Using Bayesian tip-dating method to estimate divergence times and evolutionary rates (postprint)

Authors: ZHANG Chi

Date: 2021-05-20T00:00:00+00:00

Abstract

Bayesian tip dating is a method developed in recent years for inferring clade divergence times and evolutionary rates. It overcomes the deficiencies of traditional stepwise calculations, but also involves substantially more statistical knowledge. This paper hierarchically dissects the principles and computational procedures of tip dating from a Bayesian statistical computation perspective, describing and discussing the main models and algorithms in dating calculations according to several components: prior distributions of divergence times, prior distributions of evolutionary rates, models of character state change, and Markov chain Monte Carlo algorithms. It aims to provide a reference for paleontologists analyzing empirical data to a certain extent.

Full Text

Using Bayesian Tip-Dating Method to Estimate Divergence Times and Evolutionary Rates

ZHANG Chi[1],[2] [1] Key Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing 100044, zhangchi@ivpp.ac.cn
[2] CAS Center for Excellence in Life and Paleoenvironment, Beijing 100044

Abstract

Bayesian tip dating is a recently developed method for estimating divergence times and evolutionary rates. It overcomes several drawbacks in traditional stepwise approaches. However, it also requires more statistical knowledge. This paper hierarchically explains the theory and computation in the Bayesian tip-dating approach, dividing the whole process into four key components: prior distributions for divergence times, prior distributions for evolutionary rates, models for character changes, and Markov chain Monte Carlo algorithms. The aim is to provide general guidance for paleontologists in empirical data analyses.

Keywords: Bayesian tip dating, fossilized birth-death process, relaxed clock, Mk model, MrBayes

Citation: Zhang C, in press. Using Bayesian tip-dating method to estimate divergence times and evolutionary rates. Vertebrata PalAsiatica. DOI: 10.19615/j.cnki.2096-9899.210516

Inferring phylogenetic relationships and divergence times of clades is fundamental to cladistic analysis. How to reasonably utilize fossil morphology and geological age data to complete such inferences has long been a thorny problem. Traditional inference processes often adopt a stepwise strategy. First, phylogenetic relationships are constructed from morphological data alone using parsimony. This relationship contains only topological information without temporal data, and represents only the optimal result under given search conditions (i.e., the most parsimonious tree or its consensus tree). Then, fixing this topology, divergence times of internal nodes are determined from fossil ages using either the minimum branch length method (Laurin, 2004) or the equal branch length method (Brusatte et al., 2008). The minimum branch length method pushes the time of the current node back by one million years as the ancestor node' s time, while the equal branch length method uses the midpoint between ancestor and descendant nodes as the current node' s time. Evolutionary rates can then be estimated through ancestral state reconstruction and the number of character changes on each branch, combined with the times inferred in the previous step. This stepwise computational approach is relatively intuitive and has been used in practical data analysis (Wang and Lloyd, 2016). However, this strategy has numerous drawbacks. First, each step ignores uncertainty in the inference, including tree topology, divergence times, and ancestral character states. Second, each step utilizes only partial data information—for example, morphology alone is used in tree construction while only fossil ages are used in dating. Third, the dating approach is highly subjective, sensitive to increases or decreases in fossil numbers, and not applicable to extant taxa. Finally, the entire process lacks a rigorous statistical framework for testing different model assumptions.

The Bayesian tip-dating method developed in recent years (Ronquist et al., 2012; Gavryushkina et al., 2014; Zhang et al., 2016) effectively overcomes these problems. Bayesian tip-dating integrates fossil morphology and age data in a single comprehensive computational process, maximizing data utilization while simultaneously accounting for uncertainty in tree topology, divergence times, evolutionary rates, and fossil ages. This method uses statistical models to describe processes such as character evolution, clade birth-death dynamics, and fossil sampling, and employs a relatively mature Bayesian statistical framework and computational methods for parameter estimation and model selection. However, this method is relatively complex, requiring substantial statistical knowledge that is often difficult for paleontologists to understand and implement, and there are few systematic expositions of the tip-dating computational process and parameter meanings (Gavryushkina and Zhang, 2020). This paper dissects the computational process of tip-dating layer by layer, explaining the important models used and the meaning of parameters, aiming to provide a reference for paleontologists analyzing empirical data.

This paper first introduces the fossilized birth-death process model (Stadler, 2010) for describing timetrees, then describes relaxed morphological clock models for character evolution rates, followed by the Mk model (Lewis, 2001) for character state changes. It then connects these models through Bayes' formula, and finally introduces the Markov chain Monte Carlo (MCMC) algorithm for estimating posterior distributions of parameters. The appendix provides MrBayes commands for analyzing Mesozoic bird data (Zhang and Wang, 2019).

Fossilized Birth-Death Process Model for Timetrees

A timetree represents the phylogenetic relationships and divergence times of clades, and its probability distribution can be specified through the fossilized birth-death process (Stadler, 2010). This process describes the occurrence of speciation, extinction, fossil sampling, and extant taxon sampling from the most recent common ancestor of these clades (the root), corresponding to a complete tree (Figure 1 [Figure 1: see original paper]A). However, in practical data analysis we cannot infer this complete tree, only the portion related to our sample—the sampled tree (Figure 1B). Let λ denote the speciation rate (or diversification rate) per branch,   the extinction rate,   the fossil sampling rate along each branch, and   the sampling probability (or proportion) of extant taxa. By establishing and solving a series of ordinary differential equations, we can obtain the probability distribution of the sampled timetree $T = \{\tau, t\}$ given λ, , , and , denoted as P(T | λ, , ,), where τ represents topology and t represents divergence times in millions of years.

In MrBayes software, this birth-death process uses the root age t1 as the initial condition (Figure 2 [Figure 2: see original paper]), requiring specification of a prior distribution for t1. Typically this prior is quite broad (ranging from 0 to infinity), though we can generally estimate a more precise range from the study group—for example, the lower bound cannot be older than the oldest fossil. Fossil ages can be fixed as specific values (millions of years ago) or given upper and lower bounds through a uniform distribution. The analysis also requires providing the approximate sampling proportion ( ) of extant taxa. There are two sampling strategies for extant taxa: uniform random sampling (random) and diversified sampling (diversity) (Zhang et al., 2016), which can be selected based on the actual data situation. The latter may better fit sampling patterns for higher-level taxa (e.g., taking one representative genus per family or one representative species per genus). For speciation, extinction, and fossil sampling rates, the program reparameterizes for convenient prior specification as d = λ - , v = / λ, and s = / (+). The default prior for d is an exponential distribution (range 0 to infinity), while the default priors for v and s are uniform distributions (range 0 to 1, more generally Beta distributions). Thus the prior distribution for the timetree including divergence times and other parameters is essentially determined.

It is worth noting that some datasets contain only fossils without extant taxa. In this case, we generally assume that all clades went extinct before reaching the present. Therefore, regardless of the sampling strategy and proportion for extant taxa (the program defaults to 1.0), no samples are collected. However, due to MrBayes software limitations, the youngest sample is always displayed as an extant taxon (time 0), requiring a corresponding shift of the entire tree' s timescale—in other words, divergence times must be added to the age of the most recent fossil. This is only a display issue, not a computational error (to be fixed in subsequent versions).

In addition to the fossilized birth-death process, MrBayes also provides a uniform prior for timetrees (Ronquist et al., 2012). This prior has no birth-death or sampling parameters and depends only on the root age t1, so only the prior distribution for t1 needs to be set. The uniform distribution is often considered an uninformative prior, but it actually often carries strong information that can influence divergence time estimates. The fossilized birth-death process, despite having many parameters, can actually be set quite flexibly. For example, speciation, extinction, and fossil sampling rates can all vary over time, independently in different time intervals (Gavryushkina et al., 2014; Zhang et al., 2016). This may better reflect actual biological processes while also allowing inference of how net speciation rates and fossil sampling rates change over time.

Relaxed Clock Models for Evolutionary Rates

The evolutionary rate of morphological characters refers to the expected number of changes per character per million years. For a given time interval, faster evolutionary rates lead to more expected character changes. Generally, each branch is assigned an evolutionary rate parameter, denoted as r (Figure 2). When clock models are applied to morphological data, they are called morphological clock models, analogous to molecular clock models for molecular data. The strict clock model assumes identical evolutionary rates across all branches, which is generally inappropriate for morphological data, so relaxed clock models are typically required. Relaxed clock models can be divided into two categories: independent rates and autocorrelated rates, differing in the probability distribution P(r) of r.

The independent rates model assumes that evolutionary rates on branches are independent of each other and all follow some probability distribution with the same mean. Commonly used distributions include the gamma distribution (Lepage et al., 2007) and the log-normal distribution (Drummond et al., 2006). The mean of the distribution is also called the base rate, reflecting the average evolutionary rate magnitude. The variance reflects the degree of rate variation among branches: smaller variance means rates are similar, implying no significant rate differences across the tree, while larger variance indicates more pronounced rate differences among branches.

The autocorrelated rates model assumes that rates on descendant branches depend on the rate of the immediately ancestral branch (e.g., both r2 and r5 depend on r1, while r3 and r4 depend on r2). The rate on the current branch is generally assumed to follow a log-normal distribution (Kishino et al., 2001; Thorne and Kishino, 2002) with mean equal to the rate at the adjacent ancestral node. Similarly, the variance reflects the degree of rate variation among branches.

These two rate models often affect divergence time estimates because the autocorrelated rates model tends to favor gradual rate changes, whereas the independent rates model has no such restriction and better accommodates situations with dramatic rate changes between adjacent branches. For fossil morphological data, the independent rates model may be more appropriate.

By default, all characters in the morphological data matrix share the evolutionary rate on each branch, so this rate represents the average across all characters. If heterogeneity in evolutionary rates among different characters needs to be considered, characters must be partitioned. Partitioning can generally be based on character types, body regions, or functional categories. Characters within each partition share a set of evolutionary rates, while rate patterns are independent among partitions, allowing inference of how characters related to different regions or functions change over time (Lee, 2016; Zhang and Wang, 2019). It should be noted that more partitions mean fewer characters per partition, resulting in less information to estimate evolutionary rate parameters, leading to large variances or even situations where the number of parameters exceeds the number of characters, making parameter estimation impossible. Therefore, a trade-off must be made between considering evolutionary rate heterogeneity and the number of partitions.

Mk Model for Character State Changes

Given divergence times and evolutionary rates, we can calculate the probability of morphological characters changing from one state to another within a given time interval (called transition probability). This probability is given by the Mk model (Lewis, 2001). The Mk model is the simplest model describing character state changes, assuming equal transition rates between states. Here we use a two-state character as an example: P00(r, t) represents the probability of state 0 remaining unchanged, P01(r, t) the probability of changing from 0 to 1, P10(r, t) from 1 to 0, and P11(r, t) for state 1 remaining unchanged. Then:

$$P_{00}(r,t) = P_{11}(r,t) = \frac{1}{2} + \frac{1}{2} \times \exp(-2rt)$$

$$P_{01}(r,t) = P_{10}(r,t) = \frac{1}{2} - \frac{1}{2} \times \exp(-2rt)$$

From these formulas, we can see that time t and rate r always appear as a product; therefore, without fossil age information, the two are unidentifiable. In other words, when building trees from morphological data alone, branch lengths represent the product of time and rate—that is, distance measured in expected number of changes per character. Only by simultaneously using both fossil morphology and age data can divergence times and evolutionary rates be estimated separately.

Using Figure 2 as an example, fossil F1 and F2 have character state 0, extant taxa S1 and S2 have state 1, and internal nodes have unknown states represented by x0, x1, x2. The ages of F1 and F2 are 100 Ma and 50 Ma, respectively. Then according to the Mk model, the probability of the character state column 0011 given timetree T = $\{\tau$, t$\}$ and rate r is:

$$P(0011|T,r) = \sum_{x_0}\sum_{x_1}\sum_{x_2} P_{x_{0x}1}(t_1{-}t_2, r_1) P_{x_{1x}2}(t_2{-}t_4, r_2) P_{x_{2x}1}(t_4, r_3) P_{x_{2x}0}(t_1{-}t_3, r_6)(t_4{-}t_5, r_4) P_{x_{1x}1}(t_2, r_5) P$$

where the sigma symbols represent summation over all possible states at internal nodes. Since morphological character matrices often include only variable characters, this probability must be divided by the probability of all variable states, i.e., $P(0011|T,r)/[1-P(0000|T,r)-P(1111|T,r)]$. The Mk model with this correction is called the Mkv model (Lewis, 2001).

Assuming characters in the morphological matrix are independent of each other, we can calculate the probability of each character column on the tree and multiply these probabilities together. This probability is called the likelihood function, expressed as $P(D|T,r)$, where D represents the morphological character matrix data.

In statistical inference, parameters are unknown random variables whose distributions need to be estimated from data—that is, calculating $P(T,r,\theta|D)$, called the posterior distribution, where $T=\{\tau,t\}$ is the timetree, r is the evolutionary rate, and θ represents other parameters (including λ, μ, ψ, etc.). According to hierarchical Bayes' formula:

$$P(T,r,\theta|D) = \frac{P(D|T,r)P(r)P(T|\theta)P(\theta)}{P(D)}$$

On the right side, the first term is the likelihood function from Section 4, the second term is the prior distribution for evolutionary rates from Section 3, and the third and fourth terms are prior distributions for the timetree and its parameters from Section 2. Thus all terms in the numerator can be calculated. The denominator $P(D)$ is the probability of the character data, which requires computing multiple integrals over all parameters and generally cannot be expressed analytically, only approximated through numerical algorithms. Therefore, Bayesian computation almost always uses Markov chain Monte Carlo algorithms.

Markov Chain Monte Carlo Algorithm

The Markov chain Monte Carlo (MCMC) algorithm constructs a Markov chain whose stationary distribution is the posterior distribution to be estimated. For simplicity, we illustrate with a one-dimensional parameter case (Figure 3 [Figure 3: see original paper]). In actual analysis, parameters are generally multidimensional (e.g., τ, t, r, θ), though the algorithmic principle is similar.

The MCMC employs the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970), which proceeds as follows:

a. Set arbitrary initial values for θ;

b. Propose a new value θ' based on the current value of θ, e.g., $\theta' \sim \text{uniform}(\theta - w/2, \theta + w/2)$;

c. If $\pi(\theta') > \pi(\theta)$, accept θ'; otherwise, accept θ' with probability $\alpha = \pi(\theta')/\pi(\theta)$;

d. If θ' is accepted, update $\theta = \theta'$; otherwise, keep θ unchanged;

e. Record the value of θ and return to step b.

Notably, when calculating the acceptance probability α, the ratio of posterior distributions is computed, causing the denominator terms to cancel out, leaving only the ratio of numerator terms. In other words, as long as the numerator can be expressed analytically, the MCMC algorithm can be used to estimate the posterior distribution of parameters.

After computation, we collect samples of θ. Since the initial parameter state is often poor, MCMC requires many generations to converge to regions of high posterior probability density. Therefore, when estimating the posterior distribution, initial samples (burn-in) are discarded, using only post-convergence samples. MrBayes discards the first 25% of samples by default. Additionally, the MCMC chain must iterate sufficiently to ensure enough effective samples for estimating the posterior distribution, generally requiring an effective sample size (ESS) greater than 100.

In practice, it is best to run MCMC independently at least twice to ensure consistent results. Sometimes insufficient chain length or different runs getting stuck in different posterior regions can lead to inconsistent estimates. Adjusting MCMC settings or improving the model may help the algorithm perform better. Using Metropolis-coupled MCMC is also an effective means of traversing multimodal distributions (Lakner et al., 2008). This algorithm runs multiple MCMC chains simultaneously—one cold chain and others as hot chains that can exchange states with the cold chain. MCMC samples are collected only from the cold chain, with hot chains serving only to help traverse multiple peaks. MrBayes runs two independent analyses by default, each using four chains (one cold and three hot).

Discussion

This paper has dissected the principles and computational process of tip-dating from a Bayesian statistical perspective. The Bayesian posterior distribution comprises prior distributions and the likelihood function, with two crucial components—the divergence time and evolutionary rate models—being particularly important in dating analyses as they are the main factors affecting dating accuracy.

The fossilized birth-death process, as a stochastic process describing clade speciation, extinction, and sampling, offers considerable flexibility. However, the model has areas needing improvement. Sampling strategies for extant taxa can be random or diversified, both of which may represent extremes; the true sampling pattern likely lies somewhere in between, or some clades may be randomly sampled while others are diversely sampled—though no current model supports this scenario. Until better models are developed, data may need to be adjusted to fit one of these sampling strategies. This approach generally works fine when extant taxa are few or absent. Additionally, speciation, extinction, and fossil sampling rates can vary over time in piecewise fashion, but within each time interval all branches share the same rates. For cases where different clades experience obviously different selective pressures or fossil preservation completeness varies significantly, branch-wise rather than time-wise piecewise models (Barido-Sottani et al., 2020) may be more appropriate, although such models have their own limitations. In any case, the fossilized birth-death process serves only as a prior on times; when data are substantial (including the distribution and number of fossils on the tree as well as the quantity and completeness of morphological characters), data dominate the inference and prior influence diminishes. However, actual situations are often complex and data may be less than ideal, making examination of different prior influences particularly important.

The prior on evolutionary rates—the morphological clock model—also interacts with time to affect final divergence time estimates. This influence is especially pronounced when fossils are few or unevenly distributed on the tree. This is primarily because fossil morphological data provide only distance information (expected number of changes per character), which is the product of time and rate (see Section 4). When fossils are scarce, accurate temporal information is unavailable, so for the same distance, the scenario could be either long time with slow rate or short time with fast rate—the specific case depends on the priors on time and evolutionary rate. For clades where time estimates are obviously too large or too small but lack fossil calibration, adding internal node calibration distributions can yield more reasonable time estimates (O' Reilly and Donoghue, 2016). In the complete absence of fossils with only extant taxa, node dating (another dating method) (Yang and Rannala, 2006) is performed using internal node calibration distributions.

Models describing morphological character state changes also have substantial room for improvement, involving more modeling and stochastic simulation work that is not the focus of this study. Here we briefly discuss potential impacts of the Mk model on dating. The Mk model assumes equal transition rates between all character states. These transitions can be unordered or ordered (for characters with three or more states). Unordered means a character can directly change from one state to any other (e.g., from 0 directly to 3), while ordered means changes can only occur between adjacent states (e.g., from 0 to 1, 1 to 2, then 2 to 3). Obviously, ordered characters require more changes (i.e., longer distance) to move between nonadjacent states, so assuming unordered changes for ordered characters will underestimate distances. More complex situations arise when transition rates between states are not equal or differ substantially, as in extreme cases like Dollo characters that are irreversible. Using the Mk model in such cases will also bias distance estimation. The calculation also assumes independence among characters; if some characters are strongly correlated, evolutionary distances will be overestimated. As previously mentioned, distance is the product of time and rate, so in ideal cases with abundant fossil (temporal) information, distance bias will primarily manifest in evolutionary rates with minimal impact on divergence times. However, analyzing real data is more complex and requires case-by-case analysis. Related work remains limited (Klopfstein et al., 2019) and requires more follow-up studies for detailed investigation.

Finally, we mention the MCMC algorithm used in Bayesian computation. This algorithm' s strategy differs markedly from parsimony and likelihood methods. Parsimony seeks trees with minimum tree length, while likelihood seeks parameter estimates that maximize likelihood (i.e., the maximum likelihood tree). Therefore, when designing tree search methods, the goal is simply to find optimal trees as quickly as possible. The MCMC algorithm aims to estimate the posterior distribution of parameters—not just a point but the parameter space. Thus MCMC algorithm efficiency involves both convergence and mixing. Convergence means MCMC reaches regions of high probability density, while mixing means MCMC can sample according to the probability distribution. Improving convergence speed is relatively easy, for example through parsimony-guided tree proposals to quickly find high-probability trees (Zhang et al., 2020). Improving mixing is more difficult and requires designing better proposal methods, which is both the focus and the challenge of Bayesian computation.

In summary, Bayesian tip-dating as an integrated analytical method can combine fossil morphology and age data with extant taxon morphology and molecular data to infer phylogenetic relationships, divergence times, and evolutionary rates, while accounting for uncertainty in tree topology, divergence times, evolutionary rates, and fossil ages. However, this method is still in its early stages of development, with many aspects of models and algorithms needing improvement, leaving much work to be done.

References

Barido-Sottani J, Vaughan T G, Stadler T, 2020. A Multi-Type Birth-Death model for Bayesian inference of lineage-specific birth and death rates. Syst Biol, 69: 973-986

Brusatte S L, Benton M J, Ruta M et al., 2008. Superiority, competition, and opportunism in the evolutionary radiation of dinosaurs. Science, 321: 1485-1488

Drummond A, Ho S, Phillips M et al., 2006. Relaxed phylogenetics and dating with confidence. Plos Biol, 4: e88

Gavryushkina A, Welch D, Stadler T et al., 2014. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. Plos Comput Biol, 10: e1003919

Gavryushkina A, Zhang C, 2020. Total-Evidence Dating and the Fossilized Birth-Death Model. In: Ho S Y W ed. The Molecular Evolutionary Clock. Switzerland: Springer Nature. 175-193

Hastings W K, 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika, 57: 97-

Kishino H, Thorne J L, Bruno W J, 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. Mol Biol Evol, 18: 352-361

Klopfstein S, Ryer R, Coiro M et al., 2019. Mismatch of the morphology model is mostly unproblematic in total-evidence dating: insights from an extensive simulation study. bioRxiv, 679084

Lakner C, van der Mark P, Huelsenbeck et al., 2008. Efficiency of Markov chain Monte Carlo tree proposals in Bayesian phylogenetics. Syst Biol, 57: 86-103

Laurin M, 2004. The evolution of body size, Cope’ s rule and the origin of amniotes. Syst Biol, 53: 594-622

Lee M S Y, 2016. Multiple morphological clocks and total-evidence tip-dating in mammals. Biol Lett, 12: 20160033

Lepage T, Bryant D, Philippe H et al., 2007. A general comparison of relaxed molecular clock models. Mol Biol Evol, 24:

Lewis P O, 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst Biol, 50:

Metropolis N, Rosenbluth A W, Rosenbluth M N et al., 1953. Equation of state calculations by fast computing machines. J Chem Phys, 21: 1087-1092

O’ Reilly J E, Donoghue P C J, 2016. Tips and nodes are complementary not competing approaches to the calibration of molecular clocks. Biol Lett, 12: 20150975

Ronquist F, Klopfstein S, Vilhelmsen L et al., 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. Syst Biol, 61: 973-999

Stadler T, 2010. Sampling-through-time in birth-death trees. J Theor Biol, 267: 396-404

Thorne J L, Kishino H, 2002. Divergence time and evolutionary rate estimation with multilocus data. Syst Biol, 51: 689-

Wang M, Lloyd G T, 2016. Rates of morphological evolution are heterogeneous in Early Cretaceous birds. Proc Royal Soc B Biol Sci, 283: 20160214
Yang Z, Rannala B, 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol Biol Evol, 23: 212-226
Zhang C, Huelsenbeck J P, Ronquist F, 2020. Using parsimony-guided tree proposals to accelerate convergence in Bayesian phylogenetic inference. Syst Biol, 69: 1016-1032
Zhang C, Stadler T, Klopfstein S et al., 2016. Total-evidence dating under the fossilized birth-death process. Syst Biol, 65:
Zhang C, Wang M, 2019. Bayesian tip dating reveals heterogeneous morphological clocks in Mesozoic birds. Roy Soc Open Sci, 6: 182062

Note: Figure translations are in progress. See original paper for figures.