

Response to Technical Comment ‘A cautionary note for users of linear diversification dependencies’

Hélène Morlon^{1,*}, Jonathan Rolland^{2,3,#} and Fabien L. Condamine^{4,#}

¹ Institut de Biologie de l'ENS (IBENS), Ecole Normale Supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France.

² Department of Zoology, University of British Columbia, #4200-6270 University Blvd., Vancouver, B.C., Canada

³ Department of Computational Biology, University of Lausanne, Quartier Sorge, 1015 Lausanne, Switzerland

⁴ Institut des Sciences de l'Evolution de Montpellier (Université de Montpellier | CNRS | IRD | EPHE), Place Eugène Bataillon, 34095 Montpellier, France

* corresponding author, helene.morlon@bio.ens.psl.eu

co-last authors

Abstract

The comment by Gamisch (2020) draws the attention of users of the R-package RPANDA (Morlon *et al.* 2016) on situations when properly interpreting the results of linear diversification dependencies requires caution. Here we provide clarifications to help users interpreting their results when using any type of functional diversification dependencies with time or the environment.

In our paper “*Assessing the causes of diversification slowdowns: temperature-dependent and diversity-dependent models receive equivalent support*” (Condamine *et al.* 2019), we fitted several diversification models to 218 phylogenies of tetrapod families to confirm that diversification slowdowns are common, and to show that this widespread pattern may in part be associated to cooling temperatures during the Cenozoic. Gamisch (2020) does not question these main results, but comments on the interpretation of linear diversification dependencies fitted in the R-package RPANDA (Morlon *et al.* 2016). Here we take this opportunity to provide further clarifications and recommendations for users.

As explained in Morlon *et al.* (2016), the *fit_bd* function in RPANDA provides a flexible framework for fitting a birth-death model to a phylogeny of extant taxa with any type of user-defined functional form of time-dependency for the speciation rate (λ) and the extinction rate (μ). Here we focus on the estimation of speciation rates, although our comments extend to extinction. In several of our papers, we have used constant, linear, and/or exponential time-dependent diversification models (e.g. Morlon *et al.* 2011; Morlon *et al.* 2012; Lewitus & Morlon 2018; Lewitus *et al.* 2018; Condamine *et al.* 2019; Rolland & Condamine 2019; Billaud *et al.* 2020), but a user could in principle use any other type of dependency, by specifying the function *f.lamb*. The function *f.lamb* has two arguments, time (t) and a vector of parameters to be estimated, which are the speciation rate λ_0 in the case of a constant rate model, and λ_0 and α in the cases of a linear dependency of the form $\lambda(t) = \lambda_0 + \alpha t$, or an exponential dependency of the form $\lambda(t) = \lambda_0 \times e^{\alpha t}$. The parameters are estimated by maximum likelihood, meaning that the *fit_bd* function searches the parameters that maximize an internally specified likelihood function. This search is performed using the Nelder-Mead algorithm (Nelder & Mead 1965), which does not allow defining bounds; therefore the algorithm can in principle explore parameter values that return negative values of *f.lamb*,

regardless of the functional form used (i.e. even in the case of constant rates or exponential dependencies). However, speciation rates are positive by definition, and so such parameter values should not be retained as plausible solutions. In order to speed up convergence, instead of simply rejecting the parameters that lead to negative $f.lamb$ values, we evaluate the likelihood function for $abs(f.lamb)$. This has several implications for the interpretation of the outputs of the fit_bd function, detailed below.

Starting from the case of fitting a simple constant rate model, the fit_bd function can return negative rate estimates. We have indeed explained this individually to users who have contacted us. As explained above, the optimisation algorithm can explore negative regions of parameter space, but it is in fact the absolute value of this parameter that is used in the computation of the likelihood (as should be). If the algorithm returns a negative rate estimate for λ_0 it should thus be interpreted in absolute terms $|\lambda_0|$. Similarly, in the case of an exponential time-dependency, if the algorithm returns a negative estimate for λ_0 , the time dependency should be interpreted as $|\lambda_0 e^{\alpha t}| = |\lambda_0| e^{\alpha t}$. In the case of the linear time-dependency that Gamisch (2020) focuses on, the time-dependency should similarly be interpreted as $|\lambda_0 + \alpha t|$. While in the case of an exponential dependency the interpretation of the sign of α is not affected by the transformation (a positive α reflects a slowdown in speciation), in the case of a linear dependency it can be affected. We have also explained this individually to users who have contacted us (including A. Gamisch), and there is a post about it on the RPANDA *GitHub* repository (see for example <https://github.com/hmorlon/PANDA/issues/11>).

Other RPANDA models are such that chunks of the parameter space can lead to negative rates on parts of the history of clades. For speciation and extinction rates, it concerns the

fit_env function testing the effect of past environmental variations on diversification rates (Condamine *et al.* 2013; Lewitus & Morlon 2018). For rates of phenotypic evolution, it concerns the *fit_t_env* function testing time- and environmental-dependencies in these rates (Clavel & Morlon 2017), the *fitTipData* function fitting various phenotypic models including coevolutionary models (Manceau *et al.* 2017), and the *fit_t_comp* and *fit_t_comp_subgroup* functions fitting the matching competition and diversity-dependent models of phenotypic evolution (Drury *et al.* 2016, 2017). The *fit_env* function is constructed based on the *fit_bd* function, and also uses absolute values of rates in the likelihood function, with similar implications for the interpretation of results. The other functions instead return infinite values of the likelihood for parameters that lead to negative rates, therefore simply rejecting parameters that lead to negative rates on part of the history of the group. In this case no transformation is needed, and the interpretation of the sign of the parameters output of these functions is therefore straightforward.

These clarifications being given, we agree with Gamisch (2020) that our sentence (“*A positive α (β) reflects a slowdown of speciation (extinction) towards the present, while a negative α (β) reflects a speed-up of speciation (extinction) towards the present*” p. 1902) is imprecise. Similarly, the sentence (“*A positive α (β) indicates that speciation (extinction) rates are higher under warm climatic periods, while a negative α (β) indicates that speciation (extinction) rates are higher under cold climatic periods*” p. 1903) is imprecise. For 9% of our empirical phylogenies (20/218), parameter estimates obtained with a linear time-dependency lead to speciation or extinction rates that are negative on part of the history of the corresponding clades and should thus be transformed in absolute values; in the case of a linear temperature-dependency, this occurred for 3.7% of the phylogenies (8/218). The interpretation of the sign of alpha (beta) in these cases is less straightforward. We note that

these “symptomatic” cases affect mostly extinction (13 of the 20 time-dependent cases and 7 of the 8 temperature-dependent cases), which is often difficult to estimate from phylogenies of extant taxa. Just to be on the safe side, we checked the consequences of removing linear-dependent models of both time and temperature from our original analyses. When comparing only time-dependent and constant-rate models, we found that time-dependent models are the preferred model (i.e. models with lowest AICc score) for 43% of the phylogenies (93/218). When time-dependent models were preferred, this corresponded to a diversification slowdown in 78% of the phylogenies (73/93). When adding and comparing also temperature-dependent models, we found that these models are the preferred models for 39% of the phylogenies (85/218), constant-rate models for 29.8% of the phylogenies (65/218), diversity-dependent models for 24.8% of the phylogenies (54/218), and time-dependent models for 6.4% of the phylogenies (14/218). When temperature-dependent models were preferred, the dependency with temperature was positive in 88% of the phylogenies (75/85). Hence our main conclusions that diversification slowdowns are common and that this pattern may in part be associated to cooling temperatures during the Cenozoic hold.

While the interpretation of the sign of α in the case of linear diversification dependencies can be complicated by the above considerations, we do not think that it should drive users away from using these dependencies (or other dependencies that can occasionally generate negative rates). Indeed, the *plot_fit_bd* and *plot_fit_env* functions of RPANDA directly apply the absolute transformation to avoid misinterpretations. While this might not be convenient when carrying analyses on many phylogenies such as in our study, it should efficiently guide a user in the interpretation of diversification dynamics in clades of interest. Alternatively, a user can simply decide to specify a function *f.lamb* that equals 0 in case rates become negative (e.g. $\lambda(t) = \max(0, \lambda_0 + \alpha t)$) as for example in Rabosky & Glor (2010), Morlon *et al.* (2011), or

Etienne *et al.* (2012). Both options for *f.lamb* may lead to biologically questionable scenarios with speciation rates estimated at 0 on part of the history of clades (when using max), or with speciation rates first decreasing and then increasing (in the case of abs). This is due to the simplicity of the linear term. A pragmatic approach is to fit both models, which can easily be done in RPANDA, and choose the solution with highest likelihood support.

We have edited the help files associated with the *fit_bd* and *fit_env* functions to clarify the points outlined in this comment and to provide an explicit example for a user who adopts the $\lambda(t) = \max(0, \lambda_0 + at)$ option. We hope that these clarifications will help users of the R-package RPANDA interpreting their results when using any type of functional diversification dependencies with time or the environment.

Acknowledgements

We thank Julien Clavel, Jonathan Drury and Marc Manceau for checking how potential negative rates of phenotypic evolution were handled in the *fit_t_env*, *fitTipData*, *fit_t_comp* and *fit_t_comp_subgroup* functions of the R-package RPANDA. The development of RPANDA is supported by grant ERC-CoG PANDA to HM.

References

- Billaud, O., Moen, D.S., Parsons, T.L. & Morlon, H. (2020). Estimating diversity through time using molecular phylogenies: Old and species-poor frog families are the remnants of a diverse past. *Syst. Biol.*, 69, 363-383.
- Clavel, J. & Morlon, H. (2017). Accelerated body size evolution during cold climatic periods in the Cenozoic. *Proc. Natl Acad. Sci. USA*, 114, 4183-4188.

- Condamine, F.L., Rolland, J. & Morlon, H. (2013). Macroevolutionary perspectives to environmental change. *Ecol. Lett.*, 16, 72-85.
- Condamine, F.L., Rolland, J. & Morlon, H. (2019). Assessing the causes of diversification slowdowns: temperature-dependent and diversity-dependent models receive equivalent support. *Ecol. Lett.*, 22, 1900-1912.
- Drury, J., Clavel, J., Manceau, M. & Morlon, H. (2016). Estimating the effect of competition on trait evolution using maximum likelihood inference. *Syst. Biol.*, 65, 700-710.
- Drury, J.P., Tobias, J.A., Burns, K.J., Mason, N.A., Shultz, A.J. & Morlon, H. (2018). Contrasting impacts of competition on ecological and social trait evolution in songbirds. *PLoS Biol.*, 16, e2003563.
- Etienne, R.S., Haegeman, B., Stadler, T., Aze, T., Pearson, P.N., Purvis, A. *et al.* (2012). Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proc. R. Soc. B*, 279, 1300–1309.
- Lewitus, E. & Morlon, H. (2018). Detecting environment-dependent diversification from phylogenies: a simulation study and some empirical illustrations. *Syst. Biol.*, 67, 576–593.
- Lewitus, E., Bittner, L., Malviya, S., Bowler, C. & Morlon, H. (2018). Clade-specific diversification dynamics of marine diatoms since the Jurassic. *Nat. Ecol. Evol.*, 2, 1715-1723.
- Manceau, M., Lambert, A. & Morlon, H. (2017). A unifying comparative phylogenetic framework including traits coevolving across interacting lineages. *Syst. Biol.*, 66, 551-568.
- Morlon, H., Parsons, T. & Plotkin, J.B. (2011). Reconciling molecular phylogenies with the fossil record. *Proc. Natl Acad. Sci. USA*, 108, 16327-16332.

- Morlon, H., Kempes, B.D., Plotkin, J.B. & Brisson, D. (2012). Explosive radiation of a bacterial species group. *Evolution*, 66, 2577-2586.
- Morlon, H., Lewitus, E., Condamine, F.L., Manceau, M., Clavel, J. & Drury, J. (2016). RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. *Methods Ecol. Evol.*, 7, 589–597.
- Nelder, J.A. & Mead, R. (1965). A simplex algorithm for function minimization. *Computer J.*, 7, 308–313.
- Rabosky, D.L. & Glor, R.E. (2010). Equilibrium speciation dynamics in a model adaptive radiation of island lizards. *Proc. Natl. Acad. Sci. USA*, 107, 22178-22183.
- Rolland, J. & Condamine, F.L. (2019). The contribution of temperature and continental fragmentation to amphibian diversification. *J. Biogeogr.*, 46, 1857–1873.