

1 **Prior hypotheses or regularization allow inference of diversification histories**
2 **from extant timetrees**

3

4 **Hélène Morlon^{1,*}, Florian Hartig² & Stéphane Robin³**

5

- 6 1. Institut de Biologie de l'Ecole Normale Supérieure (IBENS), Ecole Normale Supérieure,
7 CNRS, INSERM, PSL Research University, 75005 Paris, France
8 2. Theoretical Ecology, University of Regensburg, Regensburg, Germany
9 3. UMR MIA-Paris, AgroParisTech, INRA, Paris-Saclay University, 75005 Paris, France

10

11 *corresponding author. email: helene.morlon@bio.ens.psl.eu

12

13 **Abstract**

14 Phylogenies of extant species are widely used to study past diversification dynamics¹. The
15 most common approach is to formulate a set of candidate models representing evolutionary
16 hypotheses for how and why speciation and extinction rates in a clade changed over time, and
17 compare those models through their probability to have generated the corresponding
18 empirical tree. Recently, Louca & Pennell² reported the existence of an infinite number of
19 ‘congruent’ models with potentially markedly different diversification dynamics, but equal
20 likelihood, for any empirical tree (see also Lambert & Stadler³). Here we explore the
21 implications of these results, and conclude that they neither undermine the hypothesis-driven
22 model selection procedure widely used in the field nor show that speciation and extinction
23 dynamics cannot be investigated from extant timetrees using a data-driven procedure.

24

25

26 **Main text**

27 Louca & Pennell² consider the homogeneous (i.e. lineage-independent) stochastic birth-death
28 process of cladogenesis traditionally used in macroevolution to test hypotheses about how and
29 why rates of speciation (birth, λ) and extinction (death, μ) have changed over time t . They
30 show that for any given time-dependent speciation function $\lambda > 0$ and extinction function
31 $\mu \geq 0$, there exists an infinite set of alternative functions $\lambda^* > 0$ and $\mu^* \geq 0$ such that the
32 probability distribution of extant trees under the corresponding birth-death processes M and
33 M^* is identical. Consequently, M or M^* yield identical likelihood values for any given
34 empirical tree. This identifiability issue is certainly both interesting and unfortunate, but what
35 are its implications for phylogenetic-based diversification analyses?

36

37 We consider this question against two alternative philosophies to conducting science. The
38 first is a hypothesis-driven research approach, in which a small number of alternative ideas
39 about the underlying mechanism are compared against data⁴. An example is the foundational
40 study of Nee *et al.* (1992)⁵, who examined one of the first molecular phylogenies of birds, and
41 demonstrated that it was incompatible with a constant-rate diversification model. Grounded in
42 Simpson's evolutionary theory of adaptive radiations⁶, they then hypothesized that rates of
43 cladogenesis in birds might be affected by niche-filling processes. Finding that a diversity-
44 dependent model indeed fitted their data better, they concluded that diversity-dependent
45 cladogenesis was a more plausible scenario to explain the diversification of birds.

46

47 This hypothesis-driven approach has inspired more than 30 years of research in phylogenetic
48 diversification analyses¹. Exponential time-dependencies have been used, for example, to
49 mimic early burst patterns expected from adaptive radiation theory⁶, or as an approximation to
50 diversity-dependent cladogenesis⁷. In the context of the environment-dependent models

51 mentioned by Louca & Pennell², functional hypotheses have often been derived from
52 foundational theories of biodiversity, such as the metabolic theory of biodiversity⁸ and
53 MacArthur & Wilson's theory of island biogeography⁹. Phenomenological models, such as
54 simple linear time- or environmental-dependencies, have indeed also been used, but typically
55 either as null models⁷ or as the simplest way to model the effect of an explanatory
56 environmental variable on evolutionary rates^{8,10}. The primary goal of this research, however,
57 has been to fit, test and compare diversification scenarios that were defined *a priori* to reflect
58 different evolutionary hypotheses.

59
60 Louca & Pennell's congruent models M^* , on the other hand, do not correspond *a priori* to any
61 evolutionary hypotheses, and would never be considered in a well-conducted hypothesis
62 model selection procedure in the first place⁴. Fortunately, as Louca & Pennell² admit, most
63 diversification hypotheses that are compared by evolutionary biologists are distinguishable
64 from extant trees, and the ongoing effort to integrate fossil information provides even brighter
65 perspectives¹¹. The existence of a large number of congruent models therefore poses no direct
66 challenge to the traditional hypothesis-driven research approach. The only possible concern is
67 the question of model selection consistency: if the true model is not in the set of considered
68 models, do we select the correct hypothesis? This question has not been answered one way or
69 the other and would require thorough investigation in future research.

70
71 By considering all possible diversification functions, Louca & Pennell² implicitly subscribe to
72 a different method of scientific discovery, where the goal is to learn λ and μ from the data
73 alone. After finding that these quantities are not simultaneously identifiable, even for
74 infinitely large phylogenies, they suggest to instead estimate identifiable quantities such as the
75 pulled speciation rate λ_p or pulled diversification rate r_p . There are certainly advantages of

76 the pulled rates, but estimating them from limited-size phylogenies is still a challenging task
77 (SI S.1). One could address the problem by regularization, but then the question arises why
78 we should not directly use regularization to handle the identifiability issue around λ and μ .
79 Indeed, a wide body of statistical regularization techniques exist to deal with unidentifiability
80 issues¹², such as shrinkage¹³, smoothing¹⁴ or the use of Bayesian priors. Louca & Pennell² do
81 not address the possibility to use such statistical regularization methods, although these
82 methods have already been used successfully for inference of diversification rates^{15,16} (SI
83 S.2).

84
85 To provide a concrete example of these points, we perform an analysis of the diversification
86 of the Madagascan vangas in the logic that would be applied in the field¹⁷, but simplified for
87 illustrative purposes. We hypothesize that diversification followed an ‘Early Burst’ pattern¹⁸,
88 with fast speciation at the origin of the group and subsequent slowdown, rather than constant-
89 rate diversification. The Early Burst pattern, related to the idea of adaptive radiations⁶, is
90 modeled by an exponential decay of the speciation rates through time, used as an
91 approximation of diversity-dependence. We also consider the hypothesis that a substantial
92 number of extinction events occurred during the diversification of this group. Among the four
93 corresponding models, the model with exponential change and no extinction (M) is best
94 supported by the data (see Table 1 & Supplementary Information). M is characterized by
95 $\lambda(t) = \lambda_0 e^{\alpha t}$ and $\mu(t) = 0$ where t is measured from the present to the past, $\lambda_0 = 0.018$ is
96 the estimated present-day speciation rate, and $\alpha = 0.1$ measures the estimated speed of time
97 change. A positive α reflects a decline in speciation rate from the origin of the group to the
98 present (Fig. 1). We conclude that the hypothesis of Early Burst diversification with
99 negligible extinctions is the most likely of the four hypotheses we considered.

101 We now follow Louca & Pennell² by solving Eq. 2 for models congruent to our best model
102 M. First, we choose the extinction function to be a constant $\mu_1^*(t) = \mu_0$ and compute $\lambda_1^*(t)$.
103 Second, we choose the speciation function to be a constant $\lambda_2^*(t) = \lambda_0$ and compute $\mu_2^*(t)$.
104 We find (SI; see the corresponding inferred dynamics in Fig. 1):

105

$$106 \quad \lambda_1^*(t) = \frac{\lambda_0 e^{-\frac{\lambda_0}{\alpha} t} e^{(\alpha+\mu_0)t} e^{\frac{\lambda_0}{\alpha} e^{\alpha t}}}{1 + \lambda_0 e^{-\frac{\lambda_0}{\alpha} t} \int_0^t e^{(\alpha+\mu_0)s} e^{\frac{\lambda_0}{\alpha} e^{\alpha s}} ds}$$

107

108 and

$$109 \quad \mu_2^*(t) = \lambda_0 - \alpha - \lambda_0 e^{\alpha t}$$

110

111 The biological interpretation of these additional models and of their parameters is not
112 obvious. The equation for μ_2^* looks more interpretable at first, but it expresses the temporal
113 change and the extinction rate at present through the same parameter α , which means that a
114 positive extinction rate at present ($\alpha < 0$) will force extinction rates to decline over time.

115 Here M_2^* infers negative extinction rates, and is therefore not plausible (Fig. 1). M_1^* infers a
116 decline in speciation rate from the origin of the group to the present for extinction rates μ_0
117 ranging from at least 0.05 to 0.3, consistent with our previous results (Fig. 1). While rate
118 estimates do vary substantially, the general temporal trend is preserved.

119

120 Finally, we fit r_p directly with spline functions, as suggested by Louca & Pennell² (SI S.3),
121 and deduce λ (or μ) when assuming μ (or λ) constant. General diversification trends are
122 consistent with those found above; finer dynamics depend on the choices made to estimate r_p
123 and suggest interesting new hypotheses to test (SI S.3). Here we chose λ or μ to be constant to
124 simplify the problem. Data-driven approaches have relaxed this hypothesis while still

125 estimating global temporal tendencies accurately, based on the reasonable prior belief that
126 rates don't change much in a small amount of time¹⁵. We expect such or similar regularization
127 approaches to eliminate the pathological cases with markedly different diversification
128 histories shown in Louca & Pennell².

129

130 There are useful results in Louca & Pennell's paper. The pulled diversification rate is difficult
131 to interpret biologically, yet realizing that models are only distinguishable by this quantity is
132 important, and the approach might provide an interesting way to generate new hypotheses.

133 The implications of these results for diversification analyses, however, are largely
134 overinterpreted, mainly because the constraints imposed by the practice of hypothesis-driven
135 research, prior knowledge, and the possibilities to penalize complexity are not considered in
136 the paper. A very similar identifiability issue occurred 10 years ago in population genetics,
137 when it was shown that the widely-used allelic frequency spectrum (AFS) is consistent with a
138 myriad of demographic histories¹⁹. Despite this insight, the AFS – along with other data
139 sources – remains a prominent approach of demographic inference²⁰. Identifiability issues
140 naturally arise in approaches that try to infer the potentially unlimited complexity of historical
141 processes from limited contemporary data, and this is why we work hypothesis-driven,
142 develop regularization techniques, and integrate other data types. Louca & Pennell² remind us
143 that the hypotheses we formulate when we develop models influence the conclusions we
144 draw, and that these conclusions should always be taken with a grain of salt; this is always a
145 good reminder, but it does not compromise the current practice in the field.

146

147 **References**

148

- 149 1. Morlon, H. Phylogenetic approaches for studying diversification. *Ecol. Lett.* **17**, 508–
150 525 (2014).
- 151 2. Louca, S., Pennell, M.W. Extant timetrees are consistent with a myriad of
152 diversification histories. *Nature* **580**, 502–505 (2020).
- 153 3. Lambert, A. & Stadler, T. Birth–death models and coalescent point processes: the
154 shape and probability of reconstructed phylogenies. *Theor. Popul. Biol.* **90**, 113–128
155 (2013).
- 156 4. Burnhan, K., Anderson, D. R. Model Selection and Multimodel Inference. *Springer*,
157 *New York* (2002).
- 158 5. Nee, S., Mooers, A. O., Harvey, P. H. Tempo and mode of evolution revealed from
159 molecular phylogenies. *Proc. Natl Acad. Sci. USA* **89**, 8322–8326 (1992).
- 160 6. Simpson G.G The Major Features of Evolution. *Columbia Univ Press, New York*
161 (1953).
- 162 7. Rabosky, D. L., Lovette, I. J. Density-dependent diversification in North American
163 wood warblers. *Proc. Royal Soc. B* **275**, 2363–2371 (2008).
- 164 8. Condamine, F. L., Rolland, J., Morlon, H. Assessing the causes of diversification
165 slowdowns: temperature-dependent and diversity-dependent models receive
166 equivalent support. *Ecol. Lett.* **22**, 1900–1912 (2019).
- 167 9. Valente, L. et al. A simple dynamic model explains the diversity of island birds
168 worldwide. *Nature* **579**, 92–96 (2020).
- 169 10. Clavel, J., Morlon, H. Accelerated body size evolution during cold climatic periods in
170 the Cenozoic. *Proc. Natl Acad. Sci. USA* **114**, 4183–4188 (2017).
- 171 11. Heath, T. A., Huelsenbeck, J. P., Stadler, T. The fossilized birth–death process for
172 coherent calibration of divergence-time estimates. *Proc. Natl Acad. Sci. USA* **111**,
173 E2957–E2966 (2014).

- 174 12. Hastie, T., Tibshirani, R., & Friedman, J. The elements of statistical learning: data
175 mining, inference, and prediction. *Springer Science & Business Media* (2009).
- 176 13. Zou, H., & Hastie, T. Regularization and variable selection via the elastic net. *J. R.*
177 *Stat. Soc. B*, **67**, 301-320 (2005).
- 178 14. Wood, S. N., Pya, N., & Säfken, B. Smoothing parameter and model selection for
179 general smooth models. *J. Am. Stat. Assoc.*, *111*(516), 1548-1563 (2016).
- 180 15. Maliet, O., Hartig, F., Morlon, H. A model with many small shifts for estimating
181 species-specific diversification rates. *Nat. Ecol. Evol.* **7**, 1086-1092 (2019).
- 182 16. Magee, A. F., Höhna, S., Vasylyeva, T. I., Leaché, A. D., Minin, V. N. Locally
183 adaptive Bayesian birth-death model successfully detects slow and rapid rate shifts.
184 *bioRxiv*, 853960. (2019).
- 185 17. Jønsson, K. A. et al. Ecological and evolutionary determinants for the adaptive
186 radiation of the Madagascan vangas. *Proc. Natl Acad. Sci. USA* **109**, 6620–6625
187 (2012).
- 188 18. Harmon, L. J. et al. Early bursts of body size and shape evolution are rare in
189 comparative data. *Evolution* **64**, 2385–2396 (2010).
- 190 19. Myers, S., Fefferman, C., & Patterson, N. Can one learn history from the allelic
191 spectrum? *Theor. Popul. Biol.* **73**, 342–348 (2008).
- 192 20. Ragsdale, A. P., Moreau, C., & Gravel, S. Genomic inference using diffusion models
193 and the allele frequency spectrum. *Curr. Opin. Genet. Dev.* **53**, 140–147 (2018).
- 194 21. Morlon, H., Lewitus, E., Condamine, F. L., Manceau, M., Clavel, J., & Drury, J.
195 RPANDA: an R package for macroevolutionary analyses on phylogenetic trees.
196 *Methods Ecol. Evol.* **7**, 589–597 (2016).
- 197

198 **Acknowledgments** We thank Amaury Lambert and the Morlon (current and past) research
 199 team, in particular Ignacio Quintero, Isaac Overcast, Jonathan P. Drury, Odile Maliet, and
 200 Julien Clavel for discussions. We also thank Knud Jønsson for providing us with the
 201 Madagascan vanga tree. H. Morlon acknowledges funding from the ERC (grant ERC-CoG
 202 PANDA).

203

204

205 **Table 1. Diversification models fitted to the phylogeny of the Madagascan vangas**

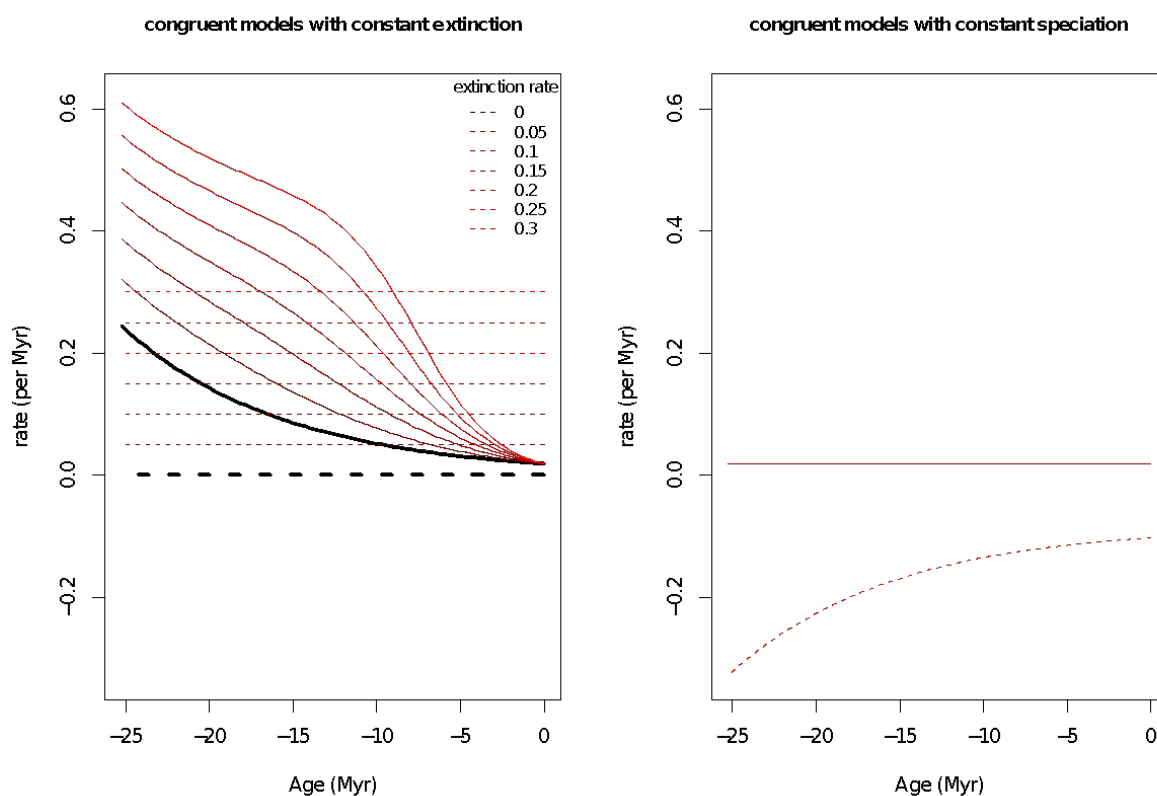
206

Model	nb	Log L	AICc	Δ AIC	λ_0	α	μ_0
Exponential change of speciation rate, no extinction (M)	2	-71.29	147.21	0	0.018	0.102	-
Exponential change of speciation rate, constant extinction	3	-71.04	149.4	2.19	0.025	0.117	0.077
Constant speciation, no extinction	1	-76.09	154.37	7.17	0.06	-	-
Constant speciation, constant extinction	2	-76.09	156.80	9.6	0.06	-	3.39e-09

207 nb denotes the number of parameters. LogL stands for the maximum log-likelihood, AICc for the second order Akaike
 208 Information Criterion⁴, and Δ AIC for the difference in AICc between the model and the best model in the set. Models are ranked based on
 209 their AICc support. λ_0 is the estimated speciation rate at present, α is the estimated rate of change of speciation with time (a positive α
 210 reflects a rate decline), and μ_0 is the estimated extinction rate. Fits were performed using the *fit_bd* function from RPANDA²¹ using the
 211 “crown” conditioning and a sampling fraction of 1 (the tree of the Madagascan vangas is complete¹⁷).

212

213



214

215 **Figure 1 Diversification of the Madagascan vangas as inferred from congruent models.**

216 The black curves on the left panel represent the dynamics of speciation (plain line) and

217 extinction (dashed line) corresponding to our best-fit model M (exponential decline in

218 speciation rate, non-significant extinctions). The colored curves illustrate the rate dynamics of

219 congruent models obtained by: fixing increasing values of a constant extinction rate (left

220 panel, M_1^*) and fixing the speciation rate to λ_0 (right panel, M_2^*). Note that M_1^* infers a

221 speciation rate decline regardless of the assumed extinction rate.

222

223

224

225

226

227 **Supplementary Information**

228

229 **S.1 Fitting the pulled rates also requires model selection or regularization**

230 In their S9, Louca & Pennell explain how to obtain non-parametric estimates of λ_p and
231 r_p from empirical timetrees by maximum likelihood. We note that, as for any non-parametric
232 approach, these estimates rely on two non-independent and somewhat arbitrary choices: a) the
233 functional basis (e.g. piecewise-polynomial, wavelets, Fourier) and b) the form of the
234 regularization (e.g. smoothness, sparsity). For phylogenies of a few hundred of species, which
235 is the data size typical for the application of the homogenous rate models considered by
236 Louca & Pennell, the final estimates will depend on both choices. In other words, although λ_p
237 and r_p are theoretically (i.e. for infinitely large phylogenies) identifiable, in practice they must
238 be estimated from limited data. Without fixed hypotheses or constraints on functional
239 complexity, the exact estimation of the pulled rates can therefore still be challenging. For
240 example, Louca & Pennell chose a spline between an arbitrary number of discrete times as the
241 functional basis. As we illustrate below, the choice of the degree of the spline and of the
242 number of discrete times can influence estimates (see S.3). Our point is that some constraints,
243 either in the form of fixed hypotheses that are chosen *a priori*, or in the form of constraints on
244 model complexity, are fundamentally unavoidable in the problem of inferring diversification
245 histories from extant time trees.

246

247 **S.2 Regularization techniques can allow inference of diversification histories despite** 248 **unidentifiability**

249 Louca & Pennell write that “common model selection methods that are based on
250 parsimony or ‘Occam’s razor’ (such as the Akaike information criterion) generally cannot
251 resolve these issues” (i.e. issues of model congruence). We first note that the Akaike

252 information criterion is designed to account for the number of parameters in a (parametric)
253 model, while most data-driven approaches are non-parametric. In the latter case, other
254 approaches that are regularly used in statistics and machine learning to deal with issues of
255 unidentifiability and over-parametrization include the use of shrinkage¹³ or smoothing¹⁴
256 estimators, or of Bayesian priors. The authors don't discuss these approaches, but we see no
257 reason why they should not be considered in this case.

258 Regarding the use of model selection, two main arguments are provided by the authors
259 for dismissing these techniques as a possible solution to the identifiability problem (section
260 S2 in Louca & Pennell's paper, the third argument is essentially the same as the first). The
261 first argument is that "There is little reason to believe that the simplest scenario in a
262 congruence class will be the one closest to the truth. Indeed, even if the true model is included
263 in a congruence class, it will almost always be the case that there are both simpler and more
264 complex scenarios within the same congruence class and, crucially, all of these alternative
265 models remain equally likely even with infinitely large datasets." Nothing about this
266 statement is specific to birth-death models adjusted to timetrees, and thus we must conclude
267 that the authors question the principle of parsimony (or Occam's razor) in general. The
268 authors are of course entitled to their opinion, as the principle of parsimony is not a
269 mathematically provable law, but we note that this opinion contradicts centuries of thinking
270 and experience from physics to machine learning, and from philosophy as well. If we follow
271 the traditional thinking in science when presented with the situation highlighted by Louca &
272 Pennell (i.e. scenarios with different degrees of complexity but equal likelihood), we should
273 select the simplest scenario, because, although there is no guarantee that this scenario is the
274 closest to the truth, the parsimony principle suggests that it is most probably true, all other
275 things equal.

276 The second argument by Louca & Pennell is that the complexity of a diversification
277 scenario cannot be quantified. Note that this argument is only relevant if one accepts the
278 parsimony principle in the first place, or else complexity is irrelevant for model selection. We
279 agree with Louca & Pennell that quantifying complexity (or its opposite: simplicity) is a hard
280 question, and one that is not always well-addressed by simply counting the number of
281 parameters (the number of parameters is in particular not meaningful in non-parametric data-
282 driven approaches). However, this again is not specific to birth-death models adjusted to
283 timetrees, and a wide body of statistical literature exists that deals with this problem¹². A
284 review of this literature is beyond the scope of this discussion, but there are various routes to
285 quantifying or penalizing complexity, such as information-theoretical measures (in particular
286 the principle of maximum entropy²²), other Bayesian and frequentist model selection
287 approaches, and shrinkage or regularization estimators that essentially amount to setting
288 priors on the parameters or smoothness of a curve. While we agree that choosing among those
289 methods has a degree of subjectivity, we see no reason why complexity cannot be quantified
290 for diversification scenarios. It seems to us that the problem is very similar to the general
291 situation in data-driven analysis (starting from linear regressions), where we can always build
292 more complex models with equal likelihood, but know for various reasons (parsimony, bias-
293 variance trade-off) that adding complexity brings disadvantages, and thus add penalizations or
294 other regularization methods to balance the complexity of the explanation. There have already
295 been a few studies that tested regularization approaches for inference of diversification
296 rates^{15,16,23}, but the study by Louca & Pennell highlights the need for more research in this
297 field, to explore what regularization methods work best.

298 Louca & Pennell also allude to the fact that the problem highlighted in their study is
299 independent of sample size. We agree that this makes the situation somewhat different from
300 some other model selection and regularization problems, but we do not see how this is

301 relevant in a practical data analysis, where the size of the timetree is fixed and typically not
302 huge (in particular if we assume homogeneous rates), meaning that the problem at hand is to
303 draw inference from limited data.

304

305 **S.3 Illustration with the Madagascan vangas**

306 Here we begin by exploring models congruent to M, the model selected among the
307 four models we considered in our illustrative hypothesis-driven model selection procedure on
308 the Madagascan vangas tree. M is characterized by $\lambda(t) = \lambda_0 e^{\alpha t}$ and $\mu(t) = 0$ where t is
309 measured from the present to the past, $\lambda_0 = 0.018$ is the estimated present-day speciation
310 rate, and $\alpha = 0.1$ measures the estimated speed of time change.

311 We first choose the extinction function to be a constant $\mu_1^*(t) = \mu_0$ and compute
312 $\lambda_1^*(t)$. Note that the corresponding model M_1^* has 3 parameters (λ_0 , α and μ_0) while M has
313 only 2 (λ_0 and α), such that it would not be selected over M based on Occam's razor. M_1^*
314 would not be selected either if compared to the model ranked second in our analysis
315 (exponential change of speciation rate and significant extinctions), as it has the same number
316 of parameters and a lower likelihood (Table 1). We compute $\lambda_1^*(t)$ using the solution to Eq.2
317 from Louca & Pennell², provided in their SI (Eq. 39 & 40), with $\eta_0 = \lambda_0$ (here the sampling
318 fraction $\rho = 1$, as the Madagascan vangas tree is complete¹⁷):

319

$$320 \quad \lambda_1^*(t) = \frac{\lambda_0 e^{\Lambda(t)}}{1 + \lambda_0 \int_0^t e^{\Lambda(s)} ds}$$

321 with

$$322 \quad \Lambda(t) = \int_0^t [r_p(s) + \mu_0] ds.$$

323

324 r_p , the 'pulled diversification rate', is given by Eq. 1 from Louca & Pennell²:

325

$$326 \quad r_p = \lambda - \mu + \frac{1}{\lambda} \frac{d\lambda}{dt} = \lambda_0 e^{\alpha t} + \alpha.$$

327 This gives:

$$328 \quad \Lambda(t) = \frac{\lambda_0}{\alpha} (e^{\alpha t} - 1) + (\alpha + \mu_0)t$$

329

330 and

$$331 \quad e^{\Lambda(t)} = e^{-\frac{\lambda_0}{\alpha}} e^{(\alpha + \mu_0)t} e^{\frac{\lambda_0}{\alpha} e^{\alpha t}}.$$

332 Finally:

$$333 \quad \lambda_1^*(t) = \frac{\lambda_0 e^{-\frac{\lambda_0}{\alpha}} e^{(\alpha + \mu_0)t} e^{\frac{\lambda_0}{\alpha} e^{\alpha t}}}{1 + \lambda_0 e^{-\frac{\lambda_0}{\alpha}} \int_0^t e^{(\alpha + \mu_0)s} e^{\frac{\lambda_0}{\alpha} e^{\alpha s}} ds}.$$

334

335 Second, we choose the speciation function to be a constant $\lambda_2^*(t) = \lambda_0$ and compute
 336 $\mu_2^*(t)$. The corresponding model M_2^* has the same number of parameters as M. Solving Eq. 2
 337 from Louca & Pennell² with $\frac{d\lambda_2^*}{dt} = 0$ gives:

338

$$339 \quad \mu_2^*(t) = \lambda_2^* - r_p = \lambda_0 - \alpha - \lambda_0 e^{\alpha t}.$$

340

341 Next, we explore congruent models found by directly fitting r_p to the tree, following
 342 the procedure outlined in Louca & Pennell's SI (section S9), and with the same choice of
 343 keeping either λ or μ constant. We used the *fit_hbd_pdr_on_grid* function from the R package
 344 *castor*²³ to fit r_p . This function fits r_p assuming a piecewise spline function on a predefined
 345 "grid", i.e. discrete times between which r_p varies as a spline. As we did not find specific
 346 recommendations on how to define the grid, we chose evenly-spaced times, and selected their
 347 number using the AIC criterion (we note though that the AIC is not a good criterion for

348 rupture detection²⁵). We followed the recommendation from the help of the
349 *fit_hbd_pdr_on_grid* function to not use a spline of degree zero (i.e. piecewise constant); we
350 used a spline of degree of 1 (piecewise linear) and 2.

351 Using a spline of degree 1, we selected 3 times, and the resulting diversification
352 dynamics are shown in Fig. S1. The model that assumes λ is constant infers negative
353 extinction rates, and is therefore not plausible. The model that assumes μ is constant infers a
354 general tendency for a decline in speciation rates, with a local peak around 7 Myrs ago. This
355 result illustrates again the stability of the general temporal trend (decline) of speciation rates
356 to different fitting approaches. It also illustrates the potential use of congruent models to
357 generate new hypotheses. Here for example, the inferred dynamics suggest to formulate a
358 model with a declining background speciation rate combined with a local burst of speciation
359 around 7 Myrs ago.

360 Using a spline of degree 2, the AIC criterion selected only 1 time on the grid (i.e a
361 constant r_p), and the resulting diversification dynamics are shown in Fig. S2. The model that
362 assumes λ is constant infers negative extinction rates, and is therefore not plausible. The
363 model that assumes μ is constant infers a general tendency for a decline in speciation rates
364 that tends to accelerate over time. This result illustrates first that the estimation of r_p (and the
365 resulting dynamics of λ when assuming μ constant) depends on specific hypotheses made
366 when fitting r_p (compare with Fig. S1). Second, it illustrates one more time the stability of the
367 general temporal trend (decline) of speciation rates to different fitting approaches/choices. It
368 also provides another illustration of the potential use of congruent models to generate new
369 hypotheses. Here, the inferred dynamics suggest to formulate a model with a slow decline in
370 speciation rates at the origin of the clade, that accelerates towards the present.

371

372

373

374 **Supplementary References**

375

376 22. Kapur, J. N. Maximum-entropy models in science and engineering. *John Wiley & Sons*.

377 (1989).

378 23. May, M. R., Höhna, S., & Moore, B. R. A Bayesian approach for detecting the impact of

379 mass-extinction events on molecular phylogenies when rates of lineage diversification

380 may vary. *Methods Ecol. Evol.* **7**, 947-959 (2016).

381 24. Louca, S. & Doebeli, M. Efficient comparative phylogenetics on large trees.

382 *Bioinformatics* **34**, 1053–1055 (2018).

383 25. Arlot, S., Celisse, A., & Harchaoui, Z. A Kernel Multiple Change-point Algorithm via

384 Model Selection. *J. Mach. Learn. Res.* **20**, 1-56 (2019).

385

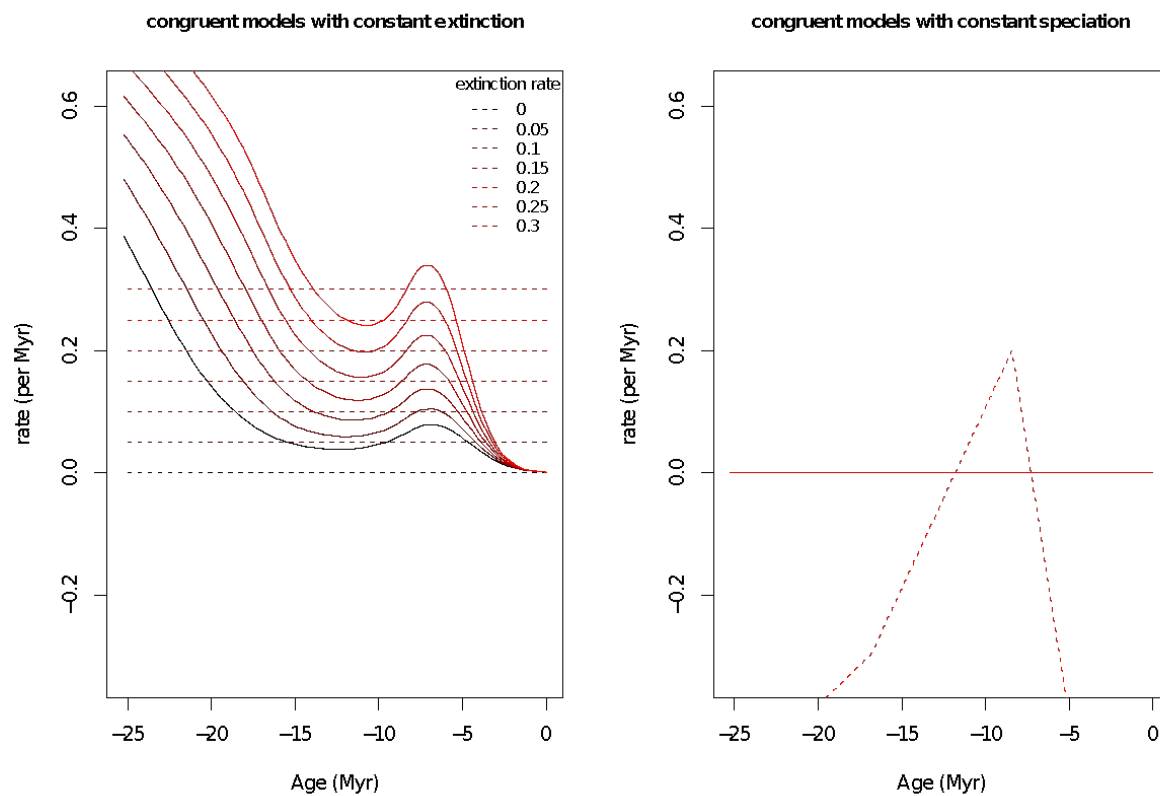
386

387

388

389

390



391

392 **Figure S1 Diversification of the Madagascan vangas as inferred by directly fitting r_p .**

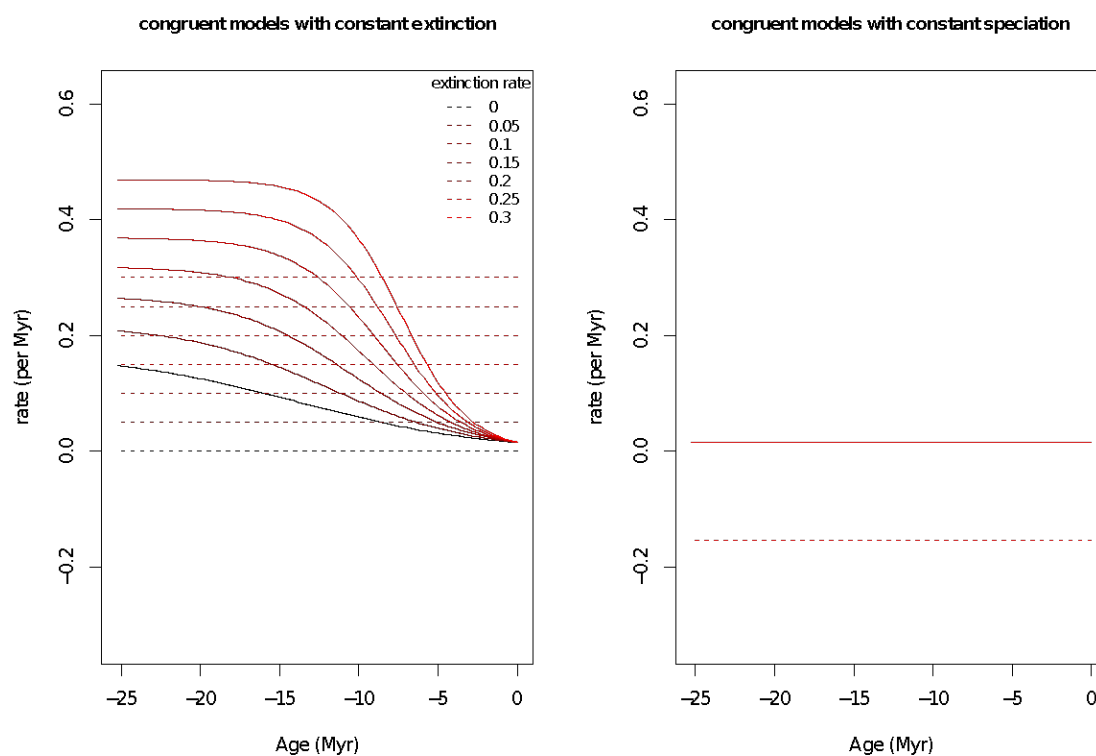
393 Here r_p is fitted with a spline of degree 1 (piecewise linear), with 3 evenly-spaced times

394 (selected based on AIC). The curves on the left panel represent the dynamics of speciation

395 (plain line) and extinction (dashed line) obtained by: fixing increasing values of a constant

396 extinction rate (left panel) and fixing the speciation rate to λ_0 (right panel).

397



398

399 **Figure S2 Diversification of the Madagascan vangas as inferred by directly fitting r_p .**

400 Here r_p is fitted with a spline of degree 2 with 1 time (selected based on AIC, i.e. r_p is

401 constant). The curves on the left panel represent the dynamics of speciation (plain line) and

402 extinction (dashed line) obtained by: fixing increasing values of a constant extinction rate (left

403 panel) and fixing the speciation rate to λ_0 (right panel).

404

405

406

407