

1 The error in Bayesian phylogenetic reconstruction  
2 when speciation is not instantaneous

3 Richèl J.C. Bilderbeek<sup>1</sup> and Rampal S. Etienne<sup>1</sup>

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

6 June 14, 2018

7 **Abstract**

8 The tools for reconstructing phylogenetic relationships between taxo-  
9 nomic units (e.g. species) have become very advanced in the last three  
10 decades. Among the most popular tools are Bayesian approaches, such as  
11 BEAST, MrBayes and RevBayes, that use efficient tree sampling routines  
12 to create a posterior probability distribution of the phylogenetic tree. A  
13 feature of these approaches is the possibility to incorporate known or hy-  
14 pothesized structure of the phylogenetic tree through the tree prior. It  
15 has been shown that the effect of the prior on the posterior distribution  
16 of trees can be substantial.

17 Currently implemented tree priors assume that speciation is instantane-  
18 ous, where we know that speciation can be a gradual process.

19 Here we explore the effects of ignoring the protractedness of the spe-  
20 ciation process with an extensive simulation study.

21 We compare the inferred tree to the simulated tree, and find that ...

22     **Keywords:** computational biology, evolution, phylogenetics, Bayesian anal-  
23     ysis, tree prior

## 24     **1    Introduction**

25     The computational tools that are currently available to the phylogeneticists  
26     go beyond the wildest imagination of those living four decades ago. Advances  
27     in computational power allowed the first cladograms to be inferred from DNA  
28     alignments in 1981 (Felsenstein 1981), and the first Bayesian tools emerged in  
29     1996 (Rannala & Yang 1996), providing unprecedented flexibility in the setup  
30     of a phylogenetic model.

31     Currently, the most popular Bayesian phylogenetics tools are  
32     BEAST (Drummond & Rambaut 2007) and its offshoot BEAST2 (Bouckaert  
33     *et al.* 2014), MrBayes (Huelsenbeck & Ronquist 2001) and RevBayes (Höhna  
34     *et al.* 2016). They allow to incorporate known or hypothesized structure of a  
35     phylogenetic tree-to-be-inferred through model priors. With these priors and  
36     an alignment of DNA, RNA or protein sequences, they create a sample of the  
37     posterior distribution of phylogenies and parameter estimates (of the models  
38     used as a prior), in which more probable combinations are represented more  
39     often. Each of these tools use efficient tree sampling routines to rapidly create  
40     an informative posterior.

41     The model priors in Bayesian phylogenetic reconstruction can be grouped  
42     into three categories: (1) site model, specifying nucleotide substitutions, (2)  
43     clock model, specifying the rate of mutation per lineage in time, and (3) tree  
44     model, constituting the speciation model underlying branching events (specia-  
45     tion) and branch termination (extinction). The choice of site model (Posada &  
46     Buckley 2004), clock model (Baele *et al.* 2012) or tree prior (Möller *et al.* 2018;  
47     Yang & Rannala 2005) is known to affect the posterior.

48 Current phylogenetic tools use tree priors that assume speciation is instan-  
49 taneous, whilst we know that, speciation is often a gradual process (Schluter  
50 2009). The (constant-rate) birth-death (BD) model is a commonly used tree  
51 prior, but it ignores this temporal aspect of speciation. The protracted birth-  
52 death (PBD) model, an extension of the BD model, does incorporate the idea  
53 that speciation takes time. In this model, a branching event does not give rise  
54 to a new species, but to a new species-to-be, called an incipient species. Such an  
55 incipient species may go extinct, finish its speciation to become a good species,  
56 or give rise to new incipient species. Protracted speciation may explain observed  
57 declines in lineage accumulation (Etienne & Rosindell 2012).

58 Unfortunately, a tree prior according to this model, providing the probability  
59 of a species tree under the PBD model, is unavailable in current Bayesian phy-  
60 logenetic tools. Whilst an approximate formula for this probability has been  
61 derived (Lambert *et al.* 2015) and the approximation is very good (Simonet  
62 *et al.* 2018), it has not been implemented as tree prior yet. There are various  
63 reasons for this. First, the computation of this probability involves solving a set  
64 of non-linear differential equations, and while this computation is quite fast, it  
65 still takes much more time than the corresponding probability of the BD model  
66 which is a simple analytical formula. In a Bayesian MCMC chain, the tree  
67 prior probability must be calculated many times, and hence the total compu-  
68 tation will take considerably longer with a PBD tree prior. Furthermore, the  
69 approximate probability is a probability for the species tree assuming an under-  
70 lying incipient species tree. It can be safely used as tree prior when only one  
71 individual per species is sampled, but if one has multiple samples per species  
72 - which is currently often the case - the methods to account for this such as  
73 the multi-species coalescent (Heled & Drummond 2009) may not be compatible  
74 with the underlying incipient species tree. More precisely, the phylogeny under

75 the PBD model may contain paraphylies, while the multi-species coalescent was  
 76 developed exactly to avoid these by explaining them as arising from incomplete  
 77 lineage sorting. Because of these paraphylies there is no such thing as a true  
 78 species tree in the PBD model. To get a species-level tree one must sample one  
 79 incipient species per species. Which incipient species is sampled may therefore  
 80 have an impact on the species tree.

81 Here we aim to explore the effect of using the BD prior on PBD simulated  
 82 phylogenies, taking into account possible sampling effects. In brief, we simulate  
 83 protracted phylogenies using the PBD process, from which we sample a species  
 84 tree in two very different ways. Given this species tree, we simulate a DNA  
 85 sequence alignment. Then, we use BEAST2 on these alignments to infer a pos-  
 86 terior of phylogenies, using a BD prior. We quantify the difference between the  
 87 (BD) posterior phylogenies and the simulated (PBD) species tree. Furthermore,  
 88 while we evidently know the clock and site models used in the simulation, us-  
 89 ing a different clock and/or site model prior in inference may compensate or  
 90 increase this difference between inferred and simulated tree. To study this, we  
 91 also explore the effect of a different clock and site model prior in inference.

92 The PBD model has five biological parameters, depicted in table 2, which we  
 93 explore in a factorial fashion, excluding - for computational reasons - the combi-  
 94 nations in which the 95% quantile of the expected number of good species is more  
 95 than 1250. This quantile is calculated with the `pbm_numspec_quantile` function  
 96 we added to the PBD package (Etienne 2015). [RSE: I think we should re-  
 97 lease the new version with this function with this manuscript][RJCB:  
 98 I would personally prefer 'release early, release often', but as lead  
 99 maintainer you get to decide][RSE: Do you want to put the deriva-  
 100 tion here?][RJCB: Yes please, either here or at the PBD package  
 101 (the functions `pbm_geom` and `pbm_numspec_quantile`) documentation.

102 **Wherever you put it, I will transfer it to the other place].** This cal-  
 103 culation assumes  $b = b_g = b_i$ , we used  $b = \max(b_g, b_i)$ . We use 1000 good  
 104 species as a threshold, to prevent overly taxon-poor and taxon-rich phylogenies  
 105 respectively. The parameter values chosen are based on the parameter sets used  
 106 by Etienne *et al.* 2014, as these parameters were shown to result in reasonably  
 107 sized phylogenies and using the same set allows us to compare results. We use  
 108 a set of speciation initiation rates,  $B = \{0.3, 0.5\}$ , of which the speciation ini-  
 109 tiation rate of good species  $b_g \in B$  and incipient species  $b_i \in B$ . [RSE: your  
 110 units are not entirely correct, as these are probability rates.][RJC: I  
 111 though I was correct in my units. If I am wrong, what are the correct  
 112 units then? In Etienne and Rosindell 2012, there is never 'probabil-  
 113 ity rate' written. To be explicit, I think these sentences are all valid:  
 114 (1) the extinction rate of good species,  $\mu_g$  is 1.2 extinction events  
 115 per time unit (2) the speciation initiation rate of good species,  $b_g$   
 116 is 2.3 speciation events per time unit, (3) the speciation completion  
 117 rate,  $\lambda$  is 3.4 speciation completion events per time unit. What  
 118 would be the better unit?][RSE: Je hebt het over de eenheden van  
 119 de rates. Dat is subtiel omdat het om een stochastisch model gaat  
 120 en dus om kansen. Voorbeeld:  $\mu * dt$  is de kans op een extinction  
 121 event per species in een tijdsinterval  $dt$ .  $1/\mu$  geeft de verwachte  
 122 tijd tot een extinctie event van een soort.  $\mu$  is dus een probability  
 123 rate. ] The speciation completion rates we use are  $\lambda = 0.1, 0.3, 1.0$  and  $10^9$ .  
 124 We use  $10^9 \approx \infty$  to mimic the BD model, because the PBD model reduces to  
 125 the BD model for  $\lambda = \infty$ . This allows us to measure the baseline error, which  
 126 is the difference between inferred tree and true species tree that arises purely  
 127 due to noise because the generating model and the model used in inference are  
 128 identical in this case. We use a set of extinction rates,  $M = \{0.0, 0.1, 0.2, \infty\}$ , of

129 which the extinction rate of good species  $\mu_g \in M$  and incipient species  $\mu_i \in M$ .

130 From each biological parameter set, we simulate a protracted birth-death  
131 tree, using the PBD package (Etienne 2015) in the R programming language (R  
132 Core Team 2013), all with a crown age of 15 million years as used in Etienne  
133 *et al.* 2014. Each protracted birth-death tree uses a different random number  
134 generator seed, which makes all runs independent, resulting in a balanced data  
135 set.

136 From each incipient species tree, we construct a species tree, by sampling one  
137 incipient/good species per good species. For example, when an incipient species  
138 branched off from its mother lineage, both of these subspecies are recognized  
139 as representing the species, and hence both can be picked as an (equally good)  
140 representative of the species. Here, we use three sampling scenarios, in which  
141 we pick the representative randomly or in such a way that this results in either  
142 the shortest or longest branch lengths. See the supplementary information for  
143 a visualization of these sampling methods.[RSE: Why not include that in  
144 the paper here; I think it is essential][RJCB: Because we are limited  
145 to two figures]

146 Based on the sampled species tree, we simulate a DNA alignment that has  
147 the same history as this species tree, using the **phangorn** package (Schliep 2011).  
148 We set the nucleotides of the DNA alignment to follow a Jukes-Cantor (Jukes  
149 *et al.* 1969) nucleotide substitution model, in which all nucleotide-to-nucleotide  
150 transitions are equally likely. In our Bayesian inference (see below) we use the  
151 same site model as the (obviously correct) site model prior, but we also explore  
152 the effect of assuming a more complex site model prior. We predict with the  
153 more complex substitution model, that there will be more noise and hence our  
154 inference error will increase. On the other hand, we dare not rule out that  
155 the inference error will decrease, due to more flexibility in the more complex

156 prior. We set the mutation rate in such a way to maximize the information  
 157 contained in the alignment. To do so, we set the mutation rate such that we  
 158 expect on average one (possibly silent) mutation per nucleotide between crown  
 159 age and present, which equates to  $\frac{1}{15}$  mutations per million years. The DNA  
 160 sequence length is chosen to provide a resolution of  $10^3$  years, that is, to have  
 161 one expected nucleotide change per  $10^3$  years per lineage on average. As one  
 162 nucleotide is expected to have on average one (possibly silent) mutation per 15  
 163 million years,  $15 \cdot 10^3$  nucleotides result in 1 mutation per alignment per  $10^3$   
 164 years (which is coincidentally the same as Möller *et al.* 2018). The simulation  
 165 of these DNA alignments follows a strict clock model, which we will specify as  
 166 one of the two clock models assumed in the Bayesian inference (see below).

167 From an alignment, we run a Bayesian analysis and create a posterior dis-  
 168 tribution of trees and parameters using the **babette** (Bilderbeek & Etienne  
 169 2018) package that sets the input parameters similar to BEAUti 2 and then  
 170 runs BEAST2. For our site model, we assume either a Jukes-Cantor or GTR  
 171 nucleotide substitution model. The Jukes-Cantor model is the correct one, as it  
 172 is used for simulating that alignment, where the GTR model is the site model  
 173 that is picked as a default by most users. For our clock model, we assume either  
 174 a strict or relaxed log-normal clock model. Also here, the strict clock model  
 175 is the correct one, as it is used for simulating the alignment, but the relaxed  
 176 log-normal clock model is the one most commonly used. We set the BD model  
 177 as a tree prior, as gauging the effect of this incorrect assumption is the goal of  
 178 this study. We assume an MRCA prior with a tight normal distribution around  
 179 the crown age, by choosing the crown age as mean, and a standard deviation of  
 180  $0.5 \cdot 10^{-3}$  time units, resulting in 95% of the crown ages inferred have the same  
 181 resolution (of  $10^{-3}$  time units) as the alignment. We ran the MCMC chain to  
 182 generate 1111 states, of which we remove the first 10% (also called the 'burn-

in'). Of the remaining 1000 MCMC states, the effective sample size (ESS) of the posterior must at least be 200 for a strong enough inference (Drummond & Bouckaert 2015). An ESS can be increased by increasing the number of samples or decreasing the autocorrelation between samples. If the ESS is less than 200, we decrease autocorrelation by doubling the MCMC sampling interval of that simulation, until the ESS exceeds 200.

We compare each posterior phylogeny to the (sampled) species tree using the nLTT statistic (Janzen *et al.* 2015), from the nLTT package (Janzen 2015). The nLTT statistic equals the area between the normalized lineages-through-time plots of two phylogenies, which has a range from zero (for identical phylogenies) to one. We use inference error and nLTT statistic interchangeably. Comparing the simulated species tree with each of the posterior species trees yields a distribution of nLTT statistics.

We produce two data sets as a comma-separated file. The general data set has 348 different combinations of biological parameter combinations, site and clock models. The data set to investigate sampling has 496 different combinations of biological parameter combinations, site models, clock models and sampling methods. The experiment is computationally intensive: pilot experiments show that the experiment takes roughly 100 days of CPU time and 20 days of wall clock time (which includes the queued waiting for computational resources) per replicate. Due to this, we choose to perform ten replicates, so that the complete experiment will take an acceptable time of roughly seven months.

For both data sets, we display the nLTT statistics distribution per biological parameter combination as a violin plot. We show combinations for which  $b_g = b_i$  and  $\mu_g = \mu_i$ , to simplify the interpretation of the results, where the other combinations are shown in the supplementary material. Additionally, we only show the nLTT distributions that were generated under the (correct) as-



Term	Definition
Phylogenetics	The inference of evolutionary relationships of groups of organisms using genetics
Model prior	Knowledge or assumptions about the ontogeny of evolutionary histories
Posterior	A collection of phylogenies and parameter estimates, in which more probable combinations (determined by the data and the model prior) are presented more frequently
Protracted speciation	The process in which speciation takes two events: a speciation-initiation event and a speciation-completion event
Speciation initiation	The start of a speciation event creating an incipient species
Speciation completion	The end of a speciation event, in which an incipient species becomes or is recognized as a good species

Table 1: Glossary

sumptions of a Jukes-Cantor site model and a strict clock model, separated per  
sampling method used. We display the nLTT statistic distributions separated  
per site or clock model in the supplementary information.

## 2 Results

## 3 Glossary

## References

- Baele, G., Li, W.L.S., Drummond, A.J., Suchard, M.A. & Lemey, P. (2012) Accurate model selection of relaxed molecular clocks in bayesian phylogenetics. *Molecular biology and evolution*, **30**, 239–243.
- Bilderbeek, R.J. & Etienne, R.S. (2018) babette: Beauti 2, beast2 and tracer for r. *bioRxiv*, p. 271866.

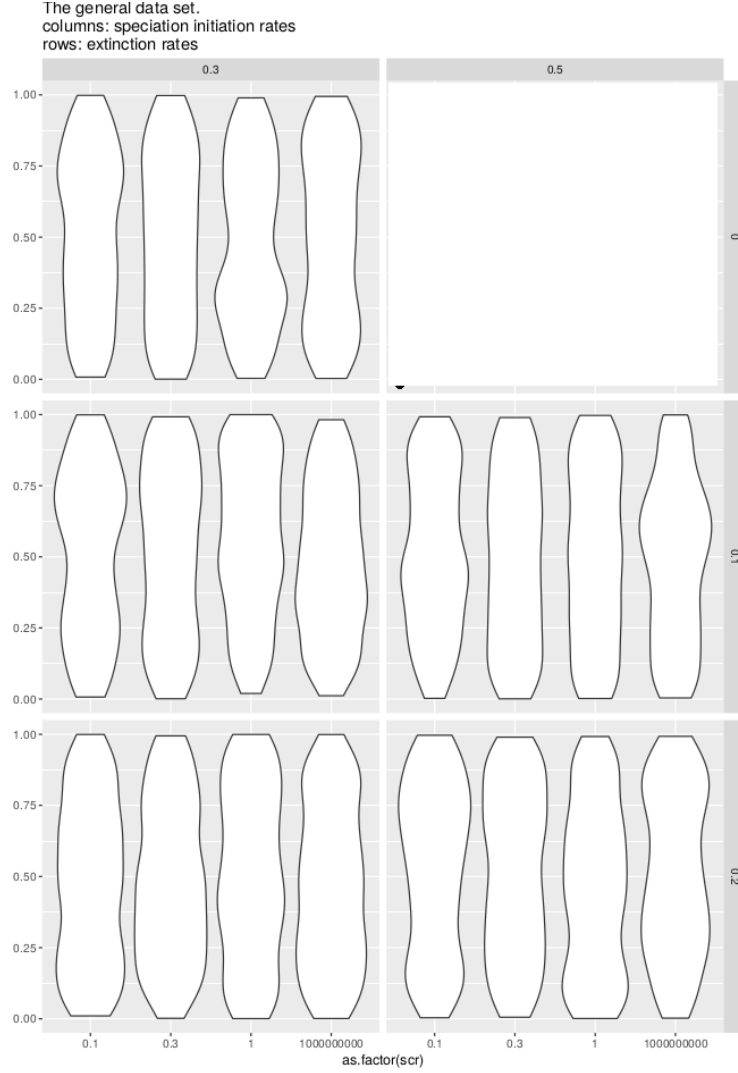


Figure 1: nLTT statistic distribution per biological parameter set, using the general data set, for the subset of combinations in which  $b_g = b_i$ ,  $\mu_g = \mu_i$ , under the (correct) assumptions of a strict clock and Jukes-Cantor site model.

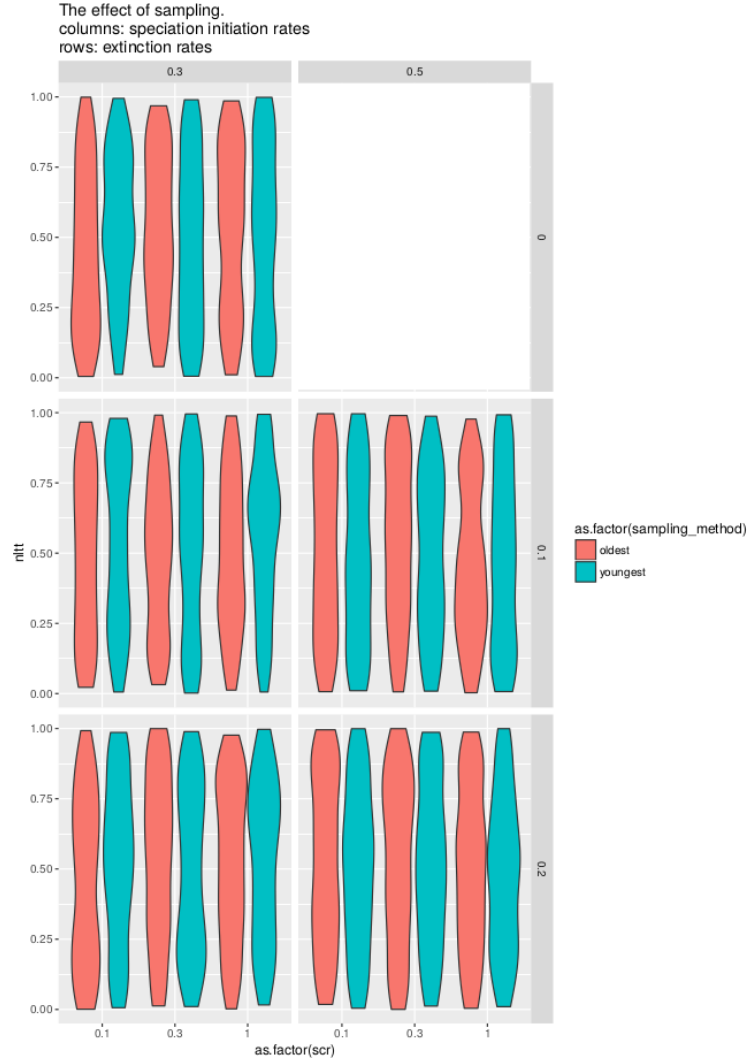


Figure 2: nLTT statistic distribution per biological parameter set per sampling regime, using the data set conditioned on sampling regime having an effect, for the subset of combinations in which  $b_g = b_i$ ,  $\mu_g = \mu_i$ , under the (correct) assumptions of a strict clock and Jukes-Cantor site model.

221 Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard,  
222 M.A., Rambaut, A. & Drummond, A.J. (2014) Beast 2: a software platform  
223 for bayesian evolutionary analysis. *PLoS computational biology*, **10**, e1003537.

224 Drummond, A.J. & Bouckaert, R.R. (2015) *Bayesian evolutionary analysis with*  
225 *BEAST*. Cambridge University Press.

226 Drummond, A.J. & Rambaut, A. (2007) Beast: Bayesian evolutionary analysis  
227 by sampling trees. *BMC evolutionary biology*, **7**, 214.

228 Etienne, R.S. (2015) *PBD: Protracted Birth-Death Model of Diversification*. R  
229 package version 1.1.

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

232 Etienne, R.S. & Rosindell, J. (2012) Prolonging the past counteracts the pull of  
233 the present: protracted speciation can explain observed slowdowns in diver-  
234 sification. *Systematic Biology*, **61**, 204–213.

235 Felsenstein, J. (1981) Evolutionary trees from dna sequences: a maximum like-  
236 lihood approach. *Journal of molecular evolution*, **17**, 368–376.

237 Heled, J. & Drummond, A.J. (2009) Bayesian inference of species trees from  
238 multilocus data. *Molecular biology and evolution*, **27**, 570–580.

239 Höhna, S., Landis, M.J., Heath, T.A., Boussau, B., Lartillot, N., Moore, B.R.,  
240 Huelsenbeck, J.P. & Ronquist, F. (2016) Revbayes: Bayesian phylogenetic  
241 inference using graphical models and an interactive model-specification lan-  
242 guage. *Systematic biology*, **65**, 726–736.

243 Huelsenbeck, J.P. & Ronquist, F. (2001) Mrbayes: Bayesian inference of phylo-  
244 genetic trees. *Bioinformatics*, **17**, 754–755.

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

246 Janzen, T., Höhna, S. & Etienne, R.S. (2015) Approximate bayesian compu-  
 247 tation of diversification rates from molecular phylogenies: introducing a new  
 248 efficient summary statistic, the nltt. *Methods in Ecology and Evolution*, **6**,  
 249 566–575.

250 Jukes, T.H., Cantor, C.R. *et al.* (1969) Evolution of protein molecules. *Mam-*  
 251 *malian protein metabolism*, **3**, 132.

252 Lambert, A., Morlon, H. & Etienne, R.S. (2015) The reconstructed tree in  
 253 the lineage-based model of protracted speciation. *Journal of mathematical*  
 254 *biology*, **70**, 367–397.

255 Möller, S., du Plessis, L. & Stadler, T. (2018) Impact of the tree prior on  
 256 estimating clock rates during epidemic outbreaks. *Proceedings of the National*  
 257 *Academy of Sciences*, p. 201713314.

258 Posada, D. & Buckley, T.R. (2004) Model selection and model averaging in  
 259 phylogenetics: advantages of akaike information criterion and bayesian ap-  
 260 proaches over likelihood ratio tests. *Systematic biology*, **53**, 793–808.

261 R Core Team (2013) *R: A Language and Environment for Statistical Computing*.  
 262 R Foundation for Statistical Computing, Vienna, Austria.

263 Rannala, B. & Yang, Z. (1996) Probability distribution of molecular evolution-  
 264 ary trees: a new method of phylogenetic inference. *Journal of molecular*  
 265 *evolution*, **43**, 304–311.

266 Schliep, K. (2011) phangorn: phylogenetic analysis in r. *Bioinformatics*, **27**,  
 267 592–593.

268 Schluter, D. (2009) Evidence for ecological speciation and its alternative. *Sci-*  
 269 *ence*, **323**, 737–741.

- 270 Simonet, C., Scherrer, R., Rego-Costa, A. & Etienne, R. (2018) Robustness of  
 271 the approximate likelihood of the protracted speciation model. *Journal of*  
 272 *evolutionary biology*, **31**, 469–479.
- 273 Tavaré, S. (1986) Some probabilistic and statistical problems in the analysis of  
 274 dna sequences. *Lectures on mathematics in the life sciences*, **17**, 57–86.
- 275 Yang, Z. & Rannala, B. (2005) Branch-length prior influences bayesian posterior  
 276 probability of phylogeny. *Systematic Biology*, **54**, 455–470.

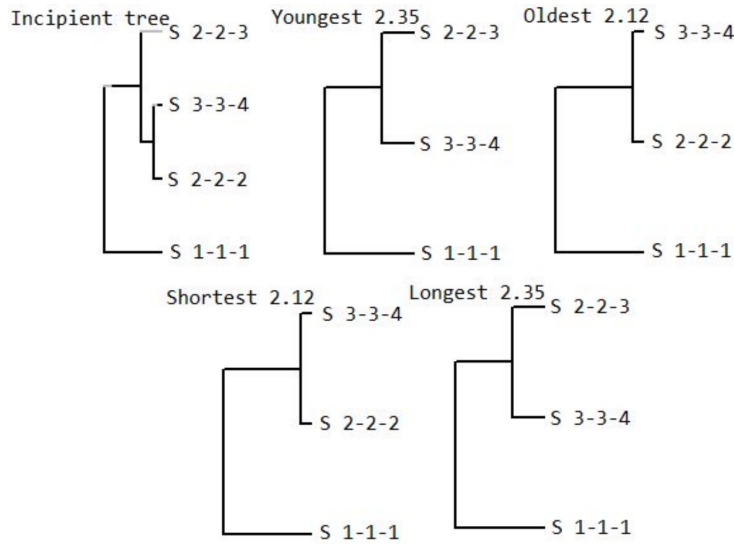


Figure 3: Sampling a species tree from an incipient species tree. At the top left, an incipient species tree is shown, of three different good species (the first and second number in the taxon label) and four different subspecies (the third number in the taxon label). The other four trees are species trees, that use a different sampling method to determine which sub-species is picked to represent a good species. These are: 'Youngest', 'Oldest', 'Shortest' and 'Longest'. With 'Youngest' the youngest sub-species is picked to represent the good species. With 'Oldest' the oldest sub-species is picked to represent the good species. 'Shortest' is the sampling method in which the sub-species are picked to assure the shortest branch lengths. 'Longest' is the sampling method in which the sub-species are picked to assure the longest branch lengths.

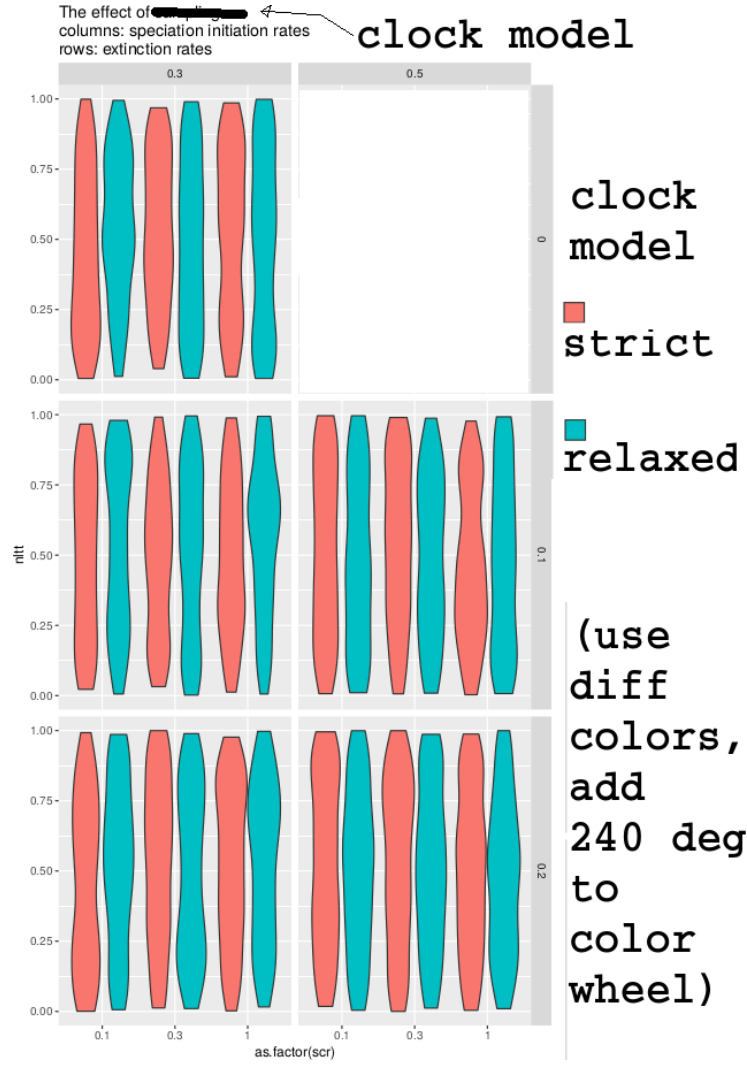


Figure 4: nLTT statistic distribution per biological parameter set per clock model, using the general data set, for the subset of combinations in which  $b_g = b_i$ ,  $\mu_g = \mu_i$ , under the (correct) assumption of a Jukes-Cantor site model.

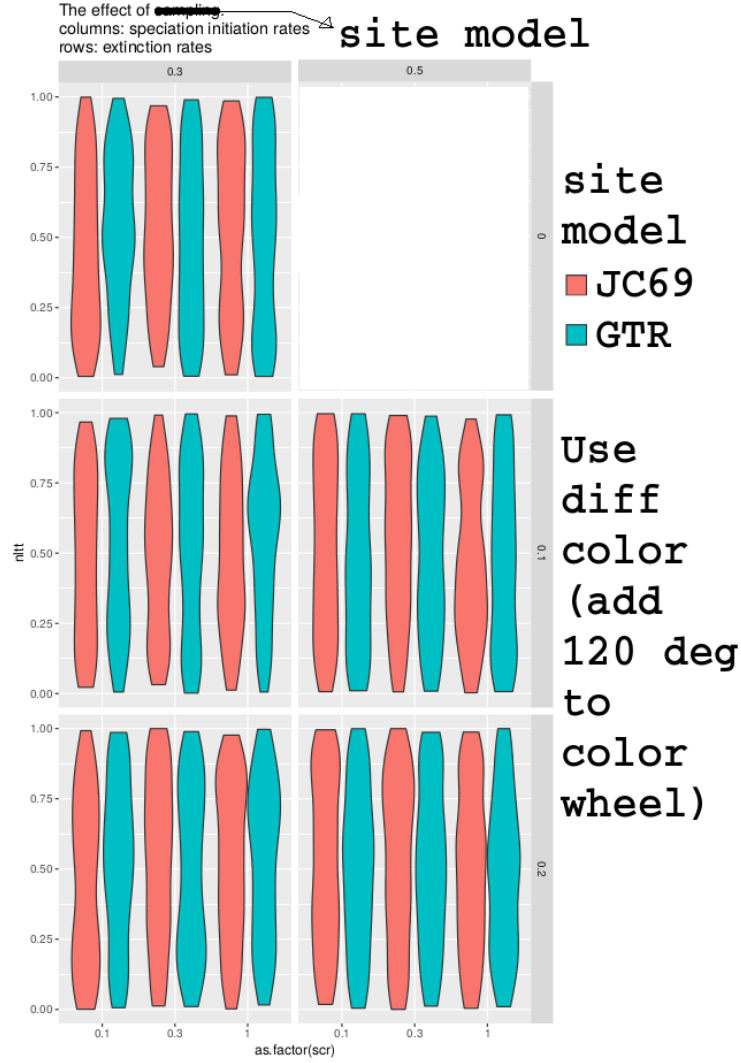


Figure 5: nLTT statistic distribution per biological parameter set per site model, using the general data set, for the subset of combinations in which  $b_g = b_i$ ,  $\mu_g = \mu_i$ , under the (correct) assumption of a strict clock model.



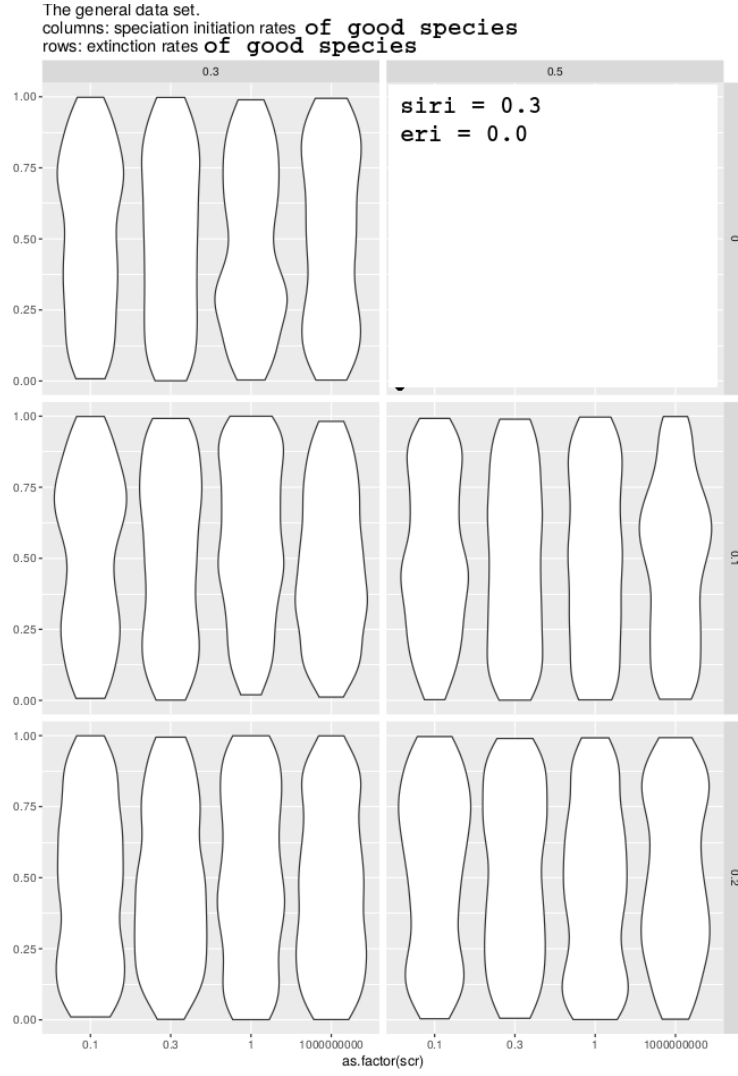


Figure 6: nLTT statistic distribution per biological parameter set, using the general data set, for  $b_i = 0.3$  and  $\mu_i = 0.0$ , under the (correct) assumptions of a strict clock and Jukes-Cantor site model.

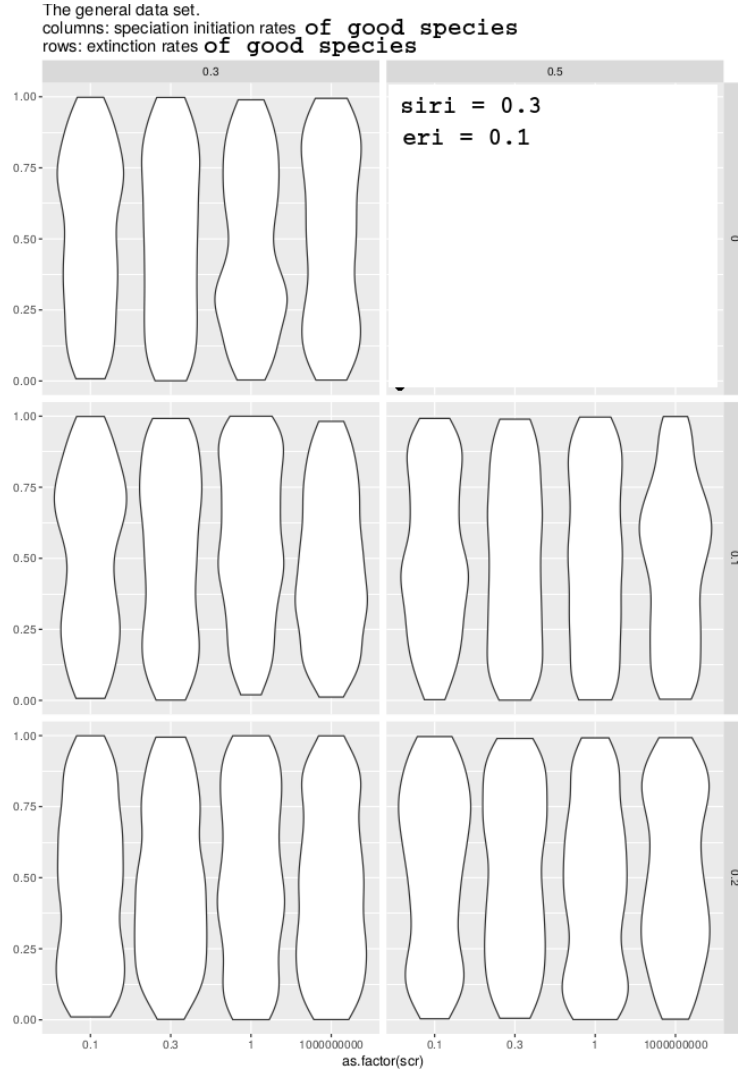


Figure 7: nLTT statistic distribution per biological parameter set, using the general data set, for  $b_i = 0.3$  and  $\mu_i = 0.1$ , under the (correct) assumptions of a strict clock and Jukes-Cantor site model.

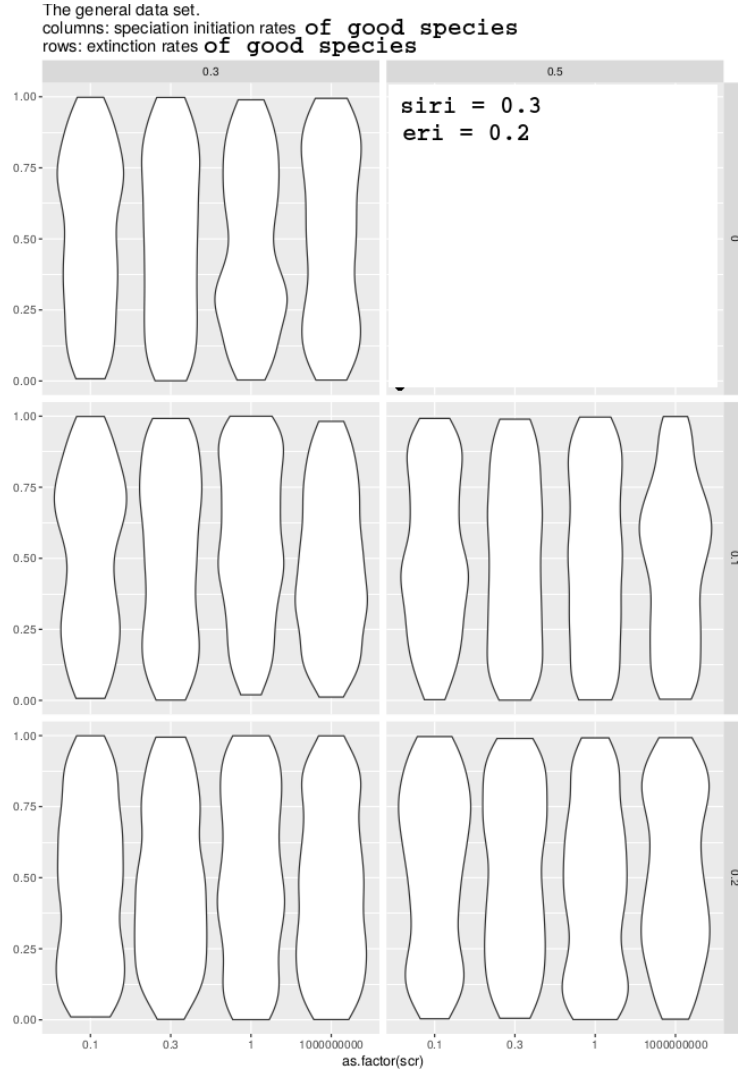


Figure 8: nLTT statistic distribution per biological parameter set, using the general data set, for  $b_i = 0.3$  and  $\mu_i = 0.2$ , under the (correct) assumptions of a strict clock and Jukes-Cantor site model.

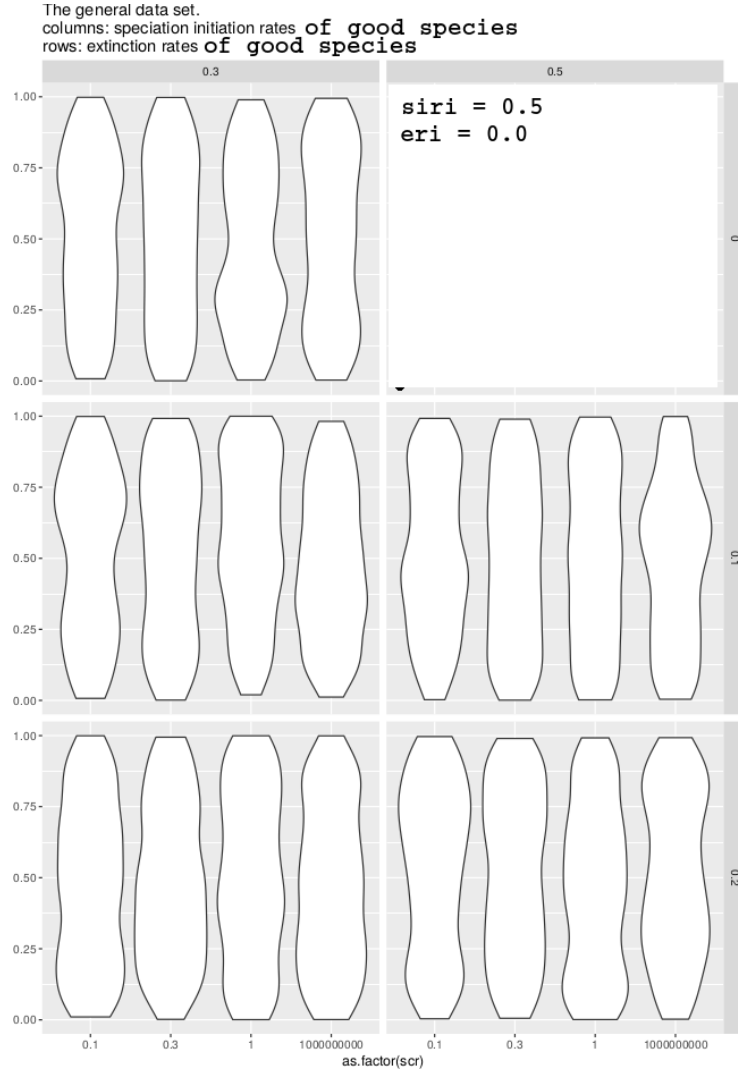


Figure 9: nLTT statistic distribution per biological parameter set, using the general data set, for  $b_i = 0.5$  and  $\mu_i = 0.0$ , under the (correct) assumptions of a strict clock and Jukes-Cantor site model.

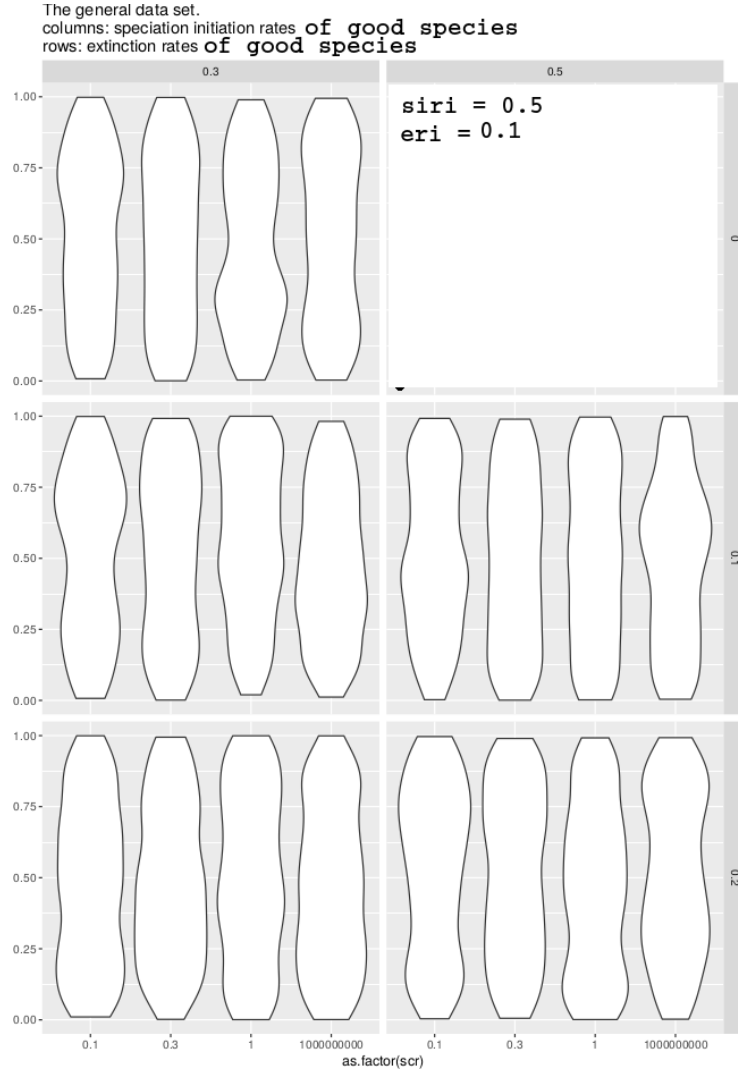


Figure 10: nLTT statistic distribution per biological parameter set, using the general data set, for  $b_i = 0.5$  and  $\mu_i = 0.1$ , under the (correct) assumptions of a strict clock and Jukes-Cantor site model.

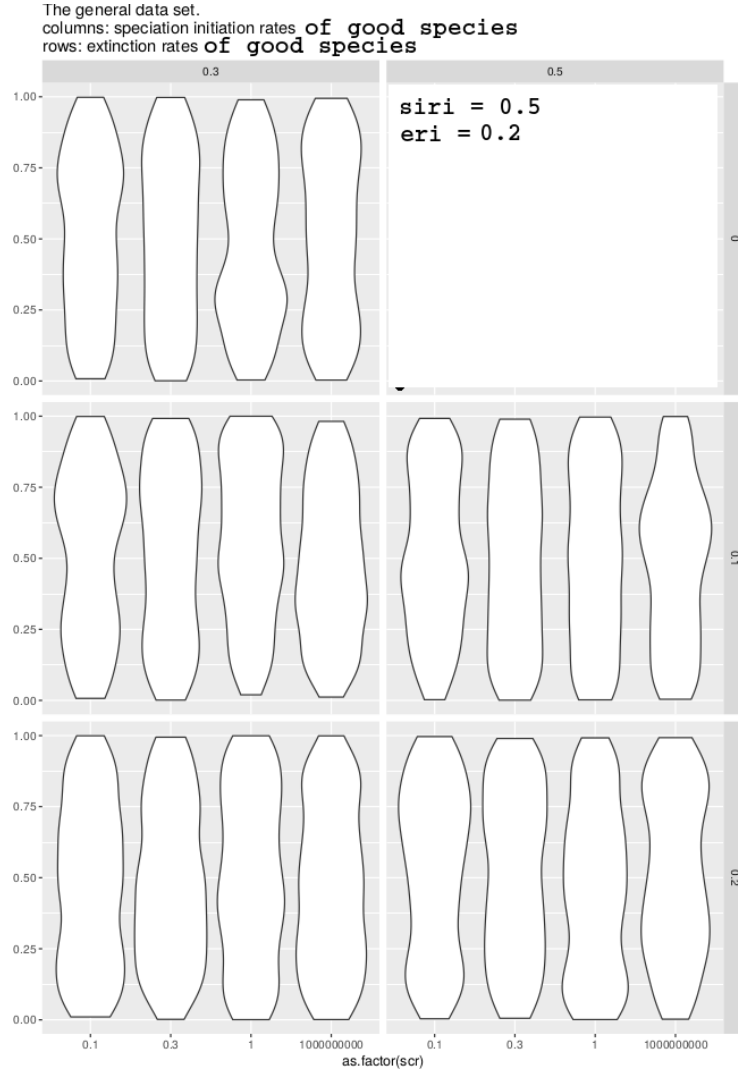


Figure 11: nLTT statistic distribution per biological parameter set, using the general data set, for  $b_i = 0.5$  and  $\mu_i = 0.2$ , under the (correct) assumptions of a strict clock and Jukes-Cantor site model.

	Description	Values
$b_g$	Speciation initiation rate of a good species	0.3, 0.5
$b_i$	Speciation initiation rate of an incipient species	0.3, 0.5
$\lambda$	Speciation completion rate	0.1, 0.3, 1.0, $\infty$
$\mu_g$	Extinction rate of a good species	0.0, 0.1, 0.2
$\mu_i$	Extinction rate of an incipient species	0.0, 0.1, 0.2
$t_c$	Crown age	15
$\sigma_c$	Standard deviation around crown age	0.001
$M_s$	Sampling method	S, L, R
$M_c$	Clock model	S, RLN
$M_t$	Site model	JC69, GTR
$r$	Mutation rate	$\frac{1}{15}$
$l_a$	DNA alignment length	15K
$f_i$	MCMC sampling interval	1K or more
$R_i$	RNG seed incipient tree and randomly sampled species tree	1, 2, ...
$R_a$	RNG seed alignment simulation	$R_i$
$R_b$	RNG seed BEAST2	$R_i$

Table 2: Overview of the simulation parameters. Above the horizontal line is the biological parameter set. The RNG seed  $R_i$  is 1 for the first simulation of the general data set, 2 for the next, and so on, up to and including 3480. The RNG seeds for the data set investigating the effect of sampling continue from there, but only those RNG seeds are used in which sampling has an effect. The sampling methods are abbreviated as such: 'R' denotes random sampling, 'S' is 'shortest' and 'L' is 'longest'. Sampling method  $M_s$  is random for the general data set. For the data set exploring the effect of sampling, we use 'shortest' and 'longest' for each value of  $R_i$  (which are random seeds in which sampling has an effect). The clock models are abbreviated as 'S' for a strict and 'RLN' for a relaxed log-normal model. The site models are abbreviated as 'JC69' for Jukes-Cantor (Jukes *et al.* 1969) and 'GTR' for the generalized time-reversible model (Tavaré 1986).

$n$	Description
12	simulation parameters, see table 2
1000	nLTT statistic values
11	ESSes of all parameters estimated by BEAST2 (see specs below)

Table 3: Specification of the data sets. Each row will contain one experiment, where the columns contain parameters, measurements and diagnostics. This table displays the content of the columns.  $n$  denotes the number of columns a certain item will occupy, resulting in a table of 1023 columns and 20K rows.

#	Description
1	posterior
2	likelihood
3	prior
4	treeLikelihood
5	TreeHeight
6	BirthDeath
7	BDBirthRate
8	BDDeathRate
9	logP.mrca
10	mrcatime
11	clockRate

Table 4: Overview of the 11 parameters estimated by BEAST2