

`datelife`'s database includes all chronograms derived from expert knowledge, and this includes those derived from secondary calibrations.

Improvements: * fossils as calibrations: Using secondary calibrations can generate biased ages when using bayesian methods, mainly because we don't know what prior to give to secondary calibrations (Schenk 2016).
* bayesian congruification * topological congruification

Phylomatic (Webb and Donoghue 2005) is a platform that synthesizes a tree containing taxa of interest from a user supplied set of trees or from its own database. It has the advantage that it allows grafting taxa of interest absent from their database, but it is limited to mostly seed plants and mammals. If your query does not match its records, it gives many polytomies back.

treePL, is a semi-parametric, rate-smoothing, penalized likelihood dating method (Smith and O'Meara 2012).

Time of lineage divergence constitutes a fundamental piece of information for evolutionary understanding in many areas of research, from developmental to conservation biology (Felsenstein 1985; Webb 2000). This information can be obtained from fossils and from molecular dating studies. The developments in DNA sequencing and phylogenetic dating inferences have encouraged/spurred the application of molecular dating methods to a very large diversity of organisms, greatly increasing the quantity of data on taxon ages across the tree of life. Along with phylogenetic relationships, it allows testing alternative evolutionary hypothesis in many areas of research, from developmental to conservation biology (Felsenstein 1985; Webb 2000). It has been shown that only a bit of information on time of divergence can improve inferences of phylogenetic distance Webb et al. (2008) and Together with geographical distribution patterns, it is a crucial component of historical biogeography studies (Posadas et al. 2006). Coupled to species number, it allows studying the tempo and mode of speciation and extinction, central for understanding how biodiversity patterns are shaped across space, time, and clades (Morlon 2014). In the past two decades, the possibility to obtain good quality DNA sequences coupled to methodological developments in phylogenetic and dating inference, allowed the application of molecular dating methods to a very large diversity of organisms, greatly increasing the quantity and quality of data on taxon ages across the tree of life. To date, there is a large amount of data on taxon ages and phylogenetic relationships in public repositories such as Dryad, TreeBASE and Open Tree of Life

(OToL). , ->

The TimeTree project (Hedges et al. 2006, 2015; Kumar et al. 2017) has aggregated chronograms from 3,163 studies, encompassing 97,085 species (Kumar et al. 2017), and continues to grow. However, even in this gold standard resource, the included taxa only encompass between 0.097 and 3.236% of total species diversity (following taxonomic expert opinion on the global, extant species numbers, which ranges from 3 to 100 million species (Mayr 2010; Mora et al. 2011)). One advantage of TimeTree is that it includes taxa from across the tree of life, versus more specialized chronogram repositories focusing on land plants (Stevens and Davis 2001) and mammals (Webb and Donoghue 2005), birds (Jetz et al. 2012), and fish (Chang et al. 2019). Users can choose between a web interface or a mobile app to receive information on divergence times for the evolutionary history of a lineage, pairs of taxa, all lineages within a taxon, or a list of taxa. As a science communication tool, TimeTree project is very powerful: it has a friendly graphical interface, with informative and colorful outputs, that allows the general public to satisfy curiosity regarding a particular organism of interest or group of them. It is of limited utility for scientific studies, however. The thousands of trees that have been entered are unavailable for examination or reuse; according to the creators (see TimeTree web FAQ), methods for allowing data downloading have been under discussion for the past several years yet the primary data remain closed. Moreover, there is no Application Programming Interface (API) allowing programmatic access to any data, greatly impairing the possibility of large-scale, automated data-mining, which is not allowed under TimeTree website's terms of use. The nearly hundred thousand taxon summary chronogram generated from TimeTree resources is not available with its publication (Kumar et al. 2017) or the TimeTree website, though the still substantial chronogram from a previous publication (Hedges et al. 2015) was made available at OToL.

Despite its great importance, analytical tools to summarize available information on taxon ages for the scientific community are still lacking. We identified several aspects that might have so far delayed the exploitation of existing data. First, original chronograms available publicly are scattered across repositories and journals supplementary data, usually with different formats too. Second, lineage names due to taxonomic idiosyncrasy can be different among studies and manual curation of that is usually necessary. Third, data

curation

Platforms that contribute to knowledge in this areas:

Supersmart is an open tool but not easy to use, still requires a lot of curation and knowledge.

Phylogenerator (???)

Generate new dates using all available DNA sequence information; Perform one global analysis using all available information;

Problems or downsides: This might be time consuming for large groups and a lot of data curation and knowledge on the group of interest is still necessary. For example, choosing correct fossils for calibration requires a lot of expertise and knowledge on the group. An incorrect use of fossils can generate severe bias in dating results (Sauquet et al. 2012). Hence, data curation is still an important part of any biological study. The research community considers it as an important or even crucial step before data analysis. Hence, automated processes for large data analysis are frequently received with skepticism.

datelife palliates this by only using information available from already published studies, which are ideally constructed using robust information, such as sequence data and thoughtfully curated fossil calibrations.

Rapidly increasing data on time of lineage divergence both from molecular and paleontological studies; the increasing importance of use of these data in distant areas of research, often not specialized enough to rapidly obtain data on their own; and the lack of an open (both the data sources and the code underlying the analyses) easy to use tool inspired the development of a prototype **datelife** service over a series of phylotastic hackathons (Stoltzfus et al. 2013) at the National Evolutionary Synthesis Center. In this paper we present the first formal description of **datelife**, featuring an improved database of chronograms, more methods to summarize trees, and new functions to visualize data, as well as comparisons of summary trees. **datelife** is the main service for scaling phylogenetic trees in Phylotastic! system (Stoltzfus et al. 2013) It

can be used through its R package , and a web interface (<http://www.datelife.org/query/>).

Barker F.K., Burns K.J., Klicka J., Lanyon S.M., Lovette I.J. 2012. Going to extremes: Contrasting rates of diversification in a recent radiation of new world passerine birds. *Systematic biology*. 62:298–320.

Barker F.K., Burns K.J., Klicka J., Lanyon S.M., Lovette I.J. 2015. New insights into new world biogeography: An integrated view from the phylogeny of blackbirds, cardinals, sparrows, tanagers, warblers, and allies. *The Auk: Ornithological Advances*. 132:333–348.

Burns K.J., Shultz A.J., Title P.O., Mason N.A., Barker F.K., Klicka J., Lanyon S.M., Lovette I.J. 2014. Phylogenetics and diversification of tanagers (passeriformes: Thraupidae), the largest radiation of neotropical songbirds. *Molecular Phylogenetics and Evolution*. 75:41–77.

Chang J., Rabosky D.L., Smith S.A., Alfaro M.E. 2019. An r package and online resource for macroevolutionary studies using the ray-finned fish tree of life. *Methods in Ecology and Evolution*. 0:1–7.

Claramunt S., Cracraft J. 2015. A new time tree reveals earth history’s imprint on the evolution of modern birds. *Science advances*. 1:e1501005.

Felsenstein J. 1985. Phylogenies and the Comparative Method. *The American Naturalist*. 125:1–15.

Gibb G.C., England R., Hartig G., McLenachan P.A., Taylor Smith B.L., McComish B.J., Cooper A., Penny D. 2015. New zealand passerines help clarify the diversification of major songbird lineages during the oligocene. *Genome biology and evolution*. 7:2983–2995.

Hedges S.B., Dudley J., Kumar S. 2006. TimeTree: A public knowledge-base of divergence times among organisms. *Bioinformatics*. 22:2971–2972.

Hedges S.B., Marin J., Suleski M., Paymer M., Kumar S. 2015. Tree of life reveals clock-like speciation and diversification. *Molecular Biology and Evolution*. 32:835–845.

Hooper D.M., Price T.D. 2017. Chromosomal inversion differences correlate with range overlap in passerine

97 birds. *Nature ecology & evolution*. 1:1526.

98 Jetz W., Thomas G., Joy J.J., Hartmann K., Mooers A. 2012. The global diversity of birds in space and
99 time. *Nature*. 491:444–448.

100 Kumar S., Stecher G., Suleski M., Hedges S.B. 2017. TimeTree: A Resource for Timelines, Timetrees, and
101 Divergence Times. *Molecular biology and evolution*. 34:1812–1819.

102 Mayr R.M. 2010. Tropical Arthropod Species, More or Less? *Science*. 329:41–42.

103 Mora C., Tittensor D.P., Adl S., Simpson A.G., Worm B. 2011. How many species are there on earth and in
104 the ocean? *PLoS Biology*. 9:1–8.

105 Morlon H. 2014. Phylogenetic approaches for studying diversification. *Ecology Letters*. 17:508–525.

106 Posadas P., Crisci J.V., Katinas L. 2006. Historical biogeography: A review of its basic concepts and critical
107 issues. *Journal of Arid Environments*. 66:389–403.

108 Price T.D., Hooper D.M., Buchanan C.D., Johansson U.S., Tietze D.T., Alström P., Olsson U., Ghosh-Harihar
109 M., Ishtiaq F., Gupta S.K., others. 2014. Niche filling slows the diversification of himalayan songbirds.
110 *Nature*. 509:222.

111 Sauquet H., Ho S.Y.W., Gandolfo M. a, Jordan G.J., Wilf P., Cantrill D.J., Bayly M.J., Bromham L., Brown
112 G.K., Carpenter R.J., Lee D.M., Murphy D.J., Sniderman J.M.K., Udovicic F. 2012. Testing the impact
113 of calibration on molecular divergence times using a fossil-rich group: the case of Nothofagus (Fagales).
114 *Systematic Biology*. 61:289–313.

115 Schenk J.J. 2016. Consequences of secondary calibrations on divergence time estimates. *PLoS ONE*. 11.

116 Smith S.A., O’Meara B.C. 2012. TreePL: Divergence time estimation using penalized likelihood for large
117 phylogenies. *Bioinformatics*. 28:2689–2690.

118 Stevens P.F., Davis H. 2001. Angiosperm phylogeny website..

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

124 Webb C.O. 2000. Exploring the Phylogenetic Structure of Ecological Communities : An Example for Rain
 125 Forest Trees. *The American Naturalist*. 156:145–155.

126 Webb C.O., Ackerly D.D., Kembel S.W. 2008. Phylocom: Software for the analysis of phylogenetic community
 127 structure and trait evolution. *Bioinformatics*. 24:2098–2100.

128 Webb C.O., Donoghue M.J. 2005. Phylomatic: Tree assembly for applied phylogenetics. *Molecular Ecology*
 129 *Notes*. 5:181–183.