

¹ Quantifying the importance of an inference model
² in Bayesian phylogenetics

³ Richèl J.C. Bilderbeek^{1,*}, Giovanni Laudanno¹, and Rampal S.
⁴ Etienne¹

⁵ ¹Groningen Institute for Evolutionary Life Sciences, University of
⁶ Groningen, Groningen, The Netherlands

⁷ *corresponding author: r.j.c.bilderbeek@rug.nl

⁸ May 27, 2020

9

Summary

10 1. Phylogenetic trees are currently routinely reconstructed from an
11 alignment of character sequences (usually nucleotide sequences). Bayesian
12 tools, such as MrBayes, RevBayes and BEAST2, have gained much popu-
13 larity over the last decade, as they allow joint estimation of the posterior
14 distribution of the phylogenetic trees and the parameters of the underlying
15 inference model. An important ingredient of these Bayesian approaches
16 is the species tree prior. In principle, the Bayesian framework allows
17 for comparing different tree priors, which may elucidate the macroevolu-
18 tionary processes underlying the species tree. In practice, however, only
19 macroevolutionary models that allow for fast computation of the prior
20 probability are used. The question is how accurate the tree estimation
21 is when the real macroevolutionary processes are substantially different
22 from those assumed in the tree prior.

23 2. Here we present **pirouette**, a free, libre and open-source R package
24 that assesses the inference error made by Bayesian phylogenetics for a
25 given macroevolutionary diversification model. **pirouette** makes use of
26 BEAST2, but its philosophy applies to any Bayesian phylogenetic infer-
27 ence tool.

28 3. We describe **pirouette**'s usage and the biological scientific question
29 it can answer, including full examples.

30 4. Last, we discuss the results obtained by the examples and their in-
31 terpretation.

32

33 **Keywords:** Bayesian model selection, BEAST2, computational biology,
34 evolution, phylogenetics, R, tree prior

35 **1 Introduction**

36 The development of new powerful Bayesian phylogenetic inference tools, such
37 as BEAST [Drummond & Rambaut 2007], MrBayes [Huelsenbeck & Ronquist
38 2001] or RevBayes [Höhna *et al.* 2016a] has been a major advance in constructing
39 phylogenetic trees from character data (usually nucleotide sequences) extracted
40 from extant (but also extinct, where possible) organisms, and hence in our
41 understanding of the main drivers and modes of diversification.

42 BEAST [Drummond & Rambaut 2007] is a typical Bayesian phylogenetics
43 tool, that needs both character data and priors to infer a posterior distribution of
44 phylogenies. Specifically, for the species tree prior - which describes the process
45 of diversification - BEAST has built-in priors such as the Yule [Yule 1925] and
46 (constant-rate) birth-death [Nee *et al.* 1994] models. These simple tree priors are
47 the most commonly used, as they represent some biologically realistic processes
48 (i.e. branching at per species constant rates), while being computationally fast.
49 BEAST's successor, BEAST2 [Bouckaert *et al.* 2019], has a package manager,
50 that allows third-party users to extend existing functionalities. For example,
51 one can add novel diversification models by writing a BEAST2 package that
52 contains the likelihood formula of a phylogeny under the novel diversification
53 model, i.e. the prior probability of a species tree. Many diversification models
54 (and their associated probability algorithms) have been developed, e.g., models
55 in which diversification is time-dependent [Nee *et al.* 1994; Rabosky & Lovette
56 2008], or diversity-dependent [Etienne *et al.* 2012], or where diversification rates
57 change for specific lineages and their descendants [Etienne & Haegeman 2012;
58 Rabosky 2014; Alfaro *et al.* 2009]. Other models treat speciation as a process
59 that takes time [Rosindell *et al.* 2010; Etienne & Rosindell 2012; Lambert *et al.*
60 2015], or where diversification rate depends on one or more traits [Maddison
61 *et al.* 2007; FitzJohn 2012]. Despite such a rich theoretical landscape, only a

62 few of these diversification models are available as tree priors in BEAST2.

63 When a novel diversification model is introduced, its performance in in-
64 ference should be tested. Part of a model’s performance is its ability to re-
65 cover parameters from simulated data with known parameters (e.g. [Etienne
66 *et al.* 2014]), where ideally the estimated parameter values closely match the
67 known/true values. Even when a diversification model passes this procedure,
68 it is not necessarily used as tree prior in Bayesian inference. Bayesian phy-
69 logenetic inference often requires that the prior probability of the phylogeny
70 according to the diversification model has to be computed millions of times.
71 Therefore, biologically interesting but computationally expensive tree priors are
72 often not implemented, and simpler priors are used instead. This is not neces-
73 sarily problematic, when the data are very informative, as this will reduce the
74 influence of the tree prior. However, the assumption that tree prior choice is of
75 low importance must first be verified.

76 There have been multiple attempts to investigate the importance of tree
77 prior choice. For example, Sarver and colleagues, [Sarver *et al.* 2019] showed
78 that the choice of tree prior does not substantially affect phylogenetic inferences
79 of diversification rates. However, they only compared current diversification
80 models to one another, and thus this does not inform us on the importance of
81 a new tree prior.

82 Here we introduce a method to quantify the importance of a novel tree
83 prior. The method starts with a phylogeny generated by the new model. Next,
84 nucleotide sequences are simulated that follow the evolutionary history of the
85 given phylogeny. Then, using BEAST2’s built-in tree priors, a Bayesian poste-
86 rior distribution of phylogenies is inferred. We then compare the inferred with
87 the original phylogenies. How to properly perform this comparison forms the
88 heart of our method. Only new diversification models that result in a large

89 discrepancy between inferred and simulated phylogenies will be worth the effort
90 and computational burden to implement a species tree prior for in a Bayesian
91 framework.

92 Our method is programmed as an R package [R Core Team 2013] called
93 **pirouette**. **pirouette** is built on **babette** [Bilderbeek & Etienne 2018], which
94 calls BEAST2 [Bouckaert *et al.* 2019].

95 2 Description

96 The goal of **pirouette** is to quantify the importance of a tree prior. It does
97 so by measuring the inference error made for a given reconstructed phylogeny,
98 simulated under a (usually novel) diversification model. We refer to the model
99 that has generated the given tree as the 'generative tree model' p_G . We define
100 the birth-death (BD) model [Nee *et al.* 1994] as the standard tree model, as many
101 (non-standard) tree models have a parameter setting for which they reduce to
102 this model. One such example is the diversity-dependent (DD) diversification
103 model [Etienne & Haegeman 2020; Etienne *et al.* 2012]: the DD model is a
104 BD model with a speciation or extinction rate that depends on the number of
105 species and a clade-level carrying capacity. For a carrying capacity of infinity,
106 the DD model reduces to the BD model. When benchmarking a novel tree
107 model, one will typically construct phylogenies for different combinations of the
108 diversification model's parameters, to assess under which scenarios the inference
109 error cannot be neglected. While we recommend many replicate simulations
110 when assessing a novel tree prior, our examples contain only one replicate, as the
111 goal is to show the workings of **pirouette**, instead of doing an extensive analysis.
112 The supplementary material, however, does show the results of replicated runs
113 under multiple settings.

114 **pirouette** allows the user to specify a wide variety of custom settings. These

115 settings can be grouped in macro-sections, according to how they operate in the
116 pipeline. We summarize them in Table 1 and Table 2.

117 2.1 pirouette's pipeline

118 The pipeline to assess the error BEAST2 makes in inferring this phylogeny
119 contains the following steps:

- 120 1. The user supplies one or (ideally) more phylogenies from a new diversifi-
121 cation model.
- 122 2. From the given phylogeny an alignment is simulated under a known align-
123 ment model A .
- 124 3. From this alignment, according to the specified inference conditions C , an
125 inference model I is chosen (which may differ from the generative model).
- 126 4. The inference model and the alignment are used to infer a posterior dis-
127 tribution of phylogenies.
- 128 5. The phylogenies in the posterior are compared with the given phylogeny
129 to estimate the error made, according to the error measure E specified by
130 the user.

131 The pipeline is visualized in Fig. 1. There is also the option to generate a
132 'twin tree', that goes through the same pipeline. The utility of this twin tree
133 will be explained below.

134 The first step simulates an alignment from the given phylogeny (Fig. 1, 1a →
135 2a). For the sake of clarity, here we will assume the alignment consists of DNA
136 sequences, but one can also use other heritable material such as amino acids.
137 The user must specify a root sequence, a mutation rate and a site model. The
138 root sequence is the DNA sequence of the shared common ancestor of all species,

Sub-argument	Description	Possible values
tree_prior	Macroevolutionary diversification model	BD, CBS, CCP, CEP, Yule
clock_model	Clock for the DNA mutation rates	RLN, strict
site_model	Nucleotide substitution model	GTR, HKY, JC, TN
mutation_rate	Pace at which mutation occurs	mutation_rate $\in \mathbb{R}_{>0}$
root_sequence	DNA sequence at the root of the tree	any combination of a, c, g, t
model_type	Criterion to select an inference model	Generative, Candidate
run_if	Condition under which an inference model is used	Always, Best candidate
do_measure_evidence	Sets whether or not the evidence of the model must be computed	TRUE, FALSE
error_fun	Specifies how to measure the error	nLTt, $ \gamma $
burn_in_fraction	Specifies the percentage of initial posterior trees to discard	burn_in_fraction $\in [0, 1]$

Table 1: Most important parameter options. BD = birth death [Nee *et al.* 1994], CBS = coalescent Bayesian skyline [Drummond *et al.* 2005], CCP = coalescent constant-population, CEP = coalescent exponential-population, Yule = pure birth model [Yule 1925], RLN = relaxed log-normal clock model [Drummond *et al.* 2006], strict = strict clock model [Zuckerkandl & Pauling 1965], GTR = Generalized time-reversible model [Tavaré 1986], HKY = Hasegawa, Kishino and Yano [Hasegawa *et al.* 1985], JC = Jukes and Cantor [Jukes *et al.* 1969], TN = Tamura and Nei [Tamura & Nei 1993].

Symbol	Macro-argument	Description
G	Generative model	The full setting to produce BEAST2 input data. Its core features are the tree prior p_G , the clock model c_G and the site model s_G .
A	Alignment model	Specifies the alignment generation, such as the clock model c_G , site model s_G and root sequence.
X_i	i -th candidate experiment	Full setting for a Bayesian inference. It is made by a candidate inference model I_i and its inference conditions C_i .
I	Inference model	The assumed phylogenetic inference model, of which the main components are the tree prior p_I , assumed clock model c_I and assumed site model s_I . Conditions under which I is used in the inference. They are composed of the model type, run condition and whether to measure the evidence.
C	Inference conditions	Errors measurement setup that can be specified providing an error function to measure the difference between the original phylogeny and the inferred posterior. The first iterations of the MCMC chain of the posterior may not be representative and can be discarded using a burn-in fraction.
E	Error measure parameters	

Table 2: Definitions of terms and relative symbols used in the main text and in Fig 1. To run the pipeline A , X and E must be specified.

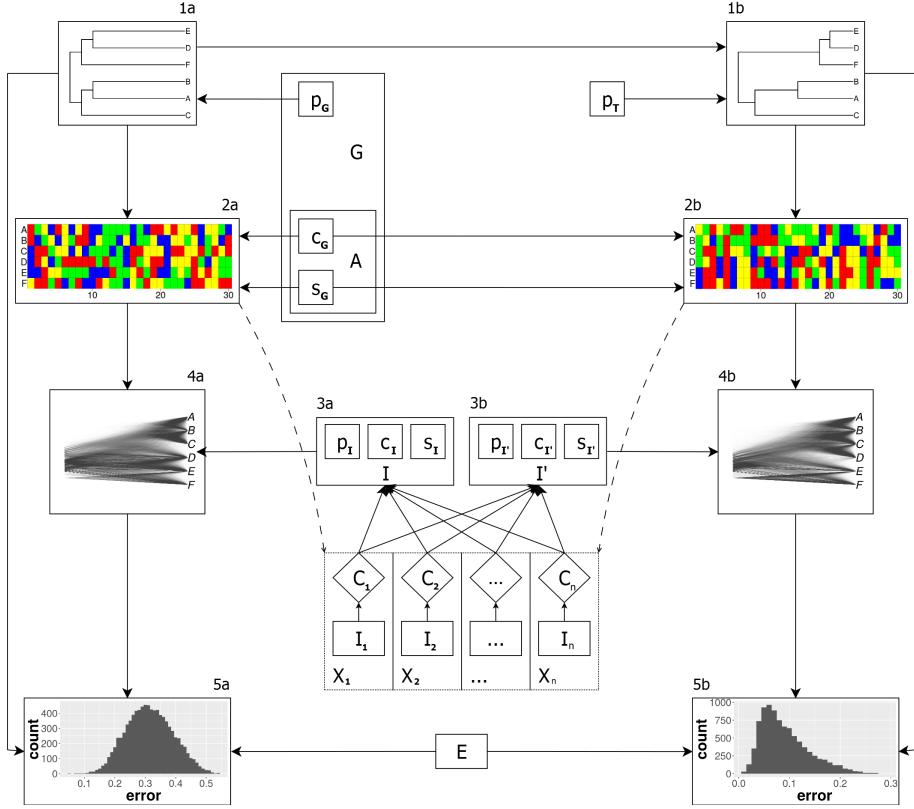


Figure 1: **pirouette** pipeline. The pipeline starts from a phylogeny (1a) simulated by the generative tree model p_G . The phylogeny is converted to an alignment (2a) using the generative alignment model $A = (c_G, s_G)$, composed of a clock model and a site model. The user defines one or more experiments. For each candidate experiment X_i (a combination of inference model I_i and condition C_i), if its condition C_i is satisfied (which can depend on the alignment), the corresponding inference model $I = I_i$ is selected to be used in the next step. The inference models (3a) of the selected experiments use the alignment (2a) to each create a Bayesian posterior of (parameter estimates and) phylogenies (4a). Each of the posteriors’ trees is compared to the true phylogeny (1a) using the error measure E , resulting in an error distribution (5a). Optionally, for each selected inference model a twin pipeline can be run. A twin phylogeny (1b) can be generated from the original phylogeny (1a) using the twin tree model p_T , selected among standard diversification models; the default option is the standard BD model, with parameters estimated from the original phylogeny. A twin alignment (2b) is then simulated from the twin phylogeny using clock model c_G and site model s_G imported from the generative model. The twin pipeline follows the procedure of the main pipeline, resulting in a twin error distribution (5b).

139 and is set to four different equally-sized mononucleotide blocks by default, which
140 helps interpreting the resulting alignment. Supported nucleotide substitution
141 models (part of a DNA site model) are JC, HKY, TN and GTR (see Table 1 for
142 the meaning of these abbreviations). JC is the default nucleotide substitution
143 model (NSM), in which the nucleotide substitution rates between all nucleotides
144 are equal and constant over time.

145 The second step (Fig. 1, 3a) selects one or more inference model(s) I from a
146 set of standard inference models I_1, \dots, I_n . For example, if the generative model
147 is known and standard, one can specify the inference model to be the same as
148 the generative model. If the tree model is unknown or non-standard - which is
149 the primary motivation for this paper -, one can pick a standard inference model
150 which is considered to be closest to the true tree model. Also, if we want to run
151 only the inference model that fits best to an alignment from a set of candidates,
152 one can specify these inference models as well (see section 'Candidate models').

153 The third step infers the posterior distributions, using the simulated align-
154 ment (Fig. 1, 2a → 4a), and the inference models that were selected in the
155 previous step (3a). For each selected experiment a posterior distribution is in-
156 ferred, using the **babette** [Bilderbeek & Etienne 2018] R package which makes
157 use of BEAST2. This step usually takes up most of the pipeline's computation
158 time.

159 The fourth step quantifies the inference error made. First the burn-in frac-
160 tion is removed, i.e. the first phase of the Markov chain Monte Carlo (MCMC)
161 run, which samples an unrepresentative part of parameter and tree space. By
162 default, **pirouette** removes the first 10% of the posterior. From the remaining
163 posterior, **pirouette** creates an error distribution, by measuring the difference
164 between the true tree and each of the posterior trees (Fig. 1, 4a → 5a). The user
165 can specify a function to quantify the differences between the true and poste-

rior trees. By default, the package uses the nLTT statistic [Janzen *et al.* 2015],
 which is the absolute difference between the normalized lineages-through-time
 plots of two trees. The nLTT statistic is chosen, as it can operate on any two
 trees (regardless of their crown ages and number of taxa) and its results have
 a clear range from zero to one. This normalized result makes it possible to
 compare trees from a distribution of trees from any tree model.

2.2 Twinning

Similar to lab experiments, `pirouette` allows to perform a control measurement,
 by use of a process we call twinning. In this context, this control results in
 an error distribution that is the baseline error of the pipeline. The difference
 between the 'true' and 'twin' error distributions is caused only by the mismatch
 between the generative tree model and the assumed tree prior.

The twinning process, T , encompasses two steps: T_1 , that generates a 'twin
 tree' (Fig. 1, 1b) and T_2 , which generates a 'twin alignment' (Fig. 1, 2b). Both
 twin tree and alignment will be analyzed in the same way as the true tree and
 alignment.

We define a phylogeny τ as the combination of branching times \vec{t} and topol-
 ogy ψ , and denote as τ_G the phylogeny produced by a (possibly non-standard)
 generative diversification model, having branching times \vec{t}_G and topology ψ_G .

The first step (T_1) of the twinning process creates a tree τ_T with branching
 times \vec{t}_T while preserving the original topology ψ_G :

$$\tau_G = (\vec{t}_G, \psi_G) \xrightarrow{T_1} \tau_T = (\vec{t}_T, \psi_G) \quad (1)$$

The default option for the diversification model p_T is the standard BD model. It
 is then possible to use the likelihood function L_T for this diversification model
 to find the parameters θ_T^* (e.g. speciation and extinction rates, in case of a BD

model) that maximize this likelihood applied to the true tree, conditioned on its number of tips n_G :

$$\max[L_T(\theta_T | \tau_G, n_G)] \rightarrow \theta_T^*. \quad (2)$$

We use θ_T^* to simulate a number $n_T = n_G$ of branching times \vec{t}_T for the twin tree τ_T , under the process p_T , while preserving the topology. We simulate the new branching times using the TESS package [Höhna 2013; Höhna *et al.* 2016b].

The second step (T_2) of the twinning process simulates the twin alignment with the same clock model, site model and mutation rate used to simulate the original alignment. The twin alignment can be simulated in any user-defined way. `pirouette` supplies the option to have it simulated with the same mutation rate as the true alignment. By default, however, not only the same mutation rate is used, but also the total number of mutations matches the true alignment. The total number of mutations is defined as the amount of different nucleotides between the (known) root sequence compared to the sequences at the tips.

2.3 Candidate models

The user has to specify exactly one standard inference model, but may be unsure which one to pick. To account for this, the user can specify a set of candidate inference models. Each of these candidate inference models is run in an initial, relatively short, analysis; the candidate model with the highest evidence (i.e., marginal likelihood) will then be used in another, longer, inference run, resulting in another error distribution. The evidence for an inference model is estimated by nested sampling [Russel *et al.* 2019], using the `NS BEAST2` package.

If twinning is used, a candidate model that has the highest evidence for the twin alignment is also used to create another error distribution.

206 **2.4 Stochasticity caused by simulating phylogenies**

207 Finally, if the goal is to evaluate BEAST2’s performance on a non-standard
208 tree model, one must also consider the last source of stochasticity: the different
209 phylogenies a tree model generates. A single phylogeny cannot be considered
210 as fully representative of the model. For this reason multiple phylogenies must
211 be considered. If the number of considered phylogenies is high enough, the
212 comparison between the main pipeline’s aggregated error distribution and its
213 twin counterpart leads to a fair evaluation of the new tree model with respect
214 to the baseline error.

215 **3 Usage**

216 We show the usage of `pirouette` on a tree generated by the non-standard
217 diversity-dependent (DD) tree model [Etienne & Haegeman 2020; Etienne *et al.*
218 2012], which is a BD model with a speciation rate that is dependent on the
219 number of species.

220 The code to reproduce our results can be found at [https://github.com/
221 richelbilderbeek/pirouette_example_30](https://github.com/richelbilderbeek/pirouette_example_30) and a simplified version is shown
222 here for convenience:

```
223 library(pirouette)  
224  
225 # Create phylogeny  
226 set.seed(314)  
227 phylogeny <- create_exemplary_dd_tree(  
228   n_taxa = 6,  
229   crown_age = 10,  
230   extinction_rate = 0.1  
231 )
```

```

232
233 # Setup pirouette
234 pir_params <- create_std_pir_params(
235   folder_name = "example_30"
236 )
237
238 # Do the runs
239 pir_out <- pir_run(
240   phylogeny = phylogeny,
241   pir_params = pir_params
242 )
243
244 # Save
245 library(ggplot2)
246 pir_outs <- list()
247 pir_outs[[1]] <- pir_out
248 pir_plots(pir_outs) +
249   ggtitle("pirouette example 30") +
250   ggsave("errors.png", width = 7, height = 7)
251
252 pir_save(
253   phylogeny = phylogeny,
254   pir_params = pir_params,
255   pir_out = pir_out,
256   folder_name = "example_30"
257 )

258 The DD tree that is generated by this code is as shown in Figure 2, which
259 has an arbitrarily chosen crown age of 10 time units and 6 tips for an extinction
260 rate of 0.1. The carrying-capacity is also set at 6. The initial speciation rate  $\lambda_0$ 

```

₂₆₁ is chosen such that the expected number of species in a constant-rate BD model
₂₆₂ would be equal to the number of tips, which amounts to $\lambda_0 = 0.63$.

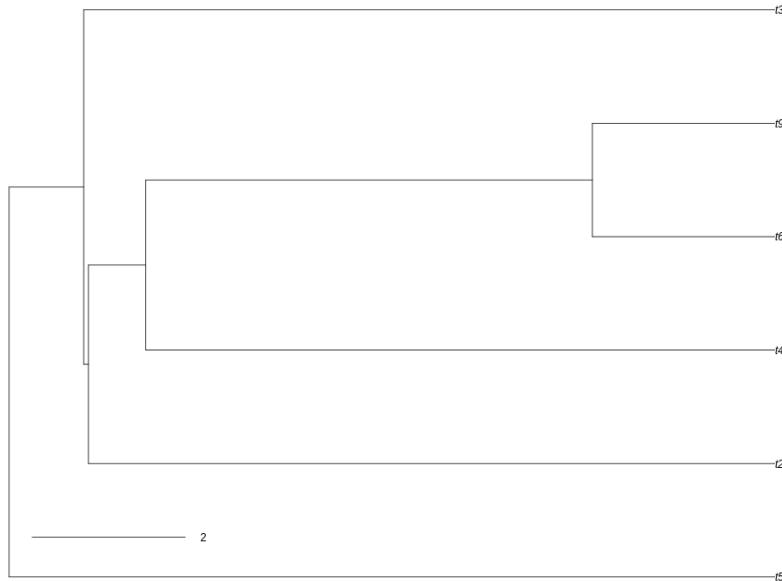


Figure 2: The example tree resulting from a diversity-dependent (DD) simulation.

₂₆₃ Using the default **pirouette** settings, the error distribution shown in Figure
₂₆₄ 3 is produced.

pirouette example 30

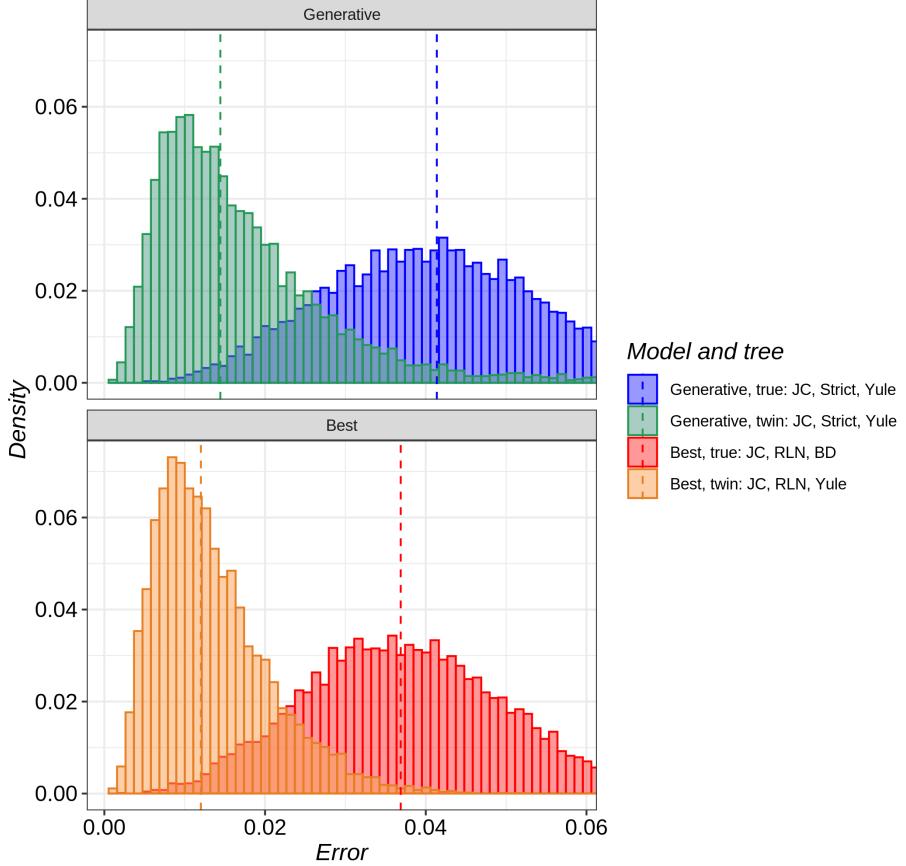


Figure 3: The inference error made for an assumed generative tree model and best candidate model, compared with the error obtained for the twin tree. The assumed generative tree model is the model assumed to be closest to the actual generative tree model that generated the true tree. The twin distributions show the baseline inference error. Vertical dashed lines show the median error value per distribution.

265 In the upper panel of Figure 3, we can see that the error distributions of
 266 the assumed generative model differ strongly between the true and twin tree.

267 This difference shows the extent of the mismatch between the actual tree model
268 (which is DD) and the (BD) tree prior used. Because these distributions are
269 distinctively different, the inference error we make when using an incorrect (that
270 is, BD) tree prior on a DD tree is profound.

271 Comparing the upper and lower panel of Figure 3, we can see that the
272 best candidate model is only slightly better at inferring the true tree, than the
273 assumed generative model, thereby showing that they cannot compensate for
274 the true generative tree model not being among the inference models.

275 The candidate model that had highest evidence given the simulated align-
276 ment, was JC, RLN, BD (see Table 1 for the meaning of these abbreviations).
277 JC was indeed the site model used when simulating the alignments. The RLN
278 clock model is an unexpected choice: the RLN clock model assumes that mu-
279 tation rates differ between lineages, whereas the alignment was simulated with
280 equal mutation rates for all lineages. The BD model is indeed closer to the DD
281 model than Yule, as both BD and DD model allow for extinction, while Yule
282 does not.

283 4 Discussion

284 We showed how to use `pirouette` to quantify the importance of a tree prior
285 in Bayesian phylogenetics, assuming the simplest standard tree model possible.
286 In principle any other standard tree model can be assumed, but we chose to
287 provide the simplest example.

288 Figure 3 illustrates the primary result of our pipeline: it shows the error
289 distributions for the true tree and the twin tree when either the (assumed)
290 generating model or the best candidate model is used in inference. The clear
291 difference between the error distributions for the true tree and the twin tree
292 suggests that the choice of tree prior does matter. We note, however, that only

293 one tree from a novel tree model is not enough to determine the impact of using
294 an incorrect tree prior. Instead, a distribution of multiple trees, generated by
295 the novel tree model, should be used. In the supplementary material we have
296 provided some examples.

297 Like most phylogenetic experiments, the setup of **pirouette** involves many
298 choices. A prime example is the length of the simulated DNA sequence. One
299 expects that the inference error decreases for longer DNA sequences. We inves-
300 tigated this superficially and confirmed this prediction (see the supplementary
301 materials). However, we note that for longer DNA sequences, the assumption
302 of constant substitution rates may become less realistic and hence longer se-
303 quences may require more parameters. Hence, simply getting longer sequences
304 will not always lead to a drastic reduction of the influence of the species tree
305 prior. Fortunately, **pirouette** provides a pipeline that works for all choices.

306 Interpreting the results of **pirouette** is up to the user; **pirouette** does not
307 answer the question whether the inference error is too large to trust the inferred
308 tree. The user is encouraged to use different statistics to measure the error. The
309 nLTT statistic is a promising starting point, as it can compare any two trees
310 and results in an error distribution of known range, but one may also explore
311 other statistics. In principle, **pirouette** allows for this, but in our example
312 we used a diversification model (DD) that only deviates from the Yule and BD
313 models in the temporal branching pattern, not in the topology.

314 As noted in the introduction, Duchêne and colleagues [Duchene *et al.* 2018],
315 also developed a method to assess the adequacy of a tree model on empirical
316 trees. They simulate trees from the posterior distribution of the parameters and
317 then compare this to the originally inferred tree using tree statistics, to deter-
318 mine whether the assumed tree model in inference indeed generates the tree as
319 inferred. This is useful if these trees match, but when they do not, this does not

320 mean that the inferred tree is incorrect; if sufficient data is available the species
321 tree prior may not be important, and hence the inference may be adequate even
322 though the assumed species tree prior is not. In short, the approach is applied
323 to empirical trees and compares the posterior and prior distribution of trees
324 (with the latter generated with the posterior parameters!). **pirouette** aims to
325 identify when assuming standard priors for the species tree leads to incorrect
326 inference if one believes more complex diversification models are operating than
327 can be currently accommodated in inference. In short, our approach applies to
328 simulated trees and compares the posterior distributions of trees generated with
329 a standard and non-standard model, but inferred with a standard one. The two
330 methods therefore complement one another.

331 However, we note that the **pirouette** pipeline is not restricted to exploring
332 the effects of a new species tree model. The pipeline can also be used to explore
333 the effects of non-standard clock or site models, such as relaxed clock models
334 with a non-standard distribution, correlated substitutions on sister lineages, or
335 elevated substitutions rates during speciation events. It is, however, beyond the
336 scope of this paper to discuss all these options in more detail.

337 In conclusion, **pirouette** can show the errors to be expected when the model
338 assumed in inference is different from the actual generative model. The user
339 can then judge whether or not this new model should be implemented in our
340 Bayesian phylogenetic tool.

341 5 Acknowledgments

342 We thank the Center for Information Technology of the University of Groningen
343 for its support and for providing access to the Peregrine high performance com-
344 puting cluster. We thank the Netherlands Organization for Scientific Research
345 (NWO) for financial support through a VICI grant awarded to RSE.

³⁴⁶ **6 Data Accessibility**

³⁴⁷ All code is archived at http://github.com/richelbilderbeek/piroquette_article, with DOI <https://doi.org/10.3456/zenodo.1234567>.

³⁴⁹ **7 Author contributions**

³⁵⁰ RJCB, GL and RSE conceived the idea for the package. RJCB created, tested
³⁵¹ and revised the package. GL provided major contributions to the package.
³⁵² RJCB wrote the first draft of the manuscript, GL and RSE contributed to
³⁵³ revisions.

³⁵⁴ **References**

³⁵⁵ Alfaro, M.E., Santini, F., Brock, C., Alamillo, H., Dornburg, A., Rabosky, D.L.,
³⁵⁶ Carnevale, G. & Harmon, L.J. (2009) Nine exceptional radiations plus high
³⁵⁷ turnover explain species diversity in jawed vertebrates. *Proceedings of the*
³⁵⁸ *National Academy of Sciences of the USA*, **106**, 13410–13414.

³⁵⁹ Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wick-
³⁶⁰ ham, H., Cheng, J. & Chang, W. (2017) *rmarkdown: Dynamic Documents*
³⁶¹ *for R*. R package version 1.8.

³⁶² Bache, S.M. & Wickham, H. (2014) *magrittr: A Forward-Pipe Operator for R*.
³⁶³ R package version 1.5.

³⁶⁴ Bilderbeek, R.J. (2019) mcbette. <https://github.com/richelbilderbeek/mcbette> [Accessed: 2019-01-21].

³⁶⁶ Bilderbeek, R.J. & Etienne, R.S. (2018) babette: BEAUti 2, BEAST 2 and
³⁶⁷ Tracer for R. *Methods in Ecology and Evolution*.

- 368 Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M.,
369 Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N. *et al.*
370 (2019) BEAST 2.5: An advanced software platform for Bayesian evolutionary
371 analysis. *PLoS Computational Biology*, **15**, e1006650.
- 372 Cotton, R. (2016) *assertive: Readable Check Functions to Ensure Code Integrity*.
373 R package version 0.3-5.
- 374 Drummond, A.J., Ho, S.Y., Phillips, M.J. & Rambaut, A. (2006) Relaxed phy-
375 logenetics and dating with confidence. *PLoS Biology*, **4**, e88.
- 376 Drummond, A.J. & Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis
377 by sampling trees. *BMC Evolutionary Biology*, **7**, 214.
- 378 Drummond, A.J., Rambaut, A., Shapiro, B. & Pybus, O.G. (2005) Bayesian
379 coalescent inference of past population dynamics from molecular sequences.
380 *Molecular Biology and Evolution*, **22**, 1185–1192.
- 381 Duchene, S., Bouckaert, R., Duchene, D.A., Stadler, T. & Drummond, A.J.
382 (2018) Phylodynamic model adequacy using posterior predictive simulations.
383 *Systematic Biology*, **68**, 358–364.
- 384 Etienne, R.S. & Haegeman, B. (2012) A conceptual and statistical framework for
385 adaptive radiations with a key role for diversity dependence. *The American
386 Naturalist*, **180**, E75–E89.
- 387 Etienne, R.S. & Haegeman, B. (2020) DDD. [https://CRAN.R-project.org/
388 package=DDD](https://CRAN.R-project.org/package=DDD).
- 389 Etienne, R.S., Haegeman, B., Stadler, T., Aze, T., Pearson, P.N., Purvis, A.
390 & Phillipmore, A.B. (2012) Diversity-dependence brings molecular phylogenies
391 closer to agreement with the fossil record. *Proceedings of the Royal Society
392 B: Biological Sciences*, **279**, 1300–1309.

- 393 Etienne, R.S., Morlon, H. & Lambert, A. (2014) Estimating the duration of
394 speciation from phylogenies. *Evolution*, **68**, 2430–2440.
- 395 Etienne, R.S. & Rosindell, J. (2012) Prolonging the past counteracts the pull of
396 the present: protracted speciation can explain observed slowdowns in diver-
397 sification. *Systematic Biology*, **61**, 204–213.
- 398 FitzJohn, R.G. (2012) Diversitree: comparative phylogenetic analyses of diver-
399 sification in R. *Methods in Ecology and Evolution*, **3**, 1084–1092.
- 400 Hasegawa, M., Kishino, H. & Yano, T.a. (1985) Dating of the human-ape split-
401 ting by a molecular clock of mitochondrial dna. *Journal of Molecular Evolu-*
402 *tion*, **22**, 160–174.
- 403 Hester, J. (2016) *lintr: Static R Code Analysis*. R package version 1.0.0.
- 404 Höhna, S. (2013) Fast simulation of reconstructed phylogenies under global
405 time-dependent birth–death processes. *Bioinformatics*, **29**, 1367–1374.
- 406 Höhna, S., Landis, M.J., Heath, T.A., Boussau, B., Lartillot, N., Moore, B.R.,
407 Huelsenbeck, J.P. & Ronquist, F. (2016a) RevBayes: Bayesian phylogenetic
408 inference using graphical models and an interactive model-specification lan-
409 guage. *Systematic Biology*, **65**, 726–736.
- 410 Höhna, S., May, M.R. & Moore, B.R. (2016b) TESS: an R package for efficiently
411 simulating phylogenetic trees and performing Bayesian inference of lineage
412 diversification rates. *Bioinformatics*, **32**, 789–791.
- 413 Huelsenbeck, J.P. & Ronquist, F. (2001) Mrbayes: Bayesian inference of phylo-
414 genetic trees. *Bioinformatics*, **17**, 754–755.
- 415 Janzen, T. (2019) nLTT. <https://github.com/thijsjanzen/nLTT> [Accessed:
416 2019-04-15].

- 417 Janzen, T., Höhna, S. & Etienne, R.S. (2015) Approximate Bayesian compu-
418 tation of diversification rates from molecular phylogenies: introducing a new
419 efficient summary statistic, the nLTT. *Methods in Ecology and Evolution*, **6**,
420 566–575.
- 421 Jukes, T.H., Cantor, C.R. *et al.* (1969) Evolution of protein molecules. *Mam-*
422 *malian Protein Metabolism*, **3**, 132.
- 423 Lambert, A., Morlon, H. & Etienne, R.S. (2015) The reconstructed tree in
424 the lineage-based model of protracted speciation. *Journal of Mathematical*
425 *Biology*, **70**, 367–397.
- 426 Maddison, W.P., Midford, P.E. & Otto, S.P. (2007) Estimating a binary char-
427 acter's effect on speciation and extinction. *Systematic Biology*, **56**, 701–710.
- 428 Maechler, M. (2019) *Rmpfr: R MPFR - Multiple Precision Floating-Point Re-*
429 *liable*. R package version 0.7-2.
- 430 Nee, S., May, R.M. & Harvey, P.H. (1994) The reconstructed evolutionary pro-
431 cess. *Philosophical Transactions of the Royal Society of London B*, **344**, 305–
432 311.
- 433 Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics
434 and evolution in R language. *Bioinformatics*, **20**, 289–290.
- 435 R Core Team (2013) *R: A Language and Environment for Statistical Computing*.
436 R Foundation for Statistical Computing, Vienna, Austria.
- 437 Rabosky, D.L. (2014) Automatic detection of key innovations, rate shifts, and
438 diversity-dependence on phylogenetic trees. *PLOS One*, **9**, e89543.
- 439 Rabosky, D.L. & Lovette, I.J. (2008) Explosive evolutionary radiations: de-
440 creasing speciation or increasing extinction through time? *Evolution: Inter-*
441 *national Journal of Organic Evolution*, **62**, 1866–1875.

- 442 Ratnakumar, S., Mick, T. & Davis, T. (2016) *rappdirs: Application Directories:*
443 *Determine Where to Save Data, Caches, and Logs.* R package version 0.3.1.
- 444 Revell, L.J. (2012) phytools: An R package for phylogenetic comparative biology
445 (and other things). *Methods in Ecology and Evolution*, **3**, 217–223.
- 446 Rosindell, J., Cornell, S.J., Hubbell, S.P. & Etienne, R.S. (2010) Protracted
447 speciation revitalizes the neutral theory of biodiversity. *Ecology Letters*, **13**,
448 716–727.
- 449 Russel, P.M., Brewer, B.J., Klaere, S. & Bouckaert, R.R. (2019) Model selection
450 and parameter inference in phylogenetics using nested sampling. *Systematic*
451 *Biology*, **68**, 219–233.
- 452 Sarver, B.A., Pennell, M.W., Brown, J.W., Keeble, S., Hardwick, K.M., Sulli-
453 van, J. & Harmon, L.J. (2019) The choice of tree prior and molecular clock
454 does not substantially affect phylogenetic inferences of diversification rates.
455 *PeerJ*, **7**, e6334.
- 456 Schliep, K. (2011) phangorn: phylogenetic analysis in R. *Bioinformatics*, **27**,
457 592–593.
- 458 Tamura, K. & Nei, M. (1993) Estimation of the number of nucleotide substitu-
459 tions in the control region of mitochondrial dna in humans and chimpanzees.
460 *Molecular biology and Evolution*, **10**, 512–526.
- 461 Tavaré, S. (1986) Some probabilistic and statistical problems in the analysis of
462 DNA sequences. *Lectures on Mathematics in the Life Sciences*, **17**, 57–86.
- 463 Wickham, H. (2009) *ggplot2: elegant graphics for data analysis.* Springer New
464 York.
- 465 Wickham, H. (2011a) The split-apply-combine strategy for data analysis. *Jour-*
466 *nal of Statistical Software*, **40**, 1–29.

- 467 Wickham, H. (2011b) *testthat: Get started with testing.* *The R Journal*, **3**,
- 468 5–10.
- 469 Wickham, H. (2015) *R packages: organize, test, document, and share your code.*
- 470 O'Reilly Media, Inc.
- 471 Wickham, H. (2017) *stringr: Simple, Consistent Wrappers for Common String*
- 472 *Operations.* R package version 1.2.0.
- 473 Wickham, H. & Chang, W. (2016) *devtools: Tools to Make Developing R Pack-*
- 474 *ages Easier.* R package version 1.12.0.9000.
- 475 Wickham, H., François, R., Henry, L. & Müller, K. (2019) *dplyr: A Grammar*
- 476 *of Data Manipulation.* R package version 0.8.1.
- 477 Wickham, H. & Henry, L. (2019) *tidyr: Easily Tidy Data with 'spread()' and*
- 478 *'gather()' Functions.* R package version 0.8.3.
- 479 Xie, Y. (2014) *testit: A Simple Package for Testing R Packages.* R package
- 480 version 0.4, <http://CRAN.R-project.org/package=testit>.
- 481 Xie, Y. (2017) *knitr: A General-Purpose Package for Dynamic Report Genera-*
- 482 *tion in R.* R package version 1.17.
- 483 Yule, G.U. (1925) A mathematical theory of evolution, based on the conclusions
- 484 of Dr. JC Willis, FRS. *Philosophical transactions of the Royal Society of*
- 485 *London Series B, containing papers of a biological character*, **213**, 21–87.
- 486 Zuckerkandl, E. & Pauling, L. (1965) Molecules as documents of evolutionary
- 487 history. *Journal of Theoretical Biology*, **8**, 357–366.

488 **8 Supplementary material**

489 This supplementary material contains additional facets of `pirouette`, such as
490 the installation of the package, an overview of `pirouette`'s main functions and
491 a guide for empiricists, based on multiple experiments that are shown here as
492 well.

493 For these experiments, we limited the number of replicates by time, aiming
494 at a duration of 24 hours per setting, when run on the Peregrine computer
495 cluster of the University of Groningen. Due to this, for example, a run of 40
496 taxa only has 6 replicates, because one run takes 4 hours. For all experiments,
497 the intermediate results can all be downloaded from their respective websites,
498 which is approximately 5 gigabyte in total.

499 All the figures shown in this section are shown without any aesthetical mod-
500ifications, with the exception that the arrangement of the sub-figures in subsec-
501tion 8.6, where we aligned parts of the figure by hand.

502 Here is an overview of the various sections:

- 503 • subsection 8.1: guidelines for empiricists
- 504 • subsection 8.2: installation
- 505 • subsection 8.3: resources, such as website, tutorials, packages used, bug
506 reporting and contributing
- 507 • subsection 8.4: citation of `pirouette`
- 508 • subsection 8.5: main functions
- 509 • subsection 8.6: code, extra figures and diagnostics regarding the main
510 example.
- 511 • subsection 8.7: the result of using multiple trees, as generated by the same
512 stochastic process as the main example

- 513 • subsection 8.8: the effect of the number of taxa
- 514 • subsection 8.9: the effect of the DNA alignment sequence length
- 515 • subsection 8.10 shows the effect when doing an inference in the simplest
- 516 use case
- 517 • subsection 8.11 shows the effect when doing an inference with an under-
- 518 parameterization
- 519 • subsection 8.13 shows the effect when the twin alignment is allowed to
- 520 have a different number of mutations
- 521 • subsection 8.14 shows the effect of different mutation rates

522 **8.1 Guidelines for empiricists**

523 From the experiments shown below, we composed some rough guidelines. These

524 guidelines should be treated as preliminary results, as the total runtime of these

525 experiments is 'only' 19 days.

- 526 • The use of 20 replicates results in decent plots.
- 527 • The use of more taxa increases the inference error
- 528 • The use of longer DNA sequences decreases the inference error.
- 529 • When we do not impose the same number of mutations between true and
- 530 twin alignment, we observe a difference in the error distributions with
- 531 respect to the standard case (presented in the main text) where they are
- 532 forced to have the same number of mutations.
- 533 • Using a mutation rate less than $1.0 / \text{crown age}$, decreases the inference
- 534 error. We predict this will increase the error in the parameter estimation.

535 **8.2 Installation**

536 `pirouette` will be made available on CRAN from which it can then be easily
537 installed:

```
538  
539    install.packages("pirouette")  
540
```

541 Until it is on CRAN, and for the most up-to-date version, one can download
542 and install the package from `pirouette`'s GitHub repository. We do first need
543 the `mcbette` and `nodeSub` packages:

```
544  
545    remotes::install_github(  
546        "richelbilderbeek/mcbette"  
547     )  
548    remotes::install_github(  
549        "thijsjanzen/nodeSub"  
550     )  
551
```

552 Now we can install `pirouette`:

```
553  
554    remotes::install_github(  
555        "richelbilderbeek/pirouette"  
556     )  
557
```

558 which also installs its dependencies from CRAN.

559 To start using `pirouette`, load its functions in the global namespace first:

```
560  
561    library(pirouette)  
562
```

563 Because `pirouette` calls BEAST2, BEAST2 must be installed. This can be
564 done from within R, using:

```
565  
566    beastier::install_beast2()  
567
```

568 For the option to select the best candidate model, `pirouette` needs the "NS"
569 BEAST2 package [Russel *et al.* 2019]. It can be installed from within R, using:

```
570  
571   mauricer::install_beast2_pkg("NS")  
572
```

573 8.3 Resources

574 **pirouette** is free, libre and open source software available at
575 <http://github.com/richelbilderbeek/pirouette>,
576 licensed under the GNU General Public License version 3. **pirouette** de-
577 pends on multiple packages, which are: **ape** [Paradis *et al.* 2004], **assertive**
578 [Cotton 2016], **babette** [Bilderbeek & Etienne 2018], **DDD** [Etienne & Haege-
579 man 2020], **devtools** [Wickham & Chang 2016], **dplyr** [Wickham *et al.* 2019],
580 **ggplot2** [Wickham 2009], **knitr** [Xie 2017], **lintr** [Hester 2016], **magrittr**
581 [Bache & Wickham 2014], **mcbette** [Bilderbeek 2019], **nLTT** [Janzen 2019], **phangorn**
582 [Schliep 2011], **phytools** [Revell 2012], **plyr** [Wickham 2011a], **rappdirs** [Rat-
583 nakumar *et al.* 2016], **rmarkdown** [Allaire *et al.* 2017], **Rmpfr** [Maechler 2019],
584 **stringr** [Wickham 2017], **TESS** [Höhna 2013], **testit** [Xie 2014], **testthat**
585 [Wickham 2011b] and **tidyR** [Wickham & Henry 2019].
586 **pirouette**'s development takes place on GitHub,
587 <https://github.com/richelbilderbeek/pirouette>,
588 which allows submitting bug reports, requesting features, and adding code.
589 To improve quality, **pirouette** uses a continuous integration service, has a code
590 coverage of above 95% and enforces the most commonly used R style guide
591 [Wickham 2015].
592 **pirouette**'s is extensively documented on its website, its documentation
593 and its vignettes. The **pirouette** website is a good starting point to learn how
594 to use **pirouette**, as it links to tutorials and videos. The **pirouette** package
595 documentation describes all functions and liberally links to related functions.
596 All exported functions show a minimal example as part of their documentation.

597 The `pirouette` vignette demonstrates extensively how to use `pirouette` in a
598 more informally written way.
599 The code used in this article and more examples that are periodically tested,
600 can be found at
601 https://github.com/richelbilderbeek/pirouette_examples.

602 8.4 Citation of `pirouette`

603 To cite `pirouette` this article from within R, use:

```
604 > citation("pirouette")
```

605 8.5 Main functions

606 An overview of `pirouette`'s main functions is shown in Table 3. All `pirouette`'s
607 functions are documented, have a useful example and sensible defaults.

Name	Description
<code>pir_run</code>	Run <code>pirouette</code>
<code>pir_plot</code>	Show the <code>pirouette</code> results as a plot
<code>create_pir_params</code>	Create the <code>pirouette</code> parameters
<code>create_alignment_params</code>	Create the alignment parameters
<code>create_twinning_params</code>	Create the twinning parameters
<code>create_experiment</code>	Create one experiment
<code>create_error_measure_params</code>	Create the error measurement parameters

Table 3: `pirouette`'s main functions and description.

608 **8.6 Main example**

609 This subsection shows the diagnostics from the main example, which uses one
610 tree. To assess if the results of the inference are meaningful one important
611 parameter is the Effective Sample Size (ESS). This quantity describes how many
612 independent trees are sampled from the posterior distributions. For reliable
613 results it is good practice to have at least $ESS = 200$ (see
614 https://beast.community/ess_tutorial).

615 In the following we present the ESS for the posterior distributions of the 4
616 cases shown in Fig. 4: "true" pipeline with generative model (Table ??), "true"
617 pipeline with best candidate model (Table ??), "twin" pipeline with genera-
618 tive model (Table ??), "twin" pipeline with best candidate model (Table ??).
619 We also report the marginal likelihood (or evidence) data for model selection
620 performed both in the "true" (Table ??) and "twin" pipeline (Table ??).

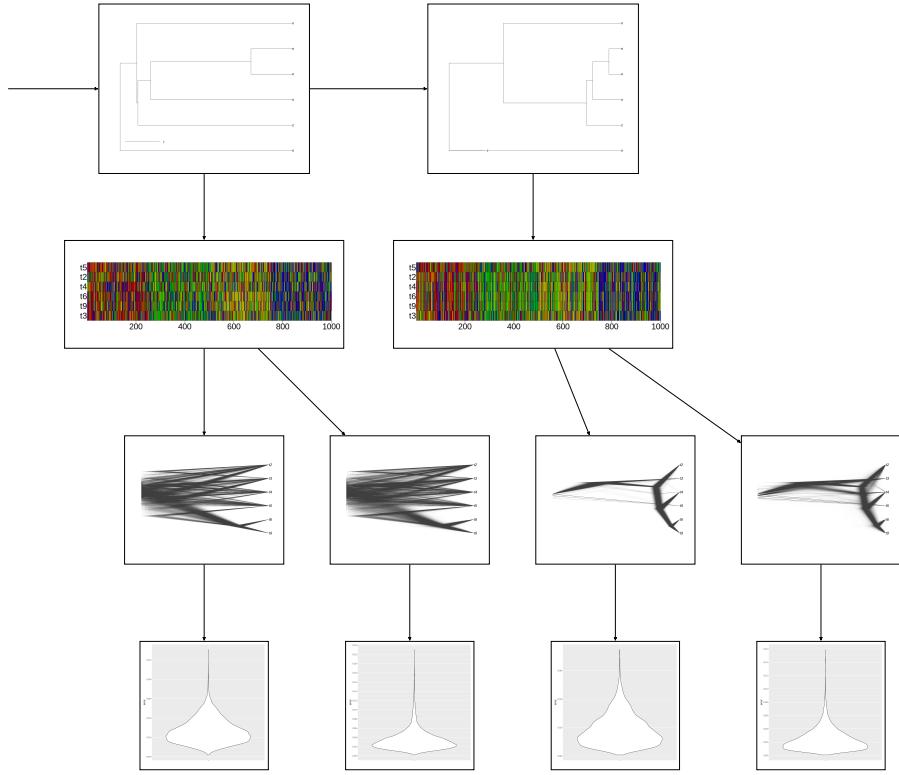


Figure 4: Full pirouette pipeline, including comparison to baseline error. The true tree (top right) is used to simulate an alignment. From this alignment two posterior distributions of trees are created: one using the generative model and another one using the inference model with the highest marginal likelihood. For each distribution of trees, a distribution of errors, measured with the nLTT statistic, between the posterior trees and the main trees is drawn. From the true tree also a twin tree is created (right side of the figure) which follows the same pipeline, leading to two additional error distributions to use as baseline errors.

.....

,,,,,,,,,,

621 8.7 Using a distribution of trees

622 This subsection extends the main example, by using multiple (instead of one)
623 trees. These trees are produced by running a DD tree simulation with the same
624 parameters as the main example.

625 The code to reproduce this figure can be found at

626 https://github.com/richelbilderbeek/piroquette_example_28.

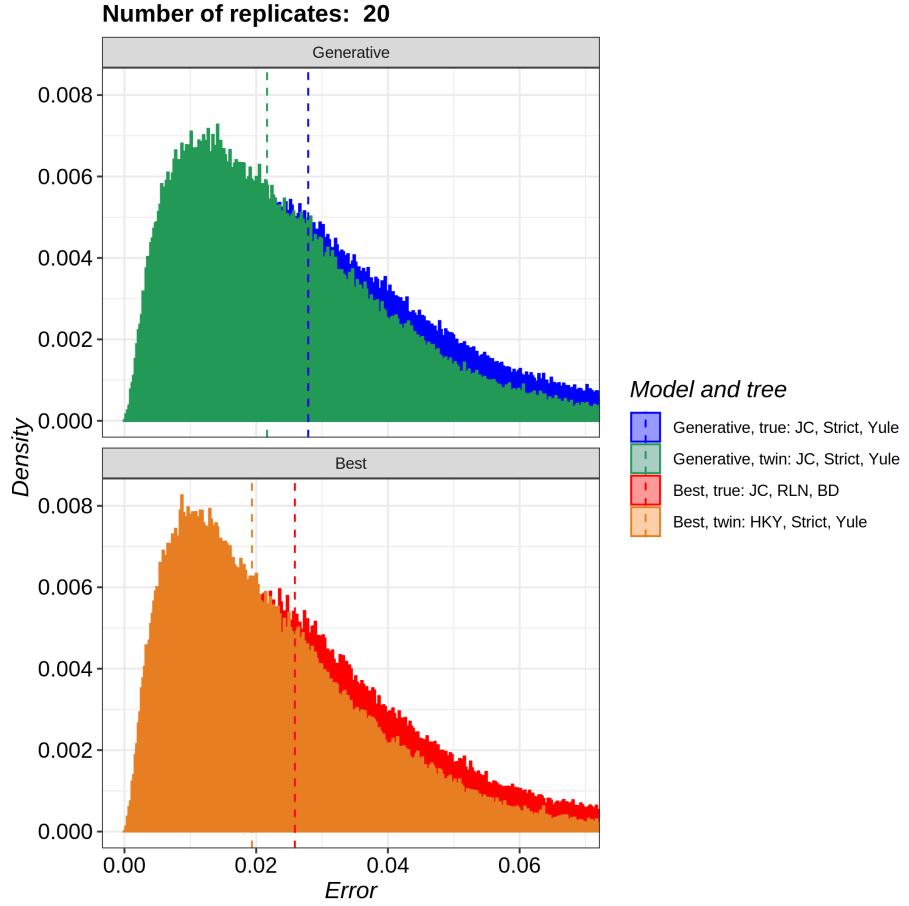


Figure 5: Aggregate error distributions, similar to Fig. 3 for the main example, but now for a collection of 20 replicate trees. For each setting (true generative, true best candidate, twin generative and twin best candidate), the resulting errors from each replicate pipeline have been merged into a single distribution.

627 **8.8 The effect of the number of taxa**

628 The main example uses 6 taxa. Here we show the same results as the main
629 example, except for a varying number of taxa. We did so, by setting the DD
630 model's carrying capacity to the desired number of taxa.

631 The code to reproduce these figure can be found at

632 https://github.com/richelbilderbeek/pirouette_example_28 (6
633 taxa, main example), https://github.com/richelbilderbeek/pirouette_example_32 (12 taxa), https://github.com/richelbilderbeek/pirouette_example_33 (24 taxa), https://github.com/richelbilderbeek/pirouette_example_41 (32 taxa), https://github.com/richelbilderbeek/pirouette_example_42 (40 taxa).

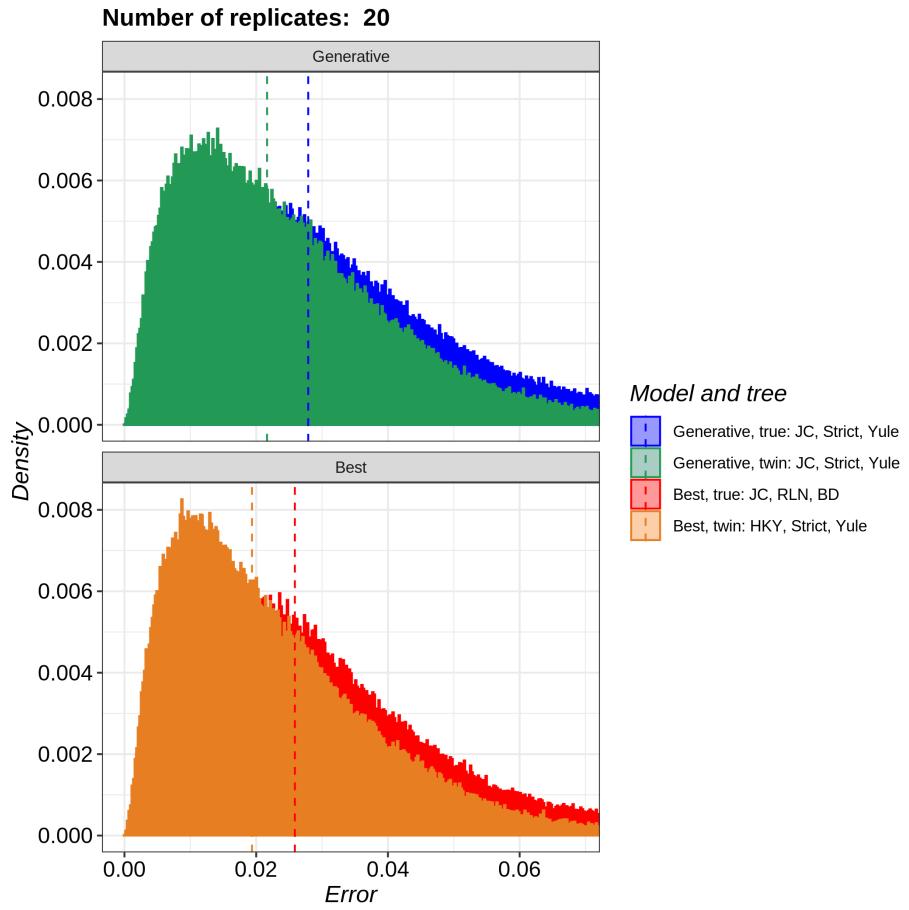


Figure 6: Aggregate error distributions for 20 replicates. Here each true tree has 6 taxa. This took 16 hours to compute.

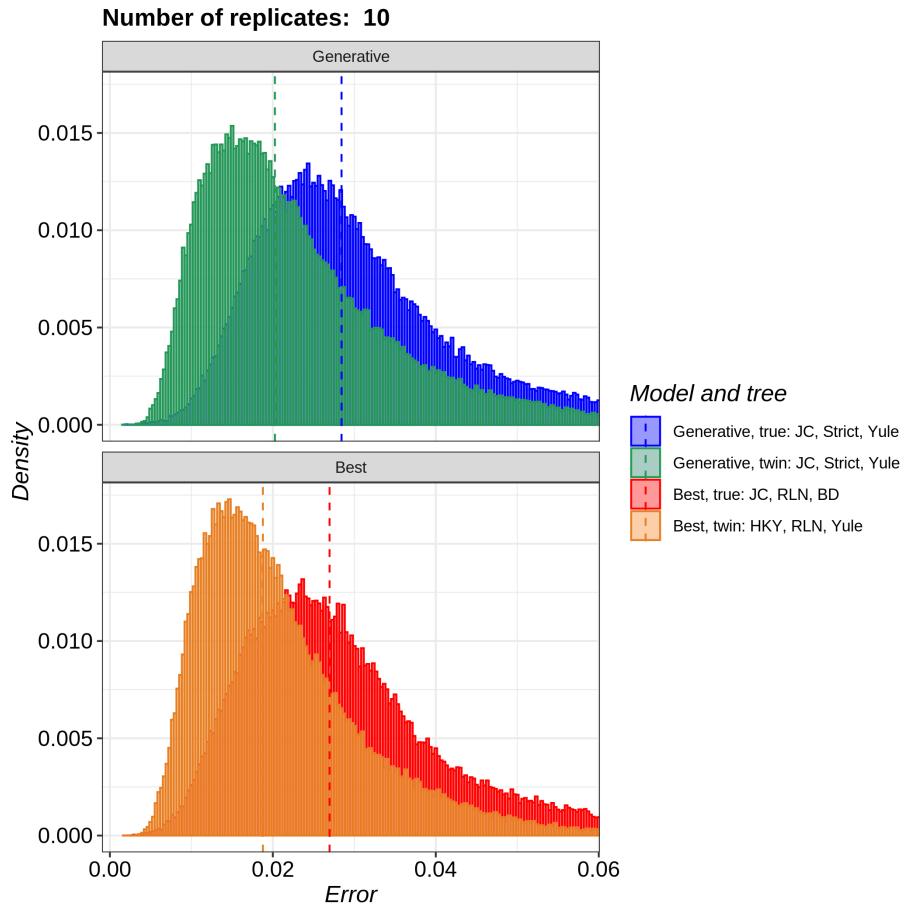


Figure 7: Aggregate error distributions for 10 replicates. Here each true tree has 12 taxa. This took 19 hours to compute.

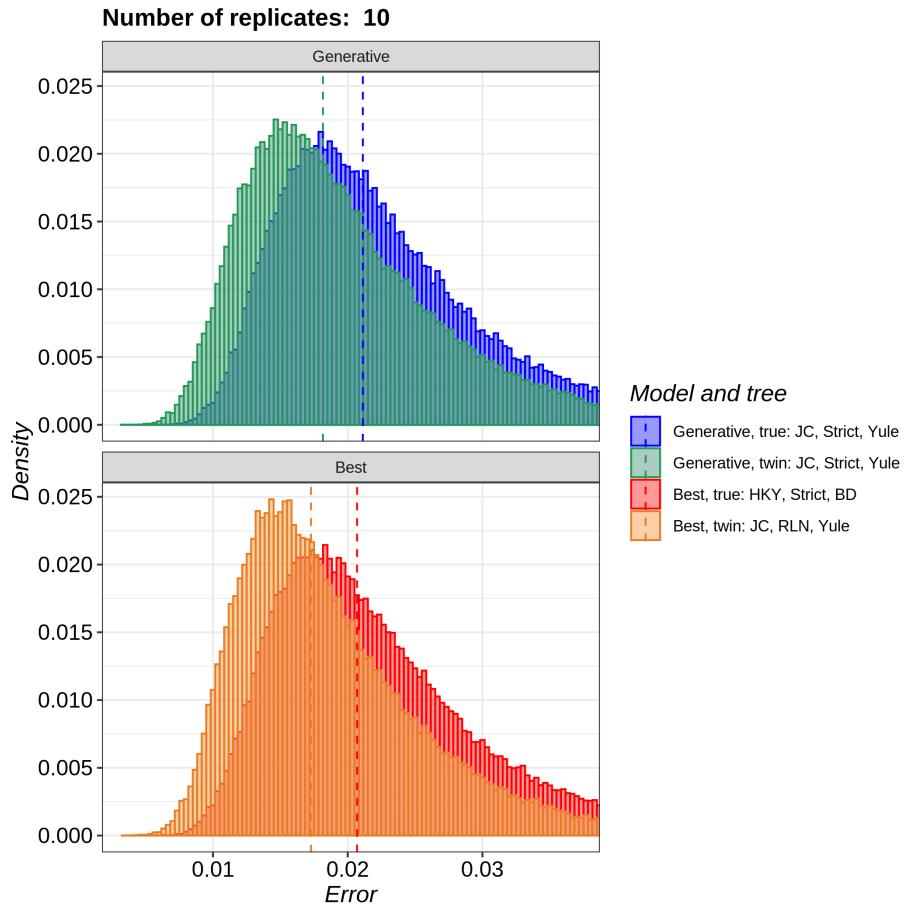


Figure 8: Aggregate error distributions for 20 replicates. Here each true tree has 24 taxa. This took 28 hours to compute.

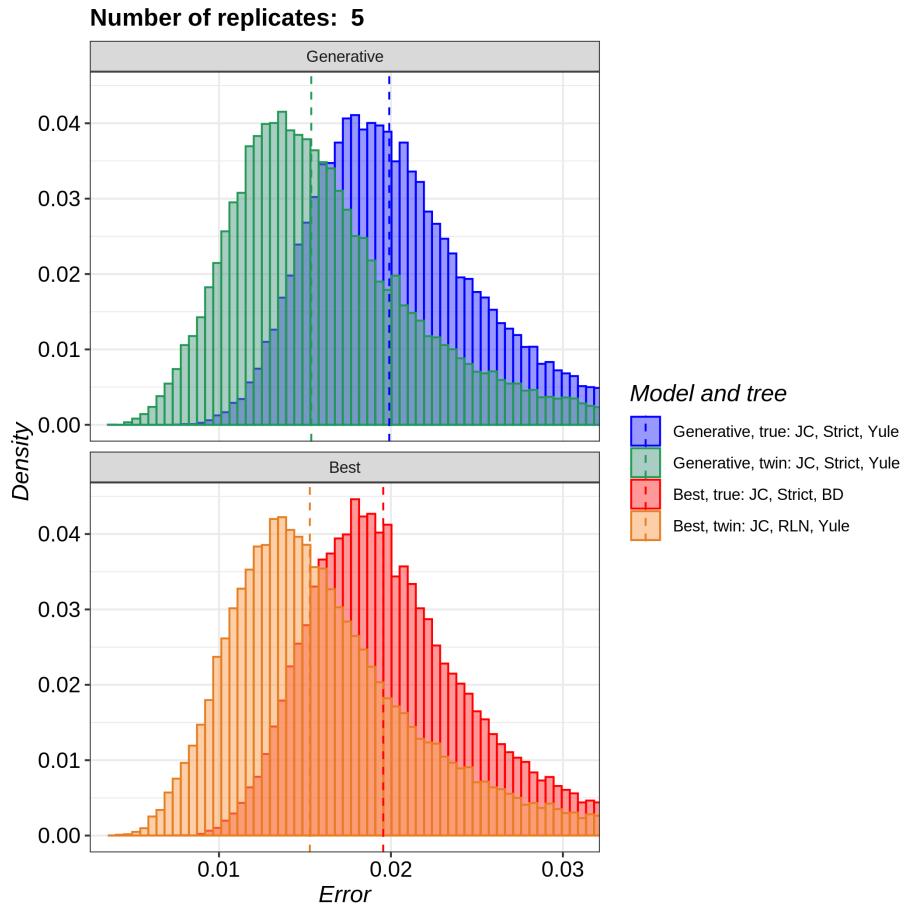


Figure 9: Aggregate error distributions for 5 replicates. Here each true tree has 32 taxa. This took 7 hours to compute.

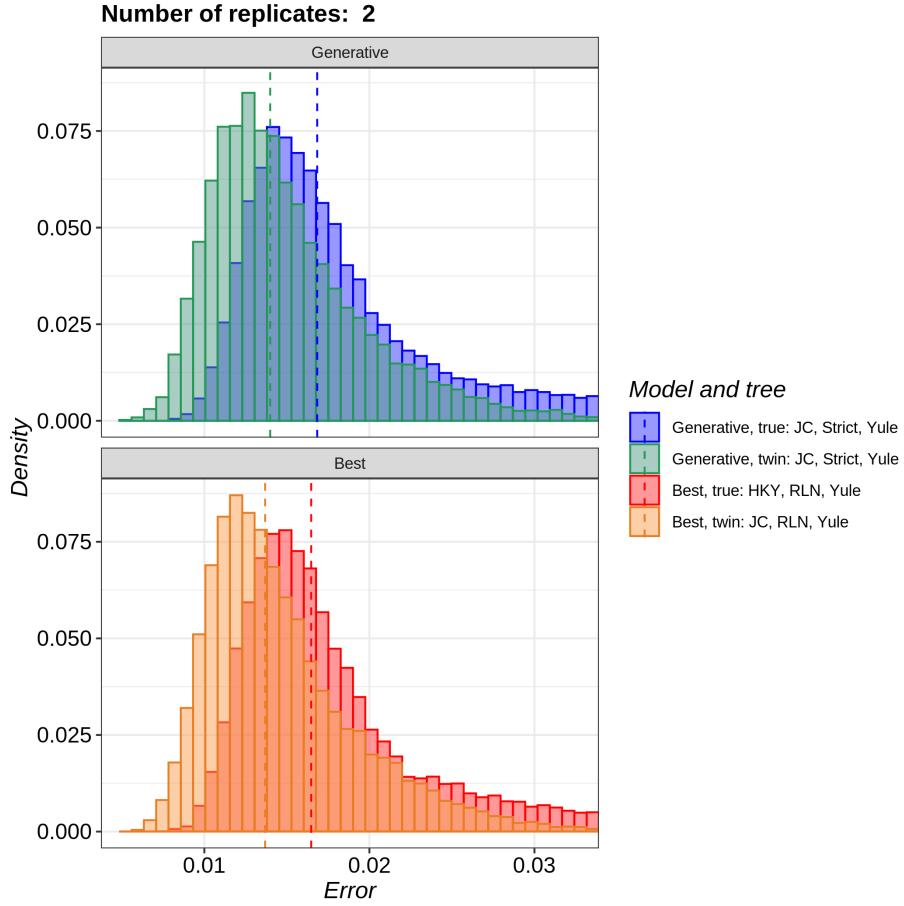


Figure 10: Aggregate error distributions for 2 replicates. Here each true tree has 40 taxa. This took 8 hours to compute.

638 8.9 The effect of DNA sequence length

639 The main example uses a DNA alignment length of 1000 nucleotides. Here, we
 640 show the same results as the main example, except for a varying DNA alignment
 641 sequence length.

642 The code to reproduce these figures can be found at

```

643     https://github.com/richelbilderbeek/pirouette\_example\_19
644     (500 nucleotides), https://github.com/richelbilderbeek/pirouette\_example\_28 (1000 nucleotides, main example), and https://github.com/richelbilderbeek/pirouette\_example\_34 (2000 nucleotides).
646

```

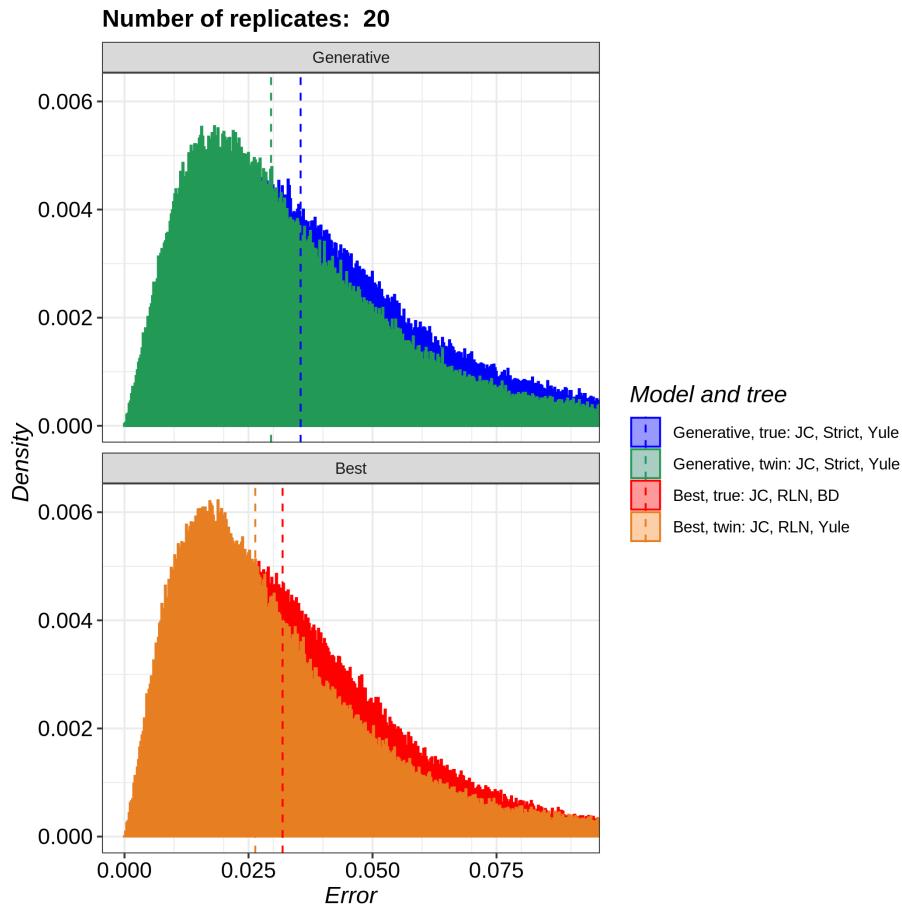


Figure 11: Aggregate error distributions for 20 replicates. Here each alignment has a sequence length of 500 nucleotides. This took 12 hours to compute.

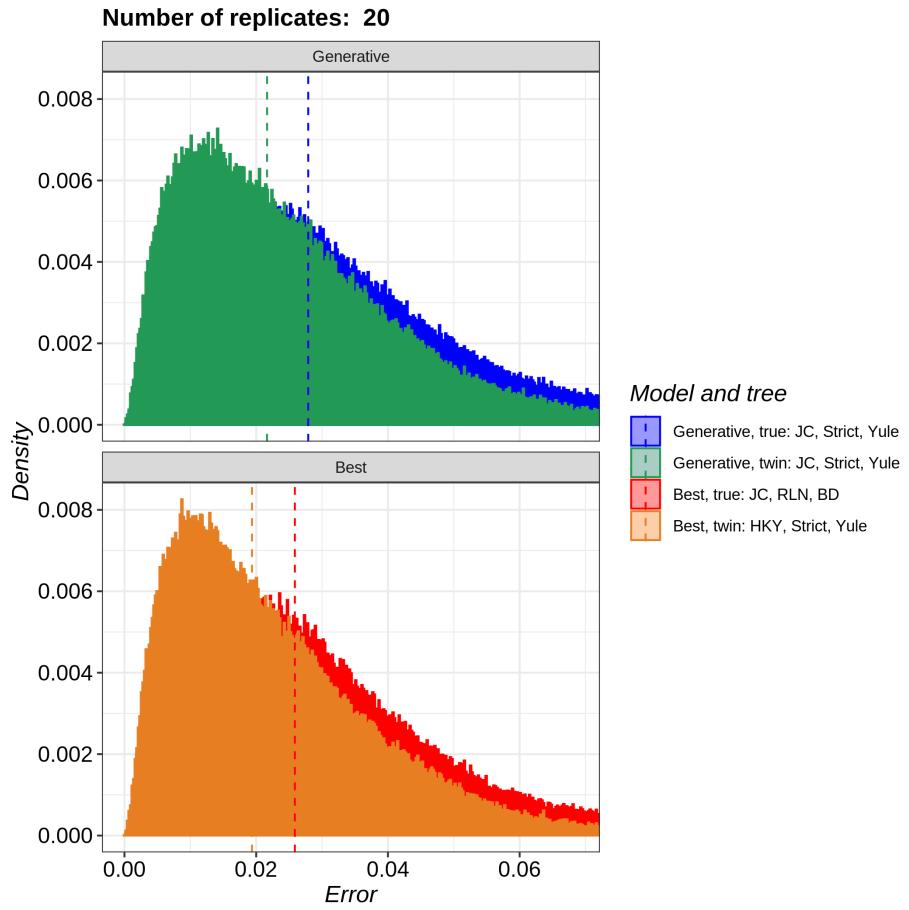


Figure 12: Aggregate error distributions for 20 replicates. Here each alignment has a sequence length of 1000 nucleotides (as used in the main example). This took 16 hours to compute.

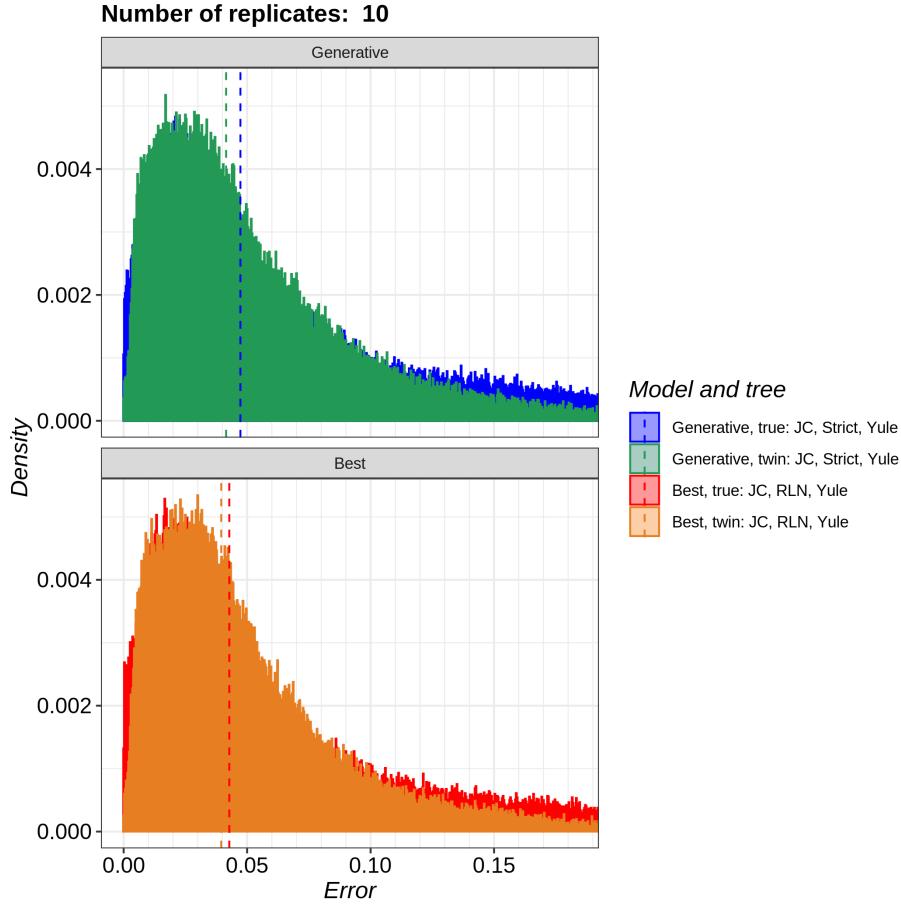


Figure 13: Aggregate error distributions for 10 replicates. Here each alignment has a sequence length of 2000 nucleotides. This took 11 hours to compute.

647 **8.10 The effect of assuming a Yule tree prior on a Yule
648 tree**

649 The main example uses a tree generated by a non-standard tree model. Here, we
650 show the same results, with the only difference that the tree used is generated
651 by simplest tree model (the Yule model), which we also assume as the (correct)

652 tree prior. This example thus shows a parameterization at the correct level for
653 the simplest case possible.

654 The code to reproduce this figure can be found at

655 https://github.com/richelbilderbeek/piroquette_example_22.

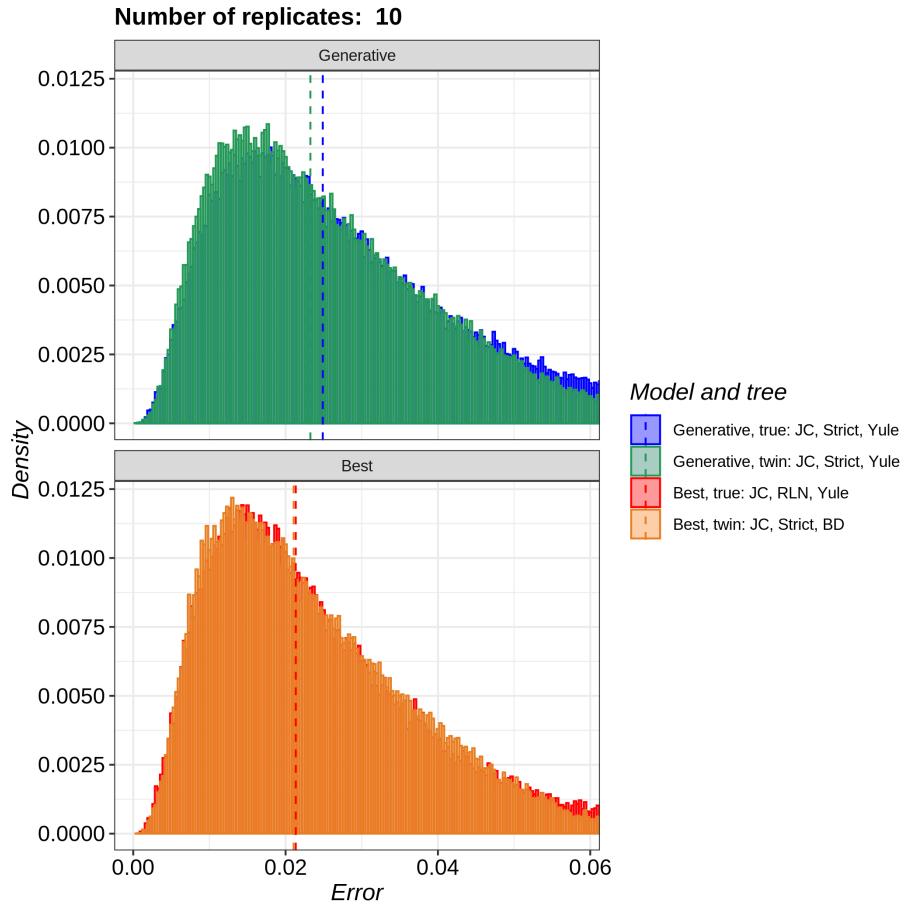


Figure 14: Aggregate error distributions for 10 replicates. Here each true tree is generated by a Yule process. For the inference we used a Yule tree prior. This took 9 hours to compute.

656 **8.11 The effect of assuming a Yule tree prior on a BD tree**

657 The main example uses a tree generated by a non-standard tree model.

658 Here, we show the same results, with the difference that the tree used is
659 generated by a birth-death (BD) tree model, where we assume it is generated
660 by a Yule (or pure-birth) model. This example thus shows the effect of under-
661 parameterization.

662 The code to reproduce this figure can be found at

663 https://github.com/richelbilderbeek/piroquette_example_26.

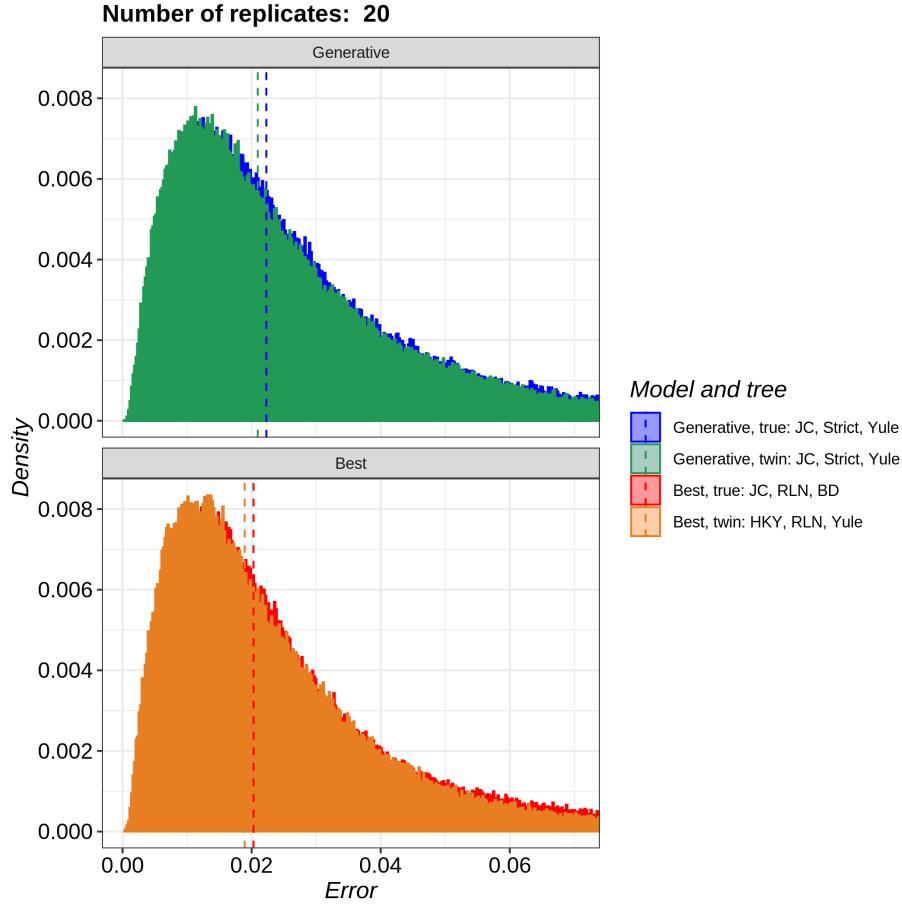


Figure 15: Aggregate error distributions for 20 replicates. Here each true tree is generated by a BD process. For the inference we used instead a Yule tree prior. This took 9 hours to compute.

664 **8.12 The effect of diversity-dependent trees differing in
665 how likely they are under the DD process**

666 Here we show the results of a `pirouette` run on a dataset of multiple DD trees
667 that we selected for having a low, median and high likelihood. In this way, we

668 effectively selected for trees that are rare, uncommon and common respectively.

669 The code to reproduce these figure can be found at

670 https://github.com/richelbilderbeek/piroquette_example_23

671 and took 120 hours in total.

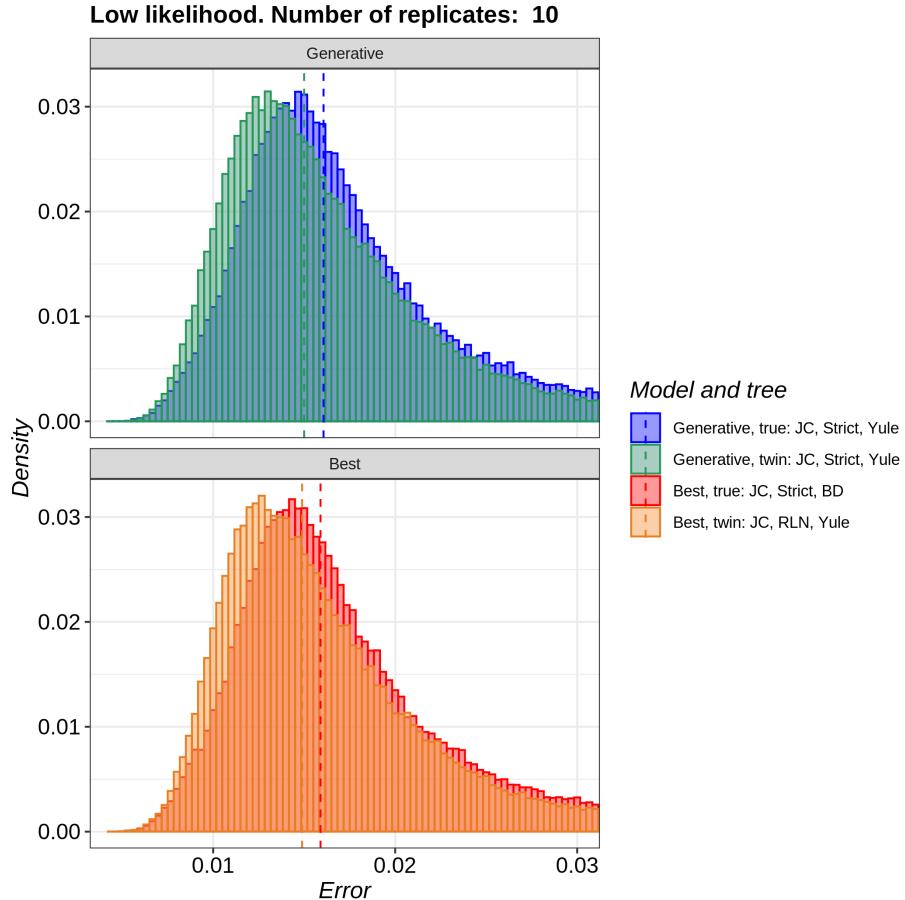


Figure 16: Aggregate error distributions for a distribution of trees, where the true trees are DD with low likelihood.

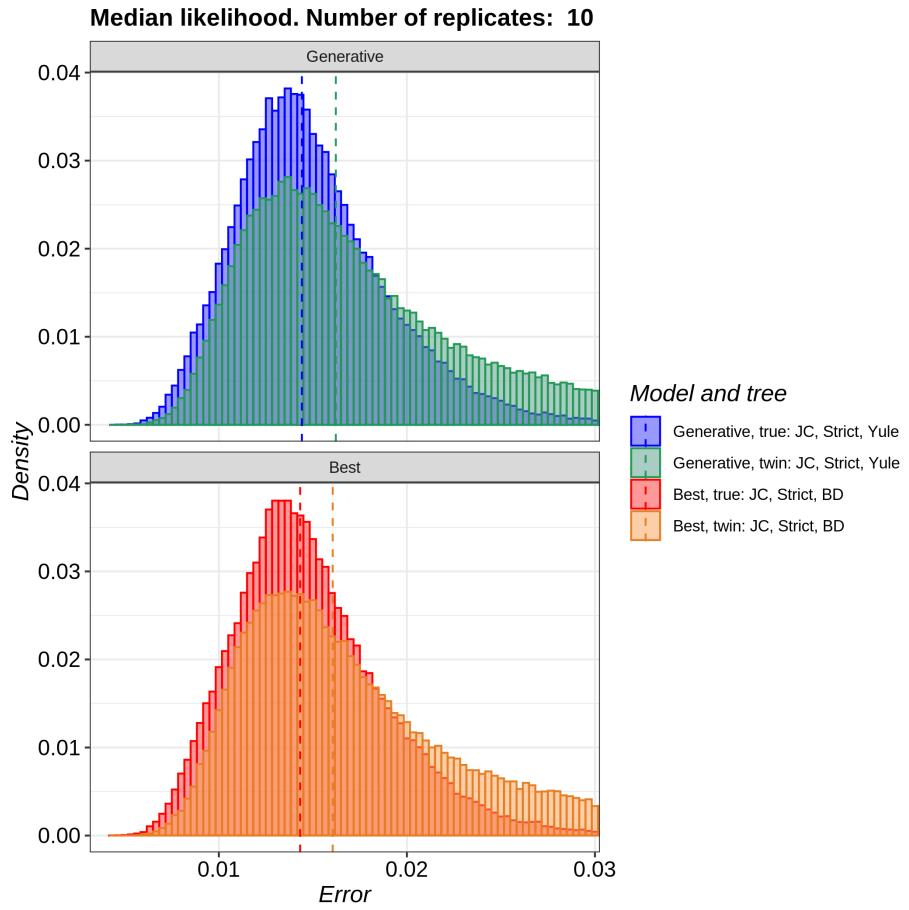


Figure 17: Aggregate error distributions for a distribution of trees, where the true trees are DD with median likelihood.

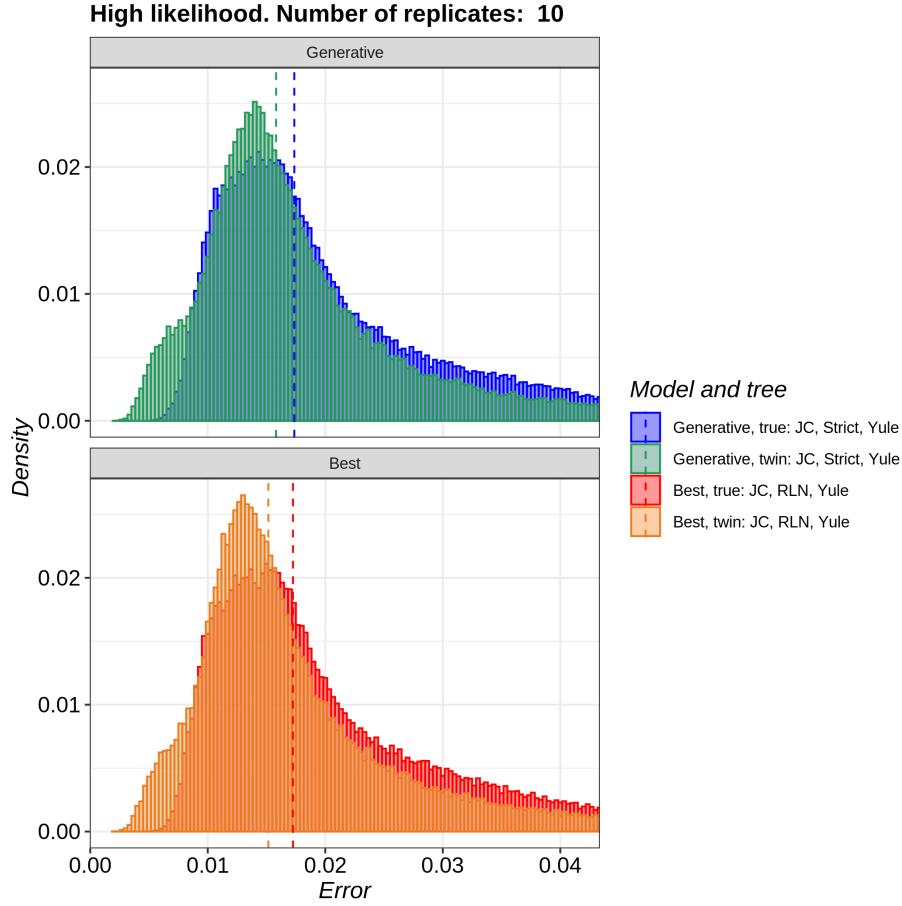


Figure 18: Aggregate error distributions for a distribution of trees, where the true trees are DD with high likelihood.

672 **8.13 The effect of equal or equalized mutation rate in the
673 twin alignment**

674 The main example uses a twin alignment that has the same number of mutations
675 (as measured from the ancestral sequence) as the true alignment. Here, we show
676 the same results, with the difference that the twin alignment uses the same

677 mutation rate, yet is not guaranteed to have the same number of mutations.

678 Any difference in the error distributions can therefore also stem from a difference

679 in the amount of information contained in the alignments.

680 The code to reproduce this figure can be found at

681 https://github.com/richelbilderbeek/pirouette_example_18 and

682 https://github.com/richelbilderbeek/pirouette_example_28.

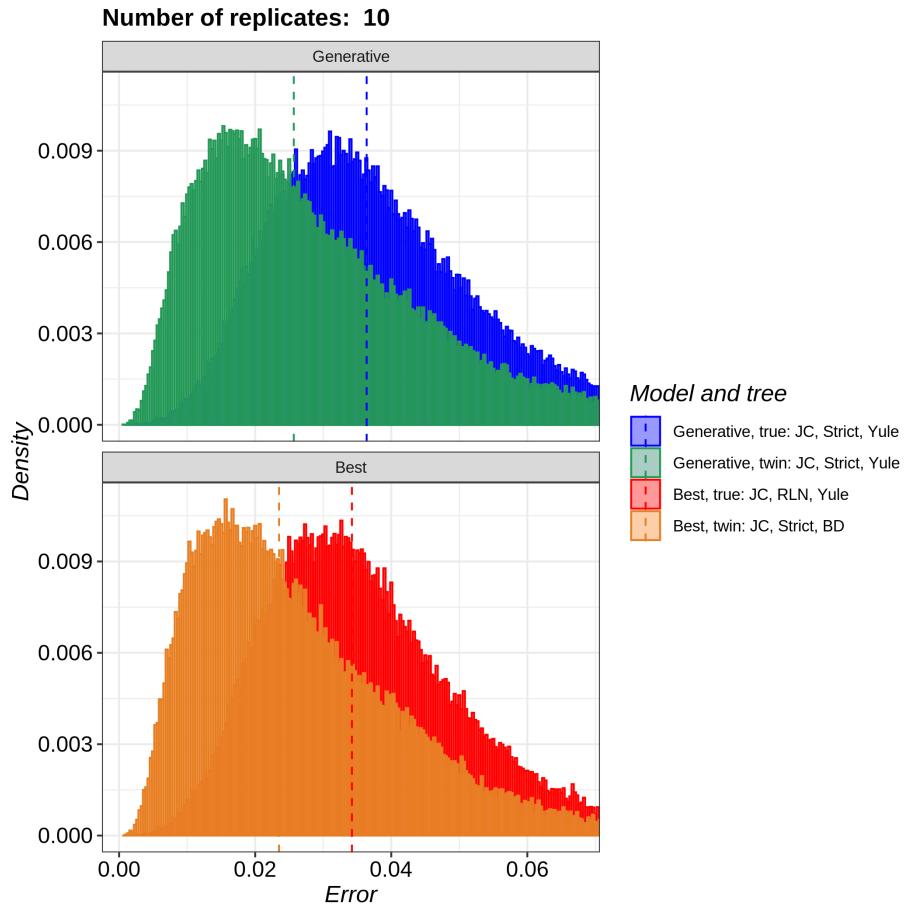


Figure 19: Aggregate error distributions similar to Fig. 5, but here the number of mutations is not imposed to be the same between true and twin alignment. Instead we just use an equal mutation rate. This took 10 hours to compute.

683 **8.14 The effect of mutation rate**

684 The main example uses a mutation rate such that all nucleotides, on average,
685 mutate once over the history going from the ancestral sequence at the crown to
686 the alignments at the tips. This value equals ‘ $1.0 / \text{crown age}$ ’. In this way, the
687 alignment is expected to contain the maximum amount of information.

688 Here, we show the same results for different mutation rates.

689 The code to reproduce this figure can be found at

690 https://github.com/richelbilderbeek/piquette_example_35 ($0.25 / \text{crown age}$),
691 https://github.com/richelbilderbeek/piquette_example_36 ($0.50 / \text{crown age}$), https://github.com/richelbilderbeek/piquette_example_37 ($0.75 / \text{crown age}$), https://github.com/richelbilderbeek/piquette_example_28 ($1.00 / \text{crown age}$, example reported in 8.7, see Fig.
692 5), https://github.com/richelbilderbeek/piquette_example_38 ($1.25 / \text{crown age}$),
693 https://github.com/richelbilderbeek/piquette_example_39 ($1.50 / \text{crown age}$), https://github.com/richelbilderbeek/piquette_example_40 ($2.00 / \text{crown age}$),

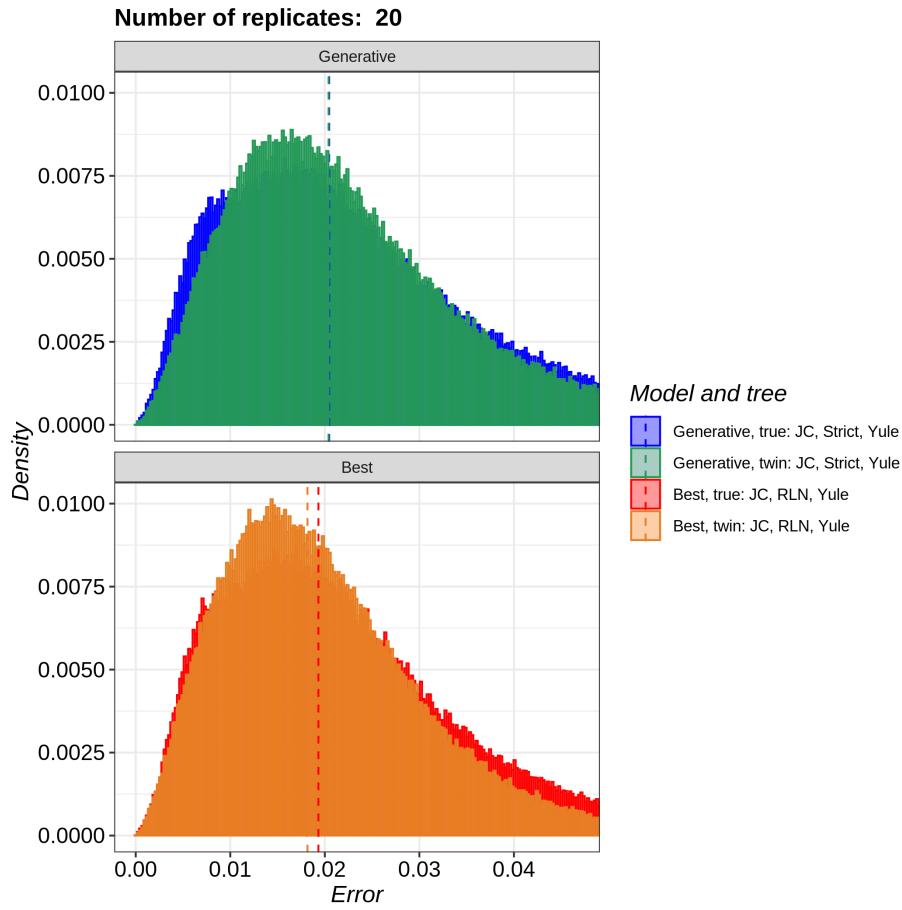


Figure 20: Aggregate error distributions for the tree distribution presented in 8.7 but with a per-nucleotide mutation rate of 0.25 / crown age. This took 11 hours to compute.

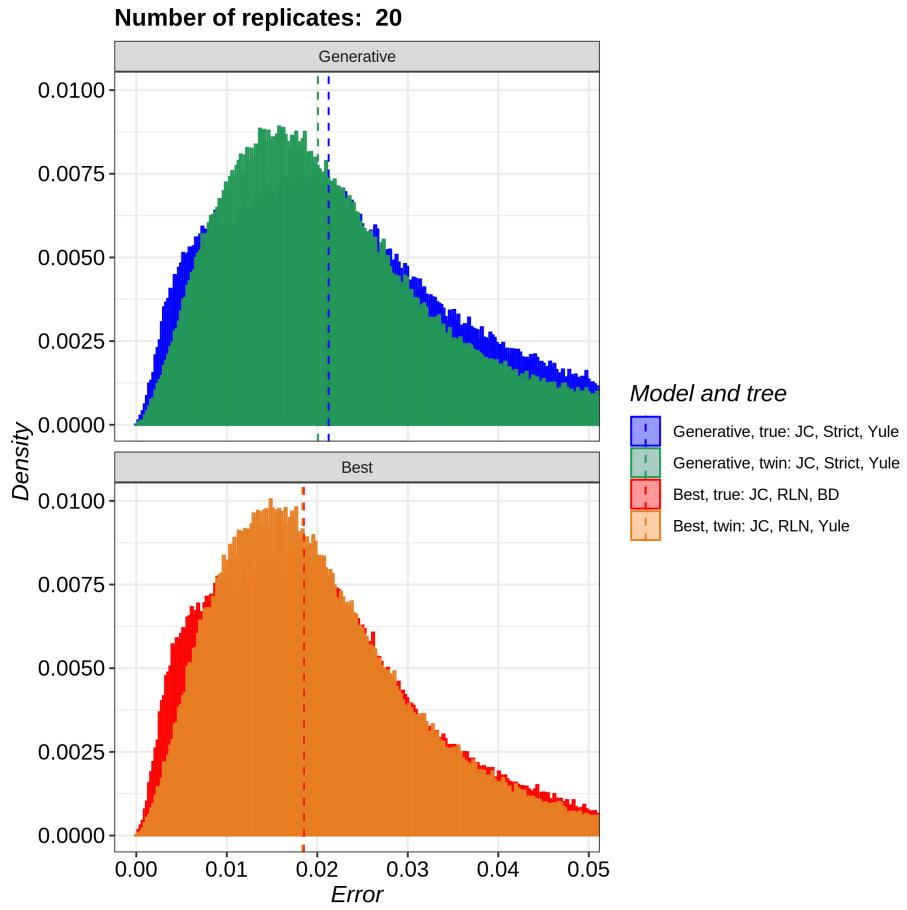


Figure 21: Aggregate error distributions for the tree distribution presented in 8.7 but with a per-nucleotide mutation rate of 0.50 / crown age. This took 14 hours to compute.

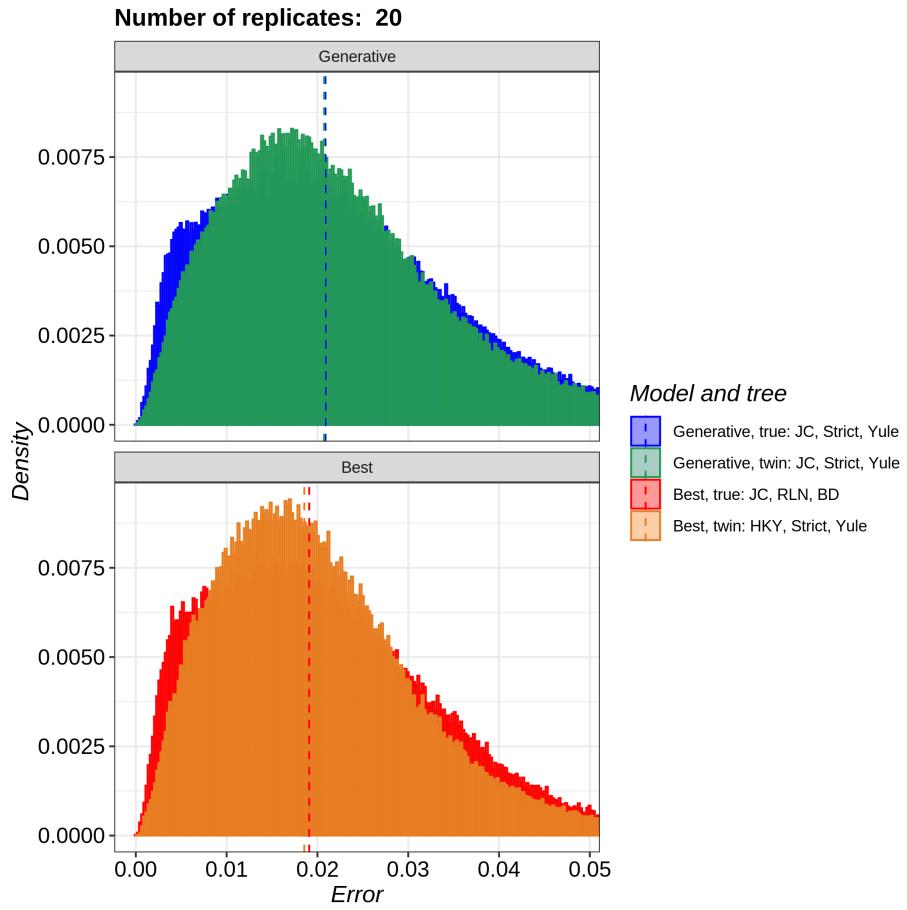


Figure 22: Aggregate error distributions for the tree distribution presented in 8.7 but with a per-nucleotide mutation rate of 0.75 / crown age. This took 16 hours to compute.

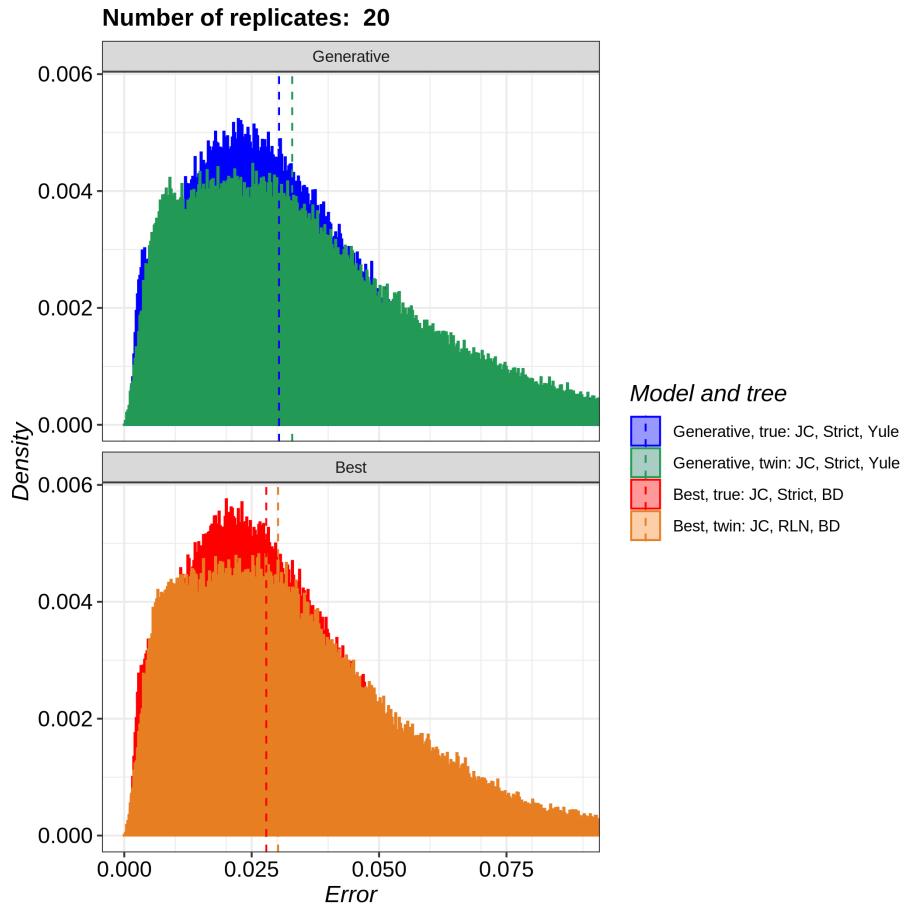


Figure 23: Aggregate error distributions for the tree distribution presented in 8.7 but with a per-nucleotide mutation rate of 1.25 / crown age. This took 17 hours to compute.

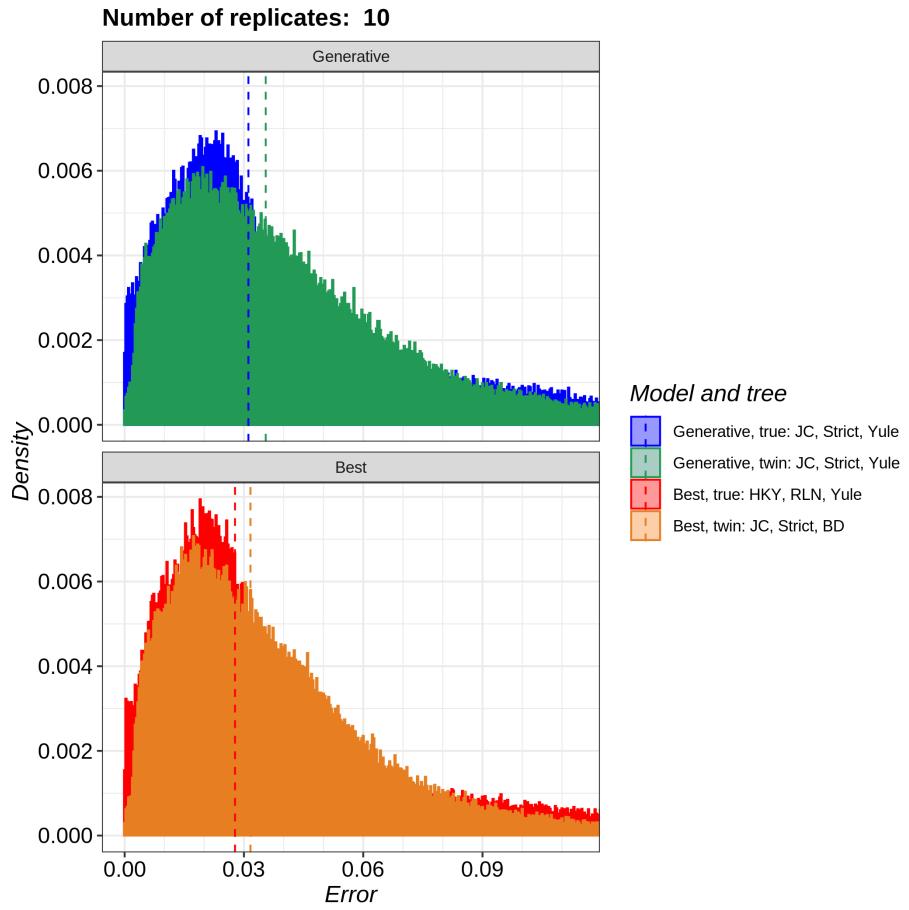


Figure 24: Aggregate error distributions for the tree distribution presented in 8.7 but with a per-nucleotide mutation rate of 1.50 / crown age. This took 10 hours to compute.

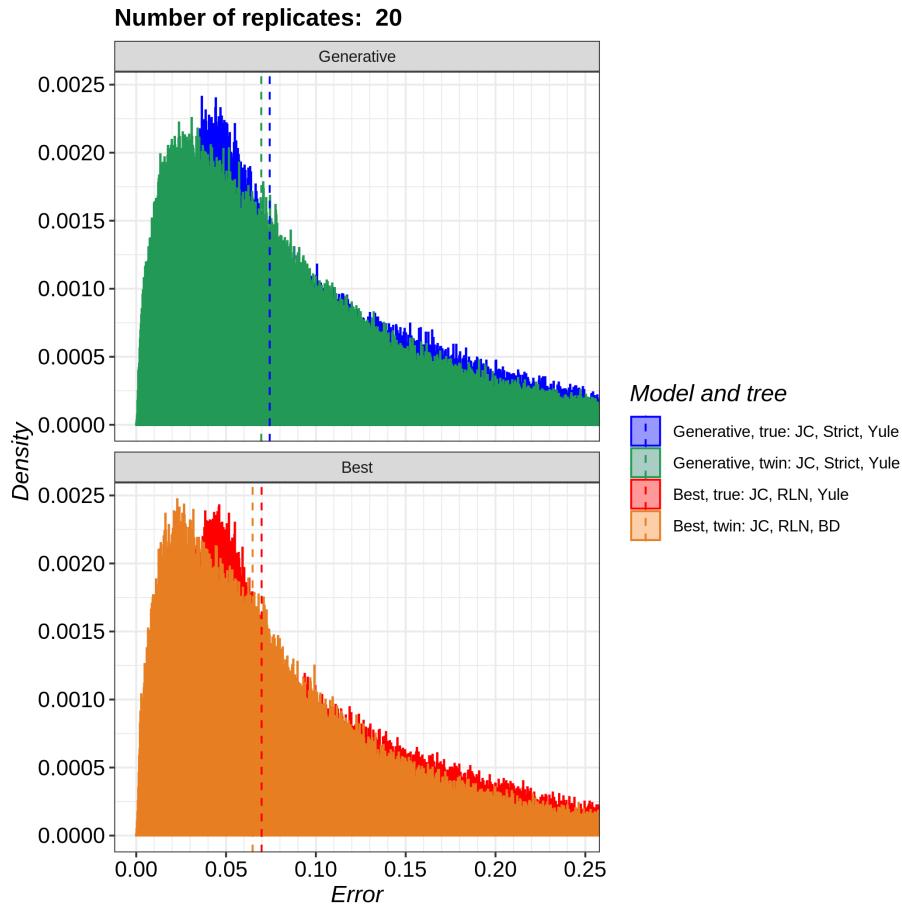


Figure 25: Aggregate error distributions for the tree distribution presented in 8.7 but with a per-nucleotide mutation rate of 2.0 / crown age. This took 19 hours to compute.

699 References

- 700 Alfaro, M.E., Santini, F., Brock, C., Alamillo, H., Dornburg, A., Rabosky, D.L.,
 701 Carnevale, G. & Harmon, L.J. (2009) Nine exceptional radiations plus high

- 702 turnover explain species diversity in jawed vertebrates. *Proceedings of the*
703 *National Academy of Sciences of the USA*, **106**, 13410–13414.
- 704 Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wick-
705 ham, H., Cheng, J. & Chang, W. (2017) *rmarkdown: Dynamic Documents*
706 for R. R package version 1.8.
- 707 Bache, S.M. & Wickham, H. (2014) *magrittr: A Forward-Pipe Operator for R*.
708 R package version 1.5.
- 709 Bilderbeek, R.J. (2019) mcbette. <https://github.com/richelbilderbeek/mcbette> [Accessed: 2019-01-21].
- 711 Bilderbeek, R.J. & Etienne, R.S. (2018) babette: BEAUTi 2, BEAST 2 and
712 Tracer for R. *Methods in Ecology and Evolution*.
- 713 Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M.,
714 Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N. *et al.*
715 (2019) BEAST 2.5: An advanced software platform for Bayesian evolutionary
716 analysis. *PLoS Computational Biology*, **15**, e1006650.
- 717 Cotton, R. (2016) assertive: Readable Check Functions to Ensure Code Integrity.
718 R package version 0.3-5.
- 719 Drummond, A.J., Ho, S.Y., Phillips, M.J. & Rambaut, A. (2006) Relaxed phy-
720 logenetics and dating with confidence. *PLoS Biology*, **4**, e88.
- 721 Drummond, A.J. & Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis
722 by sampling trees. *BMC Evolutionary Biology*, **7**, 214.
- 723 Drummond, A.J., Rambaut, A., Shapiro, B. & Pybus, O.G. (2005) Bayesian
724 coalescent inference of past population dynamics from molecular sequences.
725 *Molecular Biology and Evolution*, **22**, 1185–1192.

- 726 Duchene, S., Bouckaert, R., Duchene, D.A., Stadler, T. & Drummond, A.J.
727 (2018) Phylodynamic model adequacy using posterior predictive simulations.
728 *Systematic Biology*, **68**, 358–364.
- 729 Etienne, R.S. & Haegeman, B. (2012) A conceptual and statistical framework for
730 adaptive radiations with a key role for diversity dependence. *The American
731 Naturalist*, **180**, E75–E89.
- 732 Etienne, R.S. & Haegeman, B. (2020) DDD. [https://CRAN.R-project.org/
733 package=DDD](https://CRAN.R-project.org/package=DDD).
- 734 Etienne, R.S., Haegeman, B., Stadler, T., Aze, T., Pearson, P.N., Purvis, A.
735 & Phillipmore, A.B. (2012) Diversity-dependence brings molecular phylogenies
736 closer to agreement with the fossil record. *Proceedings of the Royal Society
737 B: Biological Sciences*, **279**, 1300–1309.
- 738 Etienne, R.S., Morlon, H. & Lambert, A. (2014) Estimating the duration of
739 speciation from phylogenies. *Evolution*, **68**, 2430–2440.
- 740 Etienne, R.S. & Rosindell, J. (2012) Prolonging the past counteracts the pull of
741 the present: protracted speciation can explain observed slowdowns in diver-
742 sification. *Systematic Biology*, **61**, 204–213.
- 743 FitzJohn, R.G. (2012) Diversitree: comparative phylogenetic analyses of diver-
744 sification in R. *Methods in Ecology and Evolution*, **3**, 1084–1092.
- 745 Hasegawa, M., Kishino, H. & Yano, T.a. (1985) Dating of the human-ape split-
746 ting by a molecular clock of mitochondrial dna. *Journal of Molecular Evolu-
747 tion*, **22**, 160–174.
- 748 Hester, J. (2016) *lintr: Static R Code Analysis*. R package version 1.0.0.
- 749 Höhna, S. (2013) Fast simulation of reconstructed phylogenies under global
750 time-dependent birth–death processes. *Bioinformatics*, **29**, 1367–1374.

- 751 Höhna, S., Landis, M.J., Heath, T.A., Boussau, B., Lartillot, N., Moore, B.R.,
752 Huelsenbeck, J.P. & Ronquist, F. (2016a) RevBayes: Bayesian phylogenetic
753 inference using graphical models and an interactive model-specification lan-
754 guage. *Systematic Biology*, **65**, 726–736.
- 755 Höhna, S., May, M.R. & Moore, B.R. (2016b) TESS: an R package for efficiently
756 simulating phylogenetic trees and performing Bayesian inference of lineage
757 diversification rates. *Bioinformatics*, **32**, 789–791.
- 758 Huelsenbeck, J.P. & Ronquist, F. (2001) Mrbayes: Bayesian inference of phylo-
759 genetic trees. *Bioinformatics*, **17**, 754–755.
- 760 Janzen, T. (2019) nLTT. <https://github.com/thijsjanzen/nLTT> [Accessed:
761 2019-04-15].
- 762 Janzen, T., Höhna, S. & Etienne, R.S. (2015) Approximate Bayesian compu-
763 tation of diversification rates from molecular phylogenies: introducing a new
764 efficient summary statistic, the nLTT. *Methods in Ecology and Evolution*, **6**,
765 566–575.
- 766 Jukes, T.H., Cantor, C.R. *et al.* (1969) Evolution of protein molecules. *Mam-
767 malian Protein Metabolism*, **3**, 132.
- 768 Lambert, A., Morlon, H. & Etienne, R.S. (2015) The reconstructed tree in
769 the lineage-based model of protracted speciation. *Journal of Mathematical
770 Biology*, **70**, 367–397.
- 771 Maddison, W.P., Midford, P.E. & Otto, S.P. (2007) Estimating a binary char-
772 acter's effect on speciation and extinction. *Systematic Biology*, **56**, 701–710.
- 773 Maechler, M. (2019) *Rmpfr: R MPFR - Multiple Precision Floating-Point Re-*
774 *liable.* R package version 0.7-2.

- 775 Nee, S., May, R.M. & Harvey, P.H. (1994) The reconstructed evolutionary pro-
776 cess. *Philosophical Transactions of the Royal Society of London B*, **344**, 305–
777 311.
- 778 Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics
779 and evolution in R language. *Bioinformatics*, **20**, 289–290.
- 780 R Core Team (2013) *R: A Language and Environment for Statistical Computing*.
781 R Foundation for Statistical Computing, Vienna, Austria.
- 782 Rabosky, D.L. (2014) Automatic detection of key innovations, rate shifts, and
783 diversity-dependence on phylogenetic trees. *PLOS One*, **9**, e89543.
- 784 Rabosky, D.L. & Lovette, I.J. (2008) Explosive evolutionary radiations: de-
785 creasing speciation or increasing extinction through time? *Evolution: Interna-*
786 *tional Journal of Organic Evolution*, **62**, 1866–1875.
- 787 Ratnakumar, S., Mick, T. & Davis, T. (2016) *rappdirs: Application Directories:*
788 *Determine Where to Save Data, Caches, and Logs*. R package version 0.3.1.
- 789 Revell, L.J. (2012) phytools: An R package for phylogenetic comparative biology
790 (and other things). *Methods in Ecology and Evolution*, **3**, 217–223.
- 791 Rosindell, J., Cornell, S.J., Hubbell, S.P. & Etienne, R.S. (2010) Protracted
792 speciation revitalizes the neutral theory of biodiversity. *Ecology Letters*, **13**,
793 716–727.
- 794 Russel, P.M., Brewer, B.J., Klaere, S. & Bouckaert, R.R. (2019) Model selection
795 and parameter inference in phylogenetics using nested sampling. *Systematic*
796 *Biology*, **68**, 219–233.
- 797 Sarver, B.A., Pennell, M.W., Brown, J.W., Keeble, S., Hardwick, K.M., Sulli-
798 van, J. & Harmon, L.J. (2019) The choice of tree prior and molecular clock

- 799 does not substantially affect phylogenetic inferences of diversification rates.
- 800 *PeerJ*, **7**, e6334.
- 801 Schliep, K. (2011) phangorn: phylogenetic analysis in R. *Bioinformatics*, **27**,
- 802 592–593.
- 803 Tamura, K. & Nei, M. (1993) Estimation of the number of nucleotide substitu-
- 804 tions in the control region of mitochondrial dna in humans and chimpanzees.
- 805 *Molecular biology and Evolution*, **10**, 512–526.
- 806 Tavaré, S. (1986) Some probabilistic and statistical problems in the analysis of
- 807 DNA sequences. *Lectures on Mathematics in the Life Sciences*, **17**, 57–86.
- 808 Wickham, H. (2009) *ggplot2: elegant graphics for data analysis*. Springer New
- 809 York.
- 810 Wickham, H. (2011a) The split-apply-combine strategy for data analysis. *Jour-*
- 811 *nal of Statistical Software*, **40**, 1–29.
- 812 Wickham, H. (2011b) testthat: Get started with testing. *The R Journal*, **3**,
- 813 5–10.
- 814 Wickham, H. (2015) *R packages: organize, test, document, and share your code*.
- 815 O'Reilly Media, Inc.
- 816 Wickham, H. (2017) *stringr: Simple, Consistent Wrappers for Common String*
- 817 *Operations*. R package version 1.2.0.
- 818 Wickham, H. & Chang, W. (2016) *devtools: Tools to Make Developing R Pack-*
- 819 *ages Easier*. R package version 1.12.0.9000.
- 820 Wickham, H., François, R., Henry, L. & Müller, K. (2019) *dplyr: A Grammar*
- 821 *of Data Manipulation*. R package version 0.8.1.

- 822 Wickham, H. & Henry, L. (2019) *tidyR: Easily Tidy Data with 'spread()' and*
823 *'gather()'* *Functions.* R package version 0.8.3.
- 824 Xie, Y. (2014) *testit: A Simple Package for Testing R Packages.* R package
825 version 0.4, <http://CRAN.R-project.org/package=testit>.
- 826 Xie, Y. (2017) *knitr: A General-Purpose Package for Dynamic Report Genera-*
827 *tion in R.* R package version 1.17.
- 828 Yule, G.U. (1925) A mathematical theory of evolution, based on the conclusions
829 of Dr. JC Willis, FRS. *Philosophical transactions of the Royal Society of*
830 *London Series B, containing papers of a biological character*, **213**, 21–87.
- 831 Zuckerkandl, E. & Pauling, L. (1965) Molecules as documents of evolutionary
832 history. *Journal of Theoretical Biology*, **8**, 357–366.