TECE progress meeting

Richèl JC Bilderbeek

University of Groningen

2017-09-13
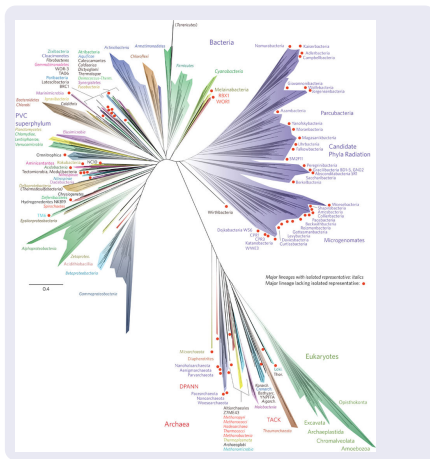
http://github.com/richelbilderbeek/Science

Table of Contents

Tree of life[2]

Uses RAxML[1]

---

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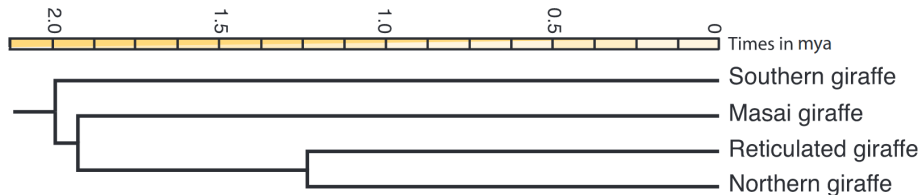[2]Hug, Laura A., et al. 2016. Nature Microbiology 1 (2016).

New giraffe species ... [34]

[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.
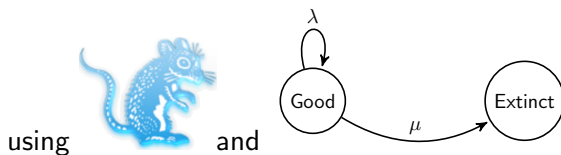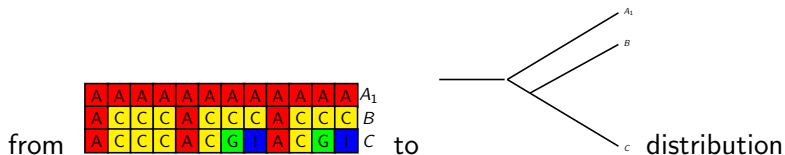
Table of Contents

Research question

What is the error we make today ...

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

Table of Contents

Common: from DNA to phylogeny[67]

from ... to ... distribution

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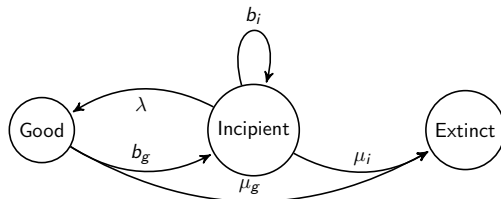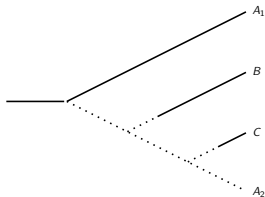
[666] Bouckaert, Remco, et al. 2014. PLoS computational biology 10.4.
[666] Nee, Sean, et al. 1994. Philos Trans R Soc Lond B Biol Sci 344.1309.

Novel: simulating DNA[89]

using

simulate and

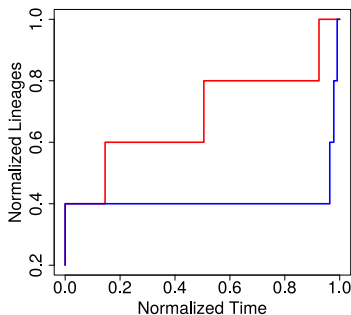[666]Etienne, Rampal S., and James Rosindell. 2012. Systematic Biology 61.2.
[666]Schliep, Klaus Peter. 2010. Bioinformatics btq706.

Measuring error using nLTT statistic[10]

compare 'true' tree   with inferred tree

using

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

Table of Contents

Longer duration of speciation

Mean nLTT statistic for different duration of speciations

Longer duration of speciation

Mean nLTT statistic for different duration of speciations

$\overline{\Delta_{nLTT}} = 0.0111 + 0.00108 \overline{t_{ds}}$ $R^2_{adj} = 0.053$

$\overline{\Delta_{nLTT}}$

$t_{\overline{ds}}$ (million years)

figure_error_expected_mean_dur_spec_mean

Effect of bad estimation

original species tree

posterior tree

Longer DNA alignments
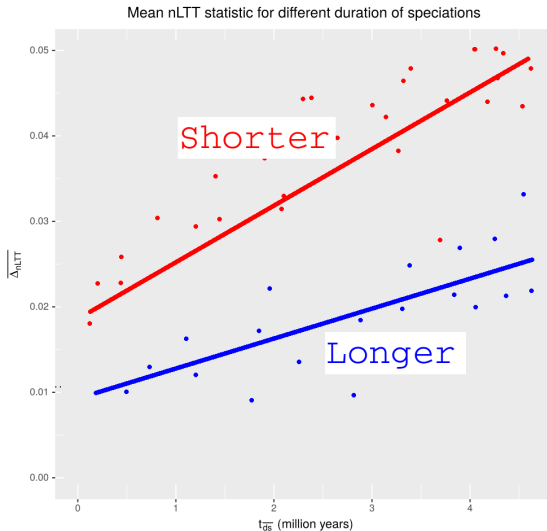
Mean nLTT statistic for different duration of speciations

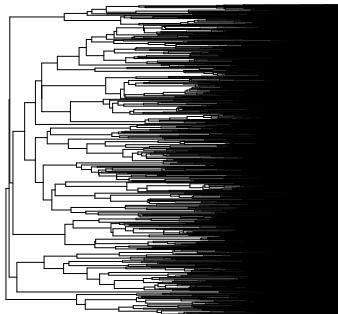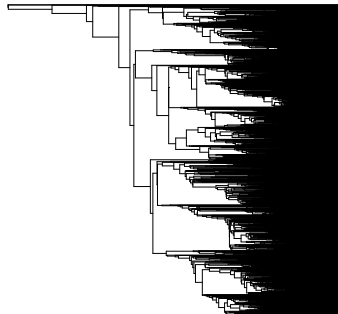# Longer duration of speciation

figure_error_expected_mean_dur_spec_mean_alignment_length

Table of Contents

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

Current problem

Incomplete documentation

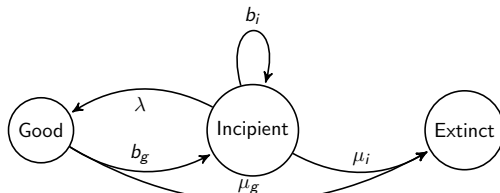# Parameter estimation

given

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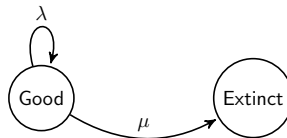
then for $\lambda \to \infty$, becomes

- λ: speciation rate
- μ: extinction rate

Parameter estimation

takes a [sequence alignment] 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

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

(a) Parameters to create phylogenies

$\hat{\lambda}$	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

(b) Expected parameter estimates

Real results

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

(c) Parameters to create phylogenies

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

(d) Parameter estimates