

<sup>1</sup> DateLife: leveraging databases and analytical tools to reveal the dated Tree of Life

<sup>2</sup> Luna L. Sánchez Reyes<sup>1,2</sup>, Emily Jane McTavish<sup>1</sup>, & Brian O'Meara<sup>2</sup>

<sup>3</sup> <sup>1</sup> University of California, Merced

<sup>4</sup> <sup>2</sup> University of Tennessee, Knoxville

<sup>5</sup> Author Note

6           School of Natural Sciences, University of California, Merced, Science and Engineering  
7   Building 1.

8           Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville,  
9   446 Hesler Biology Building, Knoxville, TN 37996, USA.

10          The authors made the following contributions. Luna L. Sánchez Reyes: Data curation,  
11   Investigation, Software, Visualization, Validation, Writing - Original Draft Preparation,  
12   Writing - Review & Editing; Emily Jane McTavish: Resources, Software, Writing - Review &  
13   Editing; Brian O'Meara: Conceptualization, Funding acquisition, Methodology, Resources,  
14   Software, Supervision, Writing - Review & Editing.

15          Correspondence concerning this article should be addressed to Luna L. Sánchez Reyes, .  
16   E-mail: [sanchez.reyes.luna@gmail.com](mailto:sanchez.reyes.luna@gmail.com)

17

## Abstract

18 Achieving a high-quality reconstruction of a phylogenetic tree with branch lengths  
19 proportional to absolute time (chronogram) is a difficult and time-consuming task. But the  
20 increased availability of fossil and molecular data, and time-efficient analytical techniques  
21 has resulted in many recent publications of large chronograms for a large number and wide  
22 diversity of organisms. Knowledge of the evolutionary time frame of organisms is key for  
23 research in the natural sciences. It also represent valuable information for education, science  
24 communication, and policy decisions. When chronograms are shared in public, open  
25 databases this wealth of expertly-curated and peer-reviewed data on evolutionary timeframe  
26 is exposed in a programmatic and reusable way, as intensive and localized efforts have  
27 improved data sharing practices, as well as incentivized open science in biology. Here we  
28 present DateLife, a service implemented as an R package and an R Shiny website application  
29 available at [www.datelife.org](http://www.datelife.org), that provides functionalities for efficient and easy finding,  
30 summary, reuse, and reanalysis of expert, peer-reviewed, public data on time frame of  
31 evolution. The main DateLife workflow constructs a chronogram for any given combination  
32 of taxon names by searching a local chronogram database constructed and curated from the  
33 Open Tree of Life Phylesystem phylogenetic database, which incorporates phylogenetic data  
34 from the TreeBASE database as well. We implement and test methods for summarizing time  
35 data from multiple source chronograms using supertree and congruification algorithms, and  
36 using age data extracted from source chronograms as secondary calibration points to add  
37 branch lengths proportional to absolute time to a tree topology. DateLife will be useful to  
38 increase awareness of the existing variation in alternative hypothesis of evolutionary time for  
39 the same organisms, and can foster exploration of the effect of alternative evolutionary  
40 timing hypotheses on the results of downstream analyses, providing a framework for a more  
41 informed interpretation of evolutionary results.

42

*Keywords:* Tree; Phylogeny; Scaling; Dating; Ages; Divergence times; Open Science;

<sup>43</sup> Congruification; Supertree; Calibrations; Secondary calibrations

<sup>44</sup> Word count: 4906

45 DateLife: leveraging databases and analytical tools to reveal the dated Tree of Life

46 **Introduction**

47 Chronograms –phylogenies with branch lengths proportional to time– provide key data  
48 on evolutionary time frame for the study of natural processes in many areas of biological  
49 research, such as developmental biology (Delsuc et al., 2018; Laubichler & Maienschein,  
50 2009), conservation biology (Felsenstein, 1985; C. Webb, 2000), historical biogeography  
51 (Posadas, Crisci, & Katinas, 2006), and species diversification (Magallon & Sanderson, 2001;  
52 Morlon, 2014).

53 Building a chronogram is not an easy task. It requires obtaining and curating data to  
54 construct a phylogeny, selecting and placing appropriate calibrations on the phylogeny using  
55 independent age data points from the fossil record or other dated events, and inferring the  
56 full dated tree; it also generally requires specialized biological training, taxonomic domain  
57 knowledge, and a non-negligible amount of research time, computational resources and  
58 funding.

59 Here we present the DateLife project which has the main goal of capturing age data  
60 from published chronograms, and making these data readily accessible to the community for  
61 reuse and reanalysis, for research, teaching, and science communication and policy.  
62 DateLife’s core software application is available as an R package (Sanchez-Reyes et al., 2022),  
63 and as an online Rshiny interactive website at [www.datelife.org](http://www.datelife.org). It features key elements for  
64 scientific reproducibility, such as a versioned, open and fully public source database  
65 (McTavish et al., 2015), data stored and available in a computer readable format (Vos et al.,  
66 2012), automated and programmatic ways of accessing the data (Stoltzfus et al., 2013) and  
67 methods to summarize and compare the data.

68

## Description

69 DateLife's core software application consists of the R package `datelife`. Its latest  
 70 stable version – v0.6.4, is available from the CRAN repository (Sanchez-Reyes et al., 2022),  
 71 and relies on functionalities from various biological R packages: ape (Paradis, Claude, &  
 72 Strimmer, 2004), bold (Chamberlain et al., 2019), geiger (Pennell et al., 2014), paleotree  
 73 (Bapst, 2012), phyloch (Heibl, 2008), phylocomr (Ooms & Chamberlain, 2018), phytools  
 74 (Revell, 2012), rotl (Michonneau, Brown, & Winter, 2016), and taxize (Chamberlain & Szöcs,  
 75 2013; Chamberlain et al., 2019). Figure 1 provides a graphical summary of the three main  
 76 steps of the DateLife algorithm: creating a search query, searching a database, and  
 77 summarizing results from the search.

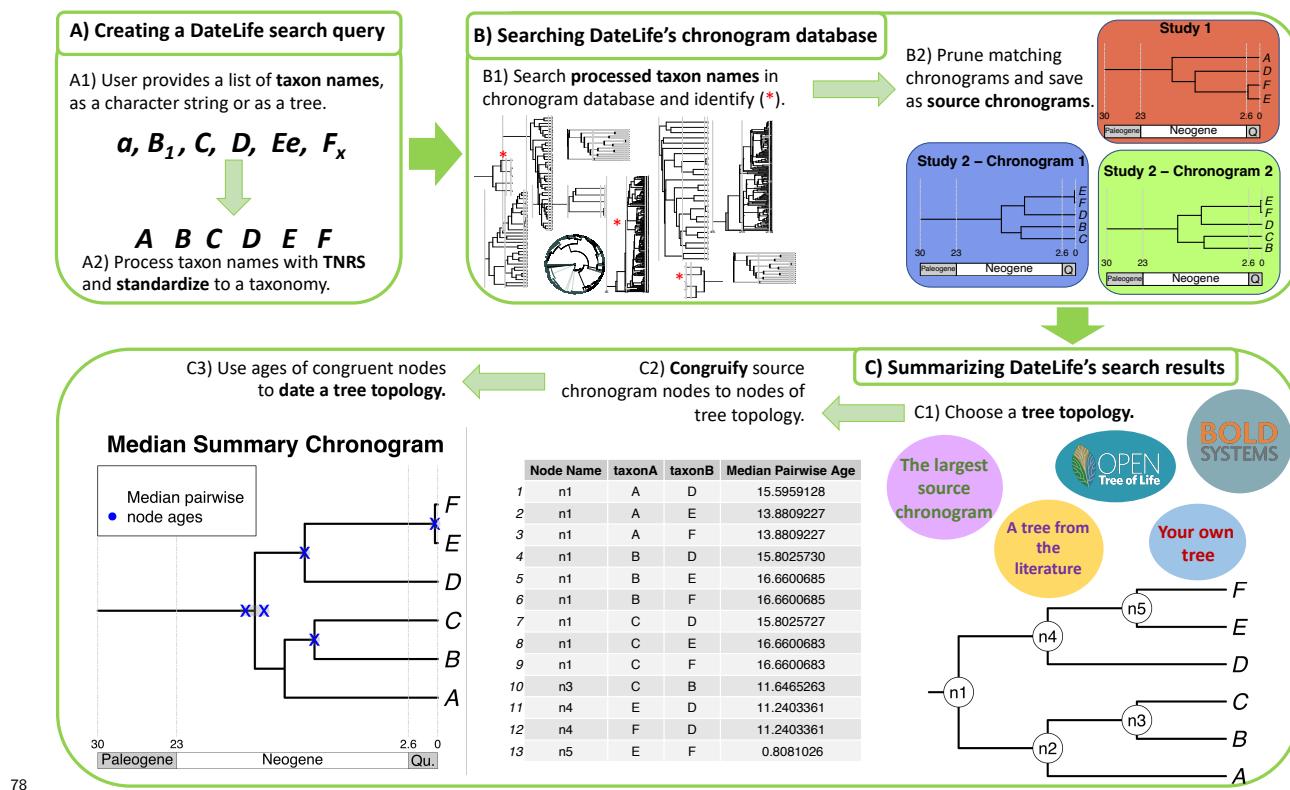


FIGURE 1. Stylized DateLife workflow. This shows the general workflows and analyses that can be performed with `datelife`, via the R package or through the website at <http://www.datelife.org/>. Details on the functions involved on each workflow are shown in `datelife`'s R package vignette.

79 **Creating a search query**

80 DateLife starts by processing an input consisting of at least two taxon names, which  
81 can be provided as a comma separated character string or as tip labels on a tree. If the input  
82 is a tree, it can be provided as a classic newick character string (Archie et al., 1986), or as a  
83 “phylo” R object (Paradis et al., 2004). The input tree is not required to have branch lengths,  
84 and its topology is used in the summary steps described in the next section.

85 DateLife accepts scientific names that can belong to any inclusive taxonomic group  
86 (e.g., genus, family, tribe, etc.) or a binomial species name. Subspecies and variants are  
87 ignored. If an input taxon name belongs to an inclusive taxonomic group, the algorithm has  
88 two alternative behaviors defined by the “get species from taxon” flag. If the flag is active,  
89 DateLife retrieves all species names within the inclusive taxonomic group (following a  
90 taxonomy of choice; current options are Open Tree of Life, NCBI, GBIF, or IRMNG  
91 taxonomies) and adds them to the input string. If the flag is inactive, DateLife excludes any  
92 taxon names above the species level from the search query.

93 DateLife processes input scientific names using a Taxonomic Name Resolution Service  
94 (TNRS), which increases the probability of correctly finding the queried taxon names in the  
95 chronogram database. TNRS detects, corrects and standardizes name misspellings and typos,  
96 variant spellings and authorities, and nomenclatural synonyms to a single taxonomic  
97 standard (Boyle et al., 2013). DateLife implements TNRS using OpenTree’s unified  
98 taxonomy as standard (Open Tree Of Life et al., 2016; Rees & Cranston, 2017), storing  
99 OpenTree’s Taxonomy identification numbers for further processing.

100 The processed input taxon names are saved as an R object of a newly defined class,  
101 `datelifeQuery`, that is used in the following steps. This object contains the standardized  
102 names, the corresponding OpenTree taxonomic id numbers, and the topology of the input  
103 tree if any was provided.

104 **Searching a chronogram database**

105 At the time of writing of this manuscript (Jun 20, 2022), DateLife's chronogram  
106 database latest version consist of 253 chronograms published in 187 different studies. It is  
107 curated from OpenTree's phylogenetic database, the Phylesystem, which constitutes an open  
108 source of expert and peer-reviewed phylogenetic knowledge with rich metadata (McTavish et  
109 al., 2015), which allows automatic and reproducible assembly of our chronogram database.  
110 Datelife's chronogram database is navigable as an R data object within the `datelife` R  
111 package.

112 A unique feature of OpenTree's Phylesystem is that the community can add new  
113 published, state-of-the-art chronograms any time through their curator application  
114 (<https://tree.opentreeoflife.org/curator>). As chronograms are added to Phylesystem, they are  
115 incorporated into an updated `datelife`'s database that is assigned a new version number,  
116 followed by a package release on CRAN. `datelife`'s chronogram database is updated as new  
117 chronogram data is added to Phylesystem, at a minimum of once a month and a maximum  
118 of every 6 months. Users can also upload and curate new chronograms to OpenTree  
119 themselves. Then, trigger an update of their local `datelife` database to incorporate the new  
120 chronograms and have them available for their current analysis.

121 A DateLife search is implemented by matching processed taxon names provided by the  
122 user to tip labels in the chronogram database. Chronograms with at least two matching  
123 taxon names on their tip labels are identified and pruned down to preserve only the matched  
124 taxa. These matching pruned chronograms are referred to as source chronograms. Total  
125 distance (in units of millions of years) between taxon pairs within each source chronogram  
126 are stored as a patristic distance matrix (Figure 1). The matrix format speeds up extraction  
127 of pairwise taxon ages of any queried taxa, as opposed to searching the ancestor node of a  
128 pair of taxa in a “phylo” object or newick string. Finally, the patristic matrices are  
129 associated to the study citation where the original chronogram was published, and stored as

130 an R object of the newly defined class `datelifeResult`.

131 **Summarizing search results**

132 Summary information is extracted from the `datelifeResult` object to inform  
133 decisions for subsequent steps in the analysis workflow. Basic summary information available  
134 to the user is:

- 135 1. The matching pruned chronograms as newick strings or “phylo” objects.
- 136 2. The ages of the root of all source chronograms. These ages can correspond to the age  
137 of the most recent common ancestor (mrca) of the user’s group of interest if the source  
138 chronograms have all taxa belonging to the group. If not, the root corresponds to the  
139 mrca of a subgroup within the group of interest.
- 140 3. Study citations where original chronograms were published.
- 141 4. A report of input taxon names matches across source chronograms.
- 142 5. The source chronogram(s) with the most input taxon names.
- 143 6. Various single summary chronograms resulting from summarizing age data, generated  
144 using the methodology described next.

145 ***Choosing a topology.***— DateLife requires a tree topology to summarize age data  
146 upon. We recommend that users provide a tree topology as input from the literature, or one  
147 of their own making. If no topology is provided, DateLife automatically obtains one from the  
148 OpenTree synthetic tree (Open Tree Of Life et al., 2019). These are known as “induced”  
149 OpenTree subtrees, which do not necessarily correspond to a monophyletic group.  
150 Alternatively, DateLife can combine topologies from source chronograms using a supertree  
151 approach. To combine topologies from source chronograms into a single summary (or  
152 supertree) topology, the DateLife algorithm starts by identifying the source chronograms  
153 that form a grove, roughly, a sufficiently overlapping set of taxa between trees, by  
154 implementing definition 2.8 for n-overlap from Ané et al. (2009). In rare cases, a group of

155 trees can have multiple groves. By default, DateLife chooses the grove with the most taxa,  
156 however, the “criterion = trees” flag allows the user to choose the grove with the most trees  
157 instead. If source chronograms do not form a grove, the supertree reconstruction will fail.

158 ***Dating the topology.***— Input topologies from OpenTree or the supertree approach  
159 described above do not include branch length estimates of any kind. Optionally, to estimate  
160 branch lengths proportional to substitution rates on these topologies, DateLife can mine the  
161 Barcode of Life Data System, BOLD (Ratnasingham & Hebert, 2007) to obtain genetic  
162 markers for the input taxa. These markers are aligned with MUSCLE (Edgar, 2004) (by  
163 default) or MAFFT (Katoh, Asimenos, & Toh, 2009). This alignment can be used to  
164 estimate branch lengths on input topologies that lack branch lengths. Currently, branch  
165 length reconstruction in DateLife is performed using parsimony and the likelihood of the  
166 phylogenetic tree given a sequence alignment is computed (Schliep, 2011). While relative  
167 branch length information provides additional data for nodes without secondary date  
168 calibrations, topologies without branch lengths can also be dated.

169 Once a topology is chosen, DateLife applies the congruification method (Eastman,  
170 Harmon, & Tank, 2013) to find nodes belonging to the same clade across source  
171 chronograms, and extract the corresponding node ages from the patristic distance matrices  
172 stored as `datelifeResult`. By definition, the matrices store total distance (time from tip to  
173 tip), hence, node ages correspond to half the values stored in the patristic distance matrices.  
174 This assumes that the terminal taxa are coeval and occur at the present. A table of  
175 congruified node ages that can be used as calibrations for a dating analysis is stored as a  
176 `congruifiedCalibrations` object.

177 For each congruent node, the pairwise distances that traverse that node are  
178 summarized into a single summary matrix using classic summary statistics (i.e., mean,  
179 median, minimum and maximum ages), and the Supermatrix Distance Method [SDM;  
180 Criscuolo, Berry, Douzery, and Gascuel (2006)], which deforms patristic distance matrices by

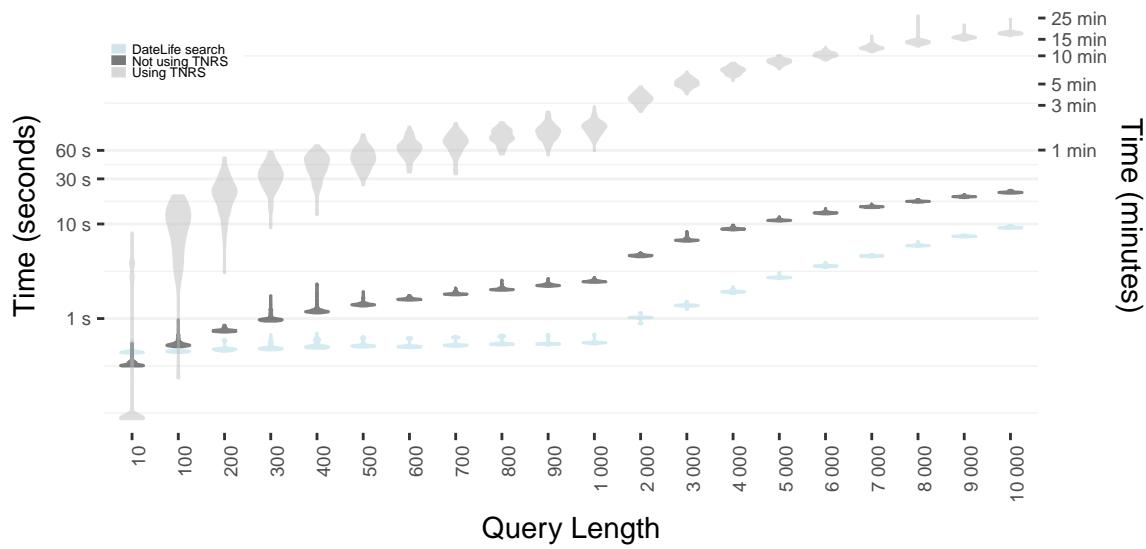
181 minimizing variance and then averaging them. These single summary taxon pair age  
182 matrices (Summarized calibrations) can be applied as calibrations to date a tree topology,  
183 using different dating methods currently supported within DateLife: MrBayes (Huelsenbeck  
184 & Ronquist, 2001; Ronquist & Huelsenbeck, 2003), PATHd8 (Britton, Anderson, Jacquet,  
185 Lundqvist, & Bremer, 2007), BLADJ (Campbell O. Webb, Ackerly, & Kembel, 2008;  
186 Campbell O. Webb & Donoghue, 2005), and treePL (Stephen A. Smith & O'Meara, 2012).

187 By default, DateLife implements the Branch Length Adjuster (BLADJ) algorithm to  
188 obtain a fully dated topology. BLADJ fixes node ages that have calibration data, and  
189 distributes time between nodes with no data evenly between nodes with calibration data.  
190 This minimizes age variance in the resulting chronogram (Campbell O. Webb et al., 2008).  
191 BLADJ does not use branch lengths even when they are present in the input tree or  
192 summarizing topology. When there is conflict in ages between nodes with calibration data,  
193 the algorithm ignores node ages that are older than the age of a parent node. BLADJ  
194 requires a root age estimate. If there is no information on the age of the root in the  
195 chronogram database, users can provide an estimate from the literature. If none is provided,  
196 DateLife assigns an arbitrary age to the root as 10% older than the oldest age available  
197 within the group.

198 Alternative options supported in DateLife (MrBayes, PATHD8, TreePL) incorporate  
199 branch length information from the input topology in combination with the calibrations.  
200 PATHd8 is a non-clock, rate-smoothing method (Britton et al., 2007) to date trees. treePL  
201 (Stephen A. Smith & O'Meara, 2012), is a semi-parametric, rate-smoothing, penalized  
202 likelihood dating method (Michael J. Sanderson, 2002). The MrBayes (Huelsenbeck &  
203 Ronquist, 2001; Ronquist & Huelsenbeck, 2003) approach in datelife uses the calibrations as  
204 priors on node ages.

205 **Visualizing results.**— Finally, users can save all source and summary chronograms in  
206 formats that permit reuse and reanalyses (newick and R “phylo” format), as well as visualize

207 and compare results graphically, or construct their own graphs using `datelife`'s chronogram  
 208 plot generation functions available from the R package `datelifeplot` (Sanchez-Reyes &  
 209 O'Meara, 2022).



210

FIGURE 2. Computation time used to process a `datelife` query and a `datelife` search across `datelife`'s chronogram database, relative to number of input taxon names. For each  $N = 10, 100, 200, \dots, 1\,000, \dots, 9\,000, 10\,000$ , we sampled  $N$  species names from the class Aves a hundred times, and then performed a `datelife` search with input processing, without implementing Taxon Names Resolution Service (TNRS; dark gray), and implementing TNRS (light gray). We also performed a search using a query that had been pre-processed with TNRS, for comparison (light blue).

211

### Benchmark

212 `datelife`'s code speed was tested on an Apple iMac with one 3.4 GHz Intel Core i5  
 213 processor. We registered variation in computing time of query processing and search through  
 214 the database relative to number of queried taxon names. Query processing time increases  
 215 roughly linearly with number of input taxon names, and increases considerably if Taxonomic  
 216 Name Resolution Service (TNRS) is activated. Up to ten thousand names can be processed

and searched in less than 30 minutes with the most time consuming settings. Once names have been processed as described in methods, a name search through the chronogram database can be performed in less than a minute, even with a very large number of taxon names (Fig. 2). `datelife`'s code performance was evaluated with a set of unit tests designed and implemented with the R package `testthat` (R Core Team, 2018) that were run both locally with the `devtools` package (R Core Team, 2018), and on a public server –via GitHub, using the continuous integration tool Travis CI (<https://travis-ci.org>). At present, unit tests cover more than 40% of `datelife`'s code (<https://codecov.io/gh/phylotastic/datelife>). Unit testing helps identify potential issues as code is updated or, more critically, as services code relies upon may change.

## Case studies

We illustrate the DateLife algorithm using a family within the Passeriform birds encompassing the true finches, Fringillidae, as case study. The first example analyses 6 bird species and shows all steps of the algorithm. The second example is an application analysing 289 species in the family Fringillidae that are included in the NCBI taxonomy.

### A small example

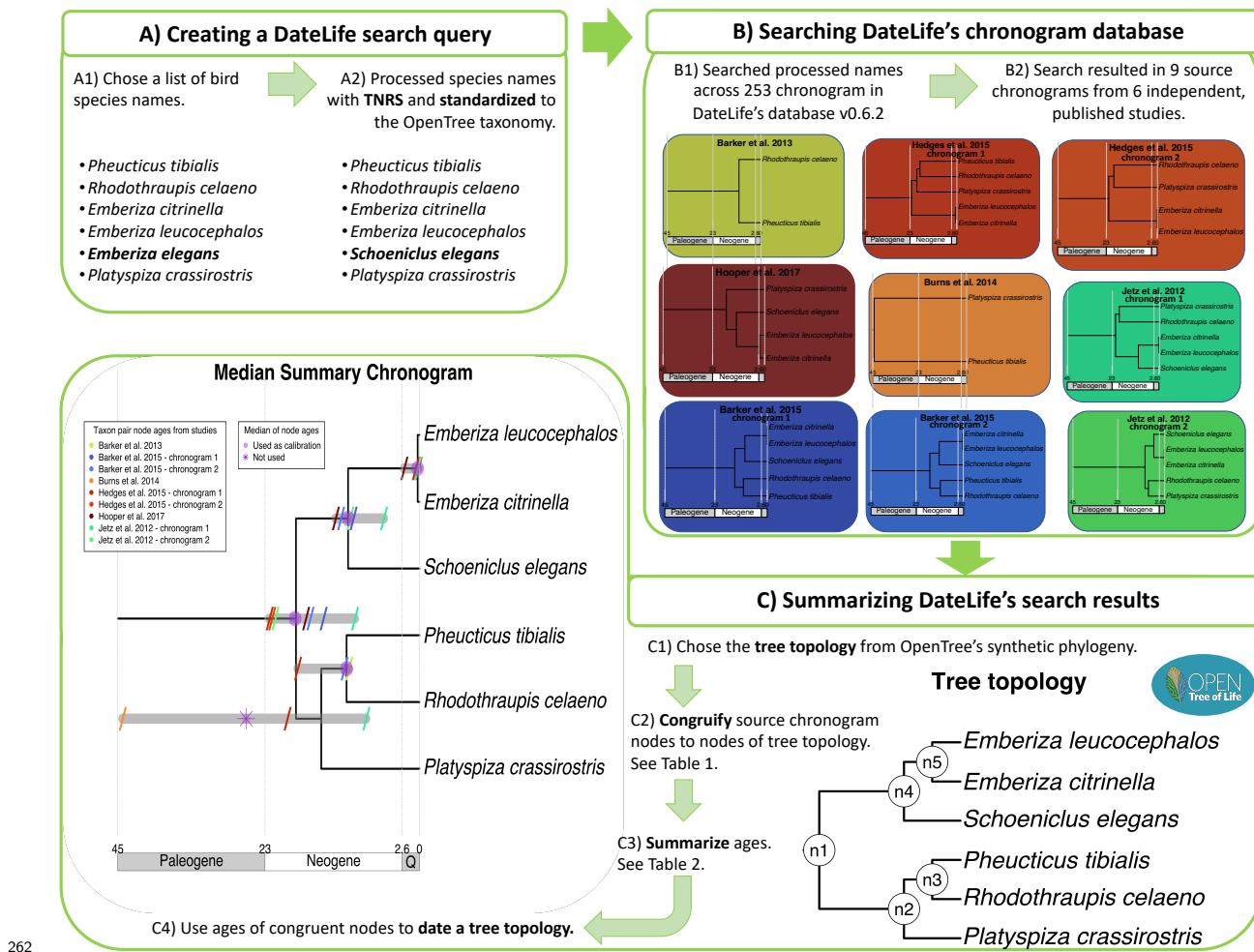
**Creating a search query.-** We chose 6 bird species within the Passeriformes. The sample includes two species of cardinals: the black-thighed grosbeak – *Pheucticus tibialis* and the crimson-collared grosbeak – *Rhodothraupis celaeno*; three species of buntings: the yellowhammer – *Emberiza citrinella*, the pine bunting – *Emberiza leucocephalos* and the yellow-throated bunting – *Emberiza elegans*; and one species of tanager, the vegetarian finch – *Platyspiza crassirostris*.

Processing input names found that *Emberiza elegans* is synonym for *Schoeniclus elegans* in the default reference taxonomy (Open Tree of Life Taxonomy v3.3, June 1, 2021). For a detailed discussion on the state of the synonym, refer to Avibase (Avibase, 2022; Lepage, 2004; Lepage, Vaidya, & Guralnick, 2014). Discovering this synonym allowed assigning five

<sup>243</sup> age data points for the parent node of *Emberiza elegans*, shown as *Schoeniclus elegans* in  
<sup>244</sup> figure 3A, which would not have had any data otherwise.

<sup>245</sup> **Searching the database.**- DateLife used the processed input names to search the local  
<sup>246</sup> chronogram database and found 9 matching chronograms in 6 different studies (Fig. 3B).  
<sup>247</sup> Three studies matched five input names (Barker, Burns, Klicka, Lanyon, & Lovette, 2015;  
<sup>248</sup> Hedges, Marin, Suleski, Paymer, & Kumar, 2015; Jetz, Thomas, Joy, Hartmann, & Mooers,  
<sup>249</sup> 2012), one study matched four input names (Hooper & Price, 2017) and two studies matched  
<sup>250</sup> two input names (Barker, Burns, Klicka, Lanyon, & Lovette, 2013; Burns et al., 2014). No  
<sup>251</sup> studies matched all input names. Together, source chronograms provide 28 unique age data  
<sup>252</sup> points, covering all nodes on our chosen tree topology to date (Table 1).

<sup>253</sup> **Summarizing search results.**- DateLife obtained OpenTree's synthetic tree topology for  
<sup>254</sup> these taxa (Fig. 3C), and congruified and mapped age data to nodes in this chosen topology  
<sup>255</sup> (Table 1). The name processing step allowed including five data points for node "n4" (parent  
<sup>256</sup> of *Schoeniclus elegans*; Fig. 3A) that would not have had any data otherwise due to name  
<sup>257</sup> mismatch. Age summary statistics per node were calculated (Table 2) and used as  
<sup>258</sup> calibrations to date the tree topology using the BLADJ algorithm. As expected, more  
<sup>259</sup> inclusive nodes (e.g., node "n1") have more variance in age data than less inclusive nodes  
<sup>260</sup> (e.g., node "n5"). Summary age data for node "n2" were excluded as final calibration  
<sup>261</sup> because they are older than age data of the more inclusive node, "n1" (Fig. 3C4).



262

FIGURE 3. DateLife analysis results for a small sample of A) 6 bird species within the Passeriformes. B) Processed species names were found across 9 chronograms within 6 independent studies (Barker et al. (2012), Barker et al. (2015), Burns et al. (2014), Hedges et al. (2015), Hooper and Price (2017), Jetz et al. (2012).) C) This revealed 28 source age data points for the queried species names. Summarized age data is used as secondary calibrations to date a tree topology obtained from OpenTree's synthetic tree, resulting in a summary chronogram of source ages.

263 **The family Fringillidae**

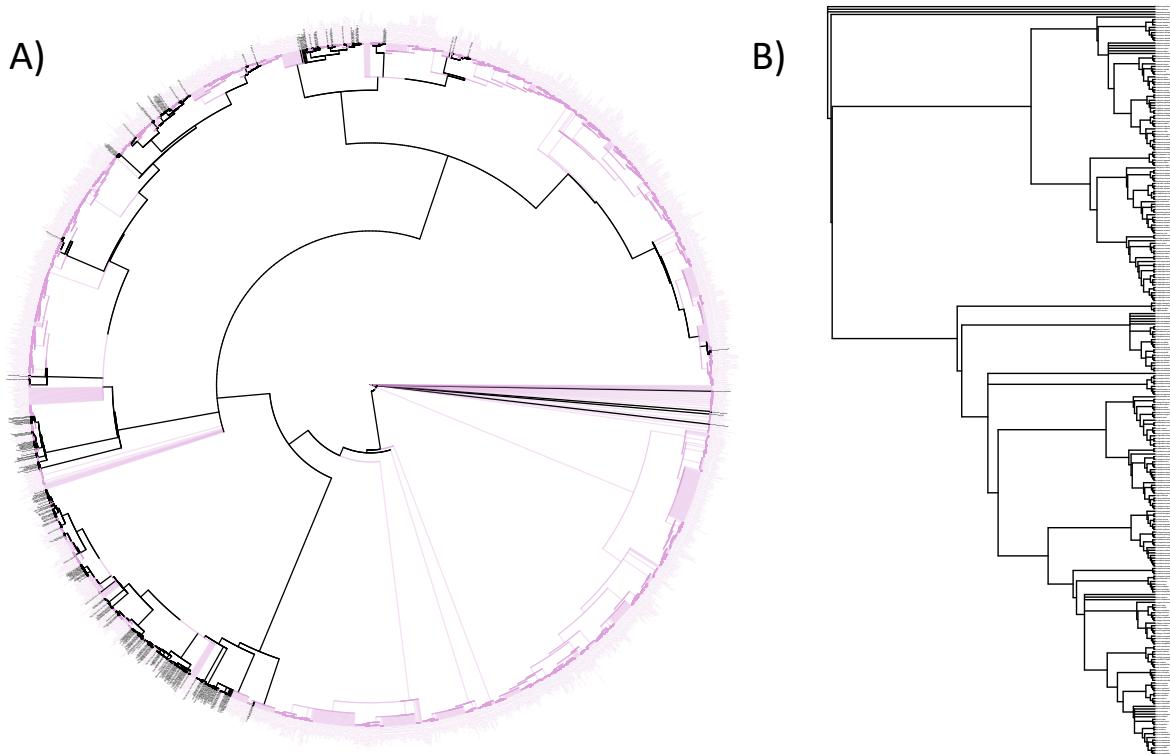
264 ***Creating a query.***- To obtain ages for all species within the family of true finches,  
265 Fringillidae, we ran a DateLife query using the “get species from taxon” flag, which gets all  
266 recognized species names within a named group from a taxonomy of choice. Following the  
267 NCBI taxonomy, our DateLife query has 289 Fringillidae species. This taxon-constrained  
268 approach implies that the final results of a full DateLife analysis will be done using a tree  
269 topology and ages for the species in a named group, which do not necessarily correspond to a  
270 monophyletic group. Users can change this behaviour by providing a monophyletic tree as  
271 input for a DateLife search, or as a tree topology for a DateLife summary.

272 ***Searching the database.***- Next, we used the processed species names in our DateLife  
273 query to identify chronograms with at least two Fringillidae species. The DateLife search  
274 identified 13 chronograms containing at least two Fringillidae species, published in 9 different  
275 studies (Barker et al., 2013, 2015; Burns et al., 2014; Claramunt & Cracraft, 2015; Gibb et  
276 al., 2015; Hedges et al., 2015; Hooper & Price, 2017; Jetz et al., 2012; Price et al., 2014).  
277 Once identified, DateLife pruned matching chronograms to keep Fringillidae species names  
278 on tips only, and transformed these pruned chronograms to pairwise distance matrices,  
279 revealing 1206 different age data points available for species within the Fringillidae  
280 (supplementray table 1).

281 ***Summarizing search results.***- The final step is to congruify and summarize the age data  
282 available for the Fringillidae species into single summary chronograms, using different types  
283 of summary ages, median and SDM. As explained in the “Description” section, a tree  
284 topology to summarize age data upon is required. By default, DateLife uses as summarizing  
285 topology the portion of OpenTree synthetic phylogeny that contains the species provided as  
286 search query. According to this phylogeny, species belonging to the family Fringillidae do not  
287 form a monophyletic group (Fig. 4).

288 Age data from source chronograms was congruified, decreasing to 818 different age data

289 points available (supplementray table 2) and then age summary statistics per congruified  
290 node are obtained and used as fixed secondary calibrations over a consensus tree topology, to  
291 obtain a fully dated phylogeny with the program BLADJ (Fig. ??). Median summary  
292 chronograms are older and have wider variation in maximum ages than chronograms  
293 obtained with SDM.



294

FIGURE 4. Tree topologies obtained from Open Tree of Life's (OpenTree) synthetic phylogenetic tree. A) Topology of 2,333 tips and 1,305 internal nodes, encompassing bird species within the family Fringillidae following the NCBI taxonomy (black), as well as all other bird species that share the same mrca node in OpenTree's synthetic tree (purple). B) Topology of 289 tips and 253 internal nodes, encompassing bird species within the Fringillidae only. This tree results from pruning species that are not assigned to the family Fringillidae (purple branches in topology A) but that form a monophyletic group with them. Bird species within the Fringillidae are paraphyletic (Alström et al. 2014, Barker, Cibois, Schikler, Feinstein, & Cracraft 2004, Barker et al. 2013, Barker 2014, Barker et al. 2015, Beresford, Barker, Ryan, & Crowe 2005, Bryson Jr et al. 2014, Burleigh, Kimball, & Braun 2015, Burns et al. 2014, Chaves, Hidalgo, & Klicka 2013, Claramunt & Cracraft 2015, Gibb et al. 2015, Hackett et al. 2008, Jetz et al. 2012, Johansson, Fjeldså, & Bowi 200, Kimball et al. 2019, Klicka et al. 2014, Lamichhaney et al. 2015, Lerner, Meyer, James, Hofreiter, & Fleischer 2011, Lovette et al. 2010, Moyle et al. 2016, Ödeen, Håstad, & Alström 2011, Oliveros et al. 2019, Päckert et al. 2012, Parchman, Benkman, & Mezquida 2007, Powell et al. 2014, Price et al. 2014, Pulgarín-R, Smith, Bryson Jr, Spellman, & Klicka 2013, Selvatti, Gonzaga, & Moraes Russo 2015, Tietze, Päckert, Martens, Lehmann, & Sun 2013, Treplin et al. 2008, Zuccon, Prŷs-Jones, Rasmussen, & Ericson 2012).

295

### Cross-validation test

- 296 We performed a cross validation analysis of the DateLife workflow using the Fringillidae  
297 chronograms. We used the individual tree topologies from each of the 19 source chronograms  
298 from 13 studies as inputs, treating their node ages as unknown. We then estimated dates for  
299 these topologies using the node ages from the chronograms from the other studies as  
300 calibrations and smoothing using BLADJ.  
301 We found that node ages from original study, and ages estimated using all other age data  
302 available are correlated (Supplementary Fig. ??).

303 For five studies, Datelife tended to underestimate ages for topologically deeper nodes (those  
304 with many descendant taxa, aka ‘closer to the root’) relative to the orginal estimate, and  
305 overestimate ages for nodes closer to the tips. Accordingly, root ages are generally older in  
306 the original study than estimated using cross-validated ages (Supplementary Fig. ??).

307 **Discussion**

308 **datelife** makes state-of-the-art data on evolutionary time frame easily accessible for  
309 comparison, reuse, and reanalysis, to researchers in all areas of science and with all levels of  
310 expertise in the matter. It is an open service that does not require any expert biological  
311 knowledge from users –besides the names of the species or group they want to work with, for  
312 any of its functionality.

313 A total of 99474 unique terminal taxa are represented in **datelife**’s database. Incorporation  
314 of more chronograms into **datelife**’s database will continue to improve its services. One  
315 option to increase the number of chronograms in the DateLife database is the Dryad data  
316 repository. Methods to automatically mine chronograms from Dryad could be designed and  
317 implemented. However, Dryad’s metadata system has no information to automatically detect  
318 branch length units, and those would still need to be determined manually by a human  
319 curator. We would like to emphasize on the importance of sharing chronogram data,  
320 including systematically curated metadata, into open repositories, such as OpenTree’s  
321 Phylesystem (McTavish et al., 2015) for the benefit of the scientific community as a whole.

322 **Primary vs Secondary calibrations in dating analyses**

323 While most chronograms in DateLife’s database are constructed using primary calibrations  
324 (molecular substitution rates or ages obtained from the fossil record or geological events),  
325 DateLife summarizes chronograms using secondary calibrations only (ages coming from other  
326 chronograms).

327 Graur and Martin (2004) cautioned on the increased error and uncertainty in estimated ages  
328 when using secondary calibrations in dating analyses. Schenk (2016) showed that, in

329 simulations, divergence times inferred using secondary calibrations are significantly younger  
330 than those inferred with primary calibrations, when obtained with Bayesian inference  
331 methods, and when priors are implemented in similar ways in both analyses. Accordingly,  
332 the scientific community seems to have more confidence in chronograms obtained from a  
333 single analysis, using fossil data as primary sources of calibrations (Schenk, 2016), and using  
334 fossils that have been widely discussed and curated as calibrations to date other trees,  
335 making sure that all data reflect a coherent evolutionary history (Sauquet, 2013), as for  
336 example done by Antonelli et al. (2017). There have been attempts to create fossil  
337 calibration databases (Ksepka et al., 2015), though these still have room to grow.

338 It seems that using several (as opposed to just a few) secondary calibrations can provide  
339 sufficient information to alleviate or even neutralize potential biases (Sauquet, 2013).  
340 Certainly, further studies are required to fully understand the effect of secondary calibrations  
341 on outputs from different tree dating methods, and on downstream analyses. It is possible  
342 that secondary calibrations can be safely used with dating methods that do not require  
343 setting priors, such as penalized likelihood (Michael J. Sanderson, 2003), with methods that  
344 do not make any assumptions on the ages and fix them to a node on a tree topology, such as  
345 BLADJ (Campbell O. Webb et al., 2008; Campbell O. Webb & Donoghue, 2005), or  
346 methods that summarize age data unto a tree topology.

347 Our cross validation analysis might provide some insight in this regard. When ages are  
348 estimated with secondary calibrations, nodes closer to the root do tend to be slightly  
349 younger than ages estimated with primary calibrations. However, nodes closer to the tip  
350 tend to be older when estimated using secondary calibrations with a dating method that  
351 does not make any prior assumptions on the nature of the calibrations themselves  
352 (Supplementary Figures 1-9), with a notable exception of cross validation of Burns et al.  
353 (2014) chronogram, which results in much younger node ages when estimated using  
354 secondary calibrations (Supplementary figs. 1 and 4).

### 355 Variation in source chronograms

356 Conflict among alternative dating studies is common in the literature (see, for example, the  
357 robust ongoing debate about crown group age of angiosperms (Barba-Montoya, Reis,  
358 Schneider, Donoghue, & Yang, 2018; Magallón, Gómez-Acevedo, Sánchez-Reyes, &  
359 Hernández-Hernández, 2015; Ramshaw et al., 1972; Michael J. Sanderson & Doyle, 2001;  
360 Sauquet, Ramírez-Barahona, & Magallón, 2021)).

361 Source chronograms may have been estimated using different types of primary calibrations  
362 (fossil record, geologic or biogeographic events, molecular substitution rates). These  
363 differences in approach can deepen the already substantial variation in time estimates  
364 between lineages, as observed from the comparison of source chronograms in the Fringillidae  
365 example. For example, the chronograms from Burns et al. (2014) were inferred using  
366 molecular substitution rate estimates across birds (Weir & Schlüter, 2008), and have much  
367 older date estimates for the same nodes than chronograms that were inferred using fossils as  
368 primary calibrations (figs. fig:fringillidages and fig:cvcXY; supplementary figs. 1 and 4).

369 Different chronograms might also imply fundamentally distinct evolutionary hypotheses. For  
370 example, two different researchers working on the same clade could both carefully select and  
371 argue for their choices of fossil calibrations. Still, if one researcher decides a fossil will  
372 calibrate the ingroup of a clade, while another researcher uses the same one to calibrate  
373 outside the clade, the resulting age estimates will often differ substantially, as the placement  
374 of calibrations as stem or crown group is proved to deeply affect estimated times of lineage  
375 divergence (Sauquet, 2013).

### 376 Summarizing chronograms

377 By default, DateLife currently summarizes all source chronograms that overlap with at least  
378 two species names. Users can exclude source chronograms if they have reasons to do so.

379 Strictly speaking, a good chronogram should reflect the real time of lineage divergence  
380 accurately and precisely. To our knowledge, there are no tested measures to determine

381 independently when a chronogram is better than another. Yet, several characteristics of the  
382 data used for dating analyses, as well as from the output chronogram itself, could be used to  
383 score the quality of source chronograms.

384 Some measures that have been proposed are the proportion of lineage sampling and the  
385 number of calibrations used Magallón et al. (2015). Some characteristics that are often cited  
386 in published studies as a measure of improved age estimates as compared to previously  
387 published estimates are: quality of alignment (missing data, GC content), lineage sampling  
388 (strategy and proportion), phylogenetic and dating inference method, number of fossils used  
389 as calibrations, support for nodes and ages, and magnitude of confidence intervals.

390 DateLife provides an opportunity to capture concordance and conflict among date estimates,  
391 which can also be used as a metric for chronogram reliability. Its open database of  
392 chronograms allows other researchers to do such analyses themselves reproducibly, and  
393 without needing permission. Though, of course, they should follow proper citation practices,  
394 especially for the source chronogram studies.

395 The exercise of summarizing age data from across multiple studies provides the opportunity  
396 to work with a more inclusive chronogram, that reflects a unified evolutionary history for a  
397 lineage, by putting together evidence from different hypotheses. The largest, and  
398 taxonomically broadest chronogram currently available from OpenTree was constructed  
399 summarizing age data from 2,274 published chronograms using NCBI's taxonomic tree as  
400 backbone (Hedges et al., 2015). A summarizing exercise may also amplify the effect of  
401 uncertainty and errors in source data, and blur parts of the evolutionary history of a lineage  
402 that might only be reflected in source chronograms and lost on the summary chronogram  
403 (Sauquet et al., 2021).

#### 404 Effects on downstream analyses

405 For downstream analyses, using alternative chronogram may deeply affect our inferences  
406 (Title & Rabosky, 2016), particularly when studying phenomena dependent on the timing of

407 species diversification events, such as macroevolutionary processes.

408 In ecology and conservation biology, incorporating at least some data on lineage divergence

409 times represents a relevant improvement for testing alternative hypothesis using phylogenetic

410 distance (Campbell O. Webb et al., 2008). Hence, DateLife’s workflow features different

411 ways of estimating node ages in the absence of calibrations and branch length information

412 for certain taxa. “Making up” branch lengths is a common practice in scientific publications:

413 Jetz et al. (2012), created a time-calibrated tree of all 9,993 bird species, where 67% had

414 molecular data and the rest was simulated; Rabosky et al. (2018) created a time-calibrated

415 tree of 31,536 ray-finned fishes, of which only 37% had molecular data; Stephen A. Smith and

416 Brown (2018) constructed a tree of 353,185 seed plants where only 23% had molecular data.

417 Notably, risks come with this practice. Taken to the extreme, one could make a fully

418 resolved, calibrated tree of all modern and extinct taxa using a single taxonomy and a single

419 calibration, using polytomy resolution and branch estimation methods. There has yet to be

420 a thorough analysis of what can go wrong when one extends inferences beyond the data in

421 this way, so we urge caution; we also urge readers to follow the example of the large tree

422 papers cited above, by carefully considering the statistical assumptions being made, and

423 assessing the consistency of the results with prior work.

## 424 Conclusions

425 Knowledge of the evolutionary time frame of organisms is key to many research areas: trait

426 evolution, species diversification, biogeography, macroecology and more. It is also crucial for

427 education, science communication and policy, but generating chronograms is difficult,

428 especially for those who want to use phylogenies but who are not systematists, or do not

429 have the time to acquire and develop the necessary knowledge and skills to construct them

430 on their own. Importantly, years of primarily public funded research have resulted in vast

431 amounts of chronograms that are already available on scientific publications, but hidden to

432 the public and scientific community for reuse.

433 The DateLife project allows for easy and fast summary of public and state-of-the-art data on  
434 time of lineage divergence. It provides a straightforward way to get an informed idea on the  
435 state of knowledge of the time frame of evolution of different regions of the tree of life, and  
436 allows identification of regions that require more research, or that have conflicting  
437 information. It is available as an R package, and as a web-based R shiny application at  
438 [www.datelife.org](http://www.datelife.org) Both summary and newly generated trees are useful to evaluate  
439 evolutionary hypotheses in different areas of research. The DateLife project helps with  
440 awareness of the existing variation in expert time of divergence data, and will foster  
441 exploration of the effect of alternative divergence time hypothesis on the results of analyses,  
442 nurturing a culture of more cautious interpretation of evolutionary results.

443 **Availability**

444 The **datelife** software is free and open source and it can be used through its R shiny web  
445 application at <http://www.datelife.org>, through the **datelife** R package, and through  
446 Phylotastic's project web portal <https://phylo.cs.nmsu.edu/>. DateLife's web application is  
447 maintained using RStudio's shiny server and the shiny package open infrastructure, as well  
448 as Docker and Open Tree of Life's infrastructure ([dates.opentreeoflife.org/datelife](https://dates.opentreeoflife.org/datelife)).  
449 **datelife**'s R package stable version is available for installation from the CRAN repository  
450 (<https://cran.r-project.org/package=datelife>) using the command `install.packages(pkgs`  
451 `= "datelife")` from within R. Development versions are available from the GitHub  
452 repository (<https://github.com/phylotastic/datelife>) and can be installed using the  
453 command `devtools::install_github("phylotastic/datelife")`.

454 **Supplementary Material**

455 Code used to generate all versions of this manuscript, the biological examples, as well as the  
456 benchmark of functionalities are available at datelifeMS1, datelife\_examples, and  
457 datelife\_benchmark repositories in LLSR's GitHub account.

458

## Funding

459 Funding was provided by the US National Science Foundation (NSF) grants ABI-1458603 to  
460 the Datelife project; DBI-0905606 to the National Evolutionary Synthesis Center (NESCent),  
461 ABI-1458572 to the Phylotastic project, and ABI-1759846 to the Open Tree of Life project.

462

## Acknowledgements

463 The DateLife project was born as a prototype tool aiming to provide these services, and was  
464 initially developed over a series of hackathons at the National Evolutionary Synthesis Center,  
465 NC, USA (Stoltzfus et al., 2013). We thank colleagues from the O'Meara Lab at the  
466 University of Tennessee Knoxville for suggestions, discussions and software testing. The late  
467 National Evolutionary Synthesis Center (NESCent), which sponsored hackathons that led to  
468 initial work on this project. The team that assembled DateLife's first proof of concept:  
469 Tracy Heath, Jonathan Eastman, Peter Midford, Joseph Brown, Matt Pennell, Mike Alfaro,  
470 and Luke Harmon. The Open Tree of Life project that provides the open, metadata rich  
471 repository of trees used to construct DateLife's chronogram database. The many scientists  
472 who publish their chronograms in an open, reusable form, and the scientists who curate them  
473 for deposition in the Open Tree of Life repository. The NSF for funding nearly all the above,  
474 in addition to the ABI grant that funded this project itself.

**References**

- 475 Alström, P., Hooper, D. M., Liu, Y., Olsson, U., Mohan, D., Gelang, M., ... Price, T.  
476 D. (2014). Discovery of a relict lineage and monotypic family of passerine birds.  
477 *Biology Letters*, 10(3), 20131067.
- 478 Ané, C., Eulensteiner, O., Piaggio-Talice, R., & Sanderson, M. J. (2009). Groves of  
479 phylogenetic trees. *Annals of Combinatorics*, 13(2), 139–167.
- 480 Antonelli, A., Hettling, H., Condamine, F. L., Vos, K., Nilsson, R. H., Sanderson, M.  
481 J., ... Vos, R. A. (2017). Toward a self-updating platform for estimating rates of  
482 speciation and migration, ages, and relationships of Taxa. *Systematic Biology*,  
483 66(2), 153–166. <https://doi.org/10.1093/sysbio/syw066>
- 484 Archie, J., Day, W. H., Felsenstein, J., Maddison, W., Meacham, C., Rohlf, F. J., &  
485 Swofford, D. (1986). The Newick tree format. Retrieved from  
486 %7B<https://evolution.genetics.washington.edu/phylip/newicktree.html>%7D
- 487 Avibase. (2022). Yellow-throated Bunting. *Avibase - The World Bird Database*,  
488 (Online Resource). Retrieved from %7B<https://avibase.bsc-eoc.org/species.jsp?lang=EN&avibaseid=82D1EE0049D8D927%7D>
- 489 Bapst, D. W. (2012). Paleotree: An R package for paleontological and phylogenetic  
490 analyses of evolution. *Methods in Ecology and Evolution*, 3(5), 803–807.  
491 <https://doi.org/10.1111/j.2041-210X.2012.00223.x>
- 492 Barba-Montoya, J., Reis, M. dos, Schneider, H., Donoghue, P. C., & Yang, Z. (2018).  
493 Constraining uncertainty in the timescale of angiosperm evolution and the  
494 veracity of a cretaceous terrestrial revolution. *New Phytologist*, 218(2), 819–834.
- 495 Barker, F. K. (2014). Mitogenomic data resolve basal relationships among passeriform  
496 and passeridan birds. *Molecular Phylogenetics and Evolution*, 79, 313–324.
- 497 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2013). Going  
498 to extremes: Contrasting rates of diversification in a recent radiation of new world  
499 passerine birds. *Systematic Biology*, 62(2), 298–320.
- 500

- 502 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2015). New  
503 insights into new world biogeography: An integrated view from the phylogeny of  
504 blackbirds, cardinals, sparrows, tanagers, warblers, and allies. *The Auk: Ornithological Advances*, 132(2), 333–348.
- 505
- 506 Barker, F. K., Cibois, A., Schikler, P., Feinstein, J., & Cracraft, J. (2004). Phylogeny  
507 and diversification of the largest avian radiation. *Proceedings of the National  
508 Academy of Sciences*, 101(30), 11040–11045.
- 509 Beresford, P., Barker, F., Ryan, Pg., & Crowe, T. (2005). African endemics span the  
510 tree of songbirds (passeri): Molecular systematics of several evolutionary  
511 “enigmas.” *Proceedings of the Royal Society B: Biological Sciences*, 272(1565),  
512 849–858.
- 513 Boyle, B., Hopkins, N., Lu, Z., Raygoza Garay, J. A., Mozzherin, D., Rees, T., ...  
514 Enquist, B. J. (2013). The taxonomic name resolution service: An online tool for  
515 automated standardization of plant names. *BMC Bioinformatics*, 14(1).  
516 <https://doi.org/10.1186/1471-2105-14-16>
- 517 Britton, T., Anderson, C. L., Jacquet, D., Lundqvist, S., & Bremer, K. (2007).  
518 Estimating Divergence Times in Large Phylogenetic Trees. *Systematic Biology*,  
519 56(788777878), 741–752. <https://doi.org/10.1080/10635150701613783>
- 520 Bryson Jr, R. W., Chaves, J., Smith, B. T., Miller, M. J., Winker, K., Pérez-Emán, J.  
521 L., & Klicka, J. (2014). Diversification across the new world within the  
522 ‘blue’cardinalids (aves: cardinalidae). *Journal of Biogeography*, 41(3), 587–599.
- 523 Burleigh, J. G., Kimball, R. T., & Braun, E. L. (2015). Building the avian tree of life  
524 using a large-scale, sparse supermatrix. *Molecular Phylogenetics and Evolution*,  
525 84, 53–63.
- 526 Burns, K. J., Shultz, A. J., Title, P. O., Mason, N. A., Barker, F. K., Klicka, J., ...  
527 Lovette, I. J. (2014). Phylogenetics and diversification of tanagers (passeriformes:  
528 Thraupidae), the largest radiation of neotropical songbirds. *Molecular*

- 529                   *Phylogenetics and Evolution*, 75, 41–77.
- 530 Chamberlain, S. A., & Szöcs, E. (2013). taxize : taxonomic search and retrieval in R  
531 [version 2; referees: 3 approved]. *F1000Research*, 2(191), 1–29.  
532 <https://doi.org/10.12688/f1000research.2-191.v2>
- 533 Chamberlain, S. A., Szöcs, E., Foster, Z., Arendsee, Z., Boettiger, C., Ram, K., ...  
534 Li, G. (2019). *taxize: Taxonomic information from around the web*. Retrieved  
535 from <https://github.com/ropensci/taxize>
- 536 Chaves, J. A., Hidalgo, J. R., & Klicka, J. (2013). Biogeography and evolutionary  
537 history of the n eotropical genus s altator (a ves: T hraupini). *Journal of  
538 Biogeography*, 40(11), 2180–2190.
- 539 Claramunt, S., & Cracraft, J. (2015). A new time tree reveals earth history's imprint  
540 on the evolution of modern birds. *Science Advances*, 1(11), e1501005.
- 541 Criscuolo, A., Berry, V., Douzery, E. J. P., & Gascuel, O. (2006). SDM: A fast  
542 distance-based approach for (super)tree building in phylogenomics. *Systematic  
543 Biology*, 55(5), 740–755. <https://doi.org/10.1080/10635150600969872>
- 544 Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M.-K., Turon, X., ...  
545 Douzery, E. J. (2018). A phylogenomic framework and timescale for comparative  
546 studies of tunicates. *BMC Biology*, 16(1), 1–14.
- 547 Eastman, J. M., Harmon, L. J., & Tank, D. C. (2013). Congruification: Support for  
548 time scaling large phylogenetic trees. *Methods in Ecology and Evolution*, 4(7),  
549 688–691. <https://doi.org/10.1111/2041-210X.12051>
- 550 Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and  
551 high throughput. *Nucleic Acids Research*, 32(5), 1792–1797.
- 552 Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American  
553 Naturalist*, 125(1), 1–15. Retrieved from <http://www.jstor.org/stable/2461605>
- 554 Gibb, G. C., England, R., Hartig, G., McLenaghan, P. A., Taylor Smith, B. L.,  
555 McComish, B. J., ... Penny, D. (2015). New zealand passerines help clarify the

- 556 diversification of major songbird lineages during the oligocene. *Genome Biology  
557 and Evolution*, 7(11), 2983–2995.
- 558 Graur, D., & Martin, W. (2004). Reading the entrails of chickens: Molecular  
559 timescales of evolution and the illusion of precision. *TRENDS in Genetics*, 20(2),  
560 80–86.
- 561 Hackett, S. J., Kimball, R. T., Reddy, S., Bowie, R. C., Braun, E. L., Braun, M. J.,  
562 ... others. (2008). A phylogenomic study of birds reveals their evolutionary  
563 history. *Science*, 320(5884), 1763–1768.
- 564 Hedges, S. B., Marin, J., Suleski, M., Paymer, M., & Kumar, S. (2015). Tree of life  
565 reveals clock-like speciation and diversification. *Molecular Biology and Evolution*,  
566 32(4), 835–845. <https://doi.org/10.1093/molbev/msv037>
- 567 Heibl, C. (2008). *PHYLOCH: R language tree plotting tools and interfaces to diverse  
568 phylogenetic software packages*. Retrieved from  
569 <http://www.christophheibl.de/Rpackages.html>
- 570 Hooper, D. M., & Price, T. D. (2017). Chromosomal inversion differences correlate  
571 with range overlap in passerine birds. *Nature Ecology & Evolution*, 1(10), 1526.
- 572 Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of  
573 phylogenetic trees. *Bioinformatics*, 17(8), 754–755.  
574 <https://doi.org/10.1093/bioinformatics/17.8.754>
- 575 Jetz, W., Thomas, G., Joy, J. J. B., Hartmann, K., & Mooers, A. (2012). The global  
576 diversity of birds in space and time. *Nature*, 491(7424), 444–448.  
577 <https://doi.org/10.1038/nature11631>
- 578 Johansson, U. S., Fjeldså, J., & Bowie, R. C. (2008). Phylogenetic relationships  
579 within passerida (aves: Passeriformes): A review and a new molecular phylogeny  
580 based on three nuclear intron markers. *Molecular Phylogenetics and Evolution*,  
581 48(3), 858–876.
- 582 Katoh, K., Asimenos, G., & Toh, H. (2009). Multiple alignment of DNA sequences

- 583 with MAFFT. In *Bioinformatics for DNA sequence analysis* (pp. 39–64).  
584 Springer.
- 585 Kimball, R. T., Oliveros, C. H., Wang, N., White, N. D., Barker, F. K., Field, D. J.,  
586 ... others. (2019). A phylogenomic supertree of birds. *Diversity*, 11(7), 109.
- 587 Klicka, J., Barker, F. K., Burns, K. J., Lanyon, S. M., Lovette, I. J., Chaves, J. A., &  
588 Bryson Jr, R. W. (2014). A comprehensive multilocus assessment of sparrow (aves:  
589 Passerellidae) relationships. *Molecular Phylogenetics and Evolution*, 77, 177–182.
- 590 Ksepka, D. T., Parham, J. F., Allman, J. F., Benton, M. J., Carrano, M. T.,  
591 Cranston, K. A., ... others. (2015). The fossil calibration database—a new  
592 resource for divergence dating. *Systematic Biology*, 64(5), 853–859.
- 593 Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M.,  
594 Martinez-Barrio, A., ... others. (2015). Evolution of darwin's finches and their  
595 beaks revealed by genome sequencing. *Nature*, 518(7539), 371–375.
- 596 Laubichler, M. D., & Maienschein, J. (2009). *Form and function in developmental  
597 evolution*. Cambridge University Press.
- 598 Lepage, D. (2004). *Avibase: The world bird database*. Bird Studies Canada.
- 599 Lepage, D., Vaidya, G., & Guralnick, R. (2014). Avibase—a database system for  
600 managing and organizing taxonomic concepts. *ZooKeys*, (420), 117.
- 601 Lerner, H. R., Meyer, M., James, H. F., Hofreiter, M., & Fleischer, R. C. (2011).  
602 Multilocus resolution of phylogeny and timescale in the extant adaptive radiation  
603 of hawaiian honeycreepers. *Current Biology*, 21(21), 1838–1844.
- 604 Lovette, I. J., Pérez-Emán, J. L., Sullivan, J. P., Banks, R. C., Fiorentino, I.,  
605 Córdoba-Córdoba, S., ... others. (2010). A comprehensive multilocus phylogeny  
606 for the wood-warblers and a revised classification of the parulidae (aves).  
607 *Molecular Phylogenetics and Evolution*, 57(2), 753–770.
- 608 Magallon, S., & Sanderson, M. J. (2001). Absolute diversification rates in angiosperm  
609 clades. *Evolution*, 55(9), 1762–1780.

- 610 Magallón, S. (2010). Using fossils to break long branches in molecular dating: A  
611 comparison of relaxed clocks applied to the origin of angiosperms. *Systematic*  
612 *Biology*, 59(4), 384–399.
- 613 Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., & Hernández-Hernández, T.  
614 (2015). A metacalibrated time-tree documents the early rise of flowering plant  
615 phylogenetic diversity. *New Phytologist*, 207(2), 437–453.
- 616 McTavish, E. J., Hinchliff, C. E., Allman, J. F., Brown, J. W., Cranston, K. A.,  
617 Holder, M. T., . . . Smith, S. A. (2015). Phylesystem: A git-based data store for  
618 community-curated phylogenetic estimates. *Bioinformatics*, 31(17), 2794–2800.
- 619 Michonneau, F., Brown, J. W., & Winter, D. J. (2016). rotl: an R package to interact  
620 with the Open Tree of Life data. *Methods in Ecology and Evolution*, 7(12),  
621 1476–1481. <https://doi.org/10.1111/2041-210X.12593>
- 622 Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecology*  
623 *Letters*, 17(4), 508–525. <https://doi.org/10.1111/ele.12251>
- 624 Moyle, R. G., Oliveros, C. H., Andersen, M. J., Hosner, P. A., Benz, B. W., Manthey,  
625 J. D., . . . Faircloth, B. C. (2016). Tectonic collision and uplift of wallacea  
626 triggered the global songbird radiation. *Nature Communications*, 7(1), 1–7.
- 627 Ödeen, A., Håstad, O., & Alström, P. (2011). Evolution of ultraviolet vision in the  
628 largest avian radiation—the passerines. *BMC Evolutionary Biology*, 11(1), 1–8.
- 629 Oliveros, C. H., Field, D. J., Ksepka, D. T., Barker, F. K., Aleixo, A., Andersen, M.  
630 J., . . . others. (2019). Earth history and the passerine superradiation.  
631 *Proceedings of the National Academy of Sciences*, 116(16), 7916–7925.
- 632 Ooms, J., & Chamberlain, S. (2018). *Phylocomr: Interface to 'phylocom'*. Retrieved  
633 from <https://CRAN.R-project.org/package=phylocomr>
- 634 Open Tree Of Life, Redelings, B., Cranston, K. A., Allman, J., Holder, M. T., &  
635 McTavish, E. J. (2016). Open Tree of Life APIs v3.0. *Open Tree of Life Project*,  
636 (Online Resources). Retrieved from

- 637                  %7Bhttps://github.com/OpenTreeOfLife/germinator/wiki/Open-Tree-of-Life-  
638                  Web-APIs%7D
- 639                  Open Tree Of Life, Redelings, B., Sánchez Reyes, L. L., Cranston, K. A., Allman, J.,  
640                  Holder, M. T., & McTavish, E. J. (2019). Open tree of life synthetic tree v12.3.  
641                  *Zenodo*. Retrieved from <https://doi.org/10.5281/zenodo.3937742>
- 642                  Päckert, M., Martens, J., Sun, Y.-H., Severinghaus, L. L., Nazarenko, A. A., Ting, J.,  
643                  ... Tietze, D. T. (2012). Horizontal and elevational phylogeographic patterns of  
644                  himalayan and southeast asian forest passerines (aves: passeriformes). *Journal of*  
645                  *Biogeography*, 39(3), 556–573.
- 646                  Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and  
647                  evolution in R language. *Bioinformatics*, 20(2), 289–290.
- 648                  Parchman, T. L., Benkman, C. W., & Mezquida, E. T. (2007). Coevolution between  
649                  hispaniolan crossbills and pine: Does more time allow for greater phenotypic  
650                  escalation at lower latitude? *Evolution*, 61(9), 2142–2153.
- 651                  Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., FitzJohn,  
652                  R. G., ... Harmon, L. J. (2014). Geiger v2. 0: An expanded suite of methods for  
653                  fitting macroevolutionary models to phylogenetic trees. *Bioinformatics*, 30(15),  
654                  2216–2218.
- 655                  Posadas, P., Crisci, J. V., & Katinas, L. (2006). Historical biogeography: A review of  
656                  its basic concepts and critical issues. *Journal of Arid Environments*, 66(3),  
657                  389–403.
- 658                  Powell, A. F., Barker, F. K., Lanyon, S. M., Burns, K. J., Klicka, J., & Lovette, I. J.  
659                  (2014). A comprehensive species-level molecular phylogeny of the new world  
660                  blackbirds (icteridae). *Molecular Phylogenetics and Evolution*, 71, 94–112.
- 661                  Price, T. D., Hooper, D. M., Buchanan, C. D., Johansson, U. S., Tietze, D. T.,  
662                  Alström, P., ... others. (2014). Niche filling slows the diversification of himalayan  
663                  songbirds. *Nature*, 509(7499), 222.

- 664 Pulgarín-R, P. C., Smith, B. T., Bryson Jr, R. W., Spellman, G. M., & Klicka, J.  
665 (2013). Multilocus phylogeny and biogeography of the new world pheucticus  
666 grosbeaks (aves: cardinalidae). *Molecular Phylogenetics and Evolution*, 69(3),  
667 1222–1227.
- 668 R Core Team. (2018). *R: a language and environment for statistical computing*.  
669 Vienna, Austria: R Foundation for Statistical Computing.
- 670 Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., ...  
671 others. (2018). An inverse latitudinal gradient in speciation rate for marine fishes.  
672 *Nature*, 559(7714), 392.
- 673 Ramshaw, J., Richardson, D., Mealyard, B., Brown, R., Richardson, M., Thompson,  
674 E., & Boulter, D. (1972). The time of origin of the flowering plants determined by  
675 using amino acid sequence data of cytochrome c. *New Phytologist*, 71(5), 773–779.
- 676 Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The barcode of life data system  
677 (<http://www.Barcodinglife.org>). *Molecular Ecology Notes*, 7(3), 355–364.
- 678 Rees, J. A., & Cranston, K. (2017). Automated assembly of a reference taxonomy for  
679 phylogenetic data synthesis. *Biodiversity Data Journal*, (5).
- 680 Revell, L. J. (2012). Phytools: An r package for phylogenetic comparative biology  
681 (and other things). *Methods in Ecology and Evolution*, 3, 217–223.
- 682 Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic  
683 inference under mixed models. *Bioinformatics*, 19(12), 1572–1574.  
684 <https://doi.org/10.1093/bioinformatics/btg180>
- 685 Sanchez-Reyes, L. L., & O'Meara, B. (2022). **datelifeplot**: Methods to plot  
686 chronograms and outputs of the **datelife** package. *R Package Release V0.2.2*.  
687 Retrieved from <https://zenodo.org/badge/latestdoi/381501451>
- 688 Sanchez-Reyes, L. L., O'Meara, B., Eastman, J., Heath, T., Wright, A., Schliep, K.,  
689 ... Alfaro, M. (2022). **datelife**: Scientific Data on Time of Lineage Divergence  
690 for Your Taxa. *R Package Release V0.6.2*. Retrieved from

- 691 https://doi.org/10.5281/zenodo.593938
- 692 Sanderson, Michael J. (2002). Estimating Absolute Rates of Molecular Evolution and  
693 Divergence Times: A Penalized Likelihood Approach. *Molecular Biology and*  
694 *Evolution*, 19(1), 101–109.
- 695 https://doi.org/10.1093/oxfordjournals.molbev.a003974
- 696 Sanderson, Michael J. (2003). r8s: Inferring absolute rates of molecular evolution and  
697 divergence times in the absence of a molecular clock. *Bioinformatics*, 19(2),  
698 301–302.
- 699 Sanderson, Michael J., & Doyle, J. A. (2001). Sources of error and confidence  
700 intervals in estimating the age of angiosperms from rbcL and 18S rDNA data.  
701 *American Journal of Botany*, 88(8), 1499–1516.
- 702 Sauquet, H. (2013). A practical guide to molecular dating. *Comptes Rendus Palevol*,  
703 12(6), 355–367.
- 704 Sauquet, H., Ramírez-Barahona, S., & Magallón, S. (2021). *The age of flowering*  
705 *plants is unknown*.
- 706 Schenk, J. J. (2016). Consequences of secondary calibrations on divergence time  
707 estimates. *PLoS ONE*, 11(1). <https://doi.org/10.1371/journal.pone.0148228>
- 708 Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in r. *Bioinformatics*, 27(4),  
709 592–593.
- 710 Selvatti, A. P., Gonzaga, L. P., & Moraes Russo, C. A. de. (2015). A paleogene origin  
711 for crown passerines and the diversification of the oscines in the new world.  
712 *Molecular Phylogenetics and Evolution*, 88, 1–15.
- 713 Smith, Stephen A., & Brown, J. W. (2018). Constructing a broadly inclusive seed  
714 plant phylogeny. *American Journal of Botany*, 105(3), 302–314.
- 715 Smith, Stephen A., & O'Meara, B. C. (2012). TreePL: Divergence time estimation  
716 using penalized likelihood for large phylogenies. *Bioinformatics*, 28(20),  
717 2689–2690. <https://doi.org/10.1093/bioinformatics/bts492>

- 718 Stoltzfus, A., Lapp, H., Matasci, N., Deus, H., Sidlauskas, B., Zmasek, C. M., ...  
719 Jordan, G. (2013). Phylotastic! Making tree-of-life knowledge accessible, reusable  
720 and convenient. *BMC Bioinformatics*, 14.  
721 <https://doi.org/10.1186/1471-2105-14-158>
- 722 Tietze, D. T., Päckert, M., Martens, J., Lehmann, H., & Sun, Y.-H. (2013). Complete  
723 phylogeny and historical biogeography of true rosefinches (aves: carpodacus).  
724 *Zoological Journal of the Linnean Society*, 169(1), 215–234.
- 725 Title, P. O., & Rabosky, D. L. (2016). Do Macrophylogenies Yield Stable  
726 Macroevolutionary Inferences? An Example from Squamate Reptiles. *Systematic  
727 Biology*, syw102. <https://doi.org/10.1093/sysbio/syw102>
- 728 Treplin, S., Siegert, R., Bleidorn, C., Thompson, H. S., Fotso, R., & Tiedemann, R.  
729 (2008). Molecular phylogeny of songbirds (aves: Passeriformes) and the relative  
730 utility of common nuclear marker loci. *Cladistics*, 24(3), 328–349.
- 731 Vos, R. A., Balhoff, J. P., Caravas, J. A., Holder, M. T., Lapp, H., Maddison, W. P.,  
732 ... others. (2012). NeXML: Rich, extensible, and verifiable representation of  
733 comparative data and metadata. *Systematic Biology*, 61(4), 675–689.
- 734 Webb, C. (2000). Exploring the Phylogenetic Structure of Ecological Communities :  
735 An Example for Rain Forest Trees. *The American Naturalist*, 156(2), 145–155.
- 736 Webb, Campbell O., Ackerly, D. D., & Kembel, S. W. (2008). Phylocom: Software for  
737 the analysis of phylogenetic community structure and trait evolution.  
738 *Bioinformatics*, 24(18), 2098–2100.  
739 <https://doi.org/10.1093/bioinformatics/btn358>
- 740 Webb, Campbell O., & Donoghue, M. J. (2005). Phylomatic: Tree assembly for  
741 applied phylogenetics. *Molecular Ecology Notes*, 5(1), 181–183.
- 742 Weir, J., & Schluter, D. (2008). Calibrating the avian molecular clock. *Molecular  
743 Ecology*, 17(10), 2321–2328.

TABLE 1

*Ages of congruified nodes obtained from a datelife search for the Small Example presented.*

*See Figure 3, step C2.*

Node Name	taxon A	taxon B	Node Age	Study chronogram
1 n1	<i>Emberiza citrinella</i>	<i>Platyspiza crassirostris</i>	9.6509881	Jetz et al. 2012 – chronogram 1
2 n1	<i>Pheucticus tibialis</i>	<i>Schoeniclus elegans</i>	14.3336520	Barker et al. 2015 – chronogram 1
3 n1	<i>Rhodothraupis celaeno</i>	<i>Schoeniclus elegans</i>	16.2984859	Barker et al. 2015 – chronogram 2
4 n1	<i>Emberiza citrinella</i>	<i>Platyspiza crassirostris</i>	16.9499615	Hooper et al. 2017
5 n1	<i>Rhodothraupis celaeno</i>	<i>Schoeniclus elegans</i>	21.5140867	Jetz et al. 2012 – chronogram 2
6 n1	<i>Emberiza leucocephalos</i>	<i>Platyspiza crassirostris</i>	22.0000000	Hedges et al. 2015 – chronogram 2
7 n1	<i>Emberiza citrinella</i>	<i>Platyspiza crassirostris</i>	22.3757277	Hedges et al. 2015 – chronogram 1
8 n2	<i>Platyspiza crassirostris</i>	<i>Rhodothraupis celaeno</i>	7.9691925	Jetz et al. 2012 – chronogram 1
9 n2	<i>Platyspiza crassirostris</i>	<i>Rhodothraupis celaeno</i>	19.7085830	Jetz et al. 2012 – chronogram 2
10 n2	<i>Platyspiza crassirostris</i>	<i>Rhodothraupis celaeno</i>	19.7085900	Hedges et al. 2015 – chronogram 2
11 n2	<i>Platyspiza crassirostris</i>	<i>Rhodothraupis celaeno</i>	19.7128363	Hedges et al. 2015 – chronogram 1
12 n2	<i>Pheucticus tibialis</i>	<i>Platyspiza crassirostris</i>	44.2958603	Burns et al. 2014
13 n3	<i>Pheucticus tibialis</i>	<i>Rhodothraupis celaeno</i>	10.5304440	Barker et al. 2015 – chronogram 1
14 n3	<i>Pheucticus tibialis</i>	<i>Rhodothraupis celaeno</i>	10.5379092	Barker et al. 2013
15 n3	<i>Pheucticus tibialis</i>	<i>Rhodothraupis celaeno</i>	11.2095375	Barker et al. 2015 – chronogram 2
16 n3	<i>Pheucticus tibialis</i>	<i>Rhodothraupis celaeno</i>	18.1570685	Hedges et al. 2015 – chronogram 1
17 n4	<i>Emberiza citrinella</i>	<i>Schoeniclus elegans</i>	5.3238969	Jetz et al. 2012 – chronogram 1
18 n4	<i>Emberiza leucocephalos</i>	<i>Schoeniclus elegans</i>	9.8622460	Barker et al. 2015 – chronogram 1
19 n4	<i>Emberiza leucocephalos</i>	<i>Schoeniclus elegans</i>	10.3391445	Jetz et al. 2012 – chronogram 2
20 n4	<i>Emberiza leucocephalos</i>	<i>Schoeniclus elegans</i>	11.7317630	Barker et al. 2015 – chronogram 2
21 n4	<i>Emberiza citrinella</i>	<i>Schoeniclus elegans</i>	12.5133870	Hooper et al. 2017
22 n5	<i>Emberiza citrinella</i>	<i>Emberiza leucocephalos</i>	0.1407015	Jetz et al. 2012 – chronogram 1
23 n5	<i>Emberiza citrinella</i>	<i>Emberiza leucocephalos</i>	0.1516230	Hedges et al. 2015 – chronogram 2
24 n5	<i>Emberiza citrinella</i>	<i>Emberiza leucocephalos</i>	0.2011990	Barker et al. 2015 – chronogram 1
25 n5	<i>Emberiza citrinella</i>	<i>Emberiza leucocephalos</i>	0.2409300	Barker et al. 2015 – chronogram 2
26 n5	<i>Emberiza citrinella</i>	<i>Emberiza leucocephalos</i>	0.2732460	Jetz et al. 2012 – chronogram 2
27 n5	<i>Emberiza citrinella</i>	<i>Emberiza leucocephalos</i>	0.5760260	Hedges et al. 2015 – chronogram 1
28 n5	<i>Emberiza citrinella</i>	<i>Emberiza leucocephalos</i>	2.2898230	Hooper et al. 2017

TABLE 2

*Summary of congruified nodes ages. See Figure 3, step C3.*

Node Name	Min Age	Q1	Median Age	Mean Age	Q3	Max Age	Variance	SD
n1	9.6509881	15.316069	16.94996	17.5889860	21.757043	22.375728	22.2431847	4.7162681
n2	7.9691925	19.708583	19.70859	22.2790124	19.712836	44.295860	177.3279940	13.3164558
n3	10.5304440	10.536043	10.87372	12.6087398	12.946420	18.157069	13.7831237	3.7125630
n4	5.3238969	9.862246	10.33914	9.9540875	11.731763	12.513387	7.8263782	2.7975665
n5	0.1407015	0.176411	0.24093	0.5533641	0.424636	2.289823	0.6079318	0.7796998