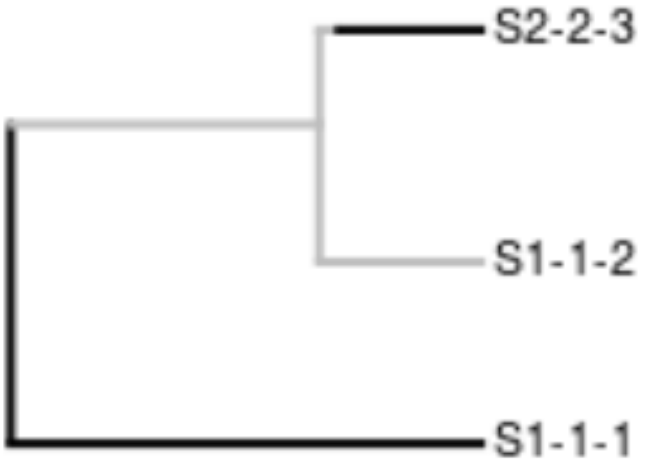
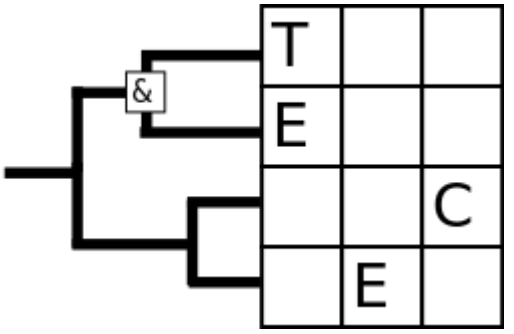


Sampling incipient trees

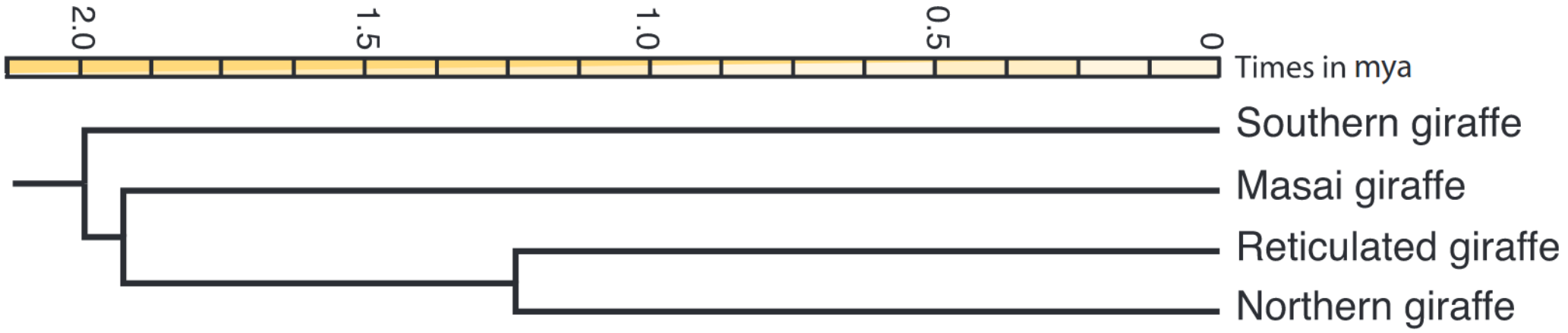
www.github.com/richelbilderbeek/Science



Speciation takes time

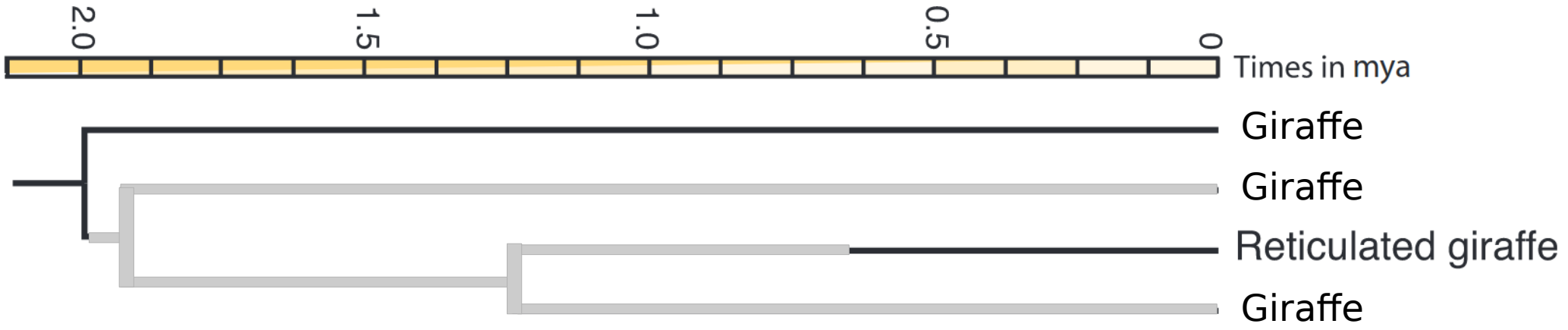
Fennessy, Julian, et al. "Multi-locus analyses reveal four giraffe species instead of one." Current Biology 26.18 (2016): 2543-2549.

Speciation takes time



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

Speciation takes time

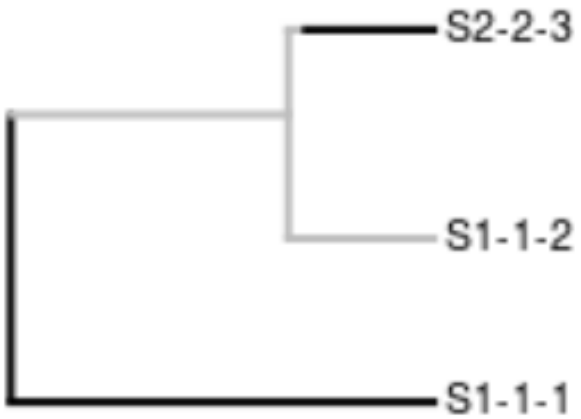


Goal

Share my naivety and insights

I will:

- Describe how incipient species trees are sampled
- When the behavior is as I expected
- When the behavior appeared unexpected
- Suggest a new sampling method

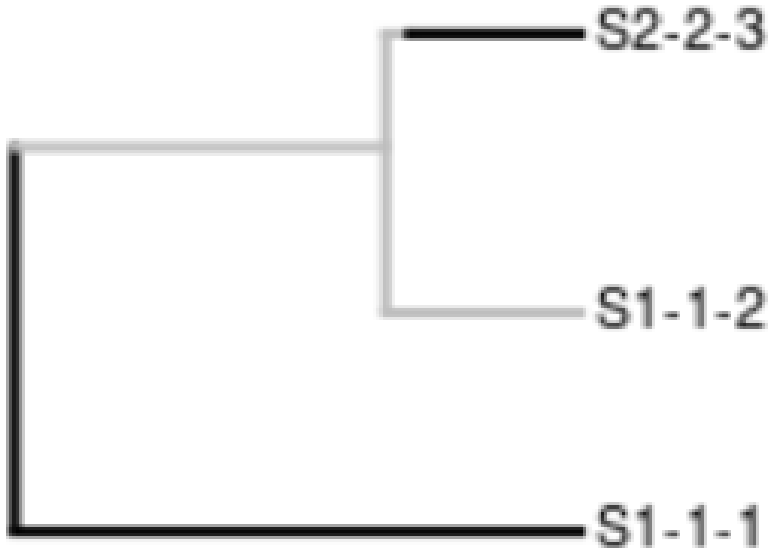


Sampling

An algorithm to pick an incipient species to represent a species

Three ways:

1) oldest
2) random
3) youngest



Can result in different species trees

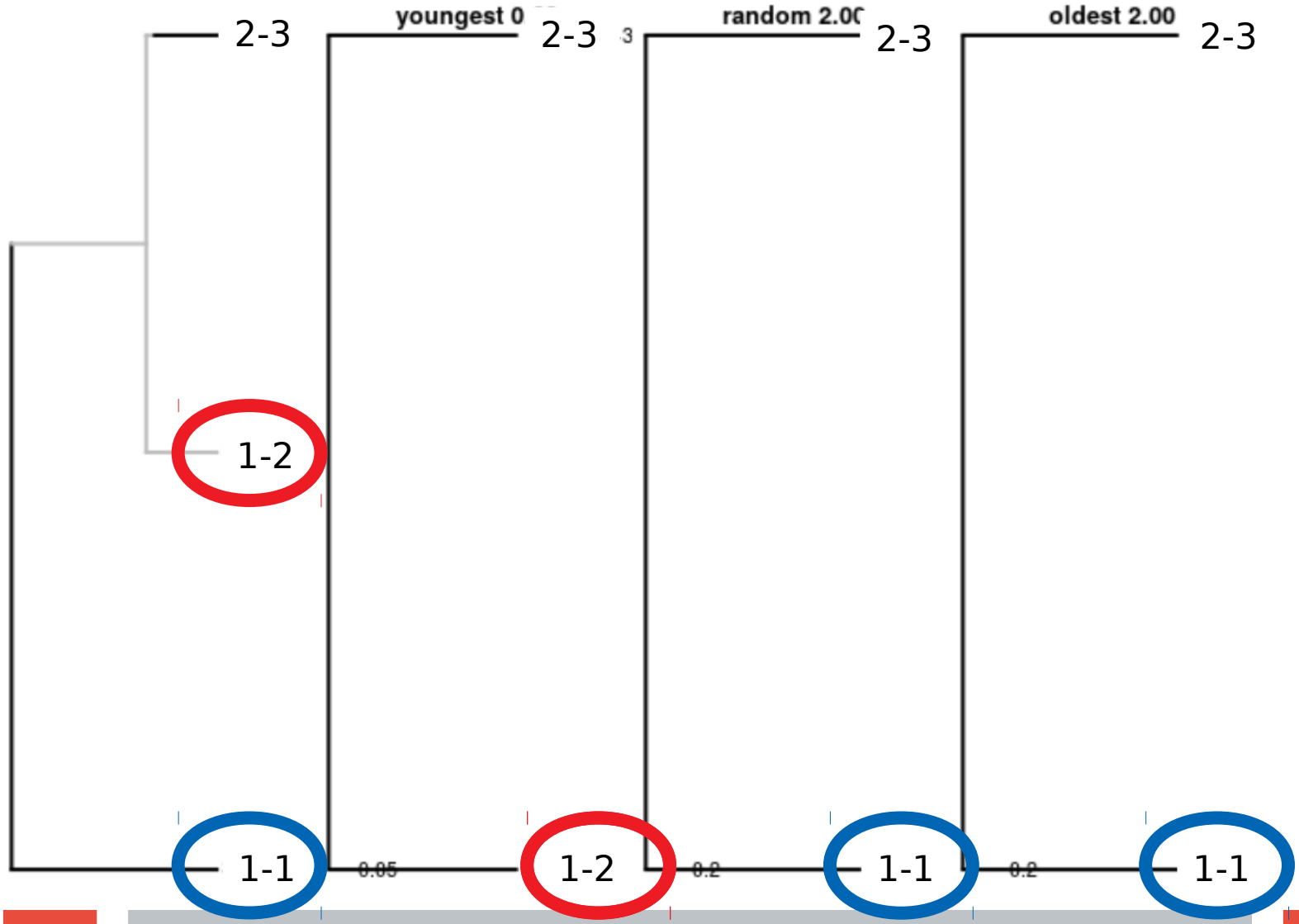
My naive branch length expectations

youngest ≤ oldest

youngest ≤ random ≤ oldest

Expected #1

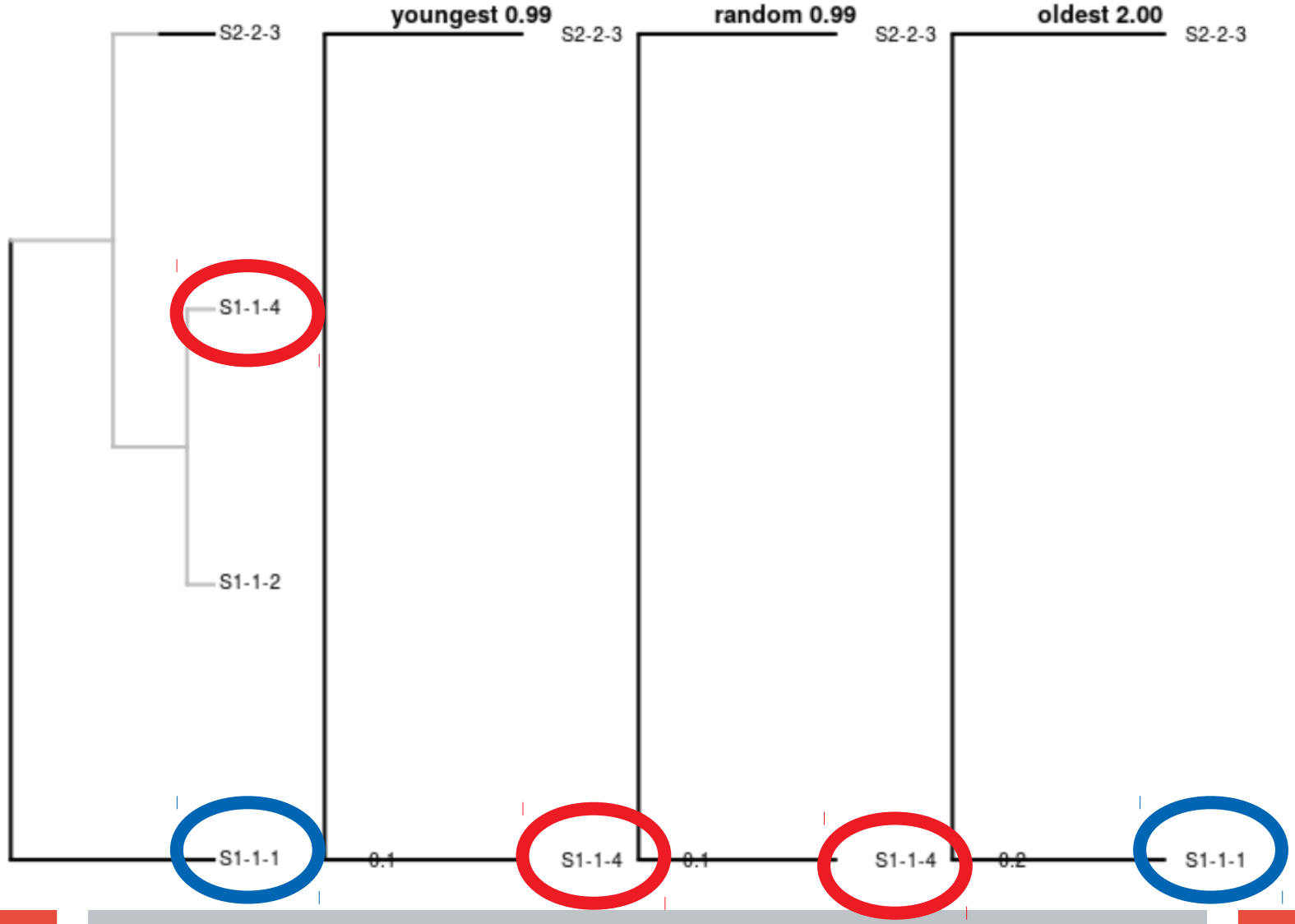
youngest ≤ random ≤ oldest



0.69 ≤ 2.0 ≤ 2.0

Expected #2

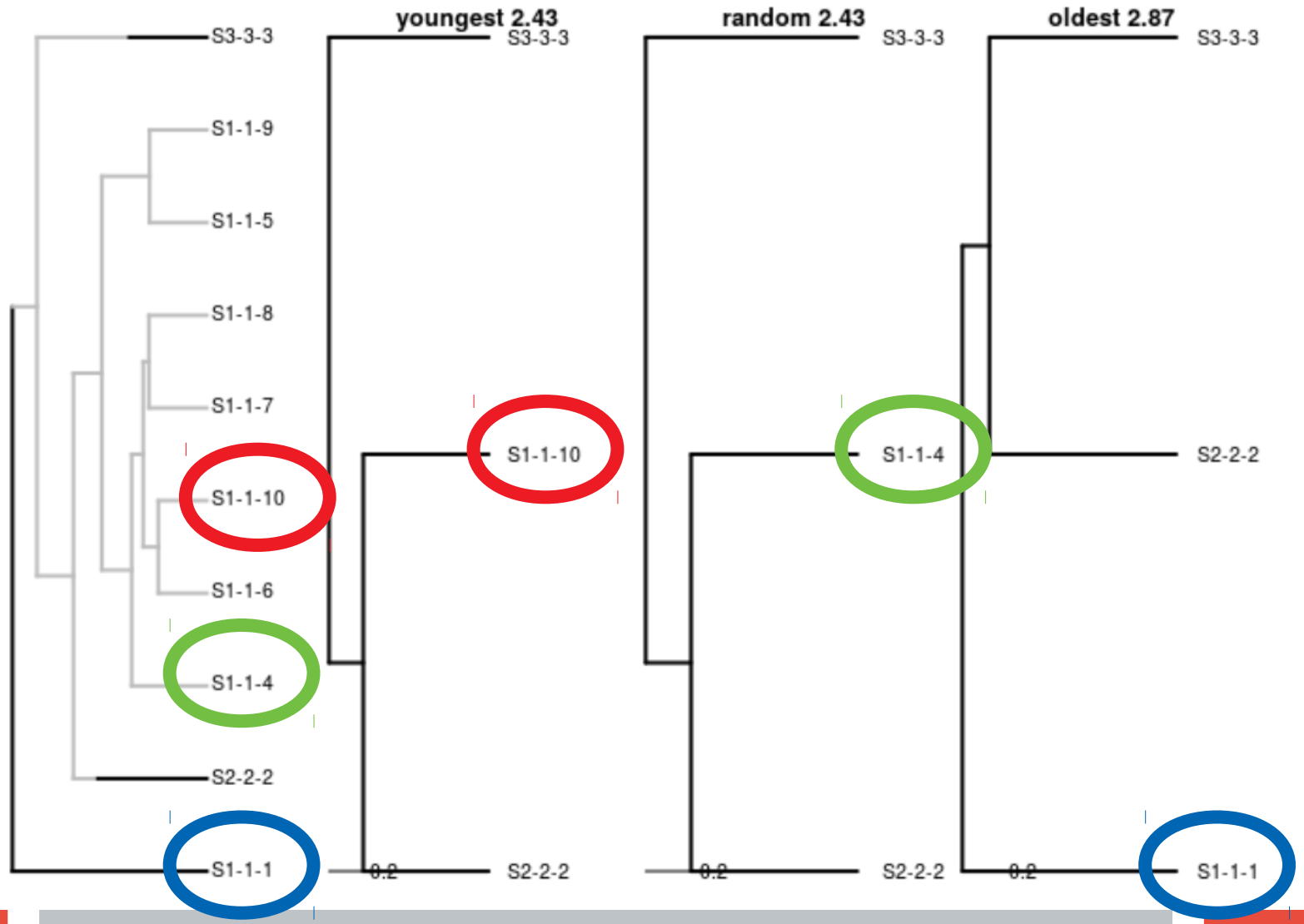
youngest ≤ random ≤ oldest



0.99 ≤ 0.99 ≤ 2.0

Expected #3

youngest ≤ random ≤ oldest



2.43 ≤ 2.43 ≤ 2.87

My naive branch length expectations

youngest ≤ oldest

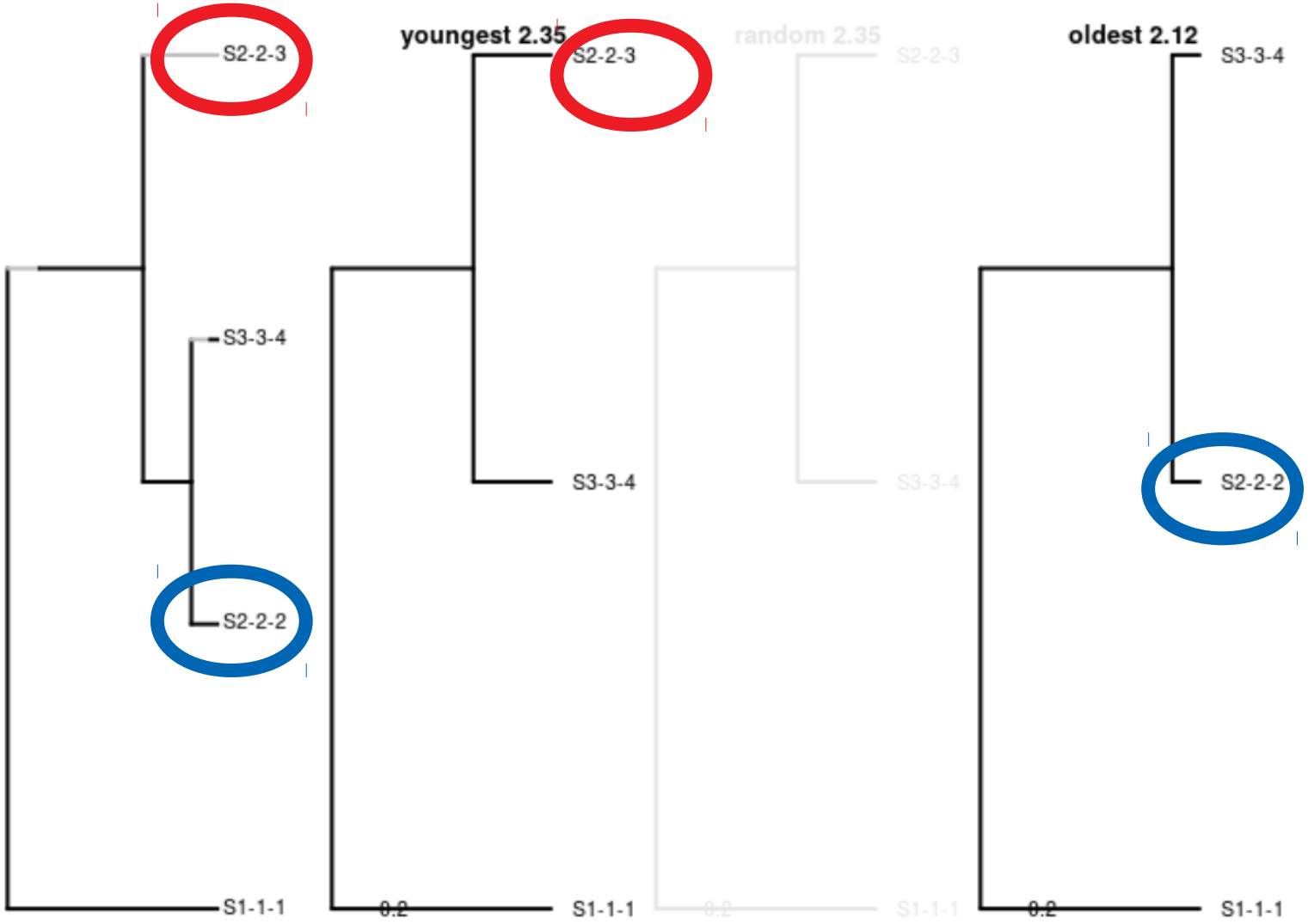
youngest ≤ random ≤ oldest

My naive branch length expectations

youngest ≤ oldest

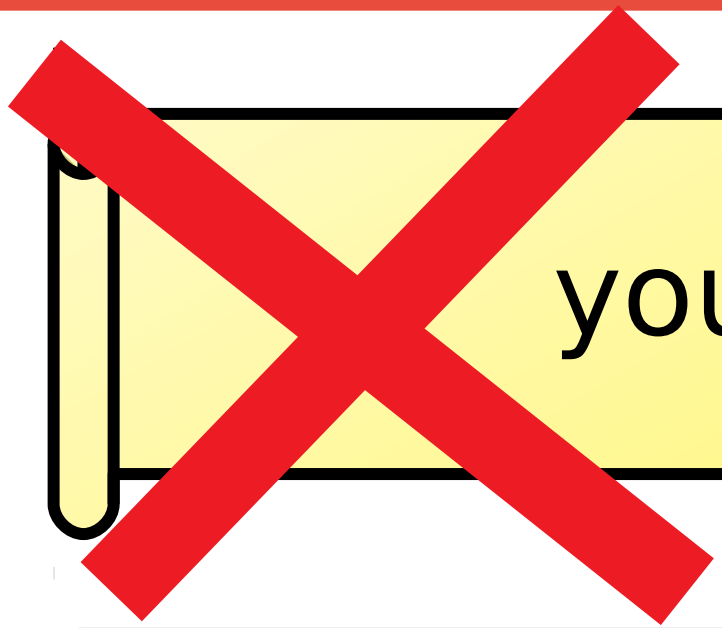
youngest ≤ random ≤ oldest

oldest ≤ youngest



2.35 ≤ 2.12

My naive branch length expectations



youngest ≤ oldest

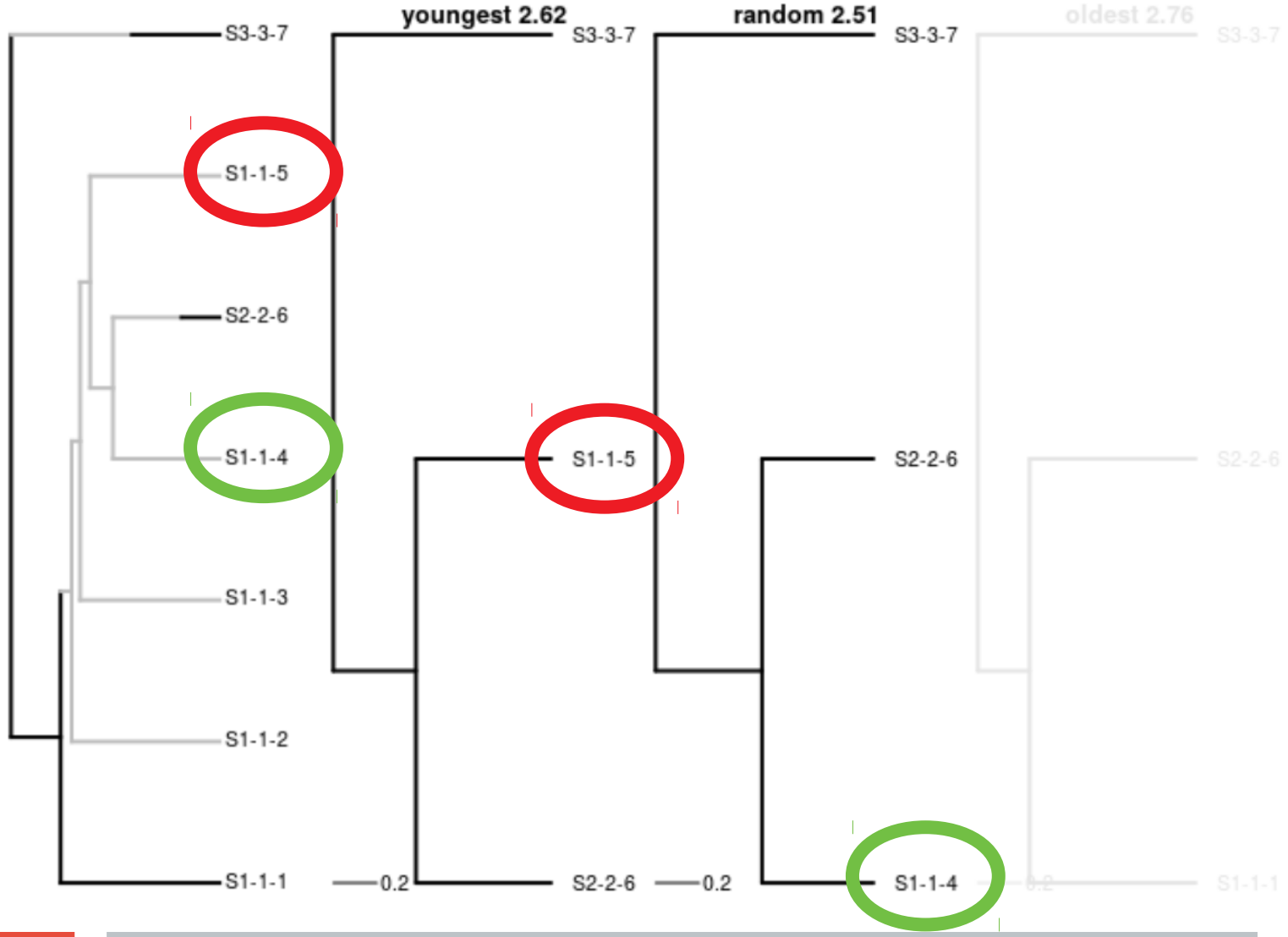
youngest ≤ random ≤ oldest

My naive branch length expectations

youngest ≤ oldest

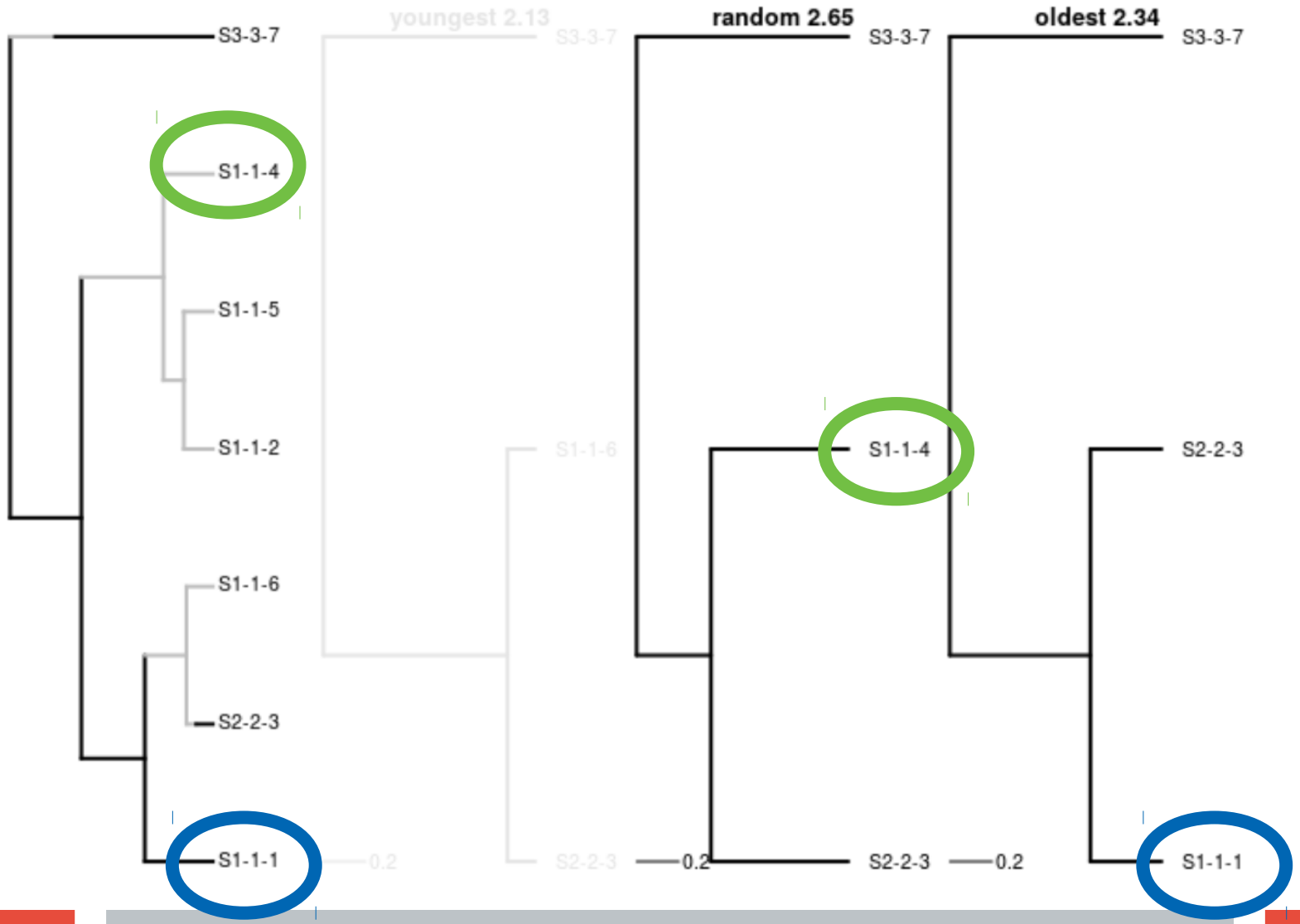
youngest ≤ random ≤ oldest

random ≤ youngest



2.62 ≰ 2.51

oldest ≤ random

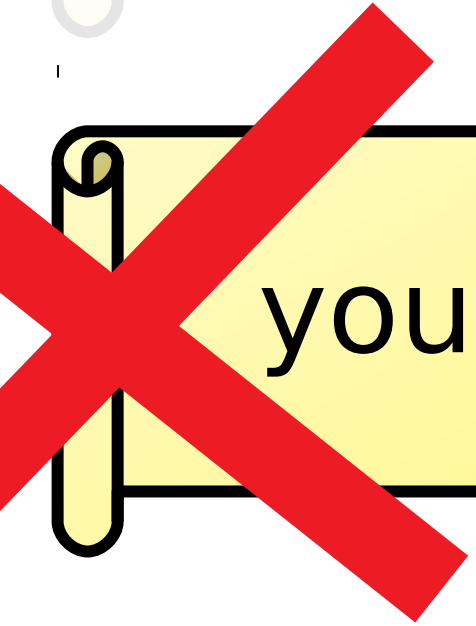


2.65 ≤ 2.34

My naive branch length expectations

youngest ≤ oldest

youngest ≤ random ≤ oldest



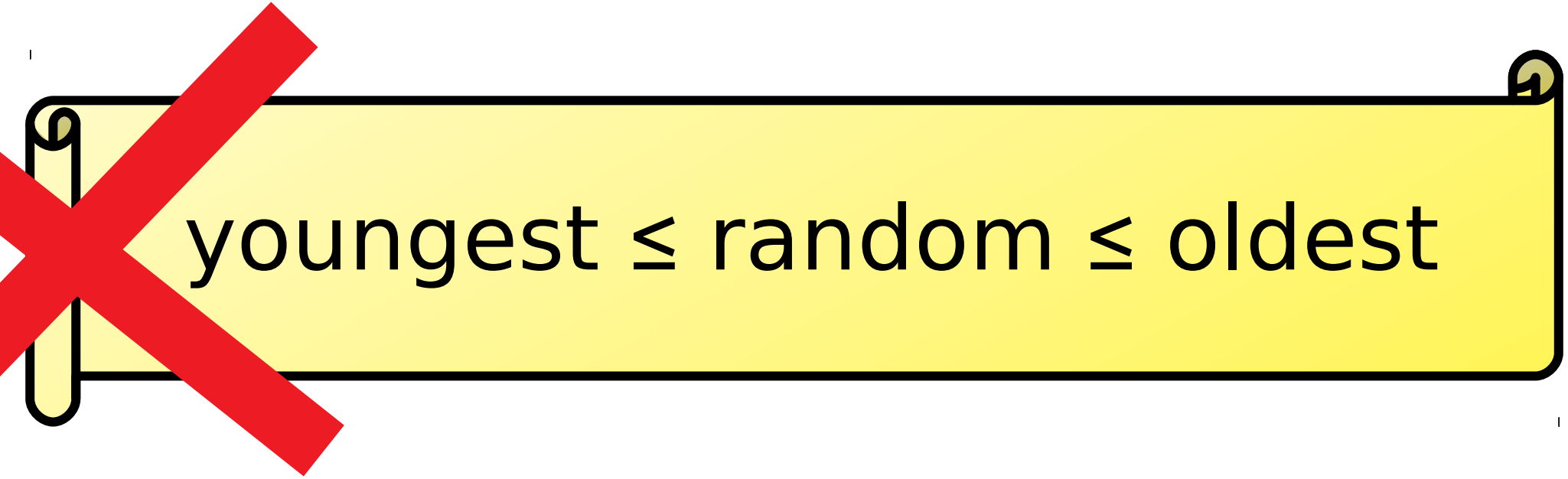
Conclusions

The names ‘youngest’ and ‘oldest’ misled me:

‘youngest’ does not give shortest branch lengths

‘oldest’ does not give longest branch lengths

‘random’ does not give intermediate branch lengths



Conclusions

Suggest to add ‘shortest’ and ‘longest’

‘shortest’ gives shortest branch lengths

‘longest’ gives longest branch lengths

‘random’ gives intermediate branch lengths

shortest ≤ random ≤ longest

Questions?

Redo out these calculations by running the vignette ‘pbd_sampling’ from:

https://github.com/richelbilderbeek/raket

```
devtools::install_package("richelbilderbeek/raket")
git clone https://github.com/richelbilderbeek/raket
```