

Master thesis proposal

Effects of root state priors

Krzysztof Bartoszek

October 8, 2025

Background — evolution models

The field of phylogenetic comparative methods emerged from the need of a rigorous framework to study inter-species data. On the between species level one cannot assume that the data sample is an independent one. The measurements are dependent through the species' phylogenetic tree, i.e. their common evolutionary history. The first traits to be studied were simple binary traits (e.g. ability to fly or not). Biologists are interested in two questions, firstly can one estimate the ancestral state(s) and secondly can the transition rates between states be obtained.

Thesis project

Estimation of transition rates and ancestral states is commonly based on observations of contemporary species, the phylogeny (estimated from molecular data) and a Markov model for the transitions between states. However, often there is additional biological knowledge available. For instance, the ability to fly is more complex than not being able to fly. Hence, we should expect the state “not fly” to be the root state. One can use such biological knowledge to aid the estimation procedure. However, often one will have a more vague idea what is the ancestral state — have a prior distribution on it. The aim of the project is to explore effects of different ancestral state prior distributions on estimation of transition rates for a binary trait.

Simulation and estimation procedures are available in e.g. `ape` or `phytools` R packages. In the scope of the project one will learn how to use appropriate functions from these packages, how to input priors for them and if time permits extend them with hyperprior (for the probabilities of the root state) models.

Goals

The below general goals are for an “ideal” thesis. Depending on the student they will be made more specific in the direction of the student’s interests. In particular the focus of the work will not be on the mathematical models (these will be “provided”) but on implementing and putting together software to do simulations, inference and explore the statistical aspects of the models.

1. Become acquainted with phylogenetics and phylogenetic comparative methods [2].
2. Become acquainted with simple Markov processes and simulating them on a phylogeny [3, 4].
3. Become acquainted with using (biological) priors and hyperpriors [1].
4. Explore prior effects and learn how to code a prior for e.g. the function `fitMk()`, `MuSSE/HiSSE` family of functions or `RevBayes`.

Data

The project will be predominantly illustrated with simulated data. Possibly data sets attached to the `ape` or `phytools` packages can be explored.

References

- [1] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. CRC Press, 2003.
- [2] E. Paradis. *Analysis of Phylogenetics and Evolution with R*. Springer, 2006.

- [3] E. Paradis, J. Claude, and K. Strimmer. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20:289–290, 2004.
- [4] Liam J. Revell. phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3:217–223, 2012.