

GOAL

This annotated bibliography briefly introduce some literature that is relevant to study: 1) The birth-death process as a model of species diversification; 2) The birth-death as a prior for calibration phylogenies' divergence times; and 3) The fossilized birth-death. Hopefully this will provide some background to the discussion of Heath et al 2014.

PART 1: Birth-Death process and diversification

Yule, 1925 - Phil. Trans. R. Soc. Lond.

A Mathematical Theory of Evolution, based on the conclusions of Dr. J.C. Willis, F.R.S.

This reference couldn't be left out of this annotated bibliography. This paper was the first to propose a model of pure birth process (or Yule process) to describe the accumulation of species through time. In this model, the older the clade the larger the number of species it is expected to have.

Kendall, 1948 - The Annals of Mathematical Statistics

On the Generalized "Birth-and-Death" Process

This paper develops the mathematical basis of the birth-death process. It provides equations and solutions to describe how population size changes through time (t) if it depends on the rates of births (λ) and deaths (μ). The specific case of constant rates was already known, in this paper Kendall generalized the solutions for scenarios where the rates are a function of time, *i.e.*, $\lambda(t)$ and $\mu(t)$. Clearly this is a paper with dense mathematics (at least for me) and it might not be a pleasant reading for everyone. However it is a very useful reference to keep in mind because it provides equations to calculate, for example, the expected mean population size and its variance in time t for a given $\lambda(t)$ and $\mu(t)$. Also there are derivations that allow one to calculate the chance of a population going extinct given the birth-death rates and considering the stochasticity of the process. Although the paper is in terms of population size the same math applies for species richness (or even higher taxa) if one considers births as originations and deaths as extinctions.

Raup, 1985 - Paleobiology - (Review)

Mathematical Models of Cladogenesis

This classic review paper starts discussing the role of stochastic and deterministic processes to explain how clades wax and wane (as observed in the fossil record). Raup explains how the simplest model of constant origination and extinction rates can be used to predict the diversity of a clade changing through time, emphasizing the stochasticity of the model. Furthermore, he discuss a few modifications of the model that have been used to study the evolutionary history of clades in paleontology, for example, to incorporate mass extinctions or time dependent rates. There is also an interesting development on the implications of using this

models to understand the patterns of diversity changing through time. I consider this to be very clear and insightful paper because it describes mathematical models in a very accessible way and applied to questions of interest to biologists and paleontologists.

Nee et al. 1994 - Phil. Trans. R. Soc. Lond.

a - Extinction rates can be estimated from molecular phylogenies; &

b - The reconstructed Evolutionary Process

I combined these two papers because they are very related, they are authored by the same people (mostly) and came out in the same issue. While paper *a* gives a graphical visualization of how molecular phylogenies can inform the birth-death rates of a clade, paper *b* develops the mathematical basis of the process. The authors derive the equations for the birth-death process in phylogenies. This derivation comes from a comparison between the “actual” history of a clade and its “reconstructed” history. The “actual” history includes all the extant and extinct lineages of a group (which isn’t known in most cases). The “reconstructed” history is the phylogeny built with lineages that survived to the present. Using this comparison the authors derive the equations to calculate the expected time between birth events (speciations) in the reconstructed process (a normal phylogeny). Furthermore, they derive equations to estimate speciation and extinction rates from time calibrated phylogenies.

Foote 1996 - Paleobiology

On the probability of ancestors in the fossil record

In this paper Foote calculates the probability of a given fossil occurrence being the direct ancestor of another fossil species, *i.e.*, the chance of finding pairs of ancestors-descendant (A-D) species in the fossil record. For that purpose he assumes the same rate of fossilization among different lineages and that this rate is constant through time. On his own words: “even under pessimistic assumptions regarding the completeness of the fossil record, the probability of finding fossil ancestor-descendant pairs is not negligible”. Furthermore, there is a comparison on the chance of finding A-D pairs among three different models of species formation: budding, cladogenesis and anagenesis. There are some differences in the expected number of A-D pairs depending on the model of species emergency, with more pairs being found in a budding model of speciation.

Maddison et al. 2007 - Syst. Biol

Estimating a Binary Character's Effect on Speciation and Extinction

* This youtube video shows Sarah Otto (one of the authors) explaining the model described in the paper: <https://www.youtube.com/watch?v=ryBzouc4TQo> and is very helpful!

This paper describes the BiSSE model, in which speciation and extinction rates are inferred from phylogenies in case those rates are affected by binary state characters (thus the acronym BiSSE: Binary-State Speciation and Extinction). The motivation for developing the model was to estimate a rate of speciation and a rate of extinction associated with each of the two

states of a binary character. Previous analysis with similar goals used a comparison of sister clades (e.g. evolution of herbivory occurred multiple times in insects and resulted in clades with higher diversity than their non-herbivore sister group). The paper derives likelihood equations for the models. With these equations, it introduces the idea of a speciation event that happened in the past leading to a lineage that is currently extinct, and how to calculate the probability of that happening. Furthermore the paper briefly explains how the method is implemented, which can be useful for people developing methods.

Note 1: After the original article this method has been generalized to Multi-state characters (MuSSE), quantitative state characters (QuaSSE) and many others.

Note 2: The *SSE group of methods have been criticized for yielding higher than expected Type I error, and some corrections have been proposed. See Rabosky & Goldberg (2015, Syst. Biol., Model Inadequacy and Mistaken Inferences of Trait-Dependent Speciation) which is not in this annotated bibliography.

Quental & Marshall 2010 - TREE

Diversity Dynamics: Molecular Phylogenies need the Fossil Record

This paper shows that researchers should use caution when interpreting diversity dynamics from molecular phylogenies only, and Cetaceans are used as an example. With the methods available up to that time, the molecular phylogeny of Cetaceans didn't refute a constant rate birth-death process. However, the fossil record of the group is fairly well known and shows a larger diversity of Cetaceans in the past. In this case, the methods that existed were not able to detect diversity declines. This and other papers (e.g. Rabosky, 2010 - Extinction rates should not be estimated from molecular phylogenies) were an impulse for the development of new methods that use molecular phylogenies to estimate speciation and extinction rates changing through time under complex scenarios.

Morlon, H. 2014 - Ecol Lett. - (Review)

Phylogenetic approaches for studying diversification

This is an excellent review of methods that use molecular phylogenies to estimate speciation and extinction rates. In total 13 types of models are described. The first one to be discussed is the constant rates birth-death models. Then many factors that could drive the variation of those rates are discussed as different models. For example, models with rates that are time-dependant or diversity-dependent, as well as models with clade-specific diversification and character-dependent diversification. The paper brings an extremely useful table (3 pages long!) with a comparison of the different methods. The review includes a very interesting section on empirical trends that have been observed after almost two decades of phylogenies being used to study the accumulation of diversity through time. For example, the rates of speciation slow down over time in a great number of phylogenies studied and diversification rates vary across lineages.

PART 2 - Birth-Death as a prior for calibrating phylogenies

Rannala & Yang, 2014 - Book chapter in The Princeton guide to Evolution Sec. II Chap. 3 Molecular Clock Dating

This textbook chapter gives a brief overview of the different steps involved in estimating node ages of a phylogeny. The text doesn't get in the details of any methods, but gives a general

idea of what exists. Even though it is in a textbook, I wouldn't say it is an easy text, and it requires some level of previous knowledge, for example knowing what is an MCMC.

Rannala & Yang, 1996 J. Mol. Evol.

Probability Distribution of Molecular Evolutionary Trees: A New Method of Phylogenetic Inference

This paper is the first (that I could find) to use a prior distribution of node ages described by the birth-death process. However, the description of the model doesn't scale the phylogeny to absolute time. Instead, they set the oldest node of the phylogeny at "1" and the tips to "0" while all other node ages are a relative estimation. [I don't understand what is going on fully, but I think:] The ages come from the likelihood function in which branch lengths are calculated according to a clock like substitution rate (calculated from a sequence alignment, and assumed equal for all lineages) that are multiplied by the birth-death prior.

Dos Reis et al 2016, Nature Review Genetics

Bayesian molecular clock dating of species divergences in the genomics era

This recent review starts with a historical perspective of how the molecular clock and fossils were used to build dated phylogenies. Then there is a general explanation of how Bayesian statistics can be used to estimate a phylogeny and its divergence times. What follows is a discussion of how sensitive the methods are to the choice of wrong priors, which is not overcome by increasing the amount of (sequence) data. There is a brief overview on the priors that can be used for *i*) relaxed rates of molecular clocks and *ii*) divergence times with fossil calibration. This paper brings an interesting figure of the interaction between the likelihood function and a prior for substitution rates (no prior for divergence times, though) to build the posterior estimate of a node age. There is also a table with a brief description of 6 different Bayesian programs to estimate phylogenies.

PART 3 - The Fossilized Birth-Death

Stadler 2010 - Journal of Theoretical Biology

Sampling-through-time in birth-death trees

This is the paper that first introduced the "fossilized" birth-death model (in the paper, she calls it sample-through-time instead of fossilized). In this process a tree diversifies with a given rate of branching and extinction at the same time that individuals are sampled randomly from lineages. Applications of this model can be the inclusion of fossils in phylogenies as well as taking into account pathogens sampled in the past. In this paper there (many) equations are derived that allow calculating the

Heath et al 2014 - PNAS

The fossilized birth-death process for coherent calibration of divergence-time estimates