

Heterogeneity in evolutionary processes : multi-type birth-death processes

Joëlle Barido-Sottani
IBENS, ENS Ulm, Paris (France)

Outline

- Introduction
- Multi-type birth-death processes
- Character-dependent processes
- Character-independent processes
- Use in practice and extensions

Tree models in Bayesian inference



$$P(\text{Molecular alignment, Substitution model, Clock model} \mid \text{Time tree, Tree model} \mid \text{ACAC... TCAC... ACAG...}) =$$

Posterior

Likelihood

Probability of the tree model

Priors

$$P(\text{ACAC... TCAC... ACAG...} \mid \text{Molecular alignment, Substitution model, Clock model, Time tree}) \cdot P(\text{Time tree} \mid \text{Tree model}) \cdot P(\text{Molecular alignment, Substitution model, Clock model} \mid \text{Tree model})$$

$$P(\text{ACAC... TCAC... ACAG...})$$

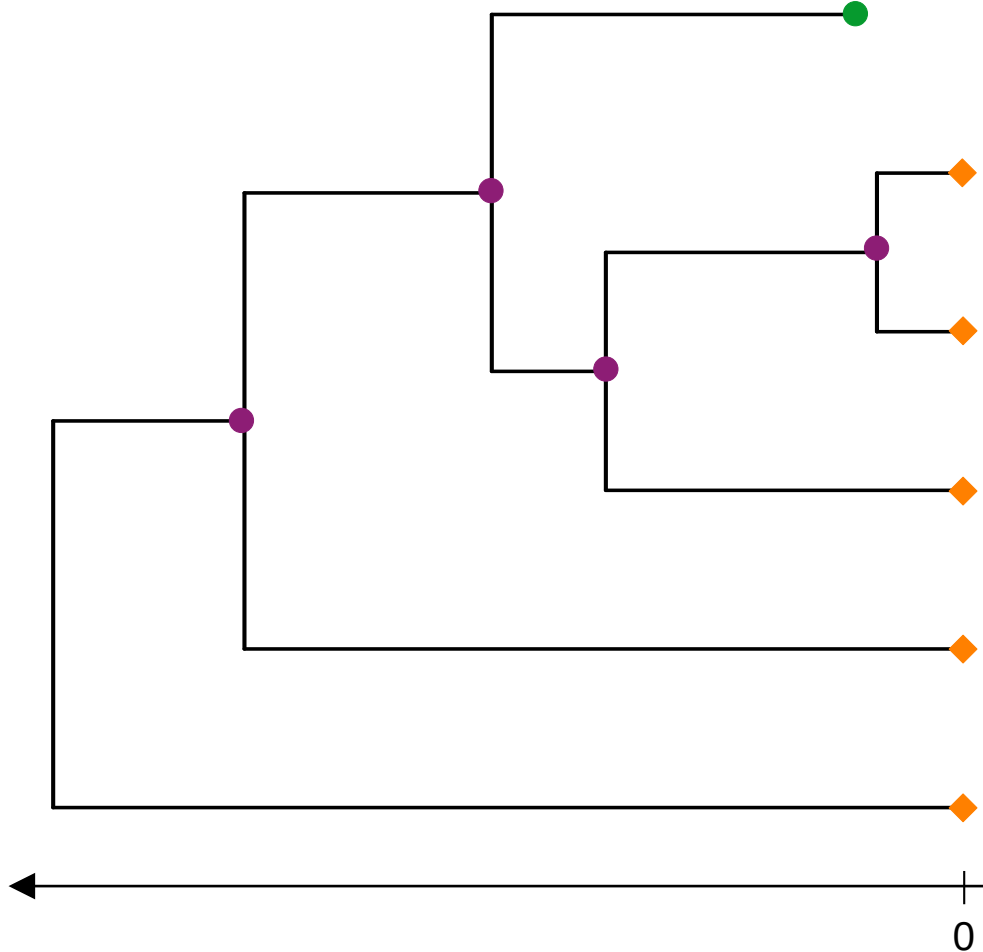
ACAC...
TCAC...
ACAG... Molecular alignment

Substitution model

Clock model

Time tree
Tree model

Simple birth-death process



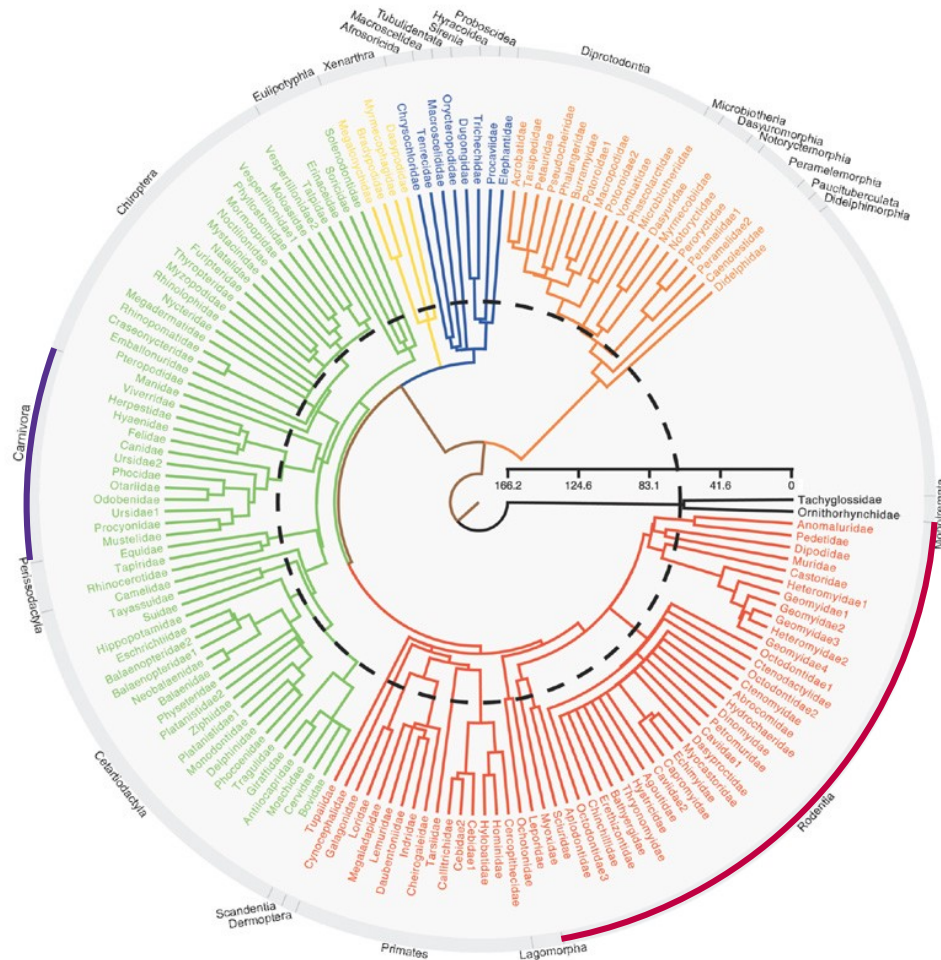
Parameters:

● λ — birth rate (= new lineage appearing)

● μ — death rate (= lineage disappearing)

◆ ρ — extant species sampling probability

Heterogeneity in evolution

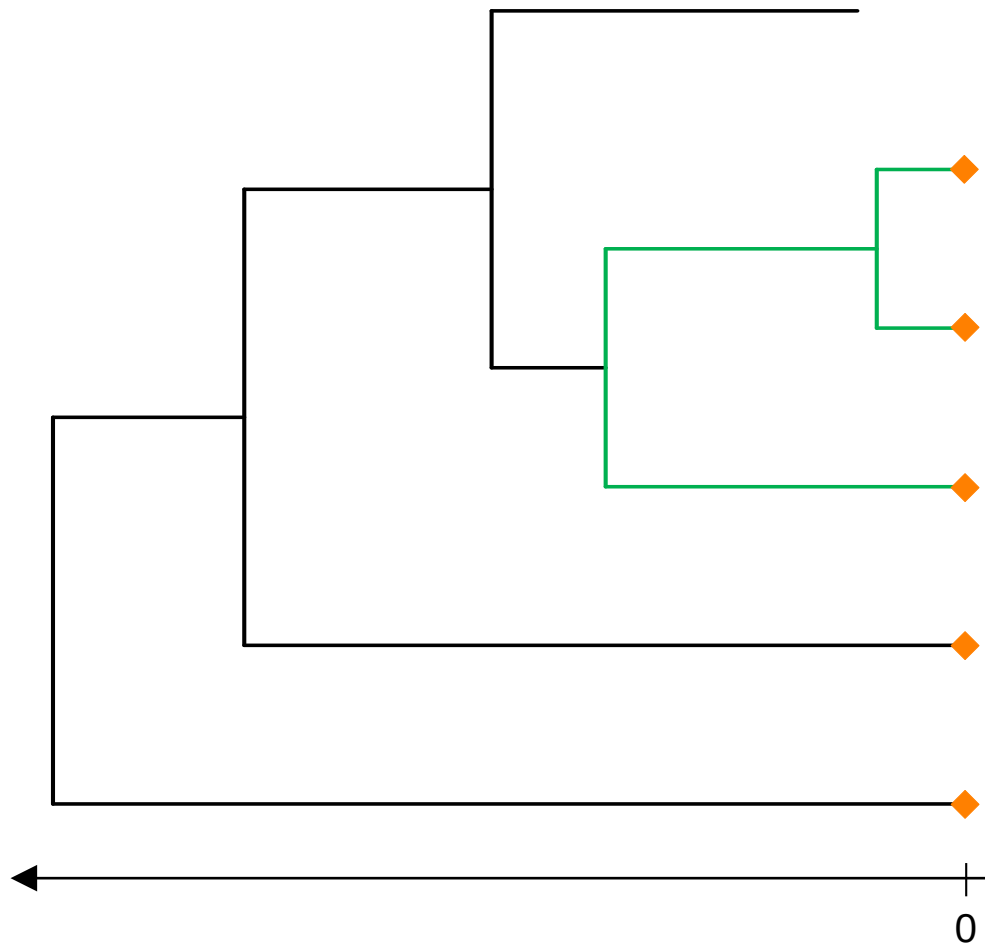


- Size discrepancies are evidence of variations in evolutionary processes
- Many traits are proposed to drive variation: for example body size, mating system, environment
- New models are needed to account for these variations

Outline

- Introduction
- Multi-type birth-death processes
- Character-dependent processes
- Character-independent processes
- Use in practice and extensions

Multi-type birth-death process



2 types, type 1 & **type 2**

λ_1 & λ_2 — birth rates

μ_1 & μ_2 — death rates

ρ — extant species
sampling probability

Character-dependent or independent ?



λ_1 λ_2 λ_3

μ_1 μ_2 μ_3

Character-dependent



λ_1 λ_2 λ_3

μ_1 μ_2 μ_3

Character-independent

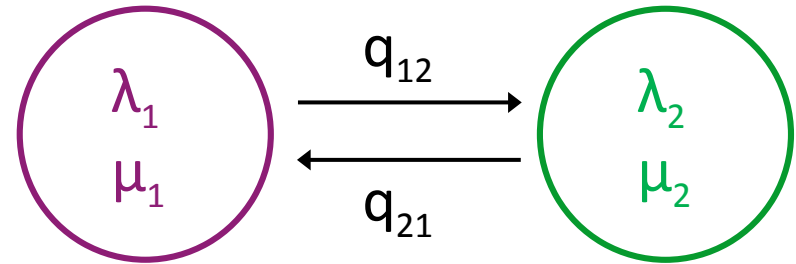
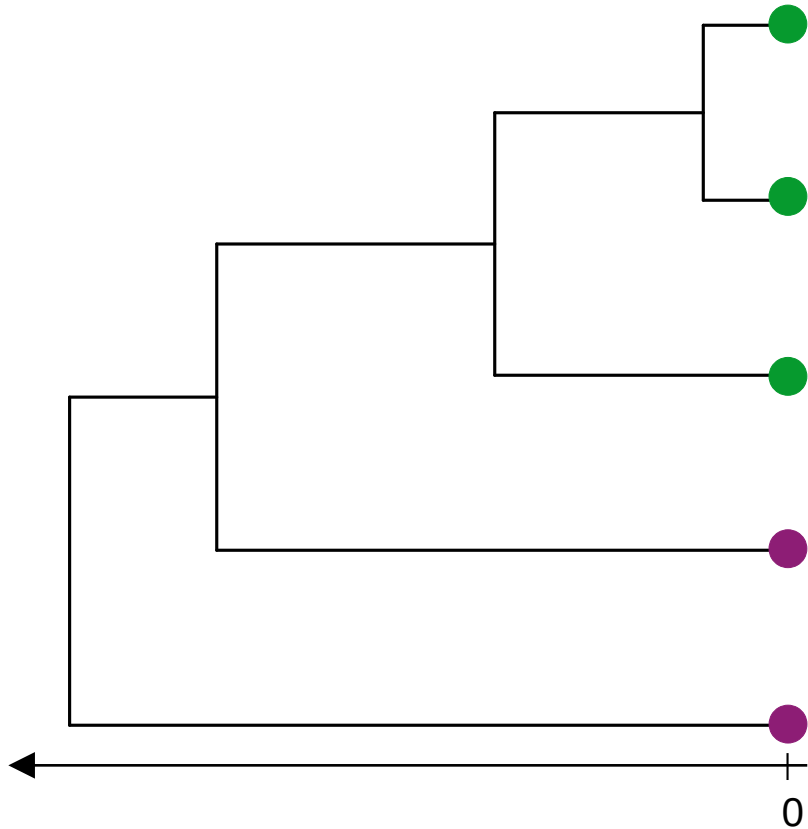
In a character-dependent model :

- The number of types is known
- The type at the tips is known

Outline

- Introduction
- Multi-type birth-death processes
- Character-dependent processes
- Character-independent processes
- Use in practice and extensions

The BiSSE model



Parameters of the model:

λ_i – birth rates

μ_i – death rates

q_{ij} – transition rates

ρ – extant species sampling probability

Maddison et al. 2007, *Sys. Bio.*

Fitzjohn et al. 2009, *Sys. Bio.*

Likelihood calculation

$$\frac{dE_i(t)}{dt} = - \left(\sum_{j \neq i} q_{ij} + \lambda_i + \mu_i \right) E_i(t) + \mu_i + \lambda_i E_i(t)^2 + \sum_{j \neq i} q_{ij} E_j(t)$$

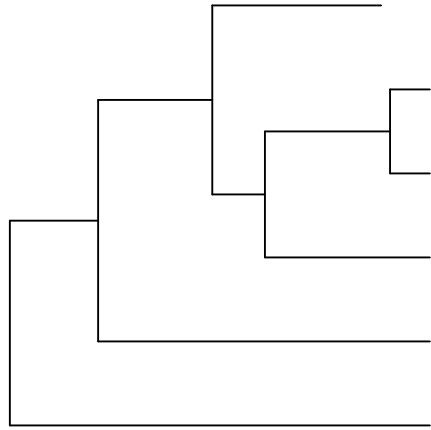
Probability of lineage extinction in type i at time t

$$\frac{dD_i(t)}{dt} = - \left(\sum_{j \neq i} q_{ij} + \lambda_i + \mu_i \right) D_i(t) + 2 \lambda_i E_i(t) D_i(t) + \sum_{j \neq i} q_{ij} D_j(t)$$

Likelihood of lineage in type i at time t

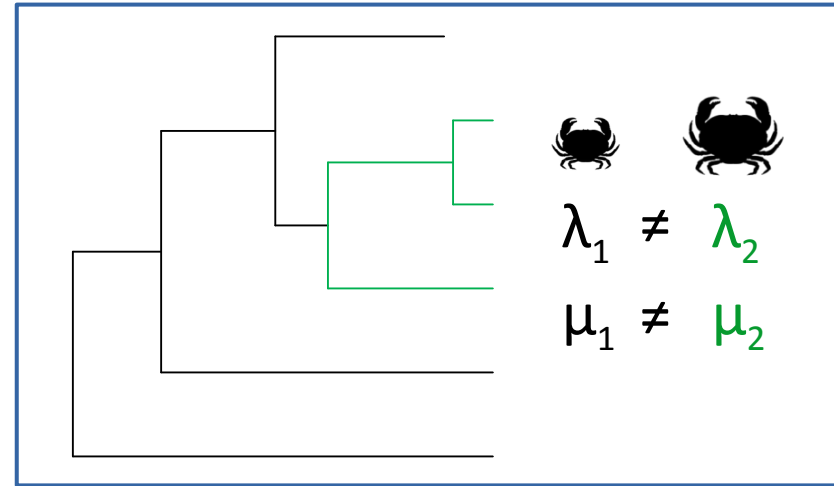
Important assumption: the evolutionary processes in the complete phylogeny (including non-sampled parts) are identical to the processes in the reconstructed phylogeny.

Model selection issues



$$\lambda_1 = \lambda_2$$

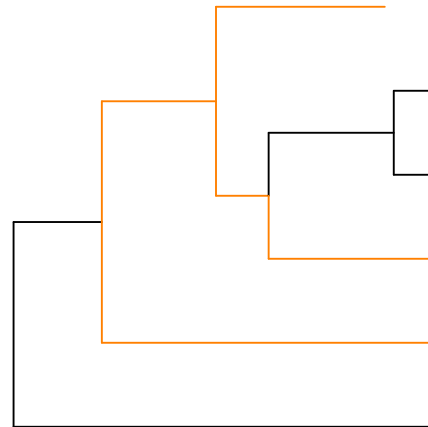
$$\mu_1 = \mu_2$$



$$\lambda_1 \neq \lambda_2$$

$$\mu_1 \neq \mu_2$$

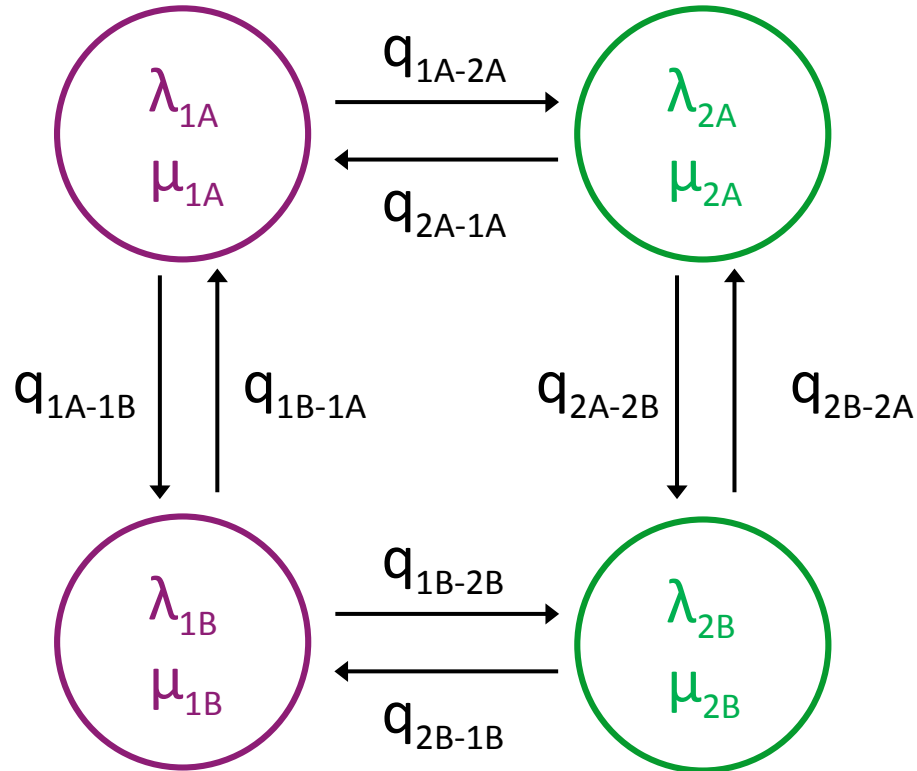
Rabosky & Goldberg 2015, Sys. Bio.



$$\lambda_1 \neq \lambda_2$$

$$\mu_1 \neq \mu_2$$

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
- Only single transitions are allowed (no diagonal)
- Remaining issues:
 - The number of values for the hidden character is chosen by the user
 - Higher complexity of the model

Examples

RESEARCH ARTICLE



Evidence linking life-form to a major shift in diversification rate in *Crassula*


Meng Lu^{1,2}  | Marc Fradera-Soler^{1,3}  | Félix Forest¹  |
Timothy G. Barraclough^{2,4}  | Olwen M. Grace¹ 

ORIGINAL ARTICLE

Ecological
Entomology



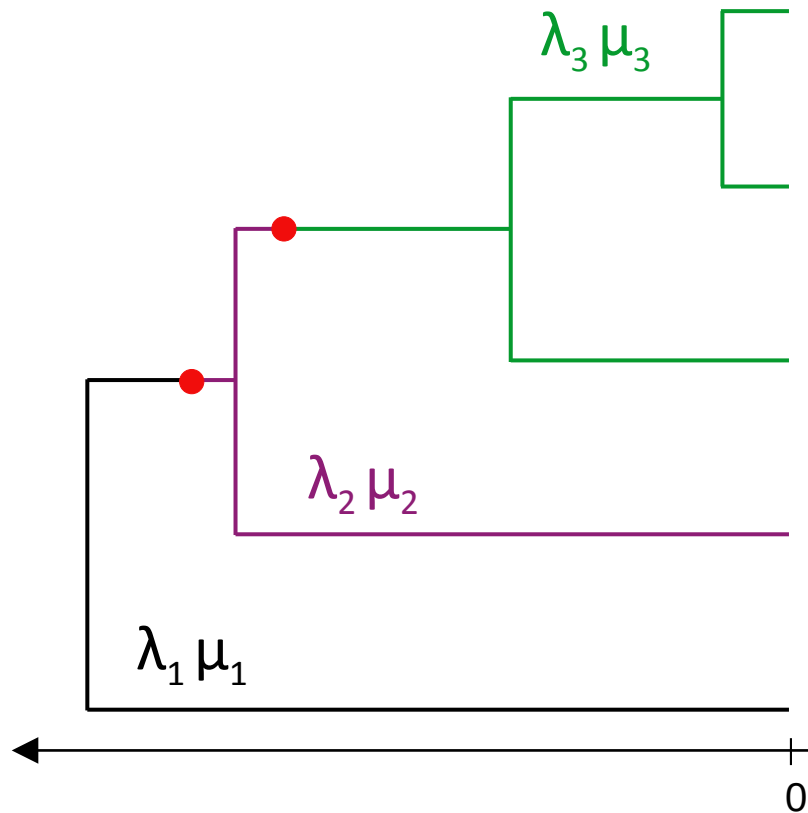
Fast species diversification among dragonflies (Anisoptera: Odonata: Insecta) inhabiting lentic environments regardless of wing pigmentation

Benjamín Padilla-Morales¹  | Paola Cornejo-Páramo¹ | Oscar García-Miranda² |
Aldo Issac Carrillo Muñoz² | Andrea Nieto López² | Daniel L. Castillo-Morales³ |
Gustavo Wappler Barragán² | Araxi O. Urrutia^{1,3} | Martín Alejandro Serrano-Meneses²

Outline

- Introduction
- Multi-type birth-death processes
- Character-dependent processes
- Character-independent processes
- Use in practice and extensions

BAMM/MSBD model



- Character-independent version of SSE
- New estimated parameters:
 - N total number of types
 - Types of edges and tips
- Simplified transition process: ●
 - Each transition is a new type (BAMM)
 - Constant transition rate γ (MSBD)
- Assumes that all types appear in the sampled tree – no unseen types

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

Simplifying the model

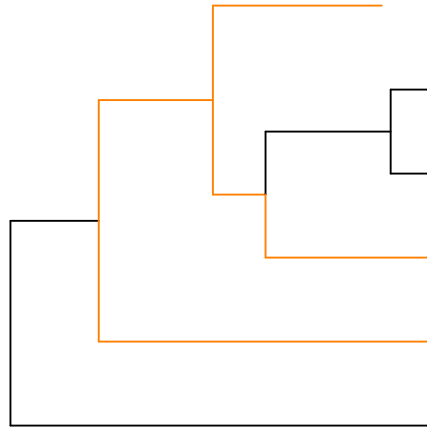
Character-dependent



$$\begin{array}{l} \lambda_1 \approx \lambda_2 \quad \lambda_3 \\ \mu_1 \approx \mu_2 \quad \mu_3 \end{array}$$



$$\begin{array}{l} \lambda_1 \quad \lambda_2 \\ \mu_1 \quad \mu_2 \end{array}$$



Character-independent



~~$$\begin{array}{l} \lambda_1 \approx \lambda_2 \quad \lambda_3 \\ \mu_1 \approx \mu_2 \quad \mu_3 \end{array}$$~~

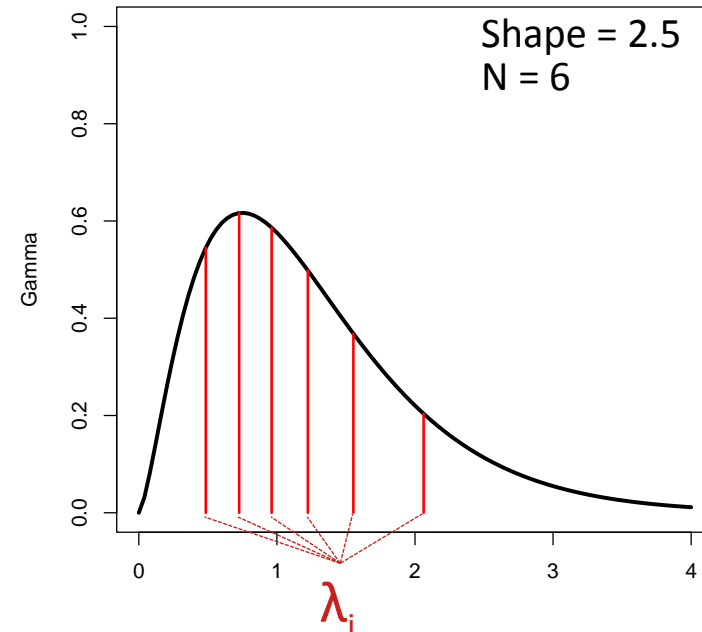
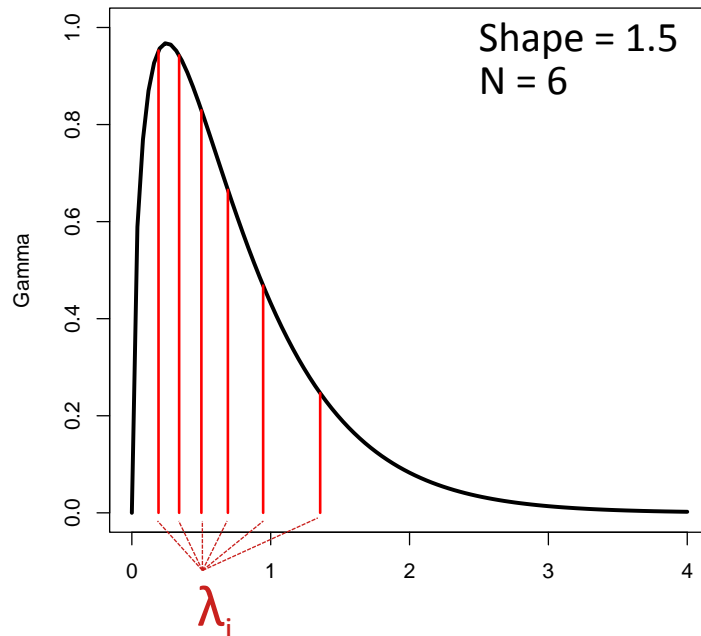


$$\begin{array}{l} \lambda_1 \quad \lambda_2 \\ \mu_1 \quad \mu_2 \end{array}$$

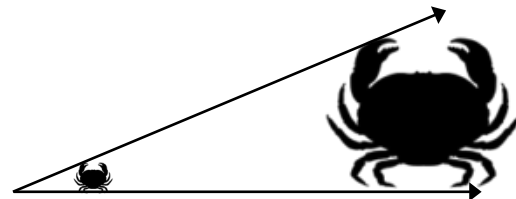
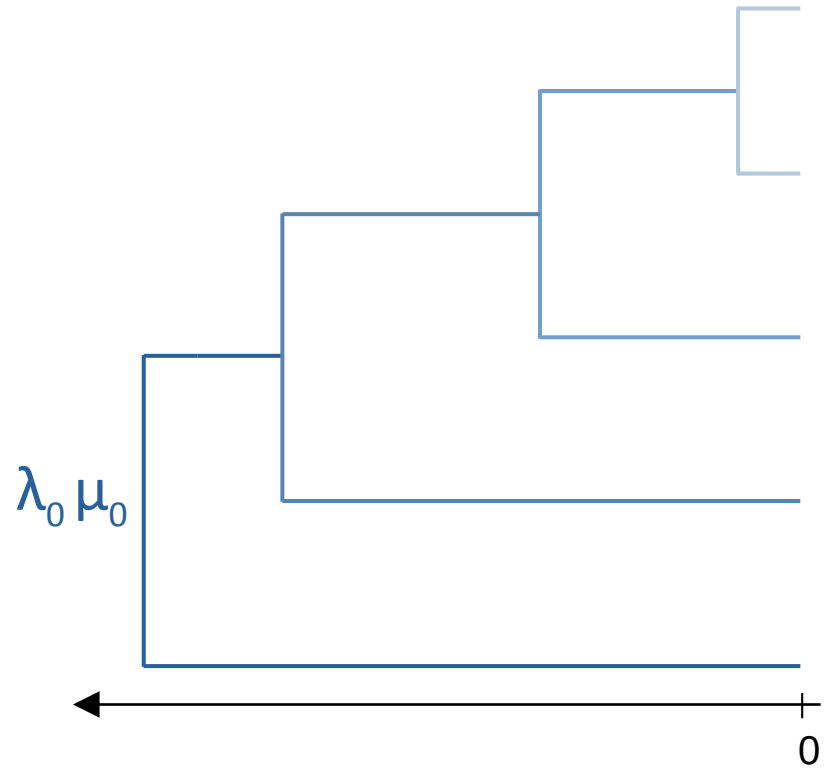
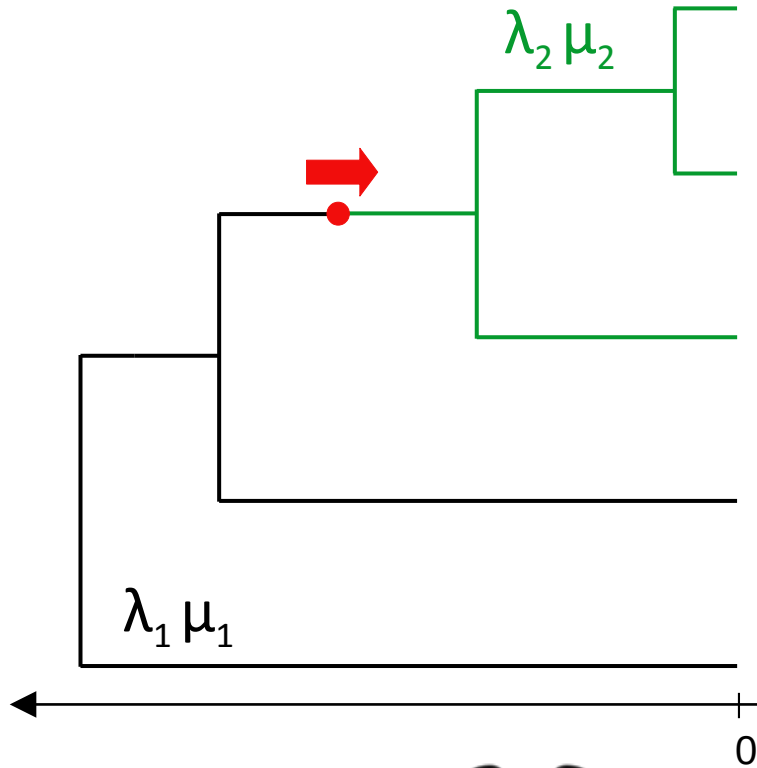
RevBayes model

- 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

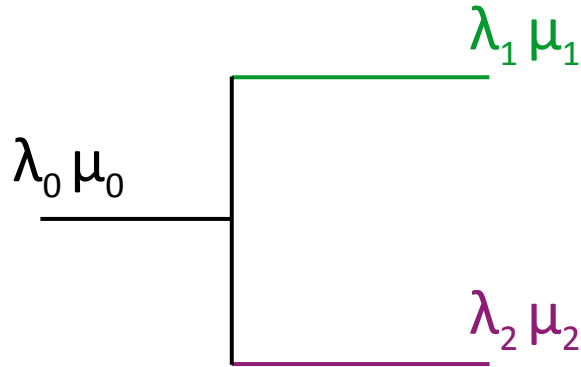
Höhna et al. 2019, BiorXiv



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.

- Continuous evolution process driven by a lognormal distribution
- 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
- Assumption of constant turnover:
$$\mu_i / \lambda_i = \mu_0 / \lambda_0$$

Examples

ARTICLE



<https://doi.org/10.1038/s41467-020-16498-w>

OPEN

Trophic innovations fuel reef fish diversification

Alexandre C. Siqueira ¹✉, Renato A. Morais ^{1,2}, David R. Bellwood ^{1,2} & Peter F. Cowman ¹

Nat. Comm. 2020

No link between population isolation and speciation rate in squamate reptiles

Sonal Singhal^{a,1}, Guarino R. Colli^b , Maggie R. Grundler^{c,d}, Gabriel C. Costa^e , Ivan Prates^{f,g}, and Daniel L. Rabosky^{f,g,1}

PNAS 2022

So – character-dependent or independent ?

Character-dependent / hybrid

- 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 ?

Outline

- Introduction
- Multi-type birth-death processes
- Character-dependent processes
- Character-independent processes
- Use in practice and extensions

The issue of phylogenetic inference

In practice, multi-type birth-death processes are frequently applied to fixed phylogenies via a two-step process:

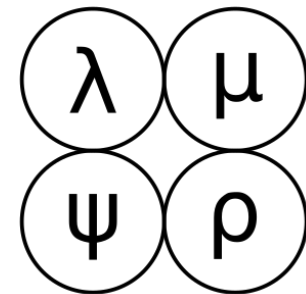
- 1) phylogenetic inference
- 2) diversification analysis

ACAC...
TCAC...
ACAG...

(1)



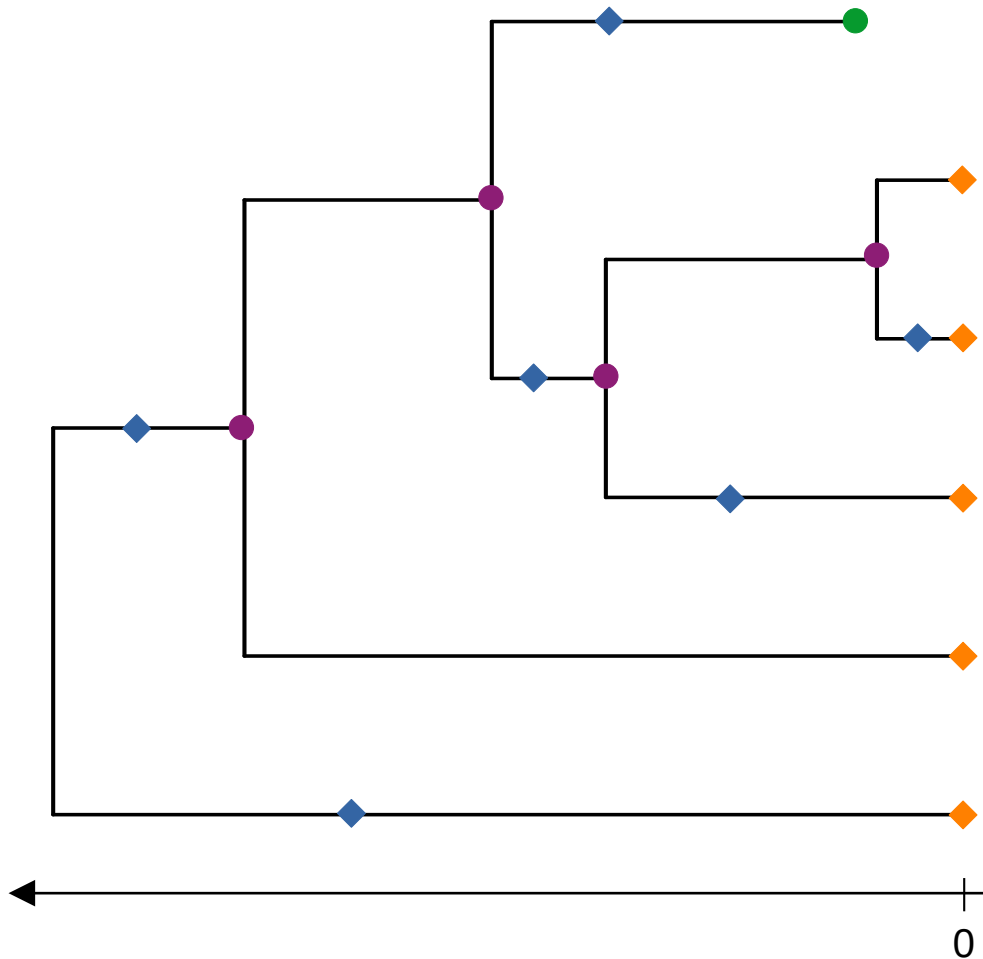
(2)



The issue of phylogenetic inference

- Why is this done ?
 - Computational cost of inference
 - Reuse of existing phylogenies to test new hypotheses
- Why is this an issue ?
 - Phylogenetic uncertainty is not properly accounted for
 - Mismatch of assumptions between step 1 (constant rates) and step 2 (rate variations)

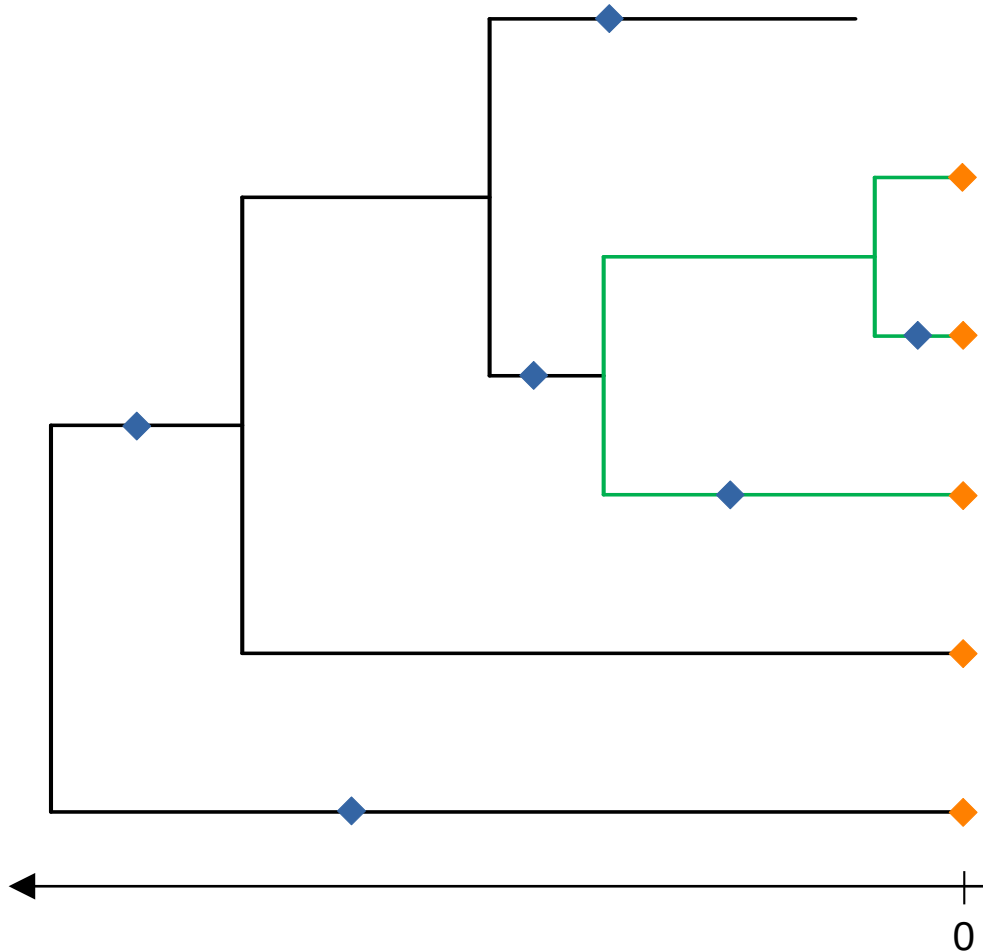
Integration of fossil data



Parameters:

- λ — birth rate
- μ — death rate
- ◆ ψ — fossilization rate
- ◆ ρ — extant species sampling probability

Multi-type FBD 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

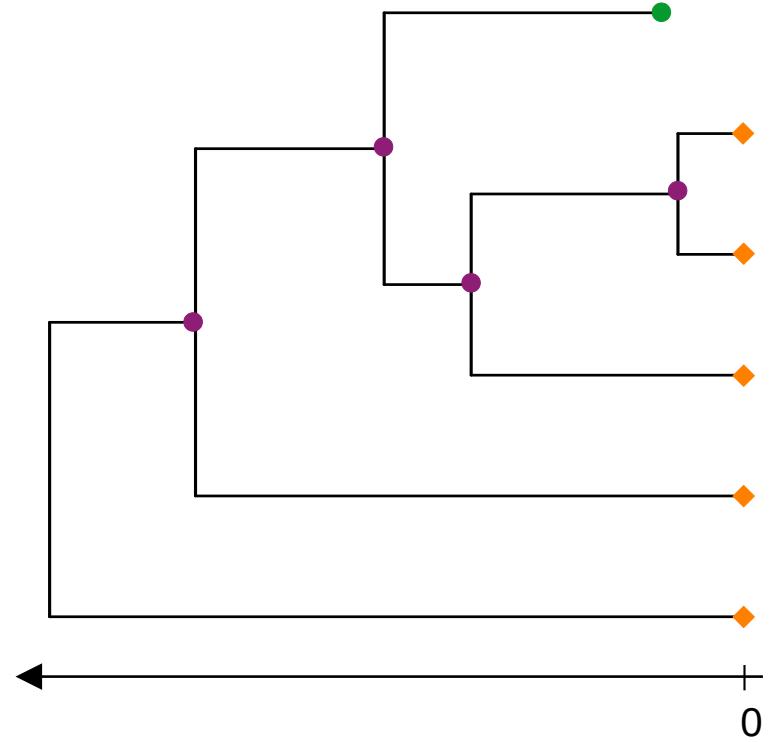
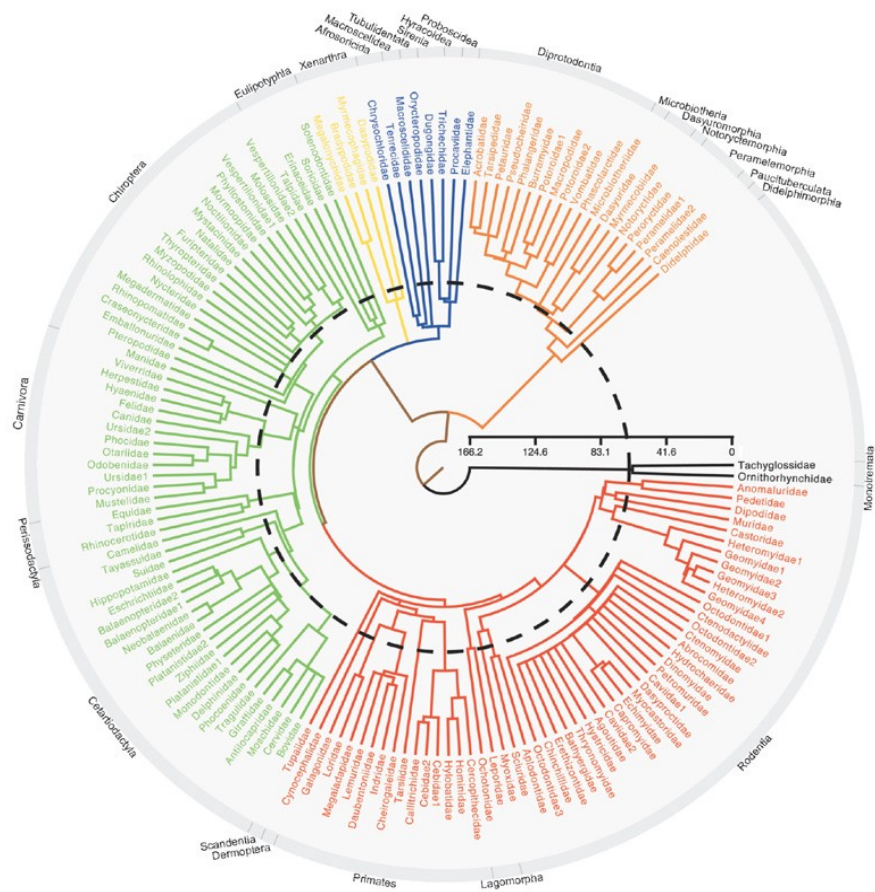
Key questions for including fossils

- Is the model identifiable or does it need additional assumptions ?
- Is the computational cost of the combined model manageable ?
- Can we distinguish between variation in fossil recovery and variation in diversification processes ?
- How many fossils & data are needed for reliable inferences ? Is that different for character-dependent vs character-independent models ?
- How to handle missing or uncertain characters for fossil tips ?

In summary

- Empirical data supports widespread variation in evolutionary processes, which can be modeled using multi-type birth-death processes
- 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
- These models are still a very active area of research and development (extension to continuous processes, integration of fossils, interpretation of results, etc.)

Questions ?



Tutorial time



Download

Tutorials

Documentation

RevScripter

Workshops

Jobs

Developer

GUI

State-dependent diversification with BiSSE and MuSSE

Inference using the binary/multiple state-dependent speciation and extinction (BiSSE/MuSSE) branching process

Sebastian Höhna, Will Freyman, and Emma Goldberg

Last modified on March 11, 2022

<https://revbayes.github.io/tutorials/sse/bisse.html>