Inferring speciation and extinction processes from extant species data

Tanja Stadler
Institut für Integrative Biologie, Eidgenössische Technische Hochschule Zürich, 8092 Zürich, Switzerland

Querying the past is hard. Speciation and extinction processes are on a scale of thousands to millions of years. Thus, they are most often studied by reconstructing the evolutionary past. This past is reconstructed using phylogenetic methods either on the basis of data from living species or by directly examining the fossil record. Robust methods for inferring the evolutionary past purely on the basis of living species would allow us to understand speciation and extinction processes for the large number of groups without a good fossil record. Generally, studies using living species infer lower extinction rates than the rates suggested by the fossil record (1, 2). A new study in PNAS (3) suggests that this mismatch is due to our use of oversimplified models of speciation and extinction.

Fifteen years ago, Nee et al. (4) presented the first method to infer speciation and extinction rates on the basis of “reconstructed” phylogenies, i.e., phylogenies inferred on only extant species (Fig. 1 A and B). This first likelihood method relied on the idea that lineages in a reconstructed phylogeny accumulate through time with rate $\lambda - \mu$ (where $\lambda$ is the speciation rate and $\mu$ is the extinction rate) and accumulate in the very recent past with rate $\lambda$. The change in rate of lineage accumulation from $\lambda - \mu$ to $\lambda$, called the “pull-of-the-present” (5), allows us to estimate both $\lambda$ and $\mu$ given only data from living species.

The method of Nee et al. has been widely used for estimating speciation and extinction rates. Unfortunately, it often produces estimates of $\mu$ near zero (1, 2). Morlon et al. (3) suggest that the low extinction rate estimates might be due to the assumption that $\lambda$ and $\mu$ are constant, which is, in particular for large groups, most likely wrong. For large clades, we expect both rate heterogeneity through time due to environmental factors and rate heterogeneity across subclades due to subclade-specific traits influencing speciation and extinction rates.

Several likelihood-based approaches now exist that infer speciation and extinction rates under each of these two scenarios of rate heterogeneity, based on reconstructed phylogenies. The basic idea is the same for the different approaches: Speciation and extinction rates are determined that maximize the likelihood of the reconstructed tree. The challenge has been to derive an analytic formula for the likelihood of the tree under the complex dynamics. Three recent PNAS studies provide more general analytic likelihood functions:

Alfaro et al. (6) provide a likelihood approach in which the speciation and extinction rates may vary across subclades, but each subclade has a constant rate (Fig. 1D). The original method derivation is provided in ref. 7 and found in the package MEDUSA. Such a model allows for detecting subclade-specific speciation and extinction processes.

Stadler (8) relaxes the assumption of constant rates by allowing for rates changing at specific points in time (Fig. 1D). Such a model allows for detecting rapid changes in speciation and extinction rates due to environmental effects like at the Cretaceous–Tertiary boundary at 65 Ma.

Morlon et al. (3) extend the two methods such that rates may change continuously through time (instead of discretely as in ref. 8), and subclades may have different speciation and extinction rates (as in ref. 6) (Fig. 1C).

The models above have in common that the speciation and extinction rates within subclades are a function of time only, meaning that the rates are changing only due to external factors (i.e., the envi-

Author contributions: T.S. wrote the paper.

The author declares no conflict of interest.

See companion article 10.1073/pnas.1102543108.

1E-mail: tanja.stadler@env.ethz.ch.