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

2   **Running title (40 characters max): Diversification method survival**

3   **Authors:** Brian C. O'Meara[1][*] and Jeremy M. Beaulieu[2]

4   **Affiliations:**
5   [1]Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville,
6   Tennessee, 37996-1610 USA
7   [2]Department of Biological Sciences, University of Arkansas, Fayetteville, Arkansas, 72701 USA
8
9   [*]**Corresponding author:** Department of Ecology and Evolutionary Biology, University of
10   Tennessee, Knoxville, Tennessee, 37996-1610 USA; email: bomeara@utk.edu
11
12

13   **Author contributions:** BCO and JMB take equal responsibility for the contents of this article.
14
15

22   **Data Availability Statement:** Code supporting this article is made freely available at
23   http://flippedcoin.info/, https://github.com/bomeara/diversificationlives, and
24   https://github.com/bomeara/CondamineEtAlExample.
25
26
27
28   **Conflict of Interest:** The authors have declared no conflict of interest.

29

30

31

32

33

34

35

36

Abstract

Models have long been used for understanding changing diversification patterns over time. The rediscovery that models with very different rates through time can fit a phylogeny equally well has led to great concern about the use of these models. We share and add to these concerns: even with time heterogeneous models without these issues, the distribution of the data means that estimates will be very uncertain. However, we argue that congruence issues such as this also occur in models as basic as Brownian motion and coin flipping. Taxon-heterogeneous models such as many SSE models appear not to have this particular issue.

45

46

Keywords: diversification, congruence, likelihood, identifiability

48

49

50

51

52

53

54

55

56

57

58

59

Introduction

60

61   For decades, molecular phylogenies have served as vital sources of historical information

62 for deciphering the birth and the death dynamics of lineages. Thousands of studies of molecular

63 phylogenies have been dedicated to investigating diversification. In theory, estimating constant

64 birth and death rates separately is possible because each has distinguishable effects on the tree

65 shape and branch length distributions (Nee et al. 1994). There are several extensions that expand

66 this simple model for characterizing diversification as a function of time or diversity (e.g., Nee et

67 al. 1992; Rabosky 2006, 2009; Bokma 2008; Rabosky and Lovette 2008; Morlon et al. 2011;

68 Etienne et al. 2012), which are used to reconstruct lines showing speciation and extinction rates

69 scrolling into the past, like the pen of a seismometer tracking vibrations through time. A sudden

70 sweep up of the extinction rate arm could mean a mass extinction. A slow, downward trajectory

71 of the speciation rate arm as time approaches the present could mean available niches have

72 become filled up, limiting the possibilities of adding new species. And, as with constant rate

73 birth-death models, we have been working under the assumption that even tiny changes in

74 speciation and/or extinction through time should leave distinct signatures on the tree shape and

75 branching structure in a molecular phylogeny.

76   In a recent paper by Louca and Pennell (2020), the entire enterprise of estimating

77 diversification rates, at least from molecular phylogenies alone, has been called into question. As

78 it turns out, for any given phylogeny there are an infinite array of congruent models each having

79 unique functions of speciation and/or extinction rates smoothly varying through time. This is

80 based on the property of both constant rate birth-death and time-varying models in which every

81 lineage at any given time-point experiences the same rates, and so sampling times for either a

82   speciation or extinction event are drawn from the same distribution (also known as a coalescent

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

106   problems do not extend to *all* models of diversification. Instead, they are limited to situations

107   where the goal is to interpret diversification rates through time using what we refer to as, "time-

108   varying, lineage homogeneous" models — again, models in which all lineages experience the

109   same variable rates at any given point in time. These would be analogous to a non-heritable trait-

110   dependent process (Lambert and Stadler, 2013), where changes in a trait occur the same in all

111   species independently (e.g., global CO_2, sea-level changes, global temperature patterns). We

112   argue that what we refer to as "lineage-specific heterogeneous" models, in which rates vary

113   among lineages across time points, perhaps due to the inheritance of a trait (e.g., state-speciation

114   and extinction, or SSE models; Maddison et al. 2007), should be immune to the issues of

115   identifiability raised above. This comes with the substantial caveat that this is true if, and

116   probably only if, the heritable rate changes are modeled as containing a single speciation and

117   extinction rate that do not vary through time. Essentially, we will show that these models do

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

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

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

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

122   distributed on trees.

123   Overall, we make four points:

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

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

126   questions are infeasible.

127   2. Time-varying, lineage homogeneous models that use just the information from a

128   lineage through time curve to estimate changing speciation, extinction,

129   diversification, turnover, or extinction fraction should be avoided due to

130   congruence issues.

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

132   2020) are identifiable, but they fail to incorporate the substantial uncertainty in

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

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

135   methods in point 2)

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

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

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

139   (1995).

140

141 **Model congruence is common**

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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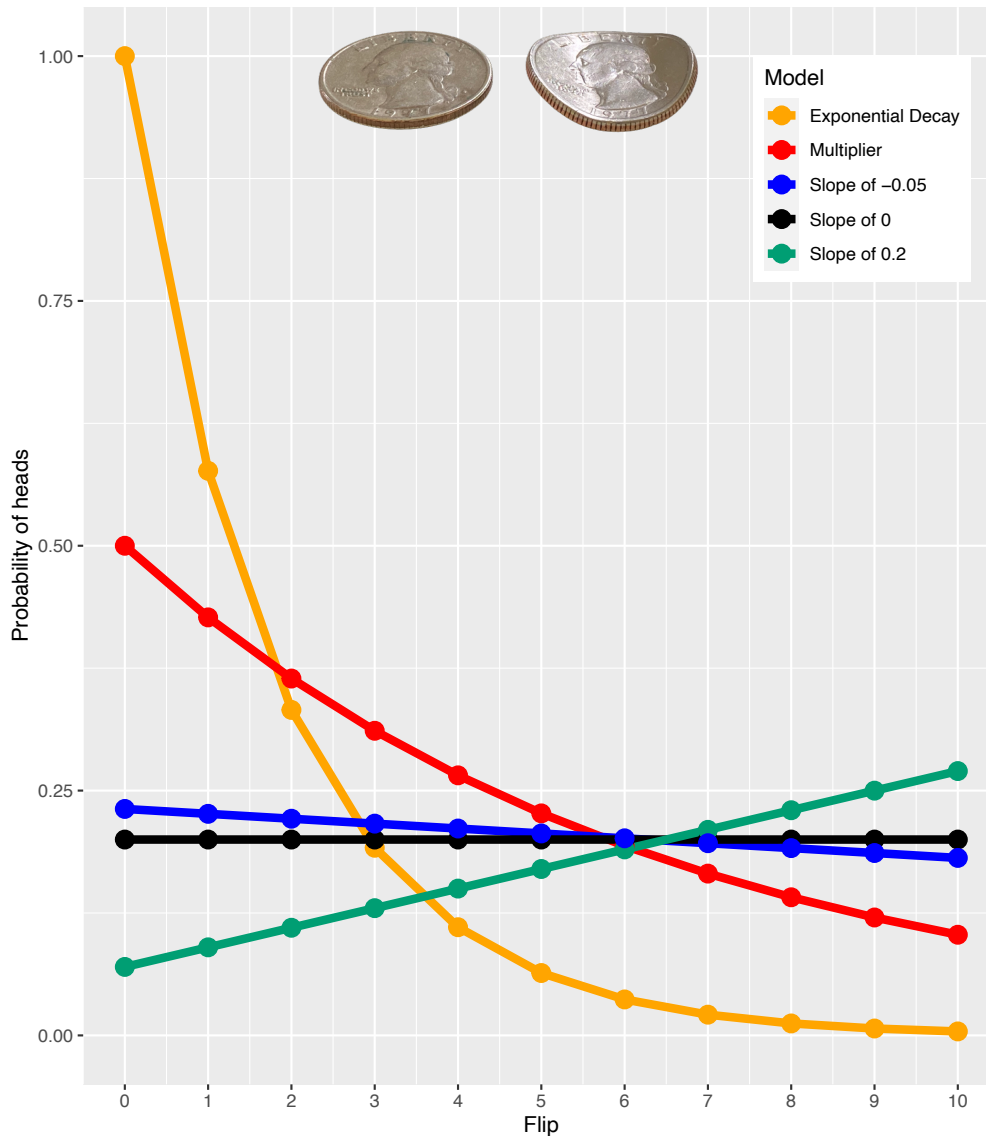
Figure 1: Probability of heads per flip on different models of coin flipping. Each of these models
can fit the same dataset of two heads, eight tails with equal likelihood but make very different
predictions about the next flip.

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

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

181   for care and analysis, not panic.

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

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

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

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

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

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

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

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

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

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

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

193   2 heads in 10 flips of the coin in each of these models is the *same* as the simple binomial model,

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

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

196   20% higher parameters ahead of time rather than fitting them, these have the same number of

197   free parameters as the homogeneous binomial model.

198

199   **Avoid inference of congruent diversification models**

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

232 (2020), do not have other congruent models with different parameters, such as models that

233 change rates by taxa rather than solely by time.

234

235 **Avoid ancestral rate reconstruction**

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

252
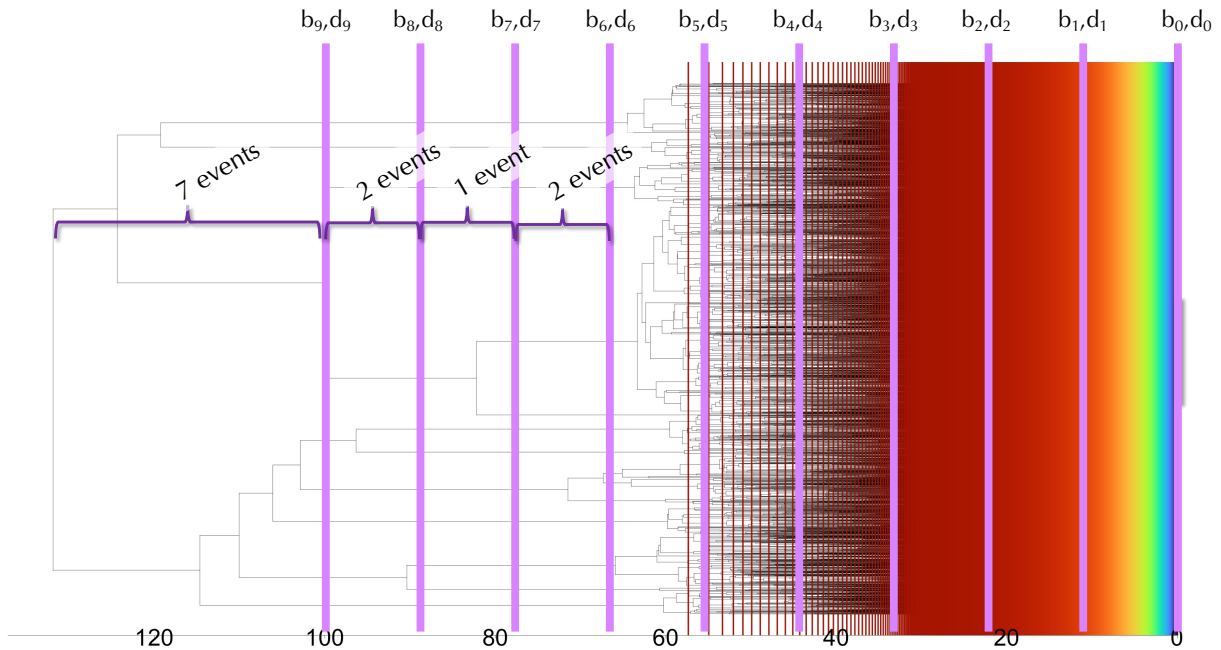
Figure 2. Million taxon tree from Louca and Pennell (2020). The purple lines separate the
regimes used to estimate rates. The thin vertical lines in a rainbow distinguish regimes with 100
events within them representing equal-sized slices of data. Half the regimes are on each side of
the green band, showing how much of the data are near the tips. The brackets show how many
events occur in each regime. Ignoring uncertainty in branch lengths or topology, this makes a 10
Myr long edge equally informative regardless of whether it ended 3 million years ago or 300
million years ago.

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

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

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

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

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

266   al 2016), and other variations of the birth-death model. Nevertheless, lineage through time plots
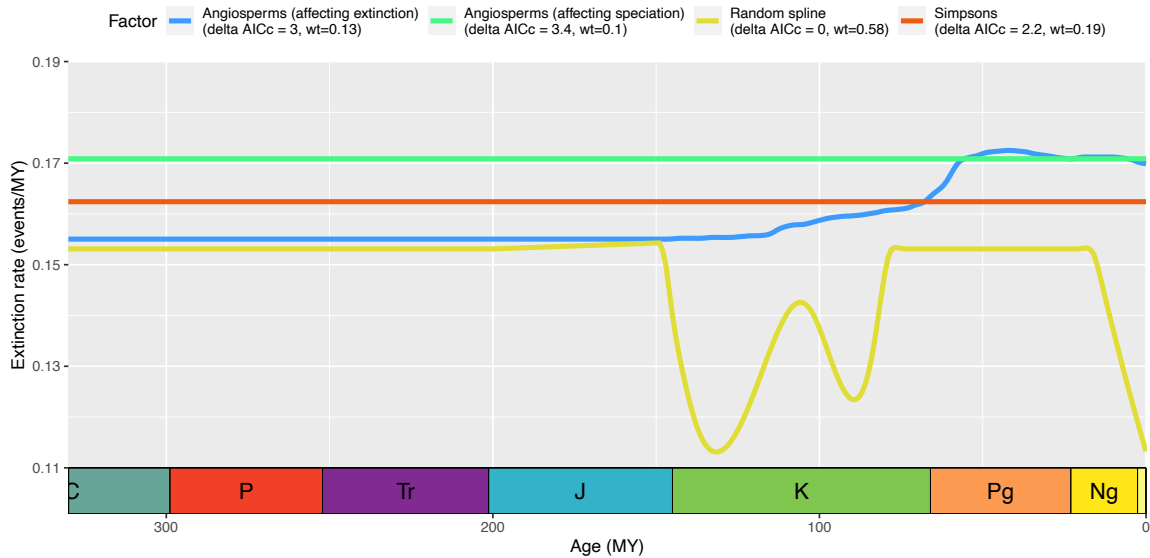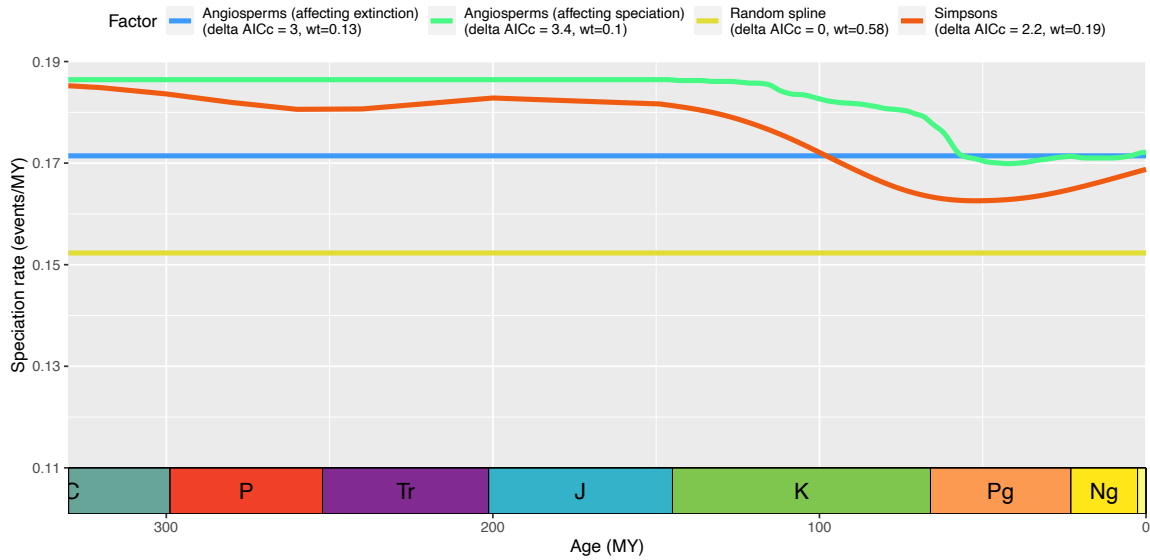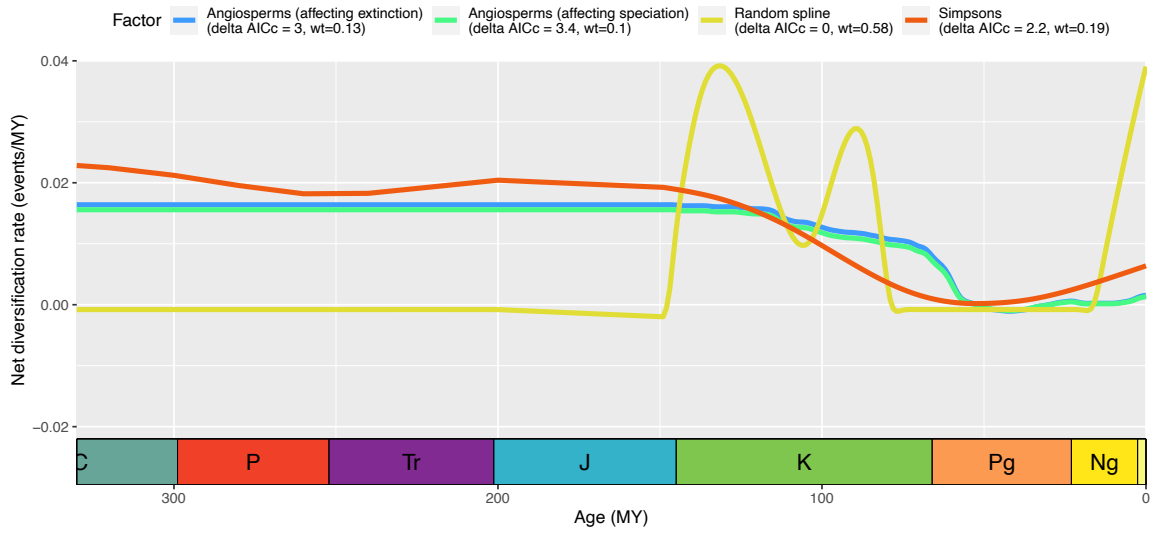
267   are commonly shown on a log scale for the number of lineages due to this nearly exponential

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Conifer diversification with various predictors

Figure 3: Comparison of net diversification, speciation, and extinction rate of conifers using as a predictor the best model from Condamine et al. (2020) in blue where only extinction rate varies with angiosperm diversity, a slightly worse model from that paper (green) where speciation rate varies with angiosperm diversity, a model (yellow) that fits the data best (at least in terms of likelihood — the number of free parameters of the spline is hard to compare), and using scaled IMDB ratings of the television program the Simpsons (red) as a predictor for speciation rate (which did a better job predicting conifer diversification than angiosperm diversity did). Not shown are numerous other attempts for other predictors using other splines, linear change models, and ratings of many other television programs. Some of these also outperformed angiosperms, but many did not.
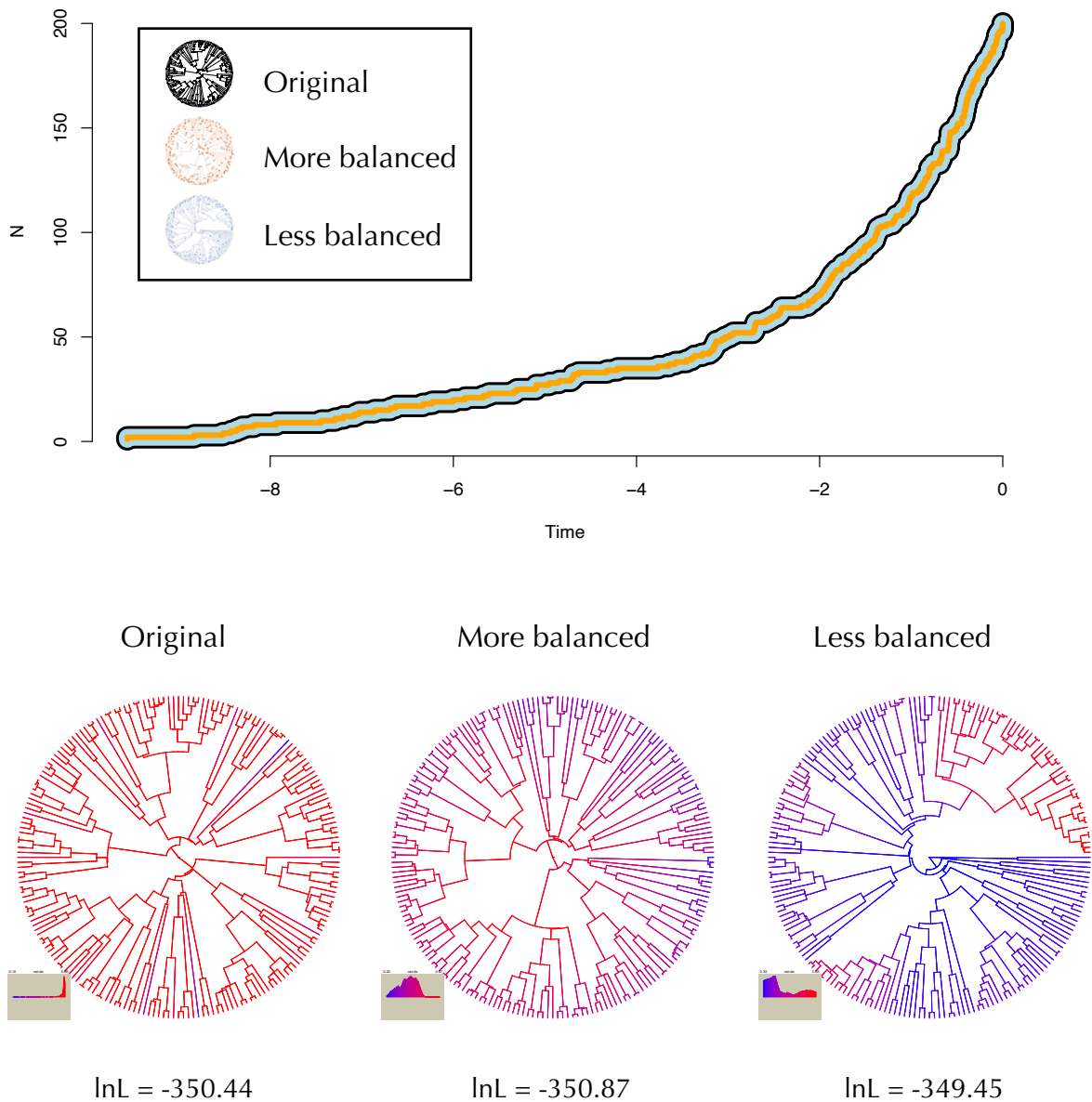
Besides unexamined uncertainty in point estimates, there is substantial uncertainty in which model fits best, even if one ignores the congruence issue. For example, Condamine et al. (2020) compared various models correlating various rates with angiosperm diversity using just a phylogenetic tree; their best model showed an exponential dependence of conifer extinction rate with the number of angiosperms. However, models nearly as good (ΔAICc < 2; see their Table S5) include an effect on speciation or both speciation and extinction (only 41% of the model weight is on variable extinction only models; 39% is on variable speciation only, and 21% on both varying). One can construct other patterns of diversification rates with very different conclusions that are better predictors. For example, in Figure 3, the yellow diversification curves predict the conifer data even better but tell a very different story of constant speciation with decreases of extinction in the Cretaceous and Neogene rather than the recovered pattern of a gradual rise of extinction in the Cretaceous onward. Even using ratings of a television show (the Simpsons, the red line) scaled for the appropriate time period predicts conifer diversification better than the postulated angiosperm mechanism. Similarly, Morlon et al. (2011) looking at a paraphyletic set of 16 cetaceans found a constant speciation but variable extinction model fit best, but there were two other models with a ΔAICc of less than 1 (including one where extinction does not vary) — this makes it hard to draw any firm conclusions from modern data

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

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

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

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

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

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

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

328 better directed elsewhere.

329

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

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

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

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

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

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

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

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

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

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

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

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

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

343 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

375   models can have congruence, as shown above), just that the identifiability issue identified by

376   Kubo and Iwasa (1995) and Louca and Pennell (2020) does not apply to them.

What are we really learning anyway?

Null hypothesis testing is intended to show whether an effect is significantly different from chance alone. At some point, though, comparing against chance becomes an uninteresting and dull exercise as the end point of a study. After several decades of studying diversification on molecular phylogenies and continually finding variation in rates across taxa and across time, favoring a complex model over a "dull" null hypothesis of simple constant birth-death is no longer surprising. No reasonable scientist will argue that diversification processes have remained perfectly constant through time, with no changes in extinction rates, no factors changing speciation rates, and more. We know the data comes from a heterogeneous, complex process and so any even somewhat reasonable more complex model will fit better than a simple model. As we have noted elsewhere (see Beaulieu and O'Meara 2016; Caetano et al. 2018), rejecting the "null" does not imply that the slightly more complex alternative is the true model. Like a hot gas moved from a simple bottle to a more complex bottle with greater volume, our complex data will happily expand to take the shape of the biggest container offered to it. Model rejection, model weighting, posterior probability of models are all ways of saying, "my cloud of data is more comfortable in this larger bottle than in this smaller bottle. Since the extra bulge on the larger bottle is called factor X, this clearly shows that factor X is important." However, a different bottle with the same volume but with a bulge for factor Y might fit as well. Good science will involve comparing different reasonable models to the data, not just comparing our slightly more complex model of interest with slightly simpler models. Much of our work on hidden rate models (e.g., Beaulieu et al. 2013; Beaulieu and O'Meara 2016; Caetano et al. 2018; Boyko and Beaulieu 2021) is motivated by this desire to give our preferred models an actual chance to lose against other models in the hope that we learn from this.

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

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

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

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

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

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

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

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

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

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

410   biological processes that are possible given available data.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

425   faster diversification or turnover rates.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

443   time-heterogeneous diversification models.

444

445

Conclusions

446   **Conclusions**

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

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

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

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

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

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

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

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

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

456   can affect diversification models.

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

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

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

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

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

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

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

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

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

466   in importance in future studies of diversification.

467

468

469   **Literature Cited**

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

471   1:108-110.

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

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

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

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

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

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

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

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

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

481   2369-2383.

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

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

484   angiosperms. Systematic Biology 62: 725:737.

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

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

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

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

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

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

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

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

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

494   dependent diversification approaches, including biogeographic models. Evolution

495   72:2308-2324.

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

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

498   of Sciences, USA 117:28867-28875.

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

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

501   support. Ecology Letters 22:1900-1912.

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

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

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

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

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

507   Royal Society, B 205:489-511.

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

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

510   156:47-58.

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

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

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

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

515   Systematics 19:445-471.

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

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

518   50:1404-1417.

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

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

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

522   modeling. Systematic Biology in press.

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

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

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

526   phylogenies. Evolution 49:694-704.

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

528   Theoretical Population Biology 90:113-128.

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

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

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

532   94.

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

534   diversification histories. Nature 580:502-506.

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

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

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

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

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

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

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

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

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

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

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

546   diversification histories from extant timetrees. bioRxiv doi:

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

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

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

550   82.

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

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

553   89:8322–8326.

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

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

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

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

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

559   Systematic Biology 48:604-611.

560 Pennell M.W., J.M. Eastman, G.J. Slater, J.W. Brown, J.C. Uyeda, R.G. FitzJohn, M.E. Alfaro,
561   and L.J. Harmon. 2014. geiger v2.0: an expanded suite of methods for fitting
562   macroevolutionary models to phylogenetic trees. Bioinformatics 30:2216-2218.

563 Pillai, A.S., S.A. Chandler, Y. Liu, A.V. Signore, C.R. Cortez-Romero, J.L.P. Benesch, A.
564   Laganowsky, J.F. Storz, G.K.A. Hochberg, and J.W. Thornton. 2020. Origin of
565   complexity in haemoglobin evolution. Nature 581:480-485.

566 Rabosky, D. L. 2006. Likelihood methods for detecting temporal shifts in diversification rates.
567   Evolution 60:1152–1164.

568 Rabosky, D. L. 2009. Heritability of extinction rates links diversification patterns in molecular
569   phylogenies and fossils. Systematic Biology 58:629–640.

570   Rabosky, D.L., and E.E. Goldberg. 2015. Model inadequacy and mistaken inferences of
571   trait-dependent speciation. Systematic Biology 64:340–355.

572 Rabosky, D. L., and I. J. Lovette. 2008. Density-dependent diversification in North American
573   wood warblers. Proceedings of the Royal Society, B 275:2363–2371.

574 Raup, D.M., and J.J. Sepkowski, Jr. 1984. Periodicity of extinction in the geological past.
575   Proceedings of the National Academy of Sciences, USA 81:801-805.

576 Ree, R.H., and S.A. Smith. 2008. Maximum likelihood inference of geographic range evolution
577   by dispersal, local evolution, and cladogenesis. Systematic Biology 57:4-14.

578 Revell, L.J. 2021. A variable-rate quantitative trait evolution model using penalized likelihood.
579   bioRxiv doi: https://doi.org/10.1101/2021.04.17.440282.

580 Schluter D., T. Price, A.Ø. Mooers, and D. Ludwig. 1997. Likelihood of ancestor states in
581   adaptive radiation. Evolution 51:1699-1711.

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

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

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

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

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

587   Royal Society, B 273:1619-1624.

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

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

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

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

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

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

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

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

596

597

598

599

600

601

602

603

604

Figure Legends

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

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

predictions about the next flip.

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

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

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

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

events occur in each regime.

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

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

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

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

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

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

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

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

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

angiosperms, but many did not.

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.