Master thesis proposal
Multiple testing in a phylogenetic setup

Krzysztof Bartoszek

October 6, 2022

One of the many aims that phylogenetic comparative methods have is to identify outlier, in terms of some trait(s), species in a group of species. If one has multiple target species or suspects (does not have a priori hypotheses on which ones) that there could be multiple outliers, then one runs into the problem of multiple testing. Classical multiple–testing correction methods, like Bonferroni or even Benjamini–Hochberg, can be expected to be too conservative, as we know that the p–values will be heavily dependent through the underlying phylogeny. In the case of related genes, where one has case and control measurements, a simulation based approach is possible [2]. However, without a case–control setup one would need to have a leave–some–out approach (non–parametric) or start modelling the p–values themselves [1]. The aim of the thesis is to develop and compare different parametric and non–parametric approaches for identifying outliers and clades of outliers for data that have a tree–based dependency structure. Some particular questions are how do the different approaches behave for different tree models and sample sizes. For the thesis simulated and from the literature data will be used.

References

[1] A. Bichat, C. Ambroise, and M. Mariadassou. Hierarchical correction of p–values via a tree running Ornstein–Uhlenbeck process. *ArXiv e-prints*, 2020.

[2] H. Xiao, K. Bartoszek, and P. Liò. Multi–omic analysis of signalling factors in inflammatory comorbidities. *BMC Bioinformatics, Proceedings from the 12th International BBCC conference*, 19(Suppl. 15):439, 2018.