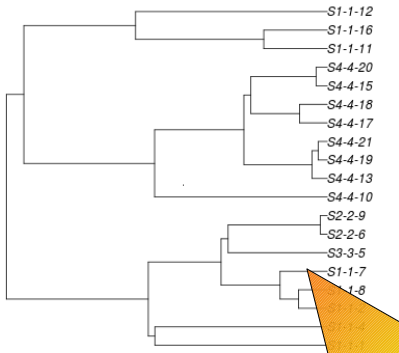# Progress

GitHub

Research question

If speciation in nature takes time, what is the error BEAST2 makes in inferring a phylogeny?

- Under which conditions does it hurt?

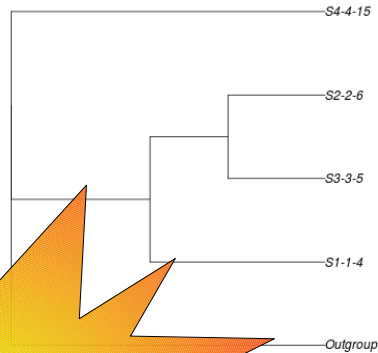- Are these conditions relevant and/or realistic?

Recap

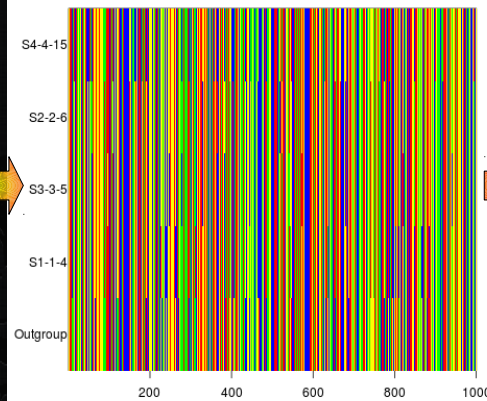**Step with highest impact**
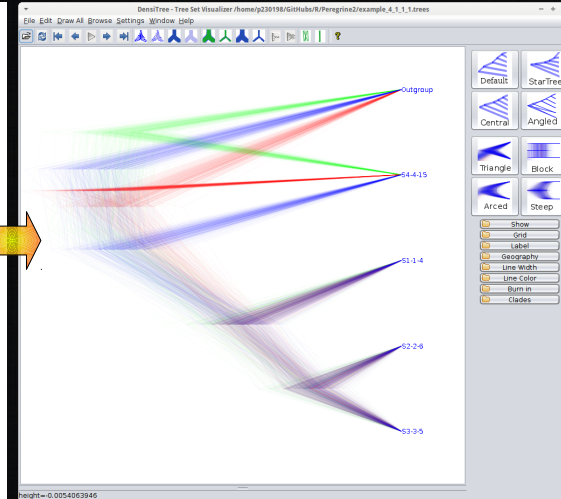
Simulate a true gene tree

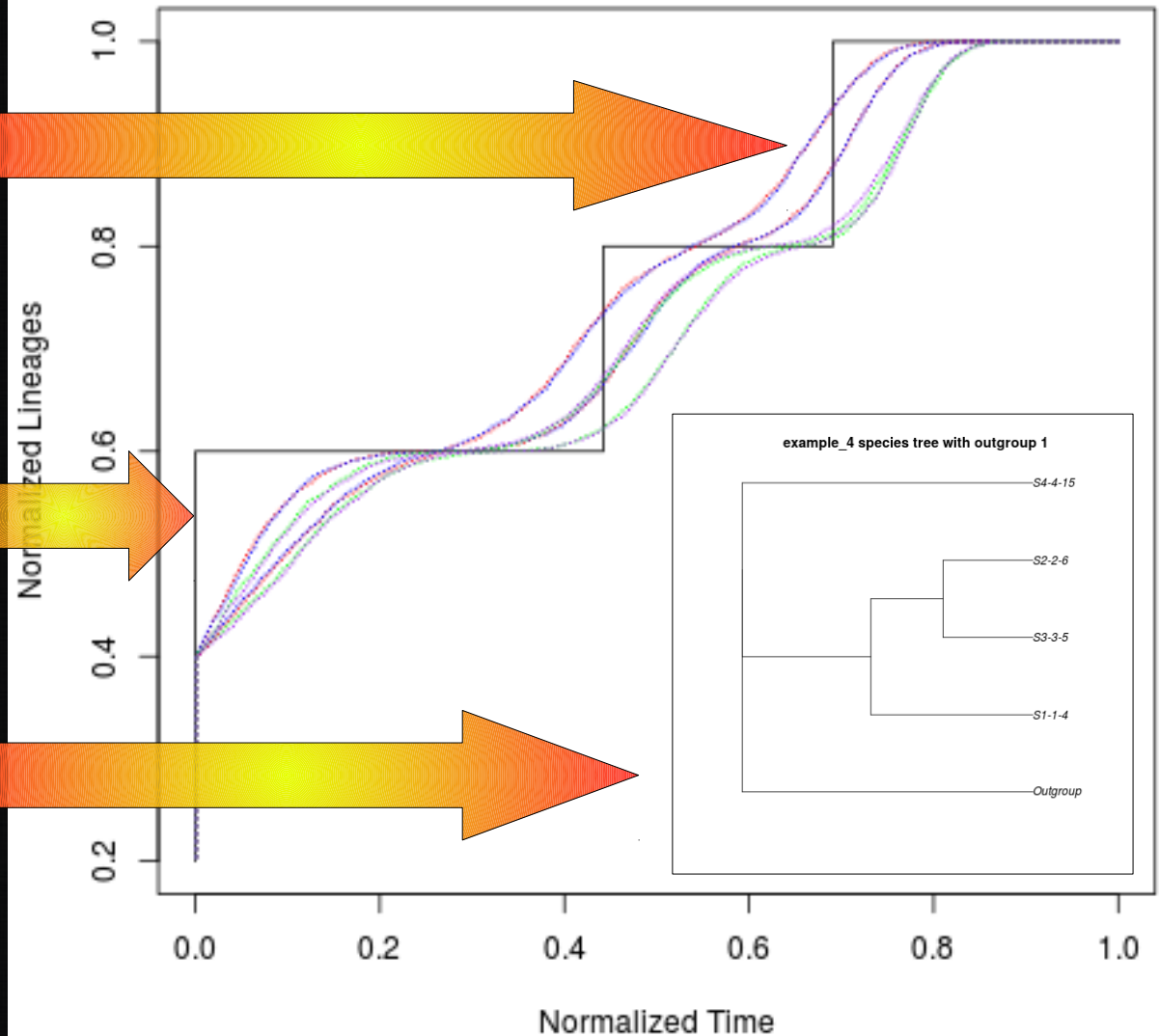Sample a species tree

Simulate a DNA alignment

Run BEAST2 to obtain a posterior
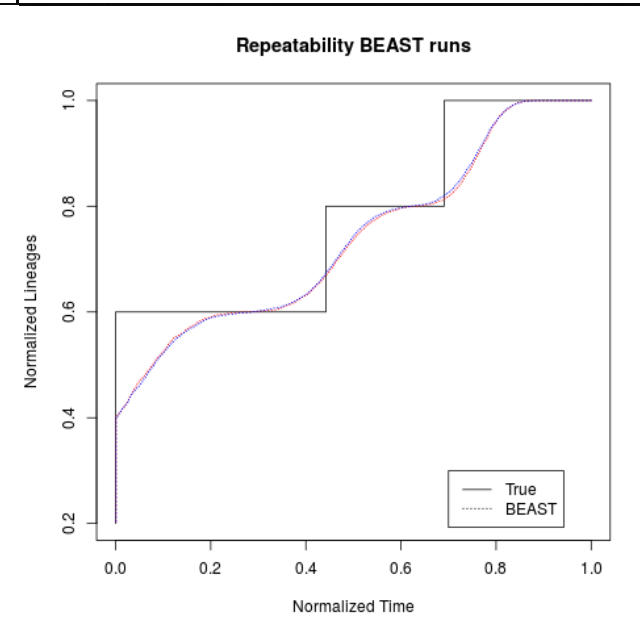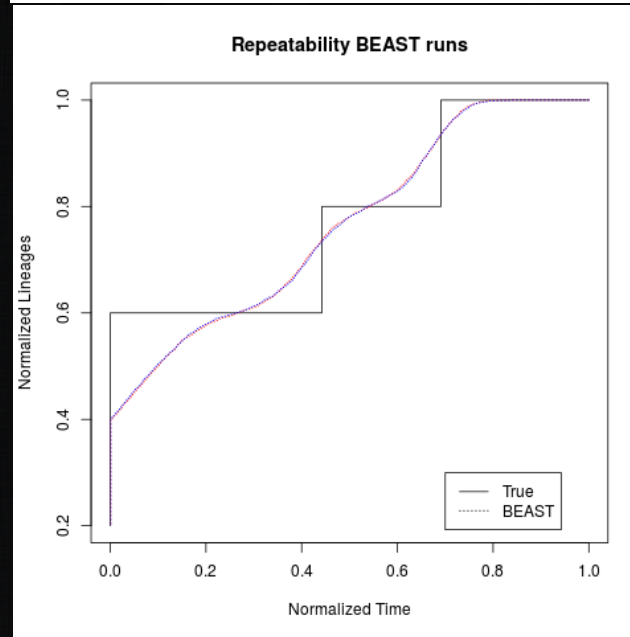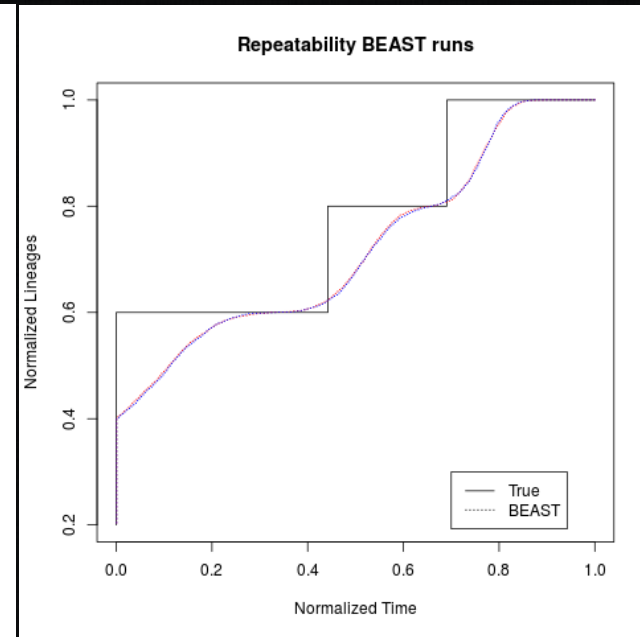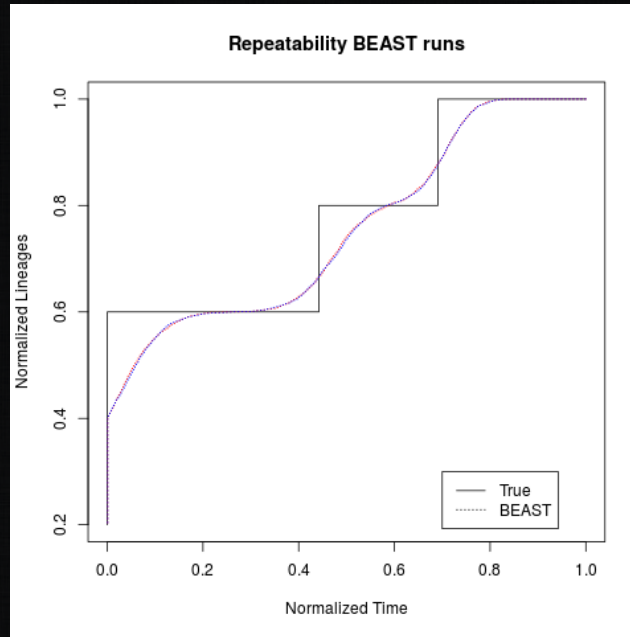
Pipeline works

Average nLTT of posteriors

nLTT of species tree

True species tree

example_4 species tree with outgroup 1

S4-4-15
S2-2-6
S3-3-5
S1-1-4
Outgroup

Normalized Lineages

Normalized Time
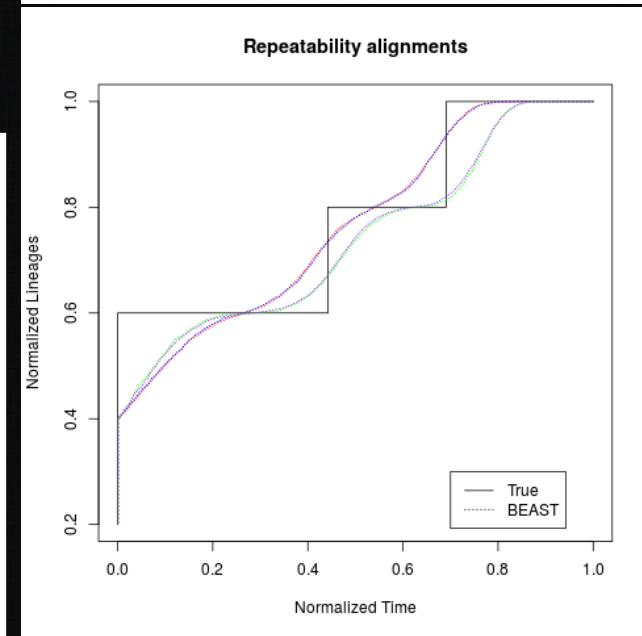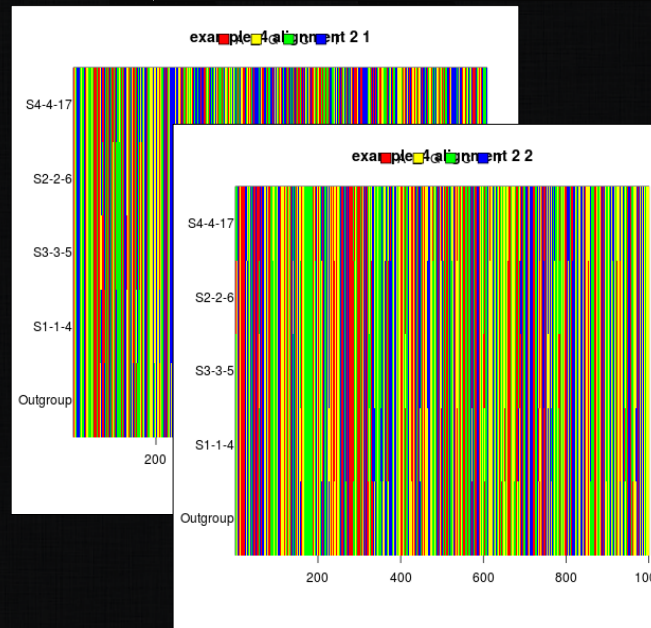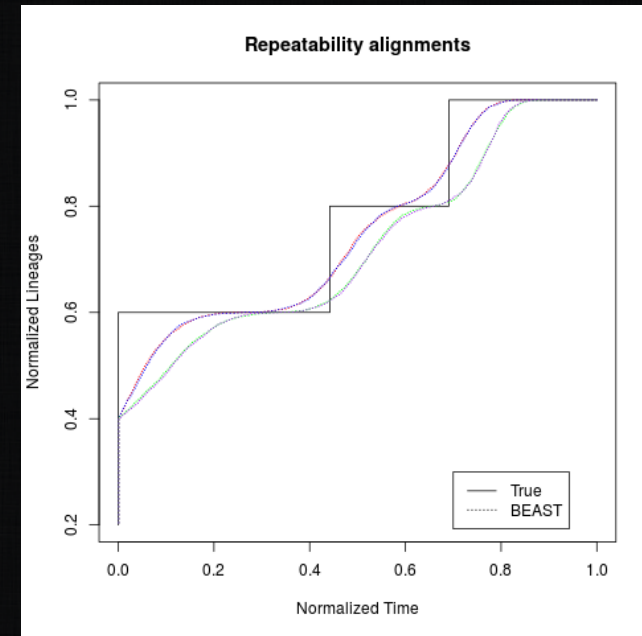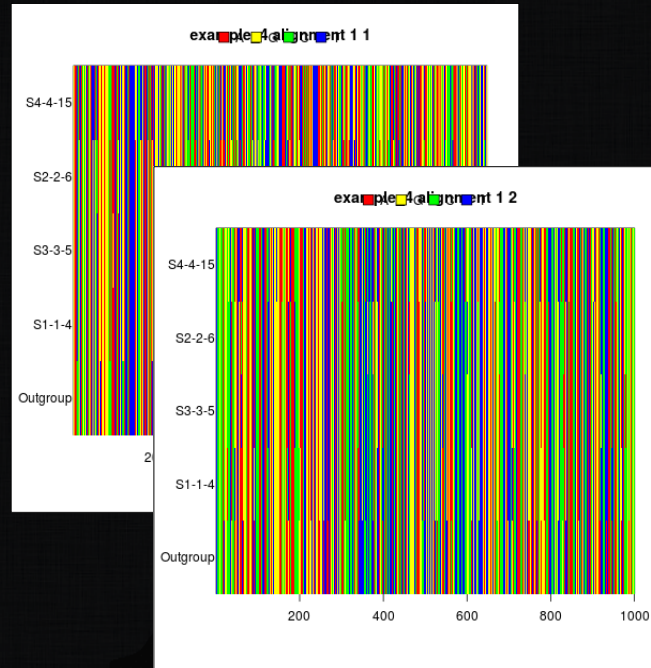
BEAST2 reps are reps

- Two BEAST2 runs on the same alignment give very similar nLTT plots

Alignment reps are less rep

- Two different alignments on the same species tree give similarish nLTT plots
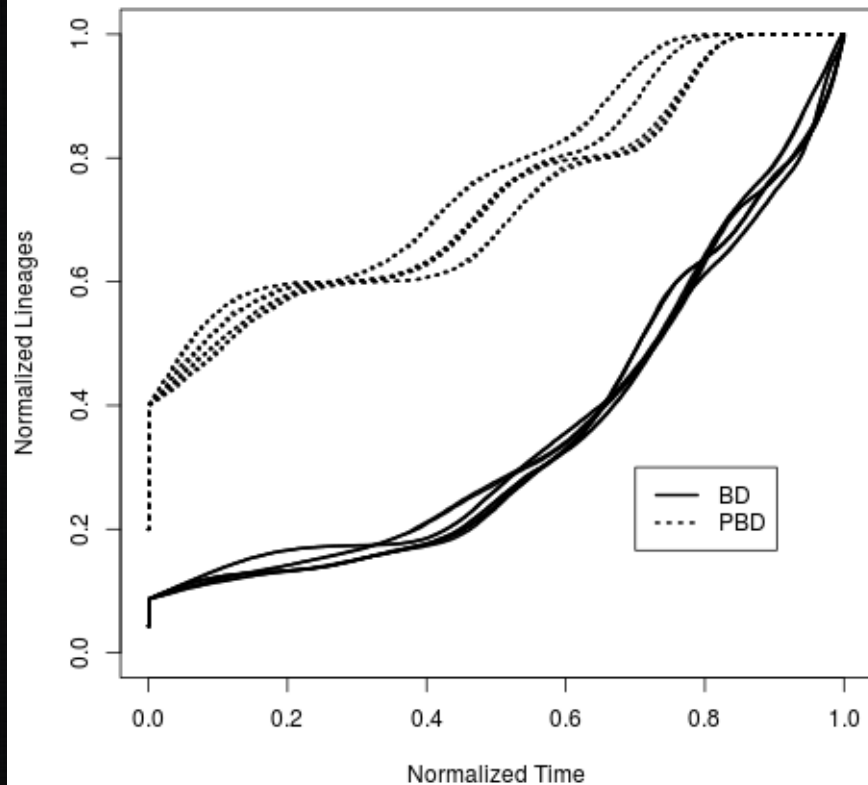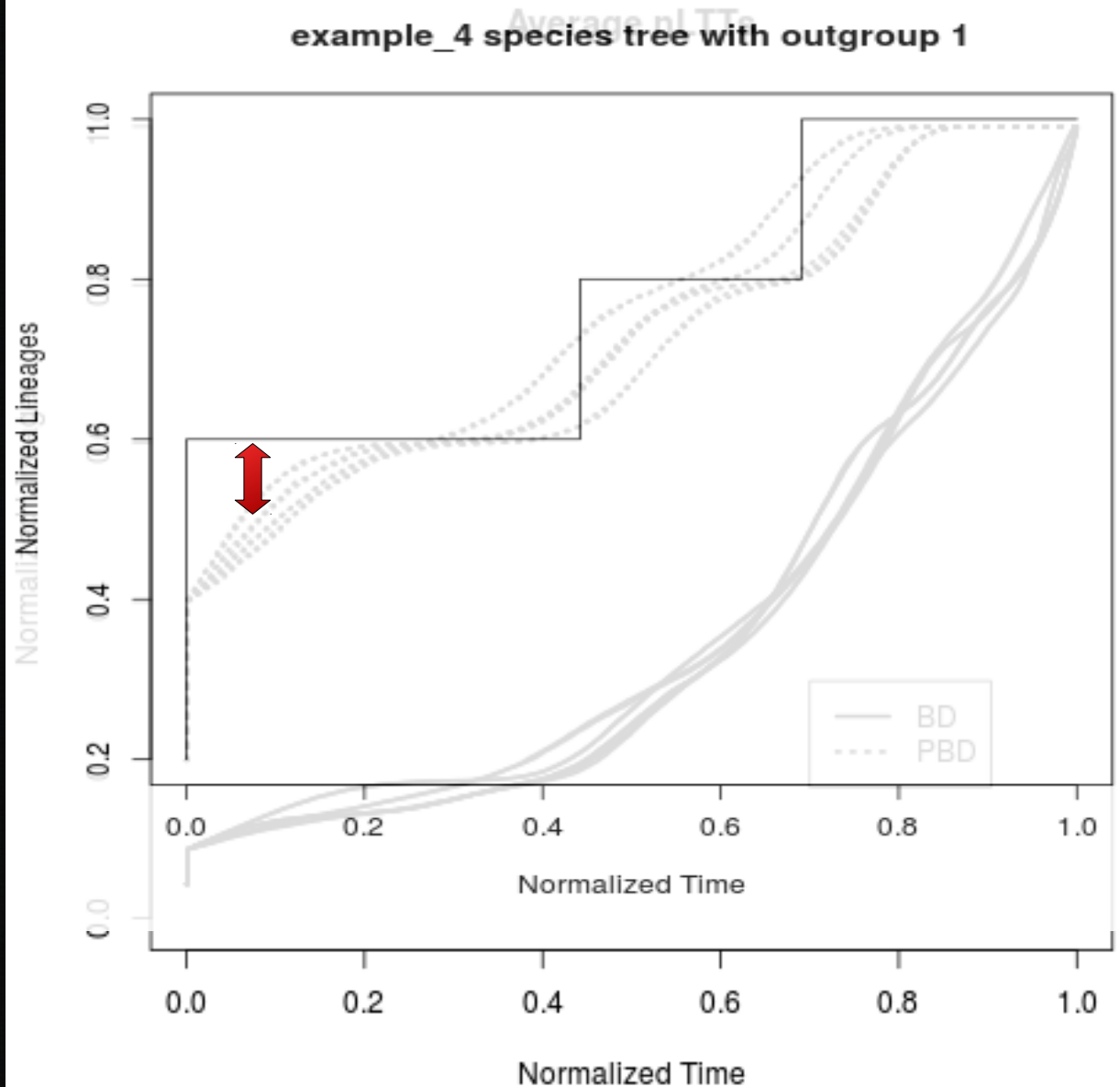
Error

- Sum of the difference between true species tree nLTT and average posterior nLTT

example_4 species tree with outgroup 1

Error BD = error PBD?

- If speciation takes time, how big will the estimation errors be?
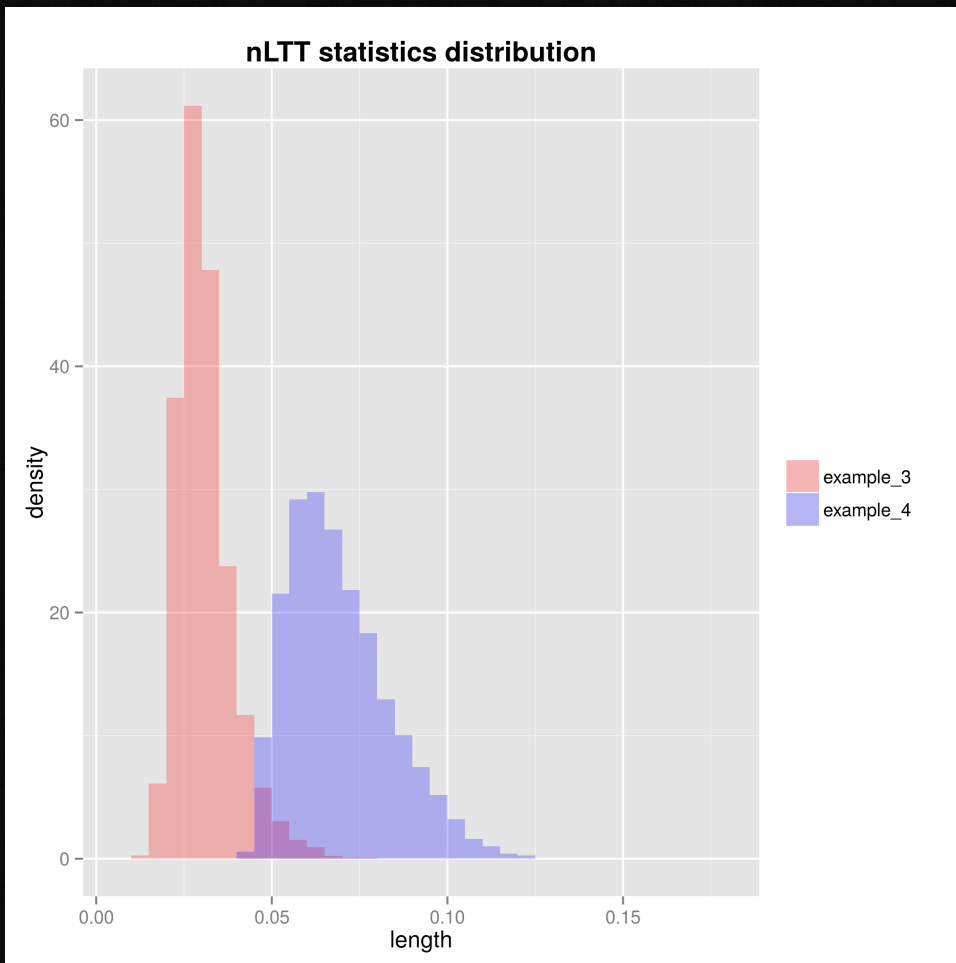
If speciation takes time (example_4), BEAST2 will have a threefold error in recovering a known phylogeny
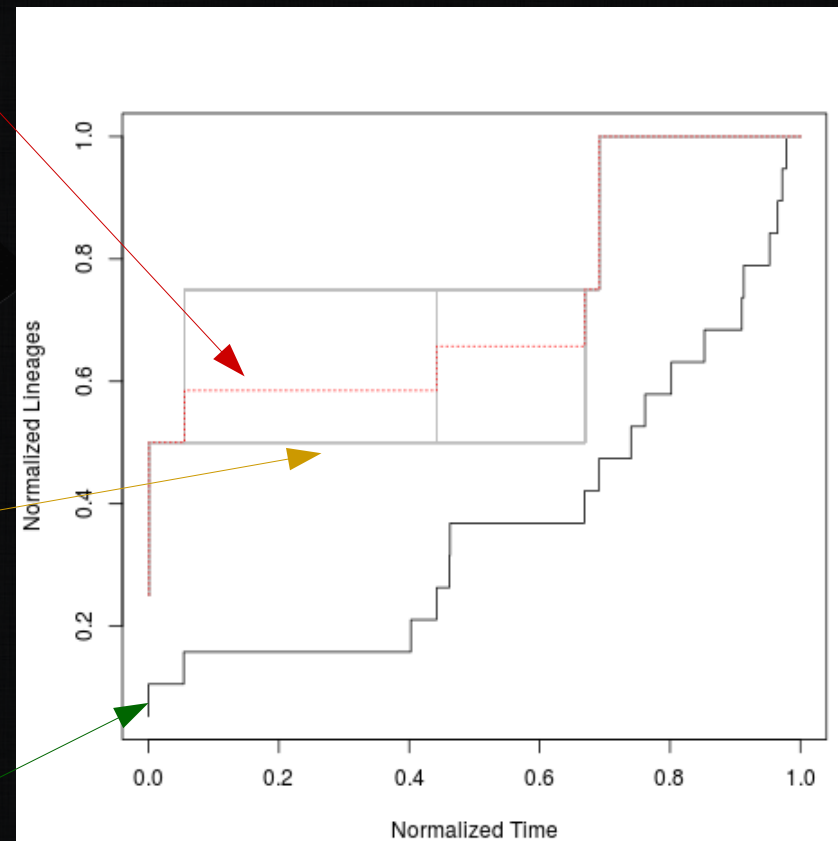
Major cause

- Most of the differences comes from the sampling of the gene tree to obtain a species tree

Average nLTT of all species trees (red line)
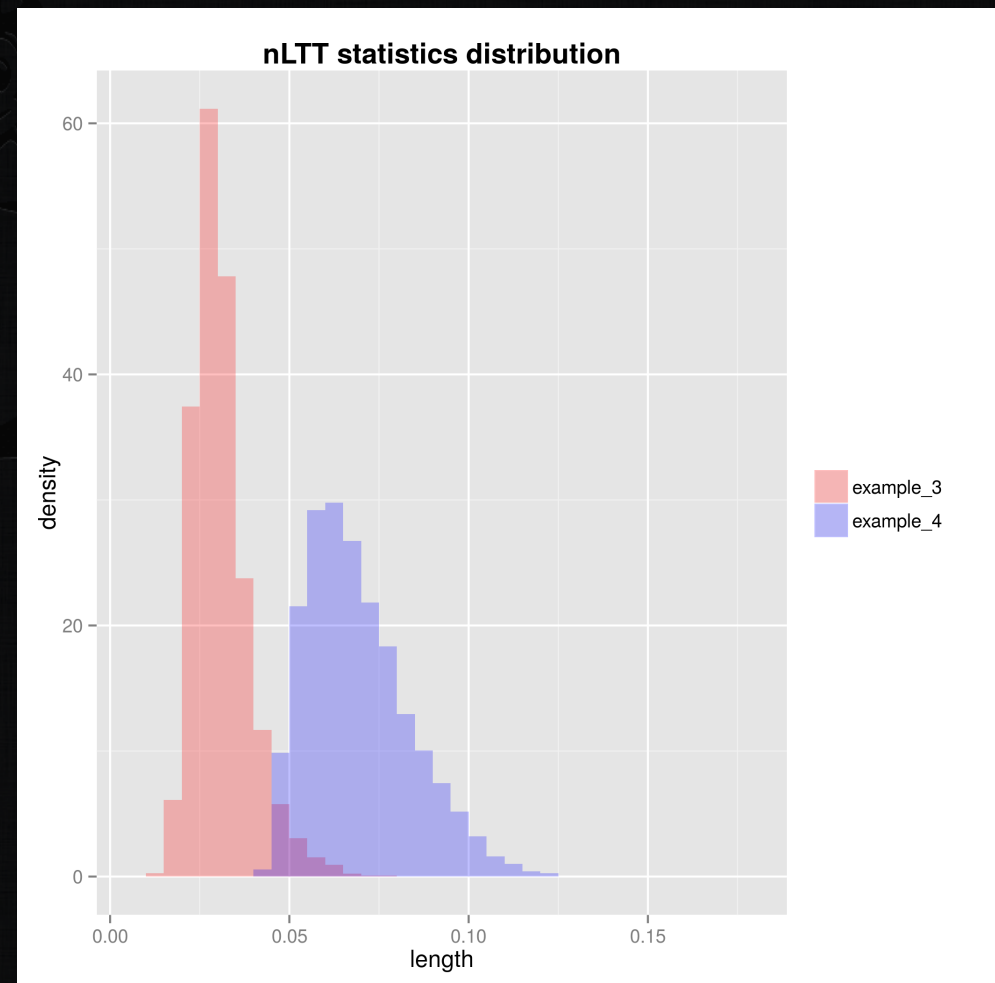
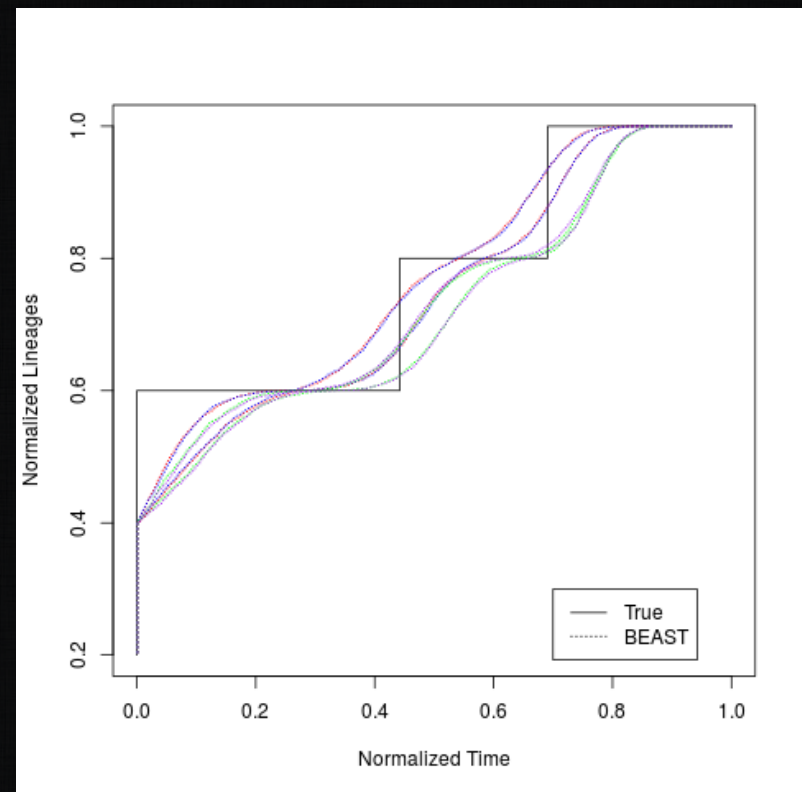All sampled species trees (grey lines)

Gene tree (black solid line)

Preliminary conclusions

- Speciation taking time causes bigger error

- The expected difference is now quantified
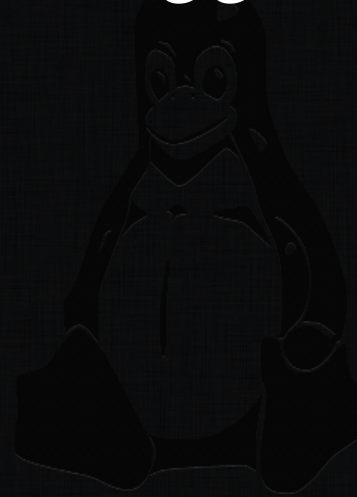

nLTT statistics distribution

Preliminary conclusions
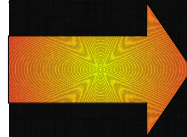
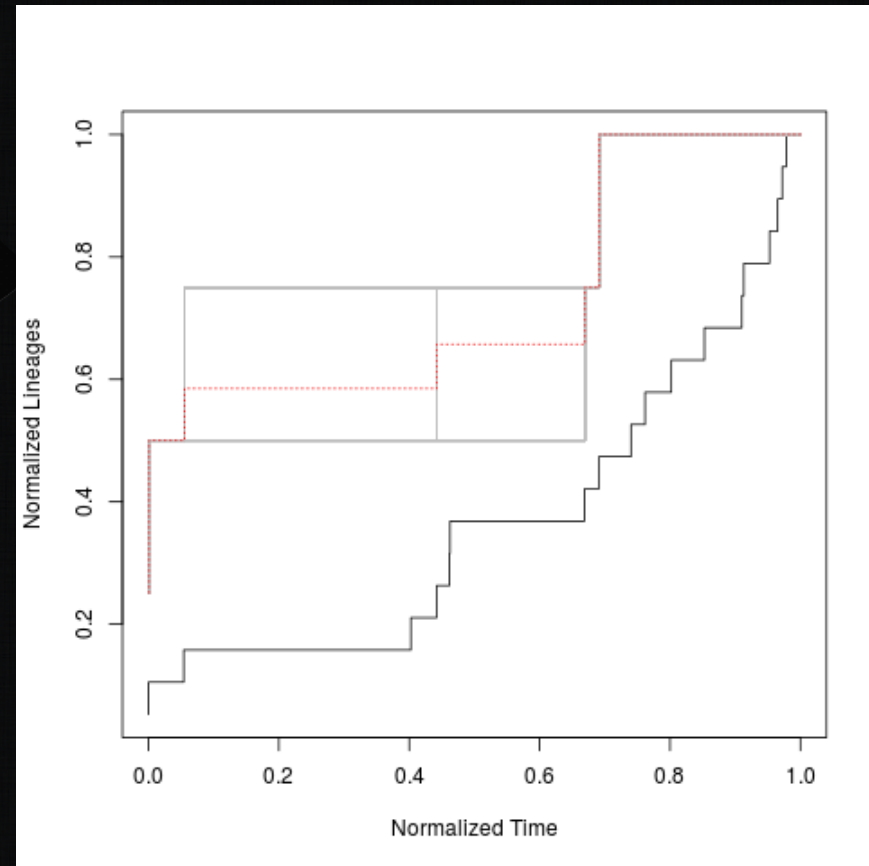- BEAST2 always does a fine job

Preliminary conclusions

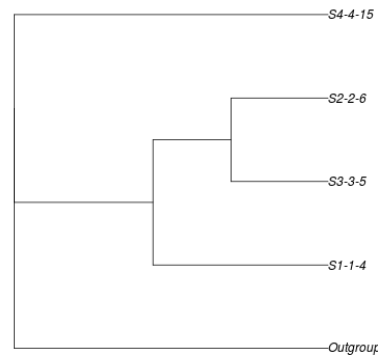- Sampling species trees from gene trees has the biggest impact
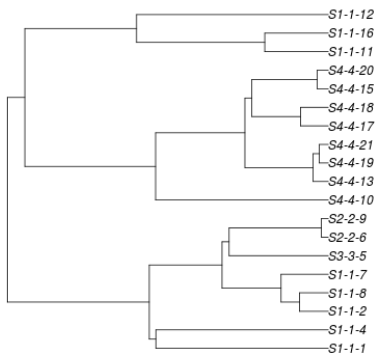
Future work

- Quantify the errors
- Use *BEAST to do a multi-individual inference