



**HAL**  
open science

## Quantifying Variation in Speciation and Extinction Rates With Clade Data

Emmanuel Paradis, Pablo A Tedesco, Bernard Hugueny

► **To cite this version:**

Emmanuel Paradis, Pablo A Tedesco, Bernard Hugueny. Quantifying Variation in Speciation and Extinction Rates With Clade Data. *Evolution - International Journal of Organic Evolution*, 2013, 67 (12), pp.3617 - 3627. 10.1111/evo.12256 . hal-01821922

**HAL Id: hal-01821922**

**<https://hal.umontpellier.fr/hal-01821922>**

Submitted on 23 Jun 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Running head: VARIATION IN SPECIATION AND EXTINCTION

## QUANTIFYING VARIATION IN SPECIATION AND EXTINCTION RATES WITH CLADE DATA

**Emmanuel Paradis<sup>1,3</sup>, Pablo A. Tedesco<sup>2</sup> and Bernard Hugueny<sup>2</sup>**

<sup>1</sup>*Institut de Recherche pour le Développement, ISEM UMR 226/5554 – UM2/CNRS/IRD, Jl. Taman Kemang 32B, Jakarta 12730, Indonesia*

<sup>2</sup>*UMR Biologie des ORganismes et des Écosystèmes Aquatiques (UMR BOREA, IRD 207-CNRS 7208-UPMC-MNHN), Département Milieux et Peuplements Aquatiques, Muséum National d'Histoire Naturelle, 43 rue Cuvier, 75231 Paris cedex, France*

<sup>3</sup>*Email: Emmanuel.Paradis@ird.fr*

1 High-level phylogenies are very common in evolutionary analyses, though they are of-  
2 ten treated as incomplete data. Here we provide statistical tools to analyze what we  
3 name 'clade data', that are the ages of clades together with their numbers of species.  
4 We develop a general approach for the statistical modeling of variation in speciation  
5 and extinction rates, including temporal variation, unknown variation, and linear and  
6 nonlinear modeling. We show how this approach can be generalized to a wide range of  
7 situations, including testing the effects of life-history traits and environmental variables  
8 on diversification rates. We report the results of an extensive simulation study to assess  
9 the performance of some statistical tests presented here as well as of the estimators of  
10 speciation and extinction rates. These latter results suggest the possibility to estimate  
11 correctly extinction rate in the absence of fossils. An example with data on fish is  
12 presented.

13 **KEY WORDS:** birth–death models, extinction, maximum likelihood, speciation, stem  
14 ages.

15 The study of the tempo and mode of evolution has experienced a new wave  
16 of interest from evolutionists using new mathematical and statistical tools  
17 to analyze molecular phylogenies (Sanderson and Donoghue 1996; Ricklefs  
18 2007). Following some initial breakthrough (e.g., Nee et al. 1992, 1994), sig-  
19 nificant progress has been achieved in biologically relevant statistical mod-  
20 eling of diversification, such as quantifying temporal variation in diversifi-  
21 cation (Paradis 2011; Hallinan 2012) or assessing the effects of biological  
22 traits on speciation and extinction rates (Maddison et al. 2007; FitzJohn  
23 et al. 2009; FitzJohn 2010). Recent advances have also been accomplished  
24 in integrating molecular and fossil data (e.g., Morlon et al. 2011; Didier et al.  
25 2012).

26 Most of these recent statistical developments have focused on analyz-  
27 ing complete phylogenies. Incomplete phylogenies are often treated as a  
28 separate case in order to take missing data into account (Pybus et al. 2002;  
29 FitzJohn et al. 2009; Stadler 2011). The most common form of such data is  
30 a phylogeny resolved at a high level accompanied by the number of species  
31 associated to each tip of the tree. On the other hand, the ages of clades  
32 together with the numbers of species (named here ‘clade data’) have been  
33 a neglected source of data in the analysis of diversification. Magallón and  
34 Sanderson (2001) provided some methods for the analysis of such data and  
35 applied them to angiosperms. They particularly developed various estima-  
36 tors of the (net) rate of diversification of a clade giving its age and number  
37 of species.

38 The relative lack of interest towards clade data may come from the fact  
39 that, for a given clade, its complete phylogeny contains more information  
40 than the pair of values ‘age + number of species’. However, for a collection  
41 of clades, such data are a valuable source of information for several reasons.

42 First, clades defined by higher-level taxa (e.g., families, orders) are clearly  
43 identified for almost all groups of living beings and their numbers of species  
44 are in many cases already known. Second, phylogenetic relationships among  
45 higher-level taxa have been much more studied than within them, so it is  
46 more straightforward to date the age of a clade rather than the divergences  
47 among its species. Third, the fossil record is generally more informative on  
48 the origin of higher-level taxa compared to species or other low-level taxa.  
49 Fourth, it is easier to examine the impact of the species concept on the  
50 definition of clade data rather than on a phylogeny since, in the former the  
51 species concept will mostly affect the number of species while in the latter it  
52 will be often hard to infer different phylogenies under those distinct species  
53 definitions. Clade data have also some disadvantages: the inherent lack  
54 of temporal resolution within each clade makes it impossible to study the  
55 variation in diversification within them.

56 In the present paper, we extend the approach presented by Magallón  
57 and Sanderson (2001) and present statistical tools for the inference of di-  
58 versification patterns and processes with clade data. Our approach assumes  
59 that each clade, instead of having its own speciation and extinction rates,  
60 comes from a ‘statistical population of clades’ so that maximum likelihood  
61 inference is straightforward. With this rationale, we show how to make  
62 inference on variation in diversification parameters among clades using dif-  
63 ferent modeling tools, including testing the effects of life-history traits and  
64 environmental variables and the case where variation is a priori unknown.  
65 We also present the results of a simulation study in order to assess the sta-  
66 tistical performance of several tests and estimators presented in this paper,  
67 and finally we apply our approach on a data set of fish.

68 *Statistical Modeling Approach*

69 Throughout this paper we assume that diversification proceeds with speci-  
70 ation ( $\lambda$ ) and extinction ( $\mu$ ) rates which are the probabilities that a species  
71 splits into two daughter-species or goes extinct during a very short time.  
72 We denote as  $X_t$  the number of species in a clade of age  $t$  where this may  
73 be either the stem age of the clade (divergence time of the clade from its  
74 sister-clade) or its crown age (time to the most common recent ancestor  
75 of the species belonging to the clade). Specifically, using equation 8 from  
76 Kendall (1948), we can write the probability that  $X_t$  takes a specific integer  
77 value  $x$ :

$$\Pr(X_t = x | \theta, X_0 = 1) = \eta_t (1 - \eta_t)^{x-1} \quad x \geq 1, \quad (1)$$

78 where  $\theta$  is a vector of parameters specifying how speciation and extinction  
79 rates vary through time and  $\eta_t$  is a function of these parameters. The  
80 conditioning on  $X_0 = 1$  emphasizes that in this paper we consider stem  
81 groups. For the case of crown groups ( $X_0 = 2$ ), the probabilities must  
82 be summed on all possible combinations. In most applications, stem groups  
83 are considered because the origin of a group is inferred from its relationships  
84 with its sister-group. On the other hand, deriving the crown age of a group  
85 requires to estimate the age of the most recent common ancestor of its species  
86 which is usually more complicated because it requires to sample all species  
87 in the clade. On the other, inferring stem ages requires one species from the  
88 clade and one from its sister-clade.

89 Various forms exist for these probabilities depending on the parameteri-  
90 zation of  $\theta$  and whether we wish to condition them on survival of the lineage  
91 until present or not. For instance, if extinction rate is zero and speciation  
92 rate is constant, then  $\eta_t = e^{-\lambda t}$ . This is the Yule (1924) model. Models with

93 a non-null extinction rate are called birth–death models (Kendall 1948).

94 The point of conditioning on no extinction is important when analyzing  
95 data on actual groups because total extinction of these groups did not occur.  
96 Thus the probabilities must be modified accordingly, otherwise this would  
97 result in underestimated extinction rates (Rabosky et al. 2007).

98 Let us consider for the moment the simple Yule model. The expected  
99 number of species at time  $t$  is given by  $E(X_t) = e^{\lambda t}$ . From this expectation,  
100 a simple estimator of  $\lambda$  based on the method of moments is  $\hat{\lambda} = \ln(x)/t$   
101 (Magallón and Sanderson 2001). When considering a single clade, and in  
102 the absence of more detailed information, it does not seem possible to go  
103 further in the inference. When considering more than one group (e.g., the  
104 families within an order or a class), researchers usually estimate  $\lambda$  separately  
105 for each group, then proceed with standard statistics (e.g., McPeck 2008).  
106 This approach assumes that each clade is characterized by its own speciation  
107 rate. On the other extreme, one may assume that speciation rate is the same  
108 in all groups so that the observed data are independent outcomes of the  
109 same diversification process. Thus, it is possible to use maximum likelihood  
110 inference using equation 1. The likelihood function is:

$$\prod_i \Pr(x_i|\lambda), \quad (2)$$

111 where  $\Pr(x|\lambda)$  is a simplified notation of equation 1. We may expect less  
112 bias in the estimates from this approach, but also the possibility to test  
113 hypotheses based on fitting alternative models.

114 The assumption of equal speciation rates among clades is, certainly in  
115 most cases, unrealistic (Purvis et al. 1995; Paradis 2005; Alfaro et al. 2009).  
116 However, since we have several observations we may model the variation  
117 in this parameter with a statistical modeling approach. We explore several

118 such approaches below. Firstly, we consider approaches based on determin-  
119 istic variation between two or more groups of clades. Secondly, we consider  
120 how temporal variation in speciation and extinction rates can be modeled  
121 and assessed. Thirdly, we develop an approach handling unknown variation  
122 based on mixture modeling, including the combination of mixtures with a  
123 linear modeling of the speciation rate. Finally, we attack the problem of  
124 estimating extinction rates.

### 125 **Variation Among Clades**

126 A simple way to model variation in diversification among clades is to assume  
127 that there are two categories: some clades diversify with speciation rate  $\lambda_1$   
128 and the others with rate  $\lambda_2$ . The data are made of  $n_1$  and  $n_2$  clades in each  
129 category, respectively. The likelihood function is:

$$\prod_{i_1=1}^{n_1} \Pr(x_{i_1}|\lambda_1) \prod_{i_2=1}^{n_2} \Pr(x_{i_2}|\lambda_2).$$

130 Note that each clade is assigned to a category a priori, although there is  
131 no assumption on whether  $\lambda_1$  is greater, or smaller, than  $\lambda_2$ . The null  
132 hypothesis  $\lambda_1 = \lambda_2$  can be tested by fitting this model and the null model  
133 whose likelihood is given by equation 2: the likelihood-ratio test (LRT)  
134 comparing these two models follows a  $\chi^2$  distribution with  $df = 1$ . An  
135 alternative is to use the Akaike information criterion (Akaike 1973).

136 The present approach is easily generalized to more than two categories:  
137 let us denote the number of categories as  $K$ , then the likelihood function  
138 would become the product of  $K$  products:

$$\prod_{j=1}^K \prod_{i_j=1}^{n_j} \Pr(x_{i_j}|\lambda_j),$$



139 where  $n_j$  is the number of clades in the  $j$ th category. The LRT comparing  
140 this model with the null model of homogeneous diversification follows a  $\chi^2$   
141 with  $df = K - 1$ .

142 These models assume, mostly for simplicity, that there is no extinction  
143 ( $\mu = 0$ ); however, variation in extinction rate can be incorporated in a  
144 straightforward way. For instance a model with two categories diversify-  
145 ing with the same  $\lambda$  but with different extinction rates has the following  
146 likelihood function:

$$\prod_{i=1}^{n_1} \Pr(x_{i_1} | \lambda, \mu_1) \prod_{i=1}^{n_2} \Pr(x_{i_2} | \lambda, \mu_2),$$

147 which could be compared with the null model with  $\mu > 0$  whose likelihood  
148 is:

$$\prod_{i=1}^N \Pr(x_i | \lambda, \mu),$$

149 with  $N = n_1 + n_2$ . This test is related, but not identical, to the tests  
150 of equal diversification using sister-clades where the ages of clades are not  
151 needed (Paradis 2012b).

152 The supplementary materials provide annotated R code explaining how  
153 to build and fit any model following the present approach.

## 154 **Linear Modeling**

155 Following the previous section, two extreme models can be defined: the sim-  
156 plest one where all clades diversify at the same rate, and the most complex  
157 one where each clade has its own parameter(s). This second model will be  
158 overparameterized for a likelihood approach. Nevertheless, it is possible to  
159 model variation in diversification parameters with linear models. For in-

160 stance, we may know a priori some variables that are likely to affect the  
161 value of speciation rate (e.g., body size), and a model that relates such ‘co-  
162 variates’ to speciation rate may be an appropriate candidate to model the  
163 variation in diversification among clades. We use here a standard strategy  
164 to model variation in a rate with respect to a covariate  $z$ :

$$g(\lambda_i) = \beta z_i + \alpha,$$

165 where  $\lambda_i$  is the speciation rate in clade  $i$ ,  $g$  is a function used to transform the  
166 rate in order to linearize the relationship, and  $\beta$  and  $\alpha$  are two parameters.  
167 Here  $\beta$  controls the effect of  $z$  on  $\lambda$ : if  $\beta > 0$  then species with large values  
168 of  $z$  will speciate faster than those with small values of  $z$  (and inversely if  
169  $\beta < 0$ ). It is possible to consider more than one predictor in which case  
170 the number of parameters is equal to the number of predictors plus one.  
171 Nonlinear models can also be considered. Each clade has its own speciation  
172 rate given by (with  $g^{-1}$  being the inverse transformation of  $g$ ):

$$\lambda_i = g^{-1}(\beta z_i + \alpha), \tag{3}$$

173 which is used to calculate the likelihood defined by equation 2: the likelihood  
174 function is then maximized to estimate  $\beta$  and  $\alpha$  (see code in the Supplemen-  
175 tary Material). A common choice for  $g$  is the logit function,  $\ln(\lambda_i/(1 - \lambda_i))$ ,  
176 so  $g^{-1}$  gives:

$$\lambda_i = \frac{1}{1 + e^{-(z_i\beta + \alpha)}},$$

177 The null model is defined by fixing  $\beta = 0$  in which case  $\lambda = 1/(1 + e^{-\alpha})$  for  
178 all clades. The logit function is well suited for parameters varying between

179 0 and 1 which is the case for speciation rates considered on geological time  
180 scales (million of years). However, speciation rates may be larger than one  
181 on shorter scales. Other transformations can be used such as the one used  
182 below.

183 It must be noted that the variation among clades as modeled in the  
184 previous section is a special case of linear models where the membership of  
185 a clade to a category is coded with a discrete variable and this variable is  
186 entered as a predictor into the linear model after coding it into binary 0/1  
187 variable(s) (see appendix in Paradis 2005, for details). Therefore, continuous  
188 and categorical predictors can be combined in the linear model.

### 189 **Temporal Variation**

190 Kendall (1948) studied the birth–death model in a very general way, in-  
191 cluding the cases where  $\lambda$  and  $\mu$  vary through time. Thus it is possible to  
192 derive the probability density of the distribution of the  $x_i$ 's when diversifi-  
193 cation changed through time. The likelihood can be defined and fit in the  
194 same way as above. Such a temporal model can be compared with the null  
195 model of constant diversification with a  $\chi^2$  test whose  $df$  will be equal to  
196 the number of additional parameters in the first model. As before, tempo-  
197 ral variation may reflect speciation and/or extinction rate(s). The simplest  
198 temporal model has two rates before and after a given time point in the past,  
199 so it has one additional parameter than the null model. Note that if the  
200 time point is unknown, it could be estimated from the data so there would  
201 be two additional parameters. However, a wide variety of temporal models  
202 can be defined in `ape` (Paradis et al. 2004) using the function `dbdTime` where  
203 the temporal variation is defined by the user with a standard R function.

204 **Unknown Variation**

205 The above models assume that diversification parameters vary in relation  
206 to some known variables, either categorical or continuous. On the other  
207 hand, it is possible that these variables are not observable. Such unknown  
208 variation can be modeled with two approaches depending on whether we  
209 assume that the diversification parameters vary in a discrete or continuous  
210 manner.

211 A mixture of distributions is based on the assumption that observations  
212 come from two or more categories each characterized by its own distribution,  
213 but the assignment of an observation to a particular category is unknown  
214 (see Flury et al. 1992, for a biological example). As a simple example,  
215 consider a mixture of two Yule processes, then the likelihood function will  
216 be:

$$\prod_{i=1}^N f \Pr(x_i|\lambda_1) + (1 - f) \Pr(x_i|\lambda_2), \quad (4)$$

217 where  $f$  is the proportion of clades in the first category. This model has  
218 three parameters ( $\lambda_1$ ,  $\lambda_2$  and  $f$ ) and can be compared with the null model  
219 of homogeneous speciation with a LRT with  $df = 2$ . The idea is easily  
220 generalized to more than two mixtures: a mixture with  $K$  Yule models  
221 would have  $2K - 1$  parameters. As above, the mixture may involve speciation  
222 and/or extinction rate(s). By contrast to the situation above where clades  
223 were assigned to categories a priori, there is here no assignment a priori.  
224 On the other hand, assignment a posteriori is possible by calculating the  
225 relative contributions to the likelihood function.

226 The idea may even be further generalized to include mixtures of linear  
227 models. Suppose we know that one variable, say body size, has a significant  
228 effect on speciation rate but there is some other, unknown, variation in this

229 parameter that we want to model with a mixture. Then it is possible to  
 230 calculate the  $\lambda_i$ 's with equation 3 and use them to compute the likelihood  
 231 with eq. 4. Each category would have its own parameters  $\beta$  and  $\alpha$ , so a  
 232 model with  $K$  categories has  $3K - 1$  parameters.

233 The second approach assumes that, in the case of a Yule model,  $\lambda$  varies  
 234 continuously across clades following a specified distribution whose parame-  
 235 ters are estimated from the data. A transformation of  $\lambda$  is useful so that it  
 236 follows a normal distribution:  $g(\lambda) \sim \mathcal{N}(\mu_\lambda, \sigma_\lambda^2)$ . A useful transformation  
 237 here is the complementary log-log transformation:  $g(\lambda) = \ln(-\ln(\lambda))$ . As  
 238 above we do not know the value of  $\lambda$  for a given clade, but this time instead  
 239 of a discrete sum we have to do a continuous integration. The likelihood  
 240 function is thus:

$$\prod_{i=1}^N \int_{-\infty}^{\infty} f_{\mathcal{N}}(u|\mu_\lambda, \sigma_\lambda^2) \Pr(x_i|g^{-1}(u))du,$$

241 where  $f_{\mathcal{N}}$  is the density function of the normal distribution. A graphical  
 242 representation of the variation in  $\lambda$  is obtained with the inverse transforma-  
 243 tion  $g^{-1}(u) = \exp(-e^u)$  with the density of  $u$  computed with the normal  
 244 distribution and the estimates  $\hat{\mu}_\lambda$  and  $\hat{\sigma}_\lambda^2$ .

## 245 **Estimating Extinction Rates**

246 The estimation of extinction rates in the absence of fossil data has appeared  
 247 to be a complicated issue (Paradis 2004, 2011; McPeck 2008; Aldous et al.  
 248 2011; Morlon et al. 2011; Didier et al. 2012; Hallinan 2012). To try to tackle  
 249 this problem, we implemented a procedure which fits a birth–death model  
 250 estimating  $\lambda$  and  $\mu$  simultaneously. These estimates are denoted as  $\hat{\lambda}_{\text{BD}}$  and  
 251  $\hat{\mu}_{\text{BD}}$ .

252 *Simulation Study*

253 The present statistical modeling approach offers many possibilities and it  
254 would take a large number of simulations to assess the statistical proper-  
255 ties of all of them. Instead, we focus on a few key questions. What is the  
256 statistical power to detect a difference in diversification between two groups  
257 of clades? How powerful is the test to detect temporal variation in diver-  
258 sification? What is the statistical power to detect unknown variation in  
259 diversification between two groups of clades using mixtures? Finally, what  
260 is the precision of the  $\lambda$  and  $\mu$  estimators?

261 To address these four questions, we ran four sets of simulations. First, we  
262 considered a simple two-category scenario with  $n_1$  and  $n_2$  clades simulated  
263 with rates  $\lambda_1$  and  $\mu_1$  and  $\lambda_2$  and  $\mu_2$ , respectively. The times of evolution  
264 were drawn from a uniform distribution:  $t_i \sim \mathcal{U}(10, 20)$ . A phylogeny was  
265 simulated under a birth–death process during a time  $t_i$  using `ape` starting  
266 from a single species. The number of species surviving at time  $t_i$ ,  $x_i$ , was  
267 extracted and the pairs  $(x_i, t_i)$  were analyzed as described above using a Yule  
268 model. The LRT testing the null hypothesis of homogeneous diversification  
269 was computed, and the rejection rate was assessed under different sets of  
270 parameter values:  $n_1 = n_2 = \{1, 3, 5, 10, 20\}$ ,  $\lambda_1 = \{0.1, 0.15, 0.2\}$ ,  $\lambda_2 = 0.1$ ,  
271  $\mu_1 = \{0, 0.05\}$ ,  $\mu_2 = \{0, 0.05\}$ .

272 Second, we performed simulations under three scenarios with different  
273 values of diversification rates before and after 30 time units. We first gen-  
274 erated 100 values of  $t$  from a uniform distribution between 10 and 50. We  
275 then simulated clades with constant, increasing, or decreasing diversification  
276 rate. The number of species was extracted as before, and two models were  
277 fitted: the null Yule model of constant diversification, and an alternative  
278 model assuming different speciation rates before and after 30 time units (as

279 above  $\mu = 0$  was assumed). The rejection rates of the LRTs comparing both  
280 models were computed.

281 Third, a scenario similar to the first one was considered: the difference  
282 is that the simulated clades were not identified to a particular category so  
283 the data were analyzed with a mixture of Yule models. We used  $K = 2$ ,  
284  $n_1 = n_2 = \{10, 20, 50\}$ , and  $t_i \sim \mathcal{U}(10, 20)$ . Four combinations of speciation  
285 and extinction rates were used: (i) the null hypothesis is true and there is  
286 no extinction:  $\lambda_1 = \lambda_2 = 0.1$ ,  $\mu_1 = \mu_2 = 0$ ; (ii) the null hypothesis is false  
287 and there is no extinction:  $\lambda_1 = 0.1$ ,  $\lambda_2 = 0.2$ ,  $\mu_1 = \mu_2 = 0$ ; (iii) the null  
288 hypothesis is false but only  $\mu$  varies:  $\lambda_1 = \lambda_2 = 0.2$ ,  $\mu_1 = 0$ ,  $\mu_2 = 0.1$ ; and  
289 (iv) same than before with stronger variation in  $\mu$ :  $\lambda_1 = \lambda_2 = 0.2$ ,  $\mu_1 = 0$ ,  
290  $\mu_2 = 0.15$ .

291 Finally, we performed an assessment of the precision of the estimators of  
292 speciation and extinction rates using five combinations of  $\lambda$  and  $\mu$ :  $(0.1, 0)$ ,  
293  $(0.1, 0.03)$ ,  $(0.1, 0.06)$ ,  $(0.2, 0.1)$ , and  $(0.2, 0.15)$ . Here  $t_i \sim \mathcal{U}(10, 30)$  and  
294  $n = 100$ .

295 The simulations were replicated 1000 times. Annotated R (R Devel-  
296 opment Core Team 2012) code is available in the Supplementary Material  
297 with guidelines on how to run these simulations so that the readers can  
298 adapt them to their own problems. Besides, we did not attempt to compare  
299 our method with previous ones because some scenarios considered here can-  
300 not be analyzed by the latter (e.g., the third scenario does not seem to be  
301 tractable with Magallón and Sanderson's method).

### 302 *Application to Fish Data*

303 We used the data from Vega and Wiens (2012) who compiled the number  
304 of species, stem age, and percentage of marine fish species for 22 orders and

305 super-orders and for 97 families. They also provided a phylogeny for the  
306 22 higher taxa which allowed to compare our estimates with those obtained  
307 from the combined analysis of phylogeny and species richness data (Paradis  
308 2003). All data were unmodified from the original publication and are avail-  
309 able at <http://dx.doi.org/10.1098/rspb.2012.0075>. With this data set, we  
310 explored the variation in diversification using different mixtures of Yule and  
311 birth–death models. We also tried to assess whether this variation is due to  
312 differences in the speciation or in the extinction rates.

### 313 *Results*

#### 314 **Simulation Study**

315 The first set of simulations showed that, overall, the LRT testing for dif-  
316 ferent diversification rates between two categories of clades had satisfactory  
317 statistical properties (Table 1). The type I error rate (rejection rate when  
318 the null hypothesis is true, i.e.,  $\lambda_1 = \lambda_2$  and  $\mu_1 = \mu_2$ ) was, as expected,  
319 close to 5% (first and seventh lines in Table 1). However, when  $\lambda - \mu$  was  
320 the same in both categories, the rejection rate was greater than 5% (eighth  
321 line in Table 1) showing that the present test does not test for equal di-  
322 versification rate. In the cases where the null hypothesis was not true, the  
323 rejection rate varied as expected: it was greater for larger sample sizes ( $n_1$ )  
324 and for larger contrast in the speciation or extinction rate. Interestingly, if  
325 one category of clades had smaller  $\mu$  while  $\lambda$  was the same, then the test was  
326 able to detect this difference; however, the statistical power was less than  
327 when the same contrast in diversification was due to different  $\lambda$  (compare  
328 the second and third lines in Table 1).

329 In the second set of simulations, the test for temporal variation rejected  
330 the null hypothesis in more than 90% when  $\mu = 0$  and  $\lambda$  varied, whether this



331 was an increase or a decrease (third to sixth lines in Table 2). On the other  
332 hand, the results were contrasted when  $\mu > 0$ . When there was no temporal  
333 variation in the parameters, the type I error rates were inflated in relation  
334 to the value of  $\mu$  (seventh and eighth lines in Table 2). When  $\mu$  varied  
335 through time, the test behaved very differently depending on the direction  
336 of this variation: it did not reject the null hypothesis in most cases when  
337  $\mu$  increased (ninth line in Table 2) while it rejected it in 68% of the cases  
338 when  $\mu$  decreased (tenth line in Table 2). To further investigate this point,  
339 we repeated some of these simulations but this time the null model was a  
340 birth–death model with  $\lambda$  and  $\mu$  constant through time, and the alternative  
341 model was with  $\lambda$  constant and  $\mu$  allowed to vary before and after 30 time  
342 units. In this situation, the test behaved as expected: the rejection rate  
343 was less than 5% when  $\mu$  was constant, whereas it varied between 8% and  
344 31% when the null hypothesis was false (Table 3). It is noteworthy that the  
345 present test to detect time-dependent extinction rate is not very powerful: it  
346 was necessary to simulate a strong contrast in  $\mu$  to reach a statistical power  
347 greater than 0.2.

348 The third set of simulations showed that the mixture-based LRT was  
349 able to detect heterogeneous diversification among two unknown categories  
350 of clades (Table 4). The test was more powerful when the contrast was  
351 due to different  $\lambda$  compared to different  $\mu$ . Otherwise, the test showed  
352 satisfactory statistical performance: its power increased with sample size  
353 and/or contrast in the parameters.

354 The distribution of the estimates of speciation rate under the Yule model,  
355  $\hat{\lambda}_{\text{Yule}}$ , shows that this estimator appeared unbiased when  $\mu = 0$  (Fig. 1A).  
356 On the other hand, when  $\mu > 0$  it was negatively biased though it can be ob-  
357 served that  $\hat{\lambda}_{\text{Yule}} > \lambda - \mu$  so this cannot be actually taken as an estimator of

358 the net diversification rate. The estimator based on the birth–death model,  
359  $\hat{\lambda}_{\text{BD}}$ , appears less biased, even though the presence of extinctions seems  
360 to induce a slightly more dispersed distribution of the estimates (Fig. 1B).  
361 The estimates of extinction rate based on the birth–death model,  $\hat{\mu}_{\text{BD}}$ , were  
362 almost unbiased (Fig. 1C).

### 363 **Application to Fish Data**

364 The fit of the Yule model to the fish data at the higher level ( $N = 22$ )  
365 resulted in a global estimate  $\hat{\lambda}_{\text{Yule}} = 0.058$  (SE = 0.002; AIC = 456). We  
366 tried to fit a birth–death model which led to a much improved fit (AIC =  
367 376); however, the likelihood function had a pronounced ridge on the line  
368  $\lambda = \mu$  (not shown). The fit of mixtures of Yule models with increasing  
369 number of categories ( $K$ ) showed that the best fit was with three categories  
370 (Table 5). The parameter estimates were:  $\hat{\lambda}_1 = 0.041$ ,  $\hat{\lambda}_2 = 0.080$ ,  $\hat{\lambda}_3 =$   
371  $0.013$ ,  $\hat{f}_1 = 0.65$ , and  $\hat{f}_2 = 0.10$ . The analysis of the combined taxonomic  
372 and phylogenetic data (Paradis 2003) gave  $\hat{\lambda} = 0.056$  and  $\hat{\mu} = 1.83 \times 10^{-7}$ .

373 The analysis at the level of the families ( $N = 97$ ) gave for the Yule  
374 model  $\hat{\lambda}_{\text{Yule}} = 0.0756$  (SE = 0.0016; AIC = 1483). Like above, the fit of  
375 the birth–death model resulted in a likelihood surface with a ridge on the  
376 line  $\lambda = \mu$ . The mixture of Yule models with the best fit had two categories  
377 (Table 5); the parameter estimates were:  $\hat{\lambda}_1 = 0.099$ ,  $\hat{\lambda}_2 = 0.036$ ,  $\hat{f} = 0.42$ .

378 The analysis with a model assuming continuous variation in  $\lambda$  across  
379 clades gave close results for both taxonomic levels. In both cases, the model  
380 fitted well and the AIC values were smaller than for any of the previous  
381 models (Table 6). Figure 2 shows the distribution of  $\lambda$  inferred with the  
382 estimated parameters. Trying to introduce  $\mu$  did not result in successful fits  
383 and the estimates of this parameter were close to zero.

384 Vega and Wiens (2012) reported the percentage of marine and freshwater  
385 species at both taxonomic levels. This was distributed very asymmetrically  
386 with most orders and families having only marine or freshwater species.  
387 Thus, we split the data into two groups whether they had more or less  
388 than 50% of marine species. A test of different speciation rates between  
389 these groups was performed. For orders, the difference was significant (LRT:  
390  $\chi_1^2 = 28.09$ ,  $P < 0.001$ ) with a larger estimate for marine orders ( $\hat{\lambda} = 0.063$ ,  
391  $SE = 0.002$ ) compared to the freshwater ones ( $\hat{\lambda} = 0.046$ ,  $SE = 0.002$ ).  
392 An examination of the data suggested that this result was dependent on  
393 Percomorpha which is one of the youngest clades in this data set and includes  
394 16,625 species (Fig. 3A). Removing this clade resulted in a non-significant  
395 test ( $\chi_1^2 = 2.10$ ,  $P = 0.147$ ,  $N = 21$ ). For families, an analogous result  
396 was found with a significant test (LRT:  $\chi_1^2 = 5.58$ ,  $P = 0.018$ ) comparing  
397 marine families ( $\hat{\lambda} = 0.079$ ,  $SE = 0.002$ ) and freshwater ones ( $\hat{\lambda} = 0.071$ ,  
398  $SE = 0.002$ ). This result was dependent on two families older than 200 Myr  
399 (Fig. 3B): the Amiidae (one species) and Polypteridae (12). Removing these  
400 two families led to a non-significant test:  $\chi_1^2 = 2.02$ ,  $P = 0.155$  ( $N = 95$ ).

#### 401 *Discussion*

402 The analysis of phylogenetic diversification with molecular data is enjoying  
403 a remarkable success in the literature. Some spectacular results have been  
404 accomplished using complete phylogenies (e.g., Goldberg et al. 2010; Hugall  
405 and Stuart-Fox 2012; Penney et al. 2012). Though complete phylogenies,  
406 possibly supplemented with fossil data, are probably the best way to inves-  
407 tigate evolutionary diversification, the goal of our study was to show the  
408 merit of an alternative approach based on the analysis of clade data.

409 Our modeling approach is based on the assumption that each clade is

410 characterized by its diversification parameters and variation among these pa-  
411 rameters can be quantified in a statistical way. Bokma (2003) and Paradis  
412 (2003) developed a method to combine information from high-level phylo-  
413 genies with clade data: both authors considered the simple constant-rate  
414 birth–death model. Alfaro et al. (2009) used similar combined data to as-  
415 sess variation among clades of vertebrates using a stepwise procedure (see  
416 details in Paradis 2012a). Thus the approach in the present paper comple-  
417 ments previous methodological developments. The possibility to quantify  
418 variation among clades with linear models seems a fruitful way to avoid  
419 overparameterization. Future applications will reinforce the relative merits  
420 of this approach.

421 Recently, Stadler and Bokma (2013) developed alternative likelihood  
422 functions with respect to the way higher taxa are defined. They showed  
423 that the estimation of speciation and extinction rates vary substantially de-  
424 pending on these definitions. While they considered only the constant-rate  
425 birth–death model, it seems possible and interesting to include their sam-  
426 pling scheme into the developments presented in the present paper.

427 Our modeling approach ignores the background phylogeny of the clades,  
428 the set of branches that link the clades together to make a higher-level  
429 phylogeny. There are two reasons for this. First, using information from  
430 the background phylogeny is straightforward when the rates of speciation  
431 and extinction are constant and homogenous, but when this assumption  
432 is relaxed it is not simple how one must assume changes in rates in the  
433 background tree. It is clear that if a well-supported background phylogeny  
434 is available, this might give additional information which can be combined  
435 with clade data (e.g., Paradis 2003). However, this extra information will  
436 in most cases require its own model since it relates to older diversification

437 events compared to clade data. On the other hand, ignoring backbone phy-  
438 logeny and assuming that the clades are independent units simplifies the  
439 definition of alternative models as done in this paper. Second, though some  
440 higher-level phylogenies are available (mammals, birds), we believe these are  
441 still exceptions rather than the rule. For instance, the basal relationships  
442 of reptiles, amphibians, or fishes are still debated. Therefore, having the  
443 possibility to analyze their clade data without the need of a background  
444 phylogeny is of some general application. Furthermore, the present ap-  
445 proach can be used when analyzing sets of clades across different phyla, for  
446 instance arthropods, echinoderms, vertebrates, etc., where the background  
447 phylogeny would not be very informative since this would branch at the  
448 origin of Metazoa.

449 The use of mixtures as an approach to analyze heterogeneity in diver-  
450 sification rates is not limited to clade data. For instance, one could model  
451 speciation and extinction rates on a fully-resolved phylogeny assuming that  
452 these parameters vary among its branches though we do not know a pri-  
453 ori which sections of the tree evolved fast and which others evolved slowly.  
454 Furthermore, the mixture approach can also be used to model variation in  
455 rates of trait evolution along a phylogeny. In that case, the variation may be  
456 among branches (as in the previous example), or among traits where some  
457 traits are assumed to evolve faster but we do not know which ones.

458 Some subtle but important facts come from the results of the simulation  
459 study. Even though most of the tests considered here assumed  $\mu = 0$ , they  
460 appeared not to be tests of equal diversification. If the net diversification  
461 rates ( $\lambda - \mu$ ) were equal among clades, the tests rejected the null hypothesis  
462 in more than 5% (see eighth row of Table 1). On the other hand, if  $\lambda$  was  
463 equal among clades, the tests detected differences in  $\mu$ . It is clear that results

464 based only on the Yule model must be interpreted with caution.

465 The tests of temporal variation showed some contrasted but interesting  
466 results. When the extinction rate was zero, these tests performed very well  
467 and were able to detect either a decrease or an increase in speciation rate.  
468 However, when extinction rate was not null, the tests based on the Yule  
469 model showed poor performance with an increased type I error rate and a  
470 high type II error rate (frequency of accepting the null hypothesis when it is  
471 false) when  $\mu$  decreased through time. These poor performances were cor-  
472 rected if the assumption  $\mu = 0$  was relaxed (i.e., if a null birth–death model  
473 was used in place of the Yule one), though the test had low power. Some  
474 of these results make sense: the increased type I error rate obtained with  
475 the Yule model is clearly due to the fact that a pattern of accelerated spe-  
476 ciation can be created under a diversification process with extinction, when  
477 old lineages are mostly extinct (e.g., Paradis 2011). On the other hand, the  
478 high type II error rate of the same model when extinction rate increased  
479 through time is somehow surprising considering the widely reported results  
480 of slowing-down diversification (Rabosky and Lovette 2008b,a; Morlon et al.  
481 2011; Etienne and Haegeman 2012, among others). Obviously, the same test  
482 was not used in these studies, so this clearly requires further investigation.  
483 Besides, the result that the test based on a birth–death model shows statis-  
484 tically consistent results (i.e., the null hypothesis was rejected in less than  
485 5% when  $\mu$  was constant and in more than 5% when this parameter varied  
486 through time) is encouraging and will also be further investigated. Interest-  
487 ingly, this test was more powerful when the extinction rate increased through  
488 time.

489 A particularly interesting result comes from the precision of the estimator  
490 of extinction rate,  $\hat{\mu}_{\text{BD}}$ , which appears to have a very small bias, even when

491 the data were simulated with a relatively large value of  $\mu$ . This contrasts  
492 with previous studies showing that the estimator of extinction rate based  
493 on complete phylogenies is, overall, inaccurate except if it is small compared  
494 to the speciation rate (Paradis 2004; Didier et al. 2012). This result is  
495 important because several authors have cast doubt on the possibility to  
496 estimate with some precision extinction rates without fossils (Paradis 2011;  
497 Aldous et al. 2011).

498 The analysis with the fish data were essentially illustrative, but the re-  
499 sults call for several comments. The present method seems successful in  
500 quantifying variation in diversification rates from a sample of clades. The  
501 difference in the results from both taxonomic levels makes sense since we  
502 expect more variation among families than among orders. The AIC values  
503 evidence that the model assuming continuous variation in  $\lambda$  across clades  
504 fits better than a model with discrete variation in this parameter. Since  
505 similar tests have not been done with other data, this clearly calls for fur-  
506 ther analyses before concluding whether diversification varies continuously  
507 or discretely across clades.

508 The apparent failure to estimate the extinction rate,  $\mu$ , of fishes is disap-  
509 pointing since our simulation study showed that this parameter can be es-  
510 timated correctly with the present approach. The fossil record shows many  
511 episodes of radiations, extinctions, and turn-over during the evolutionary  
512 history of fishes (Friedman and Sallan 2012). So the reality is very differ-  
513 ent from the homogeneous scenario used in our simulations. Our results  
514 combined with previous studies (e.g., Aldous et al. 2011) suggest that the  
515 estimators of  $\mu$  are far more complex when rate heterogeneity is present  
516 which is likely the case with most real data set.

517 Vega and Wiens (2012) addressed the paradox of equivalent species di-

518 versity between marine and freshwater fishes despite the fact that freshwater  
519 environments occupy a considerably smaller fraction of the Earth's surface  
520 than oceans. In particular they wondered whether this could be related to  
521 differences in diversification rates. Our results are in agreement with these  
522 authors' who tested their hypothesis by correlating the proportion of marine  
523 species in a clade with the method-of-moment estimator from Magallón and  
524 Sanderson (2001). We found significant differences in  $\lambda$  between marine and  
525 freshwater clades from the raw data; however, the small difference in  $\hat{\lambda}$  be-  
526 tween both groups suggested the influence of one or two clades. Hopefully,  
527 the analysis of a more comprehensive data set with the statistical tools intro-  
528 duced in this paper will help to solve the paradox of less biological diversity  
529 in the ocean (Mora et al. 2011).

### 530 *Acknowledgments*

531 We are grateful to four anonymous reviewers, the Associate Editor, and  
532 Laure Kubatko for their constructive comments on previous versions of our  
533 manuscript. Financial support was provided by grant ANR-09-PEXT-008.

### 534 *References*

- 535 Akaike, H., 1973. Information theory and an extension of the maximum  
536 likelihood principle. Pages 267–281 *in* B. N. Petrov and F. Csaki, edi-  
537 tors. Proceedings of the Second International Symposium on Information  
538 Theory. Akadémia Kiado, Budapest.
- 539 Aldous, D. J., M. A. Krikun, and L. Popovic. 2011. Five statistical questions  
540 about the tree of life. *Syst. Biol.* 60:318–328.
- 541 Alfaro, M. E., F. Santini, C. Brock, H. Alamillo, A. Dornburg, D. L. Ra-  
542 bosky, G. Carnevale, and L. J. Harmon. 2009. Nine exceptional radiations



543 plus high turnover explain species diversity in jawed vertebrates. Proc.  
544 Natl. Acad. Sci. USA 106:13410–13414.

545 Bokma, F. 2003. Testing for equal rates of cladogenesis in diverse taxa.  
546 Evolution 57:2469–2474.

547 Didier, G., M. Royer-Carenzi, and M. Laurin. 2012. The reconstructed  
548 evolutionary process with the fossil record. J. Theor. Biol. 315:26–37.

549 Etienne, R. S. and B. Haegeman. 2012. A conceptual and statistical frame-  
550 work for adaptive radiations with a key role for diversity dependence. Am.  
551 Nat. 180:E75–E89.

552 FitzJohn, R. G. 2010. Quantitative traits and diversification. Syst. Biol.  
553 59:619–633.

554 FitzJohn, R. G., W. P. Maddison, and S. P. Otto. 2009. Estimating trait-  
555 dependent speciation and extinction rates from incompletely resolved phy-  
556 logenies. Syst. Biol. 58:595–611.

557 Flury, B. D., J.-P. Airoidi, and J.-P. Biber. 1992. Gender identification of  
558 water pipits (*Anthus spinoletta*) using mixtures of distributions. J. Theor.  
559 Biol. 158:465–480.

560 Friedman, M. and L. C. Sallan. 2012. Five hundred million years of extinc-  
561 tion and recovery: a Phanerozoic survey of large-scale diversity patterns  
562 in fishes. Palaeontology 55:707–742.

563 Goldberg, E. E., J. R. Kohn, R. Lande, K. A. Robertson, S. A. Smith, and  
564 B. Igić. 2010. Species selection maintains self-incompatibility. Science  
565 330:493–495.

566 Hallinan, N. 2012. The generalized time variable reconstructed birth–death  
567 process. J. Theor. Biol. 300:265–276.

568 Hugall, A. F. and D. Stuart-Fox. 2012. Accelerated speciation in colour-  
569 polymorphic birds. Nature 485:631–634.

- 570 Kendall, D. G. 1948. On the generalized “birth-and-death” process. *Ann.*  
571 *Math. Stat.* 19:1–15.
- 572 Maddison, W. P., P. E. Midford, and S. P. Otto. 2007. Estimating a binary  
573 character’s effect on speciation and extinction. *Syst. Biol.* 56:701–710.
- 574 Magallón, S. and M. J. Sanderson. 2001. Absolute diversification rates in  
575 angiosperm clades. *Evolution* 55:1762–1780.
- 576 McPeck, M. A. 2008. The ecological dynamics of clade diversification and  
577 community assembly. *Am. Nat.* 172:E270–E284.
- 578 Mora, C., D. P. Tittensor, S. Adl, A. G. B. Simpson, and B. Worm. 2011.  
579 How many species are there on Earth and in the ocean? *PLoS Biol.*  
580 9:e1001127.
- 581 Morlon, H., T. L. Parsons, and J. B. Plotkin. 2011. Reconciling molecular  
582 phylogenies with the fossil record. *Proc. Natl. Acad. Sci. USA* 108:16327–  
583 16332.
- 584 Nee, S., R. M. May, and P. H. Harvey. 1994. The reconstructed evolutionary  
585 process. *Phil. Trans. R. Soc. Lond. B* 344:305–311.
- 586 Nee, S., A. Ø. Mooers, and P. H. Harvey. 1992. Tempo and mode of  
587 evolution revealed from molecular phylogenies. *Proc. Natl. Acad. Sci.*  
588 *USA* 89:8322–8326.
- 589 Paradis, E. 2003. Analysis of diversification: combining phylogenetic and  
590 taxonomic data. *Proc. R. Soc. Lond. B* 270:2499–2505.
- 591 Paradis, E. 2004. Can extinction rates be estimated without fossils? *J.*  
592 *Theor. Biol.* 229:19–30.
- 593 Paradis, E. 2005. Statistical analysis of diversification with species traits.  
594 *Evolution* 59:1–12.
- 595 Paradis, E. 2011. Time-dependent speciation and extinction from phyloge-  
596 nies: a least squares approach. *Evolution* 65:661–672.

- 597 Paradis, E. 2012a. Analysis of phylogenetics and evolution with R (second  
598 edition). Springer, New York.
- 599 Paradis, E. 2012b. Shift in diversification in sister-clade comparisons: a  
600 more powerful test. *Evolution* 66:288–295.
- 601 Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phyloge-  
602 netics and evolution in R language. *Bioinformatics* 20:289–290.
- 603 Penney, H. D., C. Hassall, J. H. Skevington, K. R. Abbott, and T. N. Sher-  
604 ratt. 2012. A comparative analysis of the evolution of imperfect mimicry.  
605 *Nature* 483:461–464.
- 606 Purvis, A., S. Nee, and P. H. Harvey. 1995. Macroevolutionary inferences  
607 from primate phylogeny. *Proc. R. Soc. Lond. B* 260:329–333.
- 608 Pybus, O. G., A. Rambaut, E. C. Holmes, and P. H. Harvey. 2002. New in-  
609 ferences from tree shape: numbers of missing taxa and population growth  
610 rates. *Syst. Biol.* 51:881–888.
- 611 R Development Core Team. 2012. R: a language and environment for  
612 statistical computing. R Foundation for Statistical Computing, Vienna.  
613 Available at <http://www.R-project.org>.
- 614 Rabosky, D. L., S. C. Donnellan, A. L. Talaba, and I. J. Lovette. 2007.  
615 Exceptional among-lineage variation in diversification rates during the  
616 radiation of Australia’s most diverse vertebrate clade. *Proc. R. Soc. Lond.*  
617 *B* 274:2915–2923.
- 618 Rabosky, D. L. and I. J. Lovette. 2008a. Density-dependent diversification  
619 in North American wood warblers. *Proc. R. Soc. Lond. B* 275:2363–2371.
- 620 Rabosky, D. L. and I. J. Lovette. 2008b. Explosive evolutionary radiations:  
621 decreasing speciation or increasing extinction through time? *Evolution*  
622 62:1866–1875.
- 623 Ricklefs, R. E. 2007. Estimating diversification rates from phylogenetic

- 624 information. *Trends Ecol. Evol.* 22:601–610.
- 625 Sanderson, M. J. and M. J. Donoghue. 1996. Reconstructing shifts in diver-  
626 sification rates on phylogenetic trees. *Trends Ecol. Evol.* 11:15–20.
- 627 Stadler, T. 2011. Mammalian phylogeny reveals recent diversification rate  
628 shifts. *Proc. Natl. Acad. Sci. USA* 108:6187–6192.
- 629 Stadler, T. and F. Bokma. 2013. Estimating speciation and extinction rates  
630 for phylogenies of higher taxa. *Syst. Biol.* 62:220–230.
- 631 Vega, G. C. and J. J. Wiens. 2012. Why are there so few fish in the sea?  
632 *Proc. R. Soc. Lond. B* 279:2323–2329.
- 633 Yule, G. U. 1924. A mathematical theory of evolution, based on the conclu-  
634 sions of Dr. J. C. Willis, F.R.S. *Phil. Trans. R. Soc. Lond. B* 213:21–87.

**Table 1.** Rejection rate for the test of equality of diversification rate between two categories with  $n_1$  and  $n_2 (= n_1)$  clades.

$\lambda_1$	$\mu_1$	$\lambda_2$	$\mu_2$	$n_1$				
				1	3	5	10	20
0.1	0	0.1	0	0.044	0.060	0.066	0.054	0.056
0.1	0	0.1	0.05	0.038	0.085	0.108	0.140	0.203
0.15	0	0.1	0	0.094	0.177	0.232	0.418	0.711
0.15	0	0.1	0.05	0.112	0.297	0.446	0.759	0.958
0.2	0	0.1	0	0.174	0.497	0.707	0.943	0.998
0.2	0	0.1	0.05	0.236	0.643	0.855	0.993	1.000
0.1	0.05	0.1	0.05	0.040	0.054	0.048	0.054	0.061
0.15	0.05	0.1	0	0.050	0.075	0.083	0.123	0.209
0.15	0.05	0.1	0.05	0.069	0.129	0.201	0.387	0.653
0.2	0.05	0.1	0	0.119	0.240	0.384	0.693	0.928
0.2	0.05	0.1	0.05	0.143	0.407	0.618	0.878	0.995

**Table 2.** Rejection rate for the test of temporal variation in diversification. The null model was a Yule model with constant rate, and the alternative model was a Yule model with  $\lambda$  allowed to take different values before and after 30 time units. The first two pairs of columns give the parameter values used for the simulations (Ancient and Recent: values before and after 30 time units).

Ancient		Recent		Rejection rate
$\lambda$	$\mu$	$\lambda$	$\mu$	
0.01	0	0.01	0	0.029
0.1	0	0.1	0	0.038
0.1	0	0.05	0	0.917
0.1	0	0.01	0	1.000
0.05	0	0.1	0	0.923
0.01	0	0.1	0	1.000
0.1	0.025	0.1	0.025	0.105
0.1	0.05	0.1	0.05	0.248
0.1	0.025	0.1	0.075	0.057
0.1	0.075	0.1	0.025	0.682

**Table 3.** Same than in Table 2 but the null model was a birth–death model with constant rates, and the alternative model was a model with  $\lambda$  constant and  $\mu$  allowed to take different values before and after 30 time units.

Ancient		Recent		Rejection rate
$\lambda$	$\mu$	$\lambda$	$\mu$	
0.1	0.05	0.1	0.05	0.019
0.1	0.075	0.1	0.025	0.080
0.1	0.025	0.1	0.075	0.121
0.1	0	0.1	0.08	0.313
0.1	0.08	0.1	0	0.211

**Table 4.** Rejection rate for the test of equality of diversification rate between two unknown categories using mixtures with  $n$  clades in each category.

$\lambda$	$\mu$	$n$		
		10	20	50
0.1	0	0.011	0.006	0.011
(0.1, 0.2)	0	0.235	0.548	0.929
0.2	(0, 0.1)	0.057	0.143	0.369
0.2	(0, 0.15)	0.199	0.423	0.829



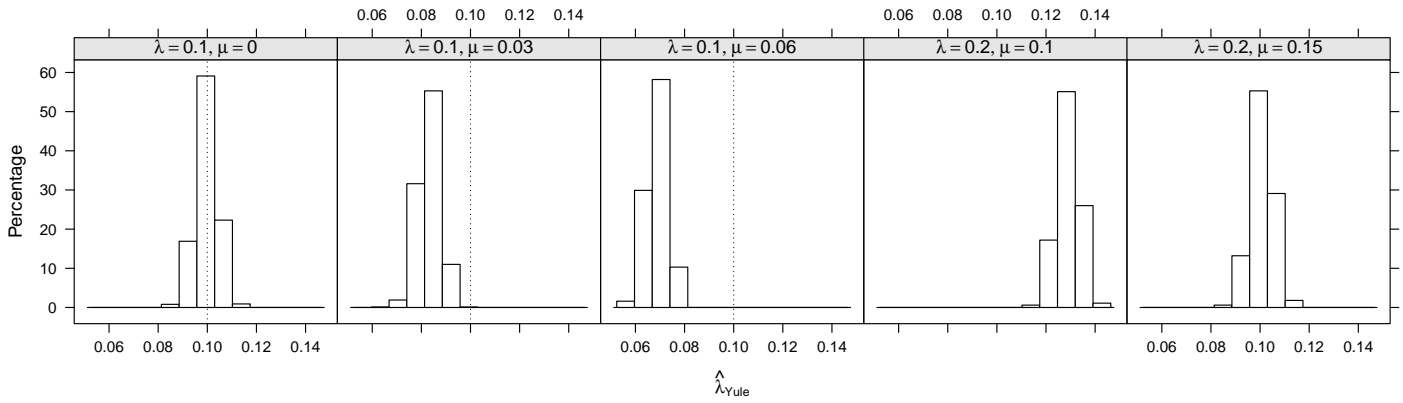
**Table 5.** Results of fitting models to the fish data using mixtures of Yule processes with  $K$  from two to seven.

$K$	Orders		Families	
	$\ln L$	AIC	$\ln L$	AIC
2	-171.207	348.414	-599.846	1205.691
3	-158.381	326.763	-599.846	1209.691
4	-171.207	356.414	-599.846	1213.691
5	-171.207	360.414	-599.846	1217.691
6	-171.207	364.414	-599.846	1221.691
7	-171.207	368.414	-599.846	1225.691

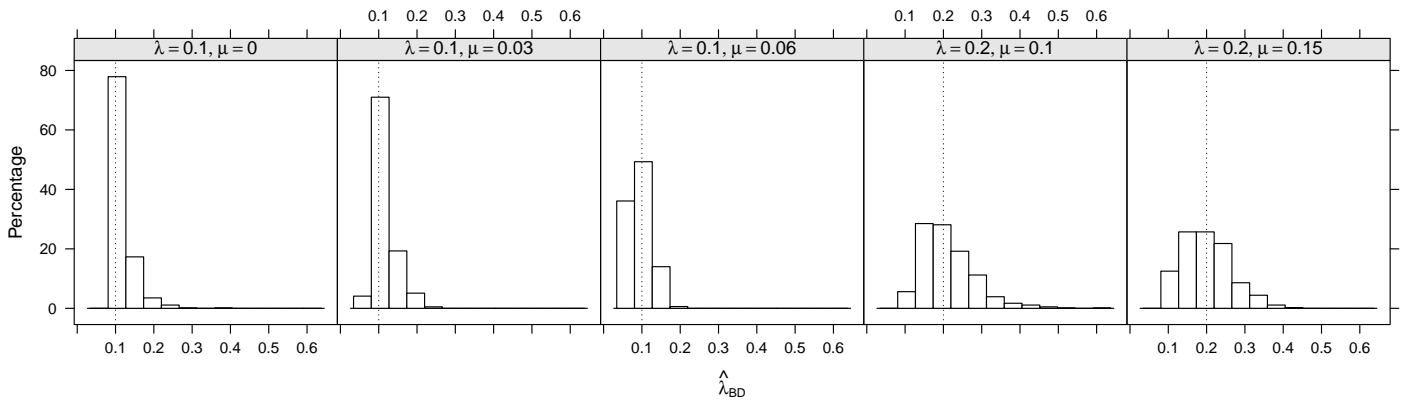
**Table 6.** Results of fitting a model of continuous variation in speciation rate across orders ( $N = 22$ ) and families ( $N = 97$ ) of fish.

	AIC	$\hat{\mu}_\lambda$ (SE)	$\hat{\sigma}_\lambda$ (SE)
Orders	320.396	1.221 (0.039)	0.163 (0.032)
Families	1137.066	1.086 (0.026)	0.224 (0.022)

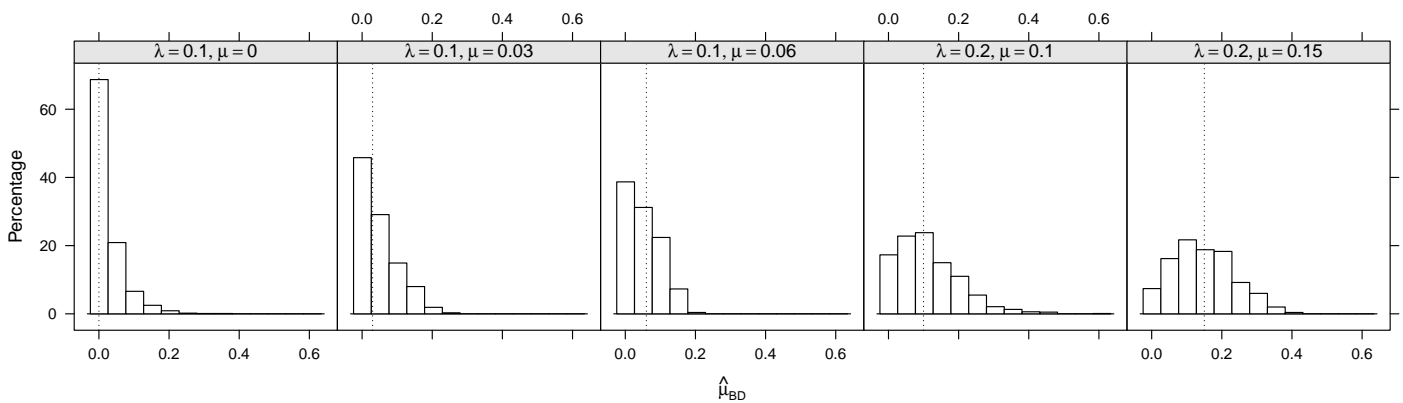
A)



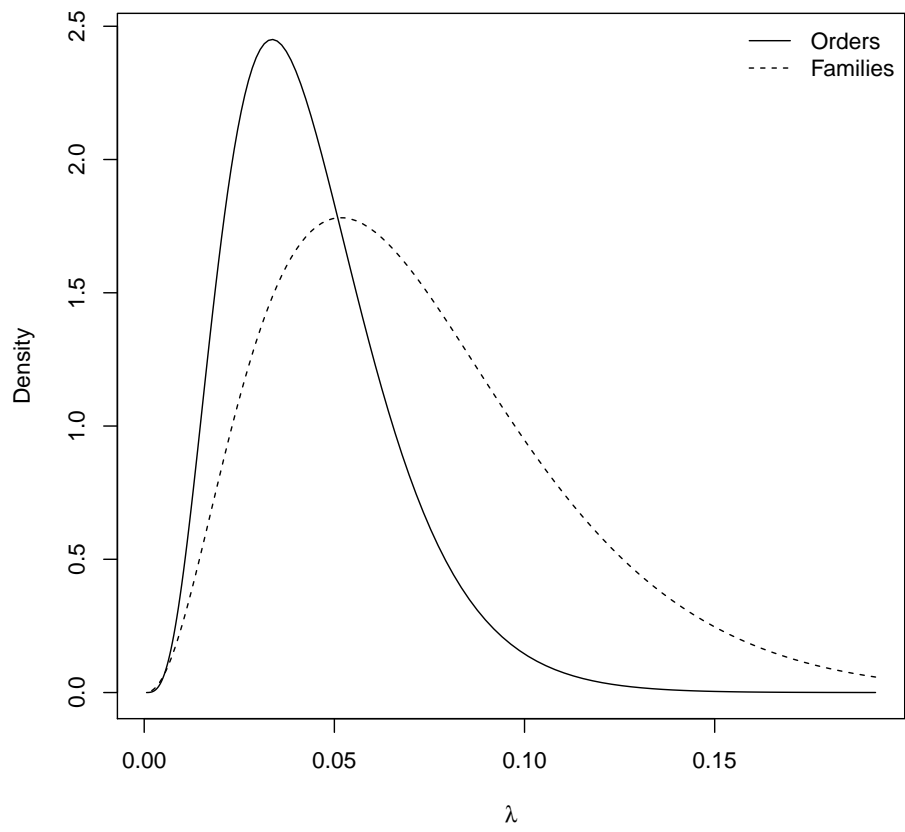
B)



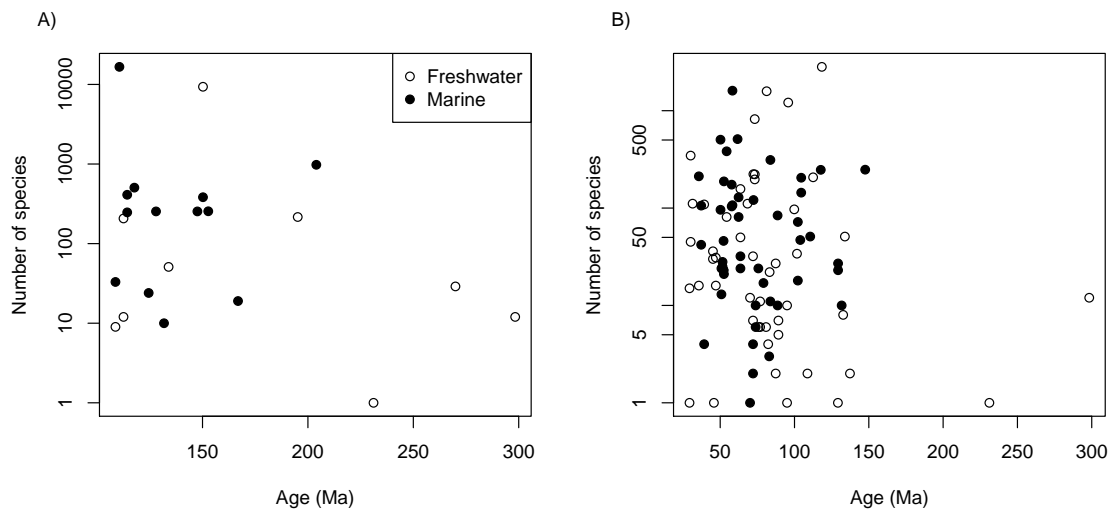
C)



**Figure 1.** Distribution of the estimates of  $\lambda$  and  $\mu$  with (A) the Yule model ( $\hat{\lambda}_{\text{Yule}}$ ) and (B and C) the birth–death model ( $\hat{\lambda}_{\text{BD}}$  and  $\hat{\mu}_{\text{BD}}$ ) under five sets of parameters (values are given in the strips). Note the different scales of the  $x$ -axes. The vertical dotted lines indicate the values of  $\lambda$  (A and B) or  $\mu$  (C) used in the simulation (not visible if outside the range of the  $x$ -axis). In all cases  $n = 100$  clades.



**Figure 2.** Inferred distribution of speciation rate among orders and families of fish.



**Figure 3.** Number of species with respect to stem clade age for (A) orders and some super-orders and (B) families of fish (the legend is the same for both plots).