

1 **RPANDA: an R package for macroevolutionary analyses on**
2 **phylogenetic trees**

3

4 Morlon, H.^{1*}, Lewitus, E.¹, Condamine, F.L.², Manceau, M.¹, Clavel, J.¹, Drury, J.¹

5

6 ¹ CNRS UMR 8197, Institut de Biologie, Ecole Normale Supérieure, 46 rue d'Ulm, 75005

7 Paris

8 ² CNRS UMR 5554, Institut des Sciences de l'Evolution, Montpellier, Place Eugène

9 Bataillon, 34000 Montpellier, France

10

11 *Corresponding author: morlon@biologie.ens.fr

12

13 **Short title:** RPANDA package for macroevolutionary analyses

14

15

16 **Summary**

17

18 1. A number of approaches for studying macroevolution using phylogenetic trees
19 have been developed in the last few years. Here, we present RPANDA, an R
20 package that implements model-free and model-based phylogenetic comparative
21 methods for macroevolutionary analyses.

22 2. The model-free approaches implemented in RPANDA are recently developed
23 approaches stemming from graph theory that allow summarizing the information
24 contained in phylogenetic trees, computing distances between trees, and
25 clustering them accordingly. They also allow identifying distinct branching
26 patterns within single trees.

27 3. RPANDA also implements likelihood-based models for fitting various
28 diversification models to phylogenetic trees. It includes birth-death models with
29 i) constant, ii) time-dependent, and iii) environmental-dependent speciation and
30 extinction rates. It also includes models with equilibrium diversity derived from
31 the coalescent process, as well as a likelihood-based inference framework to fit
32 the individual-based model of Speciation by Genetic Differentiation, which is an
33 extension of Hubbell's Neutral Theory of Biodiversity.

34 4. RPANDA can be used to: i) characterise trees by plotting their spectral density
35 profiles ii) compare trees and cluster them according to their similarities, iii)
36 identify and plot distinct branching patterns within trees, iv) compare the fit of
37 alternative diversification models to phylogenetic trees, ii) estimate rates of
38 speciation and extinction, iii) estimate and plot how these rates have varied with
39 time and environmental variables, and iv) deduce and plot estimates of species
40 richness through geological time.

41 5. RPANDA provides investigators with a set of tools for exploring patterns in
42 phylogenetic trees and fitting various models to these trees, thereby contributing
43 to the on-going development of phylogenetics in the life sciences.

44

45 **Keywords:** diversification rates, speciation, extinction, likelihood, macroevolution,
46 speciation by genetic differentiation, graph Laplacian, spectral density profiles.

47

48

49

50 **Introduction**

51 Phylogenetic approaches have become a central component of various areas of the life
52 sciences. A number of packages are available to handle and utilise phylogenetic trees
53 (e.g. 'ape' Paradis *et al.* 2004), in order to understand, for example, community assembly
54 (e.g. 'picante' Kembel *et al.* 2010, 'DAMOCLES' Pigot & Etienne 2015), trait evolution (e.g.
55 'Coevol' Lartillot & Poujol 2011, 'geiger' Pennell *et al.* 2014), and diversification (e.g.
56 'BayesRate' Silvestro *et al.* 2011, 'TreePar' Stadler 2011a, 'diversitree' Fitzjohn 2012,
57 'DDD' Etienne *et al.* 2012, 'geiger' Pennell *et al.* 2014, 'BAMM' Rabosky *et al.* 2014).

58 Here, we present and describe the R package RPANDA, which implements both
59 model-free and model-based phylogenetic approaches that are not implemented in
60 previous packages. The model-free approaches are inspired from graph theory and
61 described in detail in Lewitus & Morlon (2015). They are designed to compare and
62 classify phylogenetic trees without any *a priori* formulation of a model of cladogenesis
63 underlying tree shape. The model-based approaches aim to fit various models of
64 cladogenesis to phylogenetic trees by maximum likelihood. They are described in detail
65 in Morlon *et al.* (2010, 2011), Condamine *et al.* (2013) and Manceau *et al.* (2015).

66 Within the category of model-based approaches, the package includes three main
67 classes of models: birth-death models, coalescent models, and individual-based models
68 (see Morlon 2014 for a review of these three different types of models). Birth-death
69 models are those that were originally considered by Nee *et al.* (1992) and are at the
70 basis of most diversification models used today. The present package allows fitting of
71 birth-death models: i) with speciation and extinction rates varying as a function of time,
72 with any type of functional form and with potential periods of diversity decline (Morlon
73 *et al.* 2011), and ii) with speciation and extinction rates varying as a function of any
74 variable that has been measured through geological time (e.g., temperature, Condamine

75 *et al.* 2013). Models based on the coalescent process were designed to consider
76 scenarios of equilibrium diversity (Morlon *et al.* 2010). Finally, individual-based models
77 have been extensively studied in ecology, but likelihood-based inferences from
78 phylogenies are typically not available. The present package allows likelihood inference
79 for the model of Speciation by Genetic Differentiation (Manceau *et al.* 2015), which is an
80 extension of the Neutral Theory of Biodiversity (Hubbell 2001).

81

82 **Description**

83 RPANDA is an R package (R Development Core Team 2014) that can be installed from
84 the CRAN repository (<http://cran.r-project.org>). RPANDA relies on the R packages ‘ape’
85 (Paradis *et al.* 2004), ‘picante’ (Kembel *et al.* 2010), ‘phytools’ (Revell 2012), ‘deSolve’
86 (Soetaert *et al.* 2010), and ‘igraph’ (Csardi & Nepusz 2006).

87 The main functions of the package are listed in Table 1, classified into functions related
88 to the model-free and model-based approaches. Nearly every function is associated with
89 a ‘plot’ function that helps visualize the results of the analyses.

90 To illustrate the use of RPANDA, we analyse the phylogeny of the bat family
91 Phyllostomidae. This phylogeny is the maximum clade credibility tree used in Rolland *et*
92 *al.* (2014), which originally comes from the mammalian supertree (Bininda-Emonds *et*
93 *al.* 2007; Fritz *et al.* 2009); it contains 150 of the 165 known bat species (i.e., it is 91%
94 complete). To begin, we open an R console, and we install and load the RPANDA package
95 as well as the example datasets.

```
96 > install.packages('RPANDA', dependencies=TRUE)
```

```
97 > library(RPANDA)
```

```
98 > data(Phyllostomidae)
```

```
99 > data(Phyllostomidae_genera)
```

100 'Phyllostomidae' is the family-level phylogeny and 'Phyllostomidae_genera' is a list of 25
101 phylogenies corresponding to Phyllostomidae genera with more than one species.

102

103 **Characterising and comparing phylogenies using spectral densities**

104 We recently developed a new approach, described in detail in Lewitus & Morlon (2015),
105 to efficiently summarize the shape of a phylogenetic tree. This approach can be used to
106 measure similarities between trees and to cluster them accordingly, for example in
107 order to identify phylogenies shaped by similar versus dissimilar diversification
108 patterns. It can also help in identifying regions of a tree that have distinct branching
109 patterns, which can for example reflect shifts in modes or rates of diversification. We
110 summarize the shape of a phylogeny by its spectral density, which is a smoothed version
111 of the frequencies of eigenvalues associated with a matrix (the graph Laplacian) built
112 from the pairwise phylogenetic distances between nodes (see Lewitus & Morlon 2015
113 for a more detailed description). The function `spectR` computes the eigenvalues
114 associated with a given phylogeny, and characteristics associated with the spectrum of
115 eigenvalues, namely the principal eigenvalue, asymmetry (skewness), two measures of
116 peakedness (kurtosis and peak height), and `eigengap`. The `eigengap` is given by the
117 position of the largest difference between successive eigenvalues listed in descending
118 order. This number is related to the number of peaks in the spectral density plot and is
119 indicative of the number of modalities (i.e. distinct branching patterns) in a phylogeny.

```
120 > res<-spectR(Phyllostomidae)
```

121 returns the above information for the Phyllostomidae phylogeny. In particular,
122 `res$eigengap` returns the number of modalities, suggesting three distinct branching
123 patterns in this bat family.

```
124 > plot_spectR(res)
```

125 displays the spectral density profile and a plot of the eigenvalues ranked in descending
126 order (Figure 1).

127 Once the putative number of modalities is identified, the `BICcompare` function
128 can be used to assess the significance of these modalities and to identify their location
129 on the phylogeny. The statistical significance of the modalities is assessed by comparing
130 the Bayesian Information Criterion (BIC) for detecting i clusters in the distance matrix of
131 the empirical phylogeny and in randomly bifurcating trees parameterized on that tree
132 (Lewitus & Morlon 2015). The function also identifies the location of the distinct
133 branching patterns on the phylogeny by k-means clustering and returns the ratio of
134 between-cluster sum of squares (BSS) to total sum of squares (TSS) for the clustering
135 identified by the algorithm. The highest the BSS/TSS ratio, the more distinct the
136 modalities are from each other. Different iterations of the k-means clustering algorithm
137 can lead to different modality configurations, and BSS/TSS values allow the comparison
138 between these configurations (configurations with high BSS/TSS should be preferred).

```
139 > res<-BICcompare(Phyllostomidae,3)
```

140 returns the above information for the Phyllostomidae phylogeny. The BIC score for the
141 Phyllostomidae phylogeny is nearly a magnitude smaller than it is for the randomly
142 bifurcating trees parameterized on that phylogeny, suggesting that the three modalities
143 are significant. Typically, a BIC ratio ≤ 0.25 is deemed significant.

```
144 > plot_BICcompare(Phyllostomidae,res)
```

145 displays the Phyllostomidae phylogeny with branches coloured according to the
146 modality they belong to, as assessed by the k-means clustering algorithm (Figure 2).

147 Spectral densities are particularly useful for comparing phylogenies. The
148 `JSDtree` function computes the pairwise distances between a list of phylogenies,
149 measured as the Jensen-Shannon distance between their spectral densities. The

150 JSDtree_cluster function uses these pairwise distances (or potentially other distance
151 metrics) to cluster phylogenies into groups. The clustering is implemented using the
152 hierarchical clustering and k-medoid clustering algorithms. To illustrate this approach,
153 we compare the 25 phylogenies corresponding to Phyllostomidae genera with more
154 than one species.

```
155 > res<-JSDtree(Phyllostomidae_genera)
```

156 returns the matrix containing the pairwise Jensen-Shannon distances between the 25
157 phylogenies.

```
158 > JSDtree_cluster(res)
```

159 plots the heatmap and hierarchical cluster (Figure 3) as well as the hierarchical cluster
160 with bootstrap support (not shown here). It also returns the optimal number of clusters
161 given by the k-medoids algorithm, here suggesting that Phyllostomidae genera cluster
162 into two meaningful groups. The function returns the assignment of each phylogeny to
163 each of the two groups, as well as a measure of statistical support for this assignment.

164

165 **Fitting models of diversification to phylogenies**

166 One of the most popular approaches for analysing the diversification of clades consists
167 in fitting various models of diversification to molecular phylogenies using maximum
168 likelihood inference, comparing the likelihood support of the different models, and
169 estimating the parameters of the model (see Morlon 2014 for a review). The different
170 types of functions available in RPANDA reflect this general approach (Table 1): the
171 ‘likelihood’ functions compute the likelihood associated with different
172 diversification models, the ‘fit’ functions fit the corresponding models by maximum
173 likelihood, and the ‘plot’ functions plot estimates of how various variables (e.g.
174 speciation and extinction rates, species richness) have varied as a function of time or

175 various environmental factors (e.g. temperature). Simulating phylogenies under the
176 different models is often useful, for example, to test the power of the approach to
177 recover true parameter values or to measure type I & II error rates. We do not generally
178 provide functions to simulate phylogenies, as they are available in other packages such
179 as TreeSim (Stadler 2011b) and TESS (Höhna 2013). The one exception is the `sim_sgd`
180 function that simulates phylogenies under the model of Speciation by Genetic
181 Differentiation (Manceau *et al.* 2015).

182 Fitting a diversification model to a phylogeny consists in finding the parameters
183 that maximize the likelihood associated with the model. The ‘`fit`’ functions, therefore,
184 take as argument, at minimum, a phylogeny (`phylO`), initial parameter values (`par`), and
185 the maximization algorithm to be used (`meth`). The various likelihood expressions all
186 depend on the fraction of extant species that are sampled in the phylogeny; this fraction
187 (`f`), therefore, also needs to be specified. Finally, `tot_time` specifies the age of the
188 phylogeny. We often have access only to the crown age, in which case `tot_time` is
189 given by `max(node.age(phylO)$ages)`. If the stem age is known, specifying this older
190 age in `tot_time` can provide a different and complementary inference.

191

192 *Time-dependent diversification models*

193 RPANDA can be used to test whether (and how) diversification rates varied through
194 time (Morlon *et al.* 2011). To illustrate the approach we fit a birth-death model with
195 time-varying rates to the Phyllostomidae phylogeny. We first need to specify the
196 assumed functional form of the time-dependency. For example, if we want to fit a model
197 with an exponential variation of the speciation rate with time (`f.lamb`), and a constant
198 extinction rate (`f.mu`), we define these functions as follows:

```
199 > f.lamb<-function(t,y){y[1]*exp(y[2]*t)}
```

```
200 > f.mu<-function(t,y){y[1]}
```

201 For a linear dependency of the speciation rate, the function would be:

```
202 > f.lamb.lin<-function(t,y){y[1]+y[2]*t}
```

203 The variable `t` represents time, running from the present to the past, while the variable

204 `y` is a vector containing the different parameters involved in the definition of the

205 temporal dependency. The parameters in `y` are therefore the parameters that will be

206 estimated by maximum likelihood. We need to specify initial values for these

207 parameters, for example

```
208 > lamb_par_init<-c(0.05,0.01)
```

209 sets the initial parameter values defining the `f.lamb` speciation function, and

```
210 > mu_par_init<-c(0.005)
```

211 sets the initial parameter values defining the `f.mu` extinction function. The result of the

212 fitting procedure should not depend on the choice of the initial parameter values, which

213 can be checked by running the model with several sets of (realistic) initial values. For

214 example, the speciation rate at present (`lamb_par_init[1]`) typically takes value

215 ranging between 0.01 and 1 (event per lineage, per million years). The rate of variation

216 of the speciation rate (`lamb_par_init[2]`) can then be chosen such that the resulting

217 speciation rate (output of `f.lamb(t,lamb_par_init)`) remains within this realistic

218 range throughout the clade history (i.e. for `t` ranging from 0 to crown or stem age). And

219 finally, the initial extinction parameters set in `mu_par_init` can be chosen such that the

220 resulting extinction rate (output of `f.mu(t,mu_par_init)`) is smaller than the

221 speciation rate at the beginning of clade's history (i.e. for `t` set to the crown or stem age)

222 and remains positive throughout the clade history (i.e. for t ranging from 0 to crown or
223 stem age).

224 We can now fit the model. If we knew the stem age of Phyllostomidae, we could specify
225 this age in `tot_time`. As we do not have this information, we define:

```
226 > tot_time<-max(node.age(Phyllostomidae)$ages)
```

227 Finally, we fit the model by maximum likelihood using the following command:

```
228 > res<-fit_bd(Phyllostomidae,tot_time,f.lamb,f.mu,lamb_par_init,  
229             mu_par_init,f=150/165,expo.lamb=TRUE,cst.mu=TRUE)
```

230 The two options `expo.lamb=TRUE` and `cst.mu=TRUE` are set to `TRUE` to speed up the
231 computation by using analytical solutions; such solutions have been implemented for
232 exponential, linear, and constant functions. If the options are set to `TRUE` when the
233 time-dependency is not of the proper form (exponential, linear, or constant), the code
234 will not return the proper answer. When in doubt, it is better to prefer the default `FALSE`
235 option. There is an additional `dt` option (not used in the examples here) that can also
236 speed up the computation by using piecewise constant approximations in the
237 computation of the integrals. There is also an option that specifies whether we are
238 working with crown or stem ages, which has consequences for the conditioning that
239 should be used in the computation of the likelihood: the process should be conditioned
240 on survival when working with stem ages, and conditioned on a speciation event at the
241 crown and survival of the two descending lines when working with crown ages (Morlon
242 *et al.* 2011). The default (used here) is the `cond="crown"` option, but it should be set to
243 `cond="stem"` if `tot_time` is the stem age. The output `res` of the fit contains the
244 maximum log-likelihood value (-469.36), the corrected Akaike information criterion
245 (AICc) (944.89), and the maximum likelihood parameter estimates. For example,

```

246 > res$lamb_par[1]
247 returns the maximum parameter estimate of  $y[1]$ , which is the speciation rate at time
248  $t=0$ , i.e. the present (here 0.099).
249 > res$lamb_par[2]
250 returns the maximum parameter estimate of  $y[2]$ , which is the rate of change in
251 speciation rate, with time running from the present to the past (here 0.022). A positive
252 rate of change with time running from the present to the past – as estimated here –
253 suggests a negative rate of change (decline in speciation rate) during the clade’s history.
254 > plot_fit_bd(res,tot_time)
255 returns three plots, which represent speciation, extinction and net diversification (i.e.
256 speciation minus extinction) rates through time (Figure 4). If a model without extinction
257 is fitted:
258 > f.mu<-function(t,y){0}
259 > mu_par_init<-c()
260 > res_noext<-fit_bd(Phyllostomidae,tot_time,f.lamb,f.mu,
261 lamb_par_init,mu_par_init,f=150/165,expo.lamb=TRUE,fix.mu=TRUE)
262 > plot_fit_bd(res_noext,tot_time)
263 returns two plots, which represent speciation and net diversification rates through time
264 (in this case, these two rates are equal).
265 Once estimates of the temporal variation in speciation and extinction rates have been
266 obtained (as described above), estimates of how species richness varied through time
267 can be computed by resolving the appropriate differential equation (Morlon et al. 2011;
268 Morlon 2014).
269 > plot_dtt(res,tot_time,N0=165)

```

270 plots the result of this estimation procedure, as illustrated in Figure 5.

271

272 *Environmental-dependent diversification models*

273 RPANDA can also be used to test the potential effect that past environmental conditions

274 had on diversification (Condamine *et al.* 2013). Fitting the environmental-dependent

275 birth-death model is very similar to fitting the time-dependent birth-death model. In

276 addition to a phylogeny, this model requires knowledge (typically an estimate) of how a

277 given environmental variable varied through time. The example provided in RPANDA is

278 temperature variation through the Cenozoic, estimated using oxygen isotope ratios

279 (Zachos *et al.* 2008). We begin by loading the temperature data:

```
280 > data(InfTemp)
```

281 `InfTemp` is a two-column dataframe in which the first column reports time (measured

282 from the present to the past) and the second column reports the corresponding

283 estimated temperature at each time (Condamine *et al.* 2013). Any other (abiotic or

284 biotic) environmental variable in this format can be used in place of `InfTemp` in order to

285 test the potential effect that this variable had on diversification (e.g. sea-level

286 fluctuations, Condamine *et al.* 2015). Here, diversification rates can depend on time as

287 well as on the environmental variable; the `f.lamb` and `f.mu` functions therefore take

288 two arguments (time `t` and the environmental variable `x`) in addition to the parameters

289 `y` to be estimated. We fit a simple model with an exponential dependence of the

290 speciation rate on the environmental variable, no time dependence, and no extinction.

291 We thus define:

```
292 > f.lamb<-function(t,x,y){y[1]*exp(y[2]*x)}
```

293 and

```
294 > f.mu<-function(t,x,y){0}
```

295 as well as initial parameter estimates:

```
296 > lamb_par_init<-c(0.10,0.01)
```

```
297 > mu_par_init<- c()
```

298 Finally, we fit the model:

```
299 > res<-fit_env(Phyllostomidae,InfTemp,tot_time,f.lamb,f.mu,
```

```
300 lamb_par_init,mu_par_init,f=150/165,fix.mu=TRUE,dt=1e-3)
```

301 Note that we do not use the option `expo.lamb=TRUE`, as `f.lamb` is an exponential

302 function of the environmental variable, not an exponential function of time. Setting

303 `expo.lamb` to `TRUE` would yield spurious results. However, we can speed up the

304 computation by specifying `dt=1e-3`, which uses a piece-wise constant approximation in

305 the evaluation of integrals. The output `res` of the fit contains the maximum log-

306 likelihood value (-468.44), the AICc (940.97), and the maximum likelihood parameter

307 estimates. For example,

```
308 > res$lamb_par[1]
```

309 returns the maximum parameter estimate of `y[1]`, which is the speciation rate that

310 would correspond to a temperature of 0°C (here `0.077`).

```
311 > res$lamb_par[2]
```

312 returns the maximum parameter estimate of `y[2]`, which is the rate of change in

313 speciation rate with temperature (here `0.083`). A positive value thus suggests a

314 positive effect of the environmental variable (here temperature) on speciation rates.

315 Note that the environmental model fitted here receives a better support than the model

316 with an exponential dependency of speciation rate with time ($\Delta\text{AICc} = 944.89 - 940.97 =$

317 3.92).

318 > plot_fit_env(res, InfTemp, tot_time)

319 plots diversification rates as a function of the environmental variable and time (Figure
320 6).

321

322 *The model of Speciation by Genetic Differentiation*

323 RPANDA also contains functions to fit a modified version of Hubbell's neutral model of
324 biodiversity (Hubbell 2001): the model of Speciation by Genetic Differentiation, which is
325 described in detail in Manceau *et al.* 2015. We can fit the model following a similar
326 procedure. The model is individual-based, and involves parameters describing
327 the birth and death of individuals, as well as a per-individual mutation rate. We define
328 initial parameter values for the birth, growth (birth minus death), and mutation rates
329 (given in events per Myr if the branch lengths of the phylogeny are measured in Myrs),
330 for example:

331 > par_init<-c(1e7, 1e7-0.5, 0.8)

332 We can then fit the model (this takes more time than the previous examples, as
333 computing likelihoods requires integrating a set of coupled differential equations along
334 the tree, Manceau *et al.* 2015) :

335 > fit_sgd(Phyllostomidae, tot_time, par_init, f=150/165)

336 returns the maximum log-likelihood (-466), the AICc (938), and the estimated birth
337 (1e7), growth (0.157), and mutation (0.198) rates. As explained in Manceau *et al.*
338 (2015), the likelihood surface is quite flat with respect to the birth rate, such that this
339 parameter cannot be estimated with confidence. Birth rate estimates are thus sensitive
340 to the choice of the initial parameter values and should not be trusted.

341

342 **Resources**

343 RPANDA is an open source package available for download on the CRAN repository at
344 <https://cran.r-project.org/web/packages/RPANDA/index.html>. It includes illustrative
345 data and a detailed manual. The package is constantly evolving. Bayesian
346 implementations and functions that allow fitting new models of phenotypic evolution,
347 such as the matching competition model (Drury *et al.* 2015), are already under
348 development. Contributions are welcome; automatic tests have been implemented to
349 facilitate a collaborative development and to insure the replicability of results. The most
350 recent version is available on github at <https://github.com/hmorlon/PANDA>.

351

352 **Acknowledgement**

353 We thank members of HM's group, as well as reviewers D.S. Caetano and D. Silvestro for
354 comments on a previous version of the paper. The development of RPANDA is supported
355 by grant CHEX-ECOEVOBIO from the Agence Nationale de la Recherche and grant
356 616419-PANDA from the European Research Council awarded to HM.

357

358 **Data Accessibility**

359 All data used in this manuscript is available through the R package RPANDA, available
360 on CRAN <https://cran.r-project.org/web/packages/RPANDA/index.html>.

361

362 **References**

363 Bininda-Emonds, O.R.P., Cardillo, M., Jones, K.E., MacPhee, R.D.E., Beck, R.M.D., Grenyer,
364 R., Price, S.A., Vos, R.A., Gittleman, J.L. & Purvis, A. (2007). The delayed rise of present-
365 day mammals. *Nature*, **446**, 507–512.

366 Condamine, F.L., Rolland, J. & Morlon, H. (2013). Macroevolutionary perspectives to

367 environmental change. *Ecology Letters*, **16**, 72–85.

368 Condamine, F.L., Toussaint, E.F.A., Clamens, A.-L., Genson, G., Sperling, F.A.H. & Kergoat,
369 G.J. (2015). Deciphering the evolution of birdwing butterflies 150 years after Alfred
370 Russel Wallace. *Scientific Reports*, **5**.

371 Csardi, G. & Nepusz, T. (2006). The igraph software package for complex network
372 research. *InterJournal, Complex Systems*, **1695**, 1–9.

373 Drury, J., Clavel, J. & Morlon, H. (2015). Estimating the effect of competition on trait
374 evolution using maximum likelihood. *bioRxiv*.

375 Etienne, R.S., Haegeman, B., Stadler, T., Aze, T., Pearson, P.N., Purvis, A. & Phillimore, A.B.
376 (2012). Diversity-dependence brings molecular phylogenies closer to agreement with
377 the fossil record. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 1300–1309.

378 FitzJohn, R.G. (2012). Diversitree: comparative phylogenetic analyses of diversification
379 in R. *Methods in Ecology and Evolution*, **3**, 1084–1092.

380 Fritz, S.A., Bininda-Emonds, O.R.P. & Purvis, A. (2009). Geographical variation in
381 predictors of mammalian extinction risk: big is bad, but only in the tropics. *Ecology*
382 *Letters*, **12**, 538–549.

383 Höhna, S. (2013). Fast simulation of reconstructed phylogenies under global time-
384 dependent birth–death processes. *Bioinformatics*, **29**, 1367–1374.

385 Hubbell, S.P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*.
386 Princeton University Press, Princeton.

387 Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D.,
388 Blomberg, S.P. & Webb, C.O. (2010). Picante: R tools for integrating phylogenies and
389 ecology. *Bioinformatics*, **26**, 1463–1464.

390 Lartillot, N. & Poujol, R. (2011) A phylogenetic model for investigating correlated
391 evolution of substitution rates and continuous phenotypic characters. *Molecular Biology*

392 *and Evolution*, **28**, 729-744.

393 Lewitus, E. & Morlon, H. (2015). Characterizing and comparing phylogenies from their
394 Laplacian spectrum. *bioRxiv*.

395 Manceau, M., Lambert, A. & Morlon, H. (2015). Phylogenies support out-of-equilibrium
396 models of biodiversity. *Ecology letters*, **18**, 347–356.

397 Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecology Letters*,
398 **17**, 508–525.

399 Morlon, H., Parsons, T.L. & Plotkin, J.B. (2011). Reconciling molecular phylogenies with
400 the fossil record. *Proceedings of the National Academy of Sciences of the United States of*
401 *America*, **108**, 16327–16332.

402 Morlon, H., Potts, M.D. & Plotkin, J.B. (2010). Inferring the dynamics of diversification: a
403 coalescent approach. *PLoS Biology*, **8**, e1000493.

404 Nee, S., Mooers, A.O. & Harvey, P.H. (1992). Tempo and mode of evolution revealed from
405 molecular phylogenies. *Proceedings of the National Academy of Sciences*, **89**, 8322–8326.

406 Paradis, E., Claude, J. & Strimmer, K. (2004). APE: Analyses of Phylogenetics and
407 Evolution in R language. *Bioinformatics*, **20**, 289–290.

408 Pennell, M.W., J.M. Eastman, G.J. Slater, J.W. Brown, J.C. Uyeda, R.G. FitzJohn, M.E. Alfaro &
409 L.J. Harmon. 2014. geiger v2.0: an expanded suite of methods for fitting
410 macroevolutionary models to phylogenetic trees. *Bioinformatics*, **15**, 2216-2218.

411 Pigot, A.L. & Etienne, R.S. (2015). A new dynamic null model for phylogenetic
412 community structure. *Ecology Letters*, **18**, 153–163.

413 Rabosky, D.L. (2014). Automatic detection of key innovations, rate shifts, and diversity-
414 dependence on phylogenetic trees. *PloS One*, **9**, e89543.

415 Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and
416 other things). *Methods in Ecology and Evolution*, **3**, 217–223.

417 Rolland, J., Condamine, F.L., Jiguet, F. & Morlon, H. (2014). Faster Speciation and Reduced
418 Extinction in the Tropics Contribute to the Mammalian Latitudinal Diversity Gradient.
419 *PLoS Biology*, **12**, e1001775.

420 Silvestro, D., Schnitzler, J. & Zizka, G. A Bayesian framework to estimate diversification
421 rates and their variation through time and space. *BMC Evolutionary Biology*, **11**, 311.

422 Soetaert, K.E.R., Petzoldt, T. & Setzer, R.W. (2010). Solving differential equations in R:
423 package deSolve. *Journal of Statistical Software*, **33**.

424 Stadler, T. (2011a). Mammalian phylogeny reveals recent diversification rate shifts.
425 *Proceedings of the National Academy of Sciences of the United States of America*, **108**,
426 6187–6192.

427 Stadler, T. (2011b). Simulating Trees with a Fixed Number of Extant Species. *Systematic
428 Biology*, **60**, 676–684.

429 Zachos, J.C., Dickens, G.R. & Zeebe, R.E. (2008). An early Cenozoic perspective on
430 greenhouse warming and carbon-cycle dynamics. *Nature*, **451**, 279–283.

431

432

433

434

435 **Table 1. Major functions available in RPANDA**

Function	Description
Characterising and comparing phylogenies using spectral densities	
BICcompare	computes BIC values assessing the support of modalities in a phylogeny
JSDtree	computes the Jensen-Shannon distance between phylogenies
JSDtree_cluster	clusters phylogenies using hierarchical and k-medoids clustering
spectR	computes the eigenvalues of a phylogeny, and returns the principal eigenvalue, the skewness, and kurtosis of the spectral density profile, and the eigengap
plot_BICcompare	plots modalities on a phylogenetic tree
plot_spectR	plots the spectral density and eigenvalues of a phylogeny ranked in descending order
Fitting models of diversification to phylogenies	
fit_bd	fits a birth-death model to a phylogeny
fit_coal_cst	fits an equilibrium model with constant diversity through time to a phylogeny, using the coalescent model
fit_coal_var	fits a model with expanding diversity through time to a phylogeny, using the coalescent model
fit_env	fits a birth-death model with environmental dependency to a phylogeny

fit_sgd	fits the model of Speciation by Genetic Differentiation to a phylogeny
likelihood_bd	computes the likelihood corresponding to the birth-death model
likelihood_coal_cst	computes the likelihood corresponding to the coalescent model with constant diversity through time
likelihood_coal_var	computes the likelihood corresponding to the coalescent model with expanding diversity through time
likelihood_sgd	computes the likelihood corresponding to the model of Speciation by Genetic Differentiation
plot_fit_bd	plots diversification rates through time curves
plot_fit_env	plots diversification rates as a function of one or several environmental variables and time
plot_dtt	computes and plots diversity through time curves
sim_sgd	simulates a phylogeny arising from the model of Speciation by Genetic Differentiation

436

437

438 **Figure 1** Spectral density plot of the Phyllostomidae (left panel) and corresponding
439 eigenvalues ranked in descending order (right panel). There is a clear gap between the
440 third and fourth eigenvalue (indicated by an arrow), suggesting three modes of division
441 in the phylogeny.

442

443 **Figure 2** Phyllostomidae phylogeny with branches coloured according to the 3
444 branching patterns (or modalities) identified by the eigengap, as given by k-medoid
445 clustering.

446

447 **Figure 3** Heatmap and hierarchical cluster showing the pairwise similarities between
448 the 25 Phyllostomidae genera with more than one species.

449

450 **Figure 4** Plots showing the estimated a) speciation, b) extinction, and c) net
451 diversification rates through time for the Phyllostomidae phylogeny, output of the
452 `plot_fit_bd` function.

453

454 **Figure 5** Plot showing the estimated accumulation of species richness through time for
455 the Phyllostomidae phylogeny, output of the `plot_dtt` function.

456

457 **Figure 6** Plots showing the estimated speciation (a,b) and net diversification (c,d) rates
458 as a function of time (a,c) and temperature (b,d) for the Phyllostomidae phylogeny,
459 output of the `plot_fit_env` function.

460