Bayesian molecular dating: opening up the black box

Lindell Bromham[1,*,†], Sebastián Duchêne[2,3], Xia Hua[1], Andrew M. Ritchie[3], David A. Duchêne[1,3] and Simon Y. W. Ho[3,†]

[1]*Macroevolution & Macroecology, Division of Ecology & Evolution, Research School of Biology, Australian National University, Canberra, ACT 2601, Australia*
[2]*Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Melbourne, VIC 3010, Australia*
[3]*School of Life and Environmental Sciences, University of Sydney, Sydney, NSW 2006, Australia*

ABSTRACT

Molecular dating analyses allow evolutionary timescales to be estimated from genetic data, offering an unprecedented capacity for investigating the evolutionary past of all species. These methods require us to make assumptions about the relationship between genetic change and evolutionary time, often referred to as a 'molecular clock'. Although initially regarded with scepticism, molecular dating has now been adopted in many areas of biology. This broad uptake has been due partly to the development of Bayesian methods that allow complex aspects of molecular evolution, such as variation in rates of change across lineages, to be taken into account. But in order to do this, Bayesian dating methods rely on a range of assumptions about the evolutionary process, which vary in their degree of biological realism and empirical support. These assumptions can have substantial impacts on the estimates produced by molecular dating analyses. The aim of this review is to open the 'black box' of Bayesian molecular dating and have a look at the machinery inside. We explain the components of these dating methods, the important decisions that researchers must make in their analyses, and the factors that need to be considered when interpreting results. We illustrate the effects that the choices of different models and priors can have on the outcome of the analysis, and suggest ways to explore these impacts. We describe some major research directions that may improve the reliability of Bayesian dating. The goal of our review is to help researchers to make informed choices when using Bayesian phylogenetic methods to estimate evolutionary rates and timescales.

Key words: phylogenetics, priors, evolutionary models, molecular clocks, relaxed clocks, model selection.

CONTENTS

* Address for correspondence (E-mail: lindell.bromham@anu.edu.au)
  † Authors contributed equally to this work.

I. INTRODUCTION

(1) A brief history of molecular time

The evolutionary history of life on Earth spans several billion years, but much of this history remains unknown. Fossilization is rare and for many lineages we have no direct evidence of past organisms. Analysis of nucleotide and amino acid sequences from the genomes of modern species provides an exciting opportunity to peer into the evolutionary past of a much wider range of organisms. Molecular evolution is continuous and leads to the accumulation of genetic differences between lineages over time. Therefore, the amount of genetic divergence between species can be used to estimate the age of their last common ancestor, provided that we are willing to make assumptions about the relationship between genetic divergence and time. Models that describe this relationship are often referred to as 'molecular clocks'.

The relationship between biochemical differences and degree of evolutionary relatedness among species had been observed since the early 1900s (see: Morgan, 1998). The idea of a molecular clock emerged from early comparative studies of molecular evolution in the 1960s that showed that the proportion of amino acid differences between proteins of different species was related to the time since their evolutionary divergence (Zuckerkandl & Pauling, 1962, 1965; Margoliash, 1963; Doolittle & Blombäck, 1964). Many biologists were initially sceptical about the idea of a molecular clock (Morgan, 1998), because they expected that the pace of molecular evolution would be irregular, driven by occasional bouts of adaptation to changing environmental conditions. Nonetheless, it did not take long before molecular dates were being used to overturn ideas about human evolution (Sarich & Wilson, 1967), suggesting that humans had split from other apes more recently than was then generally believed (Wilson & Sarich, 1969). These early molecular date estimates, which placed the chimpanzee–human split around 5 million years ago, were initially shocking but are now generally accepted (Bradley, 2008).

The first molecular dating studies rested on the assumption that a rate calculated for one lineage could be used to estimate the dates of all other evolutionary divergence events in the phylogeny (Zuckerkandl & Pauling, 1962). The assumption of rate constancy was agnostic with respect to the causes and patterns of molecular evolution. In particular, it did not depend on assuming that most changes were neutral (rather than driven by natural selection), nor on rates being similar across all of the amino acid sites in a protein. Instead, the rate of evolution was assumed to average out over long time periods to leave a strikingly linear relationship between the number of amino acid differences in proteins from different species and their dates of divergence estimated from fossil evidence (Dickerson, 1971). In this context, the term 'molecular clock' is perhaps best understood as a statement about the ability to infer time from genetic sequence data, not as a claim about a regular 'tick rate' of changes in the genome.

In fact, the rate of molecular evolution not only fluctuates over time and across the genome, but it also varies among lineages. A wide range of species traits have been shown to correlate with the rate of molecular evolution, including body size, generation time, population size, behaviour, niche and social structure (Bromham, 2009). There might be additional broader-scale effects on lineage-specific rates of molecular evolution, such as an impact of environment or diversification rate (Venditti & Pagel, 2009; Gillman & Wright, 2014; Bromham *et al.*, 2015). Lineage-specific rates complicate the phylogenetic inference of divergence times, because we cannot assume that the evolutionary rate calculated for one lineage is an accurate representation of rates along other lineages in the tree.

The central problem in molecular dating is that we need to separate the branch lengths on a phylogeny (reflecting the amount of genetic change) into their two components: evolutionary rates and time durations. To do this, we need to incorporate independent information to anchor one or the other, for example using rate estimates from previous studies or temporal information from the geological record. This is

known as calibrating the molecular clock and it can have a profound impact on the date estimates that are obtained.

Since the earliest molecular dating studies over half a century ago, there has been continual modification and improvement of methods for estimating evolutionary timescales from molecular sequence data. The past two decades have seen the development of various distance-based and maximum-likelihood methods for estimating evolutionary rates and timescales [for recent reviews, see Ho & Duchêne, 2014 and Kumar & Hedges, 2016]. However, one of the most important advances has been the development of the Bayesian phylogenetic framework (Rannala & Yang, 1996; Mau & Newton, 1997; Li, Pearl & Doss, 2000). This has made it possible to implement complex, parameter-rich models of evolution, including models of rate variation across lineages. As a result, Bayesian methods are now widely used not only for molecular dating (Donoghue & Yang, 2016), but also to infer demographic history (Drummond *et al.*, 2005), phylogeographic processes (Lemey *et al.*, 2009), and viral phylodynamics (Pybus & Rambaut, 2009). The capacity for joint inference of various evolutionary and demographic quantities, while taking into account the uncertainty in the phylogeny and other model parameters, is one of the most appealing strengths of the Bayesian approach.

While there is a tendency to place faith in the latest or most complex methods, these are not always more reliable than simpler ones (Sullivan & Joyce, 2005). The more complex the method, the more assumptions we have to make about the evolutionary process. Some of these assumptions may provide a more realistic description of the complexities of evolutionary patterns and processes. But some of the assumptions in phylogenetics are made for statistical tractability rather than being necessarily based on understanding of the underlying evolutionary processes.

In Bayesian molecular dating, we have a set of fundamental evolutionary assumptions that are applied almost universally. For example, all molecular dating analyses are built upon the assumption that the sequences are accurately aligned so that only homologous sites are compared. In addition to the fundamental evolutionary assumptions, there are assumptions that are built into the model, such as the form of the substitution model. These are typically fixed in any given analysis, but we could change the assumptions and run the analysis again. Finally, and specific to Bayesian analyses, we have a set of 'priors' that encapsulate our beliefs about the evolutionary history and processes: during the analysis, we update our beliefs after considering evidence from the data. Collectively, all of these assumptions can have complex impacts on Bayesian estimates of evolutionary rates and timescales, though their influence might not always be clearly discernible.

The aim of this review is to describe the process of Bayesian molecular dating, including the range of models and priors that are commonly used, and to consider how the choices of methods and assumptions can affect date estimates. Although there are many types of genetic data that can be used profitably in molecular dating studies (Ho *et al.*, 2016), the focus of this review will be on DNA sequences, which are the dominant form of data used in Bayesian molecular dating.

(2) The Bayesian framework

The Bayesian framework is built on Bayes's theorem, which states that the posterior probability distribution is given by the product of the prior distribution and the likelihood, divided by the probability of the data. In phylogenetics, the theorem takes the following form:

$$P(\theta|D) = P(\theta) P(D|\theta) / P(D)$$

where $P(\theta|D)$ is the probability distribution of the parameters given the data (posterior probability), $P(\theta)$ is the probability distribution of the parameters (prior probability), $P(D|\theta)$ is the probability of the data given the parameters (likelihood), and $P(D)$ is the probability of the data.

The likelihood score, $P(D|\theta)$, is evaluated using the likelihood model, which includes the tree and the substitution model. It represents the probability that we would have obtained the observed sequence data, if this particular history of the sequences were true (that is, given a particular tree topology, branch lengths, and model of the substitution process). The priors represent our beliefs about the parameters in the likelihood model, without considering any information from the data. In some cases, prior distributions can be generated using models of the evolutionary process. These models can have parameters called hyperparameters, and their prior distributions are referred to as hyperpriors. The likelihood model and the priors are collectively referred to as the 'hierarchical model', because they involve multiple layers of models, including the components of the likelihood model, the priors on the parameters of the likelihood model, and the hyperpriors on the hyperparameters of the priors (Fig. 1).

The probabilities of the different sets of parameter values are then updated in light of what we learn from the data (the likelihood score), which might provide more support for some values than others. This process yields the posterior probability distributions of all of the parameters in the model. If the posterior differs from the prior, we can conclude that the DNA sequences contained information that changed our mind about the most likely explanations for the history of those sequences. If the prior and posterior probability distributions are very similar, then it is likely that the evidence from the data was weak, our prior beliefs were too strong, or a combination of both.

A key problem in Bayesian phylogenetics lies in the calculation of the posterior distribution. The denominator in Bayes's theorem, the probability of the data, cannot be obtained analytically except in trivial cases. Instead, we circumvent the need to calculate this quantity by dealing only with ratios of posterior probabilities, which allow the problematic denominator to be cancelled out. This is an important feature of the Markov chain Monte Carlo (MCMC) method, which involves estimating the posterior distributions of the model parameters by drawing a large number of samples from these distributions in an MCMC
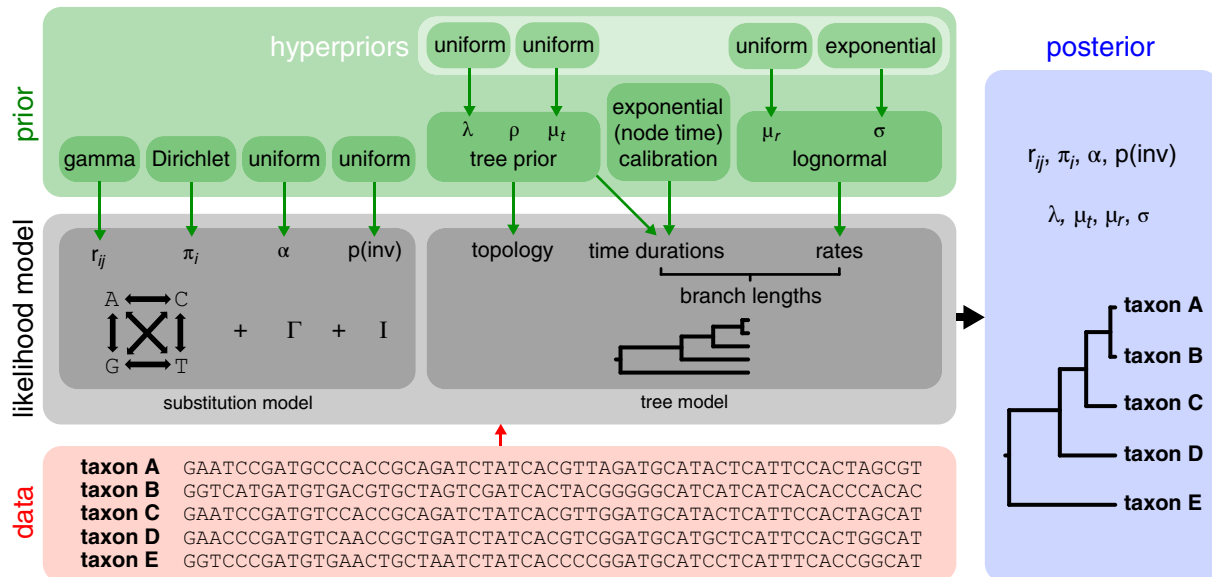
Fig. 1. Components of a typical Bayesian dating analysis, including the data, likelihood model, prior, and posterior. The data set is an alignment of five nucleotide sequences. The likelihood model consists of the substitution model [general time-reversible model with a proportion of invariable sites and gamma-distributed rates across sites (GTR + I + Γ)] and the tree model (topology and branch lengths). The substitution model includes parameters for the rates of transitions between bases (r_{ij}), base frequencies (π_i), the shape parameter for the gamma distribution of rates across sites (α), and the proportion of invariable sites [p(inv)]. In the tree model, the branch lengths are the product of the branch durations and branch rates. Prior distributions are specified for all of the parameters in the likelihood model, including the tree topology. A birth–death model is used to generate the tree prior, which provides the joint prior distribution of topology and branch durations. The birth–death model has parameters representing the speciation rate (λ), extinction rate (μ_t), and sampling probability (ρ). The speciation and extinction rates have their own priors, known as hyperpriors. The prior distribution of branch durations is also influenced by calibration densities for node times, which in this example are formulated as exponential distributions. This example uses an uncorrelated lognormal relaxed clock for the prior distribution of branch rates. The posterior distribution of parameters, including the tree, is the combination of the prior and the data (likelihood).

simulation (Yang & Rannala, 1997; Li *et al.*, 2000). This is the most common method for obtaining the posterior distribution in Bayesian phylogenetics.

The most widely used method for MCMC simulation in Bayesian phylogenetics is the Metropolis–Hastings algorithm (Metropolis *et al.*, 1953; Hastings, 1970). This algorithm starts with a tree and a set of values for the model parameters, evaluates the likelihood that the tree could have produced the observed data, given the evolutionary model, and weights the likelihood by the prior probabilities (Fig. 2). Then it proposes a change to one or more of the model parameters and calculates the likelihood of this new set of parameter values, weighted by the prior probabilities. The MCMC compares these two alternative histories of the sequence data, differing in at least one aspect of the model, each with a likelihood score and prior probability. By taking the ratio of posterior probabilities of these two alternative histories, the term describing the probability of the data neatly cancels out. This ratio is used to decide whether the MCMC stays on the first state or moves to the new state. However, unlike a likelihood hill-climbing procedure, the decision to move or stay on the current state is guided by the posterior probabilities, but also has a chance element that means that it might move even if the proposed state has a

slightly lower posterior probability (see Fig. 2). This stochastic element comes from using a random number drawn from a uniform distribution between 0 and 1. If the ratio of the posterior probability of the proposed state to the posterior probability of the current state is higher than the random number, the proposal is accepted. But if the ratio of posterior probabilities is lower than the random number, then the proposal is rejected and the chain stays on the current state.

We can picture the MCMC chain as moving through parameter space, representing the set of all possible phylogenetic solutions. Each point in this parameter space differs from its neighbours in some aspect of the topology, or a branch length, or a parameter value. We expect the MCMC to visit each tree with a frequency equal to its posterior probability. During the MCMC procedure, we keep a record of the trees and parameter values that have been visited (Fig. 2). These samples provide a means of estimating the posterior probability distribution of each parameter in our model.

Bayesian molecular dating is a complex process that involves a range of assumptions and the specification of prior beliefs about model parameters. In many cases changing the models and priors will influence the outcome of the analysis, but the impacts of these choices are often not directly investigated. If different analyses produce conflicting results,
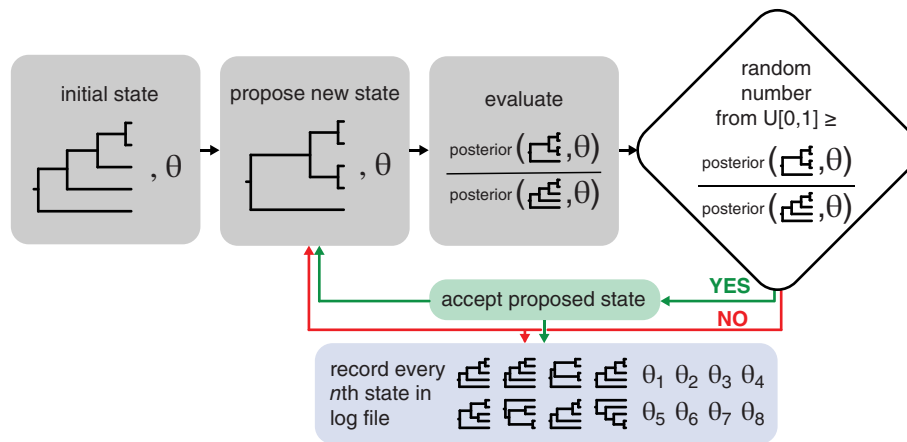
Fig. 2. Flowchart describing a simplified process of Markov chain Monte Carlo (MCMC) for Bayesian molecular dating. An initial state, consisting of a tree topology, branch lengths, and parameter values (represented by θ), is either chosen randomly or specified by the user. A new state is proposed by making incremental changes to the branch lengths, model parameters, and tree topology. Depending on the program, each step can propose a single or multiple changes. The initial and proposed states are evaluated for their unnormalized posterior probability and a ratio calculated. If the random number drawn from a uniform distribution between 0 and 1 is greater than the ratio of the two posterior probabilities, then the proposal is rejected and the chain stays on the current state, but if the random number is smaller than the ratio of the two posterior probabilities, the proposal is accepted and it becomes the current state. The process continues until it reaches an arbitrary stopping point or when the user decides that a sufficient number of samples has been drawn.

we generally do not know which set of date estimates is correct unless we have additional, uncontroversial information about the evolutionary timescale. But, by the very nature of the problem, there are relatively few data sets for which we can compare our date estimates against a known timescale. This is partly because molecular dating analyses are used to infer unknown divergence times, so are rarely applied to cases where the answer is already known. More importantly, however, Bayesian molecular dating typically incorporates all prior information about divergence times, so the resulting date estimates cannot be independently used to evaluate the estimates.

There are surprisingly few cases in which a subset of calibrating information is used to estimate the dates of other nodes for which the age is already known. One example is a Bayesian analysis of DNA from bison (*Bison priscus*), which correctly identified the ages of most of the ancient DNA sequences from samples that had already been radiocarbon-dated (Shapiro *et al.*, 2011). Bayesian date estimates can be evaluated in light of other evidence for an evolutionary timescale. For example, molecular date estimates for the origin of planktonic algae (Sánchez-Baracaldo *et al.*, 2017) have been considered to be consistent with biogeochemical evidence (Knoll, 2017).

In the absence of independent information on the evolutionary timescale with which to evaluate the veracity of Bayesian date estimates, simulation studies can be used to explore the performance of the methods under different evolutionary processes and patterns. Studies of synthetic data generated by simulation under known conditions can tell us how reliable methods are in specific (and often simplified) scenarios, but we can never be certain that those conditions

are an appropriate reflection of reality. If we lack the means to give a definitive answer on which methods are always reliable, how are we to decide which are the best choices for methods and models for a given data set? This choice will depend partly on the characteristics of the genetic data, including the nature of the markers being analysed and the sampling scheme.

II. DATA

(1) Data selection

Molecular dating analyses are based on comparing homologous sequences in order to reveal the history of populations or lineages. These are usually sequences of a number of loci, representing a miniscule fraction of the whole genome. The hope is that the chosen loci will be adequate to infer the history of the lineages from which they were sampled. Since different loci can evolve at different rates and have different evolutionary histories, the model of molecular evolution employed in a dating analysis will need to be appropriate to the data being analysed.

In a molecular dating study, DNA sequences in the alignment can represent conspecific individuals, populations, species, or higher taxa. Few data sets contain sequences representing all the descendants of the last common ancestor at the root of the tree. Instead, most molecular dating studies rely on an incomplete sample of the tips (terminal nodes) of the phylogeny. Increasing the density of taxon sampling to include more tips in the phylogeny will often improve inferences of topology (Heath, Hedtke & Hillis, 2008; Soares

& Schrago, 2012) and will also increase the accuracy of branch-length estimates, by reducing the node-density effect (Hugall & Lee, 2007). The node-density effect occurs when branch lengths are underestimated due to failure to account fully for multiple substitutions at the same site. Studies have shown that as taxon sampling increases, mean age estimates approach the values inferred from the full data set and the variance of the estimates decreases (Linder, Hardy & Rutschmann, 2005; Schulte, 2013). The impact of taxon sampling is likely to be important when there is a large degree of rate variation across branches (Soares & Schrago, 2015).

Taxon sampling is an important consideration in all molecular dating analyses, because of the potential impact on the accuracy and precision of estimates of both topology and branch lengths. In most Bayesian analyses, taxon sampling is assumed to be random with respect to phylogenetic relationships. This assumption is violated for many data sets, particularly those in which sampling has been designed to include representatives of all higher taxa, regardless of differences in their species richness. This results in non-random sampling, because tips within species-poor taxa have a proportionally greater chance of being included in the phylogeny than tips within diverse taxa. For example, in a recent bird phylogeny that included representatives of every order (Prum et al., 2015), the order Opisthocomiformes had 100% of tips sampled (because the sole species, hoatzin, was included) while the species-rich Passeriformes had fewer than 1% of its tips sampled (because 44 species out of over 5000 were included). Although such sampling biases might seem trivial, they can sometimes lead to inaccurate date estimates (Welch, Fontanillas & Bromham, 2005; Beaulieu et al., 2015; Ronquist, Lartillot & Phillips, 2016). Modifications to the tree prior are available for data sets in which taxa have been sampled in closely related clusters, or dispersed across clades so as to represent as much diversity as possible (Höhna et al., 2011; Ronquist et al., 2016).

As DNA sequences become easier to obtain in large amounts, it is becoming less common to estimate molecular dates using single-locus data sets. Combining multiple loci in a single analysis brings many advantages, including increased data and more representative sampling of the genome, but it also brings challenges. In most multi-locus studies, the loci are concatenated (joined end-to-end) and analysed together. The concatenation approach makes the implicit assumption that all loci share the same history and so can be described with a single tree, although loci are usually allowed to have different substitution models (Yang & Rannala, 1997; Pupko et al., 2002) and can be permitted to evolve at different rates (Ho & Duchêne, 2014; Yang, 2014). However, many evolutionary processes can cause loci to have different phylogenetic histories, including recombination, gene duplication, and horizontal gene transfer (Maddison, 1997; Degnan & Rosenberg, 2009).

Phylogenetic analyses can potentially be misleading if loci with incongruent evolutionary histories are assumed to share a single tree (Posada & Crandall, 2002; Mossel & Vigoda,

2005; Kubatko & Degnan, 2007). It is possible to test for congruence between gene trees, and analyse only those loci that seem to share the same evolutionary history (Leigh et al., 2008). In this case, however, the analysis is focusing on the history of one set of loci in the genome, which might not adequately represent the evolutionary relationships of the lineages being studied. An alternative approach is to analyse the loci separately to infer their gene trees, and then to combine these to infer a single phylogenetic history, or species tree (Szöllősi et al., 2015). The gene trees and the species tree can be jointly estimated in a Bayesian framework (Rannala & Yang, 2008; Heled & Drummond, 2010), but this is only computationally feasible for data sets with relatively small numbers of loci and species.

Molecular data can be combined with other forms of data in a 'total-evidence' approach to dating (Ronquist et al., 2012; O'Reilly, dos Reis & Donoghue, 2015). For example, morphological features can be coded as multistate characters or as continuous traits. However, the inclusion of morphological characters to inform branch-length estimates in a dating analysis presents some challenges. For example, while molecular analyses will typically include all of the sites in the included sequences, the selection of morphological characters is often biased towards traits that are diagnostic in taxonomy. This means that morphological data sets will tend to favour traits that are shared by multiple lineages, with relatively less sampling of autapomorphies (character states found in only a single taxon) and invariant characters (found in all lineages); this will have potential impacts on the dating analysis (Gavryushkina et al., 2017).

(2) Sequence alignment

The process of Bayesian molecular dating typically begins with an alignment of nucleotide sequences arranged so that each column of the alignment represents a site that has been copied from the same site in the ancestral sequence. For sequences that are correctly aligned, differences between them are due to nucleotide replacements, insertions, or deletions that have occurred during the history of descent of these lineages from their common ancestor. If any columns in the alignment do not represent homologous sites, then any evolutionary inference drawn from those sites is misguided. Misalignment can lead to inaccurate estimates of evolutionary divergence and therefore influence the inferred molecular dates.

Sequence alignment is often given relatively little consideration in phylogenetic studies, and is typically treated as a component of data preparation rather than a part of phylogenetic inference itself (Löytynoja & Goldman, 2009; Morrison, 2009). While the sequences themselves are observations about the world (barring errors in sequencing, assembly, or annotation), the alignment is a hypothesis about homology rather than an error-free observation (Wong, Suchard & Huelsenbeck, 2008). Irrespective of whether the sequences were aligned algorithmically or manually, or by a combination of the two (automated alignment with manual adjustment), we can rarely assign homology across sites with

certainty. Alternative alignments of the same sequences can lead to different estimates of the phylogeny, substitution rates, and branch lengths (e.g. Fontanillas *et al.*, 2007; Wong *et al.*, 2008; Liu, Linder & Warnow, 2011).

There are several possible approaches to incorporating alignment error into a molecular dating study. One is to iterate the phylogenetic analysis over many possible sequence alignments and to assess the sensitivity of the results to variations in the alignment (Wong *et al.*, 2008). Alternatively, alignments can be weighted with measures of confidence, either using metrics to assign prior probabilities to different alignment solutions or weighting individual sites in an alignment (Penn *et al.*, 2010; Wu, Chatterji & Eisen, 2012). Another approach is to move away from the classic two-step process of first optimizing the alignment then inferring the phylogeny, and instead move towards joint estimation of the tree and the sequence alignment (Lunter *et al.*, 2005).

III. COMPONENTS OF THE LIKELIHOOD MODEL

The probability of the data given a particular model is known as the likelihood. We can calculate the likelihood score using the likelihood model, comprising the tree model and the substitution model. The tree model includes the hierarchical relationships among nodes, known as the tree topology, as well as the lengths of the branches connecting the nodes. Nodes in the tree can represent sampled taxa (terminal nodes or tips) or their hypothetical ancestors (internal nodes), including the common ancestor of all of the sampled taxa (root node). Typically, the terminal nodes are fixed at the same time point (the present) but the positions of the internal nodes are estimated in the analysis. However, when DNA sequences from ancient samples are included, the depths of the terminal nodes can be either fixed at the age of the sample or treated as estimable parameters (Shapiro *et al.*, 2011). In combined total-evidence analyses of extant and fossil taxa, the terminal nodes can be millions of years old (Ronquist *et al.*, 2012; Gavryushkina *et al.*, 2014).

The substitution model allows us to estimate the amount of genetic change that has occurred since the observed sequences were copied from their last common ancestor. The simplest way to estimate genetic distance is to count the observed differences between the aligned DNA sequences. However, this 'Hamming distance' (or p-distance) will often underestimate the number of nucleotide substitutions that have actually occurred because there is a chance that some of them have occurred on top of previous substitutions, erasing evidence of past changes (Jukes & Cantor, 1969). We cannot directly observe overwritten substitutions, but we can use a model of the substitution process to infer how many of these 'multiple hits' might have occurred (Fig. 3). The substitution model uses statements about the probabilities of different kinds of substitution events to infer how many substitutions are likely to have occurred in the past, given the patterns of sequence variation that we observe in our alignment.

As with other parts of the Bayesian hierarchical model, the substitution model incorporates a range of assumptions about the molecular evolutionary process. The most commonly used substitution models are members of the general time-reversible (GTR) family (Jukes & Cantor, 1969; Hasegawa, Kishino & Yano, 1985; Tavaré, 1986), which share a number of assumptions about the data and their evolutionary history. Some of these assumptions determine the form of the model and are fixed components of the analysis, whereas other assumptions are encapsulated in the priors. A common example of a fixed assumption is that nucleotide substitutions are independent and identically distributed (i.i.d.). This means that every substitution in the alignment is considered to occur independently of all others, so a substitution at one site does not make a substitution at any other site more or less likely to happen. Actually, we have good biological reasons to suspect that substitutions at one site can affect the chance of accumulating substitutions at another site. For example, a copy of a gene with changes in one part of the sequence might be more likely to persist down the generations if it also has a compensating mutation that restores the reading frame downstream (e.g. Wernegreen, Lazarus & Degnan, 2002), keeps binding sites able to interact with each other (e.g. Nekrutenko *et al.*, 2005), or allows the gene product to keep working with other gene products (e.g. Luo *et al.*, 2013). There has been some development of models that allow site-dependency of substitution (Baele, Van de Peer & Vansteelandt, 2008).

Another fixed assumption in most analyses is that the same substitution model applies throughout the shared history of the sequences being analysed, such that there have been no changes in the frequencies of the four nucleotides in the alignment (model stationarity) and in the probabilities of different substitution types (model homogeneity). Again, these assumptions might not hold strictly true for all sequences (Lockhart & Cameron, 2001). Given that we know that nucleotide composition can vary between lineages, base frequencies must be able to evolve along the phylogeny (Gruber, Voss & Jansa, 2007).

Substitution models typically include three main components (Fig. 3): the substitution matrix that describes the relative rates of change between different nucleotides, such as the probability that an A will change to a T; the frequencies of the four nucleotides; and the way in which rates vary across sites. Within the general time-reversible family of models, the GTR model is the most complex as it allows a different rate for each kind of nucleotide change. Other models within the general time-reversible family differ in how many of the substitution rates are set to be equal to each other (Fig. 3). For example, the simplest model [the Jukes–Cantor (JC) model] assumes that all kinds of substitutions are equally likely, so can be described by a single substitution rate parameter (Jukes & Cantor, 1969). More complex models in this family can incorporate unequal patterns of substitution types, different rates of transitions to transversions (Kimura, 1980), or different nucleotide frequencies (Felsenstein, 1981). The general time-reversible family of models is most often
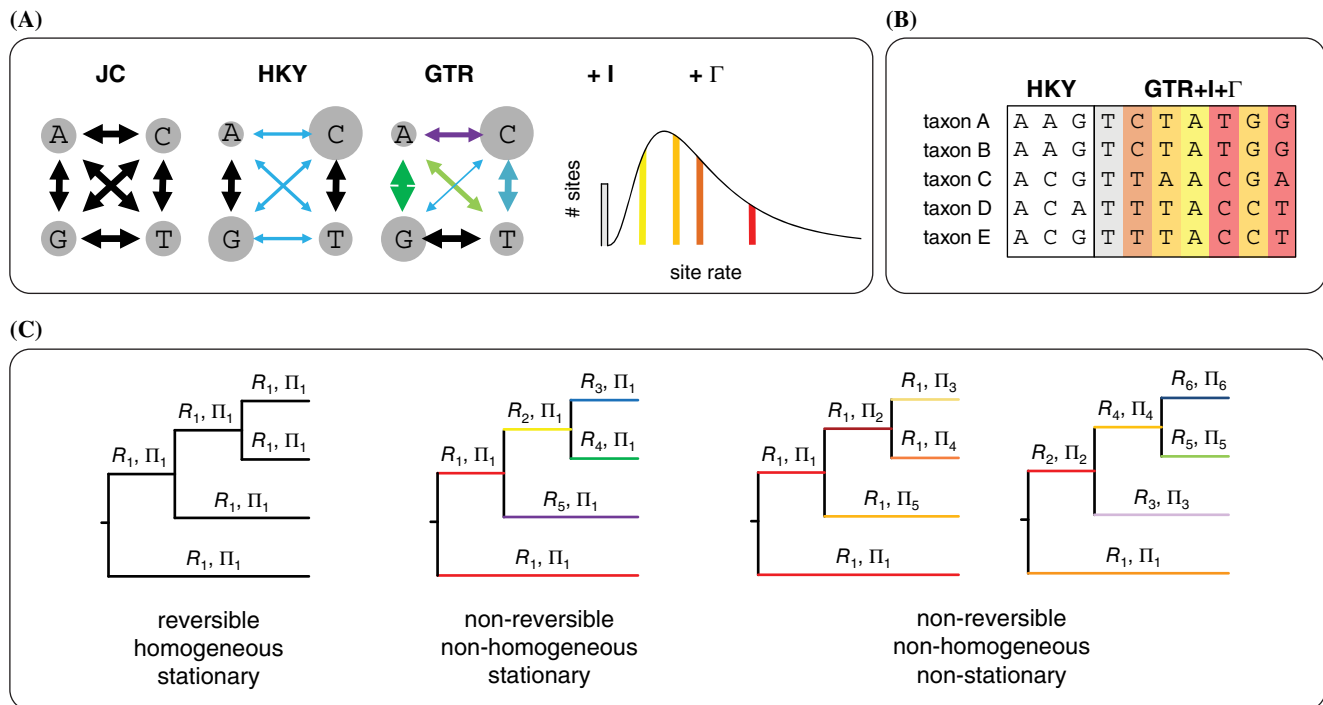
Fig. 3. Models of the nucleotide substitution process. (A) Substitution models of the general time-reversible family comprise the replacement rates between pairs of nucleotides and the equilibrium base frequencies. In the Jukes–Cantor (JC) model, all six of the nucleotide replacement rates are equal. In the HKY model, the nucleotide replacement rates can be constrained so that transitions and transversions occur at different rates. In the general time-reversible (GTR) model, the six replacement types occur at different rates. These rate matrices can be combined with a proportion of invariable sites and with a gamma distribution to model rate variation across sites. (B) An example of a five-taxon data set that has been partitioned into two subsets. An HKY (Hasegawa–Kishino–Yano) model has been assigned to the first subset, whereas a GTR + I + Γ (general time-reversible model with a proportion of invariable sites and gamma-distributed rates across sites) model has been assigned to the second subset. (C) Substitution models consist of rate matrices (R_i) and base frequency vectors ($\prod_i$). Different combinations of these model components are shown in different colours along the branches of the trees. Most substitution models assume that the evolutionary process is time-reversible, homogeneous, and stationary, as shown in the left tree. However, these assumptions can be relaxed by varying the substitution model across branches. If the rate matrix varies across the tree, but base frequencies are constant, the process is considered to be stationary, but neither reversible nor homogeneous (second tree). If base frequencies also vary through the tree, then the process is also non-stationary (third and fourth trees).

applied to nucleotide data because these models are highly computationally efficient and have been shown to be robust under a wide range of conditions in simulation studies (Sullivan & Joyce, 2005).

There are equivalent time-reversible models for amino acid data (Adachi & Hasegawa, 1996; Yang, Nielsen & Hasegawa, 1998; Huelsenbeck *et al.*, 2008), but these are rarely used because they require large numbers of parameters to be estimated in the analysis. Instead, the most common approach for protein sequences is to fix the parameter values using empirical estimates of the relative rates of change for each kind of amino acid transition, estimated from large protein data sets (e.g. Dayhoff, Schwartz & Orcutt, 1978; Jones, Taylor & Thornton, 1992; Whelan & Goldman, 2001).

In addition to variation in rates between different kinds of substitution, substitution models can allow for variation in rates across sites in an alignment (Fig. 3). Failure to account for variation in rates across sites can have negative impacts on Bayesian date estimates (Marshall, Simon & Buckley,

2006), and different models of rates across sites can result in changes to date estimates (Brandley *et al.*, 2011; Soubrier *et al.*, 2012; Jia, Lo & Ho, 2014). Variation in substitution rate across sites is usually modelled by assuming that the rates follow a parametric distribution (Yang, 1993, 1994). The gamma distribution is commonly used for this purpose because it is non-negative and simple to parameterize, and its various forms can resemble the lognormal and exponential distributions. So the popularity of the gamma distribution is due to its practical benefits rather than any empirical evidence that it is biologically realistic (Felsenstein, 2001; Jia *et al.*, 2014).

The choice of substitution model, or the form of the priors on the parameters of the model, can affect date estimates (Duchêne *et al.*, 2015; Marshall *et al.*, 2016). In particular, using an underparameterized substitution model can cause the amount of genetic change to be underestimated (Lemmon & Moriarty, 2004). However, the impact of the choice of substitution model is not always easy to predict. For

example, one case study reports that a more parameter-rich model (GTR with empirical nucleotide frequencies) resulted in strikingly younger date estimates than a simpler model [Hasegawa–Kishino–Yano (HKY) model with nucleotide frequencies estimated in the analysis (Marshall *et al.*, 2016)]. But few studies compare date estimates under different substitution models, so it is difficult to judge how important the choice of substitution model is to Bayesian molecular date estimates.

We can take one of two approaches to choosing which substitution model or models to use, and how to assign them to different subsets of the data (i.e. the partitioning scheme). One approach is to use knowledge of the loci being analysed to select an appropriate model. For example, the boundaries between genes in multi-locus alignments can be used to assign separate substitution models to different loci (or groups of loci), and protein-coding sequences might be best described by assigning separate substitution models to the three codon positions. Alternatively, we could use a model-selection procedure that compares the fit of different models, or combinations of models, to the data set (Brandley *et al.*, 2011; Lanfear *et al.*, 2012).

In a 'total evidence' dating analysis, a separate substitution model is needed for the non-molecular characters. The most commonly used is the Mk model, which is a generalization of the JC substitution model that allows different numbers of character states. An extension of this model, the Mkv model, includes a correction for ascertainment bias in cases where the data set only consists of variable characters (Lewis, 2001). The assumption of independence across characters is often problematic for morphological characters, which tend to show higher correlation with each other than do nucleotide sites (Ronquist *et al.*, 2016).

IV. PRIORS ON THE SUBSTITUTION MODEL

There are many priors and hyperpriors on the substitution model in Bayesian dating analyses, including those on the substitution rates, frequencies of the four nucleotides, and rates across sites. The choice of prior distribution depends partly on the nature of the parameter being considered. For example, the relative rates of the different kinds of substitutions must be non-negative real numbers. We might have ideas about these relative rates – for example that transitions are more common than transversions – but we generally do not have independent information about what values those rates should take for any given data set.

When we have no information about the values a parameter should take, we can specify an 'uninformative' prior distribution that gives equal weight to a broad range of possible values. A simple choice of prior for parameters like the transition rates in the substitution model is the uniform distribution. However, a uniform distribution gives the same probability to the extreme values at the edges of a chosen range as it does to intermediate values. Furthermore, reparameterization of the uniform prior can artificially inflate

the probabilities of some parameter values over others (Drummond *et al.*, 2002). This can result in a prior that is 'too informative', so that the prior overwhelms the signal from the data (Zwickl & Holder, 2004), and the date estimates might be more strongly influenced by the priors than by information from the molecular sequences. One way to avoid this problem is to use a Jeffreys prior (Jeffreys, 1961). This is a mathematical transformation that ensures that the prior distribution is invariant under rescaling, which can lead to more accurate estimates than using uniform priors (Drummond *et al.*, 2002).

Some groups of parameters add up to a fixed value, such as the nucleotide base frequencies which must add up to 1. A group of parameters with this constraint is often referred to as a simplex. A Dirichlet distribution is an appropriate prior for simplex parameters such as base frequencies, because it treats the frequency of each base as following a continuous probability distribution with values between 0 and 1. The Dirichlet prior is flexible because its parameters can be varied to reflect any expectations of biased nucleotide content.

V. PRIORS ON THE TREE MODEL

(1) Tree priors

One of the central problems in Bayesian molecular dating is how to assign prior probabilities to the phenomenally large number of possible phylogenies that could explain the sequence data we observe (Rannala & Yang, 1996). In some cases, we might have specific information about the topology (relationships) or node times (dates), but we will rarely have sufficient knowledge to place a prior probability on every possible phylogenetic history of the sequences. Instead, Bayesian phylogenetic methods typically rely on using a function for assigning prior probabilities to different tree topologies, as well as the prior distributions of node times across the phylogeny. There are two general approaches to placing priors on the distribution of node times: one is to specify a prior probability for each tree topology first and then specify a prior distribution for all node ages given each tree topology, and the other is to use a simplified model of the evolutionary process to assign probabilities to different sets of topologies and node times.

The two-step, topology-first approach starts by placing a prior probability on the particular order in which lineages split from one another and then specifies the prior distribution of the branch lengths or node times. The most common prior on topologies is the 'flat' or uniform prior, which considers every possible topology to be equally probable (Huelsenbeck *et al.*, 2002). We then need a way of assigning probabilities to different sets of branch lengths for the same topology. Two commonly used priors favour trees with divergence events spread evenly between the root and the present: the uniform prior (Lepage *et al.*, 2007) and the Dirichlet prior (Kishino, Thorne & Bruno, 2001). Although these two priors differ in the way in which node times are sampled, their resulting prior distributions are similar.

Most Bayesian dating analyses use a stylized model of the evolutionary process to derive a single distribution for both the topology and node times. The practical effect of the tree prior is to give more weight in our analysis to trees of a certain broad appearance, compatible with a particular model of the branching process. There are two classes of models that are typically used to generate tree priors for molecular dating: the birth–death process, which models speciation and extinction of lineages; and the coalescent, which models the evolutionary process within a species or population (Fig. 4).

(2) Birth–death process

The birth–death process represents the speciation and extinction of lineages in a highly simplified form (Kendall, 1948; Rannala & Yang, 1996; Gernhard, 2008). This model has parameters describing the rate of speciation ('birth') and the rate of extinction ('death'). The simplest is the Yule process, which is a pure-birth model because it only models speciation without allowing for extinction of any lineages during the history represented by the phylogeny (Yule, 1924). The Yule model assumes that the speciation rate is constant across lineages and throughout the history of the sequences being analysed. Under this process, the number of lineages builds exponentially with time, so nodes will tend to be distributed more towards the tips.

For most data sets, we expect the shapes of phylogenies to have been affected not just by speciation, but also by extinction. Birth–death models, in addition to assuming a constant rate of speciation in each lineage, allow every lineage to have the same chance of going extinct at any point in time. As the ratio of the death rate to the birth rate increases, more of the older lineages in the phylogeny will have ended in extinction. This tends to remove the deeper nodes, leading to trees with long branches towards the root and shorter branches toward the tips (Fig. 4). We rarely have any independent information about the speciation and extinction rates, so these parameters of the birth–death model are usually estimated in the analysis.

The impact of the choice of tree prior on date estimates can range from being small (e.g. Foster et al., 2017) to substantial (e.g. Ho et al., 2005; Lepage et al., 2007; Heled & Drummond, 2015; Ritchie, Lo & Ho, 2017). For example, the Yule prior has been found to give much older date estimates than the birth–death prior when applied to the same data sets (Crisp, Hardy & Cook, 2014; Condamine et al., 2015). Since the tree prior contributes to the posterior distribution of all node times, using an inappropriate tree prior or applying unreasonable constraints on the parameters potentially influences the accuracy of the date estimates for every node in the tree.

Incomplete taxon sampling tends to increase the lengths of the terminal branches of the phylogeny by failing to sample some of the recent internal nodes (Yang, 2014). Some of the models that describe the diversification process can account for incomplete taxon sampling (Yang, 2006; Heath, Huelsenbeck & Stadler, 2014), for example through the addition of a parameter representing the proportion of extant species that are included in the sampled data set (Yang & Rannala, 1997). Sampling fraction is normally fixed in an analysis, since it cannot be estimated if the other parameters are not known (Stadler, 2009), but the value chosen can have strong effects on date estimates (Welch et al., 2005; Beaulieu et al., 2015). Unlike the birth- and death-rate parameters, the sampling fraction can often be independently determined based on the estimated number of extant species from taxonomic compendia (e.g. Foster et al., 2017). In these cases, the number of missing taxa will depend on taxonomic definitions, and will be influenced by inclusion of different taxonomic levels. For example, counting only recognized species and not subspecies or incipient species might underestimate the number of recent divergence events (Etienne & Rosindell, 2012). Most birth–death models that incorporate sampling fraction assume that the included taxa represent a random sample of the contemporary taxa, an assumption that is frequently violated (see Section II.1). Modifications to the tree prior are available for data sets in which taxa have been sampled in closely related clusters, or dispersed across clades so as to represent as much diversity as possible (Höhna et al., 2011; Ronquist et al., 2016).

There have been several interesting developments of the birth–death tree prior in recent years, including priors in which the birth and death rates are allowed to change over time (Stadler & Yang, 2013). Birth–death tree priors can also be modified to incorporate information from fossil morphology through the use of dated tips. One example of this is the fossilized birth–death prior (Heath et al., 2014). Dated fossils are treated as nodes in the tree, and all lineages are modelled as having a uniform probability of producing relevant fossil evidence. For molecular dating at the species level, the distribution of fossil evidence among lineages is incorporated in the model through a 'sampling proportion' parameter that reflects the expected fraction of lineages that leave fossil evidence before they go extinct (or arrive at the present).

Fossilized birth–death models are one illustration of the changing nature of fossil evidence in molecular dating, which is increasingly being used to inform the process of phylogenetic estimation itself. For example, a total-evidence dating analysis allows the placement of extinct taxa to be inferred using morphological characters. Birth–death priors can also be modified to allow fossils to represent samples of direct ancestors to descendent lineages (Gavryushkina et al., 2014). While total-evidence dating methods are becoming increasingly popular, they can suffer from several problems, such as undue sensitivity to the tree prior (Donoghue & Yang, 2016).

(3) Coalescent tree priors

If our sequences represent individuals sampled from a single population or species, assigning probabilities to different trees is usually done using a prior based on the coalescent process for population genetics (Kingman, 1982; Drummond et al., 2002). The coalescent traces the history of the sample
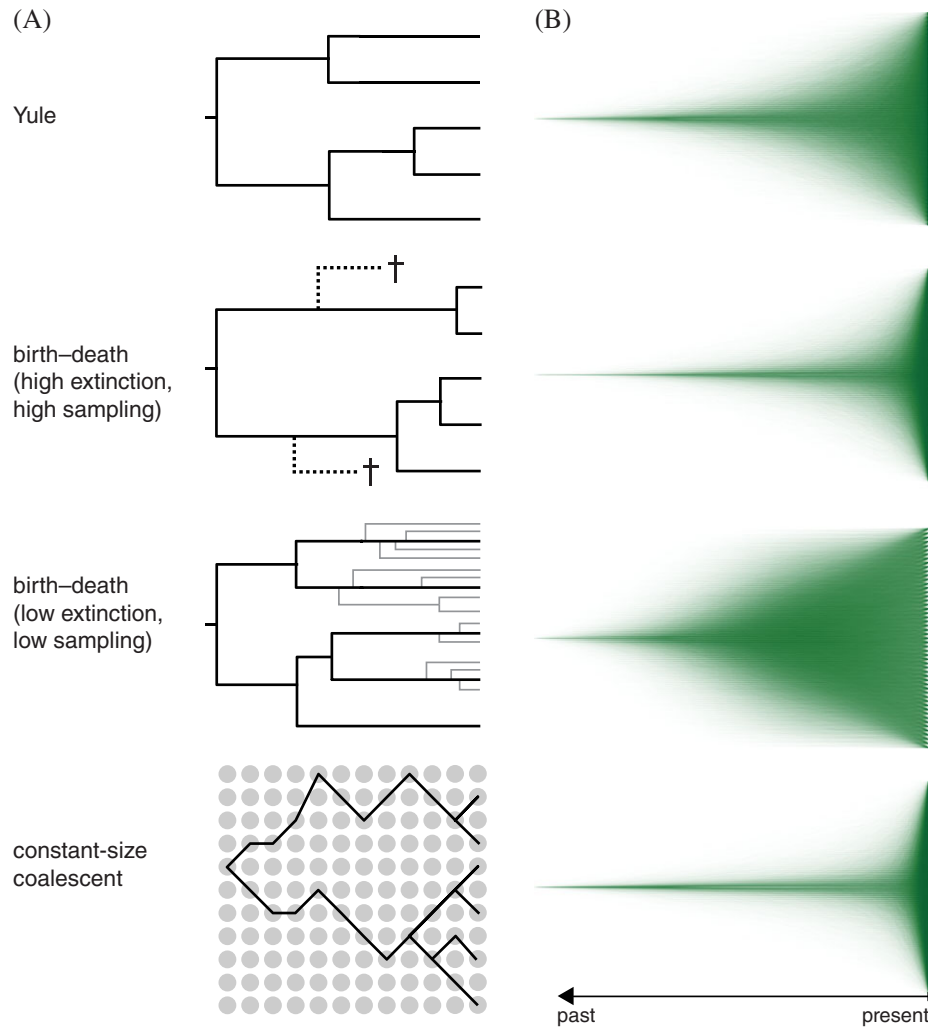
Fig. 4. Branch-length patterns produced by common models for generating tree priors. The patterns are visualized as both (A) schematic diagrams showing typical reconstructed branch-length patterns on a fixed tree topology; and (B) joint prior distributions of tree topologies and branch lengths produced by Markov chain Monte Carlo (MCMC) simulation in BEAST 2 and plotted using Densitree. Models shown include the Yule (pure-birth) process where observed species include all descendants of the common ancestor; the birth–death process with a high death rate, where older branches appear longer because of extinct lineages (dotted lines); the birth–death process with low taxon sampling, where terminal branches appear longer due to unsampled species (grey lines); and the constant-size coalescent, where the pattern of branch lengths is sensitive to the size of the population (grey circles).

backwards in time, allowing each sequence to connect to a parent in the previous generation. Coalescent priors typically assume that the data set represents a small sample of individuals from a larger population in which all individuals have the same chance of contributing offspring to the next generation, and that generations do not overlap (Fisher, 1930; Wright, 1931).

The shape of a genealogy under the coalescent is sensitive to variation in population size over time, because the number of individuals in the population in any given generation will influence the chance of sequences sampled from one generation sharing a parent in the previous generation. The simplest form of the coalescent assumes that the population maintains a constant size through time. Variations of the coalescent have been developed for scenarios in which the

population size varies according to a smooth function, such as exponential growth (Griffiths & Tavare, 1994), or in which the demographic history is divided into distinct periods with different population sizes (e.g. Drummond *et al.*, 2005; Minin, Bloomquist & Suchard, 2008). These are referred to as 'skyline plots' because the reconstructions of population size through discrete time periods resemble city skylines (Pybus, Rambaut & Harvey, 2000). However, this resemblance is lost when the skyline model is used in a Bayesian method, because the plot of population-size history is averaged over a large number of skyline reconstructions (Drummond *et al.*, 2005).

The influence of the coalescent tree prior on date estimates can be illustrated by studies where changing the prior changes the date estimates (Boskova, Bonhoeffer & Stadler, 2014).

For example, date estimates for the origin of an outbreak of hepatitis C virus obtained using a Bayesian skyline plot were approximately half those obtained using other coalescent tree priors, such as the constant-size, logistic-growth, or expansion-growth models (Golemba *et al.*, 2010). In this case, the researchers rejected the young dates produced using the Bayesian skyline plot on the basis of epidemiological evidence, but such independent information is not always available.

(4) Multi-locus tree priors

An assumption shared by most tree priors, but which is particularly problematic for coalescent tree priors, is that all sites in the alignment have evolved along a single bifurcating phylogeny. This assumption is violated under several well-known circumstances, such as when genetic material has been shared between lineages through horizontal gene transfer, or where a lineage is a hybrid of two or more divergent parental lineages (Funk & Omland, 2003). Loci also can have different histories due to recombination disconnecting parts of the genome from each other, such that the last common ancestor of one sampled locus might not coincide in time with the last common ancestor of other loci (Edwards & Beerli, 2000). Although tree priors are currently under development that allow for horizontal gene transfer and hybridization (Szöllősi *et al.*, 2015; Vaughan *et al.*, 2017), these are not yet widely implemented.

Current Bayesian software packages do have methods for resolving species trees from multiple loci that have incongruent gene trees. These involve a prior in which the histories of individual loci are assumed to have unfolded within the constraints of a species phylogeny (Rannala & Yang, 2003). The individual gene trees and the species tree can represent separate parameters and are inferred in the analysis, for example by jointly estimating them in *BEAST ('star beast') within the BEAST package (Heled & Drummond, 2010). The species-tree coalescence method in MrBayes formulates the distribution of the gene trees conditioned on the species tree as part of the prior, so the species tree is updated jointly with the gene trees (Ronquist *et al.*, 2012).

(5) Calibrations

On their own, the models used to generate the tree priors can only provide information about relative node times, for example whether more speciation events occur earlier or later in the history of the clade. In practice, the goal of most Bayesian dating analyses is to infer an absolute timescale of evolution. In some rare cases, independent estimates for the absolute rate of molecular evolution are available. For example, mutation rates for *Drosophila* calculated from laboratory breeding experiments have been used to date the diversification of endemic Hawaiian *Drosophila* lineages, resulting in much younger date estimates than those obtained using geological calibrations (Obbard *et al.*, 2012). Alternatively, previous estimates of substitution rates can be
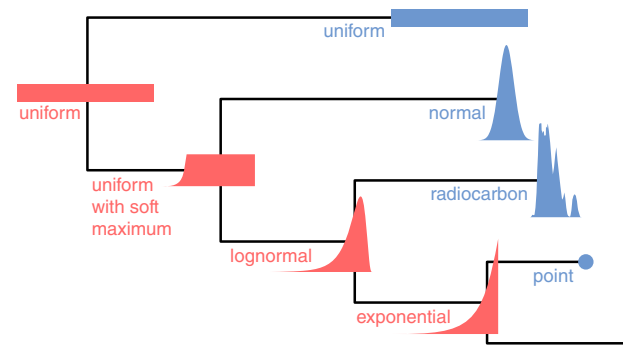

Fig. 5. Calibrations can be incorporated as probability densities for the ages of nodes in the tree. Calibration densities at internal nodes can be implemented as (top to bottom, red): uniform distribution; uniform distribution with hard minimum and soft maximum bounds; lognormal distribution; and exponential distribution. Calibration densities at terminal nodes can be implemented as (top to bottom, blue): uniform distribution; normal distribution; empirical calibrated radiocarbon distribution; and point value.

used to inform the prior distribution of branch rates (Hipsley & Müller, 2014; Nelson, Andersen & Brown, 2015; Marshall *et al.*, 2016).

In most Bayesian dating analyses, the absolute rate of change is estimated by including a temporal constraint on at least one node in the tree (Fig. 5). This is usually done either by fixing the age of the node or by specifying the prior distribution of its age (Thorne, Kishino & Painter, 1998; Drummond *et al.*, 2006; Yang & Rannala, 2006; Ho & Phillips, 2009). Although this is often referred to as the 'calibration prior', it is preferable to refer to this temporal information as a calibration density, because the time-calibrating information specified does not always strictly correspond to the prior distribution on the divergence time (Heled & Drummond, 2012). The fossilized birth–death tree prior described in Section V.2 provides a different approach to incorporating fossil evidence into the tree prior.

The use of calibrating information, such as palaeontological or biogeographic information, is subject to several sources of uncertainty (Ho & Phillips, 2009). First, there is the measurement uncertainty that is associated with the dating of fossils or geological events. For example, the uncertainty in geochemical dating, combined with that in modelling the growth of volcanoes and allowance for cooling to a habitable temperature, has given confidence limits on the ages of the Hawaiian islands (Obbard *et al.*, 2012). In analyses of molecular data from ancient samples or from rapidly evolving pathogens, the calibration densities can be chosen to reflect uncertainty in the age of the sequence itself (Shapiro *et al.*, 2011; Rieux & Balloux, 2016). For example, the uncertainty in radiocarbon dating associated with an ancient DNA sample can be taken into account when specifying the prior distributions for the ages of the terminal nodes (Molak *et al.*, 2015).

A second source of uncertainty in calibrations involves their assignment to nodes in the tree, given that the

taxonomic affinities of fossils might not be known with confidence (e.g. Lee, Oliver & Hutchinson, 2009). For example, changing the position of a key fossil to the basal node of a branch (stem) instead of the top of the same branch (crown) resulted in more than doubling of the age of lineages in phylogeny of Araucariaceae and related conifer lineages (Biffin, Hill & Lowe, 2010). One way of approaching this problem is to account for the uncertainty in fossil placement using a fossilized birth–death prior or by inferring the position of the fossil taxa in a total-evidence dating analysis (see Section V.2).

The third source of uncertainty arises from decisions about how the calibrating information relates to the age of the node. In some cases, the calibrating information is considered to provide a minimum age bound. For example, an identified fossil might be considered definitive evidence that a lineage existed at that time, so that its parent node must be at least as old as that fossil. More controversially, evidence of ancestral taxa is sometimes interpreted as placing a maximum date on a node by arguing that the descendent nodes could not have arisen before that time. Generally, absolute maximum age bounds are difficult to justify because of the limitations on proving absence of a lineage at a given time, and the potential for 'ancestors' and 'descendants' to co-exist. In practice, however, most dating analyses require specification of maximum ages either as absolute (hard) bounds or by defining a declining tail of probability on the age of a node (soft maximum). But maximum age constraints can have a strong impact on molecular date estimates, and their effects need to be evaluated carefully (Cracraft *et al.*, 2015; Mitchell, Cooper & Phillips, 2015).

The impact of maximum age bounds can be illustrated by considering molecular date estimates for the origin of animal phyla. While earlier molecular date estimates placed the origins of animal phyla long before the first undisputed animal fossils (see Bromham, 2006), more recent Bayesian date estimates have been considered more consistent with the timing of appearance of animals in the fossil record (e.g. Aris-Brosou & Yang, 2003; Peterson *et al.*, 2008; Erwin *et al.*, 2011; dos Reis *et al.*, 2015). But these estimates can be sensitive to the maximum age bounds that are specified in the analysis, which can overwhelm the signal from any other age constraints included in the analysis (Battistuzzi *et al.*, 2015). For divergence times to be compatible with the constraints imposed by these calibrations, the Bayesian analysis must infer date estimates that do not exceed the maximum age constraints, resulting in very high rates of change along the earliest lineages of the metazoan phylogeny (Lee, Soubrier & Edgecombe, 2013). In cases such as these, the Bayesian date estimates are predominantly reflecting the prior beliefs about divergence times, rather than information from the sequence data.

An alternative approach to setting a hard maximum or minimum bound on the date of a node is to describe the possible age of a calibrated node using a probability density (Fig. 5). For example, many researchers use a lognormal distribution of ages on the calibrating node, with the minimum value of the distribution defined by fossil evidence, the mean at the expected age of the node, and with a tail that describes the decreasing probability for increasingly older dates (Ho & Phillips, 2009). The lognormal distribution is described in terms of two parameters, the mean rate and its standard deviation. The exponential distribution is described by a single parameter from which both the mean rate and the standard deviation are derived. In these cases, there is a 'soft' maximum implicit in the prior distributions, because values at the tails of the distribution are given a low probability. However, the soft maximum does not exclude the possibility of older dates, which means that it is possible for the data to overwhelm the specified maximum (e.g. Sanders & Lee, 2007; Jarvis *et al.*, 2014).

Given the uncertainty in the association between calibrating information and the age of a node in the phylogeny, calibrations must be regarded as hypotheses about evolutionary history, rather than as error-free observations. It is therefore appropriate that calibration uncertainty is modelled as part of the Bayesian dating procedure. One way to do this is to use a model of fossil preservation and discovery to develop appropriate probability densities for calibrated node ages (Wilkinson *et al.*, 2011; Heath *et al.*, 2014). Alternatively, the parameters of the calibration densities can be assigned hyperparameters that are estimated during the analysis (Heath, Holder & Huelsenbeck, 2012).

In most Bayesian dating methods, the chosen calibration densities do not act in isolation simply to provide a timescale for the nodes. Calibration densities on nested nodes define hierarchical structure in the relationships among the sampled taxa, making some tree topologies less probable. So placing information on the ages of some nodes in the tree can have wider effects on phylogenetic inference (Ho & Phillips, 2009). Calibrations at internal nodes tend to have their strongest influence on estimates of rates along the descendent lineages (Duchêne, Lanfear & Ho, 2014).

Therefore, although calibrations are usually specified separately for particular nodes in the tree, the calibration densities interact with the tree prior, yielding a joint prior for the age of every node in the tree (Heled & Drummond, 2012). Furthermore, when there is temporal overlap along a branch between a pair of calibrations, each of the densities has to be truncated so that the ancestral nodes are older than the descendent nodes (Rannala, 2016). The practical upshot of this is that the joint prior distribution for the age of each calibrating node might actually end up being different from the probability density specified by the user (Warnock, Yang & Donoghue, 2012). To check for interaction between the calibration densities and tree priors, the user-specified densities and joint prior distributions of the node times should be compared in order to identify any discrepancies (Duchêne *et al.*, 2014; Warnock *et al.*, 2015; Brown & Smith, 2017). However, the problem of interaction between multiple calibration densities is more difficult to address (Heled & Drummond, 2012) and it is not always straightforward to ensure that the joint priors match the user-specified calibration priors (dos Reis, 2016; Rannala, 2016).

Adding more calibrations throughout the phylogeny should generally help to inform the estimation of model parameters, particularly those relating to rate variation across the tree (Duchêne *et al.*, 2014; Phillips, 2015). However, increasing the number of calibrations by accepting poorly constrained fossil evidence or divergence dates with questionable placement in the phylogeny might have a negative impact on accuracy (Parham *et al.*, 2011). Not surprisingly, using 'early but risky' calibrations, derived from uncertain fossil data, can result in much older date estimates than a conservative approach of using only 'late but safe' calibrations based on fossils that have been dated and identified with confidence (Sauquet *et al.*, 2012). For example, inclusion of calibrations based on microfossil data, such as pollen records, will tend to extend the basal ages of plant phylogenies by providing older calibrations than using only rarer macrofossils (Thornhill *et al.*, 2012; Christin *et al.*, 2013).

VI. PRIORS ON THE BRANCH RATES

(1) Branch-rates priors

The popularity of Bayesian dating methods has, to a large extent, been borne on their capacity to incorporate various models of rate variation across branches (dos Reis, Donoghue & Yang, 2016). There are a plethora of available priors on branch rates, usually referred to as 'clock models', and many of them have been implemented in widely used software for Bayesian phylogenetics (Heath & Moore, 2014; Ho & Duchêne, 2014). Here we will mention some key characteristics of these priors rather than aim for an exhaustive catalogue.

One way to differentiate some of the most common branch-rates models is to compare the number of different rates that are allowed (k) with the number of branches in the phylogeny (n) (Fig. 6). In a constant-rate model ('strict clock'), all branches share a single rate ($k = 1$). In a discrete-rates model (including 'local clocks'), branches are assigned rates from a number of rate categories ($1 < k < n$). In a 'relaxed clock' model, every branch can have its own rate ($k = n$). Another key difference between branch-rates models is the relationship between branches, nodes, and rate changes. Some branch-rates models assume that rate changes happen only at nodes, regardless of the length or duration of the branch. Other models allow the rate to change along a branch, for example considering the probability of rate change as dependent on the branch duration (Guindon, 2013; Huelsenbeck, Larget & Swofford, 2000; Lepage *et al.*, 2006).

(2) Constant rates ('strict clock')

The simplest branch-rates prior is a constant-rate model (also known as a 'strict clock'), which allows only one substitution rate across the whole phylogeny. If rates vary
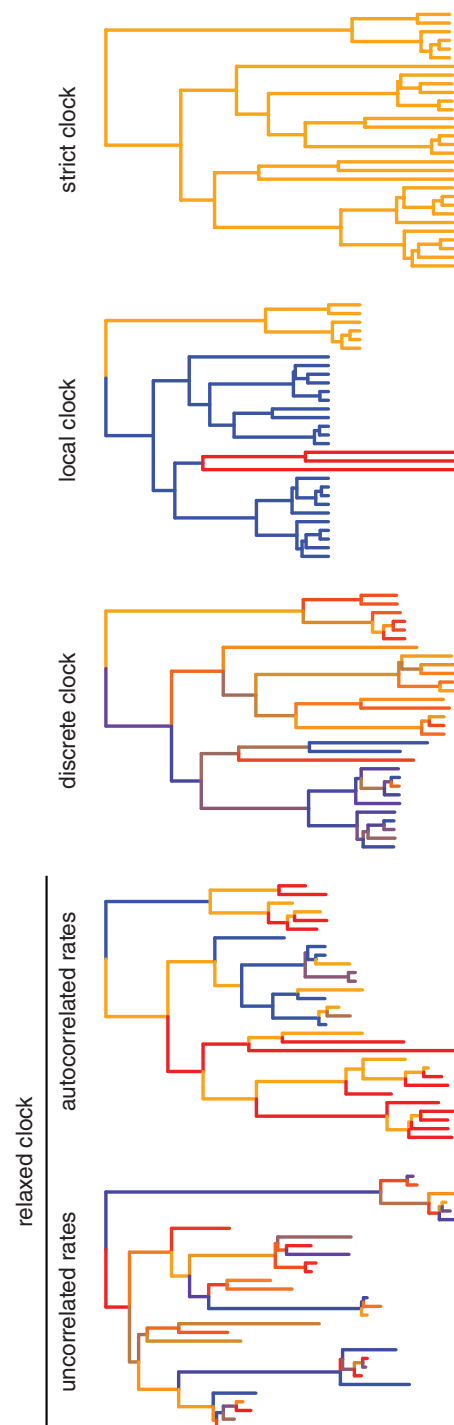

Fig. 6. Branch-rates priors differ in the number of distinct branch rates and in the pattern of rate variation. A strict clock assumes a constant rate across all branches. A local-clock model assumes a small number of distinct branch rates, clustered across the tree. A discrete-clock model also assumes a small number of distinct branch rates, but without clustering across the tree. A relaxed-clock model allows all of the branch rates to be distinct. These branch rates can either be correlated between branches (autocorrelated relaxed clock) or independently and identically distributed (uncorrelated relaxed clock).

stochastically and the variance between branches is not large, then the constant-rate model can produce date estimates with comparable accuracy and superior precision to those obtained using variable-rates models (Brown & Yang, 2011). However, constant-rate models have poor accuracy in the presence of non-negligible rate variation (Wertheim *et al.*, 2009). Given the number and variety of biological traits and environmental factors that have been implicated as influences on the substitution rate, we should be wary of assuming that all of the sampled lineages in a data set share the same underlying rate of change (Welch & Bromham, 2005).

Rate constancy is sometimes assumed for 'shallow' phylogenies where the root divergence is relatively recent. However, the forces that shape differences in the rate of molecular evolution can vary not only between closely related species, but even between conspecific populations or individuals. For example, mutation rate can be influenced by differences in average age (Kong *et al.*, 2012), growth rates (Lanfear *et al.*, 2013), physiological stress (Matsuba *et al.*, 2013), physical condition (Sharp & Agrawal, 2012), or nutritional status (Agrawal & Wang, 2008). Therefore, rate constancy should never be assumed for any data set, but should always be explicitly tested.

(3) Discrete rates ('local clocks')

Some methods allow variation in rates across the tree, but assume that some of the branches will share the same rate, such that there are fewer rate categories than there are branches in the tree (Fig. 6). There is a range of approaches for assigning branches to rate categories, and there is no single accepted terminology for referring to this set of models, but they have been variously described as local clocks, discrete-rate models, or multi-rate molecular clocks (Ho & Duchêne, 2014). The branches within a rate category might be found in different parts of the phylogeny or they may be clustered together.

Branches might be assigned to rate categories based on some independent information about the expected patterns of rate variation (Yoder & Yang, 2000), for example by defining virus lineages according to their host and allowing one rate for each host type (Worobey, Han & Rambaut, 2014). But this *ad hoc* assignment of rate categories has been rare in Bayesian molecular dating, even in cases where there is good evidence for distinct shifts in substitution rate. Instead, most researchers have favoured a stochastic approach to assigning branches to rate categories. For example, a random local clock involves testing whether each branch inherits the rate from its parent branch (Drummond & Suchard, 2010). The two key parameters in this model are the number of rate changes along the tree, which is described using a Poisson prior distribution, and the relative values of the substitution rates, drawn from a gamma prior distribution. Because each branch will inherit the rate from its parent branch unless there is evidence for a change, the random local clock will tend to favour fewer rate changes (Drummond & Suchard,

2010). The random local clock collapses to a constant-rate model (strict clock) if the number of rate changes is zero.

Some discrete-rates models require a parameter that informs the number of rate categories. For example, the number of rate categories can be determined using a Dirichlet process prior (DPP), which assigns branches to rates drawn from distinct rate categories. Using this approach, the number of rate categories is treated as a random variable, as are the value of the rate for each category and the assignment of branches to rate categories (Heath *et al.*, 2012).

(4) Relaxed molecular clocks

The model of rate variation can be relaxed further by allowing every branch in the phylogeny to have a different rate, which is the defining feature of relaxed clocks (Fig. 6). The rates can be assumed to be independently and identically distributed, as in uncorrelated relaxed clocks (e.g. Drummond *et al.*, 2006), or related between neighbouring branches, as in autocorrelated relaxed clocks (e.g. Thorne *et al.*, 1998).

In uncorrelated models of rate variation, every branch is assigned a rate independently of other branches (Drummond *et al.*, 2006; Rannala & Yang, 2007). In this case, there is no relationship between the rates along neighbouring branches, so that a branch with a low rate is no more likely to give rise to a descendent branch with a low rate than a branch with a high rate, or *vice versa*. Most uncorrelated relaxed clocks involve each branch rate being drawn independently from a parametric distribution, such as a lognormal distribution or exponential distribution (Drummond *et al.*, 2006; Rannala & Yang, 2007). There are surprisingly few empirical studies of observed patterns of rate variation across branches in a phylogeny, independent of Bayesian estimation procedures (Bininda-Emonds, 2007; Dornburg *et al.*, 2012; Phillips, 2015). As a consequence, the form of the distribution of rate values is usually chosen for statistical convenience, rather than being based on any biological understanding of evolutionary rates.

But many aspects of biology that influence rates of molecular evolution, such as life history, environment, or population structure, are likely to be more similar between related lineages. So we might expect rates to evolve along a phylogeny, just as other species traits do (Gillespie, 1984; Thorne *et al.*, 1998; Bromham, 2011). Autocorrelated relaxed clocks assume that there will be some connection between the rates along neighbouring branches. For example, in the autocorrelated exponential relaxed clock, the rate can change at each node, with the rate for the daughter branch drawn from an exponential distribution with a mean equal to the rate along the parent branch (Aris-Brosou & Yang, 2002). This provides an episodic model of rate change, where the rates are related across branches, but the degree of change is independent of the duration of each branch.

In some other autocorrelated relaxed clocks, rate changes can occur not only at nodes, but also along branches. For example, the compound Poisson process makes the assumption that branch rate can vary as a Poisson process,

so that rates can change along branches, with longer branches more likely to experience multiple rate-change events (Huelsenbeck *et al.*, 2000). Each new rate is based on the ancestral rate multiplied by a variable drawn from an appropriate distribution. The compound Poisson process can be described using three parameters, each with an appropriate prior distribution: the ancestral rate; the number of expected rate changes per unit of branch length, which usually takes a fixed value or an exponential hyperprior; and the standard deviation of the rate multipliers (the variable by which the ancestral rate is multiplied).

Autocorrelated relaxed clocks can employ continuous models of rate change, for example where rates change by Brownian motion along the tree. In these continuous models, the rate along each branch is sampled from a distribution with a mean equal to the rate along the parent branch, but the variance is proportional to the branch length. Therefore, longer branches are more likely to undergo a greater degree of rate change (Thorne *et al.*, 1998; Kishino *et al.*, 2001). These models have a variance parameter that describes the way in which rates vary from one branch to the next. If the value of the variance parameter is small, then the rate along the daughter branch is very similar to the rate along the parent branch (Thorne *et al.*, 1998). If the variance parameter is zero, the daughter rate is always the same as the parent rate and the model reduces to the constant branch-rates model (strict clock).

All of the autocorrelated relaxed-clock models assume that rate variation is stochastic, such that rate wanders up and down. One potential problem with these branch-rates models is that the random fluctuations in rate can allow substitution rates to 'wander off' into extreme values. The Cox–Ingersoll–Ross model has been proposed as a solution to this problem (Lepage *et al.*, 2006; Lepage *et al.*, 2007). In this branch-rates model, substitution rate varies continuously and stochastically over time, but there is a spring-like constraint on rate variation that draws rate values back towards a mean value. The force of the 'spring' that draws rates back is proportional to the difference between the rate on the parent branch and the mean rate of the distribution. This gives the Cox–Ingersoll–Ross model an inbuilt reluctance to allow extremely high or low rates.

A second potential limitation of stochastic models of rate variation is that all of the common models assume random changes in rate, but there are many cases in which we might expect rates to undergo directional change. For example, rates of molecular evolution are associated with body size in mammals (Welch, Bininda-Emonds & Bromham, 2008), and most placental mammal lineages increased in average size after the Cretaceous–Paleogene boundary (Slater, 2013). Therefore, we might expect a concerted slowdown in substitution rates in many mammalian lineages in the part of the phylogeny corresponding to the period of radiation into larger body sizes (Bromham, 2003; Springer *et al.*, 2003). Such concerted changes would be hard to detect using Bayesian molecular dating methods, and could result in biased date estimates (Lee & Ho, 2016).

(5) Partitioned models of branch rates

We expect that some parts of the genome will change more quickly than others, so in analyses of multi-locus data sets, each locus can be assigned a relative-rate parameter to account for differences in evolutionary rates across loci. However, such a model will not capture differences between loci in their patterns of rate variation across branches. For example, if loci have been subject to differing degrees of selection, they might show different patterns and rates of substitutions across the branches of the phylogeny (Gillespie, 1991). For example, the gene for prestin, a protein motor associated with detection of vibration by hair cells, has undergone an accelerated rate of evolution in the echo-locating lineages of bats and cetaceans (Liu *et al.*, 2010). Under these circumstances, multiple models of branch rates are required to account for the heterogeneity in branch-rate patterns across loci (Duchêne & Ho, 2014).

Each relaxed-clock model allows a different rate for each branch in the tree, so having multiple relaxed clock models brings a considerable risk of overparameterization. One way to mitigate this risk is to use a statistical approach to choose the number of clock models. For example, models of branch rates can be assigned to loci according to inferred patterns of among-lineage rate variation (Duchêne & Ho, 2014). Various methods have been developed for identifying the number of distinct patterns of among-lineage rate variation in multi-locus data sets (e.g. dos Reis *et al.*, 2012; Snir, 2014; Duchêne, Foster & Ho, 2016), but these generally involve the assumption that all loci share the same tree topology. Branch-rate models can be assigned to subsets of the data using other criteria. For example, the data set can be partitioned according to relative evolutionary rate (dos Reis *et al.*, 2012) or by codon position (Ho & Lanfear, 2010). Applying a partitioning scheme can lead to improvements in the precision of date estimates (Zhu, dos Reis & Yang, 2014). However, some studies have shown that the date estimates are relatively robust to the choice of data-partitioning method, unless there is severe model misspecification (Angelis *et al.*, 2017; Foster & Ho, 2017).

(6) Modelling branch-rate variation

The choice of branch-rates prior can have a substantial impact on estimates of divergence times and evolutionary rates (e.g. Drummond *et al.*, 2006; Lepage *et al.*, 2006; Linder, Britton & Sennblad, 2011; Crisp *et al.*, 2014). Many studies compare the results of applying several branch-rates models to their data, and it is not unusual for these different models to give very different date estimates (e.g. Berbee & Taylor, 2010; Magallón, 2010; Papadopoulou, Anastasiou & Vogler, 2010; Peterson *et al.*, 2008; Bellot & Renner, 2014; Crisp *et al.*, 2014).

Even in the relatively rare cases where there is independent knowledge of patterns of rate variation, Bayesian dating methods using different branch-rates models can produce widely different results. For example, parasitic plants typically
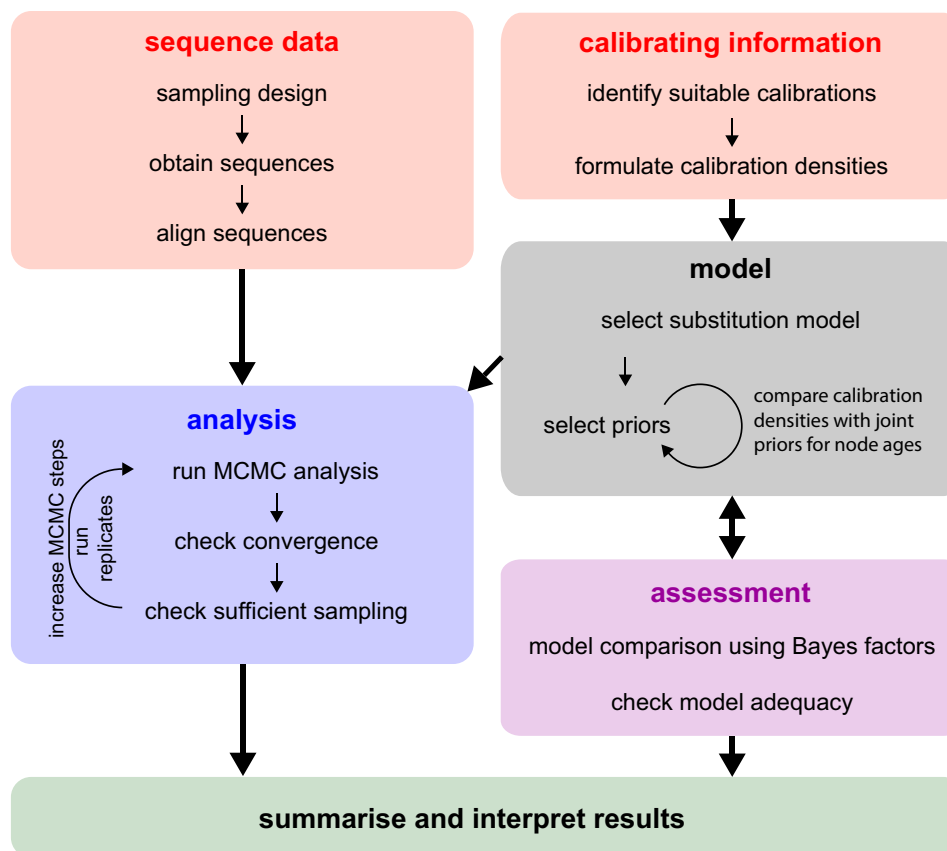
Fig. 7. Workflow of a typical Bayesian dating analysis using fossil-based calibrations. A typical workflow for a Bayesian dating analysis starts with collecting sequence data and identifying suitable calibrations. The workflow does not show the choice of a tree model, because almost all users assume a rooted binary tree in Bayesian dating analyses. Following the setting of a substitution model, the user selects the priors for the parameters of the substitution model, the tree prior, and branch-rates prior, and the node-time priors. The node-time priors are informed by the calibrating information, which can come from sources such as the fossil record or from biogeography. Each calibration density specified by the user should be compared with the joint prior for the age of the corresponding node. The sequence data and the model are then used to calculate the posterior probability, which is done by Markov chain Monte Carlo (MCMC) analysis. The performance of the MCMC analysis needs to be monitored by checking for convergence to the target distribution and for sufficient sampling. Extending the MCMC analysis or running replicates might be required. Various assessment steps should also be taken, including model comparison using Bayes factors and testing of model adequacy.

have higher substitution rates than their non-parasitic relatives (Young & dePamphilis, 2005; Bromham, Cowman & Lanfear, 2013). Encouragingly, Bayesian dating methods can detect these higher rates in parasitic plant lineages (Bellot & Renner, 2014). But the inferred distribution of rates among branches varied between models: the random local clock gave high rates within the parasitic clade (therefore younger dates), whereas uncorrelated relaxed clocks placed high rates on the stem (ancestral) lineage (therefore older dates). So even when we have reliable empirical knowledge about rate variation, we can still find ourselves without a sound basis for choosing which branch-rates model to use. There is ongoing debate about whether uncorrelated or autocorrelated relaxed clocks provide a better fit to real molecular data, but this is likely to vary from one data set to another (Lepage *et al.*, 2007; Ho, 2009; Ho, Duchêne & Duchêne, 2015; Lartillot, Phillips & Ronquist, 2016).

Relaxed clocks have led to more relaxed attitudes about molecular dating, because they allow researchers to have their cake and eat it too: recognize that rate variation is commonplace, yet still use molecular data to estimate evolutionary timescales. If different clock models yield different date estimates when applied to the same data, then we know that the disparities reflect the assumptions that we are making. Therefore, the predictive power of the analysis depends on the degree to which the chosen model provides an appropriate description of the data. If the assumptions of the method are violated by the underlying data, then there is as much potential for rate-variable Bayesian date estimates to be wrong as there is for strict-clock estimates to be wrong when the assumption of rate uniformity is violated. It is therefore critical that the sensitivity of the date estimates to the models and priors is carefully evaluated (Fig. 7).

One way to check if the model of rate change is realistic is to ask whether the inferred substitution rates in any part of the tree fall within the likely range of substitution rates for those lineages. Empirically determined rates of molecular evolution can provide a guide to the plausibility of rate estimates; a careful evaluation of the data or the assumptions is warranted whenever estimates are an order of magnitude higher or lower than any previously reported (Bromham & Hendy, 2000; Lee *et al.*, 2013; Springer *et al.*, 2013; Phillips, 2015). For example, surprisingly young estimates of the date of origin of the endangered Devil's Hole pupfish (Martin *et al.*, 2016) have been questioned on the grounds that the assumed mutation rate was approximately 20- to 50-fold greater than the typical vertebrate mutation rate (Sağlam, Baumsteiger & Miller, 2017). Although unusually high or low rates in a previously unexamined lineage cannot be ruled out, they do suggest the need for robust interrogation of the resulting date estimates.

VII. CHOOSING THE RIGHT MODEL

(1) Assessing model fit

Choices of models and priors can affect the answers produced by Bayesian dating methods, so we need a reliable means of selecting the most appropriate models to use in our analysis. Where we are uncertain about the right model, we need a way of reporting results over a range of equally plausible models.

Bayesian model fit is often measured by the marginal likelihood, which measures the probability of the data given the model, averaged across the prior distribution of the parameters in the model. Models can be compared by taking the ratio of their marginal likelihoods, which is referred to as the Bayes factor. The Bayes factor is often used as an indication of which model, averaged across the space of parameter values, is more likely to have generated the data (Kass & Raftery, 1995). A Bayes factor greater than 1 indicates that the data support the numerator model more than the denominator model. A number of guidelines have been proposed for interpreting the level of support for the numerator model. For example, support may be considered 'strong' when the Bayes factor is greater than 10 (Jeffreys, 1961) or 20 (Kass & Raftery, 1995). In phylogenetics, Bayes factors are used to compare components of the likelihood model, including the substitution model (Brandley, Schmitz & Reeder, 2005) and the tree model (Bergsten, Nilsson & Ronquist, 2013), as well the priors on trees and branch rates (Baele *et al.*, 2012).

The use of Bayes factors requires computation of the marginal likelihood, but this this would require us to integrate over all parameter values for the model, and this is not analytically possible except in some very trivial cases. So instead we have to approximate the marginal likelihood by sampling parameter values using MCMC (Sinsheimer, Lake & Little, 1996; Suchard, Weiss & Sinsheimer, 2001; Xie

et al., 2011). The basic idea is to take a random sample of possible parameter values, calculate the likelihood given those values, and use these likelihoods to approximate the marginal likelihood. MCMC methods differ in the distribution from which samples are drawn and how the estimate is calculated as a result. The simplest methods draw samples from a single distribution over parameter values, which can be done with a single MCMC run. These include harmonic-mean methods, which sample from the posterior (Newton & Raftery, 1994), and arithmetic-mean methods, which sample from the prior. The harmonic-mean estimate is easy to compute from an existing posterior sample; however, it can have an unacceptably high variance and tends to overestimate the marginal likelihood, making model selection unreliable (Lartillot & Philippe, 2006; Baele, Lemey & Suchard, 2016).

A more reliable approach to approximating the marginal likelihood is to draw samples from multiple distributions arranged along a path going from the prior to the posterior. This method is used in the path-sampling estimator (Lartillot & Philippe, 2006; Lepage *et al.*, 2007) and the stepping-stone sampling estimator (Xie *et al.*, 2011). These path-sampling methods typically consider a set of power posteriors, each of which is the prior multiplied by the likelihood raised to a power ranging from zero to 1. Samples are drawn from the prior when the power is zero, from the posterior when it is 1, and from a range of intermediate distributions in between. These path-sampling methods have been shown to select the correct model more frequently than methods based on the harmonic or arithmetic mean (Fan *et al.*, 2011). However, they also require multiple specialized MCMC runs and are therefore more computationally expensive.

It is important to bear in mind the limitations of using Bayes factors to select models for molecular dating analysis. This approach compares two models, including the specification of the priors, so a poor choice of priors can also influence the Bayes factor. For example, if the Bayes factor is used to compare two branch-rates models, the fit of the data is reflected not only in the form of the changes in rates over the tree (such as uncorrelated *versus* autocorrelated rate change), but also the specification of the prior distributions of substitution rates and the rate-change parameters. So the Bayes factor cannot in itself reveal why a particular model has a better fit to the data than another.

(2) Model averaging

In some cases, we might find that several different models all have similar marginal likelihoods, meaning that the approximate Bayes factor gives us little reason to choose between competing models. What should we do in such cases? Common practice is to accept the model with the fewest parameters, but sometimes the competing models have identical numbers of parameters. A potential solution is to report the range of date estimates across all equally acceptable models. This can be done informally, by rerunning the analysis using each of the candidate models. Alternatively, it can be carried out formally using a

model-averaging approach, which allows the data to be evaluated by a weighted average over a set of models, rather than one model at a time (Li & Drummond, 2012). This represents a fully Bayesian treatment of the problem of model uncertainty.

We can think of all the variations on the hierarchical model as occupying different positions in the space of all possible models. The Bayesian analysis can explore and compare different evolutionary models in essentially the same way that it explores and compares different phylogenies and parameter values, using stochastic sampling of the model space with MCMC (Huelsenbeck, Larget & Alfaro, 2004). In addition to being able to propose a new parameter value, tree topology or branch lengths at each step of the chain, the MCMC can propose a different model at each step. In this way, the resulting date estimates are weighted by the posterior probabilities of the models considered.

Although model averaging appears to present a healthy approach to uncertainty, its practical benefits in Bayesian dating remain unclear. There are cases in which model averaging results in a better statistical fit to the data, for example by averaging over possible substitution models and data subsets (Wu, Suchard & Drummond, 2013), or selecting an appropriate branch-rates model (Li & Drummond, 2012). However, there are also examples from both empirical and simulation studies in which model averaging did not improve phylogenetic outcomes. For example, model averaging did not increase the chances of selecting the correct topology for a data set of mammalian protein-coding sequences when compared with the two best-fitting models (Li & Drummond, 2012), and model averaging did not always correctly identify the branch-rates model used to simulate the evolution of synthetic sequence data (Duchêne & Ho, 2014).

(3) Model adequacy

Choosing models based on either model fit or model averaging can indicate the best solution given the available models, but it does not guarantee that the selected model provides an adequate description of the evolutionary process that generated the data (Huelsenbeck *et al.*, 2001; Bollback, 2002; Gelman *et al.*, 2014). We need to face the sad fact that even the best-fitting available model might not be good enough to make sound evolutionary inferences. Therefore, it is desirable to assess the adequacy of each candidate model by asking whether the data could have been generated by the chosen models.

One method for assessing model adequacy in Bayesian molecular dating is the use of posterior predictive simulations (Bollback, 2002; Duchêne *et al.*, 2015). The 'predictive' part of this label refers to the general principle that if the model and parameter values from the posterior distribution are a good description of the data, then we ought to be able to use them to predict outcomes when we 'run evolution again' using simulations. We can simulate evolution of DNA sequences using the best-fitting model and parameter values to see if we would get a similar pattern to our real data. Clearly, we cannot expect exactly the same tree

and sequences to be produced by simulation, so instead of comparing the actual phylogenies, we need some test statistic that allows us to characterize the phylogeny as a whole. Using an appropriate summary statistic to describe each phylogeny, we can ask whether the characteristics of our observed data are similar to those of data simulated using the same model and analysed in the same way (Rubin, 1984). A number of test statistics have been proposed to assess the components of the model in Bayesian dating. Some of these focus on the overall fit of the substitution model (Goldman, 1993; Huelsenbeck *et al.*, 2001; Bollback, 2002; Foster, 2004), while others examine the adequacy of the hierarchical model (Duchêne *et al.*, 2015). Potentially, statistics based on tree topology and branch lengths can be used as test statistics for evaluating the adequacy of the tree model (Blum & François, 2005; Hua & Bromham, 2016).

VIII. SENSITIVITY TO THE PRIORS

The priors play a critical role in Bayesian dating analyses and, given that they can have a large impact on the date estimates obtained, it is important to examine them carefully. The examples we have referred to show that the choice of tree prior, branch-rates prior, and other priors can have important effects on the estimates of the substitution rate and divergence times. Furthermore, the choice of priors can influence Bayesian model selection using marginal likelihoods, because these are weighted by the priors. So for any Bayesian molecular dating analysis we should investigate the impact of the priors on the outcomes of the analysis. We can do this by comparing the relative impact of the priors and the data (likelihood) on the posterior distribution (Gelman *et al.*, 2014).

One way to test the influence of the priors is to remove the impact of the likelihood by running the molecular dating analysis without any sequence data (Drummond *et al.*, 2006). As the MCMC chain draws samples of trees and parameters, there are no sequence data to evaluate the likelihood, so that we are sampling from the prior distribution. If this provides a close match to the posterior distribution that we obtain from an analysis of the data, then this suggests that the priors have had a very strong influence on the parameter estimates and that the sequence data have not changed the picture. We suggest that it is good practice to investigate different combinations of tree priors and calibration priors. Although sampling from the prior is becoming a common practice in molecular dating (e.g. Dornburg *et al.*, 2012; Christin *et al.*, 2013; Crisp *et al.*, 2014; Brown & Smith, 2017), there are as yet no established criteria for evaluating how different the prior and posterior should be in order to have confidence that the date estimates are influenced by the data and not simply a reflection of the priors.

Another way to test the influence of the prior on the date estimates is to compare the relative impact of the choice of priors on the posterior (Gelman *et al.*, 2014). For example, the effect of using different tree priors or branch-rates priors

can be examined. The greater the change in the posterior, the more sensitive is the posterior to the prior (Brown & Yang, 2010; Inoue, Donoghue & Yang, 2010; Stadler & Yang, 2013; dos Reis, Zhu & Yang, 2014; Warnock *et al.*, 2015). However, this test of the influence of the prior cannot necessarily be applied to all parameters. Parameters of this kind include the branch rates and node ages, which are semi-identifiable because there is an effectively infinite number of combinations of these that can give the same product (Rannala, 2002). Given this identifiability property, any information that is used to calibrate the molecular clock is always expected to be reflected in the posterior.

What can be done if the signal from the priors is found to speak more loudly than the signal from the data? An intuitive way to reduce the sensitivity to the priors is to increase the amount of data. However, in the case of rate estimates and divergence times, the uncertainties in the posterior and the impact of the prior will not disappear even with an infinite amount of sequence data (Rannala & Yang, 2007; Zhu, dos Reis & Yang, 2014). Irrespective of the amount of data, uncertainty in the posterior distribution of node ages is influenced by uncertainty in fossil calibrations (Yang & Rannala, 2006; dos Reis & Yang, 2013). It is worth emphasizing that the priors on node ages are among the few priors where we can gain real leverage from independently derived information (in true Bayesian fashion). Since correct calibrations can alleviate the impact of a misspecified rate prior, calibration densities have a particularly large influence on the posterior distribution of node ages (Rannala & Yang, 2007; Inoue *et al.*, 2010; Zhu *et al.*, 2014). Careful assessment of calibration densities is crucial to any molecular dating analysis.

In addition to careful evaluation of the reliability of palaeontological or biogeographic evidence used in molecular dating, assessment of calibration densities should include the two checks described above: sampling from the prior, and changing the priors. Running the dating analysis without sequence data allows us to check if the joint prior for the age on a calibrated node differs from the calibration densities that we initially specified (Warnock *et al.*, 2012; Christin *et al.*, 2013; Joyce *et al.*, 2013; Warnock *et al.*, 2015). When there is mismatch between the joint priors and the user-specified densities for calibration, Warnock *et al.* (2015) suggest reparameterizing the chosen densities until all pieces of independent calibrating information are reflected in the effective prior.

The second check for assessing calibrations is to test how sensitive the posterior distribution of node ages is to the calibration densities. This can be done by conducting the molecular dating analysis using only a subset of the calibrations. The posterior age estimates can then be compared with the calibration densities that had been omitted in the analysis (Near, Bolnick & Wainwright, 2005; Marshall, 2008; Dornburg *et al.*, 2011). However, the results of such cross-validation tests are difficult to interpret, because of the way that priors are combined in the analysis with other calibrating information and with the tree prior (see Section V.5).

Therefore, consistency in posterior estimates of node times should not be used as the only criterion to select calibrations.

IX. GETTING AN ANSWER

Rather than producing point estimates for node ages, a Bayesian molecular dating analysis produces a posterior distribution that is the combined information from the data, the likelihood model, and the priors (Fig. 1). But obtaining the posterior distribution can be a complex and difficult process. In order to use a set of samples from the MCMC to represent the posterior distribution, we need to solve four problems. First, we need to discard the initial portion of the chain where it drew samples that were poorly representative of the posterior distribution (burn-in). Second, we need to be able to confirm that the chain is sampling from the correct distribution (convergence). Third, we need to decide whether a sufficient number of samples has been drawn to allow a reliable estimate of the posterior distribution (effective sample size). Fourth, we need a meaningful way of summarizing the information in the samples from the posterior distribution (summary statistics).

In order to remove the initial samples that represent low-probability portions of the posterior distribution (burn-in), we need to identify the point at which the MCMC simulation is correctly drawing samples in proportion to their posterior probability. This is known as sampling from the stationary (or target) distribution. In practice, however, it is not always clear when the Markov chain is actually sampling from this target distribution. One approach is to plot the posterior probability (known as a trace), and discard the part of the chain where the posterior probability is still climbing upwards (Nylander *et al.*, 2008). This can be done either using a subjective judgement or using more formal convergence diagnostics (Brooks & Gelman, 1998).

Determining whether the MCMC has converged on the stationary distribution is a major practical and theoretical problem (Huelsenbeck *et al.*, 2002). Since the parameter space is effectively infinite, we cannot guarantee that the MCMC is sampling appropriately from the full range of the posterior distribution. Independent replicates of the MCMC analysis starting from different states could initially sample from different parts of the posterior distribution, which would lead to different date estimates if the analyses were stopped too early. If separate MCMC chains converge to the same distribution, this is good evidence that they are sampling from the stationary distribution (Gelman & Rubin, 1992). For this purpose, checking for convergence should be done with at least two chains, preferably more. The difference in outcome between chains can be formally compared using the potential scale reduction factor (PSRF), which compares the variation of each parameter between different runs with its average variability within runs (Gelman & Rubin, 1996). This statistic should approach 1 as chains converge on a similar result.

Checking for MCMC convergence is quite straightforward for continuous parameters of the hierarchical model, but less so for the tree itself. There are several diagnostic methods that have been implemented in the program AWTY (Nylander *et al.*, 2008; Warren, Geneva & Lanfear, 2017), and these are most effective when comparing tree samples from multiple independent MCMC analyses. Another way to check for convergence in tree topology is to run two independent MCMC simulations and report the average standard deviation of split frequencies (ASDSF) between the trees sampled in each (Ronquist *et al.*, 2012). Although this does not directly assess the difference in date estimates, high values of the ASDSF statistic would indicate dissimilarity between the two sets of tree samples. In turn, this would suggest that date estimates might also differ between chains, raising doubts about convergence to the stationary distribution.

Once we have satisfied ourselves that the MCMC has been sampling from the target distribution, we need to ensure that we have drawn enough samples to allow us to characterize the posterior distribution. We have to bear in mind that successive states of the Markov chain are autocorrelated – that is, samples drawn from close to each other in the chain are likely to be more similar to each other than samples drawn randomly from the posterior distribution. As a consequence, we would typically need to run the MCMC analysis for millions of steps to obtain enough samples to allow us to estimate the posterior distribution reliably. We can reduce the amount of disk space required to store the samples from the MCMC analysis if we only sample the chain occasionally, rather than at every step. This process is known as thinning the chain. We can decrease the frequency of sampling so that we only draw samples every kth step of the chain (Raftery & Lewis, 1996), with the value of k chosen so that the analysis produces about 10000 to 20000 samples.

To check whether we have drawn a sufficient number of independent samples from the target distribution, we can calculate the effective sample size (ESS) for each parameter, the marginal likelihood, and the posterior probability. A small ESS (<100) is typically interpreted to indicate that the samples are not sufficient to provide an adequate description of the posterior distribution (e.g. Drummond *et al.*, 2007), but much higher ESS values might be needed to obtain 95% credibility intervals on parameter estimates. If an unreasonably large number of samples is needed to achieve a satisfactory ESS, then the MCMC is considered to exhibit poor mixing. This means that the MCMC has not made sufficient changes to the model and parameters to have explored the parameter space adequately, leading to a very slow decay in the autocorrelation between successive samples. This can happen if the acceptance rate for proposals (changes to the model or parameters) is too high or too low. When too many proposals are accepted, the chain tends to stay close to its current state and does not make large enough changes that might cause it to sample from other regions of parameter space. When too many proposals are rejected, then the proposals are too 'bold', so that most proposals lead to states that have a much lower posterior probability than the current state. This is inefficient because the MCMC will spend a lot of time making proposals that it will reject.

The optimal acceptance rate can be considered to be between 20 and 50%, depending on the type of proposal being made (Roberts, Gelman & Gilks, 1997), but in practice the proposal mechanism ('operator') could be adjusted until the acceptance rate is somewhere between 15 and 70%. Diagnosing and resolving MCMC problems can be very difficult, especially in parameter-rich analyses of large data sets. Poor MCMC mixing can be due to overparameterization, inefficient proposal mechanisms, and other factors (Rannala, 2002).

Because Bayesian methods produce a posterior probability distribution for all parameters, we have a natural way to report the uncertainty in all aspects of the results. Once we have collected an appropriate sample from the target distribution, we can summarize the posterior distributions of key model parameters, such as the node times. This is typically done by presenting a measure of central tendency of the distribution of values, such as the mean or median, along with a measure of the spread, such as the 95% credibility interval (the range of values that contains the central 95% of the values in the posterior distribution). In addition to reporting the results of a particular dating analysis, it may be appropriate to report the range of date estimates that are obtained under more than one combination of models and priors. For example, if different calibration schemes give different estimates and there is no compelling way of choosing between the two, then both sets of estimates should be considered. If there are many models of equal fit to the data, then a formal model averaging approach could be used. If this is not possible, then one way to acknowledge the uncertainty in model choice is to report the range of results obtained across the candidate models.

Summarizing the whole sample of phylogenetic trees and the set of all node ages requires different techniques. If the tree topology is fixed, then the posterior distribution of each node can be treated as an ordinary continuous parameter. One approach is to select a single tree from those that have been sampled from the posterior distribution, such as that with the highest likelihood [maximum likelihood (ML) tree] or highest posterior probability [maximum *a posteriori* (MAP) tree]. An alternative approach is to find the maximum clade credibility (MCC) tree. This corresponds to the tree with the highest overall branch support, as indicated by the product of posterior probabilities across the tree (Drummond & Rambaut, 2007). This method has been shown to have good performance compared with other tree summary methods (Heled & Bouckaert, 2013). The posterior distribution for each node age in the MCC tree can be reported using only the ages of matching clades from the posterior. Consequently, the nodes for clades with low posterior probability will tend to have high uncertainty. An alternative is to compute common ancestor heights, which involves obtaining the time to the most recent common ancestor for every pair of taxa for all of the sampled trees. These values provide an estimate of

the posterior distribution for the corresponding nodes in the MCC tree (Heled & Bouckaert, 2013).

A widely used method of summarizing the posterior distribution of trees is to compute a 'majority consensus tree', which contains only clades present in at least 50% of the trees sampled from the posterior. Other nodes are treated as unresolved and are represented by polytomies (Ronquist *et al.*, 2012). Since this consensus tree has not actually been sampled from the posterior distribution, the posterior probability of the consensus tree might be very low, even though individual clades in this consensus tree might have high support (Cranston & Rannala, 2007). However, the presence of polytomies is problematic for estimating divergence dates.

If the dating analysis combines multiple gene trees without producing a single species tree, there are a range of strategies for reporting the date estimates. One approach is to provide the minimum or maximum date estimates for a specific node across all of the gene trees (Maddison & Knowles, 2006; Mossel & Roch, 2008). Alternatively, the tree topology can be summarized prior to producing the date estimates, for example using the average ranks of coalescence times as an estimate of the species tree (Maddison & Knowles, 2006; Mossel & Roch, 2008; Liu *et al.*, 2009).

X. CONCLUSIONS

(1) Bayesian phylogenetic methods are widely used to estimate evolutionary timescales using genetic sequence data. They allow the joint estimation of phylogenetic relationships, divergence times, and evolutionary (or demographic) parameters. A wide range of models have been developed for Bayesian molecular dating, creating a challenging task for users of these methods, because there are many decisions to be made in designing an analysis.

(2) Bayesian molecular dating will always carry a hefty burden of uncertainty and subjectivity. Models and priors are principled oversimplifications at best, and the results need to be carefully considered in the light of model assumptions, the influence of priors, and the conflicting claims of other evidence. But Bayesian molecular dating also provides a means of accommodating and reporting this uncertainty through making the assumptions and priors explicit, investigating and comparing a range of models, and appropriately summarizing the results.

(3) Priorities for future improvements to Bayesian molecular dating include methods for handling genome-scale data, improving ways of incorporating fossil data, increasing the biological accuracy of models, and developing reliable methods for testing model fit and model adequacy.

(4) The amount of sequence data is no longer a limiting factor for many molecular dating analyses, but massive data sets bring new analytical challenges. The efficiency of Bayesian dating methods needs to keep up with the explosive growth of genomic data, particularly for dealing with variation between gene trees in multi-locus data sets. Steps to improve the efficiency of these analyses will include effective implementations of parallel computing and multi-chain analyses, as well as the application of improved MCMC samplers.

(5) Models for molecular dating have increased markedly in their sophistication, but models are often chosen for their statistical tractability rather than biological realism. Better understanding of the patterns and rates of molecular evolution will inform the development of more realistic tree priors and models of branch rates.

(6) One of the most valuable lines of research is the improvement of methods for incorporating calibrating information from geological evidence and other sources of information on evolutionary rates. This is not simply a matter of increasing the quality and quantity of calibrations in dating analyses, but more intelligently incorporating independent temporal information in a way that allows us to get the most out of the data.

(7) There has been a growing recognition of methods for model averaging and testing model adequacy in molecular dating and phylogenetics. The field would benefit from accepted protocols for the evaluation of molecular date estimates, and convenient tests for the influence of choices of models and priors should be more widely adopted.

(8) Molecular dating is an irreplaceable and valuable tool in evolutionary biology. But like most things in biology, it turns out to be a lot more complicated than we might have hoped. In the coming years, there is likely to be continued growth in the amount of genomic data, calibrating information, computational power, and statistical methods, which should pave the way to an unprecedented level of knowledge about molecular evolution and of the history of life on earth.

XI. ACKNOWLEDGEMENTS

XII. REFERENCES

Adachi, J. & Hasegawa, M. (1996). Model of amino acid substitution in proteins encoded by mitochondrial DNA. *Journal of Evolution* **42**, 459–468.

Agrawal, A. F. & Wang, A. D. (2008). Increased transmission of mutations by low-condition females: evidence for condition-dependent DNA repair. *PLoS Biology* **6**, e30.

Angelis, K., Álvarez-Carretero, S., Dos Reis, M. & Yang, Z. (2017). An evaluation of different partitioning strategies for Bayesian estimation of species divergence times. *Systematic Biology*. https://doi.org/10.1093/sysbio/syx061.

Aris-Brosou, S. & Yang, Z. H. (2002). Effects of models of rate evolution on estimation of divergence dates with special reference to the metazoan 18S ribosomal RNA phylogeny. *Systematic Biology* **51**, 703–714.

Aris-Brosou, S. & Yang, Z. H. (2003). Bayesian models of episodic evolution support a late Precambrian explosive diversification of the Metazoa. *Molecular Biology and Evolution* **20**, 1947–1954.

Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M. A. & Alekseyenko, A. V. (2012). Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution* **29**, 2157–2167.

BAELE, G., LEMEY, P. & SUCHARD, M. A. (2016). Genealogical working distributions for Bayesian model testing with phylogenetic uncertainty. *Systematic Biology* **65**, 250–264.

BAELE, G., VAN DE PEER, Y. & VANSTEELANDT, S. (2008). A model-based approach to study nearest-neighbor influences reveals complex substitution patterns in non-coding sequences. *Systematic Biology* **57**, 675–692.

BATTISTUZZI, F. U., BILLING-ROSS, P., MURILLO, O., FILIPSKI, A. & KUMAR, S. (2015). A protocol for diagnosing the effect of calibration priors on posterior time estimates: a case study for the Cambrian explosion of animal phyla. *Molecular Biology and Evolution* **32**, 1907–1912.

BEAULIEU, J. M., O'MEARA, B. C., CRANE, P. & DONOGHUE, M. J. (2015). Heterogeneous rates of molecular evolution and diversification could explain the triassic age estimate for angiosperms. *Systematic Biology* **64**, 869–878.

BELLOT, S. & RENNER, S. S. (2014). Exploring new dating approaches for parasites: the worldwide Apodanthaceae (Cucurbitales) as an example. *Molecular Phylogenetics and Evolution* **80**, 1–10.

BERBEE, M. L. & TAYLOR, J. W. (2010). Dating the molecular clock in fungi–how close are we? *Fungal Biology Reviews* **24**, 1–16.

BERGSTEN, J., NILSSON, A. N. & RONQUIST, F. (2013). Bayesian tests of topology hypotheses with an example from diving beetles. *Systematic Biology* **62**, 660–673.

BIFFIN, E., HILL, R. S. & LOWE, A. J. (2010). Did kauri (*Agathis*: Araucariaceae) really survive the Oligocene drowning of New Zealand? *Systematic Biology* **59**, 594–602.

BININDA-EMONDS, R. (2007). Fast genes and slow clades: comparative rates of molecular evolution in mammals. *Evolutionary Bioinformatics* **3**, 59–85.

BLUM, M. G. & FRANÇOIS, O. (2005). On statistical tests of phylogenetic tree imbalance: the Sackin and other indices revisited. *Mathematical Biosciences* **195**, 141–153.

BOLLBACK, J. P. (2002). Bayesian model adequacy and choice in phylogenetics. *Molecular Biology and Evolution* **19**, 1171–1180.

BOSKOVA, V., BONHOEFFER, S. & STADLER, T. (2014). Inference of epidemiological dynamics based on simulated phylogenies using birth-death and coalescent models. *PLoS Computational Biology* **10**, e1003913.

BRADLEY, B. J. (2008). Reconstructing phylogenies and phenotypes: a molecular view of human evolution. *Journal of Anatomy* **212**, 337–353.

BRANDLEY, M. C., SCHMITZ, A. & REEDER, T. W. (2005). Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. *Systematic Biology* **54**, 373–390.

BRANDLEY, M. C., WANG, Y., GUO, X., DE OCA, A. N. M., FERÍA-ORTÍZ, M., HIKIDA, T. & OTA, H. (2011). Accommodating heterogenous rates of evolution in molecular divergence dating methods: an example using intercontinental dispersal of *Plestiodon* (*Eumeces*) lizards. *Systematic Biology* **60**, 3–15.

BROMHAM, L. (2003). Molecular clocks and explosive radiations. *Journal of Molecular Evolution* **57**, S13–S20.

BROMHAM, L. (2006). Molecular dates for the Cambrian explosion: is the light at the end of the tunnel an oncoming train? *Palaeontologia Electronica* **9**, 3.

BROMHAM, L. (2009). Why do species vary in their rate of molecular evolution? *Biology Letters* **5**, 401–404.

BROMHAM, L. (2011). The genome as a life-history character: why rate of molecular evolution varies between mammal species. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **366**, 2503–2513.

BROMHAM, L., COWMAN, P. F. & LANFEAR, R. (2013). Parasitic plants have increased rates of molecular evolution across all three genomes. *BMC Evolutionary Biology* **13**, 126.

BROMHAM, L., HUA, X., LANFEAR, R. & COWMAN, P. (2015). Exploring the relationships between mutation rates, life history, genome size, environment and species richness in flowering plants. *American Naturalist* **185**, 507–524.

BROMHAM, L. D. & HENDY, M. D. (2000). Can fast early rates reconcile molecular dates with the Cambrian explosion? *Proceedings of the Royal Society of London Series B: Biological Sciences* **267**, 1041–1047.

BROOKS, S. P. & GELMAN, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* **7**, 434–455.

BROWN, J. W. & SMITH, S. A. (2017). The past sure is tense: on interpreting phylogenetic divergence time estimates. *Systematic Biology*. https://doi.org/10.1093/sysbio/syx074.

BROWN, R. P. & YANG, Z. (2010). Bayesian dating of shallow phylogenies with a relaxed clock. *Systematic Biology* **59**, 119–131.

BROWN, R. P. & YANG, Z. (2011). Rate variation and estimation of divergence times using strict and relaxed clocks. *BMC Evolutionary Biology* **11**, 1–12.

CHRISTIN, P.-A., SPRIGGS, E., OSBORNE, C. P., STRÖMBERG, C. A., SALAMIN, N. & EDWARDS, E. J. (2013). Molecular dating, evolutionary rates, and the age of the grasses. *Systematic Biology* **63**, 153–165.

CONDAMINE, F., NAGALINGUM, N., MARSHALL, C. & MORLON, H. (2015). Origin and diversification of living cycads: a cautionary tale on the impact of the branching process prior in Bayesian molecular dating. *BMC Evolutionary Biology* **15**, 65.

CRACRAFT, J., HOUDE, P., HO, S. Y. W., MINDELL, D. P., FJELDSÅ, J., LINDOW, B., EDWARDS, S. V., RAHBEK, C., MIRARAB, S., WARNOW, T., GILBERT, M. T. P., ZHANG, G., BRAUN, E. L. & JARVIS, E. D. (2015). Response to comment on "Whole-genome analyses resolve early branches in the tree of life of modern birds". *Science* **349**, 1460.

CRANSTON, K. A. & RANNALA, B. (2007). Summarizing a posterior distribution of trees using agreement subtrees. *Systematic Biology* **56**, 578–590.

CRISP, M. D., HARDY, N. B. & COOK, L. G. (2014). Clock model makes a large difference to age estimates of long-stemmed clades with no internal calibration: a test using Australian grasstrees. *BMC Evolutionary Biology* **14**, 263–263.

DAYHOFF, M., SCHWARTZ, R. & ORCUTT, B. (1978). A model of evolutionary change in proteins. In *Atlas of Protein Sequence and Structure* (Volume 5), ed. M. O. Dayhoff, pp. 345–352. National Biomedical Research Foundation Silver Spring. National Biomedical Research Foundation, Washington DC, United States of America.

DEGNAN, J. H. & ROSENBERG, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* **24**, 332–340.

DICKERSON, R. E. (1971). The structure of cytochrome c and rates of molecular evolution. *Journal of Molecular Evolution* **1**, 26–45.

DONOGHUE, P. C. J. & YANG, Z. (2016). The evolution of methods for establishing evolutionary timescales. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **371**, 20160020.

DOOLITTLE, R. F. & BLOMBÄCK, B. (1964). Amino-acid sequence investigations of fibrinopeptides from various mammals: evolutionary implications. *Nature* **202**, 147–152.

DORNBURG, A., BEAULIEU, J. M., OLIVER, J. C. & NEAR, T. J. (2011). Integrating fossil preservation biases in the selection of calibrations for molecular divergence time estimation. *Systematic Biology* **60**, 519–527.

DORNBURG, A., BRANDLEY, M. C., MCGOWEN, M. R. & NEAR, T. J. (2012). Relaxed clocks and inferences of heterogeneous patterns of nucleotide substitution and divergence time estimates across whales and dolphins (Mammalia: Cetacea). *Molecular Biology and Evolution* **29**, 721–736.

DRUMMOND, A. J., HO, S. Y., RAWLENCE, N. & RAMBAUT, A. (2007). *A Rough Guide to BEAST 1.4*. University of Auckland, Auckland.

DRUMMOND, A. J., HO, S. Y. W., PHILLIPS, M. J. & RAMBAUT, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology* **4**, e88.

DRUMMOND, A. J., NICHOLLS, G. K., RODRIGO, A. G. & SOLOMON, W. (2002). Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* **161**, 1307–1320.

DRUMMOND, A. J. & RAMBAUT, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**, 214.

DRUMMOND, A. J., RAMBAUT, A., SHAPIRO, B. & PYBUS, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* **22**, 1185–1192.

DRUMMOND, A. J. & SUCHARD, M. a. (2010). Bayesian random local clocks, or one rate to rule them all. *BMC Biology* **8**, 114.

DUCHÊNE, D. A., DUCHÊNE, S., HOLMES, E. C. & HO, S. Y. W. (2015). Evaluating the adequacy of molecular clock models using posterior predictive simulations. *Molecular Biology and Evolution* **32**, 2986–2995.

DUCHÊNE, S., FOSTER, C. S. & HO, S. Y. W. (2016). Estimating the number and assignment of clock models in analyses of multigene datasets. *Bioinformatics* **32**, 1281–1285.

DUCHÊNE, S. & HO, S. Y. W. (2014). Using multiple relaxed-clock models to estimate evolutionary timescales from DNA sequence data. *Molecular Phylogenetics and Evolution* **77**, 65–70.

DUCHÊNE, S., LANFEAR, R. & HO, S. Y. W. (2014). The impact of calibration and clock-model choice on molecular estimates of divergence times. *Molecular Phylogenetics and Evolution* **78**, 277–289.

EDWARDS, S. & BEERLI, P. (2000). Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* **54**, 1839–1854.

ERWIN, D. H., LAFLAMME, M., TWEEDT, S. M., SPERLING, E. A., PISANI, D. & PETERSON, K. J. (2011). The Cambrian conundrum: early divergence and later ecological success in the early history of animals. *Science* **334**, 1091–1097.

ETIENNE, R. S. & ROSINDELL, J. (2012). Prolonging the past counteracts the pull of the present: protracted speciation can explain observed slowdowns in diversification. *Systematic Biology* **61**, 204–213.

FAN, Y., WU, R., CHEN, M.-H., KUO, L. & LEWIS, P. O. (2011). Choosing among partition models in Bayesian phylogenetics. *Molecular Biology and Evolution* **28**, 523–532.

FELSENSTEIN, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* **17**, 368–376.

FELSENSTEIN, J. (2001). Taking variation of evolutionary rates between sites into account in inferring phylogenies. *Journal of Molecular Evolution* **53**, 447–455.

FISHER, R. A. (1930). *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.

FONTANILLAS, E., WELCH, J.J., THOMAS, J. A. & BROMHAM, L. (2007). The influence of body size and net diversification rate on molecular evolution during the radiation of animal phyla. *BMC Evolutionary Biology* **7**, 95.

FOSTER, C. S. P. & HO, S. Y. W. (2017). Strategies for partitioning clock models in phylogenomic dating: application to the angiosperm evolutionary timescale. *Genome Biology and Evolution* **9**, 2752–2763.

FOSTER, C. S. P., SAUQUET, H., VAN DER MERWE, M., MCPHERSON, H., ROSSETTO, M. & HO, S. Y. W. (2017). Evaluating the impact of genomic data and priors on Bayesian estimates of the angiosperm evolutionary timescale. *Systematic Biology* **66**, 338–351.

FOSTER, P. (2004). Modeling compositional heterogeneity. *Systematic Biology* **53**, 485–495.

FUNK, D. J. & OMLAND, K. E. (2003). Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics* **34**, 397–423.

GAVRYUSHKINA, A., HEATH, T. A. K. D. T., STADLER, T., WELCH, D. & DRUMMOND, A. J. (2017). Bayesian total-evidence dating reveals the recent crown radiation of penguins. *Systematic Biology* **66**, 57–73.

GAVRYUSHKINA, A., WELCH, D., STADLER, T. & DRUMMOND, A. J. (2014). Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Computational Biology* **10**, e1003919.

GELMAN, A., CARLIN, J. B., STERN, H. S. & RUBIN, D. B. (2014). *Bayesian Data Analysis*. Chapman and Hall. CRC, Boca Raton, FL, USA.

GELMAN, A. & RUBIN, D. B. (1992). A single series from the Gibbs sampler provides a false sense of security. In *Bayesian Statistics 4* (eds J. M. BERNARDO, J. BERGER, A. P. DAWID and A. F. M. SMITH), pp. 625–631. Oxford University Press, Oxford.

GELMAN, A. & RUBIN, D. B. (1996). Markov chain Monte Carlo methods in biostatistics. *Statistical Methods in Medical Research* **5**, 339–355.

GERNHARD, T. (2008). The conditioned reconstructed process. *Journal of Theoretical Biology* **253**, 769–778.

GILLESPIE, J. H. (1984). Molecular evolution over the mutational landscape. *Evolution* **38**, 1116–1129.

GILLESPIE, J. H. (1991). *The Causes of Molecular Evolution*. Oxford University Press, Oxford.

GILLMAN, L. N. & WRIGHT, S. D. (2014). Species richness and evolutionary speed: the influence of temperature, water and area. *Journal of Biogeography* **41**, 39–51.

GOLDMAN, N. (1993). Statistical tests of models of DNA substitution. *Journal of Molecular Evolution* **36**, 182–198.

GOLEMBA, M. D., DI LELLO, F. A., BESSONE, F., FAY, F., BENETTI, S., JONES, L. R. & CAMPOS, R. H. (2010). High prevalence of hepatitis C virus genotype 1b infection in a small town of Argentina. Phylogenetic and Bayesian coalescent analysis. *PLoS ONE* **5**, e8751.

GRIFFITHS, R. C. & TAVARE, S. (1994). Sampling theory for neutral alleles in a varying environment. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **344**, 403–410.

GRUBER, K. F., VOSS, R. S. & JANSA, S. A. (2007). Base-compositional heterogeneity in the RAG1 locus among didelphid marsupials: implications for phylogenetic inference and the evolution of GC content. *Systematic Biology* **56**, 83–96.

GUINDON, S. (2013). From trajectories to averages: an improved description of the heterogeneity of substitution rates along lineages. *Systematic Biology* **62**, 22–34.

HASEGAWA, M., KISHINO, H. & YANO, T. (1985). Dating of the human-ape splitting by a molecular clock of mitchondrial DNA. *Journal of Molecular Evolution* **22**, 160–174.

HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

HEATH, T. A., HEDTKE, S. M. & HILLIS, D. M. (2008). Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution* **46**, 239–257.

HEATH, T. A. & MOORE, B. (2014). Bayesian inference of species divergence times. In *Bayesian Phylogenetics: Methods, Algorithms, and Applications* (eds M.-H. CHEN, L. KUO and P. O. LEWIS), pp. 277–318. Chapman & Hall/CRC, Boca Raton.

HEATH, T. A., HOLDER, M. T. & HUELSENBECK, J. P. (2012). A Dirichlet process prior for estimating lineage-specific substitution rates. *Molecular Biology and Evolution* **29**, 939–955.

HEATH, T. A., HUELSENBECK, J. P. & STADLER, T. (2014). The fossilized birth-death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E2957–E2966.

HELED, J. & BOUCKAERT, R. R. (2013). Looking for trees in the forest: summary tree from posterior samples. *BMC Evolutionary Biology* **13**, 221.

HELED, J. & DRUMMOND, A. J. (2010). Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* **27**, 570–580.

HELED, J. & DRUMMOND, A. J. (2012). Calibrated tree priors for relaxed phylogenetics and divergence time estimation. *Systematic Biology* **61**, 138–149.

HELED, J. & DRUMMOND, A. J. (2015). Calibrated birth-death phylogenetic time-tree priors for Bayesian inference. *Systematic Biology* **64**, 369–383.

HIPSLEY, C. A. & MÜLLER, J. (2014). Beyond fossil calibrations: realities of molecular clock practices in evolutionary biology. *Frontiers in Genetics* **5**, 138.

HO, S. Y. W. (2009). An examination of phylogenetic models of substitution rate variation among lineages. *Biology Letters* **5**, 421–424.

HO, S. Y. W., CHEN, A. X., LINS, L. S., DUCHÊNE, D. A. & LO, N. (2016). The genome as an evolutionary timepiece. *Genome Biology and Evolution* **8**, 3006–3010.

HO, S. Y. W. & DUCHÊNE, S. (2014). Molecular-clock methods for estimating evolutionary rates and time scales. *Molecular Ecology* **23**, 5947–5975.

HO, S. Y. W., DUCHÊNE, S. & DUCHÊNE, D. A. (2015). Simulating and detecting autocorrelation of molecular evolutionary rates among lineages. *Molecular Ecology Resources* **15**, 688–696.

HO, S. Y. W. & LANFEAR, R. (2010). Improved characterisation of among-lineage rate variation in cetacean mitogenomes using codon-partitioned relaxed clocks. *Mitochondrial DNA* **21**, 138–146.

HO, S. Y. W. & PHILLIPS, M. J. (2009). Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology* **58**, 367–380.

HO, S. Y. W., PHILLIPS, M. J., DRUMMOND, A. J. & COOPER, A. (2005). Accuracy of rate estimation using relaxed-clock models with a critical focus on the early metazoan radiation. *Molecular Biology and Evolution* **22**, 1355–1363.

HÖHNA, S., STADLER, T., RONQUIST, F. & BRITTON, T. (2011). Inferring speciation and extinction rates under different sampling schemes. *Molecular Biology and Evolution* **28**, 2577–2589.

HUA, X. & BROMHAM, L. (2016). Phylometrics: an R package for detecting macroevolutionary patterns, using phylogenetic metrics and backward tree simulation. *Methods in Ecology and Evolution* **7**, 806–810.

HUELSENBECK, J. P., JOYCE, P., LAKNER, C. & RONQUIST, F. (2008). Bayesian analysis of amino acid substitution models. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **363**, 3941–3953.

HUELSENBECK, J. P., LARGET, B. & ALFARO, M. E. (2004). Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Molecular Biology and Evolution* **21**, 1123–1133.

HUELSENBECK, J. P., LARGET, B., MILLER, R. E. & RONQUIST, F. (2002). Potential applications and pitfalls of Bayesian inference of phylogeny. *Systematic Biology* **51**, 673–688.

HUELSENBECK, J. P., LARGET, B. & SWOFFORD, D. (2000). A compound Poisson process for relaxing the molecular clock. *Genetics* **154**, 1879–1892.

HUELSENBECK, J. P., RONQUIST, F., NIELSEN, R. & BOLLBACK, J. P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* **294**, 2310–2314.

HUGALL, A. F. & LEE, M. S. Y. (2007). The likelihood node density effect and consequences for evolutionary studies of molecular rates. *Evolution* **61**, 2293–2307.

INOUE, J., DONOGHUE, P. C. J. & YANG, Z. (2010). The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. *Systematic Biology* **59**, 74–89.

JARVIS, E. D., MIRARAB, S., ABERER, A. J., LI, B., HOUDE, P., LI, C., HO, S. Y., FAIRCLOTH, B. C., NABHOLZ, B. & HOWARD, J. T. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331.

JEFFREYS, H. (1961). *The Theory of Probability*, 3rd Edition. Oxford University Press, Oxford.

JIA, F., LO, N. & HO, S. Y. W. (2014). The impact of modelling rate heterogeneity among sites on phylogenetic estimates of intraspecific evolutionary rates and timescales. *PLoS ONE* **9**, e95722.

JONES, D. T., TAYLOR, W. R. & THORNTON, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences* **8**, 275–282.

JOYCE, W. G., PARHAM, J. F., LYSON, T. R., WARNOCK, R. C. & DONOGHUE, P. C. (2013). A divergence dating analysis of turtles using fossil calibrations: an example of best practices. *Journal of Paleontology* **87**, 612–634.

JUKES, T. H. & CANTOR, C. R. (1969). Evolution of protein molecules. In *Mammalian Protein Metabolism* (Volume 3, ed. H. N. MUNRO), pp. 21–132. Academic Press, New York.

KASS, R. E. & RAFTERY, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773–795.

KENDALL, D. G. (1948). On the generalized birth death process. *Annals of Mathematical Statistics* **19**, 1–15.

KIMURA, M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequence. *Journal of Molecular Evolution* **16**, 111–120.

KINGMAN, J. F. C. (1982). The coalescent. *Stochastic Processes and their Applications* **13**, 235–248.

KISHINO, H., THORNE, J. L. & BRUNO, W. J. (2001). Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Molecular Biology and Evolution* **18**, 352–361.

KNOLL, A. H. (2017). Biogeochemistry: food for early animal evolution. *Nature* **548**, 528–530.

KONG, A., FRIGGE, M. L., MASSON, G., BESENBACHER, S., SULEM, P., MAGNUSSON, G., GUDJONSSON, S. A., SIGURDSSON, A., JONASDOTTIR, A. & JONASDOTTIR, A. (2012). Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488**, 471–475.

KUBATKO, L. S. & DEGNAN, J. H. (2007). Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology* **56**, 17–24.

KUMAR, S. & HEDGES, S. B. (2016). Advances in time estimation methods for molecular data. *Molecular Biology and Evolution* **33**, 863–869.

LANFEAR, R., CALCOTT, B., HO, S. Y. W. & GUINDON, S. (2012). PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* **29**, 1695–1701.

LANFEAR, R., HO, S. Y. W., DAVIES, T. J., MOLES, A. T. A. L., SWENSON, N. G., WARMAN, L., ZANNE, A. E. & ALLEN, A. P. (2013). Taller plants have lower rates of molecular evolution: the rate of mitosis hypothesis. *Nature Communications* **4**, 1879.

LARTILLOT, N. & PHILIPPE, H. (2006). Computing Bayes factors using thermodynamic integration. *Systematic Biology* **55**, 195–207.

LARTILLOT, N., PHILLIPS, M. J. & RONQUIST, F. (2016). A mixed relaxed clock model. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **371**, 20150132.

LEE, M., OLIVER, P. & HUTCHINSON, M. (2009). Phylogenetic uncertainty and molecular clock calibrations: a case study of legless lizards (Pygopodidae, Gekkota). *Molecular Phylogenetics and Evolution* **50**, 661–666.

LEE, M. S., SOUBRIER, J. & EDGECOMBE, G. D. (2013). Rates of phenotypic and genomic evolution during the Cambrian explosion. *Current Biology* **23**, 1889–1895.

LEE, M. S. Y. & HO, S. Y. W. (2016). Molecular clocks. *Current Biology* **26**, R387–R407.

LEIGH, J. W., SUSKO, E., BAUMGARTNER, M. & ROGER, A. J. (2008). Testing congruence in phylogenomic analysis. *Systematic Biology* **57**, 104–115.

LEMEY, P., RAMBAUT, A., DRUMMOND, A. J. & SUCHARD, M. A. (2009). Bayesian phylogeography finds its roots. *PLoS Computational Biology* **5**, e1000520.

LEMMON, A. R. & MORIARTY, E. C. (2004). The importance of proper model assumption in Bayesian phylogenetics. *Systematic Biology* **53**, 265–277.

LEPAGE, T., BRYANT, D., PHILIPPE, H. & LARTILLOT, N. (2007). A general comparison of relaxed molecular clock models. *Molecular Biology and Evolution* **24**, 2669–2680.

LEPAGE, T., LAWI, S., TUPPER, P. & BRYANT, D. (2006). Continuous and tractable models for the variation of evolutionary rates. *Mathematical Biosciences* **199**, 216–233.

LEWIS, P. O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* **50**, 913–925.

LI, S., PEARL, D. K. & DOSS, H. (2000). Phylogenetic tree construction using Markov chain Monte Carlo. *Journal of the American Statistical Association* **95**, 493–508.

LI, W. L. S. & DRUMMOND, A. J. (2012). Model averaging and Bayes factor calculation of relaxed molecular clocks in Bayesian phylogenetics. *Molecular Biology and Evolution* **29**, 751–761.

LINDER, H. P., HARDY, C. R. & RUTSCHMANN, F. (2005). Taxon sampling effects in molecular clock dating: an example from the African Restionaceae. *Molecular Phylogenetics and Evolution* **35**, 569–582.

LINDER, M., BRITTON, T. & SENNBLAD, B. (2011). Evaluation of Bayesian models of substitution rate evolution–parental guidance versus mutual independence. *Systematic Biology* **60**, 329–342.

LIU, K., LINDER, C. R. & WARNOW, T. (2011). Multiple sequence alignment: a major challenge to large-scale phylogenetics. *PLoS Currents* **2**, RRN1198.

LIU, L., YU, L. L., PEARL, D. K. & EDWARDS, S. V. (2009). Estimating species phylogenies using coalescence times among sequences. *Systematic Biology* **58**, 468–477.

LIU, Y., ROSSITER, S. J., HAN, X., COTTON, J. A. & ZHANG, S. (2010). Cetaceans on a molecular fast track to ultrasonic hearing. *Current Biology* **20**, 1834–1839.

LOCKHART, P. J. & CAMERON, S. A. (2001). Trees for bees. *Trends in Ecology & Evolution* **16**, 84–88.

LÖYTYNOJA, A. & GOLDMAN, N. (2009). Uniting alignments and trees. *Science* **324**, 1528–1529.

LUNTER, G., MIKLÓS, I., DRUMMOND, A., JENSEN, J. L. & HEIN, J. (2005). Bayesian coestimation of phylogeny and sequence alignment. *BMC Bioinformatics* **6**, 83.

LUO, D. P., XU, H., LIU, Z. L., GUO, J. X., LI, H. Y., CHEN, L. T., FANG, C., ZHANG, Q. Y., BAI, M., YAO, N., WU, H., JI, C. H., ZHENG, H. Q., CHEN, Y. L., YE, S., LI, X. Y., ZHAO, X. C., LI, R. Q. & LIU, Y. G. (2013). A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. *Nature Genetics* **45**, 573–U157.

MADDISON, W. P. (1997). Gene trees in species trees. *Systematic Biology* **46**, 523–536.

MADDISON, W. P. & KNOWLES, L. L. (2006). Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* **55**, 21–30.

MAGALLÓN, S. (2010). Using fossils to break long branches in molecular dating: a comparison of relaxed clocks applied to the origin of angiosperms. *Systematic Biology* **59**, 384–399.

MARGOLIASH, E. (1963). Primary structure and evolution of cytochrome c. *Proceedings of the National Academy of Sciences of the United States of America* **50**, 672–679.

MARSHALL, C. R. (2008). A simple method for bracketing absolute divergence times on molecular phylogenies using multiple fossil calibration points. *American Naturalist* **171**, 726–742.

MARSHALL, D. C., HILL, K. B. R., MOULDS, M., VANDERPOOL, D., COOLEY, J. R., MOHAGAN, A. B. & SIMON, C. (2016). Inflation of molecular clock rates and dates: molecular phylogenetics, biogeography, and diversification of a global cicada radiation from Australasia (Hemiptera: Cicadidae: Cicadettini). *Systematic Biology* **65**, 16–34.

MARSHALL, D. C., SIMON, C. & BUCKLEY, T. R. (2006). Accurate branch length estimation in partitioned Bayesian analyses requires accommodation of among-partition rate variation and attention to branch length priors. *Systematic Biology* **55**, 993–1003.

MARTIN, C. H., CRAWFORD, J. E., TURNER, B. J. & SIMONS, L. H. (2016). Diabolical survival in Death Valley: recent pupfish colonization, gene flow and genetic assimilation in the smallest species range on earth. *Proceedings of the Royal Society of London B: Biological Sciences* **283**, 20152334.

MATSUBA, C., OSTROW, D. G., SALOMON, M. P., TOLANI, A. & BAER, C. F. (2013). Temperature, stress and spontaneous mutation in *Caenorhabditis briggsae* and *Caenorhabditis elegans*. *Biology Letters* **9**, 20120334.

MAU, B. & NEWTON, M. A. (1997). Phylogenetic inference for binary data on dendograms using Markov chain Monte Carlo. *Journal of Computational and Graphical Statistics* **6**, 122–131.

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. & TELLER, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087–1092.

MININ, V. N., BLOOMQUIST, E. W. & SUCHARD, M. A. (2008). Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population pynamics. *Molecular Biology and Evolution* **25**, 1459–1471.

MITCHELL, K. J., COOPER, A. & PHILLIPS, M. J. (2015). Comment on "Whole-genome analyses resolve early branches in the tree of life of modern birds". *Science* **349**, 1460–1460.

MOLAK, M., SUCHARD, N. A., HO, S. Y. W., BEILMAN, D. W. & SHAPIRO, B. (2015). Empirical calibrated radiocarbon sampler: a tool for incorporating radiocarbon – date and calibration error into Bayesian phylogenetic analyses of ancient DNA. *Molecular Ecology Resources* **15**, 81–86.

MORGAN, G. (1998). Emile Zuckerkandl, Linus Pauling, and the molecular evolutionary clock, 1959–1965. *Journal of the History of Biology* **31**, 155–178.

MORRISON, D. A. (2009). Why would phylogeneticists ignore computerized sequence alignment? *Systematic Biology* **58**, 150–158.

MOSSEL, E. & ROCH, S. (2008). Incomplete lineage sorting: consistent phylogeny estimation from multiple loci. *IEEE/ACM Transactions on Computational Biology* **7**, 166–171.

MOSSEL, E. & VIGODA, E. (2005). Phylogenetic MCMC algorithms are misleading on mixtures of trees. *Science* **309**, 2207–2209.

NEAR, T. J., BOLNICK, D. I. & WAINWRIGHT, P. C. (2005). Fossil calibrations and molecular divergence time estimates in centrarchid fishes (Teleostei: Centrarchidae). *Evolution* **59**, 1768–1782.

NEKRUTENKO, A., WADHAWAN, S., GOETTING-MINESKY, P. & MAKOVA, K. D. (2005). Oscillating evolution of a mammalian locus with overlapping reading frames: an XLas/ALEX relay. *PLoS Genetics* **1**, e18.

NELSON, B. J., ANDERSEN, J. J. & BROWN, J. M. (2015). Deflating trees: improving Bayesian branch-length estimates using informed priors. *Systematic Biology* **64**, 441–447.

NEWTON, M. A. & RAFTERY, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society. Series B (Methodological)* **56**, 3–48.

NYLANDER, J. A. A., WILGENBUSCH, J. C., WARREN, D. L. & SWOFFORD, D. L. (2008). AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics* **24**, 581–583.

O'REILLY, J. E., DOS REIS, M. & DONOGHUE, P. C. (2015). Dating tips for divergence-time estimation. *Trends in Genetics* **31**, 637–650.

OBBARD, D. J., MACLENNAN, J., KIM, K.-W., RAMBAUT, A., O'GRADY, P. M. & JIGGINS, F. M. (2012). Estimating divergence dates and substitution rates in the *Drosophila* phylogeny. *Molecular Biology and Evolution* **29**, 3459–3473.

PAPADOPOULOU, A., ANASTASIOU, I. & VOGLER, A. P. (2010). Revisiting the insect mitochondrial molecular clock: the Mid-Aegean trench calibration. *Molecular Biology and Evolution* **27**, 1659–1672.

PARHAM, J. F., DONOGHUE, P. C., BELL, C. J., CALWAY, T. D., HEAD, J. J., HOLROYD, P. A., INOUE, J. G., IRMIS, R. B., JOYCE, W. G. & KSEPKA, D. T. (2011). Best practices for justifying fossil calibrations. *Systematic Biology* **61**, 346–359.

PENN, O., PRIVMAN, E., LANDAN, G., GRAUR, D. & PUPKO, T. (2010). An alignment confidence score capturing robustness to guide tree uncertainty. *Molecular Biology and Evolution* **27**, 1759–1767.

PETERSON, K. J., COTTON, J. A., GEHLING, J. G. & PISANI, D. (2008). The Ediacaran emergence of bilaterians: congruence between the genetic and the geological fossil records. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **363**, 1435–1443.

PHILLIPS, M. J. (2015). Geomolecular dating and the origin of placental mammals. *Systematic Biology* **65**, 546–557.

POSADA, D. & CRANDALL, K. A. (2002). The effect of recombination on the accuracy of phylogeny estimation. *Journal of Molecular Evolution* **54**, 396–402.

PRUM, R. O., BERV, J. S., DORNBURG, A., FIELD, D. J., TOWNSEND, J. P., LEMMON, E. M. & LEMMON, A. R. (2015). A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* **526**, 569–573.

PUPKO, T., HUCHON, D., CAO, Y., OKADA, N. & HASEGAWA, M. (2002). Combining multiple data sets in a likelihood analysis: which models are the best? *Molecular Biology and Evolution* **19**, 2294–2307.

PYBUS, O. G. & RAMBAUT, A. (2009). Evolutionary analysis of the dynamics of viral infectious disease. *Nature Reviews Genetics* **10**, 540–550.

PYBUS, O. G., RAMBAUT, A. & HARVEY, P. H. (2000). An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* **155**, 1429–1437.

RAFTERY, A. E. & LEWIS, S. M. (1996). Implementing MCMC. In *Markov Chain Monte Carlo in Practice* (eds W. R. GILKS, S. RICHARDSON and D. J. SPIEGELHALTER). Chapman and Hall/CRC, Boca Raton. 115–130.

RANNALA, B. (2002). Identifiability of parameters in MCMC Bayesian inference of phylogeny. *Systematic Biology* **51**, 754–760.

Rannala, B. (2016). Conceptual issues in Bayesian divergence time estimation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **371**, 20150134.

Rannala, B. & Yang, Z. H. (1996). Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *Journal of Molecular Evolution* **43**, 304–311.

Rannala, B. & Yang, Z. H. (2003). Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* **164**, 1645–1656.

Rannala, B. & Yang, Z. (2007). Inferring speciation times under an episodic molecular clock. *Systematic Biology* **56**, 453–466.

dos Reis, M. (2016). Notes on the birth-death prior with fossil calibrations for Bayesian estimation of species divergence times. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **371**, 20150128.

dos Reis, M., Donoghue, P. C. & Yang, Z. (2016). Bayesian molecular clock dating of species divergences in the genomics era. *Nature Reviews Genetics* **17**, 71–80.

dos Reis, M., Inoue, J., Hasegawa, M., Asher, R. J., Donoghue, P. C. & Yang, Z. (2012). Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proceedings of the Royal Society of London Series B: Biological Sciences* **279**, 3491–3500.

dos Reis, M., Thawornwattana, Y., Angelis, K., Telford, M. J., Donoghue, P. C. & Yang, Z. (2015). Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. *Current Biology* **25**, 2939–2950.

dos Reis, M. & Yang, Z. (2013). The unbearable uncertainty of Bayesian divergence time estimation. *Journal of Systematics and Evolution* **51**, 30–43.

dos Reis, M., Zhu, T. Q. & Yang, Z. H. (2014). The impact of the rate prior on Bayesian estimation of divergence times with multiple loci. *Systematic Biology* **63**, 555–565.

Rannala, B. & Yang, Z. H. (2008). Phylogenetic inference using whole genomes. *Annual Review of Genomics and Human Genetics* **9**, 217–231.

Rieux, A. & Balloux, F. (2016). Inferences from tip-calibrated phylogenies: a review and a practical guide. *Molecular Ecology* **25**, 1911–1924.

Ritchie, A. M., Lo, N. & Ho, S. Y. W. (2017). The impact of the tree prior on molecular dating of heterogeneous inter- and intraspecies data sets. *Systematic Biology* **66**, 413–425.

Roberts, G. O., Gelman, A. & Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Annals of Applied Probability* **7**, 110–120.

Ronquist, F., Lartillot, N. & Phillips, M. J. (2016). Closing the gap between rocks and clocks using total-evidence dating. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **371**, 20150136.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A. & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**, 539–542.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics* **12**, 1151–1172.

Sağlam, İ. K., Baumsteiger, J. & Miller, M. R. (2017). Failure to differentiate between divergence of species and their genes can result in over-estimation of mutation rates in recently diverged species. *Proceedings of the Royal Society B: Biological Sciences* **284**, 20170021.

Sánchez-Baracaldo, P., Raven, J. A., Pisani, D. & Knoll, A. H. (2017). Early photosynthetic eukaryotes inhabited low-salinity habitats. *Proceedings of the National Academy of Sciences of the United States of America* **114**, E7737–E7745.

Sanders, K. L. & Lee, M. S. (2007). Evaluating molecular clock calibrations using Bayesian analyses with soft and hard bounds. *Biology Letters* **3**, 275–279.

Sarich, V. M. & Wilson, A. C. (1967). Immunological time scale for hominid evolution. *Science* **158**, 1200–1203.

Sauquet, H., Ho, S. Y. W., Gandolfo, M. A., Jordan, G. J., Wilf, P., Cantrill, D.J., Bayly, M. J., Brown, G. K., Carpenter, R. J., Lee, D. M., Murphy, D. J., Sniderman, J. M. K. & Udovicic, F. (2012). Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of *Nothofagus* (Fagales). *Systematic Biology* **61**, 289–313.

Schulte, J. A. (2013). Undersampling taxa will underestimate molecular divergence dates: an example from the South American lizard clade Liolaemini. *International Journal of Evolutionary Biology* **2013**, 1–12.

Shapiro, B., Ho, S. Y. W., Drummond, A. J., Suchard, M. A., Pybus, O. G. & Rambaut, A. (2011). Bayesian phylogenetic method to estimate unknown sequence ages. *Molecular Biology and Evolution* **28**, 879–887.

Sharp, N. P. & Agrawal, A. F. (2012). Evidence for elevated mutation rates in low-quality genotypes. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 6142–6146.

Sinsheimer, J. S., Lake, J. A. & Little, R. J. A. (1996). Bayesian hypothesis testing of four-taxon topologies using molecular sequence data. *Biometrics* **52**, 193–210.

Slater, G. J. (2013). Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the Cretaceous-Palaeogene boundary. *Methods in Ecology and Evolution* **4**, 734–744.

Snir, S. (2014). On the number of genomic pacemakers: a geometric approach. *Algorithms for Molecular Biology* **9**, 26.

Soares, A. E. R. & Schrago, C. G. (2012). The influence of taxon sampling and tree shape on molecular dating: an empirical example from mammalian mitochondrial genomes. *Bioinformatics and Biology Insights* **6**, 129–143.

Soares, A. E. & Schrago, C. G. (2015). The influence of taxon sampling on Bayesian divergence time inference under scenarios of rate heterogeneity among lineages. *Journal of Theoretical Biology* **364**, 31–39.

Soubrier, J., Steel, M., Lee, M. S. Y., Der Sarkissian, C., Guindon, S., Ho, S. Y. W. & Cooper, A. (2012). The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Molecular Biology and Evolution* **29**, 3345–3358.

Springer, M. S., Meredith, R. W., Teeling, E. C. & Murphy, W. J. (2013). Technical comment on "The placental mammal ancestor and the post–K-Pg radiation of placentals". *Science* **341**, 613.

Springer, M. S., Murphy, W. J., Eizirik, E. & O'Brien, S. J. (2003). Placental mammal diversification and the cretaceous–tertiary boundary. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 1056–1061.

Stadler, T. (2009). On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology* **261**, 58–66.

Stadler, T. & Yang, Z. H. (2013). Dating phylogenies with sequentially sampled tips. *Systematic Biology* **62**, 674–688.

Suchard, M. A., Weiss, R. E. & Sinsheimer, J. S. (2001). Bayesian selection of continuous-time Markov chain evolutionary models. *Molecular Biology and Evolution* **18**, 1001–1013.

Sullivan, J. & Joyce, P. (2005). Model selection in phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* **36**, 445–466.

Szöllősi, G. J., Tannier, E., Daubin, V. & Boussau, B. (2015). The inference of gene trees with species trees. *Systematic Biology* **64**, E42–E62.

Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the. Life Sciences* **17**, 57–86.

Thorne, J. L., Kishino, H. & Painter, I. S. (1998). Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution* **15**, 1647–1657.

Thornhill, A. H., Popple, L. W., Carter, R. J., Ho, S. Y. W. & Crisp, M. D. (2012). Are pollen fossils useful for calibrating relaxed molecular clock dating of phylogenies? A comparative study using Myrtaceae. *Molecular Phylogenetics and Evolution* **63**, 15–27.

Vaughan, T. G., Welch, D., Drummond, A. J., Biggs, P. J., George, T. & French, N. P. (2017). Inferring ancestral recombination graphs from bacterial genomic data. *Genetics* **205**, 857–870.

Venditti, C. & Pagel, M. (2009). Speciation as an active force in promoting genetic evolution. *Trends in Ecology & Evolution* **25**, 14–20.

Warnock, R. C., Parham, J. F., Joyce, W. G., Lyson, T. R. & Donoghue, P. C. (2015). Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. *Proceedings of the Royal Society of London Series B: Biological Sciences* **282**, 20141013.

Warnock, R. C. M., Yang, Z. & Donoghue, P. C. J. (2012). Exploring uncertainty in the calibration of the molecular clock. *Biology Letters* **8**, 156–159.

Warren, D. L., Geneva, A. J. & Lanfear, R. (2017). RWTY (R W There Yet): an R package for examining convergence of Bayesian phylogenetic analyses. *Molecular Biology and Evolution* **34**, 1016–1020.

Welch, J. J., Bininda-Emonds, O. R. & Bromham, L. (2008). Correlates of substitution rate variation in mammalian protein-coding sequences. *BMC Evolutionary Biology* **8**, 53.

Welch, J. J. & Bromham, L. (2005). Molecular dating when rates vary. *Trends in Ecology & Evolution* **20**, 320–327.

Welch, J. J., Fontanillas, E. & Bromham, L. (2005). Molecular dates for the "Cambrian explosion": the influence of prior assumptions. *Systematic Biology* **54**, 672–678.

Wernegreen, J. J., Lazarus, A. B. & Degnan, P. H. (2002). Small genome of *Candidatus* Blochmannia, the bacterial endosymbiont of *Camponotus*, implies irreversible specialization to an intracellular lifestyle. *Microbiology* **148**, 2551–2556.

Wertheim, J. O., Sanderson, M. J., Worobey, M. & Bjork, A. (2009). Relaxed molecular clocks, the bias-variance trade-off, and the quality of phylogenetic inference. *Systematic Biology* **59**, 1–8.

Whelan, S. & Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology and Evolution* **18**, 691–699.

Wilkinson, R. D., Steiper, M. E., Soligo, C., Martin, R. D., Yang, Z. & Tavaré, S. (2011). Dating primate divergences through an integrated analysis of palaeontological and molecular data. *Systematic Biology* **60**, 16–31.

Wilson, A. C. & Sarich, V. M. (1969). A molecular timescale for human evolution. *Proceedings of the National Academy of Sciences of the United States of America* **63**, 1088–1093.

Wong, K. M., Suchard, M. A. & Huelsenbeck, J. P. (2008). Alignment uncertainty and genomic analysis. *Science* **319**, 473–476.

Worobey, M., Han, G.-Z. & Rambaut, A. (2014). A synchronized global sweep of the internal genes of modern avian influenza virus. *Nature* **508**, 254–257.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics* **16**, 97–159.

Wu, C.-H., Suchard, M. A. & Drummond, A. J. (2013). Bayesian selection of nucleotide substitution models and their site assignments. *Molecular Biology and Evolution* **30**, 669–688.

Wu, M., Chatterji, S. & Eisen, J. A. (2012). Accounting for alignment uncertainty in phylogenomics. *PLoS ONE* **7**, e30288.

Xie, W., Lewis, P. O., Fan, Y., Kuo, L. & Chen, M.-H. (2011). Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology* **60**, 150–160.

Yang, Z. (1993). Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Molecular Biology and Evolution* **10**, 1396–1401.

Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution* **39**, 306–314.

Yang, Z. (2006). *Computational Molecular Evolution*. Oxford University Press, Oxford.

Yang, Z. (2014). *Molecular Evolution: A Statistical Approach*. Oxford University Press, Oxford.

Yang, Z., Nielsen, R. & Hasegawa, M. (1998). Models of amino acid substitution and applications to mitochondrial protein evolution. *Molecular Biology and Evolution* **15**, 1600–1611.

Yang, Z. & Rannala, B. (1997). Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo Method. *Molecular Biology and Evolution* **14**, 717–724.

Yang, Z. & Rannala, B. (2006). Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution* **23**, 212–226.

Yoder, A. D. & Yang, Z. H. (2000). Estimation of primate speciation dates using local molecular clocks. *Molecular Biology and Evolution* **17**, 1081–1090.

Young, N. D. & dePamphilis, C. W. (2005). Rate variation in parasitic plants: correlated and uncorrelated patterns among plastid genes of different function. *BMC Evolutionary Biology* **5**, 16.

Yule, G. V. (1924). A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, P.R.S. *Philosophical Transactions of the Royal Society London B* **213**, 21–87.

Zhu, T., dos Reis, M. & Yang, Z. (2014). Characterization of the uncertainty of divergence time estimation under relaxed molecular clock models using multiple loci. *Systematic Biology* **64**, 267–280.

Zuckerkandl, E. & Pauling, L. (1962). Molecular disease, evolution and genic heterogeneity. In *Horizons in Biochemistry* (eds M. Kasha and B. Pullman), pp. 189–225. Academic Press, New York.

Zuckerkandl, E. & Pauling, L. (1965). Evolutionary divergence and convergence in proteins. In *Evolving Genes and Proteins* (eds V. Bryson and H. J. Vernon), pp. 97–166. Academic Press, New York.

Zwickl, D. J. & Holder, M. T. (2004). Model parameterization, prior distributions, and the general time-reversible model in Bayesian phylogenetics. *Systematic Biology* **53**, 877–888.