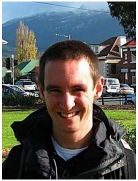
Inferring the ancestral population size under a birth-death process, from a reconstructed phylogenetic tree and a record of occurrences

Groupe de travail Math-Bio et Santé – LJLL – Paris

Marc Manceau, Antoine Zwaans, Jérémy Andréoletti, Ankit Gupta, Tim Vaughan, Rachel Warnock, Tanja Stadler

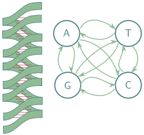
June 29, 2020

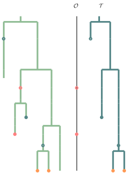


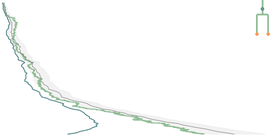



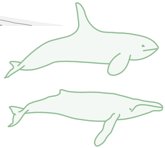
ETH zürich DBSSE cEVO

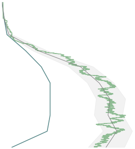
Sketch of the presentation



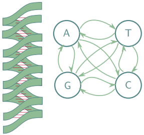








Basics of phylogenetics



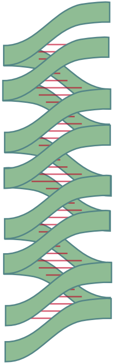
The raw data – in macroevolution

- Molecular sequences of extant species
- Morphological traits of extant species
- Morphological traits of fossil species

The raw data – in macroevolution

- Molecular sequences of extant species

Indiv. A	Indiv. B	Indiv. C
A	A	A
T	A	T
G	G	G
G	G	G
T	C	C
G	G	G
C	C	C
C	C	C
C	C	A
G	G	G
T	T	T
A	G	G
A	A	A
A	A	A
T	T	T

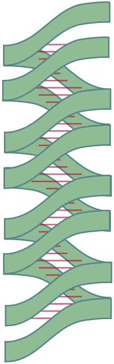


- Morphological traits of extant species
- Morphological traits of fossil species

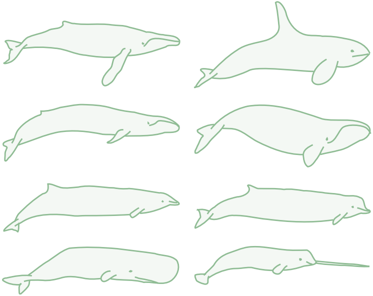
The raw data – in macroevolution

- Molecular sequences of extant species

Indiv. A	Indiv. B	Indiv. C
A	A	A
T	A	T
G	G	G
G	G	G
T	C	C
G	G	G
C	C	C
C	C	C
C	C	A
G	G	G
T	T	T
A	G	G
A	A	A
A	A	A
T	T	T



- Morphological traits of extant species

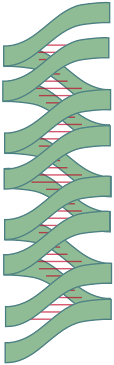


- Morphological traits of fossil species

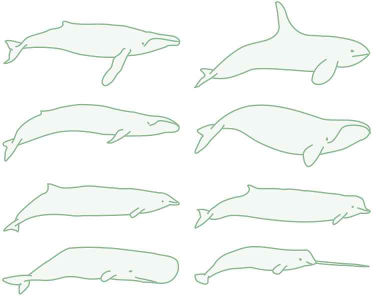
The raw data – in macroevolution

- Molecular sequences of extant species

Indiv. A	Indiv. B	Indiv. C
A	A	A
T	A	T
G	G	G
G	G	G
T	C	C
G	G	G
C	C	C
C	C	C
C	C	A
G	G	G
T	T	T
A	G	G
A	A	A
A	A	A
T	T	T



- Morphological traits of extant species



- Morphological traits of fossil species

The raw data – in epidemiology

- Infected individuals are being sampled throughout the epidemic
- Their pathogens are being sequenced
- Traits concerning pathogens and hosts can be recorded
(e.g. geographic location, viral load, gender, ...)

The raw data – in epidemiology

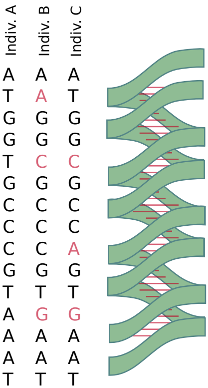
- Infected individuals are being sampled throughout the epidemic
- Their pathogens are being sequenced

- Traits concerning pathogens and hosts can be recorded
(e.g. geographic location, viral load, gender, ...)

The raw data – in epidemiology

- Infected individuals are being sampled throughout the epidemic
- Their pathogens are being sequenced

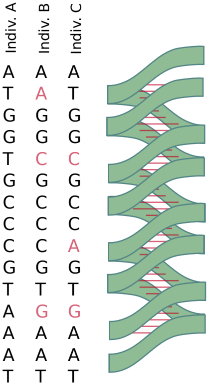
- Traits concerning pathogens and hosts can be recorded (e.g. geographic location, viral load, gender, ...)

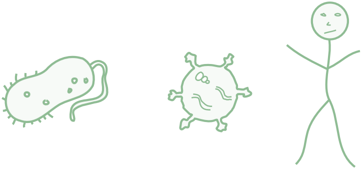


The raw data – in epidemiology

- Infected individuals are being sampled throughout the epidemic
- Their pathogens are being sequenced

- Traits concerning pathogens and hosts can be recorded (e.g. geographic location, viral load, gender, ...)





The questions

Macroevolution

1. What is the tempo of diversification ?
2. What are the drivers of trait evolution ?
3. What was the diversity in the past ?

Epidemiology

1. What is the tempo of epidemic spread ?
2. Are some pathogen traits under selection ?
3. What was the prevalence in the past ?

- There is a hidden phylogenetic/transmission tree.
- Traits evolve along the tree.
- Propose scenarios of evolution: probabilistic models with mecanistic assumptions.
- Fit these models to observed data.

The questions

Macroevolution

1. What is the tempo of diversification ?
2. What are the drivers of trait evolution ?
3. What was the diversity in the past ?

Epidemiology

1. What is the tempo of epidemic spread ?
2. Are some pathogen traits under selection ?
3. What was the prevalence in the past ?

- There is a hidden phylogenetic/transmission tree.
- Traits evolve along the tree.
- Propose scenarios of evolution: probabilistic models with mecanistic assumptions.
- Fit these models to observed data.

The questions

Macroevolution

1. What is the tempo of diversification ?
2. What are the drivers of trait evolution ?
3. What was the diversity in the past ?

Epidemiology

1. What is the tempo of epidemic spread ?
2. Are some pathogen traits under selection ?
3. What was the prevalence in the past ?

- There is a hidden phylogenetic/transmission tree.
- Traits evolve along the tree.
- Propose scenarios of evolution: probabilistic models with mecanistic assumptions.
- Fit these models to observed data.

The questions

Macroevolution

1. What is the tempo of diversification ?
2. What are the drivers of trait evolution ?
3. What was the diversity in the past ?

Epidemiology

1. What is the tempo of epidemic spread ?
2. Are some pathogen traits under selection ?
3. What was the prevalence in the past ?

- There is a hidden phylogenetic/transmission tree.
- Traits evolve along the tree.
- Propose scenarios of evolution: probabilistic models with mecanistic assumptions.
- Fit these models to observed data.

The questions

Macroevolution

1. What is the tempo of diversification ?
2. What are the drivers of trait evolution ?
3. What was the diversity in the past ?

Epidemiology

1. What is the tempo of epidemic spread ?
2. Are some pathogen traits under selection ?
3. What was the prevalence in the past ?

- There is a hidden phylogenetic/transmission tree.
- Traits evolve along the tree.
- Propose scenarios of evolution: probabilistic models with mecanistic assumptions.
- Fit these models to observed data.

The questions

Macroevolution

1. What is the tempo of diversification ?
2. What are the drivers of trait evolution ?
3. What was the diversity in the past ?

Epidemiology

1. What is the tempo of epidemic spread ?
2. Are some pathogen traits under selection ?
3. What was the prevalence in the past ?

- There is a hidden phylogenetic/transmission tree.
- Traits evolve along the tree.
- Propose scenarios of evolution: probabilistic models with mecanistic assumptions.
- Fit these models to observed data.

The questions

Macroevolution

1. What is the tempo of diversification ?
2. What are the drivers of trait evolution ?
3. What was the diversity in the past ?

Epidemiology

1. What is the tempo of epidemic spread ?
2. Are some pathogen traits under selection ?
3. What was the prevalence in the past ?

- There is a hidden phylogenetic/transmission tree.
- Traits evolve along the tree.
- Propose scenarios of evolution: probabilistic models with mecanistic assumptions.
- Fit these models to observed data.

The questions

Macroevolution

1. What is the tempo of diversification ?
2. What are the drivers of trait evolution ?
3. What was the diversity in the past ?

Epidemiology

1. What is the tempo of epidemic spread ?
2. Are some pathogen traits under selection ?
3. What was the prevalence in the past ?

- There is a hidden phylogenetic/transmission tree.
- Traits evolve along the tree.
- Propose scenarios of evolution: probabilistic models with mecanistic assumptions.
- Fit these models to observed data.

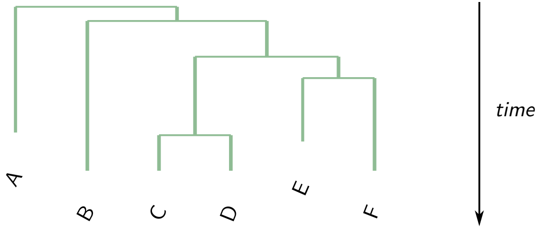
The questions

Macroevolution

1. What is the tempo of diversification ?
2. What are the drivers of trait evolution ?
3. What was the diversity in the past ?

Epidemiology

1. What is the tempo of epidemic spread ?
2. Are some pathogen traits under selection ?
3. What was the prevalence in the past ?



- There is a hidden phylogenetic/transmission tree.
- Traits evolve along the tree.
- Propose scenarios of evolution: probabilistic models with mecanistic assumptions.
- Fit these models to observed data.

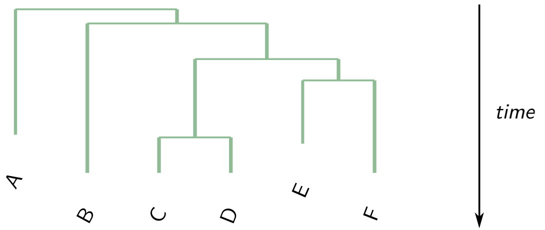
The questions

Macroevolution

1. What is the tempo of diversification ?
2. What are the drivers of trait evolution ?
3. What was the diversity in the past ?

Epidemiology

1. What is the tempo of epidemic spread ?
2. Are some pathogen traits under selection ?
3. What was the prevalence in the past ?



- There is a hidden phylogenetic/transmission tree.
- Traits evolve along the tree.
- Propose scenarios of evolution: probabilistic models with mecanistic assumptions.
- Fit these models to observed data.

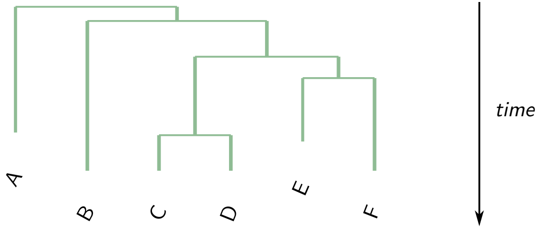
The questions

Macroevolution

1. What is the tempo of diversification ?
2. What are the drivers of trait evolution ?
3. What was the diversity in the past ?

Epidemiology

1. What is the tempo of epidemic spread ?
2. Are some pathogen traits under selection ?
3. What was the prevalence in the past ?



- There is a hidden phylogenetic/transmission tree.
- Traits evolve along the tree.
- Propose scenarios of evolution: probabilistic models with mecanistic assumptions.
- Fit these models to observed data.

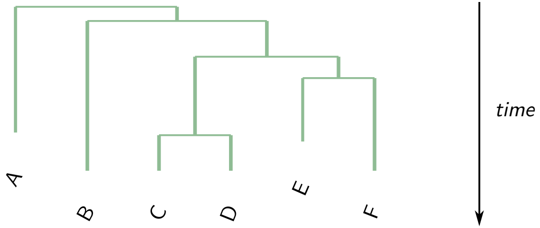
The questions

Macroevolution

1. What is the tempo of diversification ?
2. What are the drivers of trait evolution ?
3. What was the diversity in the past ?

Epidemiology

1. What is the tempo of epidemic spread ?
2. Are some pathogen traits under selection ?
3. What was the prevalence in the past ?



- There is a hidden phylogenetic/transmission tree.
- Traits evolve along the tree.
- Propose scenarios of evolution: probabilistic models with mecanistic assumptions.
- Fit these models to observed data.

The Bayesian framework

DATA

indiv. A	A	T	C	G	A	A	G	C	...
indiv. B	-	-	G	-	C	T	-	-	...
indiv. C	-	A	G	-	-	-	-	-	...
indiv. D	-	A	G	-	-	-	-	-	...
indiv. E	-	-	G	-	-	-	A	-	...
indiv. F	-	-	G	-	-	-	A	-	...

MODEL

$\mathcal{A}$ = Alignment

$\mathbb{P}(\mathcal{A} \mid \mathcal{T}, \mathcal{R})$

$\mathcal{R}$ = Substitution rate

$\mathbb{P}(\mathcal{R} \mid \mathcal{T})$

$\mathcal{T}$ = Tree

$\mathbb{P}(\mathcal{T})$

$$\mathbb{P}(\mathcal{T}, \mathcal{R} \mid \mathcal{A}) \propto \mathbb{P}(\mathcal{A} \mid \mathcal{R}, \mathcal{T})\, \mathbb{P}(\mathcal{R} \mid \mathcal{T})\, \mathbb{P}(\mathcal{T})$$

RESULT

The Bayesian framework

DATA

indiv. A	A	T	C	G	A	A	G	C	...
indiv. B	-	-	G	-	C	T	-	-	...
indiv. C	-	A	G	-	-	-	-	-	...
indiv. D	-	A	G	-	-	-	-	-	...
indiv. E	-	-	G	-	-	-	A	-	...
indiv. F	-	-	G	-	-	-	A	-	...

MODEL

$\mathcal{A}$ = Alignment

$\mathbb{P}(\mathcal{A} \mid \mathcal{T}, \mathcal{R})$

$\mathcal{R}$ = Substitution rate

$\mathbb{P}(\mathcal{R} \mid \mathcal{T})$

$\mathcal{T}$ = Tree

$\mathbb{P}(\mathcal{T})$

$$\mathbb{P}(\mathcal{T}, \mathcal{R} \mid \mathcal{A}) \propto \mathbb{P}(\mathcal{A} \mid \mathcal{R}, \mathcal{T})\, \mathbb{P}(\mathcal{R} \mid \mathcal{T})\, \mathbb{P}(\mathcal{T})$$

RESULT

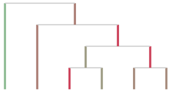
The Bayesian framework

DATA

indiv. A	A	T	C	G	A	A	G	C	...
indiv. B	-	-	G	-	C	T	-	-	...
indiv. C	-	A	G	-	-	-	-	-	...
indiv. D	-	A	G	-	-	-	-	-	...
indiv. E	-	-	G	-	-	-	A	-	...
indiv. F	-	-	G	-	-	-	A	-	...

MODEL

$\mathcal{A}$ = Alignment
$\mathbb{P}(\mathcal{A} \mid \mathcal{T}, \mathcal{R})$



$\mathcal{R}$ = Substitution rate
$\mathbb{P}(\mathcal{R} \mid \mathcal{T})$

$\mathcal{T}$ = Tree
$\mathbb{P}(\mathcal{T})$

$$\mathbb{P}(\mathcal{T}, \mathcal{R} \mid \mathcal{A}) \propto \mathbb{P}(\mathcal{A} \mid \mathcal{R}, \mathcal{T})\, \mathbb{P}(\mathcal{R} \mid \mathcal{T})\, \mathbb{P}(\mathcal{T})$$

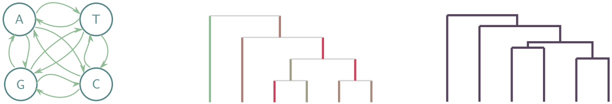
RESULT

The Bayesian framework

DATA

indiv. A	A	T	C	G	A	A	G	C	...
indiv. B	-	-	G	-	C	T	-	-	...
indiv. C	-	A	G	-	-	-	-	-	...
indiv. D	-	A	G	-	-	-	-	-	...
indiv. E	-	-	G	-	-	-	A	-	...
indiv. F	-	-	G	-	-	-	A	-	...

MODEL



$\mathcal{A}$ = Alignment $\mathbb{P}(\mathcal{A} \mid \mathcal{T}, \mathcal{R})$

$\mathcal{R}$ = Substitution rate $\mathbb{P}(\mathcal{R} \mid \mathcal{T})$

$\mathcal{T}$ = Tree $\mathbb{P}(\mathcal{T})$

$$\mathbb{P}(\mathcal{T}, \mathcal{R} \mid \mathcal{A}) \propto \mathbb{P}(\mathcal{A} \mid \mathcal{R}, \mathcal{T})\, \mathbb{P}(\mathcal{R} \mid \mathcal{T})\, \mathbb{P}(\mathcal{T})$$

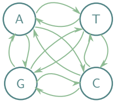
RESULT

The Bayesian framework

DATA

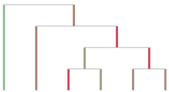
indiv. A	A	T	C	G	A	A	G	C	...
indiv. B	-	-	G	-	C	T	-	-	...
indiv. C	-	A	G	-	-	-	-	-	...
indiv. D	-	A	G	-	-	-	-	-	...
indiv. E	-	-	G	-	-	-	A	-	...
indiv. F	-	-	G	-	-	-	A	-	...

MODEL



$\mathcal{A}$ = Alignment

$\mathbb{P}(\mathcal{A} \mid \mathcal{T}, \mathcal{R})$



$\mathcal{R}$ = Substitution rate

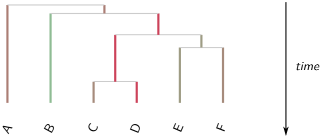
$\mathbb{P}(\mathcal{R} \mid \mathcal{T})$

$\mathcal{T}$ = Tree

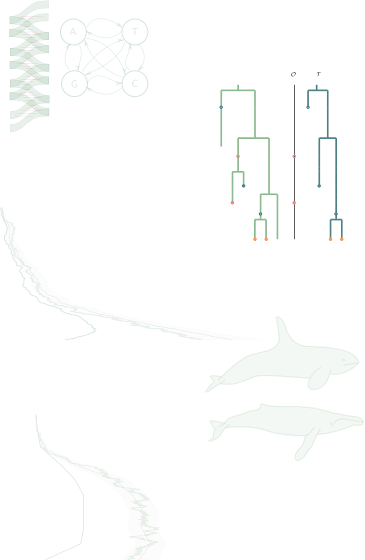
$\mathbb{P}(\mathcal{T})$

$$\mathbb{P}(\mathcal{T}, \mathcal{R} \mid \mathcal{A}) \propto \mathbb{P}(\mathcal{A} \mid \mathcal{R}, \mathcal{T}) \, \mathbb{P}(\mathcal{R} \mid \mathcal{T}) \, \mathbb{P}(\mathcal{T})$$

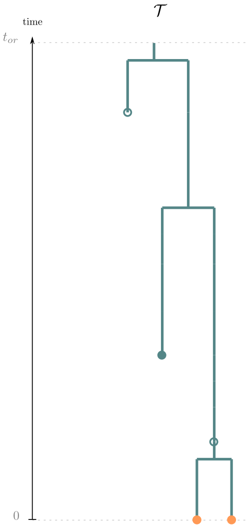
RESULT



Incorporating occurrences



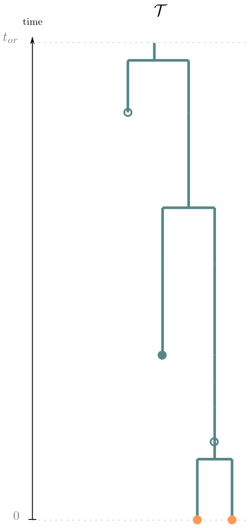
Motivation



We are given a tree $\mathcal{T}$:

- In epidemiology, samples are *infected individuals*.
- In macroevolution, samples are *species*.

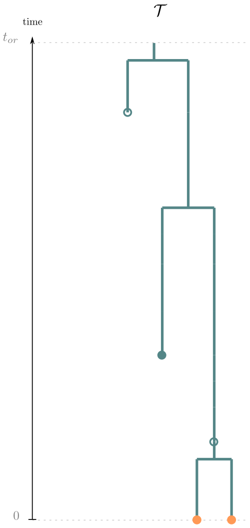
Motivation



We are given a tree $\mathcal{T}$:

- In epidemiology, samples are *infected individuals*.
- In macroevolution, samples are *species*.

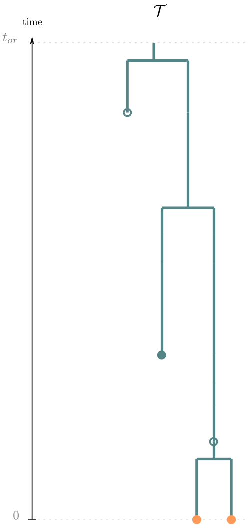
Motivation



We are given a tree $\mathcal{T}$:

- In epidemiology, samples are *infected individuals*.
- In macroevolution, samples are *species*.

Motivation



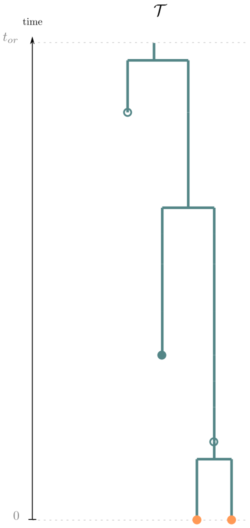
We are given a tree $\mathcal{T}$:

- In epidemiology, samples are *infected individuals*.
- In macroevolution, samples are *species*.

But additional signal could come from:

- case count data, i.e. non-sequenced individuals,
- undescribed fossils without any character data attached.

Motivation

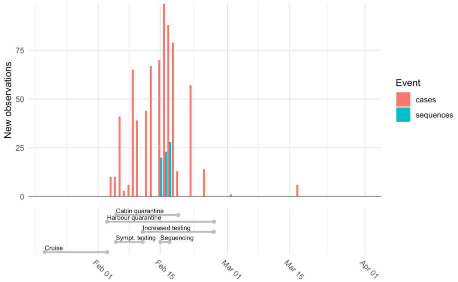


We are given a tree $\mathcal{T}$:

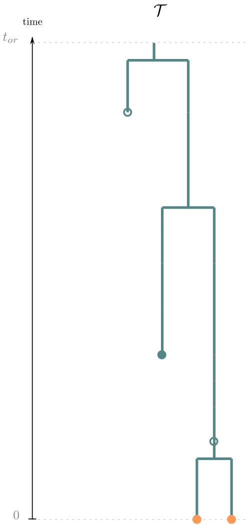
- In epidemiology, samples are *infected individuals*.
- In macroevolution, samples are *species*.

But additional signal could come from:

- case count data, i.e. non-sequenced individuals,
- undescribed fossils without any character data attached.



Motivation

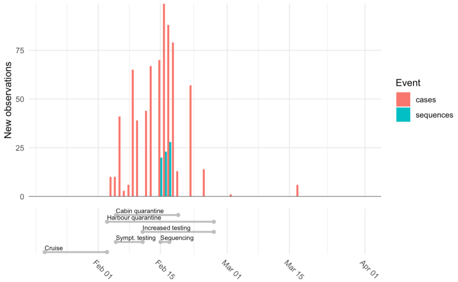


We are given a tree $\mathcal{T}$:

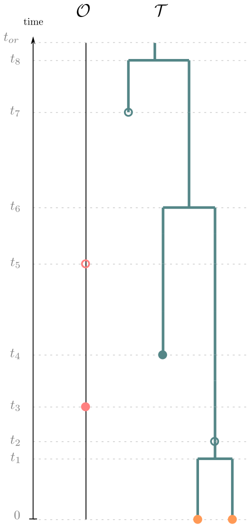
- In epidemiology, samples are *infected individuals*.
- In macroevolution, samples are *species*.

But additional signal could come from:

- case count data, i.e. non-sequenced individuals,
- undescribed fossils without any character data attached.



Motivation



We are given a tree $\mathcal{T}$:

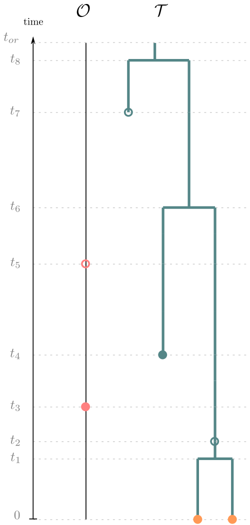
- In epidemiology, samples are *infected individuals*.
- In macroevolution, samples are *species*.

But additional signal could come from:

- case count data, i.e. non-sequenced individuals,
- undescribed fossils without any character data attached.

We call this a record of occurrences $\mathcal{O}$.

Motivation



We are given a tree $\mathcal{T}$:

- In epidemiology, samples are *infected individuals*.
- In macroevolution, samples are *species*.

But additional signal could come from:

- case count data, i.e. non-sequenced individuals,
- undescribed fossils without any character data attached.

We call this a record of occurrences $\mathcal{O}$.

What is the total number of individuals in the past ?

Model
following Vaughan et al, *MBE*, 2019

t_{or} origin

λ birth

ψ sampling and tree inclusion

μ death

ω sampling for occurrences

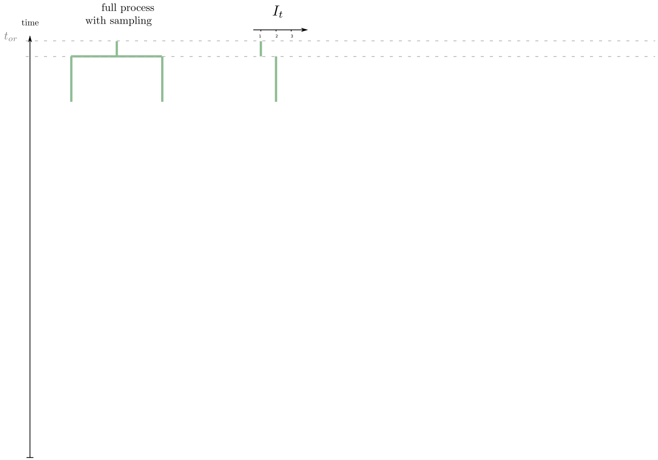
r removal at sampling

ρ sampling at present

Model

following Vaughan et al, *MBE*, 2019

- t_{or} origin
- λ birth
- ψ sampling and tree inclusion
- μ death
- ω sampling for occurrences
- r removal at sampling
- ρ sampling at present



Model

following Vaughan et al, *MBE*, 2019

t_{or} origin

λ birth

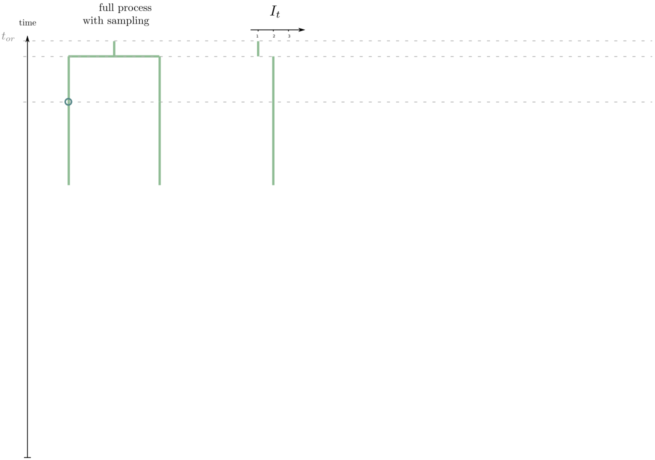
ψ sampling and tree inclusion

μ death

ω sampling for occurrences

r removal at sampling

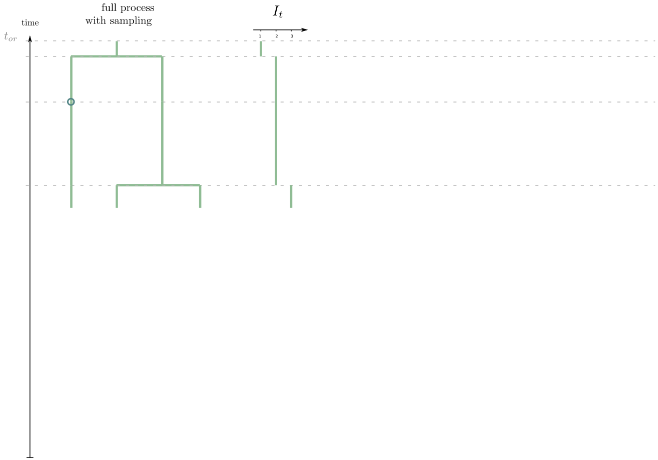
ρ sampling at present



Model

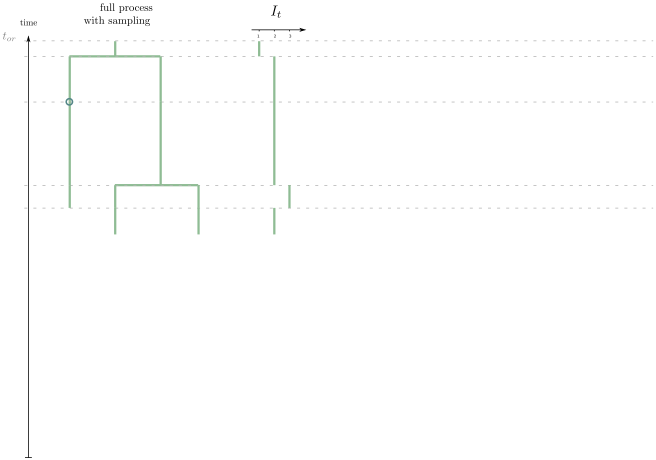
following Vaughan et al, *MBE*, 2019

- t_{or} origin
- λ birth
- ψ sampling and tree inclusion
- μ death
- ω sampling for occurrences
- r removal at sampling
- ρ sampling at present



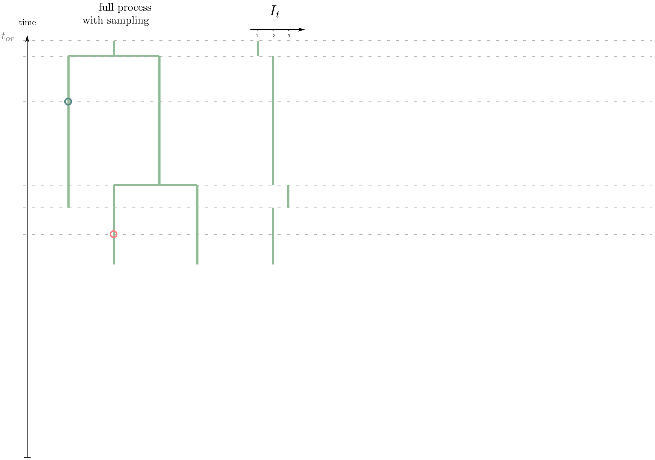
Model
following Vaughan et al, *MBE*, 2019

- t_{or} origin
- λ birth
- ψ sampling and tree inclusion
- μ death
- ω sampling for occurrences
- r removal at sampling
- ρ sampling at present



Model
following Vaughan et al, *MBE*, 2019

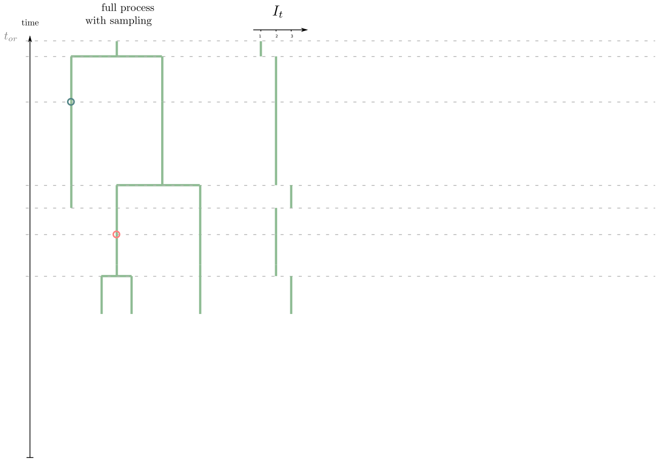
- t_{or} origin
- λ birth
- ψ sampling and tree inclusion
- μ death
- ω sampling for occurrences
- r removal at sampling
- ρ sampling at present



Model

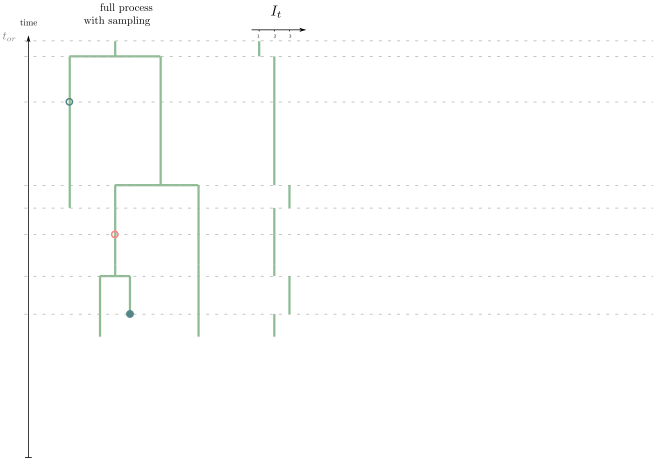
following Vaughan et al, *MBE*, 2019

- t_{or} origin
- λ birth
- ψ sampling and tree inclusion
- μ death
- ω sampling for occurrences
- r removal at sampling
- ρ sampling at present



Model
following Vaughan et al, *MBE*, 2019

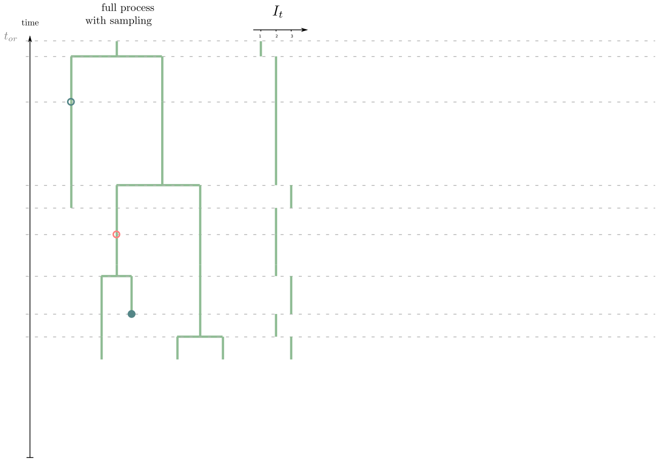
- t_{or} origin
- λ birth
- ψ sampling and tree inclusion
- μ death
- ω sampling for occurrences
- r removal at sampling
- ρ sampling at present



Model

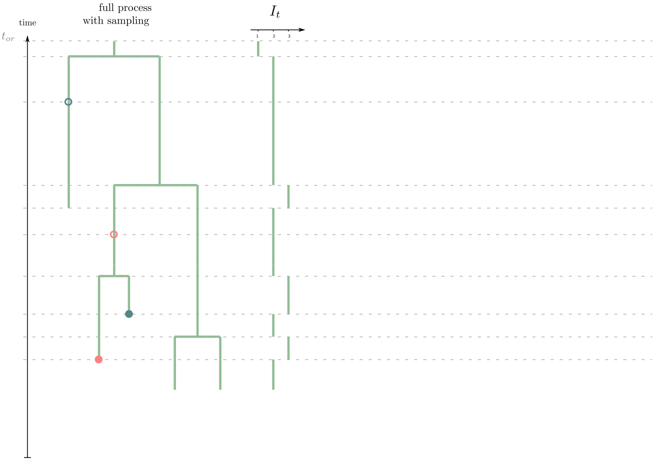
following Vaughan et al, *MBE*, 2019

- t_{or} origin
- λ birth
- ψ sampling and tree inclusion
- μ death
- ω sampling for occurrences
- r removal at sampling
- ρ sampling at present



Model
following Vaughan et al, *MBE*, 2019

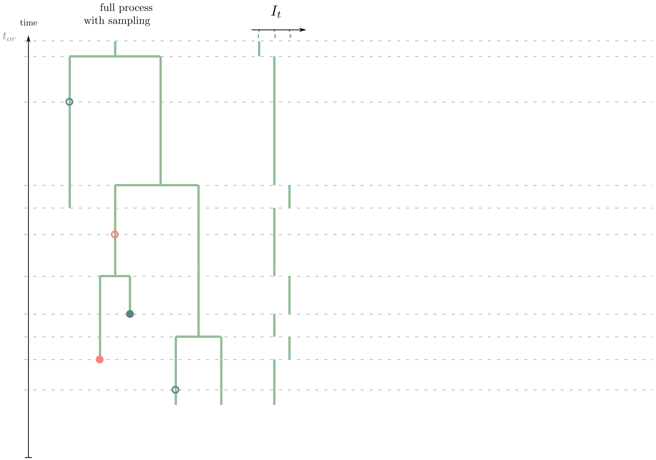
- t_{or} origin
- λ birth
- ψ sampling and tree inclusion
- μ death
- ω sampling for occurrences
- r removal at sampling
- ρ sampling at present



Model

following Vaughan et al, *MBE*, 2019

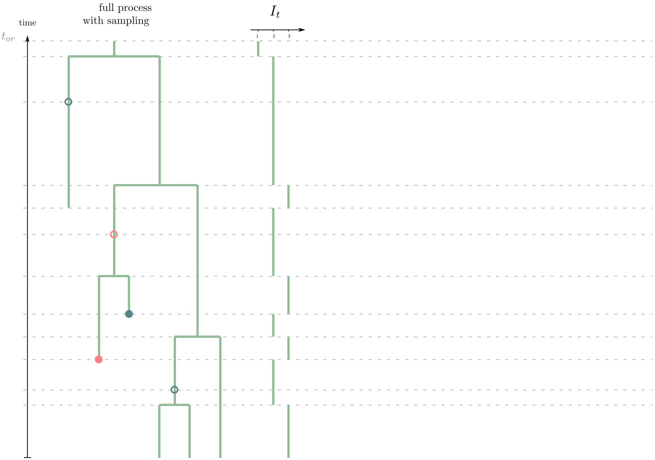
- t_{or} origin
- λ birth
- ψ sampling and tree inclusion
- μ death
- ω sampling for occurrences
- r removal at sampling
- ρ sampling at present



Model

following Vaughan et al, *MBE*, 2019

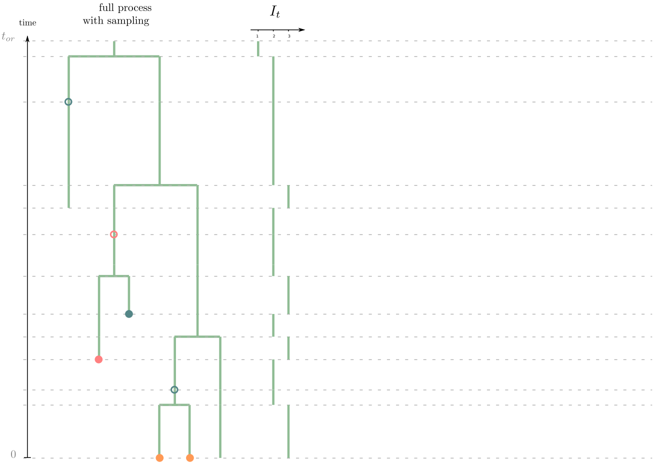
- t_{or} origin
- λ birth
- ψ sampling and tree inclusion
- μ death
- ω sampling for occurrences
- r removal at sampling
- ρ sampling at present



Model

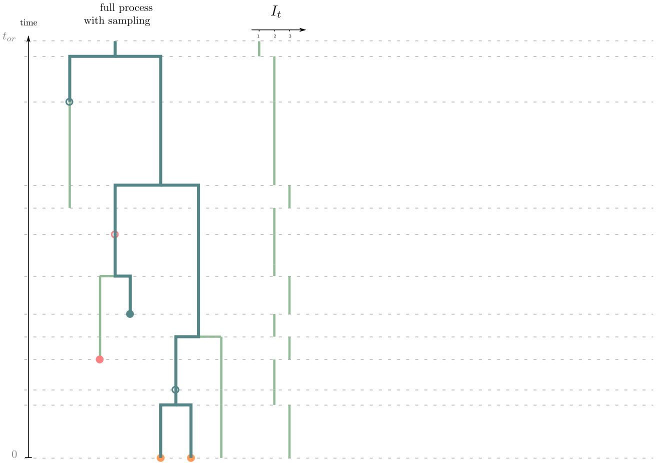
following Vaughan et al, *MBE*, 2019

- t_{or} origin
- λ birth
- ψ sampling and tree inclusion
- μ death
- ω sampling for occurrences
- r removal at sampling
- ρ sampling at present



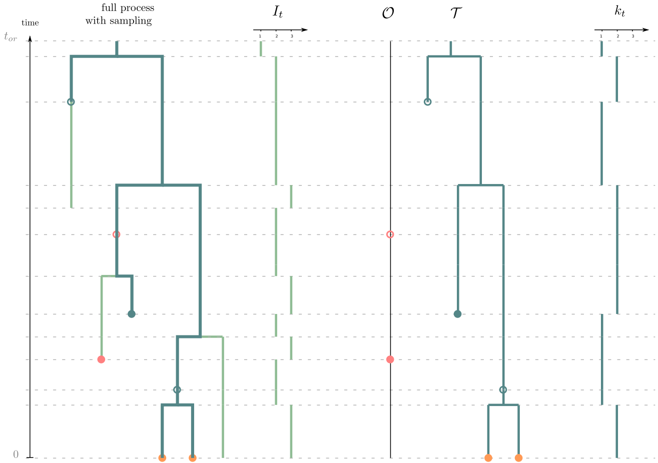
Model
following Vaughan et al, *MBE*, 2019

- t_{or} origin
- λ birth
- ψ sampling and tree inclusion
- μ death
- ω sampling for occurrences
- r removal at sampling
- ρ sampling at present



Model
following Vaughan et al, *MBE*, 2019

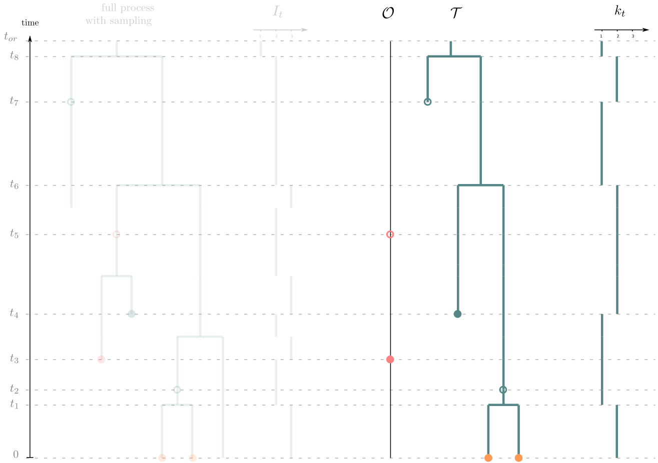
- t_{or} origin
- λ birth
- ψ sampling and tree inclusion
- μ death
- ω sampling for occurrences
- r removal at sampling
- ρ sampling at present



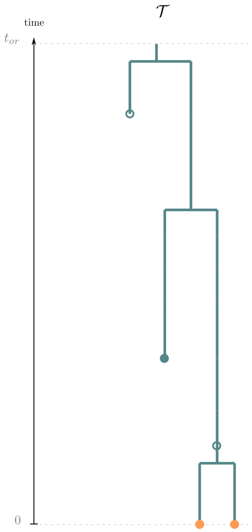
Model

following Vaughan et al, *MBE*, 2019

- t_{or} origin
- λ birth
- ψ sampling and tree inclusion
- μ death
- ω sampling for occurrences
- r removal at sampling
- ρ sampling at present



Incorporating occurrences
A bit of context



What is done, without occurrences,

- estimate $\hat{\lambda}, \hat{\mu}$ using the full tree
- compute $\mathbb{E}_{\hat{\lambda},\hat{\mu}}(I_t \mid I_{t_{or}} = 1)$.

Fast, but not accurate

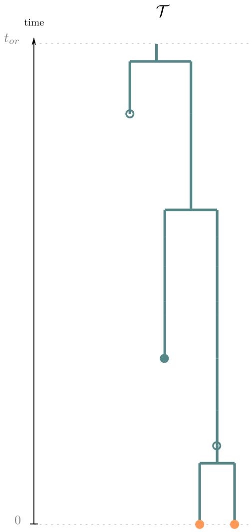
What has been proposed, with occurrences,

- simulate population size trajectories conditioned on the full tree
- to approximate $\mathbb{P}(I_t \mid \mathcal{T}, \mathcal{O})$.

Accurate, but slower

Incorporating occurrences

A bit of context



What is done, without occurrences,

- estimate $\hat{\lambda}, \hat{\mu}$ using the full tree
- compute $\mathbb{E}_{\hat{\lambda},\hat{\mu}}\left(I_t \mid I_{t_{or}} = 1\right)$.

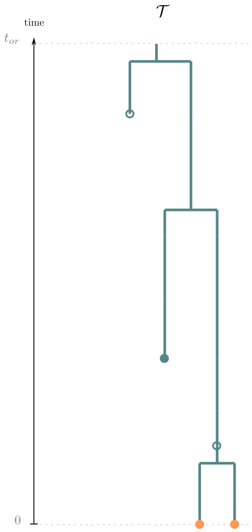
Fast, but not accurate

What has been proposed, with occurrences,

- simulate population size trajectories conditioned on the full tree
- to approximate $\mathbb{P}\left(I_t \mid \mathcal{T}, \mathcal{O}\right)$.

Accurate, but slower

Incorporating occurrences
A bit of context



What is done, without occurrences,

- estimate $\hat{\lambda}, \hat{\mu}$ using the full tree
- compute $\mathbb{E}_{\hat{\lambda},\hat{\mu}}\left(I_t \mid I_{t_{or}} = 1\right)$.

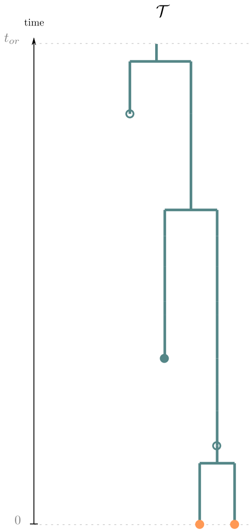
Fast, but not accurate

What has been proposed, with occurrences,

- simulate population size trajectories conditioned on the full tree
- to approximate $\mathbb{P}\left(I_t \mid \mathcal{T}, \mathcal{O}\right)$.

Accurate, but slower

Incorporating occurrences
A bit of context



What is done, without occurrences,

- estimate $\hat{\lambda}, \hat{\mu}$ using the full tree
- compute $\mathbb{E}_{\hat{\lambda},\hat{\mu}}\left(I_t \mid I_{t_{or}} = 1\right)$.

Fast, but not accurate

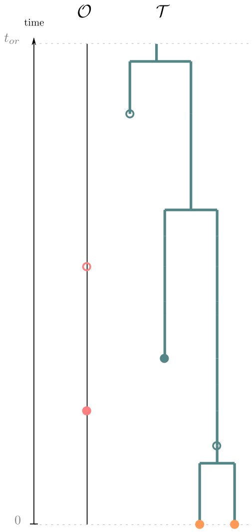
What has been proposed, with occurrences,

- simulate population size trajectories conditioned on the full tree
- to approximate $\mathbb{P}\left(I_t \mid \mathcal{T}, \mathcal{O}\right)$.

Accurate, but slower

Incorporating occurrences

A bit of context



What is done, without occurrences,

- estimate $\hat{\lambda}$, $\hat{\mu}$ using the full tree
- compute $\mathbb{E}_{\hat{\lambda},\hat{\mu}}\left(I_t \mid I_{t_{or}} = 1\right)$.

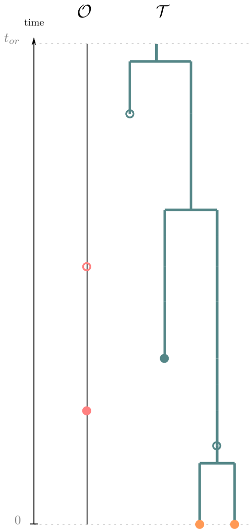
Fast, but not accurate

What has been proposed, with occurrences,

- simulate population size trajectories conditioned on the full tree
- to approximate $\mathbb{P}\left(I_t \mid \mathcal{T}, \mathcal{O}\right)$.

Accurate, but slower

Incorporating occurrences
A bit of context



What is done, without occurrences,

- estimate $\hat{\lambda}, \hat{\mu}$ using the full tree
- compute $\mathbb{E}_{\hat{\lambda},\hat{\mu}}(I_t \mid I_{t_{or}} = 1)$.

Fast, but not accurate

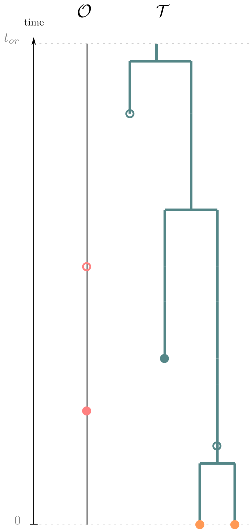
What has been proposed, with occurrences,

- simulate population size trajectories conditioned on the full tree
- to approximate $\mathbb{P}(I_t \mid \mathcal{T}, \mathcal{O})$.

Accurate, but slower

Incorporating occurrences

A bit of context



What is done, without occurrences,

- estimate $\hat{\lambda}, \hat{\mu}$ using the full tree
- compute $\mathbb{E}_{\hat{\lambda},\hat{\mu}}(I_t \mid I_{t_{or}} = 1)$.

Fast, but not accurate

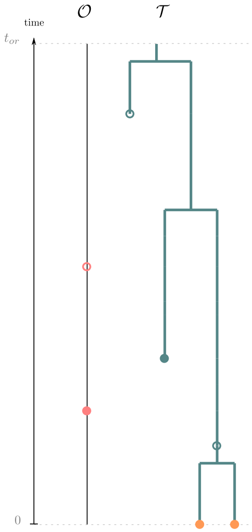
What has been proposed, with occurrences,

- simulate population size trajectories conditioned on the full tree
- to approximate $\mathbb{P}(I_t \mid \mathcal{T}, \mathcal{O})$.

Accurate, but slower

Incorporating occurrences

A bit of context



What is done, without occurrences,

- estimate $\hat{\lambda}, \hat{\mu}$ using the full tree
- compute $\mathbb{E}_{\hat{\lambda},\hat{\mu}}\left(I_t \mid I_{t_{or}} = 1\right)$.

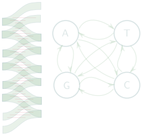
Fast, but not accurate

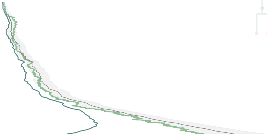
What has been proposed, with occurrences,

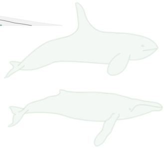
- simulate population size trajectories conditioned on the full tree
- to approximate $\mathbb{P}\left(I_t \mid \mathcal{T}, \mathcal{O}\right)$.

Accurate, but slower

The ancestral population size

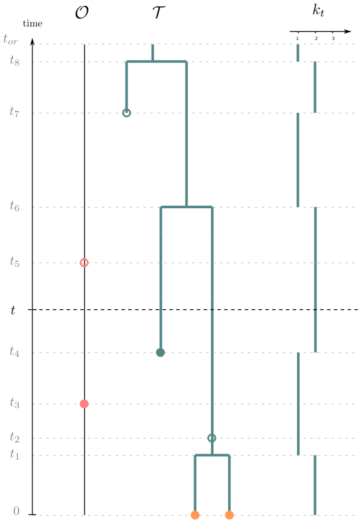






Marc Manceau, Antoine Zwaans, Jérémy Andréoletti, Ankit Gupta, Tim Vaughan, Rachel Warnock, Tanja Stadler

Sketch of the overall strategy



For any time t, we are interested in

$$K_t^{(i)} := \mathbb{P}\left(I_t = k_t + i \mid \mathcal{T}, \mathcal{O}\right)$$

We define

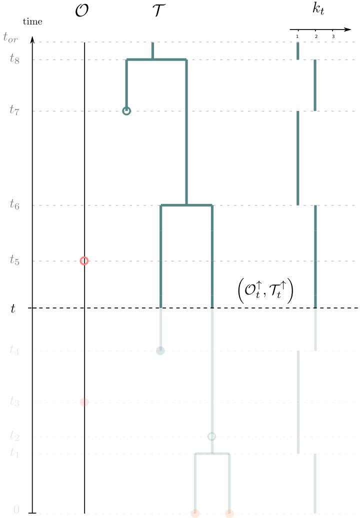
$$M_t^{(i)} := \mathbb{P}\left(T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, I_t = k_t + i\right)$$

$$L_t^{(i)} := \mathbb{P}\left(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i\right)$$

Then we get

$$\begin{aligned} K_t^{(i)} &\propto \mathbb{P}\left(I_t = k_t + i, T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, T_t^{\downarrow}, \mathcal{O}_t^{\downarrow}\right) \\ &\propto \mathbb{P}\left(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i, T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}\right) \\ &\quad \mathbb{P}\left(I_t = k_t + i, T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}\right) \\ &\propto L_t^{(i)} M_t^{(i)} \end{aligned}$$

Sketch of the overall strategy



For any time t, we are interested in

$$K_t^{(i)} := \mathbb{P}\left(I_t = k_t + i \mid \mathcal{T}, \mathcal{O}\right)$$

We define

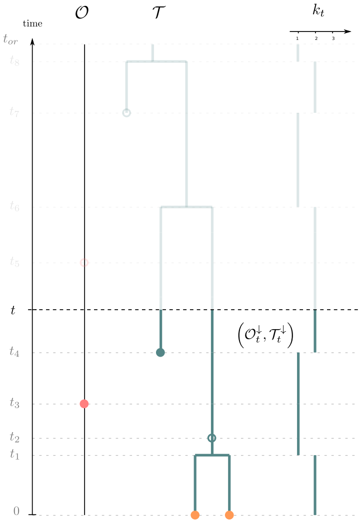
$$M_t^{(i)} := \mathbb{P}\left(T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, I_t = k_t + i\right)$$

$$L_t^{(i)} := \mathbb{P}\left(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i\right)$$

Then we get

$$\begin{aligned} K_t^{(i)} &\propto \mathbb{P}\left(I_t = k_t + i, T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, T_t^{\downarrow}, \mathcal{O}_t^{\downarrow}\right) \\ &\propto \mathbb{P}\left(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i, T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}\right) \\ &\quad \mathbb{P}\left(I_t = k_t + i, T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}\right) \\ &\propto L_t^{(i)} M_t^{(i)} \end{aligned}$$

Sketch of the overall strategy



For any time t, we are interested in

$$K_t^{(i)} := \mathbb{P}\left(I_t = k_t + i \mid \mathcal{T}, \mathcal{O}\right)$$

We define

$$M_t^{(i)} := \mathbb{P}\left(T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, I_t = k_t + i\right)$$

$$L_t^{(i)} := \mathbb{P}\left(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i\right)$$

Then we get

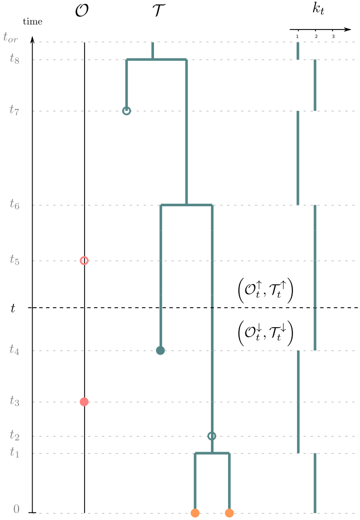
$$K_t^{(i)} \propto \mathbb{P}\left(I_t = k_t + i, T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, T_t^{\downarrow}, \mathcal{O}_t^{\downarrow}\right)$$

$$\propto \mathbb{P}\left(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i, T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}\right)$$

$$\mathbb{P}\left(I_t = k_t + i, T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}\right)$$

$$\propto L_t^{(i)} M_t^{(i)}$$

Sketch of the overall strategy



For any time t, we are interested in

$$K_t^{(i)} := \mathbb{P}\left(I_t = k_t + i \mid \mathcal{T}, \mathcal{O}\right)$$

We define

$$M_t^{(i)} := \mathbb{P}\left(T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, I_t = k_t + i\right)$$

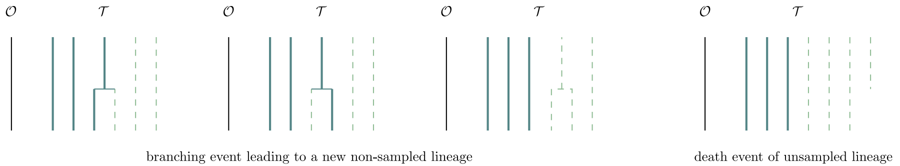
$$L_t^{(i)} := \mathbb{P}\left(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i\right)$$

Then we get

$$\begin{aligned} K_t^{(i)} &\propto \mathbb{P}\left(I_t = k_t + i, T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, T_t^{\downarrow}, \mathcal{O}_t^{\downarrow}\right) \\ &\propto \mathbb{P}\left(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i, T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}\right) \\ &\quad \mathbb{P}\left(I_t = k_t + i, T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}\right) \\ &\propto L_t^{(i)} M_t^{(i)} \end{aligned}$$

The ancestral population size

Forward-backward traversal of the tree to compute $M_t = \left(\mathbb{P}(T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, I_t = k_t + i)\right)_{i\geq 0}$



We can write the Master equation, $\forall i \in \mathbb{N}$,

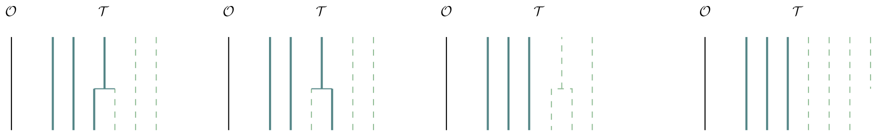
$$M_{t-\delta t}^{(i)} = (1 - (\lambda + \mu + \psi + \omega)(i + k_t)\delta t)\, M_t^{(i)} + \lambda(2k_t + i - 1)\delta t \mathbb{1}_{i>0} M_t^{(i-1)} + \mu(i+1)\delta t M_t^{(i+1)}$$

Leading to a system of ODEs,

$$\frac{dM_t^{(i)}}{dt} = (\lambda + \mu + \psi + \omega)(i + k_t)M_t^{(i)} - \lambda(2k_t + i - 1)\mathbb{1}_{i>0} M_t^{(i-1)} - \mu(i+1)M_t^{(i+1)}$$

The ancestral population size

Forward-backward traversal of the tree to compute $M_t = \left(\mathbb{P}(T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, I_t = k_t + i)\right)_{i \geq 0}$



branching event leading to a new non-sampled lineage

death event of unsampled lineage

We can write the Master equation, $\forall i \in \mathbb{N}$,

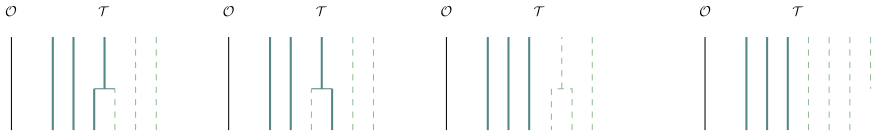
$$M_{t-\delta t}^{(i)} = (1 - (\lambda + \mu + \psi + \omega)(i + k_t)\delta t)\, M_t^{(i)} + \lambda(2k_t + i - 1)\delta t \mathbb{1}_{i>0} M_t^{(i-1)} + \mu(i+1)\delta t M_t^{(i+1)}$$

Leading to a system of ODEs,

$$\frac{dM_t^{(i)}}{dt} = (\lambda + \mu + \psi + \omega)(i + k_t)M_t^{(i)} - \lambda(2k_t + i - 1)\mathbb{1}_{i>0} M_t^{(i-1)} - \mu(i+1)M_t^{(i+1)}$$

The ancestral population size

Forward-backward traversal of the tree to compute $M_t = \left(\mathbb{P}(T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, I_t = k_t + i)\right)_{i \geq 0}$



branching event leading to a new non-sampled lineage

death event of unsampled lineage

We can write the Master equation, $\forall i \in \mathbb{N}$,

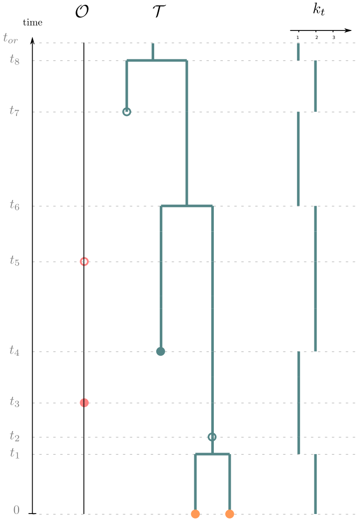
$$M_{t-\delta t}^{(i)} = (1 - (\lambda + \mu + \psi + \omega)(i + k_t)\delta t)\, M_t^{(i)} + \lambda(2k_t + i - 1)\delta t \mathbb{1}_{i>0} M_t^{(i-1)} + \mu(i+1)\delta t M_t^{(i+1)}$$

Leading to a system of ODEs,

$$\frac{dM_t^{(i)}}{dt} = (\lambda + \mu + \psi + \omega)(i + k_t) M_t^{(i)} - \lambda(2k_t + i - 1)\mathbb{1}_{i>0} M_t^{(i-1)} - \mu(i+1) M_t^{(i+1)}$$

Forward-backward traversal of the tree

A forward breadth-first traversal to compute $M_t = \left(\mathbb{P}(T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, I_t = k_t + i)\right)_{i \geq 0}$

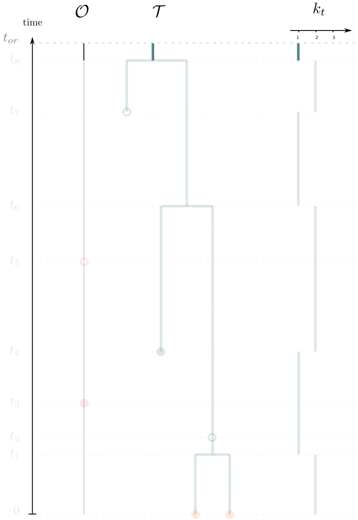


We know how to initialize M_t at the time of origin

$$M_{t_{or}}^{(i)} = \mathbb{P}(I_{t_{or}} = 1 + i) = \mathbb{1}_{i=0}$$

Forward-backward traversal of the tree

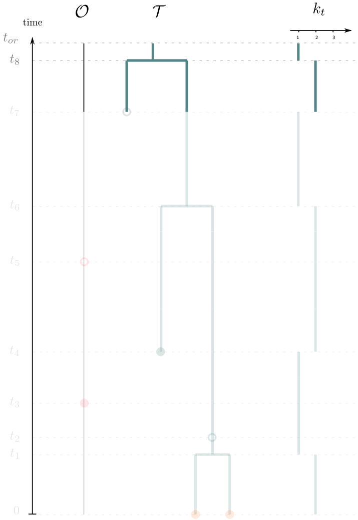
A forward breadth-first traversal to compute $M_t = \left(\mathbb{P}(T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, I_t = k_t + i)\right)_{i \geq 0}$



Between two events, M_t evolves following an ODE

Forward-backward traversal of the tree

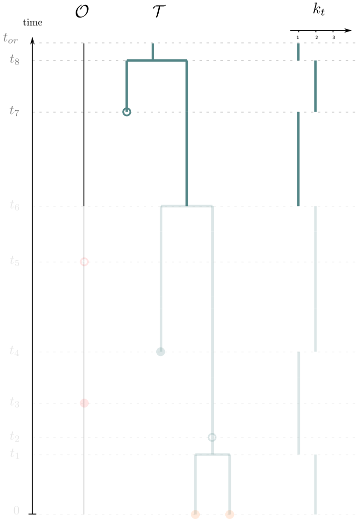
A forward breadth-first traversal to compute $M_t = \left(\mathbb{P}(T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, I_t = k_t + i)\right)_{i \geq 0}$



$$M_{t-} = \lambda M_{t+}$$

Forward-backward traversal of the tree

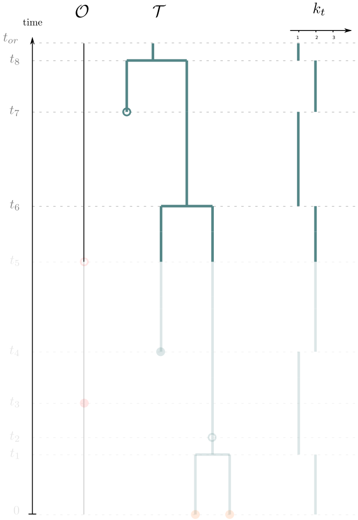
A forward breadth-first traversal to compute $M_t = \left(\mathbb{P}(T_t^\uparrow, \mathcal{O}_t^\uparrow, I_t = k_t + i)\right)_{i \geq 0}$



$$M_{t-}^{(i)} = \psi(1-r)\mathbb{1}_{i>0} M_{t+}^{(i-1)}$$

Forward-backward traversal of the tree

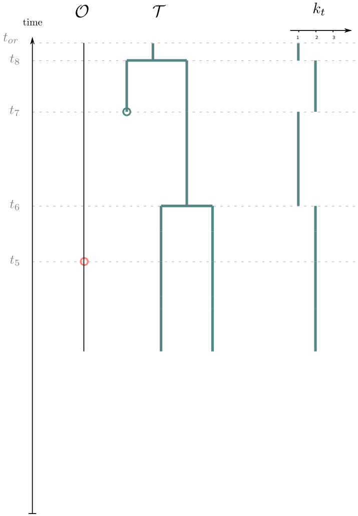
A forward breadth-first traversal to compute $M_t = \left(\mathbb{P}(T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, I_t = k_t + i)\right)_{i \geq 0}$



$$M_{t-} = \lambda M_{t+}$$

Forward-backward traversal of the tree

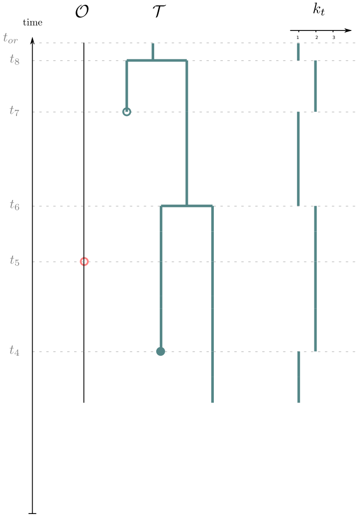
A forward breadth-first traversal to compute $M_t = \left(\mathbb{P}(T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, I_t = k_t + i)\right)_{i \geq 0}$



$$M_{t-}^{(i)} = \omega(1-r)(k_t + i)M_{t+}^{(i)}$$

Forward-backward traversal of the tree

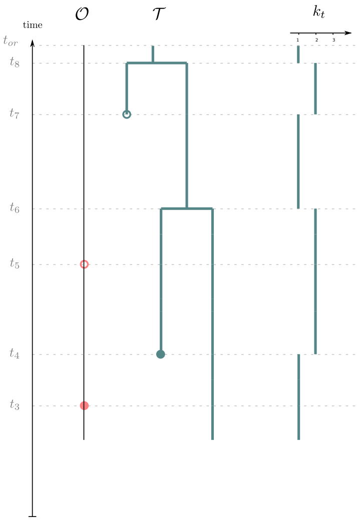
A forward breadth-first traversal to compute $M_t = \left(\mathbb{P}(T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, I_t = k_t + i)\right)_{i \geq 0}$



$$M_{t-} = \psi r M_{t+}$$

Forward-backward traversal of the tree

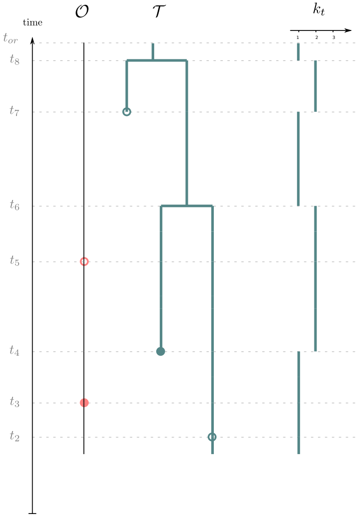
A forward breadth-first traversal to compute $M_t = \left(\mathbb{P}(T_t^\uparrow, \mathcal{O}_t^\uparrow, I_t = k_t + i)\right)_{i \geq 0}$



$$M_{t-}^{(i)} = \omega r (i+1) M_{t+}^{(i+1)}$$

Forward-backward traversal of the tree

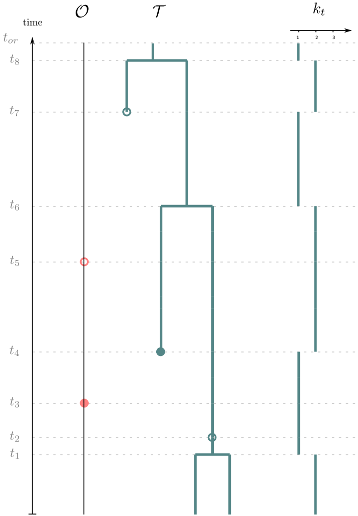
A forward breadth-first traversal to compute $M_t = \left(\mathbb{P}(T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, I_t = k_t + i)\right)_{i \geq 0}$



$$M_{t-} = \psi(1 - r)M_{t+}$$

Forward-backward traversal of the tree

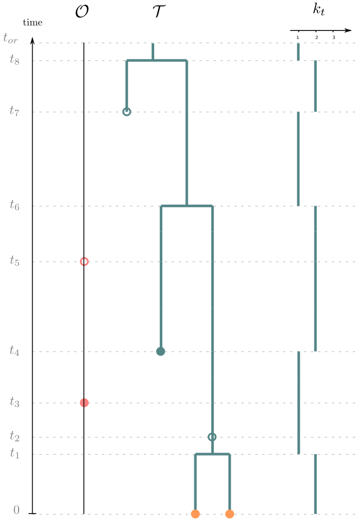
A forward breadth-first traversal to compute $M_t = \left(\mathbb{P}(T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, I_t = k_t + i)\right)_{i \geq 0}$



$$M_{t-} = \lambda M_{t+}$$

Forward-backward traversal of the tree

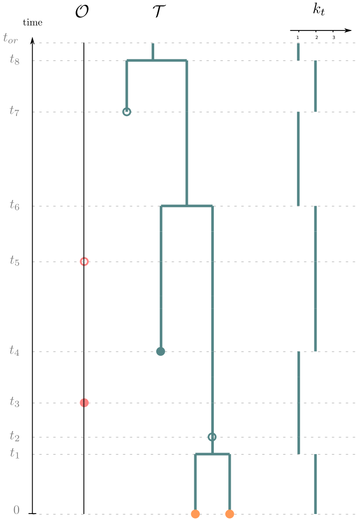
A forward breadth-first traversal to compute $M_t = \left(\mathbb{P}(T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, I_t = k_t + i)\right)_{i \geq 0}$



$$M_0^{(i)} = \rho^{k_0}(1-\rho)^i M_{0+}^{(i)}$$

Forward-backward traversal of the tree

A forward breadth-first traversal to compute $M_t = \left(\mathbb{P}(T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, I_t = k_t + i)\right)_{i \geq 0}$



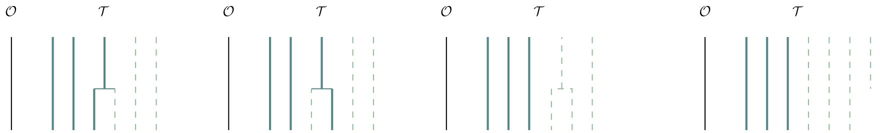
As a by-product, we get the likelihood in the end,

$$\mathcal{L} = \sum_{i=0}^{\infty} \mathbb{P}\left(\mathcal{T}, \mathcal{O}, I_{t_0} = k_0 + i\right)$$

$$= \sum_{i=0}^{\infty} M_0^{(i)}$$

The ancestral population size

Forward-backward traversal of the tree to compute $L_t = \left(\mathbb{P}(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i)\right)_{i \geq 0}$



branching event leading to a new non-sampled lineage

death event of unsampled lineage

We can write the Master equation, $\forall i \in \mathbb{N}$,

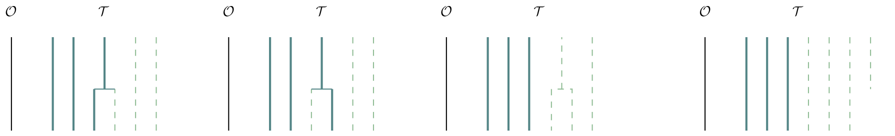
$$L_{t+\delta t}^{(i)} = (1 - (\lambda + \mu + \psi + \omega)(k_t + i)\delta t)\, L_t^{(i)} + \lambda(2k_t + i)\delta t L_t^{(i+1)} + \mu i \delta t L_t^{(i-1)}$$

Leading to a system of ODEs,

$$\frac{dL_t^{(i)}}{dt} = -(\lambda + \mu + \psi + \omega)(k_t + i)L_t^{(i)} + \lambda(2k_t + i)L_t^{(i+1)} + \mu i L_t^{(i-1)}$$

The ancestral population size

Forward-backward traversal of the tree to compute $L_t = \left(\mathbb{P}(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i)\right)_{i \geq 0}$



branching event leading to a new non-sampled lineage

death event of unsampled lineage

We can write the Master equation, $\forall i \in \mathbb{N}$,

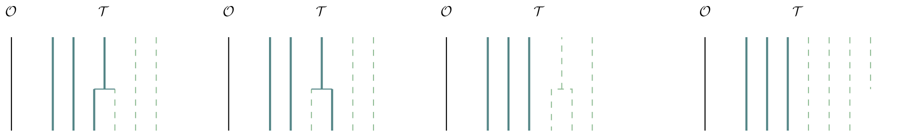
$$L_{t+\delta t}^{(i)} = (1 - (\lambda + \mu + \psi + \omega)(k_t + i)\delta t)\, L_t^{(i)} + \lambda(2k_t + i)\delta t L_t^{(i+1)} + \mu i \delta t L_t^{(i-1)}$$

Leading to a system of ODEs,

$$\frac{dL_t^{(i)}}{dt} = -(\lambda + \mu + \psi + \omega)(k_t + i)L_t^{(i)} + \lambda(2k_t + i)L_t^{(i+1)} + \mu i L_t^{(i-1)}$$

The ancestral population size

Forward-backward traversal of the tree to compute $L_t = \left(\mathbb{P}(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i)\right)_{i \geq 0}$



We can write the Master equation, $\forall i \in \mathbb{N}$,

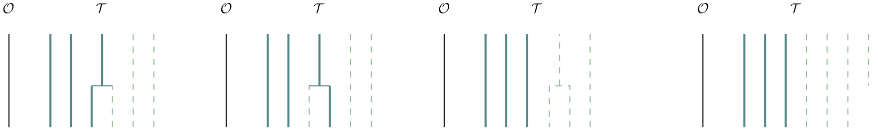
$$L_{t+\delta t}^{(i)} = (1 - (\lambda + \mu + \psi + \omega)(k_t + i)\delta t)\, L_t^{(i)} + \lambda(2k_t + i)\delta t L_t^{(i+1)} + \mu i \delta t L_t^{(i-1)}$$

Leading to a system of ODEs,

$$\frac{dL_t^{(i)}}{dt} = -(\lambda + \mu + \psi + \omega)(k_t + i)L_t^{(i)} + \lambda(2k_t + i)L_t^{(i+1)} + \mu i L_t^{(i-1)}$$

The ancestral population size

Forward-backward traversal of the tree to compute $L_t = \left(\mathbb{P}(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i)\right)_{i \geq 0}$



branching event leading to a new non-sampled lineage

death event of unsampled lineage

We can write the Master equation, $\forall i \in \mathbb{N}$,

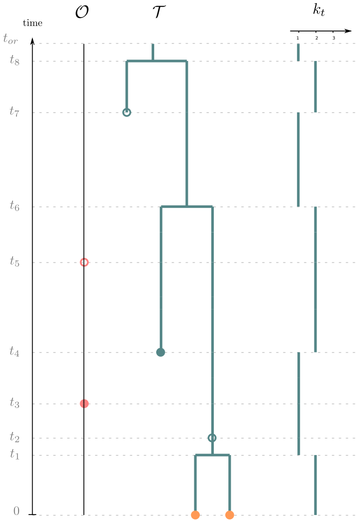
$$L_{t+\delta t}^{(i)} = (1 - (\lambda + \mu + \psi + \omega)(k_t + i)\delta t)\, L_t^{(i)} + \lambda(2k_t + i)\delta t L_t^{(i+1)} + \mu i \delta t L_t^{(i-1)}$$

Leading to a system of ODEs,

$$\frac{dL_t^{(i)}}{dt} = -(\lambda + \mu + \psi + \omega)(k_t + i) L_t^{(i)} + \lambda(2k_t + i) L_t^{(i+1)} + \mu i L_t^{(i-1)}$$

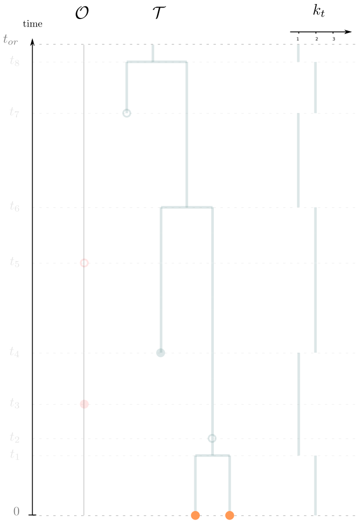
Forward-backward traversal of the tree

A backward breadth-first traversal to compute $L_t = \left(\mathbb{P}(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i)\right)_{i \geq 0}$



Forward-backward traversal of the tree

A backward breadth-first traversal to compute $L_t = \left(\mathbb{P}(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i)\right)_{i \geq 0}$

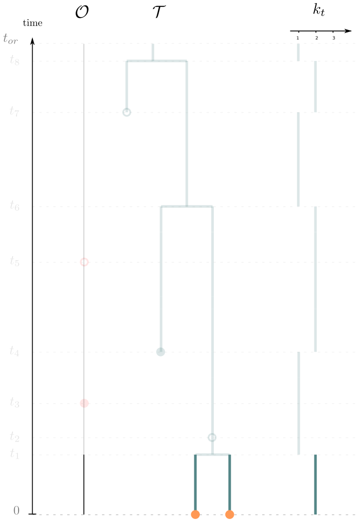


We know how to initialize L_t at present.

$$L_0^{(i)} = \rho^{k_0}(1-\rho)^i$$

Forward-backward traversal of the tree

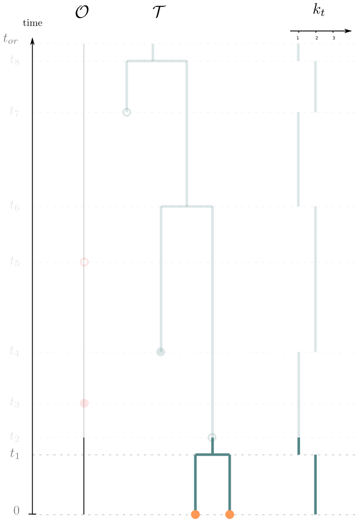
A backward breadth-first traversal to compute $L_t = \left(\mathbb{P}(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i)\right)_{i \geq 0}$



Between two events, L_t evolves following an ODE.

Forward-backward traversal of the tree

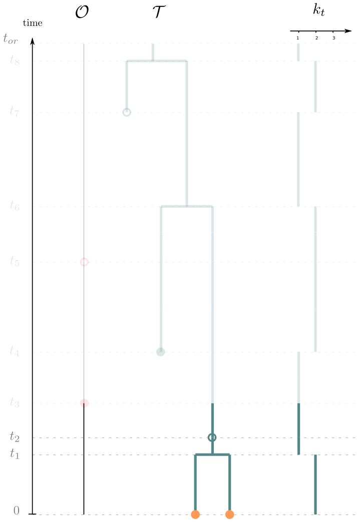
A backward breadth-first traversal to compute $L_t = \left(\mathbb{P}(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i)\right)_{i \geq 0}$



$$L_{t^+} = \lambda L_{t^-}$$

Forward-backward traversal of the tree

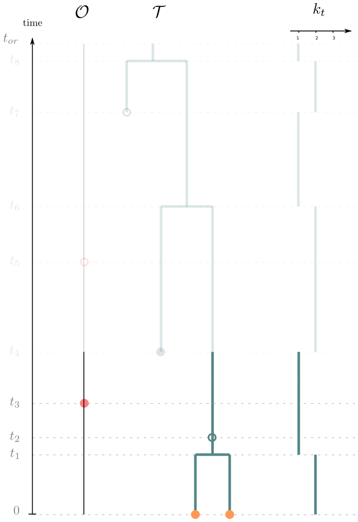
A backward breadth-first traversal to compute $L_t = \left(\mathbb{P}(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i)\right)_{i \geq 0}$



$$L_{t^+} = \psi(1-r)L_{t^-}$$

Forward-backward traversal of the tree

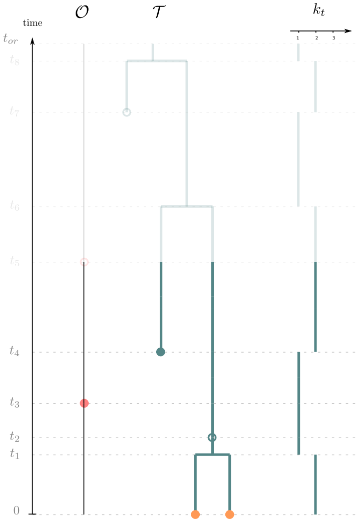
A backward breadth-first traversal to compute $L_t = \left(\mathbb{P}(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i)\right)_{i \geq 0}$



$$L_{t^+}^{(i)} = \omega r i L_{t^-}^{(i-1)}$$

Forward-backward traversal of the tree

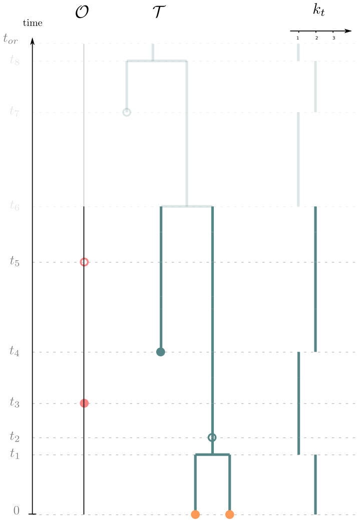
A backward breadth-first traversal to compute $L_t = \left(\mathbb{P}(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i)\right)_{i \geq 0}$



$$L_{t^+} = \psi r L_{t^-}$$

Forward-backward traversal of the tree

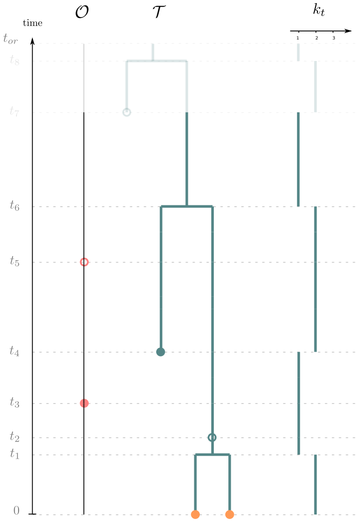
A backward breadth-first traversal to compute $L_t = \left(\mathbb{P}(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i)\right)_{i \geq 0}$



$$L_{t^+}^{(i)} = \omega(1-r)(k_t+i)L_{t^-}^{(i)}$$

Forward-backward traversal of the tree

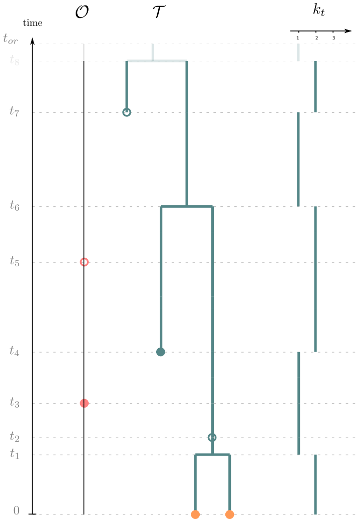
A backward breadth-first traversal to compute $L_t = \left(\mathbb{P}(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i)\right)_{i \geq 0}$



$$L_{t^+} = \lambda L_{t^-}$$

Forward-backward traversal of the tree

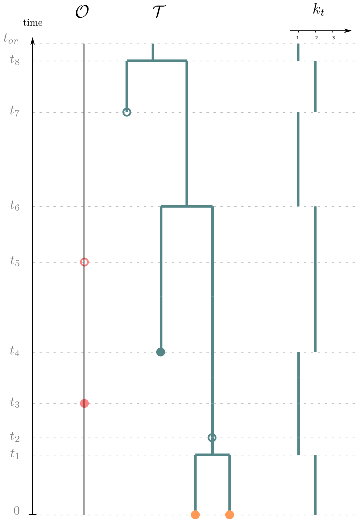
A backward breadth-first traversal to compute $L_t = \left(\mathbb{P}(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i)\right)_{i \geq 0}$



$$L_{t^+}^{(i)} = \psi(1 - r) L_{t^-}^{(i+1)}$$

Forward-backward traversal of the tree

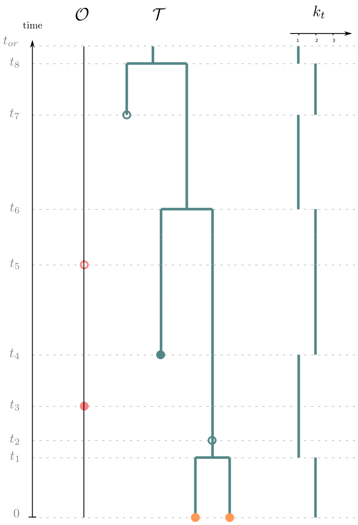
A backward breadth-first traversal to compute $L_t = \left(\mathbb{P}(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i)\right)_{i \geq 0}$



$$L_{t^+} = \lambda L_{t^-}$$

Forward-backward traversal of the tree

A backward breadth-first traversal to compute $L_t = \left(\mathbb{P}(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i)\right)_{i \geq 0}$



As a by-product, we get the likelihood in the end,

$$\mathcal{L} = \mathbb{P}(\mathcal{T}, \mathcal{O} \mid I_{t_{or}} = 1)$$
$$= L_{t_{or}}^{(0)}$$

The ancestral population size

Known corrolaries

Recall that M_t verifies:

$$\frac{dM_t^{(i)}}{dt} = \gamma(i+k)M_t^{(i)} - \lambda(2k+i-1)\mathbb{1}_{i>0}M_t^{(i-1)} - \mu(i+1)M_t^{(i+1)}$$
$$M_{t_{or}}^{(i)} = \mathbb{1}_{i=0}$$

We introduce the corresponding probability generating function:

$$\hat{M}(t,z) = \sum_{i=0}^{\infty} z^i M_t^{(i)}$$

The initial condition translates as $\forall z, \quad \hat{M}(t_{or}, z) = 1$
And the ODE translates as the following PDE:

$$\partial_t \hat{M} = -k(2\lambda z - \gamma)\hat{M} - (\lambda z^2 - \gamma z + \mu)\partial_z \hat{M}$$

- This can be solved analytically to get $\hat{M}$.
- and L_t can also be solved analytically, so far only when $r = 1$.

The ancestral population size

Known corrolaries

Recall that M_t verifies:

$$\frac{dM_t^{(i)}}{dt} = \gamma(i+k)M_t^{(i)} - \lambda(2k+i-1)\mathbb{1}_{i>0}M_t^{(i-1)} - \mu(i+1)M_t^{(i+1)}$$
$$M_{t_{or}}^{(i)} = \mathbb{1}_{i=0}$$

We introduce the corresponding probability generating function:

$$\hat{M}(t,z) = \sum_{i=0}^{\infty} z^i M_t^{(i)}$$

The initial condition translates as $\forall z, \quad \hat{M}(t_{or}, z) = 1$
And the ODE translates as the following PDE:

$$\partial_t \hat{M} = -k(2\lambda z - \gamma)\hat{M} - (\lambda z^2 - \gamma z + \mu)\partial_z \hat{M}$$

- This can be solved analytically to get $\hat{M}$.
- and L_t can also be solved analytically, so far only when $r = 1$.

The ancestral population size

Known corrolaries

Recall that M_t verifies:

$$\frac{dM_t^{(i)}}{dt} = \gamma(i+k)M_t^{(i)} - \lambda(2k+i-1)\mathbb{1}_{i>0}M_t^{(i-1)} - \mu(i+1)M_t^{(i+1)}$$
$$M_{t_{or}}^{(i)} = \mathbb{1}_{i=0}$$

We introduce the corresponding probability generating function:

$$\hat{M}(t,z) = \sum_{i=0}^{\infty} z^i M_t^{(i)}$$

The initial condition translates as $\forall z, \quad \hat{M}(t_{or}, z) = 1$

And the ODE translates as the following PDE:

$$\partial_t \hat{M} = -k(2\lambda z - \gamma)\hat{M} - (\lambda z^2 - \gamma z + \mu)\partial_z \hat{M}$$

- This can be solved analytically to get $\hat{M}$.
- and L_t can also be solved analytically, so far only when $r = 1$.

The ancestral population size

Known corrolaries

Recall that M_t verifies:

$$\frac{dM_t^{(i)}}{dt} = \gamma(i+k)M_t^{(i)} - \lambda(2k+i-1)\mathbb{1}_{i>0}M_t^{(i-1)} - \mu(i+1)M_t^{(i+1)}$$
$$M_{t_{or}}^{(i)} = \mathbb{1}_{i=0}$$

We introduce the corresponding probability generating function:

$$\hat{M}(t,z) = \sum_{i=0}^{\infty} z^i M_t^{(i)}$$

The initial condition translates as $\forall z, \quad \hat{M}(t_{or}, z) = 1$
And the ODE translates as the following PDE:

$$\partial_t \hat{M} = -k(2\lambda z - \gamma)\hat{M} - (\lambda z^2 - \gamma z + \mu)\partial_z \hat{M}$$

- This can be solved analytically to get $\hat{M}$.
- and L_t can also be solved analytically, so far only when $r = 1$.

The ancestral population size

Known corrolaries

Recall that M_t verifies:

$$\frac{dM_t^{(i)}}{dt} = \gamma(i+k)M_t^{(i)} - \lambda(2k+i-1)\mathbb{1}_{i>0}M_t^{(i-1)} - \mu(i+1)M_t^{(i+1)}$$

$$M_{t_{or}}^{(i)} = \mathbb{1}_{i=0}$$

We introduce the corresponding probability generating function:

$$\hat{M}(t,z) = \sum_{i=0}^{\infty} z^i M_t^{(i)}$$

The initial condition translates as $\forall z, \quad \hat{M}(t_{or}, z) = 1$
And the ODE translates as the following PDE:

$$\partial_t \hat{M} = -k(2\lambda z - \gamma)\hat{M} - (\lambda z^2 - \gamma z + \mu)\partial_z \hat{M}$$

- This can be solved analytically to get $\hat{M}$.
- and L_t can also be solved analytically, so far only when $r = 1$.

The ancestral population size

Known corrolaries

Recall that M_t verifies:

$$\frac{dM_t^{(i)}}{dt} = \gamma(i+k)M_t^{(i)} - \lambda(2k+i-1)\mathbb{1}_{i>0}M_t^{(i-1)} - \mu(i+1)M_t^{(i+1)}$$
$$M_{t_{or}}^{(i)} = \mathbb{1}_{i=0}$$

We introduce the corresponding probability generating function:

$$\hat{M}(t,z) = \sum_{i=0}^{\infty} z^i M_t^{(i)}$$

The initial condition translates as $\forall z, \quad \hat{M}(t_{or}, z) = 1$
And the ODE translates as the following PDE:

$$\partial_t \hat{M} = -k(2\lambda z - \gamma)\hat{M} - (\lambda z^2 - \gamma z + \mu)\partial_z \hat{M}$$

- This can be solved analytically to get $\hat{M}$.
- and L_t can also be solved analytically, so far only when $r = 1$.

The ancestral population size

Known corrolaries

I need to consider the parameter space

I want to compute

	$\omega = 0$	$r = 1$	$\omega \neq 0$ and $r \neq 1$
$\mathbb{P}(\mathcal{T}, \mathcal{O})$	Corollary 3.2	Proposition 3.3	Algorithm 1' or Algorithm 2'
K_t	Proposition 3.1 & Corollary 4.4	Proposition 3.3 & Algorithm 2	Algorithm 1 & Algorithm 2

The ancestral population size

Known corrolaries

I need to consider the parameter space

I want to compute	$\omega = 0$	$r = 1$	$\omega \neq 0$ and $r \neq 1$
$\mathbb{P}(\mathcal{T}, \mathcal{O})$	Corollary 3.2	Proposition 3.3	Algorithm 1' or Algorithm 2'
K_t	Proposition 3.1 & Corollary 4.4	Proposition 3.3 & Algorithm 2	Algorithm 1 & Algorithm 2

The ancestral population size

Known corrolaries

I need to consider the parameter space

I want to compute

	$\omega = 0$	$r = 1$	$\omega \neq 0$ and $r \neq 1$
$\mathbb{P}(\mathcal{T}, \mathcal{O})$	Corollary 3.2	Proposition 3.3	Algorithm 1' or Algorithm 2'
K_t	Proposition 3.1 & Corollary 4.4	Proposition 3.3 & Algorithm 2	Algorithm 1 & Algorithm 2

The ancestral population size

Known corrolaries

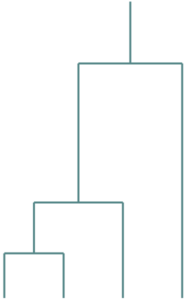
I need to consider the parameter space

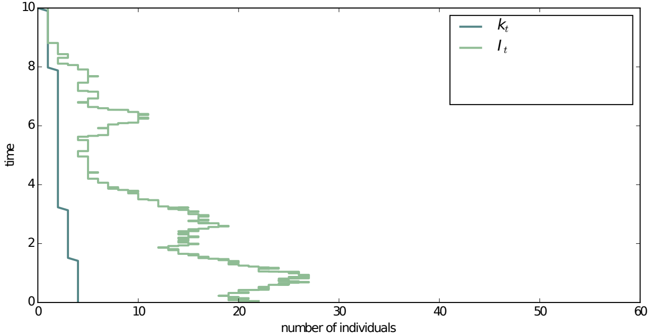
I want to compute

	$\omega = 0$	$r = 1$	$\omega \neq 0$ and $r \neq 1$
$\mathbb{P}(\mathcal{T}, \mathcal{O})$	Corollary 3.2	Proposition 3.3	Algorithm 1' or Algorithm 2'
K_t	Proposition 3.1 & Corollary 4.4	Proposition 3.3 & Algorithm 2	Algorithm 1 & Algorithm 2

Reconstructing past population size

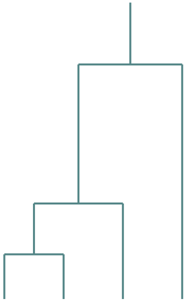
- With only ρ-sampling ($\psi, \omega = 0, 0$).
- With ρ and ψ-sampling ($\omega = 0$).
- With ρ, ψ, and ω-sampling.

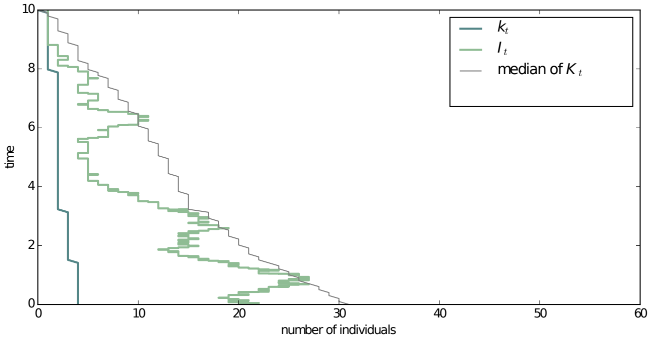




Reconstructing past population size

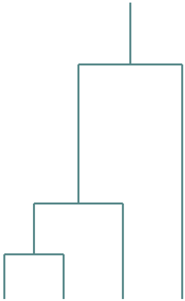
- With only ρ-sampling $(\psi, \omega = 0, 0)$.
- With ρ and ψ-sampling $(\omega = 0)$.
- With ρ, ψ, and ω-sampling.

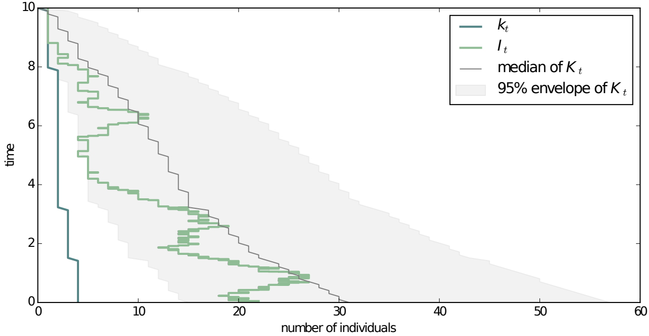




Reconstructing past population size

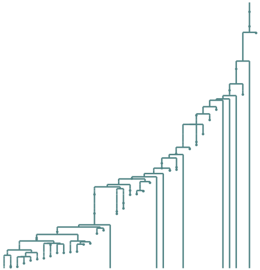
- With only ρ-sampling $(\psi, \omega = 0, 0)$.
- With ρ and ψ-sampling $(\omega = 0)$.
- With ρ, ψ, and ω-sampling.

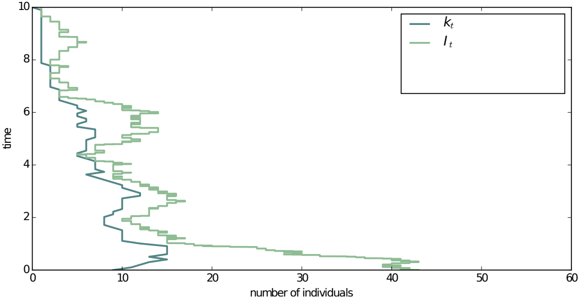




Reconstructing past population size

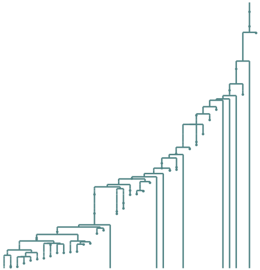
- With only ρ-sampling ($\psi, \omega = 0, 0$).
- With ρ and ψ-sampling ($\omega = 0$).
- With ρ, ψ, and ω-sampling.

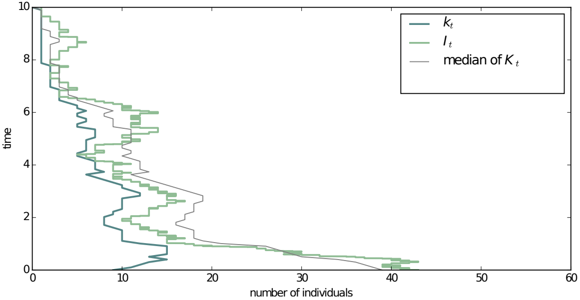




Reconstructing past population size

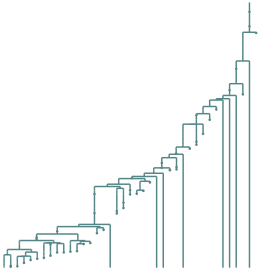
- With only ρ-sampling ($\psi, \omega = 0, 0$).
- With ρ and ψ-sampling ($\omega = 0$).
- With ρ, ψ, and ω-sampling.

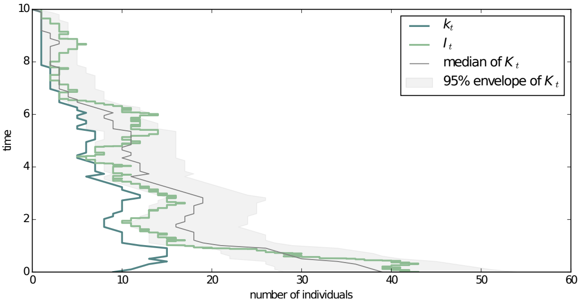




Reconstructing past population size

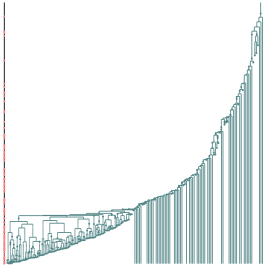
- With only ρ-sampling ($\psi, \omega = 0, 0$).
- With ρ and ψ-sampling ($\omega = 0$).
- With ρ, ψ, and ω-sampling.

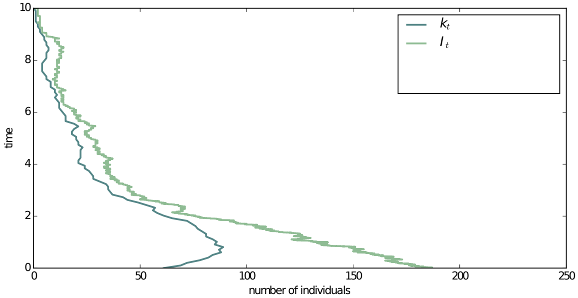




Reconstructing past population size

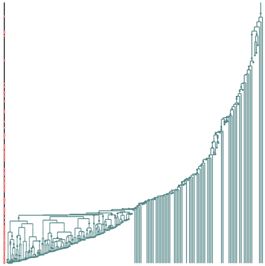
- With only ρ-sampling $(\psi, \omega = 0, 0)$.
- With ρ and ψ-sampling $(\omega = 0)$.
- With ρ, ψ, and ω-sampling.

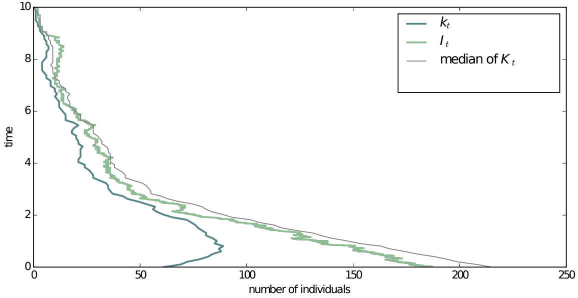




Reconstructing past population size

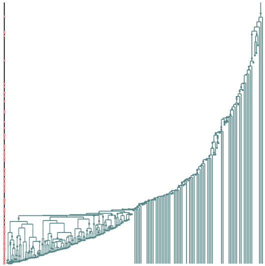
- With only ρ-sampling $(\psi, \omega = 0, 0)$.
- With ρ and ψ-sampling $(\omega = 0)$.
- With ρ, ψ, and ω-sampling.

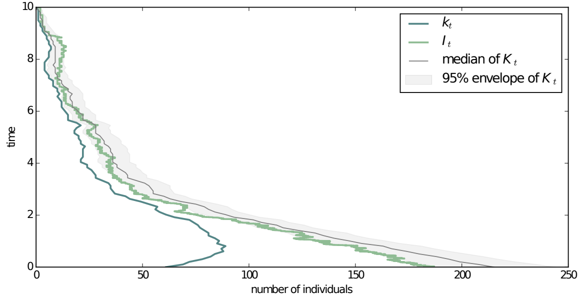




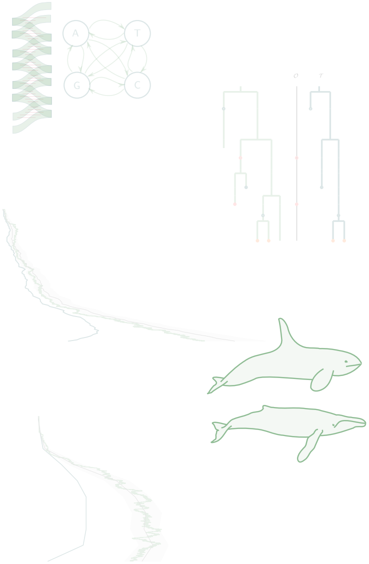
Reconstructing past population size

- With only ρ-sampling $(\psi, \omega = 0, 0)$.
- With ρ and ψ-sampling $(\omega = 0)$.
- With ρ, ψ, and ω-sampling.





Empirical case studies



Overview of the project
With Antoine Zwaans and Jérémy Andréoletti

Goals:

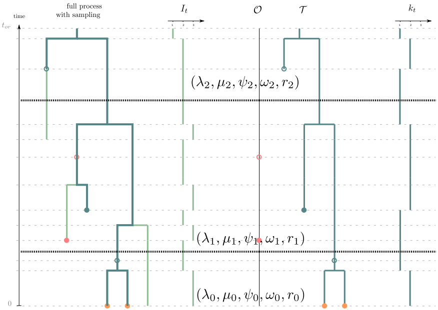
1. Work with piecewise-constant parameters.
2. Implement the method to compute $\mathbb{P}(\mathcal{T}, \mathcal{O})$.
3. Propose an easy post-analysis computation of K_t.
4. Illustrate the approach on empirical datasets.

Overview of the project

With Antoine Zwaans and Jérémy Andréoletti

Goals:

1. Work with piecewise-constant parameters.
2. Implement the method to compute $\mathbb{P}(\mathcal{T}, \mathcal{O})$.
3. Propose an easy post-analysis computation of K_t.
4. Illustrate the approach on empirical datasets.

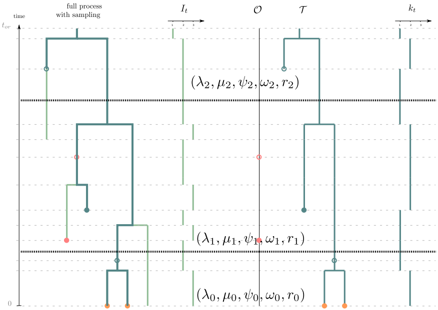


Overview of the project

With Antoine Zwaans and Jérémy Andréoletti

Goals:

1. Work with piecewise-constant parameters.
2. Implement the method to compute $\mathbb{P}(\mathcal{T}, \mathcal{O})$.
3. Propose an easy post-analysis computation of K_t.
4. Illustrate the approach on empirical datasets.

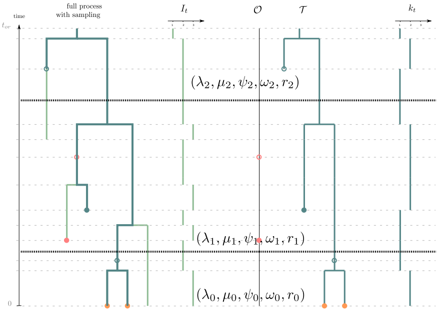


Overview of the project

With Antoine Zwaans and Jérémy Andréoletti

Goals:

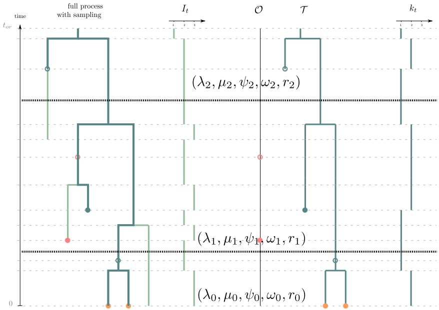
1. Work with piecewise-constant parameters.
2. Implement the method to compute $\mathbb{P}(\mathcal{T}, \mathcal{O})$.
3. Propose an easy post-analysis computation of K_t.
4. Illustrate the approach on empirical datasets.



Overview of the project
With Antoine Zwaans and Jérémy Andréoletti

Goals:

1. Work with piecewise-constant parameters.
2. Implement the method to compute $\mathbb{P}(\mathcal{T}, \mathcal{O})$.
3. Propose an easy post-analysis computation of K_t.
4. Illustrate the approach on empirical datasets.

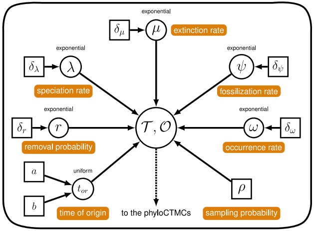


Implementation

- within the phylogenetic software revBayes,
- modular design based on graphical models,
- use to sample the Bayesian posterior.

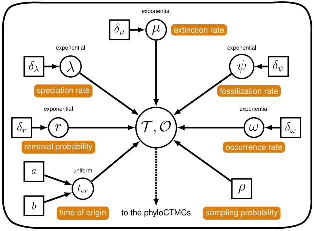
Implementation

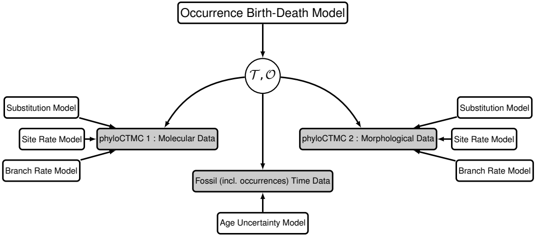
- within the phylogenetic software revBayes,
- modular design based on graphical models,
- use to sample the Bayesian posterior.



Implementation

- within the phylogenetic software revBayes,
- modular design based on graphical models,
- use to sample the Bayesian posterior.

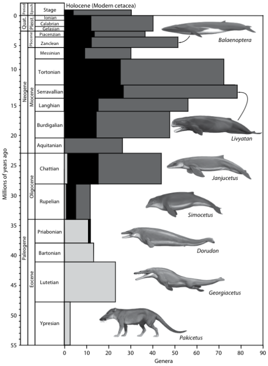




Cetacean diversity

After Marx et al. (2016) and the Paleobiology database

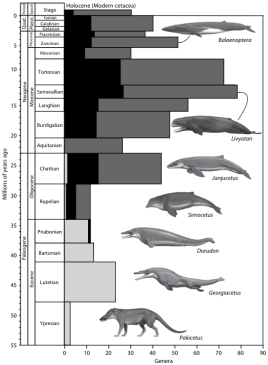
- Generic diversity
- Bias 1: Uneven sampling of time periods/localities,
- Bias 2: Species abundances,

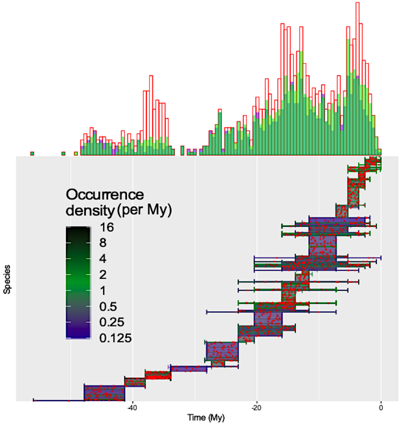


Cetacean diversity

After Marx et al. (2016) and the Paleobiology database

- Generic diversity
- Bias 1: Uneven sampling of time periods/localities,
- Bias 2: Species abundances,

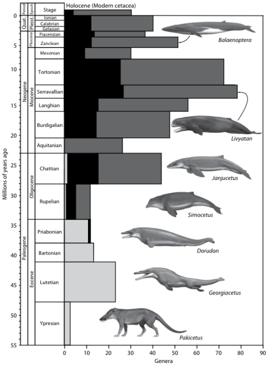




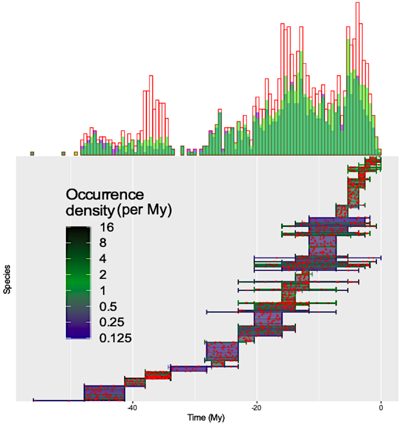
Cetacean diversity

After Marx et al. (2016) and the Paleobiology database

- Generic diversity



- Bias 1: Uneven sampling of time periods/localities,
- Bias 2: Species abundances,

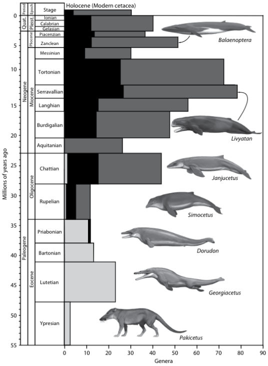


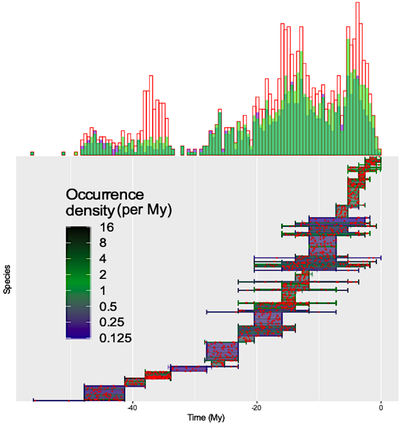
Cetacean diversity

After Marx et al. (2016) and the Paleobiology database

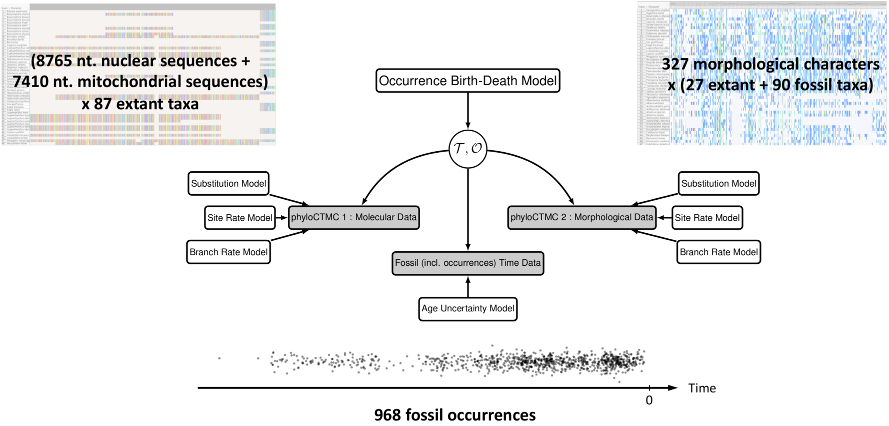
- Generic diversity

- Bias 1: Uneven sampling of time periods/localities,
- Bias 2: Species abundances,

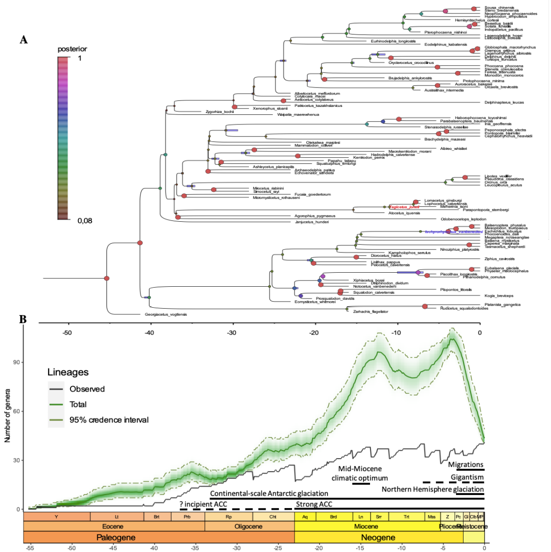




Cetacean diversity



Cetacean diversity
Preliminary results



Covid-19 prevalence on the Diamond princess

- Diamond princess cruise ship,
- Very close to the model assumptions,
- With rates varying at known time points.

Can we recover the known prevalence ?

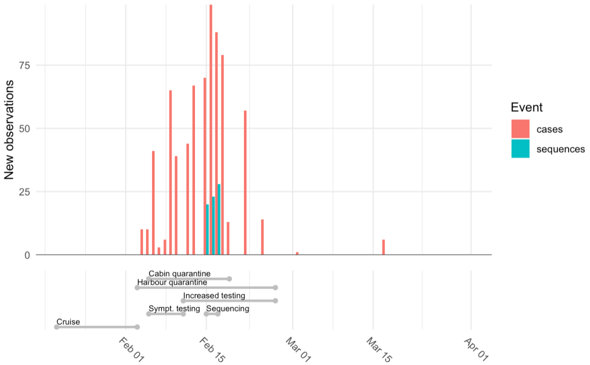
Covid-19 prevalence on the Diamond princess

- Diamond princess cruise ship,
- Very close to the model assumptions,
- With rates varying at known time points.

Can we recover the known prevalence ?

Covid-19 prevalence on the Diamond princess

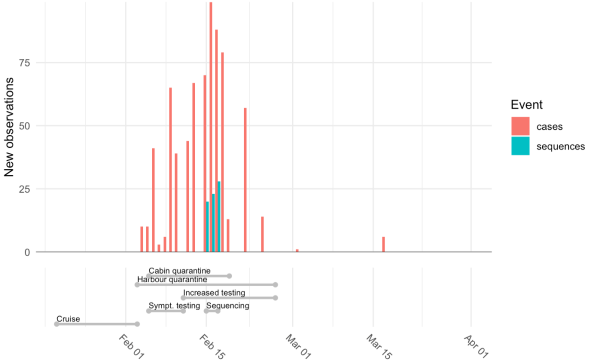
- Diamond princess cruise ship,
- Very close to the model assumptions,
- With rates varying at known time points.



Can we recover the known prevalence ?

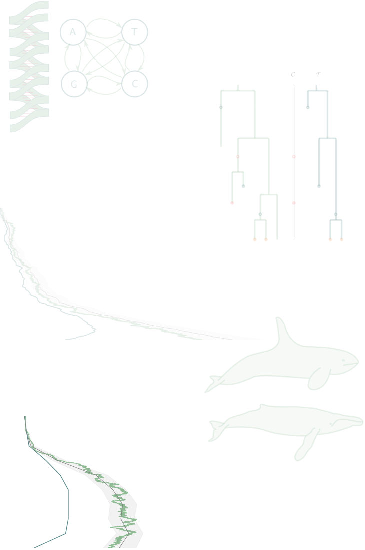
Covid-19 prevalence on the Diamond princess

- Diamond princess cruise ship,
- Very close to the model assumptions,
- With rates varying at known time points.



Can we recover the known prevalence ?

Conclusion



Perspectives
Diversity-dependent diversification

Work in progress

- extension to logistic birth-death processes, with per-capita rates either:

$$\lambda_i = \lambda - \alpha i \quad \text{or} \quad \mu_i = \mu + \beta i$$

- design methods to test hypotheses regarding diversification scenarios,
- try to fit it to empirical data, either from epidemiology or macroevolution.

Perspectives

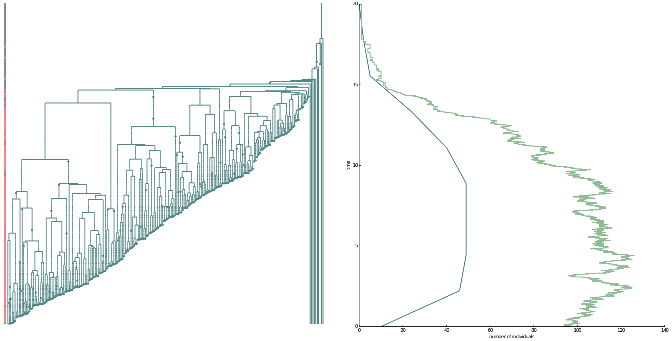
Diversity-dependent diversification

Work in progress

- extension to logistic birth-death processes, with per-capita rates either:

$$\lambda_i = \lambda - \alpha i \quad \text{or} \quad \mu_i = \mu + \beta i$$

- design methods to test hypotheses regarding diversification scenarios,
- try to fit it to empirical data, either from epidemiology or macroevolution.



Perspectives

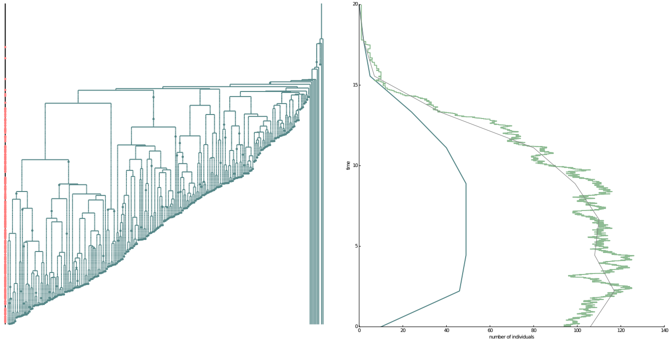
Diversity-dependent diversification

Work in progress

- extension to logistic birth-death processes, with per-capita rates either:

$$\lambda_i = \lambda - \alpha i \quad \text{or} \quad \mu_i = \mu + \beta i$$

- design methods to test hypotheses regarding diversification scenarios,
- try to fit it to empirical data, either from epidemiology or macroevolution.



Perspectives

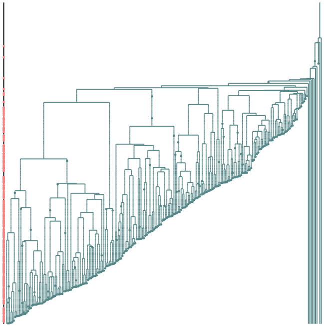
Diversity-dependent diversification

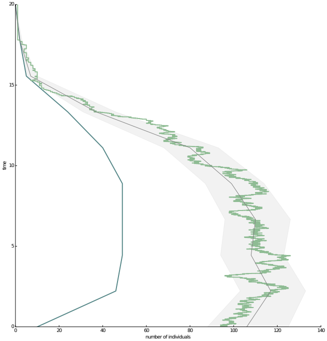
Work in progress

- extension to logistic birth-death processes, with per-capita rates either:

$$\lambda_i = \lambda - \alpha i \quad \text{or} \quad \mu_i = \mu + \beta i$$

- design methods to test hypotheses regarding diversification scenarios,
- try to fit it to empirical data, either from epidemiology or macroevolution.





Perspectives

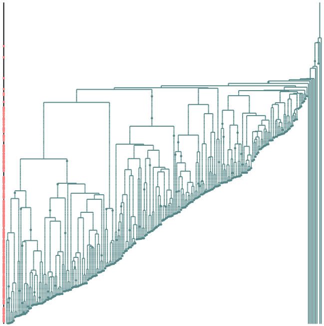
Diversity-dependent diversification

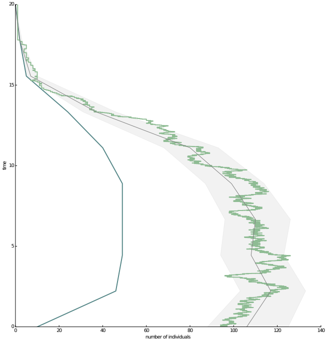
Work in progress

- extension to logistic birth-death processes, with per-capita rates either:

$$\lambda_i = \lambda - \alpha i \quad \text{or} \quad \mu_i = \mu + \beta i$$

- design methods to test hypotheses regarding diversification scenarios,
- try to fit it to empirical data, either from epidemiology or macroevolution.





Perspectives

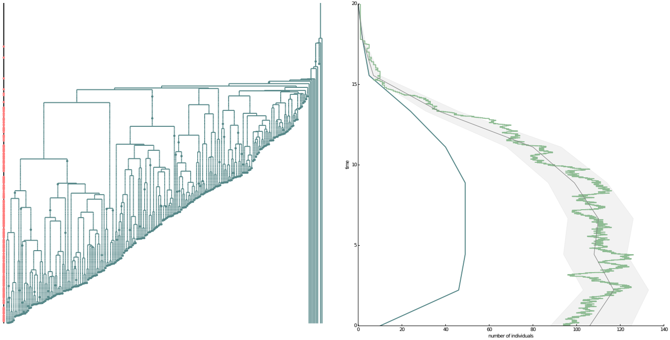
Diversity-dependent diversification

Work in progress

- extension to logistic birth-death processes, with per-capita rates either:

$$\lambda_i = \lambda - \alpha i \quad \text{or} \quad \mu_i = \mu + \beta i$$

- design methods to test hypotheses regarding diversification scenarios,
- try to fit it to empirical data, either from epidemiology or macroevolution.



Take-home messages

Model birth-death model with a specific sampling scheme through time.

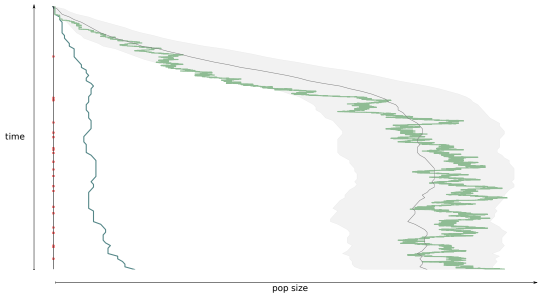
Method to get the likelihood of a tree and a record of occurrences, as well as $\mathbb{P}(I_t \mid \mathcal{O}, \mathcal{T})$.

Implementation with piecewise-constant parameters within the phylogenetic software revBayes.

Illustration on macroevolution and epidemiology datasets.

Perspectives e.g. for logistic density-dependence.

Thank you for your attention !



Take-home messages

Model birth-death model with a specific sampling scheme through time.

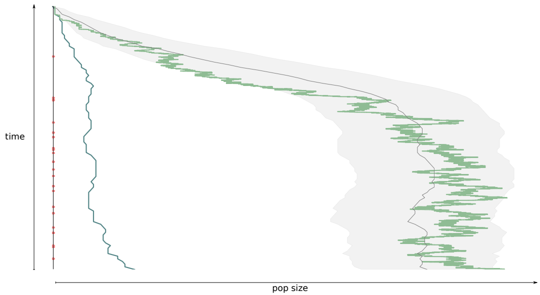
Method to get the likelihood of a tree and a record of occurrences, as well as $\mathbb{P}\left(I_t \mid \mathcal{O}, \mathcal{T}\right)$.

Implementation with piecewise-constant parameters within the phylogenetic software revBayes.

Illustration on macroevolution and epidemiology datasets.

Perspectives e.g. for logistic density-dependence.

Thank you for your attention !



Take-home messages

Model birth-death model with a specific sampling scheme through time.

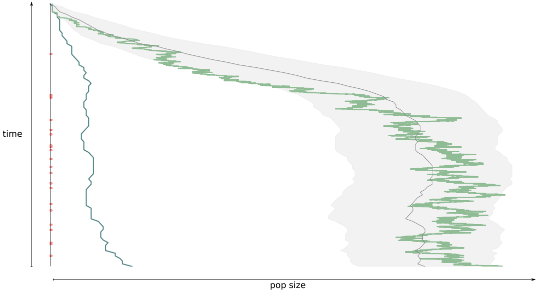
Method to get the likelihood of a tree and a record of occurrences, as well as $\mathbb{P}(I_t \mid \mathcal{O}, \mathcal{T})$.

Implementation with piecewise-constant parameters within the phylogenetic software revBayes.

Illustration on macroevolution and epidemiology datasets.

Perspectives e.g. for logistic density-dependence.

Thank you for your attention !



Take-home messages

Model birth-death model with a specific sampling scheme through time.

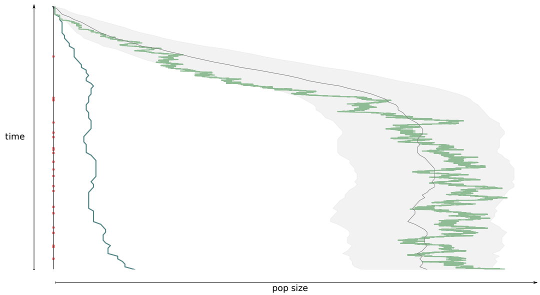
Method to get the likelihood of a tree and a record of occurrences, as well as $\mathbb{P}\left(I_t \mid \mathcal{O}, \mathcal{T}\right)$.

Implementation with piecewise-constant parameters within the phylogenetic software revBayes.

Illustration on macroevolution and epidemiology datasets.

Perspectives e.g. for logistic density-dependence.

Thank you for your attention !



Take-home messages

Model birth-death model with a specific sampling scheme through time.

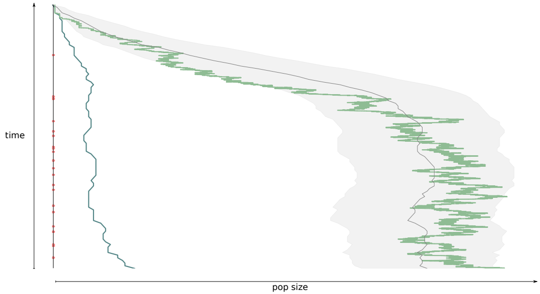
Method to get the likelihood of a tree and a record of occurrences, as well as $\mathbb{P}(I_t \mid \mathcal{O}, \mathcal{T})$.

Implementation with piecewise-constant parameters within the phylogenetic software revBayes.

Illustration on macroevolution and epidemiology datasets.

Perspectives e.g. for logistic density-dependence.

Thank you for your attention !



References

Etienne et al. (2012) used backward Kolmogorov equations to compute the likelihood of trees, under a logistic birth-death process.

Leventhal et al. (2013) used the forward Kolmogorov equations to compute the likelihood of trees, under a logistic birth-death process.

Vaughan et al. (2018) introduced the model and a Monte-Carlo method to get $\mathbb{P}\left(I_t \mid \mathcal{O}, \mathcal{T}\right)$.

Laudanno et al. (2019) did something similar to our analytical work on $\hat{M}$.

Gupta et al. (2020) analytical development to compute $\mathbb{P}(\mathcal{T}, \mathcal{O})$ when $r = 1$.

Manceau et al. (submitted) combining the forward and backward traversals to get the ancestral population size.

Andréoletti, Zwaans et al. (in prep) implementation in a Bayesian framework and application on empirical datasets.