Püschel, H., O'Reilly, J., Pisani, D., & Donoghue, P. (2019). The impact of fossil stratigraphic ranges on tip-calibration, and the accuracy and precision of divergence time estimates. *Palaeontology*. https://doi.org/10.1111/pala.12443

The impact of fossil stratigraphic ranges on tip-calibration, and the accuracy and precision of divergence time estimates

Hans P. Püschel[1,2], Joseph E. O'Reilly[1], Davide Pisani[1,3] and Philip C. J. Donoghue[1]

[1]School of Earth Sciences and [3]School of Biological Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK
[2]School of Geosciences, University of Edinburgh, Grant Institute, The King's Buildings, James Hutton Road, Edinburgh EH9 3FE, UK

ABSTRACT: The molecular clock currently provides the only viable means of establishing realistic evolutionary timescales but it remains unclear how best to calibrate divergence time analyses. Calibrations can be applied to the tips and/ or to the nodes of a phylogeny. Tip-calibration is an attractive approach since it allows fossil species to be included alongside extant relatives in molecular clock analyses. However, most fossil species are known from multiple stratigraphic horizons and it remains unclear how such age ranges should be interpreted to codify tip-calibrations. We use simulations and empirical data to explore the impact on precision and accuracy of different approaches to informing tip-calibrations. In particular, we focus on the effect of using tip-calibrations defined using the oldest versus youngest stratigraphic occurrences, the full stratigraphic range, as well as confidence intervals on these data points. The results of our simulations show that using different calibration approaches leads to different divergence-time estimates and demonstrate that concentrating tip-calibrations near the root of the dated phylogeny improves both precision and accuracy of estimated divergence times. Finally, our results indicate that the highest levels of accuracy and precision are achieved when fossil tips are calibrated based on the fossil occurrence from which the morphological data was derived. These trends were corroborated by analysis of an empirical dataset for Ursidae. Overall, we conclude that tip-dating analyses should, in particular, employ tip calibrations close to the root of the tree and they should be calibrated based on the age of the fossil used to inform the morphological data used in Total Evidence Dating.

Key words: molecular clock, simulation, fossil, divergence time, tip-calibration, Ursidae.

CALIBRATING the tree of life to geological time is a core aim of biology, facilitating temporal tests of evolutionary hypotheses, inference of evolutionary rates and patterns, as well as an understanding of the coevolution of life and the environment. The molecular clock (Zuckerkandl and Pauling 1965) affords the only viable means of establishing an evolutionary timescale, a prospect that has become increasingly tangible with the development of Bayesian relaxed clock methods (Donoghue and Yang 2016; Dos Reis *et al.* 2016; Kumar and Hedges 2016; Bromham *et al.* 2017). A key aspect of the

Bayesian divergence time estimation framework is the use of priors on divergence times which include information on the age of calibrations that is usually provided by fossil evidence. Traditionally, molecular clocks have been calibrated by node-calibration, where clade ages are constrained minimally by their oldest fossil member; clade age maxima are based on a diversity of approaches ranging from statistical analysis of occurrence data through heuristic analysis of a variety of evidence, through to arbitrary probabilities that express some visceral perception of the true time of divergence (Donoghue and Benton 2007; Parham *et al.* 2012). Node calibrations are transformed into a joint time prior for the tree which may not always reflect the original node-calibrations (Inoue *et al.* 2010; Warnock *et al.* 2012, 2015). Many view node-calibrations as unsatisfactory because of this transformation of the original fossil evidence which is affected by prior assumptions on the phylogenetic affinity of the calibrating fossils, makes no use of fossil age information in topology estimation, strongly limits the number and nature of fossil species that can be used to calibrate the clock, and because of the diverse and ad hoc approaches that have been proposed to formulating probabilistic calibrations (Gavryushkina *et al.* 2017).

More recently, tip-calibration has been introduced specifically to overcome the limitations of node-calibration (Pyron 2011; Ronquist *et al.* 2012*b*), affording a more direct approach to calibration, where fossils are included analytically *en par* with their living relatives, avoiding the need for the definition of maximum and minimum constraints in the divergence-time analysis (Ronquist *et al.* 2012*b*). Tip-calibration is achieved by supplementing the molecular sequence data from living species with morphological data from both living and fossil species. These molecular-morphological datasets are then analysed using a partitioning scheme allowing the concurrent application of molecular and morphological models to the data. The calibration information is provided by the age information associated with the fossil taxa which disambiguate the morphological distances in terms of their otherwise confounded evolutionary rates and absolute times, that spread for the rest of the internal nodes of the tree characterized by molecular data (Barba-Montoya *et al.* 2017). Tip-calibration requires no prior assumptions of phylogenetic affinity and so it allows fossils to be more readily used in the co-estimation of divergence-times and tree topology and, at least theoretically, there are no restrictions on the number of fossil species that can be included in an analysis. Therefore, this approach allows the inclusion of older and sometimes fragmentary fossils typically excluded in node-calibration analyses because of their uncertain phylogenetic placement (Ronquist *et al.* 2012*b*; O'Reilly *et al.* 2015).

Though tip-calibration overcomes some of the drawbacks of node-calibration, it presents a series of new concerns, including systematic biases in the preservation of morphological characters in the fossil record (Sansom and Wills 2013), the efficacy of the Mk model of morphological character evolution, as well as its extreme sensitivity to the prior on divergence times or the branching model used

(O'Reilly *et al.* 2015; Donoghue and Yang 2016; Dos Reis *et al.* 2016). Regarding this last aspect, it is striking that tip-calibration was initially promoted on the promise that it would deliver more accurate and precise divergence time estimates less sensitive to prior assumptions than node-calibration (Ronquist *et al.* 2012*b*) since the opposite has been shown to be the case in most tip-calibrated divergence time studies (Wood *et al.* 2013; Arcila *et al.* 2015; O'Reilly *et al.* 2015; Matzke and Wright 2016; Puttick *et al.* 2016; Ronquist *et al.* 2016). Moreover, divergence time studies employing tip-calibration have displayed a consistent tendency towards estimating older divergence-times than node-calibrated analyses (Ronquist *et al.* 2012*b*, 2016; Schrago *et al.* 2013; Slater 2013, 2015; Wood *et al.* 2013; Sharma and Giribet 2014; Tseng *et al.* 2014; Beck and Lee 2014; Arcila *et al.* 2015; Winterton and Ware 2015; Dornburg *et al.* 2015; O'Reilly *et al.* 2015; Bapst *et al.* 2016; Matzke and Wright 2016; Puttick *et al.* 2016; Saladin *et al.* 2017). This latter phenomenon has been termed Deep Root Attraction (Ronquist *et al.* 2016).

One possible explanation for the unrealistically ancient estimates from tip-calibrated divergence time analyses lies with the unresolved and little explored manner in which tip-calibrations are formulated (O'Reilly *et al.* 2015; Donoghue and Yang 2016). Most studies that have employed tip-calibration have assumed errorless point ages for the calibrating fossils based on a single age sample from the uncertainty associated with their geological age (e.g. Ronquist *et al.* 2012*b*; Schrago *et al.* 2013; Lee *et al.* 2014; Sharma and Giribet 2014; Arcila *et al.* 2015), if at all. This is surprising since it does not consider the stratigraphic uncertainty of a fossil age which is usually defined in terms of a minimum-maximum age interval (O'Reilly *et al.* 2015). Other studies have modelled tip-calibration age uncertainty as a uniform distribution between the oldest and youngest fossil stratigraphic occurrence (Wood *et al.* 2013; Dornburg *et al.* 2015; Marx and Fordyce 2015; Vea and Grimaldi 2016; Zhang *et al.* 2016; Heritage *et al.* 2016; Lee 2016; Puttick *et al.* 2016; Sallam and Seiffert 2016; Borths and Stevens 2017; Turner *et al.* 2017; Vinther *et al.* 2017; Wood 2017; Wright and Toom 2017; Gavryushkina *et al.* 2017; Harrington and Reeder 2017; Seiffert *et al.* 2017). Nevertheless, fossil species that are distributed across stratigraphic intervals imply morphological stasis which in itself informs the rate of evolution. Marx and Fordyce (2015) assert that in cases of multiple fossil occurrences of markedly different ages, the oldest fossil occurrences should be considered exclusively and all other data should be discarded. Thus, it is not at all clear which of these approaches is best to implement tip-calibrations, let alone whether choice among these models has an impact on divergence time estimation.

In an attempt to explore the impact that different approaches to formulating tip-calibration has on divergence time estimation we took a simulation approach in which the performance of competing approaches can be assessed relative to a reference generating tree. This approach overcomes the impossible challenge to reconciling competing methodological approaches when analysing empirical

data where the true timescale is unknowable (Bromham 2019). Using simulations we were able to test the credibility of the estimated diverge-times (Bromham *et al.* 2017) and to directly evaluate both accuracy and precision against known ages (Warnock *et al.* 2017). This was supplemented by analyses of an empirical dataset of Ursidae. We considered five different approaches to informing tip-calibrations and assessed their performance in terms of the absolute precision and accuracy of divergence-time estimates, as well as their coverage probability. Fossil occurrences were interpreted to inform tip-calibrations based on their (i) oldest or (ii) youngest stratigraphic occurrences, (iii) a uniform probability spanning their full stratigraphic range, or a uniform probability spanning their full stratigraphic range plus a 95% confidence interval added to their (iv) oldest or (v) youngest stratigraphic occurrences, based on their number of intervening number of fossiliferous horizons (Marshall 1990).

MATERIALS AND METHODS

Tree generation

An 18-tip tree with a Colless Index (Ic) of 0.5 was generated using the R packages ape v5.1 (Paradis *et al.* 2004), geiger v2.0.6 (Harmon *et al.* 2008) and apTreeshape v1.5.0 (Bortolussi *et al.* 2006). The Ic measures the asymmetry of the tree topology, ranging from zero to one with a value of zero representing a fully balanced tree and a value of one a fully imbalanced tree (Colless 1982). The branch lengths of this topology were manually modified to generate two different trees: a tree with fossil lineages nested along the stem and a tree with fossil lineages within the crown group. In both cases, six fossil taxa and 12 extant taxa were used (Fig. 1). The branch lengths were scaled so that the height of the tree was 250 Ma, which was selected due the ~250 Myr periodicity of a Phanerozoic Wilson cycle.

Simulation of fossil occurrences and character data generation

The R package phyclust v0.1.22 (Chen 2011) was used to simulate 1000 nucleotide sequences for each tree with the HKY nucleotide substitution model. Site-specific rate heterogeneity was modelled with a gamma distribution with $\alpha = \beta = 0.25$, the transition/transversion ratio κ was set as 0.75 and the relative state frequencies were sampled from a Dirichlet distribution with $\alpha = (1,1,1,1)$. A strict clock with a rate of 0.0025 substitutions per site per million years was applied. For fossil lineages, simulated molecular information was entirely replaced with missing data. Fossil occurrence data was simulated with the R package Fossilsim (Barido-Sottani *et al.* 2018) using a non-uniform model across 125 stratigraphic intervals using PA = 1, PD = 1 and DT = 1.5 as values for the parameters of the model (Holland 1995). Through rejection sampling, each fossil lineage was enforced to possess at least two fossil occurrences and branches were trimmed such that they did not extend beyond the oldest fossil occurrence; since morphospecies exhibit morphological consistency it would have been inappropriate to create a second set of trees in which branches were trimmed to the youngest fossil

occurrence. The trimmed trees were used to simulate morphological data using a method which matched the Mk model of morphological evolution (Lewis 2001). In order to do this, 1000 nucleotide sequences were simulated for each tree with a JC nucleotide substitution model. The site-specific rate heterogeneity and the clock rate kept the same values as in the molecular simulations. The nucleotides then were recoded as purines and pyrimidines, resulting in binary morphology-like character data (O'Reilly et al. 2016). Constant morphological characters were not assessed nor removed.

Five separate tip-calibration strategies were then applied to the simulated data: (i) a point calibration representing the oldest occurrence of a lineage (Fig. 2A); (ii) a point calibration representing the youngest occurrence of a lineage (Fig. 2B); (iii) a uniform distribution spanning the oldest and youngest occurrences (Fig. 2C); (iv) a uniform distribution spanning the oldest and youngest occurrences plus a 95% confidence interval added to the maximum age as described in Marshall (1990); and (v) a uniform distribution spanning the oldest and youngest occurrences plus the same confidence interval added to the minimum age (Fig. 2E). The span of this confidence interval depends on the number of fossil horizons and the width of the stratigraphic interval between the oldest and youngest occurrences with the only assumptions that these horizons are randomly distributed and uniformly collected over the whole stratigraphic range (Marshall 1990). For both trees this methodology was replicated 100 times for each calibration approach (R = 100).

Tip-calibration Analyses of Simulated Data

The simulated datasets were analysed in MrBayes v3.2.2 (Ronquist *et al.* 2012*a*). The generating model (HKY + Γ) was applied to analyse the molecular data, the Mk + Γ model was used to analyse morphological data (Lewis 2001), and ascertainment bias was not accounted for in the model. The topology was fixed to that of the generating tree and a strict uniform clock with an exponential prior with mean 0.025 was applied to model the rate. An exponential prior was applied to the tree height, with an offset 10 Myr less than the minimum age of 250 Ma and a mean of 312.2 Ma. For each analysis two independent runs were performed using four chains and 1000000 MCMC generations. The sampling frequency was every 200 generations, and the initial 25% of samples were discarded as burn-in. The accuracy of divergence-time estimates was determined using both the error of the estimated ages, and coverage measured as the proportion of 95% HPDs that included the true fossil age. In addition, the precision of divergence-time estimates was assessed in terms of accuracy (difference between the mean of the posterior distribution and the true age) and precision (the relative interval width which is the width of the 95% HPD range).

Tip-calibration of Empirical Data

For the empirical tip-calibration a molecular and morphological dataset of living bears (Ursidae) and extant relatives was analysed. This clade was selected because of the availability of molecular,

morphological and fossil occurrence data (Montoya *et al.* 2001; Baryshnikov 2002; Jin *et al.* 2007; Abella *et al.* 2012; Heath *et al.* 2014). The outgroup was the grey wolf (*Canis lupus*), and its molecular information was obtained from GenBank (JX013645.1, AY525044.1). The morphological information from fossil and living bears was obtained from (Heath *et al.* 2014) and the fossil stratigraphic occurrence data were obtained from the Paleobiology Database (PBDB; https://paleobiodb.org; Püschel et al. 2019: Appendix 5). Since singletons are invariant to the calibration strategies we evaluate, only extinct taxa with at least two fossil occurrences were considered. The Mk+ Γ model (Lewis 2001) was used for morphological data, and the GTR + I + Γ model was used for molecular data. Branch rate variation was modelled with the Independent Gamma Rate (IGR) relaxed-clock model (Lepage *et al.* 2007). The clock rate prior was lognormally distributed, with a mean of -4 and standard deviation of 0.1, the prior for the variance was an exponential distribution with rate 126.887 similar to that used for the Canidae phylogeny (Matzke and Wright 2016). The choice of priors was mainly based on those used in Krause *et al.* (2008) and their observation of an approximately clock-like rate. The root of the tree was calibrated using an exponential distribution, offset with the age of *Hesperocyon gregarious* with a minimal value of 37.2 Ma and a mean of 38.6 Ma, which informs the fossil node calibration for crown Carnivora (Benton *et al.* 2015). The topology was fixed to be consistent with the molecular phylogenies of Krause *et al.* (2008) and Heath *et al.* (2014), as well as with the morphological phylogeny of fossil taxa from (Abella *et al.* 2012). Constraining the analyses in this way resulted in the estimation of divergence times alone, allowing quantification of the effect of the different tip calibration approaches on these age estimates. Each analysis consisted of two runs of four chains and 10M MCMC generations with the first 25% of samples discarded. The program Tracer v1.7.1 (Rambaut *et al.* 2018) was used to assess whether the two independent runs achieved convergence and stationarity.

RESULTS

Simulation analyses

Tip calibration recovers divergence ages close to the real ages; tip calibrations clustered near the root recover more accurate and precise estimates. The divergence-time estimates in the simulation analyses tended to recover node ages close to the generating ages (especially for nodes close to the tips) if the posterior age estimates are interpreted as the median or the mean of the posterior age distribution (Fig. 3). However, accuracy and precision showed a tendency to decrease in the trees where the tip-calibrations where shallow (topologically remote from the root), in comparison to the tree in which the tip-calibrations were positioned closer to the root over all the calibration approaches, showing a minor but consistent higher error (Fig. 3; Püschel et al. 2019: Appendices 1-3). However, the percent error showed an opposite trend (Fig. 4; Püschel et al. 2019: Appendices 2, 3) and although the absolute error was lower in the most recent nodes of both trees (Fig 3; Püschel et al. 2019:

Appendices 1-3), the percent error showed the opposite trend with lower error in the deepest nodes (Fig. 4; Püschel et al. 2019: Appendices 2, 3).

Tips calibrated from oldest occurrences recover the most accurate and precise estimates. The most accurate and precise approach used the oldest stratigraphic occurrence of a lineage to inform point age tip-calibrations (Fig. 3; Püschel et al. 2019: Appendices 1-3). This result holds irrespective of whether the tip-calibrations are located topologically close to, or remote from, the root. In contrast, using the youngest occurrence of a lineage recovered the most inaccurate results, at least 6 Myr older than the estimates obtained when using other interpretations of stratigraphic range data to inform the age of tip calibrations (Fig. 3; Püschel et al. 2019: Appendices 1-3). However, considering only the percent error, the most accurate approach is using a uniform distribution spanning the oldest and youngest occurrences (Fig. 4; Püschel et al. 2019: Appendices 2, 3).

Competing approaches to tip calibration differ little in terms of coverage; clearer differences when considering the width of 95% HPD intervals. When accuracy was measured using the 95% HPD coverage (i.e. the generating age is encompassed by the 95% HPD), there are no significant differences between estimates derived from competing approaches to tip-calibration, or the topology of tip-calibrations approaches. Almost all approaches yield estimates that approximate 100% coverage, the exception using the youngest occurrence of a lineage in the tree with calibrations near to the root, which performed worst (Püschel et al. 2019: Appendices 2-4). However, the span of the 95% HPD interval differed between tip-calibration approaches and topologies (Fig. 5; Püschel et al. 2019: Appendices 2, 3). The tree with the most precise timescale (narrowest 95% HPDs on node ages) had the tip-calibrations located topologically close to the root, informed by the oldest stratigraphic occurrences (Fig. 5A). The tree with the least precise divergence times (broadest 95% HPDs on node ages) had tip-calibrations located remote from the root and informed by a uniform distribution spanning the oldest and youngest occurrences plus a 95% confidence interval added to the maximum age (Fig. 5I); this was followed very closely by tip calibrations informed by the youngest stratigraphic occurrences (Fig. 5G).

Empirical analyses: Differences between calibration approaches in the bear phylogeny.
The empirical analyses of the bear (Ursidae) dataset recovered clade age estimates that are generally more consistent than seen in the simulations (Figs. 6, 7). The most precise method used the youngest fossil occurrences as errorless tip calibrations, and tip calibrations based on the oldest occurrences recovered the oldest clade age estimates. However, these differences become more noticeable the closer the position of the node relative to the root. Nodes within crown-Ursidae showed only minor differences, especially the most recent nodes of the Ursinae subfamily (Fig. 7; Table 1).

DISCUSSION

Simulation analyses

Tip calibration does not show a general bias to older divergence-time estimates. The simulation results (Fig. 3; Püschel et al. 2019: Appendices 1-3) do not support previous inferences of a Deep Root Attraction phenomenon and the tendency of tip-calibrated analyses to recover inaccurately old divergence-time estimates (e.g. Ronquist *et al.* 2012*b*, 2016; Schrago *et al.* 2013; Slater 2013, 2015; Wood *et al.* 2013; Sharma and Giribet 2014; Tseng *et al.* 2014; Beck and Lee 2014; Arcila *et al.* 2015; Winterton and Ware 2015; Dornburg *et al.* 2015; O'Reilly *et al.* 2015; Bapst *et al.* 2016; Matzke and Wright 2016; Puttick *et al.* 2016; Saladin *et al.* 2017). Even if all the calibration approaches here tested showed different degrees of error, this error did not exhibit a general trend towards older divergence-time estimates, with the possible exception of using the oldest stratigraphic occurrences for calibrating which exhibited a small but consistent overestimation of the true ages in all nodes regardless the position of the tip-calibrations (Fig. 3A, F; Püschel et al. 2019: Appendices 1-3). In other words, the results of our simulation experiments suggest that the tendency for tip-calibrated analyses to recover erroneously ancient clade age estimates is not an intrinsic property of this approach. The Deep Root Attraction phenomenon observed in previous studies could be due to inappropriate priors in the model and the inability of the only currently available morphological model (the Mk model) to include correlations or dependencies between morphological traits (Ronquist *et al.* 2016). As our morphological data was simulated using the Mk model, there is no mismatch between the data and the model used for analysing the data, so logically it should not present the problems associated with analysing empirical data. In that sense, our results are compatible with the notion that Deep Root Attraction is a consequence of an inadequate model of morphological evolution.

Best results achieved when mismatch of the simulation model and the specified model is minimised. Considering different measurements of accuracy and precision, tips calibrated from oldest occurrences recover the most accurate and precise estimates of clade age (Figs. 3, 5; Püschel et al. 2019: Appendices 1-3). Although the percent error does not show the same (Fig. 4; Püschel et al. 2019: Appendices 2, 3), this is probably related to the fact that in very shallow nodes small differences (i.e. 1-2 Myr) produce a considerable percent error. As we are interested in the actual divergence ages of different lineages, it can be argued that the absolute error is more relevant. Nevertheless, all results agree that the worst performing approach was using the youngest stratigraphic occurrence to inform tip-calibration. Evidently, the best results are obtained when the stratigraphic fossil occurrences used to simulate the morphological data are used to analyse the data. This is because the morphological data were simulated according to the branch lengths of the input tree. In consequence, as the data were simulated using trees in which branch lengths were trimmed to the oldest fossil occurrence, the phylogenetic signal in the morphological data will tend to generate

best results when using the oldest fossil occurrence to calibrate the tree. In some sense, these results are linked to the simplicity of our simulation framework, however, they are designed to reflect stasis - the expectation that morphospecies are morphologically consistent. Regardless, these results highlight the importance of calibrating fossil tips based on the age of the fossil specimen from which the cladistic morphological data were obtained, avoiding using older or younger specimens referred to that species that could cause a mismatch in the model. This is important because often the taxonomic rank of species does not encompass the full intraspecific variation within in it, and the distinction between morphospecies is often arbitrary (Simpson 1951; Mallet 1995; Manceau and Lambert 2018).

It is important to emphasise that in terms of coverage (the 95% HPD encompassing the true node age), tip-calibration proved accurate for almost all of the calibration approaches that we explored, with the exception of using the youngest occurrence of a lineage in the tree with calibrations near to the root (Püschel et al. 2019: Appendix 4). However, this high level of accuracy is achieved because the imprecision of the most inaccurate calibration strategies is larger.

Tip-calibrations closer to the root increase the accuracy and precision of divergence-time estimates. Tip-calibration better integrates the uncertainty of fossil calibrations leading to more accurate and precise results when the calibrating fossils are located topologically close to, rather than remote from, the root (Figs. 3, 5; Püschel et al. 2019: Appendices 1-3). These results are probably related to the effect of the position of the calibrating fossils in relation to the internal nodes of tree. It has been observed that the larger the distance in time between the nodes and the calibrating fossils, the greater the error of the divergence-time estimates (Conroy and van Tuinen 2003). There are examples of this pattern in node calibration (Conroy and van Tuinen 2003; Linder *et al.* 2005; Rutschmann *et al.* 2007) and more recently in tip-calibration (Arcila *et al.* 2015). Accordingly, as the tree with the fossil tip-calibrations remote from the root does not establish any constraint on the deepest nodes of the tree, estimates of the age of these nodes tend to be older than the true nodes ages, resulting in a magnification of the error approaching the root. The reverse does not obtain because nodes ages are strongly constrained by the extant tip ages, whereas there are no such constraints towards the root. As node ages increase with proximity to the root, the uncertainty associated with their age estimates increases concomitantly (Yang and Rannala 2006; Rannala and Yang 2007). Therefore, in order to increase precision and accuracy of divergence-time estimates in the construction of deep-time phylogenies, it is important to obtain calibrations closer to the root than to the tips. This pattern has also been observed in node-calibration studies (Duchêne *et al.* 2014; Mello and Schrago 2014). The percent error (Fig. 4; Püschel et al. 2019: Appendices 2, 3) showed a different pattern probably because this error is strongly sensitive to small differences in shallow nodes.

Empirical analyses: bear divergence time estimates

Empirical and simulation results are consistent. Considering the phylogeny of bears (Ursidae) under the different calibration approaches, it is clear that the results (Fig. 6A-E) are consistent with the patterns exhibited in the simulation analyses (Fig. 3), *viz*. differing formulations of tip-calibrations based on the same stratigraphic data result in different estimates of clade age. A similar observation has been made in an attempt to estimate the evolutionary timescale of *Pinus* (Saladin *et al.* 2017). This could explain why the most precise approach in the simulations (Fig. 5A) showed to be the least precise in the empirical analysis (Fig. 6A). It is very likely that similarity in divergence-time estimates across calibration approaches within crown-Ursidae (and specifically within the subfamily Ursinae; Fig. 7) is a consequence of their very recent cladogenesis. Shallow node age estimates are heavily constrained by the recent age of living lineages and, hence, they converge readily on similar ages, in contrast to deep nodes which are much more unconstrained. The same pattern was noted above for the most recent nodes in the simulations (Fig 3).

Ursidae divergence-times show no correspondence with previous studies. Previous studies placed the radiation within the subfamily Ursinae close to the Miocene-Pliocene boundary at 5.33 Ma (Krause *et al.* 2008; Heath *et al.* 2014), in a period characterized by the expansion of C4 grass biomes (Cerling *et al.* 1998), faunal turnover and a generalized temperature drop (Van Valkenburgh 1999). In contrast, our analyses suggested this radiation to have occurred in the Pleistocene, around 2 Ma (Fig.7). These results probably differ because of the position of the fossil calibrations. It has been established that even with multiple calibrations, the choice of a particular calibration arrangement can have significant effects on node ages estimation (Sauquet *et al.* 2012), and that precision tends to increase with the proximity of the calibration to the node estimated (Conroy and van Tuinen 2003; Linder *et al.* 2005; Rutschmann *et al.* 2007; Arcila *et al.* 2015). Previous studies used fossils within the subfamily Ursinae, such as *Ursus minimus*, to calibrate this part of the tree and this had as impact on the estimation of evolutionary rate (Krause *et al.* 2008; Heath *et al.* 2014). In our analyses, there were no fossil calibrations within this subfamily (because most species are known from a single stratigraphic occurrence) and so evolutionary rate change was not as well informed for that branch of the crown group. Indeed, just one fossil taxon (*Indarctus arctoides*) phylogenetically close to the panda (*Ailuropoda melanoleuca*) was used to calibrate crown-Ursidae.

Effective morphological stasis and tip-calibration

One of the prime motivations for our study was to determine how to calibrate fossil species as tip calibrations when they are known from multiple stratigraphic horizons. At least in terms of categorical characters, such taxa exhibit morphological stasis, sometimes over protracted episodes of geological time, for which there are examples from both extinct (Eldredge *et al.* 2005) and extant (e.g. Lavoue *et al.* 2011) lineages. In such cases, molecular and morphological rates must be decoupled since, doubtless, molecular evolution will continue. Our results demonstrate that the choice of the

calibration bounds will have a considerable effect on the inferred rate of evolution on the adjacent branches, leading to over- or under-estimations of divergence ages (Ho 2009; O'Reilly *et al.* 2015). Morphological stasis has traditionally be attributed to either developmental constraints, an intrinsic factor, or to stabilizing selection, an extrinsic factor (Maynard Smith *et al.* 1985), while recent theoretical and empirical studies suggest that morphological stasis is a product of stabilizing selection from ecological interactions (Beldade *et al.* 2002; Eldredge *et al.* 2005; Estes and Arnold 2007; Davis *et al.* 2014; Haller and Hendry 2014). Thus, from a theoretical perspective, it can be argued that stasis should not be considered in inference of evolutionary rates, and such lineages should be calibrated based on the oldest stratigraphic occurrence. The results of our analyses corroborate this view.

Tip-calibrations and the Fossilised Birth-Death Process.

Our study was focussed on comparing tip-calibration approaches within a Total Evidence Dating framework, using a realistic simulation of fossil occurrence data; our results and their interpretations remain equally valid for morphological clock analyses. We anticipate that our results are also relevant to mechanistic models of cladogenesis like that applied to the prior on ages in the Fossilized Birth-Death process. Certainly, comparison between mechanistic models and non-mechanistic uniform tree priors in the same dataset have revealed significant differences in divergence-time estimates, with mechanistic models giving more realistic results (Matzke and Wright 2016). Recently, the FBD model has shown attractive improvements in a new framework in which fossil occurrence data are considered explicitly in terms of their stratigraphic range (Stadler *et al.* 2018). Whether these improvements could lead to most precise and accurate results in tip-calibration remains to be assessed in future studies.

CONCLUSIONS

Overall, our study has shown for the first time the varying effects that different tip-calibration approaches have in divergence-time estimates. By performing simulation analyses, we have demonstrated that (i) tip-calibration returns different divergence-time estimates depending on the calibration approach employed, (ii) tip-calibration using the age of the fossil occurrence used to code the cladistic morphological data, recovers the most accurate and precise divergence time estimates, and (iii) tip-calibrations located deep within phylogenetic trees tend to recover more accurate and precise results. Consequently, the best approach to calibrating the fossil tips is using fossils closer to the root and using the age of the specimen used to code the morphology to calibrate its tip. Empirical analyses based in bear phylogeny support these general conclusions, showing that the interpretation of fossil ages really matters in tip-calibrated molecular clock analyses.

Acknowledgements. This study was conducted by H.P.P under the guidance of J.O'R., D.P. and P.C.J.D, in partial fulfilment of the requirements of the MSc Palaeobiology at the School of Earth Sciences, University of Bristol. We thank Mario dos Reis and an anonymous reviewer for helpful comments on an earlier version of the manuscript. H.P.P.R. was funded by a Becas Chile scholarship (73180060, CONICYT-Chile. J.O'R., D.P. and P.C.J.D are funded by NERC (NE/P013678/1; NE/N002067/1) and BBSRC (BB/N000919/1).

DATA ARCHIVING STATEMENT

Data for this study are available in the Dryad Digital Repository: https://datadryad.org/review?doi=doi:10.5061/dryad.nk5fv1s [please note that the data for this paper are not yet published and this temporary link should not be shared without the express permission of the author].

REFERENCES

ABELLA, J., ALBA, D. M., ROBLES, J. M., VALENCIANO, A., ROTGERS, C., CARMONA, R., MONTOYA, P. and MORALES, J. 2012. *Kretzoiarctos* gen. nov., the oldest member of the Giant Panda clade. *PLoS ONE*, **7**, 48985.

ARCILA, D., ALEXANDER PYRON, R., TYLER, J. C., ORTÍ, G. and BETANCUR, R. 2015. An evaluation of fossil tip-dating versus node-age calibrations in tetraodontiform fishes (Teleostei: Percomorphaceae). *Molecular Phylogenetics and Evolution*, **82**, 131–145.

BAPST, D. W., WRIGHT, A. M., MATZKE, N. J. and LLOYD, G. T. 2016. Topology, divergence dates, and macroevolutionary inferences vary between different tip-dating approaches applied to fossil theropods (Dinosauria). *Biology Letters*, **12**, 20160237.

BARBA-MONTOYA, J., DOS REIS, M. and YANG, Z. 2017. Comparison of different strategies for using fossil calibrations to generate the time prior in Bayesian molecular clock dating. *Molecular Phylogenetics and Evolution*, **114**, 386–400.

BARIDO-SOTTANI, J., O'REILLY, J. E., PETT, W. and WARNOCK, R. C. M. 2018. fossilsim: an R package for simulating fossil occurrence data under mechanistic models of preservation and recovery in a phylogenetic framework. *Methods in Ecology and Evolution.*

BARYSHNIKOV, G. F. 2002. Late Miocene *Indarctos punjabiensis atticus* (Carnivora, Ursidae) in Ukraine with survey of Indarctos records from the former USSR. *Russian Journal of Theriology*, **1**, 83–89.

BECK, R. M. D. and LEE, M. S. Y. 2014. Ancient dates or accelerated rates? Morphological clocks and the antiquity of placental mammals. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20141278.

BELDADE, P., KOOPS, K. and BRAKEFIELD, P. M. 2002. Developmental constraints versus

exibility in morphological evolution. *Nature*, **416**, 844–847.

BENTON, M. J., DONOGHUE, P. C. J., ASHER, R. J., FRIEDMAN, M., NEAR, T. J. and VINTHER, J. 2015. Constraints on the timescale of animal evolutionary history. *Palaeontologia Electronica*, article 18.1.1FC.

BORTHS, M. R. and STEVENS, N. J. 2017. The first hyaenodont from the late Oligocene Nsungwe Formation of Tanzania: Paleoecological insights into the Paleogene-Neogene carnivore transition. *PLoS ONE*, **12**, 1–30.

BORTOLUSSI, N., DURAND, E., BLUM, M. and FRANÇOIS, O. 2006. apTreeshape: Statistical analysis of phylogenetic tree shape. *Bioinformatics*, **22**, 363–364.

BROMHAM, L., DUCHÊNE, S., HUA, X., RITCHIE, A. M., DUCHÊNE, D. A. and HO, S. Y. W. 2017. Bayesian molecular dating: opening up the black box. *Biological Reviews*, **93**, 1165–1191.

——— 2019. Six impossible things before breakfast: assumptions, models, and belief in molecular dating. *Trends in Ecology and Evolution*.

CERLING, T. E., EHLERINGER, J. R. and HARRIS, J. M. 1998. Carbon dioxide starvation, the development of the C4 ecosystem, and mammalian evolution. *Philosophical Transactions of the Royal Society, London, Series B*, **353**, 159.

CHEN, W. C. 2011. Overlapping codon model, phylogenetic clustering, and alternative partial expectation conditional maximization algorithm. PhD thesis, Iowa State University, 89pp.

COLLESS, D. H. 1982. Reviewed Work(s): Phylogenetics: The theory and practice of phylogenetic systematics. by E. O. Wiley. *Systematic Zoology*, **31**, 100–104.

CONROY, C. J. and VAN TUINEN, M. 2003. Extracting time from phylogenies: positive interplay between fossil and genetic data. *Journal of Mammalogy*, **84**, 444–455.

DAVIS, C. C., SCHAEFER, H., XI, Z., BAUM, D. A., DONOGHUE, M. J. and HARMON, L. J. 2014. Long-term morphological stasis maintained by a plant-pollinator mutualism. *Proceedings of the National Academy of Sciences*, **111**, 5914–5919.

DONOGHUE, P. C. J. and BENTON, M. J. 2007. Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends in Ecology and Evolution*, **22**, 424–431.

——— and YANG, Z. 2016. The evolution of methods for establishing evolutionary timescales. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **371**, 20160020.

DORNBURG, A., FRIEDMAN, M. and NEAR, T. J. 2015. Phylogenetic analysis of molecular and morphological data highlights uncertainty in the relationships of fossil and living species of Elopomorpha (Actinopterygii: Teleostei). *Molecular Phylogenetics and Evolution*, **89**, 205–218.

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

ELDREDGE, N., THOMPSON, J. N., BRAKEFIELD, P. M., GAVRILETS, S., JABLONSKI, D., JACKSON, J. B. C., LENSKI, R. E., LIEBERMAN, B. S., MCPEEK, M. A. and MILLER, W.

2005. The dynamics of evolutionary stasis. *Paleobiology*, **31**, 133–145.

ESTES, S. and ARNOLD, S. J. 2007. Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. *The American Naturalist*, **169**, 227–244.

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

HALLER, B. C. and HENDRY, A. P. 2014. Solving the paradox of stasis: Squashed stabilizing selection and the limits of detection. *Evolution*, **68**, 483–500.

HARMON, L. J., WEIR, J. T., BROCK, C. D., GLOR, R. E. and CHALLENGER, W. 2008. GEIGER: Investigating evolutionary radiations. *Bioinformatics*, **24**, 129–131.

HARRINGTON, S. M. and REEDER, T. W. 2017. Phylogenetic inference and divergence dating of snakes using molecules, morphology and fossils: new insights into convergent evolution of feeding morphology and limb reduction. *Biological Journal of the Linnean Society*, **121**, 379–394.

HEATH, T. A., HUELSENBECK, J. P. and STADLER, T. 2014. The fossilized birth-death process: a coherent model of fossil calibration for divergence time estimation. *Proceedings of the National Academy of Sciences*, **111**, E2957–E2966.

HERITAGE, S., FERNÁNDEZ, D., SALLAM, H. M., CRONIN, D. T., ESARA ECHUBE, J. M. and SEIFFERT, E. R. 2016. Ancient phylogenetic divergence of the enigmatic African rodent *Zenkerella* and the origin of anomalurid gliding. *PeerJ*, **4**, e2320.

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

HOLLAND, S. M. 1995. The stratigraphic distribution of fossils. *Paleobiology*, **21**, 92-109.

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

JIN, C., CIOCHON, R. L., DONG, W., HUNT, R. M., LIU, J., JAEGER, M. and ZHU, Q. 2007. The first skull of the earliest giant panda. *Proceedings of the National Academy of Sciences*, **104**, 10932–10937.

KRAUSE, J., UNGER, T., NOÇON, A., MALASPINAS, A. S., KOLOKOTRONIS, S. O., STILLER, M., SOIBELZON, L., SPRIGGS, H., DEAR, P. H., BRIGGS, A. W., BRAY, S. C. E., O'BRIEN, S. J., RABEDER, G., MATHEUS, P., COOPER, A., SLATKIN, M., PÄÄBO, S. and HOFREITER, M. 2008. Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC Evolutionary Biology*, **8**, 220.

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

LAVOUE, S., MIYA, M., ARNEGARD, M. E., MCINTYRE, P. B., MAMONEKENE, V. and

NISHIDA, M. 2011. Remarkable morphological stasis in an extant vertebrate despite tens of millions of years of divergence. *Proceedings of the Royal Society B: Biological Sciences*, **278**, 1003–1008.

LEE, M. S. Y. 2016. Multiple morphological clocks and total-evidence tip-dating in mammals. *Biology Letters*, **12**.

———, CAU, A., NAISH, D. and DYKE, G. J. 2014. Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds. *Science*, **345**, 562–566.

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

LEWIS, P. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Society of Systematic Biologists*, **50**, 913–925.

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

MALLET, J. 1995. *A species definition for the Modern Synthesis*. Vol. 10. Clarendon Press 54.

MANCEAU, M. and LAMBERT, A. 2018. The species problem from the modeler's point of view. *Bulletin of Mathematical Biology*, **81**, 878–898.

MARSHALL, C. R. 1990. Confidence intervals on stratigraphic ranges. *Paleobiology*, **16**, 1–10.

MARX, F. G. and FORDYCE, R. E. 2015. Baleen boom and bust: a synthesis of mysticete phylogeny, diversity and disparity. *Royal Society Open Science*, **2**, 140434.

MATZKE, N. J. and WRIGHT, A. 2016. Inferring node dates from tip dates in fossil Canidae: the importance of tree priors. *Biology Letters*, **12**, 1–4.

MAYNARD SMITH, J., BURIAN, R., KAUFFMAN, S., ALBERCH, P., CAMBELL, J., GOODWIN, B., LANDE, R., RAUP, D. and WOLPERT, L. 1985. Developmental constraints and evolution. *The Quarterly Review of Biology*, **60**, 265–287.

MELLO, B. and SCHRAGO, C. G. 2014. Assignment of calibration information to deeper phylogenetic nodes is more effective in obtaining precise and accurate divergence time estimates. *Evolutionary Bioinformatics*, **10**, 79–85.

MONTOYA, P., ALCALÁ, L. and MORALES, J. 2001. *Indarctos* (Ursidae, Mammalia) from the Spanish Turolian (Upper Miocene). *Scripta Geologica*, **122**, 123–151.

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

———, PUTTICK, M. N., PARRY, L. A., TANNER, A. R., TARVER, J. E., FLEMING, J., PISANI, D. and DONOGHUE, P. C. J. 2016. Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Biology Letters*, **12**, 20160081.

PARADIS, E., CLAUDE, J. and STRIMMER, K. 2004. APE: Analyses of phylogenetics and

evolution in R language. *Bioinformatics*, **20**, 289–290.

PARHAM, J. F., DONOGHUE, P. C. J., BELL, C. J., CALWAY, T. D., HEAD, J. J., HOLROYD, P. A., INOUE, J. G., IRMIS, R. B., JOYCE, W. G., KSEPKA, D. T., PATANÉ, J. S. L., SMITH, N. D., TARVER, J. E., VAN TUINEN, M., YANG, Z., ANGIELCZYK, K. D., GREENWOOD, J. M., HIPSLEY, C. A., JACOBS, L., MAKOVICKY, P. J., MÜLLER, J., SMITH, K. T., THEODOR, J. M., WARNOCK, R. C. M. and BENTON, M. J. 2012. Best practices for justifying fossil calibrations. *Systematic Biology*, **61**, 346–359.

PÜSCHEL, H. P., O'REILLY, J. E., PISANI, D. and DONOGHUE, P. C. J. 2019. Data from: The impact of fossil stratigraphic ranges on tip-calibration, and the accuracy and precision of divergence time estimates. Dryad Digital Repository. https://datadryad.org/review?doi=doi:10.5061/dryad.nk5fv1s

PUTTICK, M. N., THOMAS, G. H. and BENTON, M. J. 2016. Dating placentalia: Morphological clocks fail to close the molecular fossil gap. *Evolution*, **70**, 873–886.

PYRON, R. A. 2011. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. *Systematic Biology*, **60**, 466–481.

RAMBAUT, A., DRUMMOND, A. J., XIE, D., BAELE, G. and SUCHARD, M. A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, **67**, 901–904.

RANNALA, B. and YANG, Z. 2007. Inferring speciation tunes under an episodic molecular clock. *Systematic Biology*, **56**, 453–466.

DOS REIS, M., DONOGHUE, P. C. J. and YANG, Z. 2016. Bayesian molecular clock dating of species divergences in the genomics era. *Nature Reviews Genetics*, **17**, 71–80.

RONQUIST, F., TESLENKO, M., VAN DER MARK, P., AYRES, D. L., DARLING, A., HÖHNA, S., LARGET, B., LIU, L., SUCHARD, M. A. and HUELSENBECK, J. P. 2012*a*. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.

———, LARTILLOT, N. and PHILLIPS, M. J. 2016. Closing the gap between rocks and clocks using total-evidence dating. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **371**, 20150136.

———, KLOPFSTEIN, S., VILHELMSEN, L., SCHULMEISTER, S., MURRAY, D. L. and RASNITSYN, A. P. 2012*b*. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Systematic Biology*, **61**, 973–999.

RUTSCHMANN, F., ERIKSSON, T., SALIM, K. A. and CONTI, E. 2007. Assessing calibration uncertainty in molecular dating: the assignment of fossils to alternative calibration points. *Systematic Biology*, **56**, 591–608.

SALADIN, B., LESLIE, A. B., WÜEST, R. O., LITSIOS, G., CONTI, E., SALAMIN, N. and ZIMMERMANN, N. E. 2017. Fossils matter: improved estimates of divergence times in *Pinus* reveal older diversification. *BMC Evolutionary Biology*, **17**, 1–15.

SALLAM, H. M. and SEIFFERT, E. R. 2016. New phiomorph rodents from the latest Eocene of Egypt, and the impact of Bayesian "clock"-based phylogenetic methods on estimates of basal hystricognath relationships and biochronology. *PeerJ*, **4**, e1717.

SANSOM, R. S. and WILLS, M. A. 2013. Fossilization causes organisms to appear erroneously primitive by distorting evolutionary trees. *Scientific Reports*, **3**, 2545.

SAUQUET, H., HO, S. Y. W., GANDOLFO, M. A., JORDAN, G. J., WILF, P., CANTRILL, D. J., BAYLY, M. J., BROMHAM, L., BROWN, G. K., CARPENTER, R. J., LEE, D. M., MURPHY, D. J., SNIDERMAN, J. M. K. and UDOVICIC, F. 2012. Testing the impact of calibration on molecular divergence times using a fossil-rich group: The case of *Nothofagus* (Fagales). *Systematic Biology*, **61**, 289–313.

SCHRAGO, C. G., MELLO, B. and SOARES, A. E. R. 2013. Combining fossil and molecular data to date the diversification of New World Primates. *Journal of Evolutionary Biology*, **26**, 2438–2446.

SEIFFERT, E. R., BOYER, D. M., FLEAGLE, J. G., GUNNELL, G. F., HEESY, C. P., PERRY, J. M. G. and SALLAM, H. M. 2017. New adapiform primate fossils from the late Eocene of Egypt. *Historical Biology*, **30**, 204–226.

SHARMA, P. P. and GIRIBET, G. 2014. A revised dated phylogeny of the arachnid order Opiliones. *Frontiers in Genetics*, **5**, 255.

SIMPSON, G. G. 1951. The species concept. *Evolution. International Journal of Organic Evolution*, **5**, 285–298.

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

———. 2015. Iterative adaptive radiations of fossil canids show no evidence for diversity-dependent trait evolution. *Proceedings of the National Academy of Sciences*, **112**, 4897–4902.

STADLER, T., GAVRYUSHKINA, A., WARNOCK, R. C. M., DRUMMOND, A. J. and HEATH, T. A. 2018. The fossilized birth-death model for the analysis of stratigraphic range data under different speciation modes. *Journal of Theoretical Biology*, **447**, 41–55.

TSENG, Z. J., WANG, X., SLATER, G. J., TAKEUCHI, G. T., LI, Q., LIU, J. and XIE, G. 2014. Himalayan fossils of the oldest known pantherine establish ancient origin of big cats. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20132686–20132686.

TURNER, A. H., PRITCHARD, A. C. and MATZKE, N. J. 2017. Empirical and Bayesian approaches to fossilonly divergence times: A study across three reptile clades. *PLoS ONE*, **12**, e0169885.

VAN VALKENBURGH, B. 1999. Major patterns in the history of carnivorous mammals. *Annual Review of Earth and Planetary Sciences*, **27**, 463–493.

VEA, I. M. and GRIMALDI, D. A. 2016. Putting scales into evolutionary time: The divergence of

major scale insect lineages (Hemiptera) predates the radiation of modern angiosperm hosts. *Scientific Reports*, **6**, 23487.

VINTHER, J., PARRY, L., BRIGGS, D. E. G. and VAN ROY, P. 2017. Ancestral morphology of crown-group molluscs revealed by a new Ordovician stem aculiferan. *Nature*, **542**, 471–474.

WARNOCK, R. C. M., YANG, Z. and DONOGHUE, P. C. J. 2012. Exploring uncertainty in the calibration of the molecular clock. *Biology Letters*, **8**, 156–159.

———, YANG, Z. and DONOGHUE, P. C. J. 2017. Testing the molecular clock using mechanistic models of fossil preservation and molecular evolution. *Proceedings of the Royal Society B: Biological Sciences*, **284**, 20170227.

———, PARHAM, J. F., JOYCE, W. G., LYSON, T. R. and DONOGHUE, P. C. J. 2015. Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. *Proceedings of the Royal Society B: Biological Sciences*, **282**, 20141013.

WINTERTON, S. L. and WARE, J. L. 2015. Phylogeny, divergence times and biogeography of window flies (Scenopinidae) and the therevoid clade (Diptera: Asiloidea). *Systematic Entomology*, **40**, 491–519.

WOOD, H. M. 2017. Integrating fossil and extatn lineages. anexamination of morphological space through time (Areanea: Archaeidae). *Journal of Arachnology*, **45**, 20–29.

———, MATZKE, N. J., GILLESPIE, R. G. and GRISWOLD, C. E. 2013. Treating fossils as terminal taxa in divergence time estimation reveals ancient vicariance patterns in the palpimanoid spiders. *Systematic Biology*, **62**, 264–284.

WRIGHT, D. F. and TOOM, U. 2017. New crinoids from the Baltic region (Estonia): fossil tip-dating phylogenetics constrains the origin and Ordovician–Silurian diversification of the Flexibilia (Echinodermata). *Palaeontology*, **60**, 893–910.

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

ZHANG, C., STADLER, T., KLOPFSTEIN, S., HEATH, T. A. and RONQUIST, F. 2016. Total-Evidence Dating under the Fossilized Birth-Death Process. *Systematic Biology*, **65**, 228–249.

ZUCKERKANDL, E. and PAULING, L. 1965. *In Evolving Genes and Proteins, (ed. by V. Bryson & HJ Vogel)*. Academic Press, New York, 97–166.

FIGURE CAPTIONS

FIG. 1. True dated trees used for simulating the molecular, morphological and fossil occurrences data. A, tree with fossils in the stem group; B, tree with fossils in the crown group. Each tree has 12 extant taxa and six extinct species marked with grey lines. The grey lines also show the places in

which fossil occurrences could be simulated. The divergence times are presented over the nodes in Ma.

FIG. 2. Examples of the different approaches to establishing tip-calibrations explored in our analyses based on fossil stratigraphic range data. These examples are based on a hypothetical extinct taxon with multiple fossil occurrences (1-5) in an interval of time between 90-150 Ma. A, single point calibration in the oldest occurrence of a lineage (maximum) fixed in 150 Ma. B, single point calibration in the youngest occurrence of a lineage (minimum) fixed in 90 Ma. C, uniform distribution between the youngest (minimum) and oldest (maximum) occurrences of a lineage (90-150 Ma). D, uniform distribution between the youngest (minimum) and the oldest (maximum) occurrences of a lineage plus a 95% confidence interval described in Marshall (1990) of 66.8 million years in the maximum reaching 216 Ma. E, uniform distribution between the youngest (minimum) and the oldest (maximum) occurrences of a lineage plus a 95% confidence interval described in Marshall (1990) of 66.8 million years in the minimum reaching 23 Ma.

FIG. 3. Boxplots of divergence-time estimates of the nodes of the tree under different calibration approaches over 100 replicates. The columns represent whether fossils used in the tip-calibrations were in the stem group or in the crown group and the rows show the five tip-calibration approaches employed and described in Fig. 2 from A to E applying the same order for both columns (A-E; F-J). The triangles represent the true ages for each node and the dots the mean of the divergence-time estimates. The outliers were removed for clarity. *Abbreviation*: CI, confidence interval.

FIG. 4. Boxplots of the percent error in divergence-time estimates of the nodes of the tree under different calibration approaches over 100 replicates. The columns represent whether fossils used in the tip-calibrations were in the stem group or in the crown group and the rows show the five tip-calibration approaches employed and described in Fig. 2 from A to E applying the same order for both columns (A-E; F-J). The outliers were removed for clarity. *Abbreviation*: CI, confidence interval.

FIG. 5. Boxplots of the 95% HPD widths of each node of the tree under different calibration approaches over 100 replicates. The columns represent whether fossils used in the tip-calibrations were in the stem group or in the crown group and the rows show the five tip-calibration approaches employed and described in Fig. 2 from A to E applying the same order for both columns (A-E; F-J). The outliers were removed for clarity. *Abbreviation*: CI, confidence interval; HPD, highest posterior density.

FIG. 6. Dated phylogenies of bears (Ursidae) under five different calibration approaches. A, maximum; B, minimum; C, minimum-maximum; D, minimum-maximum 95% confidence interval in

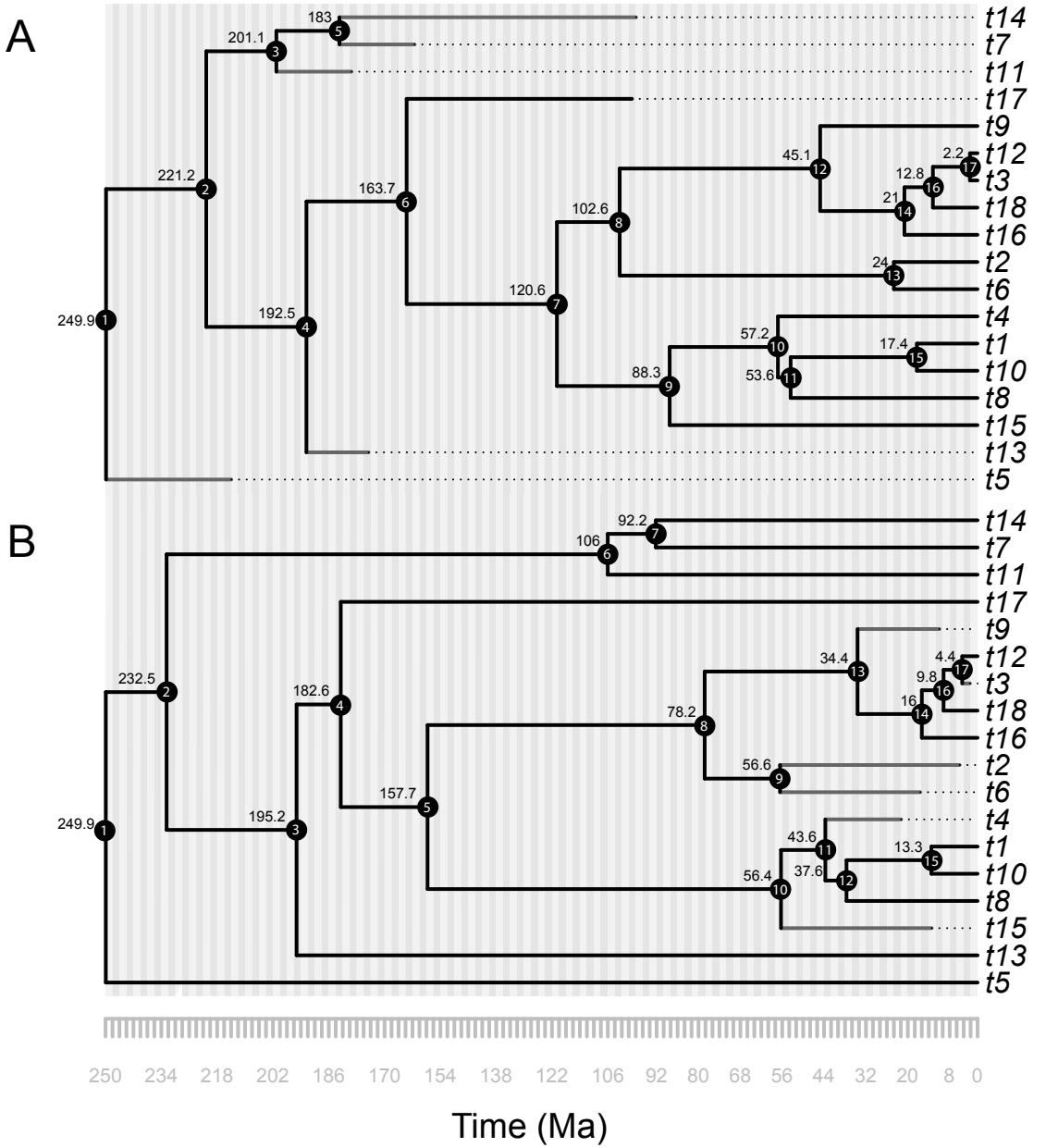
maximum; E, minimum-maximum 95% confidence interval in minimum. More details of the calibration approaches are described in Fig. 2 in the same order. Node bars represent the 95% highest posterior density (HPD) for the estimated node ages.

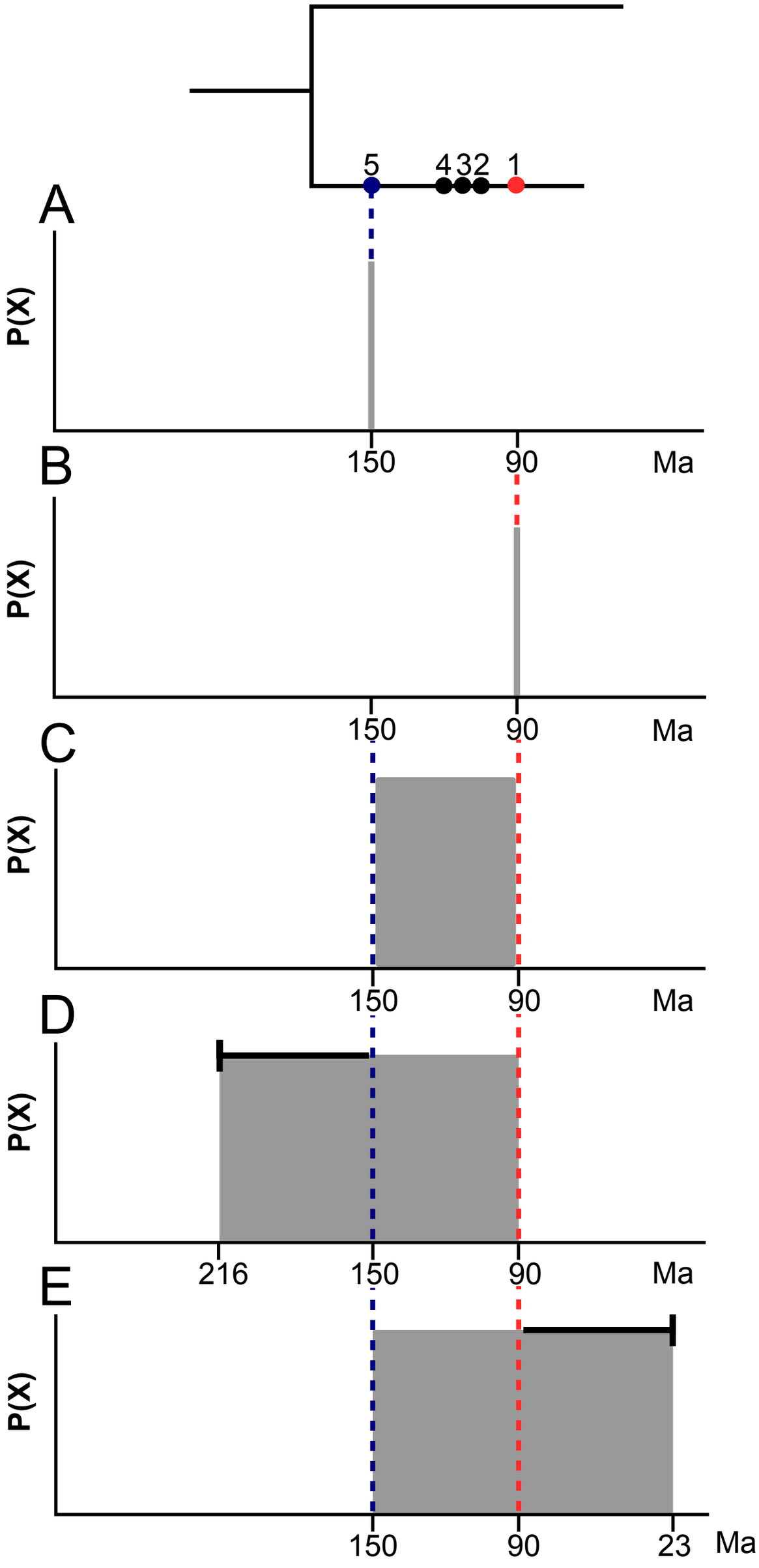
FIG. 7. Divergence-time ages measured in nodes of extant taxa of bears (Ursidae) under five different calibration approaches. The bars and dots represent the 95% highest posterior density (HPD) and the mean for the estimated node ages respectively. In colour the same calibration methods described in the Fig. 1 and in black the results from a previous analyses (Krause *et al.* 2008). Silhouette icons of ursids from phylopic.org reproduced under a Public Domain Dedication License 1.0.

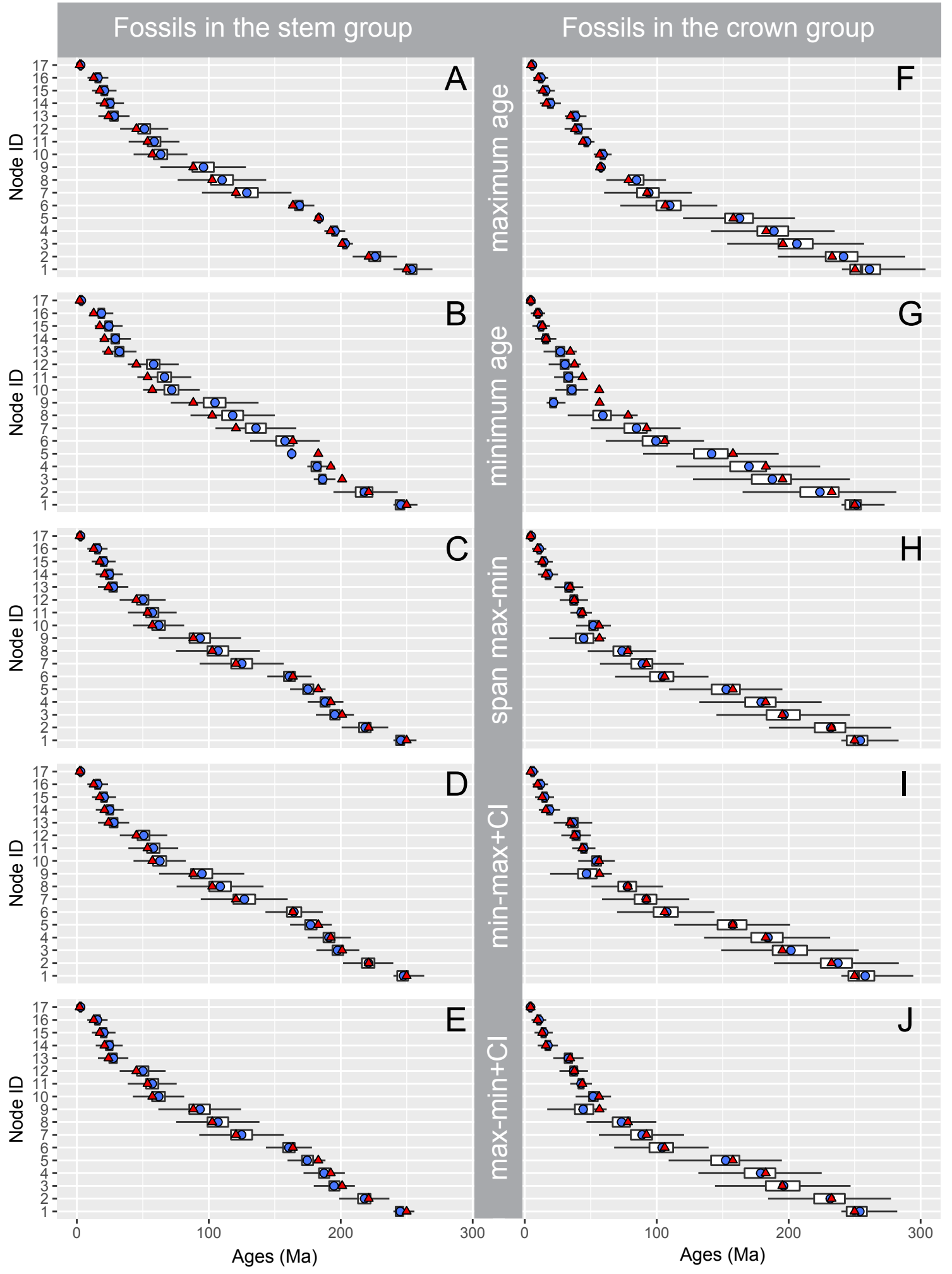
TABLE 1. Posterior estimates of divergence times and 95% HPD widths of the Ursidae phylogeny in the five calibrations approaches described in Fig. 2.

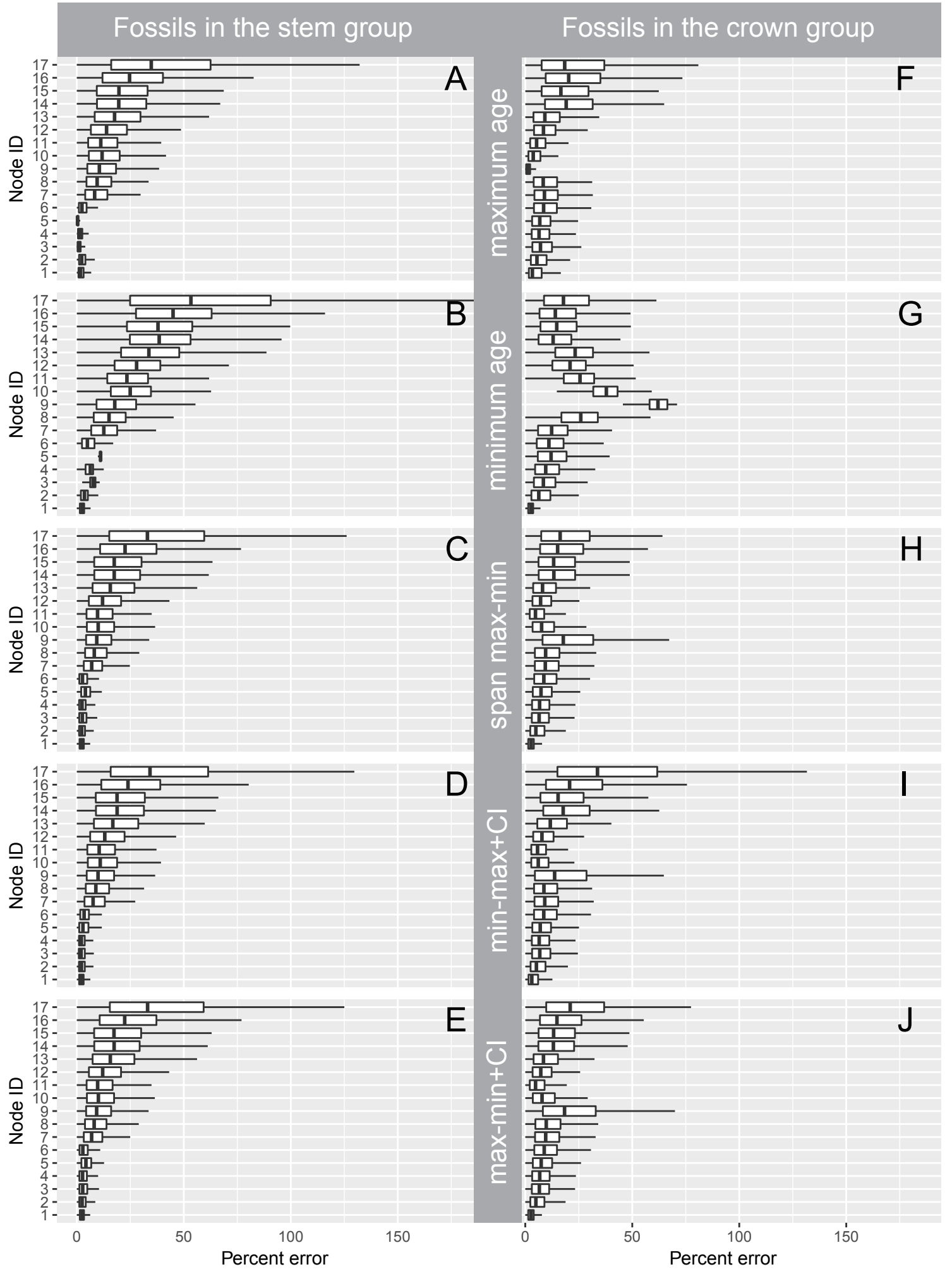
Calibration	Maximum		Minimum		Minimum-maximum		Min-max 95% CI max		Min-max 95% CI min	
	Mean	95% HPDw	Mean	95% HPDw	Mean	95% HPDw	Mean	95% HPDw	Mean	95% HPDw
Node 1	40.35	10.07	38.71	5.87	38.89	6.92	38.99	6.74	38.07	3.67
Node 2	24.9	9.34	14.24	2.84	16.5	6.6	17.31	10.85	17.31	12.3
Node 3	22.08	4.63	10.24	1.73	13.55	5.38	14.11	9.77	13.44	8.8
Node 4	21.07	2.71	9.91	0.97	12.91	5.44	13.47	9.43	11.92	8.19
Node 5	13.3	3.55	8.31	2.88	8.82	3.93	8.87	3.97	8.12	4.39
Node 6	11.97	1.44	5.62	1.21	6.15	2.28	6.15	2.33	2.51	4.72
Node 7	5.98	3.63	4.35	2.41	4.48	2.77	4.51	2.77	4.28	2.83
Node 8	2.15	1.58	1.82	1.17	1.84	1.25	1.86	1.29	1.56	1.27
Node 9	1.95	1.43	1.63	1.05	1.65	1.11	1.67	1.13	1.43	1.15
Node 10	1.72	1.3	1.44	0.96	1.45	1.01	1.47	1.03	1.26	1.03
Node 11	1.46	1.21	1.22	0.88	1.23	0.93	1.25	0.93	1.07	0.93
Node 12	0.36	0.48	0.4	0.41	0.4	0.44	0.4	0.43	0.27	0.37
Rate	0.64	0.25	0.65	0.26	0.65	0.26	0.65	0.25	0.64	0.25
Igrvar	0.08	0.04	0.01	0.11	0.01	0.11	0.01	0.03	0.01	0.03

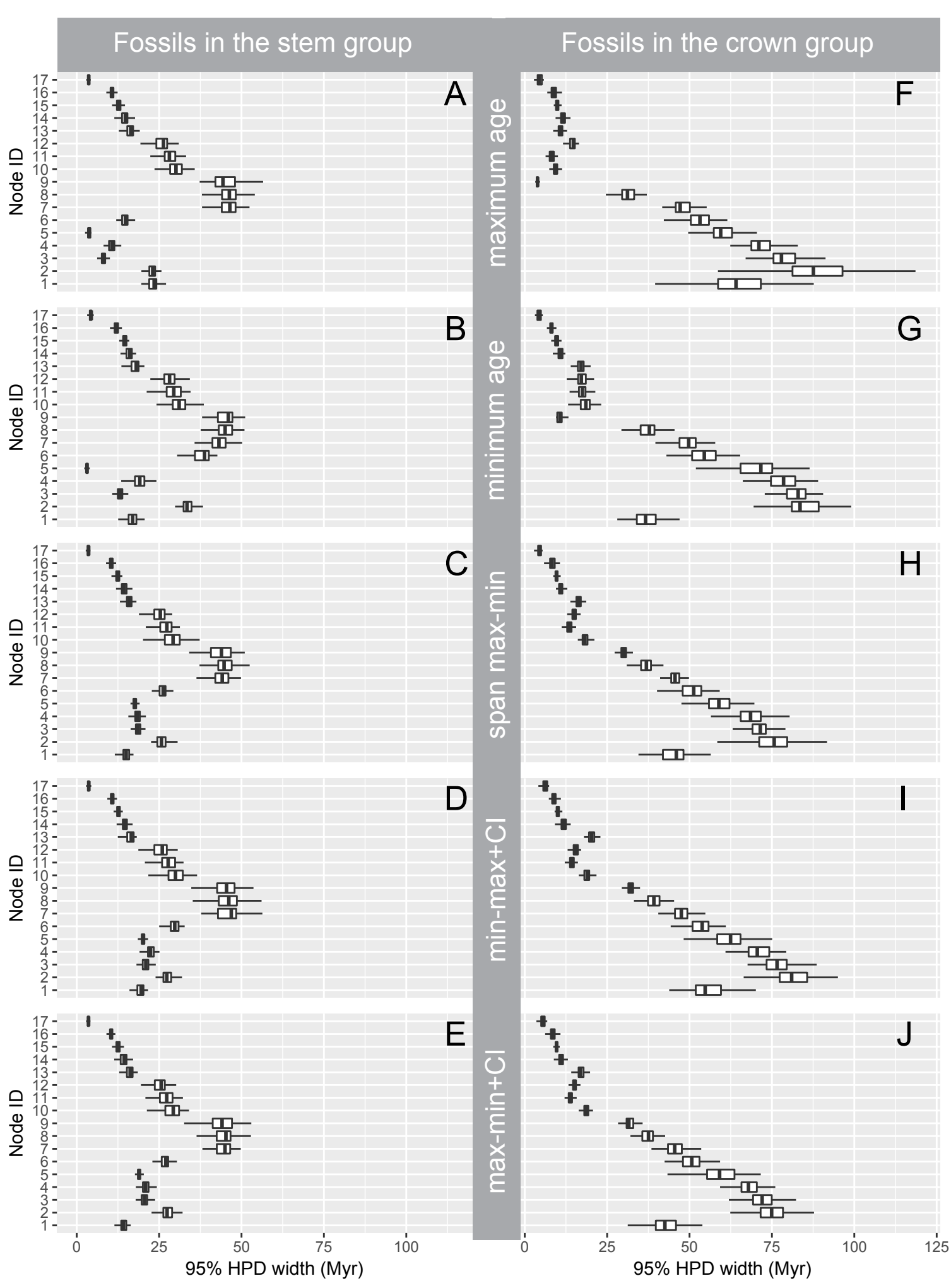
The divergence times are in millions of years ago (Ma). min, minimum; max, maximum; CI, confidence interval; HPD, highest posterior density; w, width; Rate, evolutionary rate (substitutions per site per million years); Igrvar, variation in the rate of evolution across branches from the independent branch rate (IGR) relaxed clock model.

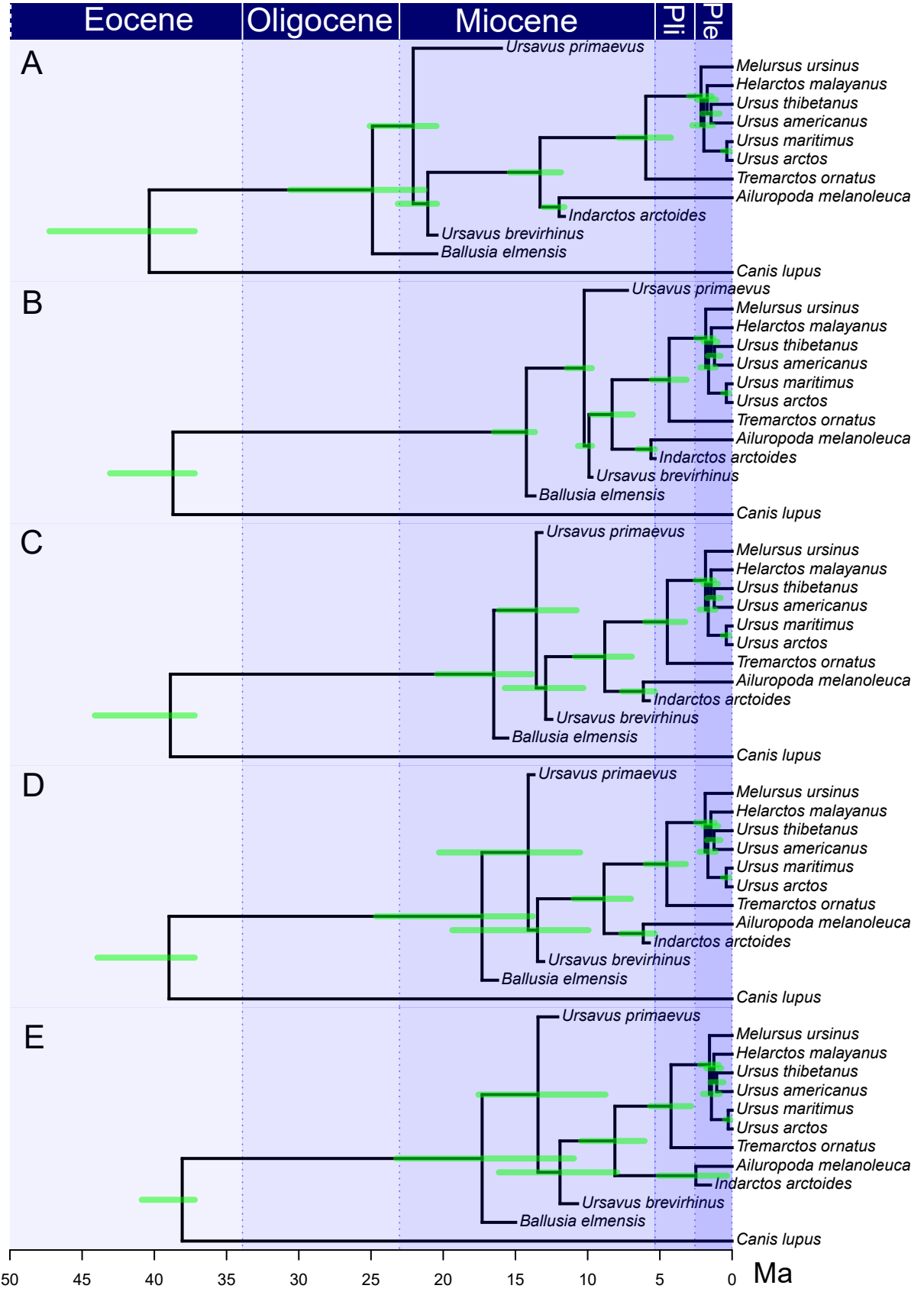

A
183
5
201.1
3
221.2
2
163.7
6
102.6
8
45.1
12
2.2
17
12.8
16
21
14
24
13
120.6
7
192.5
4
249.9
1
57.2
10
17.4
15
88.3
9
53.6
11
t14
t7
t11
t17
t9
t12
t3
t18
t16
t2
t6
t4
t1
t10
t8
t15
t13
t5
B
92.2
7
106
6
232.5
2
182.6
4
34.4
13
4.4
17
9.8
16
16
14
78.2
8
56.6
9
157.7
5
195.2
3
249.9
1
43.6
11
13.3
15
37.6
12
56.4
10
t14
t7
t11
t17
t9
t12
t3
t18
t16
t2
t6
t4
t1
t10
t8
t15
t13
t5
250
234
218
202
186
170
154
138
122
106
92
80
68
56
44
32
20
8
0
Time (Ma)

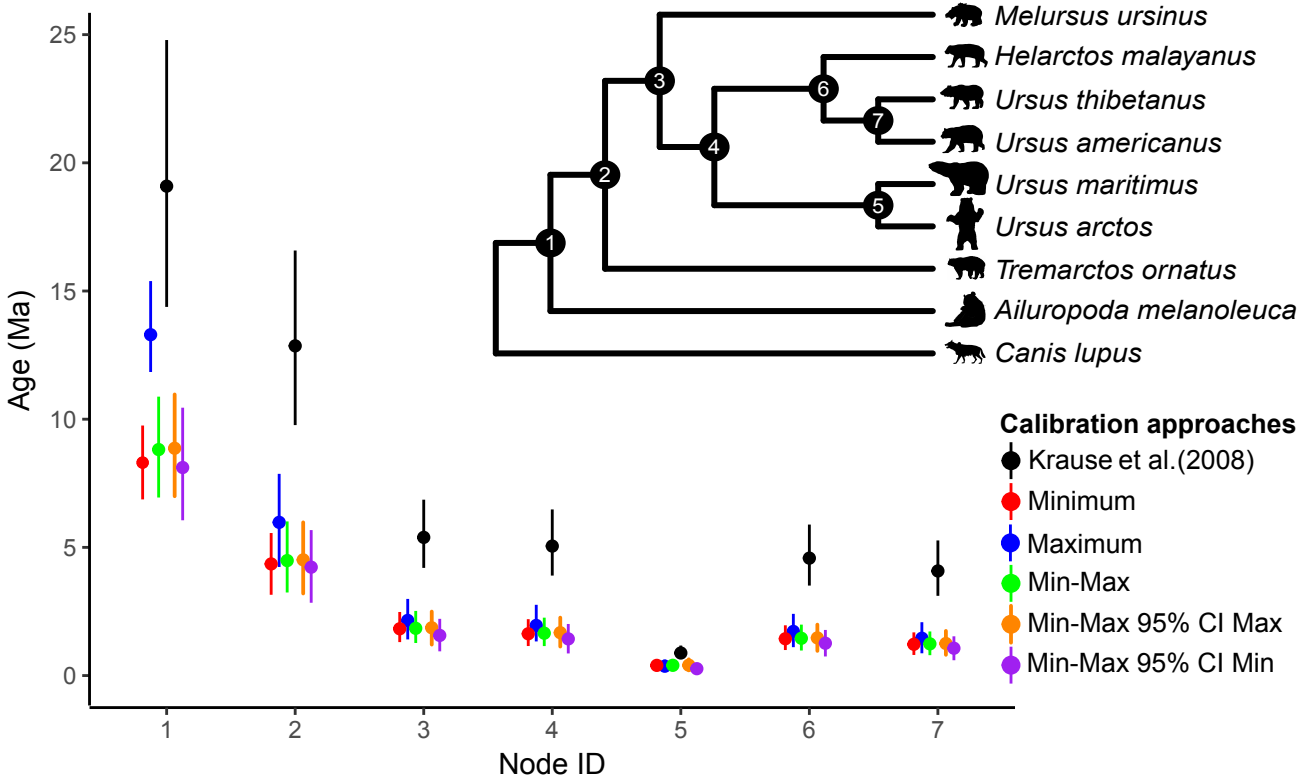

5
4
3
2
1
A
P(X)
150
90
Ma
B
P(X)
150
90
Ma
C
P(X)
150
90
Ma
D
P(X)
216
150
90
Ma
E
P(X)
150
90
23
Ma


Fossils in the stem group
Fossils in the crown group
maximum age
minimum age
span max-min
min-max+CI
max-min+CI
A
B
C
D
E
F
G
H
I
J
Node ID
Ages (Ma)
0
100
200
300


Fossils in the stem group
Fossils in the crown group
maximum age
minimum age
span max-min
min-max+CI
max-min+CI
A
B
C
D
E
F
G
H
I
J
Node ID
Percent error
0
50
100
150


Fossils in the stem group
Fossils in the crown group
maximum age
minimum age
span max-min
min-max+CI
max-min+CI
A
B
C
D
E
F
G
H
I
J
Node ID
95% HPD width (Myr)
0
25
50
75
100
125


Eocene
Oligocene
Miocene
Plio
Ple
A
B
C
D
E
Ursavus primaevus
Melursus ursinus
Helarctos malayanus
Ursus thibetanus
Ursus americanus
Ursus maritimus
Ursus arctos
Tremarctos ornatus
Ailuropoda melanoleuca
Indarctos arctoides
Ursavus brevirhinus
Ballusia elmensis
Canis lupus
50
45
40
35
30
25
20
15
10
5
0
Ma


Melursus ursinus
Helarctos malayanus
Ursus thibetanus
Ursus americanus
Ursus maritimus
Ursus arctos
Tremarctos ornatus
Ailuropoda melanoleuca
Canis lupus
1
2
3
4
5
6
7
Calibration approaches
Krause et al.(2008)
Minimum
Maximum
Min-Max
Min-Max 95% CI Max
Min-Max 95% CI Min
Age (Ma)
25
20
15
10
5
0
Node ID