OpenTree: A Python Package for Accessing and Analyzing Data from the Open Tree of Life

Emily Jane McTavish[1,*], Luna Luisa Sánchez-Reyes[1], and Mark T. Holder[2,3]

[1]*Department of Life and Environmental Sciences, University of California, Merced, CA 95343, USA;* [2]*Department of Ecology and Evolutionary Biology,*
University of Kansas, Lawrence, KS 66045, USA; [3]*Biodiversity Institute, University of Kansas, Lawrence, KS 66045, USA*
*Correspondence to be sent to: Department of Life and Environmental Sciences, University of California, Merced, CA 95343, USA;
E-mail: ejmctavish@ucmerced.ed.

Abstract.—The Open Tree of Life project constructs a comprehensive, dynamic, and digitally available tree of life by synthesizing published phylogenetic trees along with taxonomic data. Open Tree of Life provides web-service application programming interfaces (APIs) to make the tree estimate, unified taxonomy, and input phylogenetic data available to anyone. Here, we describe the Python package opentree, which provides a user friendly Python wrapper for these APIs and a set of scripts and tutorials for straightforward downstream data analyses. We demonstrate the utility of these tools by generating an estimate of the phylogenetic relationships of all bird families, and by capturing a phylogenetic estimate for all taxa observed at the University of California Merced Vernal Pools and Grassland Reserve.[Evolution; open science; phylogenetics; Python; taxonomy.]

Evolutionary history provides a framework to link all life on earth. However, it is not easy to access accurate, up-to-date phylogenetic relationships for arbitrary sets of taxa of interest, even if phylogenetic estimates for those taxa have been made and published (Drew et al. 2013; McTavish et al. 2017). Individual phylogenetic estimates are not comprehensive, and therefore seldom contain all taxa of interest. Taxonomic relationships, while comprehensive, provide a coarse, and often outdated, picture of shared ancestry. The Open Tree of Life project (OpenTree) provides a reproducible framework for accessing up-to-date and expert-knowledge-based evolutionary relationships for arbitrary sets of taxa across the entire tree of life. All data in OpenTree are freely available via application programming interfaces (API's). The package opentree provides a user friendly Python interface to access these data. In addition, opentree is packaged with a set of tutorials and scripts to make common downstream analyses straightforward.

Description

opentree is a Python package for accessing and analyzing data from the OpenTree of Life project. Open Tree of Life stores a wealth of taxonomic and phylogenetic data gathered together in an open-access interoperable framework. The current synthetic tree (OpenTreeOfLife et al. 2019a) comprises 2.4 million tips. Most of the tips of the tree represent species, but some are infraspecific taxa. The framework of this tree is provided by a unified taxonomy (Rees and Cranston 2017; OpenTreeofLife et al. 2019b). This taxonomy links unique identifiers across many online taxonomic resources, including NCBI (Federhen 2012), GBIF (GBIF Secretariat 2019), as well as user contributed taxonomic amendments contained in https://github.com/OpenTreeOfLife/amendments-1. These taxonomic relationships are refined by evolutionary estimates from 1216 published papers including 87,000 tip taxa (OpenTreeOfLife et al. 2019a; Redelings and Holder 2017). The Open Tree data store, "Phylesystem" (McTavish et al. 2015) contains 4500 published studies, including those incorporated in the latest version of OpenTree's synthetic tree. Phylogenies in the Phylesystem database have mappings between their tip names and unique OpenTree taxonomic identifiers. There are several reasons trees are included in the "Phylesystem" data store but not the synthetic tree. The complete curation of trees requires human intervention and confirmation at several steps, including vetting taxonomic name resolution service matches and correctly rooting trees, as many output files are shared as unrooted trees. In addition, some of the trees in Phylesystem describe gene tree level or within species relationships, which are not appropriate for the species (and named subspecies) level relationships captured in the synthetic tree. There is also lag between study upload, which makes studies immediately available via the API, and the time to rebuild the entire synthetic tree, which is released in versioned increments (current version, 12.3) The number of input studies increases in each successive synthesis tree release.

All of these data and tools are freely accessible via API calls, documented at https://github.com/OpenTreeOfLife/germinator/wiki/Open-Tree-of-Life-Web-APIs. opentree provides a user-friendly wrapper for calling these APIs from the command line or from a Python interactive session. In addition, it converts OpenTree data between commonly used file formats and data types. This package allows users to generate data objects for use in DendroPy, a widely cited Python phylogenetic computing library (Sukumaran and Holder 2010).

`opentree` incorporates in Python the functionality available in rotl, an R package to interact with the Open Tree of Life data (Michonneau et al. 2016). Additionally, `opentree` expands the toolset available for working with OpenTree's unified taxonomy (Rees and Cranston 2017), as well as additional downstream analysis and interoperability tools.

`rotl` has already been cited 132 times in the 4 years since its publication, demonstrating a demand for accessible user access to OpenTree data and tools. By providing a Python package to interact with OpenTree's data and services, we make it straightforward for Python users to access and analyze these data, generally improving the accessibility of OpenTree.

The `opentree` Python package for OpenTree facilitates linking all these data with the stable of other Python biodiversity informatics tools.

SERVICES PROVIDED BY OPENTREE

OpenTree APIs are divided into three main categories: synthetic tree, taxonomy and taxonomic name resolution, and study search. Many analyses integrate calls from each of these categories. The `opentree` package links across these services to facilitate calls to commonly used OpenTree APIs. Some example calls are described here, but all methods and scripts are fully documented, including examples and return formats, at https://opentree.readthedocs.io.

Synthetic Tree.—The OpenTree synthetic tree contains all 2.4 million taxa in the OpenTree taxonomy, with relationships for 87,000 taxa informed by 1216 studies. Each branch in the tree is informed by published phylogenetic relationships, where they are present in the curated data store, or by taxonomic relationships where no phylogenetic data are available. For each node in the synthetic tree, the API returns identifiers for all the phylogenies in the synthesis pipeline which support or conflict with that node. Each node is uniquely labeled. If the descendants of a node align with named taxonomic group, the taxon identifier is applied to the node. If the node does not correspond to a taxon named in the OpenTree taxonomy, the node is labeled using a phyloreference (Parr et al. 2012), a unique name identifying that node as the most recent common ancestor of two named taxa. `opentree` users can easily access evolutionary estimates for arbitrary sets of taxa. The web-service response also includes the published phylogenetic estimates which underlie those inferences. The `opentree` wrapper captures and formats these citations to make providing appropriate credit for these synthetic induced subtree estimates straightforward. Users can also access full synthetic subtrees subtending any individual node.

Taxonomy and Taxonomic Name Resolution.—The OpenTree taxonomy not only provides a scaffold for the synthetic tree but is also a valuable resource in its own right. Matching names is a key hurdle in bioinformatics. Correct taxon names change through time, and spelling and typographical errors can propagate through bioinformatic resources. Thus, demanding exact matching of names from different sources can be too stringent and fail to match the same taxon. However, different names can be very close in spelling to one another. So, tolerating misspellings makes it easy to accidentally match names that should refer to two distinct taxa. The OpenTree taxonomy (Rees and Cranston 2017; OpenTreeofLife et al. 2019b) provides a link between the unique identifiers generated by several large scale online taxonomic resources [GBIF, NCBI, Silva, Worms], as well as all known name synonymies provided by those resources. The OpenTree taxonomic name resolution service (TNRS) searches this full taxonomy and returns exact or fuzzy matches from input names string to taxa and their unique taxonomic identifiers. This TNRS forms a link between human readable name strings, and rigorous unique identifiers.

`opentree` wraps the OpenTree taxonomy and TNRS APIs for ease of integrating taxonomy and TNRS queries with downstream analyses. In addition, `opentree` provides helpers for quickly searching the text download of the taxonomy, which can be more efficient for bulk queries.

Study Search.—The OpenTree datastore contains 4468 published phylogenetic studies, including 9395 phylogenies (as of December 4, 2020). These studies and phylogenies are indexed on a number of properties including author name, curator name, and publication information. In addition, the tips of these trees are mapped to the unified OpenTree taxonomy making comparisons among estimates of relationships and searching for taxa of interest straightforward. This allows searching of studies based on taxa contained in the study. Importantly, this search does not rely on string-matching of what the taxonomic name was at the time of publication—rather, it leverages the full suite of synonymies gathered across the input taxonomies to find equivalent taxa across studies, even if the canonical name has changed between publications. The indexing of these phylogenies is taxonomically explicit. So, for example, a search for "canidae" will find phylogenies with taxa contained in the taxonomic group Canidae, even if the term "canidae" does not itself appear in the phylogeny or tips. Based on the results of these searches, studies can be viewed in a browser on the OpenTree curator site, or the phylogenies themselves can be downloaded for comparisons or other downstream use.

In addition, as the tips of each study are mapped by curators to taxonomic identifiers in the OpenTree taxonomy, comparing the relationships represented in input phylogenies to taxonomic relationships and to taxonomy is straightforward. The browser based tree viewer has a graphical visualization of this concordance and conflict. `opentree` provides a wrapper for this conflict functionality, which makes it straightforward to assess what taxon definitions and evolutionary

relationships a phylogenetic estimate agrees with and conflicts with. This functionality can also be applied to local phylogenies for which the tips have been matched to taxonomy. This allows users to assess concordance and conflict with previous inferences in prepublication phylogenies, even without sharing them to the publicly available OpenTree database (McTavish et al. 2015; Sanchez Reyes et al. 2020).

BIOLOGICAL EXAMPLES

There are a plethora of downstream applications of this linked set of resources. We highlight two examples based on user queries.

A Phylogeny of All Bird Families

A full Jupyter notebook tutorial demonstrating how to access a synthetic tree of all bird families is packaged with the software at https://github.com/OpenTreeOfLife/python-opentree /blob/main/docs/notebooks/TreeOfBirdGenera.ipynb. Capturing evolutionary information at large scales is often simplified by using arbitrary taxonomic cutoffs. While the OpenTree taxonomy is not rank focused, it does track rank information from component taxonomies. By searching the OpenTree taxonomy for families in birds, we find that there are 390 listed bird families, 196 of which are included in the synthetic tree. Groups are excluded from the synthetic tree for a few potential reasons, the most common of which is that all members of the group are extinct, and we have no curated published studies providing information about the correct evolutionary relationships. Placements of fossil taxa based only on taxonomy tend to be unstable, and the OpenTree synthesis procedure excludes taxa if the taxon is not present in at least one phylogenetic input. These families can be included in later synthetic trees if new studies or phylogenies are added to the Phylesystem database, providing information on their relationships. Other taxa are excluded from synthesis if issues have been raised about their taxonomic validity, such as if the name corresponds to a family that is "barren," that is, it contains no species in the OpenTree taxonomy, or because the name was judged to be invalid by the OpenTree taxonomy merging software (Rees and Cranston 2017).

If we request an induced subtree from OpenTree's synthetic tree for these 164 taxa, we get back an output subtree that has 150 tips. The return value also includes a list of nonmonophyletic taxa. Some of the nonmonophyletic taxa map to internal nodes on our output subtree. In those cases, input phylogenies are telling us that these "families" are paraphyletic with respect to other families. Which studies contest the monophyly of a taxonomic clade can be easily accessed through the browser (e.g., https:// tree.opentreeoflife.org/opentree/argus/ottol@603925)

or via the opentree wrapper using queries to opentree.synth_subtree. Figure 1 shows the topology of 130 monophyletic bird families plus most recent common ancestors (MRCA's) of 20 additional nonmonophyletic families as tips. The other 14 taxa are nonmonophyletic families for which the MRCA is an internal node on the output subtree.

These families contain from 1 to 920 total descendant tip taxa (species and subspecies). Across all 150 families, 10,357 descendant tip taxa are captured by the relationships shown in this subtree. Figure 1 displays the number of descendant taxa in each family as a heat map, with log of the number of descendants displayed on the left, and the actual number of descendants on the right. This display demonstrates that the use of "families" is not a very even way to break up biodiversity across birds.

Based on OpenTree's phylogenetic synthesis algorithm (Redelings and Holder 2017), each branch in the synthetic tree is supported by either taxonomy alone (where there are no input phylogenetic studies that traverse that branch), or by one or more input phylogenetic studies. The source of support for each node in the synthetic tree can be interrogated using a synth_node_info function call from the OpenTree class in the package here presented. While each branch must be supported by taxonomy or at least one input study from Phylesystem, where multiple inputs traverse a branch, there can be conflict among studies. OpenTree's synthesis algorithm is greedy, and the synthetic tree will display the branch supported by the highest ranked study included in synthesis. The synth_node_info function will return not only which studies support a branch but also which studies have relationships which conflict with that branch. In Figure 1, support or conflict for each branch is displayed by the intensity of green and red coloration, respectively. Some branches in this subtree are supported by 20 studies, and a few show conflict with up to 17 other studies. Of the 443 branches in this subtree, 422 are supported by at least one input phylogenetic study, and the other 21 are based on taxonomic relationships.

It is important to note that OpenTree's synthetic tree shows only topology. When combining taxonomy, and phylogenetic branches from across studies with vastly different data types, merging branch lengths is not meaningful. For downstream analyses requiring branch lengths, users have scaled topologies inferred from OpenTree using a variety of data types and rate-smoothing approaches (e.g. Eastman et al. 2013; Uyeda et al. 2017; Smith and Brown 2018; Allen et al. 2019; Li et al. 2019; Jantzen et al. 2019; Sánchez-Reyes and O'Meara 2019; Smith and Walker 2019; Geffroy et al. 2020).

Linking Data from the Global Biodiversity Information Facility (GBIF) with Phylogenetic Information from Open Tree of Life

The University of California (UC), Merced has a natural reserve directly adjacent to campus, which
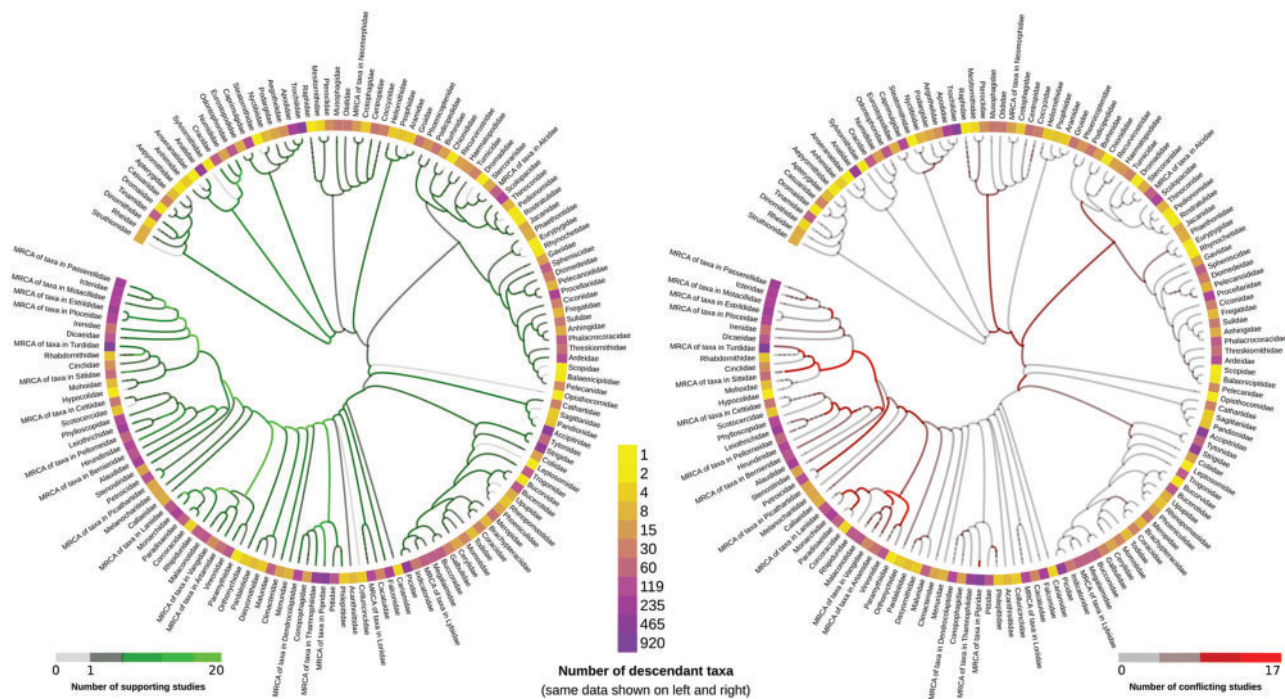
FIGURE 1. Phylogenetic relationships of 150 bird families based on the latest OpenTree synthetic tree (v12.3). For families which are not monophyletic according to published phylogenies, tips for those families are labeled with "MRCA of taxa in X family name." Heat maps show the number of tip taxa descendants in OpenTree within each tip. Branch colors show the number of input studies which support (left, green) or conflict with (right, red) each inferred branch in the synthetic tree. Branch lengths are arbitrary. A total of 64 published phylogenies underlie the relationships in this subtree (citations in supplemental information). Figure created using the interactive Tree Of Life (iTOL) v4 (Letunic and Bork 2019).

contains several vernal pools. These vernal pools create a unique habitat which allows native species to thrive, and the proximity to campus allows undergraduate classes to experience this ecosystem on field trips which can be accomplished during class time. A species list for the reserve and adjacent campus areas is available through the Global Biodiversity Information Facility (GBIF) website (GBIF Secretariat 2019). GBIF provides a repository for species occurrence data tracked in a variety of data stores, including bird observations from eBird (Sullivan et al. 2009), community science observations from iNaturalist (www.inaturalist.org), and several other resources. A full tutorial demonstrating how to access a subtree for a GBIF data download is included with this package at https://github.com/OpenTreeOfLife/python-opentree /blob/main/docs/notebooks/gbif/GBIF_to_OpenTree. ipynb.

We downloaded the full list of animal observations from the UC Merced Vernal pools reserve from GBIF (GBIF Secretariat 2019). This data download comprised 6709 records from 223 species. Using the GBIF unique taxon identifiers, 201 of these species could be directly matched to taxa in the OpenTree taxonomy using opentree.taxon_info(source_id = gbif unique identifier). This direct matching captures exact one to one relationships between these taxonomies and avoids slow and potentially error prone string matching. Nineteen taxa had updated identifiers in

GBIF since the most recent reconciliation between the GBIF taxonomy and the OpenTree taxonomy, and were assigned OpenTree taxon identifiers based on exact string matches. There were two taxa "*P. abortivum* St." and "*Ichneumon cupitus* Cresson 1877," which were not found in the OpenTree taxonomy, and were dropped from the analysis.

Using this set of 223 OpenTree unique identifiers, an induced synthetic tree for these taxa can be downloaded (Fig. 2). This synthetic tree is supported by 160 individual published trees (citations in supplemental information).

For researchers, working in the vernal pools reserve, this subtree also provides the necessary information for community phylogenetic analyses. Li et al. (2019) demonstrated that synthetic phylogenies from the OpenTree project perform well in community phylogenetic studies. By providing ready access to these estimates, based on 160 previously published phylogenies, opentree makes basing ecological analyses in an accurate evolutionary framework straightforward.

The ability to build a phylogeny of local taxa is also a valuable pedagogical tool. One of us (EJM) used this phylogeny to discuss the diversity of life of animal life as part of a class exercise on vernal pools ecology and evolution, in an undergraduate evolution class. Students visited the UC Merced Vernal Pools and Grassland Reserve, and then explored the evolutionary relationships of all the animal

FIGURE 2. Evolutionary relationships between all animal taxon records in the UC Merced Vernal Pools and Grassland Reserve. Branch lengths are arbitrary. A total of 160 published phylogenies underlie the relationships in this tree (citations in supplemental information). Figure created using iTOL v4 (Letunic and Bork 2019).

species recorded as observed in the reserve. There are several threatened and endangered species on the vernal pools reserve, including two species of fairy shrimp, *Branchinecta lynchii* (threatened) and *Branchinecta mesovallensis* (endangered). By working with a subtree of taxa found on and around campus, tree thinking examples in class can have a direct connection for students. For example, this subtree (Fig. 2) shows that the genus of dabbling ducks, *Anas*, does not form a monophyletic group. Walking the tree of life has been demonstrated to be an effective way to get students to understand the connections among different lineages of

life on earth (Ballen and Greene 2017). Walking through this subtree, and labeling major animal groups allows students to connect to the diversity of animal life based on the actual species surrounding them, rather than arbitrary textbook examples.

DISCUSSION

The OpenTree project makes available a synthetic tree across 2.4 million taxa, as well as thousands of peer-reviewed and published phylogenies that together can be reused for applications from scientific discovery to

education, conservation and outreach (Rosindell and Harmon 2012; Stoltzfus et al. 2013; McTavish et al. 2015; Wong and Rosindell 2020). While several tools can generate phylogenies from DNA sequence data mined from the GenBank genetic database (e.g., Sanderson et al. 2008; Smith et al. 2009; Antonelli et al. 2017; Sanchez Reyes et al. 2020), for regions of the tree where phylogenetic estimates already exist, relying on published peer-reviewed inferences of relationships is more efficient and accurate than generating new estimates (Owen et al. 2015; Brown et al. 2017; Smith and Brown 2018; Ewers-Saucedo et al. 2019). Broad taxon sampling coverage is key to correctly estimate phylogenetic diversity (Park et al. 2018; Jantzen et al. 2019), and genetic data are usually not available for all taxa in a community or a biological group of study.

Many researchers have used the OpenTree's synthetic tree and its phylogenetic and taxonomic databases to answer their research questions across the breadth of both the tree of life, and of biological research domains. In contrast to existing phylogenetic databases, (Piel et al. 2000), the links across studies provided by OpenTree and its synthesis algorithm make it possible to extract inferences for an arbitrary subset of taxa and combine inferences from trees with few or no overlapping tips. OpenTree services have been used to consolidate evolutionary estimates across taxa (freshwater crayfish Owen et al. 2015; birds Brown et al. 2017; barnacles Ewers-Saucedo et al. 2019; Ewers-Saucedo and Pappalardo 2019, mammals Uyeda et al. 2017). OpenTree's large scale synthetic subtrees have provide the opportunity to study a variety of trait evolution topics such as the phylogenetic distribution of organism's life history traits at large scales (Tarka et al. 2018; Healy et al. 2019; Capdevila et al. 2020), antipredator behavior associations with human interactions (Geffroy et al. 2020), effects of anthropogenic noise on animal behavior (Kunc and Schmidt 2019), phylogenetic patterns of global wildlife trade (Fukushima et al. 2020), miniaturization in insects (Polilov and Makarova 2017), shoot flammability evolution (Cui et al. 2020), and host-symbiont dependence (Fisher et al. 2017).

In molecular biology, OpenTree's large scale synthetic subtrees have been used to understand the evolution of gene families and taxa across prokaryotes and eukaryotes, to improve orthology prediction for phylogenetic applications Boeckmann et al. (2015). Herrera et al. (2015) used an OpenTree's synthetic subtree to predict RAD-seq marker cleavage site numbers across the eukaryotic tree of life to guide the design of genome-wide genotypic and sequencing projects. Miller et al. (2020) found that codon pairing biases are conserved across the diversity of the tree of life and can be used for phylogenomic analyses.

OpenTree's synthetic tree has been proved to provide a useful phylogenetic backbone for community phylogenetic analyses (Jantzen et al. 2019; Li et al. 2019).

This opentree Python toolkit expands user access to the OpenTree API's, and provides tutorials on how to easily access up-to-date evolutionary information across the entire tree of life, which will continue to facilitate biological research and outreach.

AVAILABILITY

opentree is fully open source with a CC0 license. It is available on GitHub https://github.com/OpenTreeOfLife/python-opentree. It can be installed from PyPi using pip install opentree. The code is packaged with an automated test suite which is maintained to cover at minimum 75% of the code. Testing reports are generated on travis.io and posted to codecov.io and reflected on the GitHub readme with each commit. Documentation and tutorials are available with the code and are posted to https://opentree.readthedocs.io.

REFERENCES

Allen J.M., Germain-Aubrey C.C., Barve N., Neubig K.M., Majure L.C., Laffan S.W., Mishler B.D., Owens H.L., Smith S.A., Whitten W.M., Abbott J.R., Soltis D.E., Guralnick R, Soltis P.S. 2019. Spatial phylogenetics of florida vascular plants: the effects of calibration and uncertainty on diversity estimates. IScience 11:57–70.

Antonelli A., Hettling H., Condamine F.L., Vos K., Nilsson R.H., Sanderson M.J., Sauquet H., Scharn R., Silvestro D., Töpel M., Bacon C., Oxelman B., Vos R.A. 2017. Toward a self-updating platform for estimating rates of speciation and migration, ages, and relationships of taxa. Syst. Biol. 66:152–166.

Ballen C.J., Greene H.W. 2017. Walking and talking the tree of life: why and how to teach about biodiversity. PLoS Biol. 15:e2001630.

Boeckmann B., Marcet-Houben M., Rees J.A., Forslund K., Huerta-Cepas J., Muffato M., Yilmaz P., Xenarios I., Bork P., Lewis S.E., Gabaldón T., Quest for Orthologs Species Tree Working Group. 2015. Quest for orthologs entails quest for tree of life: in search of the gene stream. Genome Biol. Evol. 7:1988–1999.

Brown J.W., Wang N., Smith S.A. 2017. The development of scientific consensus: analyzing conflict and concordance among avian phylogenies. Mol. Phylogenet. Evol. 116:69–77.

Capdevila P., Beger M., Blomberg S.P., Hereu B., Linares C., Salguero-Gómez R. 2020. Longevity, body dimension and reproductive mode

drive differences in aquatic versus terrestrial life-history strategies. Funct. Ecol. 34:1613–1625.

Cui X., Paterson A.M., Wyse S.V., Alam M.A., Maurin K.J., Pieper R., Cubino J.P., O'Connell D.M., Donkers D., Bréda J., Buckley H.L., Perry G.L.W., Curran T.J. 2020. Shoot flammability of vascular plants is phylogenetically conserved and related to habitat fire-proneness and growth form. Nat. Plants 6:355–359.

Drew B.T., Gazis R., Cabezas P., Swithers K.S., Deng J., Rodriguez R., Katz L.A., Crandall K.A., Hibbett D.S., Soltis, D.E. 2013. Lost branches on the tree of life. PLoS Biol. 11:e1001636.

Eastman J.M., Harmon L.J., Tank D.C. 2013. Congruification: support for time scaling large phylogenetic trees. Methods Ecol. Evol. 4:688–691.

Ewers-Saucedo C., Owen C.L., Pérez-Losada M., Høeg J.T., Glenner H., Chan B.K., Crandall K.A. 2019. Towards a barnacle tree of life: integrating diverse phylogenetic efforts into a comprehensive hypothesis of the costracan evolution. PeerJ 7:e7387.

Ewers-Saucedo C., Pappalardo P. 2019. Testing adaptive hypotheses on the evolution of larval life history in acorn and stalked barnacles. Ecol. Evol. 9:11434–11447.

Federhen S. 2012. The NCBI taxonomy database. Nucleic Acids Res. 40:D136–D143.

Fisher R.M., Henry L.M., Cornwallis C.K., Kiers E.T., West S.A. 2017. The evolution of host-symbiont dependence. Nat. Commun. 8:1–8.

Fukushima C.S., Mammola S., Cardoso P. 2020. Global wildlife trade permeates the tree of life. Biol. Conserv. 247:108503.

GBIF Secretariat. 2019. GBIF occurrence download. https://doi.org/10.15468/dl.9bigak

Geffroy B., Sadoul B., Putman B.J., Berger-Tal O., Garamszegi L.Z., Møller A.P., Blumstein D.T. 2020. Evolutionary dynamics in the anthropocene: life history and intensity of human contact shape antipredator responses. PLoS Biol. 18:e3000818.

Healy K., Ezard T.H., Jones O.R., Salguero-Gómez R., Buckley Y.M. 2019. Animal life history is shaped by the pace of life and the distribution of age-specific mortality and reproduction. Nat. Ecol. Evol. 3:1217–1224.

Herrera S., Reyes-Herrera P.H., Shank T.M. 2015. Predicting RAD-seq marker numbers across the eukaryotic tree of life. Genome Biol. Evol. 7:3207–3225.

Jantzen J.R., Whitten W.M., Neubig K.M., Majure L.C., Soltis D.E., Soltis P.S. 2019. Effects of taxon sampling and tree reconstruction methods on phylodiversity metrics. Ecol. Evol. 9:9479–9499.

Kunc H.P., Schmidt R. 2019. The effects of anthropogenic noise on animals: a meta-analysis. Biol. Lett. 15:20190649.

Letunic I., Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 47:W256–W259.

Li D., Trotta L., Marx H.E., Allen J.M., Sun M., Soltis D.E., Soltis P.S., Guralnick R.P., Baiser B. 2019. For common community phylogenetic analyses, go ahead and use synthesis phylogenies. Ecology 100:e02788.

McTavish E.J., Drew B.T., Redelings B., Cranston K.A. 2017. How and why to build a unified Tree of Life. BioEssays 39:1700114.

McTavish E.J., Hinchliff C.E., Allman J.F., Brown J.W., Cranston K.A., Holder M.T., Rees J.A., Smith S.A. 2015. Phylesystem: a git-based data store for community-curated phylogenetic estimates. Bioinformatics 31:2794–2800.

Michonneau F., Brown J.W., Winter D.J. 2016. rotl: an R package to interact with the Open Tree of Life data. Methods Ecol. Evol. 7:1476–1481.

Miller J.B., McKinnon L.M., Whiting M.F., Kauwe J.S., Ridge P.G. 2020. Codon pairs are phylogenetically conserved: a comprehensive analysis of codon pairing conservation across the tree of life. PLoS One 15:e0232260.

OpenTreeOfLife Redelings B., Sanchez Reyes L.L., Cranston K.A., Allman J., Holder M.T., McTavish E.J. 2019a. Open Tree of Life synthetic tree. Available from: https://tree.opentreeoflife.org/about/synthesis-release/v12.3.

OpenTreeofLife, Cranston K.A., Redelings, B., Sanchez Reyes, L.L., Allman J., McTavish E.J., Holder M.T. 2019b. Open Tree of Life taxonomy. Available from: https://tree.opentreeoflife.org/about/taxonomy-version/ott3.2

Owen C.L., Bracken-Grissom H., Stern D., Crandall K.A. 2015. A synthetic phylogeny of freshwater crayfish: insights for conservation. Philos. Trans. R. Soc. B 370:20140009.

Park D.S., Worthington S., Xi Z. 2018. Taxon sampling effects on the quantification and comparison of community phylogenetic diversity. Mol. Ecol. 27:1296–1308.

Parr C.S., Guralnick R., Cellinese N., Page R.D.M. 2012. Evolutionary informatics: unifying knowledge about the diversity of life. Trends Ecol. Evol. 27:94–103.

Piel W.H., Donoghue M., Sanderson M., Netherlands L. 2000. Treebase: a database of phylogenetic information. In: Proceedings of the 2nd International Workshop of Species, Vol. 2000. Available from: http://donoghuelab.yale.edu/sites/default/files/124_piel_shimura02.pdf.

Polilov A.A., Makarova A.A. 2017. The scaling and allometry of organ size associated with miniaturization in insects: a case study for coleoptera and hymenoptera. Sci. Rep. 7:1–7.

Redelings B.D., Holder M.T. 2017. A supertree pipeline for summarizing phylogenetic and taxonomic information for millions of species. PeerJ 5:e3058.

Rees J.A., Cranston K. 2017. Automated assembly of a reference taxonomy for phylogenetic data synthesis. Biodivers. Data J. (5):e12581l.

Rosindell J., Harmon L.J. 2012. Onezoom: a fractal explorer for the tree of life. PLoS Biol. 10:e1001406.

Sánchez-Reyes L.L., Kandziora M., McTavish E.J. 2020. Physcraper: a python package for continual update of evolutionary estimates using the Open Tree of Life. bioRxiv 2020.09.15.299156. Available from: https://www.biorxiv.org/content/early/2020/09/17/2020.09.15.299156

Sanderson M.J., Boss D., Chen D., Cranston K.A., Wehe A. 2008. The phylota browser: processing genbank for molecular phylogenetics research. Syst. Biol. 57:335–346.

Smith S.A., Beaulieu J.M., Donoghue M.J. 2009. Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. BMC Evol. Biol. 9:1–12.

Smith S.A., Brown J.W. 2018. Constructing a broadly inclusive seed plant phylogeny. Am. J. Bot. 105:302–314.

Smith S.A., Walker J.F. 2019. Pyphlawd: a python tool for phylogenetic dataset construction. Methods Ecol. Evol. 10:104–108.

Stoltzfus A., Lapp H., Matasci N., Deus H., Sidlauskas B., Zmasek C.M., Vaidya G., Pontelli E., Cranston K., Vos R., Webb C.O., Harmon L.J., Pirrung M., O'Meara B., Pennell M.W., Mirarab S., Rosenberg M.S., Balhoff J.P., Bik H.M., Heath T.A., Midford P.E., Brown J.W., McTavish E.J., Sukumaran J., Westneat M., Alfaro M.E., Steele A., Jordan G. 2013. Phylotastic! Making tree-of-life knowledge accessible, reusable and convenient. BMC Bioinformatics 14:1–17.

Sukumaran J., Holder M.T. 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics 26:1569–1571.

Sullivan B.L., Wood C.L., Iliff M.J., Bonney R.E., Fink D., Kelling S. 2009. eBird: a citizen-based bird observation network in the biological sciences. Biol. Conserv. 142:2282–2292.

Sánchez-Reyes L.L., O'Meara B.C. 2019. DateLife: leveraging databases and analytical tools to reveal the dated Tree of Life. bioRxiv 782094. Available from: https://www.biorxiv.org/content/early/2019/10/05/782094

Tarka M., Guenther A., Niemelä P.T., Nakagawa S., Noble D.W. 2018. Sex differences in life history, behavior, and physiology along a slow-fast continuum: a meta-analysis. Behav. Ecol. Sociobiol. 72:1–13.

Uyeda J.C., Pennell M.W., Miller E.T., Maia R., McClain C.R. 2017. The evolution of energetic scaling across the vertebrate tree of life. Am. Nat. 190:185–199.

Wong Y., Rosindell J. 2020. Dynamic visualisation of million-tip trees: the onezoom project. bioRxiv. Available from: https://www.biorxiv.org/content/early/2020/10/14/2020.10.14.323055