Barido-Sottani, J., Pett, W., O'Reilly, J. E., & Warnock, R. C. M. (2019). FossilSim: An r package for simulating fossil occurrence data under mechanistic models of preservation and recovery. *Methods in Ecology and Evolution*, *10*(6), 835-840. https://doi.org/10.1111/2041-210X.13170

Received: 19 July 2018 | Accepted: 16 February 2019

DOI: 10.1111/2041-210X.13170

APPLICATION

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

FossilSim: An R package for simulating fossil occurrence data under mechanistic models of preservation and recovery

Joëlle Barido-Sottani[1,2] | Walker Pett[3] | Joseph E. O'Reilly[4] | Rachel C. M. Warnock[1,2]

[1]Department of Biosystems Science & Engineering, Eidgenössische Technische Hochschule Zürich, Basel, Switzerland

[2]Swiss Institute of Bioinformatics (SIB), Switzerland

[3]Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, Iowa

[4]School of Earth Sciences, University of Bristol, Bristol, UK

Correspondence
Rachel C. M. Warnock
Email: rachel.warnock@bsse.ethz.ch

Handling Editor: Lee Hsiang Liow

Abstract

1. Key features of the fossil record that present challenges for integrating palaeontological and phylogenetic datasets include (i) non-uniform fossil recovery, (ii) stratigraphic age uncertainty and (iii) inconsistencies in the definition of species origination and taxonomy.
2. We present an R package FossilSim that can be used to simulate and visualise fossil data for phylogenetic analysis under a range of flexible models. The package includes interval-, environment- and lineage-dependent models of fossil recovery that can be combined with models of stratigraphic age uncertainty and species evolution.
3. The package input and output can be used in combination with the wide range of existing phylogenetic and palaeontological R packages. We also provide functions for converting between FossilSim and paleotree objects.
4. Simulated datasets provide enormous potential to assess the performance of phylogenetic methods and to explore the impact of using fossil occurrence databases on parameter estimation in macroevolution.

KEYWORDS
fossil preservation, macroevolution, phylogeny, sampling biases, simulation

1 | INTRODUCTION

The fossil record plays a fundamental role in understanding many aspects of evolution. Fossil occurrences have long provided the basis for inferring macroevolutionary trends, diversification rates and a timescale for events in Earth's history. It is also increasingly used to reconstruct and timescale phylogenetic relationships, motivating the development of phylogenetic models that can incorporate information from the fossil record (Heath, Huelsenbeck, & Stadler, 2014; Mitchell, Etienne, & Rabosky, 2018; Stadler, Gavryushkina, Warnock, Drummond, & Heath, 2018). However, available models necessarily make simplifying assumptions about processes leading to the generation of palaeontological data. Incorporating more complex models of fossil recovery and species evolution will be essential in developing these methods further.

A key feature of the fossil record is the incomplete and non-uniform nature of preservation and sampling, which varies over time, across space and among taxa. This aspect of palaeontological data must be considered in the interpretation of events in deep time (Holland, & Patzkowsky, 2015) but has rarely been considered in constructing models in phylogenetics. The interpretation of species and speciation in the fossil record is also not straightforward and can impact macroevolutionary parameter estimates (Ezard, Pearson, Aze, & Purvis, 2012). Similarly, the role of different speciation processes in phylogenetics and their combined interaction with fossil recovery processes remains relatively unexplored (Bapst, 2013).

Improving models in macroevolution and phylogenetics requires assessing the performance of different methods under a wide range of speciation, preservation and sampling scenarios. This assessment

is only possible in datasets where the true underlying parameters are known, thus simulations play a critical role in the exploration and refinement of inference tools. Simulations are valuable for validating software implementations, assessing model adequacy and establishing the limitations of existing methods. Indeed, simulation studies have already played an important role in demonstrating the impact of palaeontological data on parameter inference in both phylogenetics (e.g. Heath et al., 2014; Warnock, Yang, & Donoghue, 2017) and palaeobiology (e.g. Soul, & Friedman, 2017; Smiley, 2018).

Here, we present an R package FossilSim that can be used to simulate fossil and taxonomy data under a wide range of mechanistic models in a format that can be integrated easily with existing tools. The package also provides visualisation functions that allow the user to summarise information about tree topology, taxonomy and stratigraphic data in a single plot.

1.1 | Models of speciation and taxonomy

We use the term 'taxonomy' to describe the interspecific relationships among simulated fossil and extant samples obtained using any combination of the speciation modes outlined in the following text and Figure 1. This is intended to emulate the process by which a taxonomist may divide lineages into units of morphotaxa, each one representing intervals during which characters of the greatest diagnostic value appear invariant.

Species taxonomy can be generated using FossilSim for any fully resolved timescaled phylogenetic tree. Although the number of co-existing lineages is informative about the total number of species at a given moment in time, a phylogenetic tree object does not typically contain information about how individual species relate to the underlying branching process. FossilSim can be used to model the speciation process along a tree and incorporates three possible modes of speciation shown in Figure 1: budding, bifurcation and anagenesis. A budding or asymmetric speciation event gives rise to one new species and does not result in the extinction of the ancestor. A bifurcation or symmetric speciation event gives rise to two new species and results in the extinction of the ancestor. At each branching event in a phylogenetic tree, bifurcation speciation occurs with probability β. If $\beta = 0$ all speciation occurs via budding and if $\beta = 1$ all speciation occurs via bifurcation. Anagenetic speciation occurs along each branch in a phylogenetic tree with rate λ_a.

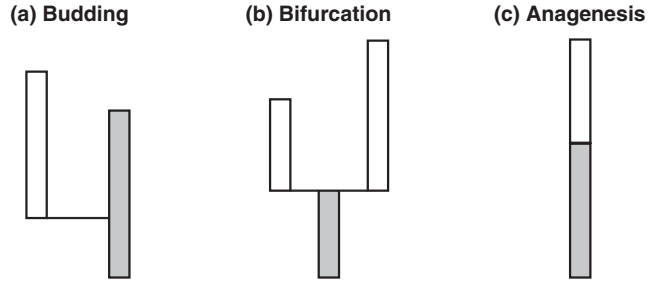

FIGURE 1 Three different modes of speciation that can be simulated using FossilSim. The shaded area represents the ancestral species, while the white area shows the descendant species. Budding and bifurcation may also be referred to as asymmetric and symmetric speciation

Cryptic speciation can also be modelled: in this case, ancestor and descendant species cannot be distinguished and FossilSim will record both the true and apparent species identity. At each speciation event, cryptic speciation occurs with probability κ. This mixed model of speciation (or taxonomy) is described in Bapst (2012) and Stadler et al. (2018).

We note that this model of speciation and the FossilSim package in general treats taxonomic and species units as equivalent, and treats speciation and extinction as point processes, reflecting the assumptions made by models available for the analysis of fossil data. Simulation and inference tools incorporating protracted speciation are not yet widely available but users should always consider the implications of the assumptions being made about the diversification process.

1.2 | Models of fossil recovery

1.2.1 | Constant and time-dependent fossil recovery

The simplest model is a Poisson process, where each species has a constant fossil recovery rate ψ, which corresponds to the exponential waiting times between fossil sampling events (Stadler, 2010). Under the time-dependent model, fossil recovery varies in a piecewise manner across discrete geological intervals. These intervals are specified by the user and can be of even or uneven lengths. During a given interval *i*, each species has a constant rate of fossil recovery ψ_i (Gavryushkina, Welch, Stadler, & Drummond, 2014). The number of fossils sampled during each interval will be drawn from a Poisson distribution with rate ψ_i. Alternatively, per-interval fossil recovery can be described using probabilities: during a given interval *i*, each species has a probability $P_{\text{collection}_i}$ of being sampled. When using probabilities, at most one fossil per species will be sampled during each interval.

1.2.2 | Environment-dependent fossil recovery

The environmental processes that control the spatial and temporal distribution of a living species also drive patterns of fossil recovery. Holland (1995) described a model of fossil recovery based on the abundance of species relative to an environmental or ecological gradient (i.e. a species response curve). The probability of sampling a fossil of species *i*, $P_{\text{collection}_i}$, is described using a symmetric unimodal species response curve with the following Gaussian distribution:

$$P_{collection_i}(e) = PA e^{-(e-PE)^2/2ET^2}, \quad (1)$$

where *e* is the current environment (i.e. the position along the gradient) of species *i*, *PE* is the species' preferred environment, *ET* is its environmental tolerance and *PA* its peak abundance. *PE*, *ET* and *PA* give respectively the mean, the standard deviation and the amplitude of the distribution. Changes in environment (*e*) can be modelled using any user defined function. The model can be applied to any environmental or depositional context, in which a measure of environmental gradient

can be provided (Patzkowsky, & Holland, 2012) but has typically been applied in a marine context. In this context, environmental changes are often linked to relative water depth.

1.2.3 | Lineage- and trait-dependent fossil recovery

Two models are available to generate variation in trait values associated with fossil recovery across species assuming traits are either autocorrelated or independent between ancestor and descendant species. These models are analogous to the relaxed molecular models with the same name. Under the autocorrelated trait values model, values evolve along lineages according to a Brownian motion process, where the strength of the relationship between ancestor and descendant values is determined by the parameter ν. If ν is small, values will be more similar between ancestor and descendants, and if ν is zero all values will be equal. For a given species i that has an ancestor with trait value κ_i', a new trait value κ_i is drawn from a lognormal distribution with $\kappa_i = LN(\ln[\kappa_i'] - \frac{\sigma^2}{2}, \sigma)$, where the standard deviation $\sigma = \nu t_i$ and t_i is the duration of the species (Thorne, & Kishino, 2002; Heath et al., 2014). Under the independent trait values model, values are drawn from a user-specified distribution and assigned independently to each species (Drummond, Ho, Phillips, & Rambaut, 2006). In addition, we incorporate the parameter P_{change}, which is the probability that there will be a change from the ancestral trait value at each speciation event.

Lineage-dependent models of fossil recovery can be coupled with both the time-homogenous Poisson process and depth-dependent models of fossil recovery. Lineage-specific trait values can be assigned to the ψ, *PE*, *ET* and *PA* parameters.

2 | FUNCTIONALITY

2.1 | Simulation

FossilSim stores the simulated taxonomy and fossils as custom objects, taking advantage of the class system in the R programming language. These objects can also be created from a dataframe containing the required information, which allows the user to easily import pre-existing datasets.

In Figure 2, we distinguish between five important ages that may be used in reference to a given fossil species (Holland, & Patzkowsky, 2002). The speciation (or origination) and extinction times reflect the endpoints of the true species duration, and fossil occurrences are sampled within this interval. The oldest and youngest fossil occurrences reflect the endpoints of the sampled duration (i.e. the first and last appearance times). Speciation and extinction times can be calculated from the `taxonomy` object using the functions `species.start` and `species.end`. Fossil occurrence times as stored as part of the `fossils` object, as `hmax` and `hmin`, which represent the oldest and youngest age of the sampling horizon, respectively. The oldest and youngest occurrence ages can be calculated from the information provided by the `fossils` object.

2.1.1 | Taxonomy objects

The `taxonomy` object is a dataframe associated with a tree, which records the positions of species on the topology as well as information about each species. It contains one row per unique combination of edge and species, that is, the same species can appear multiple times in the case of budding speciation and the same edge can appear multiple times in the case of anagenetic speciation. Each row has the following information:

- corresponding species (`sp`) and edge (`edge`)
- parent (`parent`) of the species
- mode of speciation that generated the species (`mode`): anagenetic (`a`), budding or asymmetric (`b`), bifurcating or symmetric (`s`)
- start (`start`) and end (`end`) time of the species along the edge
- optional: whether the species is cryptic (`cryptic`) and if true which species it is identified as (`cryptic.id`)

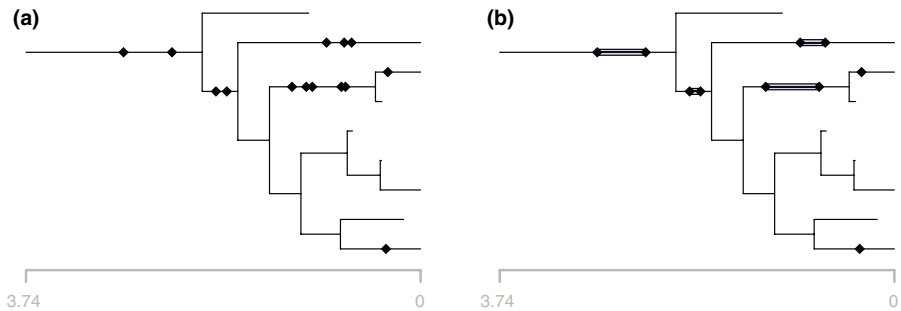

FIGURE 2 FossilSim plot illustrating the relationship between the five ages associated with fossil species described in (Holland, & Patzkowsky, 2002): speciation, extinction, fossil occurrences ages, oldest occurrence (first appearance) and youngest occurrence (last appearance). A simulated tree is shown with each branch representing a separate species; the start and end points of each branch represent the speciation and extinction times, respectively. Each fossil occurrence is represented by a diamond. In (a) all simulated fossil occurrences are shown. In (b) only the oldest and youngest occurrences are shown (i.e. the stratigraphic ranges). Blue bars represent the interval between the oldest and youngest occurrence. If a species is sampled once (i.e. the species is a singleton) only a single occurrence is shown. If the age of fossil occurrences is not known precisely, each occurrence will also be associated with a minimum and maximum age

2.1.2 | Fossil objects

The `fossils` object is a dataframe that records a series of fossil occurrences and contains the following information for each fossil:

- position of the occurrence on the tree topology (`edge`)
- species to which the sample belongs (`sp`)
- youngest (`hmin`) and oldest (`hmax`) age associated with the occurrence

Fossil simulation functions will by default record the true simulated age of the fossils, in which case `hmin` and `hmax` are equal to this true age. To better represent the dating uncertainty associated with the fossil recovery process, fossils can also be simulated using user-defined stratigraphic intervals, in which case only the minimum and maximum age associated with each occurrence will be recorded.

2.1.3 | Simulation functions

FossilSim can be used to simulate data for any user-specified fully resolved and dated phylogeny (simulated or empirical). `Taxonomy` objects are simulated from a tree topology in APE (Paradis, Claude, & Strimmer, 2004) format under the mixed model of speciation described above. `Fossils` objects are simulated from a tree topology and/or a taxonomy object under the models of fossil recovery described above. The package also contains several functions that can simulate trees, taxonomy and fossils in an integrated way.

2.2 | Interaction with R packages & other software

PALEOTREE (Bapst, 2012) is a package for simulating fossil data and timescaling phylogenetic trees. PALEOTREE simulates jointly the tree, taxonomy and fossil record, while these are separate functions in our package. This modularity allows the user to easily replace one component of the simulation with their own model. By allowing the user to specify the starting tree, FossilSim creates the opportunity to explore expected distributions of fossil occurrences under different taxonomic and sampling scenarios for empirical phylogenies. In addition, we provide a range of alternative models for simulating fossil recovery. On the other hand, PALEOTREE offers simulation models that are not currently supported by FossilSim and can filter the simulation results under multiple conditions. These conditions can only be achieved with FossilSim by using rejection sampling, although the starting tree can be conditioned on age and/or number of taxa when produced using TreeSim. Thus, we intend our simulation and plotting functions to be complementary to those provided by PALEOTREE and provide functions for converting between FossilSim and PALEOTREE objects. PALEOTREE was also used to validate FossilSim functions (see Supplementary Material).

BEAST2 (Bouckaert, et al., 2014) is a popular Bayesian inference software for phylogenetic analyses. Its add-on Sampled Ancestors package (Gavryushkina, et al., 2014) is designed to handle datasets containing fossil samples, and in particular datasets where fossils are present along the lineages of the tree and not simply treated as tips. To maintain a strictly binary structure for the tree, these sampled ancestors are stored as tips at the end of zero-length branches. FossilSim implements this format in the `SAtree` class and provides functions to convert simulated trees and fossils into this format. This allows datasets simulated by FossilSim to be used for validation or exploration of the Sampled Ancestors package. Trees formatted as `SAtree` objects are also suitable for use in the statistical computing language RevBayes (Höhna, et al., 2016).

2.3 | Data visualisation

These functions require either a tree object in APE format or a taxonomy object. If the tree is provided without taxonomy, all speciation events will be assumed to be symmetric bifurcations. When plotting trees in `SAtree` format, all speciation events are assumed to be asymmetric.

2.4 | Examples

In this section we show examples of the data and plots that can be produced using FossilSim. More information about the functions and options available can be found in the package documentation, including extensive vignettes. All examples require the package to be loaded as follows:

```
> library(FossilSim)
```

2.4.1 | Example 1: simulating taxonomy under mixed speciation

Figure 3a shows a complete tree with its full taxonomy record: each colour represents a different species and speciation events are labelled by type. This plot was produced by the following commands:

```
> t = TreeSim::sim.bd.taxa(n = 8, numbsim = 1,
lambda = 1, mu = 0.3)[[1]]
> s = sim.taxonomy(tree = t, beta = 0.5, lambda.a = 1)
> plot.taxonomy(s, tree = t)
```

2.4.2 | Example 2: simulating fossils under mixed speciation and constant fossil recovery

Figure 3b shows a complete tree with fossil occurrences where the fossils and stratigraphic ranges are colour-coded by the species to which they belong. We can see both anagenetic and budding speciation events appearing in the fossil record. This plot was produced by the following commands:

```
> t = TreeSim::sim.bd.taxa(n = 8, numbsim = 1,
lambda = 1, mu = 0.3)[[1]]
> s = sim.taxonomy(tree = t, beta = 0.5, lambda.a = 1)
> f = sim.fossils.poisson(rate = 3, taxonomy = s)
> plot.fossils(f, tree = t, taxonomy = s,
show.taxonomy = TRUE, show.ranges = TRUE)
```

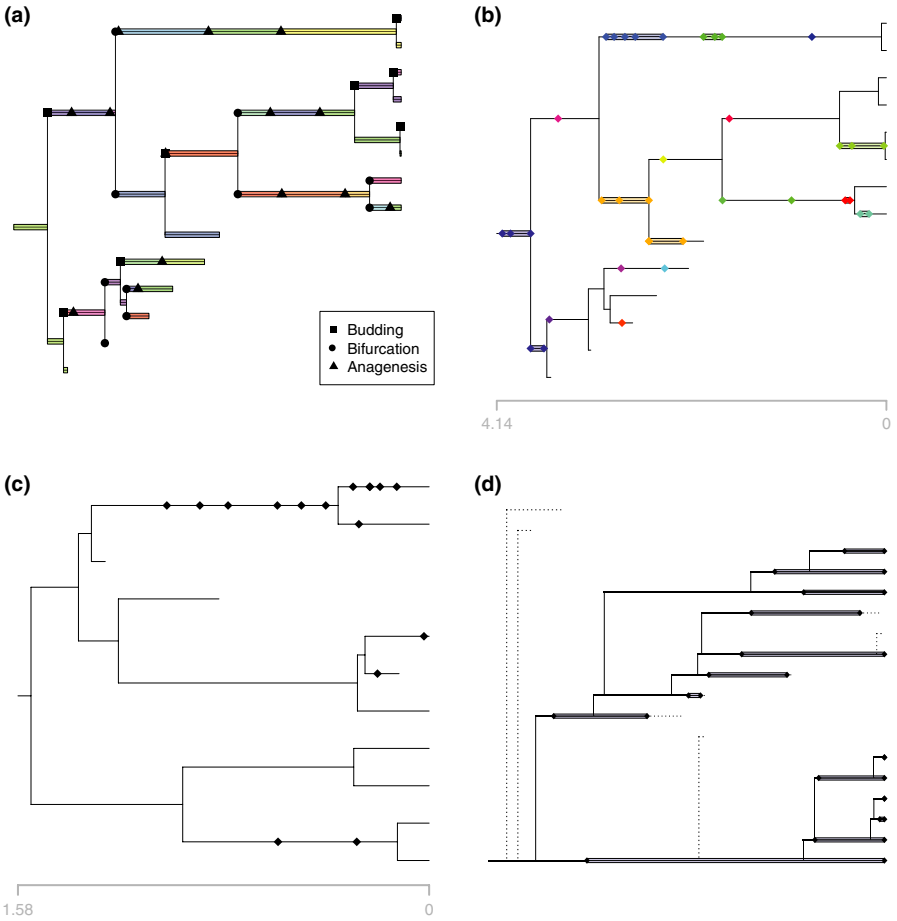

FIGURE 3 Example output produced by FossilSim. (a) Taxonomy plot produced by FossilSim, showing each species as a colour. The symbol at the start of each species range marks the speciation mode (budding, bifurcation or anagenesis) which gave rise to that species. Note that in case of budding speciation, the range of a given species can cover several edges. (b) Fossil and taxonomy plot produced by FossilSim. Each diamond represents a different fossil occurrence, coloured by the species it belongs to. When multiple samples are present for a given species, the sampled range is shown in the same colour. (c) Fossil occurrences simulated under a lineage-dependent model of fossil recovery. (d) Asymmetric plot of a fully budding tree, showing the sampled range of each species. Dashed lines represent unsampled lineages

2.4.3 | Example 3: lineage-dependent fossil recovery

Figure 3c shows a complete tree with fossil occurrences simulated under lineage-dependent fossil recovery. The function `sim.trait.values` is used to simulate fossil recovery rates under the independent trait values model. In this example, the probability that trait values change at each speciation event is 0.5 and new values are drawn from an exponential distribution with mean = 3. If a tree object is passed to the simulation function, rather than a taxonomy object, the functions assume that all speciation occurs via bifurcation.

```
> t = TreeSim::sim.bd.taxa(n = 8, numbsim = 1,
lambda = 1, mu = 0.3)[[1]]
> dist = function() { rexp(1, 1/4) }
> rates = sim.trait.values(init = 1, tree = t,
model = "independent",
 dist = dist, change.pr = 0.5)
> f = sim.fossils.poisson(t, rate = rates)
```

2.4.4 | Example 4: plotting non-bifurcating trees

Figure 3d shows a non-bifurcating representation for asymmetric speciation events. This representation is not currently compatible with mixed speciation and can only be used for fully budding trees. It was produced by the following commands:

```
> t = sim.fbd.taxa(n = 10, numbsim = 1,
lambda = 3, mu = 2, psi = 1,
          complete = TRUE)[[1]]
> rangeplot.asymmetric(t, complete = TRUE)
```

3 | CONCLUSIONS

A large number of R packages have been developed for simulating and analysing phylogenetic data (a comprehensive list is compiled by the CRAN R project https://cran.rproject.org/web/views/Phylogenetics.html). Many available packages use the APE `phylo` class. Since FossilSim also uses this framework, the output can easily be combined with data simulated using a wide range of phylogenetic packages, including molecular sequences (e.g. PhyloSim, Sipos *et al.*, 2011) or morphological character data (e.g. GEIGER, Pennell, et al., 2014). A much smaller number of packages exist for simulating palaeontological data (e.g. PALEOTREE, Bapst, 2012). By providing a package that combines flexible, mechanistic models of fossil recovery that can output data in a range of formats, we expand the scope for exploring the performance of methods in macroevolution.

ACKNOWLEDGEMENTS

We thank David Bapst, Alexei Drummond, Steve Holland and one anonymous reviewer for comments that helped improve the manuscript and package code. R.C.M.W. was funded by the ETH Zürich Postdoctoral Fellowship and Marie Curie Actions for People COFUND programme. JBS was funded by the European Research Council under the Seventh Framework Programme of the European Commission (PhyPD: grant agreement number 335529).

AUTHORS' CONTRIBUTIONS

J.B.-S. and R.C.M.W. designed the package and led the writing of the manuscript. All authors contributed to the package implementation and to the drafts, and gave final approval for publication.

DATA ACCESSIBILITY

The package FossilSim is available on CRAN (https://CRAN.Rproject.org/package=FossilSim) and on GitHub (https://github.com/fossilsim/fossilsim). The code and data used to validate the package and create the figures are archived on GitHub (https://doi.org/10.5281/zenodo.2579184).

REFERENCES

Bapst, D. W. (2013). When can clades be potentially resolved with morphology? *Plos One*, *8*, e62312. https://doi.org/10.1371/journal.pone.0062312

Bapst, D. W. (2012). paleotree: an R package for paleontological and phylogenetic analyses of evolution. *Methods in Ecology and Evolution*, *3*, 803–807. https://doi.org/10.1111/j.2041-210x.2012.00223.x

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., Suchard, M. A., Rambaut, A. & Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *10*, e1003537. https://doi.org/10.1371/journal.pcbi.1003537

Drummond, A., Ho, S., Phillips, M. & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, *4*, e88. https://doi.org/10.1371/journal.pbio.0040088

Ezard, T. H. G., Pearson, O. N., Aze, T. & Purvis, A. (2012). The meaning of birth and death (in macroevolutionary birth–death models). *Biol Lett*, *8*, 139–142. https://doi.org/10.1098/rsbl.2011.0699

Gavryushkina, A., Welch, D., Stadler, T. & Drummond, A. J. (2014). Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLOS Computational Biology*, *10*, 1–15. https://doi.org/10.1371/journal.pcbi.1003919

Heath, T. A., Huelsenbeck, J. P. & Stadler, T. (2014). The fossilized birth-death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences*, *111*, E2957–E2966. https://doi.org/10.1073/pnas.1319091111

Höhna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore, B. R., Huelsenbeck, J. P. & Ronquist, F. (2016). Revbayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology*, *65*, 726–736. https://doi.org/10.1093/sysbio/syw021

Holland, S. M. (1995). The stratigraphic distribution of fossils. *Paleobiology*, *21*, 92–109. https://doi.org/10.1017/s0094837300013099

Holland, S. M. & Patzkowsky, M. E. (2002). Stratigraphic variation in the timing of first and last occurrences. *Palaios*, *17*, 134–146.

Holland, S. M. & Patzkowsky, M. E. (2015). The stratigraphy of mass extinction. *Palaeontology*, *58*, 903–924. https://doi.org/10.1111/pala.12188

Mitchell, J. S., Etienne, R. S. & Rabosky, D. L. (2018). Inferring diversification rate variation from phylogenies with fossils. *Systematic Biology*, *68*, 1–18. https://doi.org/10.1093/sysbio/syy035

Paradis, E., Claude, J. & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*, 289–290. https://doi.org/10.1093/bioinformatics/btg412

Patzkowsky, M. E. & Holland, S. M. (2012). *Stratigraphic paleobiology: understanding the distribution of fossil taxa in time and space*. Chicago, IL: University of Chicago Press.

Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., FitzJohn, R. G., Alfaro, M. E. & Harmon, L. J. (2014). geiger v2. 0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics*, *30*, 2216–2218. https://doi.org/10.1093/bioinformatics/btu181

Sipos, B., Massingham, T., Jordan, G. E. & Goldman, N. (2011). PhyloSim-Monte Carlo simulation of sequence evolution in the R statistical computing environment. *BMC bioinformatics*, *12*, 104. https://doi.org/10.1186/1471-2105-12-104

Smiley, T. M. (2018). Detecting diversification rates in relation to preservation and tectonic history from simulated fossil records. *Paleobiology*, *44*, 1–24. https://doi.org/10.1017/pab.2017.28

Soul, L. C. & Friedman, M. (2017). Bias in phylogenetic measurements of extinction and a case study of end-permian tetrapods. *Palaeontology*, *60*, 169–185. https://doi.org/10.1111/pala.12274

Stadler, T. (2010). Sampling-through-time in birth-death trees. *Journal of Theoretical Biology*, *267*, 396–404. https://doi.org/10.1016/j.jtbi.2010.09.010

Stadler, T., Gavryushkina, A., Warnock, R. C. M., Drummond, A. J. & Heath, T. A. (2018). The fossilized birth-death model for the analysis of stratigraphic range data under different speciation modes. *Journal of Theoretical Biology*, *447*, 41–55. https://doi.org/10.1016/j.jtbi.2018.03.005

Thorne, J. L. & Kishino, H. (2002). Divergence time and evolutionary rate estimation with multilocus data. *Systematic Biology*, *51*, 689–702. https://doi.org/10.1080/10635150290102456

Warnock, R. C., Yang, Z. & Donoghue, P. C. (2017). Testing the molecular clock using mechanistic models of fossil preservation and molecular evolution. *Proc R Soc B*, *284*, 20170227. https://doi.org/10.3410/f.727752709.793536398

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Barido-Sottani J, Pett W, O'Reilly JE, Warnock RCM. FossilSim: An R package for simulating fossil occurrence data under mechanistic models of preservation and recovery. *Methods Ecol Evol*. 2019;10:835–840. https://doi.org/10.1111/2041-210X.13170