

# Computational Biology

Lecturers:

Tanja Stadler, Carsten Magnus & Tim Vaughan

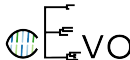
Teaching Assistants:

Jūlija Pečerska, Jérémie Sciré,  
Sarah Nadeau & Marc Manceau

Computational Evolution

Department of Biosystems Science and Engineering

HS 2019



c09-Phylogenetics

Introduction

Population dynamic models

Phylogenetic models

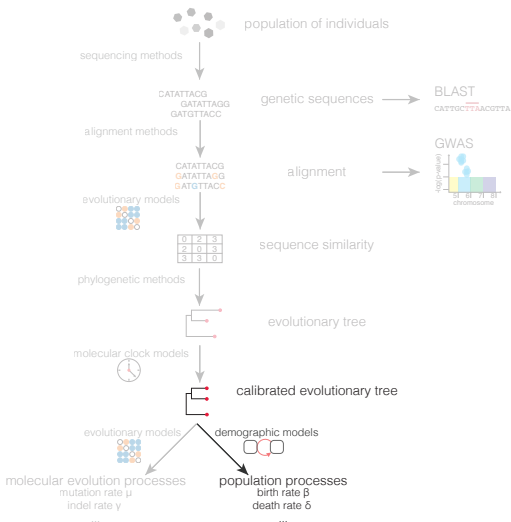
Probabilities

Applications

References

# Continuous traits and comparative methods: Discuss questions.

- ❓ In a Fisher's exact test, how would you calculate which values for one of the cells in the contingency table would lead to a rejection of the null hypothesis, given that row and column sums remain the same?
- ❓ Is the Brownian motion model a good model for all continuous traits? Could you imagine situations where this is not the case and which assumption in this model could be violated?
- ❓ Do you think it is a good strategy to first determine the species tree and then look at character evolution, or would a co-estimation of characters and the phylogeny make more sense?



## c09-Phylodynamics

### Introduction

### Population dynamic models

### Phylodynamic models

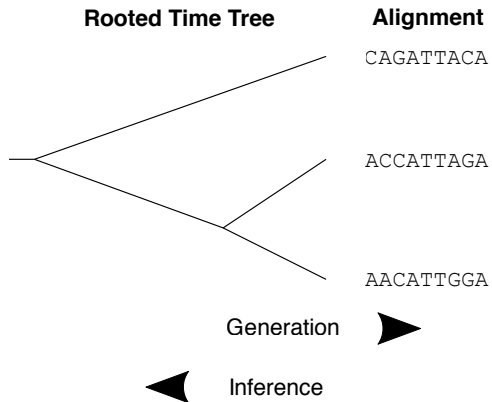
### Probabilities

### Applications

## References

# What have we learned so far?

CB



c09-Phylodynamics

Introduction

Population dynamic models

Phylogenetic models

Probabilities

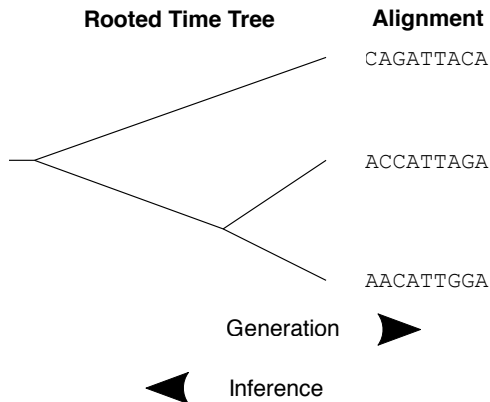
Applications

References

# What have we learned so far?

CB

c09-Phylogenetics  
Introduction  
Population dynamic models  
Phylogenetic models  
Probabilities  
Applications  
References



## Question

What is the process that generated the phylogenetic tree?

# Phylogenetic trees encode past macroevolutionary dynamics

Macroevolution: individuals = species

c09-Phylogenetics

Introduction

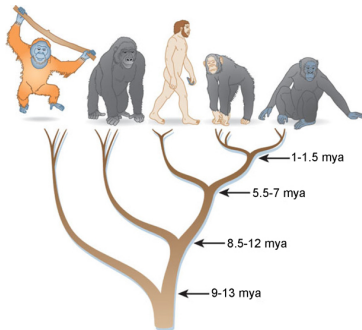
Population dynamic models

Phylogenetic models

Probabilities

Applications

References

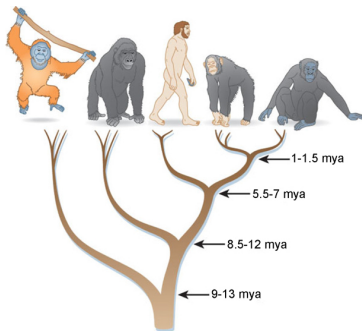


# Phylogenetic trees encode past macroevolutionary dynamics

Macroevolution: individuals = species

## ► (Molecular) Evolution

(Genetic) makeup of species changes through time.



c09-Phylogenetics

Introduction

Population dynamic models

Phylogenetic models

Probabilities

Applications

References

# Phylogenetic trees encode past macroevolutionary dynamics

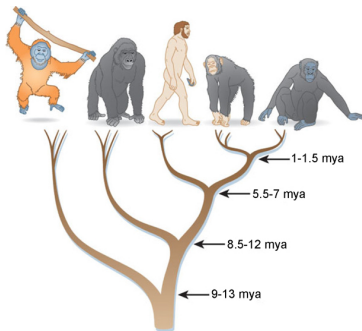
Macroevolution: individuals = species

## ► (Molecular) Evolution

(Genetic) makeup of species changes through time.

## ► Phylogenetics

Phylogeny displays species relationships.



c09-Phylogenetics

Introduction

Population dynamic models

Phylogenetic models

Probabilities

Applications

References



# Phylogenetic trees encode past macroevolutionary dynamics

c09-Phylogenetics

Introduction

Population dynamic models

Phylogenetic models

Probabilities

Applications

References

Macroevolution: individuals = species

- **(Molecular) Evolution**

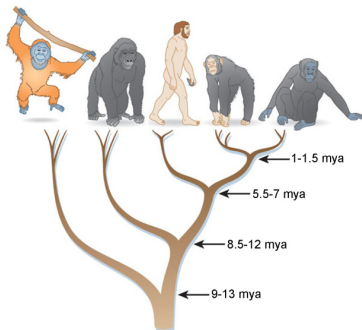
(Genetic) makeup of species changes through time.

- **Phylogenetics**

Phylogeny displays species relationships.

- **Phylogenetics**

Population dynamics is the speciation and extinction process.



# Phylogenetic trees encode past epidemiological dynamics

CB

c09-Phylodynamics

Introduction

Population dynamic models

Phylodynamic models

Probabilities

Applications

References

Epidemiology: individuals = infected hosts

# Phylogenetic trees encode past epidemiological dynamics

CB

c09-Phylodynamics

Introduction

Population dynamic models

Phylodynamic models

Probabilities

Applications

References

Epidemiology: individuals = infected hosts

## ► **Evolution**

Pathogen is evolving through time.

# Phylogenetic trees encode past epidemiological dynamics

CB

c09-Phylodynamics

Introduction

Population dynamic models

Phylodynamic models

Probabilities

Applications

References

Epidemiology: individuals = infected hosts

- ▶ **Evolution**

Pathogen is evolving through time.

- ▶ **Phylogenetics**

Phylogeny displays transmission history.

# Phylogenetic trees encode past epidemiological dynamics

CB

c09-Phylodynamics

Introduction

Population dynamic models

Phylogenetic models

Probabilities

Applications

References

Epidemiology: individuals = infected hosts

- ▶ **Evolution**

Pathogen is evolving through time.

- ▶ **Phylogenetics**

Phylogeny displays transmission history.

- ▶ **Phylodynamics**

Population dynamics is the transmission and becoming non-infectious process.

# Phylogenetic trees encode past epidemiological dynamics

CB

c09-Phylodynamics

Introduction

Population dynamic models

Phylodynamic models

Probabilities

Applications

References

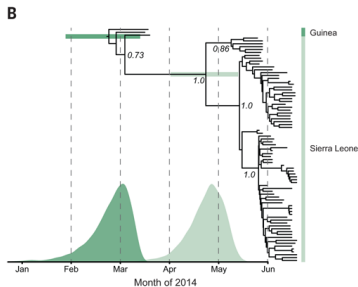


Figure adapted from [Gire et al., 2014]

Phylogenetic tree of Ebola in West Africa.

# Phylogenetic trees encode past epidemiological dynamics

CB

c09-Phylodynamics

Introduction

Population dynamic models

Phylodynamic models

Probabilities

Applications

References

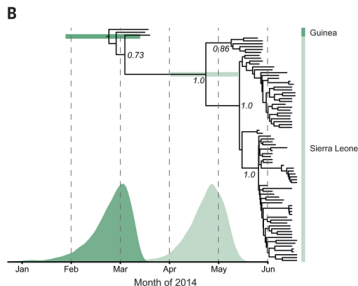


Figure adapted from [Gire et al., 2014]

Phylogenetic tree of Ebola in West Africa.

- What is the basic reproductive number  $R_0$  of an emerging epidemic? ( $R_0$  is the average number of secondary infections caused by a single infected individual at the start of an epidemic.)

- ▶ **Immunology:** individuals = B cells

c09-Phylodynamics

Introduction

Population dynamic models

Phylogenetic models

Probabilities

Applications

References



## c09-Phylodynamics

### Introduction

### Population dynamic models

### Phylodynamic models

### Probabilities

### Applications

### References

- ▶ **Immunology:** individuals = B cells
  - Phylogeny displays B cell differentiation (through somatic hypermutation).

[c09-Phylodynamics](#)[Introduction](#)[Population dynamic models](#)[Phylogenetic models](#)[Probabilities](#)[Applications](#)[References](#)

- ▶ **Immunology:** individuals = B cells
  - Phylogeny displays B cell differentiation (through somatic hypermutation).
  - Population dynamics is the B cell generation and loss process.

[c09-Phylodynamics](#)[Introduction](#)[Population dynamic models](#)[Phylodynamic models](#)[Probabilities](#)[Applications](#)[References](#)

- ▶ **Immunology:** individuals = B cells
  - Phylogeny displays B cell differentiation (through somatic hypermutation).
  - Population dynamics is the B cell generation and loss process.
- ▶ **Cancer:** individuals = cells

- ▶ **Immunology:** individuals = B cells
  - Phylogeny displays B cell differentiation (through somatic hypermutation).
  - Population dynamics is the B cell generation and loss process.
- ▶ **Cancer:** individuals = cells
  - Phylogeny displays relationship of different cancer cells and healthy cells.

- ▶ **Immunology:** individuals = B cells
  - Phylogeny displays B cell differentiation (through somatic hypermutation).
  - Population dynamics is the B cell generation and loss process.
- ▶ **Cancer:** individuals = cells
  - Phylogeny displays relationship of different cancer cells and healthy cells.
  - Population dynamics is the spread and loss of cell types.

- ▶ **Immunology:** individuals = B cells
  - Phylogeny displays B cell differentiation (through somatic hypermutation).
  - Population dynamics is the B cell generation and loss process.
- ▶ **Cancer:** individuals = cells
  - Phylogeny displays relationship of different cancer cells and healthy cells.
  - Population dynamics is the spread and loss of cell types.
- ▶ **Languages:** individuals = languages

- ▶ **Immunology:** individuals = B cells
  - Phylogeny displays B cell differentiation (through somatic hypermutation).
  - Population dynamics is the B cell generation and loss process.
- ▶ **Cancer:** individuals = cells
  - Phylogeny displays relationship of different cancer cells and healthy cells.
  - Population dynamics is the spread and loss of cell types.
- ▶ **Languages:** individuals = languages
  - Phylogeny displays language evolution.

- ▶ **Immunology:** individuals = B cells
  - Phylogeny displays B cell differentiation (through somatic hypermutation).
  - Population dynamics is the B cell generation and loss process.
- ▶ **Cancer:** individuals = cells
  - Phylogeny displays relationship of different cancer cells and healthy cells.
  - Population dynamics is the spread and loss of cell types.
- ▶ **Languages:** individuals = languages
  - Phylogeny displays language evolution.
  - Population dynamics is the gain and loss of languages.



## c09-Phylodynamics

### Introduction

### Population dynamic models

### Phylogenetic models

### Probabilities

### Applications

## References

- *Population dynamics* models the birth and death of individuals (species, infected hosts, B cells, cancer cells, languages).

## c09-Phylodynamics

### Introduction

### Population dynamic models

### Phylodynamic models

### Probabilities

### Applications

### References

- ▶ *Population dynamics* models the birth and death of individuals (species, infected hosts, B cells, cancer cells, languages).
- ▶ The birth and death process gives rise to a phylogenetic tree.

## c09-Phylodynamics

### Introduction

#### Population dynamic models

#### Phylogenetic models

#### Probabilities

#### Applications

## References

- ▶ *Population dynamics* models the birth and death of individuals (species, infected hosts, B cells, cancer cells, languages).
- ▶ The birth and death process gives rise to a phylogenetic tree.
- ▶ *Phylodynamics* aims to understand and quantify the population dynamics based on a phylogenetic tree. Today we quantify birth and death dynamics given the phylogenetic tree (and then also  $R_0$ ).

c09-Phylodynamics

Introduction

Population dynamic models

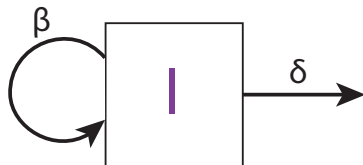
Phylogenetic models

Probabilities

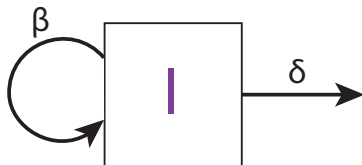
Applications

References

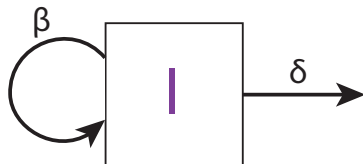
Population dynamic models.



- Rate of birth of new individuals per individual in I:  $\beta$

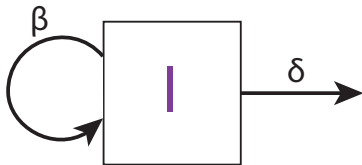


- Rate of birth of new individuals per individual in I:  $\beta$
- Rate of death per individual in I:  $\delta$



- Rate of birth of new individuals per individual in I:  $\beta$
- Rate of death per individual in I:  $\delta$

Such a process is called a (linear) birth-death process.



Consider the fate of one individual:

c09-Phylodynamics

Introduction

Population dynamic models

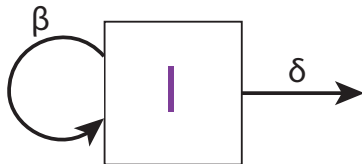
Phylodynamic models

Probabilities

Applications

References





Consider the fate of one individual:

- The probability of giving birth to another individual in a very small time step  $\Delta t$  is  $\beta \Delta t$ .

c09-Phylodynamics

Introduction

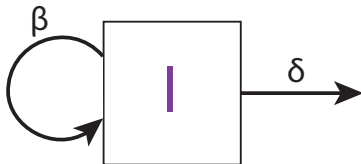
Population dynamic models

Phylogenetic models

Probabilities

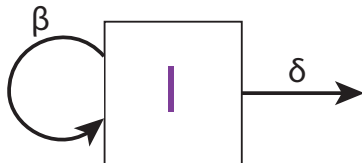
Applications

References



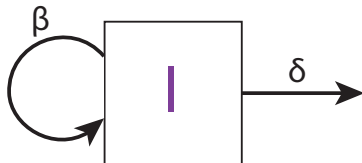
Consider the fate of one individual:

- ▶ The probability of giving birth to another individual in a very small time step  $\Delta t$  is  $\beta \Delta t$ .
- ▶ The probability of dying in a very small time step  $\Delta t$  is  $\delta \Delta t$ .



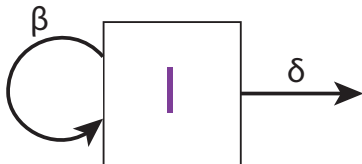
Consider the fate of one individual:

- ▶ The probability of giving birth to another individual in a very small time step  $\Delta t$  is  $\beta \Delta t$ .
- ▶ The probability of dying in a very small time step  $\Delta t$  is  $\delta \Delta t$ .
- ▶ The waiting time to a birth event is exponentially distributed with parameter  $\beta$ .

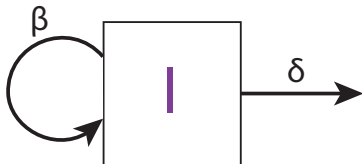


Consider the fate of one individual:

- ▶ The probability of giving birth to another individual in a very small time step  $\Delta t$  is  $\beta \Delta t$ .
- ▶ The probability of dying in a very small time step  $\Delta t$  is  $\delta \Delta t$ .
- ▶ The waiting time to a birth event is exponentially distributed with parameter  $\beta$ .
- ▶ The waiting time to the first event (birth or death) is exponentially distributed with parameter  $\beta + \delta$  (minimum of two exponentially distributed random variables with rates  $r_1, r_2$  is exponentially distributed with the rate  $r_1 + r_2$ ).



Consider the fate of  $N$  individuals:



Consider the fate of  $N$  individuals:

- The waiting time to the first event (birth or death) is exponentially distributed with parameter  $N(\beta + \delta)$ .

# An example for the dynamics of a population

CB

c09-Phylodynamics

Introduction

Population dynamic models

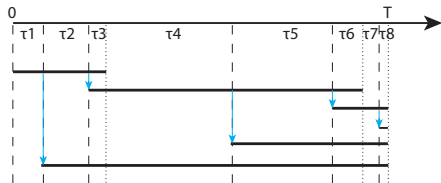
Phylodynamic models

Probabilities

Applications

References

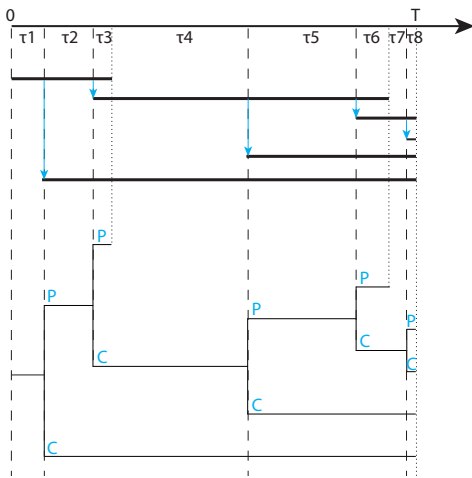
# From population dynamics to phylogenetic trees



This diagram illustrates the full population dynamics of a birth-death process which starts with one individual and is stopped after time  $T$ . Each solid black line is the lifetime of an individual. Blue arrows are birth events.



# From population dynamics to phylogenetic trees



A *complete population tree* displays the full population dynamics. The labels P and C illustrate the parent-child associations.

## c09-Phylodynamics

### Introduction

### Population dynamic models

### Phylodynamic models

### Probabilities

### Applications

### References

## c09-Phylodynamics

Introduction

Population dynamic models

Phylodynamic models

Probabilities

Applications

References

Phylodynamic models.

A phylodynamic model adds a sampling process of individuals to the population dynamics (sampling is equivalent to sequencing the individual and adding it to the phylogenetic tree). The simplest phylodynamic model is:

## c09-Phylodynamics

Introduction

Population dynamic models

Phylodynamic models

Probabilities

Applications

References

A phylodynamic model adds a sampling process of individuals to the population dynamics (sampling is equivalent to sequencing the individual and adding it to the phylogenetic tree). The simplest phylodynamic model is:

- Birth rate  $\beta$

## c09-Phylodynamics

Introduction

Population dynamic models

Phylodynamic models

Probabilities

Applications

References

A phylodynamic model adds a sampling process of individuals to the population dynamics (sampling is equivalent to sequencing the individual and adding it to the phylogenetic tree). The simplest phylodynamic model is:

- ▶ Birth rate  $\beta$
- ▶ Death rate  $\delta$

## c09-Phylodynamics

Introduction

Population dynamic models

Phylodynamic models

Probabilities

Applications

References

A phylodynamic model adds a sampling process of individuals to the population dynamics (sampling is equivalent to sequencing the individual and adding it to the phylogenetic tree). The simplest phylodynamic model is:

- ▶ Birth rate  $\beta$
- ▶ Death rate  $\delta$
- ▶ Process duration  $T$

## c09-Phylodynamics

Introduction

Population dynamic models

Phylodynamic models

Probabilities

Applications

References

A phylodynamic model adds a sampling process of individuals to the population dynamics (sampling is equivalent to sequencing the individual and adding it to the phylogenetic tree). The simplest phylodynamic model is:

- ▶ Birth rate  $\beta$
- ▶ Death rate  $\delta$
- ▶ Process duration  $T$
- ▶ Extant tip sampling probability  $\rho$

## c09-Phylodynamics

Introduction

Population dynamic models

Phylodynamic models

Probabilities

Applications

References

A phylodynamic model adds a sampling process of individuals to the population dynamics (sampling is equivalent to sequencing the individual and adding it to the phylogenetic tree). The simplest phylodynamic model is:

- ▶ Birth rate  $\beta$
- ▶ Death rate  $\delta$
- ▶ Process duration  $T$
- ▶ Extant tip sampling probability  $\rho$
- ▶ Extinct tip sampling probability  $\phi$

## c09-Phylodynamics

Introduction

Population dynamic models

Phylodynamic models

Probabilities

Applications

References

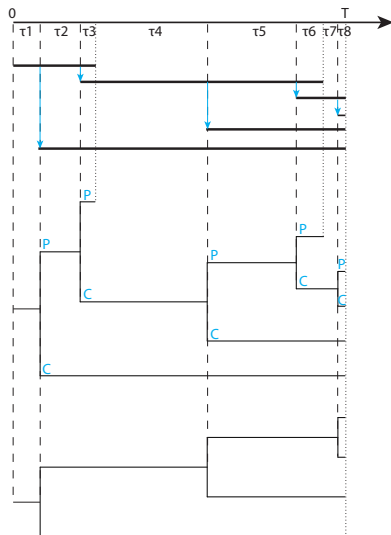


A phylodynamic model adds a sampling process of individuals to the population dynamics (sampling is equivalent to sequencing the individual and adding it to the phylogenetic tree). The simplest phylodynamic model is:

- ▶ Birth rate  $\beta$
- ▶ Death rate  $\delta$
- ▶ Process duration  $T$
- ▶ Extant tip sampling probability  $\rho$
- ▶ Extinct tip sampling probability  $\phi$

We will now assume  $\rho = 1$ ,  $\phi = 0$ . For macroevolution, that means no fossil sampling and complete extant species sampling. The subtree of the complete population tree connecting the sampled individuals, and ignoring the parent-children labels, is called the *phylogenetic tree*. The phylogenetic tree is the tree we infer from data.

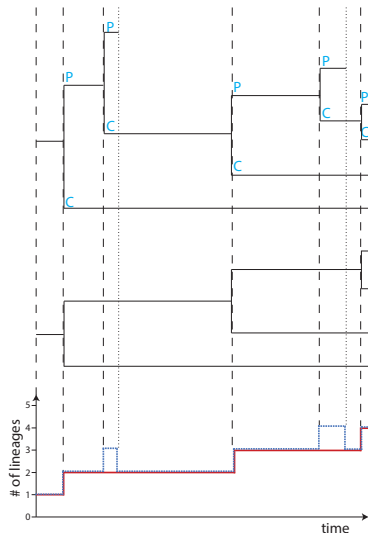
# From population dynamics to phylogenetic trees



The phylogenetic tree with  $\rho = 1, \phi = 0$  displays the dynamics giving rise to the surviving individuals at time T.

# Lineages-through-time plot

CB



Plotting the number of lineages (y-axis) vs time (x-axis) is called the lineages-through-time (LTT) plot. The LTT plot of the complete tree (blue; dashed) shows the population size through time. The LTT plot of the phylogenetic tree (red) shows the number of surviving lineages through time.

c09-Phylogenetics

Introduction

Population dynamic models

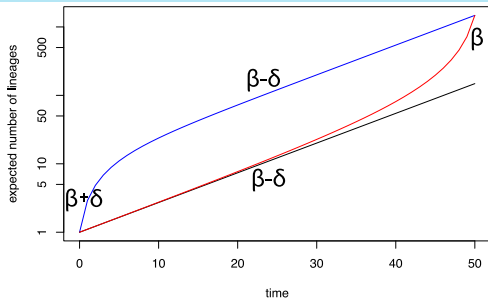
Phylogenetic models

Probabilities

Applications

References

# Can we estimate the birth and death rates from reconstructed phylogenies?



Average LTTs for a population of age  $T = 50$ :

c09-Phyldynamics

Introduction

Population dynamic models

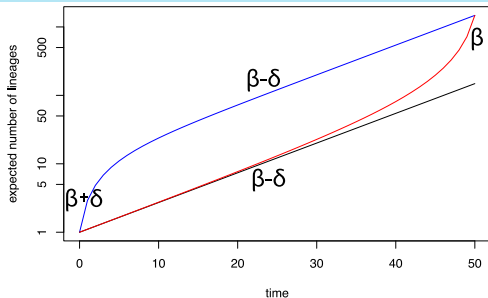
Phyldynamic models

Probabilities

Applications

References

# Can we estimate the birth and death rates from reconstructed phylogenies?



Average LTTs for a population of age  $T = 50$ :

- Red: Average number of lineages in the phylogenetic tree.

c09-Phyldynamics

Introduction

Population dynamic models

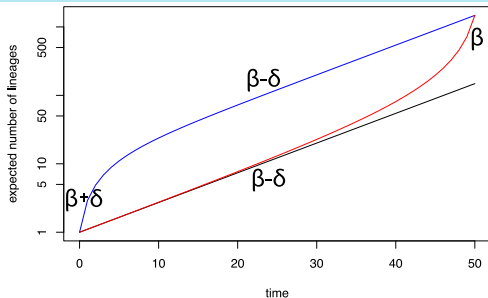
Phyldynamic models

Probabilities

Applications

References

# Can we estimate the birth and death rates from reconstructed phylogenies?



Average LTTs for a population of age  $T = 50$ :

- ▶ Red: Average number of lineages in the phylogenetic tree.
- ▶ Blue: Average of surviving population trajectories.

c09-Phylogenetics

Introduction

Population dynamic models

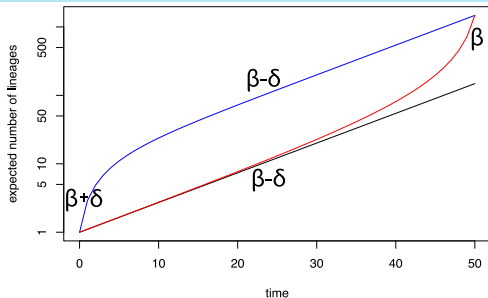
Phylogenetic models

Probabilities

Applications

References

# Can we estimate the birth and death rates from reconstructed phylogenies?



Average LTTs for a population of age  $T = 50$ :

- ▶ Red: Average number of lineages in the phylogenetic tree.
- ▶ Blue: Average of surviving population trajectories.
- ▶ Black: Exponential growth curve (linear with slope  $\beta - \delta$  on the log scale).

c09-Phyldynamics

Introduction

Population dynamic models

Phyldynamic models

Probabilities

Applications

References

# Can we estimate the birth and death rates from reconstructed phylogenies?

c09-Phylogenetics

Introduction

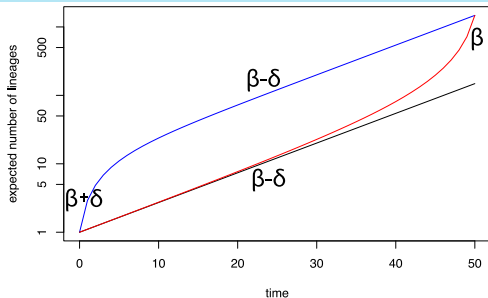
Population dynamic models

Phylogenetic models

Probabilities

Applications

References



Average LTTs for a population of age  $T = 50$ :

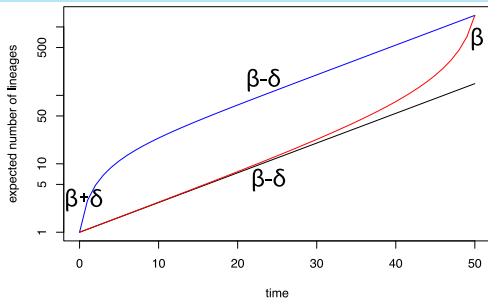
- ▶ Red: Average number of lineages in the phylogenetic tree.
- ▶ Blue: Average of surviving population trajectories.
- ▶ Black: Exponential growth curve (linear with slope  $\beta - \delta$  on the log scale).

The early blue part is called push-of-the-past, the late red part is called pull-of-the-present [Nee et al., 1994a].

- ▶ We observe a push of the past as only individuals with a



# Can we estimate the birth and death rates from reconstructed phylogenies?



- We observe a push-of-the-past as only individuals with a quick replication early on will produce surviving populations.
- We observe the pull-of-the-present (i.e. an apparent acceleration in diversification towards the present) as the very recent lineages did not yet have time to go extinct.

c09-Phylogenetics

Introduction

Population dynamic models

Phylogenetic models

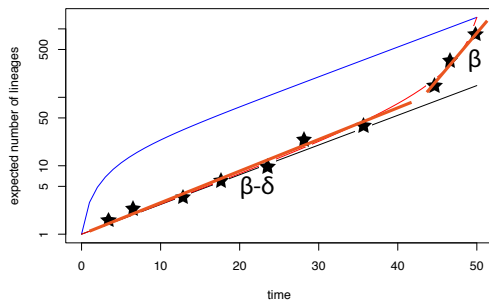
Probabilities

Applications

References

# Estimating $\beta$ and $\delta$

CB



c09-Phylodynamics

Introduction

Population dynamic models

Phylodynamic models

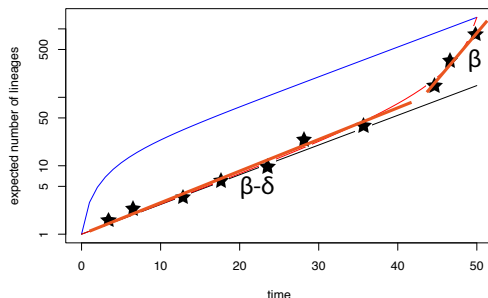
Probabilities

Applications

References

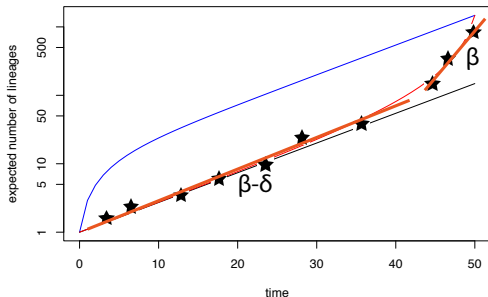
# Estimating $\beta$ and $\delta$

CB

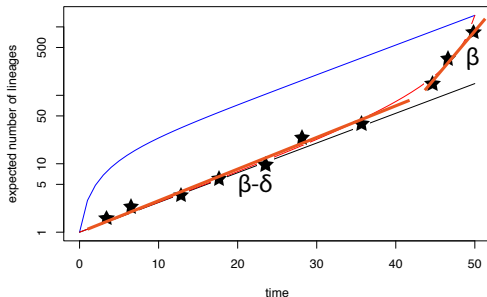


- We could fit a regression line to the early branching events and estimate its slope, this is an estimate for  $\beta - \delta$ .

c09-Phylogenetics  
Introduction  
Population dynamic models  
Phylogenetic models  
Probabilities  
Applications  
References



- We could fit a regression line to the early branching events and estimate its slope, this is an estimate for  $\beta - \delta$ .
- We could fit a regression line to the late branching events and estimate its slope, this is an estimate for  $\beta$ .



- ▶ We could fit a regression line to the early branching events and estimate its slope, this is an estimate for  $\beta - \delta$ .
- ▶ We could fit a regression line to the late branching events and estimate its slope, this is an estimate for  $\beta$ .
- ▶ Problem: It is not clear how to incorporate the variances into the regression, and how to choose the time interval for the two regression lines!

## c09-Phylogenetics

Introduction

Population dynamic models

Phylogenetic models

Probabilities

Applications

References

Probabilities.

## c09-Phylodynamics

Introduction

Population dynamic models

Phylodynamic models

Probabilities

Applications

References

- **Phylogenetic likelihood:**  $L(\mathcal{T}, Q) = P(A|\mathcal{T}, Q)$

- **Phylogenetic likelihood:**  $L(\mathcal{T}, Q) = P(A|\mathcal{T}, Q)$
- **Phylogenetic likelihood:**  $L(\eta = (\beta, \delta, T)) = P(\mathcal{T}|\eta)$



- **Phylogenetic likelihood:**  $L(\mathcal{T}, Q) = P(A|\mathcal{T}, Q)$
- **Phylodynamic likelihood:**  $L(\eta = (\beta, \delta, T)) = P(\mathcal{T}|\eta)$

We aim to determine the maximum likelihood estimate for the parameters  $(\beta, \delta, T)$ , *given a phylogenetic tree*. Note that in this lecture, the age of the process  $T$  is assumed to be fixed to a known value (this will be relaxed in lecture 11).

- **Phylogenetic likelihood:**  $L(\mathcal{T}, Q) = P(A|\mathcal{T}, Q)$
- **Phylogenetic likelihood:**  $L(\eta = (\beta, \delta, T)) = P(\mathcal{T}|\eta)$

We aim to determine the maximum likelihood estimate for the parameters  $(\beta, \delta, T)$ , *given a phylogenetic tree*. Note that in this lecture, the age of the process  $T$  is assumed to be fixed to a known value (this will be relaxed in lecture 11).

The work on the phylogenetic likelihood began in the 1990s in Oxford [Nee et al., 1994b].

- **Phylogenetic likelihood:**  $L(\mathcal{T}, Q) = P(A|\mathcal{T}, Q)$
- **Phylodynamic likelihood:**  $L(\eta = (\beta, \delta, T)) = P(\mathcal{T}|\eta)$

We aim to determine the maximum likelihood estimate for the parameters  $(\beta, \delta, T)$ , *given a phylogenetic tree*. Note that in this lecture, the age of the process  $T$  is assumed to be fixed to a known value (this will be relaxed in lecture 11).

The work on the phylodynamic likelihood began in the 1990s in Oxford [Nee et al., 1994b].

In order to do maximum likelihood estimation, we now derive  $P(\mathcal{T}|\eta)$ . This requires us to first derive the probability of a single individual after time  $t$  leaving 0 or 1 offspring, we denote this by  $p(0|t, \beta, \delta)$  and  $p(1|t, \beta, \delta)$ .

## c09-Phylodynamics

Introduction

Population dynamic models

Phylodynamic models

Probabilities

Applications

References

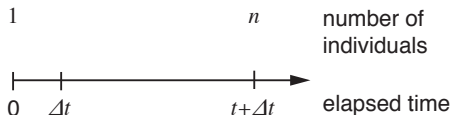
Suppose you start the birth-death process with one individual.  
What is the probability that no surviving individuals remains after time  $t$  ( $p(0|t, \beta, \delta)$ )?

Suppose you start the birth-death process with one individual.  
What is the probability that no surviving individuals remains after time  $t$  ( $p(0|t, \beta, \delta)$ )?

On the following pages we will use the abbreviated notation  
 $p(0|t) := p(0|t, \beta, \delta)$ .

# Probability of extinction $p(0|t)$

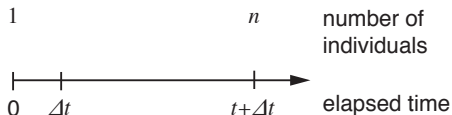
(This is very similar to the derivation of transition probabilities for the substitution models.)



- Consider a small timestep  $\Delta t$  during which only one event occurs.

# Probability of extinction $p(0|t)$

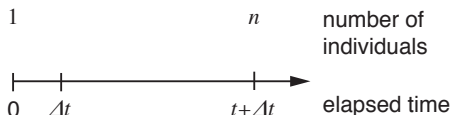
(This is very similar to the derivation of transition probabilities for the substitution models.)



- Consider a small timestep  $\Delta t$  during which only one event occurs.
- During that time step, for a single individual, a death event happens with probability  $\delta \Delta t$ , and a birth event happens with probability  $\beta \Delta t$ . No event happens with probability  $1 - (\beta + \delta) \Delta t$ .

# Probability of extinction $p(0|t)$

(This is very similar to the derivation of transition probabilities for the substitution models.)



- ▶ Consider a small timestep  $\Delta t$  during which only one event occurs.
- ▶ During that time step, for a single individual, a death event happens with probability  $\delta \Delta t$ , and a birth event happens with probability  $\beta \Delta t$ . No event happens with probability  $1 - (\beta + \delta) \Delta t$ .
- ▶ The resulting individual(s) after time  $\Delta t$  have probability  $p(0|t)$  to go extinct within time interval  $t$ .



Thus:

$$p(0|t + \Delta t) = (1 - (\beta + \delta)\Delta t)p(0|t) + \delta\Delta t + \beta\Delta tp(0|t)^2.$$

c09-Phylodynamics

Introduction

Population dynamic models

Phylodynamic models

Probabilities

Applications

References

Thus:

$$p(0|t + \Delta t) = (1 - (\beta + \delta)\Delta t)p(0|t) + \delta\Delta t + \beta\Delta t p(0|t)^2.$$

Rearranging leads to:

$$\frac{p(0|t + \Delta t) - p(0|t)}{\Delta t} = -(\beta + \delta)p(0|t) + \delta + \beta p(0|t)^2.$$

Thus:

$$p(0|t + \Delta t) = (1 - (\beta + \delta)\Delta t)p(0|t) + \delta\Delta t + \beta\Delta tp(0|t)^2.$$

Rearranging leads to:

$$\frac{p(0|t + \Delta t) - p(0|t)}{\Delta t} = -(\beta + \delta)p(0|t) + \delta + \beta p(0|t)^2.$$

Taking the limit  $\Delta t \rightarrow 0$  leads to:

$$\frac{d}{dt}p(0|t) = -(\beta + \delta)p(0|t) + \delta + \beta p(0|t)^2.$$

Thus:

$$p(0|t + \Delta t) = (1 - (\beta + \delta)\Delta t)p(0|t) + \delta\Delta t + \beta\Delta tp(0|t)^2.$$

Rearranging leads to:

$$\frac{p(0|t + \Delta t) - p(0|t)}{\Delta t} = -(\beta + \delta)p(0|t) + \delta + \beta p(0|t)^2.$$

Taking the limit  $\Delta t \rightarrow 0$  leads to:

$$\frac{d}{dt}p(0|t) = -(\beta + \delta)p(0|t) + \delta + \beta p(0|t)^2.$$

The initial condition is  $p(0|0) = 0$ .

$$\begin{aligned}\frac{d}{dt}p(0|t) &= -(\beta + \delta)p(0|t) + \delta + \beta p(0|t)^2, \\ p(0|0) &= 0.\end{aligned}$$

[c09-Phylodynamics](#)[Introduction](#)[Population dynamic models](#)[Phylodynamic models](#)[Probabilities](#)[Applications](#)[References](#)

$$\begin{aligned}\frac{d}{dt}p(0|t) &= -(\beta + \delta)p(0|t) + \delta + \beta p(0|t)^2, \\ p(0|0) &= 0.\end{aligned}$$

The solution to this differential equation is

$$p(0|t) = \frac{\delta(1 - e^{-(\beta-\delta)t})}{\beta - \delta e^{-(\beta-\delta)t}}$$

which can easily be verified by differentiating the expression to obtain  $\frac{d}{dt}p(0|t)$ , and substituting  $\frac{d}{dt}p(0|t)$  and  $p(0|t)$  into the differential equation.

In general, the probability to obtain  $n$  surviving lineages after time  $t$ ,  $p(n|t; \beta, \delta)$  (for which we again write short  $p(n|t)$ ), is [Kendall et al., 1948] (Eqn. 16),

$$\begin{aligned} p(0|t) &= \frac{\delta(1 - e^{-(\beta-\delta)t})}{\beta - \delta e^{-(\beta-\delta)t}} \\ p(1|t) &= e^{-(\beta-\delta)t} (1 - p(0|t))^2 \\ p(n|t) &= p(1|t) \left( \frac{\beta}{\delta} p(0|t) \right)^{n-1} \quad \text{for } n \geq 2. \end{aligned}$$

In general, the probability to obtain  $n$  surviving lineages after time  $t$ ,  $p(n|t; \beta, \delta)$  (for which we again write short  $p(n|t)$ ), is [Kendall et al., 1948] (Eqn. 16),

$$\begin{aligned}p(0|t) &= \frac{\delta(1 - e^{-(\beta-\delta)t})}{\beta - \delta e^{-(\beta-\delta)t}} \\p(1|t) &= e^{-(\beta-\delta)t}(1 - p(0|t))^2 \\p(n|t) &= p(1|t) \left( \frac{\beta}{\delta} p(0|t) \right)^{n-1} \quad \text{for } n \geq 2.\end{aligned}$$

We now provide a proof for  $p(1|t)$ . A proof for  $p(n|t)$  can be obtained using an induction (provided in script).



- In order to prove the equation for  $p(1|t)$ , we write the differential equation analogously to  $p(0|t)$ ,

$$\frac{d}{dt}p(1|t) = -(\beta + \delta)p(1|t) + 2\beta p(1|t)p(0|t)$$

with initial condition  $p(1|0) = 1$ .

- In order to prove the equation for  $p(1|t)$ , we write the differential equation analogously to  $p(0|t)$ ,

$$\frac{d}{dt}p(1|t) = -(\beta + \delta)p(1|t) + 2\beta p(1|t)p(0|t)$$

with initial condition  $p(1|0) = 1$ .

- The factor of 2 in the differential equation for  $p(1|t)$  accounts for either one of the descendants of the birth event leading to the surviving individual after time  $t$ .

- In order to prove the equation for  $p(1|t)$ , we write the differential equation analogously to  $p(0|t)$ ,

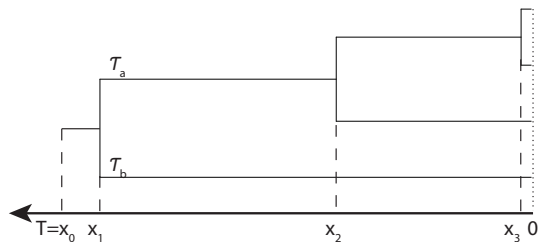
$$\frac{d}{dt}p(1|t) = -(\beta + \delta)p(1|t) + 2\beta p(1|t)p(0|t)$$

with initial condition  $p(1|0) = 1$ .

- The factor of 2 in the differential equation for  $p(1|t)$  accounts for either one of the descendants of the birth event leading to the surviving individual after time  $t$ .
- Evaluating the left- and right-hand side of the differential equation using  $p(1|t) = e^{-(\beta+\delta)t}(1 - p(0|t))^2$  shows that this function is a solution to the differential equation.

# Probability density of a tree, $P(\mathcal{T}|\mathbf{x}_0)$

CB



c09-Phylogenetics

Introduction

Population dynamic models

Phylogenetic models

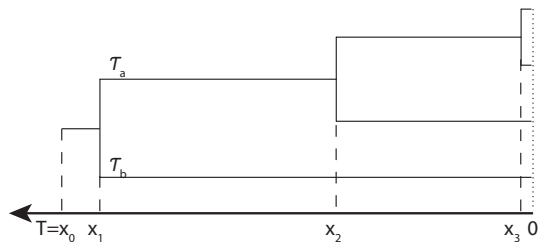
Probabilities

Applications

References

# Probability density of a tree, $P(\mathcal{T}|x_0)$

CB



- Time is measured as *age relative to the present* (i.e. where  $t = 0$ ).

c09-Phylodynamics

Introduction

Population dynamic models

Phylodynamic models

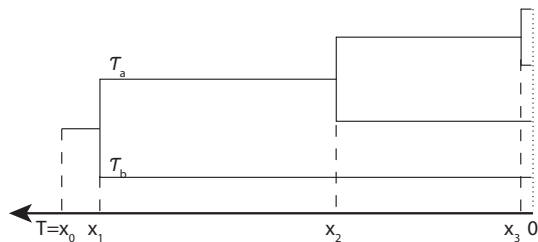
Probabilities

Applications

References

# Probability density of a tree, $P(\mathcal{T}|x_0)$

CB



- Time is measured as *age relative to the present* (i.e. where  $t = 0$ ).
- We take a dynamic programming approach.

c09-Phyldynamics

Introduction

Population dynamic models

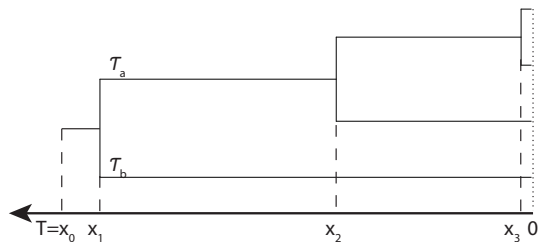
Phyldynamic models

Probabilities

Applications

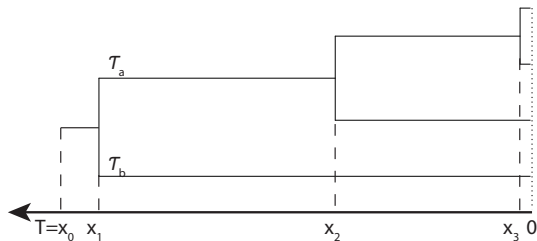
References

# Probability density of a tree, $P(\mathcal{T}|x_0)$



- Time is measured as *age relative to the present* (i.e. where  $t = 0$ ).
- We take a dynamic programming approach.
- Let  $p(x_0, x_1)$  be the probability density for a branch of length  $x_0 - x_1$  extending from an individual at time  $x_0$  in the past.

# Probability density of a tree, $P(\mathcal{T}|\mathbf{x}_0)$



- ▶ Time is measured as *age relative to the present* (i.e. where  $t = 0$ ).
- ▶ We take a dynamic programming approach.
- ▶ Let  $p(\mathbf{x}_0, \mathbf{x}_1)$  be the probability density for a branch of length  $\mathbf{x}_0 - \mathbf{x}_1$  extending from an individual at time  $\mathbf{x}_0$  in the past.
- ▶ Then, the probability density of a tree  $\mathcal{T}$  with age  $\mathbf{x}_0$  is,

$$p(\mathcal{T}|\mathbf{x}_0) = p(\mathbf{x}_0, \mathbf{x}_1)\beta p(\mathcal{T}_a|\mathbf{x}_1)p(\mathcal{T}_b|\mathbf{x}_1)$$

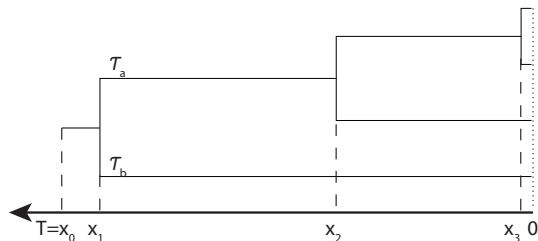
with  $p(\mathcal{T}|\mathbf{x}) := p(\mathcal{T}|\eta = (\beta, \delta, T = \mathbf{x}))$ .



# Probability density of a branch, $p(x_0, x_1)$

CB

We calculate the probability density of the branch between  $t$  and  $x_1$ ,  $p(t, x_1)$ .



c09-Phylogenetics

Introduction

Population dynamic models

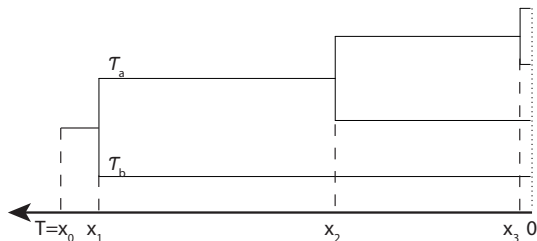
Phylogenetic models

Probabilities

Applications

References

- c09-Phylogenetics
  - Introduction
  - Population dynamic models
  - Phylogenetic models
  - Probabilities
  - Applications
- References



- $p(t + \Delta t, x_1) = (1 - (\beta + \delta)\Delta t)p(t, x_1) + 2\beta\Delta t p(t, x_1)p(0|t),$
- $p(x_1, x_1) = 1.$

- This leads to the differential equation:

$$\frac{d}{dt}p(t, x_1) = -(\beta + \delta)p(t, x_1) + 2\beta p(t, x_1)p(0|t).$$

- This leads to the differential equation:

$$\frac{d}{dt}p(t, x_1) = -(\beta + \delta)p(t, x_1) + 2\beta p(t, x_1)p(0|t).$$

- This is the same differential equation as for  $p(1|t)$ .

- ▶ This leads to the differential equation:

$$\frac{d}{dt}p(t, x_1) = -(\beta + \delta)p(t, x_1) + 2\beta p(t, x_1)p(0|t).$$

- ▶ This is the same differential equation as for  $p(1|t)$ .
- ▶ As the initial condition is different ( $p(x_1, x_1) = 1$ ), we have

$$p(x_0, x_1) = p(1|x_0)/p(1|x_1).$$

- For a tree on  $n$  present day tips, age of the process  $\mathbf{x}_0$ , and branching times  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n-1}$ , we have the probability density

$$p(\mathcal{T}|\mathbf{x}_0) = p(\mathbf{x}_0, \mathbf{x}_1) \beta p(\mathcal{T}_a|\mathbf{x}_1) p(\mathcal{T}_b|\mathbf{x}_1) = \beta^{n-1} \prod_{i=0}^{n-1} p(1|\mathbf{x}_i).$$

- For a tree on  $n$  present day tips, age of the process  $x_0$ , and branching times  $x_1, x_2, \dots, x_{n-1}$ , we have the probability density

$$p(\mathcal{T}|x_0) = p(x_0, x_1) \beta p(\mathcal{T}_a|x_1) p(\mathcal{T}_b|x_1) = \beta^{n-1} \prod_{i=0}^{n-1} p(1|x_i).$$

- Analogous strategy provides us with a tree probability density when  $\rho < 1$  (incomplete extant sampling) and  $\phi > 0$  (sampling through time).

## c09-Phylodynamics

[Introduction](#)[Population dynamic models](#)[Phylodynamic models](#)[Probabilities](#)[Applications](#)[References](#)

Applications.



# Quantifying the spread of Ebola in 2014 (Sierra Leone)

- In each epidemic, it is crucial to know the basic reproductive number  $R_0$ . This is the number of secondary infections caused by a single infected individual in a susceptible population. The number indicates the amount of public health effort for containing the epidemic, i.e. pushing the number of secondary infections below 1.

[c09-Phylodynamics](#)[Introduction](#)[Population dynamic models](#)[Phylogenetic models](#)[Probabilities](#)[Applications](#)[References](#)

# Quantifying the spread of Ebola in 2014 (Sierra Leone)

## c09-Phylodynamics

Introduction

Population dynamic models

Phylodynamic models

Probabilities

Applications

References

- ▶ In each epidemic, it is crucial to know the basic reproductive number  $R_0$ . This is the number of secondary infections caused by a single infected individual in a susceptible population. The number indicates the amount of public health effort for containing the epidemic, i.e. pushing the number of secondary infections below 1.
- ▶ By August 2014, 72 Ebola genomes from different patients in a Sierra Leone outbreak were available [Gire et al., 2014]. The publication also included the phylogenetic tree.

# Quantifying the spread of Ebola in 2014 (Sierra Leone)

- ▶ In each epidemic, it is crucial to know the basic reproductive number  $R_0$ . This is the number of secondary infections caused by a single infected individual in a susceptible population. The number indicates the amount of public health effort for containing the epidemic, i.e. pushing the number of secondary infections below 1.
- ▶ By August 2014, 72 Ebola genomes from different patients in a Sierra Leone outbreak were available [Gire et al., 2014]. The publication also included the phylogenetic tree.
- ▶ We wanted to calculate the basic reproductive number based on these trees.

# Quantifying the spread of Ebola in 2014 (Sierra Leone)

## c09-Phylodynamics

### Introduction

### Population dynamic models

### Phylodynamic models

### Probabilities

### Applications

## References

- ▶ In each epidemic, it is crucial to know the basic reproductive number  $R_0$ . This is the number of secondary infections caused by a single infected individual in a susceptible population. The number indicates the amount of public health effort for containing the epidemic, i.e. pushing the number of secondary infections below 1.
- ▶ By August 2014, 72 Ebola genomes from different patients in a Sierra Leone outbreak were available [Gire et al., 2014]. The publication also included the phylogenetic tree.
- ▶ We wanted to calculate the basic reproductive number based on these trees.
- ▶ As it was early in the epidemic, they assumed a constant  $\beta$  (transmission rate) and  $\delta$  (becoming-uninfectious rate, ending with recovery or death).

# Quantifying the spread of Ebola in 2014 (Sierra Leone)

## c09-Phylodynamics

### Introduction

### Population dynamic models

### Phylodynamic models

### Probabilities

### Applications

### References

- ▶ In each epidemic, it is crucial to know the basic reproductive number  $R_0$ . This is the number of secondary infections caused by a single infected individual in a susceptible population. The number indicates the amount of public health effort for containing the epidemic, i.e. pushing the number of secondary infections below 1.
- ▶ By August 2014, 72 Ebola genomes from different patients in a Sierra Leone outbreak were available [Gire et al., 2014]. The publication also included the phylogenetic tree.
- ▶ We wanted to calculate the basic reproductive number based on these trees.
- ▶ As it was early in the epidemic, they assumed a constant  $\beta$  (transmission rate) and  $\delta$  (becoming-uninfectious rate, ending with recovery or death).
- ▶ The data were sampled through time, i.e.  $\rho = 0$  but  $\phi = 0.7$ .

# Quantifying the spread of Ebola in 2014 (Sierra Leone)

CB

c09-Phylodynamics

Introduction  
Population dynamic models  
Phylodynamic models  
Probabilities  
Applications

References

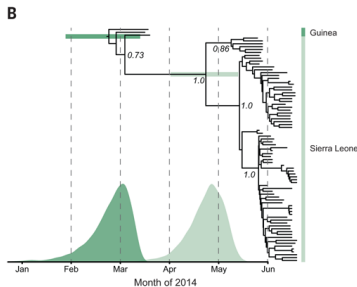


Figure adapted from [Gire et al., 2014]

Phylogenetic tree of Ebola in West Africa.

- We obtained the maximum likelihood estimates  $\hat{\beta}$  and  $\hat{\delta}$  using the phylodynamic likelihood, and thus estimated  $R_0 = \hat{\beta}/\hat{\delta} = 1.34$  (CI 1.12–1.55) [Stadler et al., 2014].

# Quantifying the spread of Ebola in 2014 (Sierra Leone)

CB

c09-Phylodynamics

Introduction

Population dynamic models

Phylodynamic models

Probabilities

Applications

References

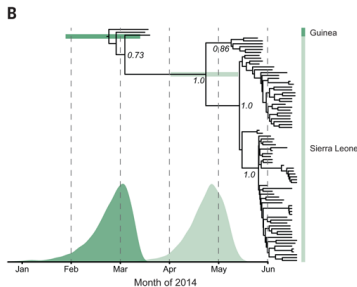


Figure adapted from [Gire et al., 2014]

Phylogenetic tree of Ebola in West Africa.

- ▶ We obtained the maximum likelihood estimates  $\hat{\beta}$  and  $\hat{\delta}$  using the phylodynamic likelihood, and thus estimated  $R_0 = \hat{\beta}/\hat{\delta} = 1.34$  (CI 1.12–1.55) [Stadler et al., 2014].
- ▶ Bayesian methods improve these estimates (lecture 11).

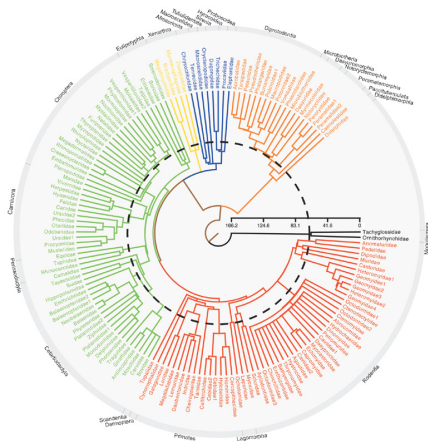


Figure adapted from [Bininda-Emonds et al., 2007]

Phylogenetic tree of mammalian families. Dinosaur extinction (65 million years ago) is marked with a black dashed circle.



- ▶ Did mammals only start diversifying rapidly after the extinction of dinosaurs?

c09-Phylodynamics

Introduction

Population dynamic models

Phylogenetic models

Probabilities

Applications

References

## c09-Phylodynamics

Introduction

Population dynamic models

Phylodynamic models

Probabilities

Applications

References

- ▶ Did mammals only start diversifying rapidly after the extinction of dinosaurs?
- ▶ The diversification of mammals after the extinction of dinosaurs is supported by most mammalian fossils dating from a time after the dinosaur extinction.

- ▶ Did mammals only start diversifying rapidly after the extinction of dinosaurs?
- ▶ The diversification of mammals after the extinction of dinosaurs is supported by most mammalian fossils dating from a time after the dinosaur extinction.
- ▶ Does the phylogeny also support elevated diversification after the dinosaur extinction?

- ▶ Did mammals only start diversifying rapidly after the extinction of dinosaurs?
- ▶ The diversification of mammals after the extinction of dinosaurs is supported by most mammalian fossils dating from a time after the dinosaur extinction.
- ▶ Does the phylogeny also support elevated diversification after the dinosaur extinction?
- ▶ Birth-death model is extended to allow for changes in parameters through time: we have constant rates until time  $t_1$ , then change to other constant rates until time  $t_2$ , etc.

- ▶ Did mammals only start diversifying rapidly after the extinction of dinosaurs?
- ▶ The diversification of mammals after the extinction of dinosaurs is supported by most mammalian fossils dating from a time after the dinosaur extinction.
- ▶ Does the phylogeny also support elevated diversification after the dinosaur extinction?
- ▶ Birth-death model is extended to allow for changes in parameters through time: we have constant rates until time  $t_1$ , then change to other constant rates until time  $t_2$ , etc.
- ▶ Speciation and extinction rates were estimated in [Stadler, 2011] using a mammal phylogeny [Bininda-Emonds et al., 2007].

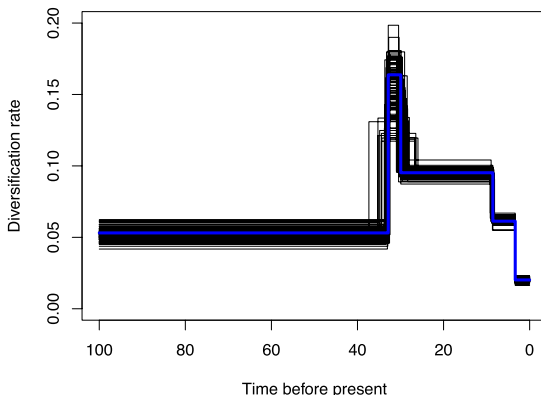
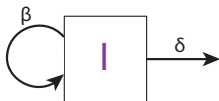


Figure adapted from [Stadler, 2011]

Maximum likelihood diversification rate  $\beta - \delta$  through time (blue). The black lines are parametric bootstrap intervals: birth-death trees were simulated using the maximum likelihood parameters, and then the diversification rate was re-estimated and plotted in black. We find no signal for elevated diversification upon dinosaur extinction (65 Ma).

## c09-Phylogenetics

- Introduction
- Population dynamic models
- Phylogenetic models
- Probabilities
- Applications
- References



## c09-Phylodynamics

Introduction

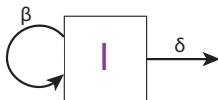
Population dynamic models

Phylodynamic models

Probabilities

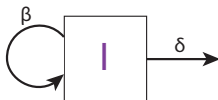
Applications

References

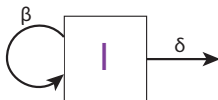


- *Aim:* Determine the birth (speciation / transmission / cell replication) and death (extinction / recovery / cell death) dynamics from a phylogenetic tree.





- *Aim:* Determine the birth (speciation / transmission / cell replication) and death (extinction / recovery / cell death) dynamics from a phylogenetic tree.
- *Method:* The phylodynamic likelihood is maximized to obtain maximum likelihood birth and death rates.



- *Aim:* Determine the birth (speciation / transmission / cell replication) and death (extinction / recovery / cell death) dynamics from a phylogenetic tree.
- *Method:* The phylodynamic likelihood is maximized to obtain maximum likelihood birth and death rates.
- Extensions allow us to incorporate complex models; here we showed rate variation through time in the mammal example.

- ❓ How does the approximate number of steps required to calculate the phylogenetic likelihood depend on the number of leaves in a phylogenetic tree? (I.e. what is the time complexity of this calculation?)
- ❓ What kind of population dynamic process could a decrease in slope in the LTT plot reflect?
- ❓ Assume a birth-death process where each individual at present is sampled with probability  $\rho$ . How is the derivation of  $p(0|t, \rho)$ , the probability of sampling no individual at present, different compared to the derivation of  $p(0|t)$ ?

- Bininda-Emonds, O. R., Cardillo, M., Jones, K. E., MacPhee, R. D., Beck, R. M., Grenyer, R., Price, S. A., Vos, R. A., Gittleman, J. L., and Purvis, A. (2007). The delayed rise of present-day mammals. *Nature*, 446(7135):507–512.
- Gire, S. K., Goba, A., Andersen, K. G., Sealfon, R. S., Park, D. J., Kanneh, L., Jalloh, S., Momoh, M., Fullah, M., Dudas, G., et al. (2014). Genomic surveillance elucidates ebola virus origin and transmission during the 2014 outbreak. *Science*, 345(6202):1369–1372.
- Kendall, D. G. et al. (1948). On the generalized" birth-and-death" process. *The Annals of Mathematical Statistics*, 19(1):1–15.
- Nee, S., Holmes, E. C., May, R. M., and Harvey, P. H. (1994a). Extinction rates can be estimated from molecular phylogenies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 344(1307):77–82.
- Nee, S., May, R. M., and Harvey, P. H. (1994b). The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 344(1309):305–311.
- Paabo, S. (2003). The mosaic that is our genome. *Nature*, 421(6921):409–12.
- Stadler, T. (2011). Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National Academy of Sciences*, 108(15):6187–6192.
- Stadler, T., Kühnert, D., Rasmussen, D. A., and du Plessis, L. (2014). Insights into the early epidemic spread of ebola in sierra leone provided by viral sequence data. *PLoS currents*, 6.

## c09-Phyldynamics

Introduction  
Population dynamic models  
Phyldynamic models  
Probabilities  
Applications

## References