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

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

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

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

Abstract

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

Main text

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

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

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

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

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

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

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

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

We now follow Louca & Pennell² by solving Eq. 2 for models congruent to our best model

M. First, we choose the extinction function to be a constant $\mu_1^*(t) = \mu_0$ and compute $\lambda_1^*(t)$.

Second, we choose the speciation function to be a constant $\lambda_2^*(t) = \lambda_0$ and compute $\mu_2^*(t)$.

We find (SI; see the corresponding inferred dynamics in Fig. 1):

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

and

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

The biological interpretation of these additional models and of their parameters is not

obvious. The equation for μ_2^* looks more interpretable at first, but it expresses the temporal change and the extinction rate at present through the same parameter α , which means that a positive extinction rate at present ($\alpha < 0$) will force extinction rates to decline over time.

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

Finally, we fit r_p directly with spline functions, as suggested by Louca & Pennell² (SI S.3),

and deduce λ (or μ) when assuming μ (or λ) constant. General diversification trends are

consistent with those found above; finer dynamics depend on the choices made to estimate r_p

and suggest interesting new hypotheses to test (SI S.3). Here we chose λ or μ to be constant to

simplify the problem. Data-driven approaches have relaxed this hypothesis while still

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

There are useful results in Louca & Pennell's paper. The pulled diversification rate is difficult to interpret biologically, yet realizing that models are only distinguishable by this quantity is important, and the approach might provide an interesting way to generate new hypotheses. The implications of these results for diversification analyses, however, are largely overinterpreted, mainly because the constraints imposed by the practice of hypothesis-driven research, prior knowledge, and the possibilities to penalize complexity are not considered in the paper. A very similar identifiability issue occurred 10 years ago in population genetics, when it was shown that the widely-used allelic frequency spectrum (AFS) is consistent with a myriad of demographic histories¹⁹. Despite this insight, the AFS – along with other data sources – remains a prominent approach of demographic inference²⁰. Identifiability issues naturally arise in approaches that try to infer the potentially unlimited complexity of historical processes from limited contemporary data, and this is why we work hypothesis-driven, develop regularization techniques, and integrate other data types. Louca & Pennell² remind us that the hypotheses we formulate when we develop models influence the conclusions we draw, and that these conclusions should always be taken with a grain of salt; this is always a good reminder, but it does not compromise the current practice in the field.

References

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

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

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

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

Model	nb	Log L	AICc	ΔAIC	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

nb denotes the number of parameters. LogL stands for the maximum log-likelihood, AICc for the second order Akaike Information Criterion⁴, and ΔAIC for the difference in AICc between the model and the best model in the set. Models are ranked based on 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 reflects a rate decline), and μ_0 is the estimated extinction rate. Fits were performed using the *fit_bd* function from RPANDA²¹ using the “crown” conditioning and a sampling fraction of 1 (the tree of the Madagascan vangas is complete¹⁷).

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

The black curves on the left panel represent the dynamics of speciation (plain line) and extinction (dashed line) corresponding to our best-fit model M (exponential decline in speciation rate, non-significant extinctions). The colored curves illustrate the rate dynamics of congruent models obtained by: fixing increasing values of a constant extinction rate (left panel, M_1^*) and fixing the speciation rate to λ_0 (right panel, M_2^*). Note that M_1^* infers a speciation rate decline regardless of the assumed extinction rate.

Supplementary Information

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

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

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

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

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

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

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

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

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

S.3 Illustration with the Madagascan vangas

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

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

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

with

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

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

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

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

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

Supplementary References

22. Kapur, J. N. Maximum-entropy models in science and engineering. *John Wiley & Sons*. (1989).
23. May, M. R., Höhna, S., & Moore, B. R. A Bayesian approach for detecting the impact of mass-extinction events on molecular phylogenies when rates of lineage diversification may vary. *Methods Ecol. Evol.* **7**, 947-959 (2016).
24. Louca, S. & Doebeli, M. Efficient comparative phylogenetics on large trees. *Bioinformatics* **34**, 1053–1055 (2018).
25. Arlot, S., Celisse, A., & Harchaoui, Z. A Kernel Multiple Change-point Algorithm via Model Selection. *J. Mach. Learn. Res.* **20**, 1-56 (2019).

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

Here r_p is fitted with a spline of degree 1 (piecewise linear), with 3 evenly-spaced times (selected based on AIC). The curves on the left panel represent the dynamics of speciation (plain line) and extinction (dashed line) obtained by: fixing increasing values of a constant extinction rate (left panel) and fixing the speciation rate to λ_0 (right panel).

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

Here r_p is fitted with a spline of degree 2 with 1 time (selected based on AIC, i.e. r_p is constant). The curves on the left panel represent the dynamics of speciation (plain line) and extinction (dashed line) obtained by: fixing increasing values of a constant extinction rate (left panel) and fixing the speciation rate to λ_0 (right panel).