

## Review

## Dating Tips for Divergence-Time Estimation

Joseph E. O'Reilly,<sup>1</sup> Mario dos Reis,<sup>2,3</sup> and Philip C.J. Donoghue<sup>1,\*</sup>

**The molecular clock is the only viable means of establishing an accurate timescale for Life on Earth, but it remains reliant on a capricious fossil record for calibration. 'Tip-dating' promises a conceptual advance, integrating fossil species among their living relatives using molecular/morphological datasets and evolutionary models. Fossil species of known age establish calibration directly, and their phylogenetic uncertainty is accommodated through the co-estimation of time and topology. However, challenges remain, including a dearth of effective models of morphological evolution, rate correlation, the non-random nature of missing characters in fossil data, and, most importantly, accommodating uncertainty in fossil age. We show uncertainty in fossil-dating propagates to divergence-time estimates, yielding estimates that are older and less precise than those based on traditional node calibration. Ultimately, node and tip calibrations are not mutually incompatible and may be integrated to achieve more accurate and precise evolutionary timescales.**

Establishing an evolutionary timescale for Life on Earth has long been a fundamental goal of evolutionary biology, providing the framework for inferring modes and rates of molecular and phenotypic evolution, as well as a means of associating intrinsic evolutionary change to extrinsic causal factors. This endeavor was originally the domain of paleontologists, but it is now widely accepted that fossil data alone are insufficient because of the incompleteness of the fossil record [1]. Molecular clock dating methodology can be used to establish an evolutionary timescale by calculating the molecular distance between species, and by estimating absolute molecular evolutionary rates based on the oldest fossil evidence for the antiquity of the living lineages [2]. This powerful combination of molecular and paleontological data sees through the gaps in the fossil record, providing the only viable means of establishing an accurate evolutionary timescale.

Molecular clock methodology has been developed to accommodate tree-wide substitution rate heterogeneity [3–6], and precision has increased with the availability of genome-scale datasets (i.e., an effectively infinite amount of sequence data) [7]. However, further increases in accuracy and precision may only be possible with a concomitant increase in the precision of calibrations [5,8–10]. Hence, recent years have witnessed attempts to constrain the uncertainties associated with fossil-based calibrations, including phylogenetic position, age interpretation, and the degree to which calibrating fossils approximate the true time of divergence for the nodes that they calibrate [1,11,12]. Controversially, this requires not only the oldest fossil records of extant clades on which minimum age constraints are established, but also interprets the absence of older fossils attributable to the clade to establish maximum age constraints [1,11]. Alternatively, simple mathematical functions are employed to express, probabilistically, a visceral perception of the degree to which fossil minima reflect the time of lineage divergence [1,13]. Fossil occurrence data can also be modeled statistically, with or without reference to a phylogeny, to determine the extent of the temporal gap between the age of a clade and its oldest fossils

## Trends

Total evidence dating constitutes a significant advance in divergence-time estimation. It overcomes problems with calibration by including fossil species on a par with their living relatives, using molecular sequence data from living species supplemented by morphological data from both living and fossil species.

The method relies on the controversial hypothesis of a morphological clock and suffers from the lack of development of realistic models of morphological evolution.

Most studies have failed to accommodate fossil age uncertainty. We present a protocol for characterizing and implementing this uncertainty, and demonstrate its impact on divergence-time estimation.

We argue that total evidence dating encompasses a suite of methods that can be used in bespoke combinations chosen to best suit the nature of specific divergence-time estimation studies.

<sup>1</sup>School of Earth Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK

<sup>2</sup>Department of Genetics, Evolution and Environment, University College London, London, WC1E 6BT, UK

<sup>3</sup>Present address: School of Biological and Chemical Sciences, Queen Mary, University of London, London, E1 4NS, UK

\*Correspondence: phil.donoghue@bristol.ac.uk (P.C.J. Donoghue).

## Box 1. Node Calibration

The development of TED has been shaped by a desire to overcome perceived shortcomings in node-calibration, the traditional means by which molecular clock analyses have been calibrated to absolute time. Node calibrations are established based on the oldest evidence for the existence of a clade and, most commonly, this is evidenced by the oldest fossil record of the clade. Thus, node calibrations require a prior phylogenetic hypothesis. This establishes a minimum age for the clades, but this must be complemented by a maximum age constraint. Deriving a maximum bound is more difficult to justify because it must, by necessity, rely on negative evidence. There are many methods for establishing maxima, including birth–death models [16] and statistical analysis of the stratigraphic distribution of fossils [81]. However, most commonly, maxima are established using taphonomic controls from the existence of outgroup taxa to interpret evidence of absence of ingroup taxa [82]. It is also necessary to establish the prior probability of the time of divergence between (and, using soft bounds [8], beyond) the minimum and maximum age constraints. The resulting probability density functions for each node calibration are ultimately combined with a stochastic branching model to derive effective priors on non-calibrated nodes in the tree, facilitating divergence-time estimates for all nodes.

Node calibrations have been considered unsatisfactory because they require a prior phylogenetic hypothesis and they fail to integrate uncertainty in the phylogenetic affinity of the calibrating fossils. This is problematic because the earliest fossil occurrences are often fragmentary, and therefore of uncertain affinity, and they are therefore ignored in favor of younger, better-known, and therefore phylogenetically-secure species. However, this leads to less-certain and less-informative calibrations – and dismisses an effectively infinite amount of other rate-informative fossil evidence. Some consider maximum age constraints based on fossil evidence or, rather, its absence, as unjustifiable, and establishing the nature of a probability density function spanning minimum and maximum constraints has little justification beyond gut-feeling. Unfortunately, arbitrary choices between competing parameters have an almost overwhelming impact on divergence-time estimates [83,84]. Finally, the node calibrations specified by users are invariably transformed in the establishment of the joint time prior, to the extent that they sometimes bear little relation to the original fossil evidence [7,83–85].

[14–16]. Attempts to constrain uncertainty with fossil calibrations must be welcomed, but they have not led to significantly increased precision in divergence-time estimation, not least because node calibrations require complex and often *ad hoc* interpretations of fossil and phylogenetic evidence to establish probabilistic calibrations, which are viewed by some as a grossly over-interpreted yet inadequate solution to a complex problem [17] (Box 1).

The recent introduction of fossil tip calibration [18,19], also known as ‘tip-dating’ or ‘total evidence dating’ (TED) has, therefore, enjoyed an enthusiastic welcome. This method requires both molecular sequence and morphological character datasets that are analyzed using molecular and morphological models of evolution, but its chief innovation is that it allows fossil species to be incorporated into divergence-time analyses on a par with their living relatives. This calibration methodology is analogous to the manner in which ancient DNA or archived viral sequences of known age are employed to infer rates of evolution among extant species or strains [20]. In this case, fossils of known age calibrate the rate of evolution based on their phylogenetic position, branch length, and an inferred rate of evolution. Phylogenetic topology may be estimated independently or co-estimated with the divergence-time analysis, and the rate of evolution may be based on independent or correlated rates of morphological and molecular evolution.

Thus, tip-calibration obviates many of the controversies associated with node-calibration. First, fossil species inform the evolutionary rate without recourse to *ad hoc* assumptions about the degree to which these species approximate the age of a living clade. Second, because time and topology can be co-estimated, it becomes possible to include older, temporally more-informative fossils that could not be used for node-calibration because their phylogenetic position is uncertain. Third, because calibrations no longer serve as prior estimates of clade age, tip-calibrations can be drawn from any and all fossil species, removing restrictions on the amount paleontological data that can be included in divergence-time studies. Finally, tip calibrations summarize the age of a single species only, avoiding the over-interpretation of negative evidence in establishing maximum constraints.

Tip-calibration was originally introduced based on empirical divergence-time analyses of insects [19] and amphibians [18], and it has since been applied to mammals [21–26], teleost fishes [27–31],

arachnid spiders [32,33], flies [34], and plants [35]. The approach has been extended to analyses of entirely extinct clades, relying exclusively on morphological data [36]. While tip-calibration was initially advocated on the basis that it was less sensitive to root time prior densities, and yielded more precise divergence-time estimates in comparison to node-calibration [19], subsequent studies have shown the reverse to be true [30,32]. Furthermore, tip-calibration has proven consistently to yield older age-estimates than traditional node-calibration [19,21–24,30,32,33]. Thus, while it is clear that in incorporating all data pertinent to divergence-time estimation, and tip-calibration is the most promising approach for establishing accurate and precise evolutionary timescales, at present it appears to be less accurate than conventional node calibration methods. Below we consider the factors biasing current methods employing tip-calibration, and suggest ways in which they can be developed to obtain more accurate divergence-time estimates.

### Models of Morphological Character Evolution and the Incompleteness of Fossils

While there are several nested models of molecular substitution, morphological models have not enjoyed much development, with only a handful proposed to date and even fewer actually implemented in popular software packages [37–42]. The Mk model of discrete character evolution has been utilized in all published tip-calibrated analyses to date [43]. The Mk model is a  $k$  states generalization of the JC69 model of molecular substitution and, inevitably, it possesses many simplifying assumptions that may not hold true for morphology [44]. Independent evolution of sites and equal equilibrium frequencies are two factors that are particularly difficult to justify for morphological evolution. Alternative models utilizing continuous characters [45] or the threshold model [46,47] are appealing alternatives, but they have yet to be implemented.

The inherently incomplete nature of fossil phenotypic data, in comparison to living species, is undoubtedly a challenge to tip-calibrated divergence-time analyses. The impact of missing sequence data on Bayesian phylogenetic topology estimation has been investigated, with the majority of studies indicating that it is unlikely to have a strong negative impact [48–52], except where there is a comparatively small number (not proportion) of non-missing sites [49]. This is clearly a problem for topology estimation based on phenotype where datasets are generally very small in comparison to molecular sequence alignments. This issue is exacerbated by the decidedly non-random nature of missing phenotype data in fossil species [53,54]. Fossil data are invariably biased towards the preservation of phenotypic characters that are manifest in, or as, mineralized skeletal structures. Even where soft tissue characters are exceptionally well preserved, some groups exhibit a phenomenon coined ‘stem-ward slippage’ in which features are lost to decay in reverse phylogenetic order, making their fossils appear artefactually to belong to more primitive evolutionary grades [53,54]. While the impact of these factors on topology estimation has been considered, it has not been investigated explicitly in the context of time and rate estimation [53].

For tip-calibrated divergence-time analyses, the likely impact is twofold: calibrating fossil species will be assigned to erroneously early-branching positions within the phylogeny, and the branch lengths will be underestimated, both owing to their lack of shared-derived and autapomorphic soft-tissue characters, missing artefactually as a consequence of non-random decay patterns. Both these phenomena will influence rate estimates and, therefore, divergence-time estimates. To minimize the negative influence of missing data, sub-sampling approaches have been proposed, allowing the use of only the most completely coded taxa or characters. While it has been argued that such approaches have minimal impact on topology and age estimation [18,19], this is unlikely to hold true for non-random missing data. Alternatively, a model of fossilization could be employed that accounts for the directed loss of characters during

preservation, but modeling this process may be entirely unrealistic given that fossilization potential varies with environment and taxonomic group.

### Dating Tips and Calibration Strategies

Almost all TED studies conducted so far have employed point age estimates for the fossil species used as tip-calibrations, assuming implicitly that the age of the fossil is known without error. This has been done on the sometimes explicit justification that the errors associated with the dating of fossils are negligible [19,33]. This approach is reminiscent of the point age estimates for node calibrations, employed when divergence-time estimation was in its infancy, and none of the lessons learned from the development of node-calibration strategies [1,11,13] have been transferred to studies that employ fossil tip-calibration. It is well established that the age of a fossil can rarely, if ever, be known without error, and this uncertainty must be accommodated regardless of whether the fossil is used in the construction of a node or tip-calibration. The age of any fossil occurrence can be constrained only to within an envelope of minimum–maximum bounds, the span of which varies depending on the attendant evidential context. Node-calibrations are based principally on the earliest secure fossil record of a clade (Box 1), and it is thus necessary to determine only the minimum age interpretation of the calibrating fossil [13,55]. At the least, the age of a tip-calibrating fossil requires establishing both its minimum and maximum age interpretations. For both the minimum and maximum age interpretations, this invariably entails a tortuous daisy-chain of litho-, bio-, chemo-, cyclo-, and/or magneto–stratigraphic correlations between the site of the fossil occurrence and another in which a geochronological absolute date has been established, at each step taking the minimum or maximum relative age interpretation, as appropriate, leading to iteratively increasing age uncertainty (Box 2 gives a worked example). It is likely that, in many instances, this uncertainty will exceed that associated with local node-calibrations, although tip calibrations may prove more palatable because they rely on fewer assumptions.

Borrowing from practice in establishing node-calibrations, the age uncertainty associated with a fossil species can be modeled as a uniform distribution if there is equal probability of the age of the fossil, per unit time, between minimum–maximum age interpretations. Alternatively, the variety of parametric distributions already implemented for node calibrations may be redeployed in instances where there is justification for focusing uncertainty closer to the minimum, maximum, or mid-range between age bounds. The range of available distributions and instances in which they may be deployed, are discussed in Box 3.

Tip-calibrations present further peculiarities that should also be considered in attempting to integrate uncertainty associated with their age. For example, many fossil species employed in the node-calibration of divergence-time analyses are not single occurrences but, rather, occur through a stratigraphic age range. This is of little relevance to node-calibration used to establish a clade age minimum; however, in establishing a tip-calibration this is much more germane. Given that, by definition, such species will exhibit little or no morphological variation, it seems appropriate that this age range should be incorporated into the age uncertainty associated with the fossil (Box 4 expands upon this idea). Ultimately, it may prove useful to integrate this information, in the form of effective stasis in the set of traits analyzed, into the inference of rate variation across the tree.

Because tip-calibration and TED have been presented as a means to achieve greater precision in divergence-time estimation [19], it is pertinent to consider whether this can be sustained while integrating the uncertainty associated with the age of fossil tips. To this end, we reanalyzed the dataset from the seminal TED study [19], in which tip-calibrations were utilized to estimate divergence times for Hymenoptera. Ronquist and colleagues focused on the theoretical and practical introduction of the method, and they did not take account of the uncertainty associated

### Box 2. The Construction of a Tip Calibration

*Palaeothalia laiangensis* was recovered from the Laiyang Formation in Liaoning, China, which can be divided into four members, the third of which has yielded most fossils. Although the Laiyang Formation contains no directly dateable elements, correlation with the base of the Yixian Formation, also of China, allows the use of radiometric dates for the base of this formation to inform the age of the Laiyang Formation. Similarly, the unit overlying the Laiyang Formation, the Houkuang Formation, contains dateable elements, allowing an age for the base of this formation to constrain the age of the top of the Laiyang Formation. Because we consider the age of the fossil species *P. laiangensis* to lie within the chronological interval between the top and base of the unit of its provenance, and without further information to constrain the limits and distribution of probability, we can use the ages of these limits to determine the bounds of our calibration. Correlation with the Yixian Formation can be made based on numerous palynological and faunal similarities, mostly with the lowermost member of the Yixian Formation, the Lujitun Bed. While these sources may not individually be considered conclusive, numerous biostratigraphic similarities strongly support this correlation [86–90]. Radiometric dates of  $128.4 \text{ Ma} \pm 0.2 \text{ Myr}$  have been acquired from the base of the Lujitun Bed, which can be used to determine the age of the base of the Laiyang Formation on the basis of the correlation between these units [90–92] (Figure I).

The Laiyang Formation is succeeded by the Qingshan Group, of which the Houkuang Formation is the lowermost member. Because the Laiyang Formation can be no younger than the overlying unit, an age for the base of the Houkuang Formation can provide a minimum age for the Laiyang Formation. U–Pb dating of zircons from the base of the Houkuang Formation has yielded dates of  $106 \text{ Ma} \pm 2 \text{ Myr}$ , which can be used to constrain the minimum age of the Laiyang Formation [93]. Because no dates are available to further constrain the limits of this formation, and without any further information regarding the manner in which the probability of the age of *P. laiangensis* should be distributed, a uniform distribution spanning the full range of uncertainty in radiometric dates across the interval (128.6–104 Ma). This tip age can be contrasted with that utilized by Ronquist *et al.* [19] of a fixed age of 140 Ma, which falls significantly outside the bounds of this calibration.

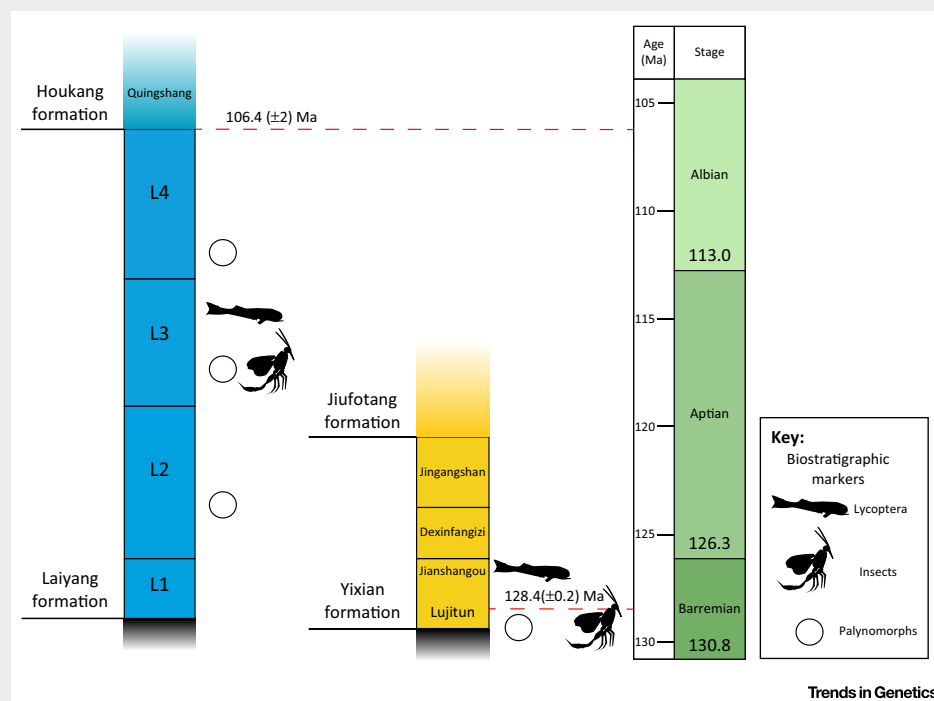


Figure I. Construction of a Tip-Calibration for *P. laiangensis* Based on Stratigraphic Correlation Between the Unit of Provenance, The Laiyang Formation, and the Yixian Formation of China.

with the fossils used in tip-calibration. We reproduced the calibrations for each fossil tip, accommodating uncertainty in the age of each fossil species using probabilistic distributions (Box 2 gives an example of this process). In contrast to previous assertions, that the uncertainties associated with tip ages would be negligible [19,33], our attempts to capture a realistic estimate of the associated uncertainty results in tip-calibrations that span tens of millions of years – in contrast to the errorless estimates of age estimates used by the original authors. To determine

the performance of node- versus tip-calibration, we also constructed node-calibrations following established best practice [11] (see the supplemental information online). On average, recalibrated node priors were 23 Myr wider than the original calibrations. In both tip- and node-calibrations, uncertainty was modeled as a uniform distribution. Analyses were performed in MrBayes 3.2.2 [41] in broadly the same manner as the original article (see the supplemental information online for details). Precision was measured as the width of the 95% confidence interval (CI) for posterior estimates of node age for 14 key in-group clades that could be resolved.

Our analyses show that when fossil age uncertainty is properly accounted for, tip-calibrated analyses do not necessarily yield divergence-time estimates that are more precise than those derived using node-calibration. Furthermore, for 27% of fossil taxa, the 95% highest posterior density (HPD) estimates of fossil tip age did not encompass the original fixed tip-calibration, demonstrating the importance of appropriate prior construction. Divergence-time estimates based on node-calibration are the most precise in all but four of the component clades (Figure 1). In line with almost all previous TED studies, tip-calibration yields clade ages that are older, in general, than like-for-like estimates based on node-calibration, the only exceptions being divergences outside Hymenoptera. These deeper divergence times are most prominent in

### Box 3. Density Distributions for Fossil Tip-Calibration

The wide range and flexibility of probability distributions has allowed the accurate incorporation of uncertainty into fossil calibrations. Unfortunately, encapsulating prior knowledge of fossil age as a density distribution is not a straightforward task, and the application of density distributions with arbitrarily assigned parameters can have profound effects on age estimates [84]. Although computational methods exist for the integration of fossil stratigraphic range and geochronological age data [94], they are rarely implemented in evolutionary studies, and in their place it is important that the construction of density distributions is justified explicitly. For tip-calibration, several distributions are applicable, depending on the context in which uncertainty manifests itself. Six distributions are presented here using the calibration of the Hymenopteran fossil *Eoxyla* (minimum = 141 Ma, maximum = 168 Ma) as an example (Figure 1).

(i) *Exponential Distribution* (Figure 1,i). Exponential distributions introduce diminishing probability over time. These calibrations are particularly useful when the weight of evidence suggests that the true age of a tip is close to the minimum bound but a much more ancient age cannot be ruled out. The rate parameter determines how far back the distribution extends to ( $\lambda$ ), with its reciprocal being equal to the mean. Here two parameterizations reflect separate assumptions of how ancient the clade may be.

(ii) *Gamma Distribution* (Figure 1,ii). The gamma distribution has two parameters, shape ( $\alpha$ ) and rate ( $\beta$ ), and is relatively flexible compared to other available distributions. For example, when  $\alpha < 1$ , the distribution is L-shaped with the mode at zero and with a long tail. When  $\alpha = 1$  the distribution reduces to the exponential distribution. Finally, when  $\alpha > 1$  the distribution has a mode away from zero. As the value of  $\alpha$  increases, the gamma converges to the normal distribution with mean of  $\alpha/\beta$  and variance  $\alpha/\beta^2$ . An offset is required to express the distribution relative to the minimum age; here an offset of 141 Ma is used.

(iii) *Normal Distribution* (Figure 1,iii). The normal distribution has seen limited use for node calibrations, but it may prove more useful in a tip-dating context. Normal distributions place equal diminishing probability (determined by the variance  $\sigma^2$ ) either side of the mean ( $\mu$ ), and may be useful when a species is known from the middle of a unit only. Here the upper and lower bounds of the species chronological distribution are set at 2 standard deviations from the mean, allowing for age estimates that violate the bounds (Figure 1,iii).

(iv) *Point Calibrations* (Figure 1,iv). Assume that the provided age is absolutely correct, disregarding any meaningful interpretation of the fossil record; therefore, erroneously inflated confidence in posterior age estimates is introduced owing to increased specificity in the prior distribution [95]. Here the assumed tip age is at the mid-point of the chronological distribution of the taxon.

(v) *Uniform Distributions* (Figure 1,v). Uniform distributions place equal probability across the interval ( $a,b$ ). This distribution is applicable when a fossil is known from a single unit in which dates can be derived for the base and top, but no additional constraints on the distribution of age can be demonstrated.

(vi) *Lognormal Distribution* (Figure 1,vi). Lognormal distributions allow the assignment of diminishing probability that the first appearance of a species is actually described by the age of the fossil specimen itself. The distribution has two parameters, the log-mean ( $\mu$ ) and log-standard deviation ( $\sigma$ ).  $\sigma$  determines the shape of the distribution; when it is close to zero, the distribution is symmetrical, and when it is large, the distribution becomes very asymmetric with a long tail and with the mode of the distribution moving towards zero.

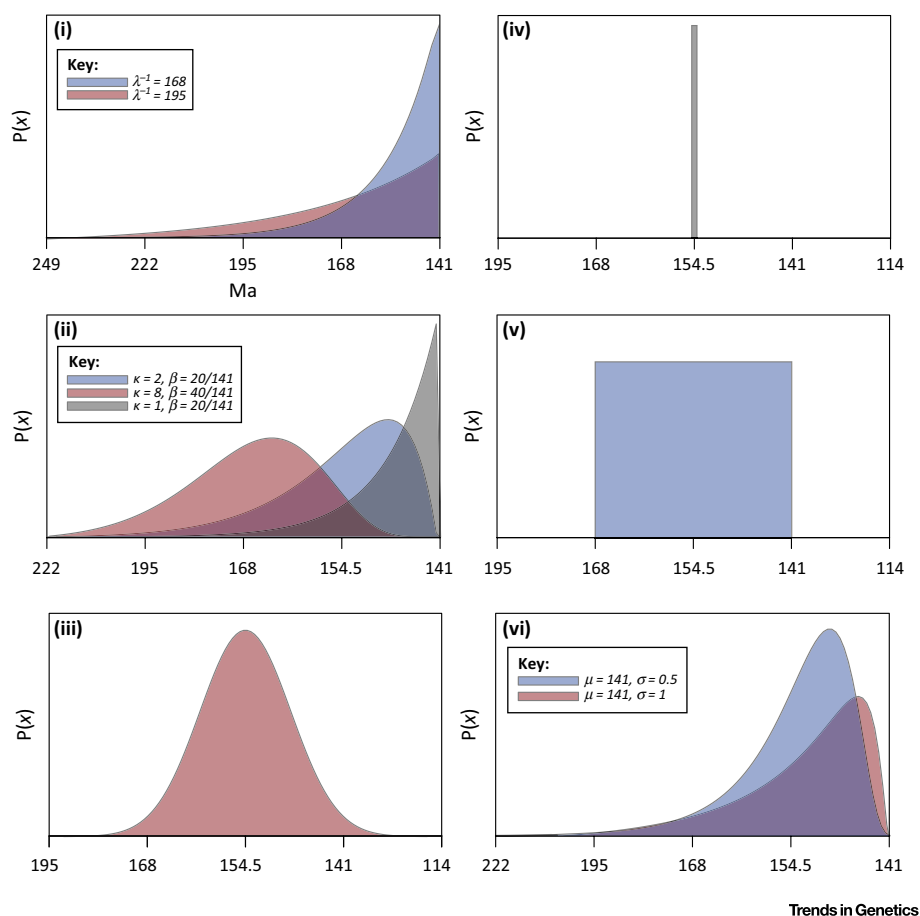


Figure I. Six Alternative Probability Density Functions Commonly Used to Encapsulate Prior Knowledge of the Chronological Distribution of a Fossil Tip. Here the calibration of the fossil taxon *Eoxyela* is used to demonstrate the characteristics of the different distributions.

Vespina, where it appears that relaxing the constraint on the age of *Mesorussus* (which was assigned to Vespina in both our analysis and the original analysis [19]) from 94 Ma to 93.7–140.3 Ma leads to the older age estimates.

While we were able to repeat the results of the original analysis using the original calibrations, we were unable to reproduce the topological resolution and/or monophyly of Xyelidae, Pamphilioidea, and the placement of fossil taxa *Palaeothalia*, *Cleistogaster*, and *Prosyntexis* when employing our revised tip-calibrations. Because the only variable between our analyses is the method of calibration construction, it appears that the more realistic age-uncertainty associated with the fossils in our revised tip-calibrations has impacted on topology estimation as part of the co-estimation of topology and time. Thus, by implication, accommodating the realistic age uncertainty associated with fossil tip-calibrations also impacts indirectly on rate and clade age estimates by contributing to topology estimation.

Claims of the superiority of tip-calibration over node-calibration appear unfounded when fossil age uncertainty is accommodated equally. Furthermore, it is not entirely clear that node calibrations are redundant in tip-calibration studies because, logically, they can still serve their purpose of constraining node age estimates and rate variation. One way to assess whether they



#### Box 4. Tip Calibrations and Apparent Morphological Stasis

The exact definition of what a fossil tip represents has not yet been defined explicitly because it is currently not clear whether calibrations should be constructed based on the age of an individual fossil, or to reflect the minimum age of the fossil species to which it is assigned, or the total known temporal range of that species. For a species with only one known fossil the situation is simple: the tip represents the evolutionary path to the first appearance of the suite of characters it possesses, and it is therefore justifiable to assign a calibration based on the provenance of that individual fossil. It is less clear how a fossil species known from several temporal intervals should be represented in terms of the tip-age. For example, consider the scenario outlined in Figure 1. A fossil species (†) with a chronological distribution of 10 Myr is recovered from two serial units (A and B), each of 5 Myr in length, with no overlap. The suite of characters at the start of the first deposit and at the end of the last deposit is the same; there is effective morphological stasis. In this scenario, morphological and molecular rates are certainly unlink because, despite the perceived evolutionary stasis, there will be molecular evolutionary change. The choice of calibration bounds in this situation can readily lead to the over- or under-estimation of rates on surrounding branches, by compressing or stretching the length of the branch subtending the fossil species. If the tip age is constrained based on the limits of the oldest occurrence, apparent morphological stasis is not accommodated; constraining tip age based on the combined time span of both temporal occurrences is likely to inference of lower rates on other branches [96]. An alternative calibration strategy might be to assign point estimates based on the statistically derived 95% CI for the lower limit of the true stratigraphic range of such fossil species, ignoring the protracted stasis but explicitly calibrating the origin of the suite of fossilized characters [97]. Is this morphological stasis a derived state that should not be extrapolated across the tree, or it is inherited from earlier members of the lineage and should therefore be used to inform rates elsewhere? Morphological stasis is hypothesized to be driven largely by either stabilizing selection [98] or developmental constraints [99], but a consensus as to which is the controlling factor has still to be reached [100]. If the latter obtains, it is likely that calibrations need to incorporate stasis as it is an inherited trait.

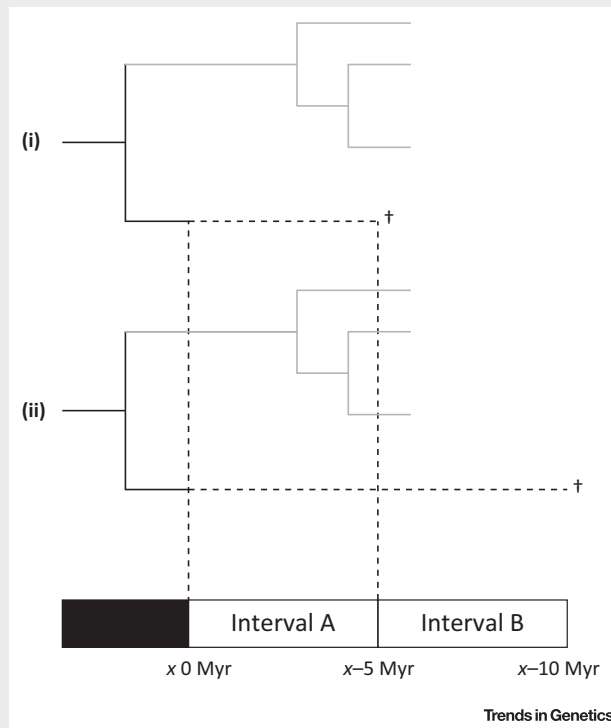
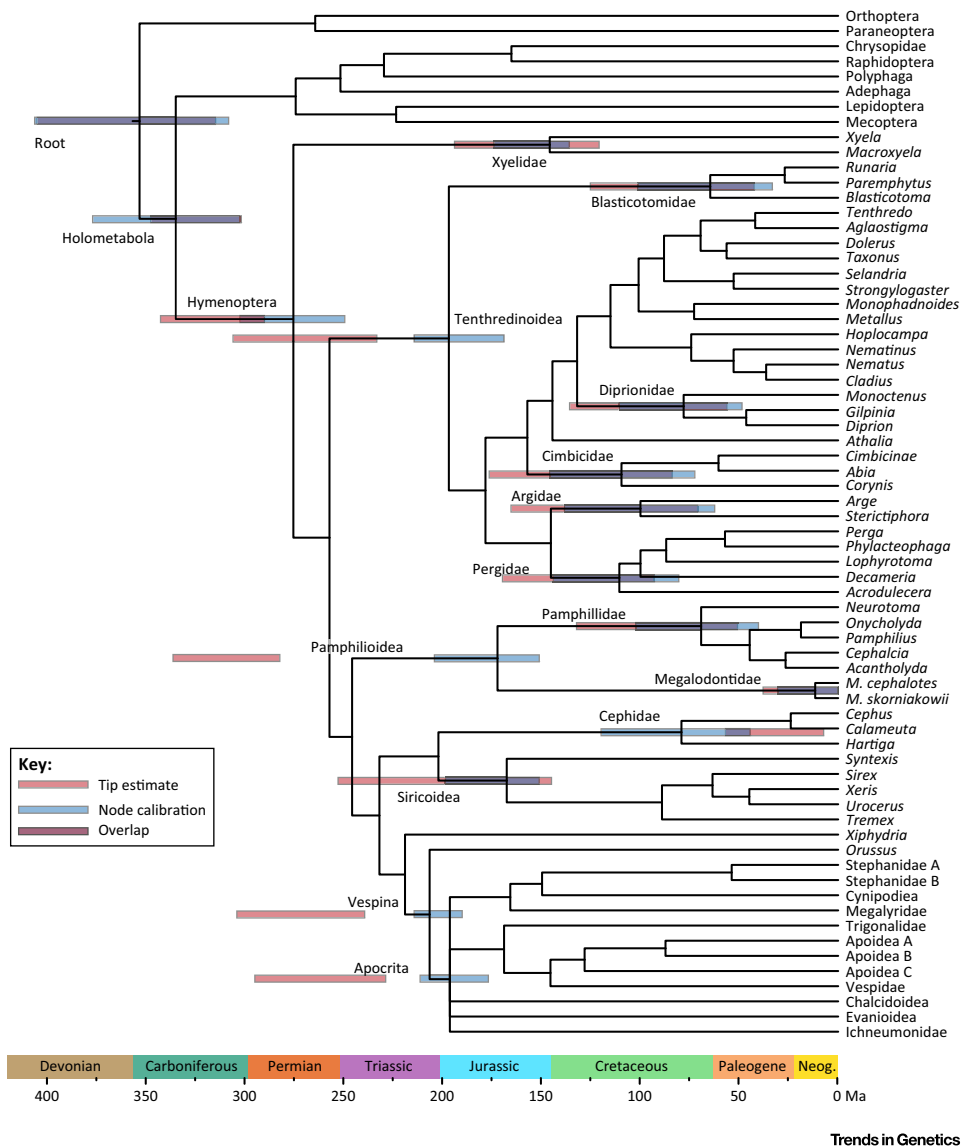


Figure 1. Potential Branch Lengths (Dotted Lines) When Fossil Tip Age Is Calibrated Based on Different Stratigraphic Limits when a Fossil Taxon (†) Is Recovered from Multiple Units (A and B). Calibrations constructed from the full stratigraphic range of the fossil taxon will incorporate stasis into the model, but may induce lower rates on nearby branches (ii). Calibrations constructed from the first appearance on the fossil taxon ignore the protracted stasis, and may induce inflated rates on surrounding branches (i).

are still useful in this role is in comparing traditional node calibrations and the posterior node-age estimates based on analyses employing tip-calibrations. We did this for the nine nodes for which we have constructed calibrations. The results (Figure 2) show that, while all of the node age estimates derived from tip-calibration are old relative to the node calibrations, four fall fully outside these node age constraints. It could be argued that this demonstrates the fallacy of



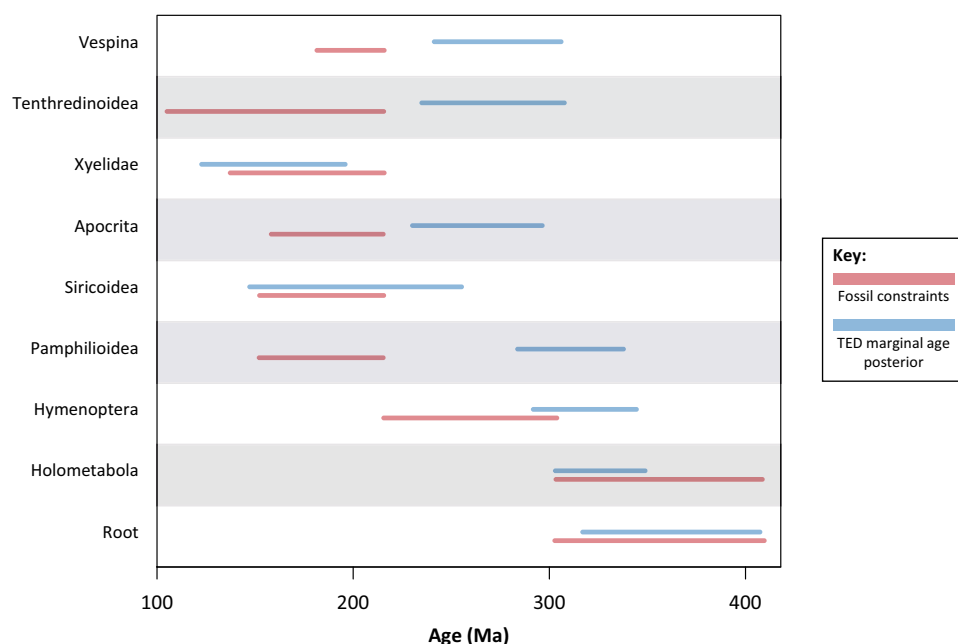


**Figure 1. A Dated Phylogeny of Hymenoptera Produced Using Node-Calibrations.** Node bars represent 95% highest posterior density (HPD) for node ages estimated with either node-calibration or total evidence dating (blue and red respectively). The dotted lines join HPD bars to the node for which they represent age estimate confidence, and do not represent an extension of the confidence interval.

fossil-based maximum age constraint, however, two of the node age estimates include age ranges that are younger than the minimum age constraints based on the empirical paleontological evidence. Evidently, there remains a role for node age constraints, even in tip calibration divergence-time analyses.

### TED – Less Than the Sum of Its Parts?

While TED has been presented as an alternative approach to conventional node-calibrated molecular clocks, this is a false dichotomy. TED is a specific combination of approaches that are neither inextricably linked, nor mutually exclusive from node-calibrated molecular clock analysis. These include: (i) the relaxed morphological clock, (ii) tip-calibration, and (iii) co-estimation of time and topology. In practice, these methods can and have been deployed in isolation in augmenting



**Figure 2.** Comparison Between Marginal Posterior Distributions on Nine Node Ages Estimated with Total Evidence Dating (TED, Blue), and Prior Clade-Age Constraints Employed for Node-Calibrated Analysis of the Same Data (Red). The calibrations for node-calibrated analysis encapsulate the fossil evidence for the possible age of each clade. A lack of overlap at any node implies that there is discordance between the TED effective prior on that node and the fossil record. Discordance between these two distributions demonstrates that TED may lead to empirically unsupportable clade age estimates.

conventional molecular clock analyses. For example, the divergence-time study of Schrago and colleagues [22] of New World primates followed a two-step protocol, using the posterior age estimates from a conventional molecular clock analysis of living species as time-priors on node ages in a morphological clock analysis including both living and fossil species. At the least, this approach obviates the problematic assumption that molecular and morphological data co-vary, following a single rate model. Lee *et al.* [56] co-estimated time and topology using dated tips and a morphological clock, eschewing molecular data altogether, in their analysis of body size evolution through the dinosaur–bird evolutionary transition. This approach will surely be adopted widely as paleontologists seek to obtain clade ages, rather than minimum ages, for entirely extinct clades. However, this enthusiasm may be short-lived given that tip-calibration approaches have consistently yielded older clade age estimates than conventional molecular clock studies – against which paleontologists have a long tradition of objecting violently [57]. Combining ancient DNA and morphological data is another possibility afforded by tip-calibration, as has been applied to studying Pantherine phylogeny [23]. This combination of ancient morphology and DNA may facilitate more accurate estimates of evolutionary rate.

While there has been enthusiasm in the application of the TED approach, not least because it provides a platform for the integration of many disparate sources of uncertainty, it is arguable that in so doing this approach serves as a black box that disengages the user from the assumptions underpinning the analysis, many of which are very difficult to justify (see Outsanding Questions). One of the most problematic, potentially, is the co-estimation of time and topology, which, as we have demonstrated, allows fossil ages to constrain their phylogenetic position and, therefore, impact on the estimation of rates and dates. This follows the common-sense

expectation that the age of a fossil species must reflect their phylogenetic position. Indeed, phylogeny estimation integrating the relative stratigraphic age of fossil species has a long tradition in paleontology, but it has been much debated [58–62] and generally abandoned in favor of phylogenetics based on phenotype, perhaps refined by stratigraphy, except in groups with exceptionally rich fossil records that are rarely if ever the focus of divergence-time studies [63]. Although there is a broad correlation between clade age and phylogenetic branching order [64], this relationship breaks down as fossil taxon sampling decreases [65]. It is complicated further by secular biases in the rock record, which serve to telescope temporally-distinct fossil species originations and extinctions [66], and in the differential preservation of fossil groups and the environments in which they lived [67]. Thus, there appears little justification for the co-estimation of time and topology where fossil ages contribute to their phylogenetic position. We strongly advocate the prior analysis of topology before divergence-time estimation. It is unfortunate that this approach precludes the integration of phylogenetic uncertainty into divergence-time estimation, but resolving phylogenetic uncertainty using tip age does not appear viable using current methods.

The majority of TED analyses model branch rates as linked across morphological and molecular partitions (i.e., the application of rate multipliers to describe inter-partition rate heterogeneity [68–70]). The suitability of this assumption for partitioned molecular data alone has been investigated, and partition-specific clocks have been developed for when this assumption is not met [68,71]. However, the effect of morphological and molecular partition-specific clocks has barely been considered [18,68,72], and most studies employ a single, partition-linked clock despite the fact that a strong covarying relationship between molecular and morphological rates has never been demonstrated [73–75]. Morphological rate heterogeneity has long been considered likely to significantly dwarf its molecular counterpart, suggesting that the assumption of phenotypic and molecular rate correlation is unjustified [76,77]. Molecular rates are interpreted as genome-wide measures of the number of substitutions accumulated per time unit, while morphological rates reflect only those aspects of the genome that specify the phenotypic traits analyzed, further diminishing any expectation of covariance between molecular and morphological evolutionary rates [73,78]. In this light, it is perhaps unsurprising that unlinked partition-specific clocks have been found to be better-fitting than a single linked clock for mixed data analyses [79].

While node and tip-based calibration have been presented as competing approaches, they are not mutually exclusive. Indeed, some temporal constraints on clade age are better suited to being implemented as node-calibrations. This is particularly true of biogeographic calibrations where, based on the modern and ancient biogeographic distributions of evolutionary lineages, it is acceptable to assume that a dateable vicariance event, such as continental fragmentation, is causal to lineage divergence. Similarly, some fossil evidence is better reflected as node-age calibrations rather than through including component fossil species as tip-calibrations. Node and tip-calibrations have already been employed together to calibrate interior nodes of the out-group, while allowing an unconstrained in-group topology, or as part of a highly constrained topology in which fossil taxa are assigned to predetermined clades [19,80]. However, this must be extended to allow node-calibrations throughout the tree. This approach requires a fixed topology (or at least backbone constraints compatible with calibrated nodes) and, thus, precludes the possibility to co-estimating time and topology, but, as we have argued, this may not be a material loss. Winterton and Ware [34] have shown that combining node and tip calibrations in this way yields younger estimates than node (or, presumably, tip) calibrations alone. Node calibrations may serve to mitigate against the propensity for tip-calibration-based studies to yield unacceptably-ancient divergence dates because it places additional constraints on the age of internal nodes of the tree, providing local checks on branch length and rate variation.

Finally, it is likely that the mismatch between divergence-time estimates based on node and tip-calibration strategies is based at least in part in the shortcomings of the Mk model in explaining the phenotypic data commonly used in tip-calibration studies. The Mk model fails to account for expected characteristics of cladistic data, including the covariation of characters that are biologically linked and logically linked through character design. Doubtless, the excitement surrounding the combined use of morphological and molecular data for divergence-time analysis will lead to the development of this and other models of evolution. However, it may also be appropriate to consider different approaches to characterizing phenotype, such as through the types of continuous variable characters obtained through morphometry of features such as skull suture patterns, tooth shape, or the dimensions of limb bones. The stochastic variation of such data is more similar to the variation seen in molecular sequence alignments and, as such, may be more readily modeled and better suited to combined data divergence-time analysis.

### Concluding Remarks

The advances inherent in TED provide an excellent platform for the further development of methods for divergence-time analysis. However, many aspects of the principal evolutionary model for phenotypic data currently employed are violated by the evolutionary process it attempts to encapsulate. The extent of these problems is so great that divergence-time estimates derived using tip-calibration cannot enjoy the same confidence as conventional node-calibrated molecular clock studies. However, with the development of evolutionary models, protocols for dating fossil species and dealing with missing data, TED encompasses a variety of powerful tools, the combination of which can be chosen to best test the hypothesis at hand. It also provides a viable framework for the best and greatest use of paleontological data that may serve as a nexus of the unification of paleontological and molecular approaches to establishing evolutionary timescales.

### Supplemental Information

Supplemental information associated with this article can be found online at <http://dx.doi.org/10.1016/j.tig.2015.08.001>.

### References

- Donoghue, P.C.J. and Benton, M.J. (2007) Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends Ecol. Evol.* 22, 424–431
- Zuckerkandl, E. and Pauling, L. (1965) Molecules as documents of evolutionary history. *J. Theor. Biol.* 8, 357–366
- Sanderson, M.J. (2002) Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* 19, 101–109
- Thorne, J.L. *et al.* (1998) Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* 15, 1647–1657
- Rannala, B. and Yang, Z. (2007) Inferring speciation times under an episodic molecular clock. *Syst. Biol.* 56, 453–466
- Drummond, A.J. *et al.* (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4, e88
- dos Reis, M. *et al.* (2012) Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc. R. Soc. B: Biol. Sci.* 279, 3491–3500
- Yang, Z. and 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
- Dos Reis, M. and Yang, Z. (2013) The unbearable uncertainty of Bayesian divergence time estimation. *J. Syst. Evol.* 51, 30–43
- Zhu, T. *et al.* (2015) Characterization of the uncertainty of divergence time estimation under relaxed molecular clock models using multiple loci. *Syst. Biol.* 64, 267–280
- Parham, J.F. *et al.* (2012) Best practices for justifying fossil calibrations. *Syst. Biol.* 61, 346–359
- Benton, M.J. and Donoghue, P.C. (2007) Paleontological evidence to date the tree of life. *Mol. Biol. Evol.* 24, 26–53
- Ho, S.Y. and Phillips, M.J. (2009) Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Syst. Biol.* 58, 367–380
- Marshall, C.R. (1994) Confidence-intervals on stratigraphic ranges – partial relaxation of the assumption of randomly distributed fossil horizons. *Paleobiology* 20, 459–469
- Wilkinson, R.D. *et al.* (2011) Dating primate divergences through an integrated analysis of palaeontological and molecular data. *Syst. Biol.* 60, 16–31
- Heath, T.A. *et al.* (2014) The fossilized birth-death process for coherent calibration of divergence-time estimates. *Proc. Natl. Acad. Sci. U.S.A.* 111, E2957–E2966
- Heads, M. (2012) Bayesian transmutation of clade divergence dates: a critique. *J. Biogeogr.* 39, 1749–1756
- Pyron, R.A. (2011) Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. *Syst. Biol.* 60, 466–481
- Ronquist, F. *et al.* (2012) A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Syst. Biol.* 61, 973–999
- Drummond, A.J. *et al.* (2003) Measurably evolving populations. *Trends Ecol. Evol.* 18, 481–488
- Slater, G.J. (2013) Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the Cretaceous–Palaeogene boundary. *Methods Ecol. Evol.* 4, 734–744

### Outstanding Questions

How adequate is the Mk model of morphological evolution for estimating divergence times? There has been little development of this model in the past 15 years. Its suitability for morphology-based divergence-time estimation remains unclear.

What is the best method for modeling the relationship between molecular and morphological evolutionary rate? Many analyses model these rates as correlated variables, but it is unclear how well this approach encapsulates their true relationship.

How congruent with the fossil record are tip-calibration node-age priors? Exploring the induced time prior is a non-trivial task for TED analyses owing to the co-estimation of time and topology. Without knowledge of the time prior it is not possible to determine whether zero probability is being assigned to age estimates that violate minima derived from the empirical evidence contained within the fossil record.

Are morphological data best characterized as categorical or continuous variable data for the purposes of divergence-time estimation?

22. Schrago, C.G. *et al.* (2013) Combining fossil and molecular data to date the diversification of New World Primates. *J. Evol. Biol.* 26, 2438–2446
23. Tseng, Z.J. *et al.* (2014) Himalayan fossils of the oldest known pantherine establish ancient origin of big cats. *Proc. R. Soc. B: Biol. Sci.* 281, 20132686
24. Slater, G.J. (2015) Iterative adaptive radiations of fossil canids show no evidence for diversity-dependent trait evolution. *Proc. Natl. Acad. Sci. U.S.A.* 201403666
25. Dembo, M. *et al.* (2015) Bayesian analysis of a morphological supermatrix sheds light on controversial fossil hominin relationships. *Proc. Natl. Acad. Sci. U.S.A.* 282, 20150943
26. Marx, F.G. and Fordyce, R.E. (2015) Baleen boom and bust: a synthesis of mysticete phylogeny, diversity and disparity. *R. Soc. Open Sci.* 2, 140434
27. Near, T.J. *et al.* (2014) Phylogenetic relationships and timing of diversification in gonorynchiform fishes inferred using nuclear gene DNA sequences (Teleostei: Ostariophysi). *Mol. Phylogenet. Evol.* 80, 297–307
28. Dornburg, A. *et al.* (2015) The impact of shifts in marine biodiversity hotspots on patterns of range evolution: evidence from the Holocentridae (squirrelfishes and soldierfishes). *Evolution* 69, 146–161
29. Alexandrou, M.A. *et al.* (2013) Genome duplication and multiple evolutionary origins of complex migratory behavior in Salmonidae. *Mol. Phylogenet. Evol.* 69, 514–523
30. Arcila, D. *et al.* (2015) An evaluation of fossil tip-dating versus node-age calibrations in tetraodontiform fishes (Teleostei: Percormphaceae). *Mol. Phylogenet. Evol.* 82, 131–145
31. Dornburg, A. *et al.* (2015) Phylogenetic analysis of molecular and morphological data highlights uncertainty in the relationships of fossil and living species of Elopomorpha (Actinopterygii: Teleostei). *Mol. Phylogenet. Evol.* 89, 205–218
32. Wood, H.M. *et al.* (2013) Treating fossils as terminal taxa in divergence time estimation reveals ancient vicariance patterns in the palpmoid spiders. *Syst. Biol.* 62, 264–284
33. Sharma, P.P. and Giribet, G. (2014) A revised dated phylogeny of the arachnid order Opiliones. *Front. Genet.* 5, 255
34. Winterton, S.L. and Ware, J.L. (2015) Phylogeny, divergence times and biogeography of window flies (Scenopinidae) and the therevoid clade (Diptera: Asiloidea). *Syst. Entomol.* 40, 491–519
35. Larson-Johnson, K. (2015) Phylogenetic investigation of the complex evolutionary history of dispersal mode and diversification rates across living and fossil Fagales. *New phytol.* Published online July 21, 2015. <http://dx.doi.org/10.1111/nph.13570>
36. Lee, M.S.Y. *et al.* (2014) Morphological clocks in paleontology, and a Mid-Cretaceous origin of crown aves. *Syst. Biol.* 63, 442–449
37. Alekseyenko, A.V. *et al.* (2008) Wagner and Dollo: a stochastic duet by composing two parsimonious solos. *Syst. Biol.* 57, 772–784
38. Bouckaert, R. *et al.* (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 10, e1003537
39. Swofford, D.L. (1998) *PAUP\*: Phylogenetic Analysis Using Parsimony (\*and Other Methods) Version 4*, Sinauer Associates
40. Felsenstein, J. (1989) PHYLIP – Phylogeny Inference Package (Version 3.2). *Cladistics* 5, 164–166
41. 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
42. Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313
43. Lewis, P.O. (2001) A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* 50, 913–925
44. Jukes, T.H. and Cantor, C.R. (1969) Evolution of protein molecules. In *Mammalian Protein Metabolism* (Munro, H.N., ed.), pp. 21–132, Academic Press
45. Felsenstein, J. (1973) Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet.* 25, 471–492
46. Felsenstein, J. (2012) A comparative method for both discrete and continuous characters using the threshold model. *Am. Nat.* 179, 145–156
47. Felsenstein, J. (2005) Using the quantitative genetic threshold model for inferences between and within species. *Philos. Trans. R. Soc. B: Biol. Sci.* 360, 1427–1434
48. Wiens, J.J. and Morrill, M.C. (2011) Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. *Syst. Biol.* 60, 719–731
49. Wiens, J. and Moen, D. (2008) Missing data and the accuracy of Bayesian phylogenetics. *J. Syst. Evol.* 46, 307–314
50. Wiens, J.J. and Tiu, J. (2012) Highly incomplete taxa can rescue phylogenetic analyses from the negative impacts of limited taxon sampling. *PLoS ONE* 7, e42925
51. Lemmon, A.R. *et al.* (2009) The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. *Syst. Biol.* 58, 130–145
52. Simmons, M.P. (2011) Misleading results of likelihood-based phylogenetic analyses in the presence of missing data. *Cladistics* 28, 208–222
53. Sansom, R.S. and Wills, M.A. (2013) Fossilization causes organisms to appear erroneously primitive by distorting evolutionary trees. *Sci. Rep.* 3, 5
54. Sansom, R.S. *et al.* (2010) Non-random decay of chordate characters causes bias in fossil interpretation. *Nature* 463, 797–800
55. Reisz, R.R. and Muller, J. (2004) Molecular timescales and the fossil record: a paleontological perspective. *Trends Genet.* 20, 237–241
56. Lee, M.S.Y. *et al.* (2014) Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds. *Science* 345, 562–566
57. Donoghue, P.C.J. and Smith, M.P., eds (2003) *Telling the Evolutionary Time: Molecular Clocks and the Fossil Record*, CRC Press
58. Smith, A.B. (2000) Stratigraphy in phylogeny reconstruction. *J. Paleontol.* 74, 763–766
59. Alroy, J. (2002) Stratigraphy in phylogeny reconstruction – reply to Smith (2000). *J. Paleontol.* 76, 587–589
60. Wagner, P.J. (2002) Testing phylogenetic hypotheses with stratigraphy and morphology – a comment on Smith (2000). *J. Paleontol.* 76, 590–593
61. Fisher, D.C. *et al.* (2002) Stratigraphy in phylogeny reconstruction – comment on Smith (2000). *J. Paleontol.* 76, 585–586
62. Sumrall, C.A. and Brochu, C.A. (2003) Resolution, sampling, higher taxa and assumptions in stratocladistic analysis. *J. Paleontol.* 77, 189–194
63. Wickström, L.M. and Donoghue, P.C.J. (2005) Cladograms, phylogenies and the veracity of the conodont fossil record. *Special Papers Paleontol.* 73, 185–218
64. Benton, M.J. *et al.* (2000) Quality of the fossil record through time. *Nature* 403, 534–537
65. Fortey, R.A. and Jefferies, R.P.S. (1982) Fossils and phylogeny – a compromise approach. In *Problems of Phylogenetic Reconstruction. Systematics Association Special Volume 21* (Joysey, K.A. and Friday, A.E., eds), pp. 197–234, Academic Press
66. Holland, S.M. (2000) The quality of the fossil record: a sequence stratigraphic perspective. *Paleobiology* 26 (Suppl.), 148–168
67. Behrensmeyer, A.K. *et al.* (2000) Taphonomy and paleobiology. *Paleobiology* 26, 103–147
68. Ho, S.Y. and Lanfear, R. (2010) Improved characterisation of among-lineage rate variation in cetacean mitogenomes using codon-partitioned relaxed clocks. *Mitochondrial DNA* 21, 138–146
69. Yang, Z. (1996) Maximum-likelihood models for combined analyses of multiple sequence data. *J. Mol. Evol.* 42, 587–596
70. Nylander, J.A. *et al.* (2004) Bayesian phylogenetic analysis of combined data. *Syst. Biol.* 53, 47–67

71. Duchene, S. and Ho, S.Y. (2014) Using multiple relaxed-clock models to estimate evolutionary timescales from DNA sequence data. *Mol. Phylogenet. Evol.* 77, 65–70
72. Thornhill, A.H. *et al.* (2012) Are pollen fossils useful for calibrating relaxed molecular clock dating of phylogenies? A comparative study using Myrtaceae. *Mol. Phylogenet. Evol.* 63, 15–27
73. Bromham, L. *et al.* (2002) Testing the relationship between morphological and molecular rates of change along phylogenies. *Evolution* 56, 1921–1930
74. Seligmann, H. (2010) Positive correlations between molecular and morphological rates of evolution. *J. Theor. Biol.* 264, 799–807
75. Davies, T.J. and Savolainen, V. (2006) Neutral theory, phylogenies, and the relationship between phenotypic change and evolutionary rates. *Evolution* 60, 476–483
76. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*, Cambridge University Press
77. Haldane, J.B.S. (1949) Suggestions as to quantitative measurement of rates of evolution. *Evolution* 3, 51–56
78. Gillespie, J.H. (1991) *The Causes of Molecular Evolution*, Oxford University Press
79. Lee, M.S.Y. *et al.* (2013) Rates of phenotypic and genomic evolution during the cambrian explosion. *Curr. Biol.* 23, 1889–1895
80. Beck, R.M.D. and Lee, M.S.Y. (2014) Ancient dates or accelerated rates? Morphological clocks and the antiquity of placental mammals. *Proc. R. Soc. B: Biol. Sci.* 281, 10
81. Marshall, C.R. (1990) Confidence-intervals on stratigraphic ranges. *Paleobiology* 16, 1–10
82. Benton, M.J. *et al.* (2009) Calibrating and constraining molecular clocks. In *The Timetree of Life* (Hedges, S.B. and Kumar, S., eds), pp. 35–86, Oxford University Press
83. Warnock, R.C.M. *et al.* (2015) Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. *Proc. R. Soc. B: Biol. Sci.* 282
84. Warnock, R.C. *et al.* (2012) Exploring uncertainty in the calibration of the molecular clock. *Biol. Lett.* 8, 156–159
85. Inoue, J. *et al.* (2010) The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. *Syst. Biol.* 59, 74–89
86. Zhang, J. and Rasnitsyn, A. (2006) New extinct taxa of Pelecinidae sensu lato (Hymenoptera:Proctotrupidea) in the Laiyang Formation, Shandong, China. *Cretaceous Res.* 27, 684–688
87. Hu, C. *et al.* (2001) *Shantungosaurus giganteus*, Beijing Geological Publishing House (in Chinese)
88. Chen, P. *et al.* (2005) Jianshangou Bed of the Yixian Formation in West Liaoning, China. *Sci. China Series D: Earth Sci.* 48, 298–312
89. Chen, P. *et al.* (2006) Geological ages of track bearing formations in China. *Cretaceous Res.* 27, 22–32
90. Zhou, Z. *et al.* (2003) An exceptionally preserved Lower Cretaceous ecosystem. *Nature* 421, 807–814
91. Zhou, Z. (2006) Evolutionary radiation of the Jehol Biota: chronological and ecological perspectives. *Geological J.* 41, 377–393
92. Wang, S. *et al.* (2001) Further discussion on geologic age of Sihetun vertebrate assemblage in western Liaoning China: evidence from Ar–Ar dating. *Petrol. Sinica* 17, 663–668
93. Ling, W. *et al.* (2007) Zircon U–Pb dating on the Mesozoic volcanic suite from the Qingshan Group stratotype section in eastern Shandong Province and its tectonic significance. *Sci. China Series D: Earth Sci.* 50, 813–824
94. Sadler, P.M. (2004) Quantitative biostratigraphy – achieving finer resolution in global correlation. *Annu. Rev. Earth Planet. Sci.* 32, 187–213
95. Graur, D. and Martin, W. (2004) Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends Genet.* 20, 80–86
96. Ho, S.Y. (2009) An examination of phylogenetic models of substitution rate variation among lineages. *Biol. Lett.* 5, 421–424
97. Marshall, C.R. (1997) Confidence intervals on stratigraphic ranges with nonrandom distributions of fossil horizons. *Paleobiology* 23, 165–173
98. Butlin, R. *et al.* (2012) What do we need to know about speciation? *Trends Ecol. Evol.* 27, 27–39
99. Maynard Smith, J. *et al.* (1985) Developmental constraints and evolution. *Q. Rev. Biol.* 60, 265–287
100. Davis, C.C. *et al.* (2014) Long-term morphological stasis maintained by a plant–pollinator mutualism. *Proc. Natl. Acad. Sci. U.S.A.* 111, 5914–5919