

“Solution to inconsistent sampling of phylogenetic trees”

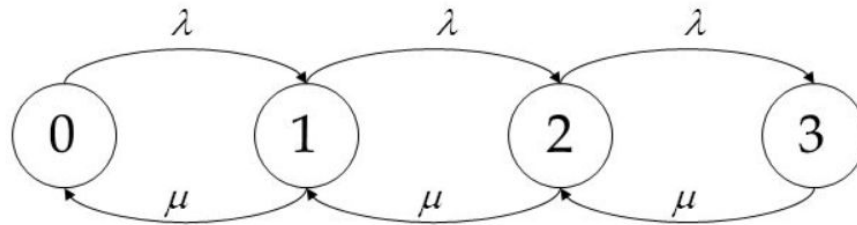


Joris Damhuis
Tom Dekker

Supervisor: Richel Bilderbeek

The birth-death model

- Rate λ at which species give rise to new species
- Rate μ at which species go extinct
- Assumes speciation is instantaneous



The protracted birth-death model

- Introduces incipient species
- Rate λ_1 at which good species give rise to new incipient species
- Rate λ_2 at which incipient species become good species
- Rate λ_3 at which incipient species to give rise new incipient species
- Rate μ_1 at which good species go extinct
- Rate μ_2 at which incipient species go extinct
- Assumes speciation takes time

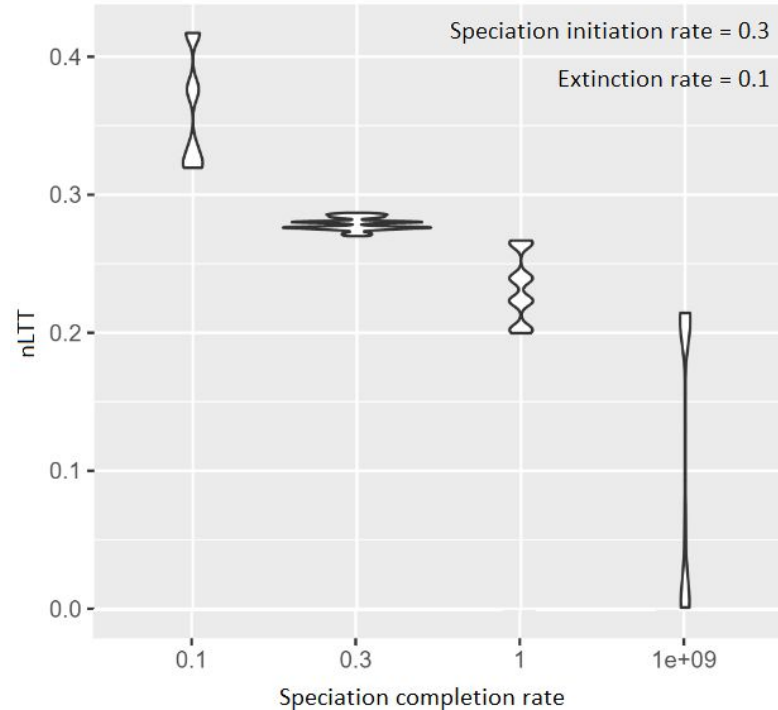
What does raket do?

Pipeline

- Parameters
- Simulation PBD
- Sample good species
- Generate DNA alignment
- Simulate BD tree
- nLTT statistics -> error

Implications of nLTT values

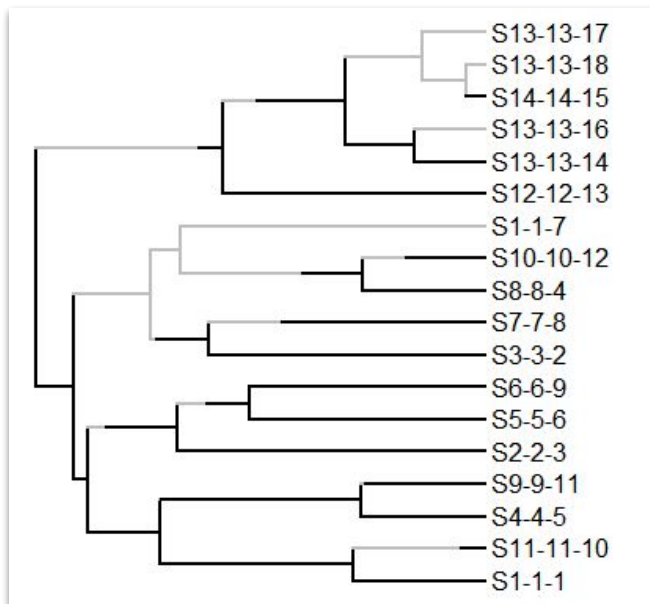
- Difference between models
- Gives indication of the error when estimating number of species



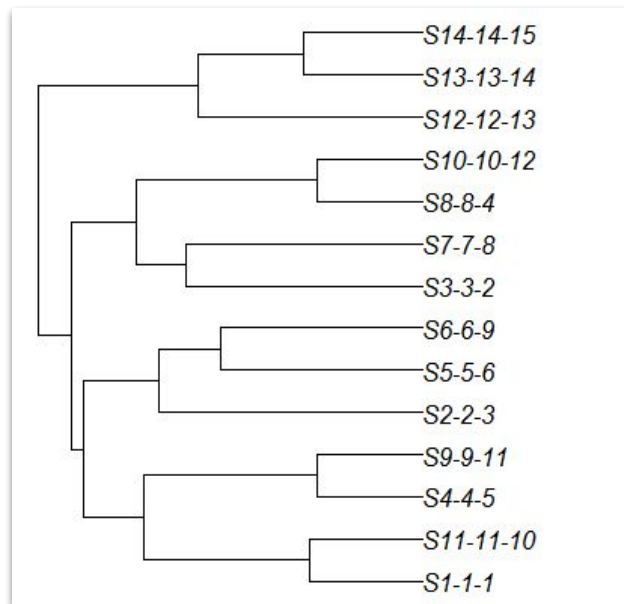
Input file

```
$sirg  
[1] 0.3  
  
$siri  
[1] 0.3  
  
$scr  
[1] 1  
  
$erg  
[1] 0  
  
$eri  
[1] 0  
  
$crown_age  
[1] 15  
  
$crown_age_sigma  
[1] 5e-04  
  
$sampling_method  
[1] "random"  
  
$mutation_rate  
[1] 0.06666667  
  
$sequence_length  
[1] 15  
  
$mcmc  
$mcmc$chain_length  
[1] 4000  
  
$mcmc$store_every  
[1] 1000
```

Incipient species tree

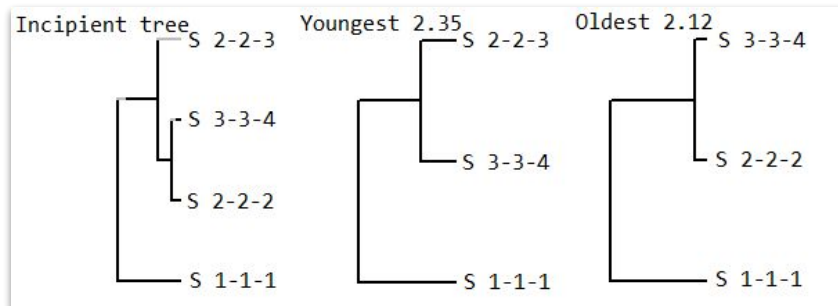


Good species tree

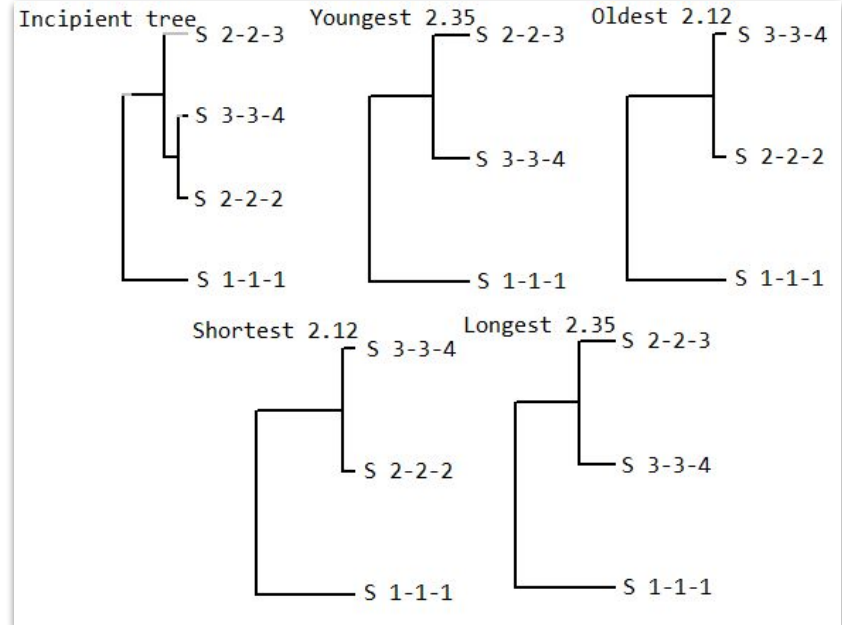
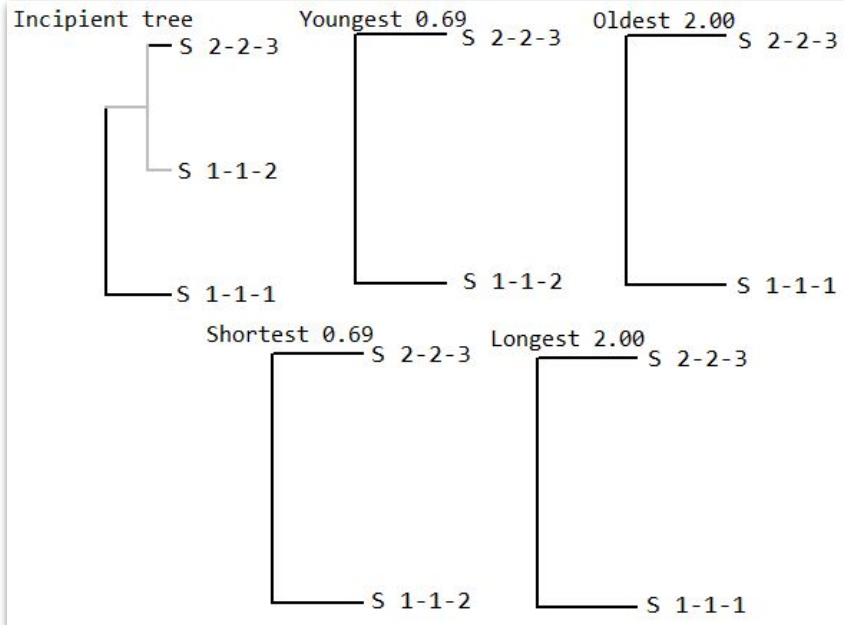


The problem with sampling methods 'youngest' and 'oldest'

- Sample to have 1 representative per species
- Random sampling has sampling error
- $\text{Youngest} \leq \text{Random} \leq \text{Oldest}$

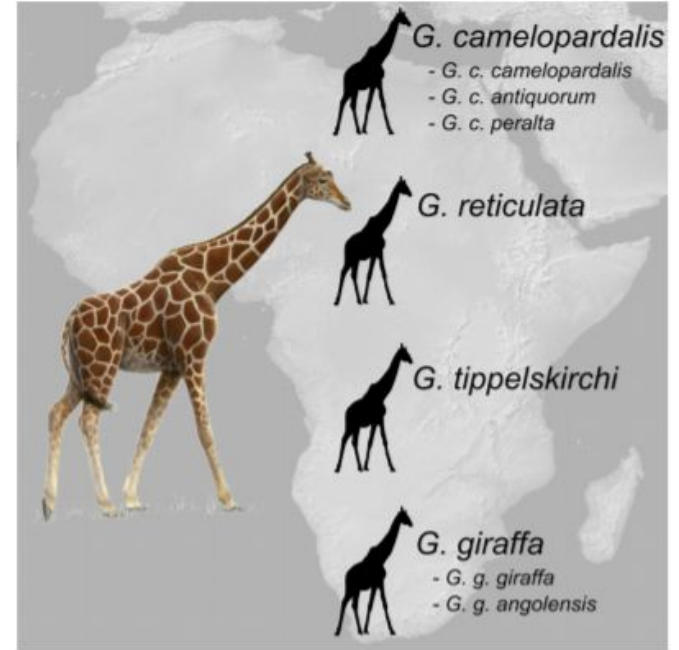


Added sampling methods 'shortest' and 'longest'



Discussion

- Indication of the error when estimating number of species
- Conservation implications



Fennessy, Julian, et al. "Multi-locus analyses reveal four giraffe species instead of one." *Current Biology* 26.18 (2016): 2543-2549