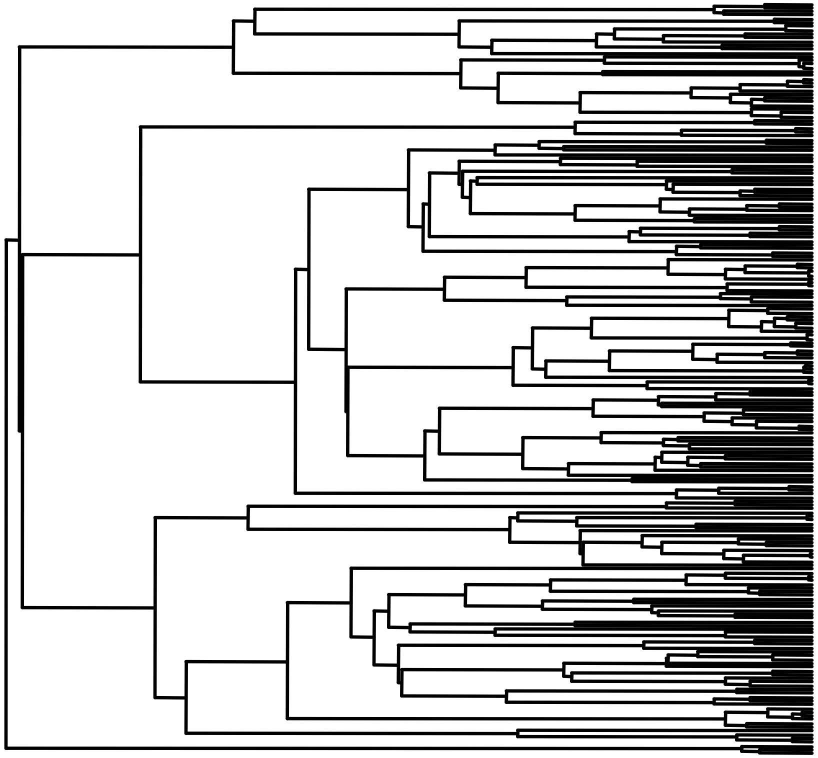
Estimating diversification rates from trees

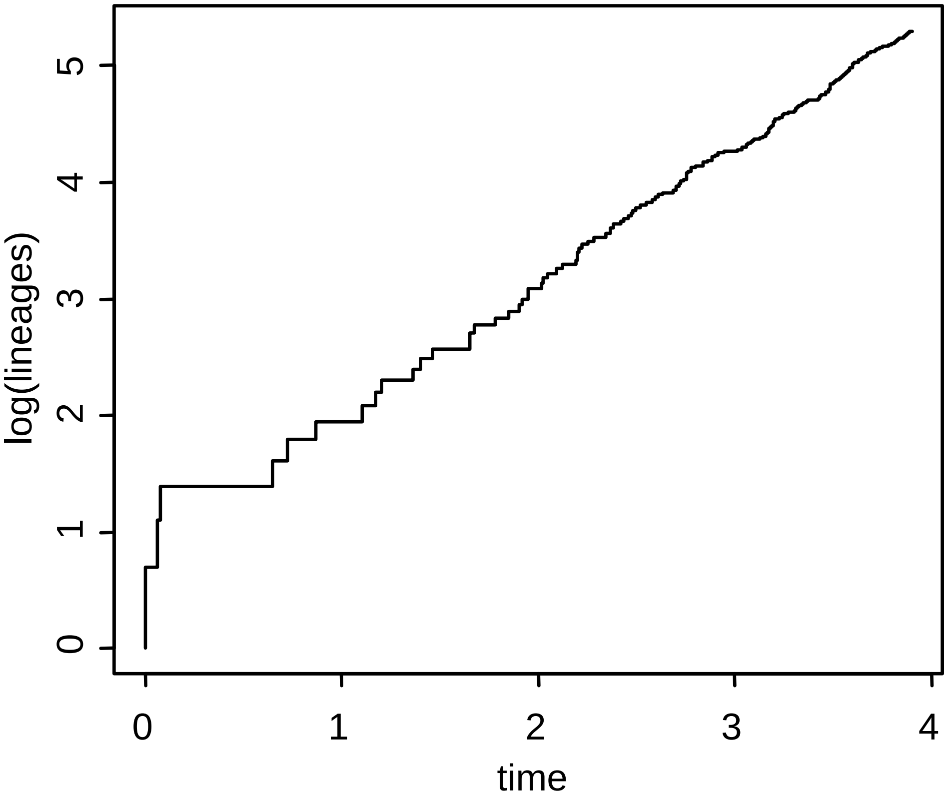
Liam J. Revell

14 December 2016

Diversification analysis: Lineage-through-time plot

Under a pure-birth model of species accumulation, the number of lineages accrues log-linearly with time.





Diversification rate analysis: The γ-statistic

A simple related statistic was devised by Pybus & Harvey (2000) and it is called the gamma(γ)-method or the gamma-statistic.

Gamma is calculated as follows:

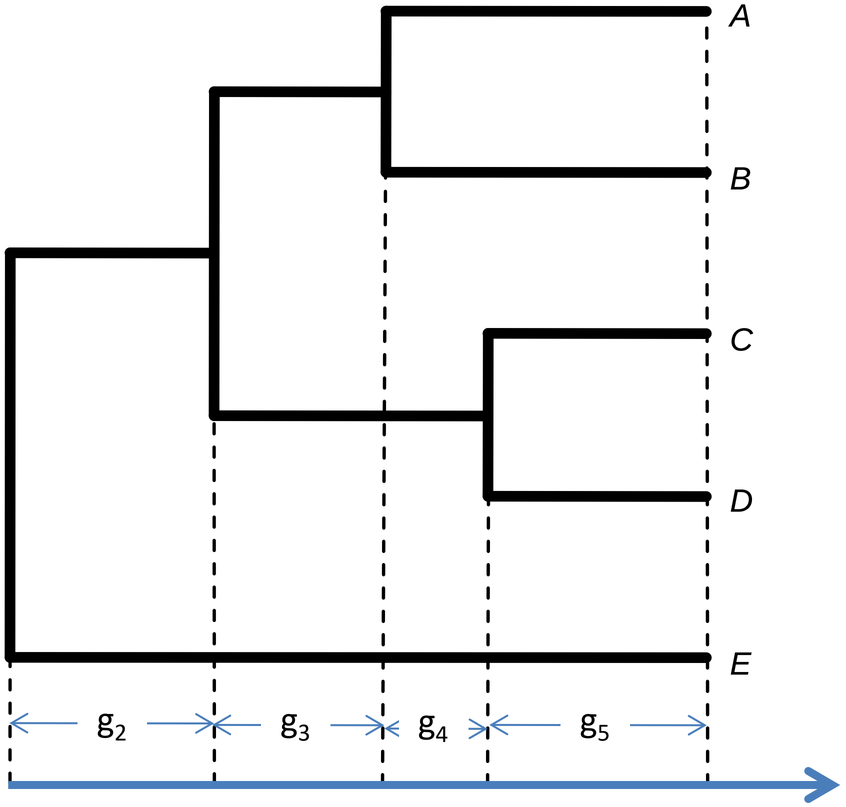
$$\gamma = \frac{\left(\frac{1}{n-2}\sum_{i=2}^{n-1}\sum_{k=2}^{i} kg_k\right) - \left(\frac{T}{2}\right)}{T\sqrt{\frac{1}{12(n-2)}}}, \; T = \sum_{j=2}^{n} jg_j$$

g is just a vector containing the internode distances from our phylogeny (on any scale).

Though complicated, this statistic is useful because it has a *standard normal distribution* for constant-rate pure-birth trees.

Diversification rate analysis

g is just a vector containing the internode distances from our phylogeny (on any scale), that is:



Diversification rate analysis

How can gamma be used?

Significantly negative gamma suggests that the speciation rate was initially high, but then slowed through time.

Significant positive gamma suggests the converse; i.e., that the speciation rate was initially slow, and increased over time.

Since the statistic has a *standard normal distribution* we can test hypotheses about gamma, by using a Z-test.

One caveat that needs to be attached to this, is that positive gamma can also be caused by *unseen past extinction*. This is due to a phenomenon that is known as the “pull of the present.”

Estimating speciation & extinction rates

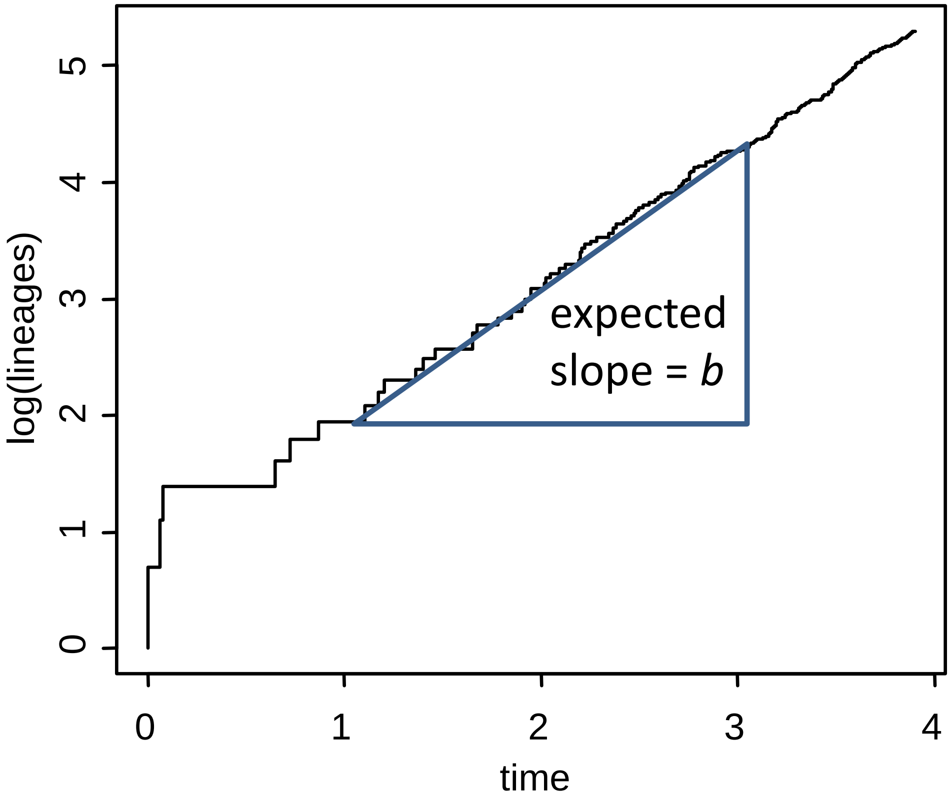
Although these methods are very useful and good a telling us something about the diversification process through time, they do not provide us with estimates of the parameters of the underlying evolutionary process.

However, it is possible to estimate the rates of speciation and extinction on a reconstructed phylogeny – even if no extinct lineages are present in the tree.

Lineage accumulation under pure-birth

Under a pure-birth model of species accumulation, the number of lineages accrues log-linearly with time.

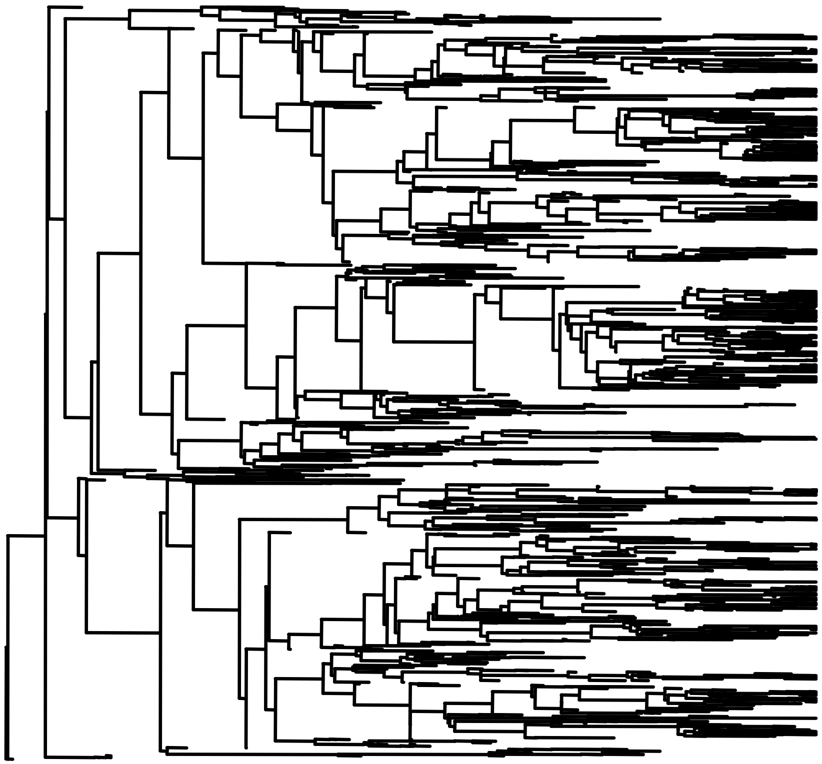
Furthermore, the slope of this line has an expectation that is equal to the *speciation rate*.

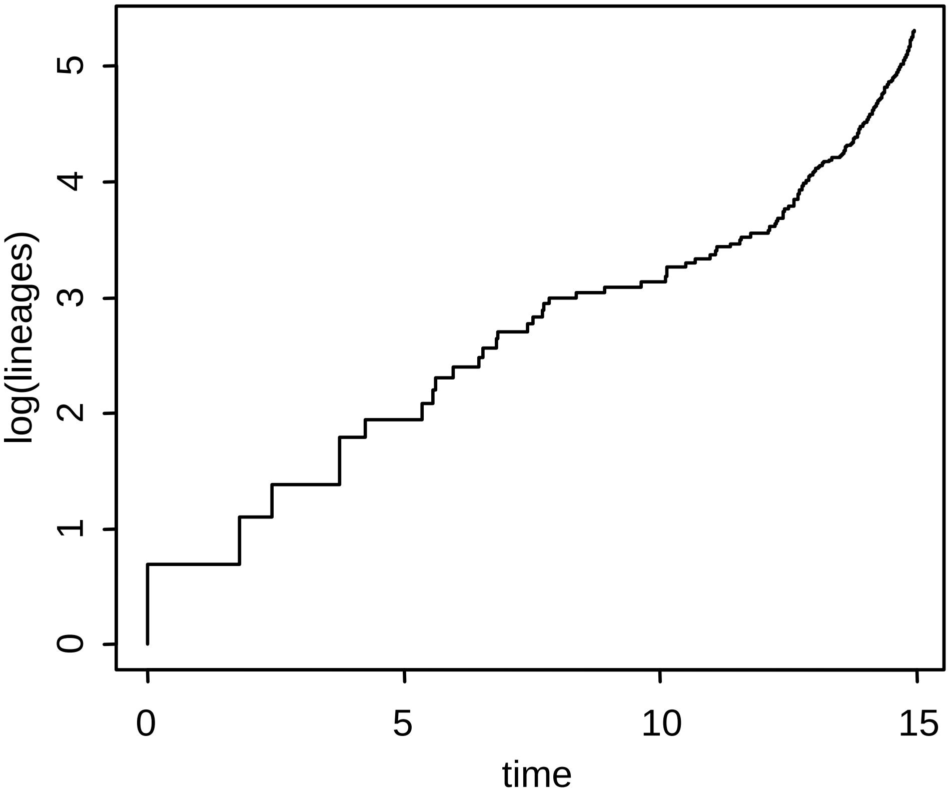


“Pull of the present”

With extinction (death) added, we see a different pattern.

Now, the shape of the graph is no longer linear:





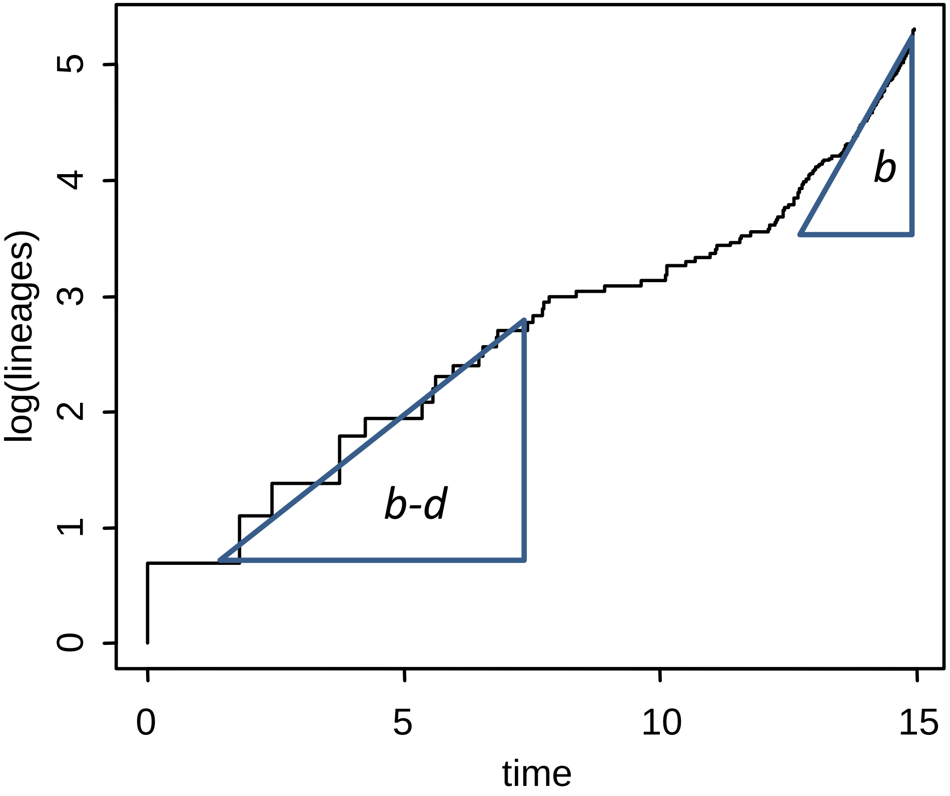
“Pull of the present”

With extinction (death) added, we see a different pattern.

Now, the shape of the graph is no longer linear:

In particular, the expected slope at the base of the tree is *b* – *d*;

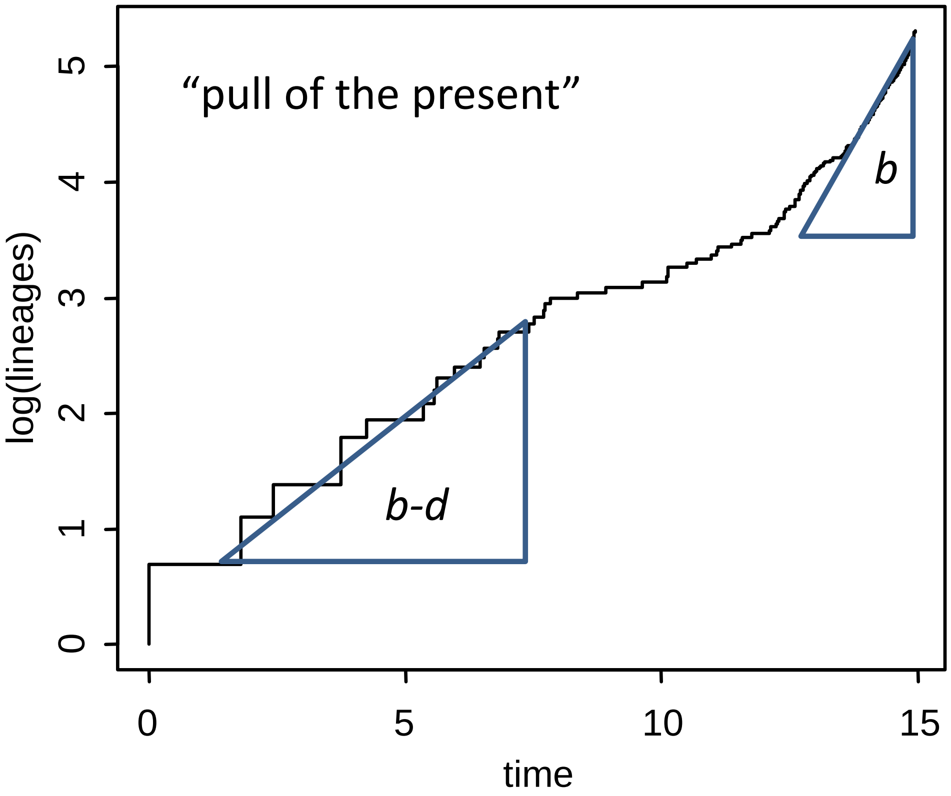
Whereas the expected slope at the tips of the tree is *b*.



Estimating speciation & extinction rates

Nee et al. (1994) realized that we could take advantage of this phenomenon, that is, the pull of the present, to estimate speciation *and* extinction rates from phylogenies with branch lengths.

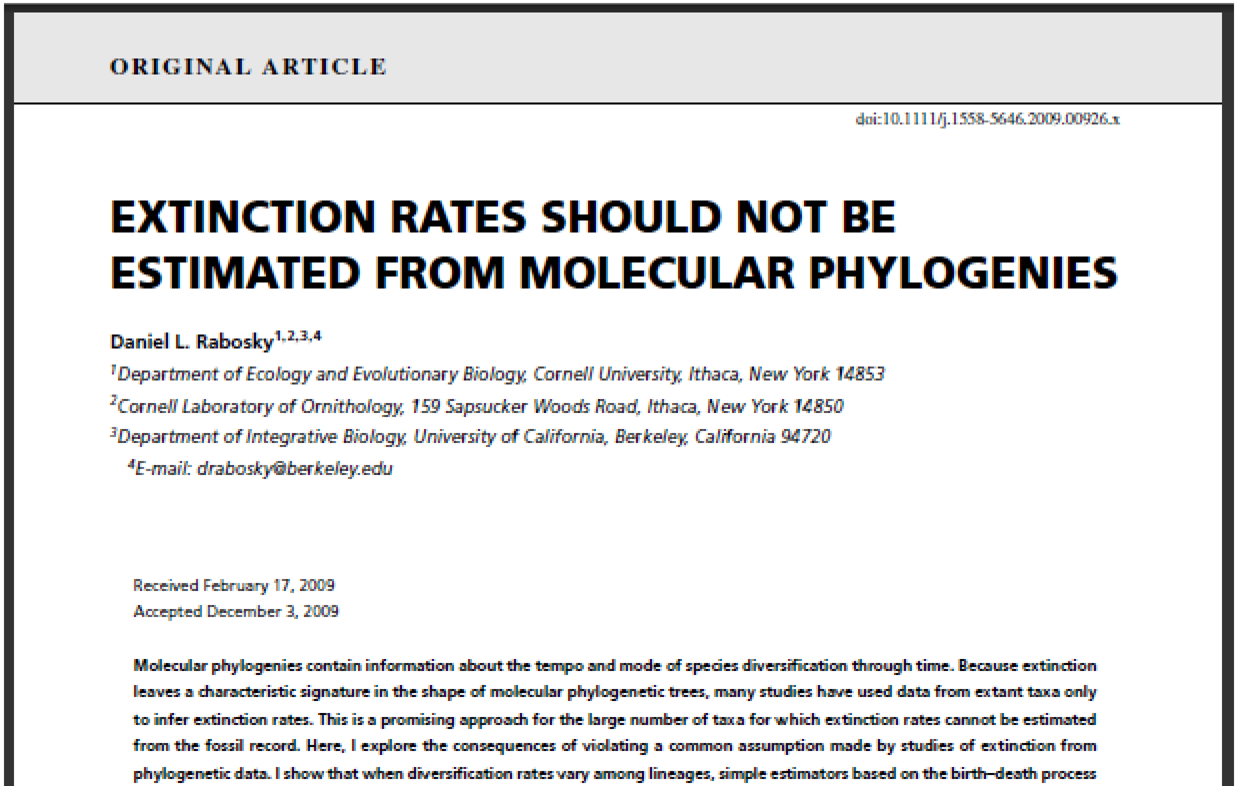
He developed a likelihood-method which uses relative branching times to estimate the birth and death rates on a tree.



Estimating speciation & extinction rates

Although Nee's likelihood method is very good at estimating speciation & extinction rates when the model assumptions hold, it has become clear in recent years that the method is *highly sensitive to model assumptions*.

In fact, Dan Rabosky recently published the following article:



ORIGINAL ARTICLE

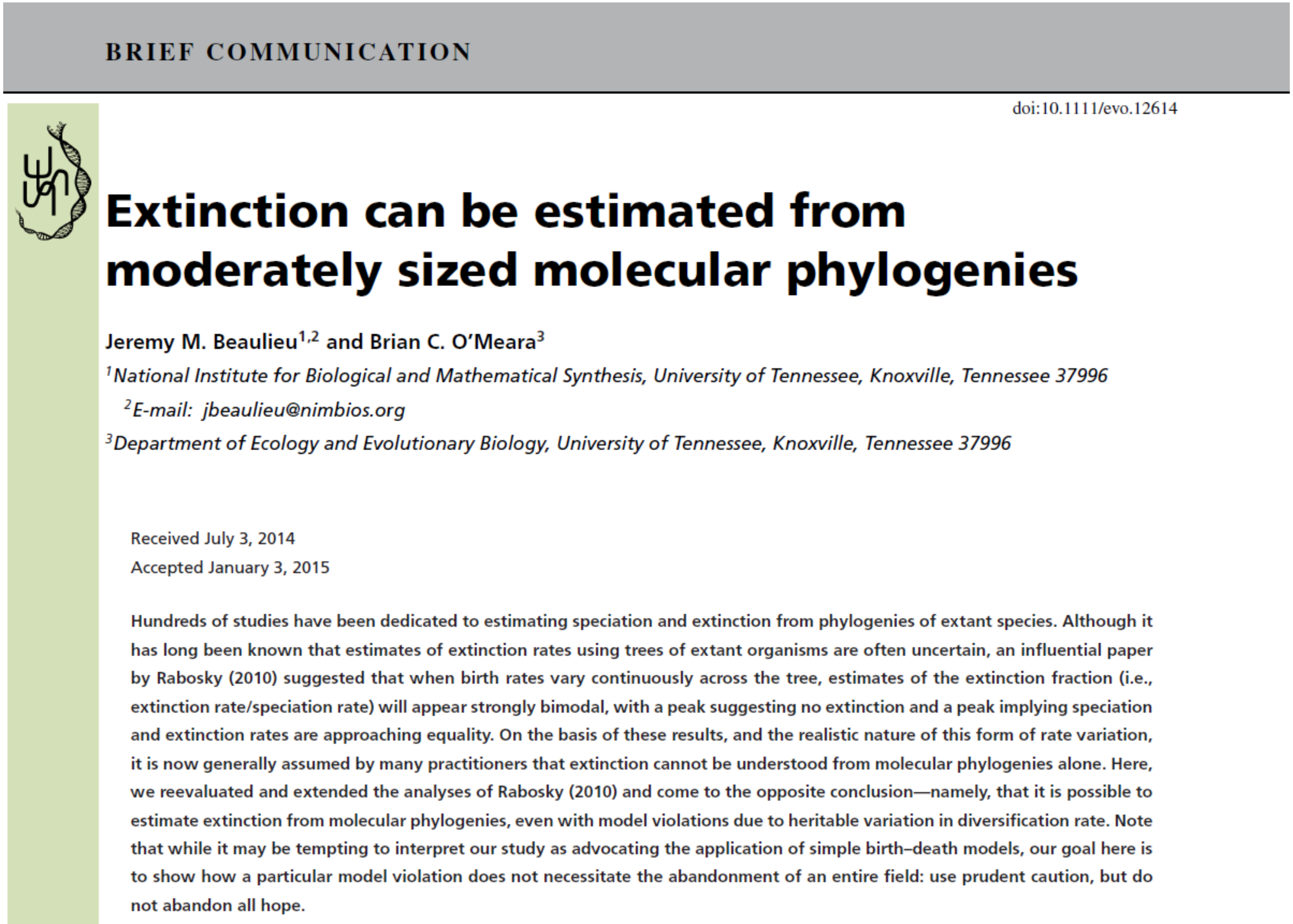
doi:10.1111/j.1558-5646.2009.00926.x

EXTINCTION RATES SHOULD NOT BE ESTIMATED FROM MOLECULAR PHYLOGENIES

Daniel L. Rabosky[1,2,3,4]
[1] *Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York 14853*
[2] *Cornell Laboratory of Ornithology, 159 Sapsucker Woods Road, Ithaca, New York 14850*
[3] *Department of Integrative Biology, University of California, Berkeley, California 94720*
[4] *E-mail: drabosky@berkeley.edu*

Received February 17, 2009
Accepted December 3, 2009

Molecular phylogenies contain information about the tempo and mode of species diversification through time. Because extinction leaves a characteristic signature in the shape of molecular phylogenetic trees, many studies have used data from extant taxa only to infer extinction rates. This is a promising approach for the large number of taxa for which extinction rates cannot be estimated from the fossil record. Here, I explore the consequences of violating a common assumption made by studies of extinction from phylogenetic data. I show that when diversification rates vary among lineages, simple estimators based on the birth–death process



BRIEF COMMUNICATION

doi:10.1111/evo.12614

Extinction can be estimated from moderately sized molecular phylogenies

Jeremy M. Beaulieu[1,2] and Brian C. O'Meara[3]

[1]*National Institute for Biological and Mathematical Synthesis, University of Tennessee, Knoxville, Tennessee 37996*

[2]*E-mail: jbeaulieu@nimbios.org*

[3]*Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee 37996*

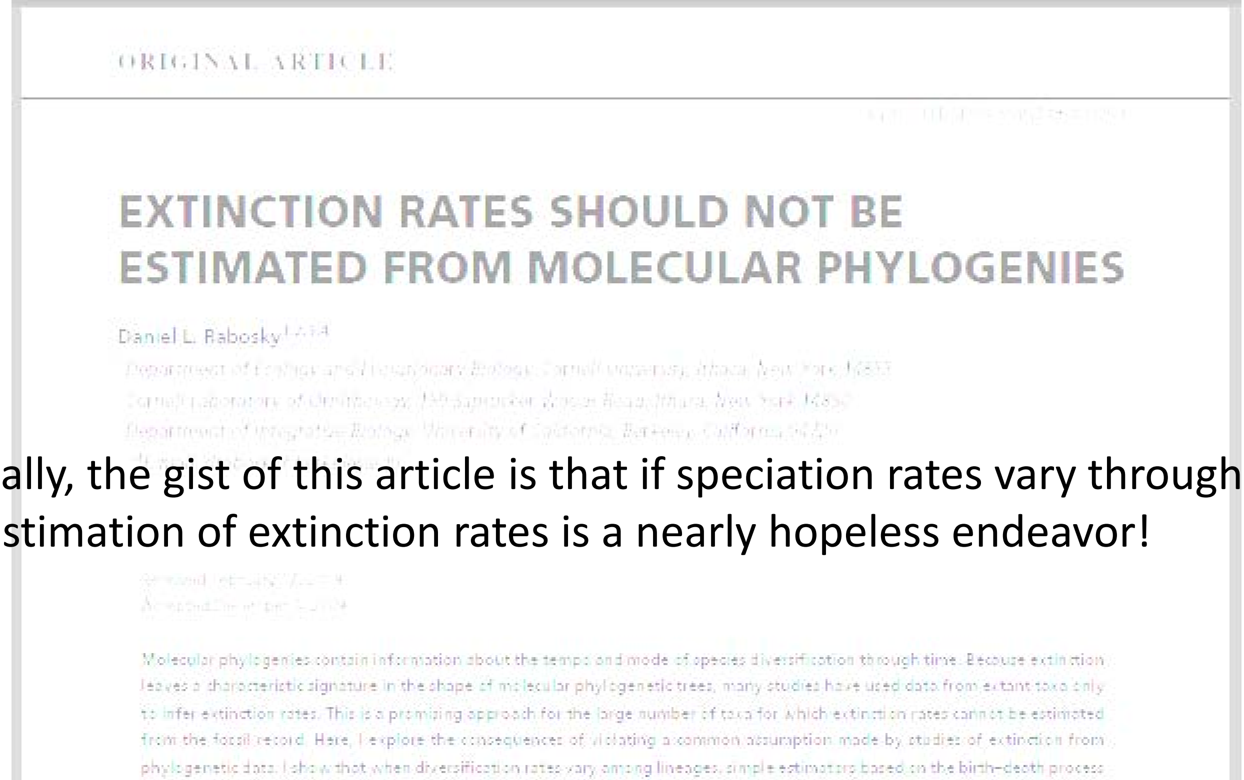
Received July 3, 2014
Accepted January 3, 2015

Hundreds of studies have been dedicated to estimating speciation and extinction from phylogenies of extant species. Although it has long been known that estimates of extinction rates using trees of extant organisms are often uncertain, an influential paper by Rabosky (2010) suggested that when birth rates vary continuously across the tree, estimates of the extinction fraction (i.e., extinction rate/speciation rate) will appear strongly bimodal, with a peak suggesting no extinction and a peak implying speciation and extinction rates are approaching equality. On the basis of these results, and the realistic nature of this form of rate variation, it is now generally assumed by many practitioners that extinction cannot be understood from molecular phylogenies alone. Here, we reevaluated and extended the analyses of Rabosky (2010) and come to the opposite conclusion—namely, that it is possible to estimate extinction from molecular phylogenies, even with model violations due to heritable variation in diversification rate. Note that while it may be tempting to interpret our study as advocating the application of simple birth–death models, our goal here is to show how a particular model violation does not necessitate the abandonment of an entire field: use prudent caution, but do not abandon all hope.

Estimating speciation & extinction rates

Although Nee’s likelihood method is very good at estimating speciation & extinction rates when the model assumptions hold, it has become clear in recent years that the method is *highly sensitive to model assumptions*.

In fact, Dan Rabosky recently published the following article:



ORIGINAL ARTICLE

EXTINCTION RATES SHOULD NOT BE ESTIMATED FROM MOLECULAR PHYLOGENIES

Daniel L. Rabosky

Molecular phylogenies contain information about the tempo and mode of species diversification through time. Because extinction leaves a characteristic signature in the shape of molecular phylogenetic trees, many studies have used data from extant taxa only to infer extinction rates. This is a promising approach for the large number of taxa for which extinction rates cannot be estimated from the fossil record. Here, I explore the consequences of violating a common assumption made by studies of extinction from phylogenetic data. I show that when diversification rates vary among lineages, simple estimators based on the birth–death process

Basically, the gist of this article is that if speciation rates vary through time, the estimation of extinction rates is a nearly hopeless endeavor!

Estimating speciation & extinction rates

- Diversification (speciation & extinction) can be estimated purely from reconstructed phylogenies.
- However, the estimation of speciation & extinction is sensitive to violations of the assumptions of the model.
- We will next learn about more flexible models in which the diversification rates (speciation & extinction) change through time or depending on the state of a discrete or continuous trait.