Chronospaces: an R package for the statistical exploration of divergence times reveals extreme dependence on molecular clocks and gene choice

Nicolás Mongiardino Koch[1], Pablo Milla Carmona[2]

[1] *Scripps Institution of Oceanography, University of California San Diego, La Jolla, California, USA. Email: nmongiardinokoch@ucsd.edu.*

[2] *Bristol Palaeobiology Group, School of Earth Sciences, University of Bristol, Bristol, UK.*

ABSTRACT

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 hinge 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 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. 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 using three empirical datasets spanning widely differing timeframes. 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. Notably, these conclusions are as valid for Cenozoic divergences as they are for the deepest eukaryote nodes. 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: time calibration, phylogenetic inference, phylogenomics, diversification, fossils

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, time scaling phylogenies to absolute (or geologic) timescales is often the first step to explore evolutionary patterns and processes at deep timescales, and one on which much of modern comparative and macroevolutionary biology rely on (Pennell & Harmon 2013).

Modern Bayesian methods of divergence time estimation have become highly sophisticated (Kumar & Hedges 2016), being able to accommodate among-species 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 discordances (Rannala *et al.* 2020). Among many achievements, these methods have spawned a new era of epidemiological research through the lens of phylodynamic modelling (Frost *et al.* 2015), and united the fields of molecular phylogenetics and paleontology by highlighting their unique contributions to understanding the tempo of biological diversification (Donoghue & Benton 2007).

As the complexity of time calibration approaches has increased, so has the number of methodological choices that the average scientist is confronted with. Dating software require users to specify which among several competing models of trait evolution, rate variation, and lineage diversification—the three distinct elements of the tripartite framework of Bayesian dating (Warnock &

Wright 2020)—are enforced. Each of these models includes a sizable number of parameters for which prior probability distributions need to be specified. Furthermore, even though progress has been made regarding computational efficiency (Dos Reis & Yang 2019; Douglas *et al.* 2022; Barba-Montoya *et al.* 2023), most methods remain limited in terms of the size of the datasets that can be run within reasonable times, forcing further decisions regarding taxon sampling and loci choice, especially when employing genome-scale datasets. These methodological decisions can have a very large effect on inferred divergence times (Warnock *et al.* 2011; Eme *et al.* 2014; Mongiardino Koch *et al.* 2022; Sauquet *et al.* 2022), producing incongruent reconstructions of biological diversification.

Two broad approaches have been explored to overcome this methodological uncertainty. First, numerous methods have been developed to justify model selection, allowing the identification of a set of parametric conditions considered optimal. These methods, however, have their own limitations: model fit statistics do not reliably discriminate between partition and mixture models (Crotty & Holland 2022), while alternative tree priors might not be amenable to comparison using standard methods such as Bayes factors (May & Rothfels 2023). General purpose approaches, such as cross-validation, are too computationally consuming to be used exhaustively (Zhou *et al.* 2007; Lartillot 2023), while methods designed to tackle specific aspects of the analysis can arrive at opposite conclusions depending on unassessed decisions (e.g., detecting autocorrelated rates can depend on which model of substitution is used to infer the tree that is 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., 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 (i.e., time-scaled trees) 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. It also provides new tools to quantify the overall impact of different methodological choices, providing an estimate of their effect sizes. We illustrate its functionalities with the use of three empirical datasets spanning widely differing evolutionary timescales.

MATERIALS AND METHODS

A) Chronospaces: the approach

Chronospaces are low-dimensional graphical representations that summarize variation in the absolute node ages 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 means for predefined groups of observations. Original observations are then projected onto the bgPCA axes, resulting in an ordination that maximizes the separation between groups through a rigid rotation of space (unlike that obtained by other potential alternatives, such as discriminant analysis), 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 data (Cardini *et al.* 2019). While bgPCA can introduce spurious patterns when the number of variables exceeds the number 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 will typically be much lower than the number of trees sampled in a Bayesian posterior distribution (see below).

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 is estimated for both sets of raw chronograms and bgPCA axes by calculating the corresponding sums of squares. This information can be used to aid in the interpretation of chronospaces as well as rank methodological choices from the most to the least impactful.

B) *chronospace:* the package

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

123 The range of functions and capabilities implemented in the package are depicted in Figure 1 and
124 further documented on the GitHub repository. The typical workflow would begin with the use of
125 function `extract_ages` which gathers node ages from posterior distributions of Bayesian timetrees
126 inferred under a constrained topology. Typically, this would include trees from multiple runs (as most of
127 the methods implemented are designed to facilitate the comparison of results obtained under different
128 analytical conditions), but some analyses can be done with a single set of trees. Chronograms are read
129 from files in newick format (one per run and/or chain), and the node ages contained in them are
130 annotated with a set of factors representing the parameters changed between different runs. The
131 constrained tree topology is also saved in order to generate graphical output. This information (ages,
132 factors, and topology) is stored as an object of class `nodeAges`. Multiple graphical outputs can be
133 directly obtained from this object: posterior distributions of node ages can be plotted using both
134 `sensitive_nodes`, which helps identify and visualize the nodes whose inferred ages change the
135 most between analyses, or `specified_node`, which plots ages for a target node defined by the user;
136 lineage-through-time plots can also be generated using function `ltt_sensitivity` to depict the
137 overall effect of methodological choices on the history of diversification. Finally, bgPCA-rotated
138 chronospaces are generated using function `chronospace`, which creates an object of the same class,
139 and its associated printing and plotting functions summarize the results of the analysis in both graphical
140 and quantitative ways. The remainder of this article focuses on highlighting the use of the latter.

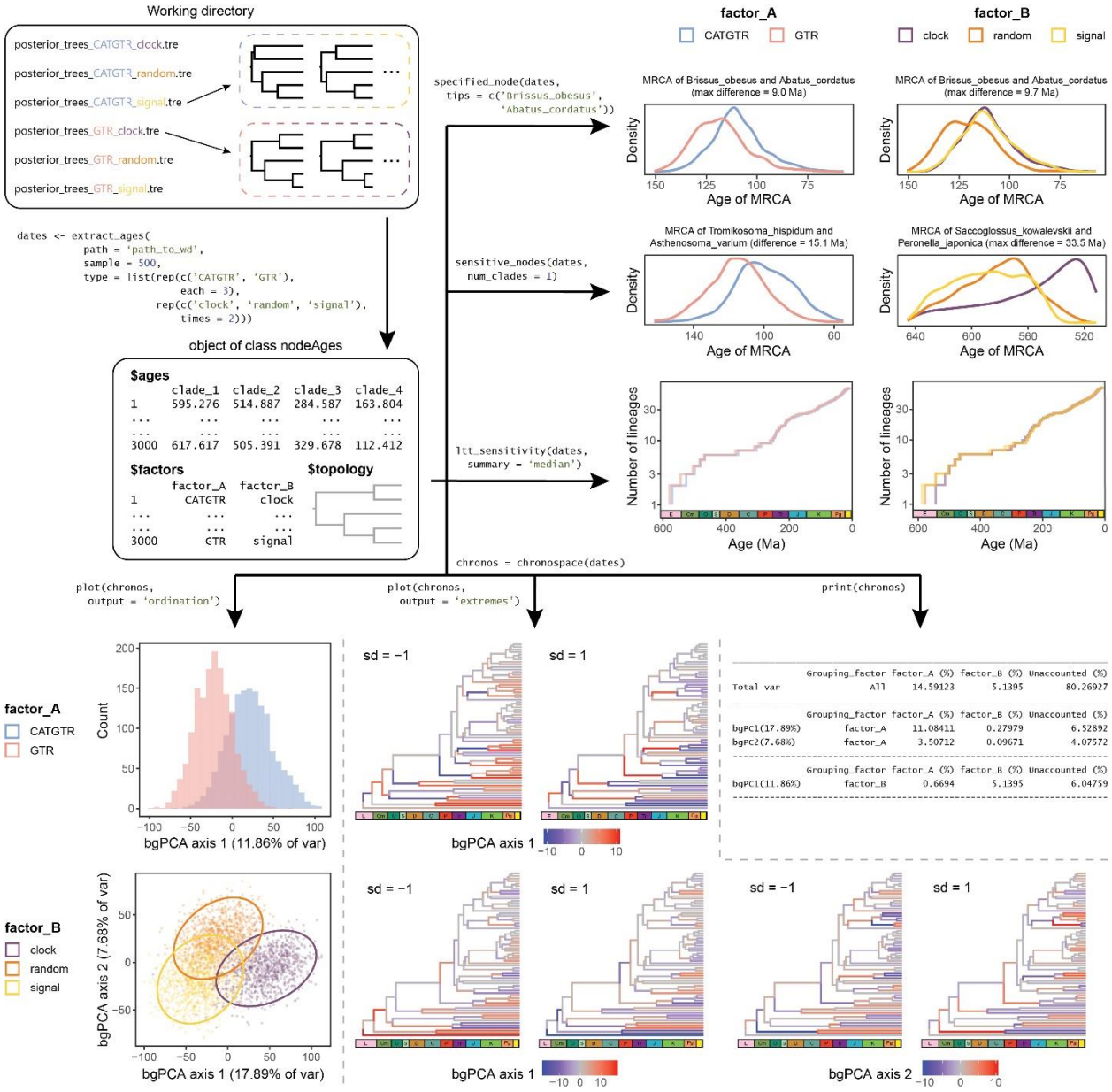

Figure 1: Structure of the R package *chronospace* and analytical workflow, represented using the data that accompanies the package. Populations of posterior topologies from time calibration analyses present in the working directory are imported, annotated, and stored as an object of class `nodeAges` 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

148 diversification (`ltt_sensitivity`). It also detects the nodes whose ages change the most between
149 analyses, and plots their posterior distributions (`sensitive_nodes`). These plots can optionally
150 include a chronostratigraphic scale (shown here in the lineage-through-time plot). Finally, a
151 `chronospace` object can be built and used to summarize the overall effect of methodological
152 decisions on the results. These can be visualized as chronospaces (which will be univariate or
153 multivariate, depending on the number of levels of each factor) and/or chronogram warping plots, which
154 depict the patterns of branch length contraction/expansion that are captured by the bgPCA axes.
155 Chronospace summaries can also printed in the form of a table reporting the partitioning of variance.

156

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

158 Three genome-scale datasets of protein-coding loci were obtained from the literature, originally
159 gathered to tackle the phylogenetic relationships and divergence times of Curculionoidea (65 terminals,
160 517 loci; Shin *et al.* 2018), Decapoda (90 terminals, 410 loci; Wolfe *et al.* 2019), and Eukaryota (136
161 terminals, 313 loci; Strassert *et al.* 2021). For each of these studies, we gathered a phylogenomic dataset
162 (including a concatenated matrix coded as amino acids and an associated partition file), a preferred tree
163 topology, and a list of node age calibrations derived from the fossil record. If studies included multiple
164 outgroups, only those forming part of the sister-group clade to the ingroup were retained. Studies were
165 selected so as to span widely differing evolutionary timespans (Figure 2).

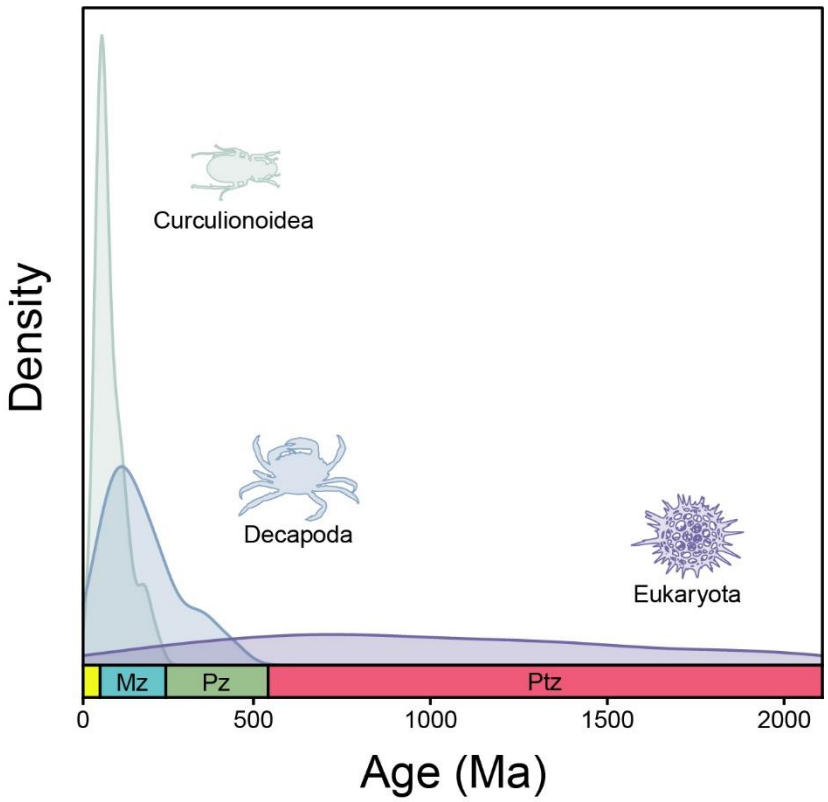

Figure 2: Geologic time in millions of years (Ma) spanned by the three empirical datasets analyzed. Density curves represent the distribution of node ages for each dataset. Ages were estimated as the median value of the distribution product of compounding the posterior trees sampled from all 40 analytical conditions used for dating (see main text). The diversification of weevils happened during the Cenozoic (unlabelled) and Mesozoic (Mz) eras, that of decapods further encompasses the Paleozoic (Pz), while node ages for the sampled eukaryotes clades fairly uniformly span the time between the present and the earlier part of the Proterozoic (Ptz) eon.

Divergence times were inferred using PhyloBayes v.4.1 (Lartillot *et al.* 2013) using a constrained tree topology. Four different methodological decisions were assessed: the type of clock, the model of molecular evolution, the prior distribution applied to calibrated nodes, and the approach used to

178 subsample loci. The first three of these are parameters specified when running the program. Divergence
179 times were obtained applying both an autocorrelated log-normal (`-ln`; Thorne *et al.* 1998) and an
180 uncorrelated gamma (`-ugam`; Drummond *et al.* 2006) clock; implementing unpartitioned site-
181 homogenous GTR + G (`-gtr`; Tavaré 1986) or site-heterogenous CAT + GTR + G (`-cat -gtr`; Lartillot &
182 Philippe 2004) models of molecular evolution; and using prior distributions of node ages that were
183 either uniformly distributed between stratigraphic bounds, or implemented a truncated Cauchy
184 minimum bound (`-lb`; Inoue *et al.* 2010). The tree prior was set to a birth-death process with free
185 parameters and soft calibration bounds were activated (Yang & Rannala 2006), allowing for 5% of prior
186 probability to lie outside of the specified bounds for calibrated nodes (`-bd -sb`). Default settings were
187 used for all other parameters.

188 Regarding the fourth and last decision evaluated (the set of loci used), phylogenomic datasets
189 were subsampled using strategies intended to achieve four distinct goals: A) minimize topological
190 conflict; B) improve levels of phylogenetic usefulness; C) enhance clock-like behavior; and D) maximize
191 the amount of information. For the first of these, loci chosen corresponded to those displaying the
192 minimum Robinson-Foulds (RF) distance (Robinson & Foulds 1981) between gene trees and the
193 constrained species tree. For the second, loci were chosen using the multivariate subsampling strategy
194 *genesortR* (Mongiardino Koch 2021; Mongiardino Koch & Thompson 2021), which identifies loci that
195 simultaneously display high phylogenetic signal and low levels of potential sources of bias. Clock-likeness
196 was estimated using the variance of root-to-tip distances displayed by gene trees. Given the positive
197 relationship between this metric and the rate of evolution of genes (Mongiardino Koch *et al.* 2022), loci
198 were first sorted by rate (estimated using the overall tree length; Telford *et al.* 2014), and both the
199 fastest- and slowest-evolving 15% of loci were excluded. Loci chosen corresponded to those with
200 minimum root-to-tip variances from among those showing relatively average evolutionary rates. Finally,
201 the loci with the highest levels of occupancy were chosen. Gene trees necessary to measure these

202 properties were inferred using ParGenes v.1.0.1 (Morel *et al.* 2018), using models that minimized the
203 Bayesian Information Criterion and estimating support with 100 bootstrap replicates. A fifth subsampled
204 dataset was derived by selecting loci at random, providing a means of estimating the overall effect
205 introduced by using targeted subsampling schemes. To standardize these different matrices in terms of
206 the overall amount of data contained, the loci necessary to first exceed a total length of 10,000 positions
207 were concatenated.

208 Dating was performed under all combinations of these four methodological decisions, resulting
209 in 40 conditions per dataset, each of which was run using two independent chains for 20,000
210 generations. The initial 25% of topologies were discarded as burn-in, and chains were thinned to every
211 two generations. Convergence was visually assessed using log-likelihood trace plots and comparing
212 median posterior estimates of node ages between replicate chains. After confirming convergence was
213 attained, time-calibrated trees from separate chains were combined into a single file. R code to
214 reproduce each of these steps is available as Supplementary Material S1. The remainder of analyses
215 were performed using package *chronospace*, summarizing each run using a random sample of 500
216 posterior trees.

217

218 **RESULTS**

219 Figure 3 summarizes the results obtained across the three empirical datasets used, showing the
220 median and 95% confidence interval (CI) of each inferred node age after combining the posterior
221 distributions of individual runs. Producing overall summaries of temporal information across populations
222 of chronograms, as depicted in both Figures 2 and 3, is straightforward once the data has been gathered
223 and formatted using the function `extract_ages`. For the ancient divergences estimated in the

224 eukaryote dataset, pooled 95% CI can span upwards of 1,000 Ma, as is the case for example of the last
225 common ancestor of Amebozoa (median: 1,291.2 Ma, 95% CI: 701.2-1,841.2 Ma). The temporal
226 uncertainty associated with younger divergences is much narrower in absolute terms, yet it can be even
227 larger in relative terms (i.e., after dividing the 95% CI width by its median value). For example, a handful
228 of divergences between closely related achelate decapods show 95% CI widths that are six to seven
229 times the value of the median estimate, with plausible dates spanning the last 142 Ma and falling
230 anywhere between the Quaternary and Early Cretaceous. Across all datasets, most nodes have 95% CI
231 widths that are larger than half their median estimate (see Fig. 3 insets).

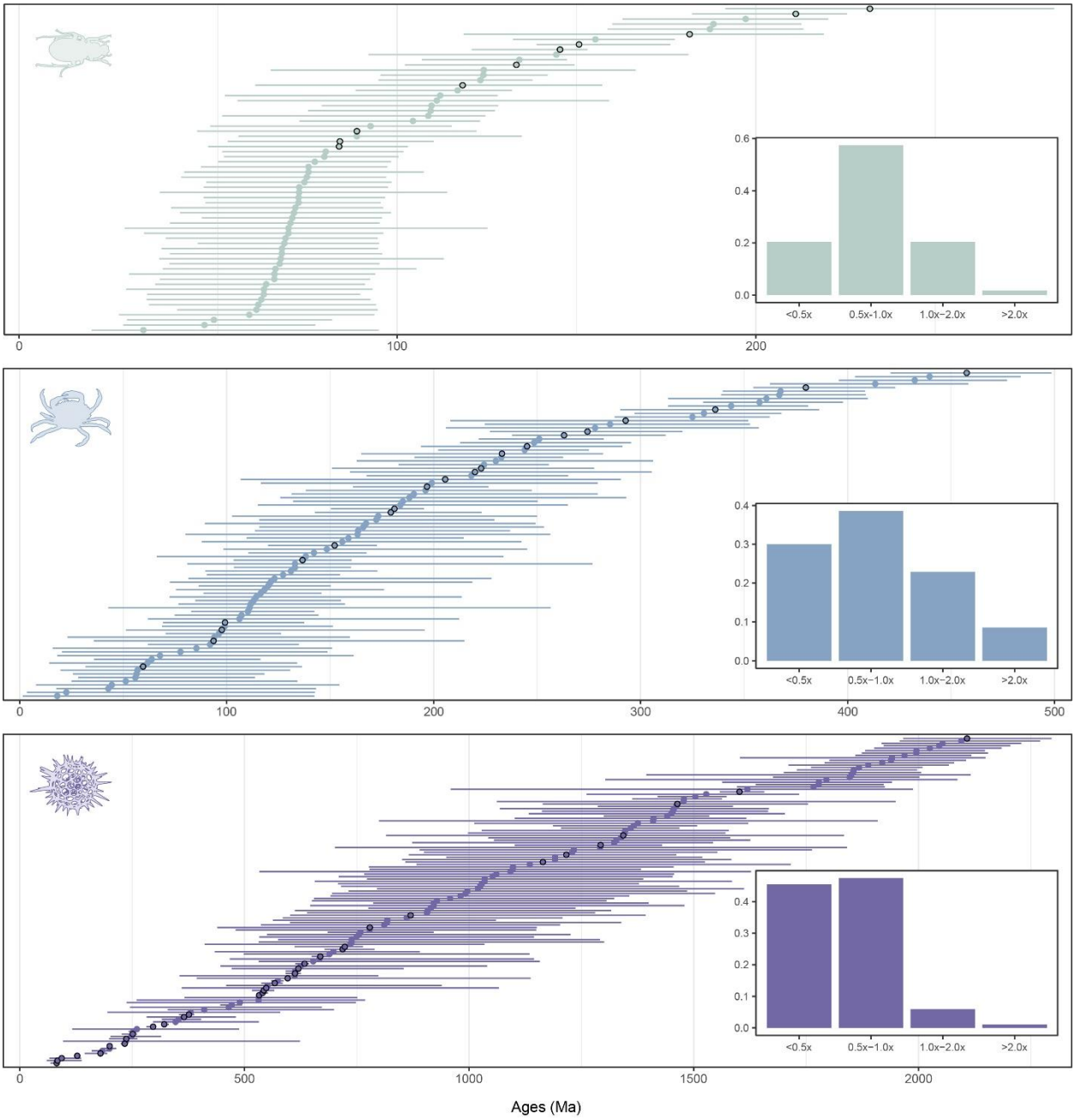

Figure 3: Uncertainty and sensitivity of node age estimates. Values shown are the median and width of pooled 95% CI per node across all analytical conditions. Prior probabilities were placed on the ages of nodes highlighted with black borders. Insets show the distribution of the relative widths of pooled 95% CIs, obtained by dividing absolute widths by the median values. Only unconstrained nodes were tallied.

The spread of posterior dates shown in Figure 3 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 4). 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 same set of decisions—namely, the type of molecular clock and the subsample of loci chosen—have bigger effect sizes, regardless of the timeframe of the phylogenetic question.

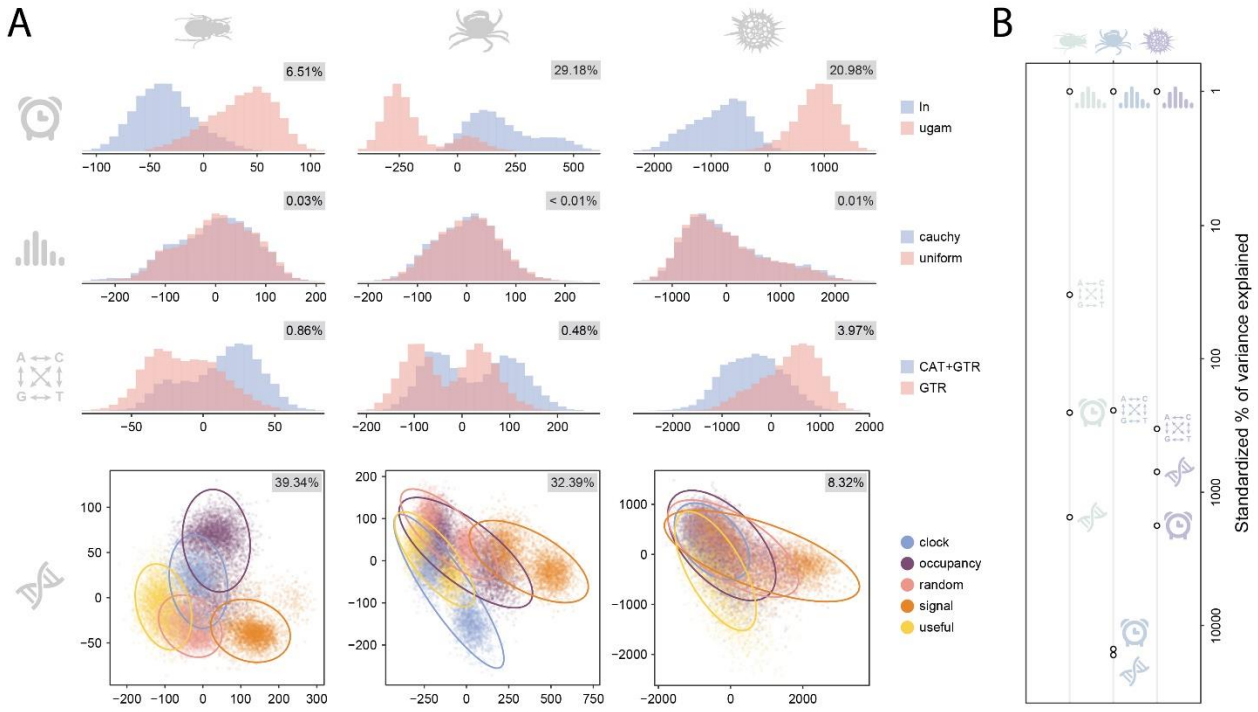

Figure 4: Sensitivity of divergence time estimates to methodological choices. **A.** Chronospaces (bgPC scores) for each factor, from top to bottom: type of molecular clock, prior distribution on constrained nodes, model of molecular evolution, and gene subsample. The proportion of total variance explained by

251 each factor is shown in grey boxes. For the loci subsample, only the first two axes (of four) are shown.
252 Ellipses include 95% of chronograms. **B.** Proportion of total variance explained by each factor, divided by
253 that of the least important factor for each dataset. Axis is log-scaled.

254

255 Across datasets, the most disparate chronograms were always obtained when loci were
256 subsampled based on phylogenetic signal (Fig. 5A), here defined as the loci supporting topologies with
257 the fewest differences to the species tree (i.e., lowest Robison-Foulds distance). This strategy also
258 resulted in the most dissimilar ages relative to those obtained through random loci selection, and largely
259 defined the first bgPCA axis (with either usefulness or clock-likeness in the opposite extreme). Employing
260 loci with high phylogenetic signal resulted in systematically longer terminal branch lengths, shorter
261 internal ones, and younger root ages (Fig. 5B).

262

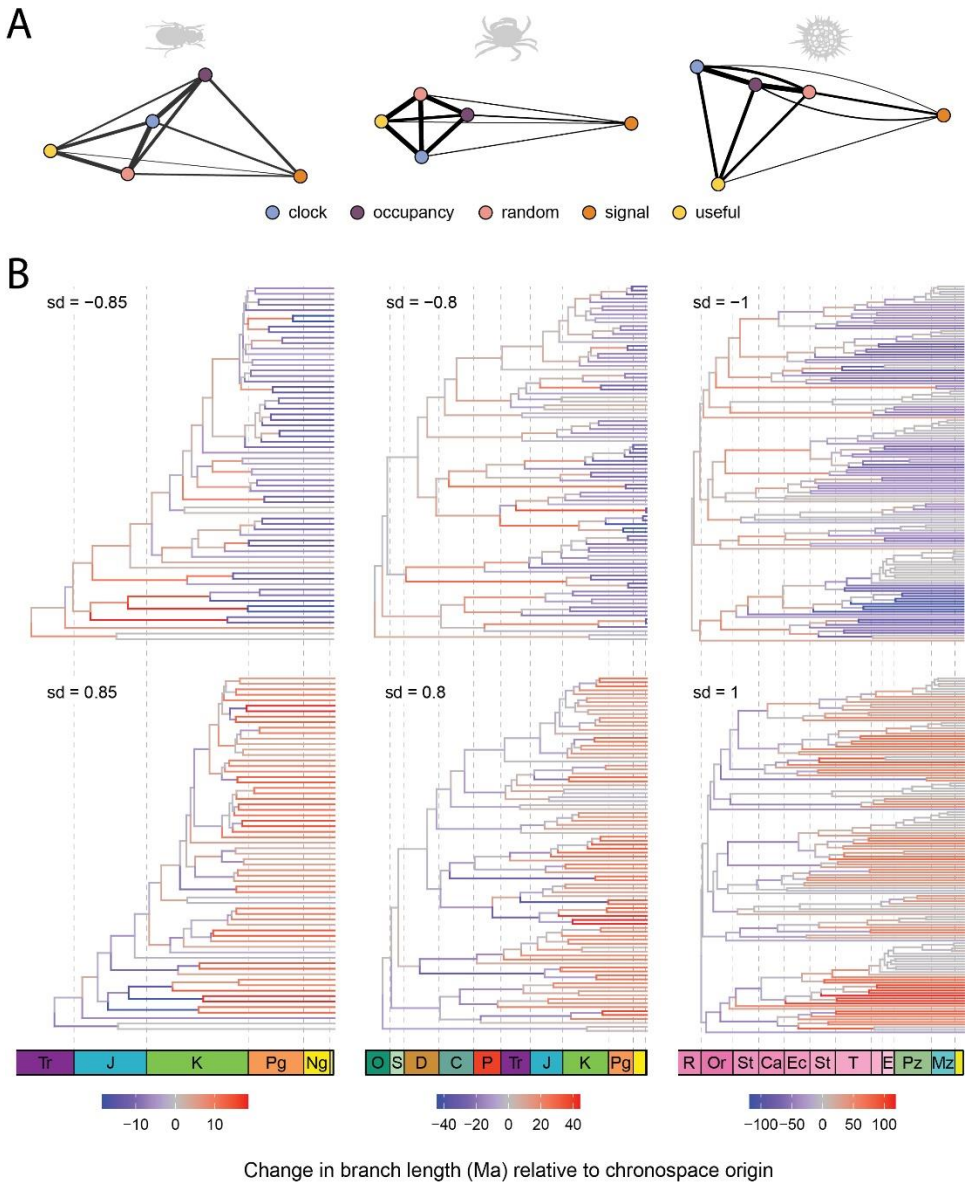

Figure 5: Effect of loci choice on inferred ages. **A.** Chronospaces built discriminating by loci subsampling strategy, as also shown in Fig. 4A. Circles depict the centroids of each distribution; the thickness of the lines connecting them is scaled to the inverse of the Euclidean distances separating 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 colored 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 warping induced by other factors is shown in Fig. SX.

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, 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).

Most of these methodological decisions are also hard to justify. Coupled with the high computational cost of analyses, empirical evolutionary timescales are generally inferred using one (or a few) among hundreds of possible parametric conditions, often times chosen arbitrarily. A literal reading

293 of the literature assessing the sensitivity of results to methodological factors suggests that every one of
294 these decisions can have large impacts on inferred dates (Hugall *et al.* 2007; Inoue *et al.* 2010; Duchêne
295 *et al.* 2014; Hirt *et al.* 2017; Barba-Montoya *et al.* 2018; Dos Reis *et al.* 2018; Smith *et al.* 2018;
296 Carruthers & Scotland 2021; Chen *et al.* 2021). Progress in pinpointing the magnitude of these effects
297 has been hampered by a general focus on a handful of nodes of interest rather than on tree-wide effects,
298 a scarcity of methods of visualization and summary, as well as of metrics that quantify the effect size of
299 individual decisions in a manner that is comparable, and a lack of testing of whether the relevance of
300 these choices scales with temporal depth.

301 Chronospaces provide novel ways of visualizing, quantifying, and exploring the sensitivity of
302 divergence time estimates, contributing to the inference of more robust evolutionary timescales
303 (Mongiardino Koch *et al.* 2022; Wolfe *et al.* 2022; Song *et al.* 2023). By representing chronograms as
304 collections of node ages, standard multivariate statistical approaches can be readily employed on
305 populations of Bayesian posterior timetrees. This representation allows for more nuanced visualizations
306 that can effectively convey the overall impact of different choices (Fig. 4A), the relative similarity of trees
307 obtained in different runs (Fig. 5A), and the patterns of branch length warping that are induced (Fig. 5B).
308 At the same time, partitioning the total variance into proportions explained by discrete factors using
309 sums of squares can result in tree-wide estimates of sensitivity with which to rank methodological
310 choices by their relative importance, helping prioritize computational resources. While the current
311 implementation is limited to analyses run under a constrained tree topology, chronospaces could
312 potentially be built using the subset of ages derived from nodes present in all (or most) of the posterior
313 trees of a simultaneous-inference analysis.

314 As shown in Fig. 4B, some decisions affect our reconstruction of diversification histories much
315 more than others, with effect sizes that vary by three to four orders of magnitude. This extreme

316 heterogeneity can help redirect the allocation of computational effort. For example, our results suggest
317 that empirical studies should either carefully validate their choice of evolutionary clock, or run
318 inferences that vary this component of the analysis, before deriving evolutionary insights (as shown
319 previously in both simulated and empirical studies; see Ho *et al.* 2005; Duchêne *et al.* 2014; Delsuc *et al.*
320 2018; Dos Reis *et al.* 2018; Mongiardino Koch *et al.* 2022; Song *et al.* 2023). Other factors can be
321 arbitrarily fixed without significant consequences, as is the case of the shape of prior distributions on
322 calibrated nodes. We note that the lack of an effect detected here applies exclusively to the use of the
323 Cauchy lower bound (Inoue *et al.* 2010) as implemented in PhyloBayes under default settings, and does
324 not mean that this factor is irrelevant when specifying other types of prior distributions or using other
325 software. Similarly, the use of complex models of substitution, represented here by the site-
326 heterogeneous CAT + GTR model (Lartillot & Philippe 2004), also seems to have relatively minor effects.
327 This robustness of divergence times to the choice of substitution model has been reported before (Du *et*
328 *al.* 2019; Tao *et al.* 2020; Mongiardino Koch *et al.* 2022), implying that the cost of using computationally
329 intensive models is not outweighed by the benefits. Notably, these conclusions are not altered by the
330 timeframe of the phylogenetic question, and apply equally to nodes of Cenozoic age as they do to
331 divergences that span the Proterozoic (Fig. 4B). Only two differences are observed as deeper timescales
332 are probed: there is a decrease in the impact of the loci subsample coupled with an increase in that of
333 the clock model (with both remaining the two most important factors overall); and the choice of
334 substitution model becomes slightly more relevant for extremely ancient divergences.

335 Loci subsampling has become a standard approach in phylogenomic dating, yet again, the effects
336 introduced by the each of numerous available approaches have not been fully characterized. While some
337 authors have advocated for the selection of clock-like genes in the hopes that reduced lineage-specific
338 rate variation will improve model fit and diminish bias, others have favored loci with less topological
339 conflict, as these are expected to harbor higher phylogenetic signal and be less affected by incomplete

340 lineage sorting (ILS). Simulations have confirmed that both of these subsampling schemes result in more
341 accurate divergence times (Smith *et al.* 2018; Carruthers *et al.* 2020; Carruthers *et al.* 2022). Other
342 sensible options include selecting genes based on phylogenetic usefulness (as estimated using
343 *genesortR*; Mongiardino Koch 2021), which should balance both aforementioned strategies by
344 discovering loci with increased clock-likeness and decreased topological incongruence; and maximizing
345 occupancy, increasing the amount of information contained within a matrix of a given size.

346 Our results show that loci choice is a major factor determining the divergence times obtained,
347 and that alternative schemes impart characteristic signatures on inferred chronograms at all
348 evolutionary scales. Minimizing topological conflict (i.e., maximizing signal) results in the most dissimilar
349 timetrees, as captured in the chronospaces by exhibiting the largest distances to those inferred with
350 randomly chosen loci, as well as the largest average distances to all other strategies (Fig. 5A).
351 Accordingly, this variation is captured by the major bgPCA axis, with clock-likeness and/or usefulness
352 occupying the opposite end. Selecting the most topologically congruent trees results in systematically
353 longer terminal branch lengths, shorter internal ones (especially for those branches subtending species
354 pairs, or cherries), and slightly younger root ages (Fig. 5B), while the opposite is true for chronograms
355 based on high usefulness/clock-likeness loci. Which of these is closer to the truth is beyond the scope of
356 empirical work; nonetheless, those studying the diversification of clades should be aware of the
357 systematic effects that result from using a given subsampling scheme. Surprisingly, the pattern seen here
358 is opposite to that reported in the simulation of Carruthers *et al.* (2022), in which the selection of
359 congruent topologies was tied to an expansion of internal branch lengths and a reduction of terminal
360 ones. This is in fact the expected behavior if topological incongruence is driven by biological factors such
361 as ILS (Mendes & Hahn 2016). While some have argued that deep recalcitrant nodes are linked to
362 ancient events of ILS (e.g., Cai *et al.* 2021; Hime *et al.* 2021; Mongiardino Koch *et al.* 2023; Song *et al.*
363 2023), it has also been suggested that topological incongruence at deep timescales is more likely to be

caused by methodological issues related to loss of signal and model inadequacy (Gatesy & Springer 2014; Richards *et al.* 2018). While our analyses were not designed to test this phenomenon, our results are contrary to those expected if ILS was the main driver of topological conflict at ancient nodes.

DATA ACCESSIBILITY

The *chronospace* package is available for installation from the GitHub repository https://github.com/mongiardino/chronospace. Empirical datasets and results can be found in the Dryad repository …

FUNDING

This study was supported by NSF grant DEB-2036186 awarded to Greg W. Rouse.

AUTHOR CONTRIBUTIONS

NMK designed the study, performed all analyses, and wrote the manuscript. PMC built the *chronospace* package from functions written by NMK, and edited the text. Both authors approved the final version.

REFERENCES

Álvarez-Carretero, S. & dos Reis, M. (2020) Bayesian phylogenomic dating. In S. Y. W. Ho (Ed.), *The Molecular Evolutionary Clock: Theory and Practice* (pp. 221-249). Springer.

Barba-Montoya, J., Sharma, S. & Kumar, S. (2023) Molecular timetrees using relaxed clocks and uncertain phylogenies. *Frontiers in Bioinformatics,* **3**, 10.3389/fbinf.2023.1225807.

Barba-Montoya, J., Dos Reis, M., Schneider, H., Donoghue, P.C. & Yang, Z. (2018) Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a Cretaceous Terrestrial Revolution. *New Phytologist,* **218,** 819-834.

Budd, G.E. & Mann, R.P. (2023) Two notorious nodes: a critical examination of relaxed molecular clock age estimates of the bilaterian animals and placental mammals. *Systematic Biology***,** syad057, 10.1093/sysbio/syad057.

Cai, L., Xi, Z., Lemmon, E.M., Lemmon, A.R., Mast, A., Buddenhagen, C.E., Liu, L. & Davis, C.C. (2021) The perfect storm: gene tree estimation error, incomplete lineage sorting, and ancient gene flow explain the most recalcitrant ancient angiosperm clade, Malpighiales. *Systematic Biology,* **70,** 491-507.

Cardini, A., O'Higgins, P. & Rohlf, F.J. (2019) Seeing distinct groups where there are none: spurious patterns from between-group PCA. *Evolutionary Biology,* **46,** 303-316.

Carruthers, T., Sanderson, M.J. & Scotland, R.W. (2020) The implications of lineage-specific rates for divergence time estimation. *Systematic Biology,* **69,** 660-670.

Carruthers, T. & Scotland, R.W. (2021) Uncertainty in divergence time estimation. *Systematic Biology*, **70**, 858-861.

Carruthers, T., Sun, M., Baker, W.J., Smith, S.A., de Vos, J.M. & Eiserhardt, W.L. (2022) The implications of incongruence between gene tree and species tree topologies for divergence time estimation. *Systematic Biology,* **71,** 1124-1146.

Chen, D., Hosner, P.A., Dittmann, D.L., O'Neill, J.P., Birks, S.M., Braun, E.L. & Kimball, R.T. (2021) Divergence time estimation of Galliformes based on the best gene shopping scheme of ultraconserved elements. *BMC Ecology and Evolution,* **21,** 209.

Crotty, S.M. & Holland, B.R. (2022) Comparing partitioned models to mixture models: Do information criteria apply? *Systematic Biology,* **71,** 1541-1548.

Culhane, A.C., Perriere, G., Considine, E.C., Cotter, T.G. & Higgins, D.G. (2002) Between-group analysis of microarray data. *Bioinformatics,* **18,** 1600-1608.

Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M.-K., Turon, X., Lopez-Legentil, S., Piette, J., Lemaire, P. & Douzery, E.J. (2018) A phylogenomic framework and timescale for comparative studies of tunicates. *BMC Biology,* **16,** 39.

Dolédec, S. & Chessel, D. (1987) Rythmes saisonniers et composantes stationnelles en milieu aquatique. I: Description d'un plan d'observation complet par projection de variables. *Acta Oecologica. Oecologia Generalis,* **8,** 403-426.

Dolédec, S. & Chessel, D. (1989) Rythmes saisonniers et composantes stationnelles en milieu aquatique. II: Prise en compte et élimination d'effets dans un tableau faunistique. *Acta Oecologica. Oecologia Generalis,* **10,** 207-232.

Donoghue, P.C. & Benton, M.J. (2007) Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends in Ecology & Evolution,* **22,** 424-431.

Dos Reis, M., Gunnell, G.F., Barba-Montoya, J., Wilkins, A., Yang, Z. & Yoder, A.D. (2018) Using phylogenomic data to explore the effects of relaxed clocks and calibration strategies on divergence time estimation: primates as a test case. *Systematic Biology,* **67,** 594-615.

Dos Reis, M., Thawornwattana, Y., Angelis, K., Telford, M.J., Donoghue, P.C. & Yang, Z. (2015) Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. *Current Biology,* **25,** 2939-2950.

428 Dos Reis, M. & Yang, Z. (2019) Bayesian molecular clock dating using genome-scale datasets. In M.
429 Anisimova (Ed.), *Evolutionary genomics. Methods in molecular biology, vol. 1910* (pp. 309-330).
430 Humana.
431 Douglas, J., Jiménez-Silva, C.L. & Bouckaert, R. (2022) StarBeast3: adaptive parallelized Bayesian
432 inference under the multispecies coalescent. *Systematic Biology,* **71,** 901-916.
433 Drummond, A.J., Ho, S.Y., Phillips, M.J. & Rambaut, A. (2006) Relaxed phylogenetics and dating with
434 confidence. *PLoS Biology,* **4,** e88.
435 Du, Y., Wu, S., Edwards, S.V. & Liu, L. (2019) The effect of alignment uncertainty, substitution models and
436 priors in building and dating the mammal tree of life. *BMC Evolutionary Biology,* **19,** 203.
437 Duchêne, S., Lanfear, R. & Ho, S.Y. (2014) The impact of calibration and clock-model choice on molecular
438 estimates of divergence times. *Molecular Phylogenetics and Evolution,* **78,** 277-289.
439 Eme, L., Sharpe, S.C., Brown, M.W. & Roger, A.J. (2014) On the age of eukaryotes: evaluating evidence
440 from fossils and molecular clocks. *Cold Spring Harbor Perspectives in Biology,* **6,** a016139.
441 Foster, P.G., Schrempf, D., Szöllősi, G.J., Williams, T.A., Cox, C.J. & Embley, T.M. (2023) Recoding amino
442 acids to a reduced alphabet may increase or decrease phylogenetic accuracy. *Systematic Biology,*
443 **72,** 723-737.
444 Frost, S.D., Pybus, O.G., Gog, J.R., Viboud, C., Bonhoeffer, S. & Bedford, T. (2015) Eight challenges in
445 phylodynamic inference. *Epidemics,* **10,** 88-92.
446 Gatesy, J. & Springer, M.S. (2014) Phylogenetic analysis at deep timescales: unreliable gene trees,
447 bypassed hidden support, and the coalescence/concatalescence conundrum. *Molecular*
448 *Phylogenetics and Evolution,* **80,** 231-266.
449 Gearty, W. (2022) deeptime: plotting tools for anyone working in deep time. R package version 0.2.2,
450 https://CRAN.R-project.org/package=deeptime.

451

Heath, T.A., Huelsenbeck, J.P. & Stadler, T. (2014) The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences,* **111,** E2957-E2966.

Hillis, D.M., Heath, T.A. & John, K.S. (2005) Analysis and visualization of tree space. *Systematic Biology,* **54,** 471-482.

Hime, P.M., Lemmon, A.R., Lemmon, E.C.M., Prendini, E., Brown, J.M., Thomson, R.C., Kratovil, J.D., Noonan, B.P., Pyron, R.A. & Peloso, P.L. (2021) Phylogenomics reveals ancient gene tree discordance in the amphibian tree of life. *Systematic Biology,* **70,** 49-66.

Hirt, M.V., Arratia, G., Chen, W.-J., Mayden, R.L., Tang, K.L., Wood, R.M. & Simons, A.M. (2017) Effects of gene choice, base composition and rate heterogeneity on inference and estimates of divergence times in cypriniform fishes. *Biological Journal of the Linnean Society,* **121,** 319-339.

Ho, S.Y., Phillips, M.J., Drummond, A.J. & Cooper, A. (2005) Accuracy of rate estimation using relaxed-clock models with a critical focus on the early metazoan radiation. *Molecular Biology and Evolution,* **22,** 1355-1363.

Howard, R.J., Giacomelli, M., Lozano-Fernandez, J., Edgecombe, G.D., Fleming, J.F., Kristensen, R.M., Ma, X., Olesen, J., Sørensen, M.V. & Thomsen, P.F. (2022) The Ediacaran origin of Ecdysozoa: integrating fossil and phylogenomic data. *Journal of the Geological Society,* **179,** jgs2021-107.

Hugall, A.F., Foster, R. & Lee, M.S. (2007) Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. *Systematic Biology,* **56,** 543-563.

Inoue, J., Donoghue, P.C. & Yang, Z. (2010) The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. *Systematic Biology,* **59,** 74-89.

Kassambara, A. (2020) ggpubr: 'ggplot2' based publication ready plots. R package version 0.4.0, https://CRAN.R-project.org/package=ggpubr.

476 Kumar, S. & Hedges, S.B. (2016) Advances in time estimation methods for molecular data. *Molecular*
477 *Biology and Evolution,* **33,** 863-869.
478 Lartillot, N. (2023) Identifying the best approximating model in Bayesian phylogenetics: Bayes factors,
479 cross-validation or wAIC? *Systematic Biology,* **72,** 616-638.
480 Lartillot, N. & Philippe, H. (2004) A Bayesian mixture model for across-site heterogeneities in the amino-
481 acid replacement process. *Molecular Biology and Evolution,* **21,** 1095-1109.
482 Lartillot, N., Rodrigue, N., Stubbs, D. & Richer, J. (2013) PhyloBayes MPI. Phylogenetic reconstruction
483 with infinite mixtures of profiles in a parallel environment. *Systematic Biology,* **62,** 611-615.
484 Li, Y., Shen, X.-X., Evans, B., Dunn, C.W. & Rokas, A. (2021) Rooting the animal tree of life. *Molecular*
485 *Biology and Evolution,* **38**, 4322-4333.
486 Lozano-Fernandez, J. (2022) A practical guide to design and assess a phylogenomic study. *Genome*
487 *Biology and Evolution,* **14,** evac129.
488 May, M.R. & Rothfels, C.J. (2023) Diversification models conflate likelihood and prior, and cannot be
489 compared using conventional model-comparison tools. *Systematic Biology,* **72,** 713-722.
490 Mendes, F.K. & Hahn, M.W. (2016) Gene tree discordance causes apparent substitution rate variation.
491 *Systematic Biology,* **65,** 711-721.
492 Mongiardino Koch, N. (2021) Phylogenomic subsampling and the search for phylogenetically reliable loci.
493 *Molecular Biology and Evolution,* **38**, 4025-4038.
494 Mongiardino Koch, N. & Thompson, J.R. (2021) A total-evidence dated phylogeny of Echinoidea
495 combining phylogenomic and paleontological data. *Systematic Biology,* **70,** 421-439.
496 Mongiardino Koch, N., Thompson, J.R., Hiley, A.S., McCowin, M.F., Armstrong, A.F., Coppard, S.E.,
497 Aguilera, F., Bronstein, O., Kroh, A. & Mooi, R. (2022) Phylogenomic analyses of echinoid
498 diversification prompt a re-evaluation of their fossil record. *eLife,* **11,** e72460.

Mongiardino Koch, N., Tilic, E., Miller, A.K., Stiller, J. & Rouse, G.W. (2023) Confusion will be my epitaph: genome-scale discordance stifles phylogenetic resolution of Holothuroidea. *Proceedings of the Royal Society B,* **290,** 20230988.

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.

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.

Richards, E.J., Brown, J.M., Barley, A.J., Chong, R.A. & Thomson, R.C. (2018) Variation across mitochondrial gene trees provides evidence for systematic error: how much gene tree variation is biological? *Systematic Biology,* **67,** 847-860.

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.

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.

546 Tan, G., Muffato, M., Ledergerber, C., Herrero, J., Goldman, N., Gil, M. & Dessimoz, C. (2015) Current
547 methods for automated filtering of multiple sequence alignments frequently worsen single-gene
548 phylogenetic inference. *Systematic Biology,* **64,** 778-791.
549 Tao, Q., Barba-Montoya, J., Huuki, L.A., Durnan, M.K. & Kumar, S. (2020) Relative efficiencies of simple
550 and complex substitution models in estimating divergence times in phylogenomics. *Molecular*
551 *Biology and Evolution,* **37,** 1819-1831.
552 Tarver, J.E., Dos Reis, M., Mirarab, S., Moran, R.J., Parker, S., O'Reilly, J.E., King, B.L., O'Connell, M.J.,
553 Asher, R.J. & Warnow, T. (2016) The interrelationships of placental mammals and the limits of
554 phylogenetic inference. *Genome Biology and Evolution,* **8,** 330-344.
555 Tavaré, S. (1986) Some probabilistic and statistical problems on the analysis of DNA sequence. *Lecture of*
556 *Mathematics for Life Science,* **17,** 57.
557 Telford, M.J., Lowe, C.J., Cameron, C.B., Ortega-Martinez, O., Aronowicz, J., Oliveri, P. & Copley, R.R.
558 (2014) Phylogenomic analysis of echinoderm class relationships supports Asterozoa. *Proceedings*
559 *of the Royal Society B,* **281,** 20140479.
560 Thorne, J.L., Kishino, H. & Painter, I.S. (1998) Estimating the rate of evolution of the rate of molecular
561 evolution. *Molecular Biology and Evolution,* **15,** 1647-1657.
562 Warnock, R.C. & Wright, A.M. (2020) *Understanding the tripartite approach to Bayesian divergence time*
563 *estimation*. Cambridge University Press.
564 Warnock, R.C., Yang, Z. & Donoghue, P.C. (2011) Exploring uncertainty in the calibration of the molecular
565 clock. *Biology Letters*, **8,** 156-159.
566 Wickham, H. (2017) tidyverse: Easily install and load 'tidyverse' packages. R package version 1.2.1.
567 https://CRAN.R-project.org/package=tidyverse.
568 Wolfe, J.M., Ballou, L., Luque, J., Watson-Zink, V.M., Ahyong, S.T., Barido-Sottani, J., Chan, T.-Y., Chu, K.H.,
569 Crandall, K.A., Daniels, S.R. Felder, D.L., Mancke, H., Martin, J.W., Ng, P.K.L., Ortega-Hernández, J.,

570 Palacios Theil, E., Pentcheff, N.D., Robles, R., Thoma, B.P., Tsang, L.M., Wetzer, R., Windsor, A.M.,
571 & Bracken-Grissom, H.D. (2022) Convergent adaptation of true crabs (Decapoda: Brachyura) to a
572 gradient of terrestrial environments. *Systematic Biology*, syad066, 10.1093/sysbio/syad066.

573 Wolfe, J.M., Breinholt, J.W., Crandall, K.A., Lemmon, A.R., Lemmon, E.M., Timm, L.E., Siddall, M.E. &
574 Bracken-Grissom, H.D. (2019) A phylogenomic framework, evolutionary timeline and genomic
575 resources for comparative studies of decapod crustaceans. *Proceedings of the Royal Society B,*
576 **286,** 20190079.

577 Yang, Z. & Rannala, B. (2006) Bayesian estimation of species divergence times under a molecular clock
578 using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution,* **23,** 212-226.

579 Yendle, P.W. & MacFie, H.J. (1989) Discriminant principal components analysis. *Journal of Chemometrics,*
580 **3,** 589-600.

581 Yu, G., Smith, D.K., Zhu, H., Guan, Y. & Lam, T.T.Y. (2017) ggtree: an R package for visualization and
582 annotation of phylogenetic trees with their covariates and other associated data. *Methods in*
583 *Ecology and Evolution,* **8,** 28-36.

584 Zhou, Y., Rodrigue, N., Lartillot, N. & Philippe, H. (2007) Evaluation of the models handling heterotachy in
585 phylogenetic inference. *BMC Evolutionary Biology,* **7,** 206.

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

588