

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

Supporting Information

Jonathan Chang, Daniel L. Rabosky, Stephen A. Smith, Michael E. Alfaro

This document is Supporting Information for: Chang, J, Rabosky, DL, Smith, SA, Alfaro, ME. (2019). An R package and online resource for macroevolutionary studies using the ray-finned fish tree of life. *Methods in Ecology and Evolution*. doi:[10.1111/2041-210x.13182](https://doi.org/10.1111/2041-210x.13182)

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

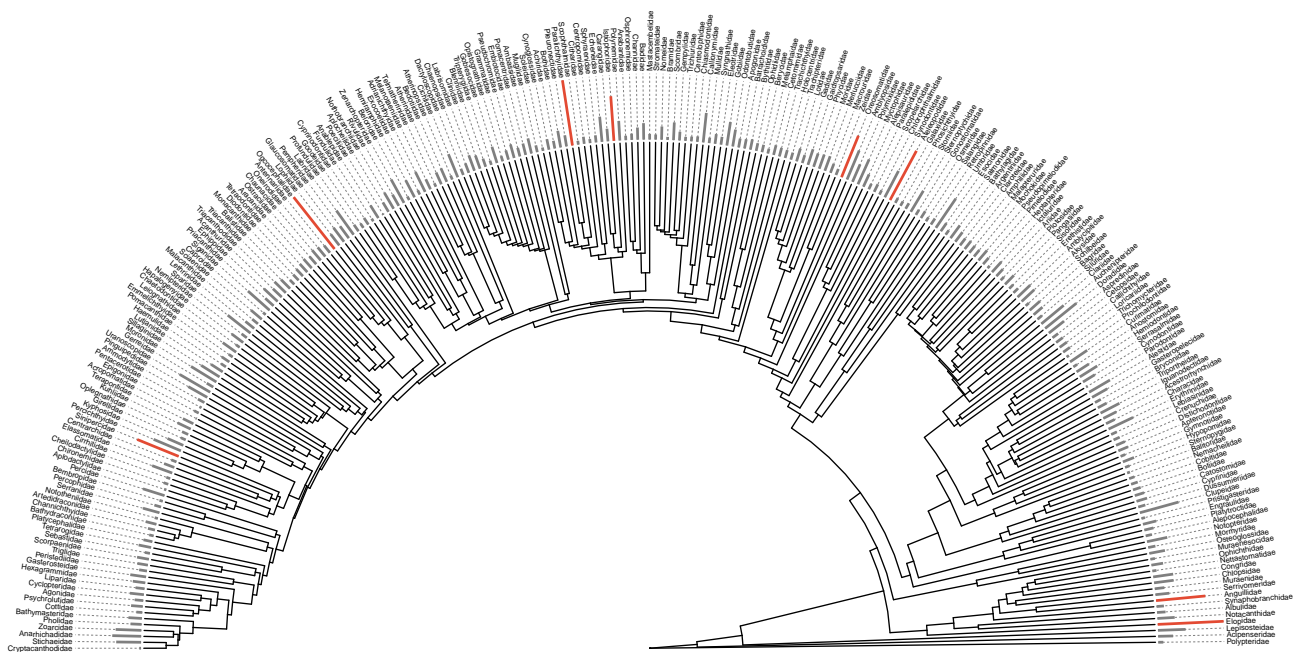


Figure S1: Skeletal family-level phylogeny of the ray-finned fish tree of life; bar lengths are the negative log of the AU test p-value. This figure was generated with the `MonoPhy` and `ggtree` packages (Schwery & O'Meara, 2016; Yu, Smith, Zhu, Guan, & Lam, 2016) and is a labelled version of Figure 2b from the main text.

Example: effects of habitat on diversification rate

A common hypothesis tested in comparative methods is whether habitat shifts drive rates of diversification in various groups. For example, Santini et al. (2013) tested whether reef-associated pufferfishes

enjoyed faster rates of speciation compared to their non-reef relatives. Here we replicate their analyses primarily to demonstrate the ease of conducting a comparative analysis, but our new dataset is also larger and new analysis techniques such as `hisse` (Beaulieu & O’Meara, 2016) could more robustly examine the original `diversitree` analyses (Maddison, Midford, & Otto, 2007).

We downloaded the subset of the Fish Tree of Life corresponding to the pufferfishes with `fishtree_phylogeny`, and matched the phylogeny to speciation rates using `fishtree_tip_rates`. We also retrieved habitat data from Fishbase using the `rfishbase` package (Boettiger, Lang, & Wainwright, 2012; Froese, Pauly, & others, 2018).

We tested the hypothesis of reef-association promoting faster rates of speciation by using `hisse` to compare 4 models: a BiSSE model (Maddison et al., 2007) where reef-association could impact diversification, a BiSSE null model, a `hisse` model where diversification depends on either reef-association or a hidden trait, and the `hisse` two-state null model where only a hidden trait affects diversification (Beaulieu & O’Meara, 2016).

The best supported model based on the corrected Akaike Information Criterion (AICc) is a `hisse` 2-state null model (Table S1), where speciation rate is independent of habitat type, a result that is concordant with the findings from Santini et al. (2013). All of the code for this analysis is available as a vignette in the R package.

Table S1: The `hisse` 2-state null model is best supported (lowest AICc), supporting the original conclusions of Santini et al. (2013).

Diversification model	Δ AICc	AICc	lnL
<code>hisse</code> CID-2 null	0	1973.516	-982.6720
<code>hisse</code> full	4.314	1977.830	-982.7325
BiSSE null	10.077	1983.593	-988.7449
BiSSE full	12.175	1985.691	-988.7594

Example: phylogenetic community ecology with reef fishes and oceans

One common use case for phylogenies is in analyses of community assembly, to test whether co-occurring species are more closely related to another (clustering) or more evenly dispersed (overdispersed) relative to a null expectation (Webb, 2000). Here we test whether communities of reef-associated fishes are more closely related than expected in three ocean basins: the Atlantic, Pacific, and Indian oceans. We downloaded the entire phylogeny of ray-finned fishes using `fishtree_phylogeny`, and matched phylogenetic data to habitat and ocean basin data from Fishbase using the `rfishbase` package (Figure S2; Boettiger et al., 2012; Froese et al., 2018).

We then use the R package `picante` (Kembel et al., 2010) to test two measures of clustering and overdispersion: the mean pairwise distance (MPD) and the mean nearest taxon distance (MNTD). The MPD is thought to be more sensitive to patterns closer to the tips, while the MNTD is thought to more closely reflect patterns towards the root of the phylogeny (Webb, Ackerly, & Kembel, 2008). We compute this

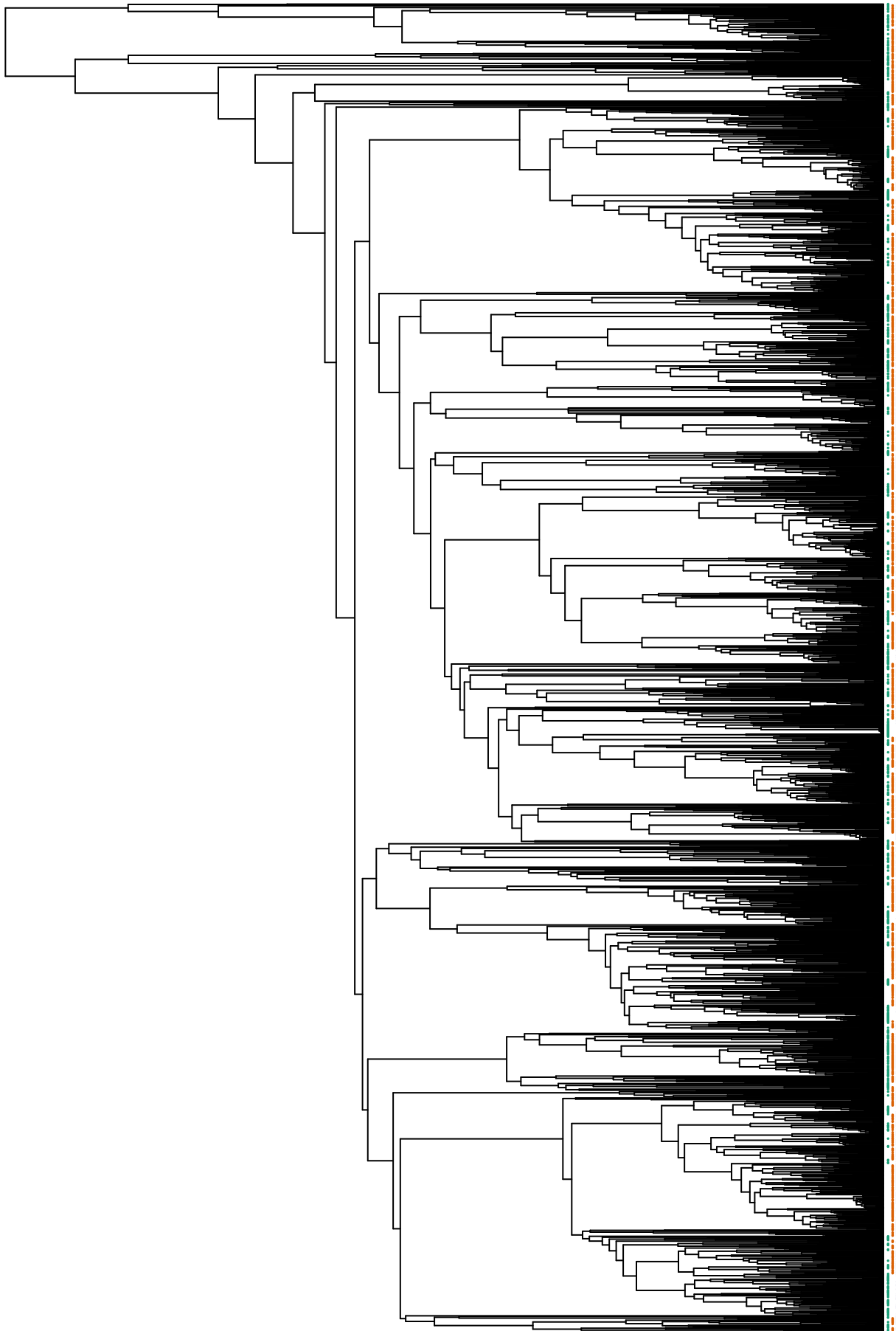


Figure S2: A pruned phylogeny showing the distribution of reef-associated fish across several oceans. The dots next to each tip represent which ocean(s) a species is present in. From left to right, the dots represent the Atlantic, Indian, and Pacific Oceans.

empirical statistic and, as a null model, perform a randomization test by shuffling the taxon labels and recomputing these statistics, and present the results in Table S2. All of the code for this analysis is available as a vignette in the R package.

Table S2: A randomization test showed that there was significant clustering with respect to the MNTD statistic for all oceans, and significant overdispersion with the MPD statistic for the Atlantic Ocean.

Area	Number of taxa	p_{MPD}	p_{MNTD}
Atlantic Ocean	637	1.00	0.01
Indian Ocean	1222	0.90	0.01
Pacific Ocean	1511	0.62	0.03

References

- Beaulieu, J. M., & O'Meara, B. C. (2016). Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Systematic Biology*, 65(4), 583–601.
- Boettiger, C., Lang, D. T., & Wainwright, P. C. (2012). rfishbase: exploring, manipulating and visualizing FishBase data from R. *Journal of Fish Biology*, 81(6), 2030–2039.
- Froese, R., Pauly, D., & others. (2018). FishBase. Fisheries Centre, University of British Columbia.
- 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.
- 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.
- Santini, F., Nguyen, M. T. T., Sorenson, L., Waltzek, T. B., Lynch Alfaro, J. W., Eastman, J. M., & Alfaro, M. E. (2013). Do habitat shifts drive diversification in teleost fishes? An example from the pufferfishes (Tetraodontidae). *Journal of Evolutionary Biology*, 26(5), 1003–1018.
- Schwery, O., & O'Meara, B. C. (2016). MonoPhy: a simple R package to find and visualize monophyly issues. *PeerJ Computer Science*, 2, e56.
- Webb, C. O. (2000). Exploring the phylogenetic structure of ecological communities: An example for rain forest trees. *The American Naturalist*, 156(2), 145–155.
- 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.
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T.-Y. (2016). 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.