TECE progress meeting

Richèl JC Bilderbeek

University of Groningen

2017-09-13

http://github.com/richelbilderbeek/Science

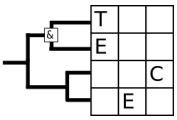
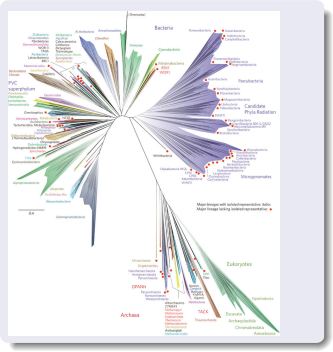


Table of Contents

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

Tree of life[2]



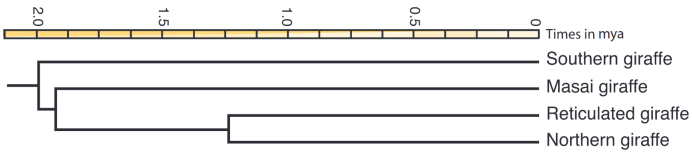
Uses RAxML[1]

[1] Stamatakis, Alexandros. 2014. Bioinformatics 30.9 (2014).
[2] Hug, Laura A., et al. 2016. Nature Microbiology 1 (2016).

New giraffe species ... [3][4]

[3] Fennessy, Julian, et al. 2016. Current Biology 26.18.
[4] Picture by Julian Fennessy

... for quite some time[5]



[5] Fennessy, 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

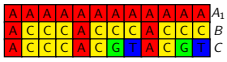
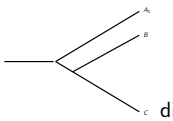
What is the error we make today ...

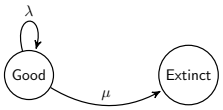
- in inferring phylogenies
- caused by ignoring that speciation takes time to complete?

Table of Contents

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

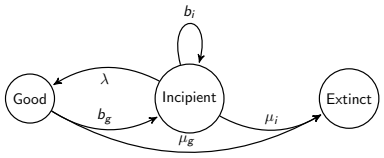
Common: from DNA to phylogeny[6][7]

from  to  distribution

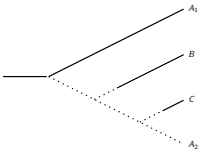
using and 

[6] Bouckaert, Remco, et al. 2014. PLoS computational biology 10.4.
[7] Nee, Sean, et al. 1994. Philos Trans R Soc Lond B Biol Sci 344.1309.

Novel: simulating DNA[8][9]



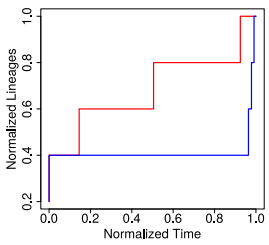
using

simulate  and

[8] Etienne, Rampal S., and James Rosindell. 2012. Systematic Biology 61.2.
[9] Schliep, Klaus Peter. 2010. Bioinformatics btq706.

Measuring error using nLTT statistic[10]

compare 'true' tree with inferred tree



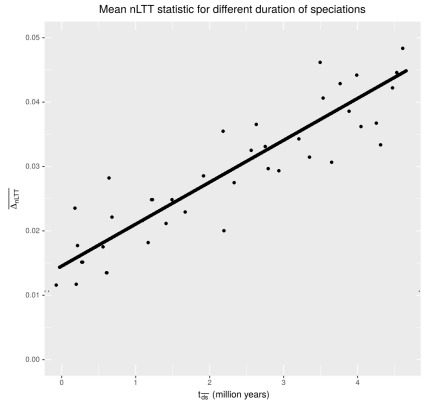
using

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

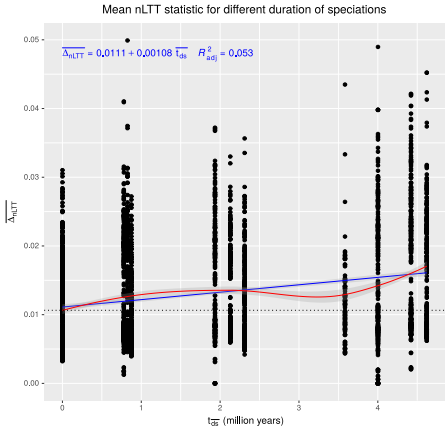
Table of Contents

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

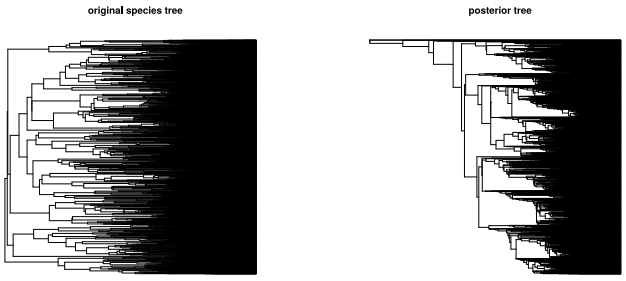
Longer duration of speciation



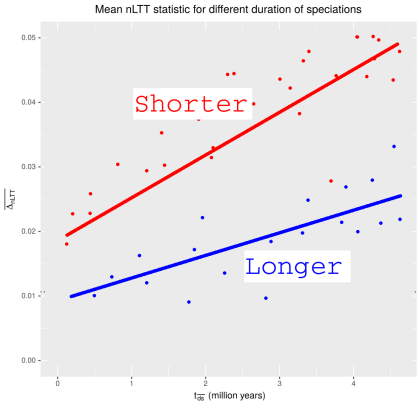
Longer duration of speciation



Effect of bad estimation



Longer DNA alignments



Longer duration of speciation

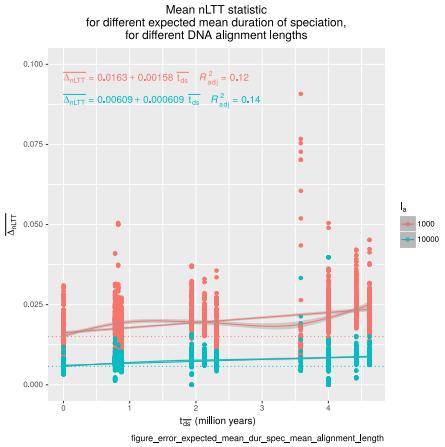


Table of Contents

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

Preliminary conclusions

It is shown that:

1. ↑ protractedness leads to ↑ error
2. ↓ DNA sequence leads to ↑ 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

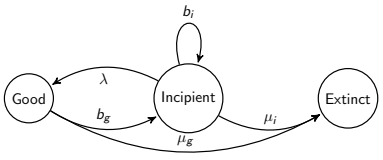
1. Problem
2. Research question
3. Introduction
4. Preliminary results
5. Preliminary conclusions
6. Work in progress

Current problem

Incomplete documentation

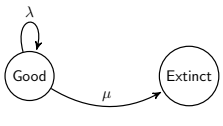
Parameter estimation

given



,

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



- λ: speciation rate
- μ: extinction rate

Parameter estimation

takes a

and creates

a distribution of $\left\{\mathcal{T}, \hat{\lambda}, \hat{\mu}, \hat{t_0}, \right\}$

- $\mathcal{T}$: 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

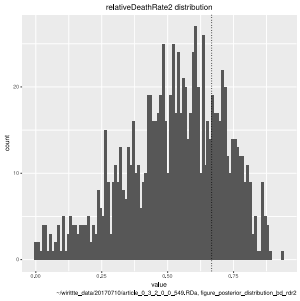
Hypothesis:

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

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"
```



birthRate2 and TreeHeight

Parameters to create phylogenies:

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

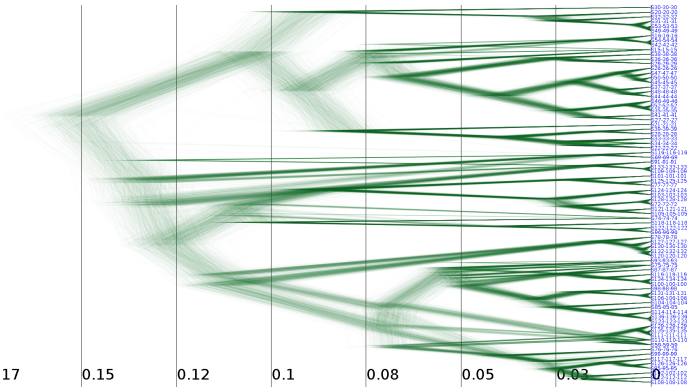
Parameter estimates:

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 ☹:

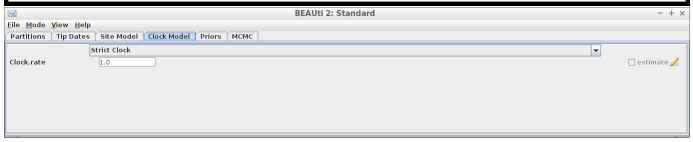
$$b = \lambda - \mu = 0.1$$

TreeHeight



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");
```

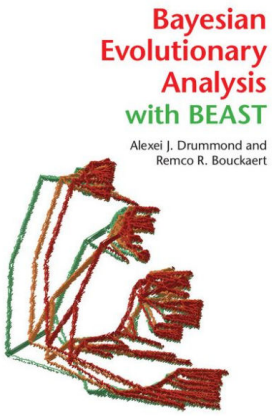


The three parameters λ, μ, ρ are non-identifyable, meaning that the probability density of a time tree is determined by two parameters $\lambda - mu$ and $\lambda \cdot \rho$ [11].

[11] Stadler, 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



Thanks

☺THANKS! ☺