PHILOSOPHICAL TRANSACTIONS B

rstb.royalsocietypublishing.org

Review

Cite this article: Donoghue PCJ, Yang Z. 2016 The evolution of methods for establishing evolutionary timescales. *Phil. Trans. R. Soc. B* **371**: 20160020.
http://dx.doi.org/10.1098/rstb.2016.0020

Accepted: 15 March 2016

One contribution of 15 to a discussion meeting issue 'Dating species divergences using rocks and clocks'.

Subject Areas:
evolution, palaeontology, taxonomy and systematics

Keywords:
molecular clock dating, divergence times, Bayesian inference, fossils, node-calibrations, tip-calibrations

Author for correspondence:
Philip C. J. Donoghue
e-mail: phil.donoghue@bristol.ac.uk

THE ROYAL SOCIETY PUBLISHING

The evolution of methods for establishing evolutionary timescales

Philip C. J. Donoghue[1] and Ziheng Yang[2]

[1]School of Earth Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK
[2]Department of Genetics, Evolution and Environment, University College London, Gower Street, London WC1E 6BT, UK

PCJD, 0000-0003-3116-7463; ZY, 0000-0003-3351-7981

The fossil record is well known to be incomplete. Read literally, it provides a distorted view of the history of species divergence and extinction, because different species have different propensities to fossilize, the amount of rock fluctuates over geological timescales, as does the nature of the environments that it preserves. Even so, patterns in the fossil evidence allow us to assess the incompleteness of the fossil record. While the molecular clock can be used to extend the time estimates from fossil species to lineages not represented in the fossil record, fossils are the only source of information concerning absolute (geological) times in molecular dating analysis. We review different ways of incorporating fossil evidence in modern clock dating analyses, including node-calibrations where lineage divergence times are constrained using probability densities and tip-calibrations where fossil species at the tips of the tree are assigned dates from dated rock strata. While node-calibrations are often constructed by a crude assessment of the fossil evidence and thus involves arbitrariness, tip-calibrations may be too sensitive to the prior on divergence times or the branching process and influenced unduly affected by well-known problems of morphological character evolution, such as environmental influence on morphological phenotypes, correlation among traits, and convergent evolution in disparate species. We discuss the utility of time information from fossils in phylogeny estimation and the search for ancestors in the fossil record.

This article is part of the themed issue 'Dating species divergences using rocks and clocks'.

1. Introduction

Approaches to inference of evolutionary history have a patchy record, punctuated as much by the discovery of new types of data, as by changing philosophies in which data are interpreted. Early phylogenies were based on comparative analysis of living species, whether based on embryology or anatomy, and guided by perceived laws of 'natural affinity', increasing complexity, or divinity [1]. Fossil species played a secondary role, providing evidence for the gradual or episodic evolution of organisms, from primitive to advanced. At the same time, perceptions of the extent of the evolutionary history of Life on Earth have been transformed, from the several million years that Darwin and the majority of his contemporaries would have perceived [2], through to the tens, hundreds and, ultimately, thousands of millions of years that were revealed by radiometric dating [3].

Calibrating the Tree of Life to geological time has traditionally been the preserve of palaeontologists, initially placing more significance on the stratigraphic distribution of fossil species than on their place within a grand Tree of Life. The goal of a universal phylogeny was unrealistic before the discovery of universal genes, and palaeontologists in the New Synthesis had a microevolutionary focus, to infer evolutionary rates on timescales that would blend with studies of living species [4]. Detailed stratigraphic analysis has demonstrated that for some fossil groups, such as the unicellular foraminifera, ancestor–descendent

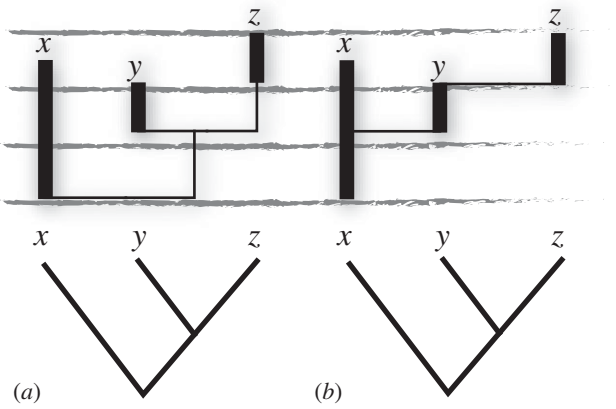

Figure 1. The relationship between cladograms, which consider only relative relationships, and phylogenies, which can represent absolute relationships. Both (*a*,*b*) are compatible with the same hypothesis of cladistic relationships; however, (*b*) represents a hypothesis of budding ancestry between (*x*,*y*), and anagenesis between (*y*,*z*). Phylogeny (*a*) implies gaps (represented by thin vertical lines subtending the thick vertical bars which reflect the stratigraphic ranges of taxa *x*–*z*) in the fossil record to accommodate the sister group relationship between lineages *x* and *y* + *z*, and between *y* and *z*. Meanwhile phylogeny (*b*) does not imply any gaps in the fossil record.

relationships can be discerned among morphospecies, as one can be traced morphing gradually into another based on morphological characters (e.g. [5]). However, this 'stratophenetic' approach is only suitable for groups with a rich fossil record (e.g. [6,7]). Indeed, even for those groups with the most complete fossil record, attempts to reconstruct time-calibrated phylogenies are confounded by gaps and the heterogeneous structure of the rock record [8]. For most groups, stratigraphic samples are smaller, there are stronger ecological and, therefore, geographical controls on distribution, and fossilization is less common. Combined with the non-uniform preservation of environments in the rock record [9], the fossil records of most groups are too non-uniform for stratophenetic approaches, and relative (rather than absolute ancestor–descendent) evolutionary relationships are the most that can be achieved with any degree of certainty. This fact is borne out by a general lack of correlation between the order of stratigraphic appearance and phylogenetic branching among fossil species, since sibling-lineages should exhibit contemporaneous first fossil occurrences [10,11] (figure 1*a*). Theoretical objections have been raised against the practice of identifying ancestors among fossil taxa since species do not beget other species but, rather, they emerge through differentiation among populations [12] and, thus, to identify an ancestral species is a category error—it effectively identifies the species as paraphyletic, only components among which can be considered as potential ancestors. And hypotheses of ancestry are also problematic because they rely upon negative evidence, i.e. the fact that putative ancestors preserve only plesiomorphic characters and lack autapomorphies that might distinguish them as distinct lineages [13,14].

Thus, for many palaeontologists, stratigraphic time has little role in phylogeny estimation, except in providing minimum ages to calibrate morphology-based cladograms to time, or in discriminating among multiple trees of equal likelihood or parsimony [14]. This is not to say that ancestors do not exist in the fossil record [15–17] and failure to accommodate fossil species will result in the perception of gaps in the fossil record where there are none [18–20]. This occurs because ancestral taxa are misrepresented as sibling-lineages of their descendants, resulting in the perception of a gap in the fossil record of the descendent, given the expectation that sibling-lineages diverge contemporaneously from their last shared ancestor [10] (figure 1). Parsimony and likelihood-based 'stratocladistic' methods have been developed that attempt to minimize perceived gaps in the fossil record, not least through the recognition of ancestor–descendent relationships among fossil morphospecies [21,22]. However, notwithstanding the likelihood of ancestors among the fossil species, the veracity of attempts to identify specific ancestral morphospecies remains questionable [13,23]. As such, stratocladistic methods may serve to conceal embarrassing gaps in the fossil record, making it appear a much better archive of evolutionary history than it really is.

Clearly, for the majority of clades, the fossil record alone is not sufficient to establish anything more than a minimum estimate for the age of living and fossil clades; only the molecular clock provides a means of approaching a true evolutionary timescale. Below, we review the history of development of molecular clock methodology, and the use of fossils and morphological data to calibrate the molecular clock. We discuss the potentials and challenges of modern Bayesian dating methods, which attempt to integrate different sources of information in one combined analysis, such as distance information in genetic sequences and time information in the fossil record. We highlight challenges confronting the latest methodological developments in divergence time estimation and show that these retread long-standing debates associated with phylogeny and timescale estimation in palaeontology, from which insights might be gained for the future development of molecular clock methodology.

2. The origin and early evolution of the molecular clock

The molecular clock hypothesis was conceived from the observation that the differences between homologous amino acid sequences from different mammal species is roughly proportional to their time of divergence [24,25]. If the time of divergence between any pair of species is known, such as based on the oldest fossil record from one of the pair of lineages, then the rate of molecular evolution can be inferred and used to date the timing of divergence between other species pairs. The molecular clock was widely employed to date species divergences in the 1990s when molecular sequence data first became available for diverse lineages. However, many of those early analyses produced extremely ancient divergence time estimates, such as a Mesoproterozoic origin of bilaterian animals [26], a Cryogenian origin of land plants [27], a deep Jurassic origin of flowering plants [28] and a deep Cretaceous origin of the ordinal level crown groups of birds and eutherian mammals [29]. These estimates challenged not only the veracity of the fossil record as an archive of evolutionary history, but also macroevolutionary hypotheses that had been based on fossil data, such as the end-Cretaceous mass extinction and its role in shaping modern biodiversity, which had effectively become philosophies in which those fossil data were interpreted.

Given the general acceptance among palaeontologists of the incompleteness of the fossil record, it might be imagined that their discipline would embrace the molecular clock

hypothesis with giddy enthusiasm and, indeed, palaeontologists were among its earliest adopters (e.g. [30,31]). However, the large mismatch between the early molecular date estimates and the fossil evidence of clade ages led many palaeontologists to reject these molecular estimates out of hand [26,28,32].

There are many reasons to expect clade ages to be older than their oldest fossil representatives. First, evidence of lineage separation cannot be manifest in the fossil record until one or both of the descendent species have acquired distinguishing anatomical characteristics that have the potential to be fossilized. By contrast, molecules evolve independently as lineages diverge. The difference between the time of lineage divergence and the age of the oldest fossil species can be even greater due to uneven species distributions and variation in the preservation of the sediments (facies) in those environments. For example, the oldest definitive records of terrestrial plant and animal lineages are approximately equivalent [33]. However, rather than an explosive radiation of terrestrial organisms, this correlated appearance more likely reflects the dearth of terrestrial sediments from which to sample fossils immediately prior to the middle Silurian [34]. There is also the challenge of correctly assigning derived species, because early diverging lineages are difficult to distinguish from their ancestors on anatomical grounds, and when derived characters are few in number it can be difficult to discern whether such characters are genuinely derived or evolved ancestrally before being lost in one of the derived lineages. This is complicated further by the incompleteness of fossil preservation, since corroborative anatomical characters are needed to distinguish between shared derived characters and convergences or parallelisms. For instance, the earliest fossil chondrichthyans are distinguished on the presence of vascular canals associated with the 'neck' of the attachment process of their scales. However, these fossils are limited to isolated scales [35,36]. How certain can we be that scale neck canals evolved only once, or that they are a primitive chondrichthyan character, or that they are not a shared primitive character lost in osteichthyans? Corroborative anatomical evidence would be useful to determine that these microremains belonged to a jawed vertebrate.

In combination, these factors result in significant differences between the timing of divergence and the age of the oldest fossil, but they cannot account for the scale of the mismatch implied by many early molecular clock studies. This is because gaps in the fossil record are largely predictable, based on the quality of the fossil record, how it varies between groups, and how fossil species' stratigraphic ranges may be influenced by secular variation in the preservation of facies in the rock record. For instance, palaeobiologists conduct gruesome decay experiments to discern the relative preservation of anatomical characters, and of taxa [37]. Further, knowledge of the sedimentary facies associations of fossil species can be exploited to predict probabilistically their occurrence through stratigraphic sequences [38,39] (figure 2). With suitable taphonomic controls [41], unfulfilled predictions of fossil stratigraphic occurrences can be interpreted as evidence for the absence of those fossil species in space and time.

Thus, despite the incomplete and non-uniform nature of the fossil record, it is safe to conclude that many of the great conflicts between the molecular time estimates and the fossil record are to a large extent due to the many limitations of the early clock dating studies [42]. These include the incorrect assumption of the strict clock or inadequate accommodation of the violation of the clock when dating deep divergences, the use of calibrations that ignore uncertainties in fossil evidence, as well as the use of secondary calibrations and substitution rates (that is, node ages and substitution rates estimated in previous molecular clock dating studies). While methods that effectively assume a strict clock continue to be employed and developed [43,44], producing unreasonably ancient divergence time estimates [45], they are hard to justify given the overwhelming evidence for violation of the clock among distant species. Nevertheless, in generating controversy, the early dating analyses were influential in shaking evolutionary biologists out of the view that evolutionary timescales could be read from the fossil record, and forced palaeontologists to review its utility and limits. They also prompted communities to reconsider their cherished hypotheses, such as the role of mass extinctions in shaping modern biodiversity. However, these timescales have not withstood the test of time, and they are now considered only in terms of their historical value in charting the history of development of this scientific method.

3. The modern molecular clock and the challenge of calibrating it to geological time

Mismatches between molecular clock estimates and clade ages based on the oldest fossil occurrences now rarely occur on the scale that they did in the 1990s. The main reason for this change is the development of analytical methods that can accommodate the violation of the clock as well as uncertainties in fossil calibrations. Although the earliest relaxed-clock methods were developed within a likelihood framework [46,47], more recent advances have been implemented within a Bayesian framework [48–54].

This is a straightforward application of Bayesian inference, in which the parameters of interest are the species divergence times ($\boldsymbol{t}$), the molecular evolutionary rates for branches on the tree ($\boldsymbol{r}$), the parameters in the substitution model and in the prior (θ), while the data are the sequence alignments at multiple gene loci (S). A Markov chain Monte Carlo algorithm is used to sample from the joint posterior

$$f(\theta, \boldsymbol{t}, \boldsymbol{r}|S) \propto f(\theta)\, f(\boldsymbol{t}|\theta)\, f(\boldsymbol{r}|\boldsymbol{t}, \theta) f(S|\boldsymbol{t}, \boldsymbol{r}, \theta), \tag{3.1}$$

where $f(\theta)$ is the prior for parameters, $f(\boldsymbol{t}|\theta)$ the time prior, specified using a branching process such as the birth–death sampling model [55,56], $f(\boldsymbol{r}|\boldsymbol{t}, \theta)$ is the prior for the rates, and $f(S|\boldsymbol{t}, \boldsymbol{r})$ is the likelihood for the sequence data. The times are shared among the multiple loci. In this formulation, the joint prior density of divergence times, $f(\boldsymbol{t}|\theta)$, incorporates fossil calibration information whenever it is available, with the distribution of ages of other nodes supplanted by the branching process such as the birth–death sampling model.

It may be important to note that the likelihood for the sequence data, $f(S|\boldsymbol{t}, \boldsymbol{r}, \theta)$, depends on the branch length, which is the expected number of changes on each branch, and is the product of the time duration for the branch and the rate for the branch. In other words, times and rates are confounded. A consequence of this confounding effect is that even if a huge amount of sequence data is analysed, the posterior of times and rates will remain sensitive to the prior on times and prior on rates [52,57,58]. Thus, having accurate fossil or temporal constraints is always important to a molecular clock dating analysis.

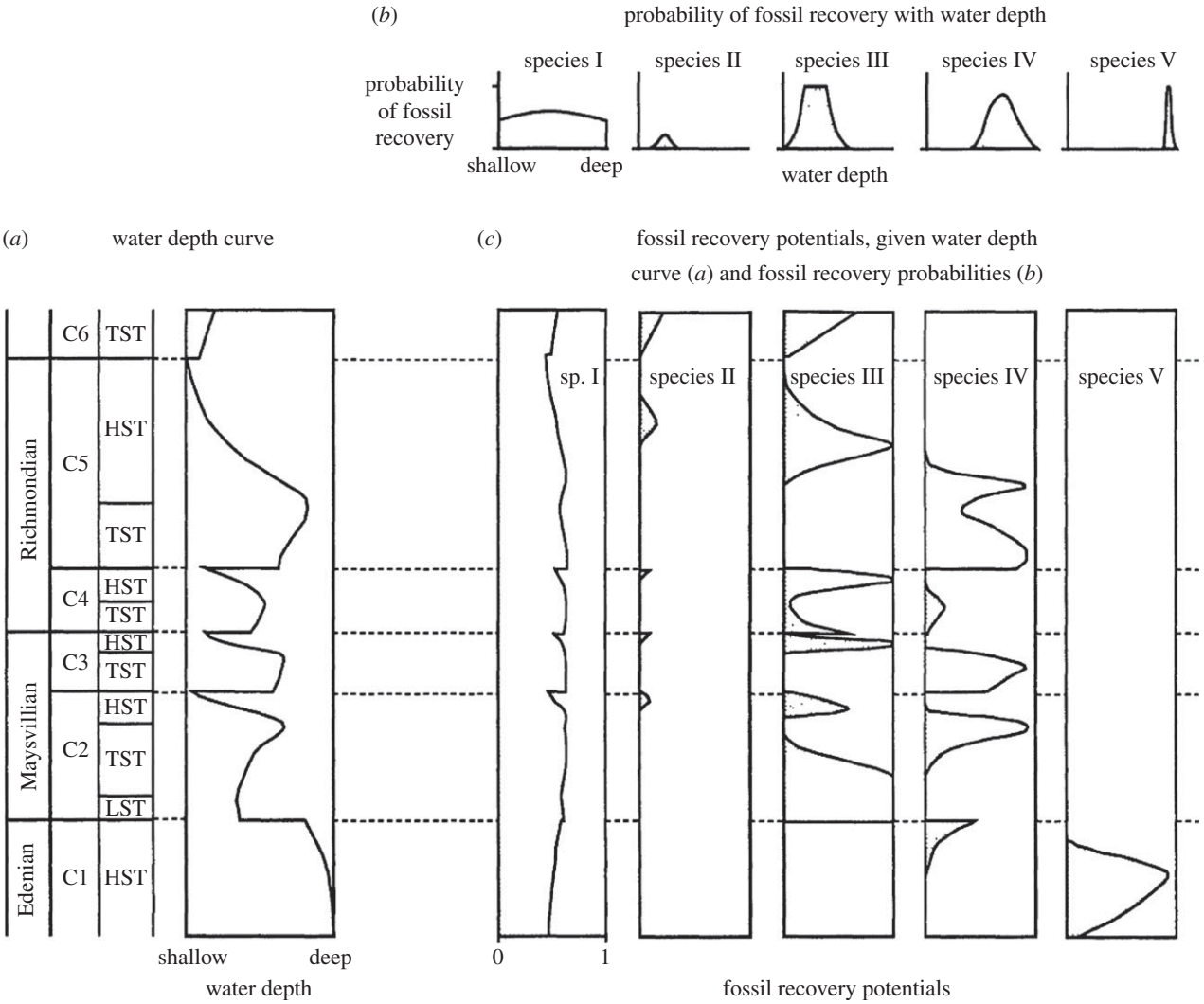

Figure 2. The predictable nature of fossil occurrences. Fossil stratigraphic occurrences are distinctly non-random, determined by the environmental controls on the distribution of the living organism and secular variation in the preservation of the environment in which the organism lived and died. Thus, fossil occurrences can be predicted based on knowledge of environmental limits of their distribution and characterization of environments and how they vary through stratigraphic sections and their global composites. (*a*) Inferred variance in water depth through a stratigraphic section; (*b*) probability of recovering a fossil based on its water depth tolerance and (*c*) fossil recovery potential given (*a*,*b*). Reproduced from Marshall [40] with the permission of the author and publisher.

In this article, we focus on the time prior $f(\boldsymbol{t}|\theta)$. Two principal approaches to calibrations have been implemented, sometimes referred to as node-dating and tip-dating; since both approaches to calibration are used to date nodes, it is preferable to refer to them as node-calibration and tip-calibration, respectively [59]. Below, we describe the different approaches and the motivations underpinning their development, before considering their relative merits and areas for their future development.

(a) Strategies to derive node-calibrations

Interpreting the fossil evidence to construct calibrations for molecular clock dating is a challenging task. The fossil record can directly inform the minimum ages of clades based on the age of their oldest fossil representative [60,61]. Establishing a maximum constraint is far more problematic since it relies on the interpretation of negative evidence—the absence of fossil evidence for a clade may be due to the vagaries of fossilization rather than simply because the clade had yet to evolve. Several pragmatic solutions have been proposed to establish maximum age constraints on clade ages. The simplest approach is to use a parametric distribution (such as the gamma, lognormal or the truncated Gaussian) to express the probability of the true divergence time relative to the minimum fossil-based age-constraint [52,62]. Without a statistical analysis of the fossil data, this approach inevitably involves some arbitrariness. While the precise parametric forms for node-calibrations (such as the gamma versus a pair of minimum and maximum bounds) were found to be unimportant in some studies (e.g. [63]), whether soft maximum bounds were included in a dating analysis has been found to have a dramatic effect on the posterior time estimates [64,65].

Another approach, phylogenetic bracketing, exploits the predictive nature of fossil occurrences based on their sedimentary facies associations and stratigraphic facies variation, to interpret the absence of fossil representatives of the clade ingroup [66,67]. It is important, however, that outgroup taphonomic control species are used, which have the same ecological and preservational characteristics of the ingroup, to discriminate absences that occur simply because the conditions required for fossilization were not met, or else because older sediments do not represent the environments that the organisms inhabited. Confidence intervals can be calculated to infer the true stratigraphic range of calibrating fossils, which

correspond to the minimum bounds for the true age of lineage divergence [40,68,69]. However, their utility in inferring the age of clades or lineage divergences is limited [70], and it is not useful for interpreting the fossil record of species known from single individuals or single stratigraphic levels.

Node-calibrations typically do not satisfy the requirement that ancestral species should be older than their descendants. If for no other reason, this occurs because the uncertainties associated with the age-priors on ancestor–descendent nodes frequently overlap. Specifying multi-dimensional priors in Bayesian inference is in general a challenging task, and in the case of specifying calibrations for multiple nodes on the tree, with some nodes being ancestral to others, this task is daunting without statistical analysis of the fossil data, as in [63,71]. Current Bayesian programs multiply the densities for the calibration nodes first, and then when the program is run, node ages that violate the biological constraint are disallowed, effectively truncating the joint prior distribution of the ages for all calibration nodes, e.g. [52]. Thus, the effective prior used by the computer program may be quite different from the user-specified priors [64,72]. Different programs may use different strategies to apply this truncation, leading to different effective priors even if the user initially specifies exactly the same calibrations [64,65,72]. Therefore, it is important that the effective priors are scrutinized (by running the analysis without sequence data) to ensure that they are compatible with the palaeontological and phylogenetic evidence on which the specified node-calibrations were originally based.

The most sophisticated among this class of methods is the probabilistic modelling of fossil preservation and discovery, and statistical analysis of the absence and presence data of fossil species in different rock strata by Tavare *et al*. [71] and Wilkinson *et al*. [63]. Bayesian analysis of the fossil data alone (the presence and absence of fossil species in the different rock strata) produces a posterior distribution of node ages, which can then be used as the prior of times in the subsequent molecular dating analysis using the sequence data. This is based on a branching-process model that describes speciation and extinction, as well as fossil preservation and discovery, assigning priors on the speciation rate, extinction rate and sampling intensity. This approach is attractive as it makes use of information in all the pertinent fossils, in contrast to other node-calibration methods that use only the oldest fossil that constrains directly the age of the extant clade.

(b) Integrative analysis of fossil and sequence data

The fossilized birth–death (FBD) model [73] is a similar to the statistical methods developed by Tavaré and co-workers [63,71], except that it analyses the fossil data jointly with the molecular sequence data. It attempts to describe both the distribution of fossils and the lineage divergence times within a clade based on an integrated diversification–fossilization model. In its original incarnation [73], fossil species are assigned to clades with varying degrees of precision, and clade ages are inferred in a conventional Bayesian molecular clock analysis. The model requires only priors on the speciation and extinction rates, the fossil recovery rate and the proportion of extant species sampled. It assumes constant speciation and extinction rates, initiating on a single lineage and identifying fossil species according to a Poisson process, and extant species at a given probability. This tree is stripped of unsampled extant and extinct species, yielding the reconstructed phylogeny of living and fossil species. However, the FBD can integrate over the uncertainty of the phylogenetic position of the fossil species and the timing of their divergence from extant lineages, including the possibility that fossil species diverged at a time equal to their age, i.e. that the fossil species is effectively ancestral to the extant lineage. This unresolved FBD model does not use any information from morphological characters or measurements for the fossil species other than its geologic age—represented by a time point sampled from its stratigraphic range and attendant age uncertainty [73]. This has both positive and negative implications. In that the model requires only qualified phylogenetic affinity and fossil age, the unresolved FBD facilitates the integration of all pertinent fossil information. However, the model always requires certainty in the phylogenetic affinity of fossil species among their extant relatives, and the impact of phylogenetic precision will propagate to the resulting divergence time estimates. Since the FBD integrates over uncertainty associated with the phylogenetic affinity and the timing of lineage divergence of extinct species, it provides insight only into the antiquity of extant clades. Nevertheless, the FBD model is an attractive objective method for deriving estimates of clade ages from phylogenetic trees and palaeontological data. Importantly, it gets around the inconsistency between the specified and effectives time priors that derive from conventional node-calibration.

(c) Tip-calibration and the joint analysis of molecular and morphological data

The desire to overcome inconsistency between specified and effective node-calibrations is one of the motivations behind the development of alternative approaches to calibrating molecular clocks through the integration of morphological data and tip-calibration methods that include fossil species as terminal taxa among their extant relatives [16,74] (figure 3). Tip-calibration is often considered synonymous with the so-called total-evidence dating approach that also facilitates the simultaneous estimation of time and topology [16]. However, because these two approaches can be employed separately, we will first consider tip-calibration before going on to appraise total-evidence dating.

Tip-calibration is achieved by analysing morphological data for both living and fossil species under a model of morphological character evolution and molecular sequence data for living species [16,74]. The formulation is similar to equation (3.1), with the differences that we also have morphological data for fossil and modern species (M), besides the sequence data for modern species (S) and that there are two sets of rates, r_S for sequences and r_M for morphology. The joint posterior is then

$$f(\theta, \boldsymbol{t}, r_S, r_M|S, M) \propto f(\theta)\, f(\boldsymbol{t}|\theta)\, f(r_S, r_M|\boldsymbol{t}, \theta) f(S|\boldsymbol{t}, r_S, \theta) f(M|\boldsymbol{t}, r_M, \theta). \quad (3.2)$$

Here $f(r_S, r_X|\boldsymbol{t}, \theta)$ is the prior for the rates, while $f(M|\boldsymbol{t}, r_M)$ is the likelihood for morphological data. The times are shared between the molecular and morphological data. In the simple case, similar rate-drift models (such as the geometric Brownian motion model) can be used to describe variable morphological rates among lineages as in the case for sequences.

Tip-calibration is theoretically attractive as it involves the simultaneous and, thus, coherent analysis of morphological

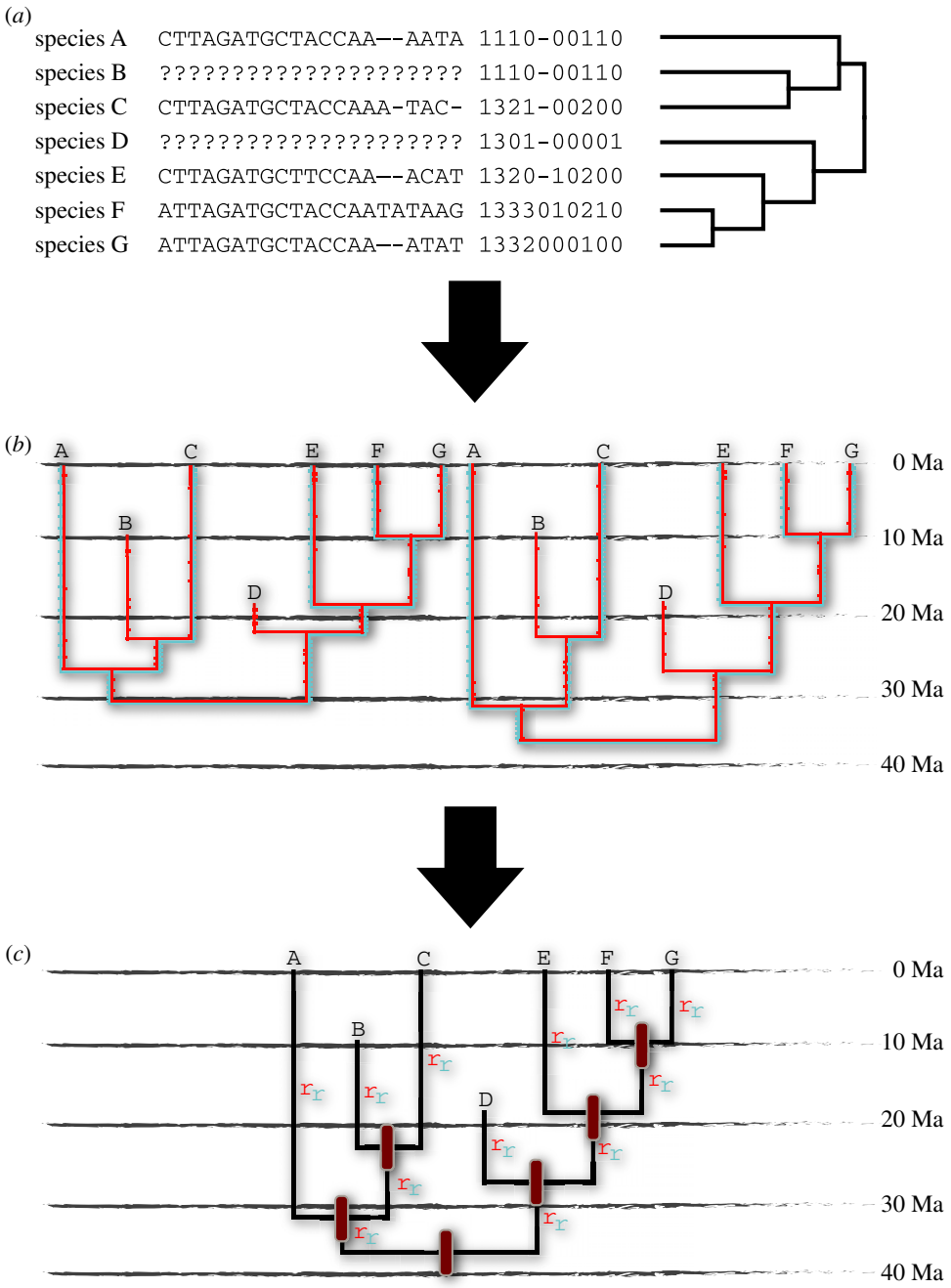

Figure 3. Tip-calibration relies upon a molecular sequence alignment from living species, a morphological character set for living and fossil species, and a prior topology (*a*); total-evidence dating co-estimates topology and timescale. Branch lengths are estimated in a Bayesian MCMC approach based on both data types for living lineages and based on morphological data alone for the extinct lineages; these are calibrated to time based on the age of the fossil species (*b*). The divergence time estimates and inferred rates of molecular and morphological evolution are based on a consensus of the MCMC analysis (*c*). (Online version in colour.)

data for both fossil and extant species and sequence data from modern species. It also integrates the time information from fossil species directly, rather than indirectly constraining the ages of clades, as in node-calibration (figure 3). Branch lengths of extant lineages are estimated based on morphological and molecular data while those of extinct lineages are estimated based on morphological data alone. Ultimately, these are calibrated to time based on the age of the fossil taxa, integrating stratigraphic age range and its attendant errors, as appropriate. It provides a means of estimating the ages of all clades, not merely those with living descendants. Indeed, the approach can be applied to fossil data alone, without sequence data, as has been done to date the timing of divergence of avian and non-avian dinosaurs [75,76], crown-Aves [77] and placental mammals [78].

(d) Total-evidence dating

The importance of fossil species in informing the relationships of extant species has long been emphasized [79,80]. The introduction of tip-calibration and the morphological clock has facilitated the development of methods for the co-estimation of phylogenetic relationships and their absolute timescales [16,74]. Thus, it is possible to integrate the phylogenetic uncertainty of living and fossil taxa in a manner that is very difficult to accommodate using node-calibrations that must make at least minimal assumptions concerning the phylogenetic relationships of living lineages and the place of fossil taxa among them. Furthermore, implementations of the total-evidence dating approach make it possible to use the age of fossil species to inform their phylogenetic position, following the expectation that fossil representatives of early

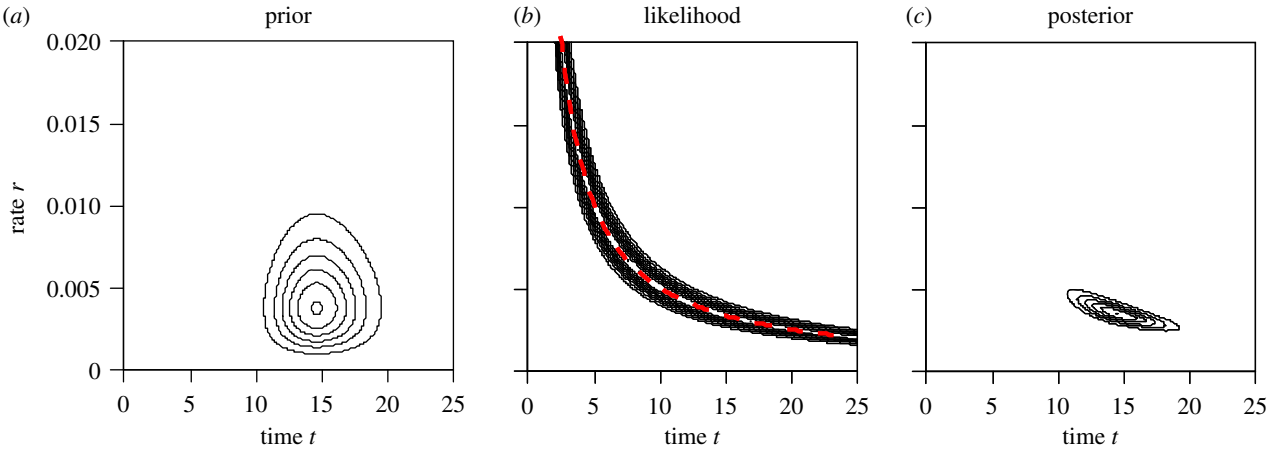

Figure 4. (*a*) Estimation of the absolute time (t) and rate (r) using the human and orangutan 12S rRNA genes from the mitochondrial genome. The data are summarized as $x = 90$ differences out of $n = 948$ aligned sites. The likelihood, calculated under the JC69 substitution model [86], depends on the distance $d = 2tr$ only, but not t and r individually. The maximum-likelihood estimate of d under the JC69 model is 0.1015, with the 95% confidence (likelihood) interval to be (0.0817, 0.1245). All points on the red dashed line in (*b*) correspond to the same likelihood value and are maximum-likelihood estimates of t and r. To generate the posterior of t and r, we assign the prior $t \sim G(40, 40/15)$, with the prior mean to be 15 Myr and the 95% equal-tail interval to be (10.7, 20.0) Myr, and the rate prior $r \sim G(4, 800)$, with the mean to be 0.005 substitution per million years and the 95% interval to be (0.14, 1.10). (*c*) Relatively, our prior knowledge of the rate is less certain than that for time. Note that to obtain sensible posterior time estimates, it is important to constrain the time from both below and above in the prior (in this case, the time is weakly constrained to be in the range 10–20 Myr). (Online version in colour.)

diverging lineages will be older than fossil representatives of lineages that diverge later in time. Indeed, even increasing uncertainty in the age of a fossil can lead to a changed phylogenetic hypothesis using total-evidence dating [59].

Extensions of the total-evidence dating method allow fossil species to be accommodated as direct ancestors of extant species, as well as the implementation of more flexible branching-process models such as the FBD model as well as those that allow diversified sampling of extant species [81,82]. This has the impact of diminishing uncertainty in the age of clades, overcoming the tendency of tip-calibration to yield ancient clade age estimates. It achieves this directly, by accommodating ancestral species that, were they identified as distinct lineages, would inflate branch lengths, the age of clades and, consequently, the extent of perceived gaps in the fossil record [81].

4. Discussion

Molecular clock methodology is currently undergoing such rapid development that it can be difficult to discriminate which methodological approach best suits the problem at hand. Indeed, the introduction of tip-calibration and the co-estimation of phylogenies and their timescales could be interpreted as replacements for conventional node-calibration and sequential analysis of phylogeny and timescales. There are strong arguments that favour the integration of all relevant lines of evidence, and their simultaneous analysis to derive time-calibrated phylogenies. Not least, these include the unchallengeable view that morphological and molecular data, living and fossil species, are all a consequence of the same evolutionary process. Meanwhile, sequential analysis of a phylogeny of extant taxa usually employs only molecular data; node-calibrations for the component clades are derived based on the phylogenetic position of fossil species within this scheme, based on morphological data. These are transformed in the assembly of the joint time prior and integrated into a molecular clock analysis that otherwise uses only molecular sequence data and the original molecular phylogeny. Implicitly, at least, this approach makes the assumption of independence among the different models and data [82]. However, despite their theoretical appeal, tip-calibration and total-evidence dating methods face a number of challenges, and exhibit strong parallels to the challenges that have long confronted the construction of time-calibrated phylogenies in palaeontology [83–85]. Many of those challenges have not yet been systematically explored.

(a) Tip-calibrations are very sensitive to the branching process or the prior for times

The most serious problem facing tip-calibration may be the extreme sensitivity of the posterior time estimates to the prior of divergence times specified by the branching process. Because the sequence data provide information about distances only, resolution of the sequence distance into absolute time and rate relies entirely on the priors on time and rate (figure 4). Most tip-calibration methods require a bound or prior on the age of the root for extant species, e.g. [16], but no calibrations are applied on the ages of other internal nodes. Thus, node ages are bounded by the ages of the fossil tips, because ancestral nodes cannot be younger than their descendent fossil tips, while there is otherwise effectively no constraint on the ages of clades except for the prior on the root age. In other words, there are multiple forces pushing up the node ages, but almost no force pushing them down. It is left to the divergence time prior or the branching-process model to keep the node ages on the tree within reasonable bounds, and that may prove to be too much burden on the time prior. A dozen or so initial studies applying the total-evidence dating approach have produced ancient time estimates, older even than those derived from the use of node-calibrations [59]. This is remedied by incorporating the FBD model into total-evidence dating, replacing the original uniform tree prior and making use of morphological character data in resolving the affinity of fossil taxa [81,82]. However, details of the FBD prior, such as the assumed sampling regime, can have a strong influence on divergence time estimates [81].

For example, when species sampling is assumed to be complete or uniform in the branching process (the BDS model), Ronquist *et al*. [16] dated crown Hymenoptera to

the late Carboniferous (309 Ma). Changing the prior model to account for diversified sampling dated the crown radiation to the Permo-Triassic (252 Ma), 57 million years younger [81]. Condamine *et al*. [87] provide a similar cautionary tale [87] in dating the origin of living cycads based on node-calibrations. While diversified sampling is arguably a more realistic prior model than uniform sampling, it may be naive to expect it to capture the main features of the complicated process of species sampling. Furthermore, there is very little information in the data to estimate the speciation rate, the extinction rate and the sampling rate in a birth–death sampling model when those rates are changing over time [88,89]. The sensitivity of the posterior time estimates to the prior branching-process model is troubling, but the effect is a general feature of the model formulation, not limited to particular datasets. Diversified sampling causes the tree to become 'bush-like', while a low fossilization rate and high extinction rate push fossils towards extinct side branches [81]. Because sequence data are not informative about those rates, the assumptions in the prior model translate directly into the posterior time estimates. Thus, we stress the importance of assessing the robustness of posterior time estimates to the prior in Bayesian relaxed-clock dating analysis, especially the prior on times (or the branching process) and the prior on substitution rates for multiple loci. In combined analysis of molecular and morphological data, it is important to assess the impact of the morphological model.

A pragmatic solution may be to combine tip and node-calibrations in the same analysis [59,90], although there are challenges to this approach. First, it would be inappropriate to use the same fossil data to inform both tip- and node-calibrations in the same analysis. Second, the prior density for divergence times under the birth–death sampling model, when some node ages are constrained by node-calibrations, may not be tractable analytically. Third, node-calibrations are contingent upon a prior hypothesis of relationships and so they are incompatible with attempts to co-estimate time and topology, as in the classic total-evidence dating approach. These concerns can, however, be overcome by not using the same fossil data to inform tip- and node-calibrations, and facilitating coestimation with a minimal backbone topology constraint compatible with the few nodes that are calibrated.

(b) A morphological clock?

In a tip-calibrated analysis, the information about absolute divergence times ultimately comes from the fossil record and the assumed clock-like evolution of morphological characters. Thus, the many weaknesses of morphological data identified in the molecules-versus-morphology debate of the early 1990s remain important to the use of morphology to estimate branch lengths and to date molecular trees. For example, phenotypes are influenced by the environment as well as by genes, and morphological characters may undergo convergent evolution in disparate lineages. Models of morphological evolution have undergone very little development and the most widely used, *Mk* model [91], is a generalization of the JC69 model of molecular sequence evolution, the inherent assumptions underpinning which are not entirely appropriate for the analysis of morphological data, including independence among sites [59]. Furthermore, the evolution of morphological characters is not clock-like, as a rule, even among closely related species [92], and so the existence of a morphological clock, no matter how relaxed, remains questionable. The non-clock-like behaviour of morphological evolution may have far greater impact on divergence time estimation than on phylogeny reconstruction.

Finally, while it is a truism to observe that morphological clock analyses are limited by the availability of morphological data, this is usually considered in terms of fossil taxa, but living taxa are particularly poorly characterized in terms of their anatomy [93]. Missing morphological data are also non-uniformly distributed, as a consequence of the work of organ- and taxon-specialists of living organisms, and of decay in fossils where non-biomineralized tissues and organs and tissues are not preserved except in the most exceptional circumstances. This can lead to a systematic bias in phylogeny estimation where fossil species lacking derived characters (as an artefact of incomplete fossilization) are resolved as less-derived phylogenetically than they really are [83,85]. This has significant implications for molecular clock calibration in general, including the formulation of node-calibrations [84]. However, this bias probably has greatest impact on tip-calibration because of its influence on topology and branch length estimation [59].

(c) Co-estimation versus sequential analysis of topology and time

Co-estimation of phylogenies and their timescales using tip-calibration promises to provide a basis for establishing a correct timescale for extinct, not merely extant, clades. However, in reality, this promise is not commonly realized when an appreciable number of fossil species are included since the phylogenetic positions of the fossil taxa are not usually well resolved (e.g. [16,81,94,95]). Rather, the fossil species serve largely to inform the age of living clades, as in node-calibration. This is likely because the (invariably incomplete) phenotypic character data are insufficiently informative on the phylogenetic position of the fossil species relative to the living species, the phylogenetic position of which are informed by both molecular and morphological data. Clearly, greater insight may be obtained into the age of extinct clades, and the timing and rates of character evolution, by instead dating a more fully resolved and separately justified tree within a sequential analysis of phylogeny and timescale.

Notwithstanding the merits of co-estimating time and topology, current implementations of the total-evidence dating approach use the ages of fossil species to inform their phylogenetic position (e.g. [16]). Indeed, even increasing uncertainty in the age of a fossil can lead to changes to the inferred phylogeny [59]. However, the underlying expectation, that the stratigraphic order of fossil species reflects their phylogenetic branching order, is contingent on the completeness of the fossil record of a clade [96] or at least its sampling within an analysis. Synoptic analyses have demonstrated significant inconsistency in stratigraphic and phylogenetic branching order, even in groups that are considered to have a rich fossil record [11,97]. Development of methods for estimating topology using fossil age information could benefit from the long-standing palaeontological debate on the topic. Indeed, methods have already been developed to rationalize phenotypic character evolution, topology and the stratigraphic range of fossil species (e.g. [22,98,99]). However, such methods must be developed to accommodate controls on the distribution of fossil species, which include their relative fossilization

potential, the impact of non-uniform preservation of phenotypic characters [83,84], as well as secular variation in the sedimentary facies preserved in the stratigraphic record [9,41]. The beginnings of such an approach, controlling for the non-uniform nature of fossil distribution, are present in the FBD model [73,81]. However, codifying these priors for the model will be a challenge, especially for analyses that must consider the global palaeontological and, therefore, stratigraphic record.

(d) In search of ancestors

Palaeontologists have long debated the possibility of inferring direct ancestors, as well as the merits and demerits of representing such absolute relationships among phylogenetic hypotheses (e.g. [100–103]). As we have discussed, there can be little doubt that ancestors occur in the fossil record, both as ancestors of other extinct species, and ancestors of living and fossil lineages [15,17]. The challenge has always been to derive an acceptable method for reconciling the homology statements that characters represent, with the potential phylogenetic informativeness of fossil ages. Much of this debate has been considered within the parsimony framework of phylogenetic inference, but the application of likelihood-based models of character evolution that better accommodate homoplasy, provides a more appropriate framework in which to develop this debate further. The FBD model provides the basis for an objective approach to the identification of ancestors in the fossil record and, thus, for overcoming the traditional criticism of cladistic approaches to palaeontology—that in failing to observe absolute relationships, they artefactually inflate the perception of the gaps in the fossil record [18–20]. The failure to consider the possibility that fossil species might even be indirect ancestors of living lineages similarly serves to distance fossil minima from perceptions of clade age and so the recognition of fossil ancestors will serve to bring divergence time estimates into a closer approximation of fossil evidence. However, in the development of such methods, it is important that they control for non-uniform fossil preservations in stratigraphic sequence [9,39], as well as non-uniform losses of anatomical data in the process of decay and preservation that led to fossilization [83–85]. Many fossil species are compatible with ancestors, i.e. they do not exhibit autapomorphies [13], but only as an artefact of incomplete preservation. Thus, the accommodation of hypotheses of ancestry might lead to molecular clock estimates that achieve an entente with fossil evidence, but only by effectively concealing real gaps in the fossil record that are otherwise indicated by the existence of fossils that are siblings, rather than ancestors of living lineages.

5. Concluding remarks

Molecular clock methodology is undergoing a period of development unparalleled in the half century since the molecular clock hypothesis was first formulated. This has been brought about principally by the introduction of Bayesian inference, which provides a powerful framework for integrating different sources of information, with the uncertainties appropriately accommodated. In our perception at least, methods are diversifying, rather than new methods superseding established approaches. Indeed, there is now a broader palate of methods and approaches to divergence time estimation than there has been at any time in the past and these may be assembled in a combination that best suits the testing of the hypothesis at hand. Many of these components, like tip-calibration, the morphological clock and the FBD model, are at an early stage of development and current applications may not stand the test of time. Nevertheless, this palate of tools is already being assembled into a toolkit (e.g. [81]) that has the promise of developing into a fully integrative framework for calibrating the Tree of Life to geologic time, including all of its branches, living and dead. Many of the challenges that confront the development of molecular clock methodology exhibit striking parallels to long-standing debates in palaeontology, such as the role of time in topology estimation, the efficacy of attempts to identify direct and indirect ancestors among fossil taxa, as well as the impact on topology estimation of the non-uniform stratgraphic distribution and preservation of fossil species. Thus, methodological advances may be more readily achieved by learning from, rather than rehearsing, these debates. Rather than perpetuating controversy between molecular systematists and palaeontologists, in its middle age the molecular clock hypothesis looks set to serve as a nexus, dissolving the artificial barriers between these disciplines and their perceptions of evolutionary history.

Data accessibility. There are no data associated with this study.

Authors' contributions. All authors made substantial contributions to (i) conception and design, or acquisition of data, or analysis and interpretation of data and (ii) drafting the article or revising it critically for important intellectual content; gave final approval of the version to be published, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Competing interests. We have no competing interests.

Funding. This research was funded by the Natural Environment Research Council (NE/N003438/1), the Biotechnology and Biological Sciences Research Council (BB/J009709/1 and BB/N000919/1) and The Royal Society (Wolfson Research Merit Award to P.C.J.D.).

Acknowledgements. We would like to thank our colleagues Joe O'Reilly and Mario dos Reis for discussion, as well as all of the attendees and contributors to the workshop on dating species divergences.

References

1. Pietsch TW. 2012 *Trees of life: a visual history of evolution*. Baltimore, MD: Johns Hopkins.
2. Bowler PJ. 1989 *Evolution—the history of an idea*. Berkeley, CA: University of California Press.
3. Lewis C. 2000 *The dating game: one man's search for the age of the Earth*. Cambridge, UK: Cambridge University Press.
4. Simpson GG. 1944 *Tempo and mode in evolution*. New York, NY: Columbia University Press.
5. Pearson PN, Ezard THG. 2013 Evolution and speciation in the Eocene planktonic foraminifer *Turborotalia*. *Paleobiology* **40**, 130–143. (doi:10.6084/m9.figshare.793840)
6. Wickström LM, Donoghue PCJ. 2005 Cladograms, phylogenies and the veracity of the conodont fossil record. *Spec. Pap. Palaeontol.* **73**, 185–218.
7. Aze T, Ezard THG, Purvis A, Coxall HK, Stewart DRM, Wade BS, Pearson PN. 2011 A phylogeny of Cenozoic macroperforate planktonic foraminifera

from fossil data. *Biol. Rev.* **86**, 900–927. (doi:10.1111/j.1469-185X.2011.00178.x)

8. Lloyd GT, Pearson PN, Young JR, Smith AB. 2012 Sampling bias and the fossil record of planktonic foraminifera on land and in the deep sea. *Paleobiology* **38**, 569–584. (doi:10.1666/11041.1)
9. Holland SM. 2016 The non-uniformity of fossil preservation. *Phil. Trans. R. Soc. B* **371**, 20150130. (doi:10.1098/rstb.2015.0130)
10. Norell MA, Novacek MJ. 1992 The fossil record and evolution: comparing cladistic and paleontologic evidence for vertebrate history. *Science* **255**, 1690–1693. (doi:10.1126/science.255.5052.1690)
11. Benton MJ, Wills MA, Hitchin R. 2000 Quality of the fossil record through time. *Nature* **403**, 534–537. (doi:10.1038/35000558)
12. Nelson G. 1989 Species and taxa: systematics and evolution. In *Speciation and its consequences* (eds D Otte, JA Endler), pp. 60–81. Sunderland, MA: Sinauer.
13. Engelmann GF, Wiley EO. 1977 The place of ancestor-descendant relationships in phylogeny reconstruction. *Syst. Zool.* **26**, 1–11. (doi:10.2307/2412861)
14. Smith AB. 1994 *Systematics and the fossil record: documenting evolutionary patterns*. Oxford, UK: Blackwell Scientific.
15. Paul CRC. 1992 The recognition of ancestors. *Hist. Biol.* **6**, 239–250. (doi:10.1080/10292389209380433)
16. Ronquist F, Klopfstein S, Vilhelmsen L, Schulmeister S, Murray DL, Rasnitsyn AP. 2012 A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Syst. Biol.* **61**, 973–999. (doi:10.1093/sysbio/sys058)
17. Foote M. 1996 On the probability of ancestors in the fossil record. *Paleobiology* **22**, 141–151.
18. Bretsky SS. 1979 Recognition of ancestor–descendant relationships in invertebrate paleontology. In *Phylogenetic analysis and paleontology* (eds N Eldredge, J Cracraft), pp. 113–163. New York, NY: Columbia University Press.
19. Wagner PJ. 1998 Phylogenetic analysis and the quality of the fossil record. In *The adequacy of the fossil record* (eds SK Donovan, CRC Paul), pp. 165–187. Chichester, UK: John Wiley & Sons.
20. Paul CRC. 2003 Ghost ranges. In *Telling the evolutionary time: molecular clocks and the fossil record* (eds PCJ Donoghue, MP Smith), pp. 91–106. London, UK: Taylor & Francis.
21. Fisher DC. 1992 Stratigraphic parsimony. In *MacClade version 3 Analaysis of phylogeny and character evolution* (eds WP Maddison, DR Maddison), pp. 124–129. Sunderland, MA: Sinaeur Associates.
22. Wagner P. 1998 A likelihood approach for evaluating estimates of phylogenetic relationship among fossil taxa. *Paleobiology* **24**, 430–449.
23. Forey PL. 2004 Systematics and paleontology. In *Milestones in systematics* (eds DM Williams, PL Forey), pp. 149–180. Boca Raton, FL: CRC Press.
24. Zuckerkandl E, Pauling L. 1965 Evolutionary divergence and convergence in proteins. In *Evolving genes and proteins* (eds V Bryson, HJ Vogel), pp. 97–166. New York, NY: Academic Press.
25. Zuckerkandl E, Pauling L. 1962 Molecular disease, evolution and genetic heterogeneity. In *Horizons in biochemistry* (eds M Kasha, B Pullman), pp. 189–225. New York, NY: Academic Press.
26. Wray GA, Levinton JS, Shapiro LH. 1996 Molecular evidence for deep precambrian divergences among metazoan phyla. *Science* **274**, 568–573. (doi:10.1126/science.274.5287.568)
27. Heckman DS, Geiser DM, Eidell BR, Stauffer RL, Kardos NL, Hedges SB. 2001 Molecular evidence for the early colonization of land by fungi and plants. *Science* **293**, 1129–1133. (doi:10.1126/science.1061457)
28. Wikström N, Savolainen V, Chase MW. 2001 Evolution of the angiosperms: calibrating the family tree. *Proc. R. Soc. Lond. B* **268**, 2211–2220. (doi:10.1098/rspb.2001.1782)
29. Hedges SB, Parker PH, Sibley CG, Kumar S. 1996 Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381**, 226–229. (doi:10.1038/381226a0)
30. Runnegar B. 1982 A molecular-clock date for the origin of the animal phyla. *Lethaia* **15**, 199–205. (doi:10.1111/j.1502-3931.1982.tb00645.x)
31. Brown RH, Richardson M, Boulter D, Ramshaw JAM, Jefferies RPS. 1972 The amino acid sequence of cytochrome c from *Helix aspera* Müeller (garden snail). *Biochem. J.* **128**, 971–974. (doi:10.1042/bj1280971)
32. Donoghue PCJ, Smith MP. 2003 *Telling the evolutionary time: molecular clocks and the fossil record*. London, UK: CRC Press.
33. Jeram AJ, Selden PA, Edwards D. 1990 Land animals in the Silurian: arachnids and myriapods from Shropshire, England. *Science* **250**, 658–661. (doi:10.1126/science.250.4981.658)
34. Kenrick P, Wellman CH, Schneider H, Edgecombe GD. 2012 A timeline for terrestrialization: consequences for the carbon cycle in the Palaeozoic. *Phil. Trans. R. Soc. B* **367**, 519–536. (doi:10.1098/rstb.2011.0271)
35. Sansom IJ, Smith MP, Smith MM. 1996 Scales of thelodont and shark-like fishes from the Ordovician. *Nature* **379**, 628–630. (doi:10.1038/379628a0)
36. Andreev PS, Coates MI, Shelton RM, Cooper PR, Smith MP, Sansom IJ, Johanson Z. 2015 Upper Ordovician chondrichthyan-like scales from North America. *Palaeontology* **58**, 691–704. (doi:10.1111/pala.12167)
37. Briggs DEG, McMahon S. 2015 The role of experiments in investigating the taphonomy of exceptional preservation. *Palaeontology* **59**, 1–11. (doi:10.1111/pala.12219)
38. Holland SM. 1995 The stratigraphic distribution of fossils. *Paleobiology* **21**, 92–109.
39. Patzkowsky ME, Holland SM. 2012 *Stratigraphic paleobiology: understanding the distribution of fossil taxa in time and space*. Chicago, IL: Chicago University Press.
40. Marshall CR. 1997 Confidence intervals on stratigraphic ranges with nonrandom distributions of fossil horizons. *Paleobiology* **23**, 165–173.
41. Holland SM. 2000 The quality of the fossil record: a sequence stratigraphic perspective. *Paleobiology* **26**, 148–168. (doi:10.1666/0094-8373(2000)26[148:TQOTFR]2.0.CO;2)
42. Graur D, Martin W. 2004 Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends Genet.* **20**, 80–86. (doi:10.1016/j.tig.2003.12.003)
43. Tamura K, Ursula Battistuzzi F, Billing-Ross P, Murillo O, Filipski A, Kumar S. 2012 Estimating divergence times in large molecular phylogenies. *Proc. Natl Acad. Sci. USA* **109**, 19 333–19 338. (doi:10.1073/pnas.1213199109)
44. Kumar S, Hedges SB. 2016 Advances in time estimation methods for molecular data. *Mol. Biol. Evol.* **33**, 863–869. (doi:10.1093/molbev/msw026)
45. Battistuzzi FU, 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. *Mol. Biol. Evol.* **32**, 1907–1912. (doi:10.1093/molbev/msv075)
46. Sanderson MJ. 1997 A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.* **14**, 1218–1231. (doi:10.1093/oxfordjournals.molbev.a025731)
47. Yang Z, Yoder AD. 2003 Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene loci and calibration points, with application to a radiation of cute-looking mouse lemur species. *Syst. Biol.* **52**, 705–716. (doi:10.1080/10635150390235557)
48. Thorne JL, Kishino H, Painter IS. 1998 Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* **15**, 1647–1657. (doi:10.1093/oxfordjournals.molbev.a025892)
49. Kishino H, Thorne JL, Bruno WJ. 2001 Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Mol. Biol. Evol.* **18**, 352–361. (doi:10.1093/oxfordjournals.molbev.a003811)
50. Thorne JL, Kishino H. 2002 Divergence time and evolutionary rate estimation with multilocus data. *Syst. Biol.* **51**, 689–702. (doi:10.1080/10635150290102456)
51. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006 Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88. (doi:10.1371/journal.pbio.0040088)
52. Yang Z, Rannala B. 2006 Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* **23**, 212–226. (doi:10.1093/molbev/msj024)
53. Lartillot N, Lepage T, Blanquart S. 2009 PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288. (doi:10.1093/bioinformatics/btp368)
54. Ronquist F *et al.* 2012 MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542. (doi:10.1093/sysbio/sys029)
55. Stadler T. 2010 Sampling-through-time in birth–death trees. *J. Theoret. Biol.* **267**, 396–404. (doi:10.1016/j.jtbi.2010.09.010)
56. Yang Z, Rannala B. 1997 Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. *Mol. Biol. Evol.* **14**, 717–724. (doi:10.1093/oxfordjournals.molbev.a025811)

57. Rannala B, Yang ZH. 2007 Inferring speciation times under an episodic molecular clock. *Syst. Biol.* **56**, 453–466. (doi:10.1080/10635150701420643)
58. Dos Reis M, Yang Z. 2013 The unbearable uncertainty of Bayesian divergence time estimation. *J. Syst. Evol.* **51**, 30–43. (doi:10.1111/j.1759-6831.2012.00236.x)
59. O'Reilly J, dos Reis M, Donoghue PCJ. 2015 Dating tips for divergence time estimation. *Trends Genet.* **31**, 637–650. (doi:10.1016/j.tig.2015.08.001)
60. Reisz RR, Muller J. 2004 Molecular timescales and the fossil record: a paleontological perspective. *Trends Genet.* **20**, 237–241. (doi:10.1016/j.tig.2004.03.007)
61. Hedges SB, Kumar S. 2004 Precision of molecular time estimates. *Trends Genet.* **20**, 242–247. (doi:10.1016/j.tig.2004.03.004)
62. Ho SYW, Phillips MJ. 2009 Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Syst. Biol.* **58**, 367–380. (doi:10.1093/sysbio/syp035)
63. Wilkinson RD, Steiper ME, Soligo C, Martin RD, Yang Z, Tavaré S. 2011 Dating primate divergences through an integrated analysis of palaeontological and molecular data. *Syst. Biol.* **60**, 16–31. (doi:10.1093/sysbio/syq054)
64. Warnock RCM, Yang Z, Donoghue PCJ. 2012 Exploring uncertainty in the calibration of the molecular clock. *Biol. Lett.* **8**, 156–159. (doi:10.1098/rsbl.2011.0710)
65. Warnock RCM, Parham JFJ, Joyce WG, Lyson TR, Donoghue PCJ. 2015 Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. *Proc. R. Soc. B* **282**, 20141013. (doi:10.1098/rspb.2014.1013)
66. Donoghue PCJ, Benton MJ. 2007 Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends Ecol. Evol.* **22**, 424–431. (doi:10.1016/j.tree.2007.05.005)
67. Müller J, Reisz RR. 2005 Four well-constrained calibration points from the vertebrate fossil record for molecular clock estimates. *BioEssays* **27**, 1069–1075. (doi:10.1002/bies.20286)
68. Wang SC, Everson PJ. 2007 Confidence intervals for pulsed mass extinction events. *Paleobiology* **33**, 324–336. (doi:10.1666/06056.1)
69. Wang SC, Everson PJ, Zhou HJ, Park D, Chudzicki DJ. 2016 Adaptive credible intervals on stratigraphic ranges when recovery potential is unknown. *Paleobiology* **42**, 240–256. (doi:10.1017/pab.2015.37)
70. Donoghue PCJ, Smith MP, Sansom IJ. 2003 The origin and early evolution of chordates: molecular clocks and the fossil record. In *Telling the evolutionary time: molecular clocks and the fossil record* (eds PCJ Donoghue, MP Smith), pp. 190–223. London, UK: CRC Press.
71. Tavaré S, Marshall CR, Will O, Soligo C, Martin RD. 2002 Using the fossil record to estimate the age of the last common ancestor of extant primates. *Nature* **416**, 726–729. (doi:10.1038/416726a)
72. Inoue JG, Donoghue PCJ, Yang Z. 2010 The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. *Syst. Biol.* **59**, 74–89. (doi:10.1093/sysbio/syp078)
73. Heath TA, Huelsenbeck JP, Stadler T. 2014 The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proc. Natl Acad. Sci. USA* **111**, E2957–E2966. (doi:10.1073/pnas.1319091111)
74. Pyron RA. 2011 Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. *Syst. Biol.* **60**, 466–481. (doi:10.1093/sysbio/syr047)
75. Lee MSY, Cau A, Naish D, Dyke GJ. 2014 Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds. *Science* **345**, 562–566. (doi:10.1126/science.1252243)
76. Gorscak E, O'Connor PM. 2016 Time-calibrated models support congruency between Cretaceous continental rifting and titanosaurian evolutionary history. *Biol. Lett.* **12**, 20151047. (doi:10.1098/rsbl.2015.1047)
77. Lee MS, Cau A, Naish D, Dyke GJ. 2014 Morphological clocks in paleontology, and a mid-Cretaceous origin of crown Aves. *Syst. Biol.* **63**, 442–449. (doi:10.1093/sysbio/syt110)
78. Beck RMD, Lee MSY. 2014 Ancient dates or accelerated rates? Morphological clocks and the antiquity of placental mammals. *Proc. R. Soc. B* **281**, 20141278. (doi:10.1098/rspb.2014.1278)
79. Donoghue MJ, Doyle J, Gauthier J, Kluge A, Rowe T. 1989 The importance of fossils in phylogeny reconstruction. *Annu. Rev. Ecol. Syst.* **20**, 431–460. (doi:10.1146/annurev.es.20.110189.002243)
80. Gauthier J, Kluge AG, Rowe T. 1988 Amniote phylogeny and the importance of fossils. *Cladistics* **4**, 105–209. (doi:10.1111/j.1096-0031.1988.tb00514.x)
81. Zhang C, Stadler T, Klopfstein S, Heath TA, Ronquist F. 2016 Total-evidence dating under the fossilized birth–death process. *Syst. Biol.* **65**, 228–249. (doi:10.1093/sysbio/syv080)
82. Gavryushkina A, Heath TA, Ksepka DT, Stadler T, Welch D, Drummond AJ. 2015 Bayesian total evidence dating reveals the recent crown radiation of penguins. (http://arxiv.org/abs/1506.04797)
83. Sansom RS. 2015 Bias and sensitivity in the placement of fossil taxa resulting from interpretations of missing data. *Syst. Biol.* **64**, 256–266. (doi:10.1093/sysbio/syu093)
84. Sansom RS, Wills MA. 2013 Fossilization causes organisms to appear erroneously primitive by distorting evolutionary trees. *Sci. Rep.* **3**, 2545. (doi:10.1038/srep02545)
85. Sansom RS, Gabbott SE, Purnell MA. 2010 Non-random decay of chordate characters causes bias in fossil interpretation. *Nature* **463**, 797–800. (doi:10.1038/nature08745)
86. Jukes TH, Cantor CR. 1969 Evolution of protein molecules. In *Mammalian protein metabolism* (ed. HN Munro), pp. 21–132. New York, NY: Academic Press.
87. Condamine FL, Nagalingum NS, Marshall CR, 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 Evol. Biol.* **15**, 65. (doi:10.1186/s12862-015-0347-8)
88. Moore BR, Hohna S, May MR, Rannala B, Huelsenbeck JP. In press. Bayesian analysis of macroevolutionary mixtures: a critical appraisal. *Proc. Natl Acad. Sci. USA*.
89. May MR, Moore BR. In press. How well can we detect lineage-specific diversification-rate shifts? A simulation study of sequential AIC methods. *Syst. Biol.*
90. O'Reilly JE, Donoghue PCJ. 2016 Tips and nodes are complementary not competing approaches to the calibration of molecular clocks. *Biol. Lett.* **12**, 20150975. (doi:10.1098/rsbl.2015.0975)
91. Lewis PO. 2001 A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* **50**, 913–925. (doi:10.1080/106351501753462876)
92. Dos Reis M, Donoghue PC, Yang Z. 2015 Bayesian molecular clock dating of species divergences in the genomics era. *Nat. Rev. Genet*. **17**, 71–80. (doi:10.1038/nrg.2015.8)
93. Guillerme T, Cooper N. 2016 Assessment of available anatomical characters for linking living mammals to fossil taxa in phylogenetic analyses. *Biol. Lett.* **12**, 20151003. (doi:10.1098/rsbl.2015.1003)
94. Dembo M, Matzke NJ, Mooers AØ, Collard M. 2015 Bayesian analysis of a morphological supermatrix sheds light on controversial fossil hominin relationships. *Proc. R. Soc. B* **282**, 20150943. (doi:10.1098/rspb.2015.0943)
95. Larson-Johnson K. 2015 Phylogenetic investigation of the complex evolutionary history of dispersal mode and diversification rates across living and fossil Fagales. *New Phytol*. **209**, 418–435. (doi:10.1111/nph.13570)
96. Fortey RA, Jefferies RPS. 1982 Fossils and phylogeny—a compromise approach. In *Problems of phylogenetic reconstruction systematics association*, vol. 21 (eds KA Joysey, AE Friday), pp. 197–234. London, UK: Academic Press.
97. Benton MJ. 1998 The quality of the fossil record of the vertebrates. In *The adequacy of the fossil record* (eds SK Donovan, CRC Paul), pp. 269–303. New York, NY: John Wiley & Sons.
98. Huelsenbeck JP, Rannala B. 1997 Maximum likelihood estimation of phylogeny using stratigraphic data. *Paleobiology* **23**, 174–180.
99. Fisher DC. 2008 Stratocladistics: integrating temporal data and character data in phylogenetic inference. *Annu. Rev. Ecol. Evol. Syst.* **39**, 365–385. (doi:10.1146/annurev.ecolsys.38.091206.095752)
100. Smith AB. 2000 Stratigraphy in phylogeny reconstruction. *J. Paleontol.* **74**, 763–766. (doi:10.1666/0022-3360(2000)074<0763:SIPR>2.0.CO;2)
101. Fisher DC, Foote M, Fox DL, Leighton LR. 2002 Stratigraphy in phylogeny reconstruction—comment on Smith (2000). *J. Paleontol.* **76**, 585–586. (doi:10.1666/0022-3360(2002)076<0585:SIPRCO>2.0.CO;2)
102. Alroy J. 2002 Stratigraphy in phylogeny reconstruction—reply to Smith (2000). *J. Paleontol.* **76**, 587–589. (doi:10.1666/0022-3360(2002)076<0587:SIPRRT>2.0.CO;2)
103. Wagner PJ. 2002 Testing phylogenetic hypotheses with stratigraphy and morphology—a comment on Smith (2000). *J. Paleontol.* **76**, 590–593. (doi:10.1666/0022-3360(2002)076<0590:TPHWSA>2.0.CO;2)