

TECE progress meeting

Richèl JC Bilderbeek

University of Groningen

2017-09-13

<http://github.com/richelbilderbeek/Science>

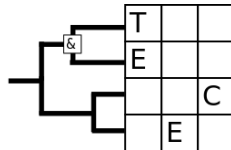
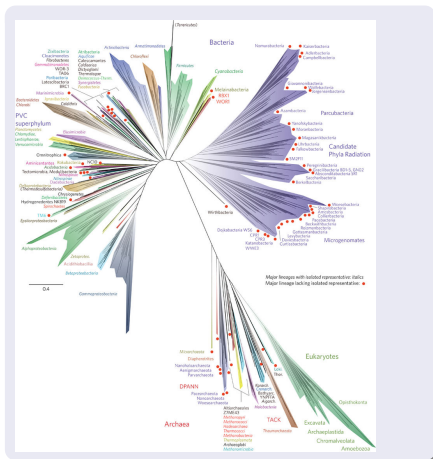


Table of Contents

- 1 Problem
- 2 Research question
- 3 Introduction
- 4 Preliminary results
- 5 Preliminary conclusions
- 6 Work in progress

Tree of life²

Uses RAxML¹



¹Stamatakis, Alexandros. 2014. Bioinformatics 30.9 (2014).

²Hug, Laura A., et al. 2016. Nature Microbiology 1 (2016).

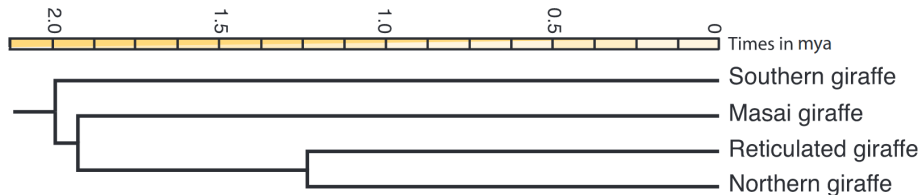
New giraffe species ... ³⁴



³Fennessy, Julian, et al. 2016. Current Biology 26.18.

⁴Picture by Julian Fennessy

... for quite some time⁵



⁵Fennessy, Julian, et al. 2016. *Current Biology* 26.18.

Table of Contents

- 1 Problem
- 2 Research question**
- 3 Introduction
- 4 Preliminary results
- 5 Preliminary conclusions
- 6 Work in progress

Research question

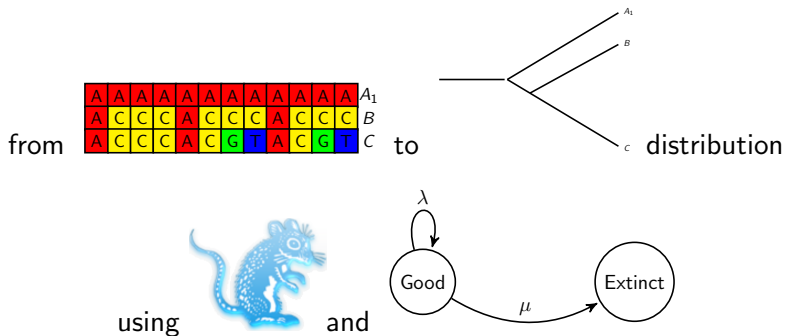
What is the error we make today ...

- in inferring phylogenies
- caused by ignoring that speciation takes time to complete?

Table of Contents

- 1 Problem
- 2 Research question
- 3 Introduction**
- 4 Preliminary results
- 5 Preliminary conclusions
- 6 Work in progress

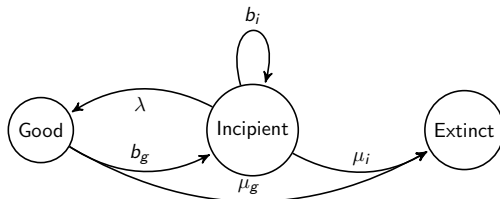
Common: from DNA to phylogeny⁶⁷



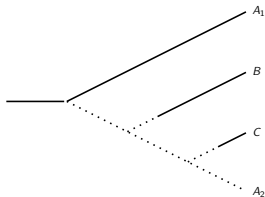
⁶Bouckaert, Remco, et al. 2014. PLoS computational biology 10.4.

⁷Nee, Sean, et al. 1994. Philos Trans R Soc Lond B Biol Sci 344.1309.

Novel: simulating DNA⁸⁹



using



simulate

and

A	A	A	A	A	A	A	A	A	A	A	A	A ₁
A	C	C	C	A	C	C	C	A	C	C	C	B
A	C	C	C	A	C	G	T	A	C	G	T	C

⁸Etienne, Rampal S., and James Rosindell. 2012. Systematic Biology 61.2.

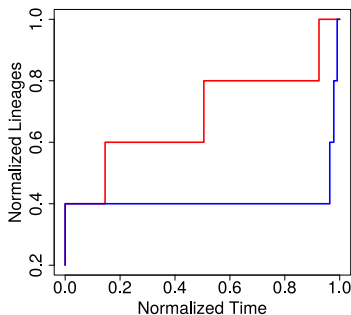
⁹Schliep, Klaus Peter. 2010. Bioinformatics btq706.

Measuring error using nLTT statistic¹⁰

compare 'true' tree



with inferred tree



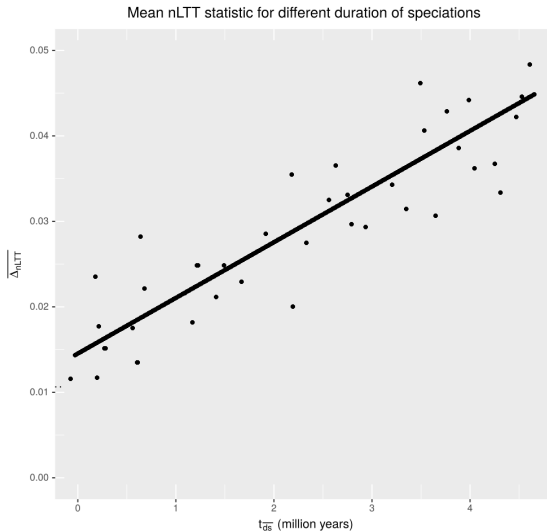
using

¹⁰ Janzen, Thijs et al. 2015. Methods in Ecology and Evolution 6.5.

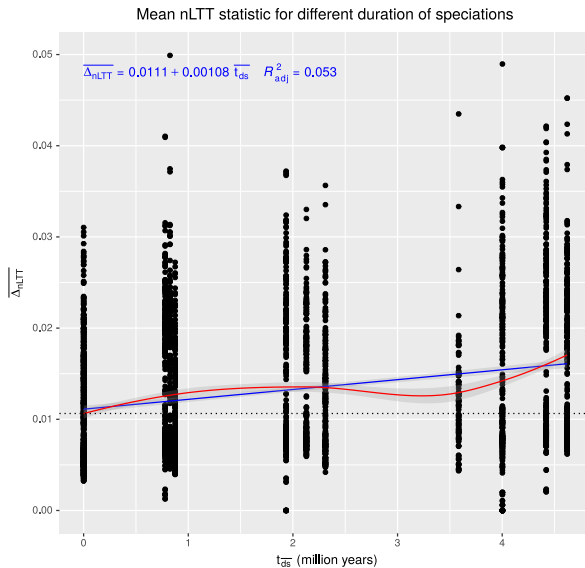
Table of Contents

- 1 Problem
- 2 Research question
- 3 Introduction
- 4 Preliminary results**
- 5 Preliminary conclusions
- 6 Work in progress

Longer duration of speciation



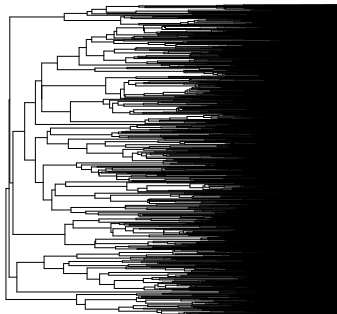
Longer duration of speciation



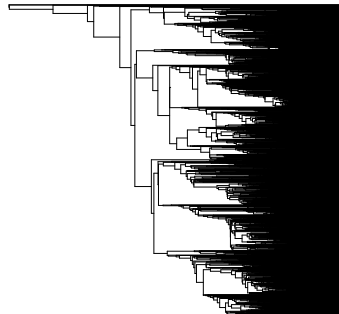
figure_error_expected_mean_dur_spec_mean

Effect of bad estimation

original species tree



posterior tree



Longer DNA alignments

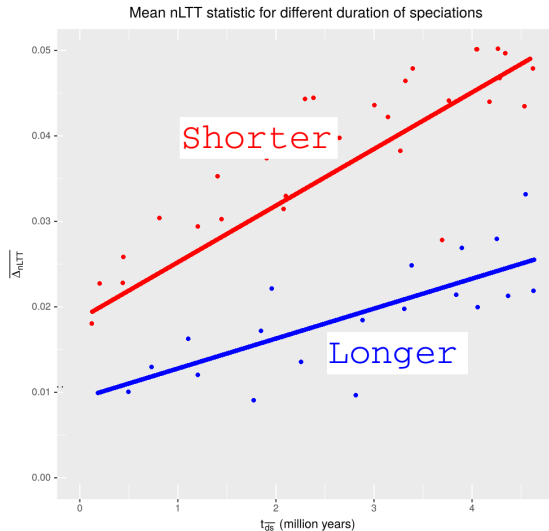


Figure 10 is a scatter plot showing the relationship between the time delay t_{ds} (in million years) on the x-axis and the normalized LTT semi-major axis Δ_{nLTT} on the y-axis. The plot includes two data series: one for $i_a = 1000$ (red) and one for $i_a = 10000$ (teal). Both series show a positive correlation, with fitted linear regression lines and R^2_{adj} values.

The fitted equations and R^2_{adj} values are:

- For $i_a = 1000$ (red): $\overline{\Delta_{nLTT}} = 0.0163 + 0.00158 \overline{t_{ds}}$, $R^2_{adj} = 0.12$
- For $i_a = 10000$ (teal): $\overline{\Delta_{nLTT}} = 0.00609 + 0.000609 \overline{t_{ds}}$, $R^2_{adj} = 0.14$

The y-axis ranges from 0.000 to 0.100, and the x-axis ranges from 0 to 4.5 million years. A legend on the right indicates the color coding for i_a .

Navigation icons: back, forward, search, and other controls.

Table of Contents

- 1 Problem
- 2 Research question
- 3 Introduction
- 4 Preliminary results
- 5 Preliminary conclusions**
- 6 Work in progress

Preliminary conclusions

It is shown that:

- ① \uparrow protractedness leads to \uparrow error
- ② \downarrow DNA sequence leads to \uparrow error
- ③ (not shown) sampling subspecies has no effect
- ④ (not shown) number of taxa has no effect

if $t \geq 16 : 00$, then done

😊 THANKS! 😊

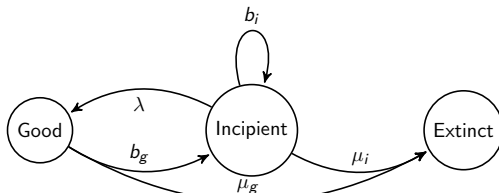
Table of Contents

- 1 Problem
- 2 Research question
- 3 Introduction
- 4 Preliminary results
- 5 Preliminary conclusions
- 6 Work in progress**

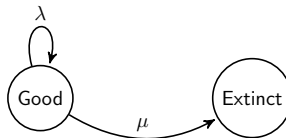
Current problem

Incomplete documentation

Parameter estimation



given



then for $\lambda \rightarrow \infty$, becomes

- λ : speciation rate
- μ : extinction rate

Parameter estimation



takes a

A	A	A	A	A	A	A	A	A	A	A	A	A	A_1
A	C	C	C	A	C	C	C	A	C	C	C	C	B
A	C	C	C	A	C	G	T	A	C	G	T	C	C

and creates

a distribution of $\{\mathcal{T}, \hat{\lambda}, \hat{\mu}, \hat{t}_0\}$

- \mathcal{T} : a phylogeny
- $\hat{\lambda}$: estimated speciation rate
- $\hat{\mu}$: estimated extinction rate
- \hat{t}_0 : estimated crown age

Naive expectation

Parameters to create phylogenies:

λ	speciation rate	0.3
μ	extinction rate	0.2
t_0	crown age	15



Expected parameter estimates:

$\hat{\lambda}$	estimated speciation rate	0.31	0.29	0.28
$\hat{\mu}$	estimated extinction rate	0.19	0.21	0.2
\hat{t}_0	estimated crown age	14.9	15.1	15.2

Real results

Parameters to create phylogenies:

λ	speciation rate	0.3
μ	extinction rate	0.2
t_0	crown age	15

Parameter estimates:

birthRate2	18.591	23.7402	16.1756	22.7706	24.0203
relativeDeathRate2	0.69492	0.3817	0.60621	0.49113	0.29777
TreeHeight	0.1515	0.14719	0.14579	0.15033	0.1514

relativeDeathRate2

Parameters to create phylogenies:

λ	speciation rate	0.3
μ	extinction rate	0.2
t_0	crown age	15

Parameter estimates:

b	birthRate2	18.59	23.740	16.176	22.771	24.020
d	relativeDeathRate2	0.6949	0.382	0.6062	0.4911	0.2978
z_0	TreeHeight	0.152	0.1472	0.1458	0.1503	0.151

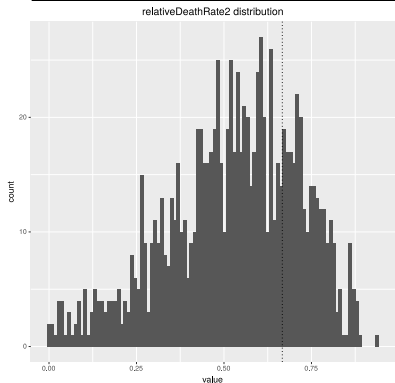
Hypothesis:

$$d = \frac{\mu}{\lambda} \approx 0.67$$

relativeDeathRate2

```
@Description("Birth Death model based on Gernhard 2008. <br/>" +
  "This derivation conditions directly on fixed N taxa. <br/>" +
  "The inference is directly on b-d (strictly positive) and d/b (constrained in [0,1)) <br/>" +
  "Verified using simulated trees generated by Klass tree sample. (http://www.klaashartmann.com/treesa)")
```

```
final public Input<RealParameter> relativeDeathRateParameterInput =
  new Input<>("relativeDeathRate", "relative death rate parameter, mu/lambda in birth death model")
```



~writttte_data/20170710/article_0_3_2_0_0_549.RDa, figure_posterior_distribution_bd_rdr2

birthRate2 and TreeHeight

Parameters to create phylogenies:

λ	speciation rate	0.3
μ	extinction rate	0.2
t_0	crown age	15

Parameter estimates:

b	birthRate2	18.59	23.740	16.176	22.771	24.020
d	relativeDeathRate2	0.6949	0.382	0.6062	0.4911	0.2978
z_0	TreeHeight	0.152	0.1472	0.1458	0.1503	0.151

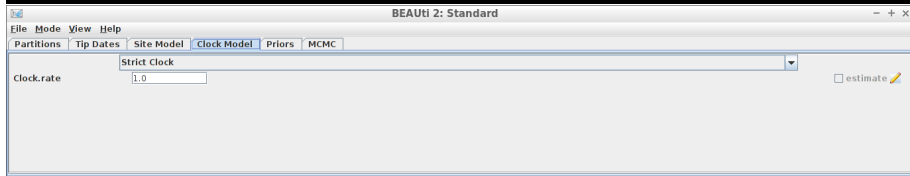
Hypothesis ☹:

$$b = \lambda - \mu = 0.1$$



birthRate2 and TreeHeight

```
@Description("Pure birth model (i.e. no deaths)")
public class YuleModel extends SpeciesTreeDistribution {
    final public Input<RealParameter> birthDiffRateParameterInput =
        new Input<>("birthDiffRate", "birth difference rate parameter, lambda - mu in birth/death model", V
    final public Input<RealParameter> originHeightParameterInput =
        new Input<>("originHeight", "the height of the point of origin of the process");
```



The three parameters λ , μ , ρ are non-identifiable, meaning that the probability density of a time tree is determined by two parameters $\lambda - \mu$ and $\lambda \cdot \rho$ ¹¹.

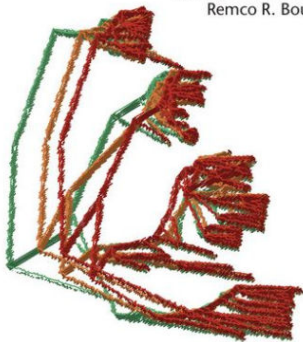
¹¹Stadler, Tanja. 2009. Journal of theoretical biology 261.1

Discussion

- Two out of three estimated parameters are unknown ...
- ... even with access to the BEAST book, online documentation and source code

Bayesian Evolutionary Analysis with BEAST

Alexei J. Drummond and
Remco R. Bouckaert



Thanks

😊 THANKS! 😊