

1 The error in Bayesian phylogenetic reconstruction  
2 when speciation co-occurs

3 Giovanni Laudanno<sup>1</sup>, Richèl J.C. Bilderbeek<sup>1</sup>, and Rampal S.  
4 Etienne<sup>1</sup>

5 <sup>1</sup>Groningen Institute for Evolutionary Life Sciences, University of  
6 Groningen, Groningen, The Netherlands

7 May 7, 2019

## Abstract

9

10

11

12

There exist millions of species on Earth, all originating from a common ancestor billions of years ago. The field of phylogenetics uses heritable material to determine which species are closest related and what are the mathematics that shape speciation.

13

14

15

16

In Bayesian phylogenetics, a DNA/RNA/protein alignment is used to infer a distribution of phylogenies and parameter estimates. To do so, we use assumptions that may be biologically unrealistic, but may give tolerable errors.

17

18

19

Contemporary inference assumes that speciation never co-occurs.

Here we show the error we make in our inference, when nature has varying degrees of co-occurring speciation.

20

21

**Keywords:** computational biology, evolution, phylogenetics, Bayesian analysis, tree prior

22

## 1 Introduction

23

24

25

26

- There are many contemporary tools that provide the possibility to infer a phylogeny from genetic data (DNA, RNA, proteins). A popular Bayesian phylogenetic tool is called BEAST (Drummond & Rambaut 2007) and its cousin BEAST2 (Bouckaert *et al.* 2014).

27

28

- BEAST is very flexible, providing the user with the option to set up all possible phylogenetic priors (e.g. site/clock/speciation model).

29

30

31

32

- However, currently available priors can be not suitable to analyze some specific datasets. With this work we aim to test whether or not the implementation of a new prior model is beneficial to study a specific kind of diversification process.

- 33 • BEAST2 gives us the possibility to introduce new tree priors to infer  
34 phylogenies based on different assumptions on how the speciation process  
35 takes place.
- 36 • One of such speciation processes is the multiple birth hypothesis, a new  
37 model (described below) and thus currently absent in BEAST.
- 38 • The Multiple birth hypothesis can be useful to explain a phenomenon  
39 that has always puzzled evolutionary biologists: what are the drivers of  
40 the diversification processes for those phylogenies that show an impressive  
41 amount of speciation events in relatively short times? The (constant-rate)  
42 birth-death (BD) model embodies the common assumption that only a  
43 single speciation event can occur at any given time. The multiple-birth-  
44 death (MBD) model relaxes this assumption, allowing events in which  
45 large-scale environmental changes lead to a great number of species in  
46 relatively short time intervals. Such a hypothesis may be a better fit to  
47 describe the burst in systems like cichlid fish diversification in the African  
48 Great Lakes: Malawi, Tanganyika and Victoria (Janzen *et al.* 2016, Janzen  
49 *et al.* 2017).
- 50 • However, it may be that current BD tree priors are good enough at de-  
51 tecting such events, with a (preferred) lower level of complexity. If this is  
52 the case one should always be more keen to adopt the simplest model.
- 53 • Here we present our study with the aim of exploring when using a more  
54 complex MBD tree prior is warranted. To do so, we simulate phylogenies  
55 using the MBD process, with varying degrees of that process. To be  
56 explicit, we define that degree,  $s$ , as the number of extinct and extant  
57 species created during a co-occurring speciation event,  $N_M$ , from the total  
58 number of extinct and extant species:

$$s = \frac{N_{\mathbb{M}}}{N_{\mathbb{M}} + N_{\mathbb{B}}} \quad (1)$$

Here,  $N_{\mathbb{B}}$  is the number of extinct and extant species created during a default single-birth speciation event.

From such phylogenies with different degrees of  $s$ , we measure the inference error we make today, would nature follow such a phylogeny. The inference error we make today is caused by the assumption of a BD process and by inherent noise in this inference.

We have the hypothesis  $\mathcal{H}_1$  that, for a higher  $s$ , the inference error  $e$  will increase:

$$e = f(s) \quad (2)$$

Where  $f$  is a monotonously increasing function of unknown shape.

The MBD process has multiple components, that may cause different values of  $e$  for identical values of  $s$ .

One MBD component that may cause different values of  $e$  for identical values of  $s$  is the MBD regime: which can be many modest multiple-speciation events, or few intense ones. As we have no prior expectations, we have the (null) hypothesis,  $\mathcal{H}_7$ , that the MBD regime has no effect on  $e$ .

Another MBD component that may cause different values of  $e$  for identical values of  $s$  is the effect of extinction. As extinctions will hit lineages created by both speciation processes equally, and we have no additional prior expectations, we have the (null) hypothesis,  $\mathcal{H}_2$ , that extinction has no effect on  $e$ .

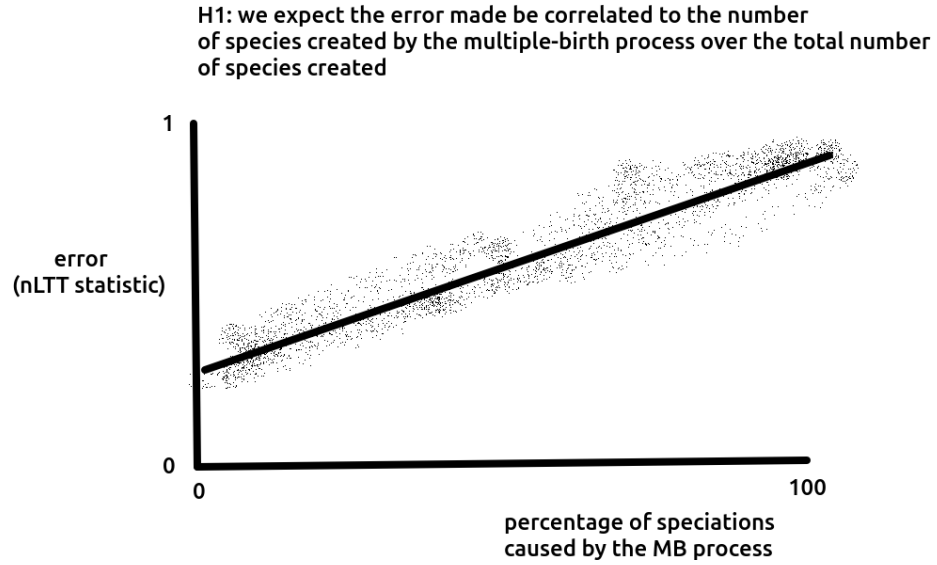


Figure 1: Hypothesis 1: we expect the error made be correlated to the number of species created by the multiple-birth process over the total number of species created

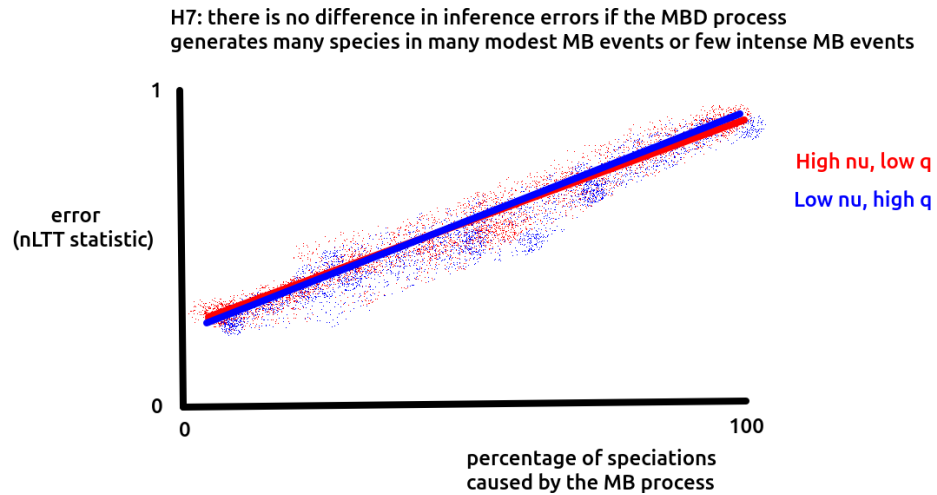


Figure 2: Hypothesis 7: there is no difference in inference errors if the MBD process generates many species in many modest MB events or few intense MB events

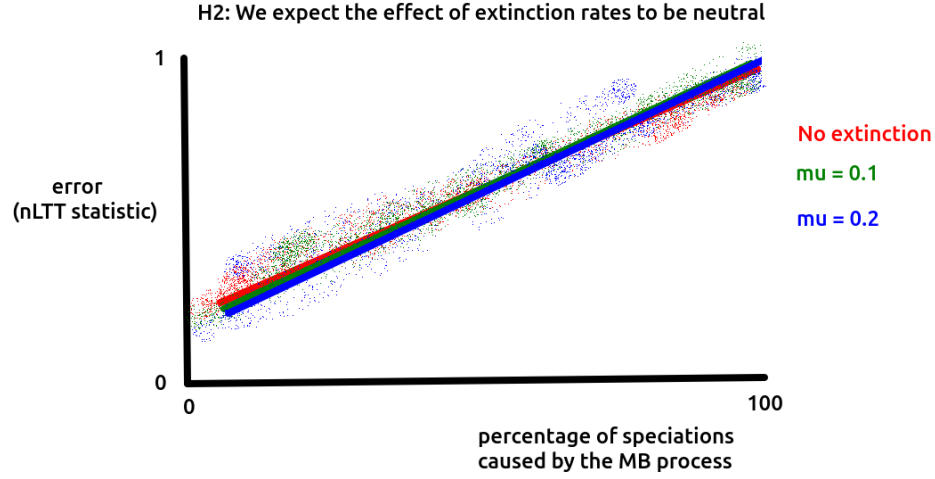


Figure 3: Hypothesis 2: the effect of extinction rates is neutral

Another MBD component that may cause different values of  $e$  for identical values of  $s$ , is the timing of a multiple-birth event: be it close to the crown age or close to the present. Compared to a late multiple birth event, an early multiple birth event may have a longer-lasting effect (as the next speciation event will be later), but it will create less new species, as there are still fewer taxa. As we have no prior expectations, we have the (null) hypothesis,  $\mathcal{H}_3$ , that the timing of multiple-birth events has no effect on  $e$ .

Another MBD component that may cause different values of  $e$  for identical values of  $s$ , is the number of taxa in a phylogeny. As there is no diversity dependency in any of the processes and we have no further prior expectations, we have the (null) hypothesis  $\mathcal{H}_5$ , that the number of taxa has no effect on  $e$ . As a higher number of taxa increases the information content in a phylogeny, we have hypothesis  $\mathcal{H}_6$  that the variance in  $e$  decreases.

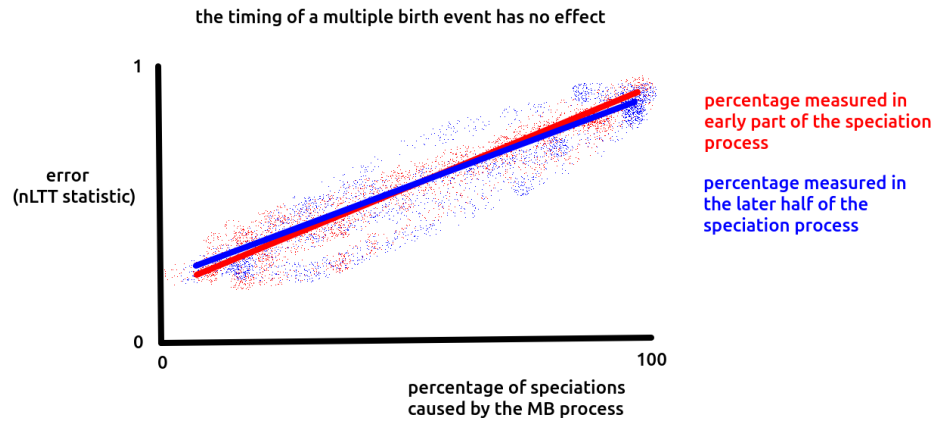


Figure 4: Hypothesis 3: the timing of a multiple birth event has no effect.

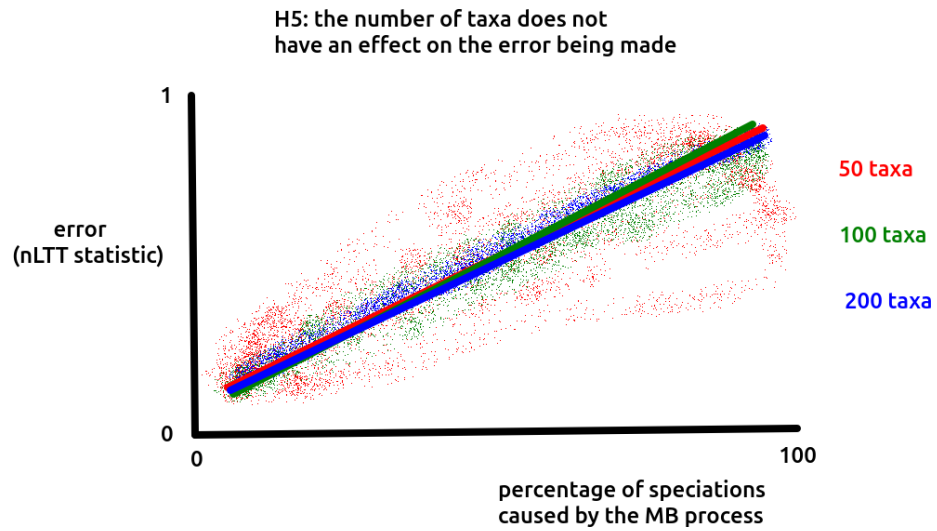


Figure 5: Hypothesis 5: the number of taxa does not have an effect on the error being made

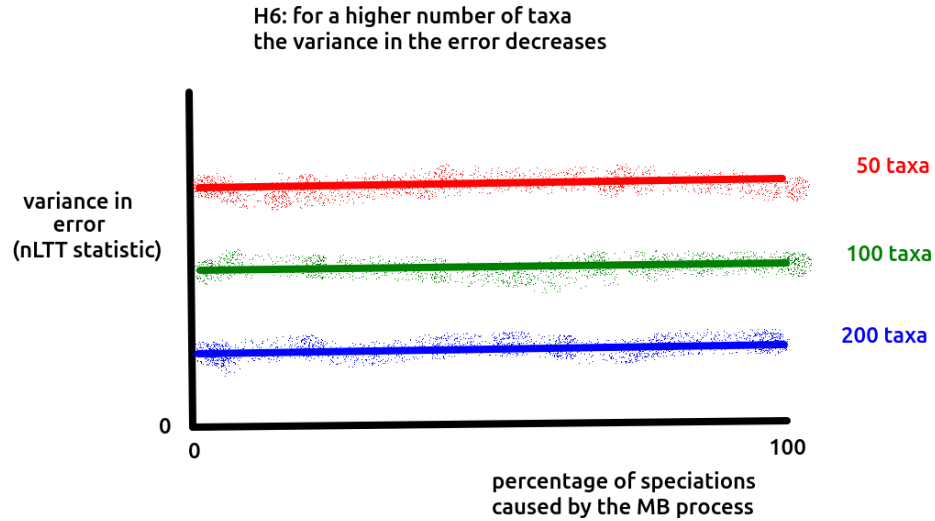


Figure 6: Hypothesis 6: for a higher number of taxa the variance in the error decreases

## 2 Methods

### 2.1 Model

- [RJCB: TODO: move to Introduction] Current phylogenetic tools assume that only a single speciation event can occur at any given time. While this assumption is useful to construct a wide variety of successful models (e.g Maddison *et al.* 2007, Valente *et al.* 2015, Etienne *et al.* 2012, Etienne *et al.* 2014), they disallow for environmental changes that trigger speciations in multiple clades at a same point in time.
- [RJCB: TODO: move to Introduction] In the MBD model, parameters  $\lambda$  and  $\mu$  correspond, respectively, to the common per-species speciation and extinction rates present also in the standard BD model. Additionally, MBD relies on two additional parameters. Parameter  $\nu$  is the rate at which an environmental change is triggered. When such event is triggered, all species present in the phylogeny at that moment have a



108 probability  $q$  to speciate at that time, which is independent on  $\lambda$ . Poly-  
 109 tomies are not allowed in such process as each species can speciate only  
 110 once at the time.

- 111 • [RJC: TODO: move to Introduction] It is also possible to write  
 112 down a likelihood function for such processes as in Laudanno 2018.

## 113 2.2 Simulations

- 114 • To investigate the effect of  $s$  on  $e$ , we simulate phylogenies for different  
 115 values of  $s$  spread equally from zero (no multiple-birth event, thus a BD  
 116 model) to one (species are created only from multiple-birth events) and  
 117 three intermediate values of 0.25, 0.5 and 0.75. Aggregating all compo-  
 118 nents of  $s$ , we show the effect of  $s$  on  $e$  in figure 9.

- 119 • To see the effect of the number of taxa on the relation between  $s$  and  $e$ , we  
 120 simulate phylogenies for different number of extant taxa  $n$ . We use values  
 121 of  $n$  of 50, 100 and 200 extant taxa, as this will result in phylogenies that  
 122 are big enough to be useful, yet small of enough to be computationally  
 123 feasible. For the different values of  $n$ , we show the relation between  $s$  and  
 124  $e$  in figure 10 and its effect on the variance of  $e$  in figure 11.

- 125 • To see the effect of the extinction rate on the relation between  $s$  and  $e$ ,  
 126 we simulate phylogenies with different extinction rates  $\mu$ . We use values  
 127 of  $\mu$  of 0.0, 0.1 and 0.2. An extinction rate of zero has two features: (1)  
 128 the model falls back to a pure multiple-birth model, (2) all multiple-birth  
 129 events are observed. For the different values of  $\mu$ , we show the relation  
 130 between  $s$  and  $e$  in figure 12.

- 131 • We start simulating  $N_S = 1000$  MBD trees, with either 50, 100 and 200  
 132 taxa.

- From each MBD tree, a DNA sequence alignment is simulated. For each sequence alignment we then perform a Bayesian analysis to recover a posterior distribution of trees, each composed of  $N_P$  phylogenies. Such analysis is performed using the 'pirouette' package (Bilderbeek 2018) to call the BEAST2 tool suite from R. We let the Bayesian analysis assume a BD prior in both cases, to investigate the extent of the error we make under this assumption.
- For each tree generated under the MBD model we aim to generate a "twin" tree under the BD model. With the word "twin" we denote a tree generated starting from the respective MBD tree, in order to perform a fair comparison with it. This operation has to be done, because we want to compare two trees that are generated by different processes. To do so we infer the parameters  $\lambda_{BD}$  and  $\mu_{BD}$  from the MBD maximizing the likelihood under a BD model. To perform this operation we use the function "bd\_ML" from the package "DDD" (Etienne *et al.* 2012).
- We then exploit such parameters to generate a BD tree using the function "tess.sim.taxa.age" from the package "TESS" (Hhna 2013). We simulate the tree in such a way the new tree has the same number of tips and the same crown age as the MBD tree. We furthermore require that the BD tree conserve the topology of the MBD tree. We have hypothesis H4 that, compared to the MBD trees, the error will be less in the BD twin tree. The difference between the errors made in MBD and twin BD trees indicates the impact the MBD process has on the error we make in inference using a contemporary BD prior.

We want the MBD and twin BD trees to contain the same amount of information, i.e. the same number of DNA mutations and the same number of taxa at the present:

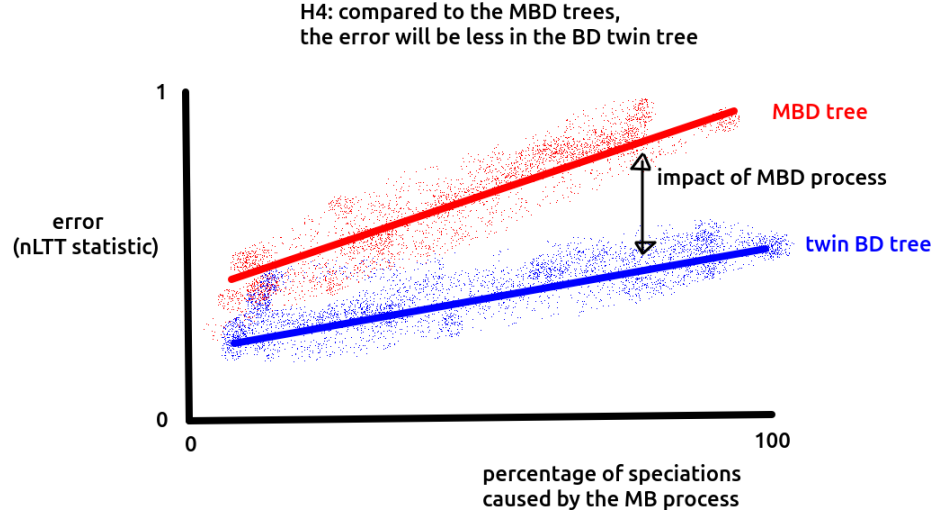


Figure 7: Hypothesis 4: compared to the MBD trees, the error will be less in the BD twin tree

$$m_{MBD} = m_{BD} \quad (3)$$

160 The expected number of mutations  $m$  of a phylogeny with crown age  $-T$   
 161 (with  $T > 0$ ) in fact is given by [RJCB: So one of use likes '-T', the  
 162 other likes 'T'. How to resolve this? ]

$$m = L \cdot \rho \cdot \int_0^T n(t) dt \quad (4)$$

163 where  $L$  is the number of DNA nucleotides,  $\rho$  is the per-site per-species  
 164 mutation rate and  $n(t)$  the number of species at each time.

165 The parameter we'll tune is  $\rho$  ... [RJCB: elaborate here :-)]

166 Since we cannot know  $n_{BD}(t)$  before running simulations we need to re-  
 167 place it with a proxy. For this reason we will use the average number of  
 168 species in time according to the BD model. It's well known that this is  
 169 equal to [GL: insert proper citation]

$$< n_{BD} > (t) = n_0 \cdot e^{(\mu_{BD} - \lambda_{BD})t} \quad (5)$$

where  $n_0 = n_{BD}(-T) = n_{MBD}(-T)$  is the initial number of species at the crown age. From 3, 4 and 5 follows:

$$m_{MBD} = L \cdot \rho \cdot \int_0^T < n_{BD} > (t) dt = L \cdot \rho \cdot n_0 \cdot \left[ \frac{e^{(\mu_{BD} - \lambda_{BD})T} - 1}{\mu_{BD} - \lambda_{BD}} \right] \quad (6)$$

If we set  $\mu_{BD} = \mu_{MBD}$  and reverse this relation we can extrapolate the value of  $\lambda_{BD}$  to use to generate BD trees.

- We explained how we set the parameters for each twin BD tree. Using this rules we generate a BD dataset. We repeat the analysis, producing alignments for each tree and subsequently using BEAST to produce a posterior for each of them.

## 2.3 Measuring the inference error

- So far we have simulated two datasets of trees under the two models:  $\{T_i^{BD}\}_{i=1}^{N_S}$  and  $\{T_i^{MBD}\}_{i=1}^{N_S}$ . We used them to generate a dataset of alignments for each model:  $\{X_i^{BD}\}_{i=1}^{N_S}$  and  $\{X_i^{MBD}\}_{i=1}^{N_S}$ . From each dataset we produced a posterior distribution from a BD prior:  $P_i(\theta|X_i^{BD}, BD)$  and  $P_i(\theta|X_i^{MBD}, BD)$ . **[GL: 1) We might want to rename the models, e.g. BD = (0) and MBD = (1). These names with capital letters are too big and ugly; ] [RJCB: I would strongly prefer MBD and BD, as I feel replacing the big ugly capital letters by short pretty numbers hurts readability even more ]**
- To compare the results for the two models we measure the inference error

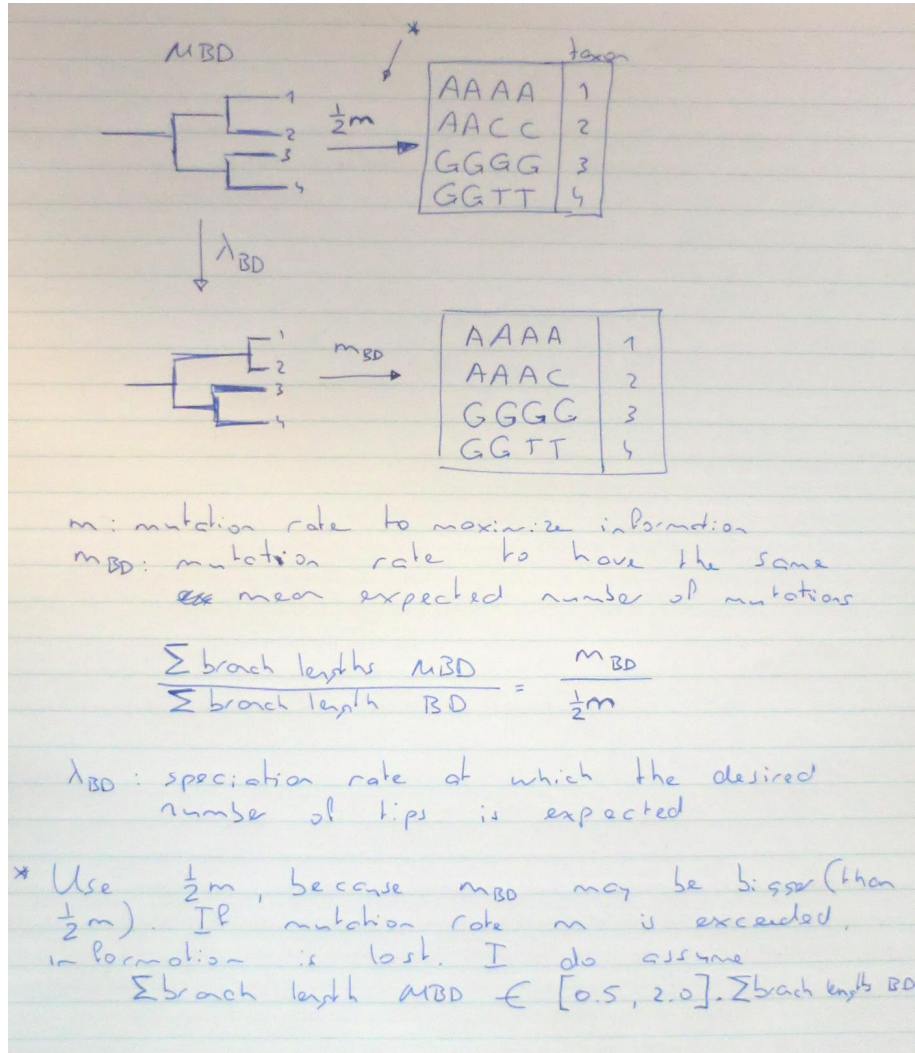


Figure 8: How to create twin trees and alignments. From a focal MBD tree, a twin tree is produced as such: (1) estimate the  $\lambda_{BD}$  to get the same expected number of tips, (2) simulate a BD tree with that amount of tips (discard trees with different number of tips), (3) estimate a mutation rate to get an alignment with the same expected number of mutations, (4) simulate alignments with that amount of mutations (discard those that don't, the picture shows an alignment that should be discarded)

using the nLTT statistic between known/true tree and posterior/inferred trees (Janzen 2015). To obtain such statistics the procedure is the following:

- From each tree  $T_{i,j}^M$  (with  $j = 1, \dots, N_S$ ) belonging to the posterior  $P_i(\theta|X_i^M, BD)$  and relative to the model  $M$ , we extrapolate the lineage-through-time (LTT), in other words we measure the number of species as a function of time  $n_{i,j}(t)$ . To allow a comparison we normalize dividing by the maximum number of species of each tree, i.e. the number of tips at the present  $N_{i,j}(t) = \frac{n_{i,j}(t)}{n_{i,j}^{max}}$ . We then define the nLTT measure as

$$nLTT_{i,j} = \int_0^T |N_{i,j}(t) - N_{T_i}| dt$$

[GL: I am running out of letters :(] [RJCB: Haha! I suggest to use the same equation and symbols as equation 1 in the nLTT article of Janzen, Hoehna and Etienne, 2015: ]

$$\Delta nLTT = \int_0^1 |nLTT_1(t) - nLTT_2(t)| dt$$

## 2.4 Model selection

We simulate alignments using the simplest nucleotide substitution model (JC69), the simplest clock model (strict). It is thus imminent to assume these models in our Bayesian inference. Nevertheless, the phylogeny the alignment was based on, could have followed either an MBD or BD tree model, where we in both cases assume a BD tree model. This will have an unknown effect on our inference: it may theoretically be that an MBD model generates (a tree that generates) an alignment in which a different site and/or clock model is favored.

We investigate this by measuring if the generative model (with the simplest

nucleotide substitution and simplest clock model) is indeed selected to be the best fitting model. To be precise, we look at the model with the highest marginal likelihood (also called evidence MacKay & Mac Kay 2003),  $f(D|M)$ , which is the probability of the data D given model M. In the context of this research, D consists of the DNA alignment, and M is the combination of site, clock and tree models.

To estimate the marginal likelihood, we use an algorithm named nested sampling Skilling *et al.* 2006. Nested sampling is attractive to use in a phylogentic context, as it gives a good estimation, requires little tuning Russel *et al.* 2018. Nested sampling is available as a BEAST2 package and can be used by babette Bilderbeek & Etienne 2018.

The nested sampling algorithm stops its run when the marginal likelihood estimation error reaches below a certain tolerance. Similar to Russel *et al.* 2018, we use a (relative) error tolerance  $\epsilon$  of  $10^{-13}$ , 1 particle to explore the parameter space and 100 active points. To achieve the latter, we use the MCMC chain length  $L_c$  of 1M (as also used in the parameter estimates), and a sub-chain length  $L_{sc}$  of 10K.

The models we use in our model comparison are the four combinations of two site models and two clock models. We use the JC69 site model, which is the (generative and) simplest model and GTR, the site model with most degrees of freedom. For the clock models, we use the strict clock model, which is the (generative and) simplest clock model, and the RLN clock model. **[RJCB: Could also just be all site models and clock models = 8 models]**

From these four marginal likelihood estimates, we calculate the weight of the generative model and plot this in figure 2. We do this for both the alignments derived from the MBD tree and the BD twin tree. We expect

239 that the generative model has the heighest weight in both the MBD and  
 240 BD alignments. We expect this weight to be higher in the BD alignments.

### 241 3 Results

- 242 • We expected the effect of  $s$  on  $e$  to be monotonously increasing function  
 243 (figure 1). Figure 9 shows the effect measured in this experiment.

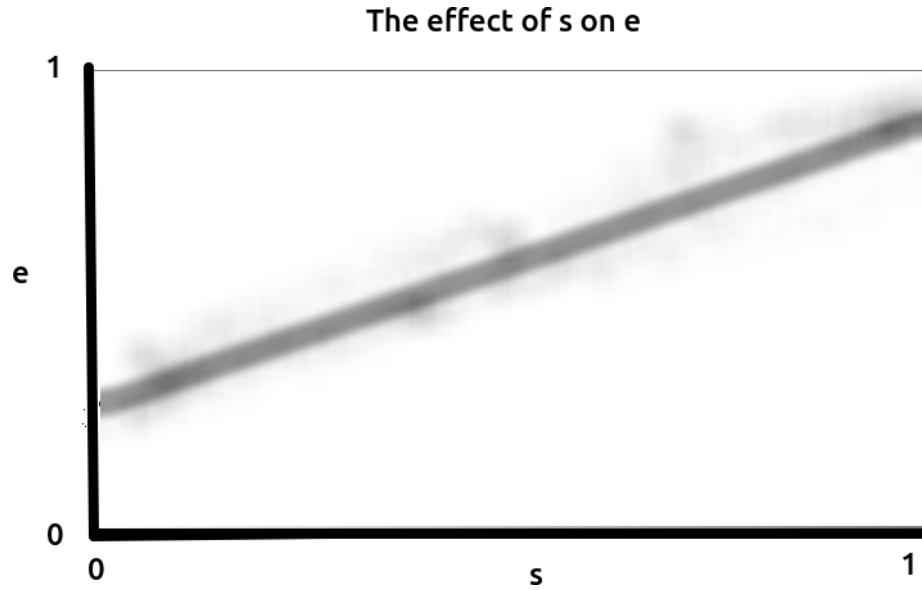


Figure 9: The effect of  $s$  on  $e$

- 244 • We expected the effect of the number of taxa on the relation between  $s$   
 245 and  $e$  to be neutral (figure 5). Figure 10 shows the effect measured in this  
 246 experiment.
- 247 • We expected the effect of the number of taxa on the variance of  $e$  for similar  
 248 values of  $s$  to decrease (figure 6). Figure 11 shows the effect measured in  
 249 this experiment.
- 250 • We expected the effect of extinction on the relation between  $s$  and  $e$  to be



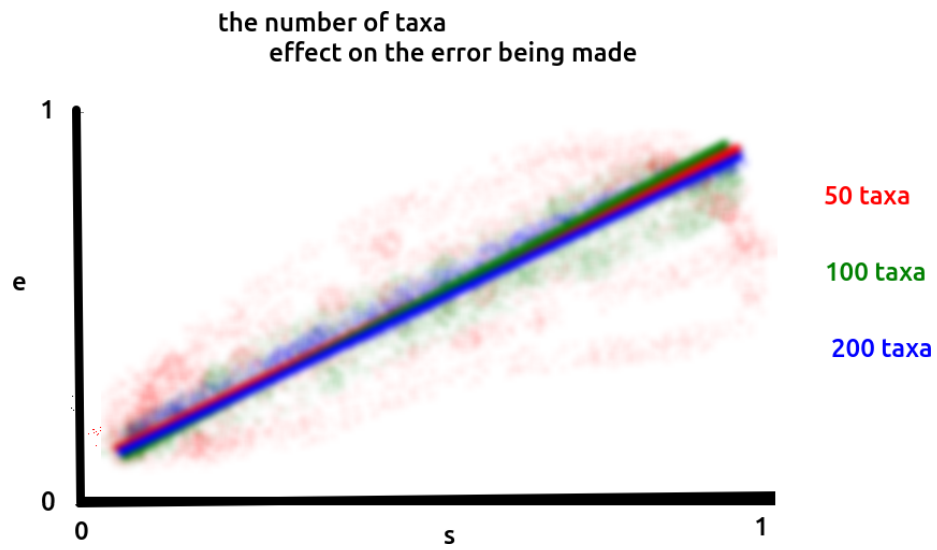


Figure 10: The effect of number of taxa on the relation between  $s$  on  $e$

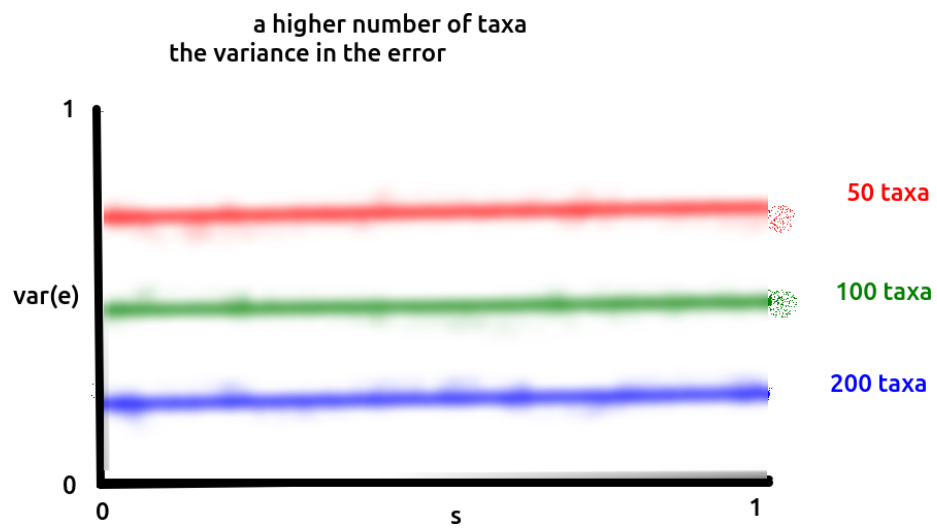


Figure 11: The effect of number of taxa on the variance in  $e$  for similar values of  $s$

neutral (figure 3). Figure 12 shows the effect measured in this experiment.

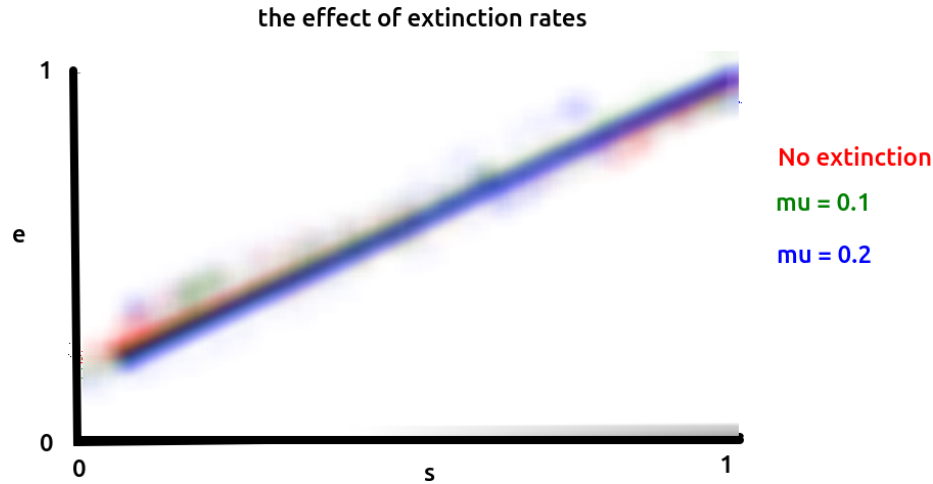


Figure 12: The effect of extinction on the relation between  $s$  on  $e$

•

## References

- Bilderbeek, R.J. (2018) *pirouette: create a posterior from a phylogeny*.
- Bilderbeek, R.J. & Etienne, R.S. (2018) babette: Beauti 2, beast 2 and tracer for r. *Methods in Ecology and Evolution*.
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A. & Drummond, A.J. (2014) Beast 2: a software platform for bayesian evolutionary analysis. *PLoS computational biology*, **10**, e1003537.
- Drummond, A.J. & Rambaut, A. (2007) Beast: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, **7**, 214.
- Etienne, R.S., Haegeman, B., Stadler, T., Aze, T., Pearson, P.N., Purvis, A. & Phillimore, A.B. (2012) Diversity-dependence brings molecular phylogenies

264 closer to agreement with the fossil record. *Proc R Soc Lond B: Biol Sci*, **279**,  
265 1300–1309.

266 Etienne, R.S., Morlon, H. & Lambert, A. (2014) Estimating the duration of  
267 speciation from phylogenies. *Evolution*, **68**, 2430–2440.

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

270 Janzen, T. (2015) *nLTT: Calculate the NLTT Statistic*. R package version 1.1.

271 Janzen, T., Alzate, A., Muschick, M., Maan, M.E., van der Plas, F. & Etienne,  
272 R.S. (2017) Community assembly in lake tanganyika cichlid fish: quantifying  
273 the contributions of both niche-based and neutral processes. *Ecology and*  
274 *Evolution*, **7**, 1057–1067.

275 Janzen, T., Alzate, A., Muschick, M., van der Plas, F. & Etienne, R.S. (2016)  
276 Stochastic processes dominate community assembly in cichlid communities in  
277 lake tanganyika.

278 Laudanno, G. (2018) *MBD: Multiple Birth Death Diversification*. R package  
279 version 0.1.

280 MacKay, D.J. & Mac Kay, D.J. (2003) *Information theory, inference and learn-*  
281 *ing algorithms*. Cambridge university press.

282 Maddison, W.P., Midford, P.E. & Otto, S.P. (2007) Estimating a binary char-  
283 acter’s effect on speciation and extinction. *Systematic Biology*, **56**, 701–710.

284 Russel, P.M., Brewer, B.J., Klaere, S. & Bouckaert, R.R. (2018) Model selection  
285 and parameter inference in phylogenetics using nested sampling. *Systematic*  
286 *Biology*, p. syy050.

- 287 Skilling, J. *et al.* (2006) Nested sampling for general bayesian computation.  
288 *Bayesian analysis*, **1**, 833–859.
- 289 Valente, L.M., Phillimore, A.B. & Etienne, R.S. (2015) Equilibrium and non-  
290 equilibrium dynamics simultaneously operate in the galápagos islands. *Ecol-*  
291 *ogy Letters*, **18**, 844–852.

Symbol	Expectation	Expected figure	Measured fig
$\mathcal{H}_1$	For a higher $s$ , $e$ will increase monotonously	1	9
$\mathcal{H}_2$	Extinction rates have no effect on $e$	3	12
$\mathcal{H}_3$	The timing of multiple-birth events has no effect on $e$	4	Not yet
$\mathcal{H}_4$	An MBD tree will have a higher $e$ than its twin BD counterpart	7	Not yet
$\mathcal{H}_5$	The number of taxa has no effect on $e$	5	10
$\mathcal{H}_6$	A higher number of taxa decreases the variance in $e$	6	11
$\mathcal{H}_7$	The MBD regime has no effect on $e$	2	Not yet

Table 1: Overview of the symbols.

	Description
$\lambda$	Speciation rate from default single-birth process
$\mu$	Extinction rate
$\nu$	Multiple-birth event trigger rate
$q$	Proportion of taxa that speciate during a multiple-birth event
$e$	Inference error
$n_{\mathbb{M}}$	Number of extant taxa created by a multiple-birth event
$n_{\mathbb{B}}$	Number of extant taxa created by a single-birth event
$N_{\mathbb{M}}$	Number of extinct and extant taxa created by a multiple-birth event
$N_{\mathbb{B}}$	Number of extinct and extant taxa created by a single-birth event
$s$	Extent/strength of the multiple birth process
$L_c$	MCMC chain length
$L_{sc}$	MCMC sub-chain length
$\epsilon$	relative error tolerance in marginal likelihood estimation

Table 2: Overview of the symbols.

	Value(s)
$s$	0.0, 0.25, 0.5, 0.75, 1.0
$\mu$	0.0, 0.1, 0.2 [TODO: can be more extreme]
$L_c$	$10^6$
$L_{sc}$	$10^4$
$\epsilon$	$10^{-13}$

Table 3: Overview of the simulation parameters.

idx	$n_{taxa}$	$f_{taxa}^{MB}$	$\mu$	MB regime	$n_\nu$ evts	$q$	$\lambda$
1	50	0	0	many modest	8	1 / 8	derive
2	50	0	0	intermediate	4	1 / 4	derive
3	50	0	0	few intense	2	1 / 2	derive
4	50	0	0.1	many modest	8	1 / 8	derive
5	50	0	0.1	intermediate	4	1 / 4	derive
6	50	0	0.1	few intense	2	1 / 2	derive
7	50	0	0.2	many modest	8	1 / 8	derive
8	50	0	0.2	intermediate	4	1 / 4	derive
9	50	0	0.2	few intense	2	1 / 2	derive
10	50	0	0.2	many modest	8	1 / 8	derive
11	50	0.25	0	intermediate	4	1 / 4	derive
12	50	0.25	0	few intense	2	1 / 2	derive
13	50	0.25	0	many modest	8	1 / 8	derive
14	50	0.25	0.1	intermediate	4	1 / 4	derive
15	50	0.25	0.1	few intense	2	1 / 2	derive
16	50	0.25	0.1	many modest	8	1 / 8	derive
17	50	0.25	0.2	intermediate	4	1 / 4	derive
18	50	0.25	0.2	few intense	2	1 / 2	derive
19	50	0.25	0.2	many modest	8	1 / 8	derive
20	50	0.25	0.2	intermediate	4	1 / 4	derive
21	50	0.5	0	few intense	2	1 / 2	derive
22	50	0.5	0	many modest	8	1 / 8	derive
23	50	0.5	0	intermediate	4	1 / 4	derive
24	50	0.5	0.1	few intense	2	1 / 2	derive
25	50	0.5	0.1	many modest	8	1 / 8	derive
26	50	0.5	0.1	intermediate	4	1 / 4	derive
27	50	0.5	0.2	few intense	2	1 / 2	derive
28	50	0.5	0.2	many modest	8	1 / 8	derive
29	50	0.5	0.2	intermediate	4	1 / 4	derive
30	50	0.5	0.2	few intense	2	1 / 2	derive
31	50	0.75	0	many modest	8	1 / 8	derive
32	50	0.75	0	intermediate	4	1 / 4	derive
33	50	0.75	0	few intense	2	1 / 2	derive
34	50	0.75	0.1	many modest	8	1 / 8	derive
35	50	0.75	0.1	intermediate	4	1 / 4	derive
36	50	0.75	0.1	few intense	2	1 / 2	derive
37	50	0.75	0.2	many modest	8	1 / 8	derive
38	50	0.75	0.2	intermediate	4	1 / 4	derive
39	50	0.75	0.2	few intense	2	1 / 2	derive
40	50	0.75	0.2	many modest	8	1 / 8	derive
41	50	1	0	intermediate	4	1 / 4	0
42	50	1	0	few intense	2	1 / 2	0
43	50	1	0	many modest	8	1 / 8	0
44	50	1	0.1	intermediate	4	1 / 4	0
45	50	1	0.1	few intense	2	1 / 2	0
46	50	1	0.1	many modest	8	1 / 8	0
47	50	1	0.2	intermediate	4	1 / 4	0
48	50	1	0.2	few intense	2	1 / 2	0
49	50	1	0.2	many modest	8	1 / 8	0
50	50	1	0.2	intermediate	4	1 / 4	0

Table 4: Overview of the MBD parameters 1/3.

idx	$n_{taxa}$	$f_{taxa}^{MB}$	$\mu$	MB regime	$n_\nu$ evts	$q$	$\lambda$
51	100	0	0	few intense	2	1 / 2	derive
52	100	0	0	many modest	8	1 / 8	derive
53	100	0	0	intermediate	4	1 / 4	derive
54	100	0	0.1	few intense	2	1 / 2	derive
55	100	0	0.1	many modest	8	1 / 8	derive
56	100	0	0.1	intermediate	4	1 / 4	derive
57	100	0	0.2	few intense	2	1 / 2	derive
58	100	0	0.2	many modest	8	1 / 8	derive
59	100	0	0.2	intermediate	4	1 / 4	derive
60	100	0	0.2	few intense	2	1 / 2	derive
61	100	0.25	0	many modest	8	1 / 8	derive
62	100	0.25	0	intermediate	4	1 / 4	derive
63	100	0.25	0	few intense	2	1 / 2	derive
64	100	0.25	0.1	many modest	8	1 / 8	derive
65	100	0.25	0.1	intermediate	4	1 / 4	derive
66	100	0.25	0.1	few intense	2	1 / 2	derive
67	100	0.25	0.2	many modest	8	1 / 8	derive
68	100	0.25	0.2	intermediate	4	1 / 4	derive
69	100	0.25	0.2	few intense	2	1 / 2	derive
70	100	0.25	0.2	many modest	8	1 / 8	derive
71	100	0.5	0	intermediate	4	1 / 4	derive
72	100	0.5	0	few intense	2	1 / 2	derive
73	100	0.5	0	many modest	8	1 / 8	derive
74	100	0.5	0.1	intermediate	4	1 / 4	derive
75	100	0.5	0.1	few intense	2	1 / 2	derive
76	100	0.5	0.1	many modest	8	1 / 8	derive
77	100	0.5	0.2	intermediate	4	1 / 4	derive
78	100	0.5	0.2	few intense	2	1 / 2	derive
79	100	0.5	0.2	many modest	8	1 / 8	derive
80	100	0.5	0.2	intermediate	4	1 / 4	derive
81	100	0.75	0	few intense	2	1 / 2	derive
82	100	0.75	0	many modest	8	1 / 8	derive
83	100	0.75	0	intermediate	4	1 / 4	derive
84	100	0.75	0.1	few intense	2	1 / 2	derive
85	100	0.75	0.1	many modest	8	1 / 8	derive
86	100	0.75	0.1	intermediate	4	1 / 4	derive
87	100	0.75	0.2	few intense	2	1 / 2	derive
88	100	0.75	0.2	many modest	8	1 / 8	derive
89	100	0.75	0.2	intermediate	4	1 / 4	derive
90	100	0.75	0.2	few intense	2	1 / 2	derive
91	100	1	0	many modest	8	1 / 8	0
92	100	1	0	intermediate	4	1 / 4	0
93	100	1	0	few intense	2	1 / 2	0
94	100	1	0.1	many modest	8	1 / 8	0
95	100	1	0.1	intermediate	4	1 / 4	0
96	100	1	0.1	few intense	2	1 / 2	0
97	100	1	0.2	many modest	8	1 / 8	0
98	100	1	0.2	intermediate	4	1 / 4	0
99	100	1	0.2	few intense	2	1 / 2	0
100	200	1	0.2	many modest	8	1 / 8	0

Table 5: Overview of the MBD parameters 2/3.

idx	$n_{taxa}$	$f_{taxa}^{MB}$	$\mu$	MB regime	$n_\nu$ evts	$q$	$\lambda$
101	200	0	0	intermediate	4	1 / 4	derive
102	200	0	0	few intense	2	1 / 2	derive
103	200	0	0	many modest	8	1 / 8	derive
104	200	0	0.1	intermediate	4	1 / 4	derive
105	200	0	0.1	few intense	2	1 / 2	derive
106	200	0	0.1	many modest	8	1 / 8	derive
107	200	0	0.2	intermediate	4	1 / 4	derive
108	200	0	0.2	few intense	2	1 / 2	derive
109	200	0	0.2	many modest	8	1 / 8	derive
110	200	0	0.2	intermediate	4	1 / 4	derive
111	200	0.25	0	few intense	2	1 / 2	derive
112	200	0.25	0	many modest	8	1 / 8	derive
113	200	0.25	0	intermediate	4	1 / 4	derive
114	200	0.25	0.1	few intense	2	1 / 2	derive
115	200	0.25	0.1	many modest	8	1 / 8	derive
116	200	0.25	0.1	intermediate	4	1 / 4	derive
117	200	0.25	0.2	few intense	2	1 / 2	derive
118	200	0.25	0.2	many modest	8	1 / 8	derive
119	200	0.25	0.2	intermediate	4	1 / 4	derive
120	200	0.25	0.2	few intense	2	1 / 2	derive
121	200	0.5	0	many modest	8	1 / 8	derive
122	200	0.5	0	intermediate	4	1 / 4	derive
123	200	0.5	0	few intense	2	1 / 2	derive
124	200	0.5	0.1	many modest	8	1 / 8	derive
125	200	0.5	0.1	intermediate	4	1 / 4	derive
126	200	0.5	0.1	few intense	2	1 / 2	derive
127	200	0.5	0.2	many modest	8	1 / 8	derive
128	200	0.5	0.2	intermediate	4	1 / 4	derive
129	200	0.5	0.2	few intense	2	1 / 2	derive
130	200	0.5	0.2	many modest	8	1 / 8	derive
131	200	0.75	0	intermediate	4	1 / 4	derive
132	200	0.75	0	few intense	2	1 / 2	derive
133	200	0.75	0	many modest	8	1 / 8	derive
134	200	0.75	0.1	intermediate	4	1 / 4	derive
135	200	0.75	0.1	few intense	2	1 / 2	derive
136	200	0.75	0.1	many modest	8	1 / 8	derive
137	200	0.75	0.2	intermediate	4	1 / 4	derive
138	200	0.75	0.2	few intense	2	1 / 2	derive
139	200	0.75	0.2	many modest	8	1 / 8	derive
140	200	0.75	0.2	intermediate	4	1 / 4	derive
141	200	1	0	few intense	2	1 / 2	0
142	200	1	0	many modest	8	1 / 8	0
143	200	1	0	intermediate	4	1 / 4	0
144	200	1	0.1	few intense	2	1 / 2	0
145	200	1	0.1	many modest	8	1 / 8	0
146	200	1	0.1	intermediate	4	1 / 4	0
147	200	1	0.2	few intense	2	1 / 2	0
148	200	1	0.2	many modest	8	1 / 8	0
149	200	1	0.2	intermediate	4	1 / 4	0
150	200	1	0.2	few intense	2	1 / 2	0

Table 6: Overview of the MBD parameters 3/3.