

APPLICATION

rOTL: an R package to interact with the Open Tree of Life data

François Michonneau^{1,2*}, Joseph W. Brown³ and David J. Winter⁴

¹Whitney Laboratory for Marine Sciences, University of Florida, St. Augustine, FL 32080, USA; ²Florida Museum of Natural History, University of Florida, Gainesville, FL 32611-7800, USA; ³Department of Ecology & Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA; and ⁴Virginia G. Piper Centre for Personalized Diagnostics, The Biodesign Institute, Arizona State University, Tempe, AZ 85287-5001, USA

Summary

1. While phylogenies have been getting easier to build, it has been difficult to reuse, combine and synthesize the information they provide because published trees are often only available as image files, and taxonomic information is not standardized across studies.
2. The Open Tree of Life (OTL) project addresses these issues by providing a digital tree that encompasses all organisms, built by combining taxonomic information and published phylogenies. The project also provides tools and services to query and download parts of this synthetic tree, as well as the source data used to build it. Here, we present rOTL, an R package to search and download data from the Open Tree of Life directly in R.
3. rOTL uses common data structures allowing researchers to take advantage of the rich set of tools and methods that are available in R to manipulate, analyse and visualize phylogenies. Here, and in the vignettes accompanying the package, we demonstrate how rOTL can be used with other R packages to analyse biodiversity data.
4. As phylogenies are being used in a growing number of applications, rOTL facilitates access to phylogenetic data and allows their integration with statistical methods and data sources available in R.

Key-words: comparative methods, macroevolution, Open Tree of Life, phylogenetics

Advances in sequencing and computing technologies have led to a revolution in systematic biology. The ability to routinely generate molecular data sets from any extant organism has allowed researchers to resolve long-standing taxonomic disputes and estimate phylogenies for previously understudied groups. In parallel, the ease with which phylogenies can be estimated has spurred the development of new phylogenetic comparative methods. These methods allow researchers to explore fundamental questions about the origin of biodiversity including the evolution of morphological and ecological traits, the spatiotemporal variation in speciation rates, or both (O'Meara 2012; Pennell & Harmon 2013).

Ideally, the ever increasing number of published phylogenies would contribute to a synthesis of phylogenetic knowledge, ultimately leading to a better understanding of the history of life while at the same time providing high-quality phylogenetic information for use in comparative analyses. However, in practice, synthesizing phylogenetic data is a difficult task. Phylogenetic information is largely scattered, often only available as image files within publications, and the lack of standardization to store and represent phylogenetic data makes it difficult for researchers to access, synthesize and integrate this information into their own research (Stoltzfus *et al.* 2012; Drew *et al.*

2013; Magee, May & Moore 2014; but see Cranston *et al.* 2014 for suggestions of best practices).

The Open Tree of Life (OTL) project aims at assembling and synthesizing our current understanding of phylogenetic relationships across all organisms on Earth while providing tools and services that facilitate access to this information (Hinchliff *et al.* 2015). OTL combines taxonomic information that serves as the backbone for the phylogenetic relationships, and published phylogenies to elucidate relationships among taxa. This combination of information is used to structure the comprehensive synthetic tree. Studies can be contributed to the synthetic tree through a curator interface (<https://tree.opentreeoflife.org/curator>), allowing the synthetic tree to be continuously updated as relationships are elucidated or re-evaluated. The current draft of the OTL synthetic tree contains 2.3 million tips. Beyond obvious applications across the life sciences to explore questions in evolution, biodiversity and conservation, the resources OTL provides are useful for education and outreach (e.g. illustrating course material, or developing outreach activities to explore relationships among species).

The R programming language is a popular tool for phylogenetics and comparative analysis. The R packages `ape` (Paradis, Claude & Strimmer 2004), `phylbase` (Bolker *et al.* 2015), `phangorn` (Schliep 2011) and `RNexML` (Boettiger *et al.* 2015b) each provide functions to import and manipulate trees within R and save the results in standard data formats.

*Correspondence author. E-mail: francois.michonneau@gmail.com

Additional packages including `phytools` (Revell 2012), `geiger` (Pennell *et al.* 2014) and `ggtree` (<https://guangchuangyu.github.io/ggtree/>) allow users to analyse and visualize data in a phylogenetic context (see <https://cran.r-project.org/web/views/Phylogenetics.html> for a comprehensive list of phylogenetics packages in R). In addition to packages for phylogenetic and comparative analyses, a growing number of R packages allow users to query and access data from the web [e.g. `rFISHBASE` (Boettiger, Lang & Wainwright 2012), `rAvis` (Varela *et al.* 2014) and `paleobioDB` (Varela *et al.* 2015)], such that data associated with taxa in a given phylogeny can be obtained directly in R.

In ecology, the development of the field of community phylogenetics (Webb *et al.* 2002) has created a need for researchers to have access to the evolutionary relationships of species making up communities. The relative contributions of the role of the environment (e.g. habitat filtering), and of biotic interactions (e.g. competitive exclusion), are inferred from the distribution of taxa on a phylogeny composed from species occurring at larger spatial scales. R packages and other software have been developed to generate phylogenies from species lists using taxonomic information or DNA sequences (e.g. Webb, Ackerly & Kembel 2008; Pearse & Purvis 2013). These phylogenies can then be used for community phylogenetics analyses (e.g. Kembel *et al.* 2010; Pearse *et al.* 2015), but they are often incomplete or not resolved enough. As OTL becomes more comprehensive, and its taxonomic resolution increases, it could become a valuable resource for ecologists seeking to use phylogenetic information in their research.

These packages, combined with the language's support for literate programming (Knuth 1984; Xie 2015), make R a comprehensive platform for reproducible research in phylogenetics and comparative biology, as they allow a complete record of the steps taken in gathering, processing and analysing a given data set to be produced.

Here, we present `rotl`, an R package that allows users to download phylogenetic and taxonomic data from the OTL directly in R. `rotl` takes advantage of OTL's Application Programming Interfaces (APIs) to access subtrees from the synthetic Open Tree, as well as the published source trees that contribute to the synthesis. By providing direct access to high-quality phylogenetic data in R, `rotl` fills a key gap in typical comparative analysis workflows and extends the degree to which R supports reproducible research in phylogenetics and comparative biology.

API services provided by OTL

The OTL project provides four resources that serve data to users through the APIs:

1. The *taxonomy* used as the backbone of the tree, the Open Tree Taxonomy (OTT).
2. The *studies* and their associated trees, some of which are chosen by curators to assemble the synthetic tree.
3. A *taxonomic name resolution service* (TNRS) used to match taxon names to the Open Tree Taxonomy identifiers.
4. The *synthetic tree* itself, the 'Open Tree'.

`rotl` gives users access to the endpoints provided by version 3 of the APIs, and other versions of the APIs can be selected by the user as they become available.

Phylogenetic trees served by the API can be imported directly into R's memory and are represented using the `ape` (Paradis, Claude & Strimmer 2004) tree structure (objects of class `phylo`), or can be written to files in the Newick, NEXUS (Maddison, Swofford & Maddison 1997) or NeXML (Vos *et al.* 2012) file formats. This allows researchers to use these trees either directly with other R packages, or to be imported in other programs that make use of phylogenetic tree files.

Currently, the synthetic tree does not have any branch lengths associated with it; therefore, parametric comparative methods cannot be used directly on the subtrees returned by OTL (although the OTL treestore contains the raw published source phylogenies, complete with branch lengths and node annotations; see below). However, resources and methods are being developed to add branch lengths to these topological subtrees (e.g. Ksepka *et al.* 2015) or use topological trees to identify phylogenetically equivalent species to increase overlap between chronograms and species trait data (Pennell, FitzJohn & Cornell 2016). Without branch lengths, these subtrees are nonetheless useful to illustrate relationships among species, or to map traits on a phylogeny.

Technical information about `rotl`

Phylogenetic information retrieved from OTL is converted into `phylo` objects by `rotl` using the NEXUS Class Library (NCL, Lewis 2003) as implemented in the `rncl` package (<https://cran.r-project.org/package=rncl>). Using NCL provides robust and efficient parsing of large trees that may contain singleton nodes labelled with taxonomic information (i.e. a monotypic taxon). Singleton nodes are collapsed after the tree has been parsed, making the resulting `phylo` object compatible with all functions from the `ape` package.

The package is well documented and includes three package vignettes (documents that demonstrate the use of the package and contain executable R code). There is also an extensive test suite that covers both the internal functions that `rotl` uses to connect to OTL, and public functions that users apply to access and process data.

Demonstrations

GETTING RELATIONSHIPS FROM A LIST OF TAXA

Before a researcher can use the Open Tree to retrieve relationships among a set of taxa, they first need to match the taxon names in their data set with records in the Open Tree Taxonomy (OTT). OTL's taxonomic name resolution service (TNRS) combines information from multiple services [e.g. National Center for Biotechnology Information (NCBI), World Register for Marine Species (WoRMS), Global Biodiversity Information Facility (GBIF)] and allows users to search for taxon names and retrieve identifiers for each matching taxon. We demonstrate the use of the TNRS within `rotl` by

searching for taxonomic records associated with several model organisms.

```
taxa<-tnrs_match_names(names=c("Escherichia coli",
"Chlamydomonas reinhardtii",
"Drosophila melanogaster",
"Arabidopsis thaliana",
"Rattus norvegicus",
"Mus musculus",
"Cavia porcellus",
"Xenopus laevis",
"Saccharomyces cerevisiae",
"Danio rerio"))
```

The function `tnrs_match_names` returns a data frame that lists the Open Tree identifiers as well as other information to help users ensure that the taxa matched are the correct ones. Here, there is no ambiguity in the taxa matched; however, as OTT includes taxa from bacteria, plants, and animals that are regulated by different nomenclatural codes (ICNP, ICN and ICZN, respectively), both OTL and `rotl` provide tools to deal with names that may represent valid taxa in more than one code. The argument `context_name` can be used to limit potential matches to a taxonomic group such as ‘Animals’ (see the function `tnrs_contexts` for a complete list of possible options). When this strategy cannot be used (as in the present example, where the tree encompasses multiple domains), the function `inspect` lists alternative matches for a taxon name and `update` replaces it in the results. An example of this approach is provided in the vignette ‘How to use `rotl`?’ that accompanies the package.

By default, approximate matching is enabled when attempting to match taxonomic names to their OTT identifiers. Additionally, taxonomic synonyms are included in OTT, allowing researchers to match correct identifiers for taxon names that might include misspellings or synonyms. These features will facilitate the tedious data cleaning process often needed when mapping taxon names. In the example provided, both *Escherichia coli* and *Saccharomyces cerevisiae* are misspelled, but OTL’s TNRS finds the correct match for these taxa.

Now that the taxon names are matched to the Open Tree identifiers, we can pass them to the function `tol_induced_subtree` to retrieve the relationships among these taxa. In turn, the tree can be plotted directly as it is returned as `phylo` object (Fig. 1).

```
tree<-tol_induced_subtree(ott_ids=ott_id(taxa))
plot(tree, cex=.8, label.offset=.1, no.margin=
TRUE)
```

GETTING TREES FROM STUDIES

`rotl` can also be used to retrieve trees accompanying studies that have been submitted through the curator interface, and identify the trees that contribute to the synthetic tree. As of March 2016, the Open Tree of Life project stores 7755 trees from 3399 studies (each having between 0 and 61 trees), and

477 of these trees are used to assemble the synthetic tree. These trees constitute a useful resource to reproduce or expand on a previously published analysis, or to explore how the elucidation of relationships within a clade has changed through time.

Criteria that can be used to search for studies or their associated trees are available through the output of the function `studies_properties`. The meaning of these properties is described at: <https://github.com/OpenTreeOfLife/phyloapi/wiki/NexSON>. Typically, users will want to search for studies or trees based on taxon names (or their OTT identifiers), but other criteria such as the title of the publication can be used. Here, we demonstrate how to look for and retrieve trees for studies focusing on the family Felidae (Fig. 2).

```
cat_studies<-studies_find_studies(property="ot:
focalCladeOTTTaxonName",
value="Felidae", exact=TRUE)
cat_studies
```

Currently only one study focused on this family is available from OTL, and a single tree is associated with it. We can then retrieve the study and tree identifiers and pass them to the function `get_study_tree` to have the tree in memory:

```
cat_tree<-get_study_tree(study_id=cat_studies
[["study_ids"]][1],
tree_id=cat_studies[["tree_ids"]][1])
cat_tree
##
##Phylogenetic tree with 38 tips and 37 internal
nodes.
##
##Tip labels:
##Neofelis_nebulosa, Panthera_tigris,
Panthera_uncia, Panthera_pardus, ...
##
##Rooted; includes branch lengths.
```

When more than one tree is available for a given study, the function `list_trees` returns a list containing the tree identifiers for each study. Alternatively, the function `get_study` returns all the trees (by default as `phylo` objects) associated with a particular study. Metadata about the study (e.g. citation information, information about the curator for the study and other technical information regarding the import of this study) can be obtained using the functions `get_study_meta` and `study_external_ids`.

HOW DOES ROTL FIT INTO THE R PACKAGE ECOSYSTEM?

In recent years, R has become an essential part of the toolbox of many researchers in evolutionary biology and ecology. R greatly facilitates the analysis of large data sets

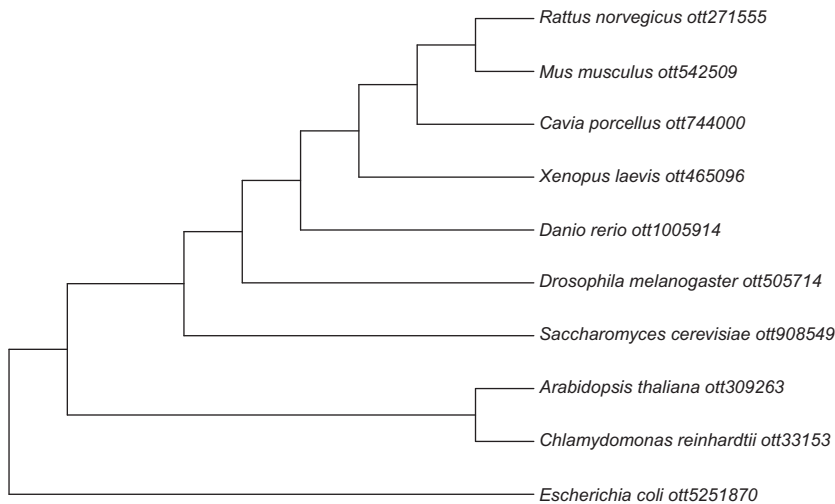


Fig. 1. The phylogenetic tree returned by OTL for the list of model species used as an example.

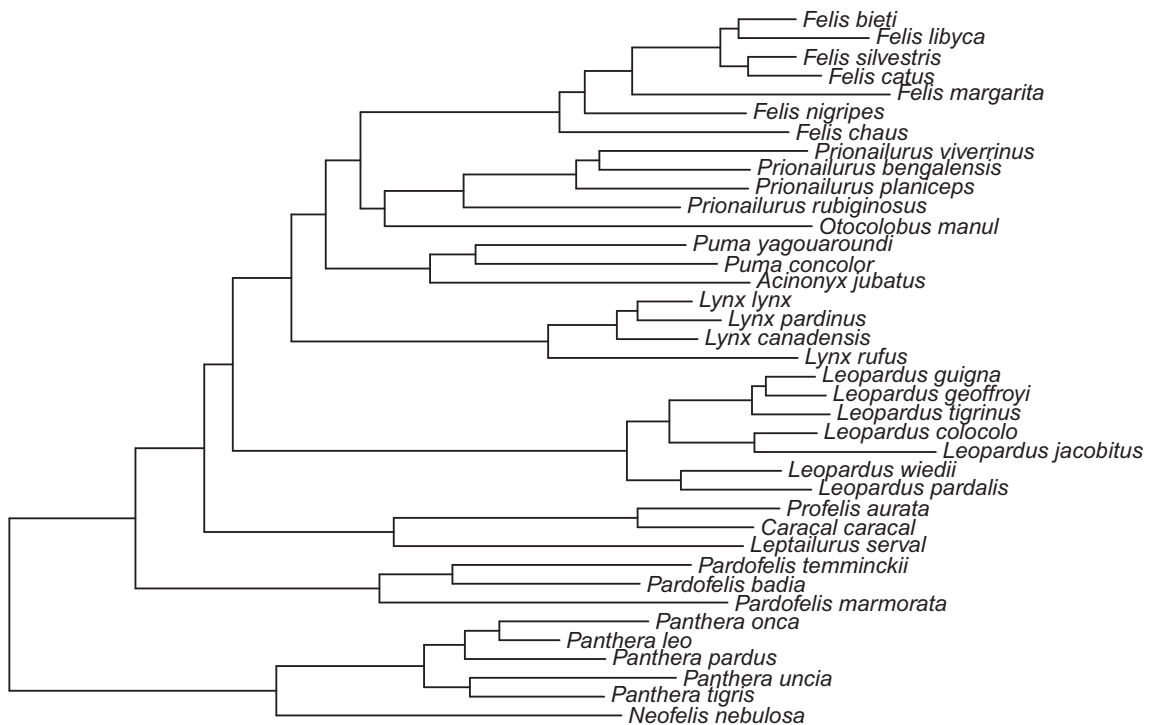


Fig. 2. Phylogeny of the Felidae published in Johnson *et al.* (2006) and retrieved from OTL using `rOTL`.

and allows researchers to combine methods in novel ways because many methods for comparative analyses are implemented, and because it is a relatively easy-to-use programming language. Additionally, as more data are made available online and accessible using web APIs, several packages have been developed to interact and download these data sets directly in R, thereby enabling direct and reproducible analyses. Notably, the organization rOpenSci (<https://ropensci.org>) has fostered a community of researchers who develop tools and methods to facilitate the use of open data as well broaden the adoption of open science practices in general (Boettiger *et al.* 2015a). For instance, the rOpenSci-developed package `TreeBase` (Boettiger

& Temple Lang 2012) allows users to access phylogenies stored in TreeBASE (<https://treebase.org>). `rOTL` contributes to this initiative and greatly extends the number of taxa for which phylogenetic data can be retrieved within R, while allowing the data from OTL to be combined with other sources easily.

Here, we show how we can obtain a map of the occurrences for a subset of the cat species that were included in the phylogeny retrieved from the Felidae study above (genus *Lynx*). We extract the species names from the phylogeny and use them to ask for the records for these species found in GBIF (Fig. 3). We include the code to reproduce this figure in Appendix S1.

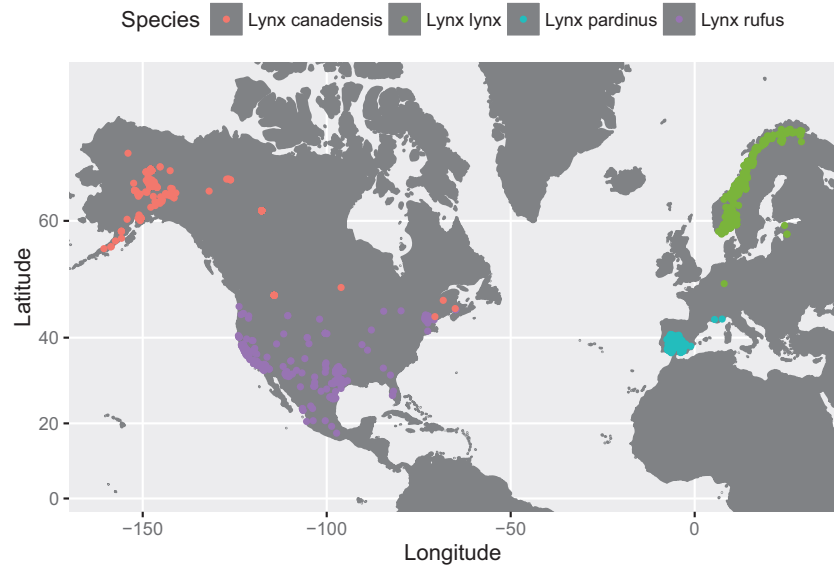


Fig. 3. GBIF records for the species in *Lynx* included in the phylogeny associated with the study by Johnson *et al.* (2006).

As trait data bases are becoming increasingly more common, and interfaces to the data they contain are being developed as R packages (e.g. the package `traits`, Chamberlain *et al.* 2016), `rotl` provides a way to easily retrieve phylogenetic information for species trait data that are available.

In addition to an introduction on how to use the package, `rotl` includes two vignettes that demonstrate how to integrate a phylogeny and data associated with the taxa it represents. Specifically, the ‘Data mashups’ vignette provides an example of how to retrieve a phylogeny for species a researcher may have data for, and visualize both the phylogeny and data associated with the species at the tips. The other vignette titled ‘Meta-analysis’ demonstrates how a complete comparative method analysis, including the gathering of data and a phylogeny, can be performed in a single R session. We reproduce a published meta-analysis testing for differential investment in male and female offsprings among 51 species of birds. As new versions of the OTL API and `rotl` are released, these vignettes will be kept up-to-date.

Concluding remarks

The recognition of the importance of phylogenies to account for the statistical non-independence of species in comparative methods, the recent development of methods to explore trait evolution or changes in diversification rates, and attempts to incorporate the evolutionary history of species forming ecological communities have driven the need for accurate phylogenies. However, there is often a discrepancy between taxa targeted by studies wanting to use phylogenetic information, and taxa for which phylogenies are available. Typically, the latter result from focused studies of taxonomic groups, while the former encompass species found in a given geographical location or ecosystem. We believe that by providing an easy-to-use interface to obtain phylogenies for an arbitrary set of taxa directly in R, `rotl` will be useful in a wide variety of contexts.

The accuracy and usefulness of the data provided by OTL relies on the community to make generated phylogenies (and their metadata) digitally available as tree files (i.e. Newick, NEXUS or NeXML). We strongly encourage researchers to submit their published phylogenies to OTL using the curator interface (<https://tree.opentreeoflife.org/curator>). By facilitating the discovery and reuse of published trees and of the synthetic Open Tree, we hope `rotl` will contribute to the wider adoption of best practices to make phylogenetic information available and reusable.

Availability

`rotl` is free, open source and released under a Simplified BSD licence. Stable versions are available from the CRAN repository (<https://cran.r-project.org/package=rotl>), and development versions are available from GitHub (<https://github.com/ropensci/rotl>). This manuscript was built using `rotl` 3.0.0 (<https://github.com/ropensci/rotl/tree/v3.0.0>). The package is under active development, and authors welcome bug reports or feature requests via the GitHub repository. The source for this manuscript is available on GitHub (<https://github.com/fmichonneau/rotl-ms>).

Python (<https://github.com/OpenTreeOfLife/pyopentree>) and Ruby (<https://github.com/SpeciesFileGroup/bark>) libraries to interact with the OTL APIs are also available.

Acknowledgments

We would like to thank the organizers of the OpenTree of Life APIs hackathon that was held at the University of Michigan, Ann Arbor, 15–19 September 2014, where the development of `rotl` was started. We would also like to thank Scott Chamberlain (rOpenSci) for providing a thorough code review, Scott Chamberlain and Ross Mounce for commenting on the pre-print version of this manuscript, Shinichi Nakagawa and Alistair Senior for their help in developing the package’s meta-analysis vignette, and Rich FitzJohn (Associate Editor) and an anonymous reviewer for comments on the manuscript. DJW was supported by NIH Grant R01-GM101352. FM was supported by 1923 Fund and by iDigBio, and therefore, this material is based upon work supported by the National Science Foundation’s Advancing Digitization of Biodiversity Collections Program (Cooperative Agreement EF-1115210).

References

- Boettiger, C., Lang, D.T. & Wainwright, P.C. (2012) Rfishbase: exploring, manipulating and visualizing FishBase data from R. *Journal of Fish Biology*, **81**, 2030–2039.
- Boettiger, C. & Temple Lang, D. (2012) Treebase: an R package for discovery, access and manipulation of online phylogenies. *Methods in Ecology and Evolution*, **3**, 1060–1066.
- Boettiger, C., Chamberlain, S., Hart, E. & Ram, K. (2015a) Building software, building community: lessons from the rOpenSci project. *Journal of Open Research Software*, **3**, p.e8.
- Boettiger, C., Chamberlain, S., Vos, R. & Lapp, H. (2015b) RNeXML: a package for reading and writing richly annotated phylogenetic, character, and trait data in R. *Methods in Ecology and Evolution*, **7**, 352–357.
- Bolker, B., Butler, M., Cowan, P., de Vienne, D., Eddelbuettel, D., Holder, M. *et al.* (2015) phylobase: Base Package for Phylogenetic Structures and Comparative Data.
- Chamberlain, S., Foster, Z., Bartomeus, I., LeBauer, D. & Harris, D. (2016) traits: Species Trait Data from Around the Web.
- Cranston, K., Harmon, L.J., O'Leary, M.A. & Lisle, C. (2014) Best practices for data sharing in phylogenetic research. *PLoS Currents*, **6**, 1–8.
- Drew, B.T., Gazis, R., Cabezas, P., Swithers, K.S., Deng, J., Rodriguez, R. *et al.* (2013) Lost branches on the tree of life. *PLoS Biology*, **11**, e1001636.
- Hinchliff, C.E., Smith, S.A., Allman, J.F., Burleigh, J.G., Chaudhary, R., Coghill, L.M. *et al.* (2015) Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proceedings of the National Academy of Sciences*, **112**, 12764–12769.
- Johnson, W.E., Eizirik, E., Pecon-Slattery, J., Murphy, W.J., Antunes, A., Teeling, E. & O'Brien, S.J. (2006) The late Miocene radiation of modern Felidae: a genetic assessment. *Science*, **311**, 73–77.
- Kemmel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P. & Webb, C.O. (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, **26**, 1463–1464.
- Knuth, D.E. (1984) Literate programming. *The Computer Journal (British Computer Society)*, **27**, 97–111.
- Ksepka, D.T., Parham, J.F., Allman, J.F., Benton, M.J., Carrano, M.T., Cranston, K.A. *et al.* (2015) The fossil calibration database—a new resource for divergence dating. *Systematic Biology*, **64**, 853–859.
- Lewis, P.O. (2003) NCL: A C++ class library for interpreting data files in NEXUS format. *Bioinformatics*, **19**, 2330–2331.
- Maddison, D.R., Swofford, D.L. & Maddison, W.P. (1997) NEXUS: an extensible file format for systematic information. *Systematic Biology*, **46**, 590–621.
- Magee, A.F., May, M.R. & Moore, B.R. (2014) The dawn of open access to phylogenetic data.
- O'Meara, B.C. (2012) Evolutionary inferences from phylogenies: a review of methods. *Annual Review of Ecology, Evolution, and Systematics*, **43**, 267–285.
- Paradis, E., Claude, J. & Strimmer, K. (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, **20**, 289–290.
- Pearse, W.D. & Purvis, A. (2013) phyloGenerator: an automated phylogeny generation tool for ecologists. *Methods in Ecology and Evolution*, **4**, 692–698.
- Pearse, W.D., Cadotte, M.W., Cavender-Bares, J., Ives, A.R., Tucker, C.M., Walker, S.C. & Helmus, M.R. (2015) pez: phylogenetics for the environmental sciences. *Bioinformatics*, **31**, 2888–2890.
- Pennell, M.W., FitzJohn, R.G. & Cornell, W.K. (2016) A simple approach for maximizing the overlap of phylogenetic and comparative data. *Methods in Ecology and Evolution*, in press.
- Pennell, M.W. & Harmon, L.J. (2013) An integrative view of phylogenetic comparative methods: connections to population genetics, community ecology, and paleobiology. *Annals of the New York Academy of Sciences*, **1289**, 90–105.
- 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.
- Revell, L.J. (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, **3**, 217–223.
- Schliep, K.P. (2011) phangorn: phylogenetic analysis in R. *Bioinformatics*, **27**, 592–593.
- Stoltzfus, A., O'Meara, B., Whitacre, J., Mounce, R., Gillespie, E.L., Kumar, S., Rosauer, D.F. & Vos, R.A. (2012) Sharing and re-use of phylogenetic trees (and associated data) to facilitate synthesis. *BMC Research Notes*, **5**, 574.
- Varela, S., González-Hernández, J., Casabella, E. & Barrientos, R. (2014) rAvis: an R-package for downloading information stored in proyecto AVIS, a citizen science bird project. *PLoS ONE*, **9**, e91650.
- Varela, S., González-Hernández, J., Sgarbi, L.F., Marshall, C., Uhen, M.D., Peters, S. & Mcclennen, M. (2015) paleobioDB: an R package for downloading, visualizing and processing data from the Paleobiology Database. *Ecography*, **38**, 419–425.
- Vos, R.A., Balhoff, J.P., Caravas, J.A., Holder, M.T., Lapp, H., Maddison, W.P. *et al.* (2012) NeXML: rich, extensible, and verifiable representation of comparative data and metadata. *Systematic Biology*, **61**, 675–689.
- Webb, C.O., Ackerly, D.D. & Kembel, S.W. (2008) Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics*, **24**, 2098–2100.
- Webb, C.O., Ackerly, D.D., McPeck, M.A. & Donoghue, M.J. (2002) Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, **33**, 475–505.
- Xie, Y. (2015) *Dynamic Documents with R and knitr*, 2nd edn. Chapman and Hall/CRC, Boca Raton, FL.

Received 4 November 2015; accepted 11 May 2016

Handling Editor: Richard Fitzjohn

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Appendix S1. Source code used to generate Fig. 3.