**Challenges in approximating species age from phylogenies of extant taxa**

Carlos Calderón del Cid1,2, Torsten Hauffe2, Juan D. Carrillo2, Rachel C. M. Warnock3, Daniele Silvestro2,4

1Laboratório de Ecologia Espacial, Instituto de Biologia, Universidade Federal da Bahia, CEP 40170-110, Salvador, Bahia, Brazil

2Department of Biology, University of Fribourg, Switzerland and Swiss Institute of Bioinformatics, Fribourg, Switzerland

3GeoZentrum Nordbayern, Friedrich-Alexander Universität Erlangen-Nürnberg, Loewenichstrasse 28, 91054, Erlangen, Germany

4Department of Biological and Environmental Sciences and Gothenburg Global Biodiversity Centre, University of Gothenburg, Sweden

**Abstract**

**Aim.** Species age, the elapsed time since origination, can give an insight into how species longevity might influence eco-evolutionary dynamics and has been hypothesized to influence extinction risks. Traditionally, species ages have been measured in the fossil record. However, recently, numerous studies have been estimating the ages of extant species from the branch lengths of time-calibrated phylogenies. This approach poses problems because phylogenetic trees contain direct information about species identity only at the tips and not along the subtending branches. Here, we show that taxon sampling, extinction, and different assumptions about speciation modes can significantly alter the relationship between true species age and phylogenetic branch lengths, leading to high error rates. We find that these biases can lead to erroneous interpretations of eco-evolutionary patterns derived from the comparison between phylogenetic age and other traits, such as the current extinction risk.

**Innovation.** For bifurcating speciation, which is the default assumption in most analyses, we propose a probabilistic approach to improve the estimation of species ages, based on the properties of a birth-death. We show that our model can reduce the error by one order of magnitude under cases of high extinction.

**Main conclusion.** Our results call for caution in interpreting the relationship between phylogenetic ages and eco-evolutionary traits, and show that, under some assumptions, it is possible to obtain better approximations of species age by combining branch lengths with the expectations of a birth-death process.

**Key words:** Evolutionary history, Extinction rates, Incomplete sampling, Simulations, Speciation modes, Phylogeny.

**Introduction**

The estimation of species age, or the elapsed time since species origin, is important to evaluate mechanisms that link species longevity with eco-evolutionary processes (Benton, 2013; Swenson, 2019). For instance, age-dependent extinction hypotheses test the relationship between species age and extinction probability, assessing whether extinction rates differ between young and old species (Balmford, 1996; Eldredge et al., 2005; Pearson, 1995). Likewise, species age could be a measure of colonization time, especially in island systems (Tanentzap *et al.* 2015) or during biotic invasions triggered by geological events, such as the formation of the Central American Isthmus for the Great American Interchange (Carrillo *et al.* 2015, 2020). Species age is measured in the fossil record through different statistical and probabilistic approaches based mostly on taxa’s stratigraphic duration (Foote, 1996; Foote & Raup, 1996). Several of these approaches consider the differences in fossil sampling and temporal resolution (Alroy et al., 2001; Silvestro et al., 2019). Species ages estimated from paleobiological data offer a reliable measure of species’ temporal duration which can be used in macroevolutionary studies (Benton, 2016; Silvestro et al., 2020; Van Valen, 1973). More recently, several studies have used the length of terminal branches in time-calibrated phylogenies as a proxy for the age of extant species, an approximation that we hereafter refer to as “phylogenetic age” (Davies et al., 2011; Gaston & Blackburn, 1997; Johnson et al., 2002; Pie & Caron, 2023; Sonne et al., 2022; Tanentzap et al., 2020; Verde Arregoitia et al., 2013). These approximate species ages have been used as the basis to test for links between age and current extinction risks [REF] and to assess various correlations with evolutionary, biogeographical, and ecological patterns in living species [REFs].

While sir Specifically, we identify three non-exclusive shortfalls that can lead to lead to over- or underestimation of species ages. First, incomplete sampling of extant species, either due to the lack of genetic data or linked to species being still unknown to science, can bias phylogenetic age estimation by artificially increasing the length of terminal branches (Heath et al., 2008; Mynard et al., 2023).

Second, extinction events will mask branching events in phylogenetic trees of extant species (Harvey et al., 1994; Nee & May, 1997). Even in phylogenetic trees that include extinct taxa, the incompleteness of the fossil record will inevitably lead to missing lineages and incorrect topologies. Unsampled extant and extinct species from the phylogeny results in an inflation of the length of terminal branches leading to sampled species (i.e., the tips of the tree), thus altering phylogenetic species ages. For instance, if the extinct species of the *Homo* genus are not included in a phylogeny, the phylogenetic age of *Homo sapiens* is approximately 10 million years, i.e., the age of the last common ancestor with its sister species, the chimpanzee (Rivas-Gonzáles et al. 2023). This estimate exceeds the age of the oldest known fossil of modern humans by two orders of magnitude (Fig. 1; Callaway 2017).

The third shortfall, is that the tree alone does not contain information about the underlying speciation mode and does not include species labels along its branches, such that only the tips can be unequivocally assigned to a named species (Losos & Glor, 2003). Alternative speciation modes have been discussed in the literature reflecting different biological processes and species concepts, including bifurcating, budding, and anagenetic speciation (Foote, 1996; Silvestro et al., 2018). These modes define the relationship between the ancestral species and its descendants, thus contributing to determining species ages (Rosenblum et al., 2012; Wagner, Erwin, & Anstey, 1995) (Fig. 2). Most phylogenetic trees are depicted in the rectangular shape where the two descending lineages split symmetrically from an ancestral lineage, thus suggesting a bifurcating speciation mode where two new species replace the ancestral lineage (Baum et al., 2005; Caetano & Quental, 2022). However, the often-unstated assumption of virtually all birth-death processes used to model phylogenetic branching times, is that speciation occurs as a budding process, with a speciation event leading to a single new species and the survival of the parent species, even though we cannot determine which descendant branch is the new species (Gernhard, 2008; Nee et al., 1994; Stadler, 2013). Anagenetic speciation, in contrast, does not lead to a branching event and is therefore not visible on a phylogenetic tree.

All speciation modes may reflect plausible biological processes, and evidence for each mode has been found in the fossil record (Foote, 1996; Simpson, 1984) and in extant species (Skeels & Cardillo, 2019). Bifurcating speciation can be linked with vicariance or allopatric speciation (Willmann & Meier, 2000), and the Hennigian species concept considers all speciation events as bifurcating (Henning, 1999). Under this speciation mode, phylogenetic ages equal species ages when there is no extinction, and all species are sampled. Budding speciation can be interpreted as the result of parapatric, peripatric, and founder-event speciation (Anacker & Strauss, 2014; Caetano & Quental, 2022) and is related to the evolutionary species concept, which considers any species as descendant of an ancestral lineage with its own evolutionary identity (Simpson 1951, but see Simpson 1961). A phylogenetic tree typically lacks information about which of the two descendent branches is the new species after a speciation event (but see Aze *et al.* 2011). Thus, in the absence of extinction, the phylogenetic age of one sister species will equal its species age while the other will be older but without the possibility to identify which one is which. Anagenetic speciation is not associated with a branching event but can be used to describe different species or morphospecies (Emerson & Patiño, 2018) delimited by substantial phenotypic change occurring along a lineage (Roopnarine et al., 1999) and will cause a higher phylogenetic age than the genuine species age.

Here we use simulations to quantify the predictability of species age from phylogenetic trees of extant taxa, under different diversification scenarios. Specifically, we performed simulations where we know the true age of species to: 1) quantify the error in phylogenetic ages under various scenarios combining different speciation modes with a range of speciation and extinction rates and incomplete sampling; 2) examine whether this error affects our ability to make qualitative decisions on the order of species ages; 3) explore whether the signal of labeled age-correlated extinction risk is preserved in the phylogenetic age of species. Finally, we propose a new method to estimate species age more accurately under the assumption of bifurcating speciation, which is the most widely used assumption in eco-evolutionary studies (e.g., Tanentzap et al. 2020), and assess its ability to improve our interpretation of age-dependent extinction risks.

**Methods**

*Simulating species ages*

We generated complete phylogenies of extant and extinct species under a stochastic birth-death process using the package TreeSim 2.4 (Stadler, 2010) for the R 4.3.0 statistical programing environment (R Core Team 2023). Then we mapped species on the complete phylogenies using the R package FossilSim 2.3.1 (Barido-Sottani et al., 2019) under different speciation modes, thus assigning species labels across all branches of the tree. We used the labels assigned to terminal extant taxa to determine the true species ages. We then dropped all extinct species from the tree and obtained the length of terminal branches, to quantify the phylogenetic age of extant species. Finally, we rescaled all phylogenetic trees to a root age of one, which ensures that the absolute errors in species ages are comparable in plots, and compared the relative true and phylogenetic ages among different simulation scenarios.

*Error in equating phylogenetic and species age*

To explore whether there is a consistent over- or underestimation of species ages and to quantify error in approximating species ages with phylogenetic ages, we simulated a range of datasets with different speciation modes and diversification rates. First, we simulated 3 sets of 100 phylogenetic trees with 100 extant species based on speciation rates equal to 0.1, 0.5, and 1, combined with 100 extinction rates ranging from 0 to 0.99 in equal increments (Beaulieu & O’Meara, 2016). Second, on each of these phylogenies, was then mapped species according to different scenarios of speciation: (1) budding speciation, (2) bifurcating speciation, (3) a combination of budding speciation and anagenetic speciation with the rate of anagenesis set to half of the birth rate, and (4) bifurcating speciation combined with anagenetic speciation with the rate of anagenesis set to half of the birth rate.

Across all trees, we obtained in total 120,000 extant species, 30,000 for each speciation scenario. For each speciation mode and extinction fraction rate (defined as death/birth Beaulieu & O’Meara, 2016), we calculated the mean absolute percentage error (MAPE) across all species for each tree as measure of the deviation between the phylogenetic ages from the true age:

 ,

where *s* is the true species age, *v* is the phylogenetic age, and *n* is the number of tips in the tree (pruned of the extinct species).

Next, for each speciation mode, we plotted the MAPE against the simulated extinction fraction.

*Impact of age error in comparing species ages*

To explore whether the error introduced by approximating species age with phylogenetic age impacts our ability to make qualitative judgements such as which of two extant species is the younger one, we simulated 1,000 phylogenetic trees with values of extinction fractions (extinction rate divided by speciation rate) of 0.9, 0.5, and 0, combined with a fixed speciation rate of 1 (3000 trees). Second, on each of these phylogenies, we mapped species according to budding and bifurcating speciation. Thus, we simulated 300,000 extant species for each speciation mode. Next, we calculated the proportion of cases where the younger of two species, according to its phylogenetic age, is, in fact, the older one given the true age of the two species. We performed this comparison from the perspective of an empirical research that can only obtain the phylogenetic ages. We made two types of comparisons for each phylogeny: (1) between the youngest and oldest species in the phylogeny, and (2) between two randomly selected species.

*Error in the phylogenetic age due to uniform incomplete sampling*

We used the intermediate extinction scenario with an extinction fraction of 0.5 to explore the error in equating phylogenetic and true age that is introduced by uniform incomplete sampling under the scenarios of bifurcating and budding speciation. We specified a fully sampled phylogeny and 25% and 50% of incomplete sampling. We simulated trees such that they included sampled 100 tips, i.e., after dropping the unsampled ones. We calculated the MAPE for each tree and compared the incomplete sampling scenarios for each speciation mode.

*A probabilistic method to infer species age*

Under the assumption of bifurcating speciation, the phylogenetic age represents the upper boundary of plausible species ages and corresponds to the true age in the absence of extinction. However, the true age could be younger if extinction led to the disappearance of recent cladogenetic events from the phylogeny of extant species. Given a phylogenic age *vi* the probability that the true species age *si* is exactly *vi* is conditional on no other speciation or extinction event having occurred between *vi* and the present. We approximate the probability of no speciation or extinction for an arbitrary small time bin *t* based on the probability that a lineage results in a single descendant, which is (Kendall, 1946):



based on a birth-death process with time-homogenous speciation rate λ and extinction rate μ. The probability that no event occurs over a time window *v, i.e.,* until the age of the observed node *i*,is approximated as:



In the absence of extinction (μ = 0), the probability of the true species age *si* to be equal to *vi* is 1, because any speciation event following the node *vi* would be observed in the tree of extant species. Thus, we calculate the normalized probability of the speciation event to occur at time *vi* as:

 (1)

We then compute the probability of a speciation time for any given time *τ* as:

, (2)

where the first term quantifies the probability of the species age a long a vector of times, while the second term normalizes it after accounting for the probability that speciation occurred exactly at the node. We use equations (1) and (2) to approximate a density describing the probability of a species origination at the observed phylogenetic age and along the branch connecting the node with tip. As point estimates of the species age, we tested whether the mean or the median of the distribution better approximate the true species age.

To evaluate the accuracy of our probabilistic species age estimator, we simulated 1,000 phylogenetic trees with extinction rates of 0.25, 0.15, and 0.05, and a fixed speciation rate of 0.3. Using these phylogenies as inputs, we mapped species on each tree based on a bifurcating speciation process. Then, we calculated the MAPE for each tree as measure of the deviation between the function’s estimated and the phylogenetic ages from the true age.

*Simulation of age-dependent species extinction risks*

To evaluate the impact of the erroneous estimation of species age due to the extinction shortfall on macroevolutionary analyses, we explored whether the assumed relationship between species age and contemporary extinction risk (e.g., Johnson et al. 2002; Tanentzap et al. 2020; Verde Arregoitia et al. 2013) is preserved in the phylogenetic ages. For this, we binned the same number of extant species according to their age in five categories reflecting the increase in extinction risk with age encapsulated by the IUCN categories: Least Concern (LC), Near Threatened (NT), Vulnerable (VU), Endangered (EN), and Critically Endangered (CR; International Union for the Conservation of Nature 2016). With this, we generated a positive effect with older species being at higher extinction risk regarding the IUCN categories, assuming bifurcating speciation.

Then, we quantified the share of the 1000 datasets where the order of the mean age per IUCN category did not match with the simulated monotonic increase when utilizing (a) phylogenetic ages, and (b) the probabilistic species age estimator. We also evaluated the effect of nonrandom incomplete sampling on the evaluation of species age-correlated extinction risks.

**Results**

*Error in equating phylogenetic and species age*

Under the assumption of bifurcating speciation and with no extinction events, phylogenetic ages matched the true age of extant species (Fig. 3). At low extinction fraction (< 0.25), 96% of the phylogenetic age estimations were congruent with the true age. At higher extinction fractions (> 0.75), this was also the case for most species (73%). However, age overestimation increased with extinction fraction and in some cases the phylogenetic age erroneously suggested that the species is as old as the root age. While under bifurcating speciation, the phylogenetic age never underestimated the true species age, both over- and underestimation occurred in the case of budding speciation. Moreover, the proportion of cases were the phylogenetic ages equal the species age was lower than in the bifurcating scenario (Fig. 3). Overestimated ages were more frequent with high extinction while underestimations occurred with low extinction but in principle both happened under the complete range of extinction rates (Fig. 3). Even at low extinction fraction, ~50% of phylogenetic ages did not match the true ages.

In datasets simulated under a mixture of anagenetic and bifurcating speciation, phylogenetic ages deviated more strongly from the true ages than under a pure bifurcating process, given that anagenetic events are unobserved in the phylogeny (Fig. SM1). With a low extinction fraction, phylogenetic ages were congruent with the true species ages in 78% of the cases instead of 96%, and with high extinction the share decreases from 73% to 62%. Datasets with mixed anagenetic and budding speciation, phylogenetic ages also deviated more that under a pure budding process; with a low extinction fraction, phylogenetic ages were congruent 40% instead of 50%, and with high extinction the accuracy decreased to 34%.

While a budding speciation mode led to a higher baseline error than bifurcation, the latter showed a stronger increase with extinction (Fig. 4). Under both modes of speciation, speciation rates did not have a substantial impact on error in age. For strictly bifurcating speciation, there was no error in the absence of extinction, but the MAPE increased to up to 150% with extinction fractions exceeding 0.75. In contrast, under budding speciation the MAPE was around 25% in the absence of extinction, increasing to 30-120% with extinction fractions exceeding 0.75. In datasets that included anagenetic speciation, the MAPE reached as high as 500% in some simulations (Fig. SM2).

*Impact of age error on comparing species ages*

For the combination of strictly bifurcating speciation and all extinction scenarios, selecting the phylogenetically youngest and oldest species never resulted in a case where the presumed older species has been in fact the younger of the two species according to their simulated age (Fig. SM3a). Thus, for this speciation mode, the risk of a qualitative error when comparing species at the extremes of the age range is minimal. Howeverspecies found to beFig. Thus, qualitative errors in comparing species ages are non-negligible under the assumption of bifurcating speciation.

In contrast, for budding speciation, the age rank of the oldest and youngest species was erroneously determined in 2.2% of the simulations in the absence of extinction, increasing to 7.5% and 12.2% for intermediate and high extinction, respectively (Fig. SM3b). Thus, under the assumption of budding speciation, there is a substantial risk of mistaking a younger and an older species for the oldest and youngest species in the clade, respectively. The error in age ranking of two randomly selected species was even higher, exceeding 25%, irrespectively of the extinction level (Fig. SM4b).

*Error on equating phylogenetic and species age given uniform incomplete sampling*

Under bifurcating speciation, the MAPE for completely sampled trees was 25 ± 20%; for trees missing 25% of the extant species the error increased by 15 orders of magnitude; and for trees missing 50 % of the extant species the error increased by 85 orders of magnitude (Fig. 5a). Under budding speciation, the MAPE for completely sampled trees was 42 ± 15 %; for trees missing 25 % of the extant species the error increased by 5 orders of magnitude; and for trees missing 50% of the extant species the error increased by 22 orders of magnitude (Fig. 5b).

*Probabilistic species age estimation*

With increasing extinction, our probabilistic estimation of species ages resulted in an increasingly lower error compared with the phylogenetic age (Figs. 6, SM5). Under low extinction the MAPE was slightly worse compared to the use of phylogenetic ages (ΔMAPE = 2.83 ± 3.9 % when using the mean of the estimated ages and 2.1 ± 1.3 % for the median across estimates). While with increasing extinction rates the MAPE of the estimated species ages increased up to 25% (Fig. SM5), the probabilistic estimation strongly reduced the error compared with phylogenetic age (ΔMAPE = -27 ± 26 % for the mean estimated ages and -32 ± 28 % for the median across estimates; Figs. 6, SM5).

*Detecting age-dependent extinction risk*

The use of phylogenetic age as an approximation of species age led to error rates of 1.3, 7.2, and 18.6 % in detecting the correlation between species ages and extinction risk for scenarios with low, intermediate, and high extinction rates, respectively (Fig. 7). Thus, even under intermediate extinction the true relationship between age and extinction risk was wrongly estimated in a significant fraction of the simulations, and higher extinction rates led to a further substantial drop in the reliability of this approach. In contrast, estimating species ages based on our probabilistic method led to much lower error rates (3 to 4-fold) that dropped to 1.3% and 4% with intermediate and high extinction, respectively. Under incomplete taxon sampling, in which the sampling probability was negatively correlated with species age, the error rates increased to 12.8%, and 49.5% for scenarios with 25%, and 50% of missing extant species, respectively (Fig. SM6).

**Discussion**

We showed that approximating speies age with phylogenetic branch lengths leads to substantial errors and that the accuracy of this approximation and that the accuracy of this approximation is hampered by three shortfalls: unobserved extinction events, unknown speciation mode, and incomplete sampling of extant species. The only instance in which phylogenetic ages correctly predict species age is under the assumption of bifurcating speciation process and in the absence of extinction (Meier & Willmann, 2000). While the prevalence of true speciation modes remains difficult to estimate [Silvestro et al 2018 and REF], the fossil record unequivocally shows that extinction occurs across virtually all clades in the tree of life [REF], thus making this a very unlikely instance. Under budding speciation, phylogenetic age shows a high error even without extinction, due to the fact that phylogenetic trees are agnostic about parent and descendant species following a branching event (Fig. 4). Similarly, anagenetic speciation, combined with budding or bifurcating speciation, resulted in high errors, which did not vary substantially with extinction. However, anagenetic speciation is virtually impossible to quantify, except perhaps in high resolution fossil time series [REF], resulting in a general debate on the use of the term in evolutionary biology (Vaux et al., 2015) and biogeography (Emerson & Patiño, 2018; Meiri et al., 2018).

Our probability age estimator performed well when compared with phylogenetic age, especially in high extinction scenarios.

For bifurcating speciation, the mean error between true and phylogenetic age was the lowest when compared with the other speciation modes. This is because taking the most recent ancestral node as phylogenetic age results in sister species to be identical in age, the same as implied by bifurcating speciation. Due to only introducing bias by extinction, there was a low risk of confusing the oldest with the youngest species (Fig. SM3) but a considerable risk to jumble the order of two random species (Fig. SM4). Therefore, under a bifurcating speciation scenario, the phylogenetic age indeed captures most of the signal of the true species age. For budding speciation, there is a high discrepancy in ages, even in no-extinction scenarios. However, the already overall high error does not increase further with augmenting extinction rates. This is due to half of the phylogenetic age estimations in no-extinction scenarios already underestimate the true age as one of the sister lineages is the older, ancestral one. When extinction rates increase, overestimation of ages start and increase until it balances the inherent underestimation at high extinction scenarios. This happens due to the phylogenetic assumption of sister species being identical in age, which is (a) wrong for budding speciation, and (b) under no extinction, only correctly estimates the youngest species of any bifurcation event. Due to these misinterpretation patterns, the qualitative error is high and increases considerably with extinction rates in the example of the oldest vs. youngest species (Fig. SM3). Thus, for clades diversifying predominantly through budding speciation, the phylogenetic age is not equal to the true age for half of the species, and some of these mismatches would be large regarding the difference between phylogenetic and true age.

Given the large inaccuracy in phylogenetic age, especially for taxa originating under budding speciation, the question is whether this affects the inferences made from the relationship between species age and eco-evolutionary variables, such as extinction risk, range size, or environmental variables (Gaston & Blackburn 1997, Johnson *et al.* 2002, Tanentzap *et al.* 2015, Pie & Caron 2023). If phylogenetic ages capture at least the relative differences in species ages, it could be argued that it is a valuable measure of species longevity. Some authors acknowledge the problems associated with measuring species age from phylogenies (Swenson, 2019), and have proposed approaches to account for them. For example, Sonne *et al.* (2022) determined young and old Andean hummingbirds by assessing the sensitivity of their results to incomplete taxon sampling by generating 1000 trees with randomly missing species. Pie & Caron (2023) accounted for taxonomic incompleteness by pruning an additional 1 – 5% of species and evaluated if their conclusions changed. Although incomplete taxon sampling caused the highest error rates in our evaluation of age-dependent extinction risk (Fig. 5), neither strategy acknowledges the other two problems of age uncertainty: the effects of extinction, which for most groups is probably high (Pimm et al., 2014), and the speciation modes. Our evaluation of the imprint of species ages on extinction risk on makes this point clear because it showed that the number of incorrect inferences is considerably high (around 18%) for high extinction scenarios (Fig. 7), even with the simulated strong extinction signal and the assumption of bifurcating speciation, which presents the lowest mismatch regarding true age.

The incomplete sampling of extant species can be caused by (a) named species that are not included in a phylogeny due to, for instance, no available DNA sequence data, or (b) species unknown to science that have yet to be collected and described. The latter refers to the Linnean shortfall, one of the seven key shortfalls of biodiversity (Diniz Filho et al., 2023; Hortal et al., 2015). The magnitude of the Linnean shortfall is unknown, but available estimates [REFs] show that it affects some clades significantly more than others, with the diversity of highly diverse groups, such as insects and fungi, likely to be a severe underestimation [REF]. ~~Obtaining a precise estimate of the number of undescribed extant species is challenging (Caley et al. 2014) and varies based on the taxonomic group and the geographical region under consideration (Hopkins 2007; Vilela et al. 2014). Moreover, the nature of the Linnean shortfall probably is nonrandom, given that is more severe for species with small body sizes, spatial distributions, and niche widths (Riddle et al. 2011).~~

Our results point out the errors associated with budding and bifurcating speciation; however, how do we know which is the predominant speciation mode of a clade? There is some debate on whether this is possible in first place (Losos & Glor, 2003). However, Anacker and Strauss (2014) proposed that budding speciation leaves its signatures on sister species: they should have overlapping or adjacent ranges, their range sizes should be asymmetrical, and specific ecological traits should differ between them. These signatures are associated with sympatric and peripatric speciation (Barraclough et al., 1998) and were used to estimate speciation modes (Skeels & Cardillo, 2018) Bifurcating speciation can be associated with allopatric speciation and the signatures it leaves on the range of sister species: ranges should not be overlapping or adjacent, range sizes should not necessarily be asymmetrical, and ecological traits should not necessarily differ between them (Barraclough & Vogler, 2000; Fitzpatrick & Turelli, 2006). Although extinction has a lower leverage on the error of species ages than incomplete taxon sampling and the unknown mode of speciation, extinction rates could be estimated from the phylogeny, the fossil record, or a combination of both (Brée, et al., 2022; López-Martínez et al., 2023; Rabosky, 2010; Silvestro et al., 2014). Thus, assessing a clade’s speciation mode (through spatial dynamics) and extinction rates before performing species age analyses could help to approximate the error of phylogenetic ages.

For instance, let us suppose that the approaches described in the previous paragraph, suggest that the studied clade speciated predominantly under bifurcating speciation and that we can reliably estimate speciation and extinction rates. In this case, we advise to use our probabilistic age estimator to correct phylogenetic ages. When compared with phylogenetic age, this estimator improved the overall accuracy of age estimation across species in a phylogenetic tree, particularly under a high extinction scenario. Additionally, the function enhanced the power to capture the proper relationship between species age and extinction risk (Fig. 7). Therefore, it can be used to test more robustly how species age influence eco-evolutionary dynamics.

**Conclusion**

This study aimed to estimate the potential deviations between true and phylogenetic age due to incomplete taxon sampling, extinction, and unknown speciation modes. Using simulations, we identified that mostly missing species and budding and anagenetic speciation cause a high mismatch between phylogenetic age and true species age. By contrast, only when a clade is fully sampled, speciated under bifurcating speciation, and possesses a low extinction rate phylogenetic age is a good proxy of species age. Thus, inferences made with phylogenetic age should be taken with caution. Besides pointing out these shortfalls, we derived a probabilistic age estimator to correct species age under the assumption of bifurcating speciation and a completely sampled tree. We showed that the probabilistic estimator has a good performance, particularly in high extinction scenarios, both in the overall age accuracy and when species ages are used in eco-evolutionary analyses. Thus, we advise its implementation when the assumptions are met. We hope this paper will stimulate discussions about the species age information in phylogenetic trees and a critical evaluation of the robustness of correlating with species traits or ecological variables.

**Acknowledgements**

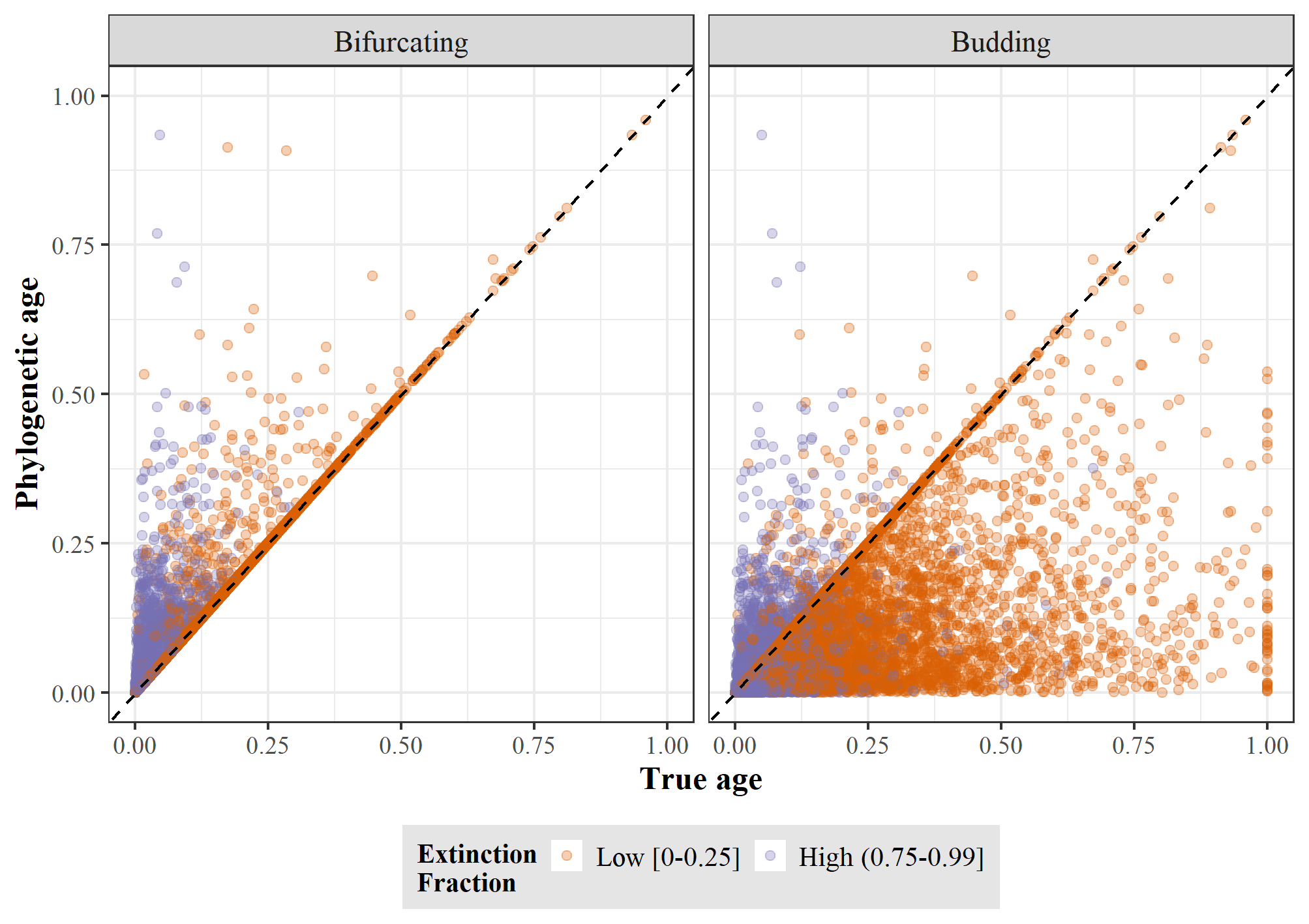
CC received a Ph.D. scholarship from CAPES (88887.814725/2023-00) and an abroad internship CAPES-Print scholarship (88887.682496/2022-00). D.S. and TH received funding from the Swiss National Science Foundation (PCEFP3\_187012). D.S. received funding the Swedish Research Council (VR: 2019-04739), and the Swedish Foundation for Strategic Environmental Research MISTRA within the framework of the research programme BIOPATH (F 2022/1448).



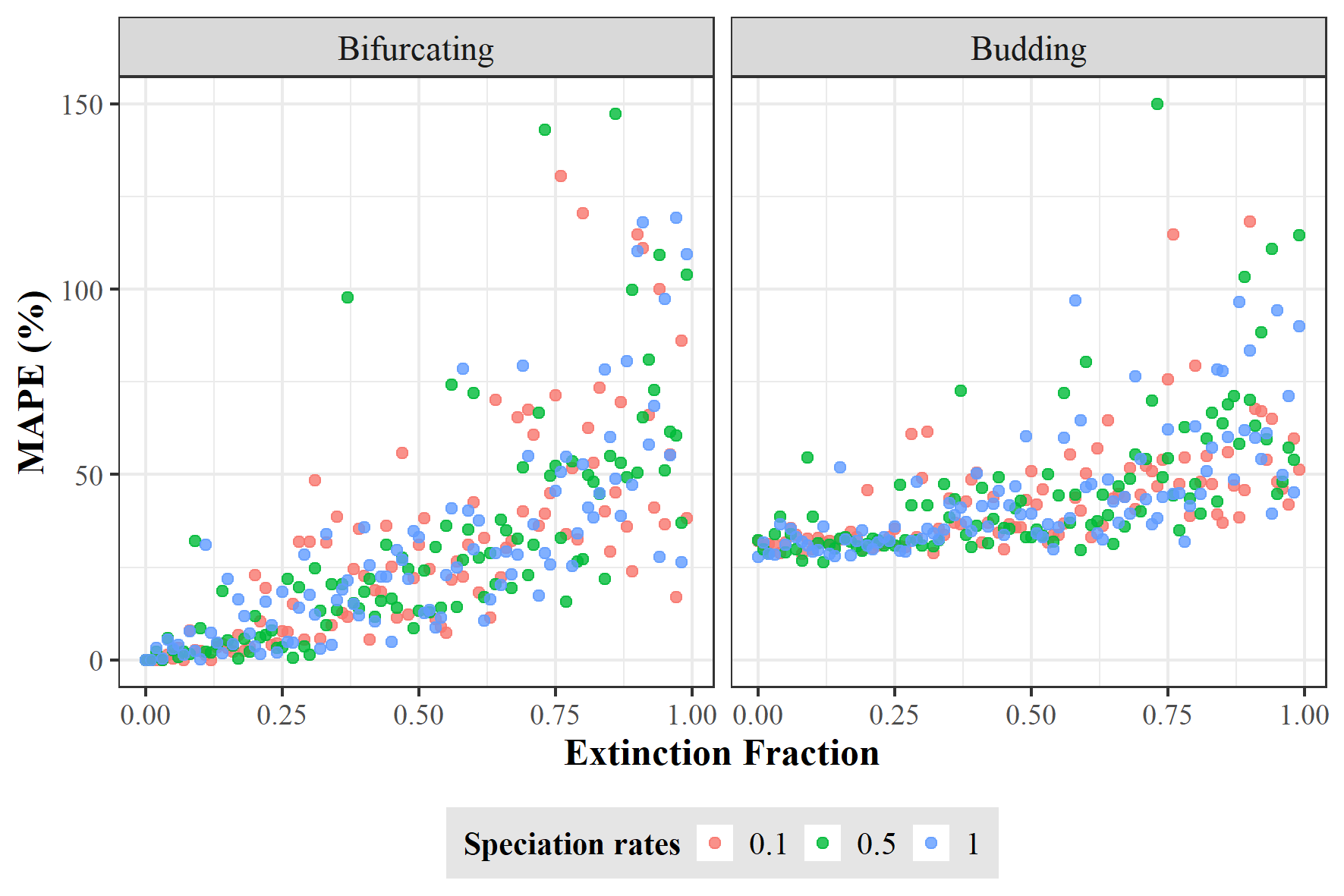
**Fig. 1.** Discrepancy between species ages. Estimations based on the length of the terminal branch in a time-calibrated phylogeny (red) and the first appearance in the fossil record (green). Shark data (*Trianenodon obesus* and *Carcharhinus obscurus*) obtained from Brée et al. (2022). Mammals’ phylogenetic data (except *Homo sapiens*) obtained from Upham et al. (2019). Mammals’ fossil data (except *Homo sapiens*) obtained from Silvestro *et al.* (2018). *Homo sapiens* fossil and phylogenetic data obtained from Callaway (2017) and Rivas-Gonzáles et al. (2023), respectively.



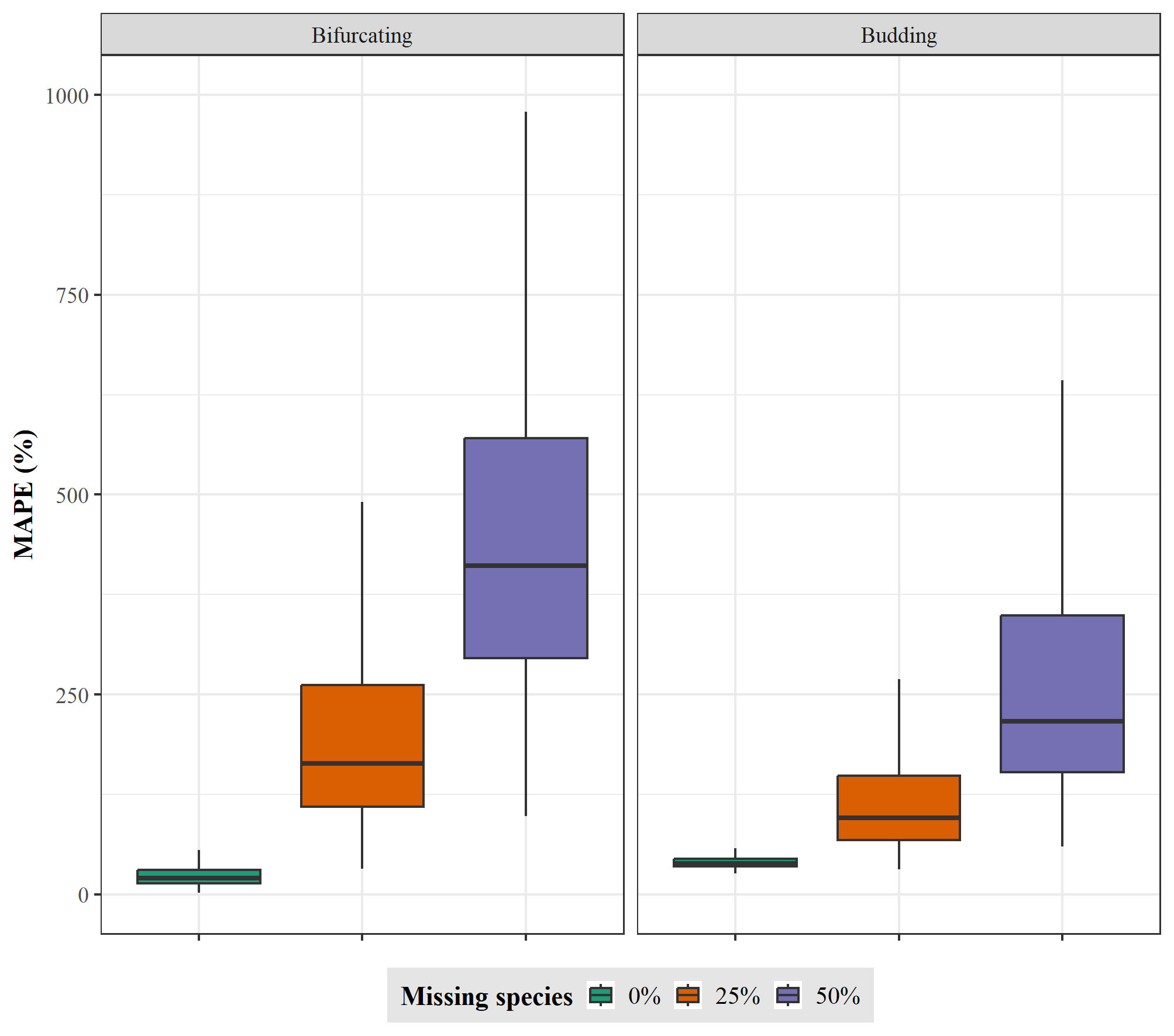
**Fig. 2.** Impact of speciation mode and extinction on species age. For the same hypothetical time-calibrated phylogeny of extant species, the continuation of the same color indicates the same species, solid and dashed lines represent extant and extinct species, respectively, under different speciation modes and extinction scenarios. Numbers display the resulting age of the respective species in million years (myr).



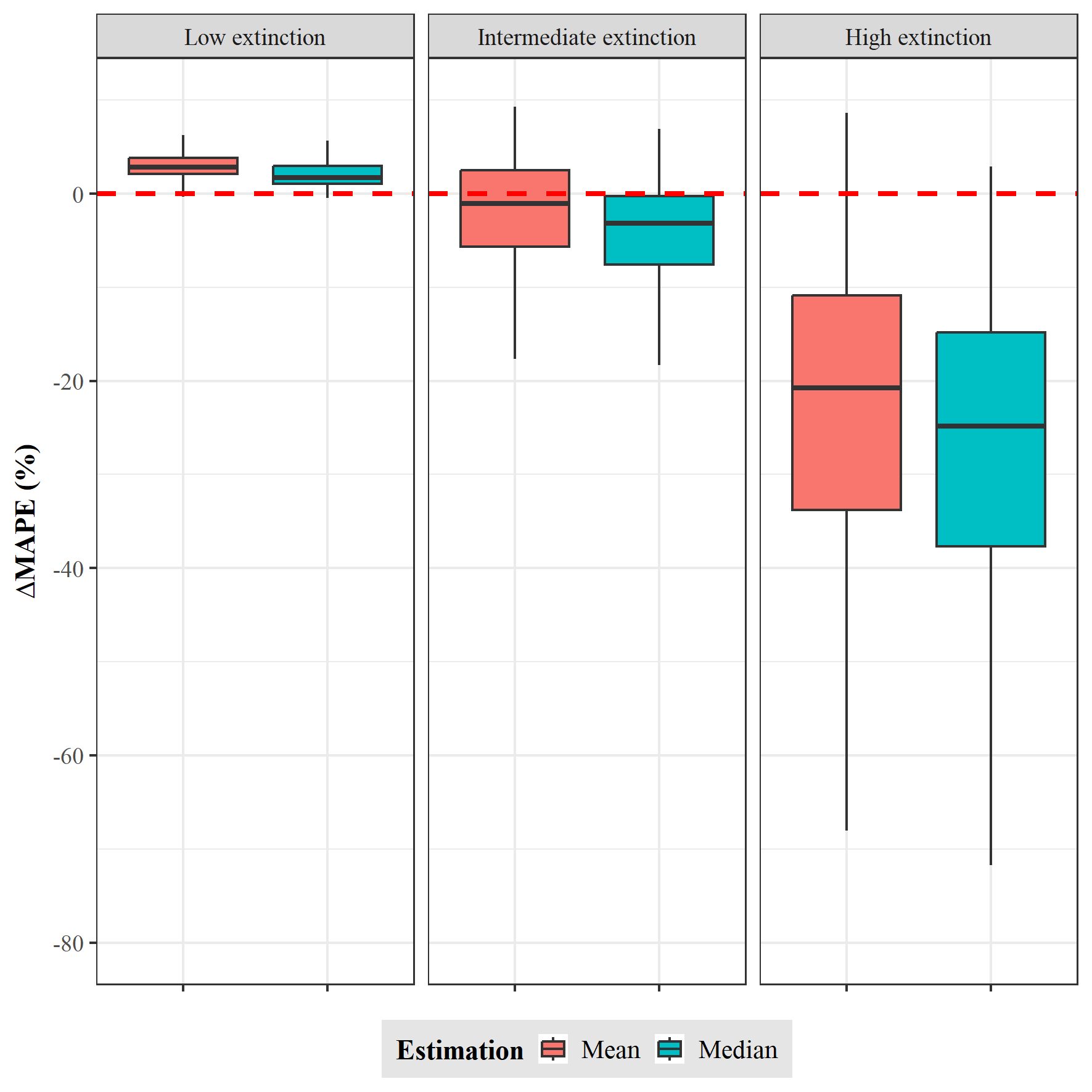
**Fig. 3.** True age versus phylogenetic age at low and high extinction fraction for bifurcating and budding speciation. Each point represents a species and both ages, true and phylogenetic, are scaled to the root age of the correspondent phylogenetic tree.



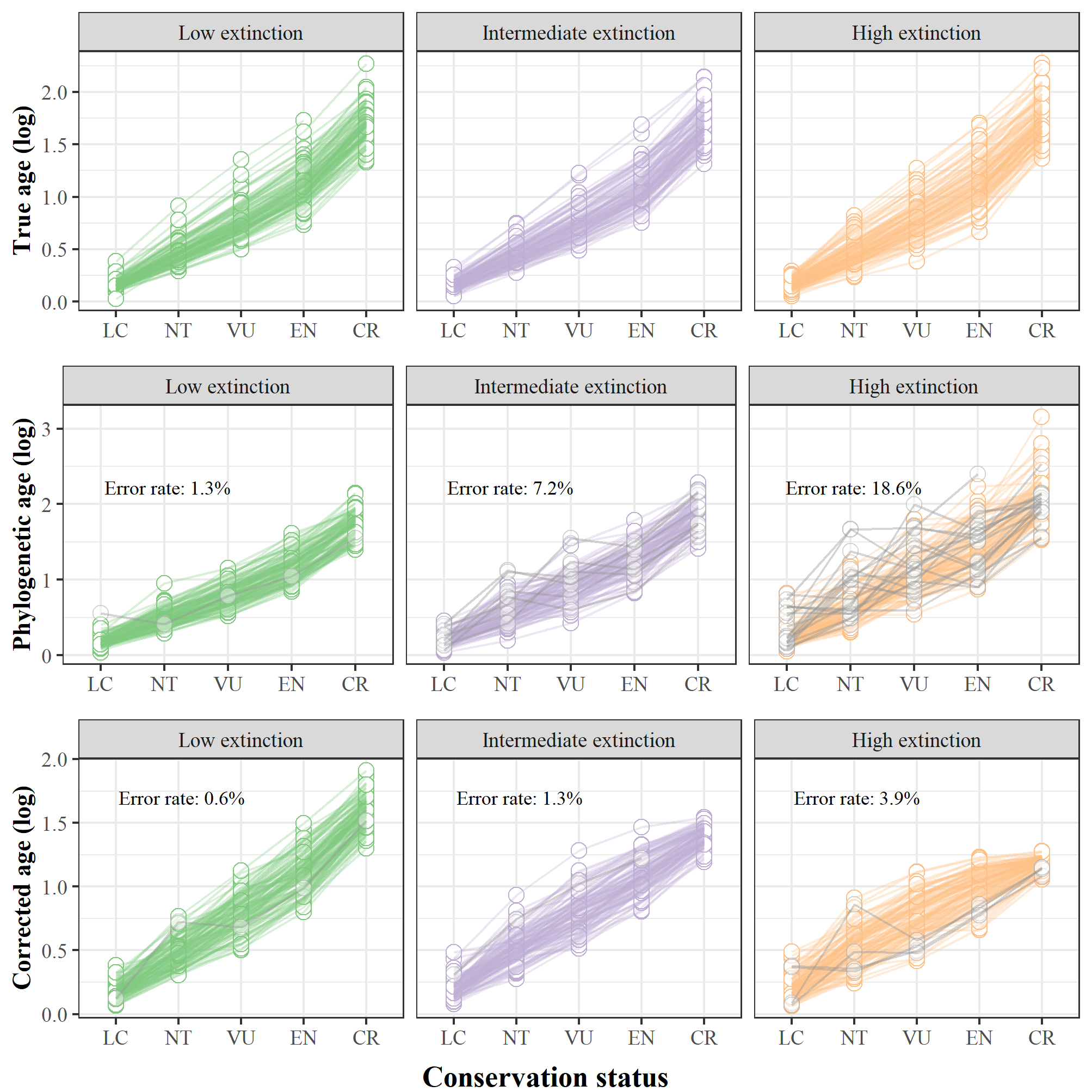
**Fig. 4.** Error in equating phylogenetic age with speciation age. The error was quantified as mean absolute percentage error (MAPE) between the true and phylogenetic ages across all species for each tree simulated under bifurcating and budding speciation. Each dot represents one replicate of the 300 trees for each speciation mode using different rates of speciation and extinction fraction.



**Fig. 5.** Effect of incomplete taxon sampling on error in species ages. Error in equating the phylogenetic age with true species age for the fully sampled phylogeny, and 25% and 50% of missing extant species, in an intermediate extinction scenario. The error was quantified as mean absolute percentage error (MAPE) between the true phylogenetic ages across 100 species for each of 1000 trees for each missing species scenario simulated under bifurcating and budding speciation.



**Fig. 6.** Performance of the probabilistic age estimator. ΔMAPE for the three extinction scenarios (low, intermediate, and high; from left to right) was quantified as the difference of the mean absolute percentage error (MAPE) of the probabilistic estimator point estimates (mean and median) and the MAPE of the phylogenetic age. The MAPE was quantified as the difference between the true and point estimates (mean and median) or phylogenetic ages across 100 species across 1000 trees for each extinction scenario simulated under bifurcating speciation. The red dashed line represents no difference between the compared MAPEs, negative ΔMAPE values indicate an improvement in the accuracy of the probabilistic estimator over the phylogenetic age.



**Fig. 7.** Power to recover an age extinction-risk relationship. Simulated species ages under three extinction scenarios and assuming bifurcating speciation were binned into conservation status categories, which represents an increase in extinction risk by age (LC = Least Concern; NT = Near Threatened; VU = Vulnerable; EN = Endangered; CR = Critically Endangered). We used the phylogenetic age and the mean age obtained from our probabilistic corrective function to calculate the mean age per conservation status category and assess if every mean age increases over the one of the previous category with lower extinction risk. The error rate is the percentage across all 1000 phylogenies where the relationship between the mean ages and the conservation status categories is not always increasing (shown by grey lines).

**References**

Alroy, J., Marshall, C. R., Bambach, R. K., Bezusko, K., Foote, M., Fürsich, et al., (2001). Effects of sampling standardization on estimates of phanerozoic marine diversification. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(11), 6261–6266. doi: 10.1073/pnas.111144698

Anacker, B. L., & Strauss, S. Y. (2014). The geography and ecology of plant speciation: Range overlap and niche divergence in sister species. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1778). doi: 10.1098/rspb.2013.2980

Aze, T., Ezard, T. H. G., Purvis, A., Coxall, H. K., Stewart, D. R. M., Wade, B. S., & Pearson, P. N. (2011). A phylogeny of Cenozoic macroperforate planktonic foraminifera from fossil data. *Biological Reviews*, *86*(4), 900–927. doi: 10.1111/j.1469-185X.2011.00178.x

Balmford, A. (1996). Extinction filters and current resilience: The significance of past selection pressures for conservation biology. *Trends in Ecology and Evolution*. doi: 10.1016/0169-5347(96)10026-4

Barido-Sottani, J., Pett, W., O’Reilly, J. E., & Warnock, R. C. M. (2019). FossilSim: An r package for simulating fossil occurrence data under mechanistic models of preservation and recovery. *Methods in Ecology and Evolution*, *10*(6), 835–840. doi: 10.1111/2041-210X.13170

Barraclough, T. G., & Vogler, A. P. (2000). Detecting the geographical pattern of speciation from species-level phylogenies. *The American Naturalist*, *155*(4), 419–434. doi: 10.1086/303332

Barraclough, T. G., Vogler, A. P., & Harvey, P. H. (1998). Revealing the factors that promote speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *353*(1366), 241–249. doi: 10.1098/rstb.1998.0206

Baum, D. A., Smith, S. D. W., & Donovan, S. S. S. (2005). The tree-thinking challenge. *Science*, *310*(5750), 979–980. doi: 10.1126/science.1117727

Beaulieu, J. M., & O’Meara, B. C. (2016). Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Systematic Biology*, *65*(4), 583–601. doi: 10.1093/sysbio/syw022

Benton, M. J. (2013). Origins of biodiversity. *Palaeontology*. doi: 10.1111/pala.12012

Benton, M. J. (2016). Origins of Biodiversity. *PLoS Biology*, *14*(11), 1–7. doi: 10.1371/journal.pbio.2000724

Brée, B., Condamine, F. L., & Guinot, G. (2022). Combining palaeontological and neontological data shows a delayed diversification burst of carcharhiniform sharks likely mediated by environmental change. *Scientific Reports*, *12*(1), 21906. doi: 10.1038/s41598-022-26010-7

Caetano, D. S., & Quental, T. B. (2022). How important is budding speciaiton for comparative studies? *BioRxiv*. doi: 10.1101/2022.05.24.493296

Caley, M. J., Fisher, R., & Mengersen, K. (2014). Global species richness estimates have not converged. Trends in Ecology & Evolution, 29(4), 187-188. doi:10.1016/j.tree.2014.02.002

Callaway, E. (2017). Oldest Homo sapiens fossil claim rewrites our species’ history. *Nature*, 546, 289-293. doi: 10.1038/nature.2017.22114

Carrillo, J. D., Faurby, S., Silvestro, D., Zizka, A., Jaramillo, C., Bacon, C. D., & Antonelli, A. (2020). Disproportionate extinction of South American mammals drove the asymmetry of the Great American Biotic Interchange. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(42), 26281–26287. doi: 10.1073/pnas.2009397117

Carrillo, J. D., Forasiepi, A., Jaramillo, C., Sánchez-villagra, M. R., & Richardson, J. E. (2015). *Neotropical mammal diversity and the Great American Biotic Interchange : spatial and temporal variation in South America ’ s fossil record*. *5*(January), 1–11. doi: 10.3389/fgene.2014.00451

Chang, J., Rabosky, D. L., & Alfaro, M. E. (2020). Estimating Diversification Rates on Incompletely Sampled Phylogenies: Theoretical Concerns and Practical Solutions. *Systematic Biology*, *69*(3), 602–611. doi: 10.1093/sysbio/syz081

Davies, T. J., Smith, G. F., Bellstedt, D. U., Boatwright, J. S., Bytebier, B., Cowling, R. M., et al., (2011). Extinction risk and diversification are linked in a plant biodiversity hotspot. *PLoS Biology*. 9(5), e1000620. doi: 10.1371/journal.pbio.1000620

Diniz Filho, J. A. F., Jardim, L., Guedes, J. J., Meyer, L., Stropp, J., Frateles, L. E. F., ... & Hortal, J. (2023). Macroecological links between the Linnean, Wallacean, and Darwinian shortfalls. *Frontiers of Biogeography*, 15(2). doi: 10.21425/F5FBG59566

Eldredge, N., Thompson, J. N., Brakefield, P. M., Gavrilets, S., Jablonski, D., Jackson, J. B. C., et al., (2005). The dynamics of evolutionary stasis. *Paleobiology* 31(S2), 133-145. doi: 10.1666/0094-8373(2005)031

Emerson, B. C., & Patiño, J. (2018). Anagenesis, Cladogenesis, and Speciation on Islands. *Trends in Ecology and Evolution*, *33*(7), 488–491. doi: 10.1016/j.tree.2018.04.006

Fitzpatrick, B. M., & Turelli, M. (2006). The Geography of Mammalian Speciation: Mixed Signals From Phylogenies and Range Maps. *Evolution*, *60*(3), 601–615. doi: 10.1111/j.0014-3820.2006.tb01140.x

Fjeldså, J., Bowie, R. C. K., & Rahbek, C. (2012). The role of mountain ranges in the diversification of birds. *Annual Review of Ecology, Evolution, and Systematics*, *43*, 249–265. doi: 10.1146/annurev-ecolsys-102710-145113

Foote, M. (1996). *On the Probability of Ancestors in the Fossil Record*. *22*(2), 141–151. doi: 10.1017/S0094837300016146

Foote, M., & Raup, D. M. (1996). *Fossil Preservation and the Stratigraphic Ranges of Taxa*. *22*(2), 121–140. doi: 10.1017/S0094837300016134

Gaston, K. J., & Blackburn, T. M. (1997). Evolutionary age and risk of extinction in the global avifauna. *Evolutionary Ecology,* 11, 557–565. doi: 10.1007/s10682-997-1511-4

Gernhard, T. (2008). The conditioned reconstructed process. *Journal of theoretical biology, 253*(4), 769-778. doi: 10.1016/j.jtbi.2008.04.005

Harvey, P. H., May, R. M., & Nee, S. (1994). Phylogenies without fossils. *Evolution*, *48*(3), 523–529. doi: 10.1111/j.1558-5646.1994.tb01341.x

Heath, T. A., Hedtke, S. M., & Hillis, D. M. (2008). Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution*, *46*(3), 239–257. doi: 10.3724/SP.J.1002.2008.08016

Henning, W. (1999). *Phylogenetic systematics*. University of Illinois Press.

Hopkins, M. J. (2007). Modelling the known and unknown plant biodiversity of the Amazon Basin. *Journal of Biogeography*, 34(8), 1400-1411. doi: 10.1111/j.1365-2699.2007.01737.x

Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 46, 523-549. doi: 10.1146/annurev-ecolsys-112414-054400

IUCN. (2016). IUCN Red List of Threatened Species. Version 2016-2. *Fourth Quarter*.

Johnson, C. N., Delean, S., & Balmford, A. (2002). *Phylogeny and the selectivity of extinction in Australian marsupials*. 135–142. doi: 10.1017/S1367943002002196

Kendall, M. G. (1946). *The advanced theory of statistics* (2nd ed.). Charles Griffin adn Co., Ltd., London.

López-Martínez, A. M., Schonenberger, J., von Balthazar, M., González-Martínez, C. A., Ramírez-Barahona, S., Sauquet, H., & Magallón, S. (2023). Integrating Fossil Flowers into the Angiosperm Phylogeny Using Molecular and Morphological Evidence. *Systematic Biology*, *syad017*. doi: 10.1093/sysbio/syad017

Losos, J. B., & Glor, R. E. (2003). Phylogenetic comparative methods and the geography of speciation. *Trends in Ecology and Evolution*, *18*(5), 220–227. doi: 10.1016/S0169-5347(03)00037-5

Meier, R., & Willmann, R. (2000). A defense of the Hennigian Species Concept. In *Species concepts and phylogenetic theory: a debate* (p. 167).

Meiri, S., Raia, P., & Santos, A. M. (2018). Anagenesis and cladogenesis are useful island biogeography terms. *Trends in Ecology & Evolution, 33*(12), 895-896. doi: 10.1016/j.tree.2018.09.005

Mynard, P., Algar, A., Lancaster, L., Bocedi, G., Fahri, F., Gubry-Rangin, C., et al., (2023). Impact of Phylogenetic Tree Completeness and Misspecification of Sampling Fractions on Trait Dependent Diversification Models. *Systematic Biology*, *72*(1), 106–119. doi: 10.1093/sysbio/syad001

Nee, S., May, R. M., & Harvey, P. H. (1994). The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 344*(1309), 305-311. doi: 10.1098/rstb.1994.0068

Nee, S., & May, R. M. (1997). Extinction and the Loss of Evolutionary History. *Science*, *278*(5338), 692–694. doi: 10.1126/science.278.5338.692

Pearson, P. N. (1995). Investigating age-dependency of species extinction rates using dynamic survivorship analysis. *Historical Biology*, 10(2), 119–136. doi: 10.1080/10292389509380516

Pie, M. R., & Caron, F. (2023). Substantial variation in species ages among vertebrate clades. *BioRxiv*, *06*. doi: 10.1101/2023.06.08.544238

Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T. M., Gittleman, J. L., Joppa, L. N., … Sexton, J. O. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, *344*(6187). doi: 10.1126/science.1246752

R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: *R Foundation for Statistical Computing*. Retrieved from https://www.r-project.org/

Rabosky, D. L. (2010). Extinction rates should not be estimated from molecular phylogenies. *Evolution*, 64(6), 1816–1824. doi: 10.1111/j.1558-5646.2009.00926.x

Riddle, B. R., Ladle, R. J., Lourie, S. A., & Whittaker, R. J. (2011). Basic biogeography: estimating biodiversity and mapping nature. *Conservation biogeography*, 45-92. doi: 10.1002/9781444390001.ch4

Rivas-Gonzáles, I., Rousselle, M., Li, F., Zhou, L., Dutheil, J., Munch, K., et al., (2023). Pervasive incomplete lineage sorting illuminates speciation and selection in primates. *Science*, *380*(6648), eabn4409. doi: 10.1126/science.abn4409

Roopnarine, P. D., Byars, G., Fitzgerald, P., Paleobiology, S., & Winter, N. (1999). Anagenetic evolution, stratophenetic patterns, and random walk models. *Paleobiology,* 25(1), 41–57. doi: 10.1666/0094-8373(1999)0252.3.CO;2

Rosenblum, E. B., Sarver, B. A. J., Brown, J. W., Des Roches, S., Hardwick, K. M., Hether, T. D., et al., (2012). Goldilocks Meets Santa Rosalia: An Ephemeral Speciation Model Explains Patterns of Diversification Across Time Scales. *Evolutionary Biology*, *39*(2), 255–261. doi: 10.1007/s11692-012-9171-x

Silvestro, D., Salamin, N., & Schnitzler, J. (2014). PyRate: a new program to estimate speciation and extinction rates from incomplete fossil data. *Methods in Ecology and Evolution2*, *5*(10), 1126–1131. doi: 10.1111/2041-210X.12263

Silvestro, D., Castiglione, S., Mondanaro, A., Serio, C., Melchionna, M., Piras, P., et al., (2020). A 450 million years long latitudinal gradient in age-dependent extinction. *Ecology Letters*, *23*(3), 439–446. doi: 10.1111/ele.13441

Silvestro, D., Salamin, N., Antonelli, A., & Meyer, X. (2019). Improved estimation of macroevolutionary rates from fossil data using a Bayesian framework. *Paleobiology*, *45*(4), 546–570. doi: 10.1017/pab.2019.23

Silvestro, D., Warnock, R. C. M., Gavryushkina, A., & Stadler, T. (2018). Closing the gap between palaeontological and neontological speciation and extinction rate estimates. *Nature Communications*, *9*(1). doi: 10.1038/s41467-018-07622-y

Simpson, George Gaylord. (1961). *Principles of animal taxonomy*. Columbia University Press.

Simpson, George Gaylord. (1984). *Tempo and mode in evolution*. Columbia University Press.

Simpson, Geroge G. (1951). The species concept. *Evolution*, *5*(4), 285–298.

Skeels, A., & Cardillo, M. (2019). Reconstructing the geography of speciation from contemporary biodiversity data. *American Naturalist*, *193*(2), 240–255. doi: 10.1086/701125

Slater, G. J., & Harmon, L. J. (2013). Unifying fossils and phylogenies for comparative analyses of diversification and trait evolution. *Methods in Ecology and Evolution*, 4(8), 699-702. doi: 10.1111/2041-210X.12091

Sonne, J., Dalsgaard, B., Borregaard, M. K., Kennedy, J., Fjeldså, J., & Rahbek, C. (2022). Biodiversity cradles and museums segregating within hotspots of endemism. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1981), 20221102. doi: 10.1098/rspb.2022.1102

Stadler, T. (2013). Recovering speciation and extinction dynamics based on phylogenies. *Journal of evolutionary biology, 26*(6), 1203-1219. doi: 10.1111/jeb.12139

Stadler T (2019). TreeSim: Simulating Phylogenetic Trees. R package version

2.4, https://CRAN.R-project.org/package=TreeSim

Swenson, N. G. (2019). *Phylogenetic ecology: A history, critique, and remodeling*. University of Chicago Press.

Tanentzap, A. J., Brandt, A. J., Smissen, R. D., Heenan, P. B., Fukami, T., & Lee, W. G. (2015). When do plant radiations influence community assembly? The importance of historical contingency in the race for niche space. *New Phytologist*, *207*(2), 468–479. doi: 10.1111/nph.13362

Tanentzap, A. J., Igea, J., Johnston, M. G., & Larcombe, M. J. (2020). Does evolutionary history correlate with contemporary extinction risk by influencing range size dynamics? *American Naturalist*, 195(3), 569–576. doi: 10.1086/707207

Thomas, G. H., Hartmann, K., Jetz, W., Joy, J. B., Mimoto, A., & Mooers, A. O. (2013). PASTIS: An R package to facilitate phylogenetic assembly with soft taxonomic inferences. *Methods in Ecology and Evolution*, *4*(11), 1011–1017. doi: 10.1111/2041-210X.12117

Upham, N. S., Esselstyn, J. A., & Jetz, W. (2019). Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. In *PLoS Biology*, 17(12), e3000494. doi: 10.1371/journal.pbio.3000494

Van Valen, L. (1973). A new evolutionary law. *Evolutionary Theory*. 1, 1–30.

Vaux, F., Trewick, S. A., & Morgan-Richards, M. (2016). Lineages, splits and divergence challenge whether the terms anagenesis and cladogenesis are necessary. Biological Journal of the Linnean Society, 117(2), 165-176. doi: 10.1111/bij.12665

Verde Arregoitia, L. D., Blomberg, S. P., & Fisher, D. O. (2013). Phylogenetic correlates of extinction risk in mammals: Species in older lineages are not at greater risk. *Proceedings of the Royal Society B: Biological Sciences* 280(1765), 20131092. doi: 10.1098/rspb.2013.1092

Vilela, B., Villalobos, F., Rodríguez, M. Á., & Terribile, L. C. (2014). Body size, extinction risk and knowledge bias in New World snakes. PloS one, 9(11), e113429. doi: 10.1371/journal.pone.0113429

Wagner, P. J., Erwin, D. H., & Anstey, R. L. (1995). Phylogenetic patterns as tests of speciation models. In *New approaches to speciaiton in the fossil record* (pp. 87–122). New York: Columbia University Press.

Willmann, R., & Meier, R. (2000). A critique from the Hennigian species concept perspective. In *Species concepts and phylogenetic theory: a debate* (pp. 101–118). New York: Columbia University Press.