

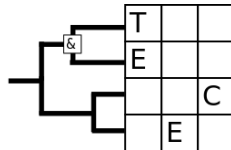
# TECE progress meeting

Richèl JC Bilderbeek

University of Groningen

2017-09-13

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

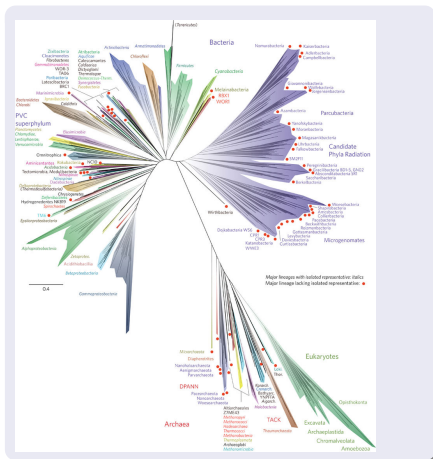


# Table of Contents

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

# Tree of life<sup>2</sup>

Uses RAxML<sup>1</sup>



<sup>1</sup>Stamatakis, Alexandros. 2014. Bioinformatics 30.9 (2014).

<sup>2</sup>Hug, Laura A., et al. 2016. Nature Microbiology 1 (2016).

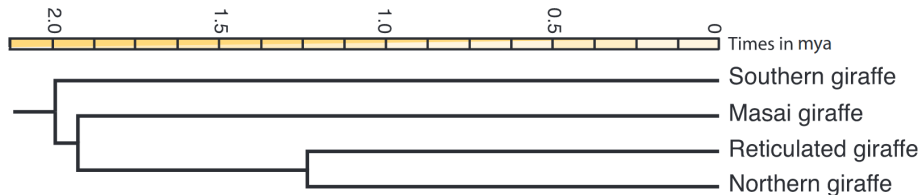
# New giraffe species ... <sup>34</sup>



<sup>3</sup>Fennessy, Julian, et al. 2016. Current Biology 26.18.

<sup>4</sup>Picture by Julian Fennessy

... for quite some time<sup>5</sup>



<sup>5</sup>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 ...

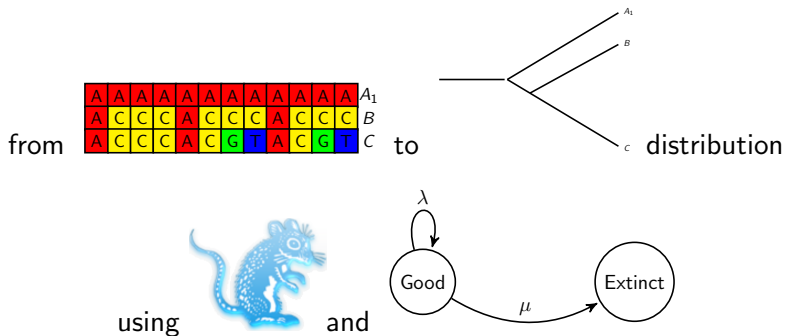
- in inferring phylogenies
- caused by ignoring that speciation takes time to complete?

# Table of Contents

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



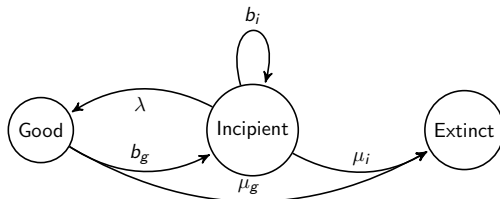
# Common: from DNA to phylogeny<sup>67</sup>



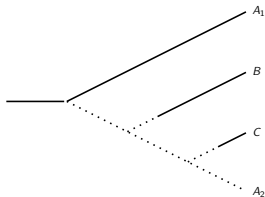
<sup>666</sup>Bouckaert, Remco, et al. 2014. PLoS computational biology 10.4.

<sup>666</sup>Nee, Sean, et al. 1994. Philos Trans R Soc Lond B Biol Sci 344.1309.

# Novel: simulating DNA<sup>89</sup>



using



simulate

and

A	A	A	A	A	A	A	A	A	A	A	A	A	A <sub>1</sub>
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

<sup>666</sup>Etienne, Rampal S., and James Rosindell. 2012. Systematic Biology 61.2.

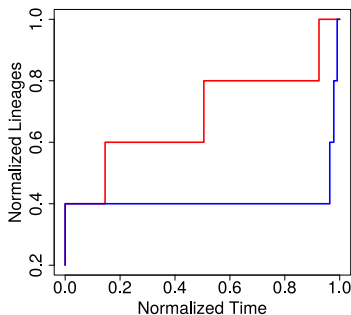
<sup>666</sup>Schliep, Klaus Peter. 2010. Bioinformatics btq706.

# Measuring error using nLTT statistic<sup>10</sup>

compare 'true' tree



with inferred tree



using

<sup>666</sup> 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

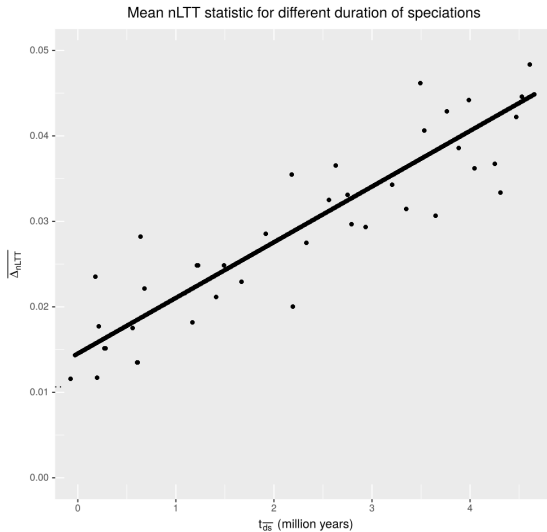
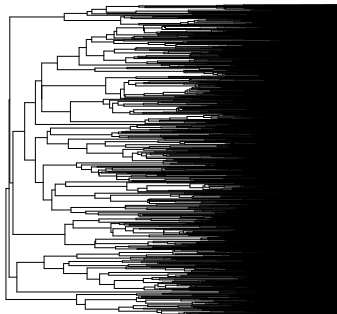


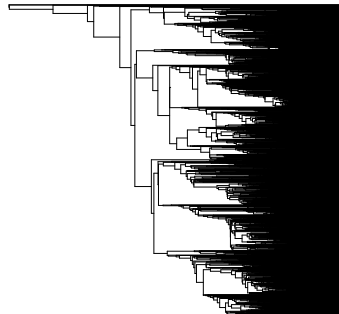
Figure 10 is a scatter plot showing the relationship between the normalized difference in luminosity function,  $\overline{\Delta_{nLTT}}$ , and the time delay,  $t_{ds}$  (in million years). The y-axis ranges from 0.00 to 0.05, and the x-axis ranges from 0 to 4.5. Black dots represent individual data points. A red line shows the linear fit, and a blue line shows the quadratic fit. A horizontal dotted line is at  $\overline{\Delta_{nLTT}} = 0.0111$ . The quadratic fit equation is  $\overline{\Delta_{nLTT}} = 0.0111 + 0.00108 t_{ds}$ , and the adjusted  $R^2$  value is 0.053.

# Effect of bad estimation

original species tree



posterior tree



# Longer DNA alignments

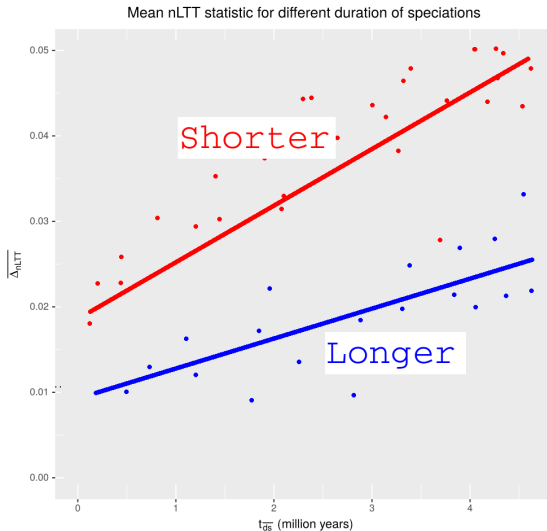




Figure 10 is a scatter plot showing the relationship between the time delay  $t_{ds}$  (in million years) on the x-axis and the normalized LTT semi-major axis  $\Delta_{nLTT}$  on the y-axis. The plot includes two data series: one for  $i_a = 1000$  (red) and one for  $i_a = 10000$  (teal). Both series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

The red series ( $i_a = 1000$ ) has a fitted line:  $\overline{\Delta_{nLTT}} = 0.0163 + 0.00158 \overline{t_{ds}}$  with  $R^2_{adj} = 0.12$ .

The teal series ( $i_a = 10000$ ) has a fitted line:  $\overline{\Delta_{nLTT}} = 0.00609 + 0.000609 \overline{t_{ds}}$  with  $R^2_{adj} = 0.14$ .

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

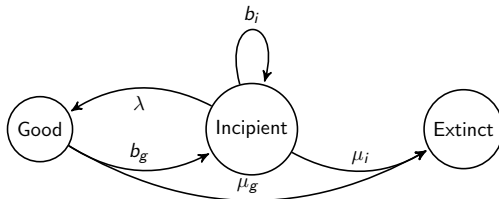
if  $t \leq 16 : 00$ , then done

😊 THANKS! 😊

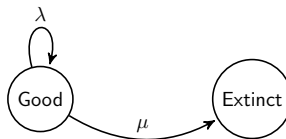
# Table of Contents

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

# Parameter estimation



given



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