Computational Biology

Lecturers:
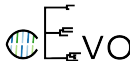Tanja Stadler, Tim Vaughan & Carsten Magnus

Teaching Assistants:
Antoine Zwaans, Adrian Lison,
James Munday & Marcus Overwater

Computational Evolution
Department of Biosystems Science and Engineering

HS 2023

ℂ𝔼vo

Continuous traits and comparative methods: Discuss questions.

CB

Phylodynamics
Introduction
Population dynamic models
Phylodynamic models
Probabilities
Applications

References

❓ 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?

Toolbox: Fisher's exact test II (Lecture 3)

Contingency table:

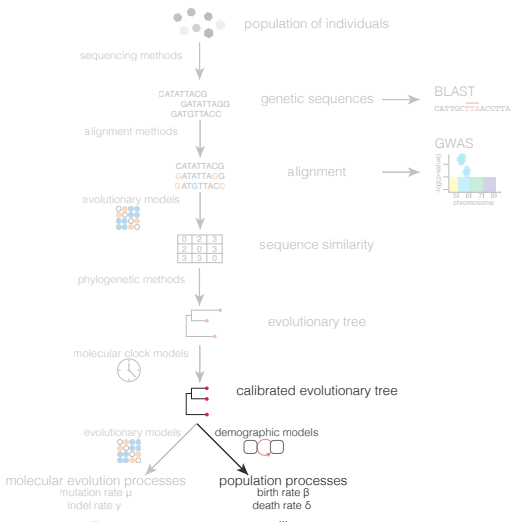| A \ B	B_1	B_2	
A_1			a+b
A_2			c+d
	a+c	b+d	n= a+b+c+d

Hypothesis:

We want to test whether class A is linked to class B

Thus our null hypothesis is:
$\mathcal{H}_0$: The number of individuals expressing both A_1 and B_1 is based on chance.

☞ The correct formulation of the null hypothesis is:
$\mathcal{H}_0$: The random variable of the number of individuals expressing both A_1 and B_1 is distributed according to a hypergeometric distribution.

☞ p-value= $\sum_{i=a}^{a+b} \frac{\binom{a+b}{i}\binom{c+d}{a+c-i}}{\binom{n}{a+c}}$

Overview

What have we learned so far?

What have we learned so far?

Rooted Time Tree

Alignment

CAGATTACA

ACCATTAGA

AACATTGGA
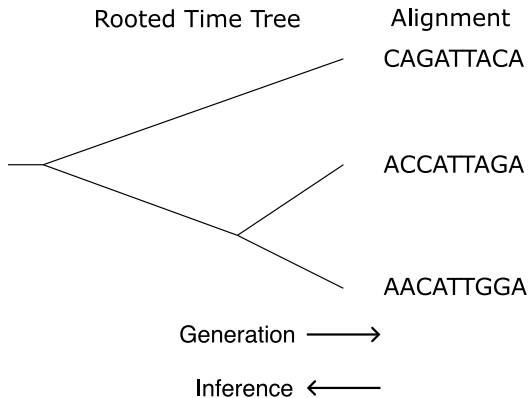
Generation $\longrightarrow$
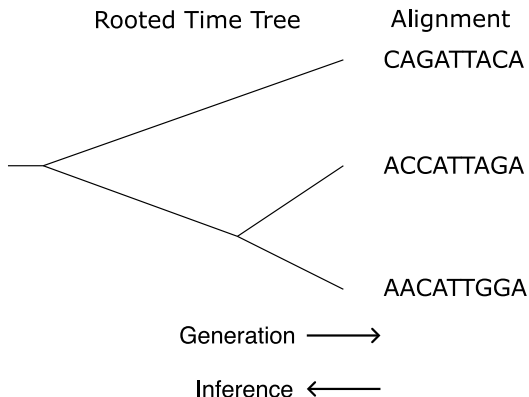
Inference $\longleftarrow$

Question

What is the process that generated the phylogenetic tree?

Phylogenetic trees encode past macroevolutionary dynamics
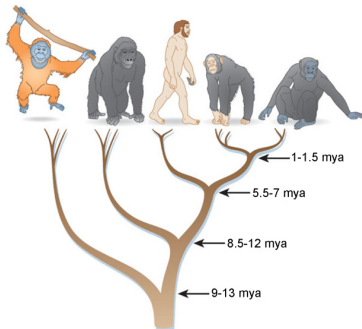
Macroevolution: individuals = species

Phylogenetic trees encode past macroevolutionary dynamics

Macroevolution: individuals = species

▶ **(Molecular) Evolution**
(Genetic) makeup of species changes through time.

Phylogenetic trees encode past macroevolutionary dynamics

Macroevolution: individuals = species

- **(Molecular) Evolution**
  (Genetic) makeup of species changes through time.
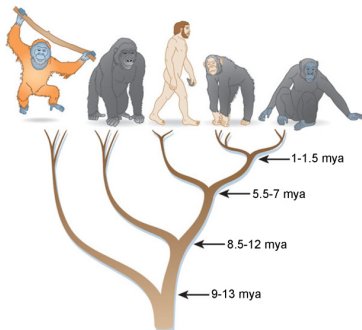- **Phylogenetics**
  Phylogeny displays species relationships.

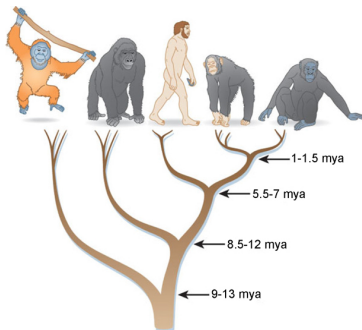# Phylogenetic trees encode past macroevolutionary dynamics

Macroevolution: individuals = species

▶ **(Molecular) Evolution**
(Genetic) makeup of species changes through time.

▶ **Phylogenetics**
Phylogeny displays species relationships.

▶ **Phylodynamics**
Population dynamics is the speciation and extinction process.
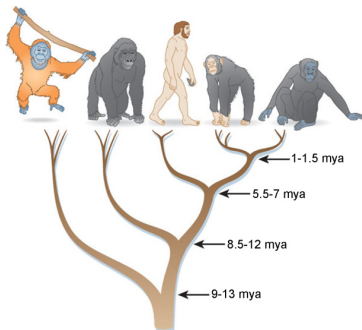
Phylogenetic trees encode past epidemiological dynamics

CB

Phylodynamics
Introduction
Population dynamic models
Phylodynamic models
Probabilities
Applications

References

Epidemiology: individuals = infected hosts

Phylogenetic trees encode past epidemiological dynamics

CB

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

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

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
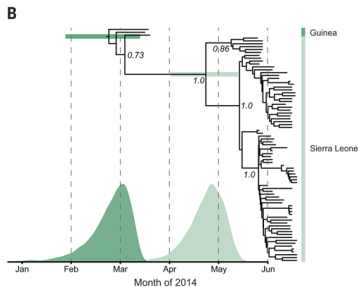
Figure adapted from [Gire et al., 2014]

Phylogenetic tree of Ebola in West Africa.

Phylogenetic trees encode past epidemiological dynamics

CB

Phylodynamics
Introduction
Population dynamic models
Phylodynamic models
Probabilities
Applications

References
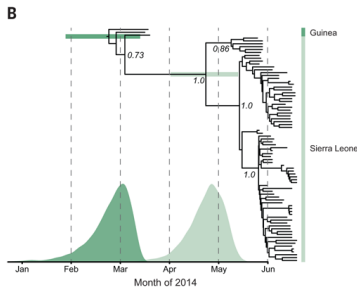
Figure adapted from [Gire et al., 2014]

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.)

Further key applications

▶ **Immunology:** individuals = B cells

Further key applications

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

Further key applications

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

Further key applications

- ▶ **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

Further key applications

▶ **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.

Further key applications

▶ **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.

Further key applications

▶ **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

Further key applications

▶ **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.

Further key applications

▶ **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.

Phylodynamics

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

Phylodynamics

▶ *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

▶ *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).

Population dynamic models

Population dynamic models.

Population dynamics

CB

Phylodynamics
Introduction
Population dynamic models
Phylodynamic models
Probabilities
Applications

References

▶ Rate of birth of new individuals per individual in I: β

Population dynamics

- ▶ Rate of birth of new individuals per individual in I: β
- ▶ Rate of death per individual in I: δ

Population dynamics

CB

Phylodynamics
Introduction
Population dynamic models
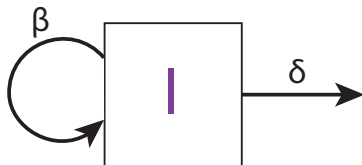Phylodynamic models
Probabilities
Applications

References
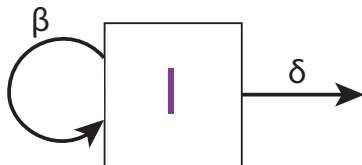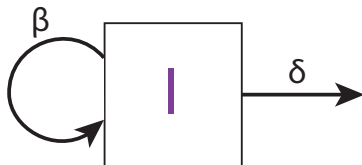
- ▶ Rate of birth of new individuals per individual in I: β
- ▶ Rate of death per individual in I: δ

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

Stochastic population dynamics

Consider the fate of one individual:

Stochastic population dynamics

CB
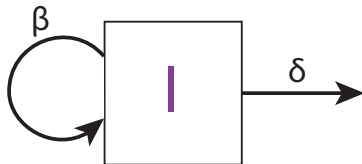
Phylodynamics
Introduction
Population dynamic models
Phylodynamic models
Probabilities
Applications

References

Consider the fate of one individual:
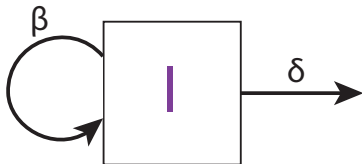
▶ The probability of giving birth to another individual in a very small time step Δt is $\beta \Delta t$.

Stochastic population dynamics

Consider the fate of one individual:

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

▶ The probability of dying in a very small time step Δt is $\delta \Delta t$.

Stochastic population dynamics

Consider the fate of one individual:

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

▶ The probability of dying in a very small time step Δt is $\delta \Delta t$.

▶ The waiting time to a birth event is exponentially distributed with parameter β.

Stochastic population dynamics
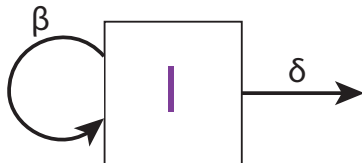
Consider the fate of one individual:

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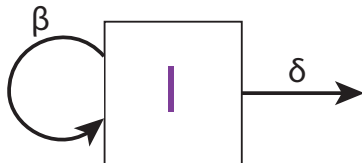
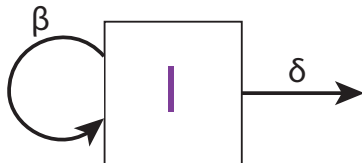▶ The probability of dying in a very small time step Δt is $\delta \Delta t$.

▶ The waiting time to a birth event is exponentially distributed with parameter β.

▶ 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 $r_1 + r_2$).

Stochastic population dynamics

Consider the fate of N individuals:

Stochastic population dynamics
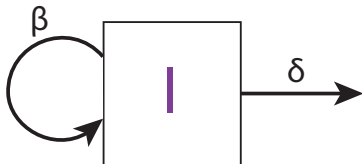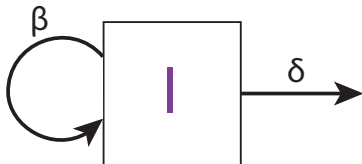
Consider the fate of N individuals:

▶ 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

From population dynamics to phylogenetic trees

CB

Phylodynamics
Introduction
Population dynamic models
Phylodynamic models
Probabilities
Applications

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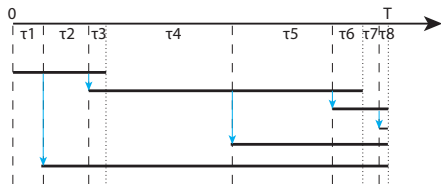
This diagram illustrates the full population dynamics of a birth-death process which starts with one individual and is stopped after time T. Each solid black line is the lifetime of an individual. Blue arrows are birth events.

From population dynamics to phylogenetic trees

CB

Phylodynamics
Introduction
Population dynamic models
Phylodynamic models
Probabilities
Applications

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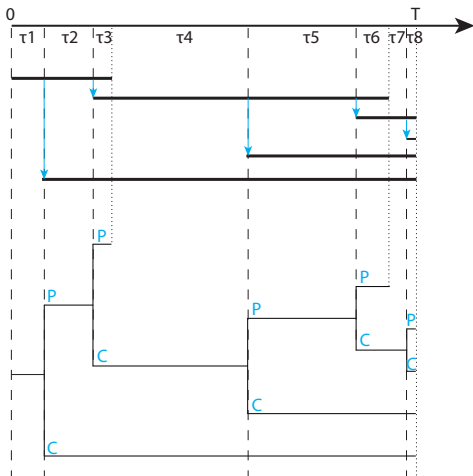
A *complete population tree* displays the full population dynamics. The labels P and C illustrate the parent-child associations.

Phylodynamic models

Phylodynamic models.

A simple phylodynamic model

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:

A simple phylodynamic model

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 β

A simple phylodynamic model

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 β
► Death rate δ

A simple phylodynamic model

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 β
- Death rate δ
- Process duration T

A simple phylodynamic model

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 β
- Death rate δ
- Process duration T
- Extant tip sampling probability ρ

A simple phylodynamic model

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 β
- Death rate δ
- Process duration T
- Extant tip sampling probability ρ
- Extinct tip sampling probability ϕ

A simple phylodynamic model

CB

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 β
- Death rate δ
- Process duration T
- Extant tip sampling probability ρ
- Extinct tip sampling probability ϕ

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

CB

Phylodynamics
  Introduction
  Population dynamic models
  Phylodynamic models
  Probabilities
  Applications

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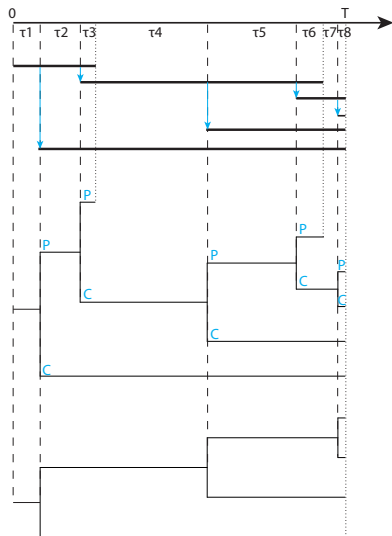
The phylogenetic tree with $\rho = 1, \phi = 0$ displays the dynamics giving rise to the surviving individuals at time T.

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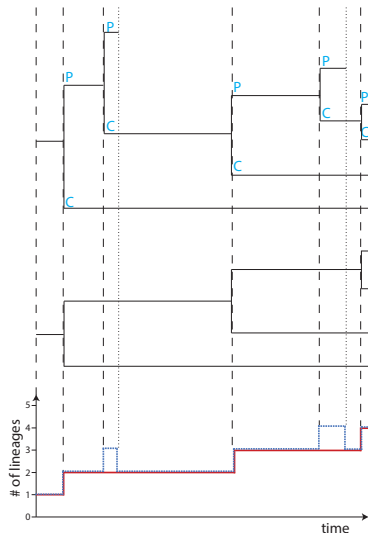
Plotting the number of lineages (y-axis) vs time (x-axis) is called the lineages-through-time (LTT) plot. The LTT plot of the complete tree (blue; dashed) shows the population size through time. The LTT plot of the phylogenetic tree (red) shows the number of surviving lineages through time.

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

Average LTTs for a population of age T = 50:

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.

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.

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).

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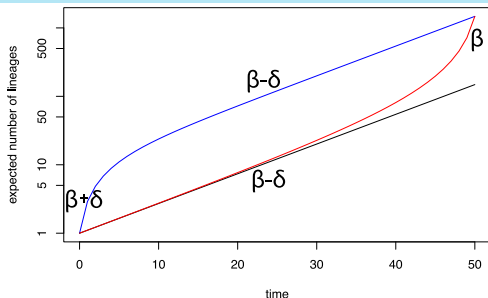
CB

Phylodynamics
  Introduction
  Population dynamic models
  Phylodynamic 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].

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

CB

Phylodynamics
Introduction
Population dynamic models
Phylodynamic models
Probabilities
Applications

References

▶ We observe a push-of-the-past as only individuals with a quick replication early on will produce surviving populations.

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

CB

Phylodynamics
  Introduction
  Population dynamic models
  Phylodynamic models
  Probabilities
  Applications
References

▶ 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.

Estimating β and δ

Estimating β and δ

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

Estimating β and δ

▶ 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 β.

Estimating β and δ

CB
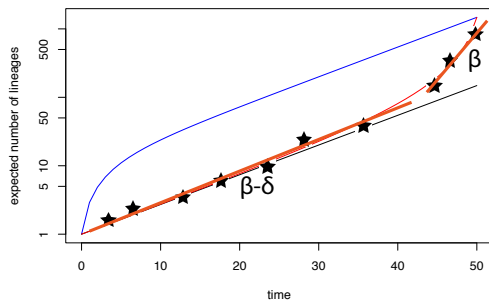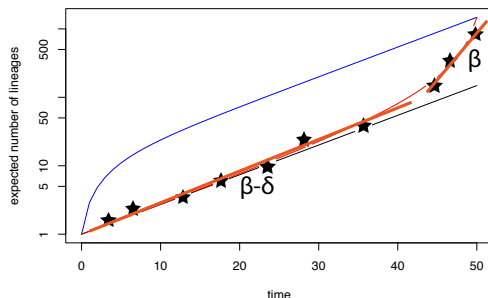
Phylodynamics
Introduction
Population dynamic models
Phylodynamic models
Probabilities
Applications

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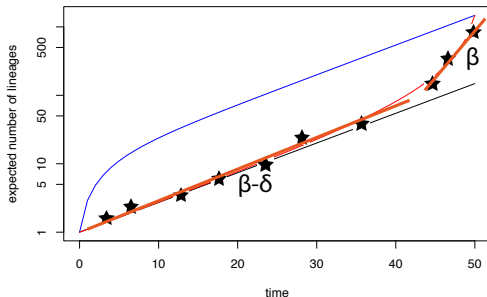
- We could fit a regression line to the early branching events and estimate its slope, this is an estimate for $\beta - \delta$.
- We could fit a regression line to the late branching events and estimate its slope, this is an estimate for β.
- <u>Problem:</u> It is not clear how to incorporate the variances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

Probabilities

Probabilities.

Probability density of a tree

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

Probability density of a tree

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

Probability density of a tree

► 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 (β, δ, 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).

Probability density of a tree

► 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 (β, δ, 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].

Probability density of a tree

► 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 (β, δ, 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 $p(1|t, \beta, \delta)$.

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

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

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

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 $p(0|t)$

CB

Phylodynamics
Introduction
Population dynamic models
Phylodynamic models
Probabilities
Applications

References

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

$$t + \Delta t \quad t \qquad\qquad\qquad 0$$

▶ Consider a small timestep Δt during which only one event occurs.

Probability of extinction $p(0|t)$
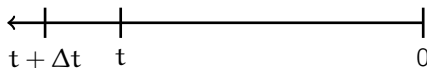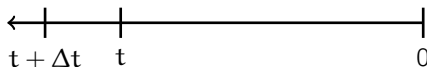
(This is very similar to the derivation of transition probabilities for the substitution models.)

- ▶ Consider a small timestep Δt 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.)

$$t + \Delta t \quad t \qquad\qquad\qquad 0$$
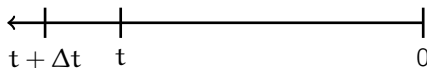
▶ Consider a small timestep Δt during which only one event occurs.

▶ 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 Δt have probability $p(0|t)$ to go extinct within time interval t.

Differential equation for $p(0|t)$

Thus:

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

Differential equation for $p(0|t)$

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:

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

Differential equation for $p(0|t)$

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:

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

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

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

Differential equation for $p(0|t)$

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:

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

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

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

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

Solution for $p(0|t)$

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

Solution for $p(0|t)$

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

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 $\frac{\mathrm{d}}{\mathrm{d}t}p(0|t)$, and substituting $\frac{\mathrm{d}}{\mathrm{d}t}p(0|t)$ and $p(0|t)$ into the differential equation.

Probability of n descendants, $p(n|t)$

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),

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

Probability of n descendants, $p(n|t)$

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),

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

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

Probability of 1 descendant, $p(1|t)$

▶ 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 $p(1|0) = 1$.

Probability of 1 descendant, $p(1|t)$

▶ 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.

Probability of 1 descendant, $p(1|t)$

▶ 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, $P(\mathcal{T}|x_0)$

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

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

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.

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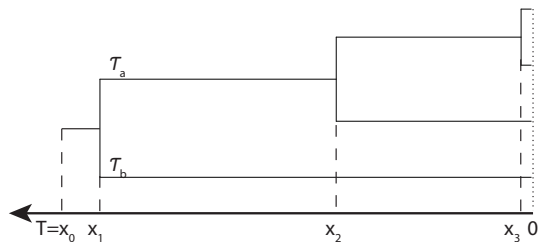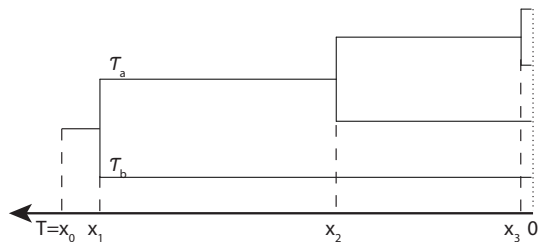
Probability density of a tree, $P(\mathcal{T}|x_0)$

- ▶ 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.
- ▶ Then, the probability density of a tree $\mathcal{T}$ with age x_0 is,

$$p(\mathcal{T}|x_0) = p(x_0, x_1)\beta p(\mathcal{T}_a|x_1)p(\mathcal{T}_b|x_1)$$

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

Probability of a branch, $p(x_0, x_1)$
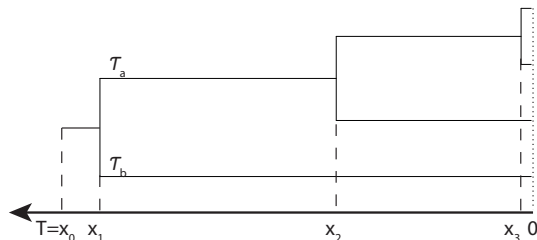
We calculate the probability of the branch between t and x_1, $p(t, x_1)$.

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

We calculate the probability of the branch between t and x_1, $p(t, x_1)$.
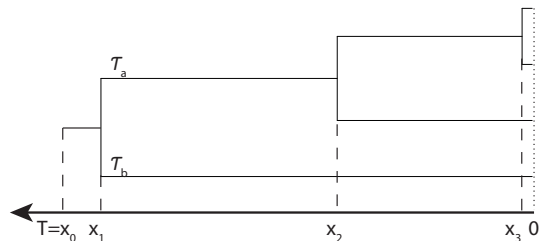
- $p(t + \Delta t, x_1) = (1 - (\beta + \delta)\Delta t)p(t, x_1) + 2\beta \Delta t p(t, x_1)p(0|t)$,
- $p(x_1, x_1) = 1$.

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

▶ This leads to the differential equation:

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

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

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

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

CB

Phylodynamics
Introduction
Population dynamic models
Phylodynamic models
Probabilities
Applications

References

▶ 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

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

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

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

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

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

▶ For a tree on n present day tips, age of the process x_0, and branching times $x_1, x_2, \ldots, x_{n-1}$, 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).

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.

Quantifying the spread of Ebola in 2014 (Sierra Leone)

CB

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)

CB

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.

Quantifying the spread of Ebola in 2014 (Sierra Leone)

CB

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 β (transmission rate) and δ (becoming-uninfectious rate, ending with recovery or death).

Quantifying the spread of Ebola in 2014 (Sierra Leone)

CB

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 β (transmission rate) and δ (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

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

Phylodynamics
Introduction
Population dynamic models
Phylodynamic models
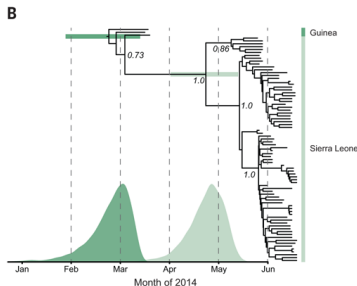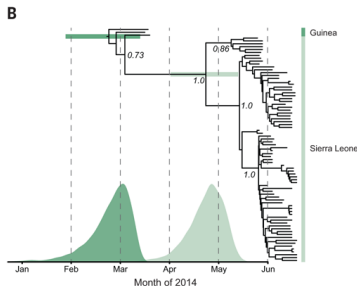Probabilities
Applications

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).

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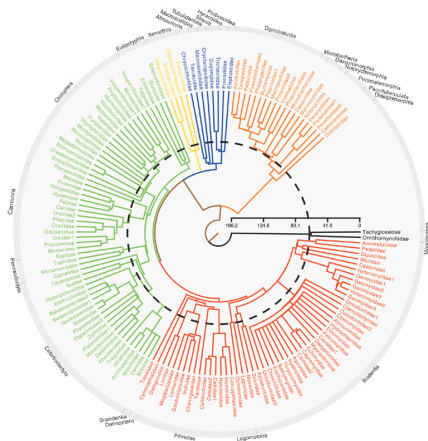
Figure adapted from [Bininda-Emonds et al., 2007]

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

Mammals and dinosaurs

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

Mammals and dinosaurs

- ▶ 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.

Mammals and dinosaurs

▶ 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?

Mammals and dinosaurs

CB

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.

▶ 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.

Mammals and dinosaurs

► 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].
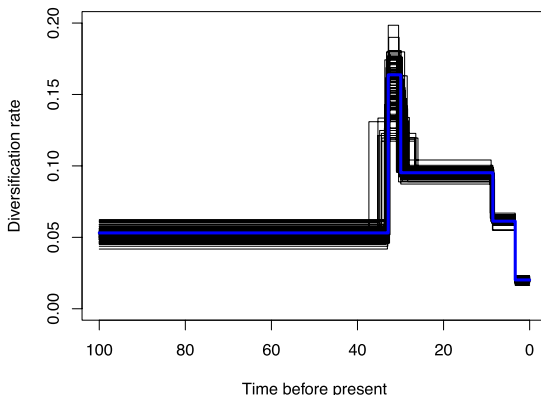
Mammals and dinosaurs

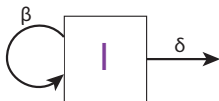Figure adapted from [Stadler, 2011]

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).

Summary

Summary

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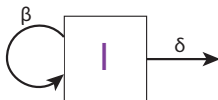
Summary

CB

Phylodynamics
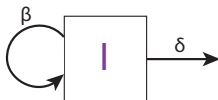Introduction
Population dynamic models
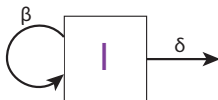Phylodynamic models
Probabilities
Applications

References

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

Summary

- ▶ *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.

Phylodynamic birth-death models: Questions

❷ How does the approximate number of steps required to calculate the phylodynamic likelihood depend on the number of leaves in a phylogenetic tree? (I.e. what is the time complexity of this calculation?)

❷ What kind of population dynamic process could a decrease in slope in the LTT plot reflect?

❷ Assume a birth-death process where each individual at present is sampled with probability ρ. How is the derivation of $p(0|t, \rho)$, the probability of sampling no individual at present, different compared to the derivation of $p(0|t)$?

References I

Phylodynamics

Introduction

Population dynamic models

Phylodynamic models

Probabilities

Applications

References

- 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.