

Fossilised Birth-Death and Total-evidence Dating

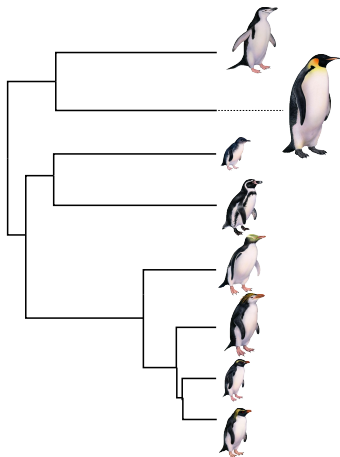
Jana Huisman

adapted from Alexandra Gavryushkina

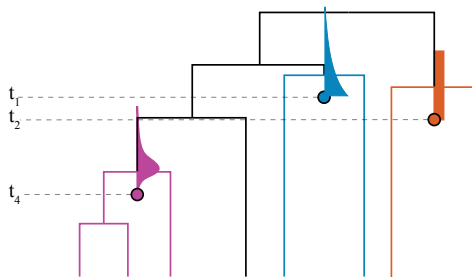
Taming the Beast Down Under

20.2.2019

Dating a phylogenetic tree



Calibration approach



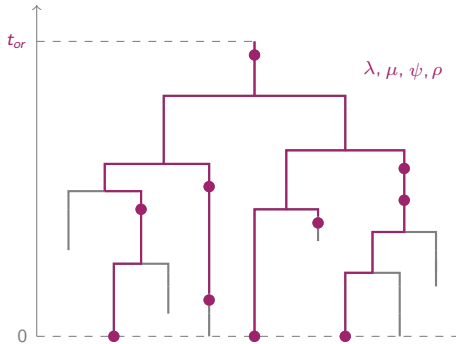
- ▶ Estimate undated phylogeny of extant species from molecular data and extract the topology.
- ▶ Based on morphological traits assign fossils to clades in the phylogeny.
- ▶ Transform the ages of assigned fossils to calibration densities and use them in a separate analysis of the molecular data to estimate the times of all divergencies.

Problems with the inference

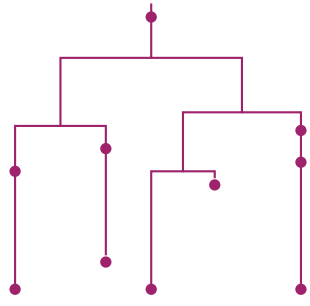
- ▶ Fossils are placed on the phylogeny using **parsimony methods**. Inference from molecular and morphological data may support different topologies. Then **fossil assignment** becomes **ambiguous**.
- ▶ Calibration approach only uses **the oldest fossils** in each clade. The ages of these fossils are directly used to specify the lower bound of a calibration density.
- ▶ Several calibration densities interact with each other and with the tree prior leading to altered prior distributions for the ages of calibration nodes.

Speciation Fossilisation model

Speciation Fossilisation model



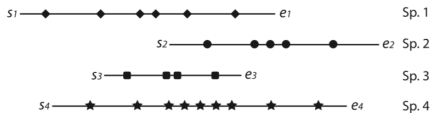
Complete phylogeny



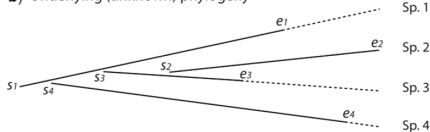
Sampled phylogeny (\mathcal{T})

What does this data tell us?

a) Fossil record and reconstructed times of speciation/extinction



b) Underlying (unknown) phylogeny



Silvestro *et al.* (2014)

- Fossil sampling dates on their own are informative about the times of the events on the phylogeny.
- Parameters of the birth-death model define the distribution of the branch lengths.
- We can estimate parameters: λ, μ, ψ, ρ from **only fossil occurrence data**.

Fossilised birth-death model (FBD)

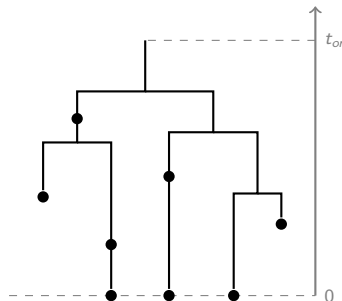
Stadler (2010), Heath *et al.* (2014).

The process starts at time $t_{or} > 0$ and ends at time zero (present time).

- ▶ birth rate λ
- ▶ death rate μ
- ▶ sampling rate ψ
- ▶ sampling at present probability ρ

Model parameters: $\eta = (t_{or}, \lambda, \mu, \psi, \rho)$.

All the parameters are identifiable.



Sampled tree

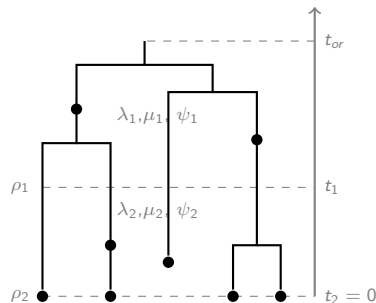
Stadler *et al.* (2012),
Gavryushkina *et al.* (2014).

There are k time intervals and parameters remain constants within the intervals but may vary from one interval to another

- ▶ birth rates $\lambda_1, \dots, \lambda_k$
- ▶ death rates μ_1, \dots, μ_k
- ▶ sampling rates ψ_1, \dots, ψ_k
- ▶ sampling at interval end points probabilities ρ_1, \dots, ρ_k

Model parameters: $\eta = (t_{or}, \bar{\lambda}, \bar{\mu}, \bar{\psi}, \bar{\rho})$

Often $\rho_1 = \dots = \rho_{k-1} = 0$



Sampled tree

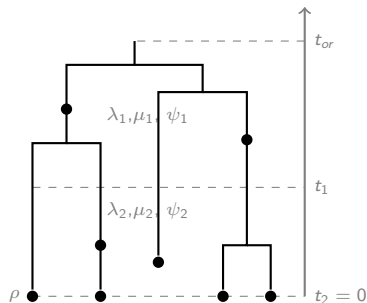
Stadler *et al.* (2012),
Gavryushkina *et al.* (2014).

There are k time intervals and parameters remain constants within the intervals but may vary from one interval to another

- ▶ birth rates $\lambda_1, \dots, \lambda_k$
- ▶ death rates μ_1, \dots, μ_k
- ▶ sampling rates ψ_1, \dots, ψ_k
- ▶ sampling at interval end points probabilities ρ_1, \dots, ρ_k

Model parameters: $\eta = (t_{or}, \bar{\lambda}, \bar{\mu}, \bar{\psi}, \bar{\rho})$

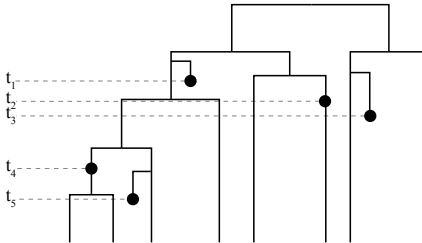
Often $\rho_1 = \dots = \rho_{k-1} = 0$



Sampled tree

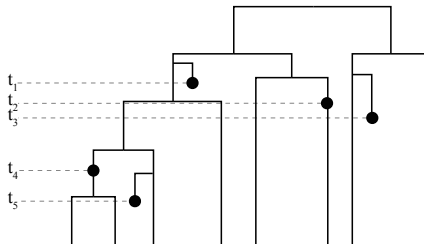
Joint/total-evidence inference

Joint inference



- ▶ Extinct species are a part of the phylogeny
- ▶ Fossil ages can be used to estimate dated phylogeny in a similar way as sampling dates of measurably evolving organisms.

Joint inference



- ▶ Extinct species are a part of the phylogeny
- ▶ Fossil ages can be used to estimate dated phylogeny in a similar way as sampling dates of measurably evolving organisms.
- ▶ Use a model that encompasses the fossil, molecular, and morphological data and then jointly estimate parameters of that model.

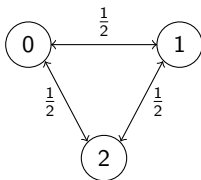


002113????3?0014?70210?...

Ksepka and Clarke (2010)

- ▶ Extant species morphology can help in estimating dated phylogenetic trees.
- ▶ We treat traits as discrete characters.
- ▶ Sequences of morphological characters of fossil and extant species are treated similar to molecular sequences.

Continuous-time Markov chain process with k states: $0, 1, 2, \dots, k - 1$.



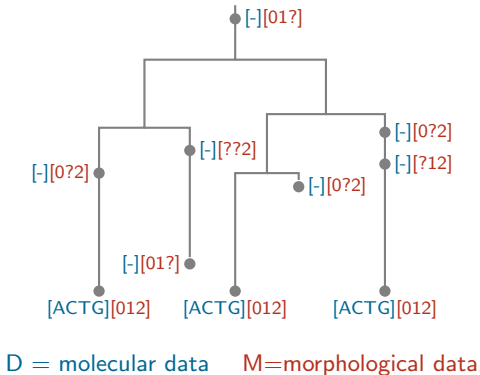
The instantaneous transition rates from any state to any other state are equal:

$$Q_k = \frac{1}{(k-1)} \begin{bmatrix} -(k-1) & 1 & 1 & \dots & 1 \\ 1 & -(k-1) & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & -(k-1) \end{bmatrix}$$

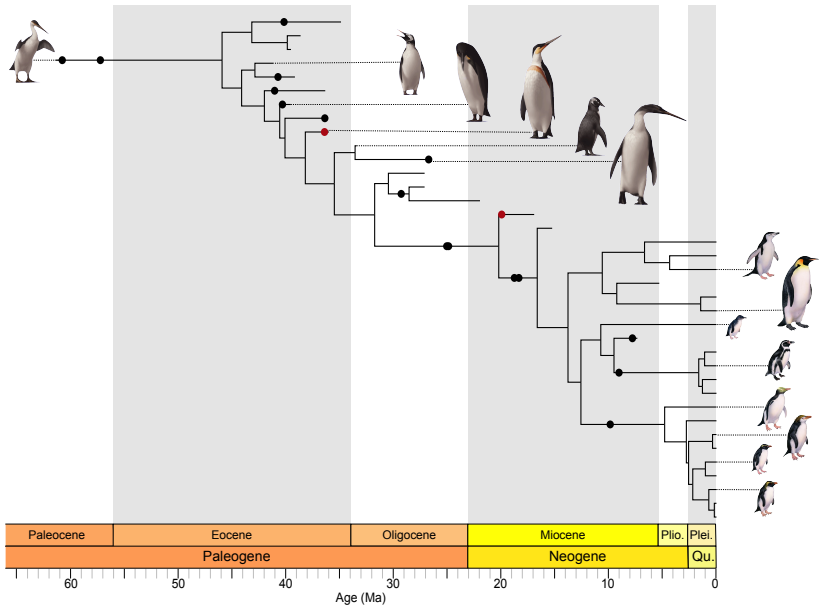
The frequencies of the characters at equilibrium are all equal:

$$\left[\frac{1}{k}, \frac{1}{k}, \dots, \frac{1}{k} \right]$$

Total-evidence analysis of molecular and morphological data



Total evidence analysis of penguins



Characters with different numbers of states

	2 states	3 states	5 states
000001001011010002	00010010110000	12	0
000041011001011002	00010110010100	12	4
000120100110020002	00101001100000	22	2
100030?00110000001	1000?001100000	01	3
??1?????1011?0110	?1?????1011011	?0	?
????1????1?00?0110	???????1?00011	?0	1

- ▶ Characters are partitioned in groups with the same number of states (greater than one).
- ▶ Models with different k are used for each group.

Gavryushkina *et al.* (2017)

<http://www.beast2.org/morphological-models/>

Advantages of the joint inference

- ▶ All available fossil data can be used.
- ▶ This approach does not require an out-group species.
- ▶ It accounts for uncertain fossil placement.
- ▶ It is model based: we directly model the fossil sampling process (e.g., with FBD).

Problems with the joint inference

Problems with the joint inference:

- ▶ Morphological data is very sparse: therefore the fossil sampling model will have a great impact on the results.
- ▶ Models of morphological evolution are very primitive (yet are better than parsimony).

Joint or total-evidence inference requires:

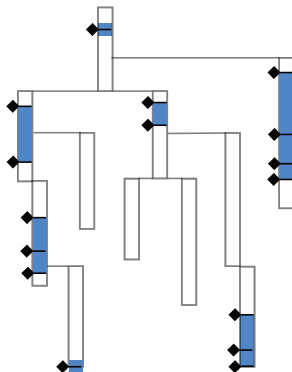
- ▶ dated fossils from several species belonging to the clade (more than five),
- ▶ enough morphological data from both fossil and extant species.

Outlook

Stratigraphic range (SR) birth-death process

Stadler *et al.* (2018)

- ▶ branching rate, λ ,
- ▶ death rate, μ ,
- ▶ fossil sampling rate, ψ ,
- ▶ probability of a symmetric branching event, β ,
- ▶ anagenetic speciation rate, λ_a , and
- ▶ probability of sampling at present, ρ .



New Results

[Comment on this paper](#)

Inferring Species Trees Using Integrative Models of Species Evolution

Huw A Ogilvie, Timothy G Vaughan, Nicholas J Matzke, Graham J Slater, Tanja Stadler, David Welch, Alexei J Drummond

doi: <https://doi.org/10.1101/242875>

This article is a preprint and has not been peer-reviewed [what does this mean?].

Questions?