An R package and online resource for macroevolutionary studies using the ray-finned fish tree of life

Jonathan Chang[1] ID | Daniel L. Rabosky[2] ID | Stephen A. Smith[2] ID | Michael E. Alfaro[3] ID

[1]School of Biological Sciences, Monash University, Clayton, VIC, Australia

[2]Museum of Zoology, Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI

[3]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA

Correspondence
Jonathan Chang
Email: jonathan.chang@monash.edu

Funding information
Encyclopedia of Life, Grant/Award Number: EOL-33066-13; National Science Foundation, Grant/Award Number: 1601830; Society of Systematic Biologists

Handling Editor: Daniele Silvestro

Abstract

1. Comprehensive, time-scaled phylogenies provide a critical resource for many questions in ecology, evolution and biodiversity. Methodological advances have increased the breadth of taxonomic coverage in phylogenetic data; however, accessing and reusing these data remain challenging.

2. We introduce the Fish Tree of Life website and associated R package `fishtree` to provide convenient access to sequences, phylogenies, fossil calibrations and diversification rate estimates for the most diverse group of vertebrate organisms, the ray-finned fishes. The Fish Tree of Life website presents subsets and visual summaries of phylogenetic and comparative data, and is complemented by the R package, which provides flexible programmatic access to the same underlying data source for advanced users wishing to extend or reanalyse the data.

3. We demonstrate functionality with an overview of the website, and show three examples of advanced usage through the R package. First, we test for the presence of long branch attraction artefacts across the fish tree of life. The second example examines the effects of habitat on diversification rate in the pufferfishes. The final example demonstrates how a community phylogenetic analysis could be conducted with the package.

4. This resource makes a large comparative vertebrate dataset easily accessible via the website, while the R package enables the rapid reuse and reproducibility of research results via its ability to easily integrate with other R packages and software for molecular biology and comparative methods.

KEYWORDS
Actinopterygii, fossil calibration, macroevolution, open data, phylogenetics, ray-finned fishes, reproducible research

1 | INTRODUCTION

Phylogenies are fundamental to comparative evolutionary biology, and their use extends to community ecology, conservation biology, ecophysiology, developmental biology and translational medical research. New phylogenetic information can illuminate open questions in biology, but this work is clouded by the difficulty in inferring phylogenies, especially for non-specialist researchers (Pearse & Purvis, 2013). To avoid these pitfalls, reusing existing phylogenies can make phylogenetic knowledge accessible without requiring researchers to collaborate with phylogenetic experts or learn these methods themselves (Arnold, Matthews, & Nunn, 2010; Magee, May, & Moore, 2014; Webb, Ackerly, & Kembel, 2008; Webb & Donoghue, 2005). However, surveys of the biological literature estimate that 60%–95% of previously published phylogenetic datasets are no longer accessible (Drew et al., 2013; Magee et al., 2014; McTavish, Drew, Redelings, & Cranston, 2017; Stoltzfus et al., 2012), highlighting the challenge of persistently

sharing data and creating a major barrier to new comparative analyses.

One alternative solution is a "tree of life" approach that centralizes research effort across large groups to create a curated and validated phylogenetic dataset, as opposed to smaller family- or genus-level analyses (Beaulieu & O'Meara, 2018; McTavish et al., 2017). These broad phylogenies, in diverse groups such as mammals, birds, squamate reptiles, fishes and angiosperms (Bininda-Emonds et al., 2007; Jetz, Thomas, Joy, Hartmann, & Mooers, 2012; Pyron, Burbrink, & Wiens, 2013; Rabosky et al., 2018; Zanne et al., 2014), represent the best target for phylogenetic re-use, as extensive sampling across these broad organismal groups is likely to cover the particular set of species that would interest a taxon-focused researcher.

Here, we present a new community resource and accompanying R package, the Fish Tree of Life, focusing on the ray-finned fishes, the most species-rich group of vertebrates with over 33,000 species. We describe this resource, which is based on a recent complete phylogeny (Rabosky et al., 2018), and provide three motivating examples, showing how this large empirical dataset could be used to investigate the common problem of long branch attraction, study a specific taxon in a phylogenetic comparative analysis and analyse a dataset using methods from phylogenetic community ecology. This work joins other resources such as birdtree.org (Jetz et al., 2012), the Open Tree of Life (Hinchliff et al., 2015), and Phylotastic (Nguyen et al., 2018). We expand on these previous offerings by also providing pre-computed taxonomic subsets with character matrices, phylogenies, fossil calibrations and diversification rate information in a website and R package.

2 | FUNCTIONALITY

2.1 | Website: fishtreeoflife.org

Our website aims to permit easy access the curated dataset introduced in Rabosky et al. (2018), including the multiple sequence alignment, the phylogram from RAxML (Stamatakis, 2014), the time-calibrated phylogeny from treePL (Smith & O'Meara, 2012) and the fossil calibrations used for divergence time estimation. We also generated pages and downloads for each rank above family in the Phylogenetic Fish Classification (Rabosky et al., 2018). Each page lists all species in that taxon, as well as taxonomy and subsets of the sequence alignments, phylogenies and fossil calibrations. Separate pages and downloads permit more focused work; for example, in conjunction with new genetic data, a researcher could use profile alignment in MAFFT (Katoh & Standley, 2013) to incorporate their new data into our existing sequence alignment. This saves time compared to a *de novo* analysis, as the rigorous validation and curation process in Rabosky et al. (2018) should reduce the amount of erroneous or misidentified sequences in combined datasets (Bridge, Roberts, Spooner, & Panchal, 2003).

We have also included a fossil section to our Fish Tree of Life website (Figure 1). This lists all 139 fossils used in our analysis, as well as the phylogenetic placement of those fossils on the phylogeny. Each page includes the taxon it calibrates (e.g. crown Acanthuridae),

as well as the minimum age, authorities for taxonomic placement and age and fossil locality. We also show the upper bound of the 95% confidence interval for the estimated age of the Hedman fossil outgroup process (Hedman, 2010), and list the fossil outgroup sequence used to calculate those bounds. Our approach explicitly integrates fossil knowledge in a phylogenetic context suitable for divergence time estimation, while some other resources, such as TimeTree (Hedges, Dudley, & Kumar, 2006) or DateLife (Nguyen et al., 2018), either do not permit reuse or lack detailed fossil taxonomy and locality data. Our compilation could provide an established starting point for analyses that for example, vary fossil calibrations to estimate their downstream effects on diversification rate inference.

2.2 | R package: `fishtree`

As the website is intended for browsing, more complex analyses should be conducted in a reproducible programming environment. We, therefore, wrote the R package `fishtree`, which facilitates access to data from the fishtreeoflife.org website. Researchers can load the alignments, phylogenies and diversification rate metrics directly into native R objects, using the `fishtree_alignment`, `fishtree_phylogeny` and `fishtree_tip_rates` functions, respectively, and can subset data by taxonomic rank, for example, by family (Labridae) or order (Labriformes). Phylogenies are classed as type `phylo` from `ape` (Paradis & Schliep, 2018) to work seamlessly in conjunction with other commonly used R packages for phylogenetics and comparative analysis. We summarize the major `fishtree` functions in Table 1.

3 | EXAMPLE APPLICATIONS

Here, we demonstrate three example studies that could be conducted with the `fishtree` R package. The first example shows how researchers could investigate a common problem in phylogenetic inference, long branch attraction. The second example shows how comparative biologists interested in a specific group (pufferfishes) could test a hypothesis related to trait-dependent diversification. We also provide a final example as a vignette in the supplement that shows a phylogenetic community ecology analysis using the R package `picante` (Kembel et al., 2010). The latter two examples are available in the Supporting Information and as vignettes in the R package.

3.1 | Example: testing long branch attraction across the fish tree of life

We demonstrate how a researcher might investigate the problem of long branch attraction (LBA). This occurs when two long branches are incorrectly grouped together as sisters (Bergsten, 2005), and is generally recognized as a problem when saturation, heterotachy or across-lineage rate variation is rampant in a sequence alignment (Philippe, Zhou, Brinkmann, Rodrigue, & Delsuc, 2005).

Here, we reanalyse the phylogeny by family to determine what portions might have been affected by LBA. If LBA artefacts are

Esox kronneri†

Taxon calibrated
Crown Esox

Minimum age
51.57 Ma

Maximum 95% CI age
114.0 Ma

Authority
Grande (1999)

Locality
Fossil Butte Member, Green River Formation (F-2 locality), Wyoming, USA

Age Authority
Smith et al. (2008)

Outgroup sequence
Esox kronneri (Crown Esox, 51.57 Ma)
Estesesox foxi (Total group Esocidae, 76.4 Ma)
Apateodus glyphodus (Total group Aulopiformes, 103.13 Ma)
Leptolepides haerteisi (Total group Euteleostei, 150.94 Ma)
Anaethalion zapporum (Crown Teleostei, 151.2 Ma)
Occithrissops willsoni (stem Teleostei 5, 166.1 Ma)
Leptolepis coryphaenoides (stem Teleostei 4, 181.7 Ma)
Dorsetichthys bechei (stem Teleostei 3, 193.81 Ma)
Knerichthys bronni (stem Teleostei 2, 221 Ma)
Prohalecites porroi (stem Teleostei 1, 236 Ma)
Watsonulus (stem Halecomorphi 1, 251.2 Ma)

Taxa this fossil might apply to:
Class: Actinopteri
Subclass: Neopterygii
Infraclass: Teleostei
Megacohort: Osteoglossocephalai
Supercohort: Clupeocephala
Cohort: Euteleosteomorpha
Subcohort: Protacanthopterygii
Order: Esociformes
Family: Esocidae

<center>(a)</center>

JSON   Raw Data   Headers

Save   Copy   Collapse All   Expand All   Filter JSON

```
0:
    clade:          "Crown Esox"
    fossil:         "Esox kronneri"
    left:           "Esox_lucius"
    right:          "Esox_niger"
    min:            "51.57"
    max:            "114.0734535"
    locality:       "Fossil Butte Member, Green River Formation
                    (F-2 locality), Wyoming, USA"
    authority:      "Grande (1999)"
    age_authority:  "Smith et al. (2008)"
```

<center>(b)</center>

FIGURE 1 (a) An example of the fossil calibration page, which includes the exact locality and authorities of the fossil, as well as the outgroup sequence used to determine the 95% upper bound on maximum ages. (b) The same data represented as Javascript Object Notation, a machine-readable data format

TABLE 1 An overview of the four major functions in the R package fishtree. For all functions that take a named taxonomic rank, any rank higher than family is accepted, including higher taxa, for example, Ostariophysi or Ovalentaria

Function	Data retrieved
fishtree_alignment	Aligned sequences for a taxonomic rank or list of species, optionally splitting by gene partition
fishtree_taxonomy	Information for a taxonomic rank, including a list of species and average diversification rates
fishtree_phylogeny	Phylogeny for a taxonomic rank or list of species. Permits downloads of paraphyletic taxa, either by dropping species that break monophyly, or by including all species descending from the most recent common ancestor of all species sampled in the taxon.
fishtree_tip_rates	Tip-specific diversification rates for a taxonomic rank or list of species, computed via BAMM (Rabosky, 2014) or DR statistic (Jetz et al., 2012)
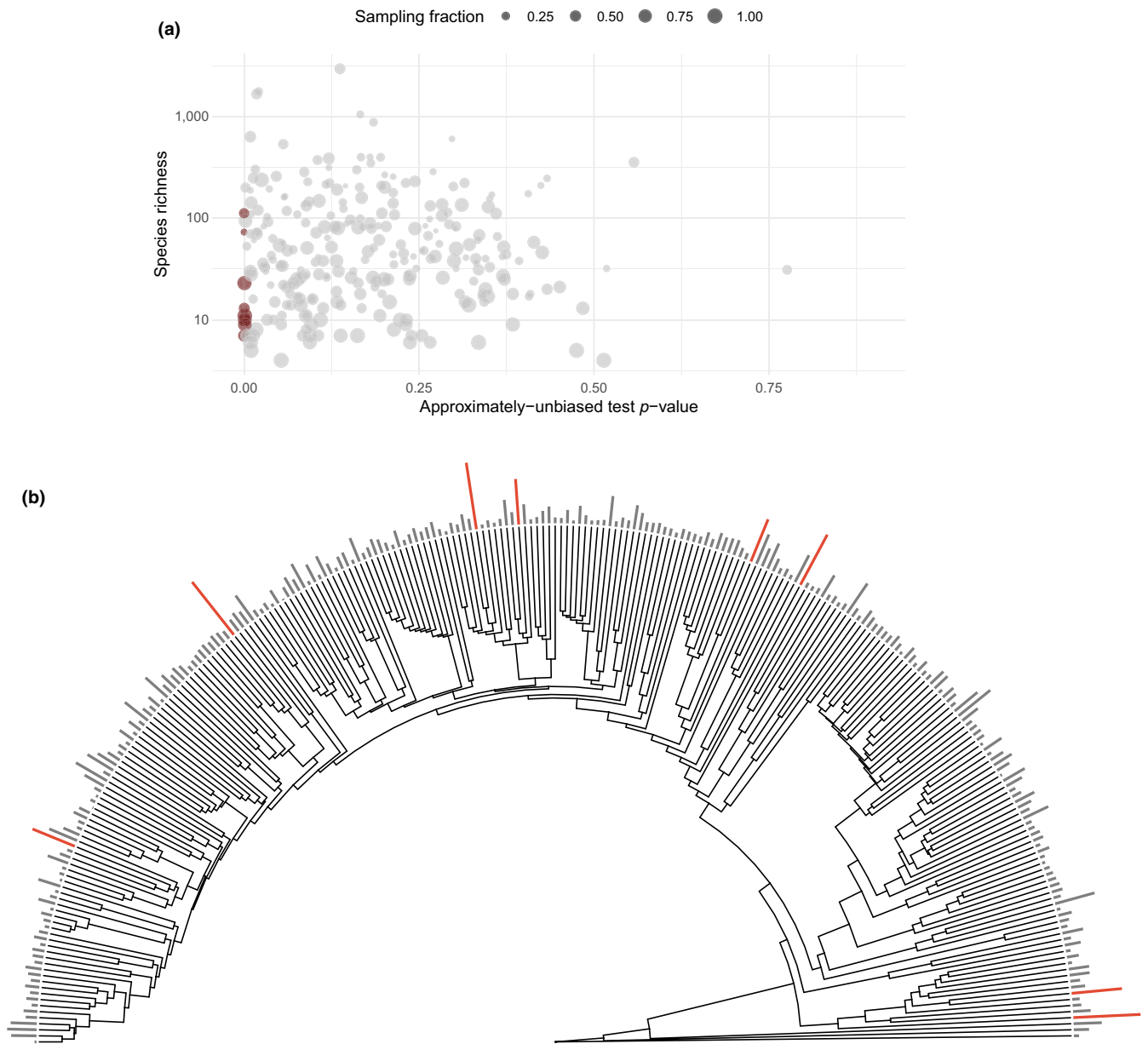
FIGURE 2 The approximately unbiased (AU) test for tree topologies significantly rejected 8 of 268 reanalysed phylogenies in favour of the original topology, coloured in dark red. (a) Alignment incompleteness, species richness, and their interaction significantly predicted the *p*-value of the AU test. (b) Skeletal family-level phylogeny of the ray-finned fish tree of life; bar lengths are the negative log of the AU test *p*-value. This figure with tips labelled by family is provided as Figure S1

present, we predict that the reanalysed topologies would be more balanced (less pectinate) than the original, globally analysed phylogeny. If saturation is causing LBA, we expect that the transition rates would also be faster in the reanalysed phylogenies. The faster transition rates may cause unrelated taxa to be recovered as sister lineages, as fast molecular evolution can lead to shared mutations that are identical by state, not by descent.

We downloaded the alignment for each family with `fishtree_alignment`, and excluded families where three or fewer species had data using `fishtree_taxonomy`. We re-estimated the topology using RAxML v8.2.11 (Stamatakis, 2014) under a partitioned GTR+GAMMA model (Yang, 1996). We refer to these as the "reanalyzed" trees. We also download the phylogeny for each family pruned from the entire phylogeny with `fishtree_phylogeny`; we refer to these as the "pruned" trees.

For each of the reanalysed and pruned topologies, we inferred the rates of molecular evolution using the `-f  e` option in RAxML. We additionally conducted an approximately unbiased (AU) test of topologies (Shimodaira & Hasegawa, 1999), using the `-f  G` option in RAxML to score per-site likelihoods in CONSEL (Shimodaira & Hasegawa, 2001). We reanalysed $n = 268$ family level phylogenies, having on average 43.16 species and the largest family (Cyprinidae) having 1,369 species. After correcting for multiple comparisons (Benjamini & Hochberg, 1995), we significantly rejected ($p_{AU} < 0.05$) the pruned topology in 8 of 268 families with the AU test (Figure 2).

We also computed the normalized Robinson-Foulds (RF) distance (Robinson & Foulds, 1981) and Yule-normalized Colless tree balance metric (Blum, François, & Janson, 2006; Colless, 1982) using `apTreeshape` and `phangorn` (Bortolussi, Durand, Blum, & François, 2006; Schliep, 2011). We fit two regression models: a full model that included alignment incompleteness, the log species richness in the family and the difference in the Colless metric and the RF distance between the pruned and reanalysed phylogenies, with all interaction terms; and a reduced model that only included the alignment incompleteness, log species richness and interaction term; both models used the AU test p-value as the response term. A likelihood ratio test supported the less complex model, with all predictors significant at $p < 0.001$.

Consistent with our prediction, we find that the reanalysed phylogenies tended to be more balanced (less pectinate) than the pruned topology, measured by the Colless metric (31 of 50, 62%). Relative to the pruned topologies, the reanalysed topologies generally had faster transition parameters and substantially different base composition frequencies (Table 2), suggesting that LBA contributed to the more balanced topologies recovered in the reanalysed phylogenies. Based on the significant predictors in the likelihood ratio test, we speculate that the larger dataset more robustly parameterizes the substitution model and leads to fewer LBA artefacts in the pruned trees.

3 | CONCLUSION

We have presented a comprehensive resource that makes a massive comparative dataset of vertebrates available for evolutionary

TABLE 2 Transition rates $A \leftrightarrow C$, $A \leftrightarrow G$, $A \leftrightarrow T$, $C \leftrightarrow G$, $C \leftrightarrow T$ tend to be faster in reanalysed phylogenies, base frequency parameters π_A, π_C, π_G, π_T have a substantially different distribution, and the α parameter of the gamma model of rate heterogeneity (Yang, 1996) suggests much less among-site rate heterogeneity in reanalysed phylogenies. Transition rate parameters were computed relative to the $G \leftrightarrow T$ transition rate

Parameter	Proportion of 268 reanalysed trees where this parameter was smaller
$A \leftrightarrow C$	0.03
$A \leftrightarrow G$	0.24
$A \leftrightarrow T$	0.05
$C \leftrightarrow G$	0.57
$C \leftrightarrow T$	0.01
π_A	0.91
π_C	0.99
π_G	0.01
π_T	0.85
α	0.97

biologists and ecologists. Our resource has numerous facilities to permit researchers to easily use subsets of an otherwise impractically large dataset. We believe that making this dataset available in both web and R package formats will unlock a massive dataset for scientific reuse and synergize well with R Notebooks and other reproducible research tools such as Docker (Boettiger, 2017), while simultaneously lowering the barrier for starting a comparative analysis for researchers of all ability levels.

To demonstrate this, we have shown three example use-cases of our resource, one examining a broad question in molecular evolution, another testing a comparative phylogenetic hypothesis, and the last using a community phylogenetics analysis. In the first example, we made an extremely time-consuming task much easier, as we were able to rapidly import and subset the relevant data into R and focus our efforts on connecting the output from different software and analysing the results. In the second example, we were able to rapidly test a hypothesis in a comparative context, since `fishtree` was designed to work well within the R phylogenetics ecosystem and all analyses could be conducted without many data cleaning tasks. In the last example, we showed how `fishtree` could also be used in a community phylogenetics analysis by testing whether reef fish communities in several ocean basis are phylogenetically clustered or overdispersed.

As concerns around data curation and cleaning in large data aggregations become increasingly visible in biological research (Franz & Sterner, 2018), the ease of use of the tooling around this large, well-curated dataset provides a framework for how concerns around data quality might be assuaged. Further development of the website and R package will focus on adding more pre-computed analyses and figures, which will provide more starting points for researchers hoping to extend and reuse these resources. Finally, our website and R package can be easily updated as new phylogenetic

knowledge becomes available. As the entire process has been standardized and automated inside of a Docker container, any newer ray-finned fish phylogeny can be added as a data file to extend the available data.

AUTHORS' CONTRIBUTIONS

J.C. drafted the manuscript, developed the methods, wrote the software and website. S.A.S., D.L.R., and M.E.A. assisted with analyses and website design. All authors planned the work and contributed to the final manuscript.

DATA ACCESSIBILITY

Our website can be accessed at https://fishtreeoflife.org. The R package is available on GitHub, https://github.com/jonchang/fishtree as well as CRAN (https://CRAN.R-project.org/package=fishtree). Source code and data for the example demonstrations are available from the Dryad Digital Repository (Chang, Rabosky, Smith & Alfaro, 2019, https://doi.org/10.5061/dryad.6vg974n) and in the Supporting Information.

ORCID

Jonathan Chang [iD] https://orcid.org/0000-0002-3811-1254

Stephen A. Smith [iD] https://orcid.org/0000-0003-2035-9531

Daniel L. Rabosky [iD] https://orcid.org/0000-0002-7499-8251

Michael E. Alfaro [iD] https://orcid.org/0000-0002-8898-8230

REFERENCES

Arnold, C., Matthews, L. J., & Nunn, C. L. (2010). The 10kTrees website: A new online resource for primate phylogeny. *Evolutionary Anthropology*, *19*(3), 114–118. https://doi.org/10.1002/evan.20251

Beaulieu, J. M., & O'Meara, B. C. (2018). Can we build it? Yes we can, but should we use it? Assessing the quality and value of a very large phylogeny of campanulid angiosperms. *American Journal of Botany*, *105*(3), 417–432. https://doi.org/10.1002/ajb2.1020

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bergsten, J. (2005). A review of long-branch attraction. *Cladistics*, *21*(2), 163–193. https://doi.org/10.1111/j.1096-0031.2005.00059.x

Bininda-Emonds, O. R. P., Cardillo, M., Jones, K. E., MacPhee, R. D. E., Beck, R. M. D., Grenyer, R., ... Purvis, A. (2007). The delayed rise of present-day mammals. *Nature*, *446*(7135), 507–512. https://doi.org/10.1038/nature05634

Blum, M. G. B., François, O., & Janson, S. (2006). The mean, variance and limiting distribution of two statistics sensitive to phylogenetic tree balance. *The Annals of Applied Probability*, *16*(4), 2195–2214. https://doi.org/10.1214/105051606000000547

Boettiger, C. (2017). A reproducible R notebook using Docker. In J. Kitzes, D. Turek, & F. Deniz (Eds.), *The practice of reproducible research: Case studies and lessons from the data-intensive sciences* (pp. 109–118). Oakland, CA: University of California Press.

Bortolussi, N., Durand, E., Blum, M., & François, O. (2006). apTreeshape: Statistical analysis of phylogenetic tree shape. *Bioinformatics*, *22*(3), 363–364. https://doi.org/10.1093/bioinformatics/bti798

Bridge, P. D., Roberts, P. J., Spooner, B. M., & Panchal, G. (2003). On the unreliability of published DNA sequences. *The New Phytologist*, *160*(1), 43–48. https://doi.org/10.1046/j.1469-8137.2003.00861.x

Colless, D. H. (1982). [Review of] Phylogenetics: The theory and practice of phylogenetic systematics. *Systematic Zoology*, *31*(1), 100–104. https://doi.org/10.2307/2413420

Chang, J., Rabosky, D. L., Smith, S.A., & Alfaro, M. E. (2019). Data from: An R package and online resource for macroevolutionary studies using the ray-finned fish tree of life. *Dryad Digital Repository*. https://doi.org/10.5061/dryad.6vg974n

Drew, B. T., Gazis, R., Cabezas, P., Swithers, K. S., Deng, J., Rodriguez, R., ... Soltis, D. E. (2013). Lost branches on the tree of life. *PLoS Biology*, *11*(9), e1001636. https://doi.org/10.1371/journal.pbio.1001636

Franz, N. M., & Sterner, B. W. (2018). To increase trust, change the social design behind aggregated biodiversity data. *Database*, *2018*, https://doi.org/10.1093/database/bax100

Hedges, S. B., Dudley, J., & Kumar, S. (2006). TimeTree: A public knowledgebase of divergence times among organisms. *Bioinformatics*, *22*(23), 2971–2972. https://doi.org/10.1093/bioinformatics/btl505

Hedman, M. M. (2010). Constraints on clade ages from fossil outgroups. *Paleobiology*, *36*(1), 16–31. https://doi.org/10.1666/0094-8373-36.1.16

Hinchliff, C. E., Smith, S. A., Allman, J. F., Burleigh, J. G., Chaudhary, R., Coghill, L. M., ... Cranston, K. A. (2015). Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proceedings of the National Academy of Sciences*, *112*(41), 12764–12769. https://doi.org/10.1073/pnas.1423041112

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*, *491*(7424), 444–448. https://doi.org/10.1038/nature11631

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., ... Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, *26*(11), 1463–1464. https://doi.org/10.1093/bioinformatics/btq166

Magee, A. F., May, M. R., & Moore, B. R. (2014). The dawn of open access to phylogenetic data. *PLoS ONE*, *9*(10), e110268. https://doi.org/10.1371/journal.pone.0110268

McTavish, E. J., Drew, B. T., Redelings, B., & Cranston, K. A. (2017). How and why to build a unified tree of life. *BioEssays*, *39*(11), https://doi.org/10.1002/bies.201700114

Nguyen, V. D., Nguyen, T. H., Tayeen, A. S. M., Dail Laughinghouse, H., Sánchez-Reyes, L. L., Pontelli, E., ... Stoltzfus, A. (2018). *Phylotastic: improving access to tree-of-life knowledge with flexible, on-the-fly delivery of trees. bioRxiv.* https://doi.org/10.1101/419143

Paradis, E., & Schliep, K. (2018). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Pearse, W. D., & Purvis, A. (2013). phyloGenerator: An automated phylogeny generation tool for ecologists. *Methods in Ecology and Evolution*, *4*(7), 692–698. https://doi.org/10.1111/2041-210X.12055

Philippe, H., Zhou, Y., Brinkmann, H., Rodrigue, N., & Delsuc, F. (2005). Heterotachy and long-branch attraction in phylogenetics. *BMC Evolutionary Biology*, *5*, 50. https://doi.org/10.1186/1471-2148-5-50

Pyron, R. A., Burbrink, F. T., & Wiens, J. J. (2013). A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evolutionary Biology*, *13*, 93. https://doi.org/10.1186/1471-2148-13-93

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

Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., ... Alfaro, M. E. (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*, *559*(7714), 392–395. https://doi.org/10.1038/s41586-018-0273-1

Robinson, D. F., & Foulds, L. R. (1981). Comparison of phylogenetic trees. *Mathematical Biosciences*, *53*(1), 131–147. https://doi.org/10.1016/0025-5564(81)90043-2

Schliep, K. P. (2011). phangorn: Phylogenetic analysis in R. *Bioinformatics*, *27*(4), 592–593. https://doi.org/10.1093/bioinformatics/btq706

Shimodaira, H., & Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, *16*(8), 1114. https://doi.org/10.1093/oxfordjournals.molbev.a026201

Shimodaira, H., & Hasegawa, M. (2001). CONSEL: For assessing the confidence of phylogenetic tree selection. *Bioinformatics*, *17*(12), 1246–1247. https://doi.org/10.1093/bioinformatics/17.12.1246

Smith, S. A., & O'Meara, B. C. (2012). treePL: Divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics*, *28*(20), 2689–2690. https://doi.org/10.1093/bioinformatics/bts492

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stoltzfus, A., O'Meara, B., Whitacre, J., Mounce, R., Gillespie, E. L., Kumar, S., ... Vos, R. A. (2012). Sharing and re-use of phylogenetic trees (and associated data) to facilitate synthesis. *BMC Research Notes*, *5*, 574. https://doi.org/10.1186/1756-0500-5-574

Webb, C. O., Ackerly, D. D., & Kembel, S. W. (2008). Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics*, *24*(18), 2098–2100. https://doi.org/10.1093/bioinformatics/btn358

Webb, C. O., & Donoghue, M. J. (2005). Phylomatic: tree assembly for applied phylogenetics. *Molecular Ecology Notes*, *5*(1), 181–183. https://doi.org/10.1111/j.1471-8286.2004.00829.x

Yang, Z. (1996). Among-site rate variation and its impact on phylogenetic analyses. *Trends in Ecology & Evolution*, *11*(9), 367–372. https://doi.org/10.1016/0169-5347(96)10041-0

Zanne, A. E., Tank, D. C., Cornwell, W. K., Eastman, J. M., Smith, S. A., FitzJohn, R. G., ... Beaulieu, J. M. (2014). Three keys to the radiation of angiosperms into freezing environments. *Nature*, *506*(7486), 89–92. https://doi.org/10.1038/nature12872

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: Chang J, Rabosky DL, Smith SA, Alfaro ME. An R package and online resource for macroevolutionary studies using the ray-finned fish tree of life. *Methods Ecol Evol*. 2019;10:1118–1124. https://doi.org/10.1111/2041-210X.13182