Can we detect decline in diversity using molecular phylogenies?

July 15, 2015

Introduction

Over the past two decades, it is possible to see the blooming and establishment of molecular phylogenies as important information used to estimate rates of diversification (Nee et al. (1994), (???), Ape, Geiger, Diversitree, BAMM). As reviewed in (???), different types of data obtained from these phylogenies (e.g. branch lengths, topology, branching times, clade richness and ages) can be used to estimate diversification rates (e.g. speciation/origination, extinction) as well as morphological/phenotypical evolution (e.g. transition rates, variance). The use of those phylogenies have allowed researchers to evaluate the diversification dynamics of several types of organisms, and especially have allowed the study of fossil-poor clades such as birs for instance ((???), Burin2015). However, some skepticism regarding the extent to which we can rely only on the molecular phylogenies was raised, due to the very small values of extinction rates commonly obtained when using only this type of data (Rabosky (2010), Quental and Marshall (2011)).

According to the fossil record most lineages that ever existed on the planet are already extinct and many of the ones that are present nowadays are in a decline phase (such as canids, rhynos, hyenas, horses) ((???), (???), MORE REFS).

Most lineages that ever existed on the planet are now extinct. It is estimated that 99% of all biodiversity is not present in current days anymore, and from the remaining few some of them (such as canids, rhynos, hyenas and horses) are in decline of diversity ((???), (???), MORE REFS). Due to this vast extinct diversity, studying past lineages can be of a great use to understand how biodiversity changed over time to get to the current picture. Paleontological studies have for years generated the most diverse results and conclusions about diversity dynamics of extinct lineages, and fossils continue to be an extremely important source of information. However, not all groups have a fossil record good enough to allow researchers to reliably observe patterns and infer processes about their life history. So another way of infering about the past history of fossil-poor clades is using molecular phylogenies.

With the advent of advanced molecular techniques, each day more high quality molecular phylogenies for all kinds of organisms are produced. Using molecular clock models the resulting trees are time calibrated and so represent the absolute time duration of lineages. This increase in the availability of phylogenetic information spurred the development of a myriad of statistical methods that aim to estimate diversification rates.

- Most lineages that existed are already extinct, and some of the extant is currently in decline

- We have a lot to gain studying past lineages. Through fossils for instance, we are able to estimate speciation, extinction and morphological evolution rates.

- However, this is not possible for all groups. Groups with bad fossil record

- Molecular phylogenies (explanation) can reconstruct part of the past history of extant groups

- Can be particularly useful for fossil-poor clades

- Increasing number of good quality phylogenies (due to new techniques) stimulated the spurr of several diversification models based on these phylogenies

- Some examples of methods

- However there are some caveats: ltt plots (common way of evaluating diversity dynamics) do not show decline (at most show deccelleration). In summary, most of the groups analyzed would still be expanding which is in disagreement with the trend suggested by the fossil record.

- Concerns about correct estimation of extinction rates from molecular phylogenies

- Simple approaches proposed: gamma statistic. Too simple, it is difficult to distinguish between decline and deccelleration.

- Additionally, Liow shows that the signature of a given diversification dynamics may change as time goes by (different phases of the dynamics would yield "false results")

- Recently developed methods claim to be able to detect decline in diversity using molecular phylogenies. However, those methods haven't been thoroughly tested.

- It is crucial to know how powerful/reliable are these new two methods in relation to detecting or not decline and also in relation to the precise estimates of rates.

- 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.

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.

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 λ_0 and decaying rate α) and one parameter for extinction rates $-\mu$. In the second scenario (BCSTDVAR), we used one parameter for speciation rates (λ) and two for extinction rates (initial extinction μ_0 and decaying rate β). 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 of 6 steps, as follows. (1) Parameter values (λ, α, μ and β) 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/gburin/labmeme/bamm_rpanda, 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 models, we analyzes only the initial and final values of the varying rate and of net diversification as well as the constant rate values.

References

Nee S., May R.M., Harvey P.H. 1994. The reconstructed evolutionary process. Philosophical transactions of the Royal Society of London. Series B, Biological sciences.

Quental T.B., Marshall C.R. 2011. The molecular phylogenetic signature of clades in decline. PloS one. 6:e25780.

Rabosky D.L. 2010. Extinction rates should not be estimated from molecular phylogenies. Evolution. 64:1816–1824.