Syst. Biol. 58(3):367–380, 2009
Copyright © Society of Systematic Biologists
DOI:10.1093/sysbio/syp035
Advance Access publication on July 3, 2009

Accounting for Calibration Uncertainty in Phylogenetic Estimation of Evolutionary Divergence Times

SIMON Y. W. HO* AND MATTHEW J. PHILLIPS

Centre for Macroevolution and Macroecology, Research School of Biology, Australian National University, Canberra, ACT 0200, Australia;

**Correspondence to be sent to: Botany and Zoology, Building 116, Daley Road, Australian National University, Canberra, ACT 0200, Australia; E-mail: simon.ho@anu.edu.au.*

The estimation of phylogenetic divergence times from sequence data is an important component of many molecular evolutionary studies. There is now a general appreciation that the procedure of divergence dating is considerably more complex than that initially described in the 1960s by Zuckerkandl and Pauling (1962, 1965). In particular, there has been much critical attention toward the assumption of a global molecular clock, resulting in the development of increasingly sophisticated techniques for inferring divergence times from sequence data. In response to the documentation of widespread departures from clocklike behavior, a variety of local- and relaxed-clock methods have been proposed and implemented. Local-clock methods permit different molecular clocks in different parts of the phylogenetic tree, thereby retaining the advantages of the classical molecular clock while casting off the restrictive assumption of a single, global rate of substitution (Rambaut and Bromham 1998; Yoder and Yang 2000).

At around the same time, Sanderson (1997) published his nonparametric rate-smoothing algorithm, which operates by minimizing the magnitude of rate changes between adjacent branches in the tree. A related method, penalized likelihood, was subsequently implemented in a maximum-likelihood framework (Sanderson 2002). In this approach, large rate changes between neighboring branches are penalized. The degree of penalization is determined by a smoothing parameter, the value of which is obtained objectively through a cross-validation procedure.

There have been a number of implementations of relaxed-clock models in Bayesian phylogenetic frameworks. Some of these models assume that substitution rates are autocorrelated between branches, which can

Downloaded from https://academic.oup.com/sysbio/article/58/3/367/1613552 by guest on 13 March 2024

be done by allowing the rate to change or "evolve" over time (Thorne et al. 1998; Kishino et al. 2001; Lepage et al. 2006; Rannala and Yang 2007; Ho 2009). Others assume that branch-specific rates are drawn from a single underlying distribution, such as a lognormal, gamma, or exponential distribution, the parameters of which are estimated from the data (Drummond et al. 2006; Lepage et al. 2007; Rannala and Yang 2007). The available relaxed-clock methods have been compared in several reviews (Magallón 2004; Welch and Bromham 2005; Lepage et al. 2006; Rutschmann 2006), and their performance has been assessed in a number of studies (e.g., Ho et al. 2005; Drummond et al. 2006; Lepage et al. 2007).

The new relaxed-clock methods have also introduced more flexible techniques for incorporating calibrations, leading to a lively discussion about approaches to calibrating estimates of divergence times (Graur and Martin 2004; Hedges and Kumar 2004; Donoghue and Benton 2007; Ho 2007). Calibrations are of central importance in divergence dating analyses because it is not possible to estimate absolute ages from molecular data alone. Observed genetic divergence is the product of 2 components (the substitution rate and the time elapsed) that cannot be separated without additional, independent information. Such data can come in 2 main forms. The first are calibrations that impart temporal information about nodes in the evolutionary tree. This category can be further divided into calibrations at terminal and internal nodes in the tree. Calibrations at terminal nodes (i.e., the tips or leaves of the tree) are only possible when the analysis involves heterochronous sequence data, such as those from serially sampled viruses or ancient DNA (Rambaut 2000), or when dated fossils are included in an analysis involving morphological characters (Lee et al. 2009). Calibrations at internal nodes (representing divergence of coalescent events) are usually based on the fossil record or dated biogeographic events, and require some a priori knowledge about the relationships among the taxa being analyzed (Fig. 1). Rate estimates can also be readily calibrated if the sequence data have been sampled from individuals with a known degree of intergenerational separation, which can be the case for data obtained from studies of pedigrees and of mutation accumulation lines (e.g., Howell et al. 2003; Haag-Liautard et al. 2007); in these cases, the age of at least one internal node is known. The second form of calibrating information is a known substitution rate that has been estimated independently. It should be noted, however, that such a rate estimate would have originally depended on some form of age calibration.

Until relatively recently, it was only possible to incorporate calibrating information into a phylogenetic analysis by fixing the age of at least one internal node to a point value, or by applying a single, errorless rate throughout the tree. This is still the case in methods such as haplotype network analysis, maximum-likelihood quartet dating (Rambaut and Bromham 1998), and some Bayesian dating techniques (Aris-Brosou and Yang 2002). In attempts to model the uncertainty in calibration points more realistically, a range of methodological developments have taken place over the past decade. The various calibration techniques are illustrated in Figure 2.

The first advance was to allow calibrations to be given in the form of minimum or maximum bounds on the ages of internal nodes. This was implemented in a variety of relaxed-clock phylogenetic methods, including nonparametric rate smoothing (Sanderson 1997), penalized likelihood (Sanderson 2002), and Bayesian relaxed-clock inference (Thorne et al. 1998), and more recently in the PATHd8 method of Britton et al. (2007; subsequently revised by Svennblad 2008). To mitigate the impact of potential errors in calibration choice, Yang and Rannala (2006) relaxed these hard bounds by assigning a nonzero probability to dates outside the bounds, leading to the introduction of soft bounds. The most recent methods, developed independently by Drummond et al. (2006) and Yang and Rannala (2006), have been presented in Bayesian frameworks and allow the user to specify the prior age distributions for selected nodes in the tree. First used in a study by Weinstock et al. (2005), this extended the method employed earlier by Barnett et al. (2005), who had assumed a lognormal *actual* distribution for the age of their calibration node. By allowing calibrating information to be represented in the form of parametric distributions, a high degree of flexibility is offered with respect to incorporating a time scale into a phylogenetic analysis.

In this paper, we compare these various calibration techniques and explain how they can be used to capture and utilize the temporal information contained in the fossil record, biogeographic events, and radiometric dates. We then discuss various issues and difficulties

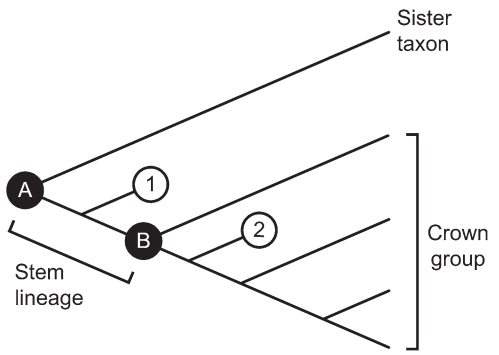

FIGURE 1. Tree illustrating the use of fossil taxa for calibrations. The crown group contains all extant members of a clade, their most recent common ancestor, and all extant and extinct descendants of that ancestor. The stem lineage is the branch that joins the crown group to the most recent common ancestor of the crown group and its existant sister clade. Fossil Taxon 1 can share synapomorphies that define the crown group and can be used to put a minimum age on Node A. Fossil Taxon 2 possesses all of the defining synapomorphies but exhibits additional apomorphies that allow it to be assigned unambiguously to one of the lineages of the crown group. Thus, Fossil Taxon 2 can be used to put a minimum constraint on the age of Node B. Note that fossil taxa assigned to the sister lineage are able to provide a minimum constraint on the age of Node A.

Downloaded from https://academic.oup.com/sysbio/article/58/3/367/1613552 by guest on 13 March 2024

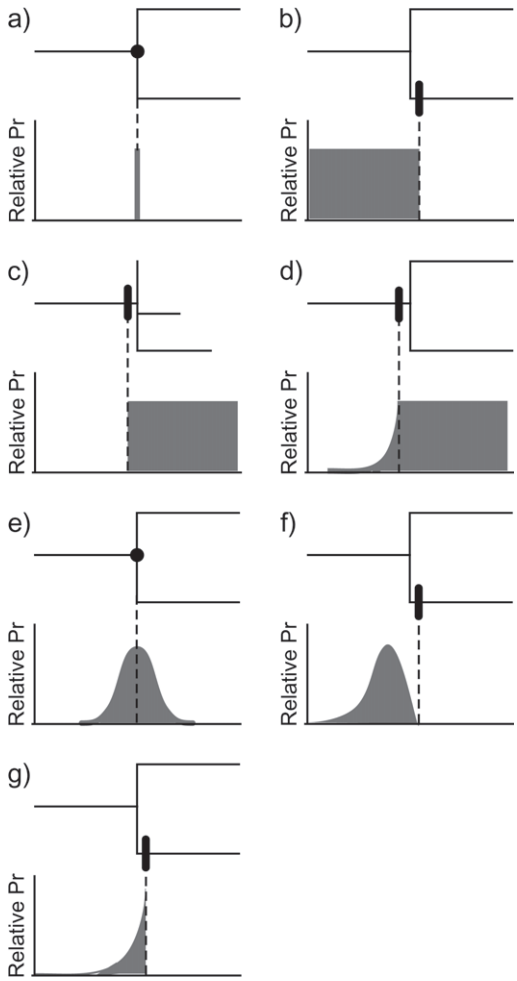

FIGURE 2. Methods for incorporating calibrations into a dating analysis: a) point calibration; b) hard minimum bound; c) hard maximum bound; d) soft maximum bound; e) normal distribution; f) lognormal distribution; g) exponential distribution. Relative probability is measured on the vertical axis of each graph.

concerning calibration methodology, particularly the use of multiple calibrations. Finally, we present a brief case study examining the effects of using different calibration techniques.

CALIBRATIONS AT INTERNAL NODES

Point Calibrations

Traditionally, divergence time estimates have been calibrated by fixing the age of at least 1 node in the tree to a point value (Fig. 2a). For example, in order to estimate the timing of mammalian globin gene duplications, Zuckerkandl and Pauling (1965) assumed that the most recent common ancestor of placental mammals existed 80 Ma ago. This calibration was not derived from specific fossil evidence but was an estimate of the basal mammalian divergences based on the Tertiary fossil record. Subsequent studies have tended to provide more explicit justifications of fossil-based calibration choices.

Converting fossil evidence into a point calibration is only appropriate if the fossil taxon represents the actual common ancestor of 2 extant lineages (e.g., Nodes A and B in Fig. 1), which is highly unlikely. In any case, the age of the fossil needs to be estimated, and this can come with errors relating to stratigraphic interpretation and radiometric dating (Magallón 2004; Gandolfo et al. 2008). For some fossils, these uncertainties can be of almost negligible magnitude because accurate stratigraphic constraints are available (Benton and Donoghue 2007), but in most instances there can be a large degree of uncertainty owing to the unavailability of isotopic or paleomagnetic dates. One notable example is the Tingamarra fossil fauna in Australia, which includes the oldest Australasian marsupial fossils. The original Early Eocene (~55 Ma) radiometric date for the site means that the fauna includes the oldest known songbirds in the world and, potentially, the oldest bat fossil (Godthelp et al. 1992; Boles 1995). These claims were placed in doubt, however, by the interpretation of Woodburne and Case (1996) that the Tingamarra fauna is Late Oligocene, less than 30 Ma old. Substantial recent corroborating evidence appears to lay this argument to rest in favor of the Early Eocene date (Beck et al. 2008).

A major deficiency of point calibrations is that their employment leads to divergence time estimates that display illusory precision. This is also a problem when secondary calibrations (i.e., molecular date estimates obtained from other, independent analyses) are converted to point values without due consideration of associated error (Graur and Martin 2004). Ignoring calibration uncertainty has profound consequences for evolutionary hypothesis testing involving time scales because it increases the probability of type I errors. The use of point calibrations appears to have declined in recent years, primarily owing to strong criticisms (Graur and Martin 2004) as well as the availability and accessibility of more sophisticated methods. Nevertheless, they remain a popular technique in divergence dating studies (for a survey, see Ho 2007).

Hard Bounds

*Hard minimum bounds (*Fig. 2b*).*—It has long been recognized that fossil evidence is only able to provide hard minimum bounds on divergence times. Correctly identified fossils can provide confirmation that a lineage existed at a certain point in time, but using them to make precise statements about divergence events would be highly inadvisable because the actual lineage could have been in existence well before the appearance of the calibrating fossil. In Figure 1, for example, this would be equivalent to assuming that Fossil 1 is positioned exactly at Node A. It would be more appropriate to use Fossil 1 to place a minimum age constraint on Node A. Prior to the development of methods that could take this characteristic into account, some key dating studies attempted to avoid the problem by using

Downloaded from https://academic.oup.com/sysbio/article/58/3/367/1613552 by guest on 13 March 2024

well-constrained divergence events to provide point calibrations (e.g., Doolittle et al. 1996; Hedges et al. 1996). Eventually, the minimum-age nature of fossil evidence was formalized in the nonparametric rate-smoothing method (Sanderson 1997). In the implementation of this method, it was possible to specify minimum age constraints on any number of nodes in the tree. In an exemplary analysis by Sanderson (1997), in which he estimated the divergence times of seed plants, 2 such constraints were enforced: a minimum age of 320 Ma for the split between conifers and *Gingko*, and a minimum age of 125 Ma for eudicot angiosperms. In effect, hard bounds assign a probability of 0 to ages that fail to satisfy the constraint, while giving equal probability to all ages that are consistent with the constraint. Hard bounds are now available in many divergence dating methods (Thorne et al. 1998; Thorne and Kishino 2002; Drummond et al. 2006; Yang and Rannala 2006; Britton et al. 2007).

One of the weaknesses of using minimum bounds alone is that they usually provide insufficient information for obtaining a unique set of date estimates. In the absence of other constraints, there are no upper limits to divergence times, which can then take arbitrarily high values while the estimated substitution rate is driven toward very low values. The need for some form of maximum constraint is especially acute for the nonparametric rate-smoothing and penalized-likelihood methods, in which there is no other information about divergence times (Sanderson 2003); it is necessary to include at least one other type of calibration, either a maximum bound or point calibration. Whereas this measure assists in obtaining a unique solution, its effects on the accuracy of date estimates are more debatable. In Bayesian divergence-dating methods, all of the nodes in the tree, including the noncalibrating nodes, have a prior age distribution. This prior can be based on a stochastic branching process (Thorne et al. 1998; Aris-Brosou and Yang 2002; Drummond et al. 2006; Rannala and Yang 2007) or in an alternative manner (Kishino et al. 2001). Even in these methods, a more explicit prior on the age of the root is desirable.

*Hard maximum bounds (*Fig. 2c*).*—Hard maximum bounds are probably the most controversial type of calibration, and the difficulty in choosing them has been well documented (Yang and Rannala 2006; Benton and Donoghue 2007; Hug and Roger 2007). A maximum age bound on a clade is equivalent to a confident assumption that the clade is absent (i.e., yet to evolve) at that time, but this cannot be explicitly supported by the fossil record. When specifying maximum age bounds, it is necessary to reach a compromise between choosing a young bound that might exclude the true divergence time, and choosing an old bound that is too uninformative (Yang and Rannala 2006; Marshall 2008). The reality of this dilemma was confirmed empirically by Hug and Roger (2007), who found that using arbitrarily large maximum bounds yielded unreasonably ancient divergence-time estimates.

Marshall (2008) recently proposed an objective method for establishing maximum age constraints using the fossil record. The technique requires an estimate of the phylogeny, along with a set of fossil calibrations that are assumed to be correctly assigned and dated. The method presents a potentially useful addition to the growing suite of quantitative methods for the objective specification of calibrations, but it requires a number of significant assumptions about random preservation and recovery of fossils as well as a reliable estimate of the phylogeny.

Phylogenetic bracketing is another technique that has been used to establish hard maximum bounds (Muller and Reisz 2005). Also known as nearest-neighbor calibration (Hug and Roger 2007), this involves placing age constraints onto a node by referring to the ages of adjacent nodes. In this framework, maximum bounds are defined by the oldest date inclusive of relatively well-sampled fossil assemblages in potential geographic regions of origin that contain no putative members of the clade of interest, but which contain stem members or ecological equivalents (Barnett et al. 2005; Benton and Donoghue 2007).

A noticeable limitation of phylogenetic bracketing is that it requires considerably more a priori knowledge about the phylogenetic relationships among the study organisms than does the direct calibration of a single node. Furthermore, the fossil record is often too limiting for phylogenetic bracketing to be effective in determining hard maximum constraints (Benton and Donoghue 2007). As an example, *Vegavis* was reported by Clarke et al. (2005) to provide a minimum bound for the divergence of ducks from the magpie-goose. However, in the absence of well-accepted neornithine (modern) birds from earlier sites, this single fossil taxon also serves to provide the oldest representative for the next 4 consecutively older crown nodes among birds, which molecular dating suggests diverged over a period of 40 Ma or more (Brown et al. 2008).

The fundamental problem with maximum bounds is that absence of evidence typically does not provide sufficient evidence of absence. When attempting to infer the timing of origination of a taxon, 2 common errors are associated with literal readings of the fossil record: local absences and taxonomic/taphonomic biases (Penny and Phillips 2004; Donoghue and Benton 2007). For example, the absence of diprotodontian marsupials (e.g., kangaroos, possums, and the koala) from the above-mentioned Tingamarra site (~55 Ma in age), which includes the earliest known Australasian marsupial fossil fauna, is not necessarily indicative of a later origination. Diprotodontians may have been restricted to Antarctica, for which there is a mammalian fossil hiatus at that time. Alternatively, stem diprotodontians might have been larger than their other marsupial contemporaries and, as noted by Beck et al. (2008), all of the Tingamarra mammal fossils are very small and so may reflect size-biased preservation.

Biogeographic and geological events, such as the formation of islands, can offer plausible instances of max-

Downloaded from https://academic.oup.com/sysbio/article/58/3/367/1613552 by guest on 13 March 2024

imum age bounds, but these bounds must be chosen carefully (Heads 2005; Renner 2005; Emerson 2007). The accuracy of biogeographic calibrations depends on the availability of reliable geological dating, and are often based on assumptions about vicariance and dispersal. In most cases, however, we cannot be confident that genetic divergence could not have antedated the origin or disappearance of a putative biogeographic barrier. On the other hand, lineage divergence may have occurred much later than geological separation, perhaps as a result of dispersal events, as appears to be the case for plants in New Caledonia and New Zealand (Pole 2001). Some of the difficulties with choosing biogeographic calibrations are discussed in subsequent sections below, and examples are given in the section on normal calibration priors.

Soft Bounds

As described above, hard bounds represent a substantial improvement on the use of point calibrations, but they still have several significant drawbacks. First, it is possible that the supporting fossil evidence has been misinterpreted, meaning that the minimum bound is incorrectly specified. This can have a large, detrimental impact on date estimates, especially if a fossil is assigned to the wrong clade (Yang and Rannala 2006). Second, owing to their speculative nature, maximum bounds generally need to be chosen in a very conservative manner. This can entail the disposal of useful paleontological data, such as the informed estimates of uncertainty that can be provided by paleontologists, in order to establish safe maximal ages for nodes. To address these shortcomings, Yang and Rannala (2006) proposed soft bounds, which assign a nonzero probability to all age values (Fig. 2d). These differ from hard bounds in that they add a diminishing tail of probability beyond the bound. This tail, which contains 2.5% of the total probability, follows a power or exponential decay, depending on whether the soft bound is a minimum or maximum (Yang and Rannala 2006).

The most obvious advantage of soft bounds, in comparison with hard bounds, is that the molecular data can overcome poor calibrations, leading to an improvement in accuracy (Yang and Rannala 2006). However, this will tend to happen only when good calibrations are also present in the analysis. Yang and Rannala (2006) described several other advantages of soft bounds, including a more reliable evaluation of estimation error. Disparities between the use of soft and hard bounds are most evident when the signal from the molecular data conflicts with that generated by the calibrations (Yang and Rannala, 2006).

Parametric Distributions

Several parametric distributions are available as priors on nodal ages in a Bayesian phylogenetic framework. The first implementation was in the method of Thorne et al. (1998), which *required* the specification of a normally-distributed prior on the age of the in-group. Subsequent implementations by Yang and Rannala (2006) and Drummond et al. (2006) enabled parametric prior distributions to be specified for the ages of other nodes in the tree. Below, we discuss several prior distributions that can be used for calibrating molecular clocks, indicating how they can be employed in different situations. We illustrate these explanations with examples drawn primarily from published case studies.

Normal distribution (Fig. 2e).—A normal distribution is symmetric, with most of the probability density around the mean and with tails of diminishing probability above and below the mean. The shape of this distribution allows for bidirectional uncertainty in estimates of divergence times. This feature makes it a useful alternative to the lognormal, exponential, and gamma distributions, all of which involve the specification of a hard, nonzero minimum bound when used for calibration purposes.

The nature of this distribution generally makes it unsuitable for summarizing fossil information, with some exceptions. A normal distribution can be used in situations where there is little justification for weighting the probability toward the minimum bound, as is done to varying extents in certain cases of the lognormal, exponential, and gamma distributions. For example, a fossil might exhibit a suite of apomorphic characters that are present in its descendent lineages, while retaining a very small number of plesiomorphies. This would imply that the fossil lies on the stem lineage, but is very close to the base of the crown group. In Figure 1, this is equivalent to Fossil Taxon 1 being very close to Node B, which implies that Node A could be much older than the fossil. In this case, we might expect that the crown group diverged from its sister group substantially earlier than the fossil date. The fossil record for the clade might also require that any maximum bound on the divergence age between the 2 groups be very conservative. In such instances, a normal distribution in which the bulk of the probability density is more central than the acknowledged conservative lower and upper bounds might reflect our prior beliefs better than a uniform distribution.

In order to provide a prior on the root of the avian tree, *Vegavis* at 66 Ma provides an appropriate minimum bound (Clarke et al. 2005; Benton and Donoghue 2007). As discussed above, this undoubted anseriform is expected to underestimate substantially the basal divergence of modern birds. Similarly, in order to allow for the possibility that some ancient fossil birds (e.g., *Gansus*) might fall within Neornithes, we can place a maximum bound at 121 Ma, before which only definite non-neornithines are known. Hence, we expect both the minimum and maximum bounds to be very conservative; a normal distribution, which places higher probability on intermediate dates, might offer an appropriate calibration prior here. An important question is whether calibrating on the basis of such considerable paleontological uncertainty is desirable at all.

Downloaded from https://academic.oup.com/sysbio/article/58/3/367/1613552 by guest on 13 March 2024

The symmetry of the normal distribution makes it suitable for some biogeographic calibrations, particularly because it can be inadvisable to specify hard bounds (Heads 2005). This reflects the prior expectation that the most likely divergence is approximately in the center of less likely, but plausible, older and younger dates. For example, following the Last Glacial Maximum around 18–20 ka ago, many European mammal species expanded their ranges after being restricted to refugia (Hewitt 2000). For monophyletic populations, the timing of glacial retreat and associated events can be used to inform calibrations (e.g., Rajabi-Maham et al. 2008). In a recent study of rates of human mitochondrial evolution, Endicott and Ho (2008) specified normally-distributed priors for the ages of Eurasian Haplogroups H1 and H3, with means of 18 ka and standard deviations of 3.5 ka (i.e., with 95% of the distribution lying between 11.1 and 24.9 ka). These 2 haplogroups are presumed to have expanded and diversified in situ, following postglacial migration into Europe. A normal prior allows for the chance that genetic diversification was associated with postglacial population growth as well as the possibility that the basal genetic divergences antedated this event.

Normal distributions are also valid for importing substitution rates from other analyses (e.g., Saarma et al. 2007). This approach is necessary in cases where other calibrating information is absent and represents a considerable improvement upon the importation of rates without proper recognition of the associated error, which can be very substantial in population-level data sets.

Lognormal distribution (Fig. 2f).—The lognormal distribution can take on a variety of skewed shapes, depending on the values of its parameters. Here we limit the discussion of this distribution to the shape shown in Figure 2f, in which the distribution has a distinct, nonzero mode. Unlike the normal distribution, it has only a single unbounded tail of declining probability. In total, a lognormal calibration prior requires 3 parameters to be chosen: the mean, the standard deviation, and the hard minimum bound. The distribution can also be described by its mode (peak of highest probability).

It is often the most appropriate distribution for summarizing paleontological information because it can assign the highest point probability for the nodal age to be somewhat older than the oldest fossil. When taking the shape shown in Figure 2f, the lognormal distribution gives zero probability to the nodal age actually being equivalent to that of the oldest fossil. It can be difficult to select objectively a value for the mode of the lognormal distribution.

Once the mean and hard minimum bound have been chosen, the standard deviation dictates the length of the distribution's tail. The value of the standard deviation can be chosen so that 95% of the probability density lies between the hard minimum bound and some "soft" maximum bound (note that this is not the same as the soft bounds described above and illustrated in Fig. 2d). The value of the maximum bound can be obtained, not without some difficulty, using an objective procedure (Marshall 2008). The relevant considerations are discussed later in this paper.

Ecologically transitional fossils can provide excellent candidates for calibrating molecular clocks. This is because the numerous characters that unequivocally define their phylogenetic placement are often expected to have evolved over a relatively short period of time. One such fossil is *Waimanu*, the earliest known penguin (Slack et al. 2006). Its geological date of ~61 Ma provides a hard minimum bound. The use of a lognormal distribution incorporates this bound as well as a tail allowing some probability of its divergence from other seabirds being as early as 74 Ma ago. This soft maximum accounts for putative fossil relatives of penguins that may be latest Cretaceous in age, while acknowledging the absence of putative close relatives before the late Campanian (~74 Ma). Instead of placing the mode in the center of this distribution, the expectation of rapid evolution in ecologically transitional penguins and the solid fossil evidence at the young end, as opposed to absence of evidence at the old end of the distribution, might lead us to place the mode close to 61 Ma. Exact timing is somewhat subjective; however, a prior belief that early penguin evolution was associated with faunal turnover at the Cretaceous–Tertiary boundary would suggest a mode at ~65 Ma before present.

If no other calibrating information is available, it may be necessary to import a secondary calibration. Typically, divergence times estimated from molecular data are lognormally distributed (Morrison, 2008); thus, a lognormal prior can be used to import secondary calibrations, while allowing the estimation error from the original date estimate to be taken into account.

Exponential distribution (Fig. 2g).—The exponential distribution shares some features with the lognormal distribution, namely a long tail of diminishing probability toward higher ages and, when used for calibration, a hard, nonzero minimum bound. The key difference between the exponential and lognormal distributions is the location of the mode, which is always at the hard minimum in the former. The exponential distribution requires 2 parameters to be chosen: the mean and the hard minimum bound. As there is one fewer parameter, usage of this distribution might be preferred over a lognormal distribution if there is inadequate paleontological information, particularly if there is no prior expectation of what the mode of the distribution might be. As the exponential distribution assigns the highest probability to a nodal age being equivalent to the oldest fossil (i.e., at the minimum bound), it can also be used when there is strong expectation that the oldest fossil lies very close to the divergence event being represented by the node, relative to a distant, soft maximum. This situation could occur when paleontological information points to a small window of time in which the divergence is most likely to have occurred, but where some

Downloaded from https://academic.oup.com/sysbio/article/58/3/367/1613552 by guest on 13 March 2024

probability of an earlier divergence must be allowed over a longer period during which the fossil record is poorer.

The divergence between archosaurs (e.g., birds and crocodiles) and lepidosaurs (e.g., tuatara and lizards) provides an example for which an exponential distribution may be appropriate. Muller and Reisz (2005) and Benton and Donoghue (2007) recently examined the utility of this divergence for calibration. A good fossil record for these groups in the Late Permian allows a hard minimum to be set at ~252 Ma ago, with an expectation that the actual divergence occurred within the previous 10 Ma. However, the fossil record for diapsid reptiles is temporally and geographically more patchy earlier in the Permian, such that our criterion for placing a maximum bound cannot be met until we go back as far as the base of the Permian (~299 Ma). Such a long-tailed prior, with substantial weighting at the younger end, can also be fitted by a lognormal distribution; however, with one fewer parameter, an exponential distribution may be justifiable.

Gamma distribution.—The gamma distribution has 2 parameters, which dictate the shape (alpha) and scale (beta) of the distribution. When used to specify calibration priors, a hard minimum bound also needs to be chosen. The gamma distribution can take on a variety of shapes, such as matching the exponential (alpha = 1) or tending towards the normal (as alpha $\rightarrow \infty$). In this respect, the gamma is similar to the lognormal, and the selection of parameter values for the gamma distribution can be guided by the same considerations as those described for the 3 distributions above.

Calibrations at Terminal Nodes

Point Calibrations

Point calibrations are routinely used in analyses of heterochronous data. For serially-sampled viral sequences, this calibration method is appropriate because the sampling dates are generally well documented in patient records. In some cases, even the transmission history of the viral infection is known (Leitner and Albert 1999). Ancient DNA sequences, however, are often obtained from prehistoric samples, with a few exceptions such as museum specimens that can usually be treated as effectively modern in age (Wandeler et al. 2007). For most samples, it is necessary to estimate their age, which is often done using some form of isotopic dating. Estimation error associated with radiocarbon dates is usually ignored in phylogenetic analyses of ancient DNA (e.g., Lambert et al. 2002; Ho et al. 2007).

In studies of morphological data, point calibrations can be associated with fossil taxa. In analyses of combined molecular and morphological data sets, the fossil taxa represent heterochronous data and can act as dated terminal nodes. This also removes the need to place fossil taxa onto the phylogenetic tree a priori, as this is done objectively using the morphological characters. A major weakness of this combined-data approach is that it requires assumptions about rates of morphological change, the estimates of which will be highly dependent on having realistic statistical models of morphological evolution.

Hard Bounds

Ancient DNA can be hundreds of thousands of years old, well beyond the 45–50 ka limit of radiocarbon dating (Willerslev and Cooper 2005). If the age of such a sample cannot be estimated by other means, it is not possible to summarize it in the form of a point calibration. Generally, radiocarbon analysis of such samples will yield a minimum age, with no indication of how much older the sample could actually be. For example, the ages of most of the known samples from 1 of the 2 major clades of Siberian woolly mammoths have not yielded finite radiocarbon dates, making it difficult to determine the exact age of disappearance of this clade (Gilbert et al. 2008). Similarly, the ages of 3 of the sequences in the ancient brown bear data set of Barnes et al. (2002) are given as >53 900 years, >56 900 years, and >59 000 years. In these cases, it is appropriate to use hard minimum bounds as calibrations for these terminal nodes, if they are to be included in a phylogenetic dating analysis at all.

Hard minimum and maximum bounds can be used simultaneously (equivalent to a uniform prior over a specified interval) if a sample has been obtained from a well-defined stratum (Korsten et al. 2009). For example, this can be the case for ancient DNA retrieved from ice cores (e.g., Coolen and Overmann 2007; Willerslev et al. 2007) or from samples associated with archaeological horizons (e.g., Edwards et al. 2007). In these situations, it is necessary to eliminate the possibility of intrusions or DNA leaching between sediment layers (Haile et al. 2007).

Parametric Distributions

For many ancient samples, radiocarbon dating errors are small relative to their absolute ages. Unfortunately, when radiocarbon dates are recalibrated according to calendar years, the errors scale in a nonlinear manner so that they can exceed 5% of the date estimate for older samples. This suggests that it might be more fitting to model the uncertainty in sample ages using parametric distributions. A common, but incorrect, assumption is that radiocarbon error is normally distributed. In reality, the error can be highly asymmetric and even multimodal (Reimer et al. 2004). Nevertheless, describing this uncertainty using a normal or gamma distribution is probably sufficient for the majority of ancient DNA studies, especially when they involve data sets with numerous ancient sequences.

Choosing the Level of Uncertainty

The opportunity to apply parametric prior distributions to calibrating molecular clocks is a welcome methodological advance, but it brings the new responsibility

Downloaded from https://academic.oup.com/sysbio/article/58/3/367/1613552 by guest on 13 March 2024

of having to decide on a level of calibration uncertainty. This represents a movement toward the quantification of our judgements about the quality of calibrations, which should also serve to make more transparent the process of choosing calibrations.

In essence, the following question needs to be considered: By what period of time could we reasonably expect the age of a node to predate the age of the oldest fossil on either of its descendent lineages? Answering this question is an exceptionally difficult task because the uncertainty is not simply a combination of quantifiable factors. Gandolfo et al. (2008) grouped the sources of uncertainty into 5 categories: (1) fossil preservation, (2) taxonomic assignment of the fossil, (3) identification of fossil homologies, (4) sampling effort, and (5) fossil age determination. As these factors have been discussed in detail elsewhere (Magallón 2004; Gandolfo et al. 2008), we merely provide brief explanations below.

The first factor concerns differential preservation probability. Rates of preservation will vary among taxa because of characteristics such as hard parts and proximity to water bodies. Furthermore, different parts of taxa can be preserved at disparate rates, leading to varying levels of uncertainty over the taxonomic affinity of fossil specimens. Some efforts have been made to develop methods for estimating preservation rates, but these depend on assumptions about species diversification (Foote 1997; Foote et al. 1999).

The next 2 factors, taxonomic assignment of the fossil and identification of fossil homologies, can both lead to incorrect placement of fossil taxa onto the phylogenetic tree, which would then yield erroneous calibrations. Neither of these factors is easily quantifiable, although one could conceivably use bootstrap values or posterior probabilities to weight competing placements of fossil taxa in the tree. For example, if 2 possible placements of a fossil taxon have posterior probabilities of 0.7 and 0.3, then these could be used to weight the relative contribution of the 2 implied calibrations. In a Bayesian setting, this can be done automatically if the dated fossil taxon is included in an analysis of a combined molecular and morphological data set (Lee et al. 2009).

The fourth factor, sampling effort, concerns the temporal and spatial completeness of the fossil record for a taxon of interest. Temporal biases in sampling effort are difficult, although not impossible, to quantify; for example, they can be assessed empirically or through phylogenetic means (Marshall 1997; Tavaré et al. 2002). Spatial biases are probably more difficult to account for.

At first glance, the final factor (fossil age determination) appears to be the most easily quantifiable of the 5. However, careful consideration of the process of determining fossil ages reveals that there are substantial sources of uncertainty that can complicate any quantification of error. Different dating methods can yield quite disparate age estimates for a single deposit. Whereas certain deposits can be precisely dated, many fossils are not well constrained stratigraphically. It is frequently the case that a fossil can only be localized to a particular geological epoch, which can span tens of millions of years. Coupled with changing definitions of geological epochs and periods over time, this creates a complex situation in which fossil age uncertainty becomes difficult to quantify (Gandolfo et al. 2008).

Considering these 5 factors, one could easily adopt a pessimistic outlook and conclude that there is no objective method for determining the level of calibration uncertainty and that the age estimation errors for individual calibrations depend on too many idiosyncratic features. At present, such an exercise clearly requires informed judgements from paleontologists and geologists. Nevertheless, there have been efforts to develop techniques for quantifying various aspects of the factors mentioned above (Strauss and Sadler 1989; Foote 1997; Marshall 1997; Foote et al. 1999), and it remains to be seen whether it is possible to combine these into a single function.

It is evident that estimating the level of uncertainty might simply be impossible for the majority of fossil calibrations. Additionally, it could be the case that the distribution of calibration error cannot be captured satisfactorily by any of the simple parametric distributions that have been hitherto implemented (Donoghue and Benton 2007).

Multiple Calibrations

A considerable number of date estimates have been obtained using single calibration points, with some prominent studies having relied exclusively on a calibration based on the bird–mammal divergence estimated at around 300 Ma ago (Lee, 1999; Hedges and Kumar, 2004). This paradigm is gradually being supplanted owing to the increasing use of relaxed-clock models, which call for the employment of multiple calibrations. By using a greater number of calibrations, a better understanding of the patterns of among-lineage rate variation can be achieved. For example, the PATHd8 method (Britton et al. 2007) splits the tree into local clocks depending on the locations of the calibrations. Presumably, improved estimation of rates throughout the tree also leads to the production of divergence time estimates that are more accurate (Ho et al. 2005).

The risks of using only a single calibration are evident. If the calibration is inaccurate then this will lead to biases in estimates of rates and dates. It is also known that it is more difficult to estimate dates for nodes that are distant from any calibrating nodes (Linder et al. 2005). Nevertheless, if one wishes to employ multiple calibrations, some data points will almost inevitably be less reliable than others. Several authors have proposed methods for assessing the quality of calibrations, with the aim of developing an objective framework for comparing candidate calibrations. Near et al. (2005) presented a cross-validation method, which can be used to test for the internal consistency of a set of proposed calibration points. The calibrations are tested individually, and any points that are poor predictors of the ages

Downloaded from https://academic.oup.com/sysbio/article/58/3/367/1613552 by guest on 13 March 2024

of other calibration nodes are discarded. Sanders and Lee (2007) proposed that the accuracy of calibrations can be tested in a Bayesian setting by comparing the posterior age distribution of a calibration node with the prior specified for it.

The method of Near et al. (2005) has several noticeable weaknesses, including the fact that calibrations are treated as point estimates rather than minimum age constraints (Parham and Irmis 2008). In this sense, calibrations that are rated poorly by this method can only be considered to be poor reflections of the estimated nodal ages, but this does not disqualify them from representing useful minimum age constraints. A second major shortcoming of the method, as pointed out by Marshall (2008), stems from its tendency to discard calibrations until the remaining ones form an internally consistent set. There is a risk that the method will treat the most informative accurate calibrations as outliers because they can be inconsistent with the majority of the remaining calibrations. Finally, it can be difficult to separate inferences about the quality of the calibration from other sources of error such as those associated with branch-length estimation and rate shifts across the tree (Hugall et al. 2007). In this regard, a greater number of calibrations can actually help to inform the evolutionary model.

We contend that problems with including less reliable calibration data are primarily limited to analyses that employ them as errorless calibration points. Indeed, the systematic exclusion of offending calibrations is likely to introduce a bias toward a molecular clock, which can be regarded as a form of calibration "linearization" (cf. tree linearization; Takezaki et al. 1995). If they are instead used to specify age constraints then the cost of including poor calibrations is minimized (but not eliminated). Overall, it is probably preferable to attempt to model among-lineage rate variation explicitly because the elimination of offending calibrations (or taxa, in the case of tree linearization) could introduce a bias into the analysis if rate heterogeneity is an attribute intrinsic to the molecular evolutionary process (Cutler 2000). One of the consequences of specifying multiple calibrations in a single analysis is that there is some degree of interaction among the calibrations, which is generally unavoidable because the ages of nodes in a tree are not mutually independent. This interaction causes the actual prior probabilities of the nodal ages to differ from the prior values specified for individual nodes. The interaction between calibrations will be most evident if they have been assigned to adjacent internal nodes. Furthermore, in a Bayesian setting, the prior on the tree (e.g., Yule or birth–death process) makes a statement about the relative ages of nodes, contributing to the prior distribution of nodal ages.

To illustrate the interaction between calibration priors, consider the simple example depicted in the tree in Figure 3. Assume that we specify 3 normally distributed calibration priors: Node x (mean 5.00, with 95% of the probability in the interval 3.04–6.96), Node y (10.00, 6.08–13.92), and the root (20.00, 12.16–27.84). These calibrations are specified separately for the 3 nodes, but when we sample from the joint prior we obtain slightly different distributions: Node x (4.89, 2.94–6.84), Node y (9.65, 6.01–13.39), and the root (17.84, 10.56–25.23). In this case, the prior ages for Nodes x and y almost match the individually specified priors, but the prior age for the root differs from the normal distribution that was originally chosen.

The combined effect of calibration priors on the other internal nodes, coupled with the tree prior, will lead to an implicit prior on the age of the root. For example, given a data set with 5 sequences (A, B, C, D, and E), assume that there are 2 calibrations in a Bayesian phylogenetic analysis (Fig. 3): (1) normal prior, with mean 10 and standard deviation 1, on the age of Node y; and (2) normal prior, with mean 5 and standard deviation 0.5, on the age of Node x. There is a Yule prior on the tree topology, which is not fixed. This configuration results in a prior age distribution for the root, with mean 28.2 and a 95% highest prior density of 8.0–74.4.

It is perhaps best to close this section with a cautionary note about the role of calibrating information. In phylogenetic analyses, calibrations (including imported rate estimates) play a unique role because they are the only source of information about absolute time. This is also the case with regard to calibration priors in a Bayesian setting, raising the question of how much influence priors should have on node times. In one Bayesian dating method, developed by Aris-Brosou and Yang (2002, 2003), the posterior estimates appeared to be strongly influenced by the priors (Ho et al. 2005; Welch and Bromham 2005). The answer to this question is likely to vary among data sets; with increasing size of the sequence alignment, the influence of the prior tends to diminish.

Calibrations and the Tree Topology

Estimating branch lengths and the tree topology are related aspects of phylogenetic inference. If the tree topology has not been fixed in a divergence dating anal-

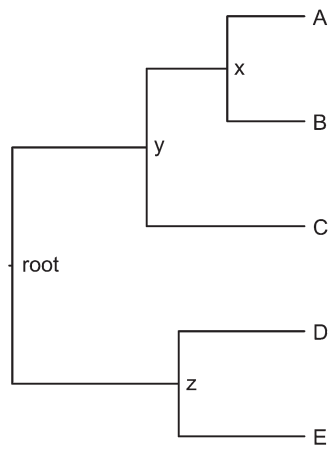

FIGURE 3. Simple phylogenetic tree used for the 3 simulation examples described in the text.

Downloaded from https://academic.oup.com/sysbio/article/58/3/367/1613552 by guest on 13 March 2024

ysis, calibrations on nodes can influence the process of topological estimation. This can be most clearly seen when priors are specified for the ages of adjacent nodes in a tree because it implies that there is an *a priori* estimate of the ordering of the nodes.

Take the example that was described in the preceding section, involving 5 sequences A, B, C, D, and E (Fig. 3). Assume that we have 3 calibrations: (1) normal prior, with mean 10 and standard deviation 1, on the age of Node y; and (2) normal prior, with mean 5 and standard deviation 0.5, on the age of Node x; and (3) normal prior, with mean 20 and standard deviation 2, on the age of the root. Upon investigating the prior distribution of trees, we find that a grouping of A and B is favored, with a prior probability of 0.66. This underlines the importance of sampling from the prior distribution in a Bayesian analysis, particularly in situations where the data set is relatively uninformative and the prior is likely to exert a greater influence on the posterior.

Uncertainties in the tree topology will tend to inflate the uncertainty on posterior date estimates, possibly leading to bimodal or multimodal posterior densities (Lee et al. 2009). This can occur if there are several competing topologies. If this is the case, one can either force groups of interest to be monophyletic or take the estimation error as being a reflection of both topological and temporal uncertainty. In general, it is probably better to avoid attempting to estimate the ages of poorly supported nodes altogether.

Case Study: Avian Diversification

One of the most debated events in avian evolution is the apparent origination of modern lineages in the early Tertiary. It has been suggested that the extinction of the dinosaurs and pterosaurs at the end of Cretaceous left open a wide range of ecological niches, allowing birds and mammals to diversify (Feduccia 1995). Most of the vacated "avian" niches were thought to have been filled by a rapid post-Cretaceous radiation of neognathous birds, which today include over 99% of all bird species and exclude only the palaeognaths (e.g., ostrich, emu, kiwi, and tinamous). This hypothesis has been extensively studied using molecular data (Hedges et al. 1996; Cooper and Penny 1997; Brown et al. 2008). Here we examine the effect of choosing different calibration techniques when estimating the age of the common ancestor of Neognathae.

The data set includes complete mitochondrial protein-coding sequences. Exclusion of the compositionally-biased third codon positions (see Harrison et al. 2004), as well as sequences with ambiguous homology, leaves 7324 nucleotide sites for 13 avian and nonavian taxa: ostrich (*Struthio camelus*, GenBank accession number NC_002785), giant tinamou (*Tinamus major*, NC_002781), chicken (*Gallus gallus*, NC_001323), magpie-goose (*Anseranas semipalmata*, NC_005933), redhead duck (*Aythya americana*, NC_000877), little blue penguin (*Eudyptula minor*, NC_004538), red-throated loon (*Gavia stellata*, NC_007007), American alligator (*Alligator mississippiensis*, AF069428), caiman (*Caiman crocodilus*, NC_002744), eastern painted turtle (*Chrysemys picta*, NC_002073), green turtle (*Chelonia mydas*, NC_000886), blue-tailed mole skink (*Eumeces egregious*, NC_000888), and iguana (*Iguana iguana*, NC_002793). Three calibrations were used for the purpose of estimating the divergence time of neognaths. The calibrations were treated differently in 3 separate analyses (Fig. 4): (1) all 3 calibrations were reduced to point-like calibrations, so that the ages of 3 nodes in the tree were constrained to intervals of 1 Ma;

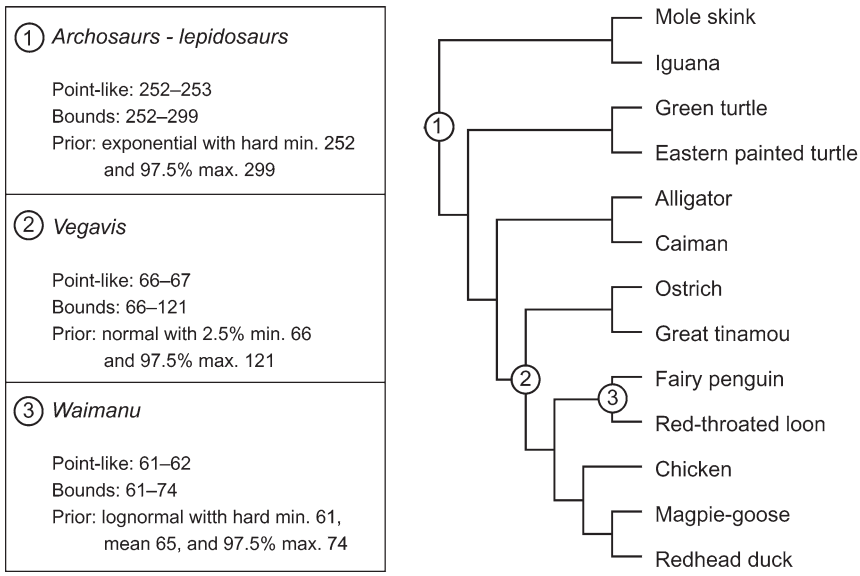

FIGURE 4. Phylogenetic tree of birds and reptiles inferred using mitochondrial protein-coding genes. Nodes used for calibration are numbered. The details of different calibration treatments for each node are given in the accompanying box, with ages given in millions of years before present. An explanation of the 3 treatments can be found in the main text.

Downloaded from https://academic.oup.com/sysbio/article/58/3/367/1613552 by guest on 13 March 2024

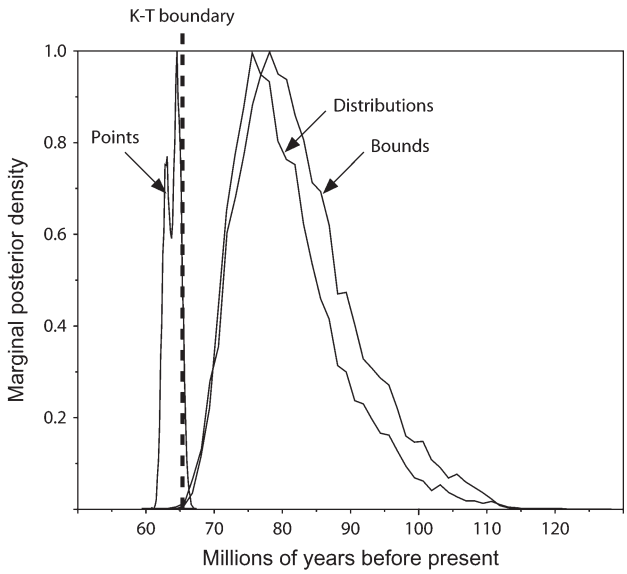

FIGURE 5. Posterior age estimates of the most recent common ancestor of neognath birds, obtained using 3 different treatments of calibration data: (1) point calibrations; (2) hard bounds; and (3) prior distributions. An explanation of the 3 treatments can be found in the main text. The Cretaceous–Tertiary (K-T) boundary is indicated on the graph.

(2) the 3 calibrations were used to provide hard minimum and maximum bounds (equivalent to a uniform prior) for the corresponding nodes; and (3) the calibrations were used to specify prior age distributions for the corresponding nodes. The derivation of these priors was discussed in the earlier sections on parametric priors.

The divergence time of Neognathae was estimated using BEAST 1.4.8 (Drummond and Rambaut 2007) for each of the 3 calibration regimes. The alignment was partitioned into first and second codon sites, and a GTR+I+Γ substitution model was employed for each partition in accordance with both the Akaike information criterion and likelihood-ratio test recommendations from ModelTest 3.06 (Posada and Crandall 1998). Rates among branches were assumed to follow a lognormal distribution (Drummond et al. 2006). Each analysis used 2 independent Markov chain Monte Carlo (MCMC) runs of 20 000 000 steps to ensure sufficient sampling of parameters. Samples were drawn from the MCMC every 2000 steps after the first 1 000 000 steps (burn-in) were discarded. The BEAST input files are available as Supplementary Material (available from http://www.sysbio.oxfordjournals.org/).

The 3 approaches to calibration yield markedly different date estimates for the most recent common ancestor of neognaths, to the extent that the 95% highest posterior density (HPD) intervals fail to overlap (Fig. 5). When point-like calibrations are used, the 95% HPD interval lies within the Tertiary and excludes the Cretaceous–Tertiary (K-T) boundary at 65.5 Ma before present. This is consistent with a model in which neognaths radiated after the extinction of dinosaurs and pterosaurs (Feduccia, 1995). However, the 95% HPD interval of the estimate also excludes the anseriform *Vegavis* at 66 Ma, which strongly implies that the divergence date inferred using point-like calibrations is too recent. Conversely, using hard bounds or prior age distributions results in the whole of the 95% HPD interval of the age estimate falling in the Cretaceous, which supports the hypothesis that neognaths, dinosaurs, and pterosaurs were contemporary (Penny and Phillips 2004).

This case study provides a concise demonstration of the problems with using point calibrations in divergence dating analyses, whereby precision comes at the cost of accuracy. Not only is there an underestimation of the level of uncertainty, but there is also a clear bias toward obtaining younger age estimates, as would be expected when minimum age information is converted into point calibrations. In contrast, the use of bounds or age priors yields estimates with wider error intervals, but which should provide a more realistic reflection of the uncertainty associated with the calibrations. The 95% HPD interval of the date estimate produced using prior age distributions is smaller than that obtained using hard bounds, which is the consequence of weighting of probability toward the lower end of the lognormal and exponential distributions. In our example, the estimates obtained using hard bounds and prior age distributions are of similar precision, but this is largely on account of the fact that we used relatively informative hard bounds.

CONCLUDING REMARKS

Calibrations clearly have a crucial role in studies of divergence times. Introducing calibrations involves selecting not only the appropriate paleontological or biogeographic data, but also the most effective manner in which to model them in the phylogenetic analysis. With the increasing availability of dating methods that are able to incorporate calibration error, it is evidently no longer justifiable to ignore this uncertainty. Although this introduces a further responsibility and burden for users of methods for estimating divergence times, it is a necessary sacrifice if we wish to move toward building a more realistic framework for estimating evolutionary time scales.

SUPPLEMENTARY MATERIAL

Supplementary material can be found at http://oxfordjournals.org/our_journals/sysbio/.

FUNDING

S.Y.W.H. and M.J.P. were funded by the Australian Research Council.

ACKNOWLEDGMENTS

We thank Philip Donoghue for comments on an early draft of this paper, and Tim Collins, Charles Marshall, Michael Lee, and an anonymous reviewer for constructive comments that led to substantial improvements to the text. Sections of this paper benefited from con-

Downloaded from https://academic.oup.com/sysbio/article/58/3/367/1613552 by guest on 13 March 2024

structive discussions in Working Group 32 at the Royal Botanic Gardens, Melbourne, which was organized by David Cantrill, Maria Gandolfo, and Peter Wilf, and funded by the Australian Research Council-New Zealand Research Network for Vegetation Function.

References

Aris-Brosou S., Yang Z. 2002. Effects of models of rate evolution on estimation of divergence dates with special reference to the metazoan 18S ribosomal RNA phylogeny. Syst. Biol. 51:703–714.

Aris-Brosou S., Yang Z. 2003. Bayesian models of episodic evolution support a Precambrian explosive diversification of the Metazoa. Mol. Biol. Evol. 20:1947–1954.

Barnes I., Matheus P., Shapiro B., Jensen D., Cooper A. 2002. Dynamics of Pleistocene population extinctions in Beringian brown bears. Science. 295:2267–2270.

Barnett R., Barnes I., Phillips M.J., Martin L.D., Harington C.R., Leonard J.A., Cooper A. 2005. Evolution of the extinct Sabretooths and the American cheetah-like cat. Curr. Biol. 15:R589–R590.

Beck R.D.M., Godthelp H., Weisbecker V., Archer M., Hand S.J. 2008. Australia's oldest marsupial fossils and their biogeographical implications. PLoS One. 3:e1858.

Benton M.J., Donoghue P.C. 2007. Paleontological evidence to date the tree of life. Mol. Biol. Evol. 24:26–53.

Boles W.E. 1995. The world's oldest songbird. Nature 374:21–22.

Britton T., Anderson C.L., Jacquet D., Lundqvist S., Bremer K. 2007. Estimating divergence times in large phylogenetic trees. Syst. Biol. 56:741–752.

Brown J.W., Rest J.S., Garcia-Moreno J., Sorenson M.D., Mindell D.P. 2008. Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. BMC Biol. 6:6.

Clarke J.A., Tambussi C.P., Noriega J.I., Erickson G.M., Ketcham R.A. 2005. Definitive fossil evidence for the extant avian radiation in the Cretaceous. Nature. 433:305–308.

Coolen M.J., Overmann J. 2007. 217 000-year-old DNA sequences of green sulfur bacteria in Mediterranean sapropels and their implications for the reconstruction of the paleoenvironment. Environ. Microbiol. 9:238–249.

Cooper A., Penny D. 1997. Mass survival of birds across the Cretaceous-Tertiary boundary: molecular evidence. Science. 275: 1109–1113.

Cutler D.J. 2000. Understanding the overdispersed molecular clock. Genetics. 154:1403–1417.

Donoghue P.C., Benton M.J. 2007. Rocks and clocks: calibrating the Tree of Life using fossils and molecules. Trends Ecol. Evol. 22: 424–431.

Doolittle R.F., Feng D.-F., Tsang S., Cho G., Little E. 1996. Determining divergence times of the major kingdoms of living organisms with a protein clock. Science. 271:470–477.

Drummond A.J., Ho S.Y.W., Phillips M.J., Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4:e88.

Drummond A.J., Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7:214.

Edwards C.J., Bollongino R., Scheu A., Chamberlain A., Tresset A., Vigne J.-D., Baird J.F., Larson G. Ho S.Y.W., Heupink T.H., Shapiro B., Czerwinski P., Arbogast R.-M., Arndt B., Bartosiewicz L., Benecke N., Budja N., Chaix L., Choyke A.M., Coqueugniot E., Döhle H.-J., Göldner H., Hartz S., Helmer D., Herzig B., Hongo H., Mashkou M. Özdogan M., Pucher E., Roth G., Schade-Lindig S., Schmölcke U., Schulting R., Stephan E., Uerpmann H.-P., Vörös I., Bradley D.G., Burger J. 2007. Mitochondrial analysis shows a Neolithic Near Eastern origin for domestic cattle and no evidence of domestication of European aurochs. Proc. R. Soc. Lond. B. 274: 1377–1385.

Emerson B.C. 2007. Alarm bells for the molecular clock? No support for Ho et al.'s model of time-dependent molecular rate estimates. Syst. Biol. 56:337–345.

Endicott P., Ho S.Y.W. 2008. A Bayesian evaluation of human mitochondrial substitution rates. Am. J. Hum. Genet. 82:895–902.

Feduccia A. 1995. Explosive evolution in tertiary birds and mammals. Science. 267:637–638.

Foote M. 1997. Estimating taxonomic durations and preservation probability. Paleobiology. 23:278–300.

Foote M., Hunter J.P., Janis C.M., Sepkoski J.J. Jr. 1999. Evolutionary and preservational constraints on origins of biologic groups: divergence times of eutherian mammals. Science. 283:1310–1314.

Gandolfo M.A., Nixon K.C., Crepet W.L. 2008. Selection of fossils for calibration of moleclar dating models. Ann. Missouri Bot. Gard. 95:34–42.

Gilbert M.T.P., Drautz D.I., Lesk A.M., Ho S.Y.W., Qi J., Ratan A., Hsu C.-H., Sher A., Dalén L., Götherström A., Tomsho L.P., Rendulic S., Packard M., Campos P.F., Kuznetsova T., Shidlovskiy F., Tikhonov A., Willerslev E., Iacumin P., Buigues B., Ericson P.G.P., Germonpré M., Kosintsev P., NikolaevV. , Nowak-Kemp M., Knight J.R., Irzyk G.P., Perbost C.S., Fredrikson K.M., Harkins T.T., Sheridan S., Miller W., Schuster S.C. 2008. Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. Proc. Natl Acad. Sci. USA. 105: 8327–8332.

Godthelp H., Archer M., Cifelli R.L., Hand S.J., Gilkeson C.F. 1992. Earliest known Australian Tertiary mammal fauna. Nature. 356: 514–516.

Graur D., Martin, W. 2004. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. Trends Genet. 20:80–86.

Haag-Liautard C., Dorris M., Maside X., Macaskill S., Halligan D.L., Houle D., Charlesworth B., Keightley P.D. 2007. Direct estimation of per nucleotide and genomic deleterious mutation rates in Drosophila. Nature. 445:82–85.

Haile J., Holdaway R., Oliver K., Bunce M., Gilbert M.T.P., Nielsen R., Munch K., Ho S.Y.W., Shapiro B., Willerslev E. 2007. Ancient DNA chronology within sediment deposits: Are paleobiological reconstructions possible and is DNA leaching a factor? Mol. Biol. Evol. 24:982–989.

Harrison G.L., McLenachan P.A., Phillips M.J., Slack K.E., Cooper A., Penny D. 2004. Four new avian mitochondrial genomes help get to basic evolutionary questions in the late cretaceous. Mol. Biol. Evol. 21:974–983.

Heads M. 2005. Dating nodes on molecular phylogenies: a critique of molecular biogeography. Cladistics. 21:62–78.

Hedges S.B., Kumar S. 2004. Precision of molecular time estimates. Trends Genet. 202:242–247.

Hedges S.B., Parker P.H., Sibley C.G., Kumar S. 1996. Continental breakup and the ordinal diversification of birds and mammals. Nature. 381:226–229.

Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. Nature. 405:907–913.

Ho S.Y.W. 2007. Calibrating molecular estimates of substitution rates and divergence times in birds. J. Avian Biol. 38:409–414.

Ho S.Y.W. 2009. An examination of phylogenetic models of substitution rate variation among lineages. Biol. Lett. 5:421–424.

Ho S.Y.W., Kolokotronis S.-O., Allaby R.G. 2007. Elevated substitution rates estimated from ancient DNA. Biol. Lett. 3: 702–705.

Ho S.Y.W., Phillips M.J., Drummond A.J., Cooper A. 2005. Accuracy of rate estimation using relaxed-clock models with a critical focus on the early metazoan radiation. Mol. Biol. Evol. 22: 1355–1363.

Howell N., Smejkal C.B., Mackey D.A., Chinnery P.F., Turnbull D.M., Herrnstadt C. 2003. The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. Am. J. Hum. Genet. 72:659–670.

Hug L.A., Roger A.J. 2007. The impact of fossils and taxon sampling on ancient molecular dating analyses. Mol. Biol. Evol. 24:1889–1897.

Hugall A.F., Foster R., Lee M.S. 2007. Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. Syst. Biol. 56:543–563.

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

Downloaded from https://academic.oup.com/sysbio/article/58/3/367/1613552 by guest on 13 March 2024

Korsten M., Ho S.Y.W., Davison J., Pähn B., Vulla E., Roht M., Tumanov I.L., Kojola I., Andersone-Lilley Z., Ozolins J., Pilot M., Mertzanis Y., Giannakopoulos A., Vorobiev A.A., Markov N.I., Saveljev A.P., Lyapunova E.A., Abramov A.V., Männil P., Valdmann H., Pazetnov S.V., Pazetnov V.S., Rõkov A.M., Saarma U. 2009. Sudden expansion of a single brown bear maternal lineage across northern continental Eurasia after the last ice age: a general demographic model for mammals? Mol. Ecol. 18: 1963–1979.

Lambert D.M., Ritchie P.A., Millar C.D., Holland B., Drummond A.J., Baroni C. 2002. Rates of evolution in ancient DNA from Adélie penguins. Science. 295:2270–2273.

Lee M.S.Y. 1999. Molecular clock calibrations and metazoan divergence dates. J. Mol. Evol. 49:385–391.

Lee M.S.Y., Oliver M., Hutchinson N. 2009. Phylogenetic uncertainty and molecular clock calibrations: a case study of legless lizards (Pygopodidae, Gekkota). Mol. Phylogenet. Evol. 50:661–666.

Leitner T., Albert J. 1999. The molecular clock of HIV-1 unveiled through analysis of a known transmission history. Proc. Natl. Acad. Sci. USA. 96:10752–10757.

Lepage T., Bryant D., Philippe H., Lartillot N. 2007. A general comparison of relaxed molecular clock models. Mol. Biol. Evol. 24: 2669–2680.

Lepage T., Tupper P., Bryant D., Lawi S. 2006. Continuous and tractable models for the variation of evolutionary rates. Math. Biosci. 199:216–233.

Linder H.P., Hardy C.R., Rutschmann F. 2005. Taxon sampling effects in molecular clock dating: an example from the African Restionaceae. Mol. Phylogenet. Evol. 35:569–82.

Magallón S. 2004. Dating lineages: molecular and paleontological approaches to the temporal framework of clades. Int. J. Plant Sci. 165:S7–S21.

Marshall C.R. 1997. Confidence intervals on stratigraphic ranges with non-random distributions of fossil horizons. Paleobiology. 23: 165–173.

Marshall C.R. 2008. A simple method for bracketing absolute divergence times on molecular phylogenies using multiple fossil calibration points. Am. Nat. 171:726–742.

Morrison D.A. 2008. How to summarize estimates of ancestral divergence times. Evol. Bioinf. 4:75–95.

Muller J., Reisz R.R. 2005. Four well-constrained calibration points from the vertebrate fossil record for molecular clock estimates. Bioessays. 27:1069–1075.

Near T.J., Meylan P.A., Shaffer H.B. 2005. Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. Am. Nat. 165:137–146.

Parham J.F., Irmis R.B. 2008. Caveats on the use of fossil calibrations for molecular dating: a comment on Near et al. Am. Nat. 171: 132–136.

Penny D., Phillips M.J. 2004. The rise of birds and mammals: are microevolutionary processes sufficient for macroevolution? Trends Ecol. Evol. 19:516–522.

Pole M.S. 2001. Can long-distance dispersal be inferred from the New Zealand plant fossil record? Aust. J. Bot. 49:357–366.

Posada D., Crandall K.A. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics. 14:817–818.

Rajabi-Maham H., Orth A., Bonhomme F. 2008. Phylogeography and postglacial expansion of *Mus musculus domesticus* inferred from mitochondrial DNA coalescent, from Iran to Europe. Mol. Ecol. 17:627–641.

Rambaut A. 2000. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. Bioinformatics. 16:395–399.

Rambaut A., Bromham L. 1998. Estimating divergence dates from molecular sequences. Mol. Biol. Evol. 15:442–448.

Rannala B., Yang Z. 2007. Inferring speciation times under an episodic molecular clock. Syst. Biol. 56:453–466.

Reimer P.J., Baillie M.G.L., Bard E., Bayliss A., Beck J.W., Bertrand C.J.H., Blackwell P.G., Buck C.E., Burr G.S., Cutler K.B., Damon P.E., Edwards R.L., Fairbanks R.G., Friedrich M., Guilderson T.P., Hogg A.G., Hughen K.A., Kromer B., McCormac G., Manning S., Bronk Ramsey C., Reimer R.W., Remmele S., Southon J.R., Stuiver M., Talamo S., Taylor F.W., van der Plicht V., Weyhenmeyer C.E. 2004. IntCal04 terrestrial radiocarbon age calibration, 0–26 cal kyr BP. Radiocarbon. 46:1029–1058.

Renner S.S. 2005. Relaxed molecular clocks for dating historical plant dispersal events. Trends Plant Sci. 10:1360–1385.

Rutschmann F. 2006. Molecular dating of phylogenetic trees: a brief review of current methods that estimate divergence times. Diversity Distrib. 12:35–48.

Saarma U., Ho S.Y.W., Pybus O.G., Kaljuste M., Tumanov I.M., Kojola I., Vorobiev A.A., Markov N.I., Saveljev A.P., Valdmann A., Lyapunova E.A., Abramov A.V., Männil P., Korsten M., Vulla E., Pazetnov S.V., Pazetnov V.S., Putchkovskiy S.V., Rõkov A.M. 2007. Mitogenetic structure of brown bears (*Ursus arctos* L.) in northeastern Europe and a new time-frame for the formation of brown bear lineages. Mol. Ecol. 16:401–413.

Sanders K.L., Lee M.S. 2007. Evaluating molecular clock calibrations using Bayesian analyses with soft and hard bounds. Biol. Lett. 3:275–279.

Sanderson M.J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. 14:1218–1231.

Sanderson M.J. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. 19:101–109.

Sanderson M.J. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics. 19:301–302.

Slack K.E., Jones C.M., Ando T., Harrison G.L., Fordyce R.E., Arnason U., Penny D. 2006. Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. Mol. Biol. Evol. 23: 1144–1155.

Strauss D., Sadler P.M. 1989. Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. Math. Geology. 21:411–427.

Svennblad B. 2008. Consistent estimation of divergence times in phylogenetic trees with local molecular clocks. Syst. Biol. 57: 947–954.

Takezaki N., Rzhetsky A., Nei M. 1995. Phylogenetic test of the molecular clock and linearized trees. Mol. Biol. Evol. 12:823–833.

Tavaré S., Marshall C.R., Will O., Soligo C., Martin R.D. 2002. Using the fossil record to estimate the age of the last common ancestor of extant primates. Nature. 416:726–729.

Thorne J.L., Kishino H. 2002. Divergence time and evolutionary rate estimation with multilocus data. Syst. Biol. 51:689–702.

Thorne J.L., Kishino H., Painter I.S. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15: 1647–1657.

Wandeler P., Hoeck P.E.A., Keller L.F. 2007. Back to the future: museum specimens in population genetics. Trends Ecol. Evol. 22: 634–642.

Weinstock J., Willerslev E., Sher A., Tong W., Ho S.Y.W., Rubinstein D., Storer J., Burns J., Martin L., Bravi C., Prieto A., Froese D., Scott E., Xulong L., Cooper A. 2005. Evolution, systematics, and phylogeography of Pleistocene horses in the New World: a molecular perspective. PLoS Biol. 3:e241.

Welch J.J., Bromham L. 2005. Molecular dating when rates vary. Trends Ecol. Evol. 20:320–327.

Willerslev E., Cappellini E., Boomsma W., Nielsen R., Hebsgaard M.B., Brand R.B., Hofreiter M., Bunce M., Poinar H.N., Dahl-Jensen D., Johnsen, S., Steffensen J.P., Bennike O., Schwenninger J.L., Nathan R., Armitage S., de Hoog C.J., Alfimov V., Christl M., Beer M., Muscheler R., Barker J., Sharp M., Penkman K.E., Haile J., Taberlet P., Gilbert M.T., Casoli A., Campani E., Collins M.J. 2007. Ancient biomolecules from deep ice cores reveal a forested southern Greenland. Science. 317:111–114.

Willerslev E., Cooper A. 2005. Ancient DNA. Proc. R. Soc. Lond. B. 272:3–16.

Woodburne M.O., Case J.A. 1996. Dispersal, vicariance, and the Late Cretaceous to early Tertiary land mammal biogeography from South America to Australia. J. Mammal. Evol. 3:121–161.

Downloaded from https://academic.oup.com/sysbio/article/58/3/367/1613552 by guest on 13 March 2024

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.

Yoder A.D., Yang Z.H. 2000. Estimation of primate speciation dates using local molecular clocks. Mol. Biol. Evol. 17:1081–1090.

Zuckerkandl E., Pauling L. 1962. Molecular disease, evolution and genic heterogeneity. In: Kasha M., Pullman B., editors. Horizons in biochemistry. New York: Academic Press. p. 189–225

Zuckerkandl E., Pauling L. 1965. Evolutionary divergence and convergence in proteins. In: Bryson V., Vogel H.J., editors. Evolving genes and proteins. New York: Academic Press. p. 97–166.

Received 15 August 2008; reviews returned 7 November 2008; accepted 21 April 2009

Associate Editor: Tim Collins

Downloaded from https://academic.oup.com/sysbio/article/58/3/367/1613552 by guest on 13 March 2024