Trends in Ecology & Evolution

CellPress REVIEWS

Review

Six Impossible Things before Breakfast: Assumptions, Models, and Belief in Molecular Dating

Lindell Bromham[1,*]

Confidence in molecular dating analyses has grown with the increasing sophistication of the methods. Some problematic cases where molecular dates disagreed with paleontological estimates appear to have been resolved with a growing agreement between molecules and fossils. But we cannot relax just yet. The growing analytical sophistication of many molecular dating methods relies on an increasingly large number of assumptions about evolutionary history and processes. Many of these assumptions are based on statistical tractability rather than being informed by improved understanding of molecular evolution, yet changing the assumptions can influence molecular dates. How can we tell if the answers we get are driven more by the assumptions we make than by the molecular data being analyzed?

Highlights

Estimating dates of divergence from DNA sequence data is a methodological and computational challenge as the amount of data explodes, and as models and methods become more statistically sophisticated.

The focus on methodological advances distracts from deeper challenges associated with historical inference: how can we construct plausible narratives for events long-past and long-term processes that we cannot directly observe?

As molecular dating methods grow more sophisticated, the number of assumptions we must make about the evolutionary process increases, so that it becomes difficult to keep track of what we must believe in order to believe the molecular dates.

Some disagreements between molecular dating studies should primarily be interpreted as a debate about assumptions and prior beliefs, such as the interpretation of fossil evidence.

Molecular Dating Analyses Rely on Assumptions

> Alice laughed: "There's no use trying," she said; "one *can't* believe impossible things."
>
> "I daresay you haven't had much practice," said the Queen. "When I was younger, I always did it for half an hour a day. Why, sometimes I've believed as many as six impossible things before breakfast."
>
> Lewis Carroll (1871) *Through the Looking-Glass: And What Alice Found There*, Macmillan

In the age of big data, there is a tendency to view molecular dating as primarily a computational challenge. How can we upgrade our bioinformatic tools to handle the ever-increasing amounts of sequence data and the growing sophistication of analytical methods? Less attention is paid to phylogenetic inference in terms of its place in the broader context of historical inference. Molecular dating – estimating dates of evolutionary events from comparative analysis of DNA or protein – shares with other 'historical sciences', including evolutionary biology, astronomy, archaeology, and geology, the challenge of reconstructing a plausible narrative of past events that we cannot directly witness, using only observations made in the present day [1]. The process of inferring dates from molecular data has special features that set it apart from other estimation procedures (Box 1). Most importantly, historical inference is dependent on assumptions about the processes that produced the data, and is not guaranteed to converge on the right answer given more data. If our assumptions are wrong, then our inference could be misleading, however much data we have.

Of course, we all know that molecular dating analyses rely on assumptions, but the nature and number of those assumptions is changing. In early molecular dating studies the assumptions were relatively few and simple (Box 2). It was possible to state what we had to believe in order

[1]Macroevolution and Macroecology, Division of Ecology and Evolution, Research School of Biology, Australian National University, Canberra, ACT 0200, Australia

*Correspondence: lindell.bromham@anu.edu.au (L. Bromham).

Box 1. Molecular Dates Are History

We commonly speak of 'molecular date estimates' and 'measuring branch length', but molecular dating is not like typical estimation or measurement. Estimating a parameter value is a classic case of induction, where we use a finite number of observations to draw general conclusions. Unlike deduction, where the conclusions flow with certainty from the observations and premises, there will always be a degree of uncertainty associated with inductive reasoning where we cannot guarantee to have sampled all instances or outcomes, but the more we sample, the more confident we grow.

Molecular dating is historical inference, where samples from a population are used to infer evolutionary history and processes. This history is not accessible to direct observation, and no amount of sampling from the population makes it so. This is abductive reasoning: using what we know, or believe, to construct a plausible explanation for the observations we have made. Historical inference requires us to make many assumptions about the processes whereby our data came to be. Because of that, our inference might not converge on the true value as we add more data. We could sample the entire genome of every individual in the population and still get the answer wrong, if our assumptions about the process are wrong, because the object of our investigation (evolutionary history) is not being directly sampled.

Consider multiple hits, which do not merely obscure past changes but erase them completely, such that there is no way to directly to recover the original sequence. We can never know for sure which substitutions occurred in the past that have since been overwritten. We can only make an informed guess at how many changes occurred based on the differences we observe and our understanding of molecular evolution. We can also never know for sure about past rates of change, and this leads to nonidentifiability of rates, dates, and branch lengths. We can constrain the possible solutions using other information, such as fossil data, but there will always be many possible ways to explain the same outcome. Branch lengths, rates, and dates are hypotheses, possible explanations of how the data we observe came to be.

to believe the dates: that the average rate of change was the same in all lineages, and that the number of differences between sequences was a reflection of the amount of change since their common ancestor [2–4]. The veracity and universality of these assumptions were challenged, leading to more complex models. The number and sophistication of the assumptions increases as the complexity of methods grows, and it becomes difficult, if not impossible, to interrogate them all [5]. In this article I look at six different assumptions ranging from

Box 2. History of Molecular Dates

The first generation of molecular clocks used genetic distance from protein sequence comparisons, immunological distances, and DNA hybridization studies as an indicator of time since divergence [2,65,66]. Constancy of rates was an observation, not a theoretical prediction [67]. It was because divergence times appeared to be related to genetic distance that a theoretical framework to understand the unexpected rate constancy was constructed [68]. The shockingly young dates for the human/chimp split (though later supported by new fossil finds) fueled skepticism of rate constancy, and led to suggestions that rates may have slowed down in hominid evolution [67].

The second generation of molecular clocks allowed for rate variation by implementing 'local clocks', grouping lineages into defined rate categories, for example estimating separate rates for higher taxa from fossil evidence [69], or grouping lineages into relevant categories such as host type [70]. In many cases, local clock dates were commensurate with earlier constant rate analyses, although at odds with paleontological evidence, for example supporting a Cretaceous radiation of mammals and a Precambrian origin of animals [64,69,71]. However, these models clearly did not capture all of the rate variation in the data [3], and dates varied depending on data partitions and calibrations [72].

The third generation of clocks allowed rates to vary over all branches in the phylogeny, whether randomly or correlated across related branches [73,74]. Allowing all branch rates to vary gives a vast number of possible solutions, so these methods require a suite of assumptions about the evolutionary process and history that produced the sequences in order to rank solutions by their plausibility. In general, the models used are flexible and arbitrary: they represent tractable statistical solutions, rather than aiming to describe a biologically realistic model of evolution [5]. Rates are drawn from a convenient distribution, but are not informed by knowledge of factors that can impact on rates of molecular evolution [33]. Because rates are fitted to information gleaned from fossils, relaxed clocks are often more in line with paleontological estimates of divergence but less in agreement with earlier uniform-rates or local-clock dates [75].

A new generation of molecular clocks is emerging that explicitly models the biological drivers of variation in rate of molecular evolution [76]. These methods still have far to go, but are an encouraging step forward that should allow us to use more of our prior knowledge of evolution in inferring molecular dates.

fundamental beliefs common to all molecular phylogenetic analyses, to specific assumptions of particular dating methods. This selection provides only a partial illustration of the complexity of the machinery underlying molecular dates, and of the difficulty in evaluating dates in light of the assumptions made.

1. The Alignment Identifies Homologous Sites

Alignment constructs a hypothesis about the history of the sequences, and is the bedrock on which all dating analyses are built [6]. Alignment begins with the assumption that all the sequences are descendants of a single ancestral sequence, and any differences between them are changes acquired since then. For molecular dating we also make the assumption that the history of the sequence tracks the history of the lineage, which may not be the case where horizontal gene transfer has combined genes with different evolutionary histories in the same lineage. Sequence alignment allows comparison of sites with a shared evolutionary history of conservation and change. Any inference from incorrectly aligned sites is spurious: it tells us nothing about history. For this reason, poorly aligned sites where we cannot be confident of homology should not be included in any analysis aiming to uncover history.

Decisions at the alignment stage can impact on inference of molecular evolutionary processes [7,8], phylogenetic inference [9–11], ancestral state reconstruction [12], tests of evolutionary hypotheses [13], and molecular dates [14–16]. Yet as datasets become larger, many researchers analyze their data with little or no direct inspection of alignment quality. Many automated alignments contain regions where homology is uncertain, or sequences that are mis-aligned for part or all of their length, but it is tacitly assumed that good signal from well-aligned sites will overcome distracting noise from mal-aligned sites. Given that we expect phylogenetic error to increase with alignment error [17], there is no guarantee that adding more data will converge on the right result [11], particularly if data quantity is prioritized over alignment quality. Alignment accuracy is influenced by topology and branch lengths [17,18], so there could be interactions between branch lengths, alignment accuracy, substitution rates, and therefore molecular dates.

A growing number of tools assist identification of regions of poor alignment [18] (although they are not guaranteed to improve inference [10]). Any such tools should target both ‘vertical’ sections of the alignment (comparable regions across all sequences) and ‘horizontal’ alignment problems (sequences that do not match the rest of the alignment for all or part of the sequence). Nonetheless, in addition to such techniques, evaluation of homology from sequence alignments requires inspection and judgment. Scientists often favor objective procedures over subjective judgment, but the objective measure – alignment score – is only an indirect reflection of the quality we seek – homology – and a high alignment score does not guarantee that all sites in the alignment represent homologous positions. Moves are afoot to explicitly account for alignment uncertainty within dating analyses [19]. The safest approach is to only analyze those alignments for which you are certain of homology for all columns and rows, resisting the temptation to analyze unverified alignments for the sake of expedience [20]. More data are worth nothing if the alignment does not reflect homology.

2. We Can Measure the Amount of Molecular Change That Has Occurred

One of the most mundane aspects of molecular evolution is also one of the most profound. Multiple hits erase historical signal (Box 1). Once a base in a sequence has changed more than once, there is no way to recover the original sequence. Overwritten changes are lost to us. The

best we can do is use the pattern of changes we can observe, combined with a model of sequence change, to guess how many changes have occurred that we can't see.

As with all models of evolution, substitution models are abstract representations of a complex process. The aim is not to describe reality in all its complexity, but to capture the salient features needed to read the historical signal. For example, many different processes are biased towards transition changes (G ↔ A, T ↔ C), including specific mutagens, repair mechanisms, and selection. We don't model the complex causes, but hope to capture the resulting patterns through simple parameters, such as the transition:transversion ratio. A balance must be found between using a model that is too simple to accurately model the process of interest, and one with so many free parameters that it risks overfitting [21,22]. The best substitution model, whether selected by preference or formal tests of model fit, may not describe the data well enough for reliable inference [23]. Model adequacy tests ask whether the chosen model could have produced these data: if not, then it is unlikely to be a fair description of the history and evolutionary processes that gave rise to these sequences [5]. Like all statistical tests, model adequacy tests vary in power and sensitivity, and a model that passes an adequacy test for a given dataset may nonetheless have poor accuracy or precision [24,25].

What do you do if all of your candidate models are rejected? In a typical statistical estimation procedure, we might report a rejection of model adequacy as evidence that the variables do not have the relationship described by the model. However, if all candidate models are rejected in a phylogenetic analysis, we can't say that the sequences do not share an evolutionary history, only that we lack the tools to tell what it is. One approach is to reject any data that do not fit available models, analyzing only those sequences that do not reject the model [26]. Even with sufficient data to be choosy, this approach runs the risk of accepting data that have insufficient power to reject the model (yet which may bias dates [27]), or listening only to those sequences that tell a particular story. Inference of past changes can be influenced not only by the model but also by the sites included in the analysis, taxon sampling, tree shape, relative branch lengths, and data partitioning strategy [28–32]. We should not lose sight of the fact that substitution models are a guide to generating plausible hypotheses about past events that we cannot directly witness, not a means for direct measurement of time, rate, or genetic distance (Box 1).

3. We Can Capture Rate Variation in the Data

The average rate of molecular evolution is shaped by a large number of biological factors, including size, generation time, longevity, and niche, and rates can therefore vary even between closely related species [33]. Rate variation creates a huge problem for inference of molecular dates. Each branch in the phylogeny can be described by a time interval, substitution rate, and number of substitutions (genetic distance). If we know two of these, we can derive the other one. If we know how many substitutions and the rate of change, then we can calculate the likely time interval represented by the branch. If we know the start and end dates of the branch, and the number of substitutions, we can calculate the rate. But since rates, times, and distance are confounded, then if we only have information on genetic distance, there are infinite combinations of rates and dates that could explain the data equally well. How do we decide between describing a branch as having a slow rate and an old date of origin, or a fast rate and a young date? 'Relaxed clock' methods use a convenient distribution of rates and a strategy for distributing these rates on the tree to optimize rate patterns during phylogeny inference [5]. As with substitution models, most branch-rates models do not aim to describe the tangled web of biological factors causing rates to vary between species, but aim to have sufficient free parameters that allow rate variation to be captured.

Box 3. When Is Best Practice Good Enough?

Science is, amongst other things, a social enterprise: members of the community set acceptable practice through feedback, peer review on papers and proposals, adoption of methods, and acceptance of results. This process is evident in molecular phylogenetics, a field characterized by long-running battles over philosophical differences and appropriate methodology [77]. In choosing a method, researchers must consider not only which method is appropriate to their data and question but also which is least likely to invoke reviewer disapproval.

Curiously, the widely-accepted principle of Ockham's razor – choose the simplest explanation with the fewest *ad hoc* assumptions until it is shown to be inadequate – is not always applied in molecular phylogenetics. Although tests are used to select effective substitution models with the fewest parameters, it is commonly assumed that complex dating methods are more reliable than simpler ones (e.g., always using a 'relaxed clock' without testing if data could be described with a single average rate). But identifying problems with earlier methods does not make newer methods automatically more reliable. For real data, we often can't verify whether more complex methods are getting closer to the right answer. In fact, you could argue that complex methods increase uncertainty in molecular dates in cases where a wide range of dates can be obtained by varying the parameters or assumptions within reasonable bounds. Published results may have relatively tight confidence intervals, but it may be difficult for the unbiased reader to favor one set of model assumptions over another.

It may seem obvious that a molecular dating study should employ 'best practice' in terms of the latest methodology. But given that we can rarely know whether our current best method is good enough, an alternative approach is to treat all molecular date estimates with healthy skepticism, and subject them to hearty interrogation. Are the dates robust to decisions made in the analysis? If not, then which assumptions are open to debate? What range of dates would be obtained if all reasonable assumptions were considered? What if you had chosen different data? Do the prior distributions around fossils really account for the uncertainty in timing, or do they make unrealistic assumptions that fail to account for geographic, temporal, or taphonomic biases? Could an alternative hypothesis also fit the data without requiring unreasonable parameter values? Are the priors determining the posterior, and if so what, if anything, have the molecular data told us that we did not already believe?

Because we expect rates to evolve over the tree as species evolve, phylogenetic methods that allow each branch of the phylogeny to have a different rate have been enthusiastically adopted. But a more complex model is not guaranteed to give a more accurate outcome (Box 3). How can we tell if our branch-rate models correctly describe the patterns of rate variation? Ideally, we would test them on cases where the patterns of rate variation are already known. For example, one of the most consistently identified patterns of rate variation is that parasitic plants have faster rates of molecular evolution than their non-parasitic relatives [34]. Encouragingly, phylogenetic analysis of one parasitic plant clade consistently inferred faster rates in the parasitic clade [35]. However, the distribution of those rates depended on the branch-rate model used: while some models pushed faster rates to the tips of the clade, some pushed them to the stem, resulting in a doubling of date estimates (Figure 1). We know that some of these dates must be wrong, but, without independent information on dates or rates, we don't know which, if any, of the solutions is correct.

4. Sequences Have Been Randomly Sampled

A default assumption of some Bayesian dating methods is that the sequences included in the alignment are an unbiased sample of all living members of that group, such that every possible 'tip' in the tree had an equal chance of inclusion in the analysis [5]. This is almost never true for real data, but falsely assuming random sampling can result in significant error in molecular dates [36].

Bayesian phylogenetic methods must begin by assigning a prior probability to any solution, including topology, branch lengths, and parameter values (Box 4), using either an arbitrary distribution (such as a flat prior) or a model of evolution that describes the type of phylogenies considered most plausible. This 'tree model' is influenced by sampling fraction: if not all of the tips of the phylogeny have been sampled, then the distribution of node heights will be affected

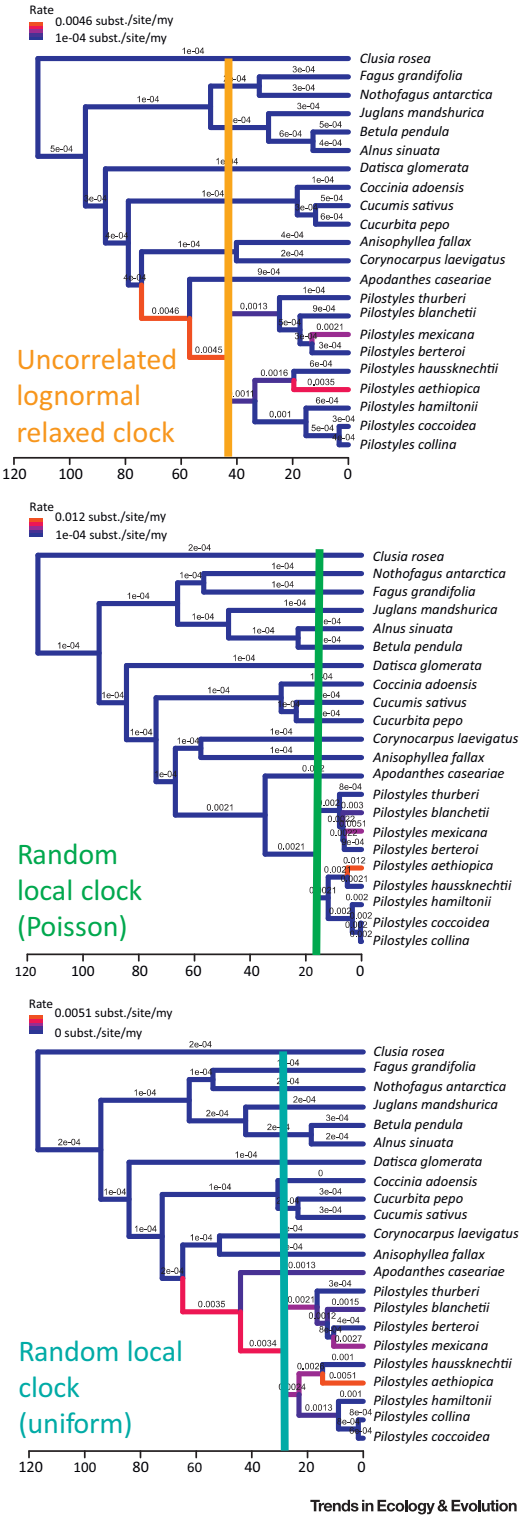

Figure 1. Rate Models Vary in Their Reconstruction of the Distribution of Faster Rates in a Parasitic Plant Clade [35] (Rightslink: 4482751040061). Horizontal scale, millions of years before present, Myr.

Box 4. Bayesian or Frequentist?

How do you view the relationship between evidence and hypothesis testing? A frequentist evaluates how likely it is that their data would have been produced if a given phylogeny was true, and a Bayesian evaluates how probable it is that a phylogeny is true, given the data they have [78]. In common practice, however, the philosophical separation between different classes of phylogenetic methods is somewhat fuzzy.

Both classes of methods rely on assumptions about the nature of the evolutionary process in order to make probability statements about the relationship between observed data and phylogenetic hypotheses. Bayesian methods are distinguished by requiring every phylogeny to have a prior probability, reflecting the degree of belief that a tree is true, without considering information from the data. The number of possible phylogenies is vast, so prior probabilities are assigned using convenient functions, such as a birth-death model. Prior probabilities can also express knowledge we already have about evolutionary history, such as using fossils to calibrate the node ages.

But not all forms of prior knowledge are used in Bayesian phylogenetics. Most researchers have prior beliefs about relationships between taxa, and could make an educated guess at how their taxa will cluster on the tree, even if they are unsure about the exact relationships between species. In practice, prior knowledge on relationships is used only informally in Bayesian phylogenetics, as a post-analysis filter: we reject the results of an analysis if they violate strongly held beliefs about relationships (e.g., if koalas cluster with bears and not with wombats). In fact, there is a social prohibition on applying knowledge of relationships as priors in a Bayesian phylogenetic analysis. Curiously, beliefs about relationships are more often applied in likelihood analyses (as 'topological constraints'), presumably due to the practical need to reduce computational burden.

Bayesian analyses do not always make use of available prior information, and they frequently employ decision-making rules that resemble frequentist approaches, such as threshold Bayes factor values for accepting one hypothesis or model over another. A 95% credible set may not technically be the same as a 95% confidence interval [79], and a significance test in a Bayesian framework may reflect a different property to a Fisherian test [80], but in practice the outputs of both approaches are interpreted in the same way. Most researchers in phylogenetics do not appear to be motivated by a strong philosophical commitment to either a Bayesian or frequentist position [78], but develop their own practical hybrid strategy (to which we might give the hybrid name 'Fresian'), adopting whichever methods or software have attractive features.

(Figure 2). Although the proportion of living species included can sometimes be estimated from taxonomies, most real datasets have uneven sampling: common species are more likely to be included than rare ones, and species-poor clades often have higher proportional representation than species-rich clades [5].

Nonrandom sampling shifts the distribution of sampled nodes. Selecting one or few sequences from each major clade preferentially samples the deeper nodes of the tree (Figure 2). In other words, prior belief about the most likely shape of the phylogeny based on random sampling will not match the underlying history of representatively sampled sequences. This would not matter if the data could overwhelm the priors, but we have reason to suspect it does not always do so. In an analysis of placental mammals, falsely assuming random sampling dated the whale/dolphin split at 16 Myr, whereas a representative sampling model gave a date (32 Myr) that was considered to be commensurate with fossil evidence [36]. Mismatch between the sampling parameters and the actual sampling strategy can also have an impact on diversification rates by changing the distribution of nodes along the tree [37].

5. Rates of Speciation and Extinction Have Been Constant

Some implementations of Bayesian dating methods assume uniform values for speciation and extinction rate over the whole phylogeny, as part of the tree model. In the real world, speciation rates can vary over time (e.g., early burst of diversification then slowdown [38]), over space (e.g., faster in the tropics [39]), and between lineages (e.g., differences in species richness between related clades [40]). It does not matter *per se* if the tree model is unrealistic: the best

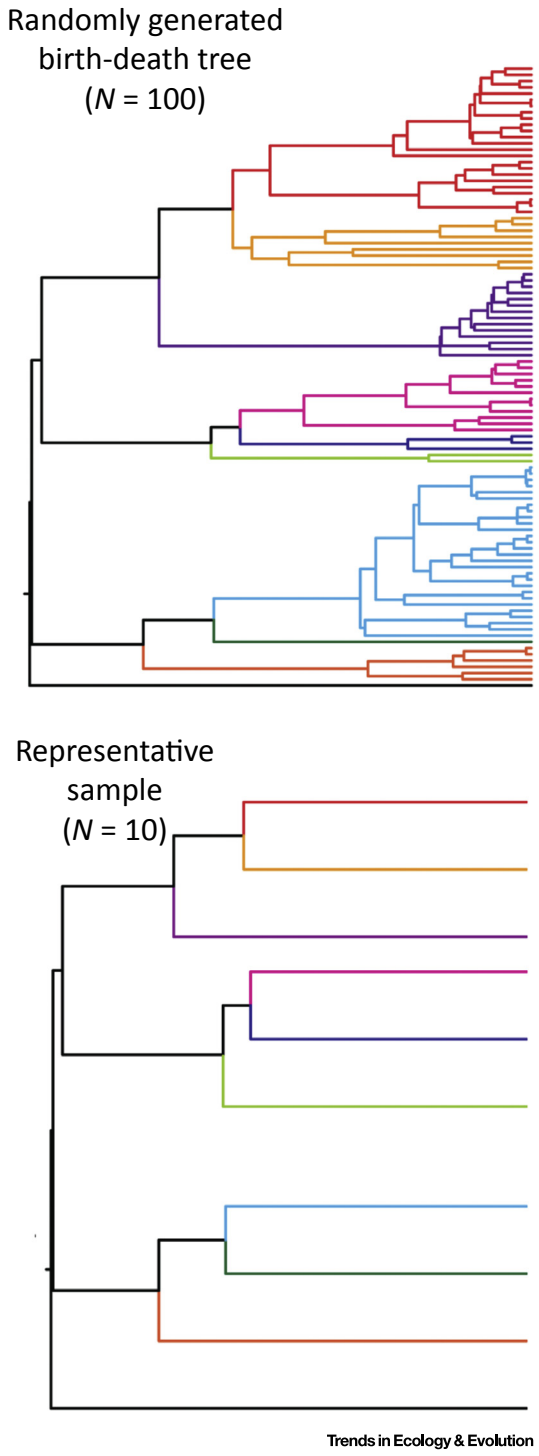

Figure 2. Representative Sampling (Including Representatives of Each Major Clade) Is a Common Strategy in Molecular Dating Studies. However, it violates the random sampling assumptions of some of the most popular Bayesian dating methods. The effect of representative sampling is to make the distribution of nodes different from that expected under the birth-death tree prior, which can have substantial effects on molecular dates [68] (Rightslink 4470580706622).

posterior solutions may be quite different from that expected under constant diversification rates [41]. But it does raise an interesting conundrum when the resulting phylogeny is used to infer differential rates of diversification between lineages, regions, or time-periods. We may not immediately recognize a phylogenetic analysis of macroevolutionary patterns as a molecular dating study, but such analyses rely critically on node heights placing speciation events in evolutionary time. Many such analyses are based on Bayesian phylogenies with the prior assumption of uniform speciation rates, and few such studies provide the reader with evidence that the phylogenetic solutions used to infer differential speciation rates reject the prior assumption of constant speciation rates.

This raises a broader issue about the disconnect between phylogenetic methods and applications. Tests of the accuracy and precision of macroevolutionary methods typically use simulated phylogenies that do not model the error in reconstructing dates and branch lengths. But realistic levels of variation in both substitution rates and speciation rates can have sufficient impact on node heights to mislead macroevolutionary inference [42]. Macroevolutionary methods can have reduced power when molecular rates vary, leading to misidentification of diversification rate trends (e.g., identifying 'slowdown' when speciation rates have increased through time) [42]. Testing a macroevolutionary model against a set of solutions from the posterior distribution allows for stochastic variability in branch length under a given model and particular set of assumptions, but what we really want to know is how confident we can be of the distribution of node heights given the uncertainty in all our assumptions, and under all reasonable regions of parameter space [41].

6. The Data Speak Louder Than the Priors

In a Bayesian molecular dating analysis, the priors provide a starting point from which to reach the posterior distribution of molecular dates. It doesn't matter if our prior beliefs are unrealistic as long as the data can overwhelm prior assumptions. There are two simple ways to check that this is the case: compare the distribution of the prior and the posterior, and ask if changing the priors changes the posterior.

An increasing number of studies provide readers with the information they need to evaluate the degree to which priors determine dates [43,44]. Sampling from the joint prior (i.e., running the analysis without the molecular data) illustrates the node ages deemed most plausible given all the assumptions about history and process [5,45]. If the posterior distribution of dates matches the prior distribution, then the molecular data have not influenced our belief in the timing of diversification beyond the assumptions we made before we started. Perhaps our assumptions exactly captured the true history of the clade, or maybe our data did not speak louder than our priors. Either way, the amount of new information we have gained from the molecular data is limited, unless the posterior has a substantially narrower range of dates than the prior. Failure to reject the priors should prompt introspection and further analysis.

We can make an assessment of the degree to which the priors determine the posterior by comparing dates inferred from the same data but with modified priors. Changing the substitution model, tree model, or 'clock model' can have a large impact on dates [35,46–48]. For example, using the same sequences and calibrations but different clock models (autocorrelated vs uncorrelated) shifted dates for the origin of animals by hundreds of millions of years [49], and dates for the origin of cyanobacteria by nearly a billion years (with nonoverlapping confidence intervals, CIs) [50].

The calibrations often have the greatest influence. Adding more but less-certain calibrations increased molecular dates for the origin of *Nothofagus* by 30 Myr (under both likelihood and Bayesian methods) [51]. Even for undisputed calibrations, the position of the calibration along an edge, or the way that the calibration densities are specified, can have a huge influence on dates [31,46]. The same set of sequences and fossil dates, but with different calibration densities, produced dates for placental mammals varying by 50 Myr or more, from mid-Cretaceous diversification to a post-dinosaur radiation [52]. Priors on the root age can have dramatic effects, with ramifications throughout the phylogeny [45], causing the marginal priors on node ages to be different from the specified calibrations [5]. For example, adding a constraint on the root age for a phylogeny of harvestmen shifted all dated nodes forward by tens of millions of years [53]. In these cases, disagreement between molecular dates is primarily a debate about how we interpret fossil evidence, rather than being about molecular evidence *per se* [54].

Because of the strong influence of calibrations on molecular dates, use of a common set of calibrations leads to a growing sense of agreement across studies by constraining the molecular dates to an interval deemed acceptable based on paleontological evidence. This is fine as long as our beliefs about history given the fossil evidence are indeed correct (although, as for all historical inference, there is an unavoidable measure of doubt in any fossil-based dates of divergence). We would be more confident of agreement across Bayesian studies if dates were robust to different priors, which would suggest that the molecular data are telling a clear story [31]. When differences between published molecular dates are primarily due to differences in calibration, we should interpret disagreement between studies as one of interpretation of fossil evidence. In this sense, molecular dates are a description of what we have to believe about the molecular data to make the data fit a particular fossil-based scenario. Accepting particular fossil dates as accurate may require us to infer very fast molecular rates for part of the history, where wider uncertainty on fossil dates might allow slower rates and deeper dates of divergence [52,55]. For example, fixing the root of the arthropod phylogeny to match fossil evidence requires inference of very fast rates of molecular change (as well as morphological change and diversification) in the early part of the history [58]. Forcing molecular data for placental mammal orders to match the first fossils requires faster rates of molecular evolution in the early history of placentals than at any subsequent time [56]. On the other hand, failing to account for the possibility that rates were faster early in the diversification when mammals were smaller pushes dates back into the Cretaceous [57]. Which dates we consider more reliable depends on which assumptions about history and process we accept.

Concluding Remarks and Future Prospects

> We again emphasize that the general question of how closely these [phylogenies] approximate what actually happened cannot be answered directly, that is, by checking it against what actually happened. We do not know what actually happened; if we did, there would be little, if any, point in doing the molecular studies.
>
> Sarich, V.M. and Cronin, J.E. (1976) Molecular systematics of the primates. In *Molecular Anthropology, Genes and Proteins in the Evolutionary Ascent of the Primates* (Goodman, M. *et al*. eds), pp. 141–170, Springer.

There is temporal information in molecular data. But we know that, in most cases, genetic distance alone does not provide an accurate timeline of history. Instead, we infer history and

Outstanding Questions

How can we more effectively incorporate prior information on variation in substitution rate between lineages, aspects of evolutionary history, and macroevolutionary processes into molecular dating?

How do we balance model generality and tractability against biological realism?

What effective means do we have for investigating the influence of our assumptions on molecular dates, and how can we report results over the range of reasonable assumptions?

How can we use molecular dates more effectively in a hypothesis-testing framework that takes into account the uncertainty arising from the challenges of selecting appropriate models, parameters, and assumptions?

processes by generating plausible explanations of how our sequences came to be, given what we know about molecular evolution. If we view molecular dates as something we measure directly from our data, we will expect that more DNA sequences and greater computational power will give us increasing accuracy and precision. But this would be to ignore the more challenging aspects of molecular dating, some of which are specific to the nature of the problem (such as the nonidentifiability of dates, rates, and distances), and some of which are general features of historical inference (such as reliance on models of process to evaluate alternative hypotheses).

All molecular dating analyses require us to make statements of belief, irrespective of whether we call them models, assumptions, or priors. We need to recognize that while some disagreements between molecular dates are due to differences in data or method, many are due to different beliefs codified in the analyses, such as the interpretation of fossil evidence. Disputes about molecular dates are sometimes more about the veracity of these beliefs than about the molecular evidence *per se*. While increasing the amount of data may improve some parameter estimates, it will not, in and of itself, solve the problem of molecular dating. If different calibration strategies, clock models, or tree priors result in different dates, then unless we can be certain which are true, we should report our uncertainty as encompassing all iterations of the analysis, under all reasonable assumptions (see Outstanding Questions).

Are molecular dates becoming more reliable? The awkward truth is that, unless we already know the truth, we just don't know. Dates are now not limited by amount of data [59], but by our understanding of evolutionary processes, and by the degree to which the stories told from DNA can be integrated with other forms of historical information. Narrowing confidence intervals are sometimes falsely interpreted to indicate growing reliability, yet may fail to contain the true date if the assumptions are violated, or if the prior beliefs misguided. Confidence intervals are a statement about precision not accuracy. Wide confidence intervals are often interpreted as a sign of poor quality of inference, but may still be accurate if the CIs contain the true date. For example, as molecular dating of the diversification of animal phyla has grown from single sequences to hundreds of loci and methods have grown in sophistication, CIs have narrowed, yet the dates for the divergence of the major lineages of animals still differ between studies by hundreds of millions of years [50,60–62]. The observed pattern of genetic distance is described using different combinations of rates and dates, such as inferring few accelerations in rate (and old dates) or many decelerations (and younger dates) [45,63]. However wide or narrow, CIs can be used for hypothesis testing by asking whether they contain a key hypothesized date [64].

Naturally, researchers may favor some results as being more consistent with their interpretation of evolutionary history or processes, but it is not possible to say which molecular dates are more accurate than others unless we are already party to the truth (in which case we would not need molecular dates). We weigh the plausibility of different narratives, whether informally or statistically, but none of the dates are assumption-free or unconditioned by prior belief. Molecular dating is one of the most fabulously useful tools we have in evolutionary biology, with an ever-extending reach from recent viral outbreaks to deep macroevolutionary dynamics. But we may need to accept that sometimes both paleontological evidence and molecular dates paint history with a broad brush not fine penwork.

Acknowledgments

I am grateful to Simon Ho, Matt Hahn, David Bryant, Rob Lanfear, Xia Hua, and Kim Sterelny for their wisdom and perspicacity.

References

1. Bromham, L. (2016) Testing hypotheses in macroevolution. *Stud. Hist. Philos. Sci. A*, 55, 47–59
2. Doolittle, R.F. and Blomback, B. (1964) Amino-acid sequence investigations of fibrinopeptides from various mammals: evolutionary implications. *Nature*, 202, 147–152
3. Zuckerkandl, E. and Pauling, L. (1965) Evolutionary divergence and convergence in proteins. *Evol. Genes Proteins*, 97, 97–166
4. Dickerson, R.E. (1971) The structure of cytochrome c and rates of molecular evolution. *J. Mol. Evol.* 1, 26–45
5. Bromham, L. *et al.* (2018) Bayesian molecular dating: opening up the black box. *Biol. Rev.* 93, 1165–1191
6. Boussau, B. and Daubin, V. (2010) Genomes as documents of evolutionary history. *Trends Ecol. Evol.* 25, 224–232
7. Schneider, A. *et al.* (2009) Estimates of positive Darwinian selection are inflated by errors in sequencing, annotation, and alignment. *Genome Biol. Evol.* 1, 114–118
8. Fletcher, W. and Yang, Z. (2010) The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection. *Mol. Biol. Evol.* 27, 2257–2267
9. Talavera, G. and Castresana, J. (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56, 564–577
10. Tan, G. *et al.* (2015) Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. *Syst. Biol.* 64, 778–791
11. Philippe, H. *et al.* (2011) Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol.* 9, e1000602
12. Vialle, R.A. *et al.* (2018) Alignment modulates ancestral sequence reconstruction accuracy. *Mol. Biol. Evol.* 35, 1783–1797
13. Levy Karin, E. *et al.* (2014) Alignment errors strongly impact likelihood-based tests for comparing topologies. *Mol. Biol. Evol.* 31, 3057–3067
14. Blackburne, B.P. and Whelan, S. (2013) Class of multiple sequence alignment algorithm affects genomic analysis. *Mol. Biol. Evol.* 30, 642–653
15. Md Mukarram Hossain, A.S. *et al.* (2015) Evidence of statistical inconsistency of phylogenetic methods in the presence of multiple sequence alignment uncertainty. *Genome Biol. Evol.* 7, 2102–2116
16. Gatesy, J. and Springer, M.S. (2017) Phylogenomic red flags: homology errors and zombie lineages in the evolutionary diversification of placental mammals. *Proc. Natl. Acad. Sci. U. S. A.* 201715318
17. Ogden, T.H. and Rosenberg, M.S. (2006) Multiple sequence alignment accuracy and phylogenetic inference. *Syst. Biol.* 55, 314–328
18. Penn, O. *et al.* (2010) GUIDANCE: a web server for assessing alignment confidence scores. *Nucl. Acids Res.* 38, W23–W28
19. Ashkenazy, H. *et al.* (2018) Multiple sequence alignment averaging improves phylogeny reconstruction. *Syst. Biol.* 68, 117–130
20. Springer, M.S. and Gatesy, J. (2018) On the importance of homology in the age of phylogenomics. *Syst. Biodivers.* 16, 210–228
21. Rannala, B. (2002) Identifiability of parameters in MCMC Bayesian inference of phylogeny. *Syst. Biol.* 51, 754–760
22. Lemmon, A.R. and Moriarty, E.C. (2004) The importance of proper model assumption in bayesian phylogenetics. *Syst. Biol.* 53, 265–277
23. Gatesy, J. (2007) A tenth crucial question regarding model use in phylogenetics. *Trends Ecol. Evol.* 22, 509–510
24. Baele, G. *et al.* (2012) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol. Biol. Evol.* 29, 2157–2167
25. Duchêne, D.A. *et al.* (2018) Differences in performance among test statistics for assessing phylogenomic model adequacy. *Genome Biol. Evol.* 10, 1375–1388
26. Ho, S.Y. (2014) The changing face of the molecular evolutionary clock. *Trends Ecol. Evol.* 29, 496–503
27. Bromham, L. *et al.* (2000) The power of relative rates tests depends on the data. *J. Mol. Evol.* 50, 296–301
28. Heath, T.A. *et al.* (2008) Taxon sampling and the accuracy of phylogenetic analyses. *J. Syst. Evol.* 46, 239–257
29. Kück, P. *et al.* (2012) Long branch effects distort maximum likelihood phylogenies in simulations despite selection of the correct model. *PLoS One*, 7, e36593
30. Foster, C.S. *et al.* (2016) Evaluating the impact of genomic data and priors on Bayesian estimates of the angiosperm evolutionary timescale. *Syst. Biol.* 66, 338–351
31. Matschiner, M. (2019) Gondwanan vicariance or trans-Atlantic dispersal of cichlid fishes: a review of the molecular evidence. *Hydrobiologia*, 832, 9–37
32. Kainer, D. and Lanfear, R. (2015) The effects of partitioning on phylogenetic inference. *Mol. Biol. Evol.* 32, 1611–1627
33. Bromham, L. (2009) Why do species vary in their rate of molecular evolution? *Biol. Lett.* 5, 401–404
34. Bromham, L. *et al.* (2013) Parasitic plants have increased rates of molecular evolution across all three genomes. *BMC Evol. Biol.* 13, 126
35. Bellot, S. and Renner, S.S. (2014) Exploring new dating approaches for parasites: the worldwide Apodanthaceae (Cucurbitales) as an example. *Mol. Phylogenet. Evol.* 80, 1–10
36. Ronquist, F. *et al.* (2016) Closing the gap between rocks and clocks using total-evidence dating. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 371, 20150136
37. Höhna, S. *et al.* (2011) Inferring speciation and extinction rates under different sampling schemes. *Mol. Biol. Evol.* 28, 2577–2589
38. Moen, D. and Morlon, H. (2016) Why does diversification slow down? *Trends Ecol. Evol.* 29, 190–197
39. Cardillo, M. (1999) Latitude and rates of diversification in birds and butterflies. *Proc. R. Soc. Lond. B*, 266, 1221–1225
40. Bromham, L. *et al.* (2015) Exploring the relationships between mutation rates, life history, genome size, environment and species richness in flowering plants. *Am. Nat.* 185, 507
41. Sarver, B.A.J. *et al.* (2018) The choice of tree prior and molecular clock does not substantially affect phylogenetic inferences of diversification rates. *bioRxiv* Published online June 29, 2018. http://dx.doi.org/10.1101/358788 Published online June 29, 2018
42. Duchêne, D.A. *et al.* (2017) Phylogenetic estimates of diversification rate are affected by molecular rate variation. *J. Evol. Biol.* 30, 1884–1897
43. dos Reis, M. *et al.* (2012) Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc. R. Soc. B Biol. Sci.* 279, 3491–3500
44. Tarver, J.E. *et al.* (2016) The interrelationships of placental mammals and the limits of phylogenetic inference. *Genome Biol. Evol.* 8, 330–344
45. Lozano-Fernandez, J. *et al.* (2017) RelTime rates collapse to a strict clock when estimating the timeline of animal diversification. *Genome Biol. Evol.* 9, 1320–1328
46. Crisp, M.D. *et al.* (2014) Clock model makes a large difference to age estimates of long-stemmed clades with no internal calibration: a test using Australian grasstrees. *BMC Evol. Biol.* 14, 263
47. Marshall, D.C. *et al.* (2016) Inflation of molecular clock rates and dates: molecular phylogenetics, biogeography, and diversification of a global cicada radiation from Australasia (Hemiptera: Cicadidae: Cicadettini). *Syst. Biol.* 65, 16–34
48. Ritchie, A.M. *et al.* (2016) The impact of the tree prior on molecular dating of data sets containing a mixture of inter-and intraspecies sampling. *Syst. Biol.* 66, 413–425
49. Dohrmann, M. and Wörheide, G. (2017) Dating early animal evolution using phylogenomic data. *Sci. Rep.* 7, 3599

50. Betts, H.C. *et al.* (2018) Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nat. Ecol. Evol.* 2, 1556
51. Sauquet, H. *et al.* (2012) Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of *Nothofagus* (Fagales). *Syst. Biol.* 61, 289–313
52. dos Reis, M. *et al.* (2014) Neither phylogenomic nor palaeontological data support a Palaeogene origin of placental mammals. *Biol. Lett.* 10, 20131003
53. Oberski, J.T. *et al.* (2018) A dated molecular phylogeny of mite harvestmen (Arachnida: Opiliones: Cyphophthalmi) elucidates ancient diversification dynamics in the Australian Wet Tropics. *Mol. Phylogenet. Evol.* 127, 813–822
54. Morris, J.L. *et al.* (2018) Reply to Hedges *et al.*: Accurate timetrees do indeed require accurate calibrations. *Proc. Natl. Acad. Sci. U. S. A.* 115, E9512–E9513
55. O'Leary, M.A. *et al.* (2013) The placental mammal ancestor and the post-K–Pg radiation of placentals. *Science*, 339, 662–667
56. Springer, M.S. *et al.* (2013) Technical comment on 'The placental mammal ancestor and the post-K–Pg radiation of placentals'. *Science*, 341, 613
57. Phillips, M.J. and Fruciano, C. (2018) The soft explosive model of placental mammal evolution. *BMC Evol. Biol.* 18, 104
58. Lee, M.S.Y. *et al.* (2013) Rates of phenotypic and genomic evolution during the Cambrian explosion. *Curr. Biol.* 23, 1889
59. dos Reis, M. *et al.* (2018) Using phylogenomic data to explore the effects of relaxed clocks and calibration strategies on divergence time estimation: primates as a test case. *Syst. Biol.* 67, 594–615
60. dos Reis, M. *et al.* (2015) Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. *Curr. Biol.* 25, 2939–2950
61. Erwin, D.H. *et al.* (2011) The Cambrian conundrum: early divergence and later ecological success in the early history of animals. *Science*, 334, 1091–1097
62. Lee, M.S. *et al.* (2013) Rates of phenotypic and genomic evolution during the Cambrian explosion. *Curr. Biol.* 23, 1889–1895
63. Battistuzzi, F.U. *et al.* (2018) RelTime relaxes the strict molecular clock throughout the phylogeny. *Genome Biol. Evol.* 10, 1631–1636
64. Bromham, L. *et al.* (1998) Testing the Cambrian explosion hypothesis by using a molecular dating technique. *Proc. Natl. Acad. Sci. U. S. A.* 95, 12386–12389
65. Sarich, V.M. and Wilson, A.C. (1967) Immunological time scale for hominid evolution. *Science*, 158, 1200
66. Margoliash, E. (1963) Primary structure and the evolution of cytochrome c. *Proc. Natl. Acad. Sci. U. S. A.* 50, 672–679
67. Sarich, V.M. and Cronin, J.E. *et al.* (1976) Molecular systematics of the primates. In *Molecular Anthropology, Genes and Proteins in the Evolutionary Ascent of the Primates* (Goodman, M., ed.), pp. 141–170, Springer
68. Kimura, M. (1968) Evolutionary rate at the molecular level. *Nature*, 217, 624–626
69. Rambaut, A. and Bromham, L. (1998) Estimating divergence dates from molecular sequences. *Mol. Biol. Evol.* 15, 442–448
70. Worobey, M. *et al.* (2014) A synchronized global sweep of the internal genes of modern avian influenza virus. *Nature*, 508, 254
71. Murphy, W.J. *et al.* (2001) Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science*, 294, 2348–2351
72. Douzery, E.J. *et al.* (2003) Local molecular clocks in three nuclear genes: divergence times for rodents and other mammals and incompatibility among fossil calibrations. *J. Mol. Evol.* 57, S201–S213
73. Drummond, A.J. *et al.* (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4, e88
74. Thorne, J.L. *et al.* (1998) Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* 15, 1647–1657
75. Goswami, A. (2012) A dating success story: genomes and fossils converge on placental mammal origins. *EvoDevo*, 3, 18
76. Lartillot, N. and Delsuc, F. (2012) Joint reconstruction of divergence times and life-history evolution in placental mammals using a phylogenetic covariance model. *Evolution*, 66, 1773–1787
77. Felsenstein, J. (2004) *Inferring Phylogenies,* Sinauer
78. Sober, E. (2008) *Evidence and Evolution: The Logic behind the Science,* Cambridge University Press
79. Huelsenbeck, J.P. and Rannala, B. (2004) Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Syst. Biol.* 53, 904–913
80. Kass, R.E. and Raftery, A.E. (1995) Bayes factors. *J. Am. Stat. Assoc.* 90, 773–795