PHYLOGENETICS

Species-specific diversification

A model-based approach allows quantification of lineage-specific speciation and extinction rates on the basis of phylogenetic trees.

Tanja Stadler

Speciation and extinction rates may vary both across lineages and through time. In particular, as reconstructed phylogenies are becoming larger, the assumption that these rates will remain constant becomes unrealistic. An accurate quantification of this rate variation remains challenging but would allow us to determine the underlying factors influencing these rates and deepen our understanding of macroevolutionary processes. In this issue, Maliet et al.[1] introduce a new method to quantify variable speciation and extinction rates.

Quantifying rates that are assumed to vary through time but not between lineages has been done successfully for some time[2,3]. The underlying mathematical theory is relatively straightforward. It assumes that lineages are 'exchangeable' at any point in time, meaning that all co-existing lineages undergo the same speciation and extinction dynamics. However, allowing for non-exchangeability means that some co-existing lineages may speciate or go extinct faster than others (that is, co-existing lineages exhibit rate variation). This means that the process of speciation and extinction is assumed to be non-neutral, with some lineages being under selection on the macroevolutionary level. Assuming such selective processes are at work when quantifying macroevolutionary dynamics is far from trivial. Up until now, there were two main approaches for quantifying rate variation across lineages. Both approaches rely on a stochastic birth–death model: each lineage may undergo birth (speciation) or death (extinction) with some lineage-specific rate. This rate may change through time, causing rate variation across co-existing lineages.

In the first approach, only the reconstructed phylogenetic tree (Fig. 1, right) is considered. On the basis of the branching pattern, speciation and extinction rates are assigned to the lineages. Intuitively, parts of the tree with many branching events have higher diversification rates (calculated as speciation rate – extinction rate) than parts with few branching events. The first
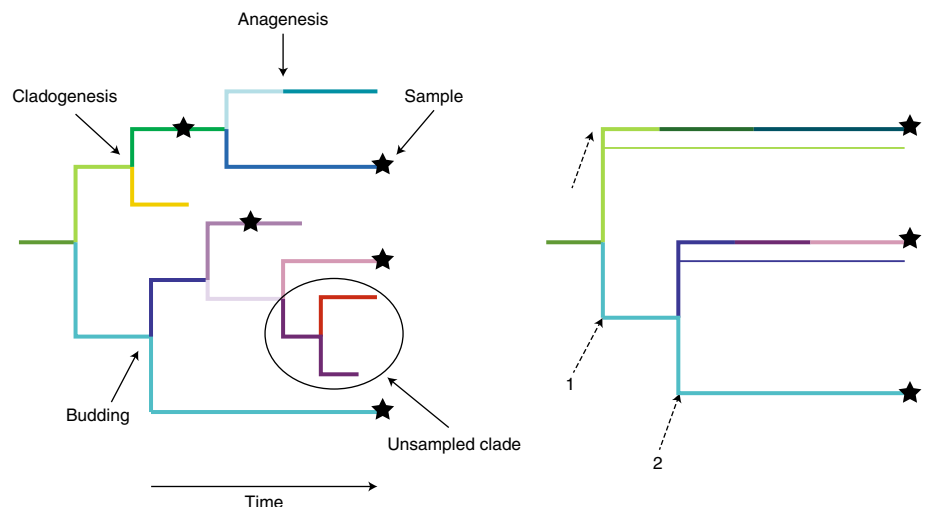

Fig. 1 | Phylogenetic tree with lineage-specific diversification rates. Left, An example of a complete phylogeny. The branching pattern of all five extant species that evolved from a single ancestor is depicted. Each branch colour indicates a unique pair of speciation and extinction rates for that branch. The stars indicate samples — three present-day samples and two fossil samples. The circle encloses an unsampled clade with one unsampled extinct and one unsampled extant tip; common methods for estimating speciation and extinction rates assume that no rate change happens in such clades. Three modes of speciation — cladogenesis, anagenesis and budding — are illustrated. Right, A reconstructed phylogeny based only on the extant species. Dashed arrows indicate the time points at which the speciation and extinction rates are estimated using the method by Maliet et al.[1]. The descending branches (thin lines) are returned in the colour corresponding to the rate at the corresponding time point. Thus, changes along branches are not estimated; however, any potential changes along branches are correctly taken into account by the method. Since Maliet et al. attribute all changes in rates to unobserved cladogenetic events, the method assumes the colours associated with arrows 1 and 2 are different even if they actually are not.

popular software that used this idea was modelling evolutionary diversification using stepwise AIC (MEDUSA)[4], with Bayesian analysis of macroevolutionary mixtures (BAMM)[5] being a subsequent improvement. BAMM has had problems in accurately determining diversification rates[6], although some of this criticism has been addressed[7]. However, all of these methods assume that there are no changes in speciation and extinction rates in extinct or non-sampled parts of the trees (Fig. 1 left, circle). Thus, these methods implicitly assume that there are few drastic shifts along the tree, with a priori assumptions regarding the number

of these shifts[4,5] or the number of rate categories[7]. These methods are expected to perform poorly in the presence of many small changes in speciation and extinction rates, as such changes would also affect the extinct or non-sampled parts of a tree. In the second approach, the reconstructed phylogenetic tree is considered together with some traits assigned to the sampled species. Although this approach is ideal for quantifying rates, as it uses all data available (that is, the tree and traits of sampled species) and takes into account changes in extinct or unsampled subtrees, problems occur because we typically do not know

the traits that determine speciation and extinction rates[8]. The hidden state speciation and extinction (HiSSE)[9] model has been introduced as an attempt to overcome such problems.

Maliet et al.[1] improve on the first set of methods. They assume that the two daughter species each inherit the speciation and extinction rates of the mother species, altered by some noise. Thus, the two daughter species will have rates that are different from each other and the mother species. The authors derive the probability density of the reconstructed phylogeny together with the speciation and extinction rates at the start of each branch (Fig. 1, right, dashed arrows). This probability density calculation is done without any approximations, in particular overcoming the assumption that there are no rate changes in extinct and non-sampled parts of the phylogeny (Fig. 1, left, circle). The probability density calculation is implemented within a Markov chain Monte Carlo (MCMC) framework to estimate the posterior distribution of the branch-specific speciation and extinction rates (that is, the rates at the start of each branch) for a given reconstructed phylogenetic tree. In that way, the authors can quantify diversification rates that potentially undergo many small changes on the basis of phylogenies only. In their simulation study, the authors show that the method can also successfully determine if large shifts are present in a phylogeny. An analysis of a bird phylogeny reveals that empirical data indeed suggest there are many small changes within clades and few large shifts between clades. In particular, for some of the bird clades, within-clade rate variation is estimated to be as high as between-clade rate variation.

The study could stimulate a number of downstream methodological developments to overcome some of its current limitations. First, as in previous work, extinction rates are still hard to quantify. The intuitive reason is that we do not see any extinction events in the reconstructed phylogeny, as the method does not include fossil samples. Conceptually, it should be straightforward to extend the method to reconstructed phylogenies with fossils[10] (Fig. 1, left), allowing disentanglement of speciation and extinction rates. Second, throughout the paper, a cladogenetic view on speciation is assumed. A cladogenetic speciation event means that the mother lineage terminates and two daughter lineages start. However, anagenetic speciation (when the mother species evolves into a daughter species without a branching event) or budding speciation (when the mother species gives rise to a daughter species and continues to exist) modes may be as important[11,12] (Fig. 1, left). The paper by Maliet et al. provides a great starting point to extend speciation and extinction rate estimations to these general macroevolutionary scenarios. Such extensions would not only model the trait evolution process more accurately, but also allow assessment of the relative importance of the different speciation modes. Finally, through these generalizations, I envision the presented method to be of great use in areas beyond macroevolution. For example, when quantifying transmission rates in epidemiology, a budding birth mode is required (as one host transmits to a second host), potentially with anagenesis (as one host may move to a different location and thus potentially change its associated rates), and samples are collected sequentially through time like fossils[13].

Future methodological developments along these lines will allow us to more reliably quantify selection in phylogenetic trees across a range of applications beyond macroevolution, which in turn will allow us to determine the traits responsible for fitness differences. Knowledge of these traits not only allows us to obtain an improved understanding of evolution, but can further help in setting up improved policies in areas such as conservation biology or public health. ❐

Tanja Stadler [ID]

Department of Biosystems Science and Engineering, ETH Zürich, Zürich, Switzerland.
e-mail: tanja.stadler@bsse.ethz.ch

References

1. Maliet, O., Hartig, F. & Morlon, H. *Nat. Ecol. Evol.* http://doi.org/10.1038/s41559-019-0908-0 (2019).
2. Morlon, H., Parsons, T. L. & Plotkin, J. B. *Proc. Natl Acad. Sci. USA* **108**, 16327–16332 (2011).
3. Stadler, T. *Proc. Natl Acad. Sci. USA* **108**, 6187–6192 (2011).
4. Alfaro, M. E. et al. *Proc. Natl Acad. Sci. USA* **106**, 13410–13414 (2009).
5. Rabosky, D. L. *PLoS One* **9**, e89543 (2014).
6. Moore, B. R., Höhna, S., May, M. R., Rannala, B. & Huelsenbeck, J. P. *Proc. Natl Acad. Sci. USA* **113**, 9569–9574 (2016).
7. Barido-Sottani, J., Vaughan, T.G. & Stadler, T. Preprint at https://doi.org/10.1101/440982 (2018).
8. Rabosky, D. L. & Goldberg, E. E. *Syst. Biol.* **64**, 340–355 (2015).
9. Beaulieu, J. M. & O'Meara, B. C. *Syst. Biol.* **65**, 583–601 (2016).
10. Heath, T. A., Huelsenbeck, J. P. & Stadler, T. *Proc. Natl Acad. Sci. USA* **111**, E2957–E2966 (2014).
11. Foote, M. *Paleobiology* **22**, 141–151 (1996).
12. Silvestro, D., Warnock, R. C. M., Gavryushkina, A. & Stadler, T. *Nat. Commun.* **9**, 5237 (2018).
13. Kühnert, D., Stadler, T., Vaughan, T. G. & Drummond, A. J. *Mol. Biol. Evol.* **33**, 2102–2116 (2016).