Inferring speciation and extinction processes from extant species data

Tanja Stadler[1]

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 λ is the speciation rate and μ is the extinction rate) and accumulate in the very recent past with rate λ. The change in rate of lineage accumulation from $\lambda - \mu$ to λ, called the "pull-of-the-present" (5), allows us to estimate both λ and μ 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 μ near zero (1, 2). Morlon et al. (3) suggest that the low extinction rate estimates might be due to the assumption that λ and μ 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
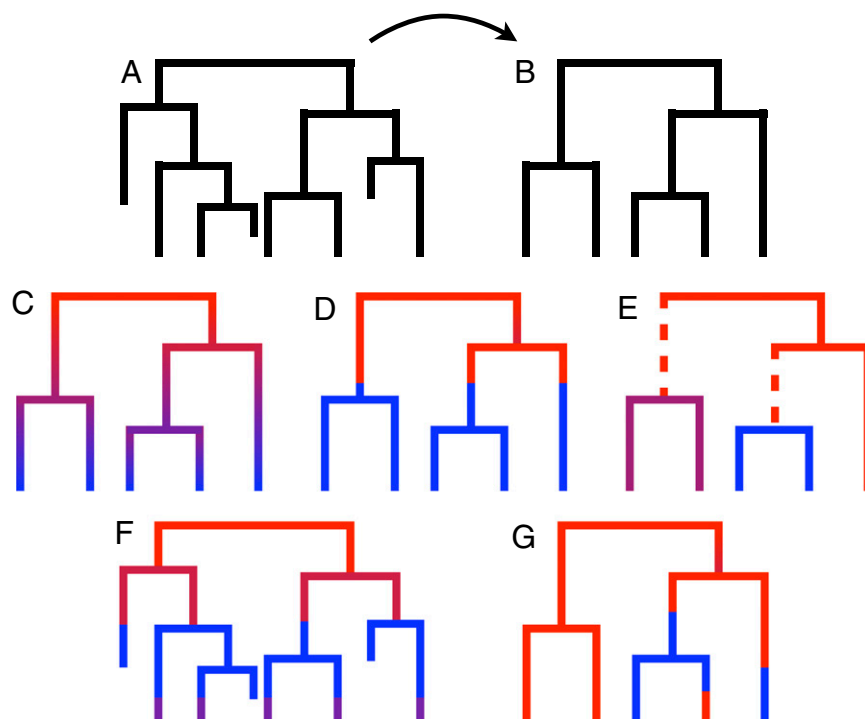
Fig. 1. (*A* and *B*) Complete phylogeny (*A*) with associated reconstructed phylogeny (*B*), which is obtained by suppressing all extinct lineages. (*C–G*) Models for speciation and extinction. Red denotes a fast rate, purple an intermediate rate, and blue a slow rate of speciation. *C–E* indicate the three models accounting for rate heterogeneity through time and across subclades: (*C*) Morlon et al. (3) (continuous change over time), (*D*) Stadler (8) (change at one time point), and (*E*) Alfaro et al. (6) (change in the dashed edges above subclades). (*F*) Density-dependent model (the complete phylogeny instead of the reconstructed phylogeny is plotted as the speciation rate depends on the total number of lineages). (*G*) Trait-dependent model by Maddison et al. (14) (changes between the two rates at arbitrary times).

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. 1*E*). 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. 1*D*). 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. 1*C*).

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-

ronment). In particular in ref. 6, the rates within subclades are constant.

Are the Speciation and Extinction Rates Governed by Environmental Effects?

For cetaceans (whales, dophins, and porpoises), Morlon et al. (3) show that incorporating rate heterogeneity produces speciation and extinction rate estimates in good agreement with the fossil record. In particular, the authors infer an exponentially decreasing speciation rate and a constant extinction rate for the cetacean subclade Balaenopteridae.

An immediate follow-up question is: Why is the speciation rate declining in this group? Is the speciation rate change really caused by the environment (which is implicitly assumed by the underlying model)?

An alternative explanation for a decreasing speciation rate is density-dependent speciation (e.g., ref. 9), meaning that the speciation rate depends on the number of species instead of time (Fig. 1F): A clade begins to radiate quickly into non-occupied niches, and the process slows down as the niches fill. The initial adaptive radiation is typically caused by a key innovation or a move to a new habitat of the ancestral species. There is no closed-form likelihood expression for density-dependent speciation models available yet. A first step is taken in ref. 10 where a likelihood function is provided for scenarios when the extinction rate is zero.

Density-dependent speciation yields a declining speciation rate through time; analog, density-dependent extinction yields an increasing extinction rate through time. A likelihood approach will allow us to formally contrast density-dependent against time-dependent models for clades such as the baleen whales.

Unfortunately, neither time-dependent nor density-dependent models can explain the speciation process completely, as the resulting trees induced by these models are too balanced: A reconstructed phylogeny is fully characterized by the times of lineage accumulation (continuous part)

and the ranked tree shape (11) (discrete part). Time-dependent and density-dependent models assume exchangeable species, meaning that at each point in time each species undergoes the same speciation and extinction mechanism. Aldous (12) showed that all exchangeable species models induce the same distribution on ranked tree shape (but can induce a wide variety of lineage accumulation patterns). Because empirical trees are less balanced than trees under exchangeable species models (13), the exchangeable species models lack an important feature.

Imbalance can be explained by trait-dependent speciation and extinction. Two PNAS studies (3, 6) take a first step to account for trait-dependent speciation by assuming heterogeneity across subclades. However, the change in trait is not explicitly modeled; a trait change (and thus a change in speciation and extinction rates) is added if the likelihood increases sufficiently. Maddison et al. (14) introduced a likelihood approach assuming discrete traits evolving under a Markov model (Fig. 1G); Fitzjohn (15) generalized the approach by allowing for continuous traits. However, these approaches lack the rate heterogeneity through time.

Are Models with Environmental-, Density-, and Trait-Dependent Speciation and Extinction Rates Sufficient?

Morlon et al. (3) obtain speciation and extinction rate estimates for whales that are in agreement with the fossil record. In particular, the extinction rates are not underestimated. This result suggests that we might be able to reconcile molecular phylogenies with the fossil record in general.

However, the bias of underestimating extinction rates in reconstructed phylogenies will not completely vanish with the methods accounting for environmental, density-dependent, and trait-dependent effects: Underestimating extinction is primarily due to the fact that very few reconstructed phylogenies have a pronounced pull-of-the-present effect; i.e., often the lineages do not accumulate

faster in the recent past. In fact, often the most recent lineage accumulation is slower instead of faster [very pronounced, e.g., for mammals (16) and birds (17)].

It is well recognized that this apparent slowdown can be due to simply not sampling all of the species, and approaches dealing with this missing data problem have recently become available (18, 19).

However, we observe the slowdown even in complete phylogenies like mammals. The reason for this slowdown may be that very recent speciation events have not yet been identified (20). The delay of recognizing species after two populations started diverging needs to be recognized by our methods to avoid biases in speciation and extinction rate estimates.

Future Challenges

In large clades, speciation and extinction rates were most likely not constant throughout evolutionary time. Rates change due to (i) a changing environment, (ii) density-dependent speciation and extinction, and (iii) trait-dependent speciation and extinction. The study in PNAS (3) improves our understanding of scenario i, but we require more work on scenario ii to test scenario i against ii. Scenarios i and ii will help account for the mode of lineage accumulation. With an improved understanding of scenario iii, we may also be able to account for tree shape. As lineage accumulation and tree shape fully describe a reconstructed phylogeny, a combination of models i–iii may have the power to describe the process of speciation and extinction.

However, in addition, we need to take into account the mode of data collection to avoid biases: We have to be aware that we may not be able to recognize very young species and we may not sample all extant species of a clade.

It is a puzzle for the future to combine the time-dependent, density-dependent, and trait-dependent models with the data collection biases to produce one unified approach that can be used in a likelihood framework. Morlon et al. (3) give us an elegant version of the time-dependent piece to fit in.

1. Nee S (2006) Birth-death models in macroevolution. *Annu Rev Ecol Evol Syst* 37:1–17.
2. Purvis A (2008) Phylogenetic approaches to the study of extinction. *Annu Rev Ecol Evol Syst* 39:301–319.
3. Morlon H, Parsons TL, Plotkin JB (2011) Reconciling molecular phylogenies with the fossil record. *Proc Natl Acad Sci USA*, 10.1073/pnas.1102543108.
4. Nee S, May RM, Harvey PH (1994) The reconstructed evolutionary process. *Philos Trans R Soc Lond B Biol Sci* 344:305–311.
5. Nee S, Holmes EC, May RM, Harvey PH (1994) Extinction rates can be estimated from molecular phylogenies. *Philos Trans R Soc Lond B Biol Sci* 344:77–82.
6. Alfaro ME, et al. (2009) Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proc Natl Acad Sci USA* 106:13410–13414.
7. Rabosky DL, Donnellan SC, Talaba AL, Lovette IJ (2007) Exceptional among-lineage variation in diversification

rates during the radiation of Australia's most diverse vertebrate clade. *Proc Biol Sci* 274:2915–2923.
8. Stadler T (2011) Mammalian phylogeny reveals recent diversification rate shifts. *Proc Natl Acad Sci USA* 108: 6187–6192.
9. Phillimore AB, Price TD (2008) Density-dependent cladogenesis in birds. *PLoS Biol* 6:e71.
10. Rabosky DL, Lovette IJ (2008) Density-dependent diversification in North American wood warblers. *Proc Biol Sci* 275:2363–2371.
11. Gernhard T (2008) The conditioned reconstructed process. *J Theor Biol* 253:769–778.
12. Aldous DJ (2001) Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. *Stat Sci* 16:23–34.
13. Blum MG, François O (2006) Which random processes describe the tree of life? A large-scale study of phylogenetic tree imbalance. *Syst Biol* 55:685–691.

14. Maddison WP, Midford PE, Otto SP (2007) Estimating a binary character's effect on speciation and extinction. *Syst Biol* 56:701–710.
15. FitzJohn RG (2010) Quantitative traits and diversification. *Syst Biol* 59:619–633.
16. Bininda-Emonds ORP, et al. (2007) The delayed rise of present-day mammals. *Nature* 446:507–512.
17. Klicka J, Zink RM (1997) The importance of recent ice ages in speciation: A failed paradigm. *Science* 277: 1666–1669.
18. Cusimano N, Renner SS (2010) Slowdowns in diversification rates from real phylogenies may not be real. *Syst Biol* 59:458–464.
19. Höhna S, Stadler T, Ronquist F, Britton T (2011) Inferring speciation and extinction rates under different species sampling schemes. *Mol Biol Evol* 28:2577–2589.
20. Avise JC, Walker D (1998) Pleistocene phylogeographic effects on avian populations and the speciation process. *Proc Biol Sci* 265:457–463.