Heterogeneity in evolutionary processes : structured birth-death processes
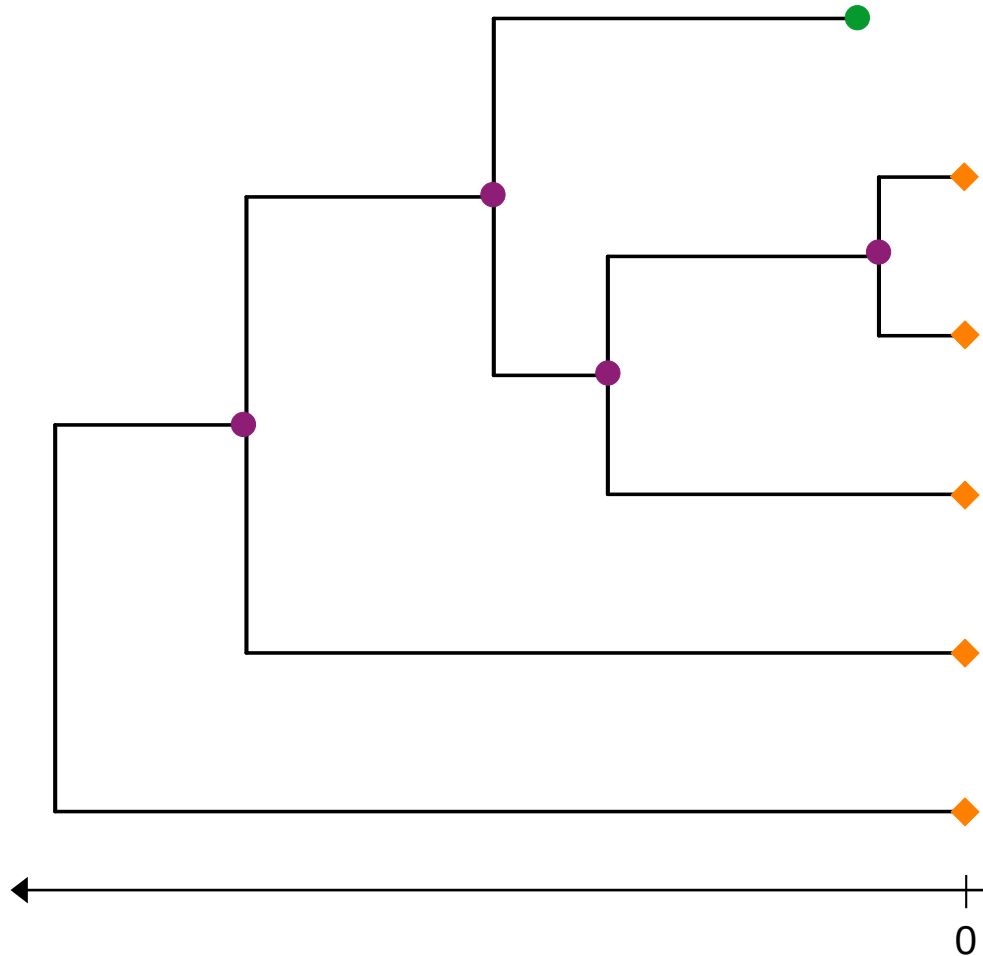
Joëlle Barido-Sottani

Simple birth-death process
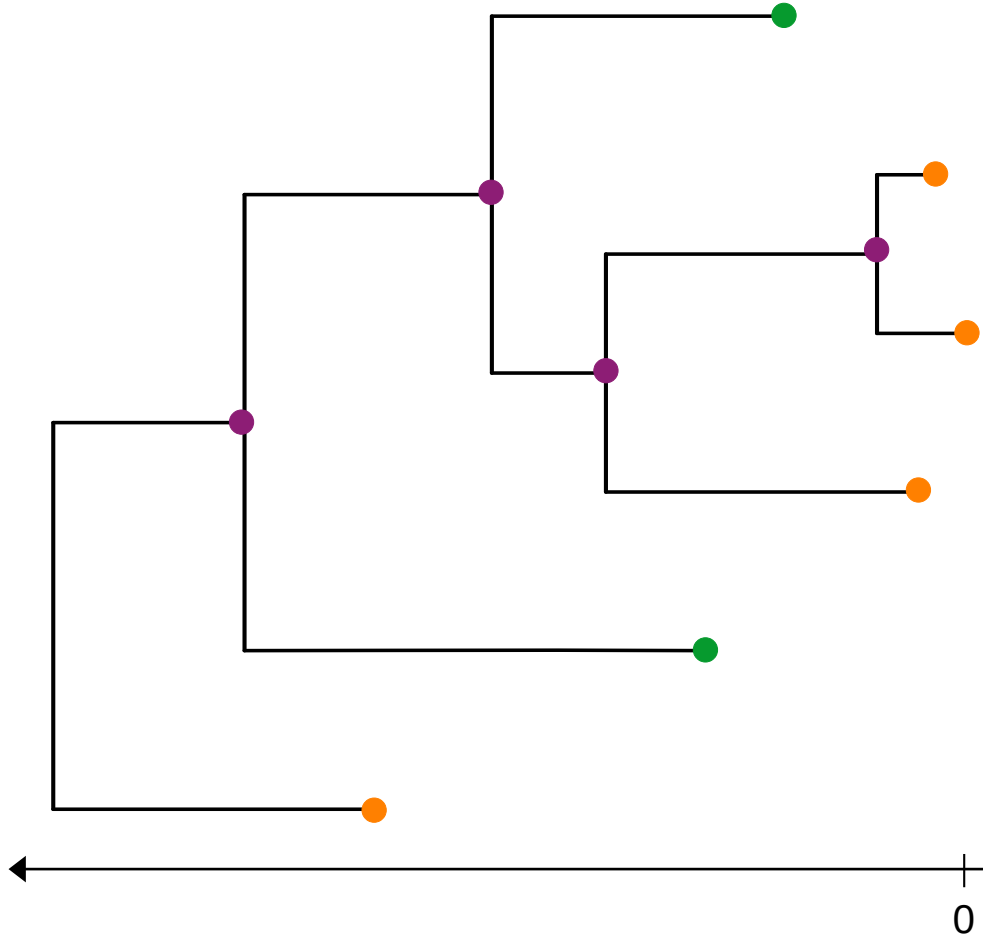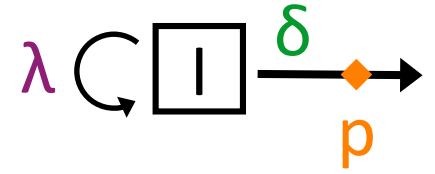

Parameters:

● λ — birth rate (= new lineage appearing)

● μ — death rate (= lineage disappearing)
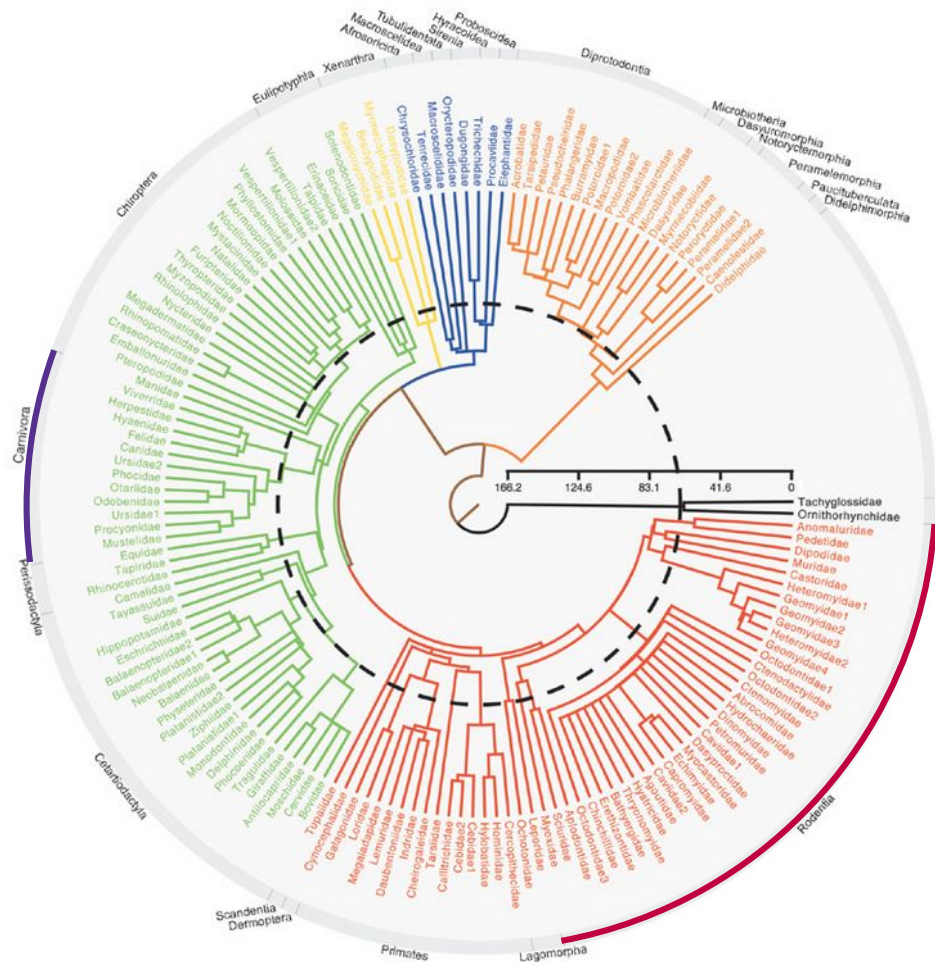
◆ ρ — extant species sampling probability

0

Birth-death for epidemiology

$\lambda \, \circlearrowleft \, \boxed{\text{I}} \xrightarrow{\delta} \quad$
p

Processes:

- ● λ — transmission rate

- ● $\mu = \delta(1-p)$ — rate of recovery without sampling
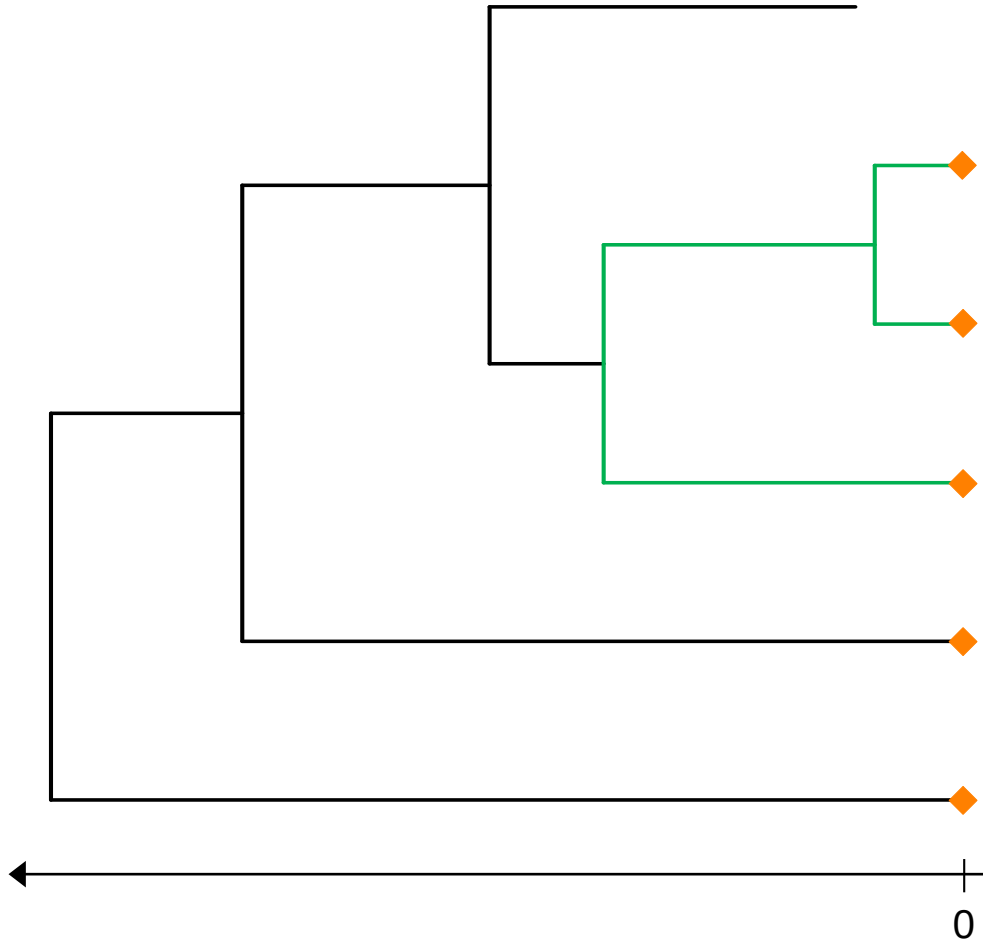
- ● $\psi = \delta p$ — rate of recovery with sampling

0

Heterogeneity in evolution

Bininda-Emonds et al. 2007, Nature

- Size discrepancies are evidence of variations in evolutionary processes

- Many traits are proposed to drive variation:
  - body size, mating system, environment, etc.
  - host location, pathogen strain, host behaviour, etc.
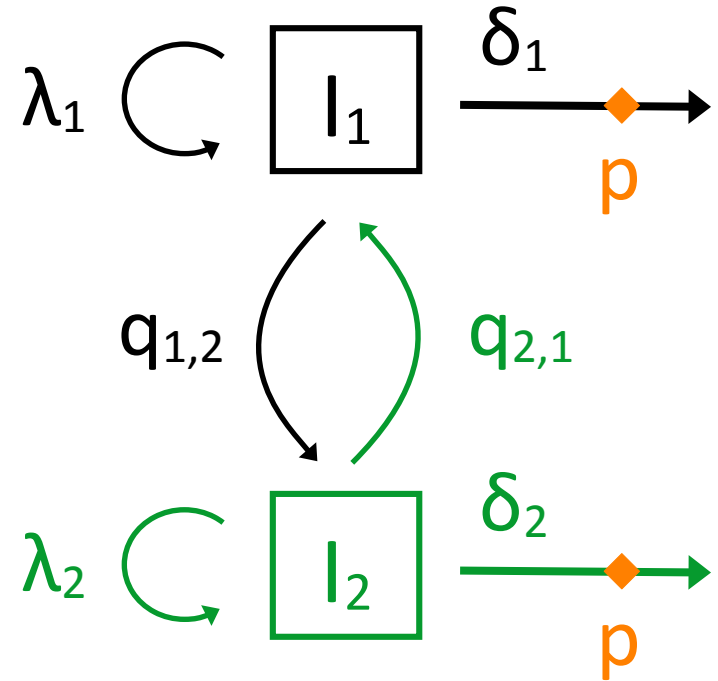
Multi-type birth-death (MTBD) process

2 types, type 1 & type 2

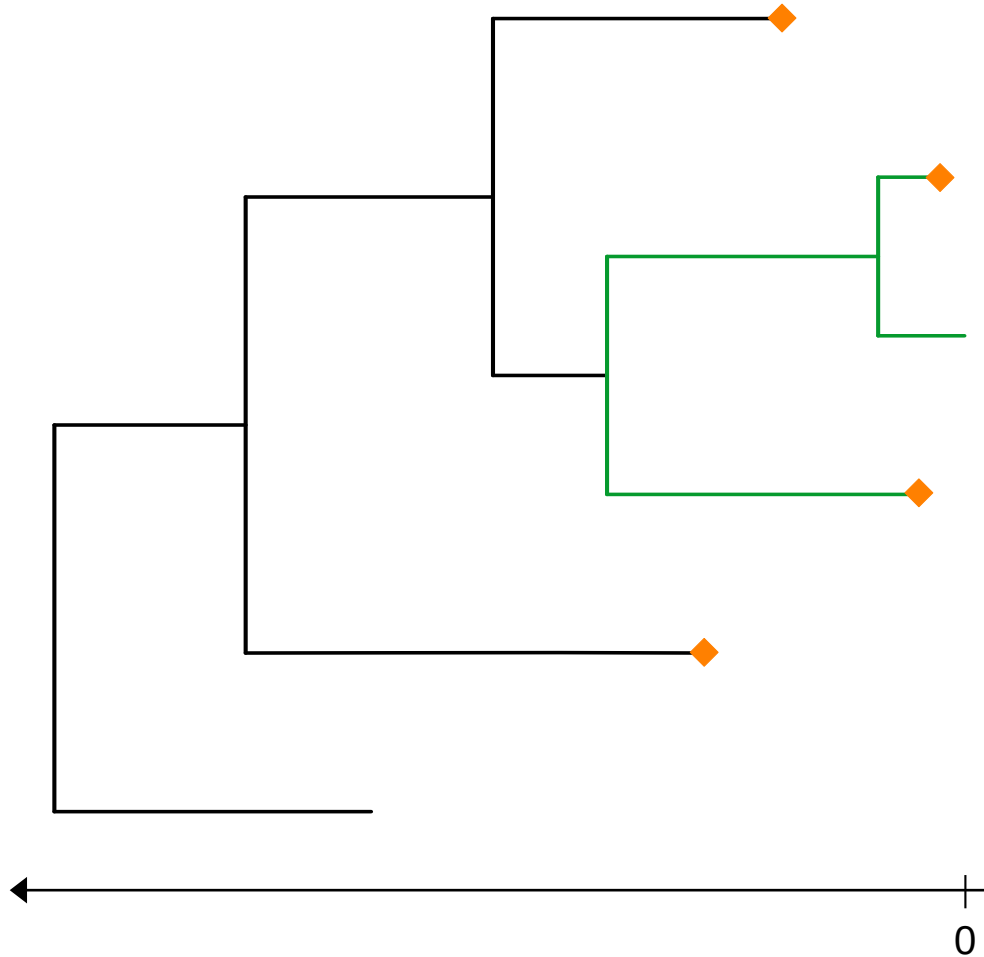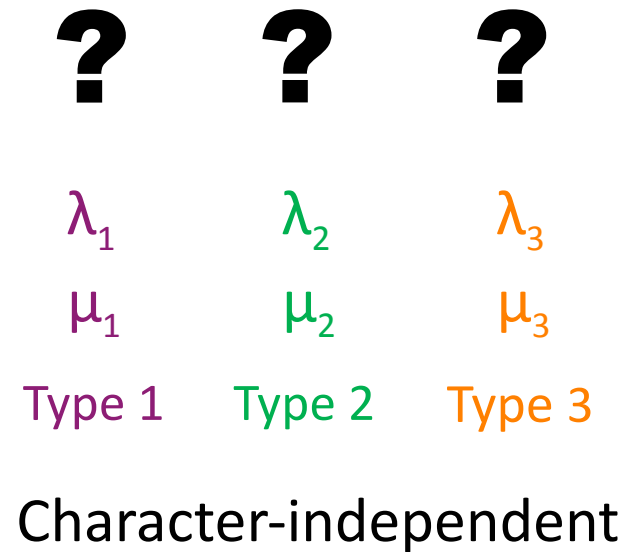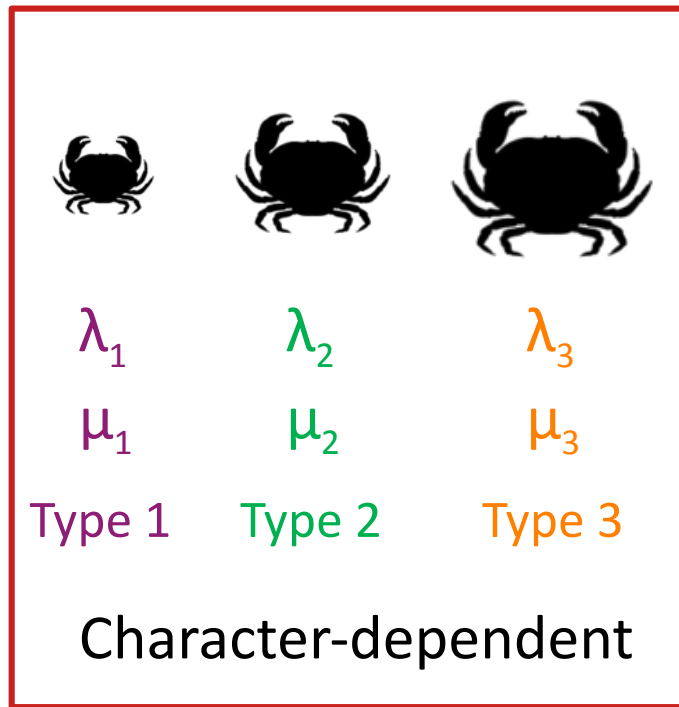λ_1 & λ_2 — birth rates

μ_1 & μ_2 — death rates

ρ — extant species
sampling probability

MTBD process (epidemiology)

Character-dependent or independent ?

λ₁ → λ_1 λ₂ → λ_2 λ₃ → λ_3

λ_1 λ_2 λ_3

μ_1 μ_2 μ_3

Type 1 Type 2 Type 3

Character-dependent

λ_1 λ_2 λ_3

μ_1 μ_2 μ_3

Type 1 Type 2 Type 3
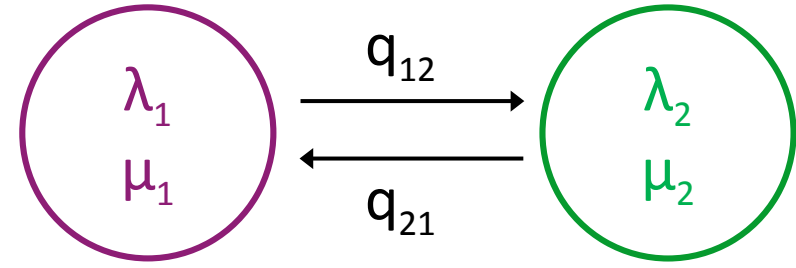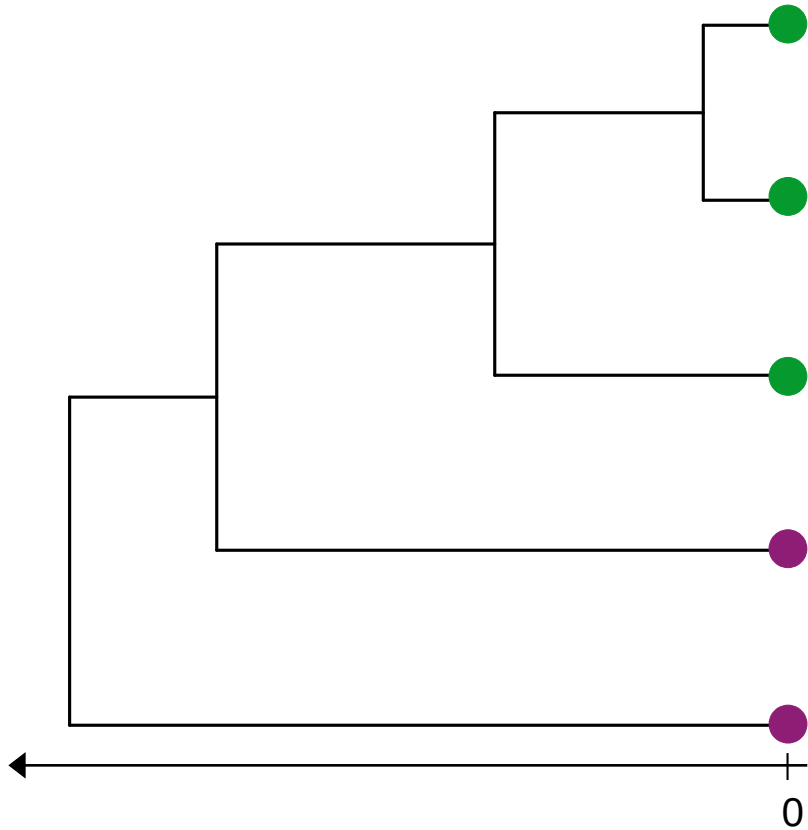
Character-independent

In a character-dependent model :
- The number of types is known
- The type at the tips is known

The BiSSE/MuSSE/BDMM model

Parameters of the model:

λ_i – birth rates
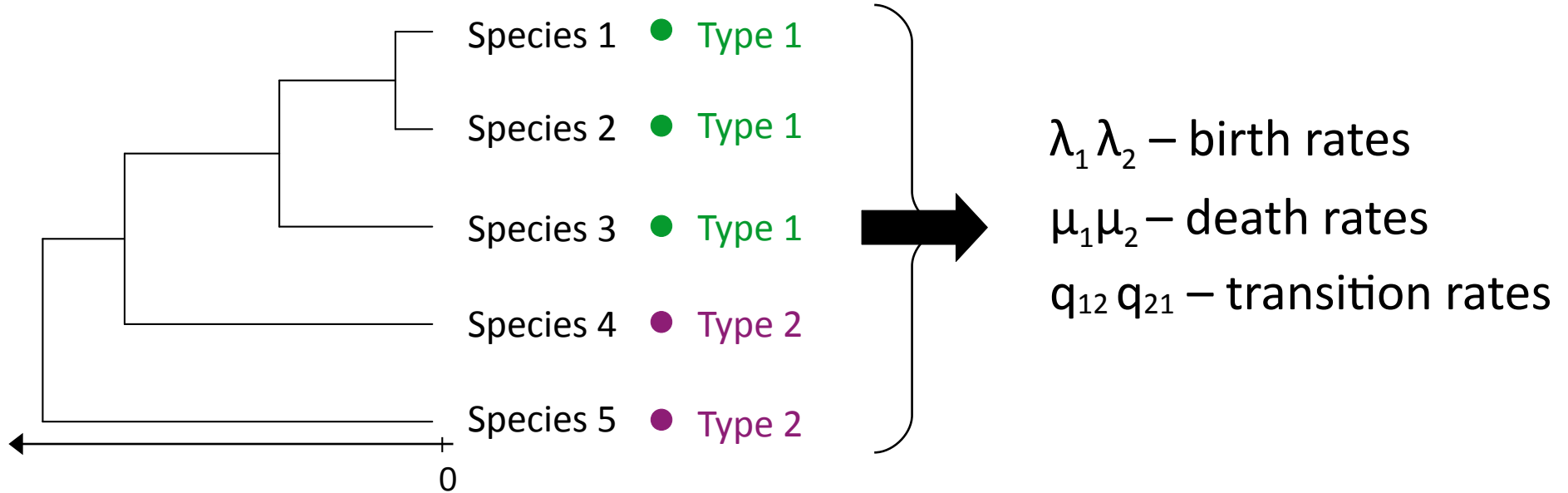
μ_i – death rates

q_{ij} – transition rates

ρ/p – sampling probability

Maddison et al. 2007, Sys. Bio.
Fitzjohn et al. 2009, Sys. Bio.
Kühnert et al. 2016 MBE

SSE/BDMM inference

$\lambda_1\,\lambda_2$ – birth rates

$\mu_1\mu_2$ – death rates
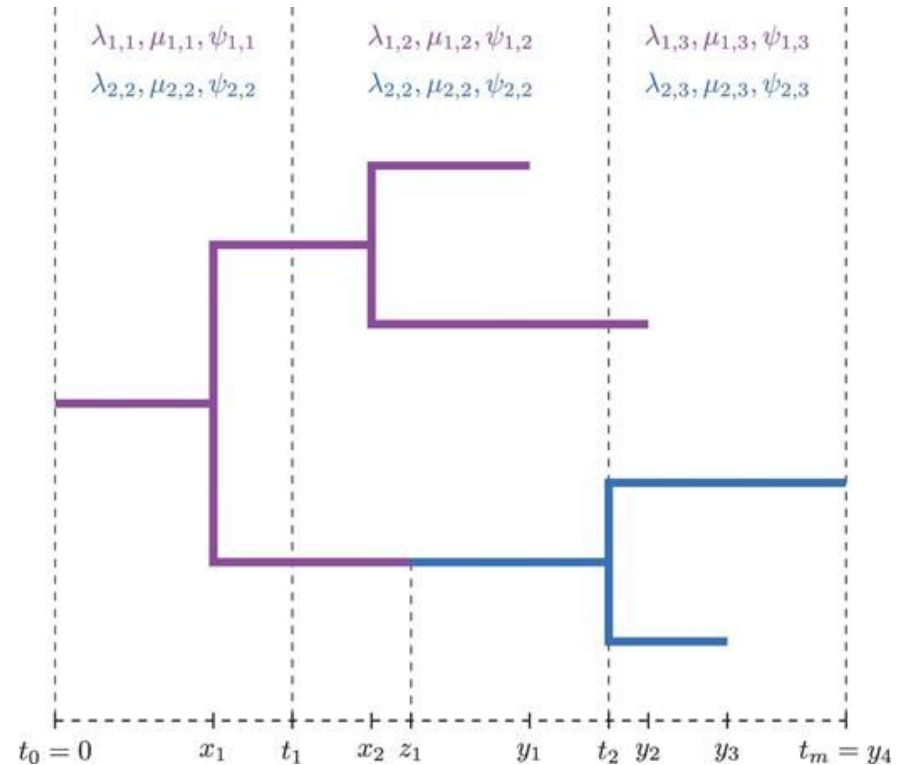
$q_{12}\,q_{21}$ – transition rates

Important assumption: the evolutionary processes in the complete phylogeny (including non-sampled parts) are identical to the processes in the reconstructed phylogeny.
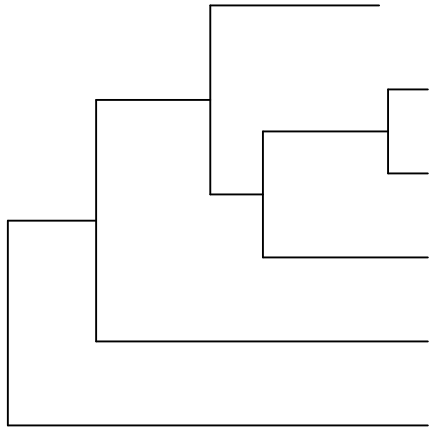
BDMM extensions

- Integration with the skyline model: piecewise-constant rates per type

- Sampling proportion per type: p_i
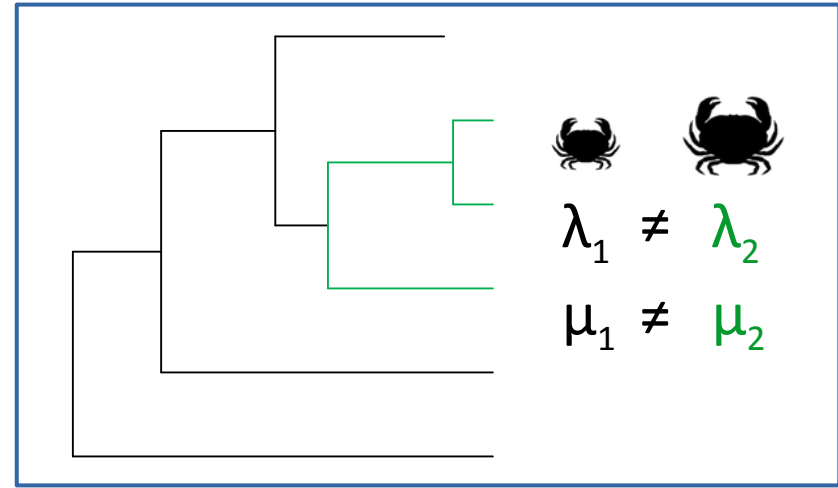
- Cross-type birth events: $\lambda_{i,j}$

Model selection issues
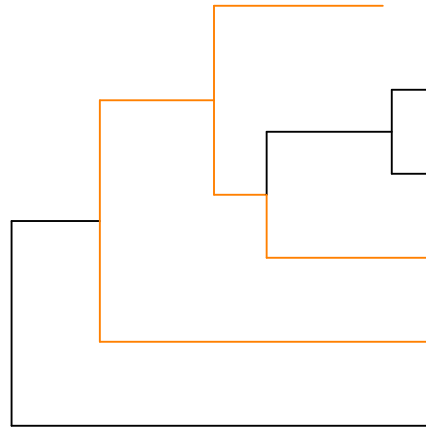

$\lambda_1 = \lambda_2$

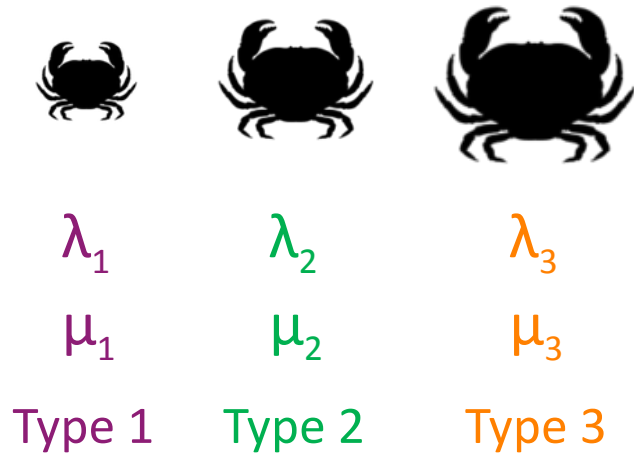$\mu_1 = \mu_2$

$\lambda_1 \neq \lambda_2$

$\mu_1 \neq \mu_2$

Rabosky & Goldberg 2015, Sys. Bio.

$\lambda_1 \neq \lambda_2$

$\mu_1 \neq \mu_2$

Character-dependent or independent ?

λ_1 λ_2 λ_3

μ_1 μ_2 μ_3

Type 1   Type 2   Type 3

Character-dependent

λ_1 λ_2 λ_3

μ_1 μ_2 μ_3
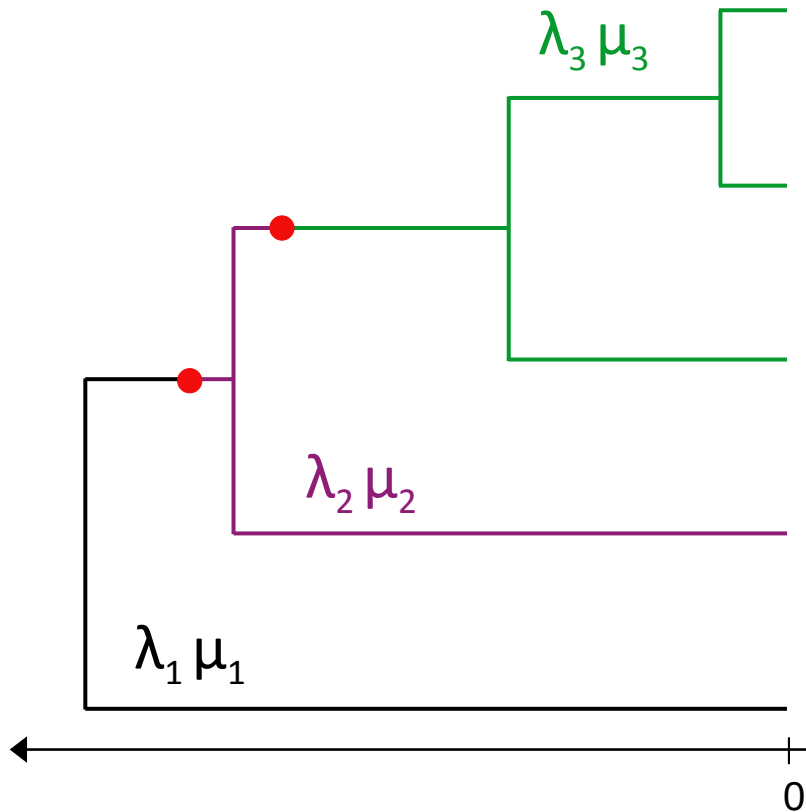
Type 1   Type 2   Type 3

Character-independent

In a character-dependent model :
• The number of types is known
• The type at the tips is known

BAMM/MSBD model

- Character-independent version of SSE/BDMM

- New estimated parameters:
  - N total number of types
  - Types of edges and tips

- Simplified transition process: ●
  => Constant transition rate γ

Rabosky et al. 2013, Nat. Comm.
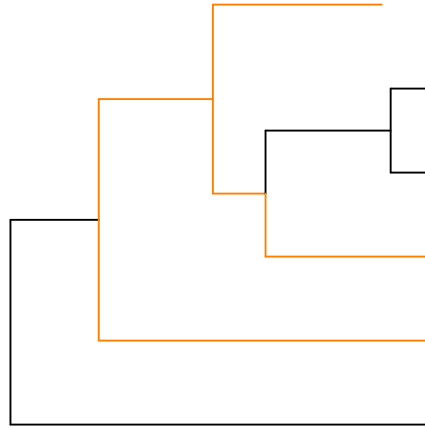Barido-Sottani et al. 2020, Sys. Bio.

Simplifying the model

Character-dependent

$\lambda_1 \approx \lambda_2 \quad \lambda_3$
$\mu_1 \approx \mu_2 \quad \mu_3$

$\lambda_1 \quad \lambda_2$
$\mu_1 \quad \mu_2$

Character-independent

$\lambda_1 \approx \lambda_2 \quad \lambda_3$
$\mu_1 \approx \mu_2 \quad \mu_3$

$\lambda_1 \quad \lambda_2$
$\mu_1 \quad \mu_2$
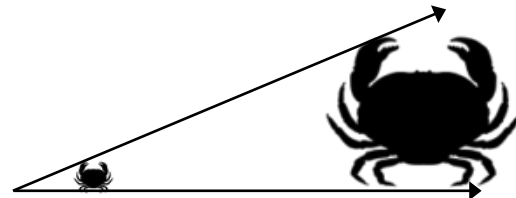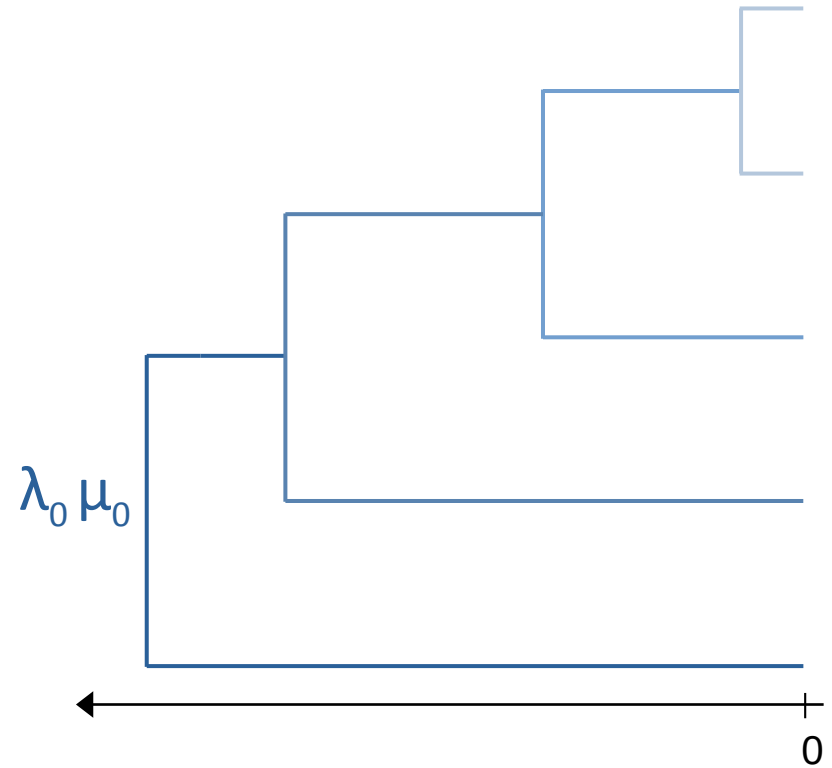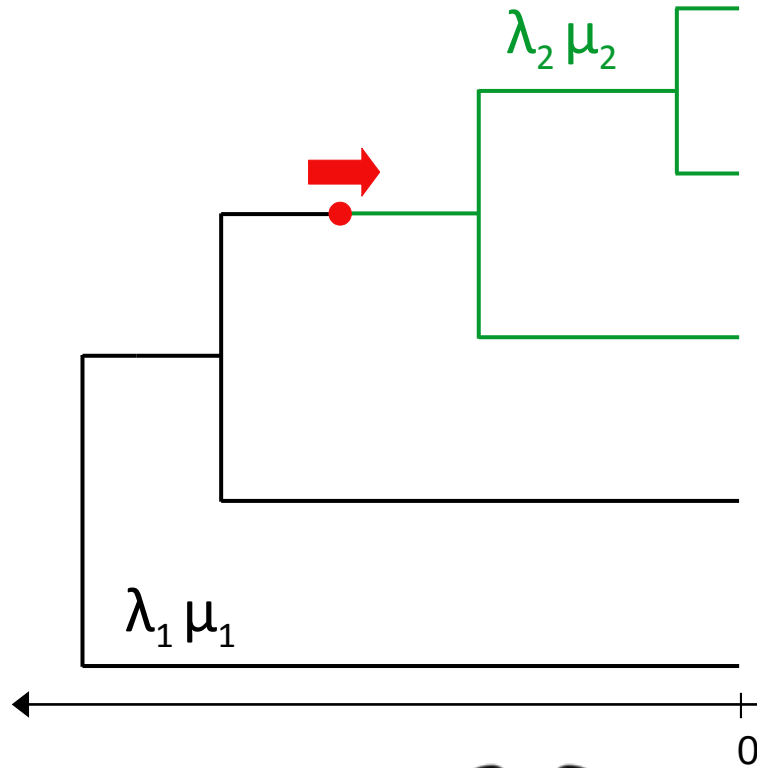
LSBDS/PESTO model

- Ordered types based on a Gamma distribution

- Fixed number of types N

- Simplified model: rates are not estimated, but determined by the shape of the Gamma distribution

Höhna et al. 2019, BiorXiv
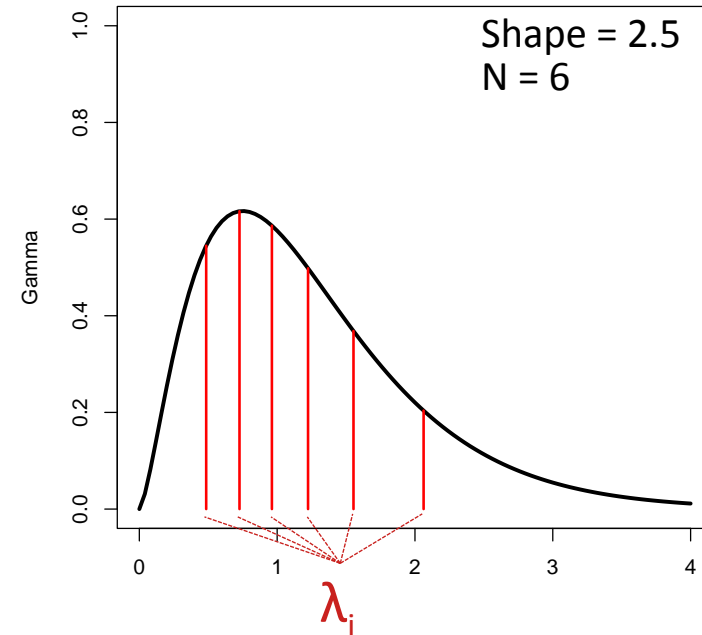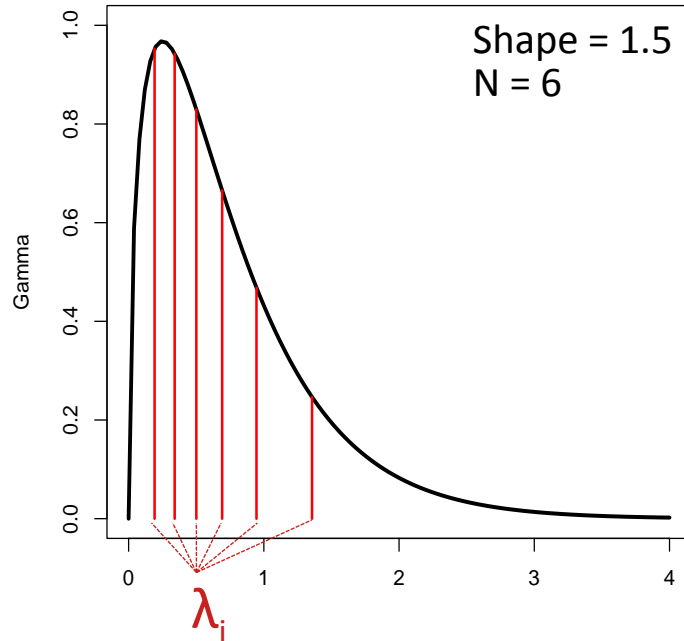
Going beyond types

ClaDS model
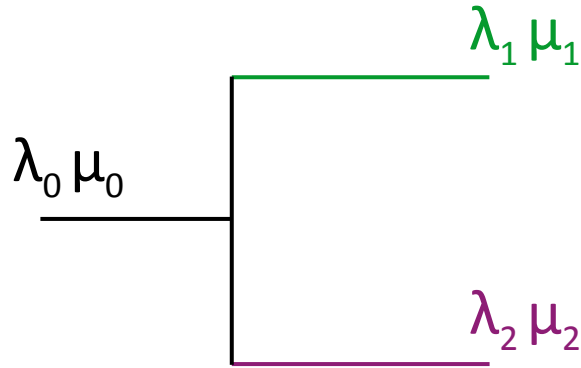
$\lambda_1 = \text{LogNormal}(\lambda_0 \times \alpha, \sigma)$

$\lambda_2 = \text{LogNormal}(\lambda_0 \times \alpha, \sigma)$

Maliet et al. 2019, Nat. Eco. Evo.
Maliet & Morlon 2021, Sys. Bio.
Barido-Sottani & Morlon 2023, Sys. Bio.

- Continuous evolution process driven by a lognormal distribution

  => similar in principle to an autocorrelated clock model

- New estimated parameters:
  - Initial rates at the root λ_0 and μ_0
  - Lognormal parameters α and σ
  - Birth rates for each edge λ_i

So – character-dependent or independent ?

Character-dependent

· Allow more complex transition processes

· Are more easily interpreted

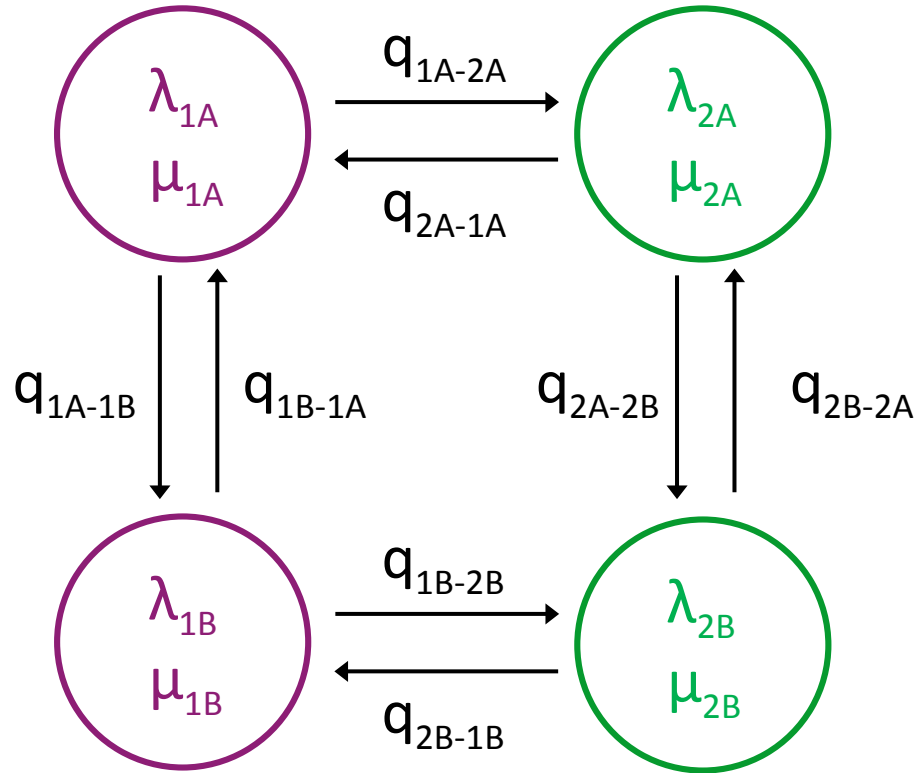· Are very dependent on the choice and accuracy of trait

Character-independent

· Usually have to make simplifying assumptions

· Do not give direct answers

· Are not constrained by trait information or hypothesis

What is your hypothesis ?
What are you trying to find out ?

The HiSSE model

Beaulieu & O'Meara 2016, Sys. Bio.

- Hybrid model with a hidden character (A/B) added to the user-chosen trait (1/2)

- Allows to distinguish whether the user-chosen character is linked to the rate variation

- Remaining issues:
  - Fixed number of states for the hidden character
  - Higher complexity of the model

Integration of fossil/SA data
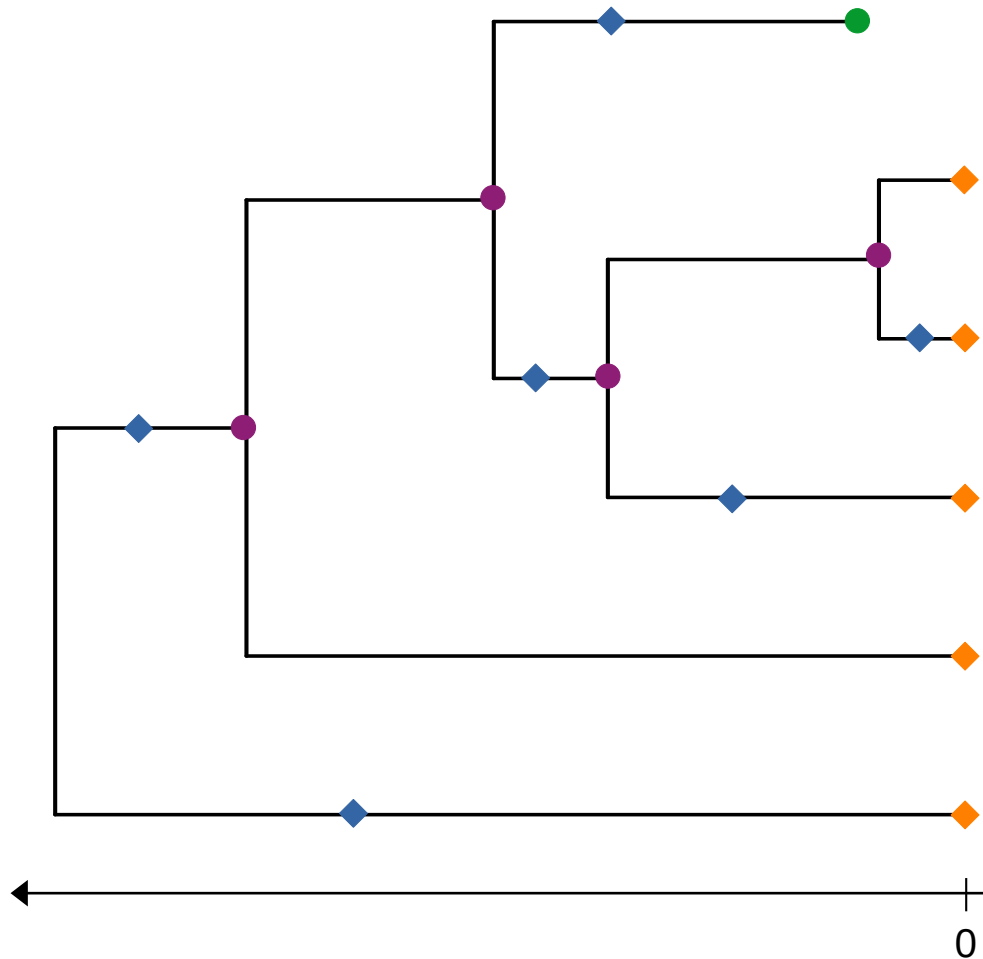
Parameters:

● λ — birth rate

● μ — death rate

◆ ψ — fossilization rate

◆ ρ — extant species sampling probability

Multi-type FBD/SA process

2 types, type 1 & type 2
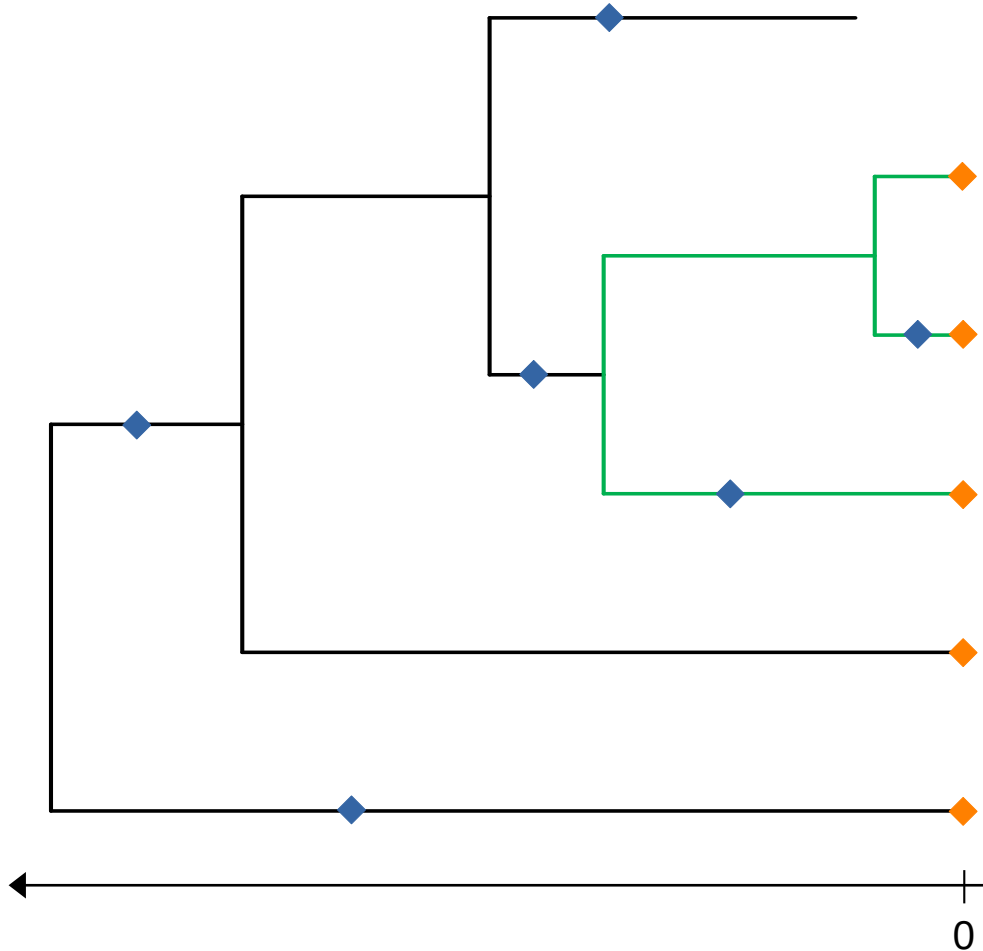
λ_1 & λ_2 — birth rates

μ_1 & μ_2 — death rates

ψ_1 & ψ_2 — fossilization rates

ρ — extant species
      sampling probability

0

In summary

- Multi-type birth-death processes come in two main categories:

  - Character-dependent: uses more information **but** subject to model selection issues

  - Character-independent: more powerful, more expensive and more difficult to interpret

- Multi-type birth-death models can be integrated with many other models (skyline, FBD, etc)

In summary (BEAST2)

- **Birth-Death-Migration Model (BDMM) package**
  - Character-dependent SSE
  - Includes time-dependent changes (skyline model)
  - Includes sampled ancestors

- **Multi-State Birth-Death (MSBD) package**
  - Character-independent SSE
  - Includes sampled ancestors, starting from v1.3.0

- **Cladogenetic Diversification rate Shift (ClaDS) package**
  - Progressive autocorrelated rate variations
  - Inclusion of sampled ancestors in development

Tutorial time

Character-dependent model (BDMM)
https://taming-the-beast.org/tutorials/Structured-birth-death-model/

Character-independent models (MSBD, ClaDS)
https://taming-the-beast.org/tutorials/MSBD-tutorial/
https://taming-the-beast.org/tutorials/ClaDS-tutorial/