**Discussion**

The use of phylogenetic branch lengths as an approximation of species ages or evolutionary ages is becoming central to an increasing number of studies that use them to explain macro-ecological and evolutionary patterns and current extinction risks [REFs]. Here we showed that this approximation leads to substantial errors and that its 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 a bifurcating speciation process in the absence of extinction and with all living species included in the phylogenetic tree. 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 high (Pimm et al., 2014) [REF], and there is substantial evidence that many living species remain unknown to science and are therefore absent from empirical phylogenetic trees [REF]. Thus, the scenario under which phylogenetic age correctly predicts species age is a very unlikely instance.

Some authors acknowledged 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. Yet, the magnitude of the error associated with a literal reading of phylogenetic branches as species ages remains under-appreciated as shown by the many studies implementing this approach.

We showed that the largest error in in approximating species ages with phylogenetic ages is linked with incomplete sampling of extant species. This is a problem that in principle can be solved by extending the scope of the sampling to include all species in the phylogenetic inference. Yet, despite the advances in the scalability of DNA sequencing, this remains impractical for large clades, including some of the best sampled ones such as vertebrate groups, in which many species still lack genetic data [REF]. In addition, a substantial proportion of species might be missing from the phylogenetic trees because they are still unknown to science, a problem often identified as the Linnean shortfall (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]. (Caley et al. 2014) (Hopkins 2007; Vilela et al. 2014 (Riddle et al. 2011). Thus, this

Our simulations showed that under some scenarios the extrapolation of accurate species ages is essentially impossible from phylogenetic trees. Under the assumption of budding speciation, the error is high even without extinction and with complete sampling, because phylogenetic trees are agnostic about parent and descendant species following a branching event (Fig. 4). Phylogenetic ages are, by construction, identical for sister species, which is necessarily wrong within a budding speciation scenario. Similarly, anagenetic speciation also leads to high error, 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). Thus, species age is virtually unidentifiable under the assumption of speciation modes that deviate from a strictly bifurcating scenario.

The lowest error in species age estimation was observed under scenarios of bifurcating speciation. This is the implicit assumption of most studies using approximations of species ages [REF], even though it is at odds with the assumption of all birth-death models used in the molecular clock analyses that estimate the phylogenetic trees in the first place [REF]. Despite the lower error, our simulations showed that both extinction and missing lineages can lead to a substantially decrease in accuracy (Fig. 4-5, SM3-4) that can even lead to qualitative misinterpretations of general patterns such as age-dependent extinction risks (Fig. 7). Given the large inaccuracy in phylogenetic age, 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).

We found our probabilistic approach to efficiently reduce the biases associated with extinction. It substantially improved the accuracy of the estimation of species ages, leading to much lower error rates even under scenarios of high extinction fractions. The performance of our estimator is however contingent on the ability of birth-death models to correctly estimate speciation and extinction rates from phylogenies of extant species. The accuracy of these methods has been shown to be high under several simulation settings [REF] but their accuracy has been questioned for empirical datasets and under complex models with rate variation [REFS]. One commonly observed pattern is the estimation of 0-extinction rate from empirical phylogeny [REF], which would make corrected ages identical to the phylogenetic ages. Yet, the fossil record decisively shows that extinction and speciation rates vary within the same order of magnitude across virtually all clades [REF]. The robustness of estimated extinction rates can be increased through the use of fossil data in the analyses [REF]. Alternatively, our probabilistic approach could be applied across a range of plausible values of speciation and extinction rates as a way to evaluate the robustness of conclusions drawn from the patterns of estimated species ages.

**Conclusion**

This study provides a quantification of the deviations between true species age and phylogenetic age expected due to incomplete taxon sampling, extinction, and unknown speciation modes. We found that phylogenetic age is a good proxy of species age only in a rather unlikely case in which 1) all species in a clade are known to science and included in the phylogenetic tree, 2) speciation occurs as a strictly bifurcating process, and 3) extinction rates are zero of orders of magnitudes lower than speciation rates. Using simulations, we identified that incomplete taxon sampling and budding and anagenetic speciation modes cause the highest mismatch between phylogenetic age and true species age and might not be corrected for in the absence of additional information for instance derived from fossil data. In contrast, under a scenario of bifurcating speciation and full sampling, we proposed a probabilistic approach based on the properties of the birth-death process that can drastically improve the accuracy of the estimated species ages by reducing the bias linked with extinction. We note that, even in this case, the robustness of such estimates will be contingent on the accuracy of the estimated speciation and extinction rates and of course of the underlying phylogenetic tree and dating of the branching times. In the light of our findings, we caution against a literal reading of branch lengths as proxies for species ages or evolutionary ages. We recommend the application of a correction of phylogenetic ages to account for extinction and an explicit acknowledgement of the assumptions made about the implied speciation modes. We hope our results will stimulate a discussion about the use of phylogenetic trees in inferring species age and help to a critical evaluation of the robustness of inferences linking species age with traits, ecological variables, and extinction risks.

.