To assess the robustness of our conclusions on model selection, we carried out a three-step procedure for each empirical dataset. 1) Using the parameters that maximize the likelihood for each of the three models (Constant-rate (CR), Concealed Trait-Dependent (CTD) and Examined Trait-Dependent (ETD) diversification) we simulated 100 datasets (phylogenies and traits; details below) that are structurally similar to the empirical data. 2) We ran SecSSE analyses under the three models for each simulated data set. 3) Finally, we compared models and counted the instances where the generating model was proven to be the best model according to Akaike weight.

For the simulations, we considered a trait with three examined states and three concealed states, as used for the parameter inference on the empirical data (i.e., 1, 2, 3 and A, B, C respectively), leading to a nine-state system (i.e., 1A, 2A, 3A, 1B, 2B, 3B, 1C, 2C, 3C). The simulation starts with two species with the same trait state; this state can shift to the other states at a transition rate Q. Our simulation is conditioned on survival of these two species; if one of the species becomes extinct, the simulation is discarded. Speciation and extinction rates depend on trait states, and similar to BiSSE, HiSSE and MuSSE, the switch from one state another does not cause speciation. Moreover, after a speciation event, both daughter species will have inherited the same trait state than the parental species. The simulation is performed for a period of time equal to the crown age of the empirical data. A phylogeny is reconstructed from the record of speciations and extinctions during the simulation. Then, we re-labeled trait states to merge them into a trait with only three states. The traits 1A, 1B and 1C were coded as 1; 2A, 2B and 2C were coded as 2 whereas 3A, 3B and 3C were coded as 3.

In here, we archive the simulated datasets.

Files are named like:

“study6dataModel1.RData”

Indicating that the simulated data set is structurally similar to empirical dataset 6, simulated under Constant Rate model (Model1); Concealed Trait-Dependent (Model2) or Examined Trait-Dependent (Model3)

After loading the .RData file, one can access the *i* dataset (phylogenetic tree plus trait data) with:

phy <- phylotrees[ [ i ] ]

traits <- traitdata[ [ i ] ]

Note that traits are already sorted according tree tips.