Chronospaces: an R package for the statistical exploration of divergence times promotes the assessment of methodological sensitivity

ABSTRACT

[1] Much of our understanding of the history of life hinges upon time calibration, the process of assigning absolute times to cladogenetic events. Bayesian approaches to time scaling phylogenetic trees have dramatically grown in complexity, and depend today upon numerous methodological choices. Arriving at objective justifications for all of these is difficult and time consuming. Thus, divergence times are routinely inferred under only one or a handful of parametric conditions, often times chosen arbitrarily. Progress towards building robust biological timescales necessitate the development of better methods to visualize and quantify the sensitivity of results to these decisions.

[2] Here, we present an R package that assists in this endeavor through the use of chronospaces, i.e., graphical representations summarizing variation in the node ages contained in time-calibrated trees. We further test this approach by estimating divergence times for three empirical datasets—spanning widely differing evolutionary timeframes—using the software PhyloBayes.

[3] Our results reveal large differences in the impact of many common methodological decisions, with the choice of clock (uncorrelated vs. autocorrelated) and loci having strong effects on inferred ages. Other decisions have comparatively minor consequences, including the use of the computationally intensive site-heterogeneous model CAT-GTR, whose effect might only be discernible for exceedingly old divergences (e.g., the deepest eukaryote nodes).

[4] The package *chronospace* implements a range of graphical and analytical tools that assist in the exploration of sensitivity and the prioritization of computational resources in the inference of divergence times.

Keywords: chronospaces, divergence times, diversification, fossils, phylogenetic inference, phylogenomics, sensitivity analysis, time calibration

INTRODUCTION

The discovery of the "molecular evolutionary clock" by Zuckerkandl and Pauling (1965), who first described the relative constancy in the rates of change of molecular sequences both through time and across clades, is a hallmark of the advent of molecular biology. By establishing a proportionality between the amount of amino acid or nucleotide differences that can be directly observed and the time that must have elapsed since those sequences last shared a common ancestor, the molecular clock greatly expands the insights that can be drawn from phylogenetic trees. As such, scaling phylogenies to absolute (or geologic) time is often the first step to explore macroevolutionary patterns and processes at deep timescales, and one on which modern comparative biology relies on (Pennell & Harmon 2013).

Bayesian methods of divergence time estimation have become highly sophisticated (Kumar & Hedges 2016), being able to accommodate among-lineage rate variation through the use of relaxed clocks (Thorne *et al.* 1998; Drummond *et al.* 2006), exploit diverse sources of both character and temporal information (Ronquist *et al.* 2012; O'Reilly *et al.* 2015), and use complex tree-building models that account for fossilization and sampling rates (Heath *et al.* 2014) and accommodate gene tree

42 discordances (Rannala *et al.* 2020). While this increased complexity has allowed for an improved
43 modelling of diversification, it has also drastically increased the number of methodological choices that
44 the average scientist is confronted with. Dating software require users to specify which among several
45 competing models of diversification, trait evolution, and among-lineage rate variation—the three distinct
46 elements of the tripartite framework of Bayesian dating (Warnock & Wright 2020)—are enforced. Each
47 of these models includes a sizable number of parameters for which prior probability distributions need
48 to be specified. Furthermore, even though progress has been made regarding computational efficiency
49 (Dos Reis & Yang 2019; Douglas *et al.* 2022; Barba-Montoya *et al.* 2023), most methods remain limited in
50 terms of the size of the datasets that can be run within reasonable times, forcing further decisions
51 regarding taxon sampling and locus choice. The choice of number, location, and shape of fossil-derived
52 temporal constraints adds yet another layer of complexity to these analyses. These methodological
53 decisions can have a sizable effect on the inference of Bayesian time-calibrated trees (Warnock *et al.*
54 2011; Magallón *et al.* 2013; Eme *et al.* 2014; Mongiardino Koch *et al.* 2022; Sauquet *et al.* 2022)—also
55 known as chronograms—producing potentially incongruent reconstructions of biological diversification.

56 Two broad approaches have been explored to overcome this methodological uncertainty. First,
57 numerous methods have been developed to justify model selection, allowing the identification of a set
58 of parametric conditions considered optimal. These methods, however, have their own limitations:
59 model fit statistics do not reliably discriminate between partition and mixture models (Crotty & Holland
60 2022), while alternative tree priors might not be amenable to comparison using standard methods such
61 as Bayes factors (May & Rothfels 2023). General purpose approaches, such as cross-validation, are too
62 computationally consuming to be used exhaustively (Zhou *et al.* 2007; Lartillot 2023), while methods
63 designed to tackle specific aspects of the analysis can arrive at opposite conclusions depending on
64 unassessed decisions (e.g., detecting autocorrelated rates can depend on which model of substitution is
65 used to infer the tree being evaluated; Mongiardino Koch *et al.* 2022). The alternative to the use of

model selection algorithms involves sensitivity analyses, which explore the robustness of results to a range of conditions (e.g., Sanderson & Doyle 2001; Britton *et al.* 2007; Warnock *et al.* 2011; Howard *et al.* 2022; Mongiardino Koch *et al.* 2022). However, there is a dearth of tools that provide efficient ways of summarizing multiple chronograms in either quantitative or visual ways, as well as a lack of criteria for gauging the relative impact that different parameters have on the final results.

Here, we present *chronospace*, an R package (R Core Team 2022) for the statistical exploration of time calibration. The package implements a set of visualization techniques that allow for a better assessment of the sensitivity of divergence times to methodological decisions. It also provides new tools to quantify the overall impact of each individual methodological choice, providing an estimate of its effect size. We illustrate its functionalities by inferring divergence times for three empirical datasets spanning widely differing evolutionary timescales. We do so using the program PhyloBayes v.4.1 (Lartillot *et al.* 2013) which implements the complex site-heterogeneous model CAT-GTR (Lartillot & Philippe 2004).

MATERIALS AND METHODS

A) Chronospaces: the approach

Chronospaces are low-dimensional graphical representations that summarize variation in the absolute node ages of populations of chronograms (Mongiardino Koch *et al.* 2022). They are similar to treespaces (Hillis *et al.* 2005) in that they provide a visual summary of populations of phylogenetic trees, but focus exclusively on temporal information. To do so, the chronograms analyzed must share the same set of nodes, representing posterior distributions of Bayesian time-calibrated analyses run under a fixed topology. While this limits their application, a two-step approach in which a preferred topology is first

inferred and then time scaled has become common in the phylogenomic literature due to the computational burden of simultaneous inference.

Whereas treespaces are generated from distance matrices that capture topological similarity (Smith 2022)—and are thus representations that are disconnected from the original variables—chronospaces are built using ordination methods that reorganize variation, capturing major patterns in a few synthetic axes. Here, we rely on between-group principal component analysis (bgPCA; Dolédec & Chessel 1987; Dolédec & Chessel 1989; Yendle & MacFie 1989), an alternative to traditional PCA that ordinates the covariance matrix of multivariate means for predefined groups of observations. Within the context of a chronospace, posterior topologies constitute the 'observations', and the ages of each of their internal nodes are the 'variables' (see also Fig. 1). When analyzing populations of topologies inferred under different analytical conditions (e.g., alternative clock models), these constitute distinct levels of a grouping variable. The estimated bgPCA axes will be those capturing the largest variation in group (i.e., level) means, and onto which individual chronograms are projected. The result is an ordination that maximizes the separation between groups through a rigid rotation of space, preserving information regarding which nodes are more variable in age, and which factors this variation is associated with. The approach will result in $g - 1$ bgPCA axes, where g is the number of groups, with the remainder of multidimensional structure occurring on residual axes. Within the biological literature, bgPCA has been mostly used to explore the discrimination of groups of observations using microarray (Culhane *et al.* 2002) and geometric morphometric (Cardini *et al.* 2019) data. While bgPCA can introduce spurious patterns when the number of variables exceeds the number of observations (Cardini *et al.* 2019), this does not represent an issue within the context of chronospaces as the number of nodes in a phylogenetic tree (i.e., the number of variables) will typically be much lower than the number of trees sampled in a Bayesian posterior distribution (i.e., the number of observations).

bgPCA axes capture the main directions of change in divergence times that are associated with a given methodological decision. If the chronograms being analyzed are derived from a set of inferences varying multiple parametric conditions, separate bgPCA rotations can be applied to the same chronospace, deriving a set of eigenvectors and eigenvalues for each one of the factors being varied. The proportion of the total variance accounted for by each factor, as well as the proportion that remains unaccounted, is estimated both for the entire chronospace as well as for each individual bgPCA axis by calculating the corresponding sums of squares for the balanced model. Chronospaces thus provide a means of summarizing the entire temporal information of individual chronograms, and exploring the effects (i.e., direction and magnitude of change) associated with alternative dating approaches. This information can be used to aid in the interpretation of the overall sensitivity of divergence times, as well as rank methodological choices from the most to the least impactful.

B) *chronospace:* the package

chronospace is an R package (R Core Team 2022) available for installation from the GitHub repository (removed for anonymity). It relies on functions from phylogenetic packages *ape* (Paradis & Schliep 2018), *phangorn* (Schliep 2011), and *phytools* (Revell 2012), visualization packages *deeptime* (Gearty 2022), *ggpubr* (Kassambara 2020), *ggtree* (Yu *et al.* 2017), and *patchwork* (Pedersen 2020), as well as data handling functionalities from the *tidyverse* (Wickham 2017).

The range of functions and capabilities implemented in the package are depicted in Figure 1 and further documented on the GitHub repository. The typical workflow would begin with the use of function `extract_ages` which gathers node ages from posterior distributions of Bayesian timetrees inferred under a constrained topology. Typically, this would include trees from multiple runs (as most of

133 the methods implemented are designed to facilitate the comparison of results obtained under different
134 analytical conditions), but some analyses can be done with a single set of trees. Chronograms are read
135 from files in Newick format (one per run and/or chain). While not all phylogenetic software generate this
136 type of output, trees can be first converted and saved to file using functions from other R packages, such
137 as *treeio* (Wang *et al.* 2020). Function `extract_ages` can then import the trees, extract and format
138 the node ages contained in them, and annotate the data with a set of factors representing the
139 parameters changed between different runs. The constrained tree topology is also saved in order to
140 generate graphical output. This information (ages, factors, and topology) is stored as an object of class
141 `nodeAges`. Multiple graphical outputs can be directly obtained from this object: posterior distributions
142 of node ages can be plotted using both `sensitive_nodes`, which helps identify and visualize the
143 nodes whose inferred ages change the most between analyses, or `specified_node`, which plots ages
144 for a target node defined by the user; lineage-through-time plots can also be generated using function
145 `ltt_sensitivity` to depict the overall effect of methodological choices on the inferred history of
146 diversification (including different options for portraying uncertainty). Finally, bgPCA-rotated
147 chronospaces are generated using function `chronospace`, which creates an object of the same class,
148 and its associated printing and plotting functions summarize the results of the analysis in both graphical
149 and quantitative ways. The remainder of this article focuses on highlighting the use of the latter.

150

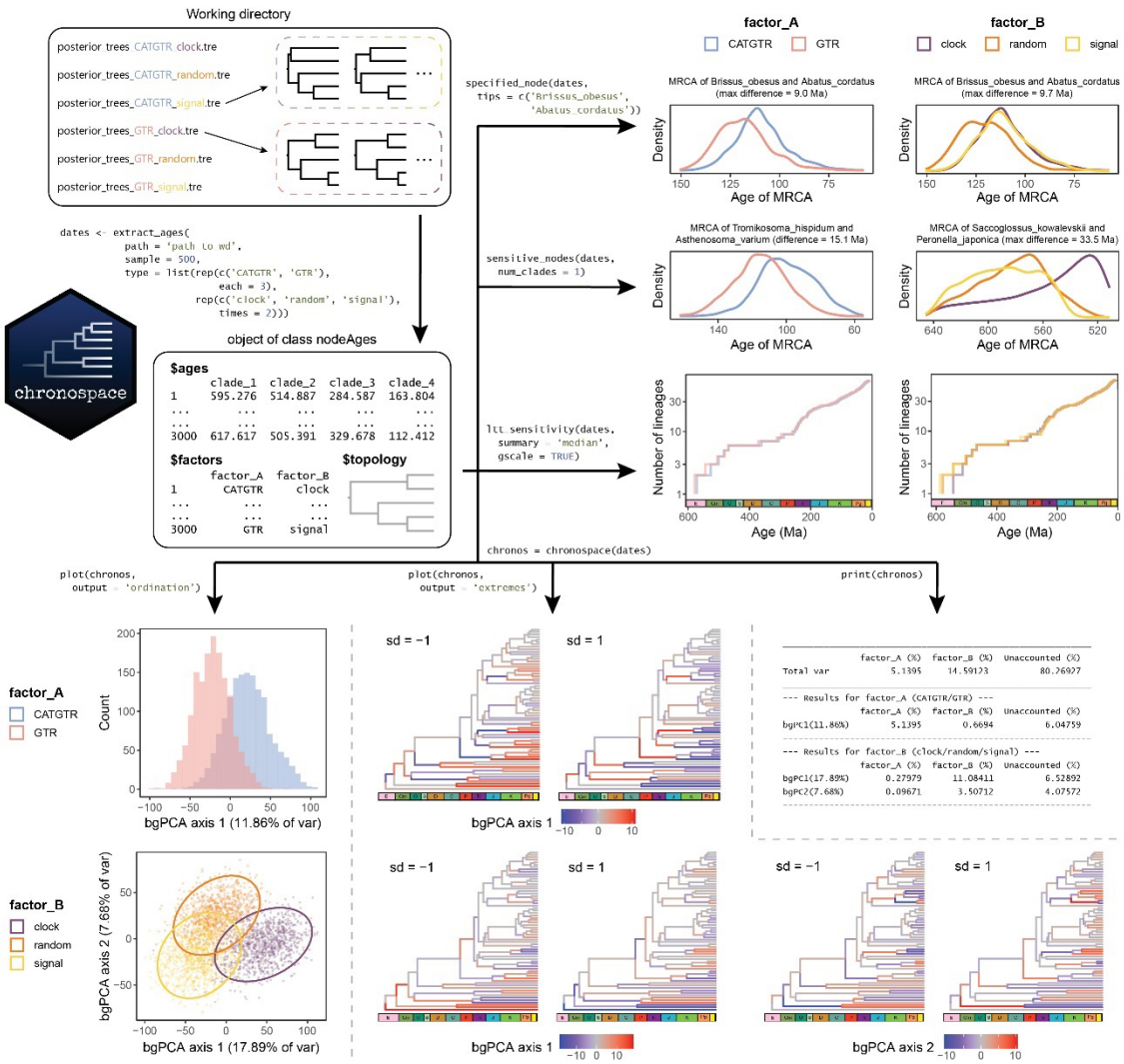


Figure 1: Structure of the R package *chronospace* and analytical workflow, represented using the data that accompanies the package. Populations of posterior topologies from time calibration analyses present in the working directory are imported, and their node ages annotated and stored as an object of class `nodeAges` using function `extract_ages`. The package provides functions to visualize the effect induced by analytical conditions on a node of interest (`specified_node`), and on the overall history of diversification (`ltt_sensitivity`). It also detects the nodes whose ages change the most between analyses, and plots their posterior distributions (`sensitive_nodes`). These plots can optionally include a chronostratigraphic scale (shown here in the lineage-through-time plot). Finally, a

160 `chronospace` object can be built and used to summarize the overall effect of methodological
161 decisions on the results. These can be visualized as bgPCA-summarized chronospaces (which will be
162 univariate or multivariate, depending on the number of levels of each factor) and/or chronogram
163 warping plots. The latter depict the patterns of branch length contraction/expansion that are captured
164 by each bgPCA axis, and are akin to the landmark deformation plots that typically accompany the
165 morphospace axes in a geometric morphometric analysis (hence, their name). All plots are generated
166 with *ggplot2* (Wickham 2016) and several extensions, providing the user with control over basic
167 visualization aspects (object colors, sizes, transparencies, etc.). Chronospace summaries can also be
168 printed in the form of a table reporting the partitioning of variance.

169

170 **C) Sensitivity of divergence times across three empirical datasets**

171 Three genome-scale datasets of protein-coding loci were obtained from the literature, designed
172 to tackle the phylogenetic relationships and divergence times of Curculionoidea (65 terminals, 517 loci;
173 Shin *et al.* 2018), Decapoda (90 terminals, 410 loci; Wolfe *et al.* 2019), and Eukaryota (136 terminals, 313
174 loci; Strassert *et al.* 2021). Studies were selected so as to span widely differing evolutionary time frames
175 (Figures 2). For each of these, we gathered a phylogenomic dataset (including a concatenated matrix
176 coded as amino acids and an associated partition file), the rooted tree topology preferred by the
177 authors, and a list of node age calibrations derived from the fossil record. If studies included multiple
178 outgroups, only those forming part of the sister-group clade to the ingroup were retained. The eukaryote
179 dataset lacked outgroups altogether.

180

181

182

183 Divergence times were inferred using PhyloBayes v.4.1 (Lartillot *et al.* 2013) using a constrained
184 tree topology. Four different methodological decisions were assessed: the type of clock, the model of
185 molecular evolution, the prior distribution applied to calibrated nodes, and the approach used to
186 subsample loci. The first three of these are parameters specified when running the program. Divergence
187 times were obtained applying both an autocorrelated log-normal (`-ln`; Thorne *et al.* 1998) and an
188 uncorrelated gamma (`-ugam`; Drummond *et al.* 2006) clock; implementing unpartitioned site-
189 homogenous GTR + G (`-gtr`; Tavaré 1986) or site-heterogenous CAT-GTR + G (`-cat -gtr`; Lartillot &
190 Philippe 2004) models of molecular evolution; and using prior distributions of node ages that were
191 either uniformly distributed between stratigraphic bounds (`-ilb`), or implemented a truncated Cauchy
192 minimum bound (`-lb`; Inoue *et al.* 2010). Default parameter values were used for the latter (p = 0.1, c =
193 1), which results in the prior probability peaking at 110% the age of the minimum bound. The tree prior
194 was set to a birth-death process with free parameters and soft calibration bounds were activated (Yang
195 & Rannala 2006), allowing for 5% of prior probability to lie outside of the specified bounds for calibrated
196 nodes (`-bd -sb`). Default settings were used for all other parameters.

197 Regarding the fourth and last decision evaluated (the set of loci used), phylogenomic datasets
198 were subsampled using strategies intended to achieve four distinct goals: A) minimize topological
199 conflict; B) improve levels of phylogenetic usefulness; C) maximize the amount of information; and D)
200 enhance clock-like behavior. For the first of these, loci chosen corresponded to those displaying the
201 minimum Robinson-Foulds (RF) distance (Robinson & Foulds 1981) between gene trees and the
202 constrained species tree. For the second, loci were chosen using the multivariate subsampling strategy
203 *genesortR* (Mongiardino Koch 2021; Mongiardino Koch & Thompson 2021), which identifies loci that
204 simultaneously display high phylogenetic signal and low levels of potential sources of bias. Subsampled

205 matrices with maximal amount of information were built from loci with the highest levels of occupancy.
206 Finally, clock-likeness was estimated using the variance of root-to-tip distances displayed by gene trees.
207 In the presence of outgroups, gene trees were rooted ahead of estimation using the node corresponding
208 to the last common ancestor of all ingroup taxa; otherwise, gene trees were midpoint rooted along a
209 random internal edge. This is expected to cause few problems for the curculionoid and decapod datasets
210 for which outgroup occupancy was 98.6% and 81.0%, respectively, but could pose an issue for the
211 eukaryote dataset. To explore this caveat, we estimated the variance of root-to-tip distances for all
212 outgroup-containing gene trees using both outgroup and random midpoint rooting. Correlations
213 between the two were high (Peason's $\rho = 0.83$ for Curculionoidea and 0.92 for Decapoda, both $p <$
214 0.001). When used to select the top fifth most clock-like loci, both approaches agreed on 75.5%
215 (Curculionoidea) and 78.8% (Decapoda) of them, further reinforcing that loci displaying clock-like
216 evolution can be recognized in the absence of outgroups. Given the positive relationship between the
217 variance of root-to-tip distances and the overall rate of evolution of genes (Mongiardino Koch *et al.*
218 2022; see Fig. S1), loci were first sorted by rate (estimated using the overall tree length; Telford *et al.*
219 2014), and both the fastest- and slowest-evolving 15% of loci were excluded. Loci chosen corresponded
220 to those with minimum root-to-tip variances from among those showing relatively average evolutionary
221 rates. Gene trees necessary to measure these properties were inferred using ParGenes v.1.0.1 (Morel *et*
222 *al.* 2018), using models that minimized the Bayesian Information Criterion and estimating support with
223 100 bootstrap replicates. A fifth subsampled dataset was derived by selecting loci at random, providing a
224 means of estimating the overall effect of targeted subsampling schemes. To standardize these different
225 matrices in terms of the overall amount of data contained, the loci necessary to first exceed a total
226 length of 10,000 positions were concatenated.

227 Dating was performed under all combinations of these four methodological decisions, resulting
228 in 40 conditions per dataset, each of which was run using two independent chains for at least 20,000

229 generations. Posterior topologies were obtained using the verbose output option of the program
230 readdiv, setting a 50% burn-in and thinning chains to every two generations (`-x 10000 2 -v`), for a
231 total of 5,000 posterior trees per chain. Convergence was visually assessed using log-likelihood trace
232 plots (Fig. S2) and comparing median posterior estimates of node ages and model parameters between
233 replicate chains (Figs. S3-S6). After confirming convergence was attained, time-calibrated trees from
234 separate chains were combined into a single file. The remainder of analyses were performed using
235 package *chronospace*, summarizing each run using a random sample of 500 posterior trees. R code to
236 reproduce each of these steps is available as Supplementary Material S1.

237

238 **RESULTS**

239 Figure 2 summarizes the results obtained across the three empirical datasets used, showing the
240 median and 95% confidence interval (CI) of each inferred node age after combining the posterior
241 distributions of individual runs. Producing such overall summaries of temporal information across
242 populations of chronograms is straightforward once the data has been gathered and formatted using the
243 function `extract_ages`. For the ancient divergences estimated in the eukaryote dataset, pooled 95%
244 CI can span upwards of 1,000 Ma, as is the case for example of the last common ancestor of Amebozoa
245 (median: 1,291.2 Ma, 95% CI: 701.2-1,841.2 Ma). The temporal uncertainty associated with younger
246 divergences is much narrower in absolute terms, yet it can be even larger in relative terms (i.e., after
247 dividing the 95% CI width by its median value). For example, a handful of divergences between closely
248 related achelate decapods show 95% CI widths that are six to seven times the value of the median
249 estimate, with plausible dates spanning the last 142 Ma and falling anywhere between the Quaternary
250 and Early Cretaceous.

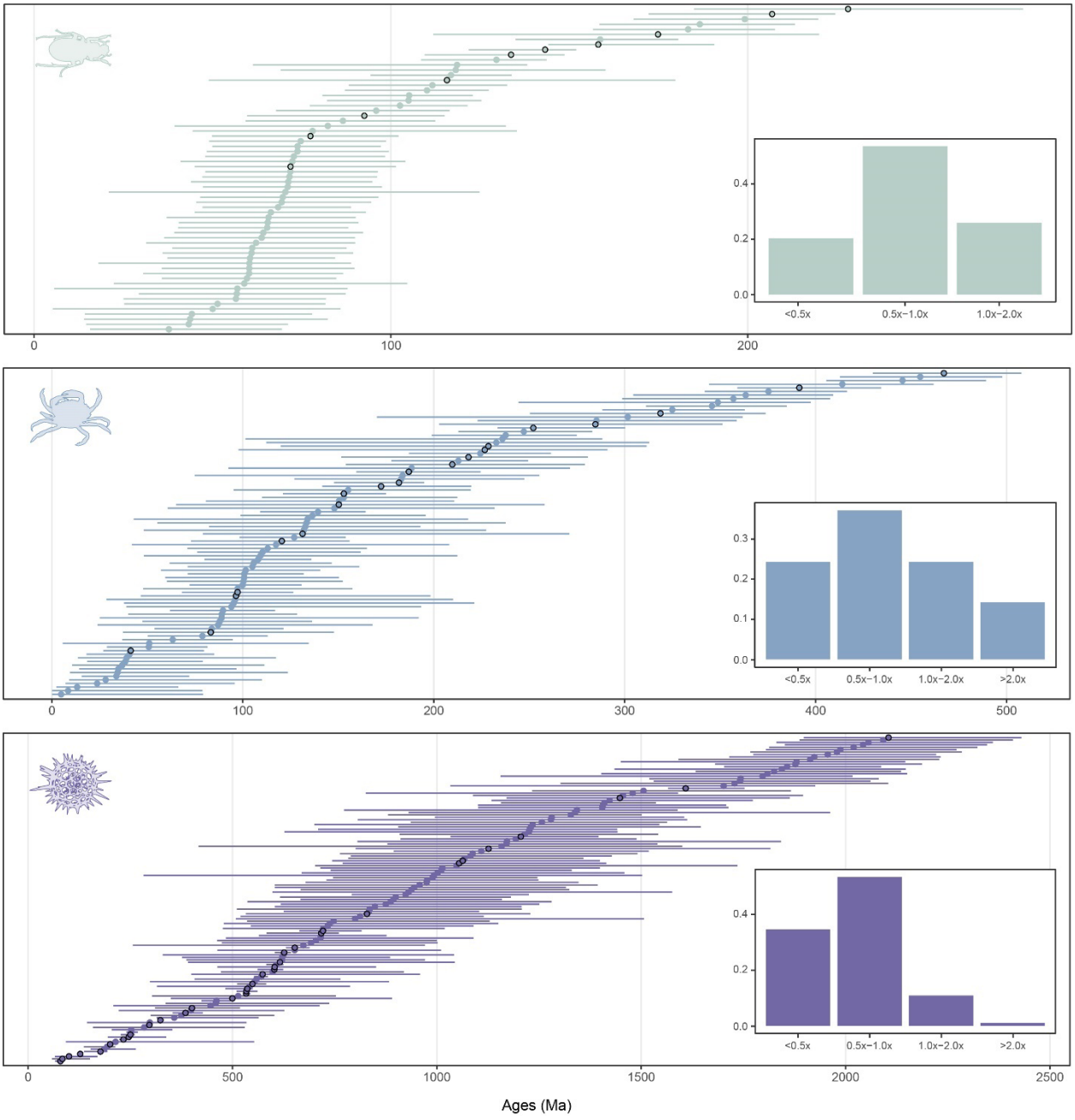


Figure 2: Uncertainty and sensitivity of node age estimates. Values shown are the median and width of pooled 95% CI per node across all analytical conditions. Fossil calibrations were placed on nodes highlighted with black margins. Insets show the distribution of the relative widths of pooled 95% CIs, obtained by dividing absolute widths by the median values. Only unconstrained nodes were tallied.

The spread of posterior dates shown in Figure 2 is a function of the width of the individual CI of each run (uncertainty) and the degree of overlap of the CIs obtained under different conditions (sensitivity). Chronospaces can help disentangle the two by visualizing and quantifying the contribution of each individual methodological choice to overall levels of sensitivity (Figure 3). In the case of the empirical datasets analyzed, this approach reveals that: A) there is a remarkable difference (over 1,000x) in the overall impact of the parameters explored; and, B) the type of molecular clock and the subsample of loci have larger effect sizes for the all three datasets. The implementation of the complex site-heterogeneous CAT-GTR model seems to have a strong effect only for the eukaryote dataset. The most sensitive nodes per dataset to each of the four factors evaluated are found in Figures S7-S9.

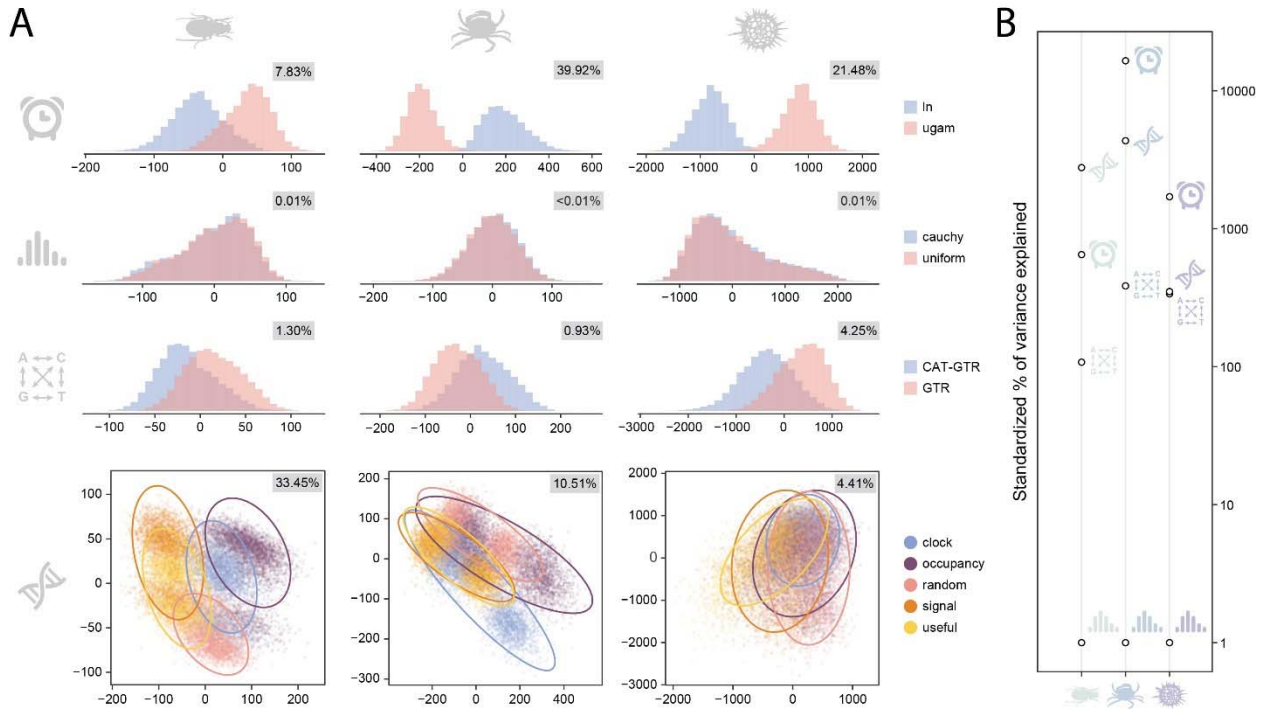


Figure 3: Sensitivity of divergence time estimates to methodological choices. **A.** bgPCA-summarized chronospaces for each factor, from top to bottom: type of molecular clock, prior distribution on constrained nodes, model of molecular evolution, and gene subsample. The proportion of total variance explained by each factor is shown in grey boxes. For the loci subsample, only the first two axes (of four) are shown. Ellipses include 95% of chronograms. **B.** Proportion of total variance explained by each factor, divided by that of the least important factor for each dataset. Axis is log-scaled.

Regarding loci subsampling, chronograms built using high signal and high usefulness loci were very similar across datasets, defining one extreme of the first (major) bgPCA axis (Fig. 4A). The other end of this axis was always populated by chronograms derived from high occupancy loci, and occasionally (for example, in the case of the eukaryote dataset) also by random and clock-like schemes. The branch-length warping plots shown in Figure 4B reveal that the selection of high signal/usefulness loci results in systematically longer internal branches and shorter terminal ones.

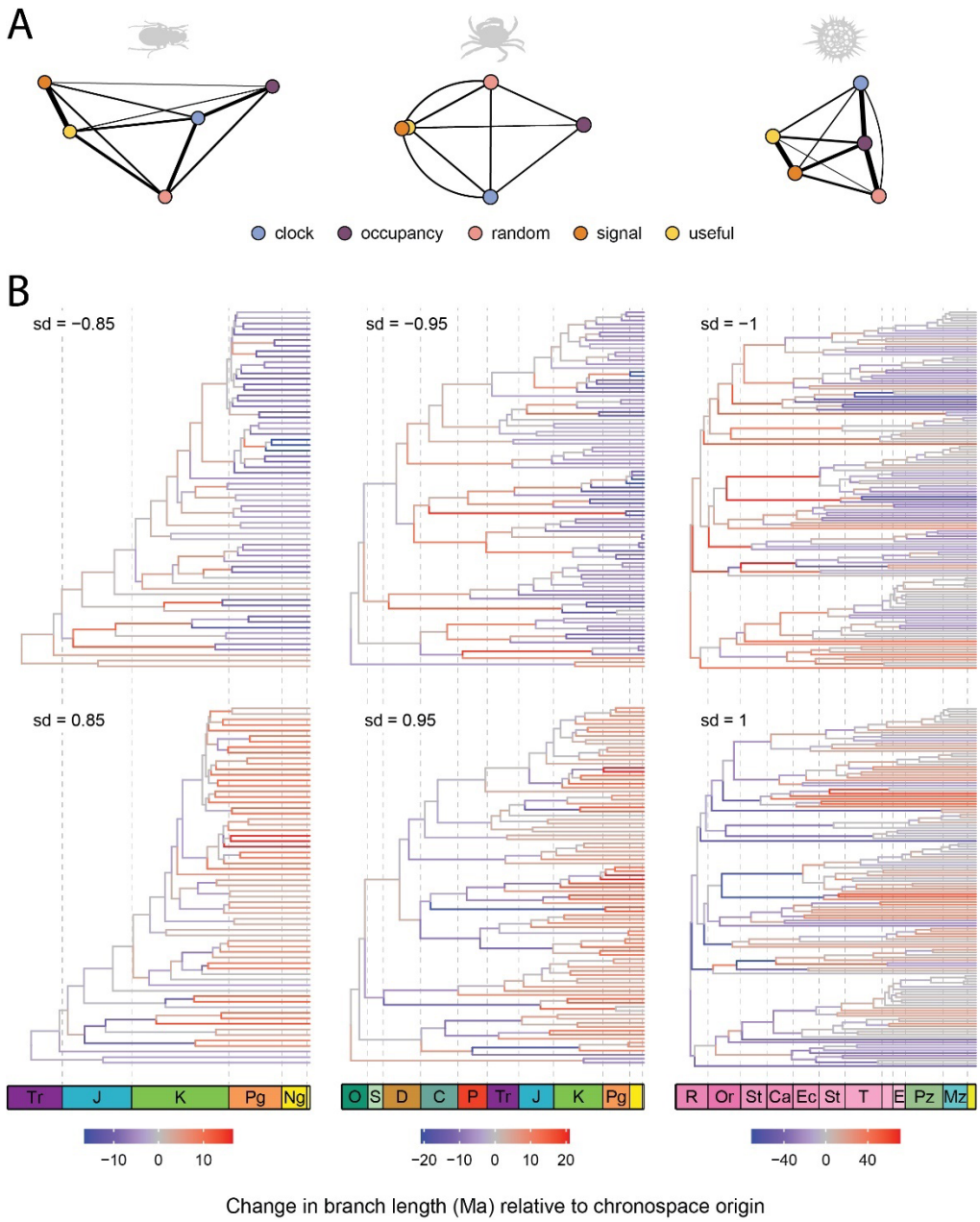


Figure 4: Effect of locus choice on inferred ages. **A.** Chronospaces built discriminating by loci subsampling strategy, as also shown in Fig. 3A. Circles depict the centroids of each distribution; the thickness of the lines connecting them is scaled to the inverse of the Euclidean distances separating them. Locations of centroids are shown on the first two bgPCA axes only, distances were calculated using all dimensions. **B.** Graphical representation of chronograms representing negative (top) and positive

(bottom) extremes of bgPC1. The expansion/contraction of branches relative to a chronogram at the origin of bgPC1 is color coded (red/blue, respectively). sd = standard deviation along bgPCA 1; values < 1 are automatically selected by the package in order to avoid generating negative branch lengths. Chronogram warpings induced by other factors are shown in Figures S10-S12.

DISCUSSION

Understanding evolutionary processes across geologic timescales cannot be done without first establishing an accurate timeline of biological diversification. While phylogenomic datasets, and methods able to handle them, have allowed us to tackle phylogenetic questions that were once intractable (e.g., Philippe *et al.* 1994; Poe & Chubb 2004), they have also introduced a plethora of methodological steps and decisions with downstream effects that we are only beginning to explore (Tan *et al.* 2015; Li *et al.* 2021; Mongiardino Koch 2021; Lozano-Fernandez 2022; Foster *et al.* 2023). In the context of modern Bayesian relaxed-clock dating, these decisions relate to the selection of one among several models that determine how the tree is built, how the characters evolve, and how a clock is enforced—adding up to dozens of parameters that require user-specified prior probability distributions (Álvarez-Carretero & dos Reis 2020; Warnock & Wright 2020). Further decisions need to be taken regarding the character information that is employed, the taxa that are sampled, and the temporal constraints that are enforced. These choices can have strong consequences: they can determine whether we think clades survived mass extinctions (Budd & Mann 2023), emerged through sudden bursts of innovation (Dos Reis *et al.* 2015), participated in events of co-diversification with mutualistic partners (Barba-Montoya *et al.* 2018), and diverged—or not—as a direct consequence of continental drift (Tarver *et al.* 2016).

311 Most of these methodological decisions are also hard to justify. Coupled with the high
312 computational cost of analyses, empirical evolutionary timescales are generally inferred using one (or a
313 few) among hundreds of possible parametric conditions, often times chosen arbitrarily. A literal reading
314 of the literature assessing the sensitivity of results to methodological factors suggests that every one of
315 these decisions can have large impacts on inferred dates (Hugall *et al.* 2007; Inoue *et al.* 2010; Clark et al.
316 2011; Duchêne *et al.* 2014; Hirt *et al.* 2017; Barba-Montoya *et al.* 2018; Dos Reis *et al.* 2018; Smith *et al.*
317 2018; Carruthers & Scotland 2021; Chen *et al.* 2021), providing little *a priori* guidance regarding which
318 should be prioritized given limited computational resources. These assessments are likely also
319 compounded by the choice of dating software, as estimates of sensitivity might be driven by the range of
320 options available, their implementation, or even their default parameterization (which is not always
321 tuned, as is the case of the present study). Progress in pinpointing the magnitude of these effects has
322 been hampered by a general focus on a handful of nodes of interest rather than on tree-wide effects, a
323 scarcity of methods of visualization and summary, as well as of metrics that quantify the effect size of
324 individual decisions in a manner that is comparable, and a lack of testing of whether the relevance of
325 these choices scales with temporal depth.

326 Chronospaces provide novel ways of visualizing, quantifying, and exploring the sensitivity of
327 divergence time estimates, contributing to the inference of more robust evolutionary timescales
328 (Mongiardino Koch *et al.* 2022; Song *et al.* 2023; Wolfe *et al.* 2023). By representing chronograms as
329 collections of node ages, standard multivariate statistical approaches can be readily employed on
330 populations of Bayesian posterior timetrees. This representation allows for more nuanced visualizations
331 that can effectively convey the overall impact of different choices (Fig. 3A), the relative similarity of trees
332 obtained across runs (Fig. 4A), and the patterns of branch length warping that are induced (Fig. 4B). At
333 the same time, partitioning the total variance into proportions explained by discrete factors using sums
334 of squares can result in tree-wide estimates of sensitivity with which to rank methodological choices by

335 their relative importance, helping prioritize computational resources. While the current implementation
336 is limited to analyses run under a constrained tree topology, chronospaces could potentially be built
337 using the subset of ages derived from nodes present in all (or most) of the posterior trees of a
338 simultaneous-inference analysis.

339 As shown in Fig. 3B, some decisions affect our reconstruction of diversification histories much
340 more than others, with effect sizes that vary by three to four orders of magnitude. This extreme
341 heterogeneity can help redirect the allocation of computational effort. For example, our results suggest
342 that empirical studies should either carefully validate their choice of evolutionary clock, or run
343 inferences that vary this component of the analysis, before deriving evolutionary insights (as shown
344 previously in both simulated and empirical studies; see Ho *et al.* 2005; Duchêne *et al.* 2014; Delsuc *et al.*
345 2018; Dos Reis *et al.* 2018; Mongiardino Koch *et al.* 2022; Song *et al.* 2023). Other factors can be
346 arbitrarily fixed without significant consequences, as is the case of the shape of prior distributions on
347 calibrated nodes (as noted previously by Betts *et al.* 2018, although see Clarke *et al.* 2011). We note that
348 the lack of an effect detected here applies exclusively to the use of the Cauchy lower bound (Inoue *et al.*
349 2010) as implemented in PhyloBayes under default settings, and does not mean that the fossil
350 calibration scheme is of little relevance—in fact, ample evidence points to the opposite (e.g., Near *et al.*
351 2005; Magallón *et al.* 2013; Montagna *et al.* 2019). The use of prior distributions with other alternative
352 shapes, the enforcement of hard or soft bounds, the choice of different calibration points (as well as
353 their overall number and relative depths), are all major components of dating analyses that were not
354 evaluated here. Similarly, the software of choice can impart idiosyncratic effects (Bapst *et al.* 2016),
355 meaning that all conclusions drawn in this study could also be software-specific. The main rationale for
356 choosing PhyloBayes is that it remains the only dating software implementing the site-heterogeneous
357 CAT-GTR model (Lartillot & Philippe 2004). Despite ongoing controversies, this model holds preeminence
358 in the inference of phylogenomic topologies (Whelan & Halanych 2017; Li et al. 2021; Bujaki & Rodrigue

359 2022; Al Jewari & Baldauf 2023; Szánthó *et al.* 2023), yet its impact on dating had so far remained
360 unexplored. Our results suggest that the CAT-GTR model induces relatively minor effects. Such
361 robustness of divergence times to the choice of substitution model has been reported before (Du *et al.*
362 2019; Tao *et al.* 2020; Mongiardino Koch *et al.* 2022), implying that the cost of using computationally
363 intensive models is not outweighed by the benefits. Notably, the relatively high effect of the choice of
364 loci and clock are not affected by the timeframe of the phylogenetic question, and affect equally nodes
365 of Cenozoic age as well as divergences spanning the Proterozoic (Fig. 3B). Only two differences were
366 observed as deeper timescales are probed: there is a decrease in the impact of the loci subsample
367 coupled with an increase in that of the clock model (with both remaining the two most important factors
368 overall); and the choice of substitution model becomes slightly more relevant for extremely ancient (e.g.,
369 Precambrian) divergences. The generality of these effects remains to be tested with a broader sample of
370 datasets.

371 Loci subsampling has become a standard approach in phylogenomic dating, yet again, the effects
372 introduced by the each of numerous available approaches have not been fully characterized. While some
373 authors have advocated for the selection of clock-like genes in the hopes that reduced lineage-specific
374 rate variation will improve model fit and diminish bias, others have favored loci with less topological
375 conflict, as these are expected to harbor higher phylogenetic signal and be less affected by incomplete
376 lineage sorting (ILS). Simulations have confirmed that both of these subsampling schemes result in more
377 accurate divergence times (Smith *et al.* 2018; Carruthers *et al.* 2020; Carruthers *et al.* 2022). Other
378 sensible options include selecting genes based on phylogenetic usefulness (as estimated using
379 *genesortR*; Mongiardino Koch 2021), which should balance both aforementioned strategies by
380 discovering loci with increased clock-likeness and decreased topological incongruence; and maximizing
381 occupancy, increasing the amount of information contained within a matrix of a given size.

382 Our results show that locus choice is a major factor determining the divergence times obtained,
383 and that alternative schemes impart characteristic signatures on inferred chronograms. Topologically
384 congruent loci (targeted by both the signal and usefulness schemes) produce shorter terminal branches
385 and longer internal ones (particularly those subtending species pairs; Fig. 4B). The opposite effect is
386 induced by maximizing occupancy, and to some extent also by favoring clock-like loci. These same
387 patterns of branch length warping were found in the simulation study of Carruthers *et al.* (2022), in
388 which the selection of congruent topologies was tied to an expansion of internal branch lengths and a
389 reduction of terminal ones. This is in fact the expected behavior if topological incongruence is driven by
390 biological factors such as ILS (Mendes & Hahn 2016). While our analyses were not designed to test this
391 phenomenon, the results are consistent with ILS driving topological incongruence even at exceedingly
392 ancient evolutionary timescales. Even if targeting clock-like sequences can improve the modelling of rate
393 variation, doing so in a manner that does not also account for topological incongruence can result in the
394 inference of distorted evolutionary timescales.

395

396 **REFERENCES**

397 Al Jewari, C. & Baldauf, S.L. (2023). Conflict over the eukaryote root resides in strong outliers, mosaics
398 and missing data sensitivity of site-specific (CAT) mixture models. *Systematic Biology*, **72**, 1-16.
399 Álvarez-Carretero, S. & dos Reis, M. (2020) Bayesian phylogenomic dating. In S. Y. W. Ho (Ed.), *The*
400 *Molecular Evolutionary Clock: Theory and Practice* (pp. 221-249). Springer.
401 Bapst, D.W., Wright, A.M., Matzke, N.J. & Lloyd, G.T. (2016). Topology, divergence dates, and
402 macroevolutionary inferences vary between different tip-dating approaches applied to fossil
403 theropods (Dinosauria). *Biology Letters*, **12**, 20160237.

404 Barba-Montoya, J., Sharma, S. & Kumar, S. (2023) Molecular timetrees using relaxed clocks and uncertain
405 phylogenies. *Frontiers in Bioinformatics,* **3**, 10.3389/fbinf.2023.1225807.
406 Barba-Montoya, J., Dos Reis, M., Schneider, H., Donoghue, P.C. & Yang, Z. (2018) Constraining uncertainty
407 in the timescale of angiosperm evolution and the veracity of a Cretaceous Terrestrial Revolution.
408 *New Phytologist,* **218,** 819-834.
409 Betts, H.C., Puttick, M.N., Clark, J.W., Williams, T.A., Donoghue, P.C. & Pisani, D. (2018) Integrated
410 genomic and fossil evidence illuminates life's early evolution and eukaryote origin. Nature
411 Ecology & Evolution, **2**, 1556-1562.
412 Britton, T., Anderson, C.L., Jacquet, D., Lundqvist, S. & Bremer, K. (2007). Estimating divergence times in
413 large phylogenetic trees. *Systematic biology*, **56**, 741-752.
414 Budd, G.E. & Mann, R.P. (2023) Two notorious nodes: a critical examination of relaxed molecular clock
415 age estimates of the bilaterian animals and placental mammals. *Systematic Biology***,** syad057,
416 10.1093/sysbio/syad057.
417 Bujaki, T. & Rodrigue, N. (2022). Bayesian cross-validation comparison of amino acid replacement
418 models: contrasting profile mixtures, pairwise exchangeabilities, and gamma-distributed rates-
419 across-sites. *Journal of Molecular Evolution*, **90**, 468-475.
420 Cardini, A., O'Higgins, P. & Rohlf, F.J. (2019) Seeing distinct groups where there are none: spurious
421 patterns from between-group PCA. *Evolutionary Biology,* **46,** 303-316.
422 Carruthers, T., Sanderson, M.J. & Scotland, R.W. (2020) The implications of lineage-specific rates for
423 divergence time estimation. *Systematic Biology,* **69,** 660-670.
424 Carruthers, T. & Scotland, R.W. (2021) Uncertainty in divergence time estimation. *Systematic Biology*, **70**,
425 858-861.

426 Carruthers, T., Sun, M., Baker, W.J., Smith, S.A., de Vos, J.M. & Eiserhardt, W.L. (2022) The implications of
427 incongruence between gene tree and species tree topologies for divergence time estimation.
428 *Systematic Biology,* **71,** 1124-1146.
429 Chen, D., Hosner, P.A., Dittmann, D.L., O'Neill, J.P., Birks, S.M., Braun, E.L. & Kimball, R.T. (2021)
430 Divergence time estimation of Galliformes based on the best gene shopping scheme of
431 ultraconserved elements. *BMC Ecology and Evolution,* **21,** 209.
432 Clark, J.T., Warnock, R.C.M. & Donoghue, P.C.J. (2011) Establishing a time-scale for plant evolution. *New*
433 *Phytologist*, **192**, 266-301.
434 Crotty, S.M. & Holland, B.R. (2022) Comparing partitioned models to mixture models: Do information
435 criteria apply? *Systematic Biology,* **71,** 1541-1548.
436 Culhane, A.C., Perriere, G., Considine, E.C., Cotter, T.G. & Higgins, D.G. (2002) Between-group analysis of
437 microarray data. *Bioinformatics,* **18,** 1600-1608.
438 Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M.-K., Turon, X., Lopez-Legentil, S., Piette, J.,
439 Lemaire, P. & Douzery, E.J. (2018) A phylogenomic framework and timescale for comparative
440 studies of tunicates. *BMC Biology,* **16,** 39.
441 Dolédec, S. & Chessel, D. (1987) Rythmes saisonniers et composantes stationnelles en milieu aquatique.
442 I: Description d'un plan d'observation complet par projection de variables. *Acta Oecologica.*
443 *Oecologia Generalis,* **8,** 403-426.
444 Dolédec, S. & Chessel, D. (1989) Rythmes saisonniers et composantes stationnelles en milieu aquatique.
445 II: Prise en compte et élimination d'effets dans un tableau faunistique. *Acta Oecologica.*
446 *Oecologia Generalis,* **10,** 207-232.
447 Dos Reis, M., Gunnell, G.F., Barba-Montoya, J., Wilkins, A., Yang, Z. & Yoder, A.D. (2018) Using
448 phylogenomic data to explore the effects of relaxed clocks and calibration strategies on
449 divergence time estimation: primates as a test case. *Systematic Biology,* **67,** 594-615.

450 Dos Reis, M., Thawornwattana, Y., Angelis, K., Telford, M.J., Donoghue, P.C. & Yang, Z. (2015) Uncertainty
451 in the timing of origin of animals and the limits of precision in molecular timescales. *Current*
452 *Biology,* **25,** 2939-2950.
453 Dos Reis, M. & Yang, Z. (2019) Bayesian molecular clock dating using genome-scale datasets. In M.
454 Anisimova (Ed.), *Evolutionary genomics. Methods in molecular biology, vol. 1910* (pp. 309-330).
455 Humana.
456 Douglas, J., Jiménez-Silva, C.L. & Bouckaert, R. (2022) StarBeast3: adaptive parallelized Bayesian
457 inference under the multispecies coalescent. *Systematic Biology,* **71,** 901-916.
458 Drummond, A.J., Ho, S.Y., Phillips, M.J. & Rambaut, A. (2006) Relaxed phylogenetics and dating with
459 confidence. *PLoS Biology,* **4,** e88.
460 Du, Y., Wu, S., Edwards, S.V. & Liu, L. (2019) The effect of alignment uncertainty, substitution models and
461 priors in building and dating the mammal tree of life. *BMC Evolutionary Biology,* **19,** 203.
462 Duchêne, S., Lanfear, R. & Ho, S.Y. (2014) The impact of calibration and clock-model choice on molecular
463 estimates of divergence times. *Molecular Phylogenetics and Evolution,* **78,** 277-289.
464 Eme, L., Sharpe, S.C., Brown, M.W. & Roger, A.J. (2014) On the age of eukaryotes: evaluating evidence
465 from fossils and molecular clocks. *Cold Spring Harbor Perspectives in Biology,* **6,** a016139.
466 Foster, P.G., Schrempf, D., Szöllősi, G.J., Williams, T.A., Cox, C.J. & Embley, T.M. (2023) Recoding amino
467 acids to a reduced alphabet may increase or decrease phylogenetic accuracy. *Systematic Biology,*
468 **72,** 723-737.
469 Gearty, W. (2022) deeptime: plotting tools for anyone working in deep time. R package version 0.2.2,
470 https://CRAN.R-project.org/package=deeptime.
471 Heath, T.A., Huelsenbeck, J.P. & Stadler, T. (2014) The fossilized birth–death process for coherent
472 calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences,* **111,**
473 E2957-E2966.

474 Hillis, D.M., Heath, T.A. & John, K.S. (2005) Analysis and visualization of tree space. *Systematic Biology,*
475 **54,** 471-482.
476 Hirt, M.V., Arratia, G., Chen, W.-J., Mayden, R.L., Tang, K.L., Wood, R.M. & Simons, A.M. (2017) Effects of
477 gene choice, base composition and rate heterogeneity on inference and estimates of divergence
478 times in cypriniform fishes. *Biological Journal of the Linnean Society,* **121,** 319-339.
479 Ho, S.Y., Phillips, M.J., Drummond, A.J. & Cooper, A. (2005) Accuracy of rate estimation using relaxed-
480 clock models with a critical focus on the early metazoan radiation. *Molecular Biology and*
481 *Evolution,* **22,** 1355-1363.
482 Howard, R.J., Giacomelli, M., Lozano-Fernandez, J., Edgecombe, G.D., Fleming, J.F., Kristensen, R.M., Ma,
483 X., Olesen, J., Sørensen, M.V. & Thomsen, P.F. (2022) The Ediacaran origin of Ecdysozoa:
484 integrating fossil and phylogenomic data. *Journal of the Geological Society,* **179,** jgs2021-107.
485 Hugall, A.F., Foster, R. & Lee, M.S. (2007) Calibration choice, rate smoothing, and the pattern of tetrapod
486 diversification according to the long nuclear gene RAG-1. *Systematic Biology,* **56,** 543-563.
487 Inoue, J., Donoghue, P.C. & Yang, Z. (2010) The impact of the representation of fossil calibrations on
488 Bayesian estimation of species divergence times. *Systematic Biology,* **59,** 74-89.
489 Kassambara, A. (2020) ggpubr: 'ggplot2' based publication ready plots. R package version 0.4.0,
490 https://CRAN.R-project.org/package=ggpubr.
491 Kumar, S. & Hedges, S.B. (2016) Advances in time estimation methods for molecular data. *Molecular*
492 *Biology and Evolution,* **33,** 863-869.
493 Lartillot, N. (2023) Identifying the best approximating model in Bayesian phylogenetics: Bayes factors,
494 cross-validation or wAIC? *Systematic Biology,* **72,** 616-638.
495 Lartillot, N. & Philippe, H. (2004) A Bayesian mixture model for across-site heterogeneities in the amino-
496 acid replacement process. *Molecular Biology and Evolution,* **21,** 1095-1109.

Lartillot, N., Rodrigue, N., Stubbs, D. & Richer, J. (2013) PhyloBayes MPI. Phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Systematic Biology,* **62,** 611-615.

Li, Y., Shen, X.-X., Evans, B., Dunn, C.W. & Rokas, A. (2021) Rooting the animal tree of life. *Molecular Biology and Evolution,* **38**, 4322-4333.

Lozano-Fernandez, J. (2022) A practical guide to design and assess a phylogenomic study. *Genome Biology and Evolution,* **14,** evac129.

Magallón, S., Hilu, K.W. & Quandt, D. (2013). Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *American Journal of Botany*, **100**, 556-573.

May, M.R. & Rothfels, C.J. (2023) Diversification models conflate likelihood and prior, and cannot be compared using conventional model-comparison tools. *Systematic Biology,* **72,** 713-722.

Mendes, F.K. & Hahn, M.W. (2016) Gene tree discordance causes apparent substitution rate variation. *Systematic Biology,* **65,** 711-721.

Mongiardino Koch, N. (2021) Phylogenomic subsampling and the search for phylogenetically reliable loci. *Molecular Biology and Evolution,* **38**, 4025-4038.

Mongiardino Koch, N. & Thompson, J.R. (2021) A total-evidence dated phylogeny of Echinoidea combining phylogenomic and paleontological data. *Systematic Biology,* **70,** 421-439.

Mongiardino Koch, N., Thompson, J.R., Hiley, A.S., McCowin, M.F., Armstrong, A.F., Coppard, S.E., Aguilera, F., Bronstein, O., Kroh, A. & Mooi, R. (2022) Phylogenomic analyses of echinoid diversification prompt a re-evaluation of their fossil record. *eLife,* **11,** e72460.

Montagna, M., Tong, K.J., Magoga, G., Strada, L., Tintori, A., Ho, S.Y. & Lo, N. (2019). Recalibration of the insect evolutionary time scale using Monte San Giorgio fossils suggests survival of key lineages through the End-Permian Extinction. *Proceedings of the Royal Society B*, **286**, 20191854.

Morel, B., Kozlov, A.M. & Stamatakis, A. (2018) ParGenes: a tool for massively parallel model selection and phylogenetic tree inference on thousands of genes. *Bioinformatics,* **35,** 1771-1773.

Near, T.J., Bolnick, D.I. & Wainwright, P.C. (2005). Fossil calibrations and molecular divergence time estimates in centrarchid fishes (Teleostei: Centrarchidae). *Evolution*, **59**, 1768-1782.

O'Reilly, J.E., dos Reis, M. & Donoghue, P.C. (2015) Dating tips for divergence-time estimation. *TRENDS in Genetics,* **31,** 637-650.

Paradis, E. & Schliep, K. (2018) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics,* **35,** 526-528.

Pedersen, T. (2020) patchwork: the composer of plots. R package version 1.1.1, https://CRAN.R-project.org/package=patchwork.

Pennell, M.W. & Harmon, L.J. (2013) An integrative view of phylogenetic comparative methods: connections to population genetics, community ecology, and paleobiology. *Annals of the New York Academy of Sciences,* **1289,** 90-105.

Philippe, H., Chenuil, A. & Adoutte, A. (1994) Can the Cambrian explosion be inferred through molecular phylogeny? *Development,* **1994,** 15-25.

Poe, S. & Chubb, A.L. (2004) Birds in a bush: five genes indicate explosive evolution of avian orders. *Evolution,* **58,** 404-415.

R Core Team (2022) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rannala, B., Leache, A., Edwards, S. & Yang, Z. (2020) The multispecies coalescent model and species tree inference. In C. Scornavacca, F. Delsuc & N. Galtier (Eds.), *Phylogenetics in the genomic era* (pp. 3:3:1-3:3:21). Self published.

Revell, L.J. (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution,* **3,** 217-223.

Robinson, D.F. & Foulds, L.R. (1981) Comparison of phylogenetic trees. *Mathematical Biosciences,* **53,** 131-147.

Ronquist, F., Klopfstein, S., Vilhelmsen, L., Schulmeister, S., Murray, D.L. & Rasnitsyn, A.P. (2012) A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Systematic Biology,* **61,** 973-999.

Sanderson, M.J. & Doyle, J.A. (2001). Sources of error and confidence intervals in estimating the age of angiosperms from rbcL and 18S rDNA data. *American Journal of Botany*, **88**, 1499-1516.

Sauquet, H., Ramírez-Barahona, S. & Magallón, S. (2022) What is the age of flowering plants? *Journal of Experimental Botany*, **73**, 3840-3853.

Schliep, K.P. (2011) phangorn: phylogenetic analysis in R. *Bioinformatics,* **27,** 592-593.

Shin, S., Clarke, D.J., Lemmon, A.R., Moriarty Lemmon, E., Aitken, A.L., Haddad, S., Farrell, B.D., Marvaldi, A.E., Oberprieler, R.G. & McKenna, D.D. (2018) Phylogenomic data yield new and robust insights into the phylogeny and evolution of weevils. *Molecular Biology and Evolution,* **35,** 823-836.

Smith, M.R. (2022) Robust analysis of phylogenetic tree space. *Systematic Biology,* **71,** 1255-1270.

Smith, S.A., Brown, J.W. & Walker, J.F. (2018) So many genes, so little time: A practical approach to divergence-time estimation in the genomic era. *PloS one,* **13,** e0197433.

Song, H., Wang, Y., Shao, H., Li, Z., Hu, P., Yap-Chiongco, M.K., Shi, P., Zhang, T., Li, C. & Wang, Y. (2023) Scaphopoda is the sister taxon to Bivalvia: Evidence of ancient incomplete lineage sorting. *Proceedings of the National Academy of Sciences,* **120,** e2302361120.

Strassert, J.F., Irisarri, I., Williams, T.A. & Burki, F. (2021) A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. *Nature Communications,* **12,** 1879.

Szánthó, L.L., Lartillot, N., Szöllősi, G.J. & Schrempf, D. (2023). Compositionally constrained sites drive long-branch attraction. *Systematic Biology*, **72**, 767-780.

568 Tan, G., Muffato, M., Ledergerber, C., Herrero, J., Goldman, N., Gil, M. & Dessimoz, C. (2015) Current
569 methods for automated filtering of multiple sequence alignments frequently worsen single-gene
570 phylogenetic inference. *Systematic Biology,* **64,** 778-791.
571 Tao, Q., Barba-Montoya, J., Huuki, L.A., Durnan, M.K. & Kumar, S. (2020) Relative efficiencies of simple
572 and complex substitution models in estimating divergence times in phylogenomics. *Molecular*
573 *Biology and Evolution,* **37,** 1819-1831.
574 Tarver, J.E., Dos Reis, M., Mirarab, S., Moran, R.J., Parker, S., O'Reilly, J.E., King, B.L., O'Connell, M.J.,
575 Asher, R.J. & Warnow, T. (2016) The interrelationships of placental mammals and the limits of
576 phylogenetic inference. *Genome Biology and Evolution,* **8,** 330-344.
577 Tavaré, S. (1986) Some probabilistic and statistical problems on the analysis of DNA sequence. *Lecture of*
578 *Mathematics for Life Science,* **17,** 57.
579 Telford, M.J., Lowe, C.J., Cameron, C.B., Ortega-Martinez, O., Aronowicz, J., Oliveri, P. & Copley, R.R.
580 (2014) Phylogenomic analysis of echinoderm class relationships supports Asterozoa. *Proceedings*
581 *of the Royal Society B,* **281,** 20140479.
582 Thorne, J.L., Kishino, H. & Painter, I.S. (1998) Estimating the rate of evolution of the rate of molecular
583 evolution. *Molecular Biology and Evolution,* **15,** 1647-1657.
584 Wang, L.G., Lam, T.T.Y., Xu, S., Dai, Z., Zhou, L., Feng, T., Guo, P., Dunn, C.W., Jones, B.R., Bradley, T., Zhu,
585 H., Guan, Y., Jiang, Y. & Yu, G. (2020) *treeio*: an R package for phylogenetic tree input and output
586 with richly annotated and associated data. *Molecular Biology and Evolution*, **37,** 599-603.
587 Warnock, R.C. & Wright, A.M. (2020) *Understanding the tripartite approach to Bayesian divergence time*
588 *estimation*. Cambridge University Press.
589 Warnock, R.C., Yang, Z. & Donoghue, P.C. (2011) Exploring uncertainty in the calibration of the molecular
590 clock. *Biology Letters*, **8,** 156-159.

591 Whelan, N.V. & Halanych, K.M. (2017). Who let the CAT out of the bag? Accurately dealing with
592 substitutional heterogeneity in phylogenomic analyses. *Systematic biology*, **66**, 232-255.
593 Wickham, H. (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag: New York.
594 Wickham, H. (2017) tidyverse: Easily install and load 'tidyverse' packages. R package version 1.2.1.
595 https://CRAN.R-project.org/package=tidyverse.
596 Wolfe, J.M., Ballou, L., Luque, J., Watson-Zink, V.M., Ahyong, S.T., Barido-Sottani, J., Chan, T.-Y., Chu, K.H.,
597 Crandall, K.A., Daniels, S.R. Felder, D.L., Mancke, H., Martin, J.W., Ng, P.K.L., Ortega-Hernández, J.,
598 Palacios Theil, E., Pentcheff, N.D., Robles, R., Thoma, B.P., Tsang, L.M., Wetzer, R., Windsor, A.M.
599 & Bracken-Grissom, H.D. (2023) Convergent adaptation of true crabs (Decapoda: Brachyura) to a
600 gradient of terrestrial environments. *Systematic Biology*, syad066, 10.1093/sysbio/syad066.
601 Wolfe, J.M., Breinholt, J.W., Crandall, K.A., Lemmon, A.R., Lemmon, E.M., Timm, L.E., Siddall, M.E. &
602 Bracken-Grissom, H.D. (2019) A phylogenomic framework, evolutionary timeline and genomic
603 resources for comparative studies of decapod crustaceans. *Proceedings of the Royal Society B,*
604 **286,** 20190079.
605 Yang, Z. & Rannala, B. (2006) Bayesian estimation of species divergence times under a molecular clock
606 using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution,* **23,** 212-226.
607 Yendle, P.W. & MacFie, H.J. (1989) Discriminant principal components analysis. *Journal of Chemometrics,*
608 **3,** 589-600.
609 Yu, G., Smith, D.K., Zhu, H., Guan, Y. & Lam, T.T.Y. (2017) ggtree: an R package for visualization and
610 annotation of phylogenetic trees with their covariates and other associated data. *Methods in*
611 *Ecology and Evolution,* **8,** 28-36.
612 Zhou, Y., Rodrigue, N., Lartillot, N. & Philippe, H. (2007) Evaluation of the models handling heterotachy in
613 phylogenetic inference. *BMC Evolutionary Biology,* **7,** 206.

614 Zuckerkandl, E. & Pauling, L. (1965) Evolutionary divergence and convergence in proteins. In V. Bryson &
615 H. J. Vogel (Eds.), *Evolving genes and proteins* (pp. 97-166). Academic Press.

616

617

SUPPLEMENTARY MATERIAL - Chronospaces: an R package for the statistical exploration of divergence times promotes the assessment of methodological sensitivity

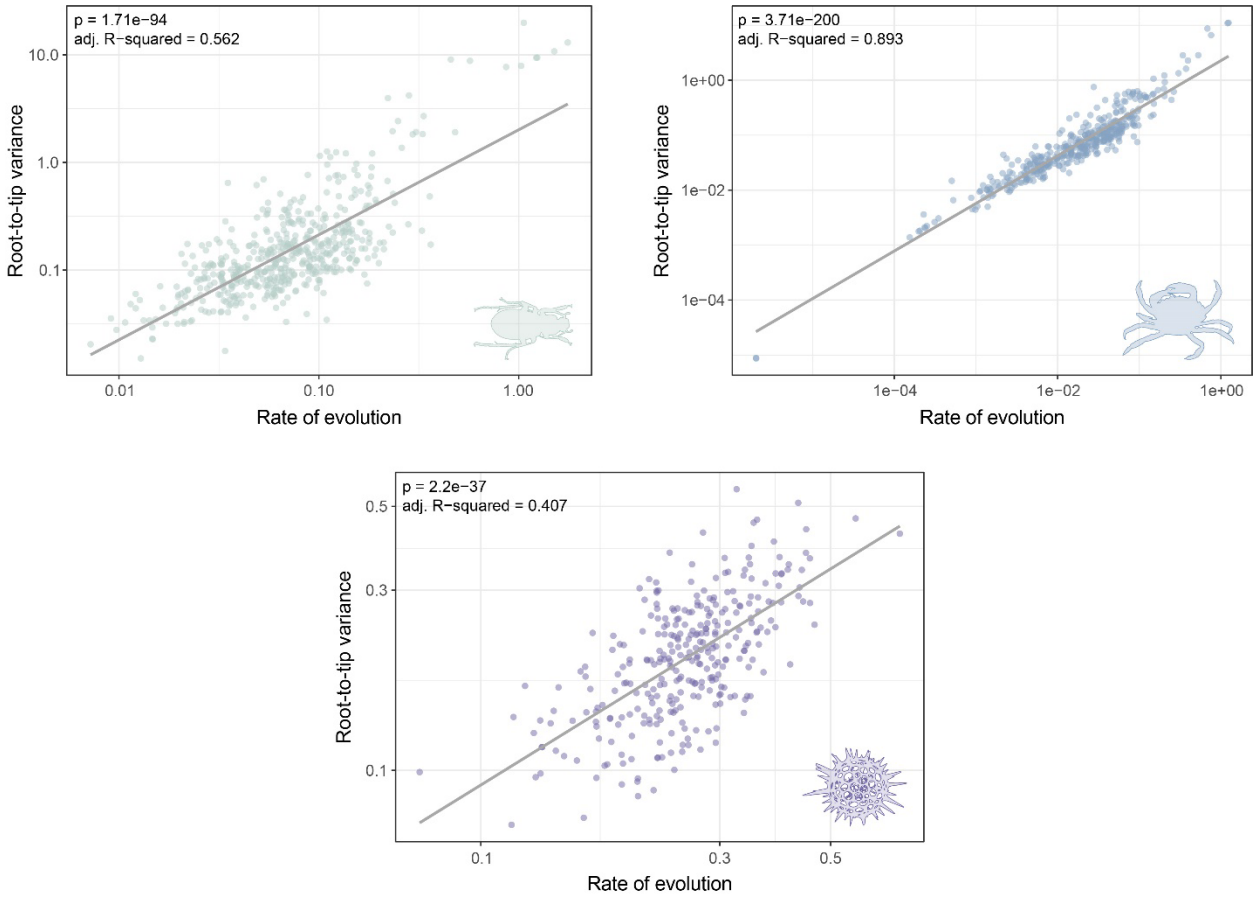


Figure S1: Root-to-tip variance (a proxy for clock-likeness) is strongly determined by the evolutionary rate of loci. Minimizing this variable, a common approach when subsampling matrices for time calibration, also results in selecting genes that evolve at minimum rates. To avoid this confounding effect, clock-like genes from these three matrices were selected after filtering the fastest and slowest evolving 15% of loci. Lines correspond to linear regressions; both axes are log-scaled.

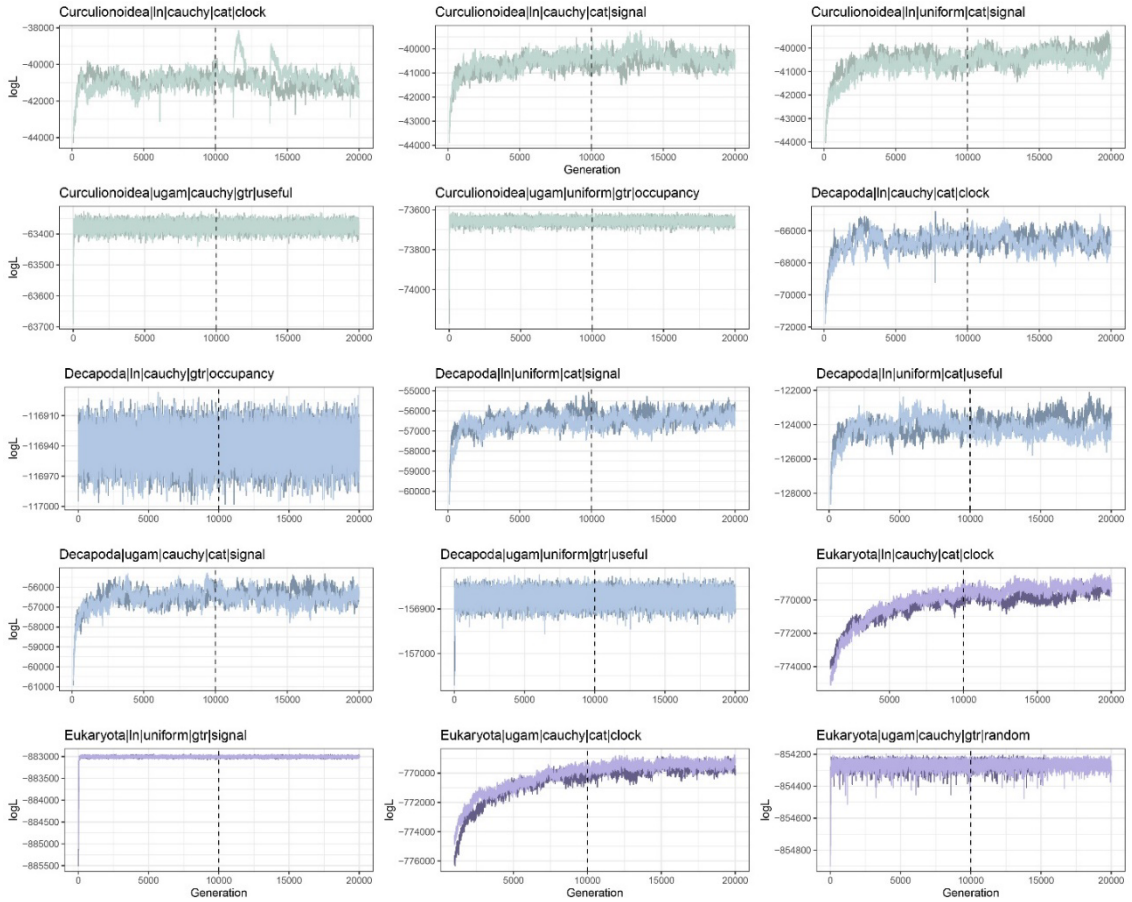


Figure S2: Trace plots of the log-likelihood (logL) of chains. Results correspond to fifteen randomly selected analyses. Dashed lines correspond to the burn-in threshold imposed (10,000 generations). Convergence and stationarity vary between conditions and datasets, with mixing being more challenging when using the site-heterogeneous model CAT + GTR + G, and convergence taking longer for the eukaryote dataset. Nonetheless, both chains of each analysis seem to have converged to the same high posterior region of parameter space by the time the burn-in generations have elapsed. Note that the first few logL values were discarded in order to arrive at a meaningful scaling of the y-axis. The number of excluded generations was 10 for analyses under the GTR + G model, and either 100 (Curculionoidea, Decapoda) or 1,000 (Eukaryota) for analyses under the CAT + GTR + G model. Surprisingly, runs under GTR + G converged within the first 50 generations (sometimes, as is the case of the second decapod plot, within less than 10).

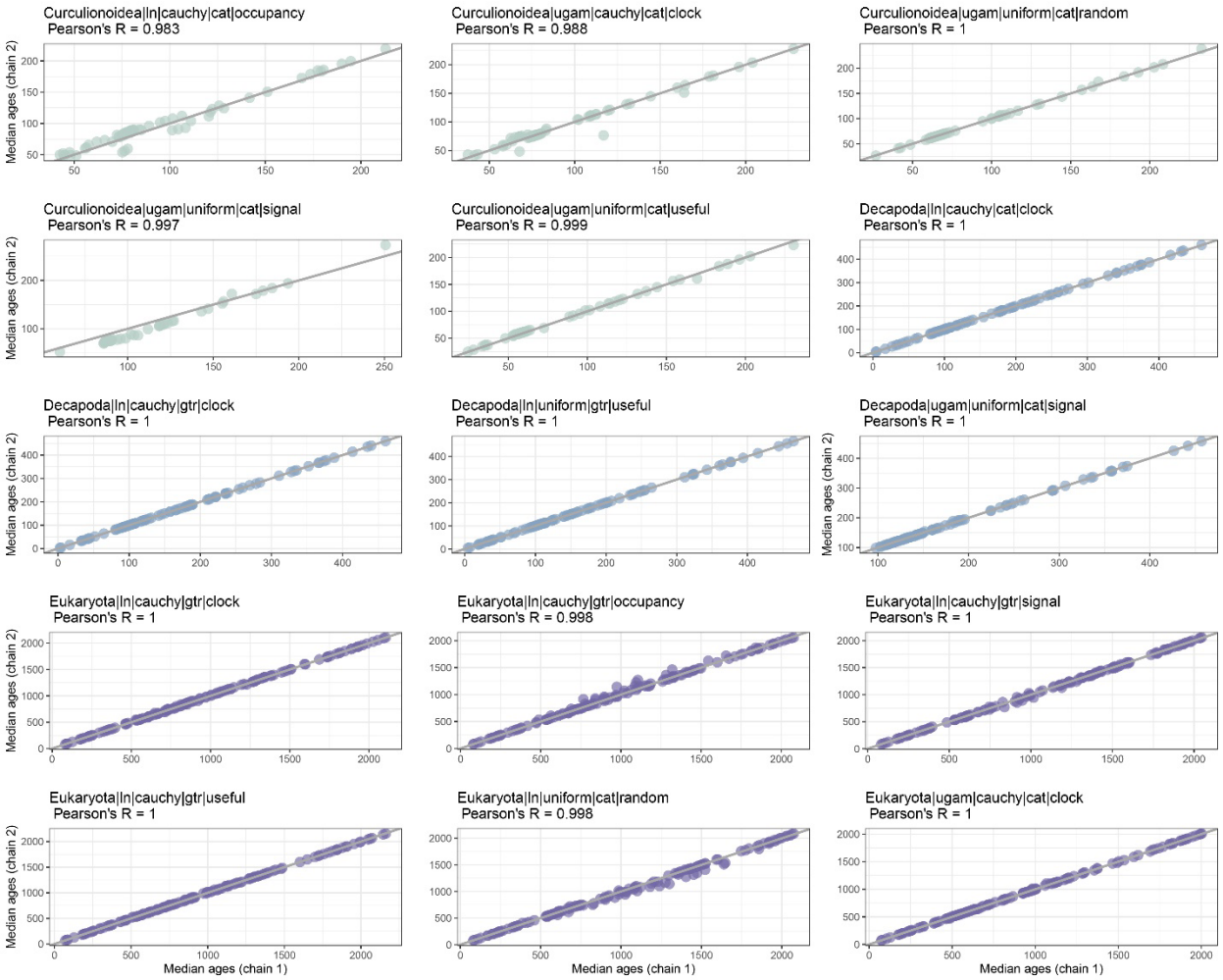


Figure S3: Median node ages for pairs of chains showing appropriate levels of convergence. Results correspond to fifteen randomly selected analyses. Grey lines show 1-to-1 lines. The minimum Pearson correlation coefficient (R) found across all analyses was 0.973.

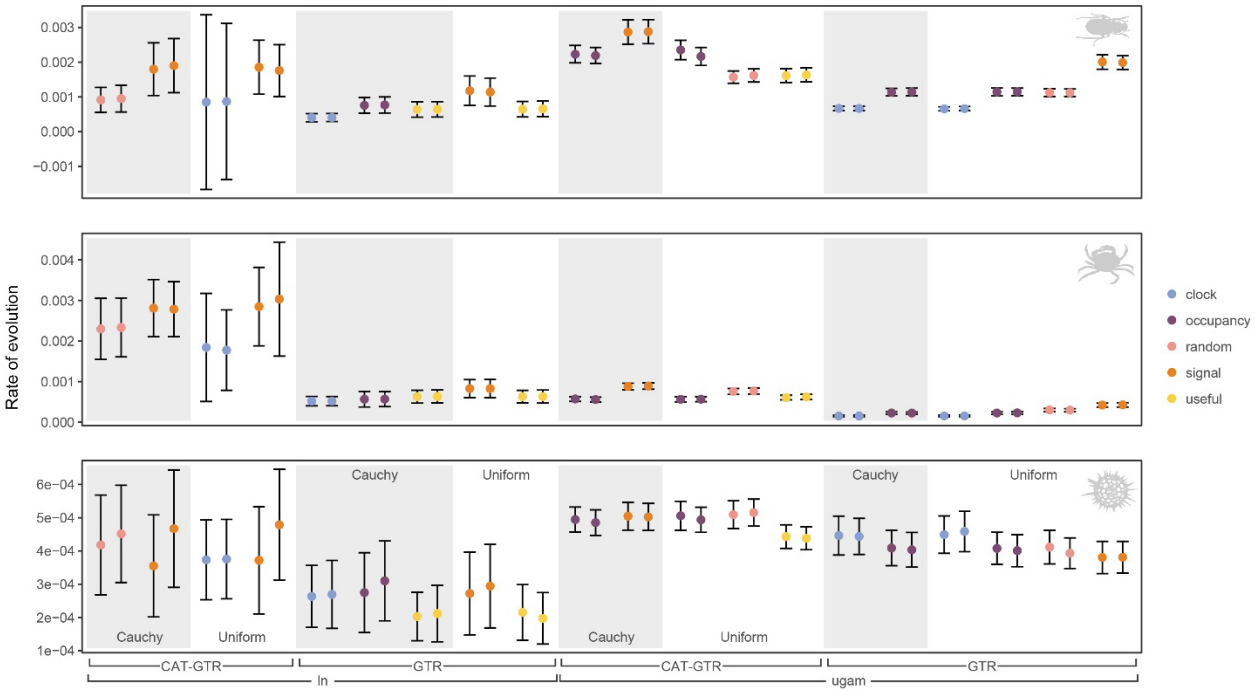


Figure S4: Comparison of the posterior estimates of evolutionary rates across a random sample of 20 analyses per dataset (i.e., 60 analyses total, 2 chains each). Pairs of chains are shown next to each other. Estimates correspond to posterior means ± 1 standard deviation.

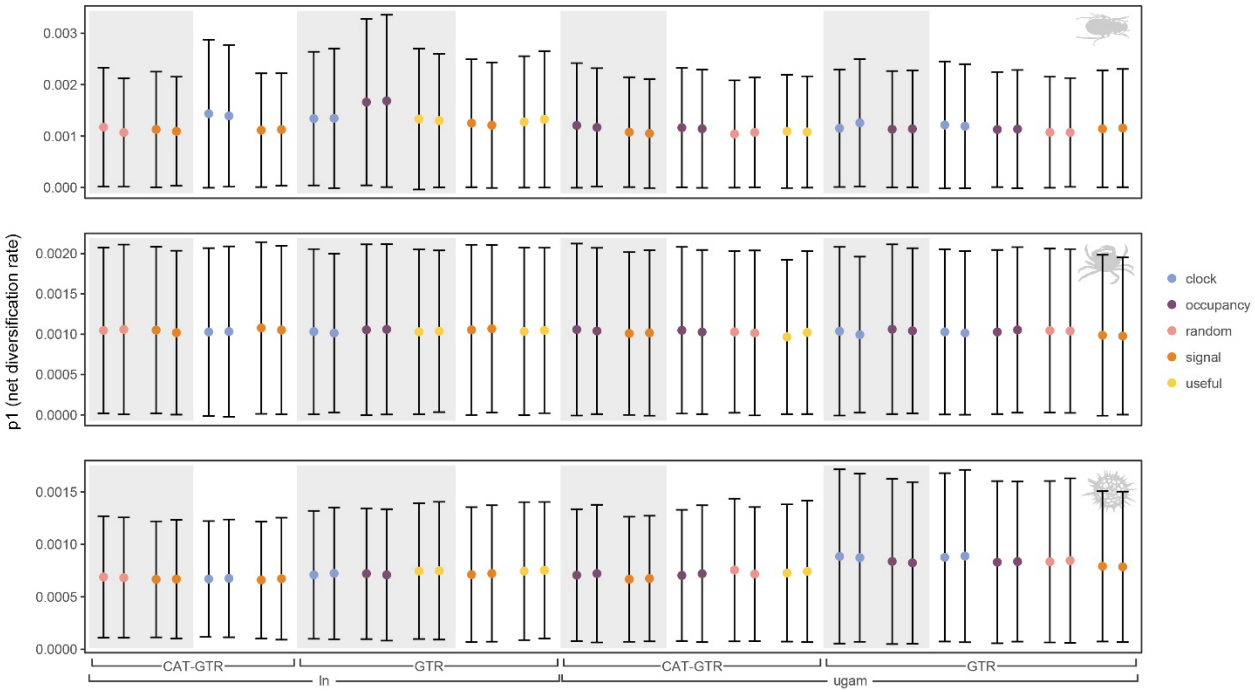


Figure S5: Comparison of the posterior estimates of p1, one of the two parameters of the birth-death tree prior, across a random sample of 20 analyses per dataset (i.e., 60 analyses total, 2 chains each). Pairs of chains are shown next to each other. Estimates correspond to posterior means ± 1 standard deviation. p1 is corresponds to the birth rate (λ) minus the death rate (μ), and is thus the net diversification rate. Data is for the same analyses shown in Fig. S4.

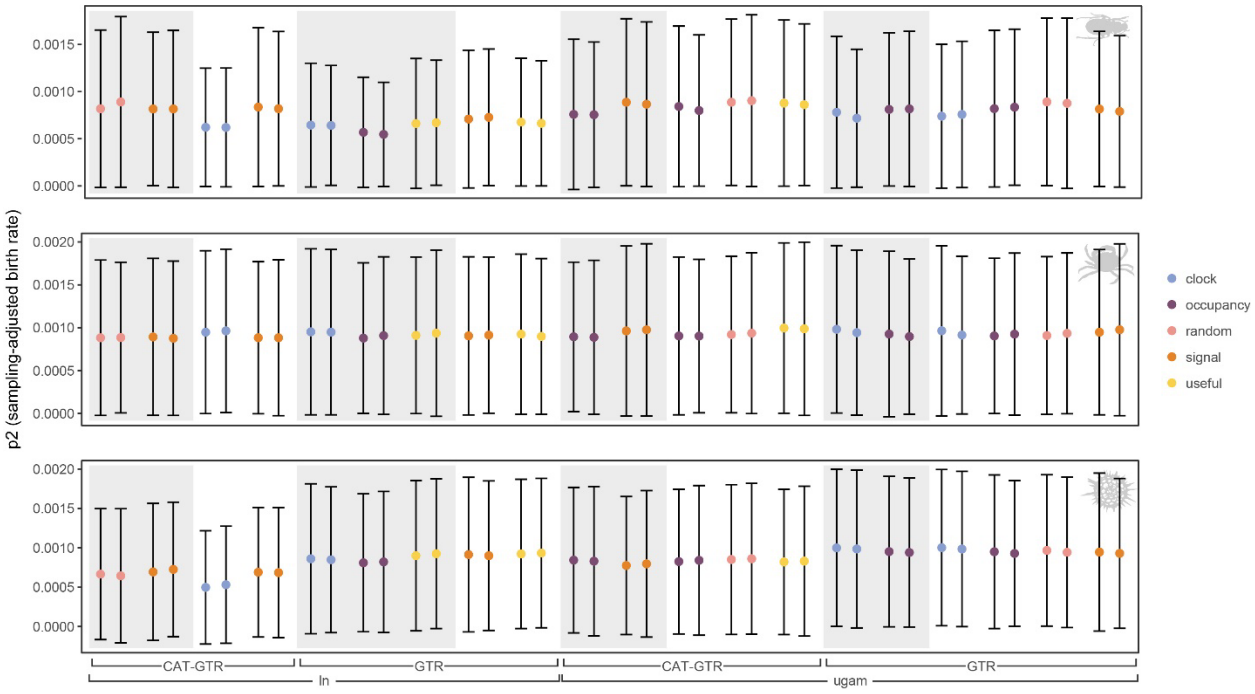


Figure S6: Comparison of the posterior estimates of p2, one of the two parameters of the birth-death tree prior, across a random sample of 20 analyses per dataset (i.e., 60 analyses total, 2 chains each). Pairs of chains are shown next to each other. Estimates correspond to posterior means ± 1 standard deviation. p2 is corresponds to the birth rate (λ) multiplied by the sampling fraction (ρ). Data is for the same analyses shown in Fig. S4.

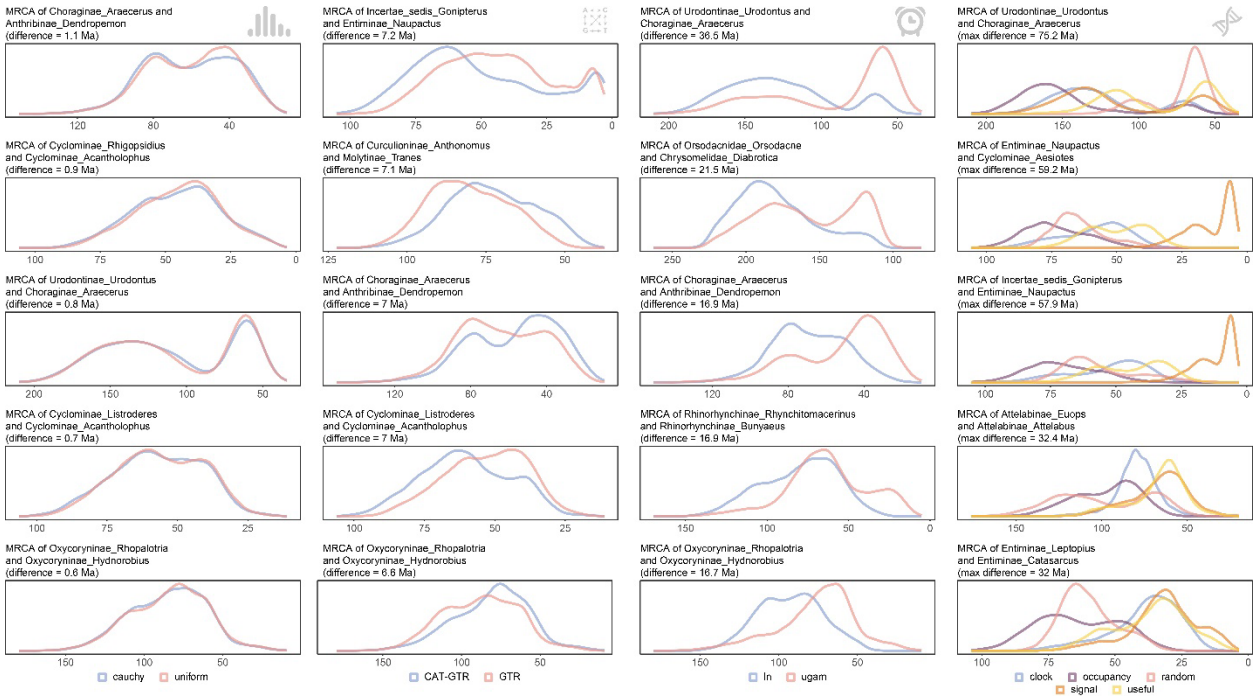


Figure S7: Posterior distributions of ages for the most sensitive nodes of the curculionid dataset. For each methodological choice being varied, the top five most sensitive nodes are depicted. Factors are ordered from the least impactful on the left, to the most impactful on the right (according to the proportions of explained variance, see Fig. 3). Nodes are identified using two randomly-selected terminals from each descendant clade.

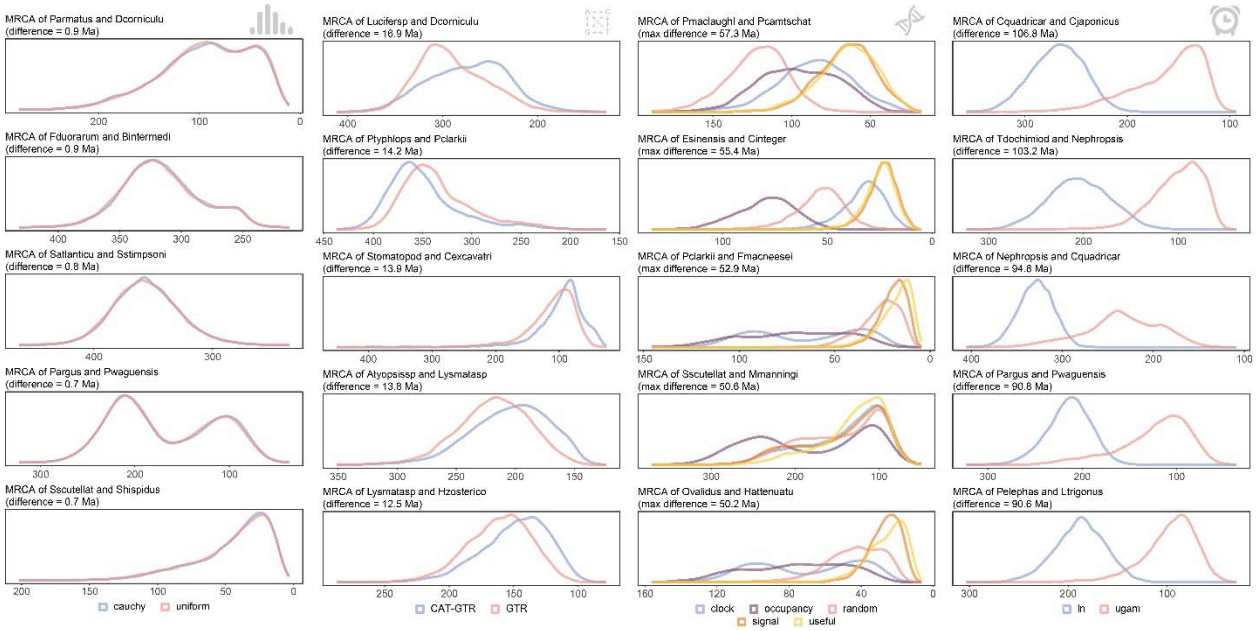


Figure S8: Posterior distributions of ages for the most sensitive nodes of the decapod dataset. For each methodological choice being varied, the top five most sensitive nodes are depicted. Factors are ordered from the least impactful on the left, to the most impactful on the right (according to the proportions of explained variance, see Fig. 3). Nodes are identified using two randomly-selected terminals from each descendant clade.

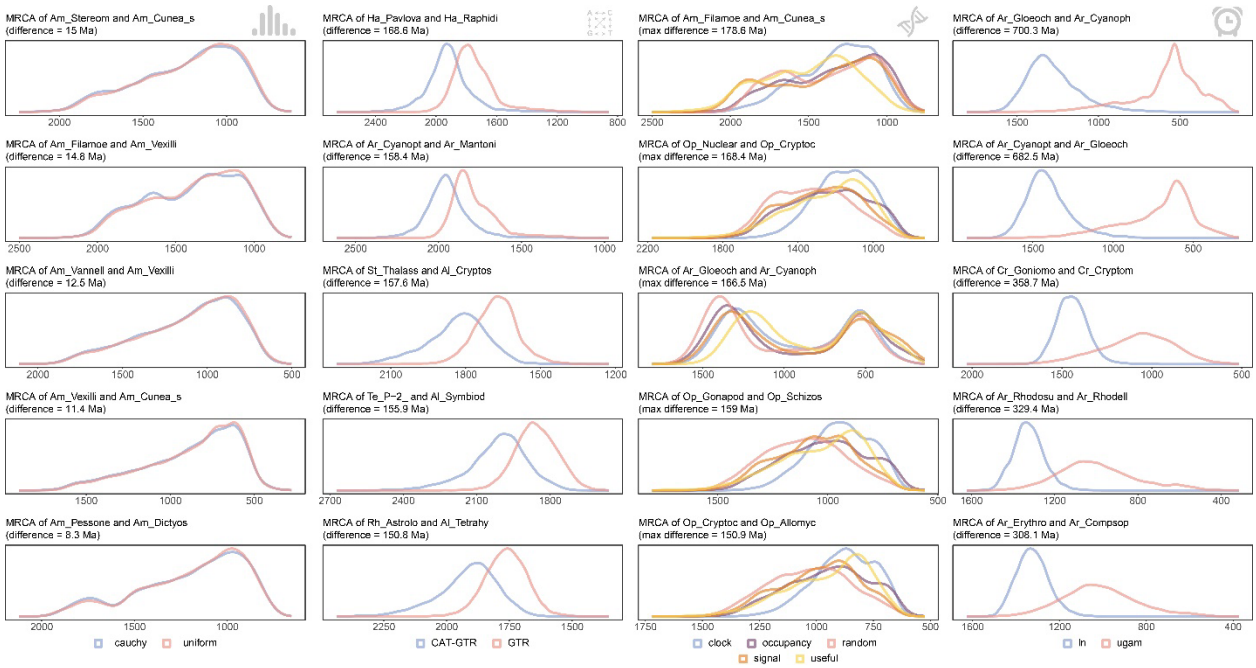
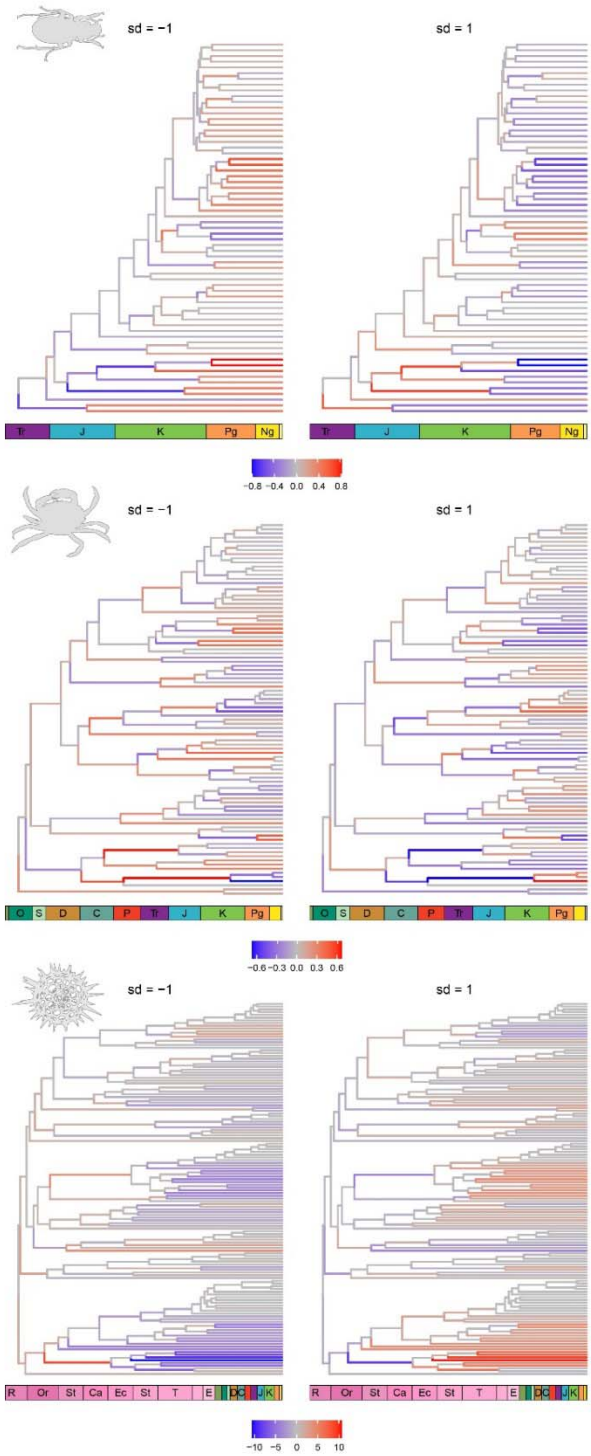


Figure S9: Posterior distributions of ages for the most sensitive nodes of the eukaryote dataset. For each methodological choice being varied, the top five most sensitive nodes are depicted. Factors are ordered from the least impactful on the left, to the most impactful on the right (according to the proportions of explained variance, see Fig. 3). Nodes are identified using two randomly-selected terminals from each descendant clade.



681

682 **Figure S10:** Branch length warping along the bgPCA maximizing the discrimination of chronograms

683 inferred using different node age prior distributions (i.e., the least impactful decision across all datasets;

684 see Fig. 3). This decision seems to induce heterogeneous effects across subclades.

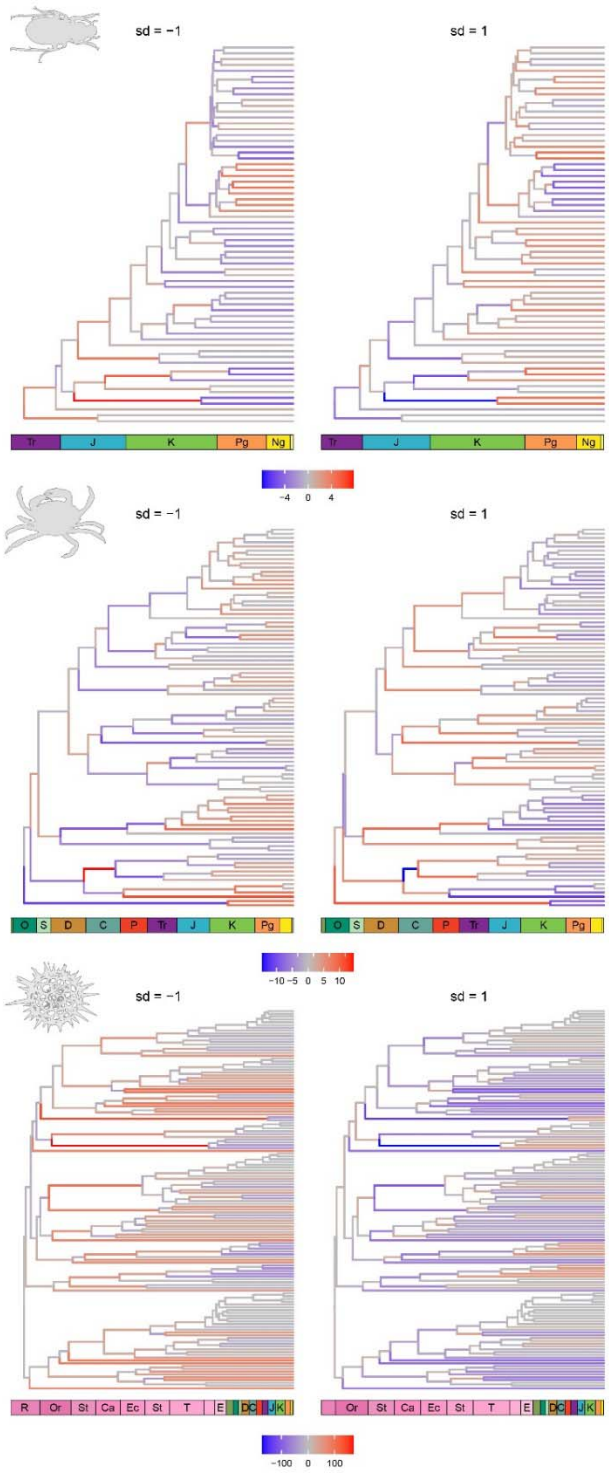


Figure S11: Branch length warping along the bgPCA maximizing the discrimination of chronograms inferred using different models of molecular evolution (i.e., the second least impactful decision across all datasets; see Fig. 3). This decision seems to induce heterogeneous effects across subclades.

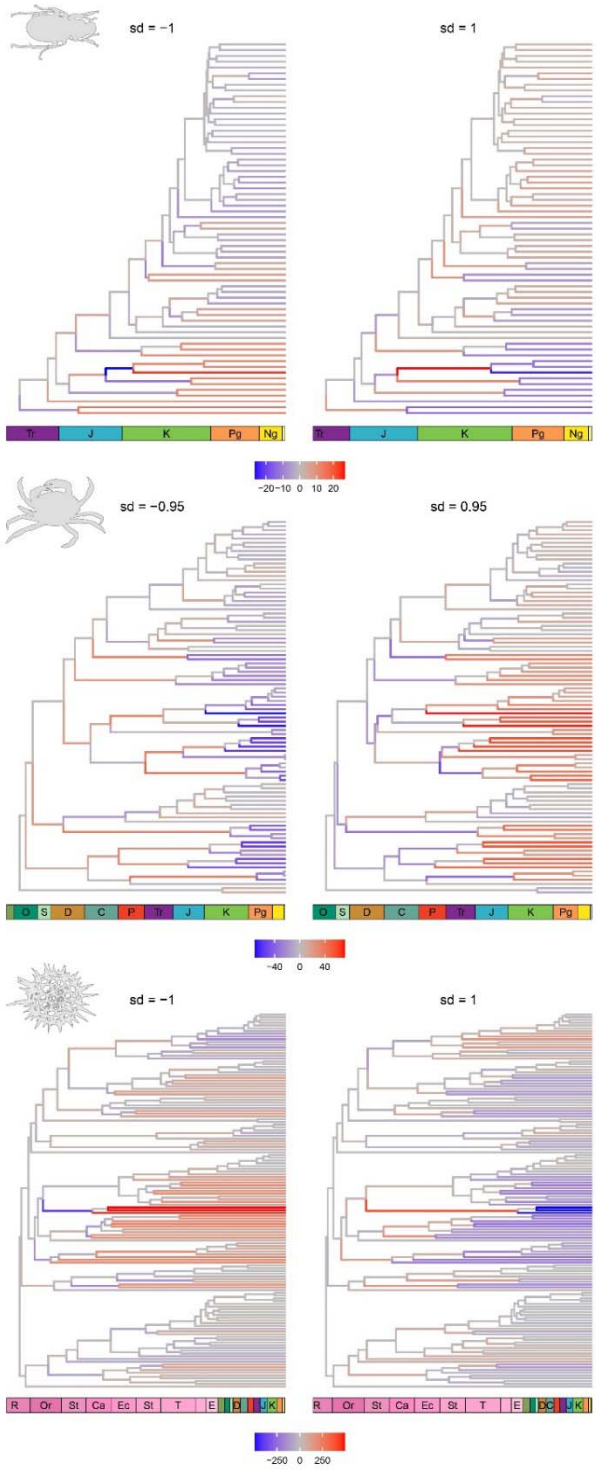


Figure S12: Branch length warping along the bgPCA maximizing the discrimination of chronograms inferred using different clock models (i.e., the first or second most impactful decision across all datasets; see Fig. 3). This decision seems to induce effects that are strong, yet clade specific.