**Teasing apart mountain uplift, climate change and biotic drivers of species diversification**

*Fabien L. Condamine\*, Alexandre Antonelli, Laura P. Lagomarsino, Carina Hoorn and Lee Hsiang Liow*

***Author for correspondence (\*):*** fabien.condamine@gmail.com

**Manual to use the R script provided in GitHub**

October 15, 2015 – Fabien L. Condamine

The following R script aims at testing the prevalence of the *Red Queen* and the *Court Jester* using time-calibrated phylogenies only. The script basically takes a tree (or multiple trees), and performs several birth-death models as explained in the chapter, and detailed below.

The Court Jester analyses are represented by the *PaleoEnv model* (Condamine *et al.*, 2013) in which one can set an environmental variable that itself varies through time (e.g. temperature, sea-level). The model will thus estimate whether the speciation/extinction varied according to the past environment, and to which extent.

The Red Queen analyses are made with the *diversity-dependent (DDD) model* that estimates whether a clade has reached its ecological limits, i.e. its carrying capacity (Etienne *et al.*, 2012). Also, one can easily change the *PaleoEnv model* into a *multiple clade diversity-dependent model* if he has past diversity curve of an interacting clade with the focal clade. This will estimate the correlation between the focal clade’s speciation/extinction and the diversity dynamics of the extra clade.

In the book chapter, Condamine *et al.* (2017) used this general approach, although a simplified one, on the Andean clade of hummingbirds. They tested whether the Andean orogeny, or change in global past temperatures, or the role of intra-clade ecological interactions best explain the diversification of the clade using that framework.

The framework is highly flexible, as explained in the chapter, and may include other models to test for the prevalence of the Red Queen and Court Jester. For instance, the time-dependent models of Morlon et al. (2011) and Stadler (2011) are included, but were not tested on the Andean clade of hummingbirds. The time-dependent models, PaleoEnv models, and DDD models are comparable based on their corrected Akaike Information Criterion (AICc, see the chapter for an explanation). Note that the TreePar model is not comparable to the others.

**Input files/data**

1. A *tree file* or sample of trees in *NEXUS* format (e.g. from BEAST)
2. The *sampling fraction*, i.e. proportion of sampled species out of the total diversity of the clade (e.g. 0.1 = 10% sampled diversity, 1 = complete phylogeny)
3. One or several *paleo-environmental data* (e.g. temperature, sea level, Andean orogeny; see folder “*PaleoEnv*” for the paleo-environmental variables)

**Requirements**

1. **R** (get the latest version at: https://cran.r-project.org/)
2. **R libraries**: *DDD*, *picante*, *pspline*, *TreePar* (the R script will automatically check for that and will install if not yet installed):
   1. *install.packages(“DDD”, dependencies=T)*
   2. *install.packages(“picante”, dependencies=T)*
   3. *install.packages(“pspline”, dependencies=T)*
   4. *install.packages(“TreePar”, dependencies=T)*

**Run the analysis**

1. Create a folder with input data and the script “***run\_diversification\_analyses.R***”, and directories “*diversification\_library*” and “*PaleoEnv*”.
2. Open R and change your working directory to your folder (e.g. using ***setwd(“~/path\_to\_your\_files/”)***), or navigate to the directory (e.g. using ctrl D on Windows, or Cmd D on Apple).
3. Type: ***source("run\_diversification\_analyses.R")***
4. Choose among the following functions:
   1. *run\_Morlon\_models* (it tests 6 time-dependent models with exponential variation of the rates, and 6 with linear variation)
   2. *run\_PaleoEnv* (it tests 4 paleoenvironment-dependent models with exponential variation of the rates, and 4 with linear variation)
   3. *run\_DDD* (it tests six DDD models)
   4. *run*\_*TreePar* (it tests up to 4 shifts of diversification)

Each function requires the name of the tree file (argument: *tree\_file*), the sampling fraction (argument: *sampling\_fraction*; default=1) or for DDD the initial richness (argument: *total\_richness*; default=*Ntip(tree\_file)*), and the number of trees for analysis (argument: *number\_of\_trees*; default = 1). Note that for DDD, analyzing more than one tree is highly time consuming, so we usually set the default value.

1. The *run\_Morlon\_models* is the most straightforward analysis, and is ran as follows:

Example: ***run\_Morlon\_models("Test\_treeM.tre", sampling\_fraction=0.9, number\_of\_trees=2)***

1. The *run\_PaleoEnv* requires an additional argument, named *env\_data\_file*, specifying the file containing paleo-environmental variables (e.g. files in PaleoEnv directory).

Example: ***run\_PaleoEnv("Test\_treeM.tre", env\_data\_file="./PaleoEnv/PastSeaLevel.txt", sampling\_fraction=0.9, number\_of\_trees=2)***

1. The *run\_DDD* requires an additional argument, named *total\_richness*, denoting the contemporary species diversity of the focal clade (i.e. the number of described species). By default, this assumes that the tree contains the total diversity of the clade, so *total\_richness=Ntip(tree\_file)*.

Example: ***run\_DDD("Test\_tree.tre", total\_richness=50, number\_of\_trees=1)***

1. The *run\_TreePar* function has an additional argument named *grid* (default = 0.1), type: “*?bd.shifts.optim*” for additional information.

Example: ***run\_TreePar("Test\_treeM.tre", sampling\_fraction=0.9, grid=1, number\_of\_trees=3)***

**Results**

The analyses will save two output files in the same directory as the input tree file:

1. R data file (with the complete analysis, which can be opened in R using *attach(file\_name.Rdata)*)
2. Table with all results (a tab-separated *txt* file, which can be opened in Excel): model names, number of parameters, likelihood values, AICc scores, and maximum likelihood estimates of the parameters (mean values and standard errors are calculated when running analysis on a sample of trees):
   1. Model name
   2. Number of parameters
   3. Log-Likelihood
   4. AICc score (smallest value = best model)
   5. Lambda: speciation rate
   6. Alpha: correlation parameter between speciation rate and environmental data
   7. Mu: extinction rate
   8. Beta: correlation parameter between extinction rate and environmental data
   9. K: the estimated carrying capacity for the clade, i.e. number of niches
   10. r: ratio of linear dependencies in speciation and extinction rates
   11. DivRate – diversification rate
   12. Turnover – extinction fraction (lambda/mu)
   13. ShiftTime – times of rate shift

Please, cite the book chapter and appropriate references (see below) when using this script and approach. Also, don’t hesitate to contact me if you have any trouble or question.

**Acknowledgements**

Thanks to Andrea S. Meseguer for beta testing the pipeline.

**References**

Etienne R.S., Haegeman B., Stadler T., Aze T., Pearson P.N., Purvis A., Phillimore A.B. 2012. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proc. Roy. Soc. Lond. B*, **279**, 1300–1309.

Morlon H., Parsons T.L., Plotkin, J. 2011. Reconciling molecular phylogenies with the fossil record. *Proc. Natl. Acad. Sci. USA,* **108,** 16327-16332.

Condamine F.L., Rolland J., Morlon H. (2013) Macroevolutionary perspectives to environmental change. *Ecol. Lett.,* **16,** 72-85.

Condamine F.L., Antonelli A., Lagomarsino L.P., Hoorn C., Liow L.H. 2017. Teasing apart mountain uplift, climate change and biotic drivers of species diversification. In: *Mountains, Climate, and Biodiversity* (eds. Hoorn C., Antonelli A.). Wiley Blackwell, pp. xx-xx.

Stadler T. 2011. Mammalian phylogeny reveals recent diversification rate shifts. *Proc. Natl. Acad. Sci. USA,* **108,** 6187-6192.