and Sampling](#)[Serial sampling](#)[Contemporaneous sampling](#)[Simulating coalescent trees](#)

## Example: SEIR trees

[Introduction](#)[Background theory](#)[MASTER](#)[MASTER population simulation examples](#)[Linear birth/death](#)[Stochastic logistic model](#)[Stochastic epidemic model](#)[Within-host virus dynamics](#)[Simulating trees](#)[Birth/death trees](#)[SEIR trees](#)[Moran trees](#)[Conditioning and Sampling](#)[Serial sampling](#)[Contemporaneous sampling](#)[Simulating coalescent trees](#)

# Example: Moran trees

- ▶ Continuous time Moran model is a standard model of population genetics (continuous time analogue of the Wright-Fisher model).

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Example: Moran trees

- ▶ Continuous time Moran model is a standard model of population genetics (continuous time analogue of the Wright-Fisher model).
- ▶ Preserves total population number.

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

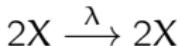
Serial sampling

Contemporaneous sampling

Simulating coalescent trees

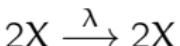
# Example: Moran trees

- ▶ Continuous time Moran model is a standard model of population genetics (continuous time analogue of the Wright-Fisher model).
- ▶ Preserves total population number.
- ▶ Without genetic component, chemical reaction description looks very strange:

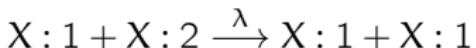
[Introduction](#)[Background theory](#)[MASTER](#)[MASTER population simulation examples](#)[Linear birth/death](#)[Stochastic logistic model](#)[Stochastic epidemic model](#)[Within-host virus dynamics](#)[Simulating trees](#)[Birth/death trees](#)[SEIR trees](#)[Moran trees](#)[Conditioning and Sampling](#)[Serial sampling](#)[Contemporaneous sampling](#)[Simulating coalescent trees](#)

# Example: Moran trees

- ▶ Continuous time Moran model is a standard model of population genetics (continuous time analogue of the Wright-Fisher model).
- ▶ Preserves total population number.
- ▶ Without genetic component, chemical reaction description looks very strange:



- ▶ Considering parent-child relationships however, this becomes useful:



(Simultaneous birth and death.)

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Example: Moran trees

MASTER input file:

```
<model spec='Model'>
  <population spec='Population' populationName='X' id='X' />
  <reaction spec='Reaction' reactionName='BirthDeath' rate="1.7">
    X:1 + X:2 -> 2X:1
  </reaction>
</model>

<initialState spec='InitState'>
  <populationSize spec='PopulationSize' population="@X" size='99' />
  <lineageSeed spec='Individual' population="@X" />
</initialState>

<output spec='NexusOutput' fileName='${filebase}_output.nexus' />
<output spec='JsonOutput' fileName='${filebase}_output.json' />
```

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

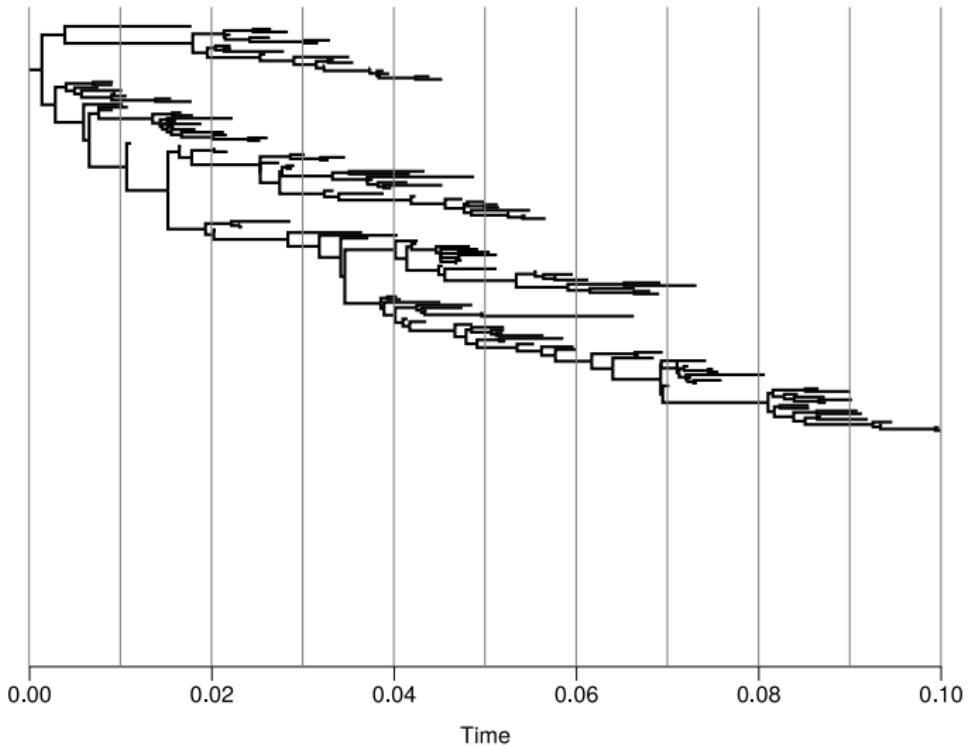
Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

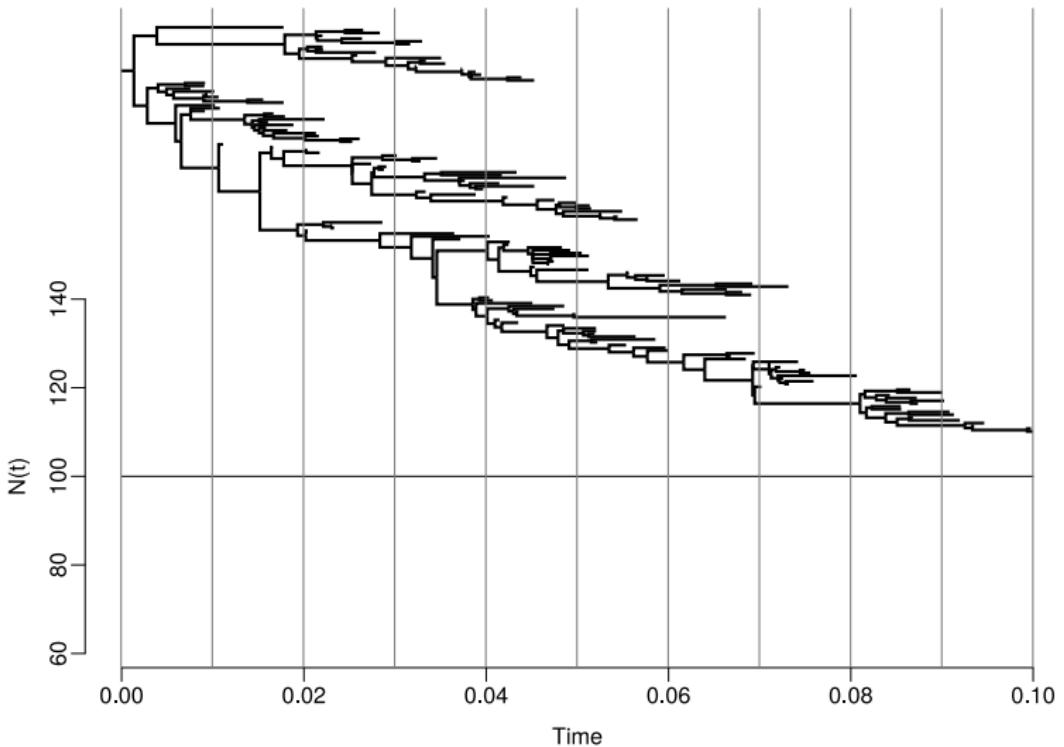
Simulating coalescent  
trees

# Example: Moran trees



- Introduction
- Background theory
- MASTER
- MASTER population simulation examples
  - Linear birth/death
  - Stochastic logistic model
  - Stochastic epidemic model
  - Within-host virus dynamics
- Simulating trees
  - Birth/death trees
  - SEIR trees
  - Moran trees
- Conditioning and Sampling
  - Serial sampling
  - Contemporaneous sampling
- Simulating coalescent trees

# Example: Moran trees



- Introduction
- Background theory
- MASTER
- MASTER population simulation examples
  - Linear birth/death
  - Stochastic logistic model
  - Stochastic epidemic model
  - Within-host virus dynamics
- Simulating trees
  - Birth/death trees
  - SEIR trees
  - Moran trees
- Conditioning and Sampling
  - Serial sampling
  - Contemporaneous sampling
- Simulating coalescent trees

# Conditioning and Sampling

- ▶ Often want to condition on some predicate:

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# Conditioning and Sampling

- ▶ Often want to condition on some predicate:
  - ▶ Population must not go extinct before end of simulation.

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# Conditioning and Sampling

- ▶ Often want to condition on some predicate:
  - ▶ Population must not go extinct before end of simulation.
  - ▶ At least 10 individuals must remain.

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# Conditioning and Sampling

- ▶ Often want to condition on some predicate:
  - ▶ Population must not go extinct before end of simulation.
  - ▶ At least 10 individuals must remain.
  - ▶ Exactly 67 individuals must remain.

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# Conditioning and Sampling

- ▶ Often want to condition on some predicate:
  - ▶ Population must not go extinct before end of simulation.
  - ▶ At least 10 individuals must remain.
  - ▶ Exactly 67 individuals must remain.
- ▶ Additionally, often want to simulate a *sampled tree*.

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# Conditioning and Sampling

- ▶ Often want to condition on some predicate:
  - ▶ Population must not go extinct before end of simulation.
  - ▶ At least 10 individuals must remain.
  - ▶ Exactly 67 individuals must remain.
- ▶ Additionally, often want to simulate a *sampled* tree.
- ▶ Sampling may be contemporaneous or through time.

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# Conditioning and Sampling

- ▶ Often want to condition on some predicate:
  - ▶ Population must not go extinct before end of simulation.
  - ▶ At least 10 individuals must remain.
  - ▶ Exactly 67 individuals must remain.
- ▶ Additionally, often want to simulate a *sampled* tree.
- ▶ Sampling may be contemporaneous or through time.
- ▶ May be probabilistic or a fixed number of samples.

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# Conditioning and Sampling

- ▶ Often want to condition on some predicate:
  - ▶ Population must not go extinct before end of simulation.
  - ▶ At least 10 individuals must remain.
  - ▶ Exactly 67 individuals must remain.
- ▶ Additionally, often want to simulate a *sampled* tree.
- ▶ Sampling may be contemporaneous or through time.
- ▶ May be probabilistic or a fixed number of samples.
- ▶ Sampling may be specific to one or more populations.

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# Conditioning and Sampling

- ▶ Often want to condition on some predicate:
  - ▶ Population must not go extinct before end of simulation.
  - ▶ At least 10 individuals must remain.
  - ▶ Exactly 67 individuals must remain.
- ▶ Additionally, often want to simulate a *sampled tree*.
- ▶ Sampling may be contemporaneous or through time.
- ▶ May be probabilistic or a fixed number of samples.
- ▶ Sampling may be specific to one or more populations.

## Warning

Sampling and conditioning is the most confusing aspect of MASTER.

*Hopefully we can improve this in future!*

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# The 3 phases of a MASTER simulation

## MASTER simulations involve the following phases

1. The simulation itself (end condition)
2. Post-processing (filtering, sampling)
3. Post-simulation conditioning

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# The 3 phases of a MASTER simulation

## MASTER simulations involve the following phases

1. The simulation itself (end condition)
  2. Post-processing (filtering, sampling)
  3. Post-simulation conditioning
- 
- ▶ End conditions at step 1 may result in early termination of the simulation or a re-do of the simulation.

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# The 3 phases of a MASTER simulation

## MASTER simulations involve the following phases

1. The simulation itself (end condition)
  2. Post-processing (filtering, sampling)
  3. Post-simulation conditioning
- 
- ▶ End conditions at step 1 may result in early termination of the simulation or a re-do of the simulation.
  - ▶ Post-processing performs modifications to the resulting tree.

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# The 3 phases of a MASTER simulation

## MASTER simulations involve the following phases

1. The simulation itself (end condition)
2. Post-processing (filtering, sampling)
3. Post-simulation conditioning

- ▶ End conditions at step 1 may result in early termination of the simulation or a re-do of the simulation.
- ▶ Post-processing performs modifications to the resulting tree.
- ▶ Post-simulation conditioning can cause the simulation to be repeated if some condition is not fulfilled.

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# Population size, leaf and lineage count conditions

- ▶ Ensure population does not go extinct:

```
<populationEndCondition spec="PopulationEndCondition"  
    threshold="0"  
    exceedCondition="false"  
    isRejection="true">  
    <!-- Population elements -->  
</populationEndCondition>
```

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Population size, leaf and lineage count conditions

- ▶ Ensure population does not go extinct:

```
<populationEndCondition spec="PopulationEndCondition"  
    threshold="0"  
    exceedCondition="false"  
    isRejection="true">  
    <!-- Population elements -->  
</populationEndCondition>
```

- ▶ Tree has at least 10 leaves:

```
<postSimCondition spec="LeafCountPostSimCondition" nLeaves="10"  
    exceedCondition="true"/>
```

[Introduction](#)[Background theory](#)[MASTER](#)[MASTER population simulation examples](#)[Linear birth/death](#)[Stochastic logistic model](#)[Stochastic epidemic model](#)[Within-host virus dynamics](#)[Simulating trees](#)[Birth/death trees](#)[SEIR trees](#)[Moran trees](#)[Conditioning and Sampling](#)[Serial sampling](#)[Contemporaneous sampling](#)[Simulating coalescent trees](#)

# Population size, leaf and lineage count conditions

- ▶ Ensure population does not go extinct:

```
<populationEndCondition spec="PopulationEndCondition"  
    threshold="0"  
    exceedCondition="false"  
    isRejection="true">  
    <!-- Population elements -->  
</populationEndCondition>
```

- ▶ Tree has at least 10 leaves:

```
<postSimCondition spec="LeafCountPostSimCondition" nLeaves="10"  
    exceedCondition="true"/>
```

- ▶ Tree does not terminate before 'simulationTime':

```
<lineageEndCondition spec='LineageEndCondition' nLineages="0"  
    isRejection="true"/>
```

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Population size, leaf and lineage count conditions

- ▶ Ensure population does not go extinct:

```
<populationEndCondition spec="PopulationEndCondition"  
    threshold="0"  
    exceedCondition="false"  
    isRejection="true">  
    <!-- Population elements -->  
</populationEndCondition>
```

- ▶ Tree has at least 10 leaves:

```
<postSimCondition spec="LeafCountPostSimCondition" nLeaves="10"  
    exceedCondition="true"/>
```

- ▶ Tree does not terminate before 'simulationTime':

```
<lineageEndCondition spec='LineageEndCondition' nLineages="0"  
    isRejection="true"/>
```

See these Wiki pages for full information:

<http://github.com/CompEvol/MASTER/wiki/populationEndCondition>

<http://github.com/CompEvol/MASTER/wiki/leafCountPostSimCondition>

<http://github.com/CompEvol/MASTER/wiki/lineageEndCondition>

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

Surprise!

We already know how to do this.

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

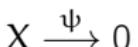
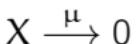
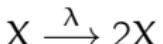
Contemporaneous sampling

Simulating coalescent  
trees

## Surprise!

We already know how to do this.

For example, the following reaction scheme:



This will generate birth/death trees where leaves are terminated by either death events or sampling events.

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

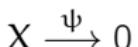
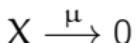
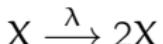
Contemporaneous sampling

Simulating coalescent trees

## Surprise!

We already know how to do this.

For example, the following reaction scheme:



This will generate birth/death trees where leaves are terminated by either death events or sampling events.

*How do we filter out the death event leaves?*

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# LineageFilter: serial sampling

The lineage filter has these options (and others):

```
<inheritancePostProcessor spec='LineageFilter'  
    reactionName="STRING"  
    populationName="STRING"  
    discard="BOOLEAN"  
    leavesOnly="BOOLEAN"/>
```

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# LineageFilter: serial sampling

The lineage filter has these options (and others):

```
<inheritancePostProcessor spec='LineageFilter'  
    reactionName="STRING"  
    populationName="STRING"  
    discard="BOOLEAN"  
    leavesOnly="BOOLEAN"/>
```

- ▶ Use 'reactionName' to select leaves caused by a specific reaction

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# LineageFilter: serial sampling

The lineage filter has these options (and others):

```
<inheritancePostProcessor spec='LineageFilter'  
    reactionName="STRING"  
    populationName="STRING"  
    discard="BOOLEAN"  
    leavesOnly="BOOLEAN"/>
```

- ▶ Use 'reactionName' to select leaves caused by a specific reaction
- ▶ Use 'populationName' to select leaves associated with a specific population

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# LineageFilter: serial sampling

The lineage filter has these options (and others):

```
<inheritancePostProcessor spec='LineageFilter'  
    reactionName="STRING"  
    populationName="STRING"  
    discard="BOOLEAN"  
    leavesOnly="BOOLEAN"/>
```

- ▶ Use 'reactionName' to select leaves caused by a specific reaction
- ▶ Use 'populationName' to select leaves associated with a specific population
- ▶ Use 'discard' to discard instead of keep matched leaves

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# LineageFilter: serial sampling

The lineage filter has these options (and others):

```
<inheritancePostProcessor spec='LineageFilter'  
    reactionName="STRING"  
    populationName="STRING"  
    discard="BOOLEAN"  
    leavesOnly="BOOLEAN"/>
```

- ▶ Use 'reactionName' to select leaves caused by a specific reaction
- ▶ Use 'populationName' to select leaves associated with a specific population
- ▶ Use 'discard' to discard instead of keep matched leaves
- ▶ Use 'leavesOnly' to remove only leaf nodes, not their ancestors.

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Example: Birth/death tree with $\psi$ sampling

MASTER input file:

```
<model spec='Model'>
  <population spec='Population' populationName='X' id='X' />
  <reaction spec='Reaction' reactionName='Birth' rate="2">
    X -> 2X
  </reaction>
  <reaction spec='Reaction' reactionName='Death' rate="0.4">
    X -> 0
  </reaction>
  <reaction spec='Reaction' reactionName='Sampling' rate="0.5">
    X -> 0
  </reaction>
</model>

<inheritancePostProcessor spec="LineageFilter"
                           reactionName="Sampling"/>
<postSimCondition spec="LeafCountPostSimCondition"
                   nLeaves="10"
                   exceedCondition="true"/>
```

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

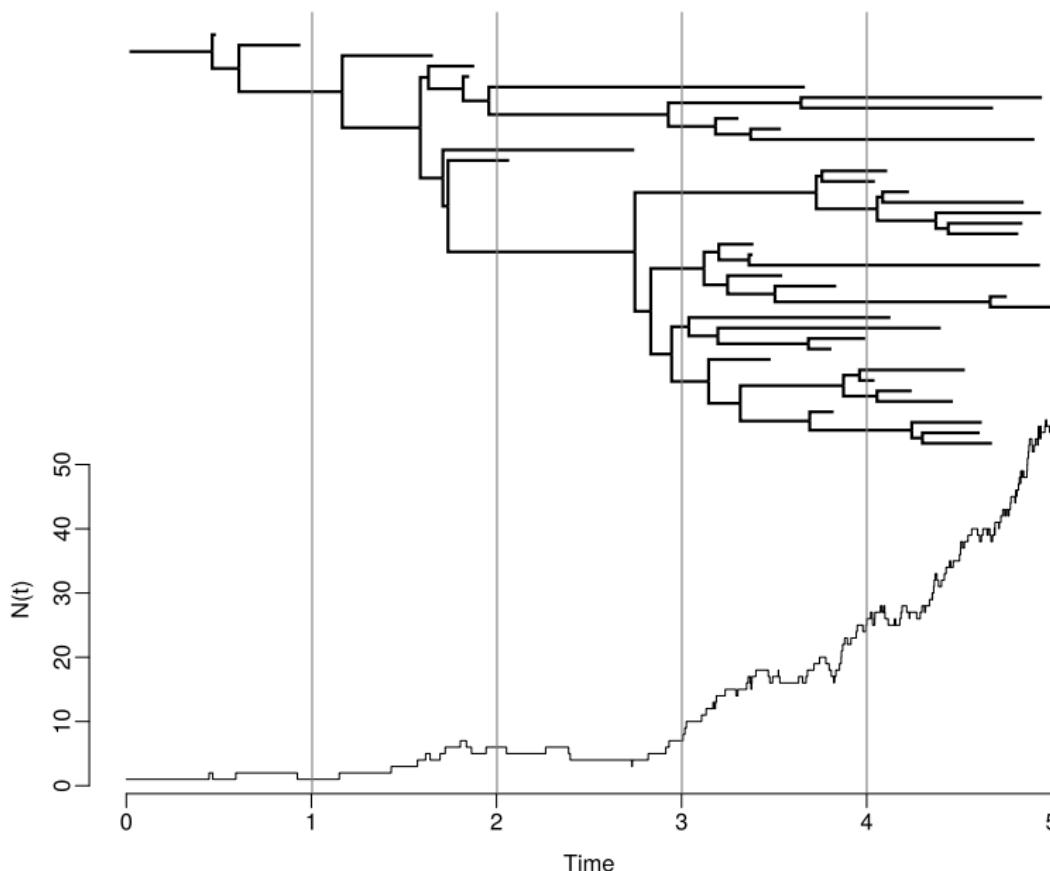
Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

Example: Birth/death tree with  $\psi$  sampling[Introduction](#)[Background theory](#)[MASTER](#)[MASTER population simulation examples](#)[Linear birth/death](#)[Stochastic logistic model](#)[Stochastic epidemic model](#)[Within-host virus dynamics](#)[Simulating trees](#)[Birth/death trees](#)[SEIR trees](#)[Moran trees](#)[Conditioning and Sampling](#)[Serial sampling](#)[Contemporaneous sampling](#)[Simulating coalescent trees](#)

# LineageSampler: contemporaneous sampling

Simulations and  
Phylogeography

The lineage sampler has these options (and others):

```
<inheritancePostProcessor spec='LineageSampler'  
    samplingTime="DOUBLE"  
    nSamples="INTEGER"  
    pSample="DOUBLE"/>
```

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# LineageSampler: contemporaneous sampling

Simulations and  
Phylogeography

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

The lineage sampler has these options (and others):

```
<inheritancePostProcessor spec='LineageSampler'  
    samplingTime="DOUBLE"  
    nSamples="INTEGER"  
    pSample="DOUBLE"/>
```

- ▶ Use 'samplingTime' to choose a single time to perform the sampling.

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

The lineage sampler has these options (and others):

```
<inheritancePostProcessor spec='LineageSampler'  
    samplingTime="DOUBLE"  
    nSamples="INTEGER"  
    pSample="DOUBLE"/>
```

- ▶ Use 'samplingTime' to choose a single time to perform the sampling.
- ▶ Use 'nSamples' to choose an exact number of lineages to sample.

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

The lineage sampler has these options (and others):

```
<inheritancePostProcessor spec='LineageSampler'  
    samplingTime="DOUBLE"  
    nSamples="INTEGER"  
    pSample="DOUBLE"/>
```

- ▶ Use 'samplingTime' to choose a single time to perform the sampling.
- ▶ Use 'nSamples' to choose an exact number of lineages to sample.
- ▶ Use 'pSample' to sample each extant lineage with this probability.

# Example: Birth/death tree with ρ sampling

MASTER input file:

```
<model spec='Model'>
  <population spec='Population' populationName='X' id='X' />
  <reaction spec='Reaction' reactionName='Birth' rate="2">
    X -> 2X
  </reaction>
  <reaction spec='Reaction' reactionName='Death' rate="0.9">
    X -> 0
  </reaction>
</model>

<inheritancePostProcessor spec="LineageSampler"
                           samplingTime="5.0"
                           pSample="0.2"/>
<postSimCondition spec="LeafCountPostSimCondition"
                   nLeaves="5"
                   exceedCondition="true"/>
```

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

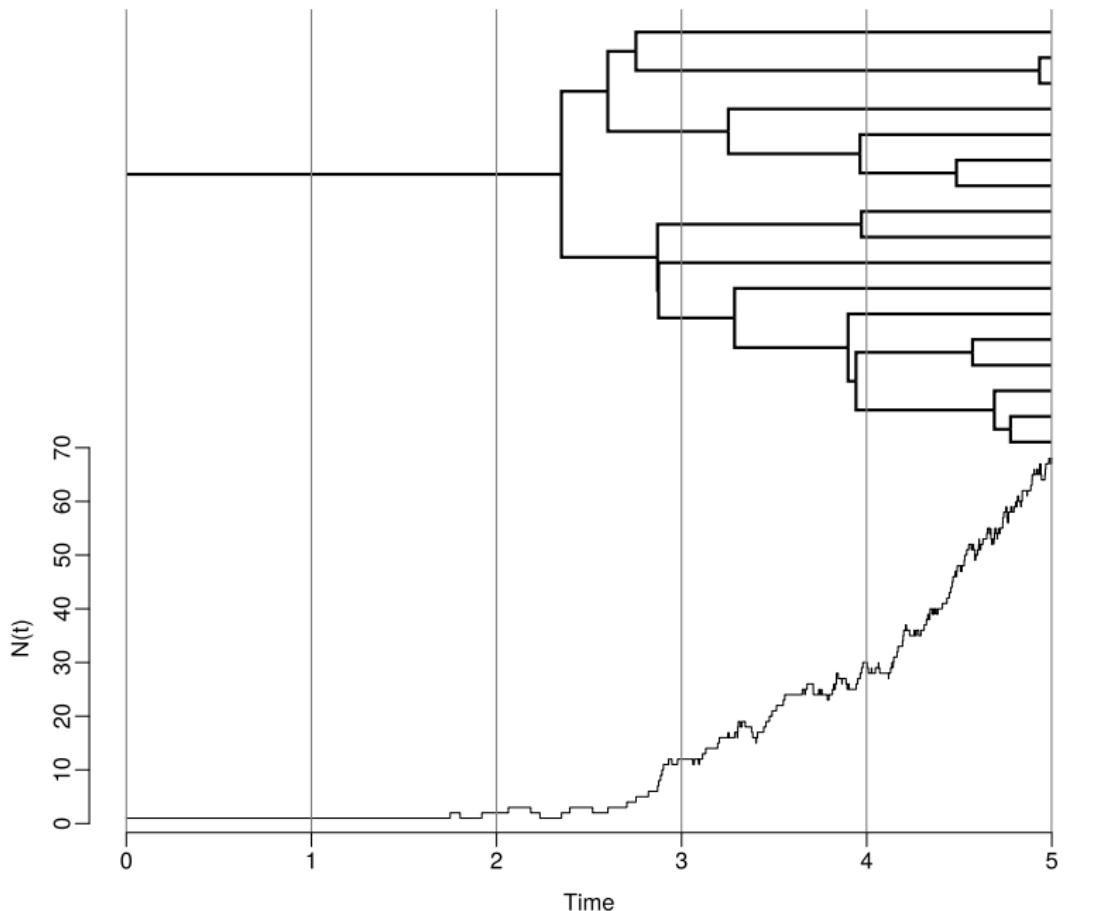
Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

Example: Birth/death tree with  $\rho$  sampling

- Introduction
- Background theory
- MASTER
- MASTER population simulation examples
  - Linear birth/death
  - Stochastic logistic model
  - Stochastic epidemic model
  - Within-host virus dynamics
- Simulating trees
  - Birth/death trees
  - SEIR trees
  - Moran trees
- Conditioning and Sampling
  - Serial sampling
  - Contemporaneous sampling
- Simulating coalescent trees

# Example: Sampled Moran tree

MASTER input file:

```
<model spec='Model'>
  <population spec='Population' populationName='X' id='X' />
  <reaction spec='Reaction' reactionName='BirthDeath' rate="1.7">
    X:1 + X:2 -> 2X:1
  </reaction>
</model>

<inheritancePostProcessor spec='LineageSampler'
  nSamples="10"
  samplingTime="0.1"/>

<postSimCondition spec='LeafCountPostSimCondition' nLeaves="10"/>
```

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

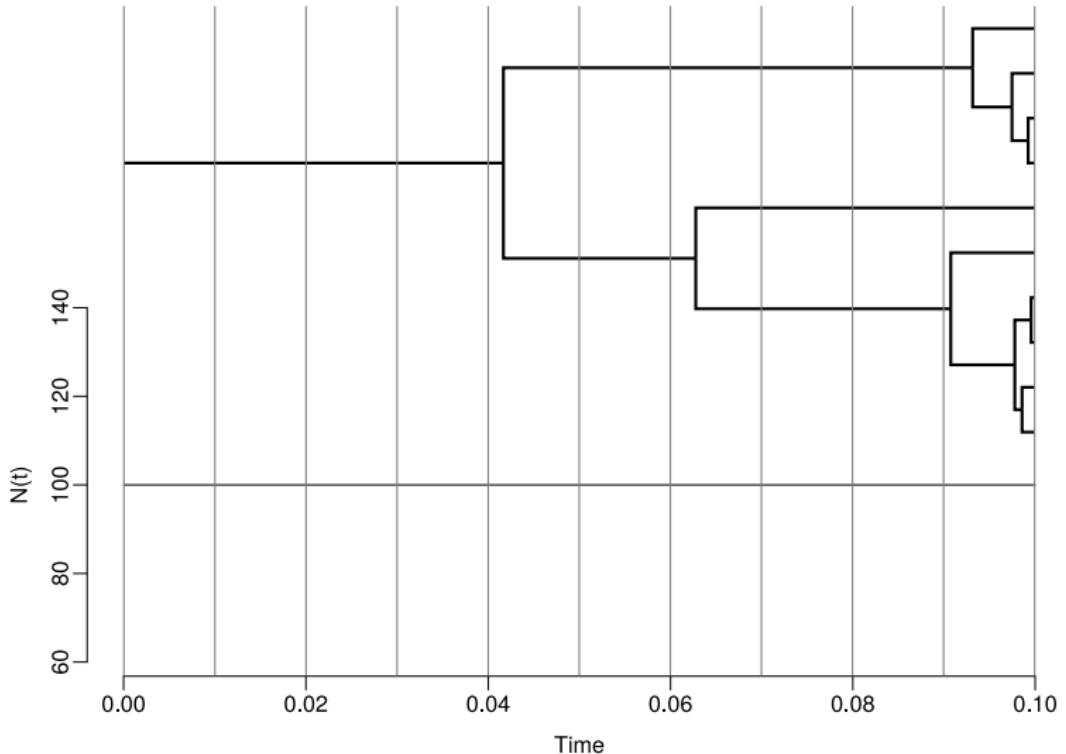
Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# Example: Sampled Moran tree



Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Simulating coalescent trees

The coalescent process is a *backwards-in-time* continuous time Markov jump process.

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Simulating coalescent trees

The coalescent process is a *backwards-in-time* continuous time Markov jump process.

## Problem

Trees in forwards time are no longer trees in reverse time.

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Simulating coalescent trees

Simulations and  
Phylogeography

The coalescent process is a *backwards-in-time* continuous time Markov jump process.

## Problem

Trees in forwards time are no longer trees in reverse time.

## Solution

Use the fact that MASTER allows for the simulation of *networks*.

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# Simulating coalescent trees

The coalescent process is a *backwards-in-time* continuous time Markov jump process.

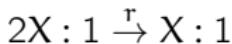
## Problem

Trees in forwards time are no longer trees in reverse time.

## Solution

Use the fact that MASTER allows for the simulation of *networks*.

Use the following reaction scheme:



Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Simulating coalescent trees

The coalescent process is a *backwards-in-time* continuous time Markov jump process.

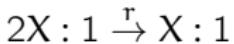
## Problem

Trees in forwards time are no longer trees in reverse time.

## Solution

Use the fact that MASTER allows for the simulation of *networks*.

Use the following reaction scheme:



Associates a single product (parent) with two reactants (children). Relation to Wright-Fisher population size N is:

$$r = \frac{1}{2N}$$

Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

# Simulating coalescent trees

Simulations and  
Phylogeography

MASTER input file:

```
<model spec='Model'>
    <population spec='Population' populationName='X' id='X' />
    <reaction spec='Reaction' reactionName='Coalescence' rate="1.0">
        2X:1 -> X:1
    </reaction>
</model>

<initialState spec='InitState'>
    <lineageSeedMultiple spec='MultipleIndividuals'
                            population="@X" copies="10"/>
</initialState>

<lineageEndCondition spec='LineageEndCondition' nLineages="1"/>

<output spec='NexusOutput' fileName='${filebase}_output.nexus'
           reverseTime="true"/>
<output spec='JsonOutput' fileName='${filebase}_output.json'/>
```

Introduction

Background theory

MASTER

MASTER population  
simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

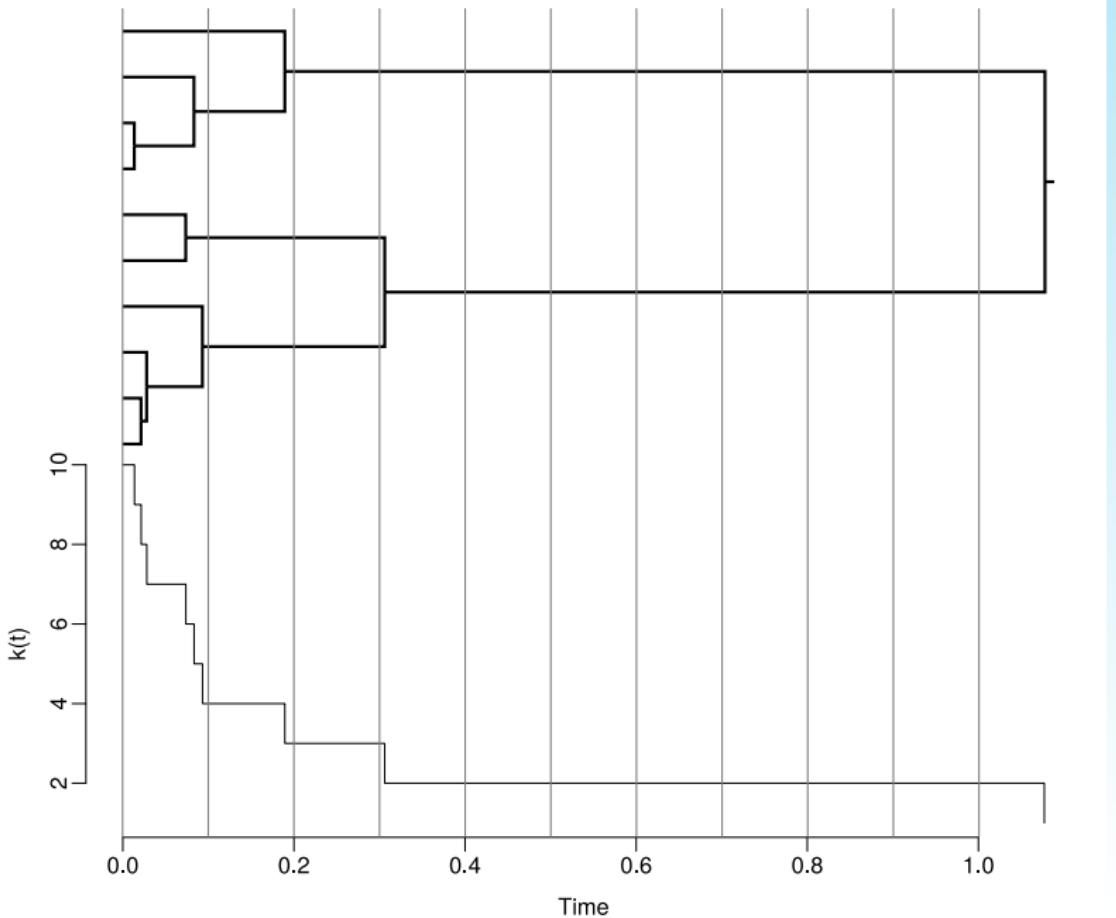
Conditioning and  
Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent  
trees

# Simulating coalescent trees



Introduction

Background theory

MASTER

MASTER population simulation examples

Linear birth/death

Stochastic logistic model

Stochastic epidemic model

Within-host virus dynamics

Simulating trees

Birth/death trees

SEIR trees

Moran trees

Conditioning and Sampling

Serial sampling

Contemporaneous sampling

Simulating coalescent trees

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

## Part II

# Computational Phylogeography

# Structured population

## What is a structured population?

A structured population is able to be partitioned into groups (subpopulations) between which gene flow is limited.

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Structured population

## What is a structured population?

A structured population is able to be partitioned into groups (subpopulations) between which gene flow is limited.

- ▶ Population structure can dramatically influence the shape of the tree.

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Structured population

## What is a structured population?

A structured population is able to be partitioned into groups (subpopulations) between which gene flow is limited.

- ▶ Population structure can dramatically influence the shape of the tree.
- ▶ Structure can be produced by
  - ▶ Spatial segregation with slow migration,

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Structured population

## What is a structured population?

A structured population is able to be partitioned into groups (subpopulations) between which gene flow is limited.

- ▶ Population structure can dramatically influence the shape of the tree.
- ▶ Structure can be produced by
  - ▶ Spatial segregation with slow migration,
  - ▶ Distinct phases of an infection during which a pathogen is more or less contagious,

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Structured population

## What is a structured population?

A structured population is able to be partitioned into groups (subpopulations) between which gene flow is limited.

- ▶ Population structure can dramatically influence the shape of the tree.
- ▶ Structure can be produced by
  - ▶ Spatial segregation with slow migration,
  - ▶ Distinct phases of an infection during which a pathogen is more or less contagious,
  - ▶ *et cetera!*

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

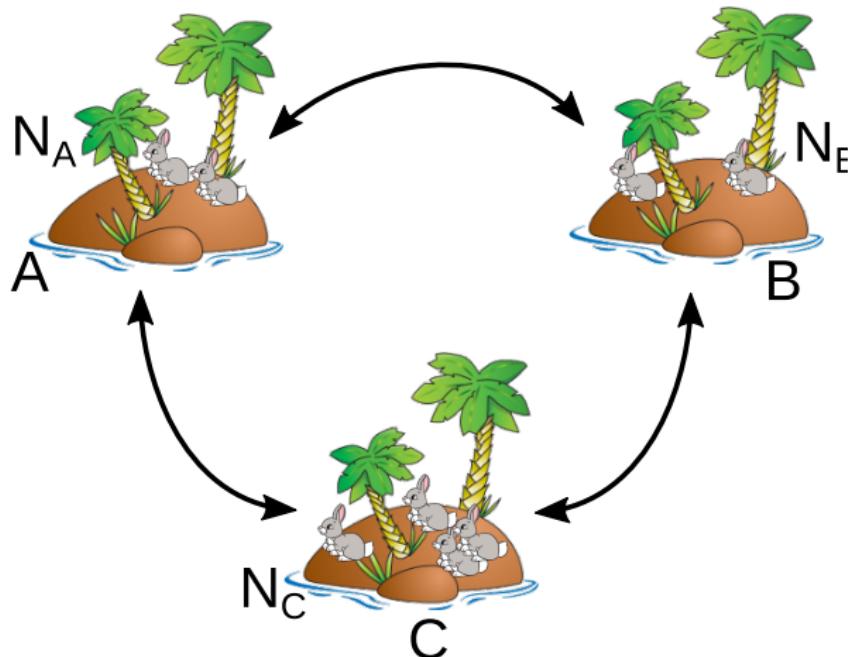
Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Generalized island models and demes

The island model is the prototypical discrete model of spatial structure:



Locations are referred to as *demes*.

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

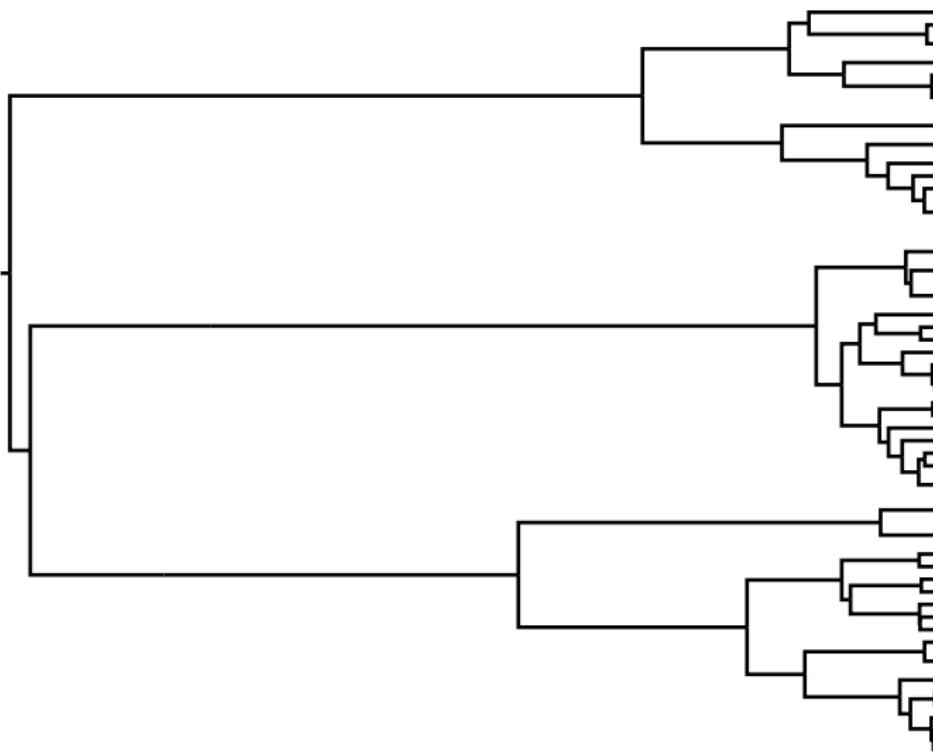
Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Effect of population structure on trees

Population structure can have a very strong effect on the shape of the trees sampled from that population:



Structured  
Populations

Structured Moran  
models

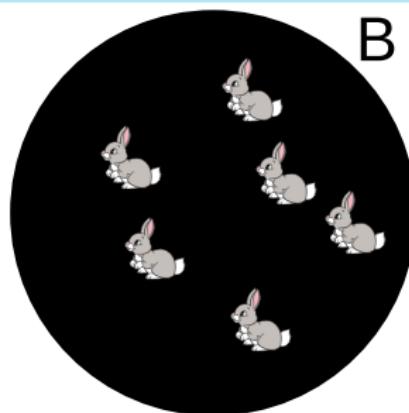
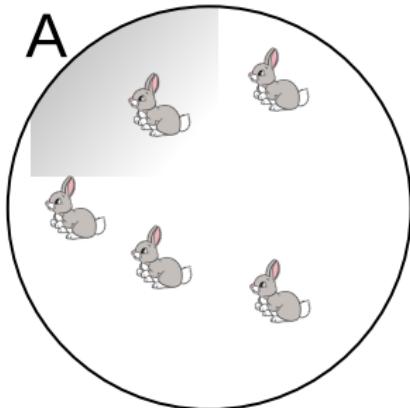
Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# A Structured Moran model



Structured  
Populations

Structured Moran  
models

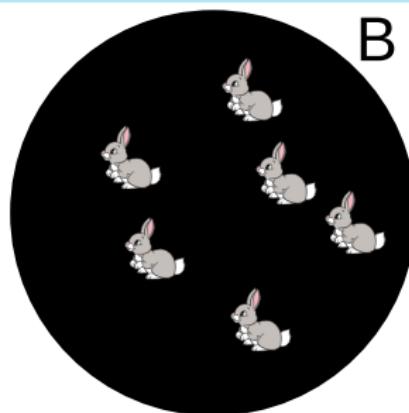
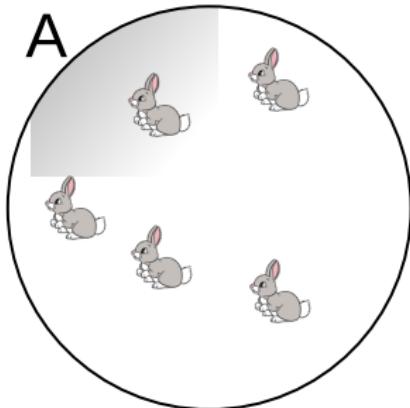
Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# A Structured Moran model



Structured  
Populations

Structured Moran  
models

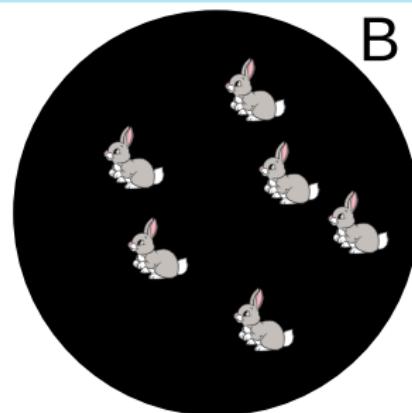
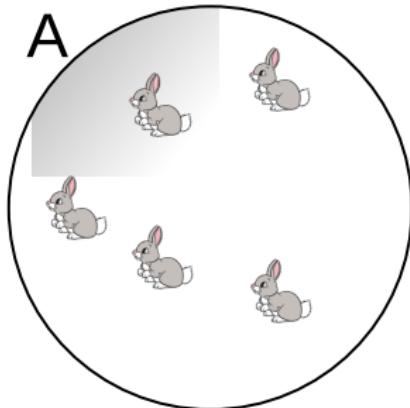
Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# A Structured Moran model



Structured  
Populations

Structured Moran  
models

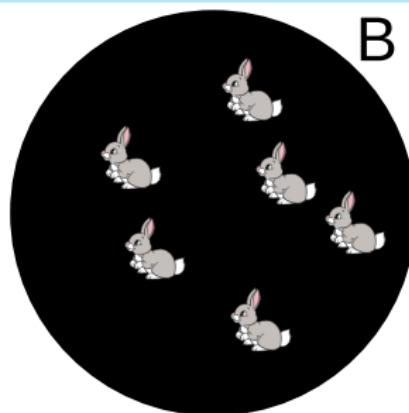
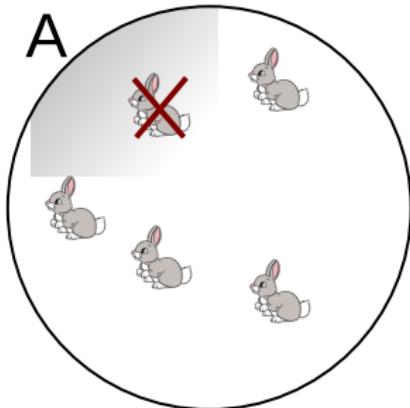
Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# A Structured Moran model



Structured  
Populations

Structured Moran  
models

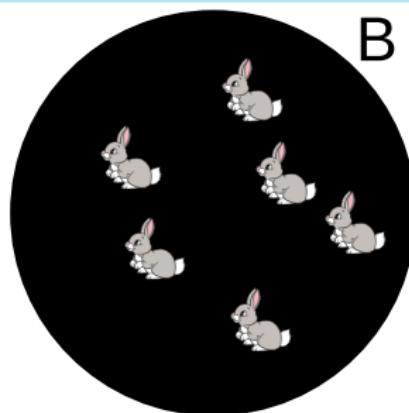
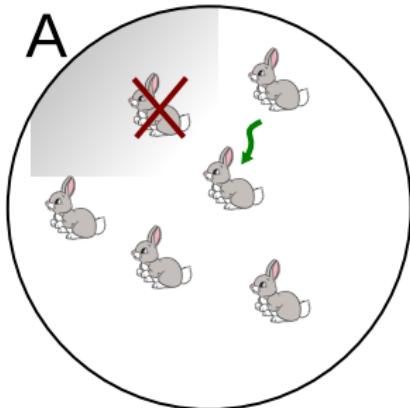
Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# A Structured Moran model



Structured  
Populations

Structured Moran  
models

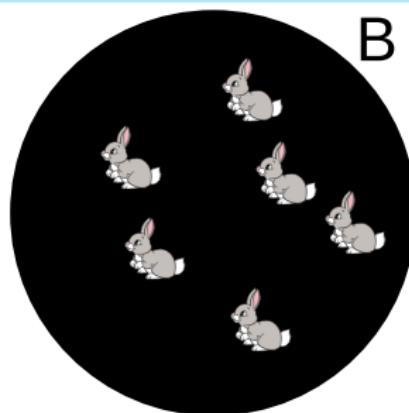
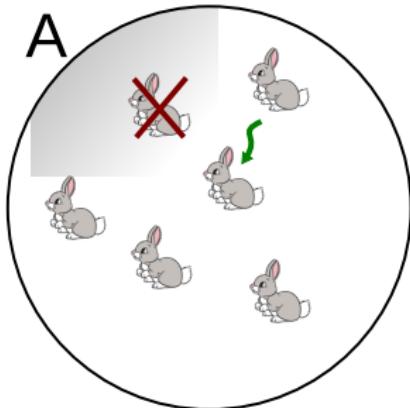
Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# A Structured Moran model



Structured  
Populations

Structured Moran  
models

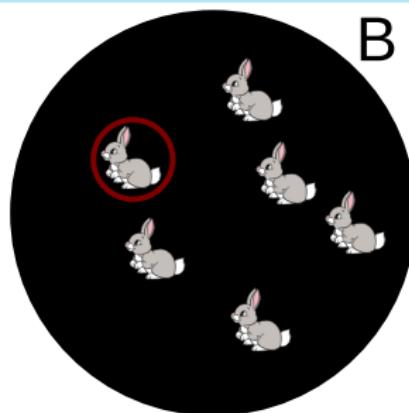
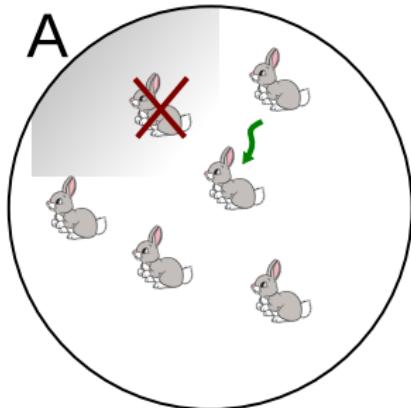
Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# A Structured Moran model



Structured  
Populations

Structured Moran  
models

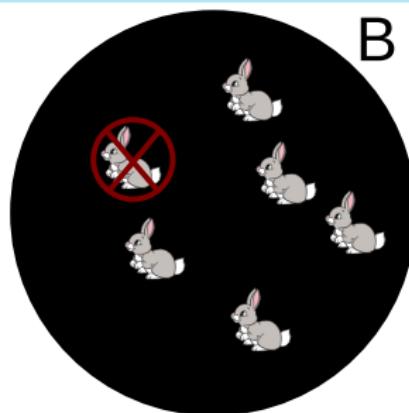
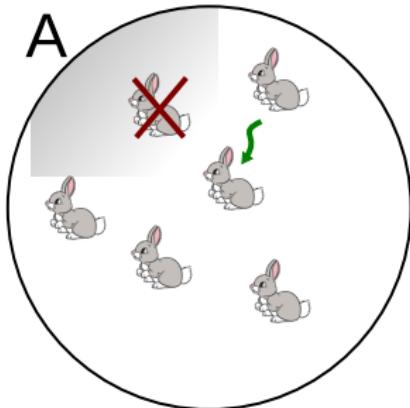
Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# A Structured Moran model



Structured  
Populations

Structured Moran  
models

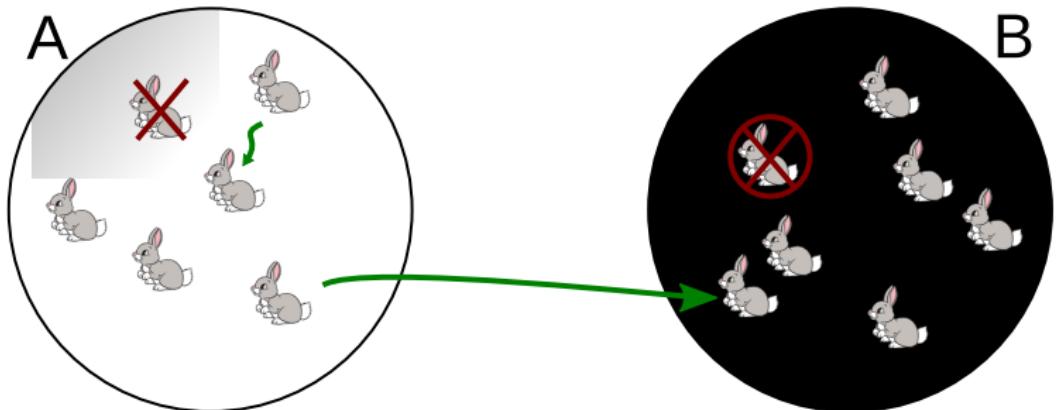
Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# A Structured Moran model



Structured  
Populations

Structured Moran  
models

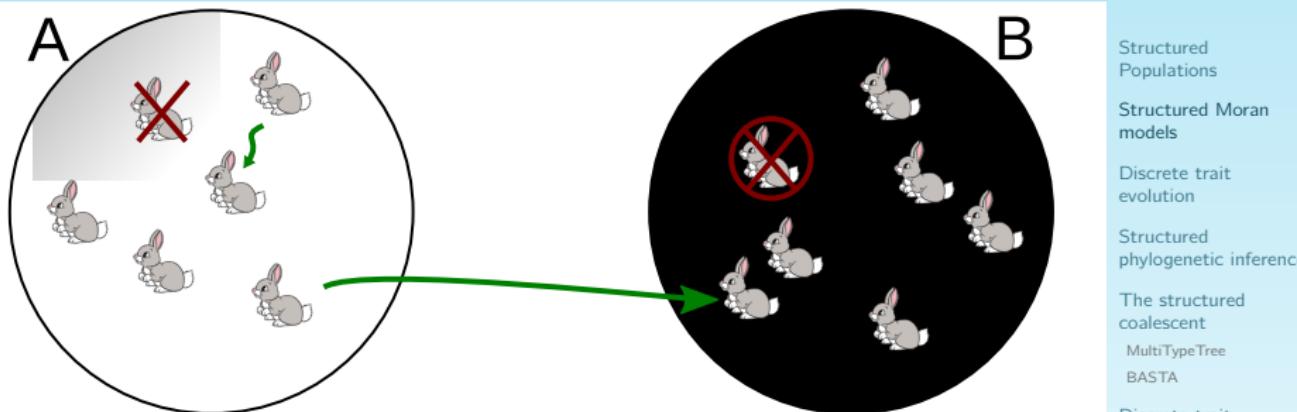
Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# A Structured Moran model



$$X_i : 1 + X_i : 2 \xrightarrow{\lambda} X_i : 1 + X_i : 1$$

$$X_i : 1 + X_j : 2 \xrightarrow{m_{ij}} X_i : 1 + X_j : 1 \text{ for all } i \neq j$$

Structured  
Populations

Structured Moran  
models

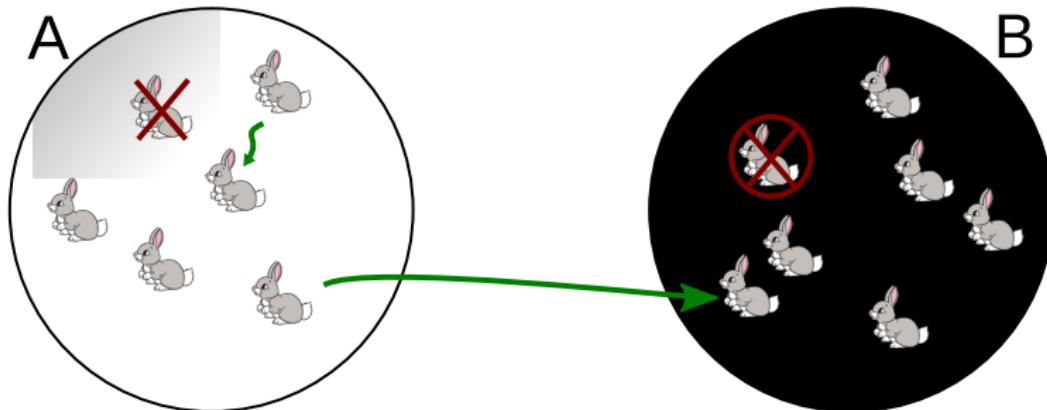
Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# A Structured Moran model



$$X_i : 1 + X_i : 2 \xrightarrow{\lambda} X_i : 1 + X_i : 1$$

$$X_i : 1 + X_j : 2 \xrightarrow{m_{ij}} X_i : 1 + X_j : 1 \text{ for all } i \neq j$$

- ▶ Children born into different locations to their parents at a rate proportional to  $m_{ij}$ .

Structured  
Populations

Structured Moran  
models

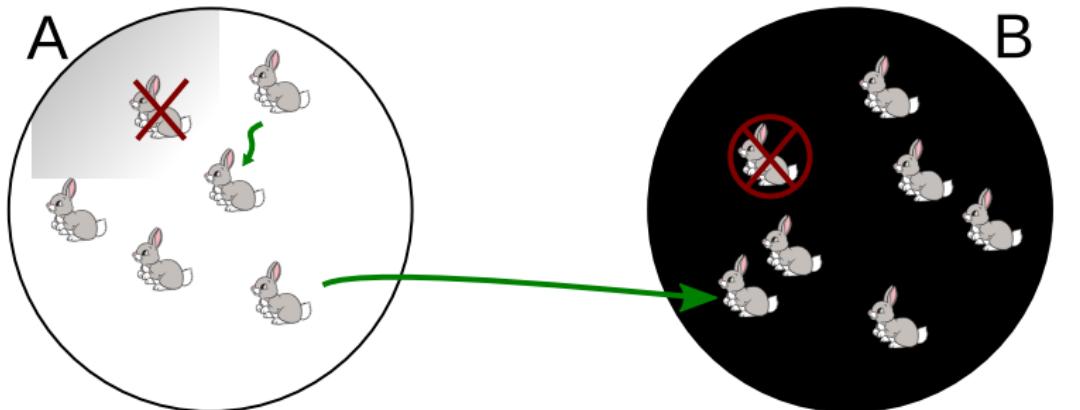
Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# A Structured Moran model



$$X_i : 1 + X_i : 2 \xrightarrow{\lambda} X_i : 1 + X_i : 1$$

$$X_i : 1 + X_j : 2 \xrightarrow{m_{ij}} X_i : 1 + X_j : 1 \text{ for all } i \neq j$$

- ▶ Children born into different locations to their parents at a rate proportional to  $m_{ij}$ .
- ▶ Preserves sub-population sizes  $N_i$ .

Structured  
Populations

Structured Moran  
models

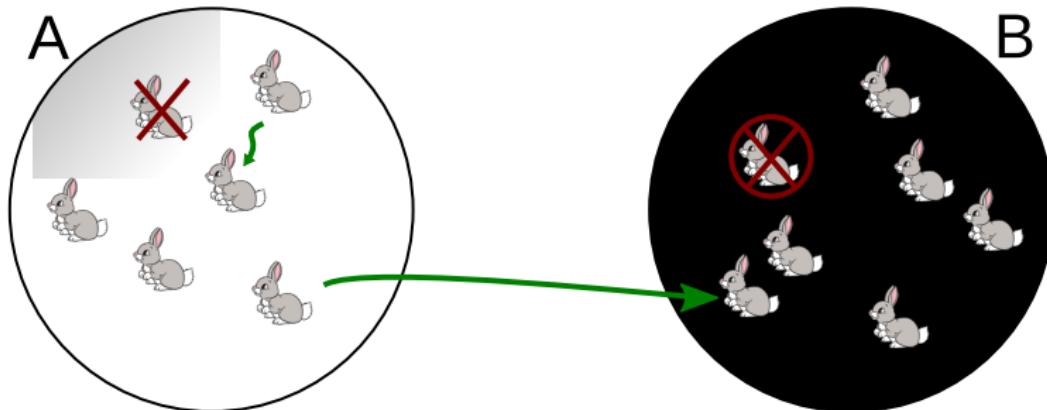
Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# A Structured Moran model



$$X_i : 1 + X_i : 2 \xrightarrow{\lambda} X_i : 1 + X_i : 1$$

$$X_i : 1 + X_j : 2 \xrightarrow{m_{ij}} X_i : 1 + X_j : 1 \text{ for all } i \neq j$$

- ▶ Children born into different locations to their parents at a rate proportional to  $m_{ij}$ .
- ▶ Preserves sub-population sizes  $N_i$ .
- ▶ Assumes that subpopulations are in demographic equilibrium: valid only for  $m_{ij} \ll \lambda$ .

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# A structured Moran model

MASTER input file:

```
<model spec='Model'>
  <populationType spec='PopulationType' typeName='X' dim='2' id='X'/>
  <reaction spec='Reaction' reactionName='BDMigration' rate="1.0">
    X[i]:1 + X[j]:2 -> X[i]:1 + X[j]:1
    <rateMultiplier spec="RateMultiplier" value="i==j ? 1.0 : 0.1"/>
  </reaction>
</model>

<initialState spec='InitState'>
  <populationSize spec='PopulationSize' size='49'>
    <population spec='Population' type="@X" location="0"/>
  </populationSize>
  <populationSize spec='PopulationSize' size='50'>
    <population spec='Population' type="@X" location="1"/>
  </populationSize>
  <lineageSeed spec='Individual'>
    <population spec='Population' type="@X" location="0"/>
  </lineageSeed>
</initialState>

<postSimCondition spec='LeafCountPostSimCondition' nLeaves="20"/>
```

Structured  
Populations

Structured Moran  
models

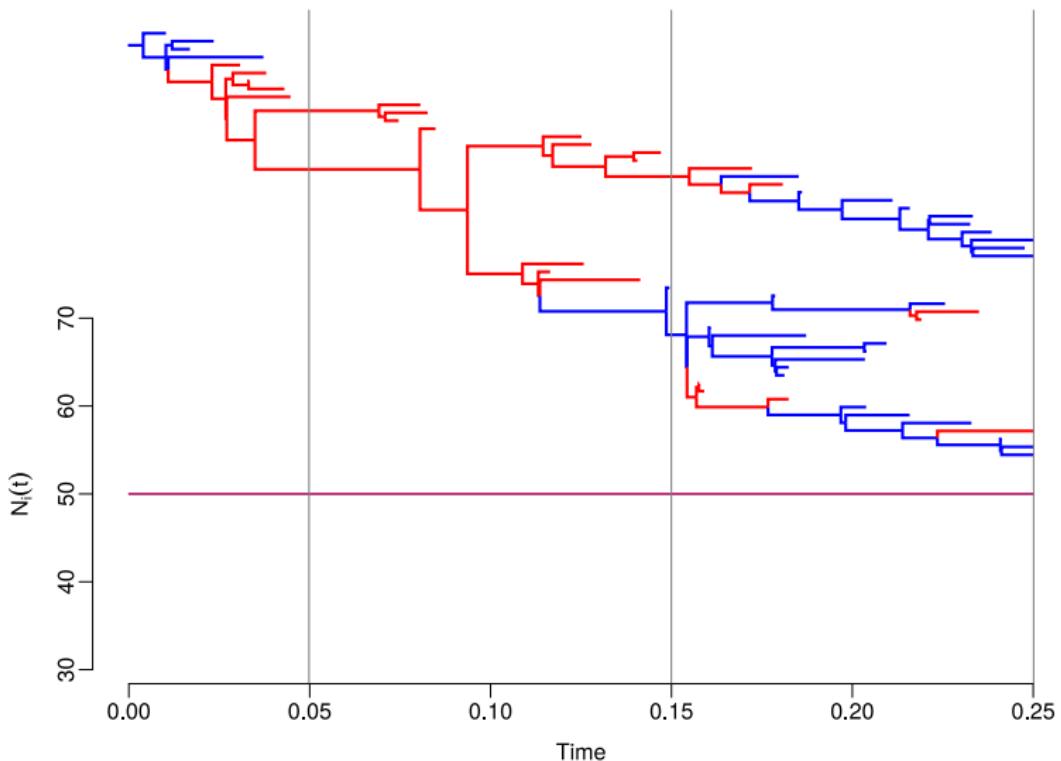
Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# A structured Moran model



Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

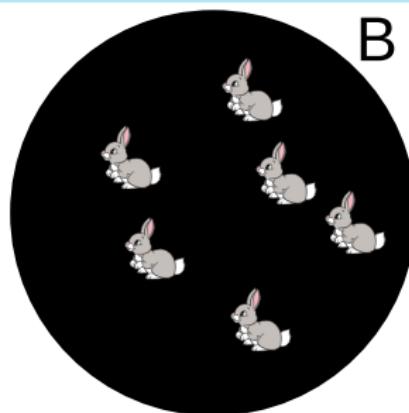
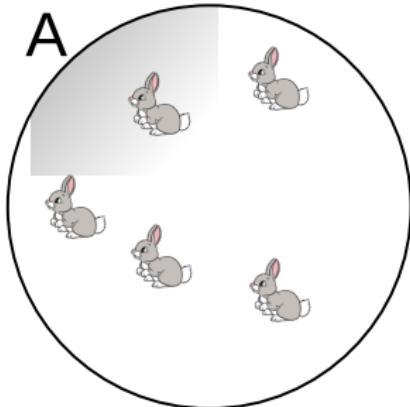
Structured  
phylogenetic inference

The structured  
coalescent

MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Discrete trait evolution



Structured  
Populations

Structured Moran  
models

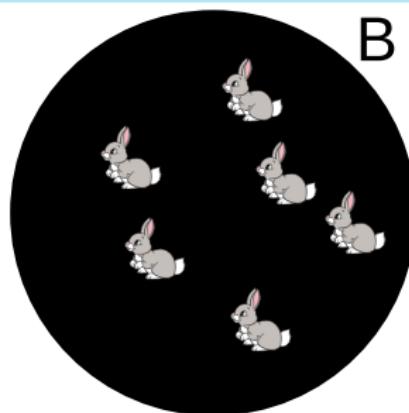
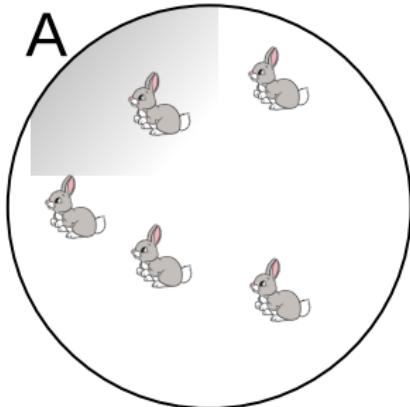
Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Discrete trait evolution



Structured  
Populations

Structured Moran  
models

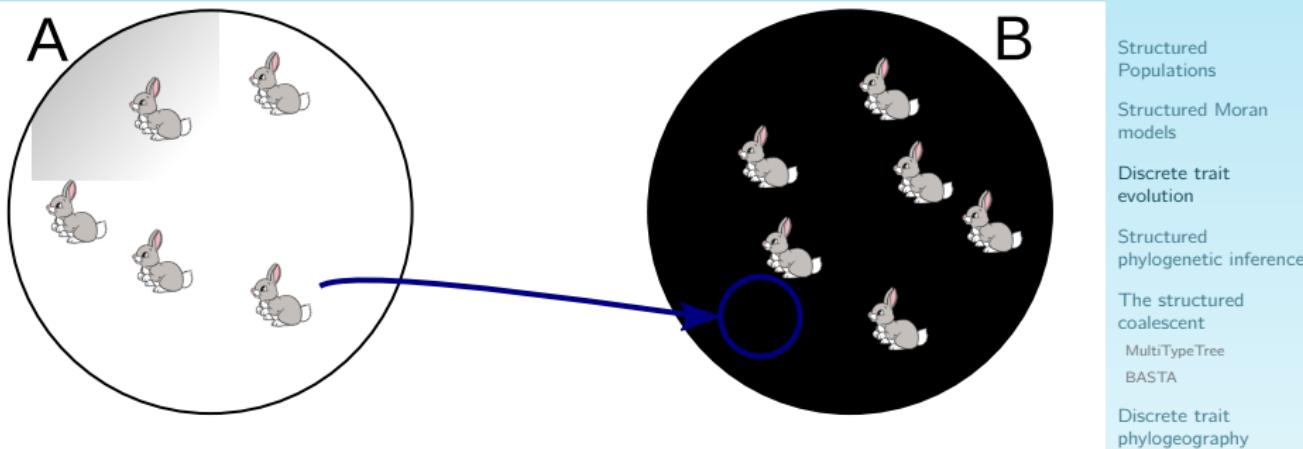
Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Discrete trait evolution



Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

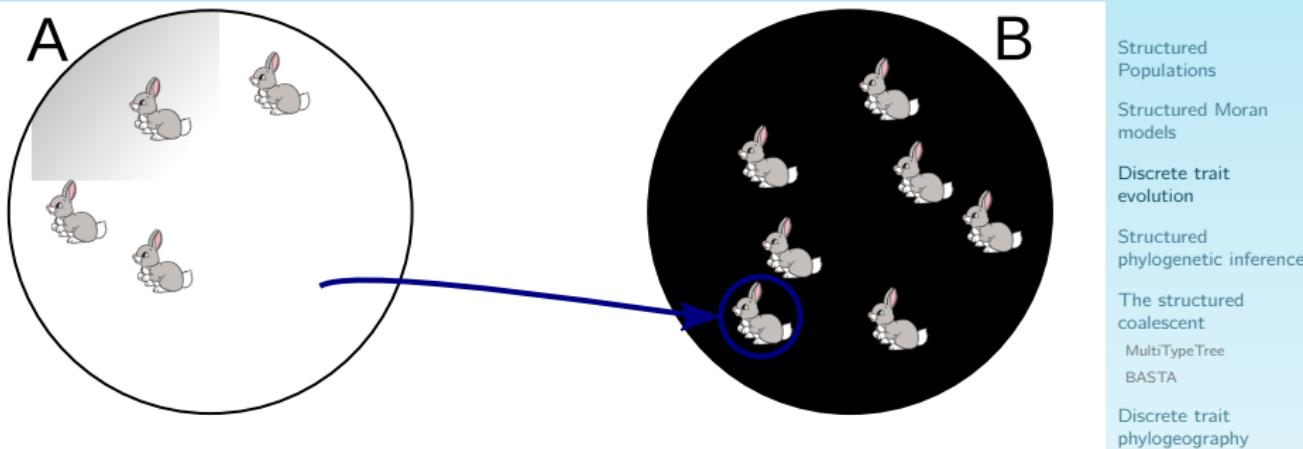
Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Discrete trait evolution

Simulations and  
Phylogeography



Structured  
Populations

Structured Moran  
models

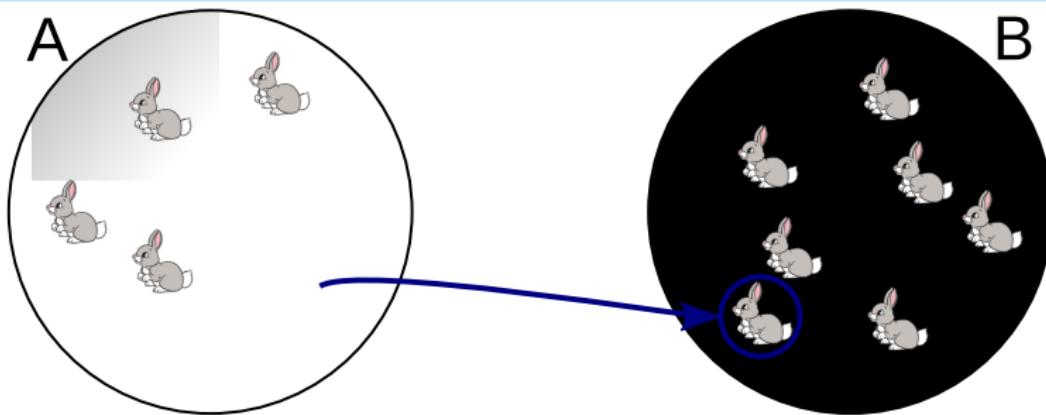
Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Discrete trait evolution



Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

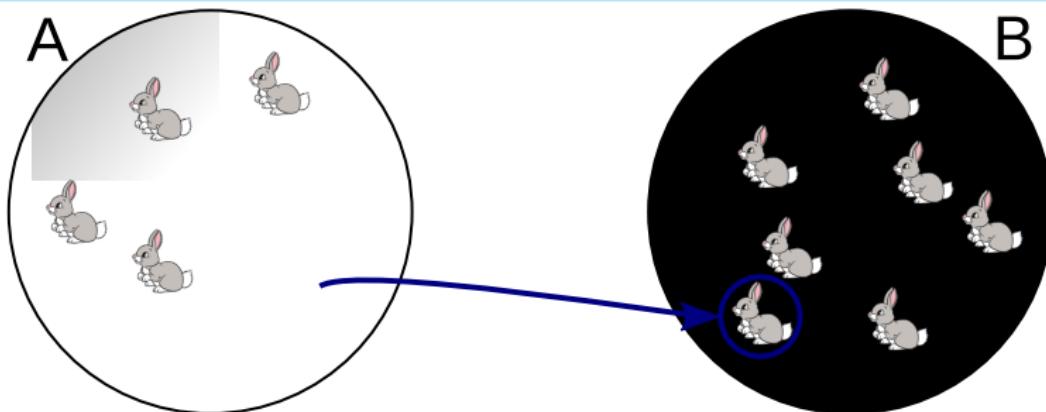
The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

$$X_i : 1 + X_j : 2 \xrightarrow{\lambda} X_i : 1 + X_i : 1$$

$$X_i : 1 \xrightarrow{m_{ij}} X_j : 1 \text{ for all } i \neq j$$

# Discrete trait evolution



Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

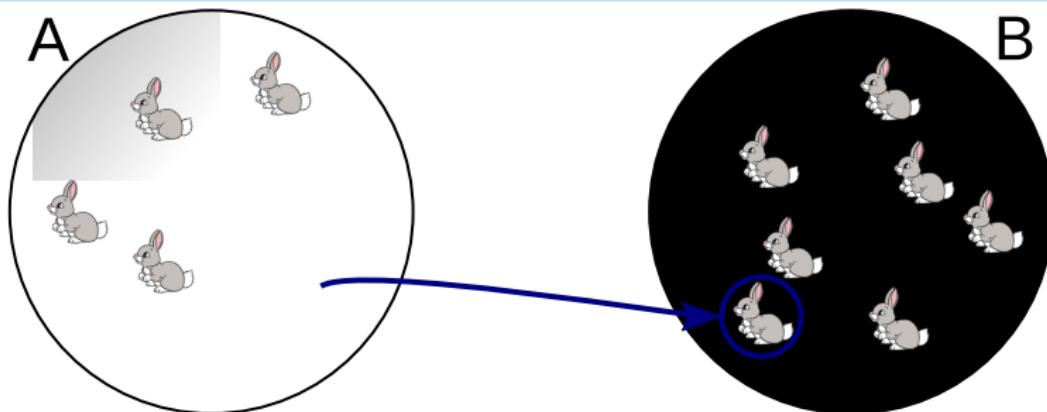
Discrete trait  
phylogeography

$$X_i : 1 + X_j : 2 \xrightarrow{\lambda} X_i : 1 + X_i : 1$$

$$X_i : 1 \xrightarrow{m_{ij}} X_j : 1 \text{ for all } i \neq j$$

- ▶ Birth rate independent of location.

# Discrete trait evolution



$$X_i : 1 + X_j : 2 \xrightarrow{\lambda} X_i : 1 + X_i : 1$$

$$X_i : 1 \xrightarrow{m_{ij}} X_j : 1 \text{ for all } i \neq j$$

- ▶ Birth rate independent of location.
- ▶ Migration process occurs at rate  $m_{ij}$  and does *not* preserve sub-population sizes  $N_i$ .

Structured  
Populations

Structured Moran  
models

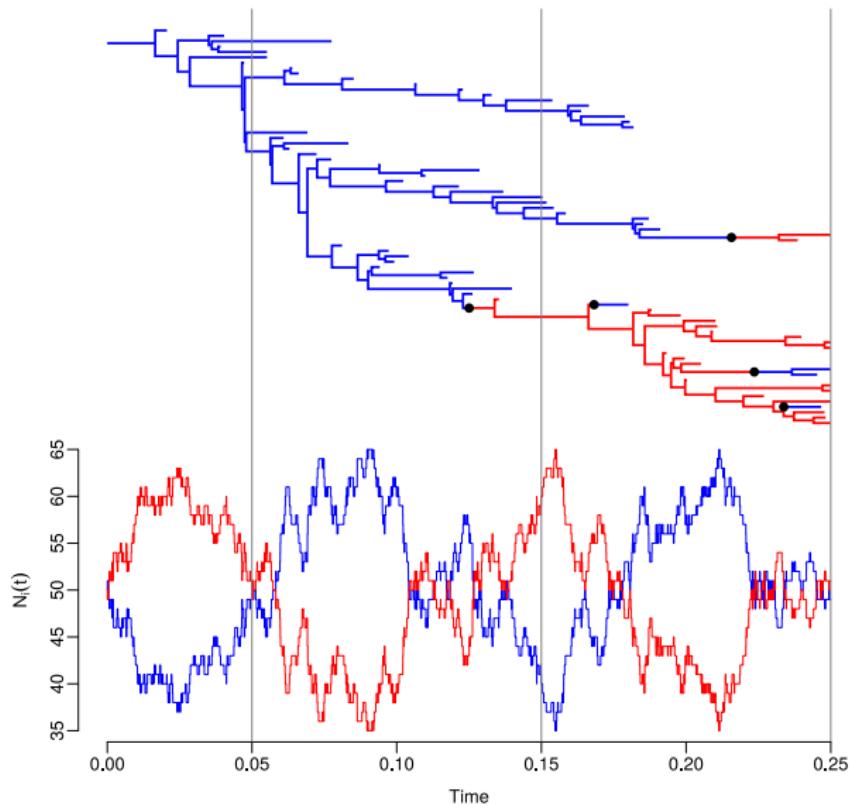
Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Discrete trait evolution



Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Accounting for spatial structure in phylogenetic inference

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

There are several advantages to doing this:

# Accounting for spatial structure in phylogenetic inference

Structured Populations

Structured Moran models

Discrete trait evolution

Structured phylogenetic inference

The structured coalescent  
MultiTypeTree  
BASTA

Discrete trait phylogeography

There are several advantages to doing this:

- ▶ Reduction in analysis bias due to model misspecification.

# Accounting for spatial structure in phylogenetic inference

Structured Populations

Structured Moran models

Discrete trait evolution

Structured phylogenetic inference

The structured coalescent  
MultiTypeTree  
BASTA

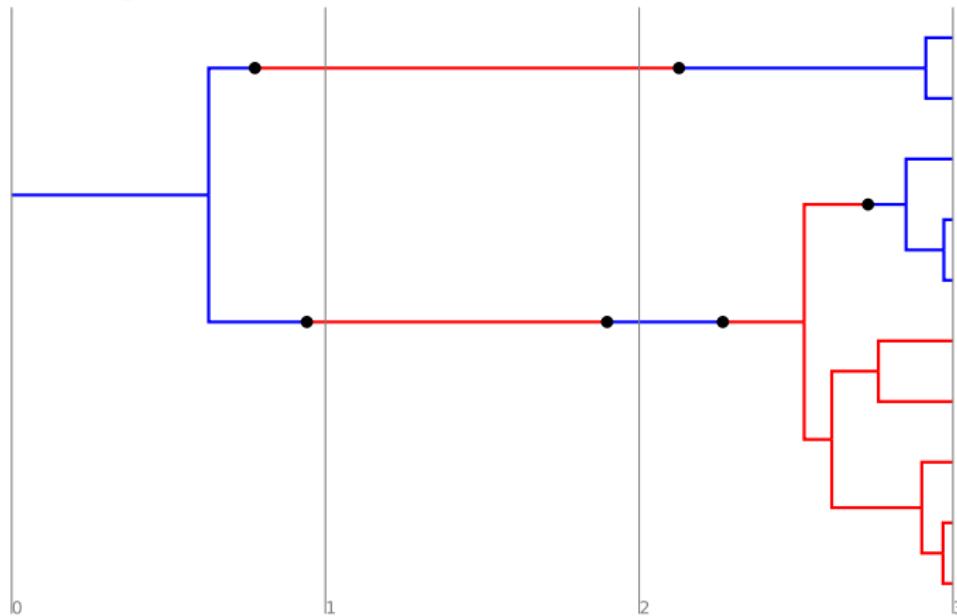
Discrete trait phylogeography

There are several advantages to doing this:

- ▶ Reduction in analysis bias due to model misspecification.
- ▶ Potential to learn about spatial aspects of population dynamics:
  - ▶ Subpopulation sizes  $N_i$ ,
  - ▶ Migration rates  $m_{ij}$ .

# The sampled structured Moran model

Sample a given number of individuals from each sub-population:



Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# The structured coalescent

- ▶ Backward-in-time process for structured genealogies.

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# The structured coalescent

- ▶ Backward-in-time process for structured genealogies.
- ▶ Introduced in its present form by [Notohara, 1990] and [Hudson, 1990].

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent

MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# The structured coalescent

- ▶ Backward-in-time process for structured genealogies.
- ▶ Introduced in its present form by [Notohara, 1990] and [Hudson, 1990].
- ▶ Arises as a limit of a number of forward-time structured population models including the structured Wright-Fisher model and the structured Moran model.

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent

MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# The structured coalescent

- ▶ Backward-in-time process for structured genealogies.
- ▶ Introduced in its present form by [Notohara, 1990] and [Hudson, 1990].
- ▶ Arises as a limit of a number of forward-time structured population models including the structured Wright-Fisher model and the structured Moran model.
- ▶ Explicitly conditions on the number and location of samples.

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

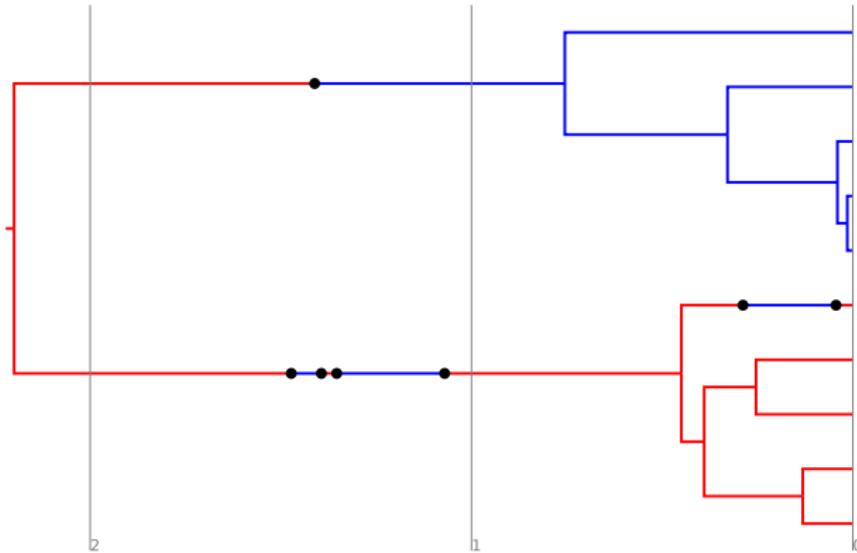
$$\begin{aligned}2X_i : 1 &\xrightarrow{r_i} X_i : 1 \\ X_i : 1 &\xrightarrow{M_{ij}} X_j : 1\end{aligned}$$

where  $r_i = 1/2N_i$  and

$$M_{ij} = \frac{N_j}{N_i} q_{ji}$$

where  $q_{ji}$  is the forward-in-time rate of movement of individuals from  $j$  to  $i$  per individual in  $j$ .

# The structured coalescent



Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

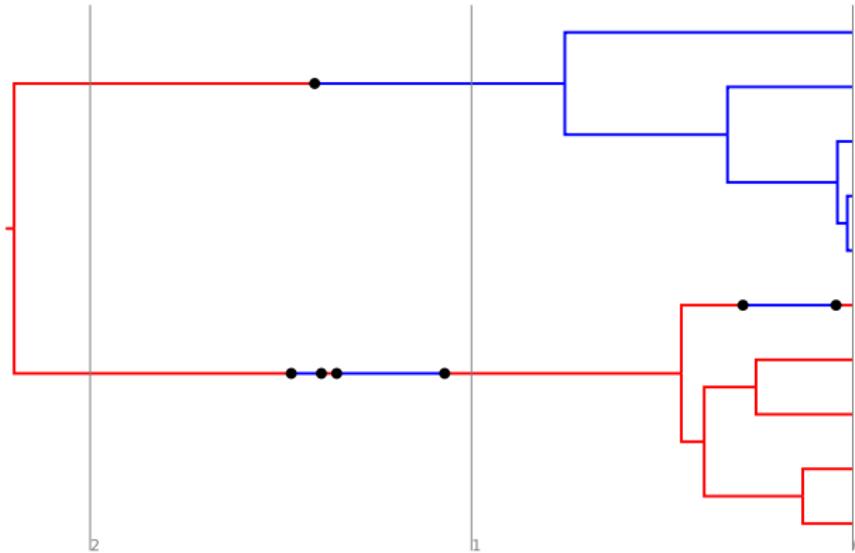
Structured  
phylogenetic inference

The structured  
coalescent

MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# The structured coalescent



Can easily write down the probability for the tree:

$$P(T|M, \vec{N}, \vec{L}, \vec{t})$$

where  $M$  are the backward-time migration rates,  $\vec{N}$  are the subpopulation sizes,  $\vec{L}$  and  $\vec{t}$  are the sample locations and times.

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

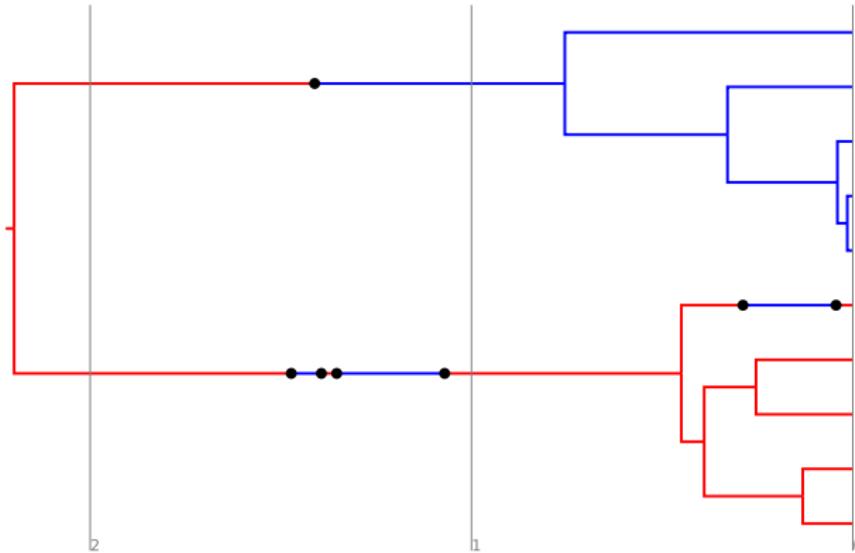
Structured  
phylogenetic inference

The structured  
coalescent

MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# The structured coalescent



Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent

MultiTypeTree  
BASTA

Discrete trait  
phylogeography

Can easily write down the probability for the tree:

$$P(T|M, \vec{N}, \vec{L}, \vec{t})$$

where  $M$  are the backward-time migration rates,  $\vec{N}$  are the subpopulation sizes,  $\vec{L}$  and  $\vec{t}$  are the sample locations and times.

*T is a tree which includes ancestral migration events.*

# Structured coalescent inference

Phylogenetic inference under the structured coalescent is traditionally done by sampling from the following posterior:

$$P(T, M, \vec{N}, \mu | A, \vec{t}, \vec{L}) \propto P(A|T, \mu) P(T|M, \vec{N}, \vec{t}, \vec{L}) P(\mu, M, \vec{N})$$

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

Phylogenetic inference under the structured coalescent is traditionally done by sampling from the following posterior:

$$P(T, M, \vec{N}, \mu | A, \vec{t}, \vec{L}) \propto P(A|T, \mu) P(T|M, \vec{N}, \vec{t}, \vec{L}) P(\mu, M, \vec{N})$$

Some points:

- ▶ Trees with full migration history may be inferred.

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

Phylogenetic inference under the structured coalescent is traditionally done by sampling from the following posterior:

$$P(T, M, \vec{N}, \mu | A, \vec{t}, \vec{L}) \propto P(A|T, \mu) P(T|M, \vec{N}, \vec{t}, \vec{L}) P(\mu, M, \vec{N})$$

Some points:

- ▶ Trees with full migration history may be inferred.
- ▶ Just like the standard coalescent, the structured coalescent takes on the role of a *tree prior* for the multi-type tree T.

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

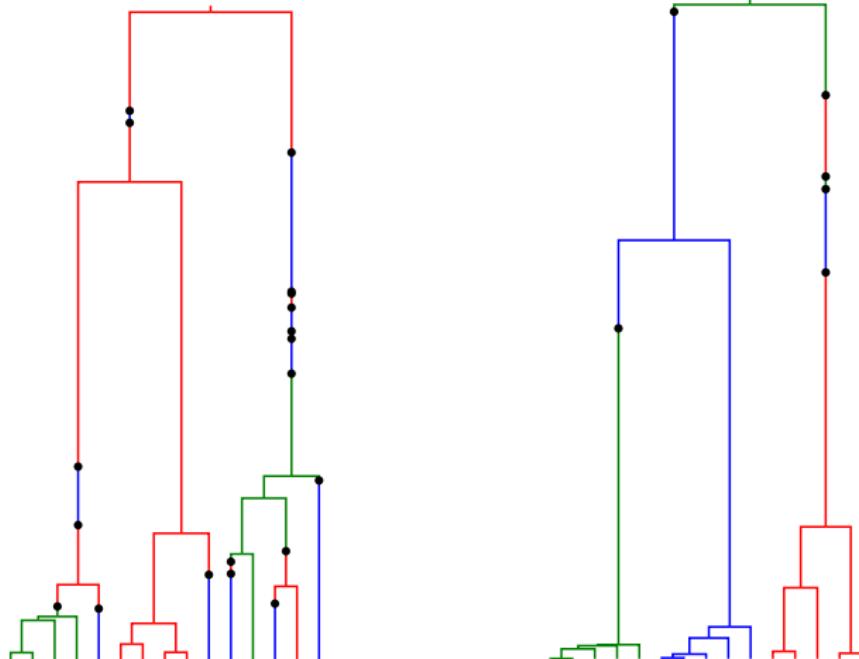
Phylogenetic inference under the structured coalescent is traditionally done by sampling from the following posterior:

$$P(T, M, \vec{N}, \mu | A, \vec{t}, \vec{L}) \propto P(A|T, \mu) P(T|M, \vec{N}, \vec{t}, \vec{L}) P(\mu, M, \vec{N})$$

Some points:

- ▶ Trees with full migration history may be inferred.
- ▶ Just like the standard coalescent, the structured coalescent takes on the role of a *tree prior* for the multi-type tree  $T$ .
- ▶ There are an INFINITE number of possible migration histories for a given number of samples. (No upper bound on the number of migration events.)

# Potential inference problems



- ▶ Likelihood allows infinite  $N_i$  when no coalescence in deme  $i$ .  
∴ Place strict upper bound on  $N_i$ .
- ▶ Likelihood has support for  $M_{ij} = 0$  when no  $i \rightarrow j$  migration.  
∴ Place lower bound on  $M_{ij}$ .

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent

MultiTypeTree  
BASTA

Discrete trait  
phylogeography

The screenshot shows a web browser window for 'MultiTypeTree by Compevol' at 'compevol.github.io/MultiTypeTree/'. The page has a blue header with the title 'MultiTypeTree' and 'Structured population inference for BEAST 2'. Below the header is a yellow navigation bar with links for 'Beginner's Tutorial', 'Wiki', and 'Source code'. The main content area has a white background. On the left, there is an 'Overview' section containing text about the package and its features. On the right, there is a large phylogenetic tree diagram with multiple colored edges (red, green, blue, purple) representing different population types. The tree is rooted and shows a complex branching pattern.

**Overview**

MultiTypeTree is a BEAST 2 package for performing phylogenetic inference under models incorporating some form of population structure. It implements a "multi-type" tree object, representing a phylogenetic tree with every time on each edge associated with one of a finite number of discrete types. These types often represent locations, but can in principle represent any well-defined sub-population.

In addition to the basic object, MultiTypeTree implements a full suite of proposal operators for these objects, as well as the structured coalescent probability density and a structured coalescent simulator. Together, these components allow for inference of trees, migration rates and population sizes under the structured coalescent model. However, we stress that the system can (and will) be easily extended to cope with other structured-population genealogical models.

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Practical MultiTypeTree advice

Simulations and  
Phylogeography

It is best to have:

- ▶ a roughly even distribution of samples across demes,

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

It is best to have:

- ▶ a roughly even distribution of samples across demes,
- ▶ a good number (30 say) samples per deme.

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

It is best to have:

- ▶ a roughly even distribution of samples across demes,
- ▶ a good number (30 say) samples per deme.

Additionally:

- ▶ the posterior distribution for  $T$ ,  $M$  and  $\bar{N}$  may be multi-modal, and

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Practical MultiTypeTree advice

It is best to have:

- ▶ a roughly even distribution of samples across demes,
- ▶ a good number (30 say) samples per deme.

Additionally:

- ▶ the posterior distribution for  $T$ ,  $M$  and  $\bar{N}$  may be multi-modal, and
- ▶ slow mixing is generally a problem for more than 4 demes.

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

It is best to have:

- ▶ a roughly even distribution of samples across demes,
- ▶ a good number (30 say) samples per deme.

Additionally:

- ▶ the posterior distribution for  $T$ ,  $M$  and  $\vec{N}$  may be multi-modal, and
- ▶ slow mixing is generally a problem for more than 4 demes.

Thus MultiTypeTree works best for well-sampled coarsely structured populations. (Improvements can be gained by constraining  $M$  in some way - e.g. to be symmetric.)

# BEAST 2 SC inference 2: BASTA

Alternative approach to SC inference by [De Maio et al., 2015].

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# BEAST 2 SC inference 2: BASTA

Alternative approach to SC inference by [De Maio et al., 2015].

- ▶ Integrates over individual migration histories - better scaling with increasing numbers of demes. (> 8 possible.)

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

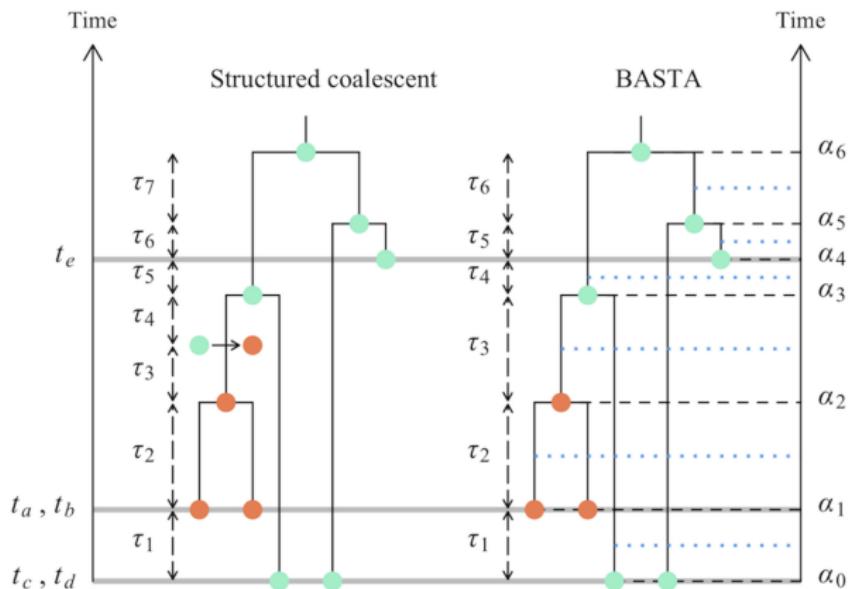
The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# BEAST 2 SC inference 2: BASTA

Alternative approach to SC inference by [De Maio et al., 2015].

- ▶ Integrates over individual migration histories - better scaling with increasing numbers of demes. (> 8 possible.)
- ▶ Approximate integration used.



Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Setting up a BASTA analysis

- ▶ BASTA currently lacks support for BEAUTi.

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Setting up a BASTA analysis

- ▶ BASTA currently lacks support for BEAUTi.
- ▶ Analyses can be constructed by setting up a MultiTypeTree analysis and modifying the resulting XML.

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Setting up a BASTA analysis

- ▶ BASTA currently lacks support for BEAUTi.
- ▶ Analyses can be constructed by setting up a MultiTypeTree analysis and modifying the resulting XML.
- ▶ See the tutorial at  
<http://github.com/compevol/MultiTypeTree/wiki/> for details on how to do this.

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

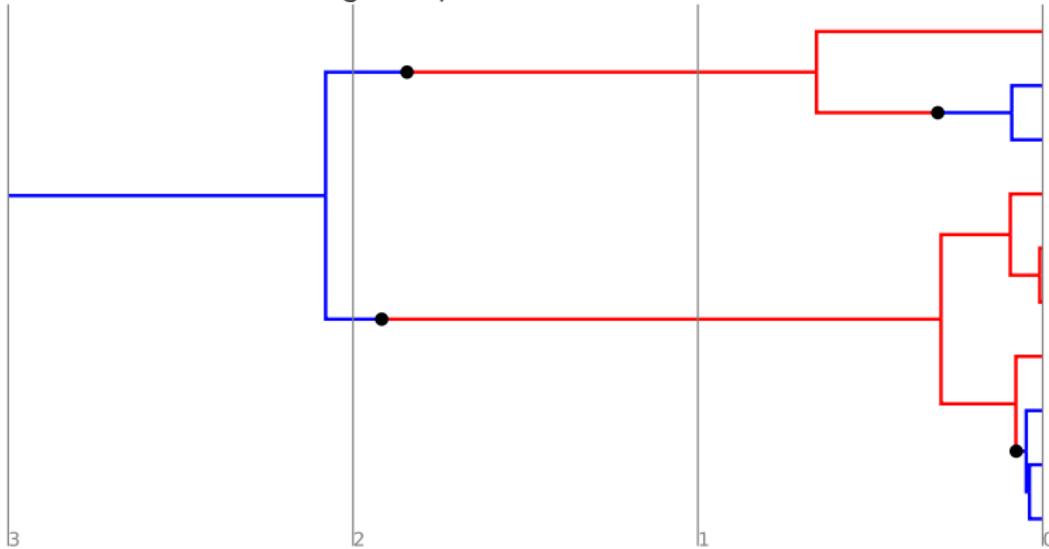
Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Discrete trait phyogeography

Consider the following sampled discrete trait evolution tree:



Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

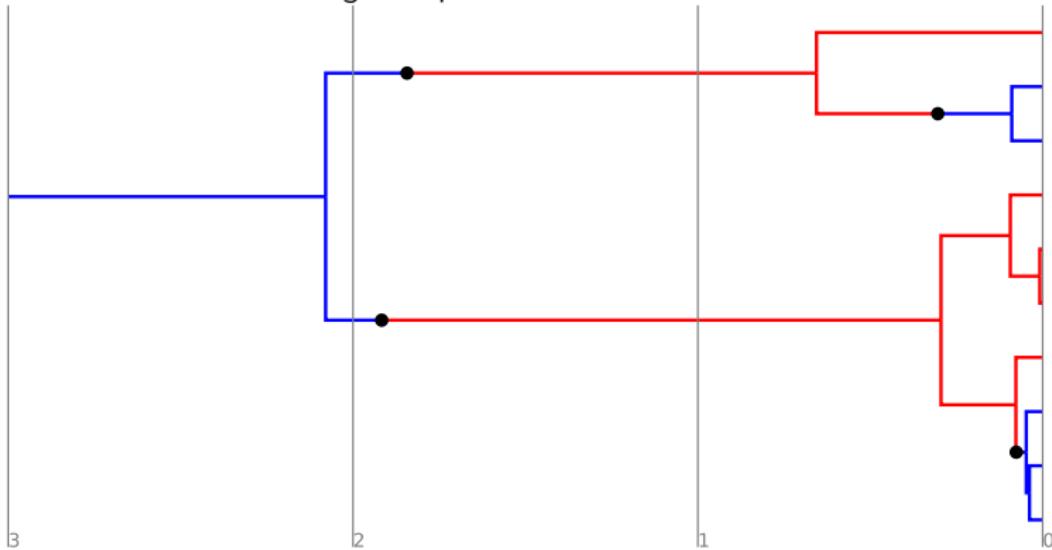
Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phyogeography

# Discrete trait phyogeography

Consider the following sampled discrete trait evolution tree:



- If sampling is done ignoring location, the tree has an unstructured coalescent distribution.

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

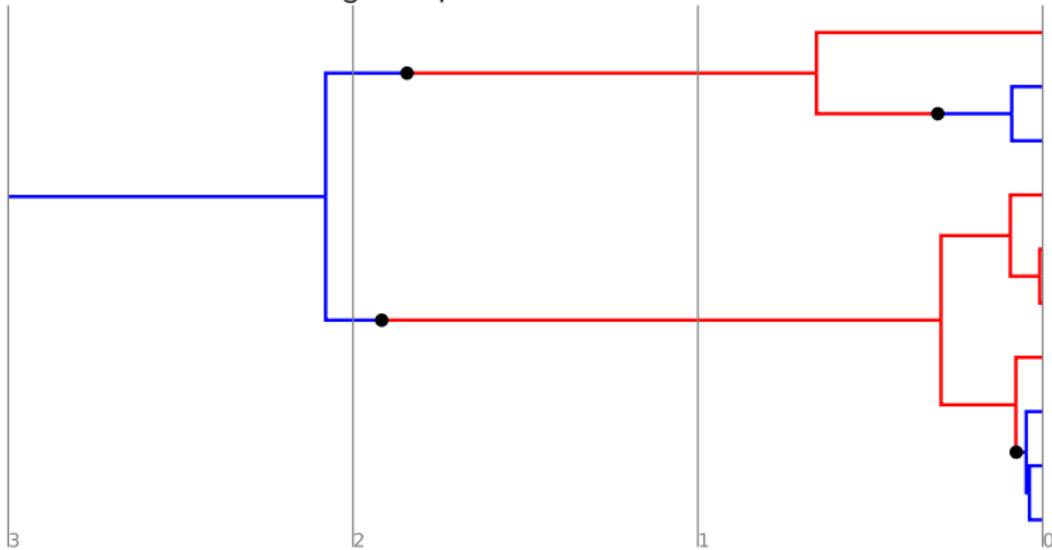
Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phyogeography

# Discrete trait phylogeography

Consider the following sampled discrete trait evolution tree:



- ▶ If sampling is done ignoring location, the tree has an unstructured coalescent distribution.
- ▶ Migration events occur as a Poisson process down the coalescent tree (cf. substitutions with the neutrality assumption).

Structured Populations  
Structured Moran models  
Discrete trait evolution  
Structured phylogenetic inference  
The structured coalescent  
MultiTypeTree  
BASTA  
Discrete trait phylogeography

# Discrete trait phylogeography inference

Inference under the discrete trait phylogeography model of [Lemey et al., 2009] is done by sampling from the following posterior:

$$P(T, M, Q, \mu | A, \vec{t}, \vec{L}) \propto P(A|T, \mu) P(L|T, Q) P(T|N, \vec{t}) P(\mu, Q, N)$$

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Discrete trait phylogeography inference

Inference under the discrete trait phylogeography model of [Lemey et al., 2009] is done by sampling from the following posterior:

$$P(T, M, Q, \mu | A, \vec{t}, \vec{L}) \propto P(A|T, \mu) P(L|T, Q) P(T|N, \vec{t}) P(\mu, Q, N)$$

Some points:

- ▶ Migration history is integrated out.

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Discrete trait phylogeography inference

Inference under the discrete trait phylogeography model of [Lemey et al., 2009] is done by sampling from the following posterior:

$$P(T, M, Q, \mu | A, \vec{t}, \vec{L}) \propto P(A|T, \mu) P(L|T, Q) P(T|N, \vec{t}) P(\mu, Q, N)$$

Some points:

- ▶ Migration history is integrated out.
- ▶ The discrete trait model contributes an additional *likelihood* term for the spatial transition matrix  $Q$ .

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent

MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Discrete trait phylogeography inference

Inference under the discrete trait phylogeography model of [Lemey et al., 2009] is done by sampling from the following posterior:

$$P(T, M, Q, \mu | A, \vec{t}, \vec{L}) \propto P(A|T, \mu) P(L|T, Q) P(T|N, \vec{t}) P(\mu, Q, N)$$

Some points:

- ▶ Migration history is integrated out.
- ▶ The discrete trait model contributes an additional *likelihood* term for the spatial transition matrix  $Q$ .
- ▶ The tree prior is the standard unstructured coalescent tree prior:  $N$  is the size of the metapopulation.

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent

MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Discrete trait phylogeography inference

Inference under the discrete trait phylogeography model of [Lemey et al., 2009] is done by sampling from the following posterior:

$$P(T, M, Q, \mu | A, \vec{t}, \vec{L}) \propto P(A|T, \mu) P(L|T, Q) P(T|N, \vec{t}) P(\mu, Q, N)$$

Some points:

- ▶ Migration history is integrated out.
- ▶ The discrete trait model contributes an additional *likelihood* term for the spatial transition matrix  $Q$ .
- ▶ The tree prior is the standard unstructured coalescent tree prior:  $N$  is the size of the metapopulation.
- ▶ Tends to scale easily to large numbers of demes: less complicated probability landscape than the structured coalescent.

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

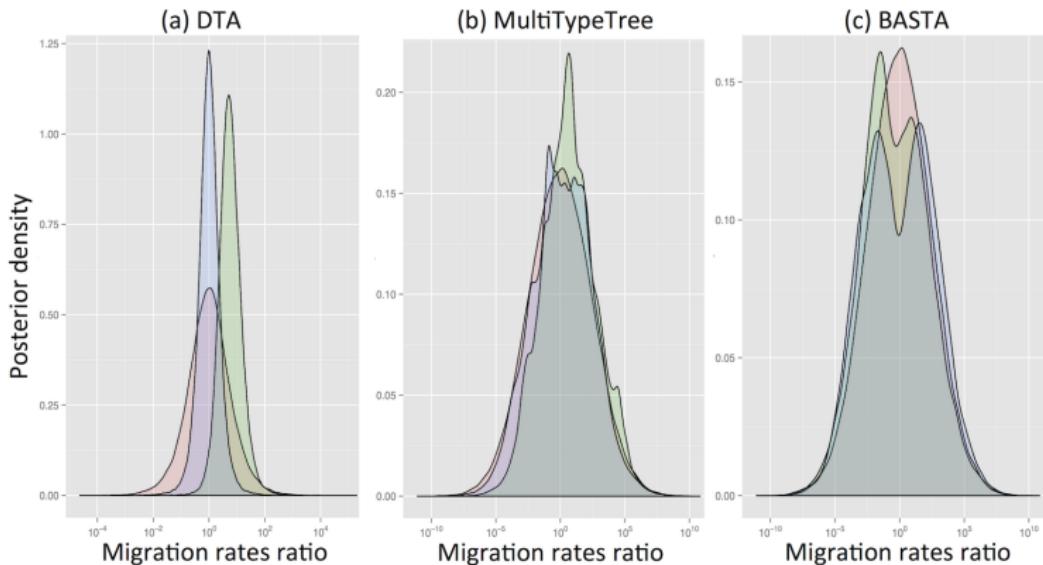
The structured  
coalescent

MultiTypeTree  
BASTA

Discrete trait  
phylogeography

# Dependence on sampling strategy

- Prior distribution
- Posterior distribution, uneven sampling (10-190)
- Posterior distribution, even sampling (100-100)



[De Maio et al., 2015]

Structured  
Populations

Structured Moran  
models

Discrete trait  
evolution

Structured  
phylogenetic inference

The structured  
coalescent  
MultiTypeTree  
BASTA

Discrete trait  
phylogeography

Bacterial population  
genetics

Coalescent with gene  
conversion

The ClonalOrigin  
model

Summarizing ARGs

Example

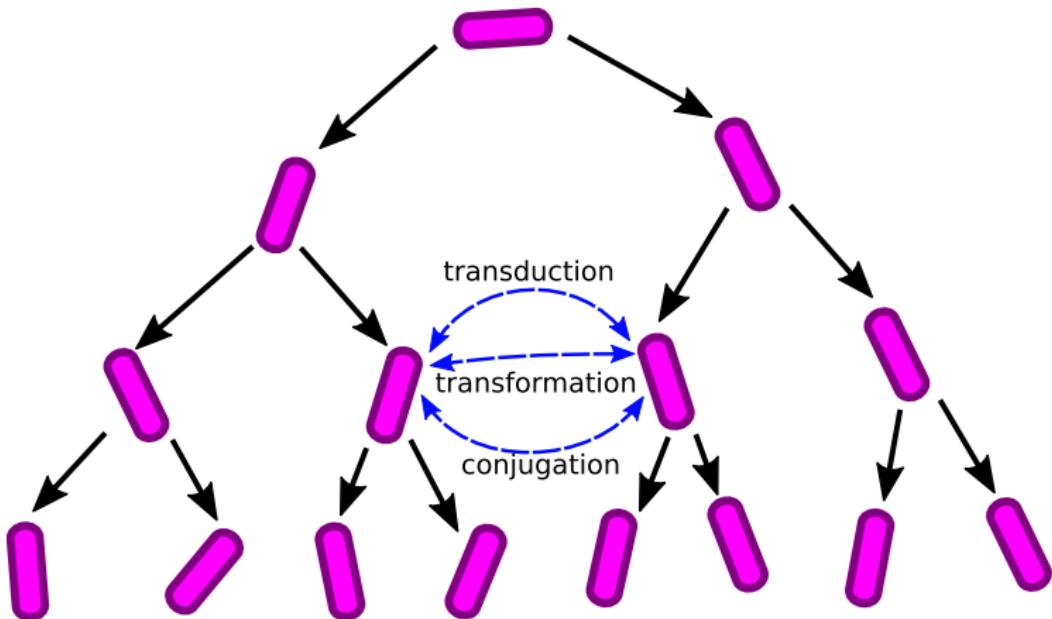
References

## Part III

### Bacter

# Cartoon bacterial population genetics

Simulations and  
Phylogeography



Bacterial population  
genetics

Coalescent with gene  
conversion

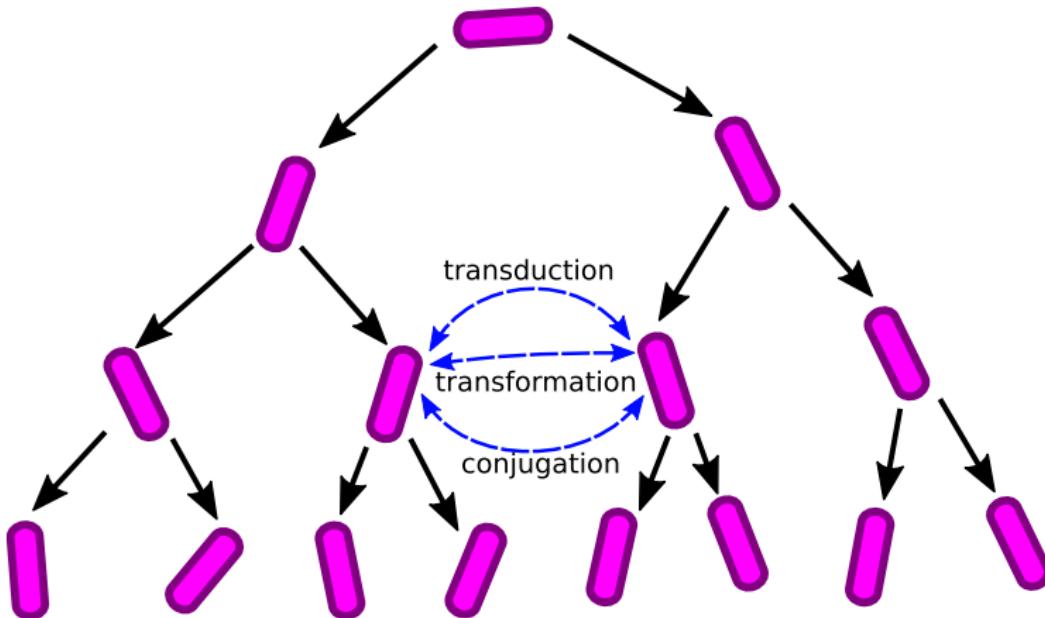
The ClonalOrigin  
model

Summarizing ARGs

Example

References

# Cartoon bacterial population genetics



- ▶ Relative recombination rates vary and are subject to selection.

Bacterial population genetics

Coalescent with gene conversion

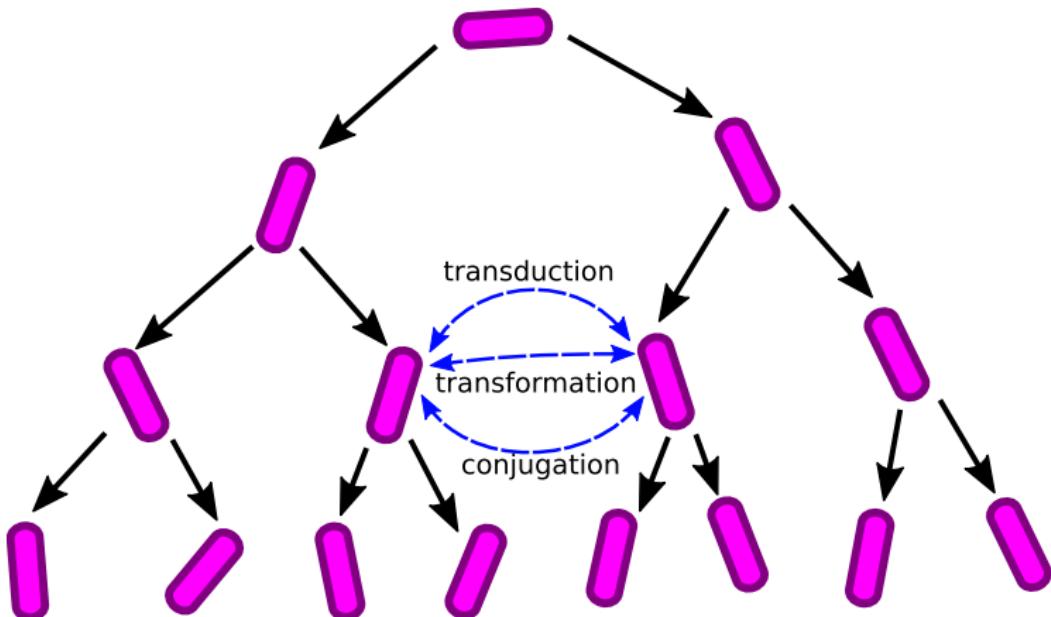
The ClonalOrigin model

Summarizing ARGs

Example

References

# Cartoon bacterial population genetics



Bacterial population  
genetics

Coalescent with gene  
conversion

The ClonalOrigin  
model

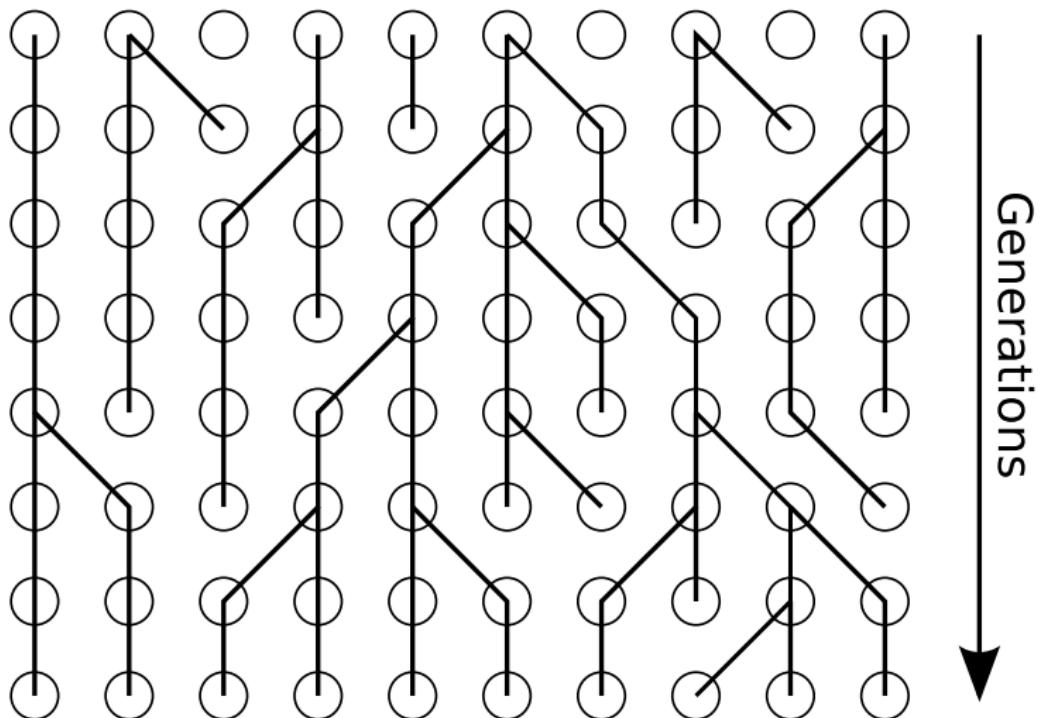
Summarizing ARGs

Example

References

- ▶ Relative recombination rates vary and are subject to selection.
- ▶ Ancestral process modelled using the *coalescent with gene conversion*.

# Coalescent with gene conversion



[Wiuf, 2000, Wiuf and Hein, 2000]

Bacterial population genetics

Coalescent with gene conversion

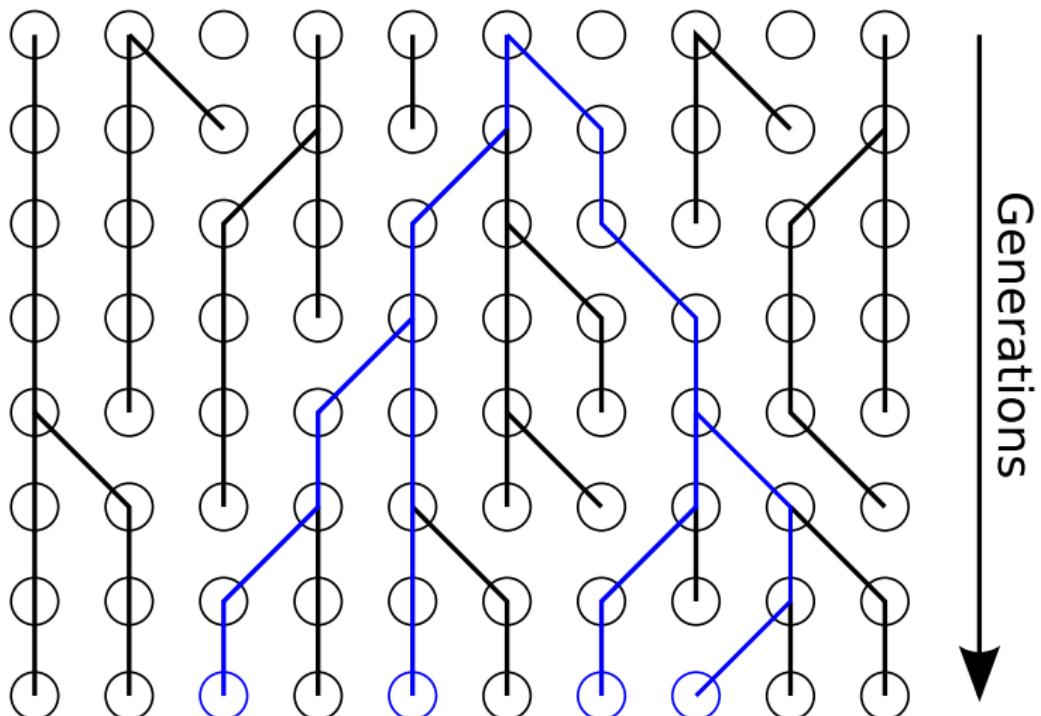
The ClonalOrigin model

Summarizing ARGs

Example

References

# Coalescent with gene conversion



[Wiuf, 2000, Wiuf and Hein, 2000]

Bacterial population genetics

Coalescent with gene conversion

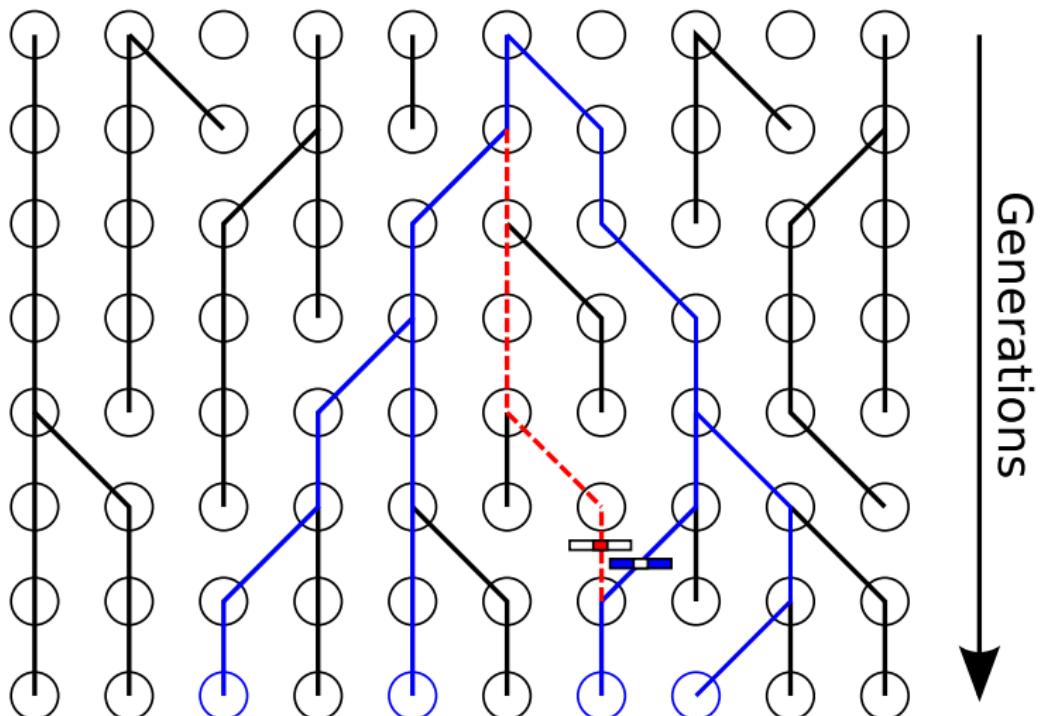
The ClonalOrigin model

Summarizing ARGs

Example

References

# Coalescent with gene conversion



[Wiuf, 2000, Wiuf and Hein, 2000]

Bacterial population  
genetics

Coalescent with gene  
conversion

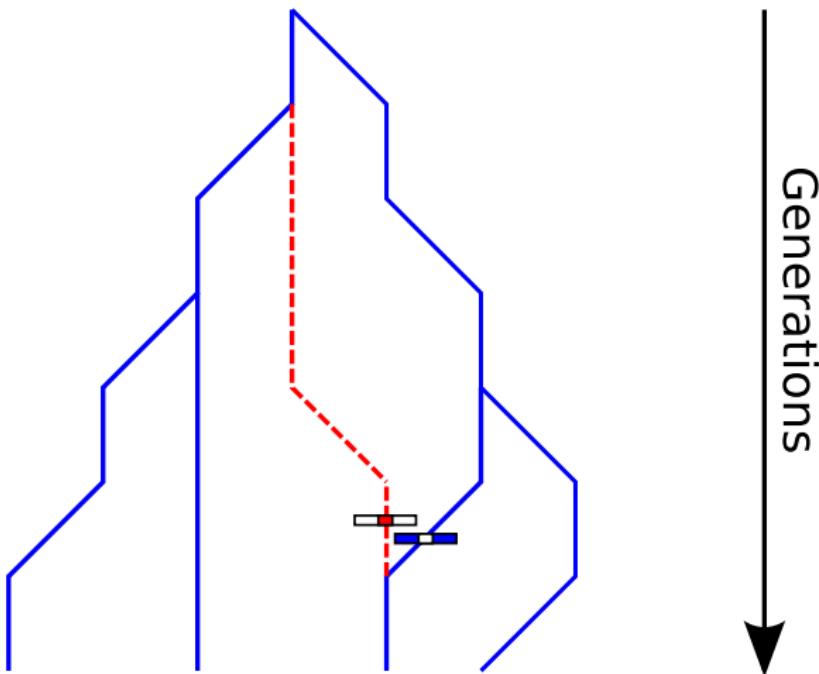
The ClonalOrigin  
model

Summarizing ARGs

Example

References

# Coalescent with gene conversion



[Wiuf, 2000, Wiuf and Hein, 2000]

Bacterial population  
genetics

Coalescent with gene  
conversion

The ClonalOrigin  
model

Summarizing ARGs

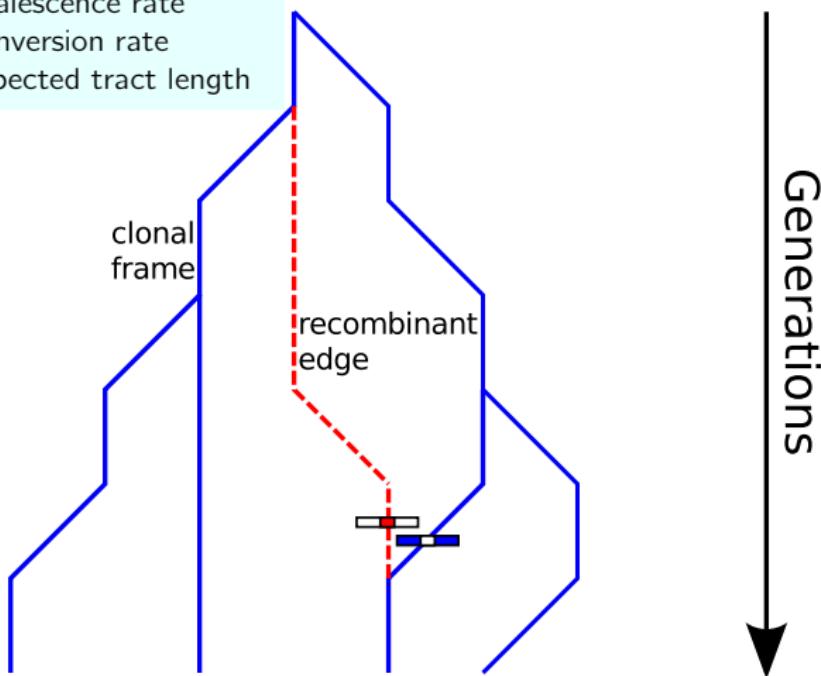
Example

References

# Coalescent with gene conversion

## Parameters

$N(t)g$	Coalescence rate
$\rho_s$	Conversion rate
$\delta$	Expected tract length



[Wiuf, 2000, Wiuf and Hein, 2000]

Bacterial population  
genetics

Coalescent with gene  
conversion

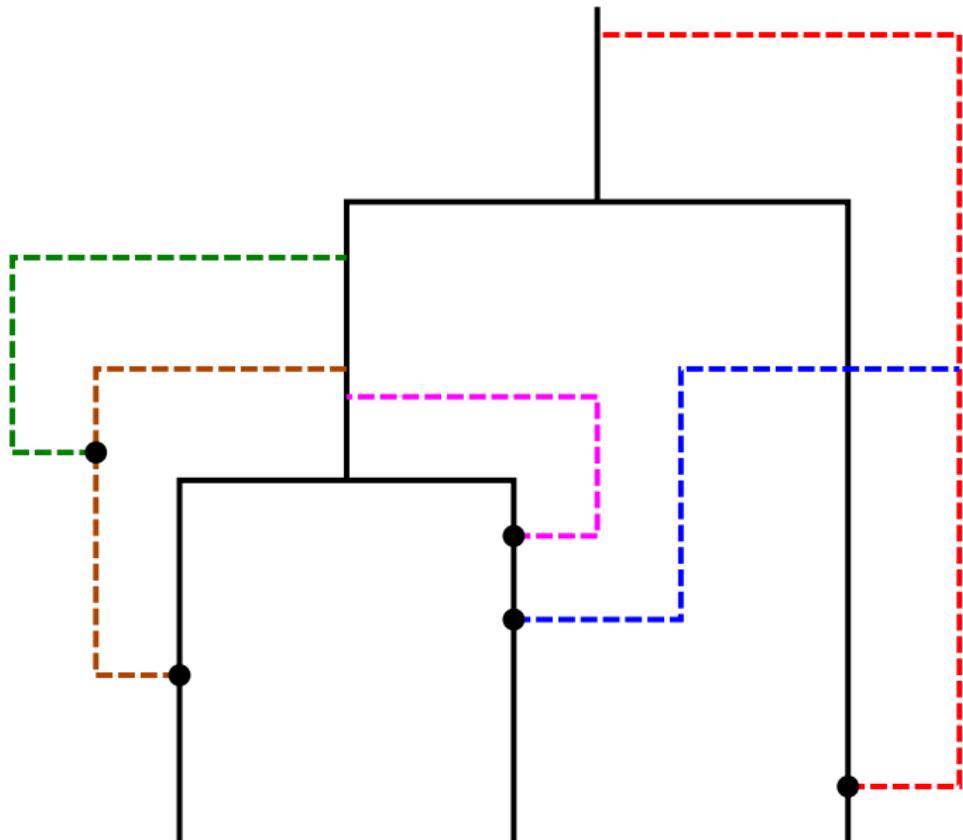
The ClonalOrigin  
model

Summarizing ARGs

Example

References

# The ClonalOrigin approximation



[Didelot et al., 2010]

Bacterial population  
genetics

Coalescent with gene  
conversion

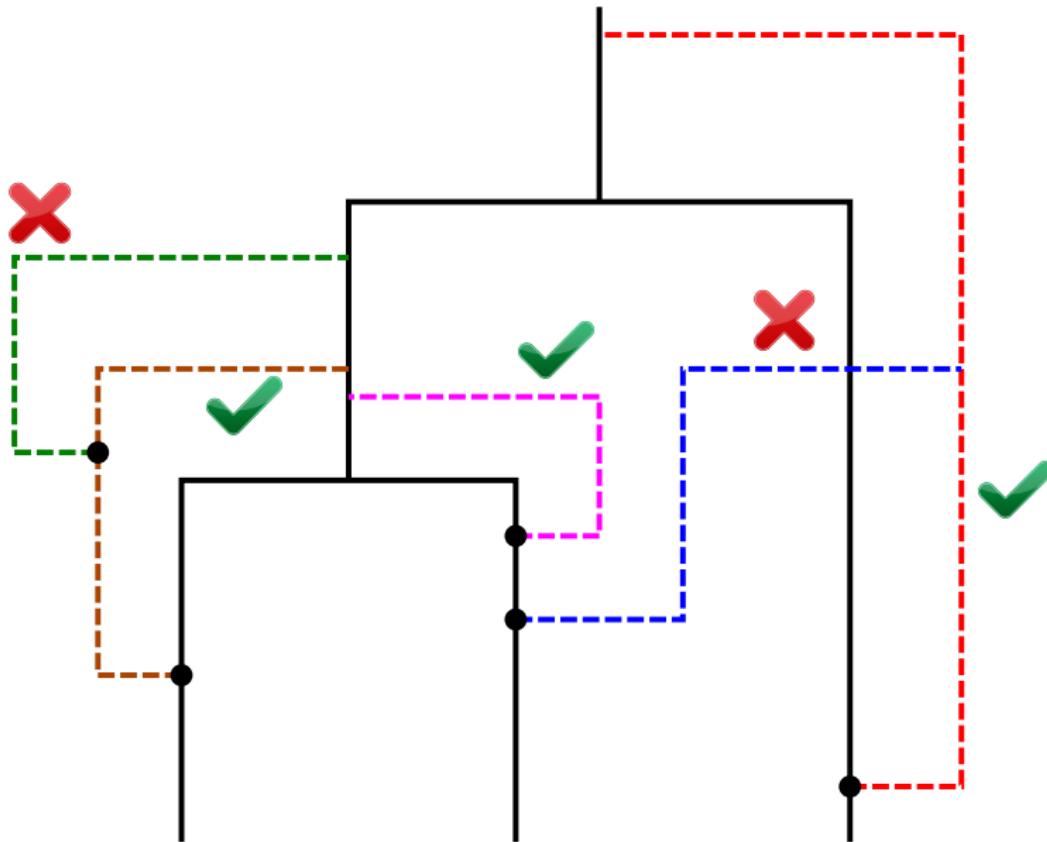
The ClonalOrigin  
model

Summarizing ARGs

Example

References

# The ClonalOrigin approximation



[Didelot et al., 2010]

Bacterial population genetics

Coalescent with gene conversion

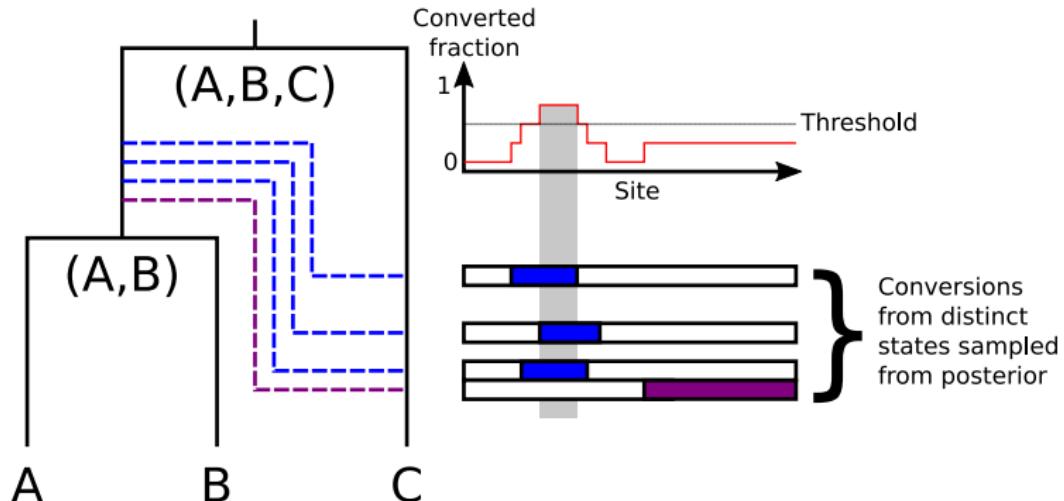
The ClonalOrigin model

Summarizing ARGs

Example

References

# Summarizing the ARG posterior



Bacterial population genetics

Coalescent with gene conversion

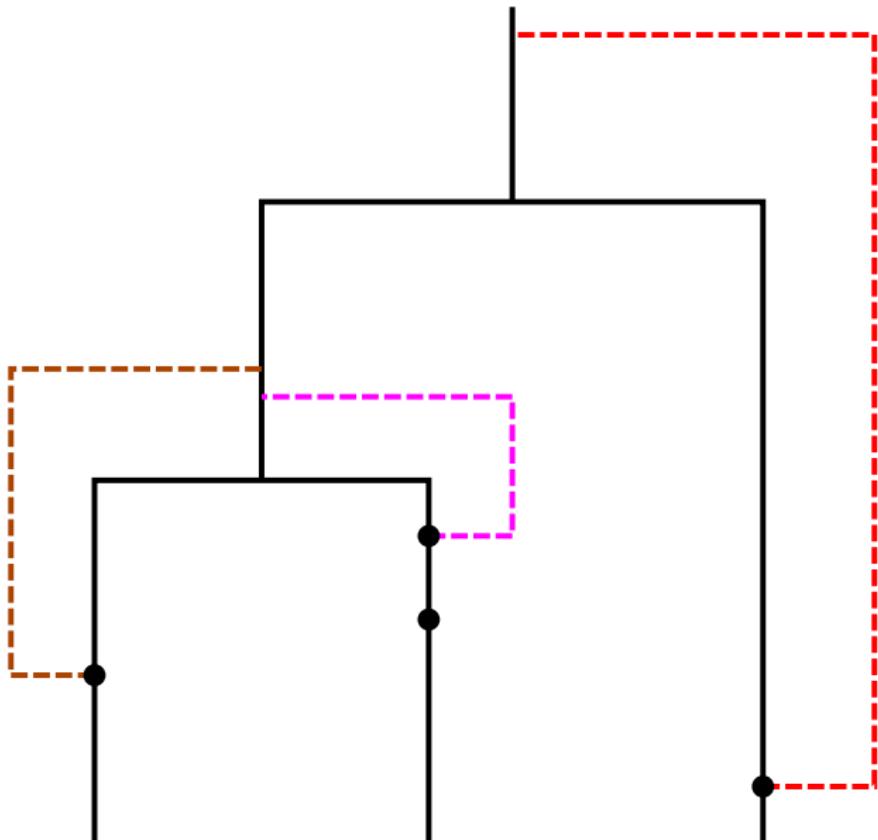
The ClonalOrigin model

Summarizing ARGs

Example

References

# The ClonalOrigin approximation



[Didelot et al., 2010]

Bacterial population  
genetics

Coalescent with gene  
conversion

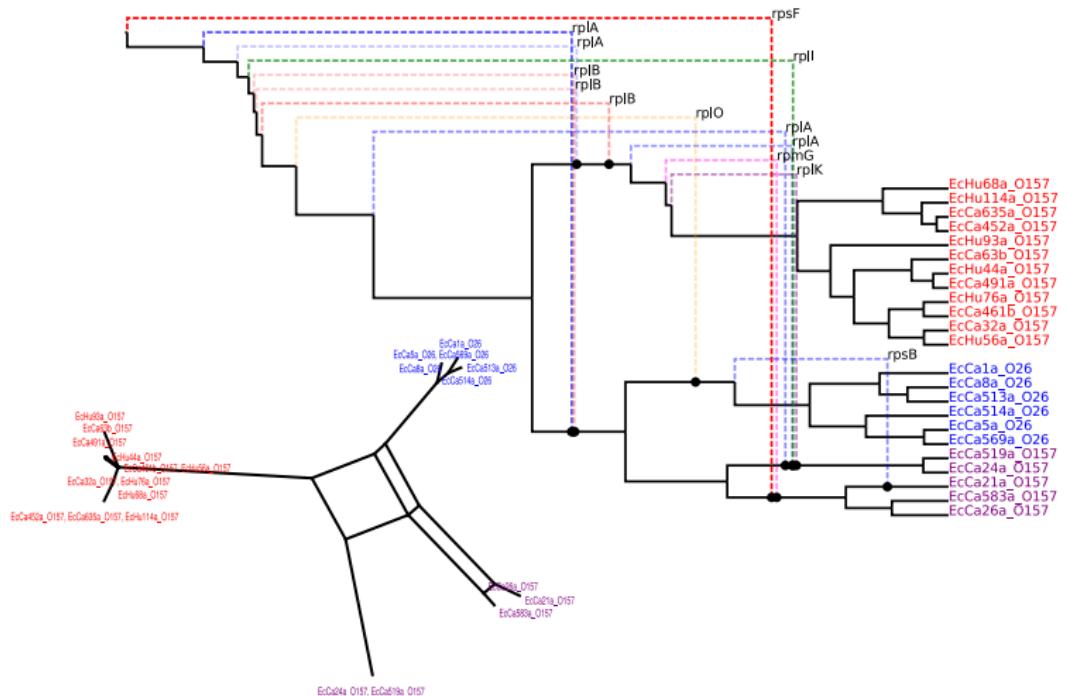
The ClonalOrigin  
model

Summarizing ARGs

Example

References

## Application to *Escherichia coli*



## Bacterial population genetics

## Coalescent with gene conversion

## The ClonalOrigin model

## Summarizing ARGs

## Example

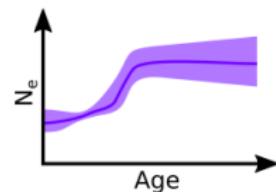
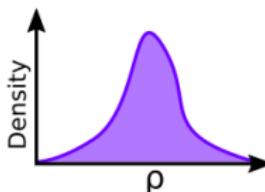
## References

# bacter

## Bacterial phylogenetics in BEAST 2

[Installation & Usage Tutorial](#)[Source Code](#)

Bacter is a [BEAST 2](#) package which facilitates joint Bayesian inference of ancestral recombination graphs (ARG) and related parameters from bacterial sequence alignments under the [ClonalOrigin](#) model.



### ARGs

Bacter allows you to infer the ARG that gave rise to your sampled data, including both clonal frame and the recombinant edges. Loci and sites affected by each conversion are also inferred.

### Parameters

For informative data sets, Bacter can sample posterior distributions for parameters such as the recombination rate and expected converted tract length jointly with the ARG. For serially sampled data, joint estimation of the molecular clock rate is also possible.

### Population dynamics

Bacter includes support for producing Bayesian Skyline Plot (BSP) inferences of bacterial population dynamics. This inference may actually be more powerful when recombination is present in your data set.

<http://tgvaughan.github.io/bacter>

Bacterial population  
genetics

Coalescent with gene  
conversion

The ClonalOrigin  
model

Summarizing ARGs

Example

References

# References |

- De Maio, N., Wu, C.-H., O'Reilly, K. M., and Wilson, D. (2015). New routes to phylogeography: A bayesian structured coalescent approximation. *PLoS Genet*, 11(8):e1005421.
- Didelot, X., Lawson, D., Daarling, A., and Falush, D. (2010). Inference of homologous recombination in bacteria using whole-genome sequences. *Genetics*, 186:1435.
- Hudson, R. R. (1990). Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology*, 7:1.
- Lemey, P., Rambaut, A., Drummond, A. J., and Suchard, M. A. (2009). Bayesian phylogeography finds its roots. *PLoS Comput Biol*, 5(9):e1000520.
- Notohara, M. (1990). The coalescent and the genealogical process in geographically structured population. *J Math Biol*, 29(1):59–75.
- Wiuf, C. (2000). A coalescence approach to gene conversion. *Theor Popul Biol*, 57(4):357–367.
- Wiuf, C. and Hein, J. (2000). The coalescent with gene conversion. *Genetics*, 155(1):451–462.

Bacterial population genetics

Coalescent with gene conversion

The ClonalOrigin model

Summarizing ARGs

Example

References