Received: 10 May 2021 | Accepted: 1 October 2021

DOI: 10.1111/2041-210X.13750

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

APPLICATION

RevGadgets: An R package for visualizing Bayesian phylogenetic analyses from RevBayes

Carrie M. Tribble[1,2] | **William A. Freyman**[3] | **Michael J. Landis**[4] | **Jun Ying Lim**[5] | **Joellë Barido-Sottani**[6] | **Bjørn Tore Kopperud**[7,8] | **Sebastian Höhna**[7,8] | **Michael R. May**[1,2]

[1]Department of Integrative Biology, University of California, Berkeley, CA, USA

[2]University Herbarium, University of California, Berkeley, CA, USA

[3]23andMe, Inc., Sunnyvale, CA, USA

[4]Department of Biology, Washington University in St. Louis, MO, USA

[5]Department of Biological Sciences, National University of Singapore, Singapore City, Singapore

[6]Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA, USA

[7]GeoBio-Center, Ludwig-Maximilians-Universität München, Munich, Germany

[8]Department of Earth and Environmental Sciences, Paleontology & Geobiology, Ludwig-Maximilians-Universität München, Munich, Germany

Correspondence
Carrie M. Tribble
Email: ctribble09@gmail.com

Present address
Carrie M. Tribble, School of Life Sciences, University of Hawai'i at Mānoa, Honolulu, HI, USA

Handling Editor: Samantha Price

Abstract

1. Statistical phylogenetic methods are the foundation for a wide range of evolutionary and epidemiological studies. However, as these methods grow increasingly complex, users often encounter significant challenges with summarizing, visualizing and communicating their key results.
2. We present `RevGadgets`, an `R` package for creating publication-quality figures from the results of a large variety of phylogenetic analyses performed in `RevBayes` (and other phylogenetic software packages).
3. We demonstrate how to use `RevGadgets` through a set of vignettes that cover the most common use cases that researchers will encounter.
4. `RevGadgets` is an open-source, extensible package that will continue to evolve in parallel with `RevBayes`, helping researchers to make sense of and communicate the results of a diverse array of analyses.

KEYWORDS

Bayesian phylogenetics, data visualization, `R`, `RevBayes`

1 | INTRODUCTION

Beyond being a graphical representation of the Tree of Life, phylogenetic trees provide a rigorous basis for a wide range of evolutionary and epidemiological inferences. Phylogenetic methods allow researchers to understand how molecular and morphological traits evolve (Felsenstein, 1985; Harvey & Pagel, 1991; Nei, 1987; Yang, 2014), how lineages disperse over geographical space (Ronquist & Sanmartín, 2011) and how lineages diversify over time (Morlon, 2014), among other evolutionary phenomena. Additionally, phylogenetic methods can be used to inform conservation decisions (Faith, 1992) and are powerful epidemiological tools (Baele et al., 2017; Volz et al., 2013).

Phylogenetic methods are increasingly based on explicit probabilistic models with parameters that describe underlying evolutionary processes. As datasets grow and evolutionary hypotheses become more nuanced, these models necessarily become more complex. `RevBayes` (Höhna et al., 2016) is a Bayesian phylogenetic inference program that was developed to accommodate this increasing complexity and allows users to explore a vast space of phylogenetic models. Models in `RevBayes` are specified as probabilistic graphical models (Höhna et al., 2014), which are graphical representations of the underlying dependencies among parameters (and their corresponding prior distributions), similar to individual Legos being used to build a complex city. Using this graphical modelling framework, users can design customized models and tailor analyses to their particular datasets and research questions. However, this flexibility comes at a cost: because of the nearly infinite variety of possible models (and model combinations) that users can explore in `RevBayes`, the results of these analyses are often challenging to summarize and visualize using standard software. This is a significant limitation for `RevBayes` users because, in addition to being the primary method for reporting results of phylogenetic analyses, graphical summaries are a valuable tool for making sense of scientific results (Tufte, 2001), and for diagnosing modelling and analytical problems (Kerman et al., 2008).

Historically, `RevBayes` users have had to process and plot their results using ad hoc scripts written for each analysis, which imposed a significant barrier to entry for users not familiar with the structure of `RevBayes` output or comfortable with developing their own graphical summaries. To address these challenges, we developed `RevGadgets`. `RevGadgets` is an `R` package (R Core Team, 2020) that adds to the diverse ecosystem of phylogenetic visualization tools—for example, `ape` (Paradis & Schliep, 2019), `Tracer` (Rambaut et al., 2018), `phytools` (Revell, 2012), `ggtree` (Yu et al., 2017), `FigTree` (Rambaut, 2014), `IcyTree` (Vaughan, 2017), among many others—but is specialized for output produced by `RevBayes`. `RevGadgets` serves as a bridge between `RevBayes` analyses and existing tools for phylogenetic data processing and plotting in `R`, especially the `ggtree` package suite, which includes the `ggtree`, `tidytree` and `treeio` packages (Wang et al., 2020; Yu et al., 2017). `RevGadgets` provides tools for plotting summary trees (including summaries of parameters for each branch), ancestral-state estimates and posterior distributions of parameters for a variety of models. Using the general framework of `ggplot2`, the `tidyverse` and associated packages (Wickham, 2011; Wickham et al., 2019), plotting functions return plot objects with default, but customizable, aesthetics. Here, we present five vignettes demonstrating how to use `RevGadgets` to summarize results for a variety of phylogenetic analyses.

2 | PHYLOGENIES

Phylogenies are central to all analyses in `RevBayes`, so accurate and information-rich visualizations of evolutionary trees are critical. In this case study, we demonstrate the tree-plotting functionality of `RevGadgets`, with methods to visualize phylogenies and their associated posterior probabilities, divergence-time estimates and branch-specific parameter estimates.

`RevGadgets` provides paired functions for (1) reading in and processing data and (2) summarizing and visualizing results. For phylogenies, the function `readTrees()` loads trees (either individual trees or sets of trees) in either Newick or `NEXUS` (Maddison et al., 1997) formats, then processes associated branch or node annotations, and finally stores the tree(s) as `treedata` object(s) (as defined by `treeio`; Wang et al., 2020). Users can then visualize the `treedata` object using either `plotTree()` or `plotFBDTree()`, as we demonstrate below. Alternatively, users may choose to write custom plotting code using existing `ggtree` functions.

`RevGadgets` can plot both unrooted and rooted trees and create plots that are compatible with plotting options from `ggtree`. Additionally, `RevGadgets` provides extensive functionality for plotting trees with non-contemporaneous tips, such as those produced by total-evidence analyses under the fossilized birth-death (FBD) process (Heath et al., 2014; Zhang et al., 2016). The fossilized birth-death process (and the related serially sampled birth-death process; Stadler, 2010) produces sampled ancestors (samples that are directly ancestral to another

```
# specify the tree file
file <- "bears.mcc.tre"

# read the tree
tree <- readTrees(paths = file)

# plot the tree
plotFBDTree(tree = tree,
            timeline = TRUE,
            geo_units = "epochs",
            tip_age_bars = TRUE,
            node_age_bars = TRUE,
            age_bars_colored_by = "posterior",
            label_sampled_ancs = TRUE) +
  ggplot2::theme(legend.position=c(0.05, 0.55))
```

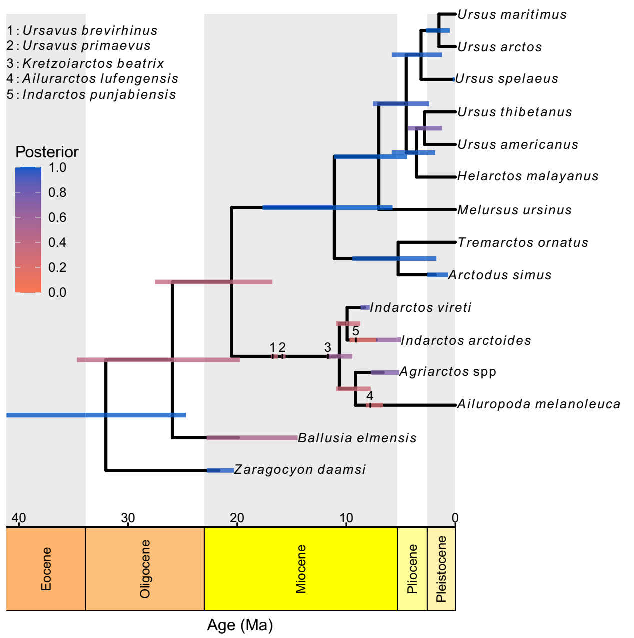


FIGURE 1 Plotting a time-calibrated phylogeny of extinct and extant taxa. (Top) `RevGadgets` code for reading in and plotting a time-calibrated phylogeny of extant and extinct bears. We use the `theme()` function from `ggplot2` to add the posterior-probability legend. (Bottom) The maximum sampled-ancestor clade-credibility tree for the bears. Sampled ancestors are indicated by numbers along the branches (legend, top left). Bars represent the 95% credible interval of the age of the node, tip or sampled ancestor in millions of years (geological timescale, *x*-axis); the colour of the bar corresponds to the posterior probability (legend, middle left) that a clade exists, the posterior probability that a fossil is a sampled ancestor, or the posterior probability that a tip is not a sampled ancestor. (Data from Abella et al., 2012; Heath et al., 2014.)

sampled taxon and thus are not represented as tips in the tree), and the ages of the samples are often subject to uncertainty (e.g. because of imperfect knowledge about the age of the strata from which the samples were collected). As a consequence, conventional tree plotting tools are unsuitable for plotting FBD trees. We demonstrate how to use `RevGadgets` to plot the results of an FBD analysis of living and extinct bears (Figure 1; data from Abella et al., 2012; Heath et al., 2014). We include age bars coloured by the posterior probability of the corresponding node, a geological time-scale and labelled epochs from the package `deeptime` (Gearty, 2021), and fossils estimated to be direct ancestors of other samples (i.e. sampled ancestors).

In addition to visualizing trees themselves, `RevGadgets` allows researchers to visualize branch-specific parameters, for example rates of evolution or diversification for each branch in the phylogeny. In Figure 2, we demonstrate how to use `plotTree()` to visualize the estimated optimal body size as it varies across the cetacean phylogeny under a relaxed Ornstein–Uhlenbeck process (Butler & King, 2004; Uyeda & Harmon, 2014; data from Slater et al., 2010; Steeman et al., 2009). Under this model, a quantitative character evolves towards an adaptive optimum that changes along the branches of the tree, and thus the optimum associated with each branch is a focal inference.

The `plotTree()` function can also visualize unrooted or circular phylogenies, and users may add text annotations to denote posterior probabilities or other quantities. Users can apply `ggtree` functions to modify the `RevGadgets` plot, for example, to highlight certain clades with `geom_hilight()` or to add phylopics (http://phylopic.org/) using `geom_phylopic()`. Together, these functions provide user-friendly and customizable tree-plotting functionality for a variety of core research questions in evolutionary biology.

3 | POSTERIOR ESTIMATES OF NUMERICAL PARAMETERS

`RevGadgets` provides several tools to visualize posterior distributions of numerical parameters. The output produced by most `RevBayes` analyses is a (typically tab-delimited) text file where rows correspond to samples from sequential iterations of an MCMC analysis, and columns correspond to parameters in the model. Most information of interest to researchers—for example, most probable parameter values (maximum a posteriori, or MAP, estimates), 95% credible intervals (CIs) or full posterior distributions—requires processing this raw MCMC output. Here, we demonstrate methods for processing and visualizing MCMC output for both quantitative and qualitative parameters.

We illustrate the core functions for reading, summarizing and visualizing posterior distributions of specific parameters with an example analysis of chromosome number evolution (Figure 3; data from Freyman & Höhna, 2018). We use `readTrace()` to read in parameters sampled during one or more MCMC analyses. We then use `summarizeTrace()` to calculate posterior point estimates and the 95% credible interval for the focal parameters. Finally, we plot the marginal posterior distributions of the focal parameters using `plotTrace()`.

```
# specify the annotated tree file
file <- "relaxed_OU_MAP.tre"

# read the tree
tree <- readTrees(paths = file)

# plot the tree
plotTree(tree = tree,
         color_branch_by = "branch_thetas",
         line_width = 1.7) +
  ggplot2::theme(legend.position=c(0.1, 0.9))
```

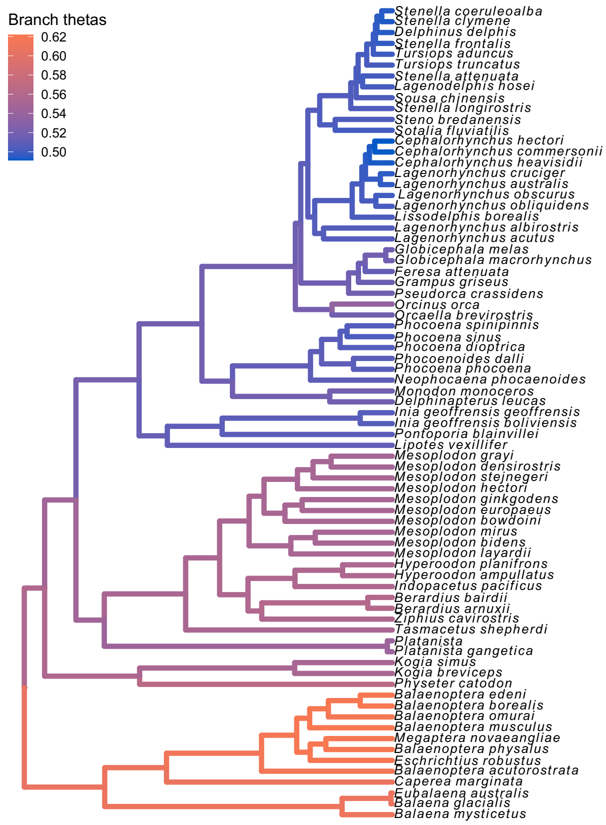


FIGURE 2 Plotting branch-specific parameter values across a phylogeny. (Top) `RevGadgets` code for reading in and plotting the cetacean phylogeny that has been annotated with branch-specific adaptive optima (θ) inferred under a relaxed Ornstein–Uhlenbeck model. (Bottom) The cetacean phylogeny with branches coloured according to the posterior-mean estimate of the inferred branch-specific optimum body size, θ (legend, top left). (Phylogeny from Steeman et al., 2009; body-size data in units of natural log-transformed meters from Slater et al., 2010.)

Plots of the posterior distributions of parameter values are key to a thorough understanding of the results of any Bayesian analysis. These tools encourage users to explore their results thoroughly rather than relying on single summary statistics. These summaries and plots may also be useful as tools for science communication and education on statistical phylogenetics, as they can easily be used to demonstrate differences in parameter estimates that result from changes to basic phylogenetic models. Additionally, the output of `readTrace()` may be passed to `R` packages specializing in MCMC diagnosis, for example, `convenience`

```
# specify the log files with rates of
# chromosome evolution
files <- c("chromevol_simple.log")

# read the trace
trace <- readTrace(path = files)

## Reading in log file 1

# summarize the traces of the parameters
# gamma: chromosome gain rate
# delta: chromosome loss rate
summary <- summarizeTrace(
    trace = trace,
    vars  = c("gamma", "delta"))

# report the summary for the gain rate
summary[["gamma"]]

## $trace_1
##          mean        median           MAP
##    0.28853914    0.23994560    0.13586156
##   quantile_2.5 quantile_97.5
##    0.03347161    0.80123940

# plot the posteriors of each parameter
plotTrace(trace = trace,
    vars  = c("gamma", "delta"))[[1]]
```

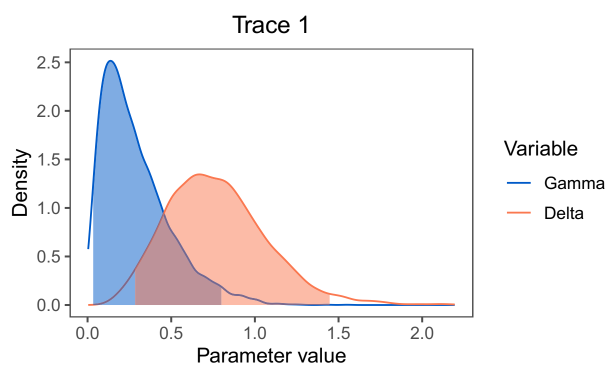


FIGURE 3 Plotting posterior distributions of numerical parameter values. (Top) `RevGadgets` code for reading in and plotting the posterior distributions of rates of chromosome evolution in *Aristolochia*. (Bottom) Marginal posterior distributions of the two rate parameters. Shaded regions represent the 95% credible interval of each posterior distribution. (Data from Freyman & Höhna, 2018.)

(Fabreti & Höhna, 2021) or `coda` (Plummer et al., 2006). These functions are compatible with any delimited text file of MCMC samples, and can be used with the output of most Bayesian phylogenetic programs.

4 | ANCESTRAL-STATE ESTIMATES

In addition to making inferences about the underlying process of evolution, researchers may be interested in studying how particular characters evolved across the branches of the phylogeny. Ancestral-state estimation is a method for inferring that history.

`RevGadgets` offers two different types of summaries for ancestral-state estimates: (1) maximum a posteriori (MAP) estimates, that is, the state with the highest posterior probability at each node and (2) pie charts that represent each state in proportion to its probability at each node. Ancestral-state estimates may be represented as text annotations rather than coloured symbols. Additionally, `RevGadgets` can summarize and visualize ancestral-state estimates at internal nodes and at the 'shoulders', that is, at the beginning of each branch. Plotting the states at internal nodes is appropriate for standard evolutionary models of anagenetic (within-lineage) change. However, models of evolution that include a cladogenetic component (e.g. models of biogeographical or chromosome-number evolution; Freyman & Höhna, 2018; Goldberg & Igić, 2012; Ree & Smith, 2008) also allow states to change at speciation events. In this case, researchers may also want to plot the shoulder states, which represent the ancestral-state estimates for each daughter lineage immediately following the speciation event.

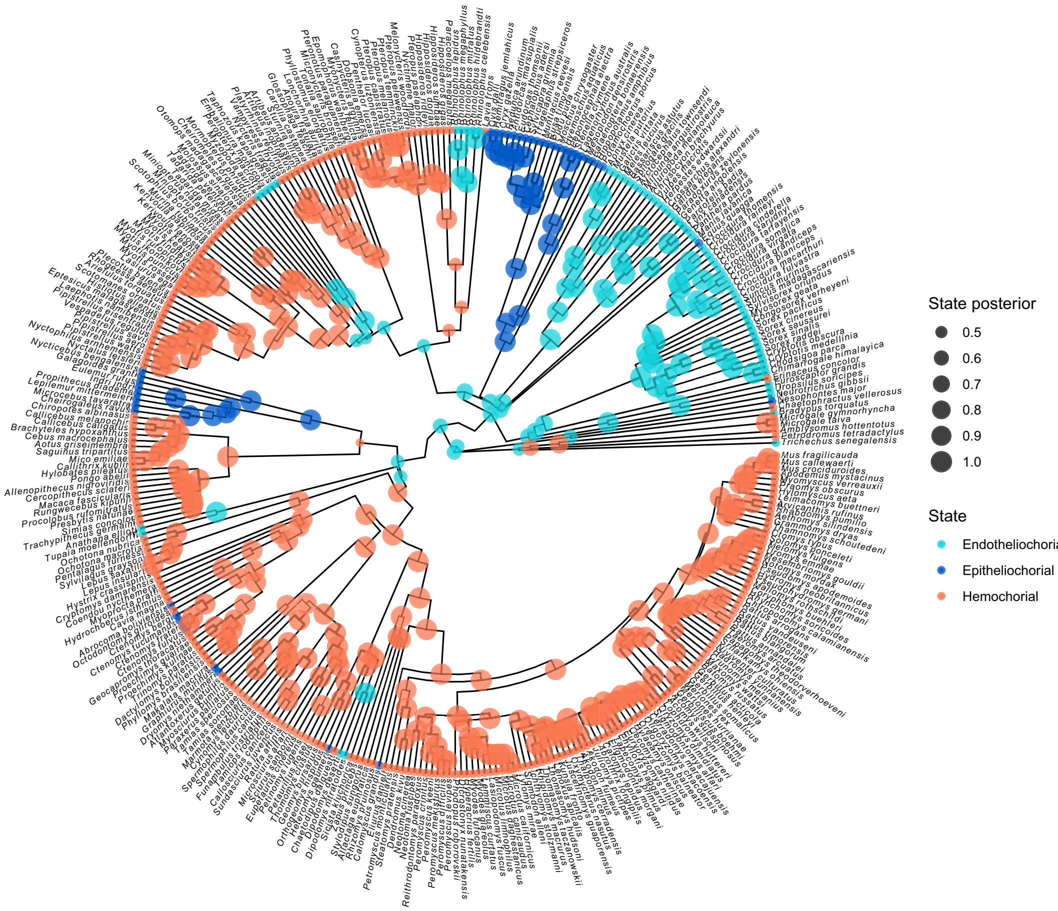
We demonstrate how to plot ancestral-state estimates of placenta type across the mammal phylogeny under an asymmetric model of character evolution (Figure 4; data from Elliot & Crespi, 2006). First, we use `processAncStates()` to read in and parse the phylogeny and ancestral-state estimates inferred using `RevBayes`. Second, we use `plotAncStatesMAP()` to colour each node symbol according to the state with the highest posterior probability and make the area of the symbol proportional to that state's posterior probability. Because of the size of the phylogeny, we choose to plot the estimates on a circular tree by changing the tree layout parameter.

Next, we demonstrate plotting estimates of ancestral ranges of the Hawaiian silversword alliance that were generated by a dispersal-extinction-cladogenesis (DEC) model (Figure 5; data from Landis et al., 2018). Since the DEC model features a cladogenetic component, we include shoulder-state estimates. Because of the large number of states in this analysis (15 possible ranges and one 'other' category), more pre-processing is necessary. As before, we pass the appropriate state names to `processAncStates()`; however, in this case, we plot pie charts representing the probability of each state using `plotAncStatesPie()` and plot states at shoulders using `cladogenetic = TRUE`.

Beyond the above examples, these versatile plotting tools can visualize any discrete ancestral-state estimates produced by `RevBayes`, including the results of chromosome-count estimates (Freyman & Höhna, 2018) and discrete state-dependent speciation-and-extinction (SSE) models (Freyman & Höhna, 2019; Zenil-Ferguson et al., 2019).

5 | DIVERSIFICATION RATES

The processes of speciation and extinction (i.e. lineage diversification) are of great interest to evolutionary biologists (Morlon, 2014). Rates of speciation and extinction may be modelled as constant

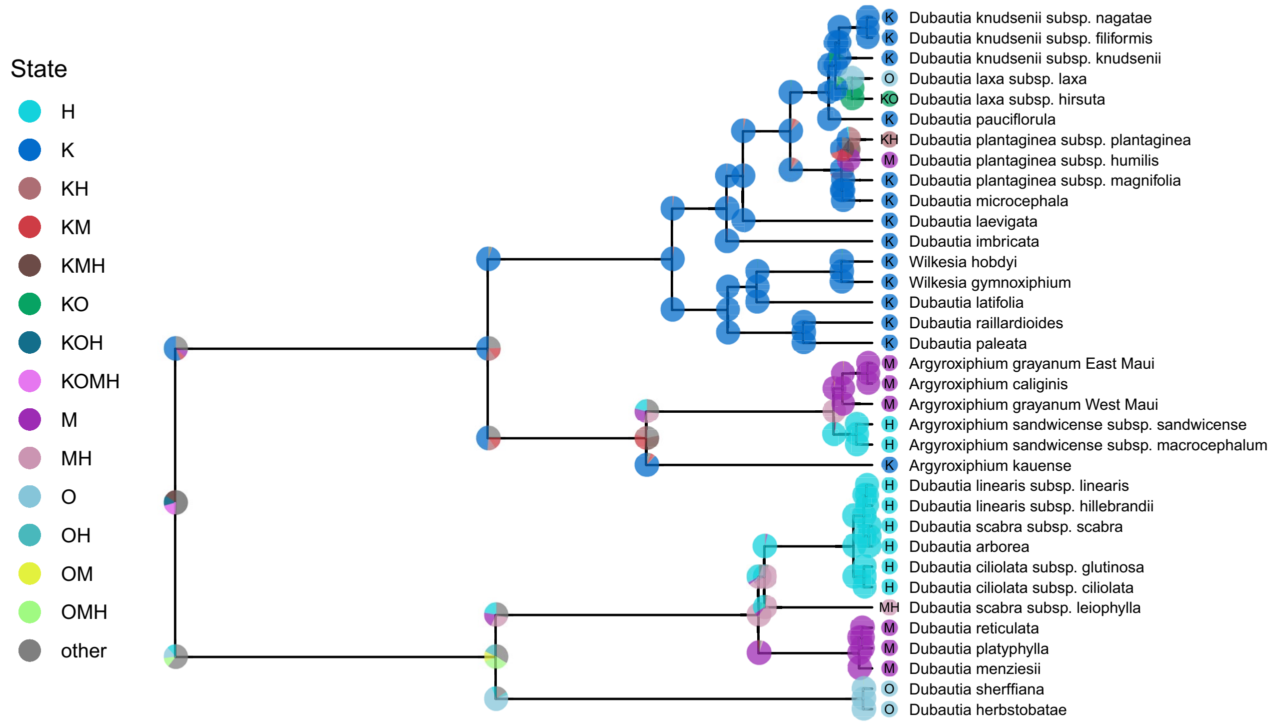



```
# specify the tree file
file <- "ase_freeK.tree"

# read in the tree annotated with ancestral states
freeK <- processAncStates(file, state_labels = c("1" = "Epitheliochorial",
                                                 "2" = "Endotheliochorial",
                                                 "3" = "Hemochorial"))

# plot the tree with MAP ancestral states
plotAncStatesMAP(t = freeK, tree_layout = "circular") +
    ggplot2::theme(legend.position = c(0.96,0.5))
```

FIGURE 4 Plotting maximum a posteriori (MAP) estimates of ancestral states on a circular phylogeny. (Top) MAP estimates of ancestral placental states across the phylogeny of mammals. Each node is coloured by the MAP state (legend, bottom right); the size of each symbol is proportional to the posterior probability of the MAP state (legend, top right). (Bottom) `RevGadgets` code for reading in and plotting the MAP estimates for ancestral placental states across the mammalian phylogeny. (Data from Elliot & Crespi, 2006.)

over time and among branches (as in a constant-rate birth-death process; Kendall, 1948; Nee et al., 1994), or allowed to vary over time (May et al., 2016; Stadler, 2011), across branches of a phylogeny (Höhna et al., 2019; Rabosky, 2014), or based on the character states of the evolving lineages (Freyman & Höhna, 2019; Maddison et al., 2007). For example, rates that vary across branches of the phylogeny can be visualized using `plotTree()` to colour the branches by their inferred rate. State-dependent diversification models provide estimates of the speciation and extinction rates associated with each character state, and may also be used to estimate ancestral states. `plotTrace()` or specific processing and plotting functions for diversification rates—`processSSE()`, `plotMuSSE()`, and `plotHiSSE()`—may be used to visualize the estimated rates. `plotAncStatesMAP()` or `plotAncStatesPie()` may be used to visualize the ancestral-state estimates.

We demonstrate how to plot the results of a time-varying model—the episodic birth-death process (Höhna, 2015; Stadler, 2011)—applied to the primate phylogeny (Figure 6; Springer et al., 2012). The episodic birth-death analysis in `RevBayes` produces separate trace files for each type of rate. We read these




```
# specify the annotated tree file
file <- "simple_ase.tre"

# define the state labels
labs <- c("1"  = "K",   "2"  = "O",   "3"  = "M",  "4"  = "H",  "5"  = "KO",  "6"  = "KM",
          "7"  = "OM",  "8"  = "KH",  "9"  = "OH", "10" = "MH", "11" = "KOM", "12" = "KOH",
          "13" = "KMH", "14" = "OMH", "15" = "KOMH")

# read the annotated tree file
dec_example <- processAncStates(file, state_labels = labs)

# plot the tree with pie charts for ancestral states
plotAncStatesPie(t = dec_example, cladogenetic = TRUE, tip_labels_states = TRUE,
                 tip_label_italics = FALSE, tip_labels_offset = 0.2,
                 tip_pie_nudge_x = 0.15, tip_pie_size = 1.0,
                 node_pie_size = 1.5, tip_labels_states_offset = 0.05) +
    ggplot2::theme(legend.position = c(-0.05, 0.5), plot.margin = unit(c(0,0,0,2),"cm"))
```

FIGURE 5 Plotting posterior distributions of ancestral states under a cladogenetic model. (Top) The posterior estimates of ancestral biogeographical states of the Hawaiian silverswords estimated under a DEC model. The size of each pie slice is proportional to the posterior probability of a given state (legend, top left) for a particular lineage. Pies at nodes represent the state of the ancestral lineage immediately before speciation; pies at 'shoulders' represent the state of each daughter lineage immediately following the speciation event. (Bottom) `RevGadgets` code for reading in and plotting the posterior estimates of ancestral geographical range across the phylogeny of Hawaiian silverswords. (Data from Landis et al., 2018.)

output files using `processDivRates()` and plot the resulting parameter estimates over time using `plotDivRates()`.

Together with the aforementioned functions for plotting diversification parameter estimates, `plotDivRates()` allows users to visualize the outputs of nearly all diversification analyses available in `RevBayes`. Stochastic character mapping of diversification estimates, in which the timing and location of diversification rate shifts are painted along the branches of the tree, will be added in the future (Freyman & Höhna, 2019; Höhna et al., 2019).

6 | MODEL ADEQUACY

In addition to visualizing the results of phylogenetic inferences with a specific model, `RevGadgets` provides tools for exploring the adequacy of the model (i.e. whether the model provides an adequate description of the data-generating process; Bollback, 2002; Brown, 2014; Gelman et al., 2013; Höhna et al., 2018). Posterior-predictive analysis tests whether a fitted model simulates (predicts) data that are similar to the observed data. This process is distinct

```
# specify the log files with diversification
# rates and rate-change times
speciation_time_file <- "speciation_times.log"
speciation_rate_file <- "speciation_rates.log"
extinction_time_file <- "extinction_times.log"
extinction_rate_file <- "extinction_rates.log"

# read the log files
rates <- processDivRates(
    speciation_time_log = speciation_time_file,
    speciation_rate_log = speciation_rate_file,
    extinction_time_log = extinction_time_file,
    extinction_rate_log = extinction_rate_file)

# plot the diversification rates
plotDivRates(rates = rates)
```

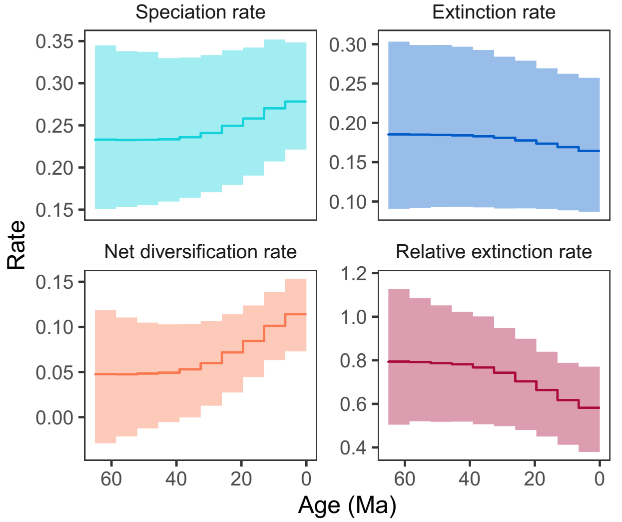


FIGURE 6 Plotting posterior distributions of diversification rates over time. (Top) `RevGadgets` code for reading in and plotting the posterior estimates of diversification rates over time inferred from the primate phylogeny. (Bottom) Posterior distributions of speciation and extinction rates over time, as well as the net-diversification rate (speciation minus extinction) and the relative extinction rate (extinction divided by speciation). Dark lines correspond to the posterior-mean estimate of each parameter for each time interval and shaded regions correspond to the 95% credible interval. (Data from Springer et al., 2012.)

from model testing, in which one model is chosen from a set of possible models, as the best model of the set may still provide an inadequate description of the underlying process.

First, users analyse their data with the model of interest and then use the inferred posterior distribution to simulate a number of new datasets. The user then selects test statistics that describe important features of the data (e.g. the number of invariant sites in a nucleotide alignment) and calculates these statistics for both the observed data and the simulated data. If the statistic from the empirical data is reasonably included within the distribution of statistics from the simulated datasets (posterior-predictive p-value > 0.05), the model is considered an adequate description of the process that produced the tested data feature.

Here, we demonstrate the workflow for a posterior-predictive analysis to test model adequacy of the Jukes–Cantor model for nucleotide-sequence evolution (Jukes & Cantor, 1969) in a single gene across a sample of 23 primates (Figure 7; data from Springer et al., 2012). First, we perform an analysis in `RevBayes` under a Jukes–Cantor model of nucleotide-sequence evolution. Second, we use `RevBayes` to simulate datasets under the joint posterior distribution estimated in the first step. Third, we use `RevBayes` to calculate statistics from the simulated and empirical datasets. These statistics should describe aspects of the data that we hope capture a meaningful aspect of model performance. Finally, we use `RevGadgets` to plot those statistics and compute posterior-predictive p-values.

Despite being computationally inexpensive compared to Bayesian model comparison methods (i.e. Bayes factors), posterior-predictive approaches remain relatively uncommon in empirical phylogenetic studies. As genome-scale datasets and increasingly complex statistical methods become more accessible to researchers, posterior-predictive simulation will be critical for testing how well our models describe the underlying generative processes. This component of `RevGadgets` functionality and the associated clear workflows for performing and interpreting posterior-predictive tests will hopefully increase the adoption of this important tool.

7 | CONCLUSIONS

`RevBayes` is a flexible platform for performing Bayesian phylogenetic evolutionary inferences. Because of the almost endless possibilities for building unique combinations of models in `RevBayes`, these analyses are often challenging to visualize using standard plotting software. We have developed an `R` package, `RevGadgets`, to produce publication-quality visualizations of phylogenetic analyses performed in `RevBayes`. The case studies described above illustrate some of the core functionality available in `RevGadgets` and demonstrate how to produce plots of the most commonly performed `RevBayes` analyses. `RevBayes` is open-source software that is actively maintained and developed. Likewise, `RevGadgets` is also open source and will continue to provide new plotting tools to meet new visualization challenges as they arise. `RevGadgets` and any future updates will be available on CRAN (https://cran.r-project.org/web/packages/RevGadgets/index.html) and on GitHub at https://github.com/revbayes/RevGadgets. Additionally, we provide thorough documentation for all functionality in the package and maintain numerous tutorials demonstrating how to use `RevGadgets` on the `RevBayes` website at https://revbayes.github.io/tutorials/. Together, the modular modelling tools from `RevBayes` and the visualization gadgets in `RevGadgets` will help researchers make sense of and communicate the results of a diverse array of sophisticated phylogenetic analyses.

7.1 | Dependencies

`RevGadgets` depends on many `R` packages, in particular: `ape` (Paradis & Schliep, 2019), `phangorn` (Schliep, 2011), `phytools` (Revell, 2012),

```
# specify the simulated statistics file
sim <- "simulated_data_pps.csv"

# specify the empirical statistics file
emp <- "empirical_data_pps.csv"

# read the statistics files
stats <- processPostPredStats(path_sim = sim,
                              path_emp = emp)

# create the posterior-predictive plots
plots <- plotPostPredStats(data = stats)

# plot some of the statistics
plots[c(1,3,5,7)]
```

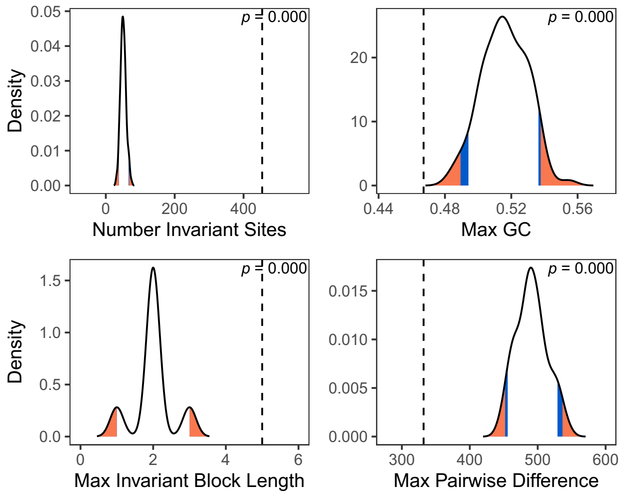


FIGURE 7 Plotting simulated posterior-predictive distributions to assess model adequacy. (Top) `RevGadgets` code for reading in and plotting the distributions of summary statistics generated using posterior-predictive simulation. (Bottom) Posterior-predictive distributions (black curves) of four statistics simulated under the Jukes–Cantor model fit to primate *cytb*, compared to the same statistics computed on the observed data (dashed vertical lines). The posterior-predictive *p*-value (upper right of each panel) is the fraction of simulated statistics that are as or more extreme than the observed statistic. If the observed statistic falls in or beyond the orange region, we deem the model as inadequate at the 5% significance level; if the observed statistic falls in the blue region, the model is marginally adequate at the 10% significance level. In this case, the Jukes–Cantor model provides an inadequate description of the true generating process according to every summary statistic. (Data from Springer et al., 2012.)

`ggplot2` (Wickham, 2011), `ggtree` (Yu et al., 2017), `treeio` (Wang et al., 2020), `deeptime` (Gearty, 2021), `dplyr` (Wickham et al., 2021), `treeplyr` (Uyeda & Harmon, 2020), `tidytree` (Yu, 2021b), `reshape` (Wickham, 2007), `ggthemes` (Arnold, 2021), `tidyr` (Wickham, 2021), `tibble` (Müller & Wickham, 2021), `gginnards` (Aphalo, 2021a), `ggimage` (Yu, 2020), `ggplotify` (Yu, 2021a), `png` (Urbanek, 2013) and `ggpp` (Aphalo, 2021b).

ACKNOWLEDGEMENTS

We would like to acknowledge Carl J. Rothfels, Benjamin K. Blackman, David D. Ackerly and Chelsea D. Specht for feedback on initial stages of this manuscript. Ixchel González Ramírez, Jenna T. B. Ekwealor, Isaac Lichter-Marck and members of the Rothfels Lab at UC Berkeley provided valuable feedback on usability and legibility of figures and code. Klaus Schliep and an anonymous reviewer provided important feedback on the package structure and stability. Andrew Magee, Kengo Nagashima, Klaus Schliep and Josef Uyeda contributed code. This research was supported by the Deutsche Forschungsgemeinschaft (DFG) Emmy Noether Program HO 6201/1-1 awarded to S.H., a National Science Foundation (NSF) award (DEB 2040347) to M.J.L. and an NSF GFRP award to C.M.T.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

AUTHORS' CONTRIBUTIONS

C.M.T. and M.R.M. designed the `R` package. All authors contributed code and examples. C.M.T. and M.R.M. drafted the manuscript. All authors revised and approved the final version of the manuscript.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/2041-210X.13750.

DATA AVAILABILITY STATEMENT

`RevGadgets` is hosted on CRAN (https://cran.r-project.org/web/packages/RevGadgets/index.html) and available on GitHub (https://github.com/revbayes/RevGadgets (Tribble et al., 2021)). All example datasets are freely available on the `RevBayes` website (https://revbayes.github.io/tutorials/intro/revgadgets) and/or are bundled with the `R` package.

ORCID

Carrie M. Tribble https://orcid.org/0000-0001-7263-7885
William A. Freyman https://orcid.org/0000-0002-3463-2044
Michael J. Landis https://orcid.org/0000-0002-8672-6966
Jun Ying Lim https://orcid.org/0000-0001-7493-2159
Joellë Barido-Sottani https://orcid.org/0000-0002-5220-5468
Bjørn Tore Kopperud https://orcid.org/0000-0002-7360-7087
Sebastian Höhna https://orcid.org/0000-0001-6519-6292
Michael R. May https://orcid.org/0000-0002-5031-4820

REFERENCES

Abella, J., Alba, D. M., Robles, J. M., Valenciano, A., Rotgers, C., Carmona, R., Montoya, P., & Morales, J. (2012). *Kretzoiarctos* gen. nov., the oldest member of the giant panda clade. *PLoS ONE*, *7*(11), e48985.
Aphalo, P. J. (2021a). *gginnards: Explore the innards of 'ggplot2' objects*.
Aphalo, P. J. (2021b). *ggpp: Grammar extensions to 'ggplot2'*.
Arnold, J. B. (2021). *ggthemes: Extra themes, scales and geoms for 'ggplot2'*.
Baele, G., Suchard, M. A., Rambaut, A., & Lemey, P. (2017). Emerging concepts of data integration in pathogen phylodynamics. *Systematic Biology*, *66*(1), e47–e65.

Bollback, J. P. (2002). Bayesian model adequacy and choice in phylogenetics. *Molecular Biology and Evolution*, *19*(7), 1171–1180.

Brown, J. M. (2014). Predictive approaches to assessing the fit of evolutionary models. *Systematic Biology*, *63*(3), 289–292.

Butler, M. A., & King, A. A. (2004). Phylogenetic comparative analysis: A modeling approach for adaptive evolution. *The American Naturalist*, *164*(6), 683–695.

Elliot, M. G., & Crespi, B. J. (2006). Placental invasiveness mediates the evolution of hybrid inviability in mammals. *The American Naturalist*, *168*(1), 114–120.

Fabreti, L. G., & Höhna, S. (2021). Convergence assessment for Bayesian phylogenetic analysis using MCMC simulation. *Methods in Ecology and Evolution*, 1–14. https://doi.org/10.1111/2041-210X.13727

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation*, *61*(1), 1–10.

Felsenstein, J. (1985). Phylogenies and the comparative method. *The American Naturalist*, *125*(1), 1–15.

Freyman, W. A., & Höhna, S. (2018). Cladogenetic and anagenetic models of chromosome number evolution: A Bayesian model averaging approach. *Systematic Biology*, *67*(2), 195–215.

Freyman, W. A., & Höhna, S. (2019). Stochastic character mapping of state-dependent diversification reveals the tempo of evolutionary decline in self-compatible Onagraceae lineages. *Systematic Biology*, *68*(3), 505–519.

Gearty, W. (2021). *deeptime: Plotting tools for anyone working in deep time*.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. CRC Press.

Goldberg, E. E., & Igić, B. (2012). Tempo and mode in plant breeding system evolution. *Evolution: International Journal of Organic Evolution*, *66*(12), 3701–3709.

Harvey, P. H., & Pagel, M. D. (1991). *The comparative method in evolutionary biology* (Vol. 239). Oxford University Press.

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 of the United States of America*, *111*(29), E2957–E2966.

Höhna, S. (2015). The time-dependent reconstructed evolutionary process with a key-role for mass-extinction events. *Journal of Theoretical Biology*, *380*, 321–331.

Höhna, S., Coghill, L. M., Mount, G. G., Thomson, R. C., & Brown, J. M. (2018). P3: Phylogenetic posterior prediction in RevBayes. *Molecular Biology and Evolution*, *35*(4), 1028–1034.

Höhna, S., Freyman, W. A., Nolen, Z., Huelsenbeck, J., May, M. R., & Moore, B. R. (2019). A Bayesian approach for estimating branch-specific speciation and extinction rates. *bioRxiv*, 555805.

Höhna, S., Heath, T. A., Boussau, B., Landis, M. J., Ronquist, F., & Huelsenbeck, J. P. (2014). Probabilistic graphical model representation in phylogenetics. *Systematic Biology*, *63*(5), 753–771.

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*(4), 726–736.

Jukes, T. H., & Cantor, C. R. (1969). Evolution of protein molecules. In H. N. Munro (Ed.), *Mammalian protein metabolism* (Vol. 3, 21st ed., p. 132). Academic Press.

Kendall, D. G. (1948). On the generalized 'birth-and-death' process. *The Annals of Mathematical Statistics*, *19*(1), 1–15.

Kerman, J., Gelman, A., Zheng, T., & Ding, Y. (2008). Visualization in Bayesian data analysis. In C.-H. Chen, W. Härdle, & A. Unwin (Eds.), *Handbook of data visualization* (pp. 709–724). Springer.

Landis, M. J., Freyman, W. A., & Baldwin, B. G. (2018). Retracing the Hawaiian silversword radiation despite phylogenetic, biogeographic, and paleogeographic uncertainty. *Evolution*, *72*(11), 2343–2359.

Maddison, D. R., Swofford, D. L., & Maddison, W. P. (1997). NEXUS: An extensible file format for systematic information. *Systematic Biology*, *46*(4), 590–621.

Maddison, W. P., Midford, P. E., & Otto, S. P. (2007). Estimating a binary character's effect on speciation and extinction. *Systematic Biology*, *56*(5), 701–710.

May, M. R., Höhna, S., & Moore, B. R. (2016). A Bayesian approach for detecting the impact of mass-extinction events on molecular phylogenies when rates of lineage diversification may vary. *Methods in Ecology and Evolution*, *7*(8), 947–959.

Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecology Letters*, *17*(4), 508–525.

Müller, K., & Wickham, H. (2021). *tibble: Simple data frames*. R package version 3.1.2.

Nee, S., May, R. M., & Harvey, P. H. (1994). The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *344*(1309), 305–311. https://doi.org/10.1098/rstb.1994.0068

Nei, M. (1987). *Molecular evolutionary genetics*. Columbia University Press.

Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*, 526–528.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, *6*(1), 7–11.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rabosky, D. L. (2014). Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS ONE*, *9*(2), e89543. https://doi.org/10.1371/journal.pone.0089543

Rambaut, A. (2014). *FigTree 1.4.2 software*. Institute of Evolutionary Biology, The University of Edinburgh.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904.

Ree, R. H., & Smith, S. A. (2008). Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, *57*(1), 4–14.

Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217–223.

Ronquist, F., & Sanmartín, I. (2011). Phylogenetic methods in biogeography. *Annual Review of Ecology, Evolution, and Systematics*, *42*(1), 441–464. https://doi.org/10.1146/annurev-ecolsys-102209-144710

Schliep, K. (2011). phangorn: Phylogenetic analysis in R. *Bioinformatics*, *27*(4), 592–593.

Slater, G. J., Price, S. A., Santini, F., & Alfaro, M. E. (2010). Diversity versus disparity and the radiation of modern cetaceans. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1697), 3097–3104.

Springer, M. S., Meredith, R. W., Gatesy, J., Emerling, C. A., Park, J., Rabosky, D. L., Stadler, T., Steiner, C., Ryder, O. A., Janečka, J. E., Fisher, C. A., & Murphy, W. J. (2012). Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. *PLoS ONE*, *7*(11), e49521. https://doi.org/10.1371/journal.pone.0049521

Stadler, T. (2010). Sampling-through-time in birth-death trees. *Journal of Theoretical Biology*, *267*(3), 396–404.

Stadler, T. (2011). Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(15), 6187–6192.

Steeman, M. E., Hebsgaard, M. B., Fordyce, R. E., Ho, S. Y., Rabosky, D. L., Nielsen, R., Rahbek, C., Glenner, H., Sørensen, M. V., & Willerslev, E. (2009). Radiation of extant cetaceans driven by restructuring of the oceans. *Systematic Biology*, *58*(6), 573–585.

Tribble, C. M., May, M. R., Barido-Sottani, J., Kopperud, B. T., Freyman, W. A., Magee, A. F., Lim, J. Y., & Landis, M. R. (2021). *cmt2/RevGadgets: DOI release*.

Tufte, E. (2001). *The visual display of quantitative information*. Graphic Press.

Urbanek, S. (2013). *png: Read and write PNG images*.

Uyeda, J. C., & Harmon, L. J. (2014). A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. *Systematic Biology*, *63*(6), 902–918.

Uyeda, J., & Harmon, L. (2020). *treeplyr: 'dplyr' functionality for matched tree and data objects*.

Vaughan, T. G. (2017). IcyTree: Rapid browser-based visualization for phylogenetic trees and networks. *Bioinformatics*, *33*(15), 2392–2394.

Volz, E. M., Koelle, K., & Bedford, T. (2013). Viral phylodynamics. *PLoS Computational Biology*, *9*(3), e1002947.

Wang, L.-G., Lam, T.-T.-Y., Xu, S., Dai, Z., Zhou, L., Feng, T., Guo, P., Dunn, C. W., Jones, B. R., Bradley, T., Zhu, H., Guan, Y., Jiang, Y., & Yu, G. (2020). treeio: An R package for phylogenetic tree input and output with richly annotated and associated data. *Molecular Biology and Evolution*, *37*(2), 599–603.

Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, *21*(12), 1–20.

Wickham, H. (2011). ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*, *3*(2), 180–185.

Wickham, H. (2021). *tidyr: Tidy messy data*.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686.

Wickham, H., François, R., Henry, L., & Müller, K. (2021). *dplyr: A grammar of data manipulation*.

Yang, Z. (2014). *Molecular evolution: A statistical approach*. Oxford University Press.

Yu, G. (2020). *ggimage: Use image in 'ggplot2'*.

Yu, G. (2021a). *ggplotify: Convert plot to 'grob' or 'ggplot' object*.

Yu, G. (2021b). *tidytree: A tidy tool for phylogenetic tree data manipulation*.

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T.-T.-Y. (2017). ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28–36.

Zenil-Ferguson, R., Burleigh, J. G., Freyman, W. A., Igić, B., Mayrose, I., & Goldberg, E. E. (2019). Interaction among ploidy, breeding system and lineage diversification. *New Phytologist*, *224*(3), 1252–1265.

Zhang, C., Stadler, T., Klopfstein, S., Heath, T. A., & Ronquist, F. (2016). Total-evidence dating under the fossilized birth-death process. *Systematic Biology*, *65*(2), 228–249.

How to cite this article: Tribble, C. M., Freyman, W. A., Landis, M. J., Lim, J. Y., Barido-Sottani, J., Kopperud, B. T., Höhna, S., & May, M. R. (2022). `RevGadgets`: An `R` package for visualizing Bayesian phylogenetic analyses from `RevBayes`. *Methods in Ecology and Evolution*, 13, 314–323. https://doi.org/10.1111/2041-210X.13750