Potential survival of some, but not all, diversification methods

1 **Potential survival of some, but not all, diversification methods**

2

3 **Abstract**

4 Models have long been used for understanding changing diversification patterns over time. The

5 rediscovery that models with very different rates through time can fit a phylogeny equally well

6 has led to great concern about the use of these models. We share and add to these concerns: even

7 with time heterogeneous models without these issues, the distribution of the data means that

8 estimates will be very uncertain. However, we argue that congruence issues such as this also

9 occur in models as basic as Brownian motion and coin flipping. Taxon-heterogeneous models

10 such as many SSE models appear not to have this particular issue.

11

12

13 **Keywords:** diversification, congruence, likelihood, identifiability

14

15

16

17

18

19

20

21

22

23

Introduction

For decades, molecular phylogenies have served as vital sources of historical information for deciphering the birth and the death dynamics of lineages. Thousands of studies of molecular phylogenies have been dedicated to investigating diversification. In theory, estimating constant birth and death rates separately is possible because each has distinguishable effects on the tree shape and branch length distributions (Nee et al. 1994). There are several extensions that expand this simple model for characterizing diversification as a function of time or diversity (e.g., Nee et al. 1992; Rabosky 2006, 2009; Bokma 2008; Rabosky and Lovette 2008; Morlon et al. 2011; Etienne et al. 2012), which are used to reconstruct lines showing speciation and extinction rates scrolling into the past, like the pen of a seismometer tracking vibrations through time. A sudden sweep up of the extinction rate arm could mean a mass extinction. A slow, downward trajectory of the speciation rate arm as time approaches the present could mean available niches have become filled up, limiting the possibilities of adding new species. And, as with constant rate birth-death models, we have been working under the assumption that even tiny changes in speciation and/or extinction through time should leave distinct signatures on the tree shape and branching structure in a molecular phylogeny.

In a recent paper by Louca and Pennell (2020), the entire enterprise of estimating diversification rates, at least from molecular phylogenies alone, has been called into question. As it turns out, for any given phylogeny there are an infinite array of congruent models each having unique functions of speciation and/or extinction rates smoothly varying through time. This is based on the property of both constant rate birth-death and time-varying models in which every lineage at any given time-point experiences the same rates, and so sampling times for either a speciation or extinction event are drawn from the same distribution (also known as a coalescent

47   point process or CPP; see Lambert and Stadler, 2013). Under such conditions, the likelihood of a

48   tree under a given birth-death model can be inferred simply in terms of the lineage-through-time

49   (LTT) curve, which is a retrospective counting of the number of lineages that led to a set of

50   species observed today, and there are always multiple qualitatively different models that can

51   produce the same curves with the same probability. For example, one model may infer the

52   observed diversity of Cetaceans (i.e., whales, dolphins, and relatives) is a product of dramatic

53   changes in the rate of speciation and extinction rates over time, whereas another, *equally likely*

54   model, may infer modern whale diversity is the product of no extinction and ever so slight

55   changes to the speciation rate. In other words, two diametrically opposed models, particularly

56   with regards to the role of extinction, provide *equally* valid explanations for the mode and tempo

57   of Cetacean diversification. In some cases, such as our example above, these models will have

58   the same number of parameters, rendering them truly indistinguishable.

59   It should come as no surprise, then, that one popular interpretation of these findings is

60   that any attempt to learn anything about diversification rates from molecular phylogenies is a

61   completely futile enterprise. A different response, which we also have seen, is the continued and

62   uncritical use of these suspect methods sanitized with a "but see Louca and Pennell (2020)"

63   citation. It is also worth noting that the findings of Louca and Pennell (2020) are substantially

64   similar, though much more detailed, to the ones presented by Kubo and Iwasa (1995) a quarter

65   century ago. These authors also described an infinite array of birth and death models fitting the

66   data equally well, which has been effectively ignored by most later workers.

67   The issues raised by Louca and Pennell (2020) and Kubo and Iwasa (1995) do represent

68   substantial methodological problems for comparative biology. However, this does not signal the

69   end of studying diversification rates on molecular phylogenies, as some have claimed, as these

70 problems do not necessarily extend to *all* models of diversification. Instead, they are, so far,

71 shown in situations where the goal is to interpret diversification rates through time using what

72 we refer to as, "time-varying, lineage homogeneous" models — again, models in which all

73 lineages experience the same variable rates at any given point in time. These would be analogous

74 to a non-heritable trait-dependent process (Lambert and Stadler, 2013), where changes in a trait

75 occur the same in all species independently (e.g., global CO_2, sea-level changes, global

76 temperature patterns). We argue that what we refer to as "lineage-specific heterogeneous"

77 models, in which rates vary among lineages across time points, perhaps due to the inheritance of

78 a trait (e.g., state-speciation and extinction, or SSE models; Maddison et al. 2007), should be

79 immune to the issues of identifiability raised above. This comes with the substantial caveat that

80 this is true if, and probably only if, the heritable rate changes are modeled as containing a single

81 speciation and extinction rate that do not vary through time. Essentially, we will show that these

82 models do "work" if we limit the model space to those with single rates at any time point.

83   We also address some of the other procedures proposed, explicitly or implicitly, by

84 Louca and Pennell (2020): continuing with pulled diversification rate reconstruction, focusing on

85 a point estimate only, no longer penalizing for model complexity, and how information is

86 distributed on trees.

87   Overall, we make four points:

88   1. Model congruence can occur in areas as different as coin flipping and Brownian

89   motion: it does not mean these models must be given up, only that certain

90   questions are infeasible.

91   2. Time-varying, lineage homogeneous models that use information that could be

92   reduced to a lineage through time curve (even if the input is a full phylogeny) to

93   estimate changing speciation, extinction, diversification, turnover, or extinction

94   fraction should be avoided due to congruence issues.

95   3. Pulled speciation and pulled diversification rate analyses (Louca and Pennell

96   2020) are identifiable, but they fail to incorporate the substantial uncertainty in

97   reconstructions that come as a result of typically exponentially decreasing number

98   of data points (lineages) as one approaches the root of a tree (this also plagues the

99   methods in point 2)

100   4. Some SSE methods, and likely other methods that investigate heterogeneity

101   across taxa, use information beyond that in a lineage through time curve and their

102   utility remains intact in the face of Louca and Pennell (2020) and Kubo and Iwasa

103   (1995).

104

105 **Model congruence is common**

106   It may come as a surprise that this issue of two models fitting data equally well is not new

107 to comparative methods. Take, for instance, the inference of evolutionary trends, which, broadly

108 defined, are identifiable patterns of trait evolution in a given direction through time. Using only

109 extant species, can we detect horses getting bigger and with fewer digits, or increases in the

110 mean seed size in flowering plants since the Cretaceous (e.g., Tifney 1984; Eriksson et al. 2000),

111 or, more generally, uncover an evolutionary arms race between predator and prey (e.g., Dawkins

112 and Krebs 1979; Abrams 1986)? It is trivial to extend a simple Brownian motion model to

113 include a parameter that allows for the focal trait to evolve along a trend, and this is available in

114 popular software like the R package *geiger* (Pennell et al. 2014). The likelihood for these models

115 given the data is finite, and the simple no trend model is even nested within the trend model, so

116   comparisons between the two are straightforward. However, as Felsenstein (1988) and Hansen

117   and Martins (1996) have pointed out, even though trait values move in a given direction under a

118   Brownian motion with a trend model, this does not affect the expected covariances among

119   species trait values. That is, the expected trait differences among species is still linearly

120   dependent on time, meaning closely related species are still expected to be more phenotypically

121   similar than more distantly related species, which is an identical assumption under a standard

122   Brownian motion model. Consequently, the two models have identical likelihoods when fitted to

123   extant species only, making them indistinguishable based on their probability alone.

124   One might argue that in cases of clear non-identifiability any careful scientist would

125   avoid fitting a degenerate model such as Brownian motion with a trend with just coeval

126   terminals. However, the problem of identifiability between Brownian motion models with and

127   without a trend is further compounded when considering the potential for dramatic effects on

128   ancestral state reconstructions. For example, the ancestor of a clade of taxa with body sizes

129   ranging from 10-12 kg might have a reconstructed state near 11 kg under a no trend model but

130   could have a reconstructed state of 50 kg under a model with a trend of an incremental trait

131   decrease through time. Such ancestral state reconstruction remains widely popular. Nevertheless,

132   it is still a rather large leap to assert that, because these models are unidentifiable, models using

133   Brownian motion are generally invalid for use on trees containing only modern taxa. We can still

134   compare Brownian motion models with more complex models, such as Ornstein-Uhlenbeck

135   models (e.g., Butler and King 2004; Beaulieu et al. 2012), Brownian models with more than one

136   rate (e.g., O'Meara et al. 2006; Thomas et al. 2006), or models where the Brownian motion rate

137   itself changes over time (e.g., Revell, 2021). In other words, while Brownian motion with a trend

138   model is unidentifiable with modern taxa only, we would not, for instance, say that any model
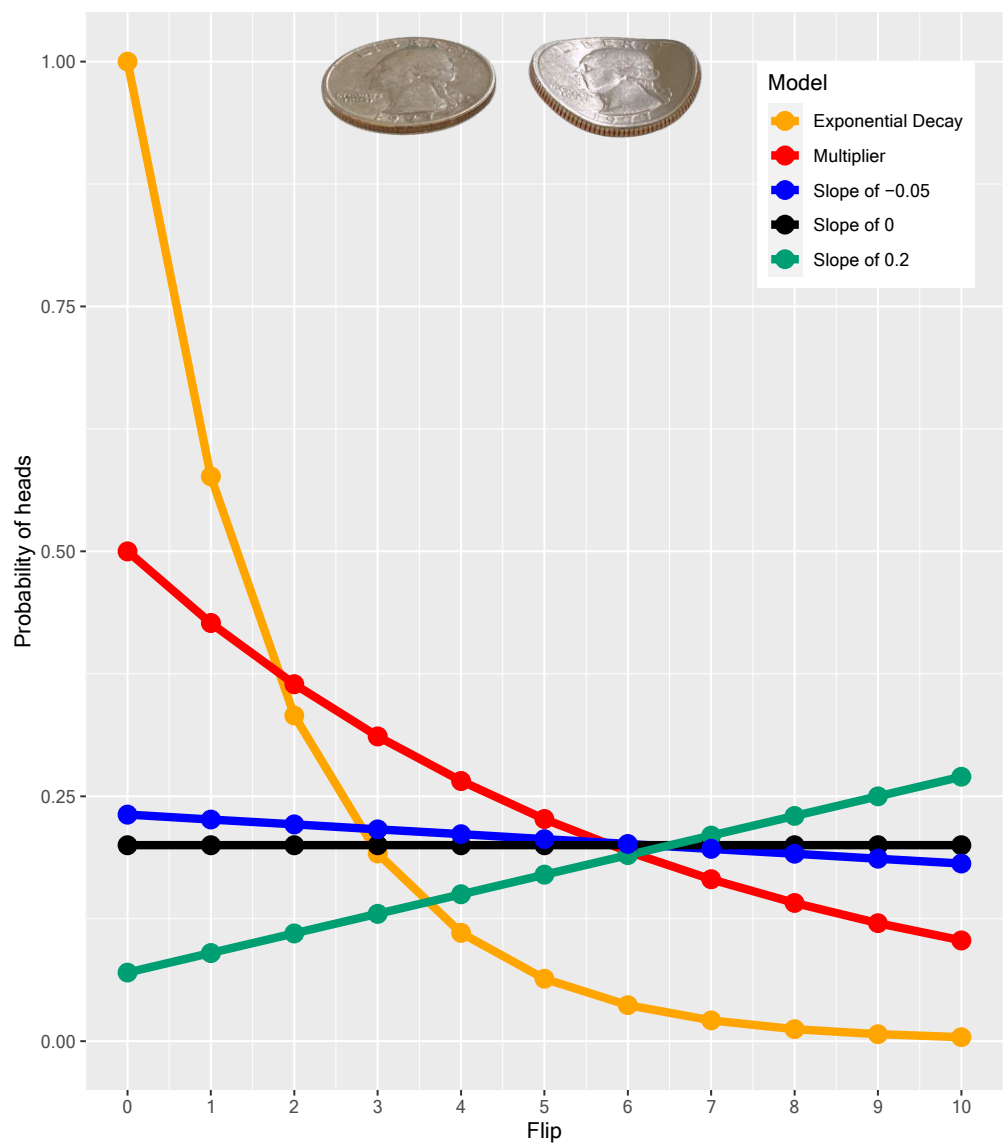
139   **Figure 1**: Probability of heads per flip on different models of coin flipping. Each of these models
140   can fit the same dataset of two heads, eight tails with equal likelihood but make very different
141   predictions about the next flip.

142

143   that attempts to estimate rates of evolution on such trees is uninterpretable. Some models in this

144   space give the same likelihoods and cannot be distinguished, but many others can, which calls

145   for care and analysis, not panic.

146   We also point out that model congruence occurs in other statistically based disciplines.

147   Consider the classic coin-flipping example. Suppose we toss a coin 10 times, and 2 of those

148   tosses come up heads. The most straightforward fitted binomial model indicates that the

149   probability of observing 2 heads in 10 flips is 0.3 for a biased coin with each flip having a 20%

150   chance of landing on heads. Now suppose that every time we touch the coin, it gets slightly

151   dented, or a bit of metal is worn away, and it becomes less and less likely to land on one side

152   than the other. We can devise several models that have different slopes to alter the probability of

153   heads after a set of coin flips (Figure 1). For instance, the probability of heads can linearly

154   increase with each flip, such that by the end the probability of heads is 10% higher than when we

155   started flipping, and a model where the probability of heads decreases with each flip so that by

156   the end it is 5% lower than when it started (Figure 1). Interestingly, the probability of observing

157   2 heads in 10 flips of the coin in each of these models is the *same* as the simple binomial model,

158   though the linear change models infer different initial probability of heads before any flips are

159   made as well as what the probability of the next flip being heads is. If we pre-set the 5% lower or

160   20% higher parameters ahead of time rather than fitting them, these have the same number of

161   free parameters as the homogeneous binomial model.

162

163   **Avoid inference of congruent diversification models**

164   While millions of students struggling with their statistics homework might cheer the

165   destruction of the concept of estimating the probability of heads from a set of coin flips, it is

166   important to emphasize that even though these models are functionally congruent, each provides

167   different predictions after a new set of coin flips are made (e.g., what is the likeliest outcome of

168   the eleventh flip?). That is, even though they are indistinguishable from a probabilistic point of

169   view, we can still distinguish them when new data becomes available. Of course, with

170   comparative methods we cannot simply "flip" evolution more times to distinguish among a set of

171 congruent models. The emphasis, then, as Morlon et al. (2020) recently pointed out, becomes

172 what we are trying to learn about the world, given what we know about how it works. It is

173 generally true that with coins, we have a good idea that the probability of heads does not change

174 meaningfully over flips, so we may be willing to assume a standard binomial model and then

175 question the fairness of a coin, perhaps as a way of extrapolating to other coins (i.e., if this Euro

176 coin has a probability of heads of 0.502, is that true for other Euro coins?). In other words, the

177 parameter can be of interest because the model is not really in question.

178   With many diversification models, the central question is about which model fits best,

179 which is at odds with a general lack of knowledge about any system to clearly know which kind

180 of model is appropriate ahead of time. Even with diversification models that explicitly link rates

181 to abiotic variables such as temperature or sea level changes (e.g., Condamine et al. 2013; 2019)

182 the goal seems more focused on which model fits best. In our view, we are not yet at the stage

183 where we can confidently rule out a congruent model where extinction rates are driven by the

184 position of a hypothetical dwarf star outside our solar system, which triggers periods of increased

185 comet activity on Earth (e.g., Raup and Sepkowski 1984), over a more "sensible" model of, say,

186 temperature clearly affecting speciation but not extinction rates. In such cases, asking questions

187 about which of several indistinguishable models fit does not seem to us a good use of our time.

188   It is also important to emphasize that our argument here is not that the issues Louca and

189 Pennell (2020) point out are trivial. In fact, there are many papers, and even entire research

190 programs, dedicated to the development of time-varying, lineage homogeneous models of

191 diversification, and trying to draw conclusions based on which models fit best. But, as with coin

192 flipping or Brownian motion, knowing what conclusions can be made given the models and data

193 and limiting our work to those areas can be important. Moreover, if even coin flipping has

194   congruent models, there is no guarantee that even models that currently seem to avoid the

195   congruence issue, such as pulled diversification rates recommended by Louca and Pennell

196   (2020), do not have other congruent models with different parameters, such as models that

197   change rates by taxa rather than solely by time. Work on non-parametric identifiability (Stoudt,

198   2020; Louca and Pennell, *pers. comm.*) may be fruitful for determining which models will end up

199   being useful.

200

201   **Avoid ancestral rate reconstruction**

202   Ancestral state reconstruction of characters remains one of the most popular and widely

203   used approaches in phylogenetic comparative methods, despite the occasional discussion to

204   dampen enthusiasm in them (e.g., Cunningham et al. 1998; Omland 1999; Oakley and

205   Cunningham, 2000). Ancestral state reconstruction is useful for formulating testable hypotheses,

206   such as the synthesis and performance evaluation of putative ancestral proteins (e.g., Thornton et

207   al. 2003; Pillai et al. 2020), biogeographic history and movements of clades through time (e.g.,

208   Ree and Smith, 2008; Landis et al. 2020), and the order and timing of character state changes

209   (e.g., Schluter et al. 1997; Ackerly et al. 2006). Reconstructing diversification rates through time

210   has a similar appeal, in that they too can point to testable hypotheses about the intrinsic and

211   extrinsic factors that drive species diversity among groups. Armed with only a phylogeny of

212   modern taxa, we can reconstruct the seismograph tracing of how speciation rate, extinction rate,

213   net diversification rate, or the new pulled diversification or pulled speciation rates, have changed

214   through time. With the reconstruction of discrete or continuous characters, state information at

215   the extant tips is generally less and less informative about states at nodes as one traverses deeper

216   in the tree towards the root. For diversification rate models, the data are not arrayed along the

217   tips of a tree, but rather, come from the distribution of branching events across the phylogeny.
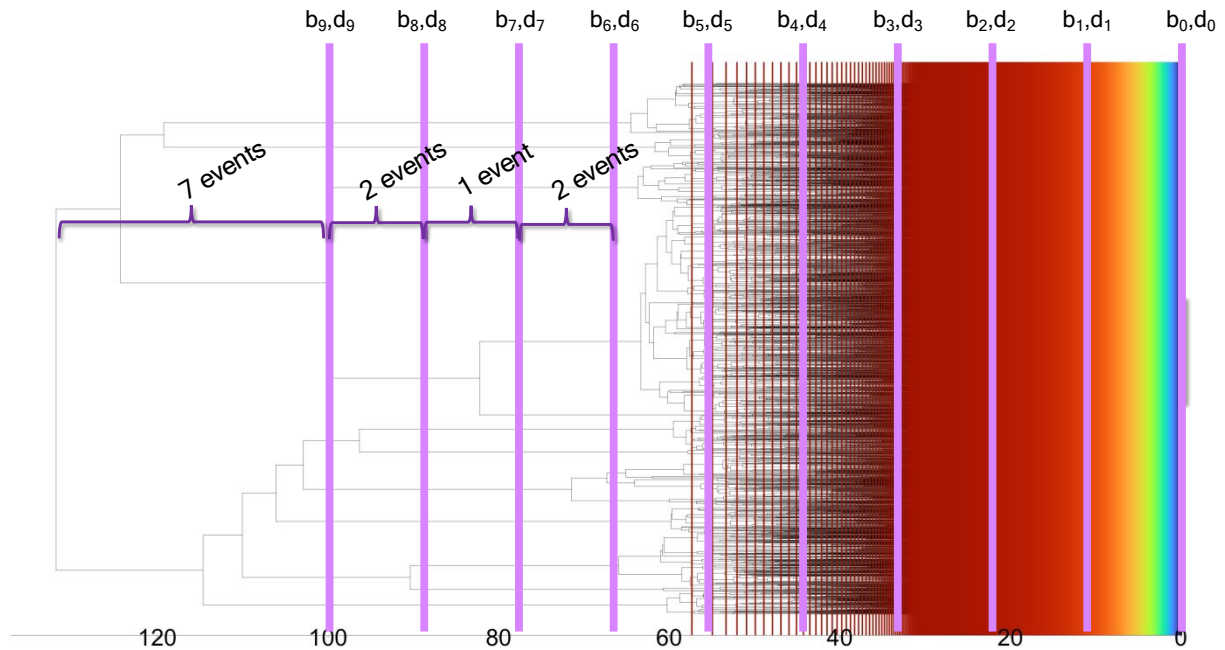
218

219   **Figure 2**. Million taxon tree from Louca and Pennell (2020). The purple lines separate the
220   regimes used to estimate rates. The thin vertical lines in a rainbow distinguish regimes with 100
221   events within them representing equal-sized slices of data. Half the regimes are on each side of
222   the green band, showing how much of the data are near the tips. The brackets show how many
223   events occur in each regime. Ignoring uncertainty in branch lengths or topology, this makes a 10
224   Myr long edge equally informative regardless of whether it ended 3 million years ago or 300
225   million years ago.

226

227   As Maddison and FitzJohn (2015) noted, our field does not yet think in terms of the

228   curvature of biodiversity-time, and so our expectations about the distribution of these branching

229   events are often wrong. The number of edges on trees, under most models, increases

230   approximately exponentially with time, although extinction complicates this, as would models

231   with carrying capacity (e.g., Rabosky and Lovette 2008), age-dependent extinction (Alexander et

232   al 2016), and other variations of the birth-death model. Nevertheless, lineage through time plots

233   are commonly shown on a log scale for the number of lineages due to this nearly exponential

234   growth. Visualizing the raw number of lineages would make the dynamics in the early parts of

235   the plot virtually invisible due to the massive growth of the line near the present. Importantly, the

236   midpoint of the data is the point at which half the number of lineages has accumulated, which is

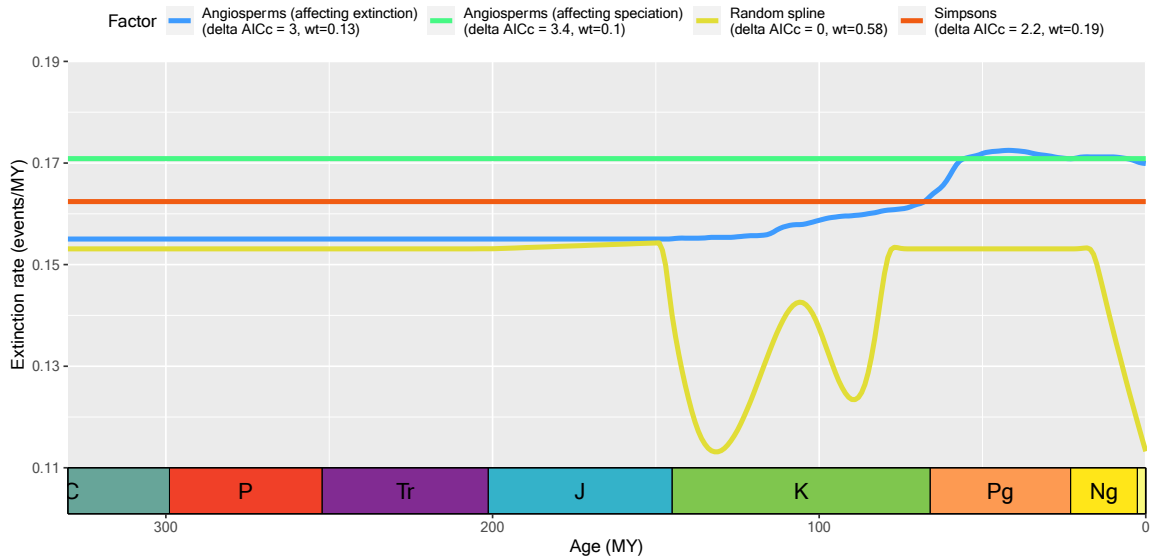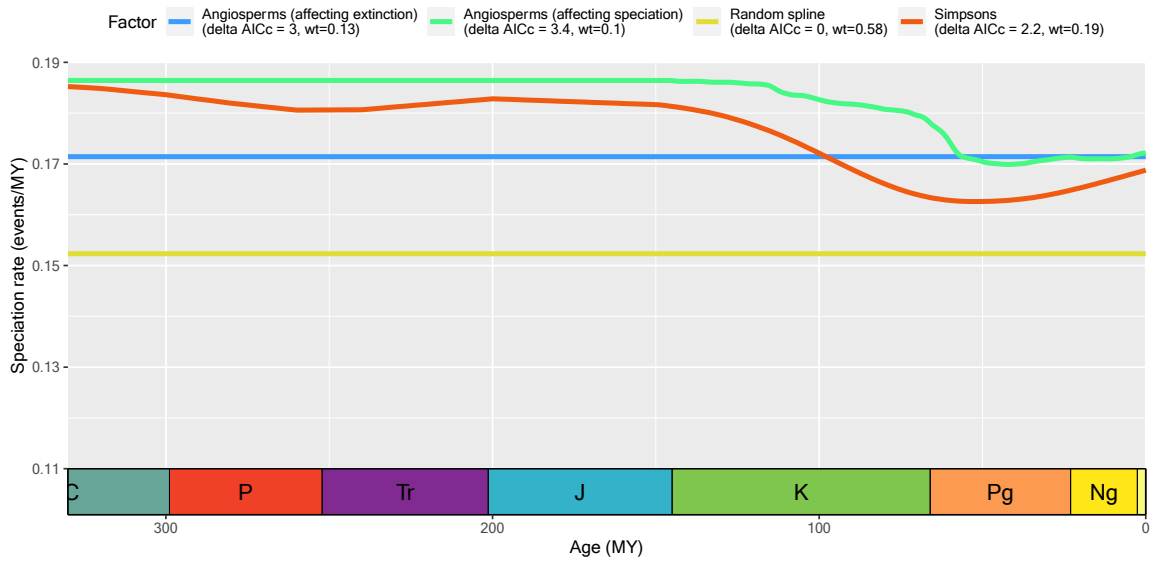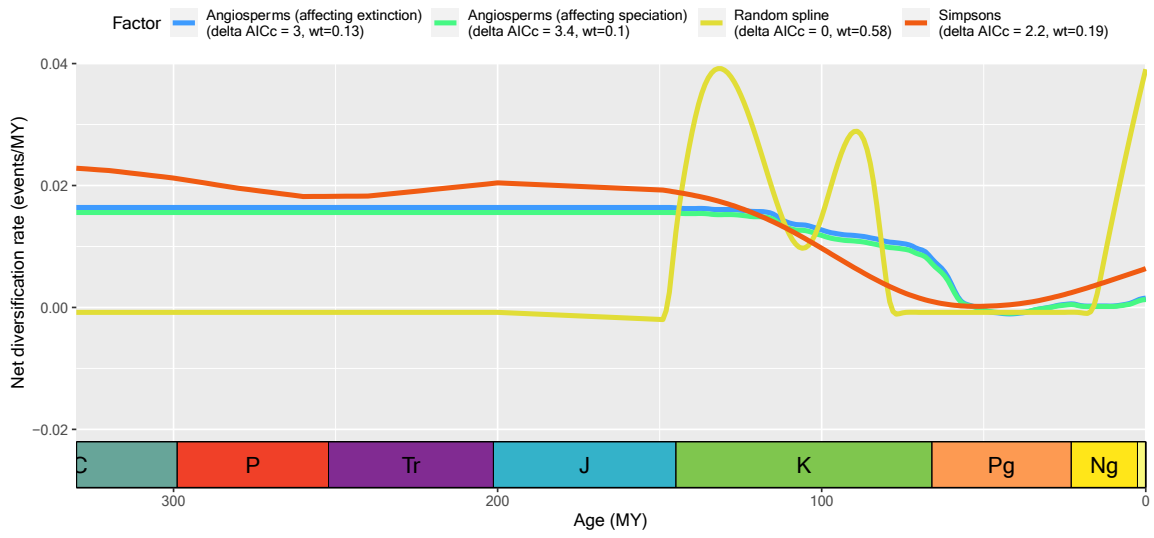237   *not* the halfway point along the time axis.

238   Consider a tree split into equal-sized chunks according to some time interval, as Louca

239   and Pennell (2020) and others have. The number of edges within a given bin naturally decreases

240   as one moves towards the root. Now, take the extreme example from Louca and Pennell (2020)

241   where they analyzed a tree with a million taxa (Figure 2). Even though the tree is far larger than

242   any published study of diversification, they only estimate rates along 10-time intervals and for

243   many of these bins there is only a trivial amount of data. For example, at the start of the 100 Myr

244   to 90 Myr interval, there are just seven lineages, and by the end of that interval, there are only

245   ten. The lineage through time plot, which is the data that goes into these methods, thus jumps just

246   three times over that ten million years. This is clearly not a lot of data points for estimating

247   speciation or extinction rates, or even a single pulled diversification rate. Each of the next several

248   intervals have a *single* jump. That is, it goes from 10 to 11 lineages from 90 to 80 Myr, and from

249   just 11 to 12 from 80 Myr to 90 Myr. It is no wonder that these methods perform poorly; a single

250   event on a 12-taxon tree does not contain much information about rates, whether pulled or not.

251   Put another way, these methods are starving for data across large portions of the tree.

252   A natural corollary, then, is that seismographic reconstructions of rates will contain

253   increasing levels of uncertainty as one moves deeper in time. Nee et al. (1994) showed clearly

254   that even rates from a constant birth-death model can carry substantial uncertainty. Yet most

255   analyses doing the sort of work Louca and Pennell (2020) criticize, and even their examples,

256   return a single point estimate for each parameter at a given time period. In a few cases, point

257   estimates are summarized together across a set of trees, which is better, but still likely reflects

258   substantially less uncertainty than what is truly present in any single estimate.

Conifer diversification with various predictors

259 **Figure 3**: Comparison of net diversification, speciation, and extinction rate of conifers using as a
260 predictor the best model from Condamine et al. (2020) in blue where only extinction rate varies
261 with angiosperm diversity, a slightly worse model from that paper (green) where speciation rate
262 varies with angiosperm diversity, a model (yellow) that fits the data best (at least in terms of
263 likelihood — the number of free parameters of the spline is hard to compare), and using scaled
264 IMDB ratings of the television program the Simpsons (red) as a predictor for speciation rate
265 (which did a better job predicting conifer diversification than angiosperm diversity did). Not
266 shown are numerous other attempts for other predictors using other splines, linear change
267 models, and ratings of many other television programs. Some of these also outperformed
268 angiosperms, but many did not.
269

270   Besides unexamined uncertainty in point estimates, there is substantial uncertainty in

271 which model fits best, even if one ignores the congruence issue. For example, Condamine et al.

272 (2020) compared various models correlating various rates with angiosperm diversity using just a

273 phylogenetic tree; their best model showed an exponential dependence of conifer extinction rate

274 with the number of angiosperms. However, models nearly as good (ΔAICc < 2; see their Table

275 S5) include an effect on speciation or both speciation and extinction (only 41% of the model

276 weight is on variable extinction only models; 39% is on variable speciation only, and 21% on

277 both varying). One can construct other patterns of diversification rates with very different

278 conclusions that are better predictors. For example, in Figure 3, the yellow diversification curves

279 predict the conifer data even better but tell a very different story of constant speciation with

280 decreases of extinction in the Cretaceous and Neogene rather than the recovered pattern of a

281 gradual rise of extinction in the Cretaceous onward. Even using ratings of a television show (the

282 Simpsons, the red line) scaled for the appropriate time period predicts conifer diversification

283 better than the postulated angiosperm mechanism. Similarly, Morlon et al. (2011) looking at a

284 paraphyletic set of 16 cetaceans found a constant speciation but variable extinction model fit

285 best, but there were two other models with a ΔAICc of less than 1 (including one where

286 extinction does not vary) — this makes it hard to draw any firm conclusions from modern data

287   alone. Careful biologists, as shown in the studies above, will limit themselves to only feasible

288   mechanisms, but as we know from other diversification models (Rabosky and Goldberg 2015;

289   Beaulieu and O'Meara, 2016), if presented with a very simple model and more complex

290   alternatives only, methods using our messy, complex empirical data will leap to use the more

291   complex predictors. That is, if the only way to incorporate the very real heterogeneity of a

292   process is to ascribe it to some varying predictor, methods will choose that. Whether it is 16

293   modern taxa or a million, it is unclear what we learn from such exercises. Our energies might be

294   better directed elsewhere.

295

296   **The state of SSE models and other approaches**

297   Louca and Pennell (2020) speculate that state-speciation and extinction models (SSE)

298   may have similar identifiability issues. This is not an unreasonable concern. Beaulieu and

299   O'Meara (2016) demonstrated that if a trait has no effect on speciation and/or extinction rates,

300   the likelihood of any SSE model becomes the product of the likelihoods of the Nee et al. (1994)

301   tree likelihood and the character model likelihood (or the sum of the log-likelihoods in log

302   space), so the models are clearly related. One could certainly alter the SSE model to include

303   realistic factors like mass extinctions and secular changes in rates through time, and any one of

304   these features will undoubtedly lead to a set of models with identical likelihoods. However, in

305   other ways, strict SSE models can be immune, because they do not split the tree into time bins.

306   Instead, they approximately treat a tree as a series of discrete chunks — that is, a chunk in one

307   part of the tree is in state 0, and so is impacted by the instantaneous speciation rate, λ_0, and

308   extinction rate, μ_0, while another chunk in another part of the tree is in state 1 and so is impacted

309   by speciation rate, λ_1, and extinction rate, μ_1 (in reality, they average over these paintings based
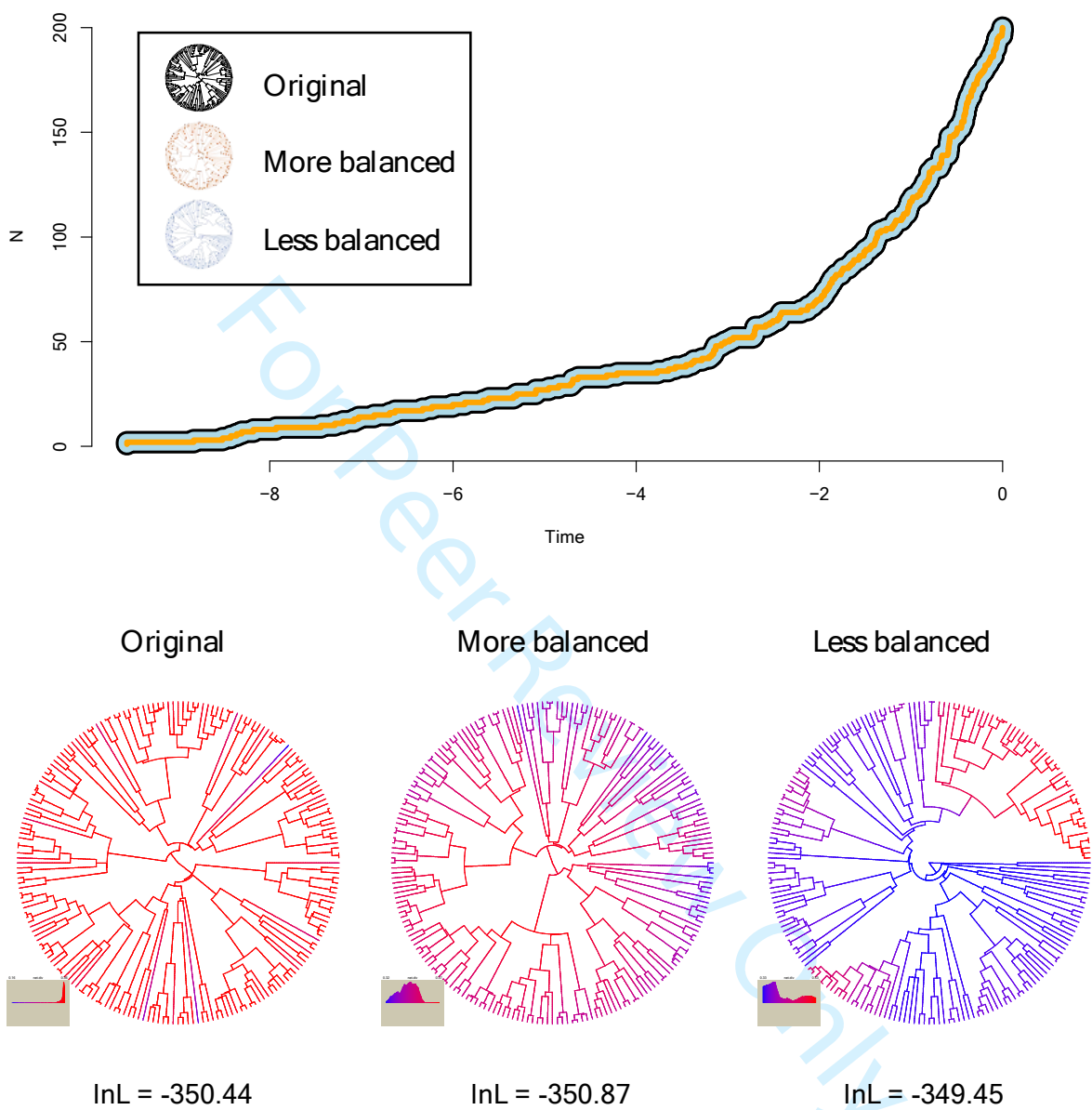
Figure 4: (A) Depicts the identical lineage through time (LTT) plots for three trees that differ in terms of tree balance. The procedure takes a simulated tree, then makes swaps across branches to either increase balance or decrease it but maintain the same lineage through time curve. (B) Depicts the log-likelihood score among the three trees under a two-rate MiSSE model. These trees produce identical log-likelihoods under taxon-homogeneous, time-heterogeneous models that use LTT data. However, this is not the case here because allowing rates to vary among clades, as our MiSSE models do, avoids the trap of having an infinite array of congruent models. Helmstetter et al. (2021) reach similarly positive conclusions about the possibility of learning about diversification from SSE models.

320   on their probabilities). Within each of these chunks the speciation and extinction rates are

321   invariant, and as Nee et al. (1994) showed, constrained in this way there is a single maximum

322   likelihood estimate of each rate. If one limits the model space to where rates are dependent on

323   states (observed or hidden or some combination of both), then SSE models should be

324   identifiable, though not immune to all the practical difficulties of estimating rates in the presence

325   of extinction, finite data, errors in branch lengths and topology, and more.

326   We can at least empirically demonstrate that SSE models are immune to the issues of

327   model congruence based on information in the lineage through time plot: SSE models use more

328   information than this. In Figure 4, there are three trees with identical lineage through time

329   curves, but different arrangements of topology. Under a constant rate Yule or birth-death model

330   the likelihoods of these three trees are identical, as one would expect given the findings of Louca

331   and Pennell (2020). However, if we allow for multiple rates to be inferred across the tree by

332   fitting a hidden states only model (which we call MiSSE; see Vasconcelos et al. 2021) the three

333   trees have different likelihood. This is because the MiSSE model uses information not accessible

334   to LTT methods, namely, the tree topology. Other methods that fit rate heterogeneity across taxa,

335   such as MSBD (Barido-Sottani et al. 2018) and ClaDS (Maliet et al. 2019), may also not be

336   bound by the issues that make different LTT models congruent. Even an approach as simple as

337   sister group comparisons (e.g., Slowinski and Guyer 1993) can detect differences in net

338   diversification rate across pairs of clades in a way that depends on topology alone: identical

339   lineage through time plots would have no effect on this. Taken together, this does not mean that

340   clade-specific models of diversification could not have their own issues (even coin flipping

341   models can have congruence, as shown above, and issues like the best estimate for extinction

342   rate being negative (Louca & Pennell (2021)) could still have an effect), just that the

343   identifiability issue identified by Kubo and Iwasa (1995) and Louca and Pennell (2020) does not

344   apply to them.

345   **What are we really learning anyway?**

346   Null hypothesis testing is intended to show whether an effect is significantly different

347   from chance alone. At some point, though, comparing against chance becomes an uninteresting

348   and dull exercise as the end point of a study. After several decades of studying diversification on

349   molecular phylogenies and continually finding variation in rates across taxa and across time,

350   favoring a complex model over a "dull" null hypothesis of simple constant birth-death is no

351   longer surprising. No reasonable scientist will argue that diversification processes have remained

352   perfectly constant through time, with no changes in extinction rates, no factors changing

353   speciation rates, and more. We know the data comes from a heterogeneous, complex process and

354   so any even somewhat reasonable more complex model will fit better than a simple model. As

355   we have noted elsewhere (see Beaulieu and O'Meara 2016; Caetano et al. 2018), rejecting the

356   "null" does not imply that the slightly more complex alternative is the true model. Like a hot gas

357   moved from a simple bottle to a more complex bottle with greater volume, our complex data will

358   happily expand to take the shape of the biggest container offered to it. Model rejection, model

359   weighting, posterior probability of models are all ways of saying, "my cloud of data is more

360   comfortable in this larger bottle than in this smaller bottle. Since the extra bulge on the larger

361   bottle is called factor X, this clearly shows that factor X is important." However, a different bottle

362   with the same volume but with a bulge for factor Y might fit as well. Good science will involve

363   comparing different reasonable models to the data, not just comparing our slightly more complex

364   model of interest with slightly simpler models. Much of our work on hidden rate models (e.g.,

365   Beaulieu et al. 2013; Beaulieu and O'Meara 2016; Caetano et al. 2018; Boyko and Beaulieu

366   2021) is motivated by this desire to give our preferred models an actual chance to lose against

367   other models in the hope that we learn from this.

368   In our view, an important aspect of the work of Louca and Pennell (2020) was showing

369   that even this limited, careful approach might not work for time-heterogeneous diversification

370   rates: there are multiple diversification bottle shapes that fit the cloud of branching times from a

371   tree equally well. Furthermore, approaches that seek to track the wiggles of the diversification

372   seismograph through time tell us very little, if anything, about the past. However, we would add

373   that instead of tracing the wiggles of a single pulled diversification rate pen on a diversification

374   seismograph, or even take the extreme step of stopping analyses of diversification using modern

375   phylogenies altogether, we should use the valid methods we do have to answer biological

376   questions, in the same way we can use Brownian motion even though different parameterizations

377   can give identical likelihoods. Focus on analyses that lead to discoveries or confirmations of

378   biological processes that are possible given available data.

379   On the whole, it is important to recognize that *our methods are better suited for using the*

380   *past to learn about the present survivors, not using the present survivors to learn about the past*.

381   Phylogenies of extant taxa convey an enormous amount of information about species and their

382   direct ancestors, but they also necessarily miss much of the history of a particular clade.

383   Therefore, there will never be a clever analysis of a phylogeny of extant archosaurs (crocodilians

384   and birds) that will result in an inference of the dynamics of the rise and fall of sauropod

385   dinosaurs, even though they are firmly nested in that clade and must have had a huge effect on

386   the lineages that survived while all were interacting. Yet this is exactly what we are asking of our

387   diversification seismograph analyses of modern taxa — that is, we think we are understanding

388   something about diversification dynamics of archosaurs in the Cretaceous from a study of their

389   weird, few surviving lineages. However, phylogenies of extant taxa can give us information

390   about what led to present diversity, what traits are associated with modern diversity patterns,

391   and, perhaps, even when certain modern lineages took off. We can understand something about

392   diversification patterns of extant birds, for example, including what traits are associated with

393   faster diversification or turnover rates. Moreover, even in cases where we have samples through

394   time, model congruence can still be an issue (Louca et al. 2021).

395   Perhaps the best example of procedures that illustrate where we think the field needs to

396   reconsider are classic sister group comparisons (Mitter et al., 1988). These explicitly are about

397   comparing modern clades and so are by their nature lineage-heterogeneous and limited to

398   examining factors leading to modern diversity. They do not claim to allow inference about rate

399   shifts in the past, since they attempt to control for the effect of time. There can be important

400   corrections for even these methods (Käfer and Mousset, 2014) but they prevent scientists from

401   spinning tales from limited information about the past. They should also be far more robust to

402   the concerns raised by Maddison and FitzJohn (2015) than even hidden rate models. Of course,

403   they are not without their own limitations: it can be hard to find enough comparisons; they only

404   allow comparison of the direction of net diversification differences due to some pre-specified

405   factor, while many of our hypotheses might relate to speciation rate, extinction rate, or, as we

406   have advocated turnover rate (Beaulieu and O'Meara, 2016; Vasconcelos et al. 2021); they

407   typically require only discrete characters (though see Harvey et al. 2020 and the bomeara/sisters

408   package on github); and they require ancestral state reconstruction to find sister pairs differing by

409   a character state. There are also questions completely inaccessible to these methods; however,

410   accepting these limitations at the outset may have prevented years of work that relied on methods

411   that felt scientific but gave ultimately meaningless results given the issues now understood about

412   time-heterogeneous diversification models.

413

414

415   **Conclusions**

416   The reconstruction of diversification rates through time, whether of pulled or classic

417   rates, is appealing but flawed in the same way that inference of ancestral states is appealing but

418   also flawed. Multiple indistinguishable models give very different estimates about the past, and

419   even for large trees, what matters is the branches and branching events at the times of interest,

420   often when the mighty tree was a mere sapling. Moreover, this only looks at branches with

421   modern descendants. What information it does provide is about what those lineages may have

422   been doing, not what the clade as a whole may have been doing. Thus, approaches that seek to

423   paint pictures about potential past diversification regimes at very incremental time periods are

424   certainly suspect, with Louca and Pennell (2020) pointing to additional congruence issues that

425   can affect diversification models.

426   Some feel that, even in the face of these congruence issues, understanding

427   macroevolution remains an exciting and promising endeavor (Helmstetter et al. 2021). We are

428   not nearly as optimistic. We can certainly learn about diversification processes from trees, but we

429   need to recognize that what we can understand largely relates *only* to the surviving tips. Current

430   SSE models and other models that infer rate heterogeneity across taxa, rather than across time,

431   may provide additional information that lets them fit different parameters and likelihood for trees

432   with identical lineage through time curves, avoiding the particular issue raised by Kubo and

433   Iwasa (1995) and Louca and Pennell (2020). However, as with Brownian motion and coin

434 flipping, congruent models can likely be found for these as well. Sister group analyses may grow

435 in importance in future studies of diversification.

436

437

438 **Literature Cited**

439 Abrams, P.A. 1986. Is predator-prey coevolution an arms race? Trends in Ecology and Evolution

440   1:108-110.

441 Ackerly, D.D., D.W. Schwilk, and C.O. Webb. 2006. Niche evolution and adaptive radiation:

442   testing the order of trait divergence. Ecology 87:S50-S61.

443 Alexander, H.K., A. Lambert, and T. Stadler. 2016. Quantifying age-dependent extinction from

444   species phylogenies. Systematic Biology 65:35-50.

445 Barido-Sottani, J., T.G. Vaughan, and T. Stadler. 2018. Detection of HIV transmission clusters

446   from phylogenetic trees using a multi-state birth-death model. Journal of the Royal

447   Society Interface 15:https://doi.org/10.1098/rsif.2018.0512

448 Beaulieu, J.M., D.-C. Jhwueng, C. Boettiger, and B.C. O'Meara. 2012. Modeling stabilizing

449   selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. Evolution 66:

450   2369-2383.

451 Beaulieu, J.M., B.C. O'Meara, M.J. Donoghue. 2013. Identifying hidden rate changes in the

452   evolution of a binary morphological character: the evolution of plant habit in campanulid

453   angiosperms. Systematic Biology 62: 725:737.

454 Beaulieu, J.M., and B.C. O'Meara. 2016. Detecting hidden diversification shifts in models of

455   trait-dependent speciation and extinction. Systematic Biology 65:583-601.

456 Boyko, J.D., and J.M. Beaulieu. 2021. Generalized hidden Markov models for phylogenetic

457   comparative methods. Methods in Ecology and Evolution 12: 468:478.

458 Butler, M.A., and A.A. King. 2004. Phylogenetic comparative analysis: a modeling approach for

459   adaptive evolution. The American Naturalist 164:683-695.

460 Bokma, F. 2008. Bayesian estimation of speciation and extinction probabilities from

461   (in)complete phylogenies. Evolution 62:2441–2445.

462 Caetano, D.S., B.C. O'Meara, and J.M. Beaulieu. 2018. Hidden state models improve state-

463   dependent diversification approaches, including biogeographic models. Evolution

464   72:2308-2324.

465 Condamine, F.L., D. Silvestro, E.B. Koppelhus, and A. Antonelli. 2020. The rise of angiosperms

466   pushed conifers to decline during global cooling. Proceedings of the National Academy

467   of Sciences, USA 117:28867-28875.

468 Condamine, F.L., J. Rolland, and H. Morlon. 2019. Assessing the causes of diversification

469   slowdowns: temperature-dependent diversity-dependent models receive equivalent

470   support. Ecology Letters 22:1900-1912.

471 Condamine, F.L., J. Rolland, and H. Morlon. 2013. Macroevolutionary perspectives to

472   environmental change. Ecology Letters 16:72-85.

473 Cunningham, C.W., K.E. Omland, and T.H. Oakley. 1998. Reconstructing ancestral character

474   states: a critical reappraisal. Trends in Ecology and Evolution 13:361-366.

475 Dawkins R., and J.R. Krebs. 1979. Arms races between and within species. Proceedings of the

476   Royal Society, B 205:489-511.

477   Eriksson, O. E.M. Friis, and P. Lofgren. 2000. Seed size, fruit size, and dispersal systems in

478   angiosperms from the Early Cretaceous to the Late Tertiary. The American Naturalist

479   156:47-58.

480   Etienne, R.S., B. Haegeman, T. Stadler, T. Aze, P. Pearson, A. Purvis, and A.Phillimore. 2012.

481   Diversity-dependence brings molecular phylogenies closer to agreement with the fossil

482   record. Proceedings of the Royal Society, B 279:1300–1309.

483   Felsenstein J. 1988. Phylogenies and quantitative characters. Annual Review of Ecology and

484   Systematics 19:445-471.

485   Hansen, T.F., and E.P. Martins. 1996. Translating between microevolutionary process and

486   macroevolutionary patterns: the correlation structure of interspecific data. Evolution

487   50:1404-1417.

488   Helmstetter, A.J., S. Glemin, J. Käfer, R. Zenil-Ferguson, H. Sauquet, H. de Boer, L-P. M.J.

489   Dagallier, N. Mazet, E.L. Reboud, T.L.P. Couvreur, F.L. Condamine. 2021. Pulled

490   diversification rates, lineages-through-time plots and modern macroevolutionary

491   modeling. Systematic Biology in press.

492   Käfer, J., and S. Mousset. 2014. Standard sister clade comparison fails when testing derived

493   character states. Systematic Biology 63:601-609.

494   Kubo, T., and Y. Iwasa. 1995. Inferring the rates of branching and extinction from molecular

495   phylogenies. Evolution 49:694-704.

496   Lambert, A., and T. Stadler. 2013. Birth-death models and coalescent point processes.

497   Theoretical Population Biology 90:113-128.

498   Landis M.J., D.A.R Eaton, W.L. Clement, B. Park, E.L. Spriggs, P.W. Sweeney, E.J. Edwards,

499   and M.J. Donoghue. 2020. Joint phylogenetic estimation of geographic movements and

500   biome shifts during the global diversification of Viburnum. Systematic Biology 70:76–

501   94.

502 Louca, S., and M.W. Pennell. 2020. Extant timetrees are consistent with a myriad of

503   diversification histories. Nature 580:502-506.

504 Louca, S. and M.W. Pennell. 2021. Why extinction estimates from extant phylogenies are so

505   often zero. Current Biology 31:1-6.

506 Louca, S., A. McLaughlin, A. MacPherson, J.B. Joy, M.W. Pennell. 2021. Fundamental

507   identifiability limits in molecular epidemiology. Molecular Biology and Evolution

508   38:4010-4024.

509 Maddison, W.P., and R.G. FitzJohn. 2015. The unsolved challenge to phylogenetic correlation

510   tests for categorical characters. Systematic Biology 64: 127-136.

511 Maddison, W.P., P.E. Midford, and S.P. Otto. 2007. Estimating a binary character's effect on

512   speciation and extinction. Systematic Biology 56:701–710.

513 Maliet, O., F. Hartig, and H. Morlon. 2019. A model with many small shifts for estimating

514   species-specific diversification rates. Nature Ecology and Evolution 3:1086-1092.

515 Mitter, C. B. Farrell, B. Wiegmann. 1988. The phylogenetic study of adaptive zones: has

516   phytophagy promoted insect diversification? American Naturalist 132: 107-128.

517 Morlon, H., T.L. Parsons, and J.B. Plotkin. 2011. Reconciling molecular phylogenies with the

518   fossil record. Proceedings of the National Academy of Sciences, USA 108:16327–16332.

519 Morlon, H., F. Hartig, and S. Robin. 2020. Prior hypotheses or regularization allow inference of

520   diversification histories from extant timetrees. bioRxiv doi:

521   https://doi.org/10.1101/2020.07.03.185074

522   Nee, S., E. C. Holmes, R. M. May, and P. H. Harvey. 1994. Extinction rates can be estimated

523   from molecular phylogenies. Philosophical Transactions of the Royal Society B344:77–

524   82.

525   Nee, S., A.Ø. Mooers, and P.H. Harvey. 1992. The tempo and mode of evolution revealed from

526   molecular phylogenies. Proceedings of the National Academy of Sciences, USA

527   89:8322–8326.

528   Oakley T.H., and C.W. Cunningham. 2000. Independent contrasts succeed where ancestor

529   reconstruction fails in a known bacteriophage phylogeny. Evolution 54:397-405.

530   O'Meara, B.C., C. Ané, M.J. Sanderson, and P.C. Wainwright. 2006. Testing for different rates

531   of continuous trait evolution using likelihood. Evolution 60:922-933.

532   Omland, K.E. 1999. The assumptions and challenges of ancestral state reconstructions.

533   Systematic Biology 48:604-611.

534   Pennell M.W., J.M. Eastman, G.J. Slater, J.W. Brown, J.C. Uyeda, R.G. FitzJohn, M.E. Alfaro,

535   and L.J. Harmon. 2014. geiger v2.0: an expanded suite of methods for fitting

536   macroevolutionary models to phylogenetic trees. Bioinformatics 30:2216-2218.

537   Pillai, A.S., S.A. Chandler, Y. Liu, A.V. Signore, C.R. Cortez-Romero, J.L.P. Benesch, A.

538   Laganowsky, J.F. Storz, G.K.A. Hochberg, and J.W. Thornton. 2020. Origin of

539   complexity in haemoglobin evolution. Nature 581:480-485.

540   Rabosky, D. L. 2006. Likelihood methods for detecting temporal shifts in diversification rates.

541   Evolution 60:1152–1164.

542   Rabosky, D. L. 2009. Heritability of extinction rates links diversification patterns in molecular

543   phylogenies and fossils. Systematic Biology 58:629–640.

544   Rabosky, D.L., and E.E. Goldberg. 2015. Model inadequacy and mistaken inferences of

545   trait-dependent speciation. Systematic Biology 64:340–355.

546   Rabosky, D. L., and I. J. Lovette. 2008. Density-dependent diversification in North American

547   wood warblers. Proceedings of the Royal Society, B 275:2363–2371.

548   Raup, D.M., and J.J. Sepkowski, Jr. 1984. Periodicity of extinction in the geological past.

549   Proceedings of the National Academy of Sciences, USA 81:801-805.

550   Ree, R.H., and S.A. Smith. 2008. Maximum likelihood inference of geographic range evolution

551   by dispersal, local evolution, and cladogenesis. Systematic Biology 57:4-14.

552   Revell, L.J. 2021. A variable-rate quantitative trait evolution model using penalized likelihood.

553   bioRxiv doi: https://doi.org/10.1101/2021.04.17.440282.

554   Schluter D., T. Price, A.Ø. Mooers, and D. Ludwig. 1997. Likelihood of ancestor states in

555   adaptive radiation. Evolution 51:1699-1711.

556   Slowinski, J.B., C. Guyer. 1993. Testing whether certain traits have caused amplified

557   diversification: an improved method based on a model of random speciation and

558   extinction. The American Naturalist 142:1019-1024.

559   Stoudt, S. 2020. A Statistical Investigation of Species Distribution Models and Communication

560   of Statistics Across Disciplines. UC Berkeley. https://escholarship.org/uc/item/2zq81799.

561   Thomas, G.H., R.P. Freckleton, and T. Szekely. 2006. Comparative analyses of the influence of

562   developmental mode on phenotypic diversification rates in shorebirds. Proceedings of the

563   Royal Society, B 273:1619-1624.

564   Thornton, J.W., E. Need, and D. Crews. 2003. Resurrecting the ancestral steroid receptor:

565   ancient origin of estrogen signaling. Science 301:1714-1717.

566   Tifney, B.H. 1984. Seed size, dispersal syndromes, and the rise of angiosperms: evidence and

567   hypothesis. Annals of the Missouri Botanical Garden 71:551-576.

568   Vasconcelos, T., B.C. O'Meara, and J.M. Beaulieu. 2021. Hidden state-only speciation

569   and extinction models provide accurate tip-estimates of diversification rates. Submitted.

570   Vasconcelos, T. B.C. O'Meara, and J.M. Beaulieu. 2021. Retiring "cradles" and

571   "museums" of biodiversity. American Naturalist in press. https://doi.org/10.1086/717412

572

573

574

575

576

577

578

579

580

581   **Figure Legends**

582   **Figure 1**: Probability of heads per flip on different models of coin flipping. Each of these models

583   can fit the same dataset of two heads, eight tails with equal likelihood but make very different

584   predictions about the next flip.

585

586   **Figure 2**: Million taxon tree from Louca and Pennell (2020). The purple lines separate the

587   regimes used to estimate rates. The thin vertical lines in a rainbow distinguish regimes with 100

588   events within them representing equal-sized slices of data. Half the regimes are on each side of

589   the green band, showing how much of the data are near the tips. The brackets show how many

590   events occur in each regime.

591

592   **Figure 3**: Comparison of net diversification, speciation, and extinction rate of conifers using as a

593   predictor the best model from Condamine et al. (2020) in blue where only extinction rate varies

594   with angiosperm diversity, a slightly worse model from that paper (green) where speciation rate

595   varies with angiosperm diversity, a model (yellow) that fits the data best (at least in terms of

596   likelihood — the number of free parameters of the spline is hard to compare), and using scaled

597   IMDB ratings of the television program the Simpsons (red) as a predictor for speciation rate

598   (which did a better job predicting conifer diversification than angiosperm diversity did). Not

599   shown are numerous other attempts for other predictors using other splines, linear change

600   models, and ratings of many other television programs. Some of these also outperformed

601   angiosperms, but many did not.

602

603   **Figure 4**: (A) Depicts the identical lineage through time (LTT) plots for three trees that differ in

604   terms of tree balance. The procedure takes a simulated tree, then makes swaps across branches to

605   either increase balance or decrease it but maintain the same lineage through time curve. (B)

606   Depicts the log-likelihood score among the three trees under a two-rate MiSSE model. These

607   trees produce identical log-likelihoods under taxon-homogeneous, time-heterogeneous models

608   that use LTT data. However, this is not the case here because allowing rates to vary among

609   clades, as our MiSSE models do, avoids the trap of having an infinite array of congruent models.

610   Helmstetter et al. (2021) reach similarly positive conclusions about the possibility of learning

611   about diversification from SSE models.