Introduction

* Over the last decade researchers are using molecular phylogenies of extant lineages to estimate rates of diversification (speciation and extinction). Skepticism comes from Rabosky 2010, Quental & Marshall 2011, due to extinction rates being frequently estimated with very small values using molecular phylogenies.
* However, according to the fossil record most lineages are already extinct and many of them are in a declining phase (examples: canids, rhinos, hyena, horses).
* Lineages are characterized by an expansion and a decline phase and it is still an open question whether we can detect it or not using molecular phylogenies (Quental & Marshall 2011 suggest with a simple approach – gamma statistics – that it is difficult to distinguish between decline and stable diversity; Morlon et al. 2011, with more sophisticated methods were able to detect decline; but discuss the limitation of their results).
* However, how general are the results from Morlon et al.? Lack of papers that have used these methods, and even greater lack of papers that actually found decline with these methods.
* Curiously BiSSE-like models suggest that some traits might be associated with negative diversification rates (Goldberg et al. 2010, Burin et al. 2015, among others).
* Additionally, more recently Beaulieau & O’Meara shown by revisiting Rabosky’s 2010 that extinction rates can be reasonably estimated in some scenarios, although these scenarios are limited by not including decline in diversification.
* New methods (Morlon et al. (2011) and Rabosky (2014)) allow for extinction to be higher than speciation although none of these two methods were tested thoroughly in their capacity of detecting declines in different points in time. Liow et al. 2010 show that the signature of a given diversity dynamics might change as time goes by.
* Here we present a broad analysis to compare their ability to detect diversity decline in different evolutionary scenarios over different viable parameter combinations within a comprehensive parameter space. Mention that our aim is not only to assess the potential of both methods in estimating the true rates, but also how do both perform when evaluating the diversity trajectory over time.

In summary, most of the groups analyzed would still be expanding which is in disagreement with the trend suggested by the fossil record.

Goals

Our goal was to explore the performance of two recently proposed methods,

Morlon et al. (2011) and Rabosky (2014) (BAMM) in detecting diversity declines using molecular phylogenies.

Methods

* We used scripts provided by Dr. Hélène Morlon to simulate 2000 in each evolutionary scenario.
* Parameter space was explored in the following way:
  + Randomly sampling parameter values (lambda, alpha, mu, beta)
  + Calculating expected time of decline (sp = ext)
  + Estimate time for losing 80% of maximum diversity
  + If this value lies between 10 and 500 species, then run the simulation
  + If either the tree dies out before the simulation time or the diversity exceeds 20000 species at any time the parameters values are re-sampled and the simulations are re-run.
  + All the parameters combinations (valid, extinct or explode) were stored in different files to help us analyze the parameter space
* The first scenario had an exponential decline on speciation rates and constant extinction rates and the second had constant speciation rates and exponential increase on extinction rates.
* Equations
* For all valid parameter combinations we numerically estimated the expected time to lose 50 and 20% of peak diversity and used these times to slice each tree to evaluate how the methods perform in different decline phase lengths.
* To check for false positives we also estimated the time expected to reach 80% of peak diversity but on the rising phase.
* We fitted the model published in Morlon et al. (2011) and the method BAMM (Rabosky et al. 2014) to all the simulated trees in all four time-slices. The fitting was run in a cluster in parallel.
* Morlon’s model estimate via maximum-likelihood the parameter values that best describe the estimated behavior of diversification rates, while BAMM returns instantaneous diversification rates using RJMCMC.

Figures

Table 1 – Parameters space intervals

Table 2 – (or figure) Trees statistics: n tip and time (min, max, mean and median)

Figure 1 – Cartoon LTT to illustrate diversity trajectories and simulated scenarios

Figure 2 – Scatterplots fitted by simulated values for lambda, alpha, mu, Beta and net diversification for each method / bar graph of differences between estimated and simulated for both methods.

Material and Methods

Parameter space exploration

We divided our simulations into two scenarios: the first scenario had exponential decline on speciation rates and constant extinction rates through time (BVARDCST), whereas the second scenario had constant speciation rates and exponential increase on extinction rates over time (BCSTDVAR). For these two scenarios, four parameters were combined according to each scenario: for BVARDCST, we used two parameters for speciation rates (initial speciation LAMBDA\_0 and ALPHA) and one parameter for extinction rates – MU. In the second scenario (BCSTDVAR), we used one parameter for speciation rates (LAMBDA) and two for extinction rates (initial extinction MU\_0 and BETA). These values were sampled randomly for each of the simulations (simulation process described below) from uniform distributions bounded by the values presented in table 1.

Diversification scenarios and simulations

The simulation process of the 2000 simulated trees for each scenario consisted in 6 steps, as follows. (1) Parameter values (LAMBDA\_0, ALPHA, MU\_0 and BETA) were randomly samples from a uniform distribution with the limits indicated in table 1. (2) The expected time of the initiation of the decline phase was calculated (*t\_max* when speciation = extinction). We also calculated the expected peak species diversity at *t\_max*. (3) We then estimated the time necessary for losing 80% of the peak diversity. The final species diversity values were forced to lie between 10 and 500 species at the end of the simulation, due to limitations imposed by working with trees too small (low statistical power) or too big (high computational memory/time demand). If the expected final diversity was outside of these limits, the sampled parameter values were stored and discarded, and steps 1-3 were repeated. The time needed for the species diversity to drop to 80% of peak diversity was obtained by numerical approximation, since it is not possible to analytically integrate the speciation and extinction functions. (4) The time calculated in the previous step was then used in the simulation function. The function is available at <http://github.com/labmeme>, and its initial version was kindly provided by Dr. Helene Morlon. The function simulates trees according to time-varying speciation and/or extinction rates, and stores the full resulting phylogeny that contains both extant and extinct species. We modified it to limit the maximum total diversity to 20000 species; if at some point the simulation reached this limit the simulation was interrupted, the parameters were stored and discarded and steps 1-4 were repeated; additionally, some simulated trees went fully extinct before the set time: in this case, the parameter values were also stored and discarded, and the steps 1-4 were repeated. All parameter combinations (valid, extinct or “exploded” were stored in separate files, and used to explore the properties of the parameter space.

After simulating all trees for each scenario, three final steps were performed. (5) We used the estimated diversity estimates for the whole life of each simulated tree to calculate, analogous to step 3, the times needed for the loss of 50% and 20% of peak species diversity. Furthermore, to check for false positives, we also calculated the time needed to reach 20% less of peak diversity but still on the increase phase, were the methods are expected to not detect decline in diversity. (6) All trees had then their extinct species pruned to give us the corresponding molecular phylogenies.

Model fitting and Parameter estimation

We used two recently described methods (RPANDA and BAMM) to retrieve information about speciation and extinction rates for the resulting molecular trees. RPANDA estimates the parameters of the functions that describe the variation of speciation and extinction rates through time, whereas BAMM provides average instantaneous rates for each time step within the duration of the tree. Both models were thus fitted to each one of the trees in both scenarios. RPANDA models were fitted within the R environment, whereas BAMM is coded in python. The two models provide distinct types of information: while RPANDA returns parameter estimates for the model(s) one is evaluating, BAMM generates point mean rate estimates for discrete time intervals along a tree history. Thus, in order to compare both results, we estimated the maximum likelihood parameters for each negative exponential model to BAMM estimates (speciation in BVARDCST and extinction in BCSTDVAR), and compared these parameters with the corresponding values estimated by RPANDA. Results were analyzed in the light of the differences between the simulated and the estimated corresponding parameters.