

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

DR JONATHAN CHANG (Orcid ID : 0000-0002-3811-1254)

DR STEPHEN SMITH (Orcid ID : 0000-0003-2035-9531)

Article type : Application

Editor : Dr Daniele Silvestro

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

Jonathan Chang^{1*}, Daniel L. Rabosky², Stephen A. Smith², Michael E. Alfaro³

¹ School of Biological Sciences, Monash University, Victoria 3800, Australia

² Museum of Zoology and Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, 48109, USA

³ Department of Ecology and Evolutionary Biology, University of California, Los Angeles, 90095, USA

* Author for correspondence: Jonathan Chang; E-mail: jonathan.chang@monash.edu

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.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/2041-210X.13182](https://doi.org/10.1111/2041-210X.13182)

This article is protected by copyright. All rights reserved

26

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

33

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

39

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

44

45 **Keywords:** websites, Actinopterygii, reproducible research, open data

46

47 **Running head:** Reproducible resources for phylogenetics in fishes

48

49 **Introduction**

50 Phylogenies are fundamental to comparative evolutionary biology, and their use extends to
51 community ecology, conservation biology, ecophysiology, developmental biology, and
52 translational medical research. New phylogenetic information can illuminate open questions in
53 biology, but this work is clouded by the difficulty in inferring phylogenies, especially for non-

54 specialist researchers (Pearse & Purvis, 2013). To avoid these pitfalls, reusing existing
55 phylogenies can make phylogenetic knowledge accessible without requiring researchers to
56 collaborate with phylogenetic experts or learn these methods themselves (Webb & Donoghue,
57 2005; Webb, Ackerly, & Kembel, 2008; Arnold, Matthews, & Nunn, 2010; Magee, May, & Moore,
58 2014). However, surveys of the biological literature estimate that 60-95% of previously-
59 published phylogenetic datasets are no longer accessible (Stoltzfus et al., 2012; Drew et al., 2013;
60 Magee et al., 2014; McTavish, Drew, Redelings, & Cranston, 2017), highlighting the challenge of
61 persistently sharing data and creating a major barrier to new comparative analyses.

62
63 One alternative solution is a "tree of life" approach that centralizes research effort across large
64 groups to create a curated and validated phylogenetic dataset, as opposed to smaller family- or
65 genus-level analyses (McTavish et al., 2017; Beaulieu & O'Meara, 2018). These broad
66 phylogenies, in diverse groups such as mammals, birds, squamate reptiles, angiosperms, and
67 fishes (Bininda-Emonds et al., 2007; Jetz, Thomas, Joy, Hartmann, & Mooers, 2012; Pyron,
68 Burbrink, & Wiens, 2013; Zanne et al., 2014; Rabosky et al., 2018), represent the best target for
69 phylogenetic re-use, as extensive sampling across these broad organismal groups is likely to
70 cover the particular set of species that would interest a taxon-focused researcher.

71
72 Here we present a new community resource and accompanying R package, the Fish Tree of
73 Life, focusing on the ray-finned fishes, the most species-rich group of vertebrates with over
74 33,000 species. We describe this resource, which is based on a recent complete phylogeny
75 (Rabosky et al., 2018), and provide three motivating examples, showing how this large
76 empirical dataset could be used to investigate the common problem of long branch attraction,
77 study a specific taxon in a phylogenetic comparative analysis, and analyze a dataset using
78 methods from phylogenetic community ecology. This work joins other resources such as
79 birdtree.org (Jetz et al., 2012), the Open Tree of Life (Hinchliff et al., 2015), and Phylotastic
80 (Nguyen et al., 2018). We expand on these previous offerings by also providing pre-computed
81 taxonomic subsets with character matrices, phylogenies, fossil calibrations, and diversification
82 rate information in a website and R package.

83 **Functionality**

84 **Website: fishtreeoflife.org**

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

97
98 We have also included a fossil section to our Fish Tree of Life website (Figure 1). This lists all
99 139 fossils used in our analysis, as well as the phylogenetic placement of those fossils on the
100 phylogeny. Each page includes the taxon it calibrates (e.g., crown Acanthuridae), as well as the
101 minimum age, authorities for taxonomic placement and age, and fossil locality. We also show
102 the upper bound of the 95% confidence interval for the estimated age of the Hedman fossil
103 outgroup process (Hedman, 2010), and list the fossil outgroup sequence used to calculate those
104 bounds. Our approach explicitly integrates fossil knowledge in a phylogenetic context suitable
105 for divergence time estimation, while some other resources, such as TimeTree (Hedges, Dudley,
106 & Kumar, 2006) or DateLife (Nguyen et al., 2018), either do not permit reuse or lack detailed
107 fossil taxonomy and locality data. Our compilation could provide an established starting point
108 for analyses that e.g., vary fossil calibrations to estimate their downstream effects on
109 diversification rate inference.

110

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: [Euteleostomorpha](#)

Subcohort: [Protacanthopterygii](#)

Order: [Esociformes](#)

Family: [Esocidae](#)

(a)

```
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)"
```

(b)

111
112 Figure 1. (a) An example of the fossil calibration page, which includes the exact locality and
113 authorities of the fossil, as well as the outgroup sequence used to determine the 95% upper

114 bound on maximum ages. (b) The same data represented as Javascript Object Notation, a
115 machine-readable data format.

116 **R package: `fishtree`**

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

126

127

Function	Data retrieved
<code>fishtree_alignment</code>	Aligned sequences for a taxonomic rank or list of species, optionally splitting by gene partition
<code>fishtree_taxonomy</code>	Information for a taxonomic rank, including a list of species and average diversification rates
<code>fishtree_phylogeny</code>	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.
<code>fishtree_tip_rates</code>	Tip-specific diversification rates for a taxonomic rank or list of species, computed via BAMM (Rabosky, 2014) or DR statistic (Jetz

	et al., 2012)
--	---------------

128 Table 1: An overview of the four major functions in the R package `fishtree`. For all functions that
129 take a named taxonomic rank, any rank higher than family is accepted, including higher taxa,
130 e.g., `Ostariophysi` or `Ovalentaria`.

131 **Example applications**

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

140 **Example: testing long branch attraction across the fish tree of life**

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

146
147 Here we reanalyze the phylogeny by family to determine what portions might have been
148 affected by LBA. If LBA artifacts are present, we predict that the reanalyzed topologies would
149 be more balanced (less pectinate) than the original, globally-analyzed phylogeny. If saturation is
150 causing LBA, we expect that the transition rates would also be faster in the reanalyzed
151 phylogenies. The faster transition rates may cause unrelated taxa to be recovered as sister

152 lineages, as fast molecular evolution can lead to shared mutations that are identical by state, not
153 by descent.

154

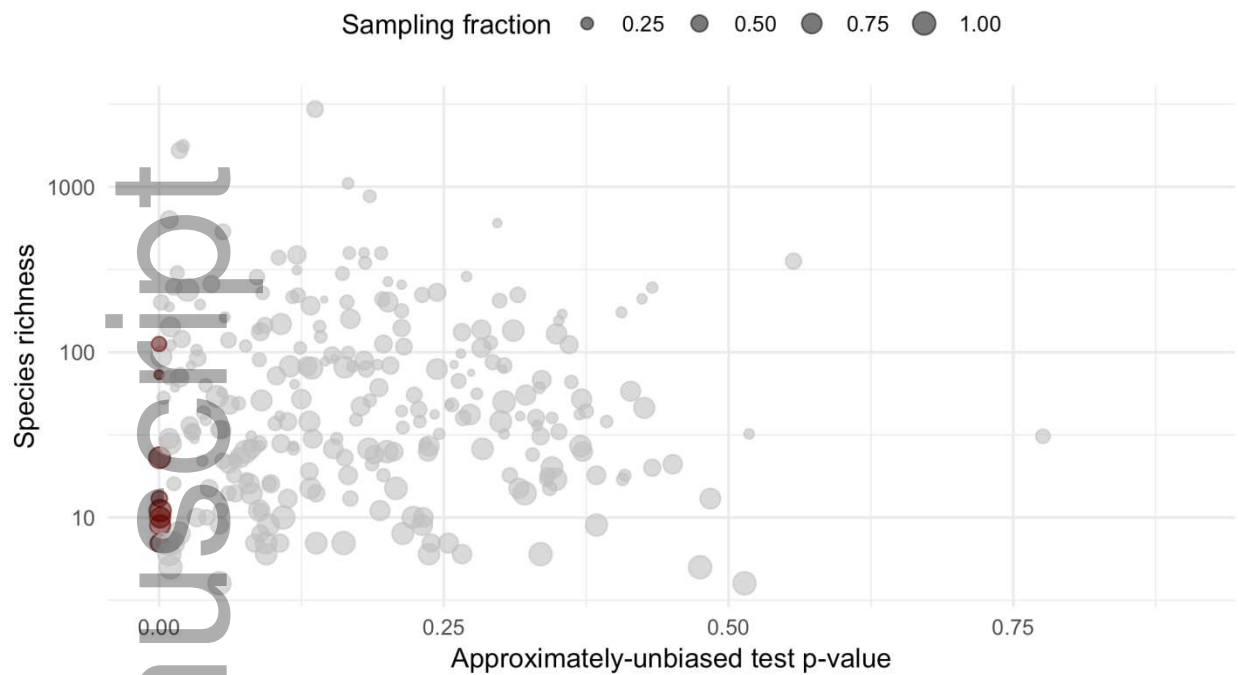
155 We downloaded the alignment for each family with `fishtree_alignment`, and excluded
156 families where three or fewer species had data using `fishtree_taxonomy`. We re-estimated
157 the topology using RAxML v8.2.11 (Stamatakis, 2014) under a partitioned GTR+GAMMA
158 model (Yang, 1996). We refer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

161

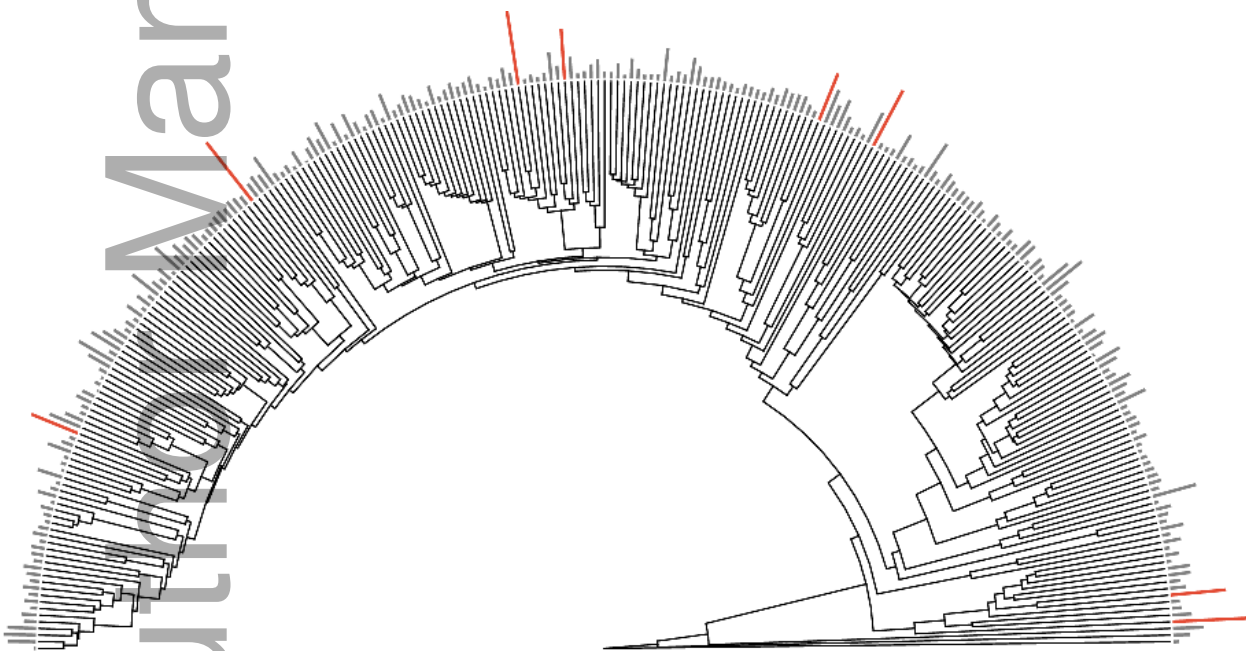
162 For each of the reanalyzed and pruned topologies, we inferred the rates of molecular evolution
163 using the `-f e` option in RAxML. We additionally conducted an approximately-unbiased (AU)
164 test of topologies (Shimodaira & Hasegawa, 1999), using the `-f G` option in RAxML to score
165 per-site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

170

Author Manuscript



171



172

173 Figure 2: The approximately unbiased (AU) test for tree topologies significantly rejected 8 of 268
 174 reanalyzed phylogenies in favor of the original topology, colored in dark red. (top) Alignment
 175 incompleteness, species richness, and their interaction significantly predicted the p-value of the
 176 AU test. (bottom) Skeletal family-level phylogeny of the ray-finned fish tree of life; bar lengths
 177 are the negative log of the AU test p-value. This figure with tips labeled by family is provided
 178 as Figure S1.

179

180 We also computed the normalized Robinson-Foulds (RF) distance (Robinson & Foulds, 1981)
181 and Yule-normalized Colless tree balance metric (Colless, 1982; Blum, François, & Janson, 2006)
182 using `apTreeshape` and `phangorn` (Bortolussi, Durand, Blum, & François, 2006; Schliep,
183 2011). We fit two regression models: a full model that included alignment incompleteness, the
184 log species richness in the family, and the difference in the Colless metric and the RF distance
185 between the pruned and reanalyzed phylogenies, with all interaction terms; and a reduced
186 model that only included the alignment incompleteness, log species richness, and interaction
187 term; both models used the AU test p-value as the response term. A likelihood ratio test
188 supported the less complex model, with all predictors significant at $p < 0.001$.

189

190 Consistent with our prediction, we find that the reanalyzed phylogenies tended to be more
191 balanced (less pectinate) than the pruned topology, measured by the Colless metric (31 of 50,
192 62%). Relative to the pruned topologies, the reanalyzed topologies generally had faster
193 transition parameters and substantially different base composition frequencies (Table 2),
194 suggesting that LBA contributed to the more balanced topologies recovered in the reanalyzed
195 phylogenies. Based on the significant predictors in the likelihood ratio test, we speculate that
196 the larger dataset more robustly parameterizes the substitution model and leads to fewer LBA
197 artifacts in the pruned trees.

198

Parameter	Proportion of 268 reanalyzed 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

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

204 Conclusion

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

212
213 To demonstrate this, we have shown three example use-cases of our resource, one focusing on a
214 broad question in molecular evolution, and the other focused on a comparative phylogenetic
215 hypothesis. In the first example, we made an extremely time-consuming task much easier, as we
216 were able to rapidly import and subset the relevant data into R and focus our efforts on
217 connecting the output from different software and analyzing the results. In the second example,
218 we were able to rapidly test a hypothesis in a comparative context, since `fishtree` was

219 designed to work well within the R phylogenetics ecosystem and all analyses could be
220 conducted without many data cleaning tasks. In the last example, we showed how `fish`
221 `tree` could also be used in a community phylogenetics analysis by testing whether reef fish
222 communities in several ocean basins are phylogenetically clustered or overdispersed.

223
224 As concerns around data curation and cleaning in large data aggregations become increasingly
225 visible in biological research (Franz & Sterner, 2018), the ease of use of the tooling around this
226 large, well-curated dataset provides a framework for how concerns around data quality might
227 be assuaged. Further development of the website and R package will focus on adding more pre-
228 computed analyses and figures, which will provide more starting points for researchers hoping
229 to extend and reuse these resources. Finally, our website and R package can be easily updated
230 as new phylogenetic knowledge becomes available. As the entire process has been standardized
231 and automated inside of a Docker container, any newer ray-finned fish phylogeny can be added
232 as a data file to extend the available data.

233 **Data availability**

234 Our website can be accessed at <https://fishtreeoflife.org>. The R package is available on GitHub,
235 <https://github.com/jonchang/fishtree> as well as CRAN ([https://CRAN.R-](https://CRAN.R-project.org/package=fishtree)
236 [project.org/package=fishtree](https://CRAN.R-project.org/package=fishtree)). Source code and data for the example demonstrations are
237 available from the Dryad Digital Repository (<https://doi.org/10.5061/dryad.6vg974n>) and in the
238 Supporting Information.

239 **Acknowledgements**

240 We thank Matt McGee and two anonymous reviewers for their insightful comments on this
241 manuscript, and Peter Cowman and Tom Near for testing early versions of the website and R
242 package. This work was supported by an Encyclopedia of Life Rubenstein Fellowship (EOL-
243 33066-13) and an NSF Doctoral Dissertation Improvement Grant (DEB-1601830) to JC. Travel
244 support to disseminate this research was provided to JC by UCLA and the Society of Systematic

245 Biologists. This research used computational and storage services associated with the Hoffman2
246 Shared Cluster provided by UCLA Institute for Digital Research and Education's Research
247 Technology Group, as well as computational resources provided by Advanced Research
248 Computing at the University of Michigan, Ann Arbor.

249 **Author Contributions**

250 JC drafted the manuscript, developed the methods, wrote the software and website. SAS, DLR,
251 and MEA assisted with analyses and website design. All authors planned the work and
252 contributed to the final manuscript.

253 **References**

- 254 Arnold, C., Matthews, L. J., & Nunn, C. L. (2010). The 10kTrees website: A new online resource
255 for primate phylogeny. *Evolutionary Anthropology*, 19(3), 114–118.
- 256 Beaulieu, J. M., & O'Meara, B. C. (2018). Can we build it? Yes we can, but should we use it?
257 Assessing the quality and value of a very large phylogeny of campanulid angiosperms.
258 *American Journal of Botany*, 105(3), 417–432.
- 259 Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and
260 Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B*,
261 *Statistical Methodology*, 57(1), 289–300.
- 262 Bergsten, J. (2005). A review of long-branch attraction. *Cladistics*, 21(2), 163–193.
- 263 Bininda-Emonds, O. R. P., Cardillo, M., Jones, K. E., MacPhee, R. D. E., Beck, R. M. D., Grenyer,
264 R., ... Purvis, A. (2007). The delayed rise of present-day mammals. *Nature*, 446(7135), 507–
265 512.
- 266 Blum, M. G. B., François, O., & Janson, S. (2006). The mean, variance and limiting distribution of
267 two statistics sensitive to phylogenetic tree balance. *The Annals of Applied Probability*, 16(4),
268 2195–2214.
- 269 Boettiger, C. (2017). A reproducible R notebook using Docker. In J. Kitzes, D. Turek, & F. Deniz
270 (Eds.), *The Practice of Reproducible Research: Case Studies and Lessons from the Data-Intensive*

271 *Sciences* (pp. 109–118). University of California Press.

272 Bortolussi, N., Durand, E., Blum, M., & François, O. (2006). apTreeshape: statistical analysis of
273 phylogenetic tree shape. *Bioinformatics*, 22(3), 363–364.

274 Bridge, P. D., Roberts, P. J., Spooner, B. M., & Panchal, G. (2003). On the unreliability of
275 published DNA sequences. *The New Phytologist*, 160(1), 43–48.

276 Colless, D. H. (1982). [Review of] Phylogenetics: The Theory and Practice of Phylogenetic
277 Systematics. *Systematic Zoology*, 31(1), 100–104.

278 Drew, B. T., Gazis, R., Cabezas, P., Swithers, K. S., Deng, J., Rodriguez, R., ... Soltis, D. E. (2013).
279 Lost branches on the tree of life. *PLoS Biology*, 11(9), e1001636.

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

282 Hedges, S. B., Dudley, J., & Kumar, S. (2006). TimeTree: a public knowledge-base of divergence
283 times among organisms. *Bioinformatics*, 22(23), 2971–2972.

284 Hedman, M. M. (2010). Constraints on clade ages from fossil outgroups. *Paleobiology*, 36(1), 16–
285 31.

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

289 Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of
290 birds in space and time. *Nature*, 491(7424), 444–448.

291 Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7:
292 improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780.

293 Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., ...
294 Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*,
295 26(11), 1463–1464.

296 Magee, A. F., May, M. R., & Moore, B. R. (2014). The dawn of open access to phylogenetic data.
297 *PloS One*, 9(10), e110268.

298 McTavish, E. J., Drew, B. T., Redelings, B., & Cranston, K. A. (2017). How and Why to Build a
299 Unified Tree of Life. *BioEssays*, 39(11). doi:10.1002/bies.201700114

300 Nguyen, V. D., Nguyen, T. H., Tayeen, A. S. M., Dail Laughinghouse, H., Sánchez-Reyes, L. L.,
301 Pontelli, E., ... Stoltzfus, A. (2018, September 17). *Phylotastic: improving access to tree-of-life*
302 *knowledge with flexible, on-the-fly delivery of trees. bioRxiv. doi:10.1101/419143*

303 Paradis, E., & Schliep, K. (2018). ape 5.0: an environment for modern phylogenetics and
304 evolutionary analyses in R. *Bioinformatics. doi:10.1093/bioinformatics/bty633*

305 Pearse, W. D., & Purvis, A. (2013). phyloGenerator: an automated phylogeny generation tool for
306 ecologists. *Methods in Ecology and Evolution / British Ecological Society, 4(7), 692–698.*

307 Philippe, H., Zhou, Y., Brinkmann, H., Rodrigue, N., & Delsuc, F. (2005). Heterotachy and long-
308 branch attraction in phylogenetics. *BMC Evolutionary Biology, 5, 50.*

309 Pyron, R. A., Burbrink, F. T., & Wiens, J. J. (2013). A phylogeny and revised classification of
310 Squamata, including 4161 species of lizards and snakes. *BMC Evolutionary Biology, 13, 93.*

311 Rabosky, D. L. (2014). Automatic detection of key innovations, rate shifts, and diversity-
312 dependence on phylogenetic trees. *PloS One, 9(2), e89543.*

313 Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., ... Alfaro, M. E.
314 (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature,*
315 *559(7714), 392–395.*

316 Robinson, D. F., & Foulds, L. R. (1981). Comparison of phylogenetic trees. *Mathematical*
317 *Biosciences, 53(1), 131–147.*

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

319 Shimodaira, H., & Hasegawa, M. (1999). Multiple Comparisons of Log-Likelihoods with
320 Applications to Phylogenetic Inference. *Molecular Biology and Evolution, 16(8), 1114–1114.*

321 Shimodaira, H., & Hasegawa, M. (2001). CONSEL: for assessing the confidence of phylogenetic
322 tree selection. *Bioinformatics, 17(12), 1246–1247.*

323 Smith, S. A., & O'Meara, B. C. (2012). treePL: divergence time estimation using penalized
324 likelihood for large phylogenies. *Bioinformatics, 28(20), 2689–2690.*

325 Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of
326 large phylogenies. *Bioinformatics, 30(9), 1312–1313.*

327 Stoltzfus, A., O'Meara, B., Whitacre, J., Mounce, R., Gillespie, E. L., Kumar, S., ... Vos, R. A.
328 (2012). Sharing and re-use of phylogenetic trees (and associated data) to facilitate synthesis.

- 329 *BMC Research Notes*, 5, 574.
- 330 Webb, C. O., Ackerly, D. D., & Kembel, S. W. (2008). Phylocom: software for the analysis of
331 phylogenetic community structure and trait evolution. *Bioinformatics*, 24(18), 2098–2100.
- 332 Webb, C. O., & Donoghue, M. J. (2005). Phylomatic: tree assembly for applied phylogenetics.
333 *Molecular Ecology Notes*, 5(1), 181–183.
- 334 Yang, Z. (1996). Among-site rate variation and its impact on phylogenetic analyses. *Trends in*
335 *Ecology & Evolution*, 11(9), 367–372.
- 336 Zanne, A. E., Tank, D. C., Cornwell, W. K., Eastman, J. M., Smith, S. A., FitzJohn, R. G., ...
337 Beaulieu, J. M. (2014). Three keys to the radiation of angiosperms into freezing
338 environments. *Nature*, 506(7486), 89–92.

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](#)

(a)

[JSON](#) [Raw Data](#) [Headers](#)

[Save](#) [Copy](#) [Collapse All](#) [Expand All](#)

0:

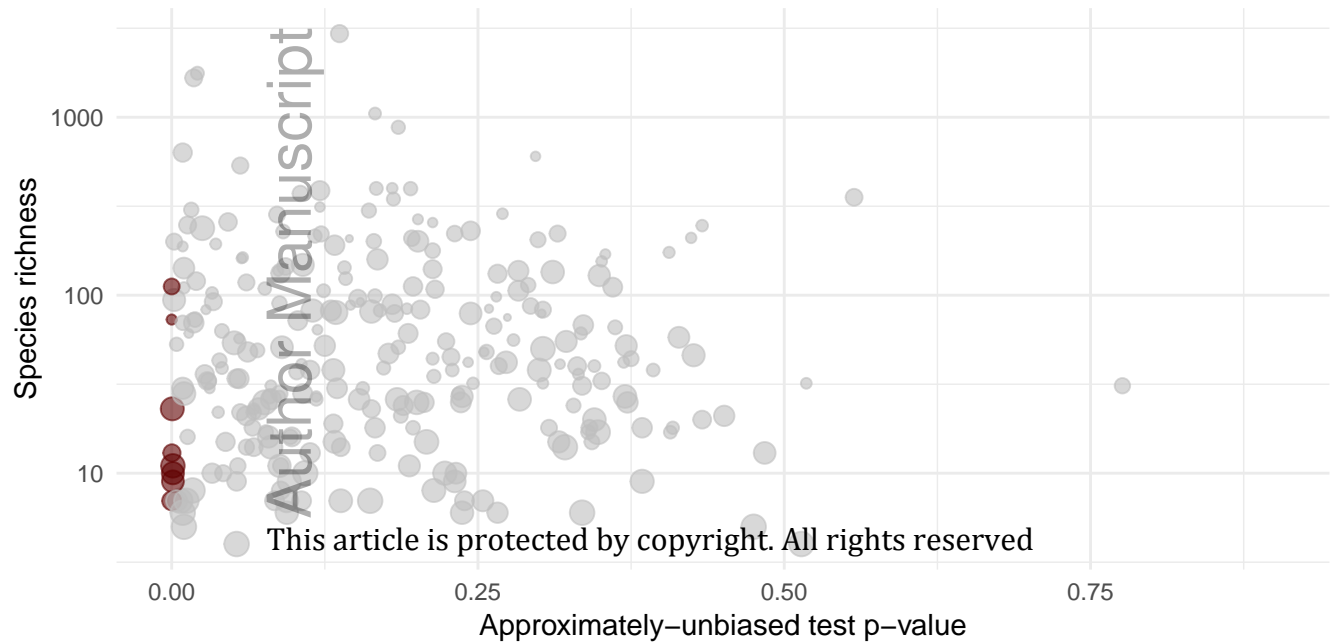
```
clade: "Crown Esox"
fossil: "Esox kronneri"
left: "Esox_lucius"
right: "Esox_niger"
min: "51.57"
max: "114.0734535"
locality: "Fossil Butte Member, Green River Formation (F-2 locality), Wyoming, USA"
authority: "Grande (1999)"
age_authority: "Smith et al. (2008)"
```

(b)

Author Manuscript

This article is prote

mee3_13182_f2a.pdf
Sampling fraction - 0.25 0.50 0.75 1.00



mee3_13182_f2b.pdf

Author Manuscript

This article is protected by copyright. All rights reserved

