

Should protractedness be included in speciation models?

Richel Bilderbeek

November 17, 2015

Big questions

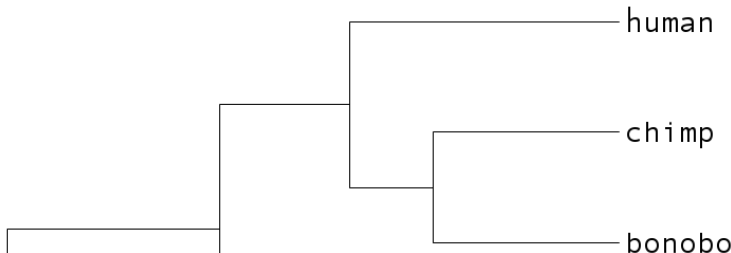
- What is the mechanism of speciation?
- Can we describe and predict the presence/absence of species in time?
- Can we possibly conclude the correct model from nature?

Smaller questions

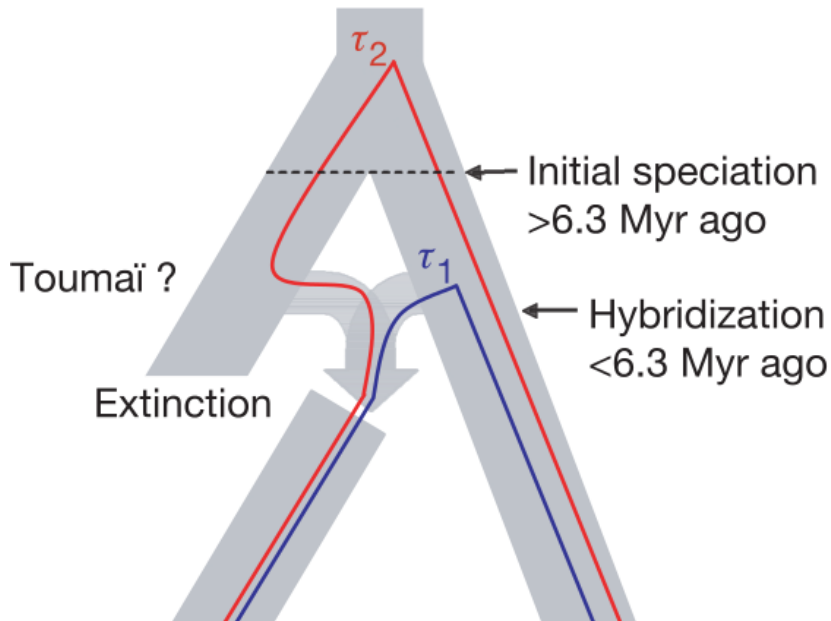
- How are species related to each other?
- When were which species present?
- When were which species being formed?

Smaller questions

- How are species related to each other?
- E.g. humans are closer related to chimps than to gorillas
- When were which species present?
- E.g. humans split of from chimps 6-7 Ma [1] or 5.4 Ma [2]
- When were which species being formed?
- E.g. humans and chips may have exchanged genes for 1.2 Myr [2]
- E.g. the first human species are recognized at 1.78-1.95 Ma [3]



Big questions



Is it warranted to add protractedness to speciation models?

Birth-death (BD) speciation model

- Assumes each species has a constant rate to cause a speciation and extinction
- Parameters:
 - speciation rate
 - extinction rate
- Assumes speciation is instantaneous
- Creates monophyletic phylogenies

[1] Nee et al., 1994

Problems with BD model

- LTT plot: exponential increase in number of lineages [1]

$$E(N) = x \quad [1]$$

- Paraphylies are common in nature [2]

[1] Nee et al., 1994 [2] Omland & Funk

Solutions: other speciation models

- Time-dependent BD models
- Diversity-dependent BD models
- Protracted BD models

Protracted speciation

- Extension of BD model
- For high speciation completion rates, PBD falls back to BD model
- Assumes speciation takes time
- Embraces paraphylies

Effect of PS on LTT plot

- gene tree
- species tree

Parameter recovery by ML by Etienne

Drawbacks of ML: no uncertainty

Alternative: Bayesian parameter estimation

- Phylogeny and parameters are jointly inferred
- Uncertainty of estimates

[1] Janzen: joint inference does not outperform two-step

- Heavily used Bayesian phylogenetic tool
- Input: DNA alignment and priors
- Output: posterior with a distribution of phylogenies and jointly-inferred
- Has implemented BD, but not PBD
- Will recover correct edge lengths, given enough data 'Let the data speak for itself'

When does BEAST2 already suffice to recover a (simulated) protracted phylogeny using a BD tree prior?

- When is the data strong enough?
- Is this biologically relevant?
- Thus: should this be added to BEAST2?

Experiment

Parameters used: * From [1]

For a parameters combination [2]:

- 1/5 Create a simulated 'true' protracted phylogeny (with possible paraphylies)
- 2/5 Sample a random (monophyletic) species tree, add an outgroup
- 3/5 Convert this to a DNA alignment
- 4/5 Let BEAST2 infer a posterior
- 5/5 Measure distance of the sampled 'true' phylogeny with the posteriors

[1] Etienne [2] Replicates omitted

1/5 Create a simulated 'true' protracted phylogeny (with possible paraphylies)

- Use PBD R package []

[Etienne]

2/5 Sample a random (monophyletic) species tree, add an outgroup

3/5 Convert this to a DNA alignment

Use simSeq R package []

[]

4/5 Let BEAST2 infer a posterior

5/5 Measure distance of the sampled 'true' phylogeny with the posteriors

Conclusion

- Unknown if PBD needs to be added to e.g. BEAST2 to recover edge lengths
- Adding this to BEAST also allows to jointly infer parameters, especially the estimation of duration of speciation

- Simple DNA substitution model: cleaner or too simple?

?constant rate clock?