

Congress of the European Society for Evolutionary Biology

August 14–19, 2022
Prague Congress Centre



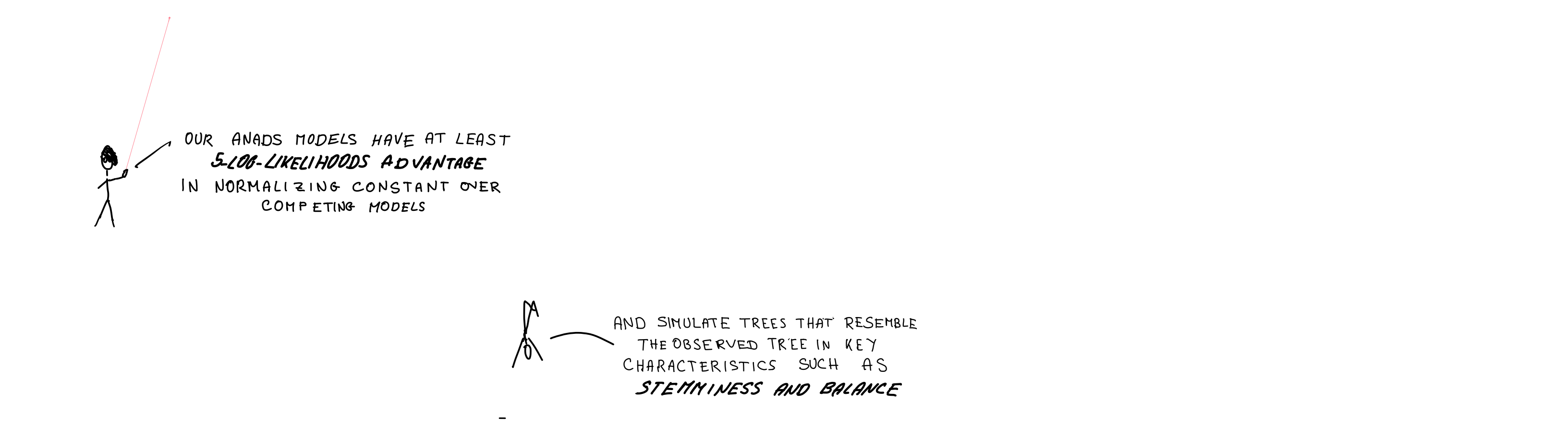
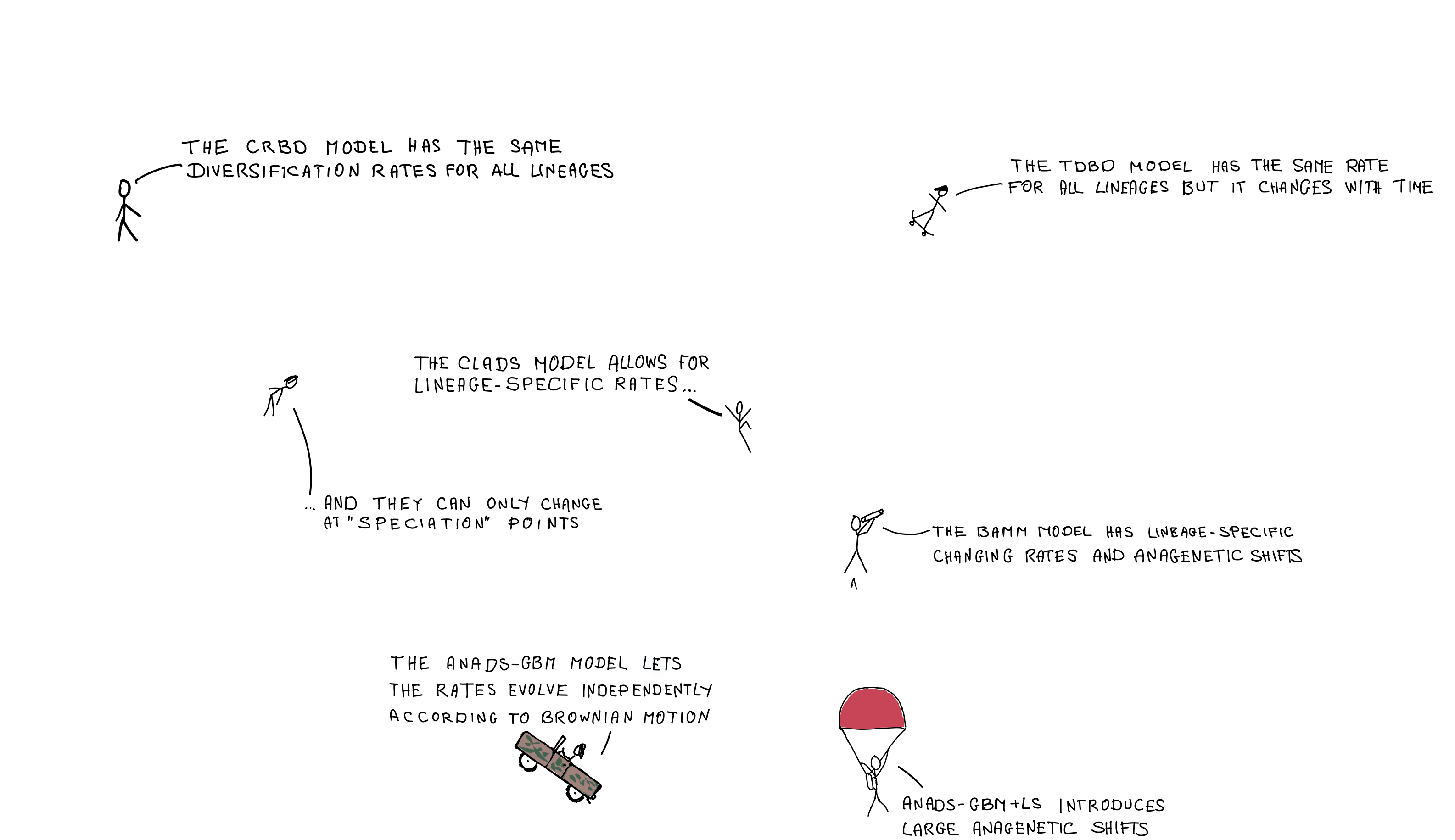
Probabilistic Modeling of Bursts and Gradual Changes of Diversification Rates in Bird Evolution

Viktor Senderov^{1,4,5}, Jan Kudlicka², Nicolas Lartillot³, Daniel Lundén⁴, Ignacio Quintero⁵,
Hélène Morlon⁵, David Broman⁴, Fredrik Ronquist¹

¹Swedish Museum of Natural History; ²BI Norwegian Business School; ³Université Claude Bernard Lyon; ⁴KTH Royal Institute of Technology; ⁵L'École normale supérieure

Diversification rate evolution under different models

Model comparison on Muscicapidae+



Background

In our previous work, the PhyPPL concept paper [1], we used the sequential Monte-Carlo algorithm in a probabilistic programming language framework to do a Bayes factor comparison for the diversification models CRBD, TDBD, BAMM, LSBDS (not shown, a weaker variation of BAMM), and the novel ClaDS on the 40 bird clades having more than 50 species, taken from <https://birdtree.org/>. We identified four patterns:

- Group 1 (small trees such as Alcedinidae): simple models are adequate (CRBD as good as the others);
- Group 2 (e.g. Muscicapidae+): evidence for slowing down of diversification (TDBD outperforms others);
- Group 3 (e.g. Accipitridae): cladogenetic changes in diversification (ClaDS outperforms others);
- Group 4 (e.g. Anatinae): cladogenetic and punctuated anagenetic lineage-specific changes. (ClaDS and BAMM/LSBDS best).

We introduce two new models (AnaDS-GBM and AnaDS-GBM+LS) to investigate the effect of gradual and punctuated *anagenetic* changes in speciation rates. For this poster we tested AnaDS-GBM on one tree per group.

Probabilistic programming languages

Probabilistic programming languages (PPL) are computer languages, whose programs (a) specify a probability distribution (model) as a forward stochastic simulation from a set of parameters—sampled from prior probability distributions—to an outcome (response), and (b) condition that outcome on observed data. Finally, (c) the inference engine that a PPL compiler provides, inverts the program so that its execution results in the posterior distribution over the parameters. Results in this poster were obtained using RootPPL provided within the Miking platform, <https://miking.org/>. We are currently (upcoming release) working on a domain-specific PPL for phylogenetics called TreePPL, <https://treeppl.org/>. Here, an example function simulating the extinct lineages under CRBD:

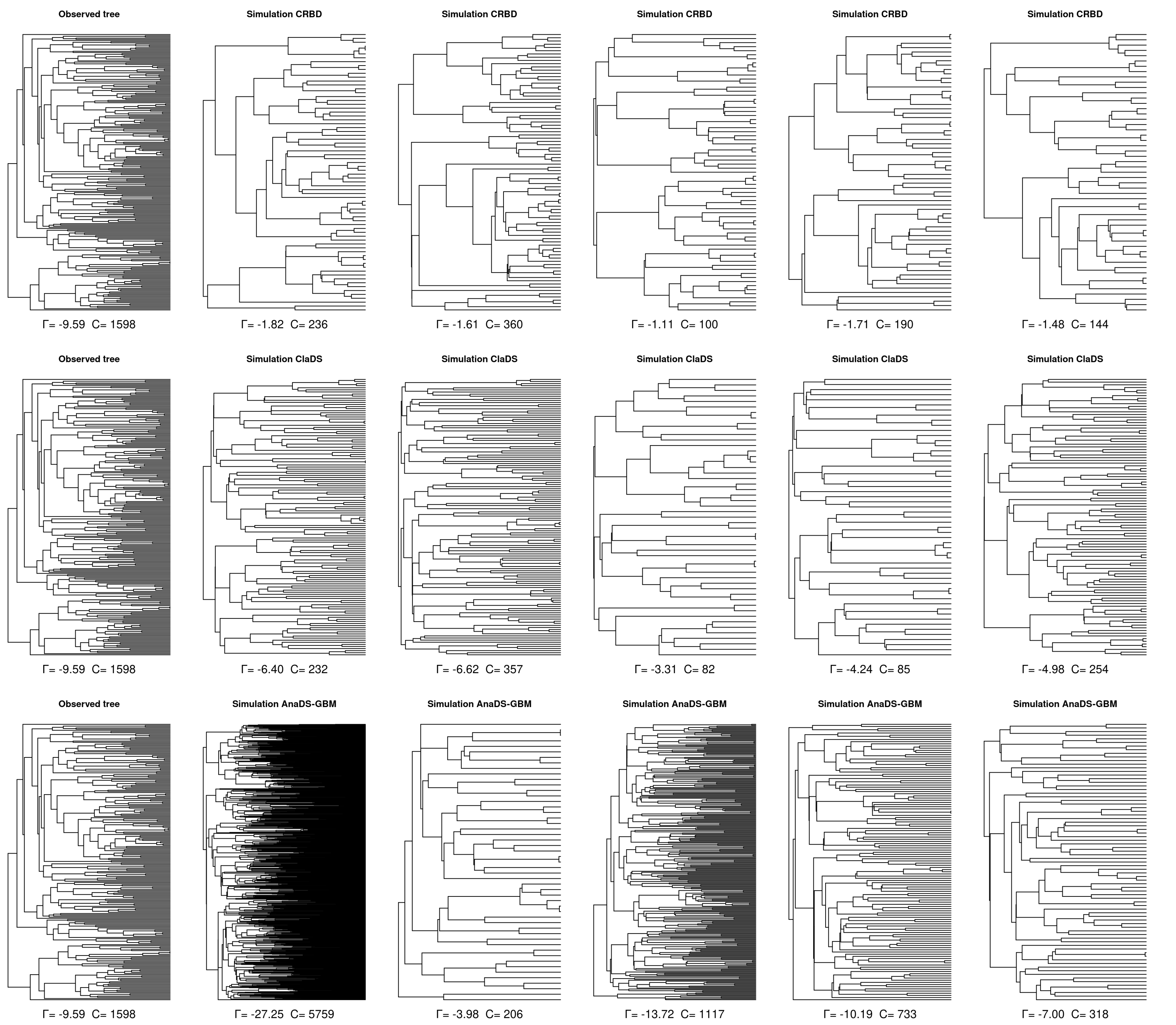
```
01 function simulate_side_lineage(start_time: Real, lambda: Real, mu: Real)
02 {
03   assume waiting_time ~ Exponential(lambda + mu) // Declare a random variable
04   let current_time = start_time - waiting_time
05   if current_time < 0 {
06     weight(0) // Condition on no lineage surviving to present
07   } else {
08     if flip(lambda/(lambda + mu)) {
09       simulate_side_lineage(current_time, lambda, mu)
10       simulate_side_lineage(current_time, lambda, mu)
11     }
12   }
13 }
```

Results and Discussion

- The gradual anagenetic model **AnaDS-GBM is as good as or better** than all other tested models for all four groups based on Bayes factor comparison.
- On the group 2 tree, Muscicapidae+, with evidence for decaying diversification rates, **the AnaDS-GBM model is better than both TDBD and ClaDS** based both on Bayes factor comparison, and on posterior predictive simulations.
- We are currently evaluating (upcoming paper), whether large anagenetic shifts, while accounting for geometric Brownian motion (AnaDS-GBM+LS) is even a better fit for the whole bird tree.

Posterior simulations using Muscicapidae+ data

The following simulations were conditioned to have at least 50 nodes.



All models tend to produce smaller trees than the test tree possibly due to selection bias. The test tree has long terminal branches, which can be explained by a decrease in diversification rates, and an unbalanced shape. CRBD produces balanced trees with shorter branches and ClaDS produces some stemmier but balanced trees. Our AnaDS-GBM produces trees that are close to the observed tree in these key characteristics.

Further reading and acknowledgments

[1] Ronquist, F., Kudlicka, J., Senderov, V. et al. Universal probabilistic programming offers a powerful approach to statistical phylogenetics. Commun Biol 4, 244 (2021). <https://doi.org/10.1038/s42003-021-01753-7>

This study has been funded under European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement PhyPPL No. 898120. You can find an extended version of the poster under <https://github.com/phypppl/eseb-2022-poster> or if you scan the QR code.