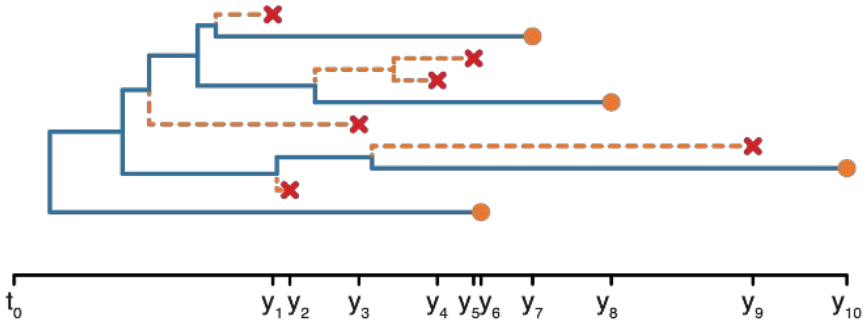
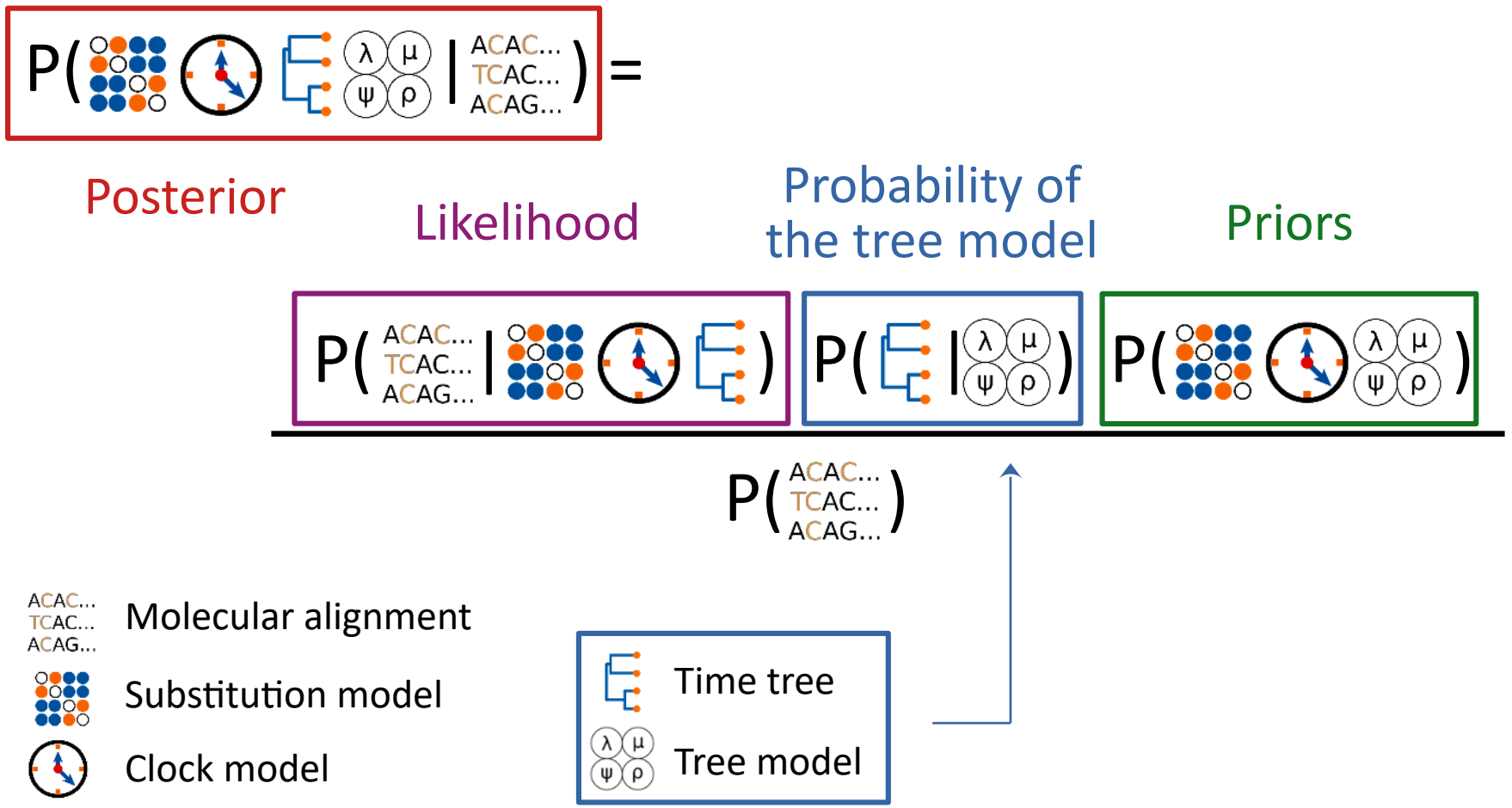


Heterogeneity in evolutionary processes : structured processes

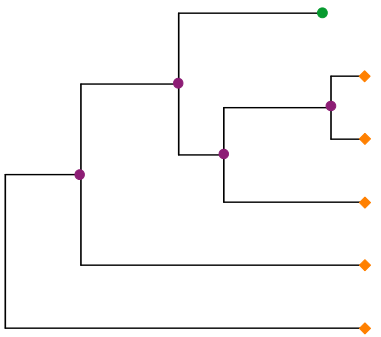
Joëlle Barido-Sottani

Tree models in Bayesian inference

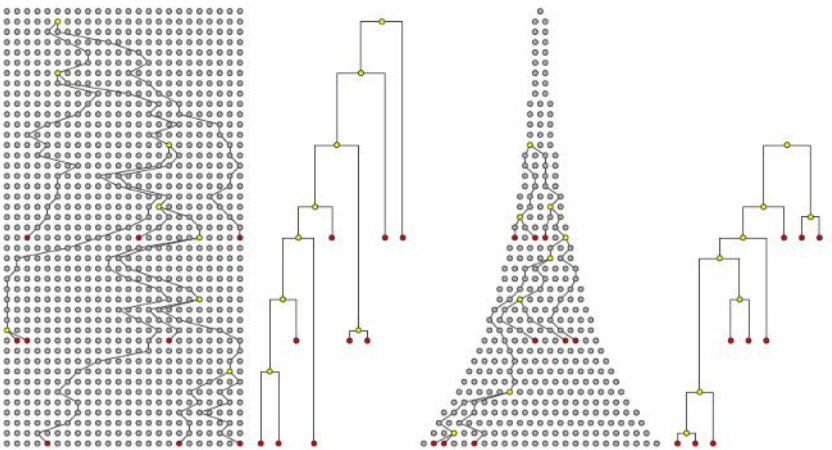


Phylodynamic models

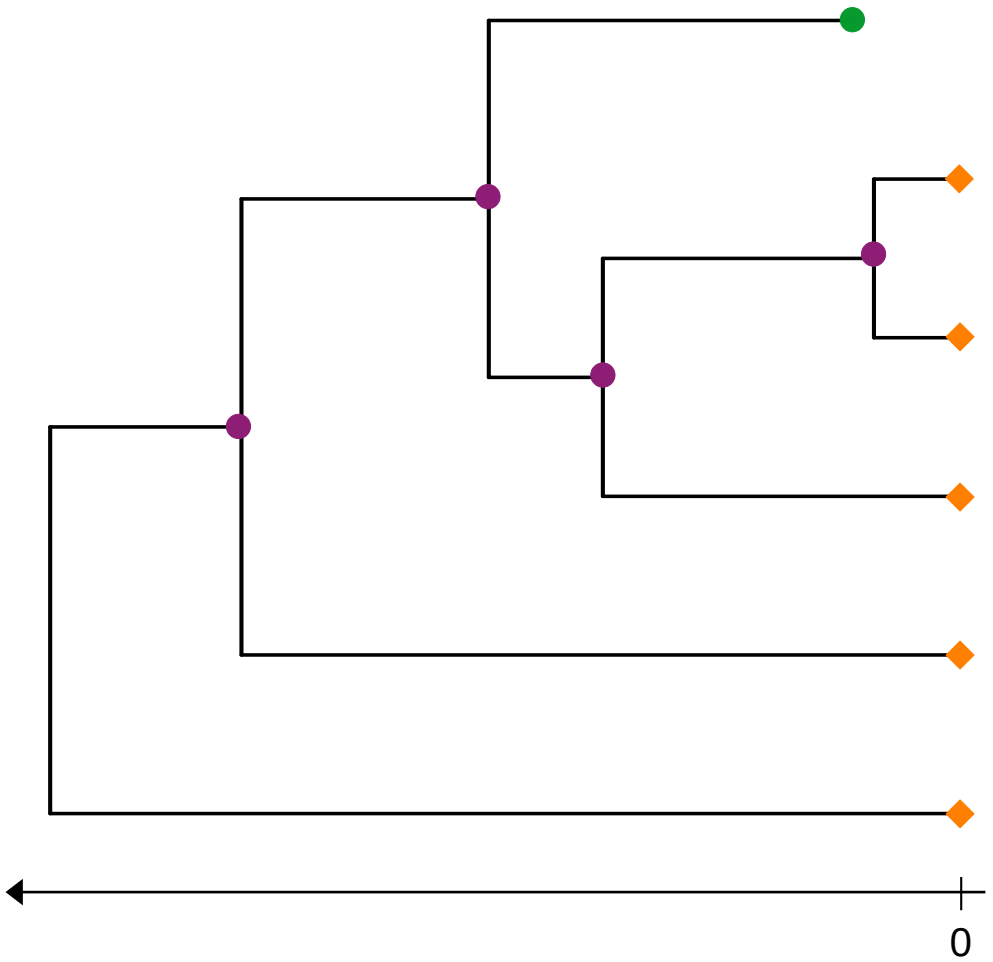
- Birth-death models



- Coalescent models



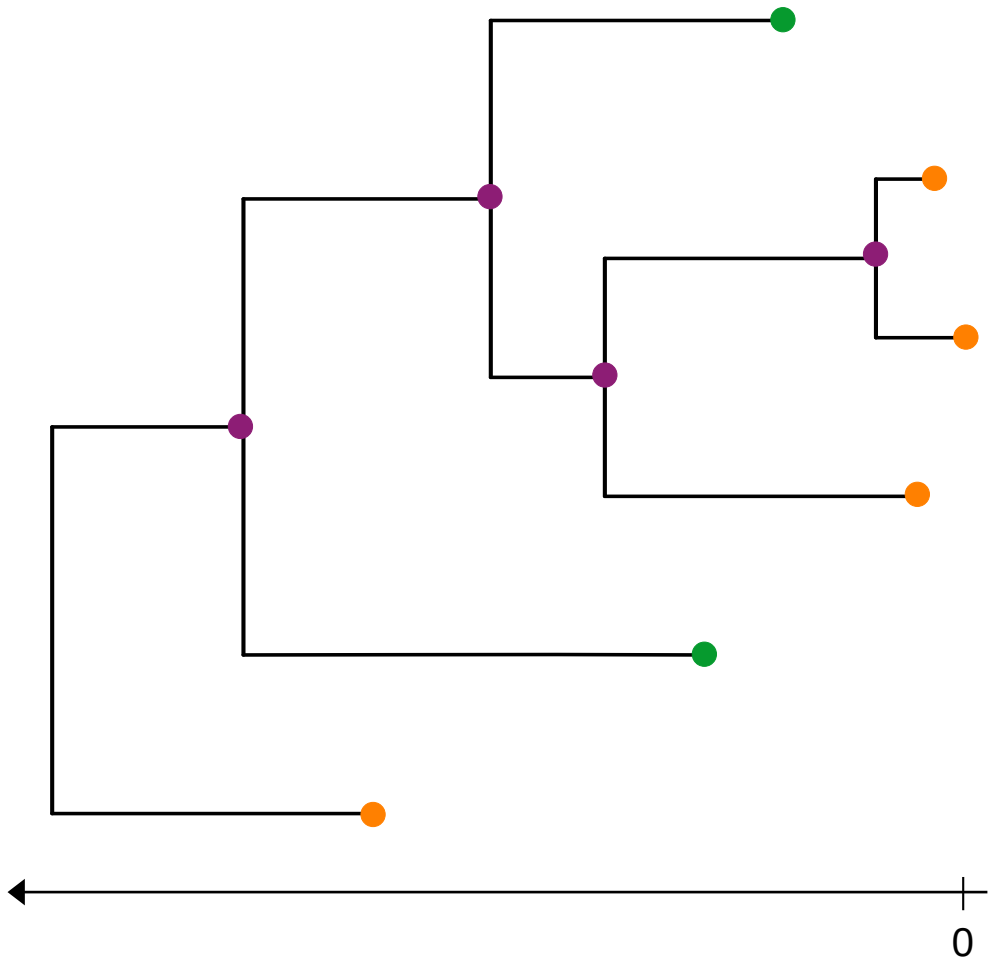
Simple birth-death process

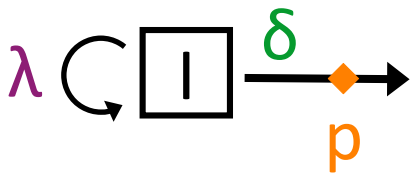


Parameters:

- λ — birth rate (= new lineage appearing)
- μ — death rate (= lineage disappearing)
- ρ — extant species sampling probability

Birth-death for epidemiology

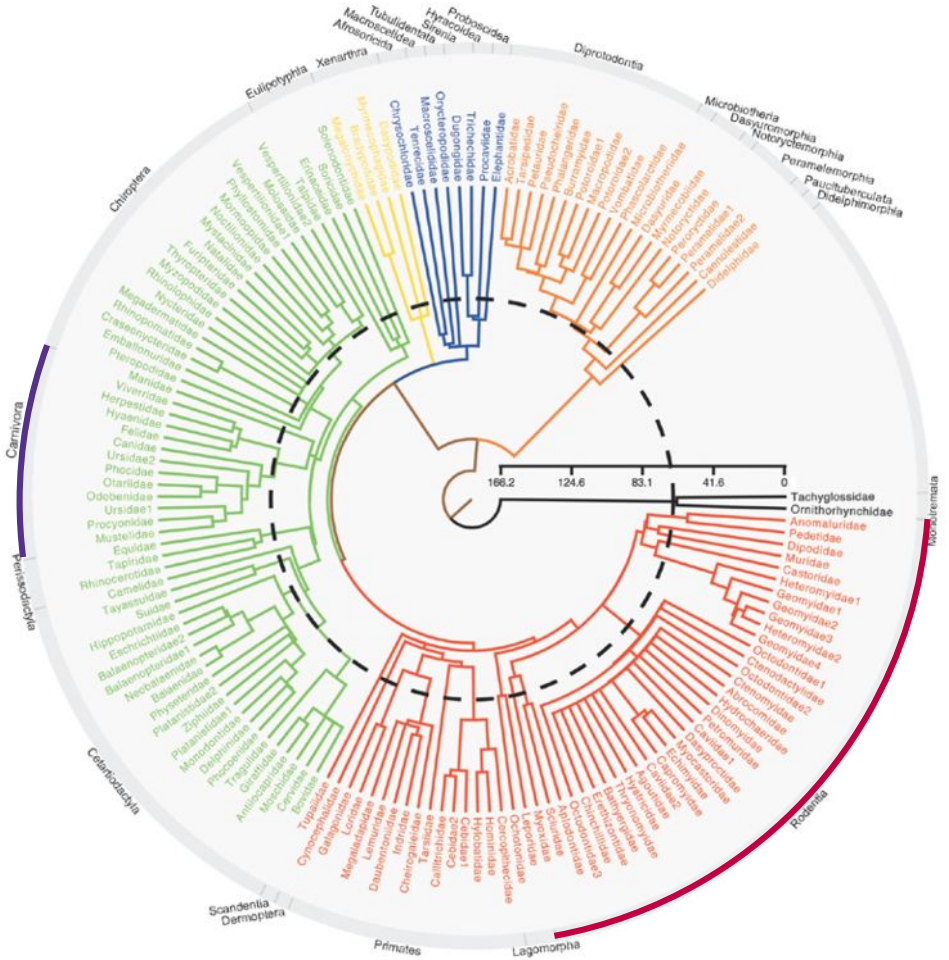




Processes:

- λ — transmission rate
- μ = δ(1–p) — rate of recovery without sampling
- ψ = δp — rate of recovery with sampling

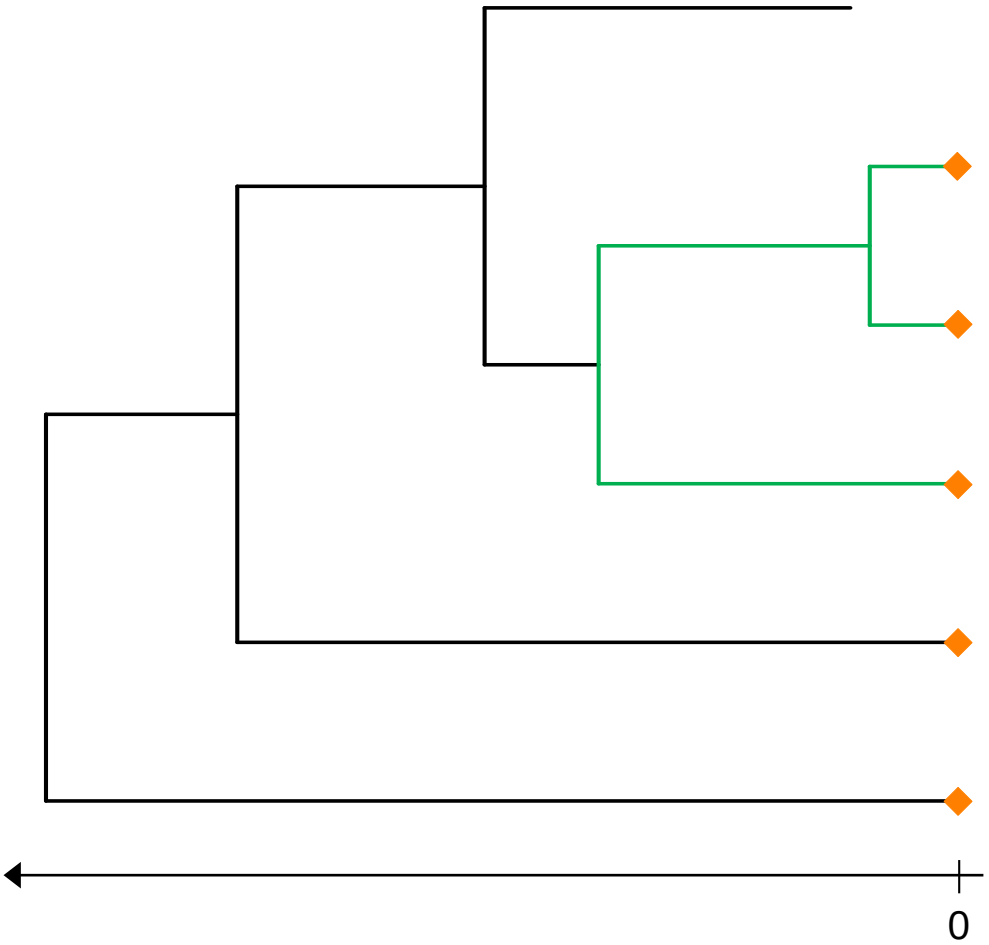
Heterogeneity in evolution



Bininda-Emonds *et al.* **Nature 2007**

- 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



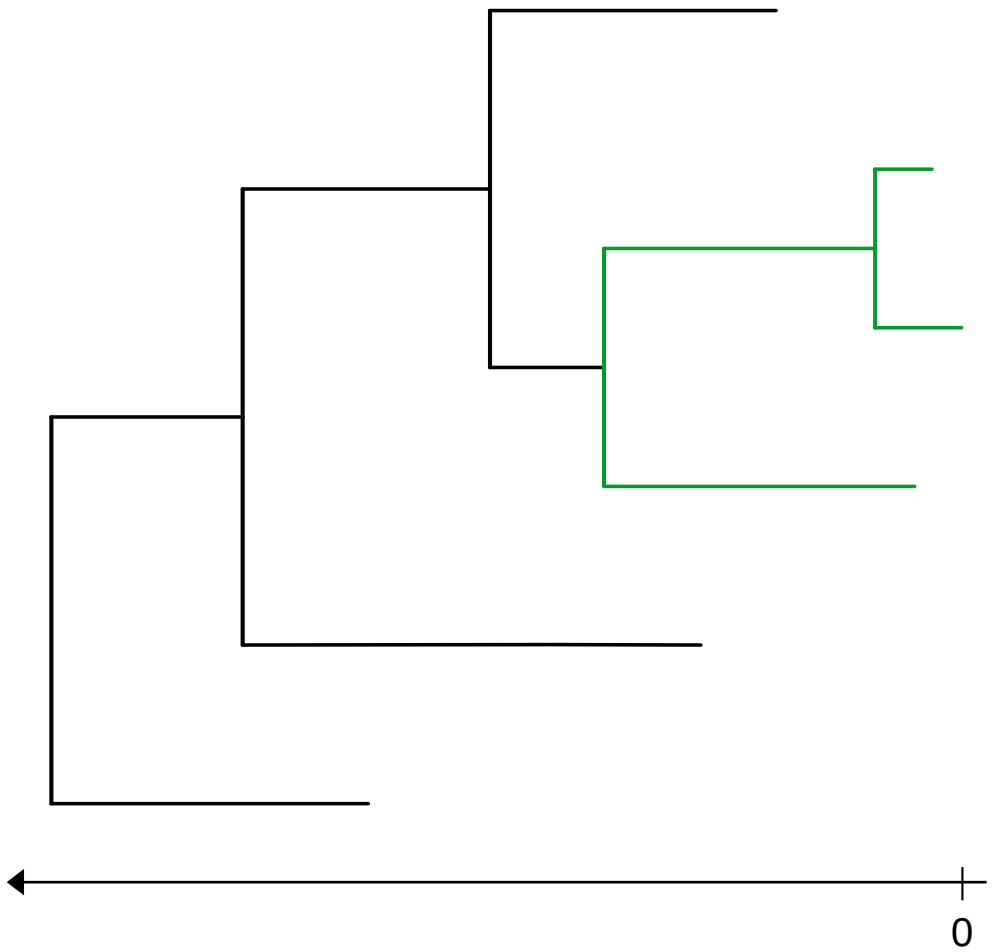
2 types, type 1 & type 2

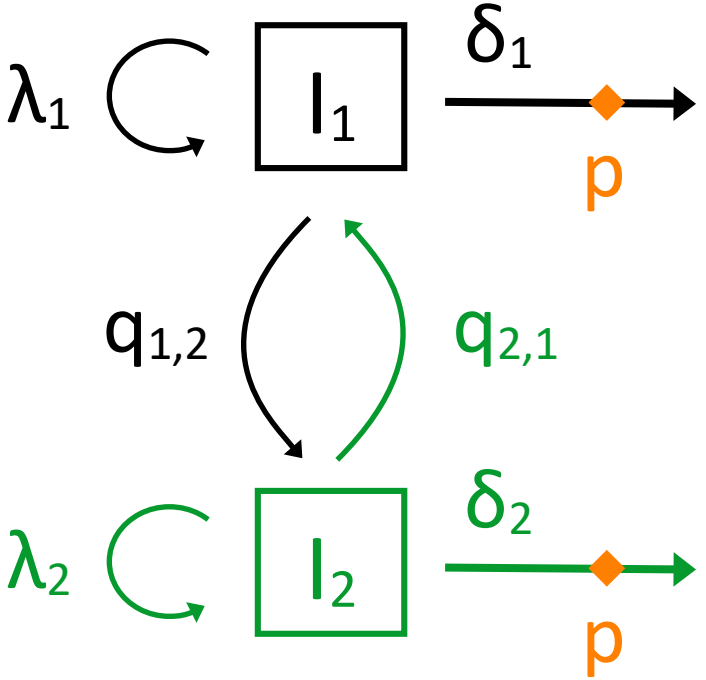
λ_1 & λ_2 — birth rates

μ_1 & μ_2 — death rates

ρ — extant species sampling probability

MTBD process (epidemiology)





Character-dependent or independent ?

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

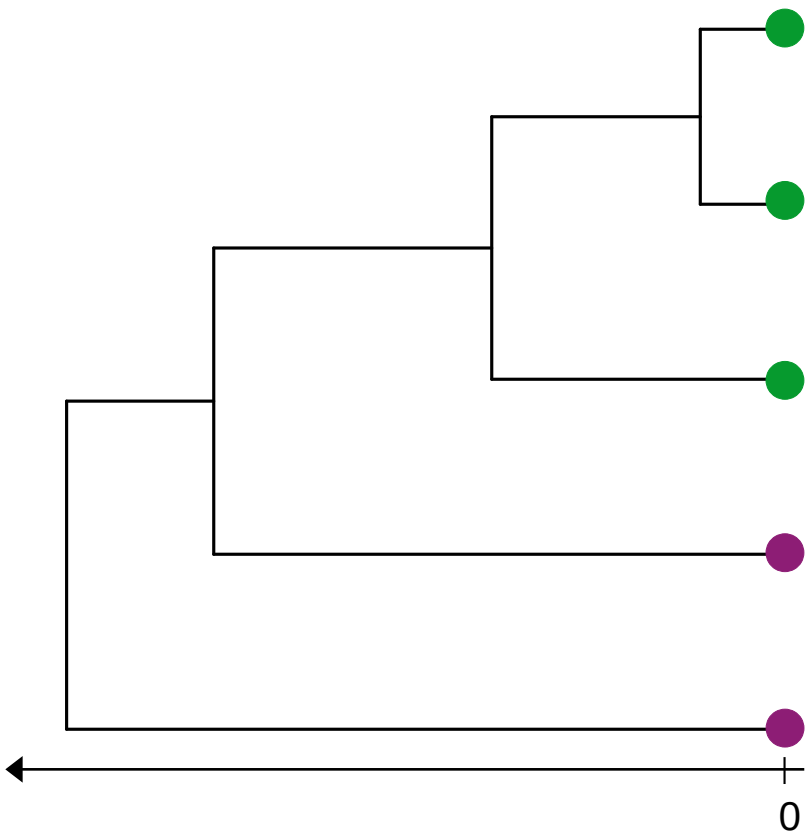
Type 1 Type 2 Type 3

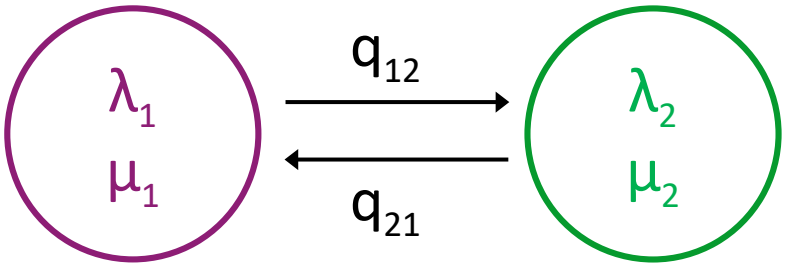
Character-independent

In a character-dependent model :

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

The BiSSE/MuSSE/BDMM model



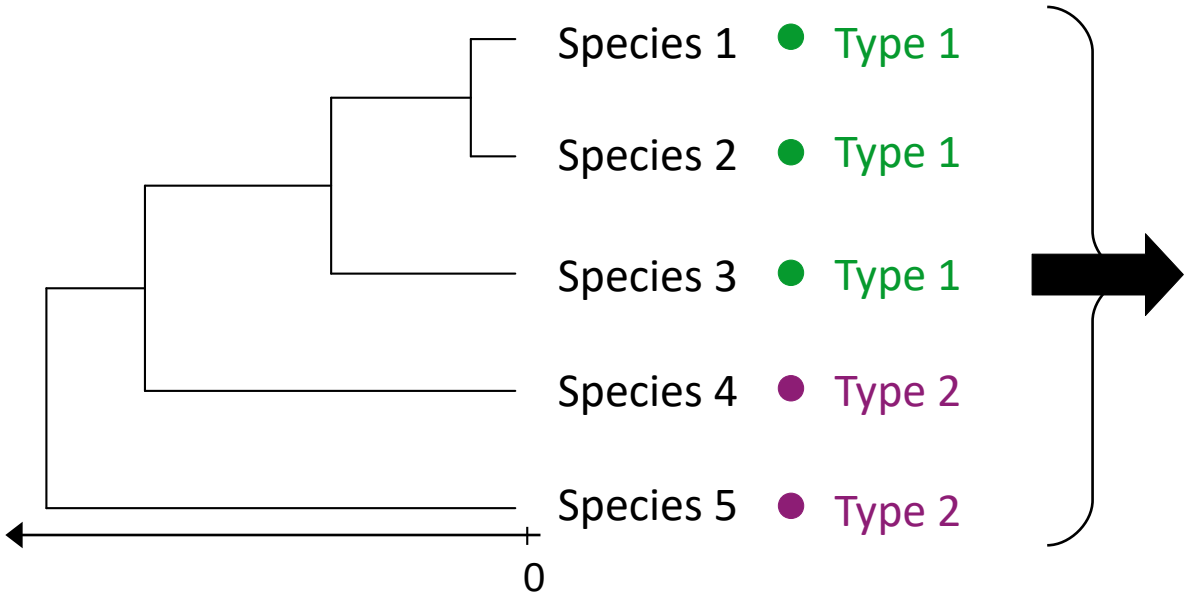


Parameters of the model:

- λ_i – birth rates
- μ_i – death rates
- q_{ij} – transition rates
- ρ/p – sampling probability

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

SSE/BDMM inference



$\lambda_1 \lambda_2$ – birth rates

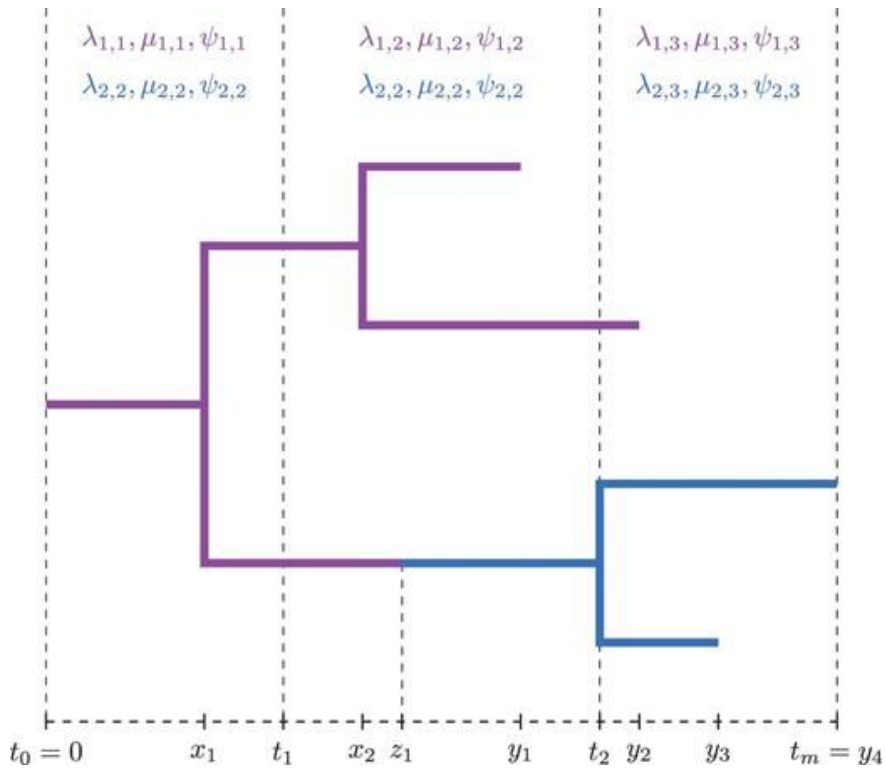
$\mu_1 \mu_2$ – death rates

$q_{12}\, q_{21}$ – 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: $\lambda_{i,j}$



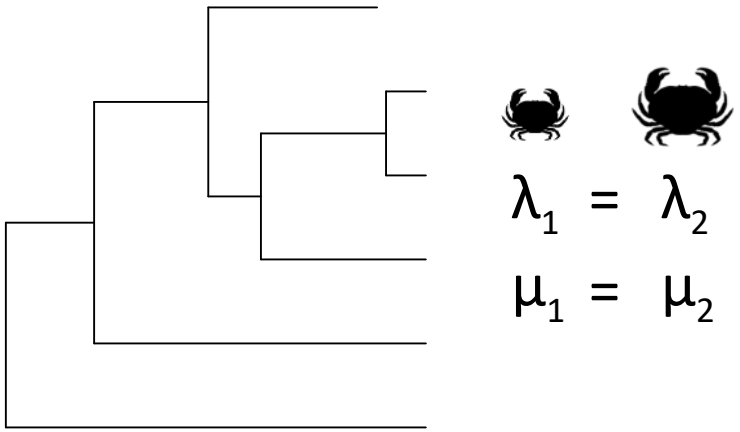
New: BDMM-Prime

- Can estimate the full (unsampled) LTT
- Uses advanced computational techniques for faster performance
- Can estimate the event history or integrate over it
- New improved BEAUti interface for setup

Bayesian Phylodynamic Inference of Multitype Population Trajectories Using Genomic Data

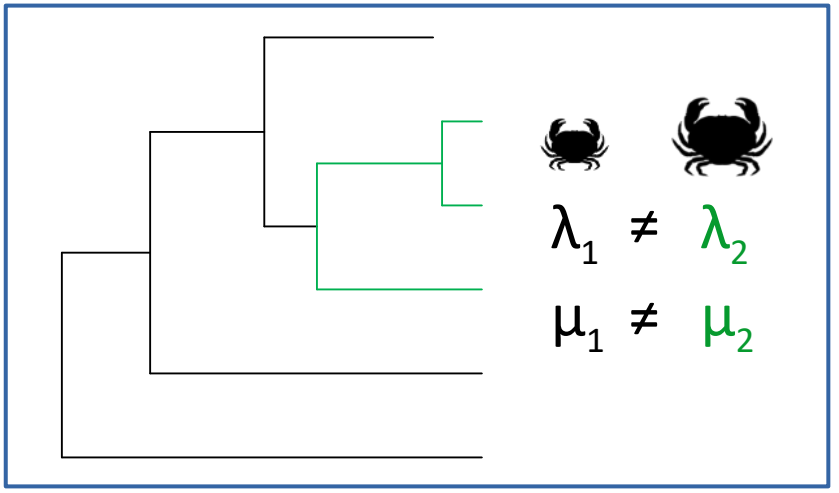
Timothy G. Vaughan [1,2,*] Tanja Stadler [1,2]

Model selection issues



$\lambda_1 = \lambda_2$

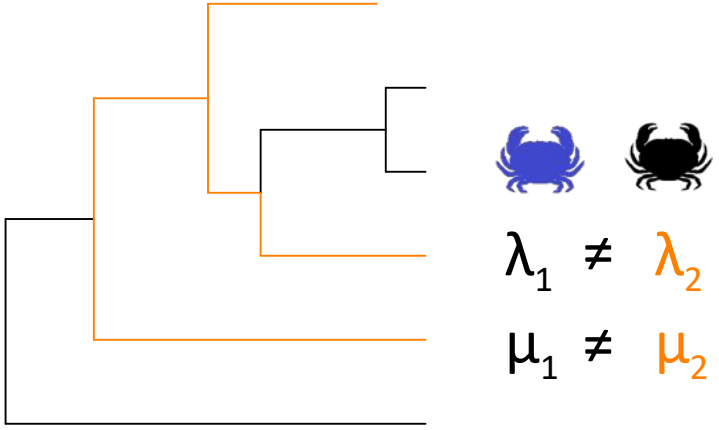
$\mu_1 = \mu_2$



$\lambda_1 \neq \lambda_2$

$\mu_1 \neq \mu_2$

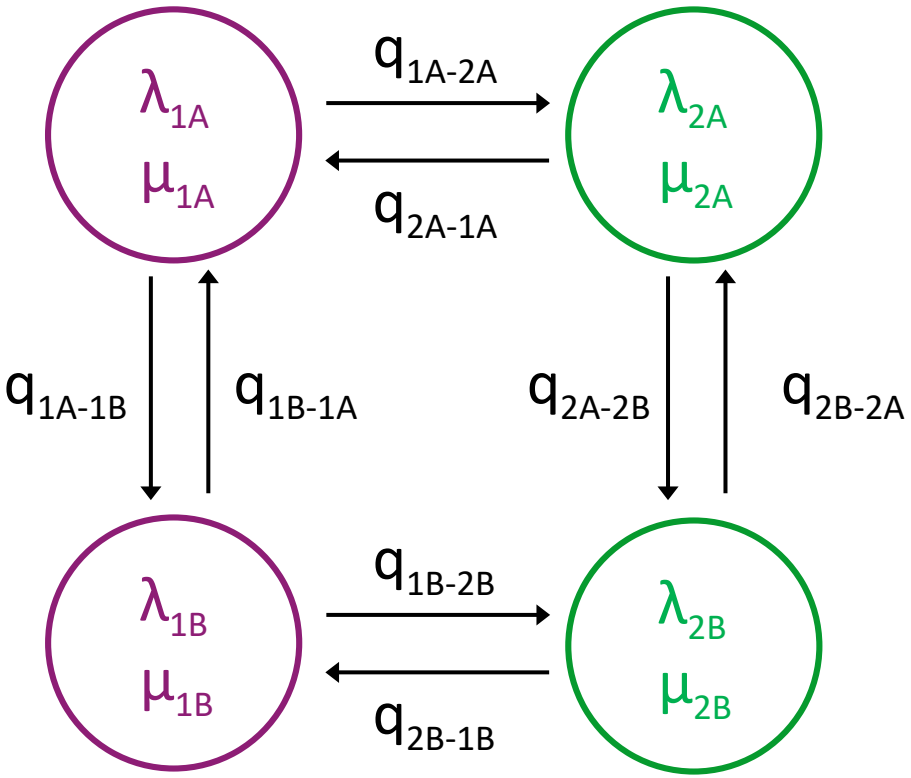
Rabosky & Goldberg 2015, Sys. Bio.



$\lambda_1 \neq \lambda_2$

$\mu_1 \neq \mu_2$

The HiSSE model



- 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

Beaulieu & O'Meara **Sys. Bio. 2016**

Examples: character-driven diversification

RESEARCH ARTICLE

Botany

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

Meng Lu[1,2] | Marc Fradera-Soler[1,3] | Félix Forest[1] |
Timothy G. Barraclough[2,4] | Olwen M. Grace[1]

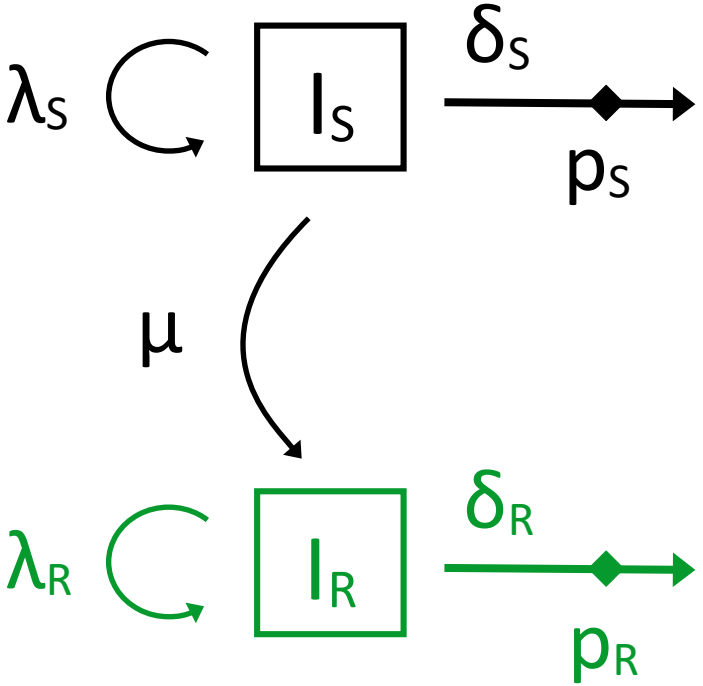
ORIGINAL ARTICLE

Ecological Entomology

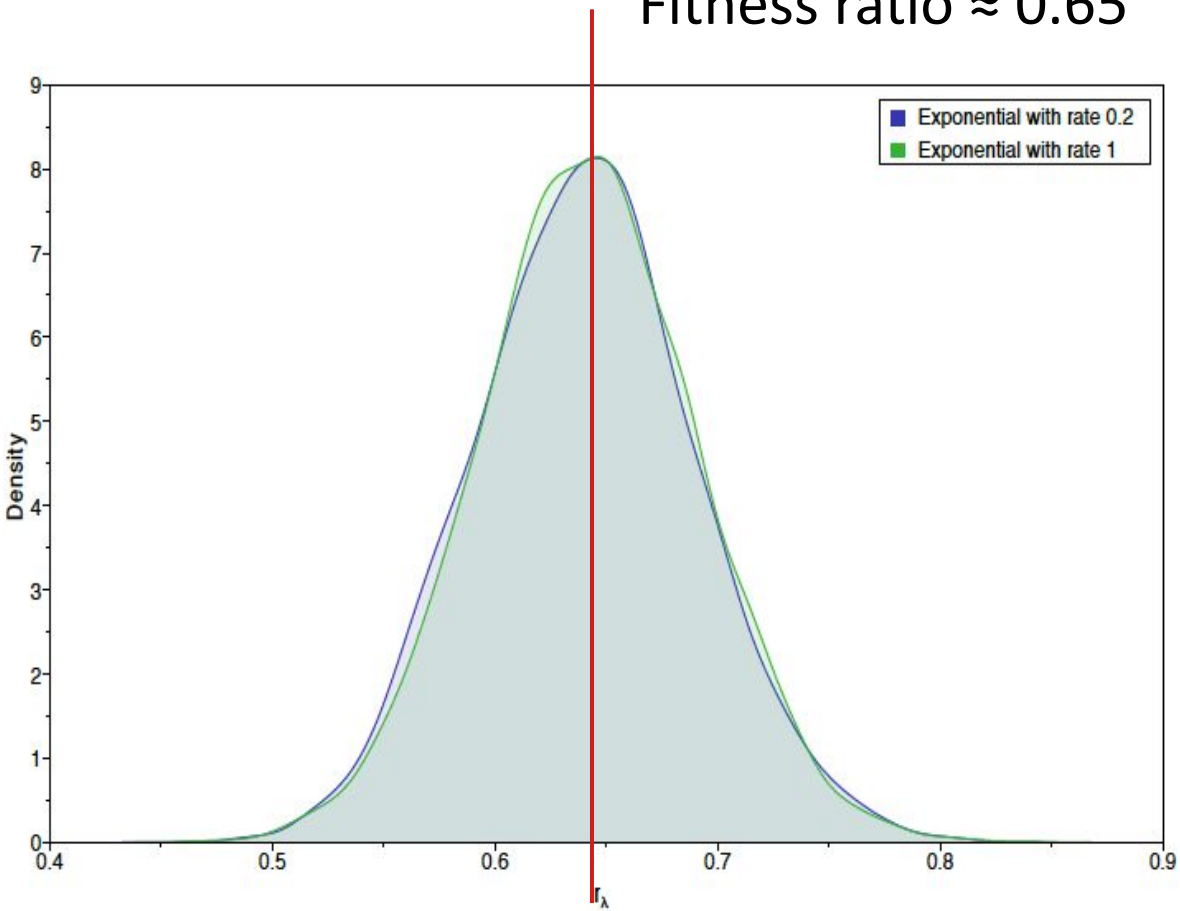
Fast species diversification among dragonflies (Anisoptera: Odonata: Insecta) inhabiting lentic environments regardless of wing pigmentation

Benjamín Padilla-Morales[1] | Paola Cornejo-Páramo[1] | Oscar García-Miranda[2] |
Aldo Issac Carrillo Muñoz[2] | Andrea Nieto López[2] | Daniel L. Castillo-Morales[3] |
Gustavo Wappler Barragán[2] | Araxi O. Urrutia[1,3] | Martín Alejandro Serrano-Meneses[2]

Example: fitness of resistant tuberculosis



Fitness ratio ≈ 0.65



Pečerska *et al.* **Epidemics 2021**

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

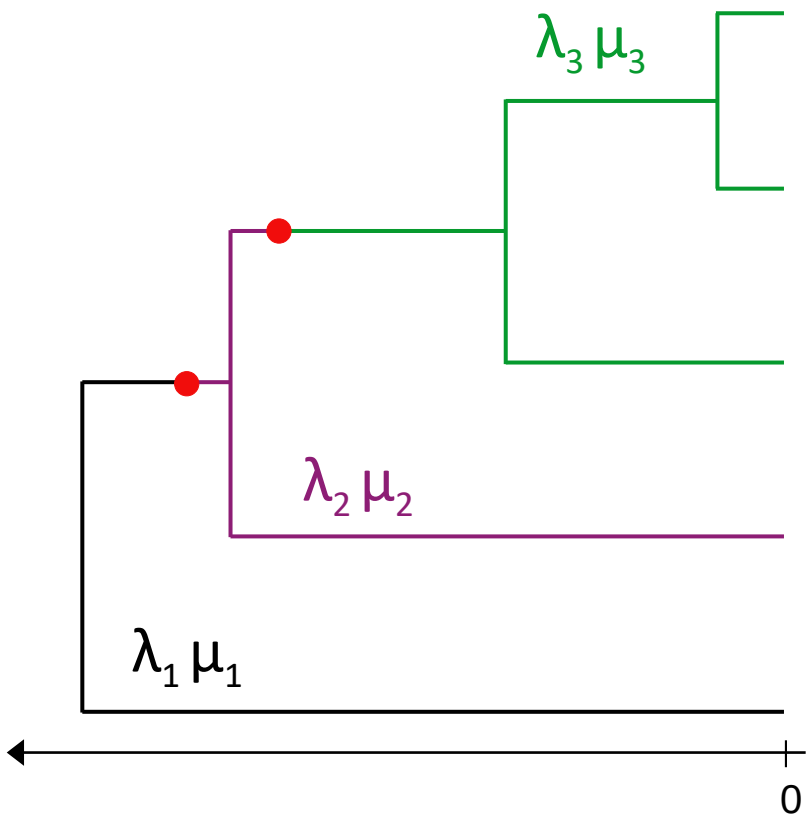
Type 1 Type 2 Type 3

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
- 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 ɣ (MSBD)
- Assumes that all types appear in the sampled tree – no unseen types

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

Simplifying the model

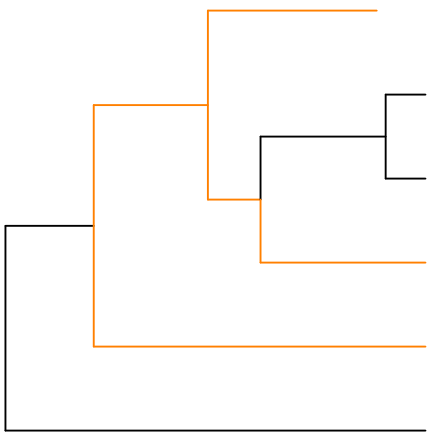
Character-dependent

$\lambda_1 \approx \lambda_2 \quad \lambda_3$

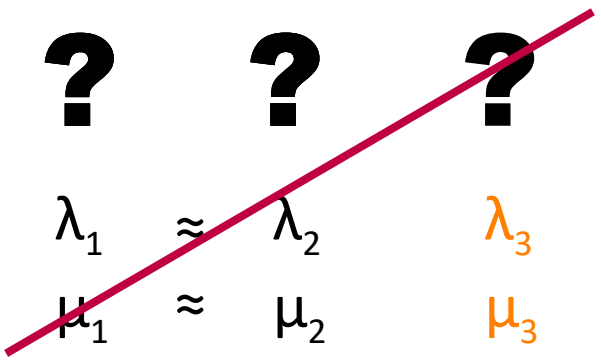
$\mu_1 \approx \mu_2 \quad \mu_3$

$\lambda_1 \quad \lambda_2$

$\mu_1 \quad \mu_2$



Character-independent



$\lambda_1 \approx \lambda_2 \quad \lambda_3$

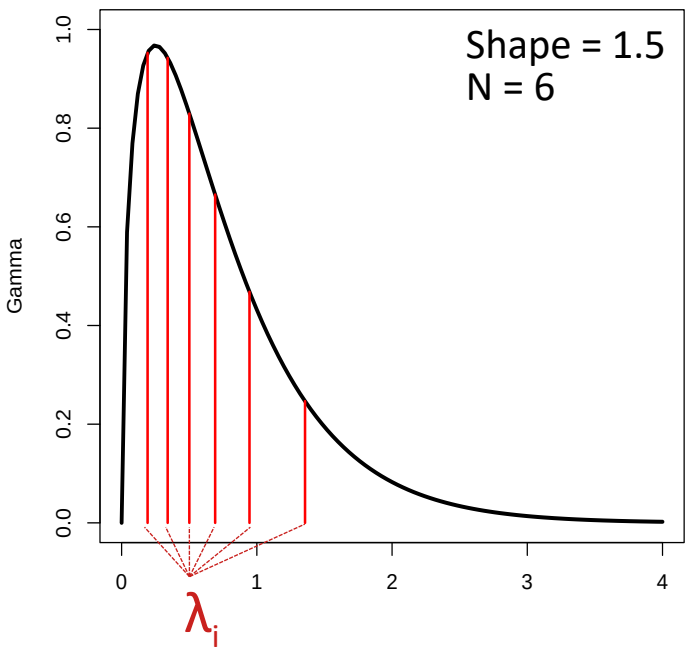
$\mu_1 \approx \mu_2 \quad \mu_3$

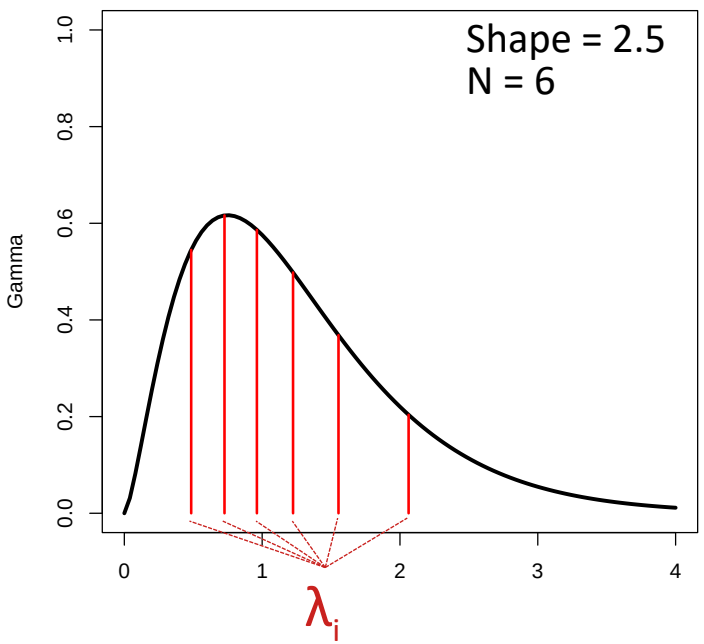
$\lambda_1 \quad \lambda_2$

$\mu_1 \quad \mu_2$

RevBayes model

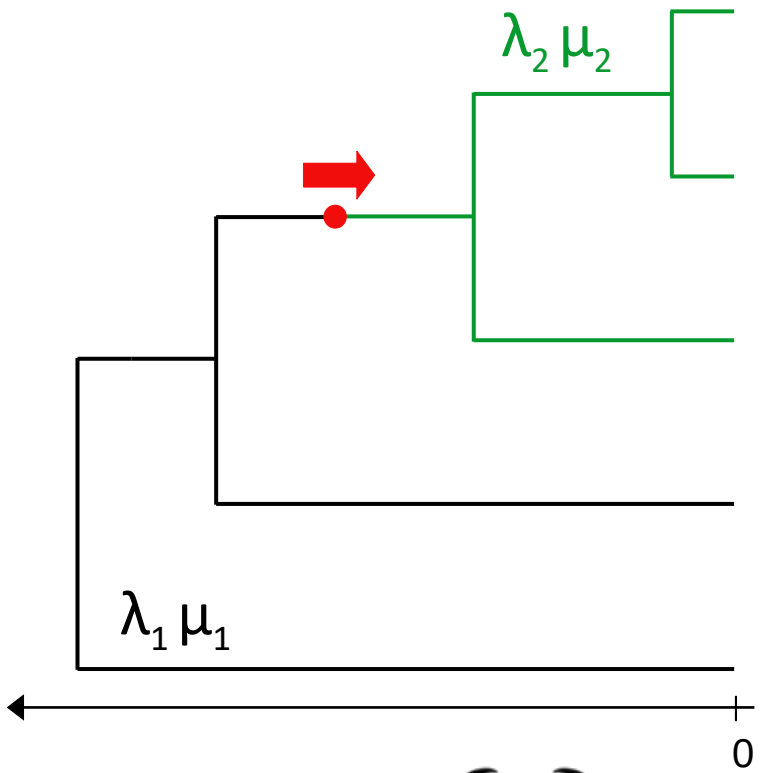
- Ordered types based on a Gamma distribution
- Fixed number of types N
- Simplified model: rates are not estimated, but determined by the shape of the Gamma distribution

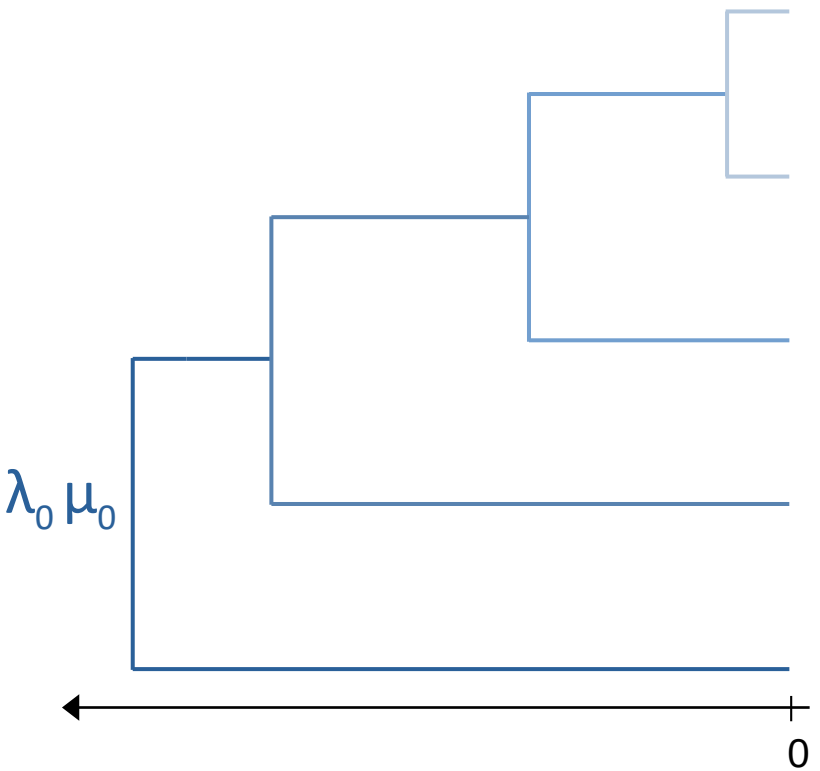


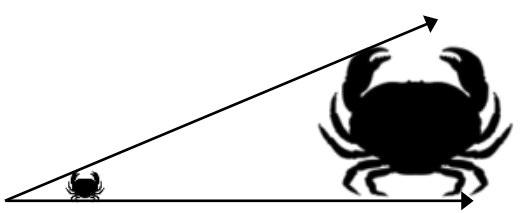


Höhna *et al.* **BiorXiv 2019**

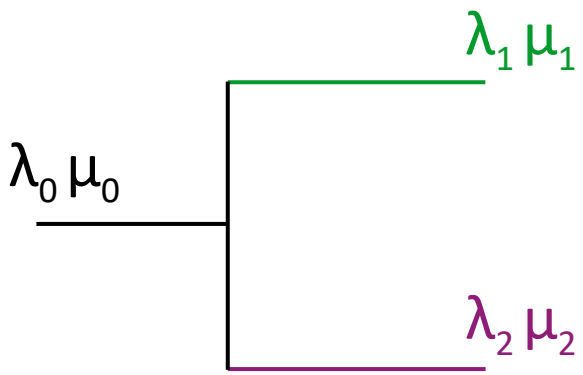
Going beyond types







ClaDS model



λ_1 = LogNormal (λ_0 x α, σ)

λ_2 = LogNormal (λ_0 x α, σ)

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

- 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
- Two parameterizations for μ
  - Lognormal process with α_μ and σ_μ
  - Assumption of constant turnover: $\mu_i / \lambda_i = \mu_0 / \lambda_0$

Examples: character-driven diversification

ARTICLE

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

Check for updates

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

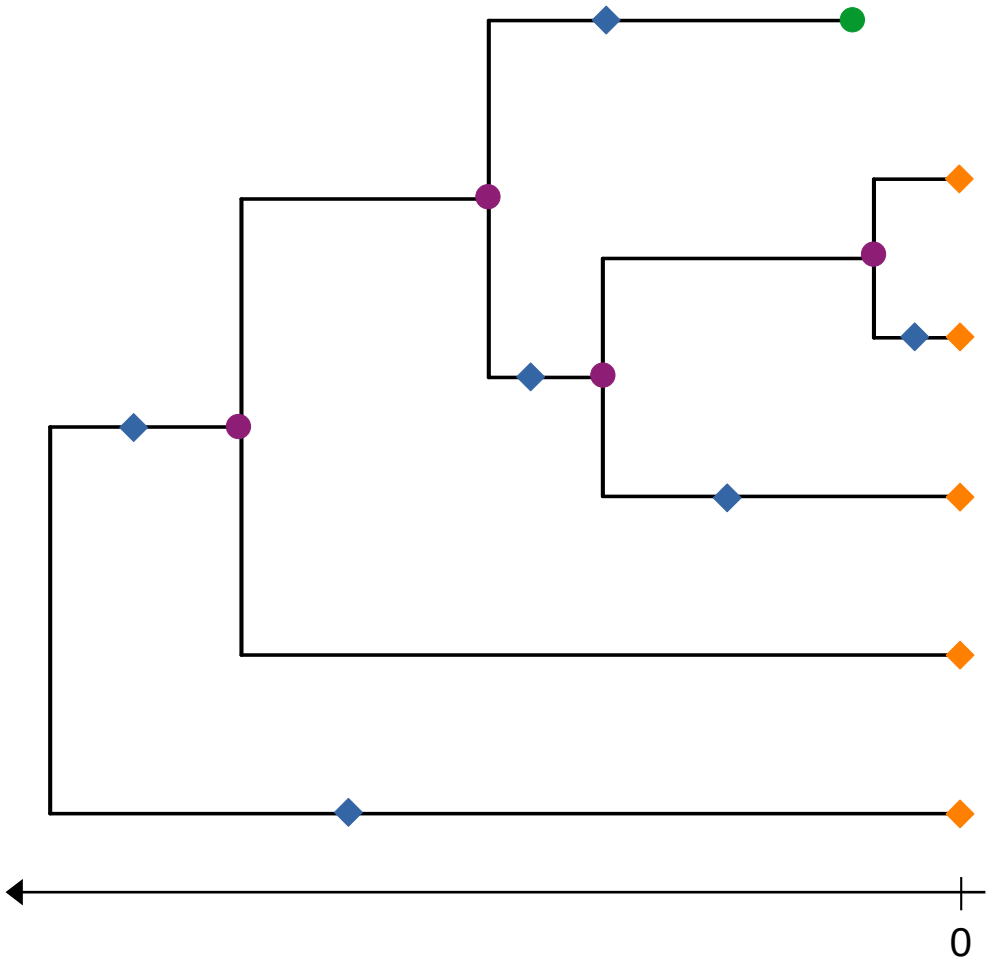
- Allow more complex transition processes
- Are more easily interpreted
- Are very dependent on the choice and accuracy of trait

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 ?

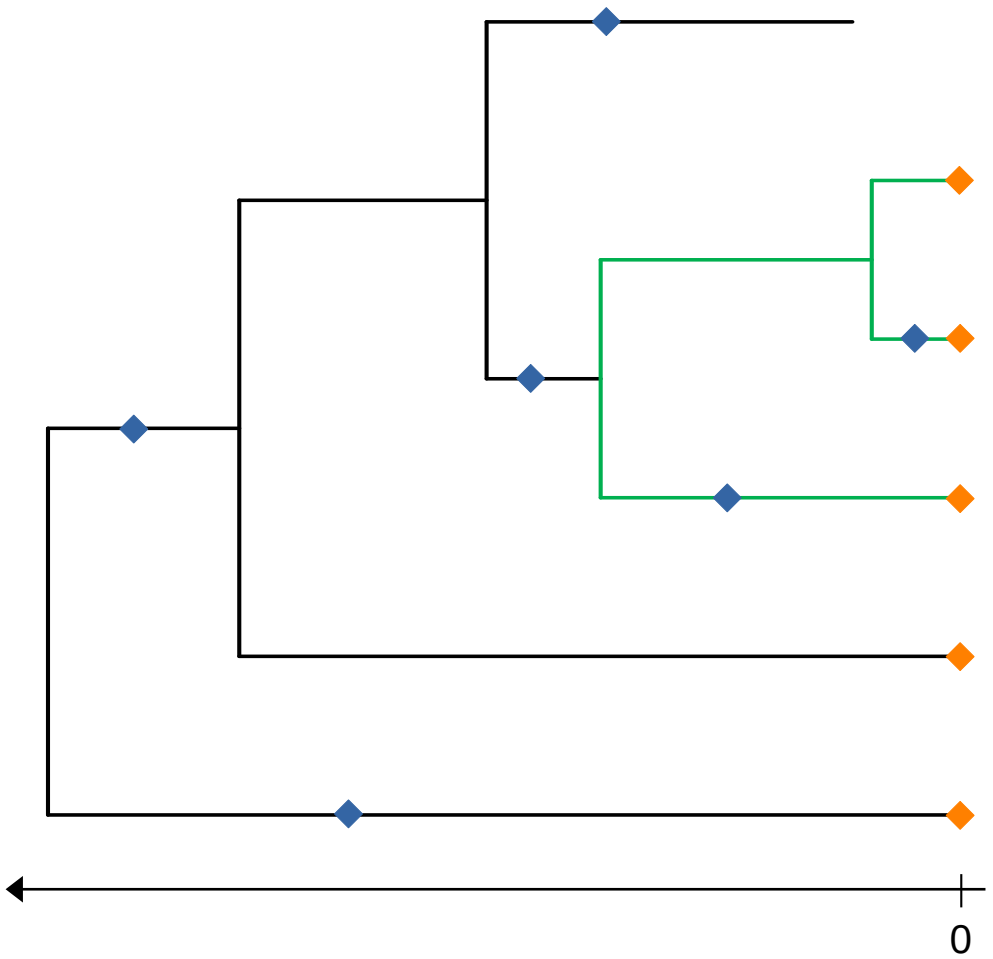
Integration of fossil/SA data



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

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.)

In summary (BEAST2)

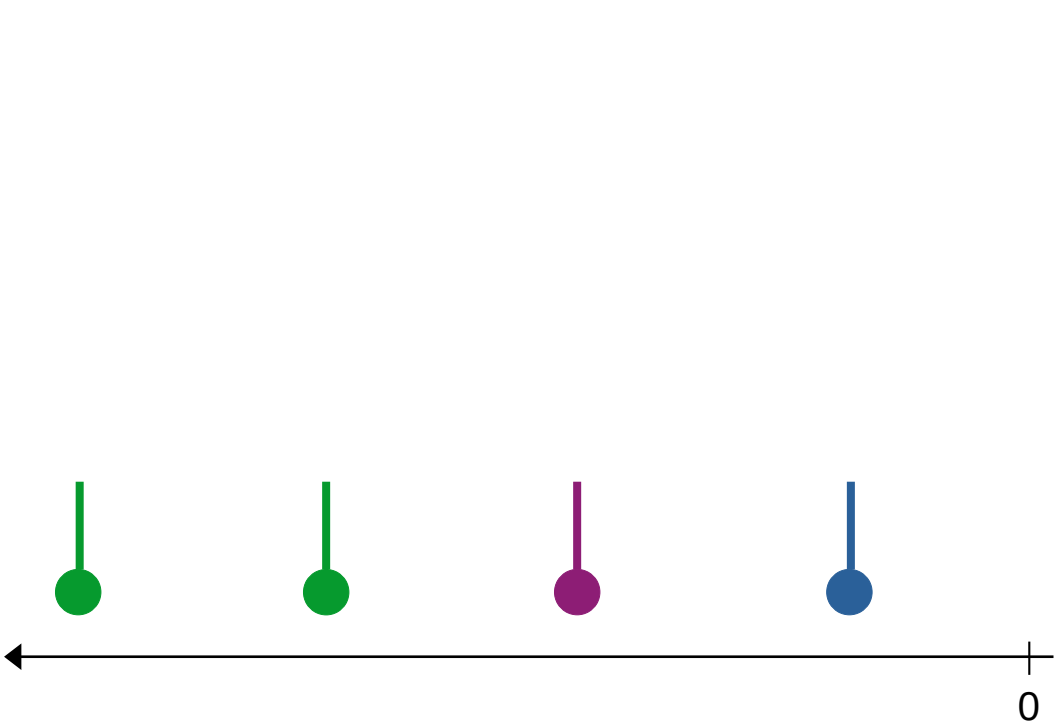
- **Birth-Death-Migration Model (BDMM) package**
  - Character-dependent SSE
  - Includes time-dependent changes (skyline model)
  - Includes sampled ancestors
- **Multi-State Birth-Death (MSBD) package**
  - Character-independent SSE
  - Includes sampled ancestors, starting from v1.3.0
- **Cladogenetic Diversification rate Shift (ClaDS) package**
  - Progressive autocorrelated rate variations
  - Inclusion of sampled ancestors in development

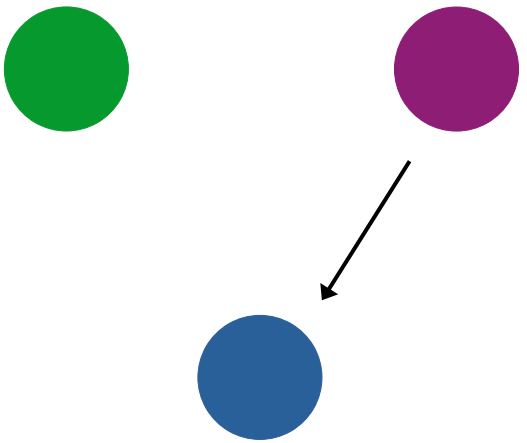
What about coalescent models?

- Basic coalescent model assumes **exchangeability**: all lineages can coalesce with each other

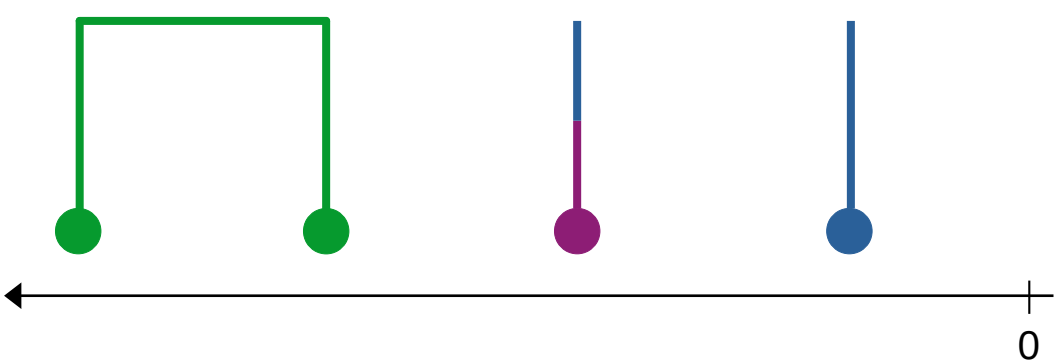
- Structured coalescent model:
  - n subpopulations with sizes N_{e1}, … , N_{en}
  - only lineages of the same subpopulation coalesce
  - adds migration events: one lineage moves from i to j

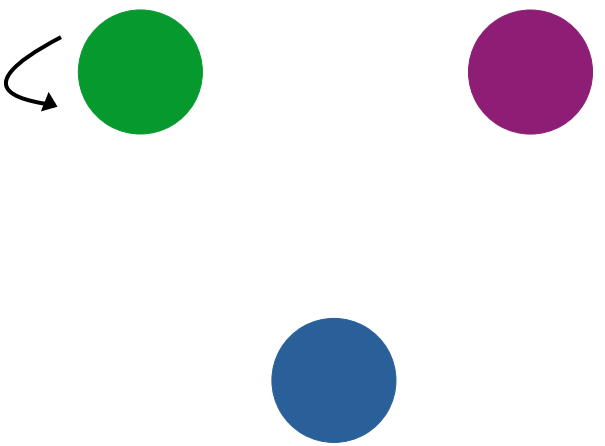
Structured coalescent in action



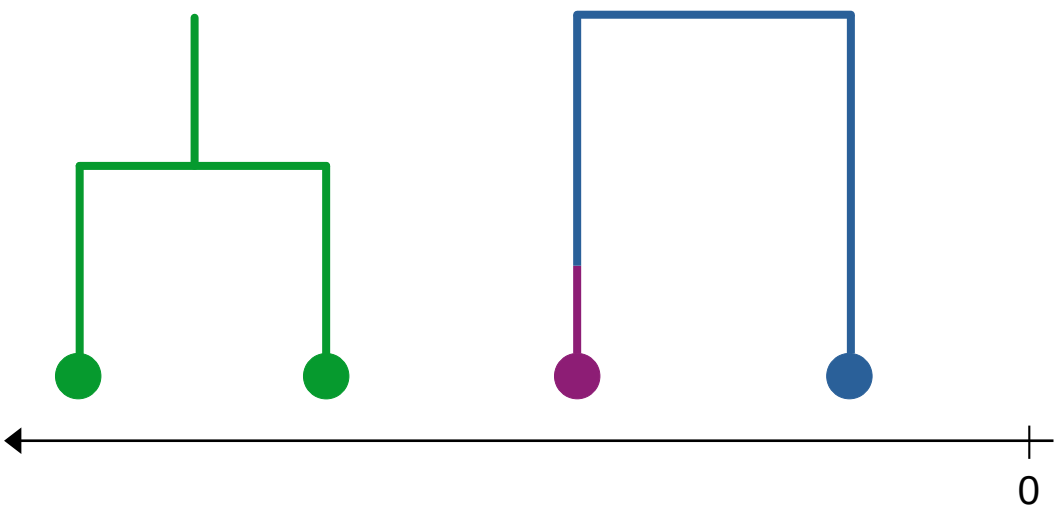


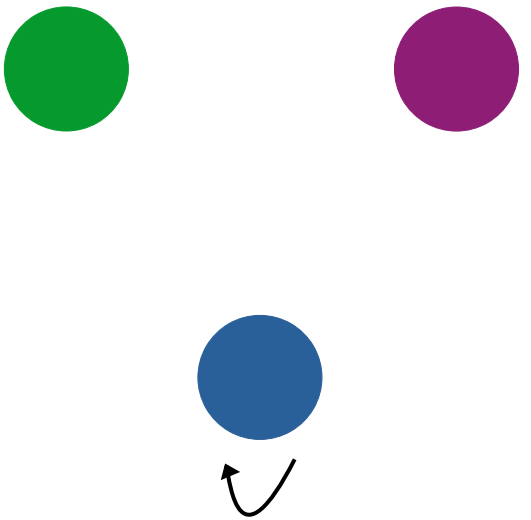
Structured coalescent in action



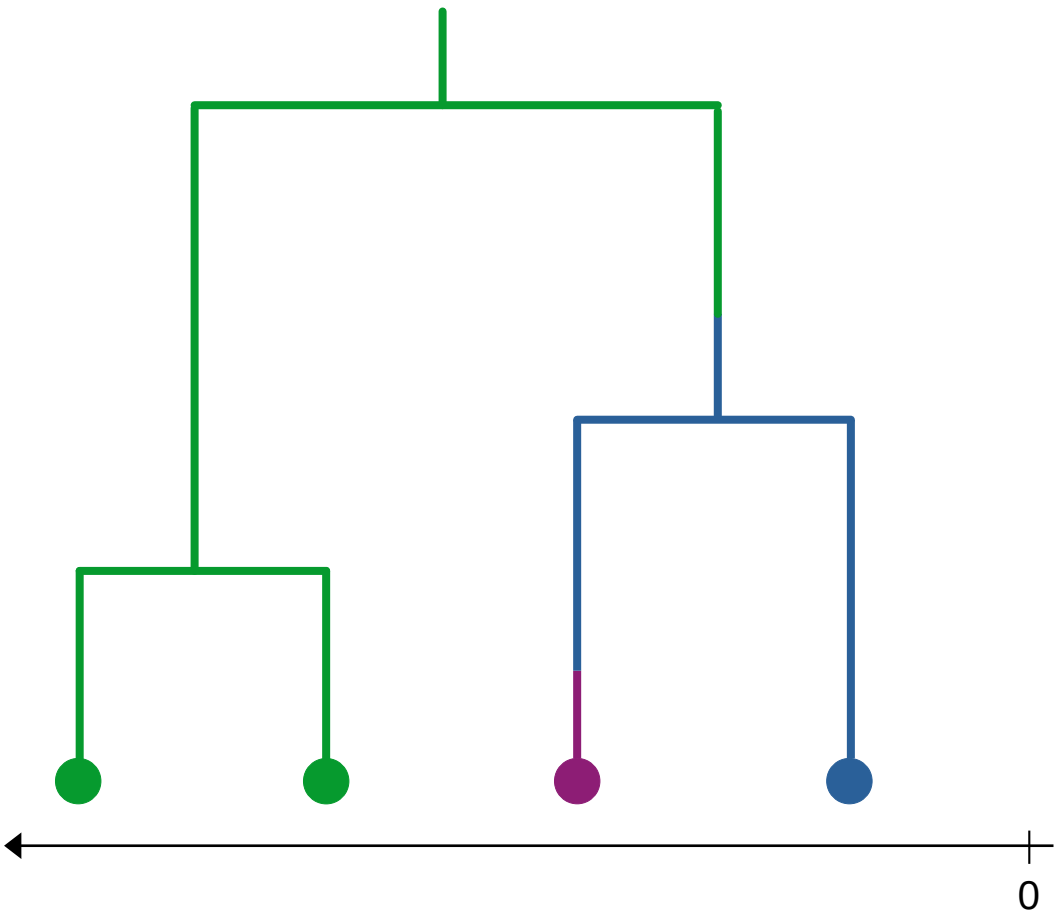


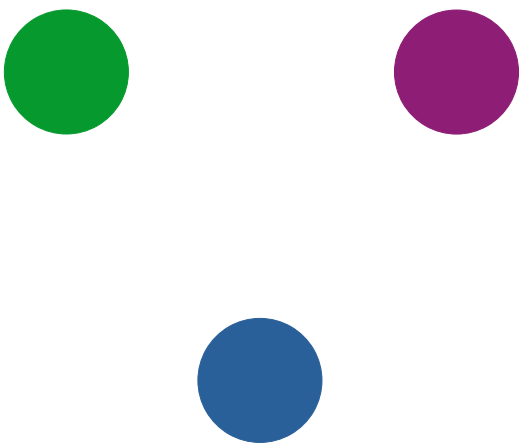
Structured coalescent in action



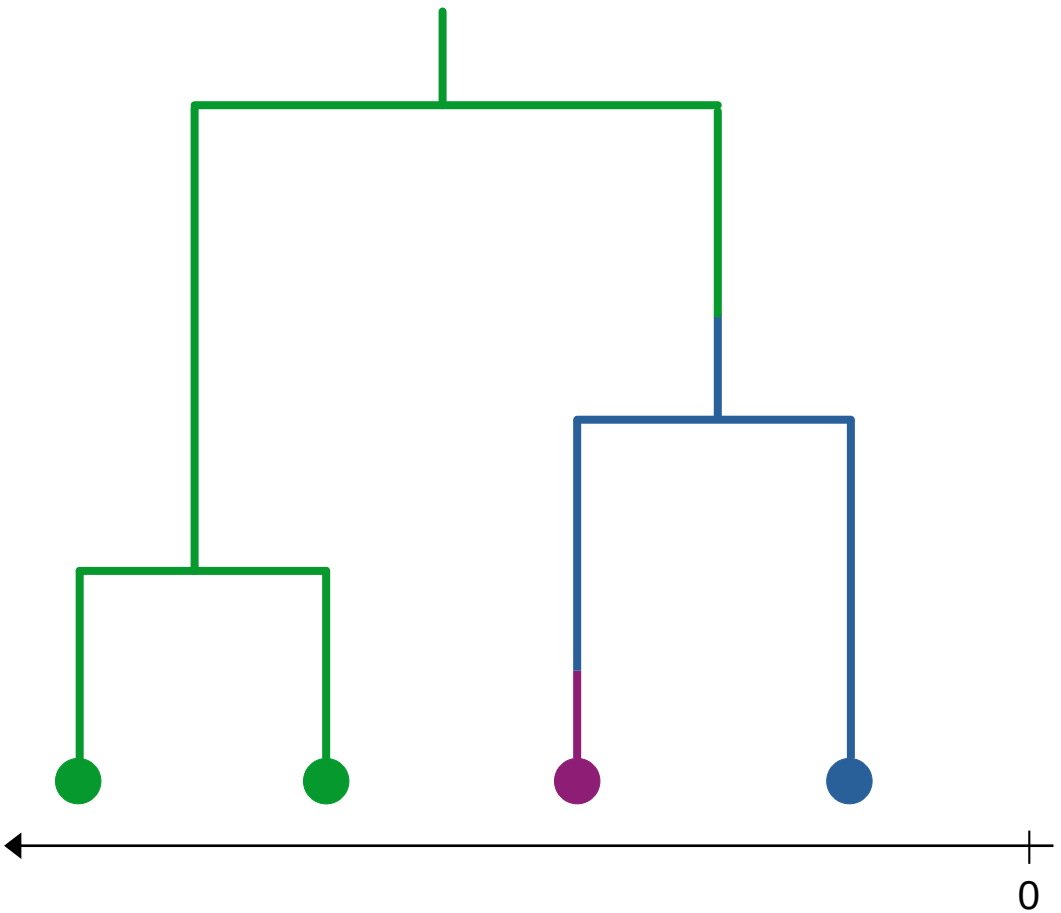


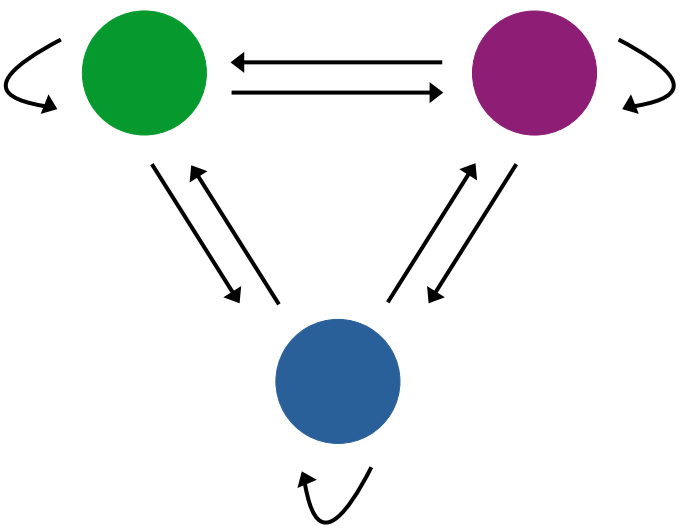
Structured coalescent in action





Structured coalescent in action



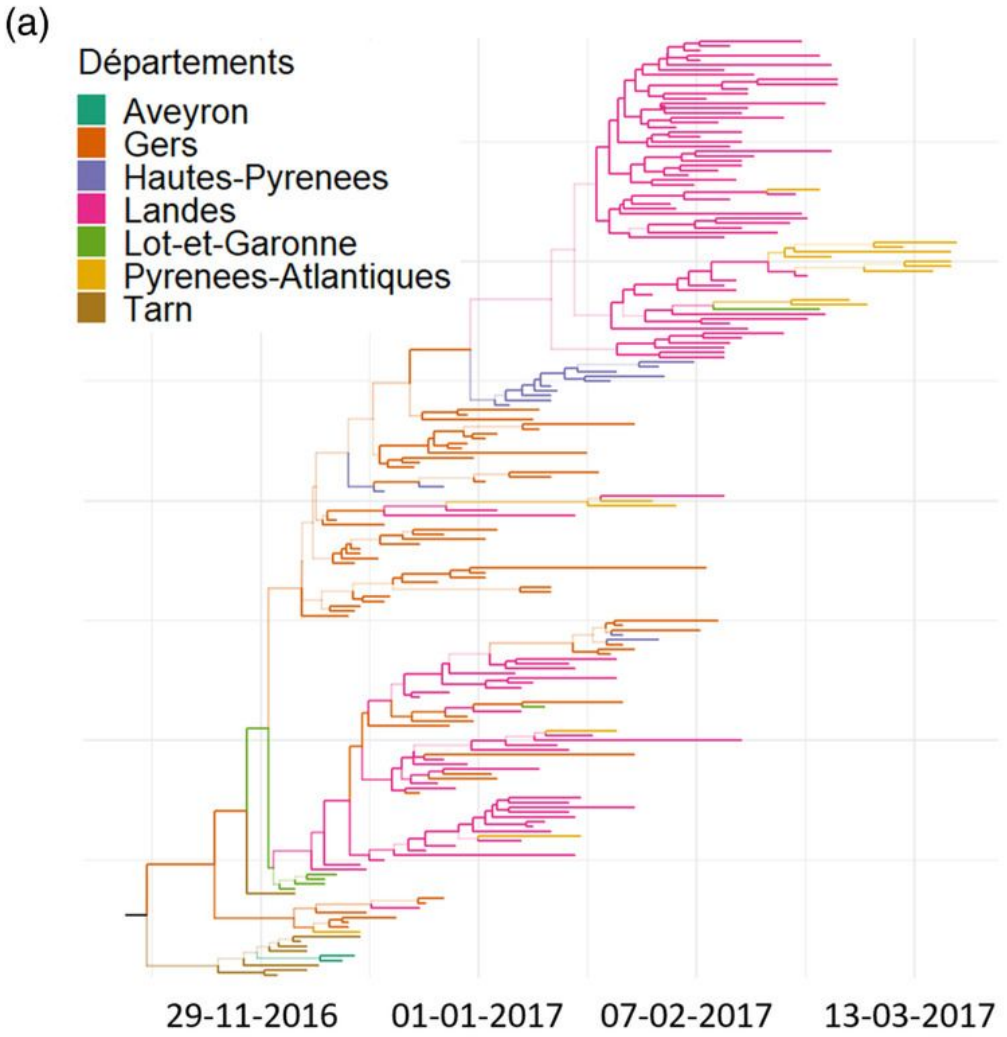


Example: source of a local outbreak

Outbreak of Influenza A/H5N8 in the south of France 2016-2017

Tarn is inferred as the most likely origin of the outbreak

Chakraborty *et al.* **TBED 2022**



In BEAST2

- MultiTypeTree (MTT): exact structured coalescent

  Vaughan *et al.* **Bioinformatics 2014**

- Structured COalescent Transmission Tree Inference (SCOTTI): approximate structured coalescent

  De Maio *et al.* **PLoS Genetics 2015**

  - assumes independence of lineages
  - assumes identical population sizes in subpopulations

- Marginal Approximation of the Structured CoalescenT (MASCOT): approximate structured coalescent

  Müller *et al.* **Bioinformatics 2018**

  - assumes independence of lineages

Key points to remember

- Structured models are designed to represent sub-populations within our dataset
- Structured birth-death models
  - Focus on differences in dynamics between sub-populations
  - Can be character-dependent or independent
  - Can be integrated with other BD models (skyline, FBD, etc)
- Structured coalescent models
  - Focus on the lack of interactions between sub-populations
  - Existing implementations are character-dependent

BD vs coalescent: the revenge

- Clearer choice due to differences in assumptions and underlying process
- Both structured models are sensitive to sampling
  - BD relies on defined sampling process
  - Coalescent is sensitive to sampling biases
- Estimating migration rates
  - Disease outbreak: BD is more accurate
  - Endemic disease: both accurate, coalescent more precise

Seidel *et al.* **Epidemics 2024**

Tutorial time

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

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

Structured coalescent (MASCOT)
https://taming-the-beast.org/tutorials/Mascot-Tutorial/