

<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, USA

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

<sup>5</sup> Author Note

<sup>6</sup> School of Natural Sciences, University of California, Merced, 258 Science and  
<sup>7</sup> Engineering Building 1, Merced, CA 95340, USA.

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

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

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

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

18 Abstract

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

<sup>43</sup> *Keywords:* Tree; Phylogeny; Scaling; Dating; Ages; Divergence times; Open Science;

<sup>44</sup> Congruification; Supertree; Calibrations; Secondary calibrations

<sup>45</sup> Word count: 5393

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

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

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

## DESCRIPTION

DateLife's core software application consists of the R package **datelife**. Its current stable version – v0.6.8, is available from The Comprehensive R Archive Network (CRAN) repository (Sanchez-Reyes et al., 2022), and relies on functionalities from various biological

71 R packages: ape (Paradis, Claude, & Strimmer, 2004), bold (Chamberlain et al., 2019),  
72 geiger (Pennell et al., 2014), paleotree (Bapst, 2012), phyloch (Heibl, 2008), phylocomr  
73 (Ooms & Chamberlain, 2018), phytools (Revell, 2012), rotl (Michonneau, Brown, & Winter,  
74 2016), and taxize (Chamberlain & Szöcs, 2013; Chamberlain et al., 2019). Figure 1 provides  
75 a graphical summary of the three main steps of the DateLife workflow: creating a search  
76 query, searching a database, and summarizing results from the search.

77 *Creating a Search Query*

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

83 DateLife accepts scientific names that can belong to any inclusive taxonomic group  
84 (e.g., genus, family, tribe, etc.) or a binomial species name. Subspecies and variants are  
85 ignored. If an input taxon name belongs to an inclusive taxonomic group, DateLife has two  
86 alternative behaviors defined by the “get species from taxon” flag. If the flag is active,  
87 DateLife retrieves all species names within the inclusive taxonomic group following a  
88 standard taxonomy of choice, and adds them to the input string. Taxonomies currently  
89 supported by DateLife are Open Tree of Life (OpenTree) unified Taxonomy (OTT, Rees &  
90 Cranston, 2017), the National Center of Biotechnology Information (NCBI) taxonomic  
91 database (Schoch et al., 2020), the Global Biodiversity Information Facility (GBIF)  
92 taxonomic backbone (GBIF Secretariat, 2022), and the Interim Register of Marine and  
93 Nonmarine Genera (IRMNG) database (Rees, Vandeputte, Decock, & Vanhoorne, 2017). If  
94 the flag is inactive, DateLife excludes any taxon names above the species level from the  
95 search query.

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

103 The processed input taxon names are saved as an R object of a newly defined class,  
104 `datelifeQuery`, that is used in the following steps. This object contains the standardized  
105 names, the corresponding OTT identification numbers, and the topology of the input tree if  
106 any was provided.

107 *Searching a Chronogram Database*

108 At the time of writing of this manuscript (Jun 22, 2022), DateLife's chronogram  
109 database latest version consist of 253 chronograms published in 187 different studies. It is  
110 curated from OpenTree's phylogenetic database, the Phylesystem, which constitutes an open  
111 source of expert and peer-reviewed phylogenetic knowledge with rich metadata (McTavish et  
112 al., 2015), which allows automatic and reproducible assembly of our chronogram database.  
113 Datelife's chronogram database is navigable as an R data object within the `datelife` R  
114 package.

115 A unique feature of the Phylesystem is that any user can add new published,  
116 state-of-the-art chronograms any time, through their curator application  
117 (<https://tree.opentreeoflife.org/curator>). As chronograms are added to Phylesystem, they are  
118 incorporated into the chronogram database of the `datelife` package. The updated database  
119 is assigned a new version number, followed by a package release on CRAN. `datelife`'s  
120 chronogram database is updated as new chronogram data is added to Phylesystem, at a

121 minimum of once a month and a maximum of every 6 months. Users can also implement  
122 functions from the `datelife` R package to trigger an update of the local chronogram  
123 database, to incorporate any new chronograms to the user's DateLife analysis before an  
124 official database update is released on CRAN.

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

135 *Summarizing Search Results*

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

- 139 1. The matching pruned chronograms as newick strings or "phylo" objects.
- 140 2. The ages of the root of all source chronograms. These ages can correspond to the age  
141 of the most recent common ancestor (mrca) of the user's group of interest if the source  
142 chronograms have all taxa belonging to the group. If not, the root corresponds to the  
143 mrca of a subgroup within the group of interest.
- 144 3. Study citations where original chronograms were published.
- 145 4. A report of input taxon names matches across source chronograms.

- 146 5. The source chronogram(s) with the most input taxon names.

147 6. Various single summary chronograms resulting from summarizing age data, generated

148 using the methodology described next.

## *Choosing a Topology*

150 DateLife requires a tree topology to summarize age data upon. We recommend that  
151 users provide as input a tree topology from the literature, or one of their own making. If no  
152 topology is provided, DateLife automatically extracts one from the OpenTree synthetic tree,  
153 a phylogeny encompassing 2.3 million taxa across all life, assembled from 1, 239 published  
154 phylogenetic trees and OpenTree’s unified Taxonomy, OTT (Open Tree Of Life et al., 2019).  
155 Alternatively, DateLife can combine topologies from source chronograms using a supertree  
156 approach. To combine topologies from source chronograms into a single summary (or  
157 supertree) topology, the DateLife workflow identifies the source chronograms that form a  
158 grove, roughly, a sufficiently overlapping set of taxa between trees, by implementing  
159 definition 2.8 for n-overlap from Ané et al. (2009). In rare cases, a group of trees can have  
160 multiple groves. By default, DateLife chooses the grove with the most taxa, however, the  
161 “criterion = trees” flag allows the user to choose the grove with the most trees instead. If  
162 source chronograms do not form a grove, the supertree reconstruction will fail.

Dating the Topology

164 Input topologies from OpenTree or the supertree approach described above do not  
165 include branch length estimates of any kind. Optionally, to estimate branch lengths  
166 proportional to substitution rates on these topologies, DateLife can mine the Barcode of Life  
167 Data System, BOLD (Ratnasingham & Hebert, 2007) to obtain genetic markers for the  
168 input taxa. These markers are aligned with MUSCLE (Edgar, 2004) (by default) or MAFFT  
169 (Katoh, Asimenos, & Toh, 2009). This alignment can be used to estimate branch lengths on  
170 input topologies that lack branch lengths. Currently, branch length reconstruction in

171 DateLife is performed using parsimony and the likelihood of the phylogenetic tree given a  
172 sequence alignment is computed (Schliep, 2011). While relative branch length information  
173 provides additional data for nodes without secondary date calibrations, topologies without  
174 branch lengths can also be dated.

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

183 For each congruent node, the pairwise distances that traverse that node are  
184 summarized into a single summary matrix using classic summary statistics (i.e., mean,  
185 median, minimum and maximum ages), and the Supermatrix Distance Method [SDM;  
186 Criscuolo, Berry, Douzery, and Gascuel (2006)], which deforms patristic distance matrices by  
187 minimizing variance and then averaging them. These single summary taxon pair age  
188 matrices (Summarized calibrations) can be applied as calibrations to date a tree topology,  
189 using different dating methods currently supported within DateLife: MrBayes (Huelsenbeck  
190 & Ronquist, 2001; Ronquist & Huelsenbeck, 2003), PATHd8 (Britton, Anderson, Jacquet,  
191 Lundqvist, & Bremer, 2007), BLADJ (Webb, Ackerly, & Kembel, 2008; Webb & Donoghue,  
192 2005), and treePL (Smith & O'Meara, 2012).

193 By default, DateLife implements the Branch Length Adjuster (BLADJ) algorithm to  
194 obtain a fully dated topology. BLADJ fixes node ages that have calibration data, and  
195 distributes time between nodes with no data evenly between nodes with calibration data.  
196 This minimizes age variance in the resulting chronogram (Webb et al., 2008). BLADJ does

197 not use branch lengths even when they are present in the input tree or summarizing  
198 topology. When there is conflict in ages between nodes with calibration data, BLADJ  
199 ignores node ages that are older than the age of a parent node. BLADJ requires a root age  
200 estimate. If there is no information on the age of the root in the chronogram database, users  
201 can provide an estimate from the literature. If none is provided, DateLife assigns an  
202 arbitrary age to the root as 10% older than the oldest age available within the group.

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

210 *Visualizing Results*

211 Finally, users can save all source and summary chronograms in formats that permit  
212 reuse and reanalyses (newick and R “phylo” format), as well as visualize and compare results  
213 graphically, or construct their own graphs using DateLife’s chronogram plot generation  
214 functions available from the R package `datelifeplot` (Sanchez-Reyes & O’Meara, 2022).

215

## BENCHMARK

216        `datelife`'s R package code speed was tested on an Apple iMac with one 3.4 GHz Intel  
217    Core i5 processor. We registered variation in computing time of query processing and search  
218    through the database relative to number of queried taxon names. Query processing time  
219    increases roughly linearly with number of input taxon names, and increases considerably if  
220    Taxonomic Name Resolution Service (TNRS) is activated. Up to ten thousand names can be  
221    processed and searched in less than 30 minutes with the most time consuming settings. Once  
222    names have been processed as described in methods, a name search through the chronogram  
223    database can be performed in less than a minute, even with a very large number of taxon  
224    names (Fig. 2).

225        `datelife`'s code performance was evaluated with a set of unit tests designed and  
226    implemented with the R package `testthat` (R Core Team, 2018) that were run both locally  
227    with the `devtools` package (R Core Team, 2018), and on a public server using the continuous  
228    integration tool of GitHub actions (<https://docs.github.com/en/actions>). At present, unit  
229    tests cover more than 40% of `datelife`'s code (<https://codecov.io/gh/phylotastic/datelife>).  
230    Unit testing helps identify potential issues as code is updated or, more critically, as services  
231    code relies upon may change.

232

## CASE STUDIES

233        We illustrate the DateLife workflow using a family within the Passeriform birds  
234    encompassing the true finches, Fringillidae, as case study. On a small example, we analysed 6  
235    bird species, and results from each step of the workflow are shown in Fig. 3. As a second  
236    example, we analysed 289 bird species in the family Fringillidae that are included in the  
237    NCBI taxonomy. The resulting summary chronogram is shown in Fig. 5, and results from  
238    previous steps of the workflow are available as Supplementary Figures.

239

*A Small Example*

**240** ***Creating a search query.***— We chose 6 bird species within the Passeriformes. The

**241** sample includes two species of cardinals: the black-thighed grosbeak – *Pheucticus tibialis*

**242** and the crimson-collared grosbeak – *Rhodothraupis celaeno*; three species of buntings: the

**243** yellowhammer – *Emberiza citrinella*, the pine bunting – *Emberiza leucocephalos* and the

**244** yellow-throated bunting – *Emberiza elegans*; and one species of tanager, the vegetarian finch –

**245** *Platyspiza crassirostris*. Processing of input names found that *Emberiza elegans* is synonym

**246** for *Schoeniclus elegans* in the default reference taxonomy (OTT v3.3, June 1, 2021). For a

**247** detailed discussion on the state of the synonym, refer to Avibase (Avibase, 2022; Lepage,

**248** 2004; Lepage, Vaidya, & Guralnick, 2014). Discovering this synonym allowed assigning five

**249** age data points for the parent node of *Emberiza elegans*, shown as *Schoeniclus elegans* in

**250** figure 3A, which would not have had any data otherwise.

251        ***Searching the database.***— DateLife used the processed input names to search the

252 local chronogram database and found 9 matching chronograms in 6 different studies (Fig.

253 3B). Three studies matched five input names (Barker, Burns, Klicka, Lanyon, & Lovette,

254 2015; Hedges, Marin, Suleski, Paymer, & Kumar, 2015; Jetz, Thomas, Joy, Hartmann, &

255 Mooers, 2012), one study matched four input names (Hooper & Price, 2017) and two studies

256 matched two input names (Barker, Burns, Klicka, Lanyon, & Lovette, 2013; Burns et al.,

257 2014). No studies matched all input names. Together, source chronograms provide 28 unique

258 age data points, covering all nodes on our chosen tree topology to date (Table 1).

259        ***Summarizing search results.***— DateLife obtained OpenTree’s synthetic tree

260 topology for these taxa (Fig. 3C), and congruified and mapped age data to nodes in this

261 chosen topology (Table 1). The name processing step allowed including five data points for

262 node “n4” (parent of *Schoeniclus elegans*; Fig. 3A) that would not have had any data

263 otherwise due to name mismatch. Age summary statistics per node were calculated (Table 2)

264 and used as calibrations to date the tree topology using the BLADJ algorithm. As expected,

265 more inclusive nodes (e.g., node “n1”) have more variance in age data than less inclusive

266 nodes (e.g., node “n5”). Summary age data for node “n2” were excluded as final calibration

267 because they are older than age data of the more inclusive node, “n1” (Fig. 3C4).

268

*An Example with the Family of True Finches*

269

***Creating a query.***— To obtain ages for all species within the family of true finches,

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

277

***Searching the database.***— Next, we used the processed species names in our

278 DateLife query to identify chronograms with at least two Fringillidae species. The DateLife  
279 search identified 13 chronograms containing at least two Fringillidae species, published in 9  
280 different studies (Barker et al., 2013, 2015; Burns et al., 2014; Claramunt & Cracraft, 2015;  
281 Gibb et al., 2015; Hedges et al., 2015; Hooper & Price, 2017; Jetz et al., 2012; Price et al.,  
282 2014). Once identified, DateLife pruned matching chronograms to keep Fringillidae species  
283 names on tips only, and transformed these pruned chronograms to pairwise distance  
284 matrices, revealing 1, 206 different age data points available for species within the  
285 Fringillidae (Supplementray Table S1).

286

***Summarizing search results.***— The final step is to congruify and summarize the

287 age data available for the Fringillidae species into single summary chronograms, using  
288 different types of summary ages, median and SDM. As explained in the “Description”  
289 section, a tree topology to summarize age data upon is required. By default, DateLife uses  
290 the topology from OpenTree’s synthetic tree that contains the species in the search query to  
291 summarize age data upon. According to OpenTree’s synthetic tree, species belonging to the  
292 family Fringillidae do not form a monophyletic group (Fig. 4).

Age data from source chronograms was congrified to OpenTree's topology (Fig. 4B), reducing the age data set to 818 different data points (Supplementray Table S2). For each congruent node, age summary statistics were calculated and used as fixed secondary calibrations over the chosen tree topology, to obtain a fully dated phylogeny with the program BLADJ (Fig. 5).

## CROSS-VALIDATION TEST

We performed a cross validation analysis of the DateLife workflow using the Fringillidae chronograms. We used the individual tree topologies from each of the 19 source chronograms from 13 studies as inputs, treating their node ages as unknown. We then estimated dates for these topologies using the node ages from the chronograms from the other studies as calibrations and smoothing using BLADJ. We found that node ages from original study, and ages estimated using all other age data available are correlated (Fig. 6). For five studies, Datelife tended to underestimate ages for topologically deeper nodes (those with many descendant taxa, aka ‘closer to the root’) relative to the original estimate, and overestimate ages for nodes closer to the tips. Accordingly, root ages are generally older in the original study than estimated using cross-validated ages (Supplementary Fig. S1).

## DISCUSSION

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

315 A total of 99,474 unique terminal taxa are represented in DateLife's database.  
316 Incorporation of more chronograms into the database will continue to improve DateLife's  
317 services. One option to increase the number of chronograms in the DateLife database is the

318 Dryad data repository. Methods to automatically mine chronograms from Dryad could be  
319 designed and implemented. However, Dryad's metadata system has no information to  
320 automatically detect branch length units, and those would still need to be determined  
321 manually by a human curator. We would like to emphasize on the importance of sharing  
322 chronogram data, including systematically curated metadata, into open repositories, such as  
323 OpenTree's Phylesystem (McTavish et al., 2015) for the benefit of the scientific community  
324 as a whole.

325 *Age Variation in Source Chronograms*

326 Conflict in estimated ages among alternative studies is common in the literature. See,  
327 for example, the robust ongoing debate about crown group age of angiosperms  
328 (Barba-Montoya, Reis, Schneider, Donoghue, & Yang, 2018; Magallón, Gómez-Acevedo,  
329 Sánchez-Reyes, & Hernández-Hernández, 2015; Ramshaw et al., 1972; Sanderson & Doyle,  
330 2001; Sauquet, Ramírez-Barahona, & Magallón, 2021). Source chronograms available for the  
331 same organisms have potentially been estimated implementing calibrations very differently.  
332 For example, the chronograms from Burns et al. (2014) were inferred using molecular  
333 substitution rate estimates across birds (Weir & Schluter, 2008), and have much older age  
334 estimates for the same nodes than chronograms that were inferred using fossils as  
335 calibrations (Figs. 5, 6; Supplementary Figs. S1, S5).

336 Different calibration implementations might also imply fundamentally distinct  
337 evolutionary hypotheses (Antonelli et al., 2017). For example, two independent researchers  
338 working on the same clade should both carefully select and justify their choices of fossil  
339 calibration placement. Yet, if one researcher concludes that a fossil should calibrate the  
340 ingroup of a clade, while another researcher concludes that the same fossil should calibrate  
341 the outgroup of the clade, the resulting age estimates will differ, as the placement of  
342 calibrations as stem or crown group has been proven to significantly affect time of lineage  
343 divergence estimates (Sauquet, 2013).

344                   *Primary vs Secondary Calibrations*

345                   While most chronograms in DateLife's database are constructed using primary  
346                   calibrations (molecular substitution rates or ages obtained from the fossil record or geological  
347                   events), DateLife summarizes chronograms using secondary calibrations (ages coming from  
348                   other chronograms). Graur and Martin (2004) cautioned on the increased error and  
349                   uncertainty in estimated ages when using secondary calibrations in dating analyses. Schenk  
350                   (2016) showed that, in simulations, divergence times inferred using secondary calibrations are  
351                   significantly younger than those inferred with primary calibrations, when obtained with  
352                   Bayesian inference methods, and when priors are implemented in similar ways in both  
353                   analyses. Accordingly, the scientific community seems to have more confidence in  
354                   chronograms obtained from a single analysis, using fossil data as primary sources of  
355                   calibrations (Schenk, 2016), and using fossils that have been widely discussed and curated as  
356                   calibrations to date other trees, making sure that all data reflect a coherent evolutionary  
357                   history (Sauquet, 2013), as for example done by Antonelli et al. (2017). There have been  
358                   attempts to create fossil calibration databases (Ksepka et al., 2015), though these still have  
359                   room to grow.

360                   It seems that using several (as opposed to just a few) secondary calibrations can  
361                   provide sufficient information to alleviate or even neutralize potential biases (Sauquet, 2013).  
362                   Certainly, further studies are required to fully understand the effect of secondary calibrations  
363                   on outputs from different tree dating methods, and on downstream analyses. It is possible  
364                   that secondary calibrations can be safely used with dating methods that do not require  
365                   setting priors, such as penalized likelihood (Sanderson, 2003), with methods that do not  
366                   make any assumptions on the ages and fix them to a node on a tree topology, such as  
367                   BLADJ (Webb et al., 2008; Webb & Donoghue, 2005), or methods that summarize age data  
368                   unto a tree topology.

369                   Our cross validation analysis might provide some insight in this regard. When ages are

370 estimated with secondary calibrations, nodes closer to the root do tend to be slightly younger  
371 than ages estimated with primary calibrations. However, nodes closer to the tip tend to be  
372 older when estimated using secondary calibrations with a dating method that does not make  
373 any prior assumptions on the nature of the calibrations themselves (Supplementary Figures  
374 S2-S20). The only exception to this was observed on results of the cross validation analysis of  
375 the Burns et al. (2014) chronogram, which results in much younger node ages when  
376 estimated using secondary calibrations (Supplementary Figs. S1, S5).

377 *Summarizing Chronograms*

378 By default, DateLife currently summarizes all source chronograms that overlap with at  
379 least two species names. Users can exclude source chronograms if they have reasons to do so.  
380 Strictly speaking, a good chronogram should reflect the real time of lineage divergence  
381 accurately and precisely. To our knowledge, there are no tested measures to determine  
382 independently when a chronogram is better than another. Yet, several characteristics of the  
383 data used for dating analyses, as well as from the output chronogram itself, could be used to  
384 score the quality of source chronograms.

385 Some measures that have been proposed are the proportion of lineage sampling and the  
386 number of calibrations used (Magallón, 2010; Magallón et al., 2015). Some characteristics  
387 that are often cited in published studies as a measure of improved age estimates as compared  
388 to previously published estimates are: quality of alignment (missing data, GC content),  
389 lineage sampling (strategy and proportion), phylogenetic and dating inference method,  
390 number of fossils used as calibrations, support for nodes and ages, and magnitude of  
391 confidence intervals.

392 DateLife provides an opportunity to capture concordance and conflict among date  
393 estimates, which can also be used as a metric for chronogram reliability. Its open database of  
394 chronograms allows other researchers to do such analyses themselves reproducibly, and

395 without needing permission. Though, of course, they should follow proper citation practices,  
396 especially for the source chronogram studies.

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

406 *Effects on Downstream Analyses*

407 For downstream analyses, using alternative chronogram may deeply affect our  
408 inferences (Title & Rabosky, 2016), particularly when studying phenomena dependent on the  
409 timing of species diversification events, such as macroevolutionary processes.

410 In ecology and conservation biology, incorporating at least some data on lineage  
411 divergence times represents a relevant improvement for testing alternative hypothesis using  
412 phylogenetic distance (Webb et al., 2008). Hence, DateLife's workflow features different ways  
413 of estimating node ages in the absence of calibrations and branch length information for  
414 certain taxa.

415 Making up branch lengths is a common practice in scientific publications: Jetz et al.  
416 (2012), created a chronogram of all 9, 993 bird species, where 67% had molecular data and  
417 the rest was simulated; Rabosky et al. (2018) created a chronogram of 31, 536 ray-finned  
418 fishes, of which only 37% had molecular data; Smith and Brown (2018) constructed a  
419 chronogram of 353, 185 seed plants where only 23% had molecular data.

420 Notably, risks come with this practice. Taken to the extreme, one could make a fully  
421 resolved, calibrated tree of all modern and extinct taxa using a single taxonomy and a single  
422 calibration, using polytomy resolution and branch estimation methods. There has yet to be  
423 a thorough analysis of what can go wrong when one extends inferences beyond the data in  
424 this way, so we urge caution; we also urge readers to follow the example of the large tree  
425 papers cited above, by carefully considering the statistical assumptions being made, and  
426 assessing the consistency of the results with prior work.

427 CONCLUSIONS

428 Knowledge of the evolutionary time frame of organisms is key to many research areas:  
429 trait evolution, species diversification, biogeography, macroecology and more. It is also  
430 crucial for education, science communication and policy, but generating chronograms is  
431 difficult, especially for those who want to use phylogenies but who are not systematists, or  
432 do not have the time to acquire and develop the necessary knowledge and skills to construct  
433 them on their own. Importantly, years of primarily public funded research have resulted in  
434 vast amounts of chronograms that are already available on scientific publications, but hidden  
435 to the public and scientific community for reuse.

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

446

## AVAILABILITY

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

457

## SUPPLEMENTARY MATERIAL

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

461

## FUNDING

462 Funding was provided by the US National Science Foundation (NSF) grants  
463 ABI-1458603 to the Datelife project; DBI-0905606 to the National Evolutionary Synthesis  
464 Center (NESCen), ABI-1458572 to the Phylotastic project, and ABI-1759846 to the Open  
465 Tree of Life project.

466

## ACKNOWLEDGEMENTS

467 The DateLife project was born as a prototype tool aiming to provide these services,  
468 and was initially developed over a series of hackathons at the National Evolutionary  
469 Synthesis Center, NC, USA (Stoltzfus et al., 2013). We thank colleagues from the O'Meara

<sup>470</sup> Lab at the University of Tennessee Knoxville for suggestions, discussions and software testing.

<sup>471</sup> The late National Evolutionary Synthesis Center (NESCent), which sponsored hackathons

<sup>472</sup> that led to initial work on this project. The team that assembled DateLife's first proof of

<sup>473</sup> concept: Tracy Heath, Jonathan Eastman, Peter Midford, Joseph Brown, Matt Pennell,

<sup>474</sup> Mike Alfaro, and Luke Harmon. The Open Tree of Life project that provides the open,

<sup>475</sup> metadata rich repository of trees used to construct DateLife's chronogram database. The

<sup>476</sup> many scientists who publish their chronograms in an open, reusable form, and the scientists

<sup>477</sup> who curate them for deposition in the Open Tree of Life repository. The NSF for funding

<sup>478</sup> nearly all the above, in addition to the ABI grant that funded this project itself.

**References**

- 479
- 480 Alström, P., Hooper, D. M., Liu, Y., Olsson, U., Mohan, D., Gelang, M., ... Price, T.  
481 D. (2014). Discovery of a relict lineage and monotypic family of passerine birds.  
482 *Biology Letters*, 10(3), 20131067.
- 483 Ané, C., Eulenstein, O., Piaggio-Talice, R., & Sanderson, M. J. (2009). Groves of  
484 phylogenetic trees. *Annals of Combinatorics*, 13(2), 139–167.
- 485 Antonelli, A., Hettling, H., Condamine, F. L., Vos, K., Nilsson, R. H., Sanderson, M.  
486 J., ... Vos, R. A. (2017). Toward a self-updating platform for estimating rates of  
487 speciation and migration, ages, and relationships of Taxa. *Systematic Biology*,  
488 66(2), 153–166. <https://doi.org/10.1093/sysbio/syw066>
- 489 Archie, J., Day, W. H., Felsenstein, J., Maddison, W., Meacham, C., Rohlf, F. J., &  
490 Swofford, D. (1986). The Newick tree format. Retrieved from  
491 %7B<https://evolution.genetics.washington.edu/phylip/newicktree.html>%7D
- 492 Avibase. (2022). Yellow-throated Bunting. *Avibase - The World Bird Database*,  
493 (Online Resource). Retrieved from %7B<https://avibase.bsc-eoc.org/species.jsp?lang=EN&avibaseid=82D1EE0049D8D927%7D>
- 494 Bapst, D. W. (2012). Paleotree: An R package for paleontological and phylogenetic  
495 analyses of evolution. *Methods in Ecology and Evolution*, 3(5), 803–807.  
496 <https://doi.org/10.1111/j.2041-210X.2012.00223.x>
- 497 Barba-Montoya, J., Reis, M. dos, Schneider, H., Donoghue, P. C., & Yang, Z. (2018).  
498 Constraining uncertainty in the timescale of angiosperm evolution and the  
499 veracity of a cretaceous terrestrial revolution. *New Phytologist*, 218(2), 819–834.
- 500 Barker, F. K. (2014). Mitogenomic data resolve basal relationships among passeriform  
501 and passeridan birds. *Molecular Phylogenetics and Evolution*, 79, 313–324.
- 502 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2013). Going  
503 to extremes: Contrasting rates of diversification in a recent radiation of new world  
504 passerine birds. *Systematic Biology*, 62(2), 298–320.
- 505

- 506 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2015). New  
507 insights into new world biogeography: An integrated view from the phylogeny of  
508 blackbirds, cardinals, sparrows, tanagers, warblers, and allies. *The Auk: Ornithological Advances*, 132(2), 333–348.
- 510 Barker, F. K., Cibois, A., Schikler, P., Feinstein, J., & Cracraft, J. (2004). Phylogeny  
511 and diversification of the largest avian radiation. *Proceedings of the National  
512 Academy of Sciences*, 101(30), 11040–11045.
- 513 Beresford, P., Barker, F., Ryan, Pg., & Crowe, T. (2005). African endemics span the  
514 tree of songbirds (passeri): Molecular systematics of several evolutionary  
515 “enigmas.” *Proceedings of the Royal Society B: Biological Sciences*, 272(1565),  
516 849–858.
- 517 Boyle, B., Hopkins, N., Lu, Z., Raygoza Garay, J. A., Mozzherin, D., Rees, T., ...  
518 Enquist, B. J. (2013). The taxonomic name resolution service: An online tool for  
519 automated standardization of plant names. *BMC Bioinformatics*, 14(1).  
520 <https://doi.org/10.1186/1471-2105-14-16>
- 521 Britton, T., Anderson, C. L., Jacquet, D., Lundqvist, S., & Bremer, K. (2007).  
522 Estimating Divergence Times in Large Phylogenetic Trees. *Systematic Biology*,  
523 56(788777878), 741–752. <https://doi.org/10.1080/10635150701613783>
- 524 Bryson Jr, R. W., Chaves, J., Smith, B. T., Miller, M. J., Winker, K., Pérez-Emán, J.  
525 L., & Klicka, J. (2014). Diversification across the new world within the  
526 ‘blue’cardinalids (aves: cardinalidae). *Journal of Biogeography*, 41(3), 587–599.
- 527 Burleigh, J. G., Kimball, R. T., & Braun, E. L. (2015). Building the avian tree of life  
528 using a large-scale, sparse supermatrix. *Molecular Phylogenetics and Evolution*,  
529 84, 53–63.
- 530 Burns, K. J., Shultz, A. J., Title, P. O., Mason, N. A., Barker, F. K., Klicka, J., ...  
531 Lovette, I. J. (2014). Phylogenetics and diversification of tanagers (passeriformes:  
532 Thraupidae), the largest radiation of neotropical songbirds. *Molecular*

- 533                   *Phylogenetics and Evolution*, 75, 41–77.
- 534                   Chamberlain, S. A., & Szöcs, E. (2013). taxize : taxonomic search and retrieval in R  
535                   [version 2; referees: 3 approved]. *F1000Research*, 2(191), 1–29.  
536                   <https://doi.org/10.12688/f1000research.2-191.v2>
- 537                   Chamberlain, S. A., Szöcs, E., Foster, Z., Arendsee, Z., Boettiger, C., Ram, K., ...  
538                   Li, G. (2019). *taxize: Taxonomic information from around the web*. Retrieved  
539                   from <https://github.com/ropensci/taxize>
- 540                   Chaves, J. A., Hidalgo, J. R., & Klicka, J. (2013). Biogeography and evolutionary  
541                   history of the n eotropical genus s altator (a ves: T hraupini). *Journal of  
542                   Biogeography*, 40(11), 2180–2190.
- 543                   Claramunt, S., & Cracraft, J. (2015). A new time tree reveals earth history's imprint  
544                   on the evolution of modern birds. *Science Advances*, 1(11), e1501005.
- 545                   Criscuolo, A., Berry, V., Douzery, E. J. P., & Gascuel, O. (2006). SDM: A fast  
546                   distance-based approach for (super)tree building in phylogenomics. *Systematic  
547                   Biology*, 55(5), 740–755. <https://doi.org/10.1080/10635150600969872>
- 548                   Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M.-K., Turon, X., ...  
549                   Douzery, E. J. (2018). A phylogenomic framework and timescale for comparative  
550                   studies of tunicates. *BMC Biology*, 16(1), 1–14.
- 551                   Eastman, J. M., Harmon, L. J., & Tank, D. C. (2013). Congruification: Support for  
552                   time scaling large phylogenetic trees. *Methods in Ecology and Evolution*, 4(7),  
553                   688–691. <https://doi.org/10.1111/2041-210X.12051>
- 554                   Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and  
555                   high throughput. *Nucleic Acids Research*, 32(5), 1792–1797.
- 556                   Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American  
557                   Naturalist*, 125(1), 1–15. Retrieved from <http://www.jstor.org/stable/2461605>
- 558                   GBIF Secretariat. (2022). GBIF Backbone Taxonomy. *Checklist Dataset*, (Online  
559                   Resource accessed via GBIF.org). Retrieved from

- 560 {<https://doi.org/10.15468/39omei> }
- 561 Gibb, G. C., England, R., Hartig, G., McLenachan, P. A., Taylor Smith, B. L.,  
562 McComish, B. J., ... Penny, D. (2015). New zealand passerines help clarify the  
563 diversification of major songbird lineages during the oligocene. *Genome Biology  
564 and Evolution*, 7(11), 2983–2995.
- 565 Graur, D., & Martin, W. (2004). Reading the entrails of chickens: Molecular  
566 timescales of evolution and the illusion of precision. *TRENDS in Genetics*, 20(2),  
567 80–86.
- 568 Hackett, S. J., Kimball, R. T., Reddy, S., Bowie, R. C., Braun, E. L., Braun, M. J., et  
569 al.others. (2008). A phylogenomic study of birds reveals their evolutionary history.  
570 *Science*, 320(5884), 1763–1768.
- 571 Hedges, S. B., Marin, J., Suleski, M., Paymer, M., & Kumar, S. (2015). Tree of life  
572 reveals clock-like speciation and diversification. *Molecular Biology and Evolution*,  
573 32(4), 835–845. <https://doi.org/10.1093/molbev/msv037>
- 574 Heibl, C. (2008). *PHYLOCH: R language tree plotting tools and interfaces to diverse  
575 phylogenetic software packages*. Retrieved from  
576 <http://www.christophheibl.de/Rpackages.html>
- 577 Hooper, D. M., & Price, T. D. (2017). Chromosomal inversion differences correlate  
578 with range overlap in passerine birds. *Nature Ecology & Evolution*, 1(10), 1526.
- 579 Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of  
580 phylogenetic trees. *Bioinformatics*, 17(8), 754–755.  
581 <https://doi.org/10.1093/bioinformatics/17.8.754>
- 582 Jetz, W., Thomas, G., Joy, J. J. B., Hartmann, K., & Mooers, A. (2012). The global  
583 diversity of birds in space and time. *Nature*, 491(7424), 444–448.  
584 <https://doi.org/10.1038/nature11631>
- 585 Johansson, U. S., Fjeldså, J., & Bowie, R. C. (2008). Phylogenetic relationships  
586 within passerida (aves: Passeriformes): A review and a new molecular phylogeny

- 587 based on three nuclear intron markers. *Molecular Phylogenetics and Evolution*,  
588 48(3), 858–876.
- 589 Katoh, K., Asimenos, G., & Toh, H. (2009). Multiple alignment of DNA sequences  
590 with MAFFT. In *Bioinformatics for DNA sequence analysis* (pp. 39–64).  
591 Springer.
- 592 Kimball, R. T., Oliveros, C. H., Wang, N., White, N. D., Barker, F. K., Field, D. J.,  
593 et al.others. (2019). A phylogenomic supertree of birds. *Diversity*, 11(7), 109.
- 594 Klicka, J., Barker, F. K., Burns, K. J., Lanyon, S. M., Lovette, I. J., Chaves, J. A., &  
595 Bryson Jr, R. W. (2014). A comprehensive multilocus assessment of sparrow (aves:  
596 Passerellidae) relationships. *Molecular Phylogenetics and Evolution*, 77, 177–182.
- 597 Ksepka, D. T., Parham, J. F., Allman, J. F., Benton, M. J., Carrano, M. T.,  
598 Cranston, K. A., et al.others. (2015). The fossil calibration database—a new  
599 resource for divergence dating. *Systematic Biology*, 64(5), 853–859.
- 600 Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M.,  
601 Martinez-Barrio, A., et al.others. (2015). Evolution of darwin’s finches and their  
602 beaks revealed by genome sequencing. *Nature*, 518(7539), 371–375.
- 603 Laubichler, M. D., & Maienschein, J. (2009). *Form and function in developmental  
604 evolution*. Cambridge University Press.
- 605 Lepage, D. (2004). *Avibase: The world bird database*. Bird Studies Canada.
- 606 Lepage, D., Vaidya, G., & Guralnick, R. (2014). Avibase—a database system for  
607 managing and organizing taxonomic concepts. *ZooKeys*, (420), 117.
- 608 Lerner, H. R., Meyer, M., James, H. F., Hofreiter, M., & Fleischer, R. C. (2011).  
609 Multilocus resolution of phylogeny and timescale in the extant adaptive radiation  
610 of hawaiian honeycreepers. *Current Biology*, 21(21), 1838–1844.
- 611 Lovette, I. J., Pérez-Emán, J. L., Sullivan, J. P., Banks, R. C., Fiorentino, I.,  
612 Córdoba-Córdoba, S., et al.others. (2010). A comprehensive multilocus phylogeny  
613 for the wood-warblers and a revised classification of the parulidae (aves).

- 614                  *Molecular Phylogenetics and Evolution*, 57(2), 753–770.
- 615                  Magallon, S., & Sanderson, M. (2001). Absolute diversification rates in angiosperm  
616                  clades. *Evolution*, 55(9), 1762–1780.
- 617                  Magallón, S. (2010). Using fossils to break long branches in molecular dating: A  
618                  comparison of relaxed clocks applied to the origin of angiosperms. *Systematic  
619                  Biology*, 59(4), 384–399.
- 620                  Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., & Hernández-Hernández, T.  
621                  (2015). A metacalibrated time-tree documents the early rise of flowering plant  
622                  phylogenetic diversity. *New Phytologist*, 207(2), 437–453.
- 623                  McTavish, E. J., Hinchliff, C. E., Allman, J. F., Brown, J. W., Cranston, K. A.,  
624                  Holder, M. T., . . . Smith, S. (2015). Phylesystem: A git-based data store for  
625                  community-curated phylogenetic estimates. *Bioinformatics*, 31(17), 2794–2800.
- 626                  Michonneau, F., Brown, J. W., & Winter, D. J. (2016). rotl: an R package to interact  
627                  with the Open Tree of Life data. *Methods in Ecology and Evolution*, 7(12),  
628                  1476–1481. <https://doi.org/10.1111/2041-210X.12593>
- 629                  Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecology  
630                  Letters*, 17(4), 508–525. <https://doi.org/10.1111/ele.12251>
- 631                  Moyle, R. G., Oliveros, C. H., Andersen, M. J., Hosner, P. A., Benz, B. W., Manthey,  
632                  J. D., . . . Faircloth, B. C. (2016). Tectonic collision and uplift of wallacea  
633                  triggered the global songbird radiation. *Nature Communications*, 7(1), 1–7.
- 634                  Ödeen, A., Håstad, O., & Alström, P. (2011). Evolution of ultraviolet vision in the  
635                  largest avian radiation—the passerines. *BMC Evolutionary Biology*, 11(1), 1–8.
- 636                  Oliveros, C. H., Field, D. J., Ksepka, D. T., Barker, F. K., Aleixo, A., Andersen, M.  
637                  J., et al.others. (2019). Earth history and the passerine superradiation.  
638                  *Proceedings of the National Academy of Sciences*, 116(16), 7916–7925.
- 639                  Ooms, J., & Chamberlain, S. (2018). *Phylocomr: Interface to 'phylocom'*. Retrieved  
640                  from <https://CRAN.R-project.org/package=phylocomr>

- 641 Open Tree Of Life, Redelings, B., Cranston, K. A., Allman, J., Holder, M. T., &  
642 McTavish, E. J. (2016). Open Tree of Life APIs v3.0. *Open Tree of Life Project*,  
643 (Online Resources). Retrieved from  
644 %7Bhttps://github.com/OpenTreeOfLife/germinator/wiki/Open-Tree-of-Life-  
645 Web-APIs%7D
- 646 Open Tree Of Life, Redelings, B., Sánchez Reyes, L. L., Cranston, K. A., Allman, J.,  
647 Holder, M. T., & McTavish, E. J. (2019). Open tree of life synthetic tree v12.3.  
648 *Zenodo*. Retrieved from <https://doi.org/10.5281/zenodo.3937742>
- 649 Päckert, M., Martens, J., Sun, Y.-H., Severinghaus, L. L., Nazarenko, A. A., Ting, J.,  
650 ... Tietze, D. T. (2012). Horizontal and elevational phylogeographic patterns of  
651 himalayan and southeast asian forest passerines (aves: passeriformes). *Journal of*  
652 *Biogeography*, 39(3), 556–573.
- 653 Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and  
654 evolution in R language. *Bioinformatics*, 20(2), 289–290.
- 655 Parchman, T. L., Benkman, C. W., & Mezquida, E. T. (2007). Coevolution between  
656 hispaniolan crossbills and pine: Does more time allow for greater phenotypic  
657 escalation at lower latitude? *Evolution*, 61(9), 2142–2153.
- 658 Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., FitzJohn,  
659 R. G., ... Harmon, L. J. (2014). Geiger v2. 0: An expanded suite of methods for  
660 fitting macroevolutionary models to phylogenetic trees. *Bioinformatics*, 30(15),  
661 2216–2218.
- 662 Posadas, P., Crisci, J. V., & Katinas, L. (2006). Historical biogeography: A review of  
663 its basic concepts and critical issues. *Journal of Arid Environments*, 66(3),  
664 389–403.
- 665 Powell, A. F., Barker, F. K., Lanyon, S. M., Burns, K. J., Klicka, J., & Lovette, I. J.  
666 (2014). A comprehensive species-level molecular phylogeny of the new world  
667 blackbirds (icteridae). *Molecular Phylogenetics and Evolution*, 71, 94–112.

- 668 Price, T. D., Hooper, D. M., Buchanan, C. D., Johansson, U. S., Tietze, D. T.,  
669 Alström, P., et al.others. (2014). Niche filling slows the diversification of  
670 himalayan songbirds. *Nature*, 509(7499), 222.
- 671 Pulgarín-R, P. C., Smith, B. T., Bryson Jr, R. W., Spellman, G. M., & Klicka, J.  
672 (2013). Multilocus phylogeny and biogeography of the new world pheucticus  
673 grosbeaks (aves: cardinalidae). *Molecular Phylogenetics and Evolution*, 69(3),  
674 1222–1227.
- 675 R Core Team. (2018). *R: a language and environment for statistical computing*.  
676 Vienna, Austria: R Foundation for Statistical Computing.
- 677 Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., et  
678 al.others. (2018). An inverse latitudinal gradient in speciation rate for marine  
679 fishes. *Nature*, 559(7714), 392.
- 680 Ramshaw, J., Richardson, D., Meatyard, B., Brown, R., Richardson, M., Thompson,  
681 E., & Boulter, D. (1972). The time of origin of the flowering plants determined by  
682 using amino acid sequence data of cytochrome c. *New Phytologist*, 71(5), 773–779.
- 683 Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The barcode of life data system  
684 (<http://www.Barcodinglife.org>). *Molecular Ecology Notes*, 7(3), 355–364.
- 685 Rees, & Cranston, K. (2017). Automated assembly of a reference taxonomy for  
686 phylogenetic data synthesis. *Biodiversity Data Journal*, (5).
- 687 Rees, Vandepitte, L., Decock, W., & Vanhoorne, B. (2017). IRMNG 2006–2016: 10  
688 Years of a Global Taxonomic Database. *Biodiversity Informatics*, 12.
- 689 Revell, L. J. (2012). Phytools: An r package for phylogenetic comparative biology  
690 (and other things). *Methods in Ecology and Evolution*, 3, 217–223.
- 691 Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic  
692 inference under mixed models. *Bioinformatics*, 19(12), 1572–1574.  
693 <https://doi.org/10.1093/bioinformatics/btg180>
- 694 Sanchez-Reyes, L. L., & O'Meara, B. (2022). **datelifeplot**: Methods to plot

695                   chronograms and outputs of the `datelife` package. *R Package Release V0.2.2.*  
696                   Retrieved from <https://zenodo.org/badge/latestdoi/381501451>

697                   Sanchez-Reyes, L. L., O'Meara, B., Eastman, J., Heath, T., Wright, A., Schliep, K.,  
698                   ... Alfaro, M. (2022). `datelife`: Scientific Data on Time of Lineage Divergence  
699                   for Your Taxa. In *R package version 0.6.6*. Retrieved from  
700                   <https://CRAN.R-project.org/package=datelife> and  
701                   <https://doi.org/10.5281/zenodo.593938>

702                   Sanderson, M. J. (2002). Estimating Absolute Rates of Molecular Evolution and  
703                   Divergence Times: A Penalized Likelihood Approach. *Molecular Biology and*  
704                   *Evolution*, *19*(1), 101–109.  
705                   <https://doi.org/10.1093/oxfordjournals.molbev.a003974>

706                   Sanderson, M. J. (2003). r8s: Inferring Absolute Rates of Molecular Evolution and  
707                   Divergence Times in the Absence of a Molecular Clock. *Bioinformatics*, *19*(2),  
708                   301–302.

709                   Sanderson, M. J., & Doyle, J. A. (2001). Sources of error and confidence intervals in  
710                   estimating the age of angiosperms from rbcL and 18S rDNA data. *American*  
711                   *Journal of Botany*, *88*(8), 1499–1516.

712                   Sauquet, H. (2013). A practical guide to molecular dating. *Comptes Rendus Palevol*,  
713                   *12*(6), 355–367.

714                   Sauquet, H., Ramírez-Barahona, S., & Magallón, S. (2021). *The age of flowering*  
715                   *plants is unknown*.

716                   Schenk, J. J. (2016). Consequences of secondary calibrations on divergence time  
717                   estimates. *PLoS ONE*, *11*(1). <https://doi.org/10.1371/journal.pone.0148228>

718                   Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in r. *Bioinformatics*, *27*(4),  
719                   592–593.

720                   Schoch, C. L., Ciufo, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R.,  
721                   et al.others. (2020). NCBI Taxonomy: a Comprehensive Update on Curation,

- 722 Resources and Tools. *Database*, 2020.
- 723 Selvatti, A. P., Gonzaga, L. P., & Moraes Russo, C. A. de. (2015). A paleogene origin  
724 for crown passerines and the diversification of the oscines in the new world.  
725 *Molecular Phylogenetics and Evolution*, 88, 1–15.
- 726 Smith, S., & Brown, J. (2018). Constructing a broadly inclusive seed plant phylogeny.  
727 *American Journal of Botany*, 105(3), 302–314.
- 728 Smith, S., & O'Meara, B. (2012). TreePL: Divergence time estimation using  
729 penalized likelihood for large phylogenies. *Bioinformatics*, 28(20), 2689–2690.  
730 <https://doi.org/10.1093/bioinformatics/bts492>
- 731 Stoltzfus, A., Lapp, H., Matasci, N., Deus, H., Sidlauskas, B., Zmasek, C. M., ...  
732 Jordan, G. (2013). Phylotastic! Making tree-of-life knowledge accessible, reusable  
733 and convenient. *BMC Bioinformatics*, 14.  
734 <https://doi.org/10.1186/1471-2105-14-158>
- 735 Tietze, D. T., Päckert, M., Martens, J., Lehmann, H., & Sun, Y.-H. (2013). Complete  
736 phylogeny and historical biogeography of true rosefinches (aves: carpodacus).  
737 *Zoological Journal of the Linnean Society*, 169(1), 215–234.
- 738 Title, P. O., & Rabosky, D. L. (2016). Do Macrophylogenies Yield Stable  
739 Macroevolutionary Inferences? An Example from Squamate Reptiles. *Systematic  
740 Biology*, syw102. <https://doi.org/10.1093/sysbio/syw102>
- 741 Treplin, S., Siegert, R., Bleidorn, C., Thompson, H. S., Fotso, R., & Tiedemann, R.  
742 (2008). Molecular phylogeny of songbirds (aves: Passeriformes) and the relative  
743 utility of common nuclear marker loci. *Cladistics*, 24(3), 328–349.
- 744 Vos, R. A., Balhoff, J. P., Caravas, J. A., Holder, M. T., Lapp, H., Maddison, W. P.,  
745 et al.others. (2012). NeXML: Rich, extensible, and verifiable representation of  
746 comparative data and metadata. *Systematic Biology*, 61(4), 675–689.
- 747 Webb, C. (2000). Exploring the Phylogenetic Structure of Ecological Communities :  
748 An Example for Rain Forest Trees. *The American Naturalist*, 156(2), 145–155.

- 749 Webb, C., Ackerly, D., & Kembel, S. (2008). Phylocom: Software for the analysis of  
750 phylogenetic community structure and trait evolution. *Bioinformatics*, 24(18),  
751 2098–2100. <https://doi.org/10.1093/bioinformatics/btn358>
- 752 Webb, C., & Donoghue, M. (2005). Phylomatic: Tree assembly for applied  
753 phylogenetics. *Molecular Ecology Notes*, 5(1), 181–183.
- 754 Weir, J., & Schluter, D. (2008). Calibrating the avian molecular clock. *Molecular  
755 Ecology*, 17(10), 2321–2328.
- 756 Zuccon, D., Prŷs-Jones, R., Rasmussen, P. C., & Ericson, P. G. (2012). The  
757 phylogenetic relationships and generic limits of finches (fringillidae). *Molecular  
758 Phylogenetics and Evolution*, 62(2), 581–596.

759 TABLE 1. Ages of congruified nodes. See Figure 3, step C2.

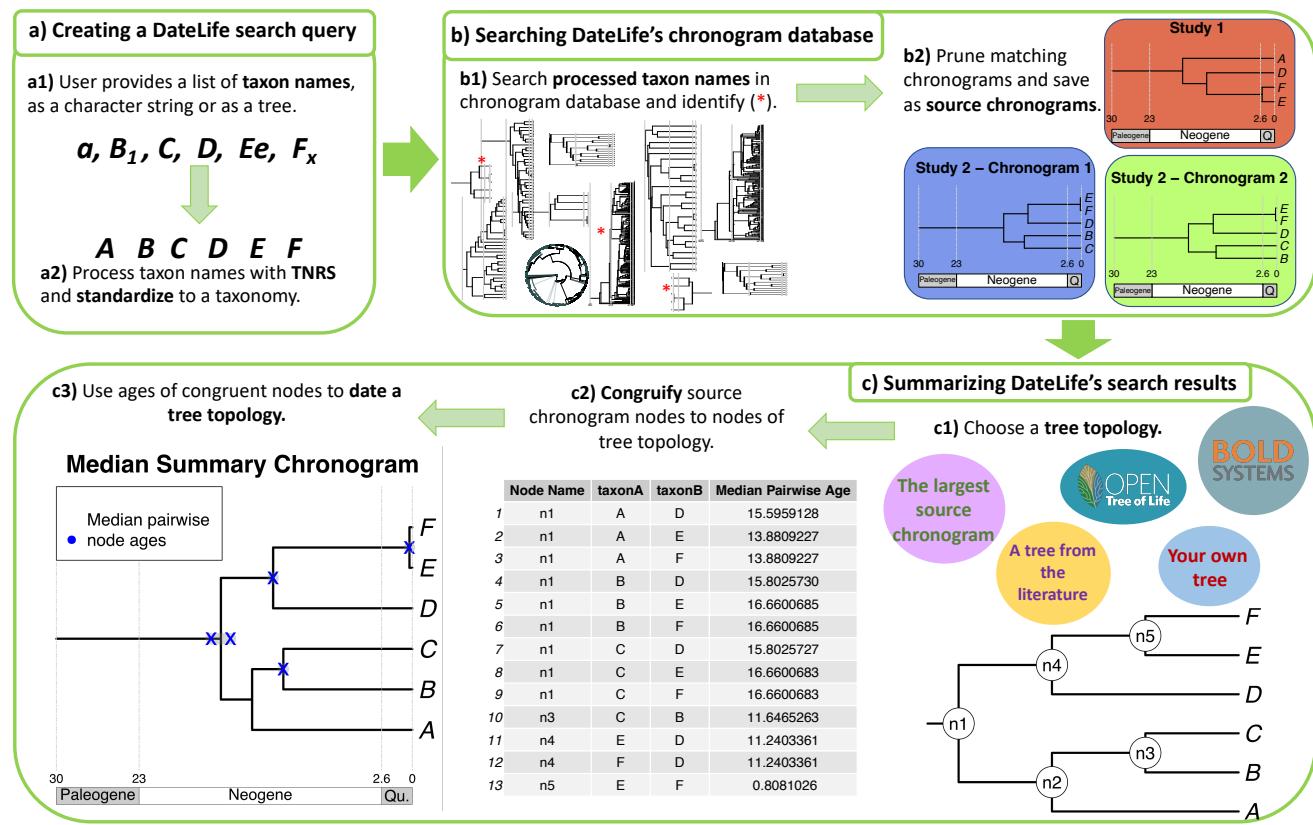
	<b>Node Name</b>	<b>taxon A</b>	<b>taxon B</b>	<b>Node Age</b>	<b>Study chronogram</b>
1	n1	<i>Emberiza citrinella</i>	<i>Platyspiza crassirostris</i>	9.6509881	Jetz et al. 2012 – chronogram 1
2	n1	<i>Pheucticus tibialis</i>	<i>Schoeniclus elegans</i>	14.3336520	Barker et al. 2015 – chronogram 1
3	n1	<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

760

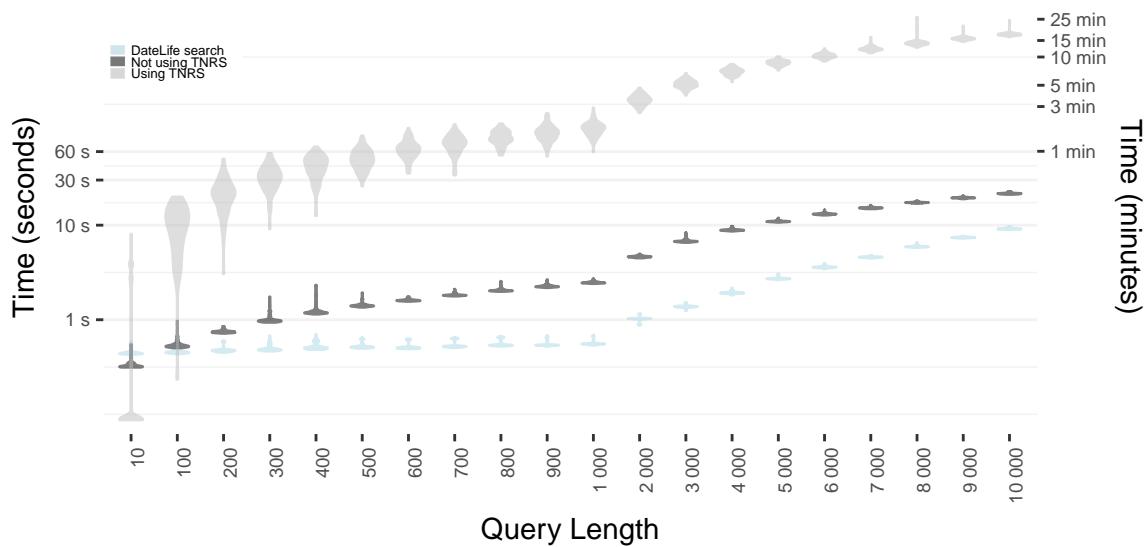
761 TABLE 2. Summary of congruified nodes ages. See Figure 3, step C3.

<b>Node Name</b>	<b>Min Age</b>	<b>Q1</b>	<b>Median Age</b>	<b>Mean Age</b>	<b>Q3</b>	<b>Max Age</b>	<b>Variance</b>	<b>SD</b>
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

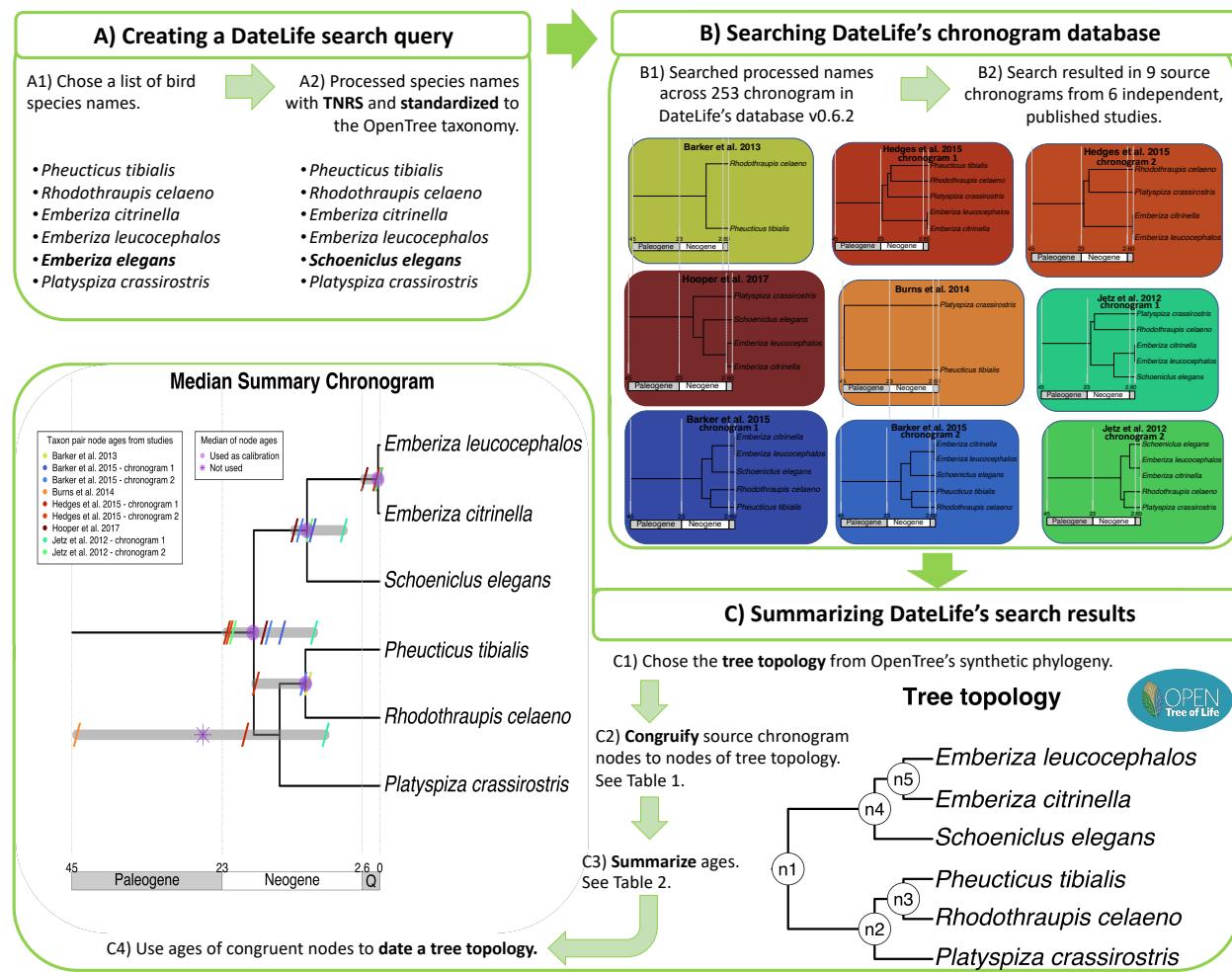
762



763 FIGURE 1. Main DateLife workflow. Analyses can be performed via DateLife's interactive  
 764 website at [www.datelife.org](http://www.datelife.org), or using the *datelife* R package. Details on the R functions used  
 765 to perform the analyses are available from *datelife*'s R package vignettes at  
 766 <https://phylotastic.org/datelife>.



769 FIGURE 2. *DateLife's* benchmarking results. Computation time used to process a query and  
 770 a search across *datelife's* chronogram database, relative to number of input taxon names.  
 771 For each  $N = \{10, 100, 200, \dots, 1\,000, \dots, 9\,000, 10\,000\}$ , we sampled  $N$  species names  
 772 from the class Aves a hundred times, and then performed a *datelife* search processing the  
 773 input names with Taxon Names Resolution Service (TNRS; light gray), and without  
 774 processing names (dark gray). For comparison, we performed a search using an input that  
 775 had been pre-processed with TNRS (light blue).



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

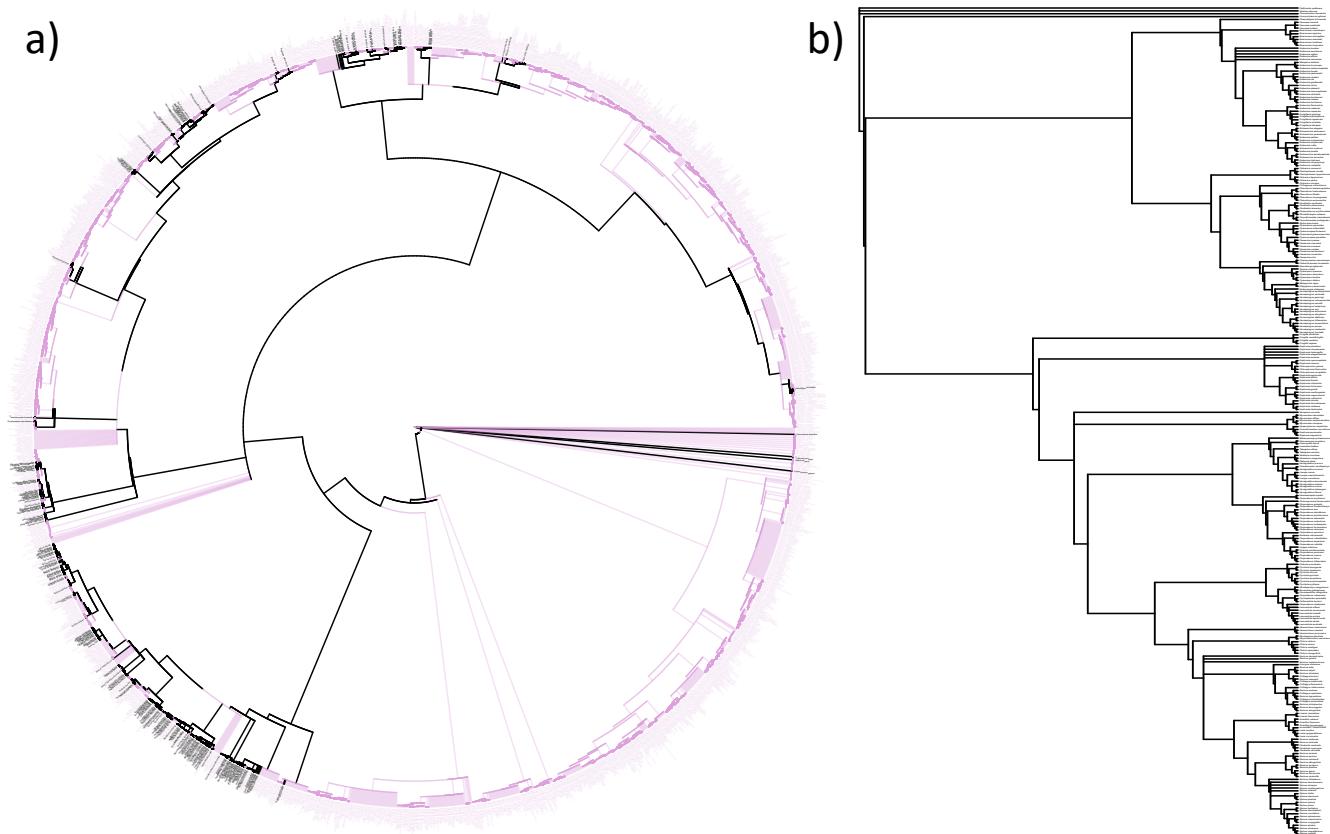


FIGURE 4. Tree topologies extracted from Open Tree of Life's (OpenTree) synthetic phylogenetic tree.  
 A) Topology of 2,333 tips and 1,305 internal nodes, encompassing bird species within the family Fringillidae following the NCBI taxonomy (black), as well as all other bird species that share the same Most Recent Common Ancestor (MRCA) node in OpenTree's synthetic tree (purple). B) Subtree topology of 289 tips and 253 internal nodes, resulting from pruning species that do not belong to the family Fringillidae according to the NCBI taxonomy (purple branches in topology A). 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).

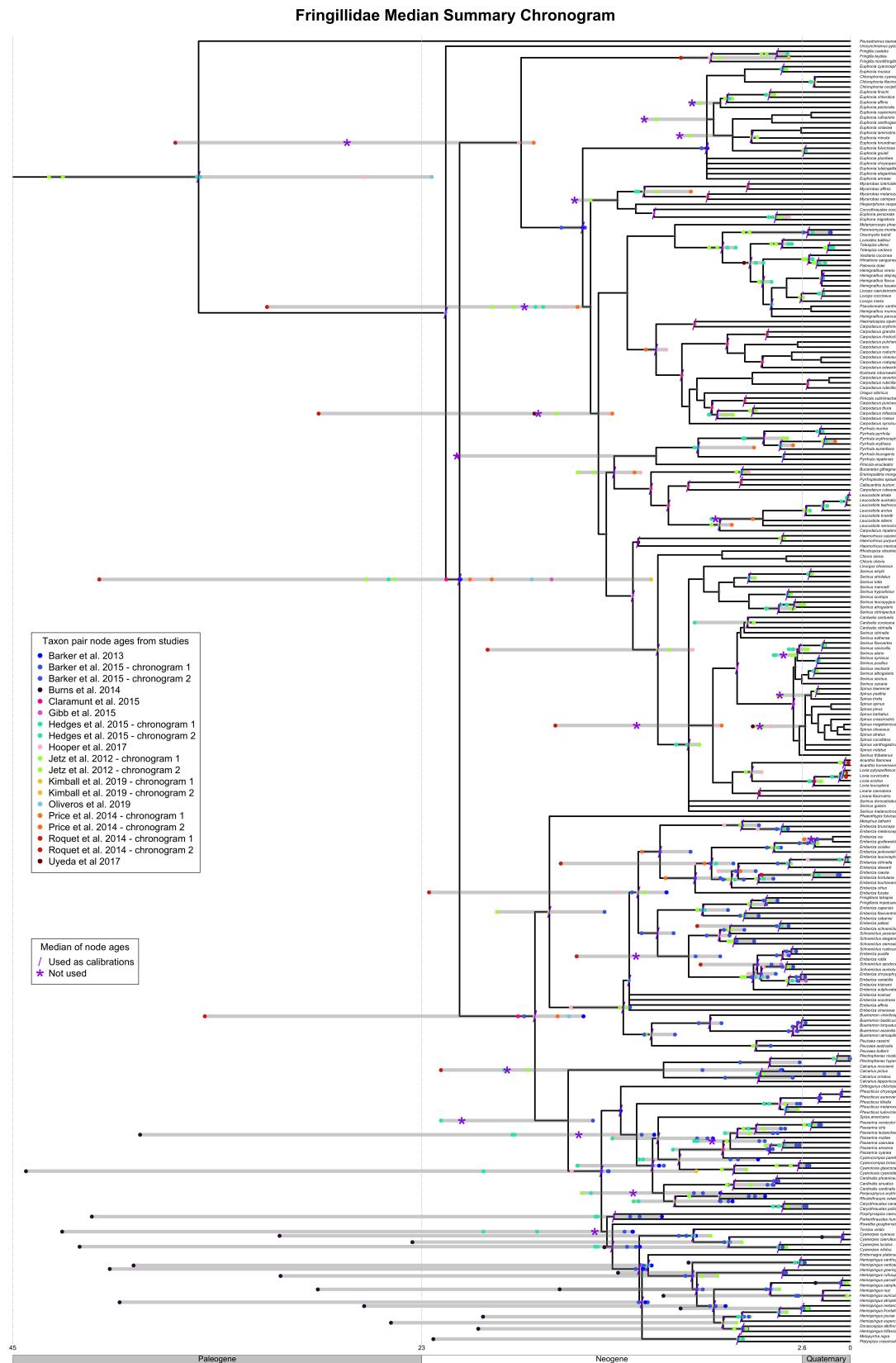


FIGURE 5. *Fringillidae* median summary chronogram generated with DateLife. It has 289 tips and 253 nodes, from which 212 have age data from at least one published chronogram. These nodes have been assigned to one of 10 different studies for 12 different taxonomic clades (subfamilies).

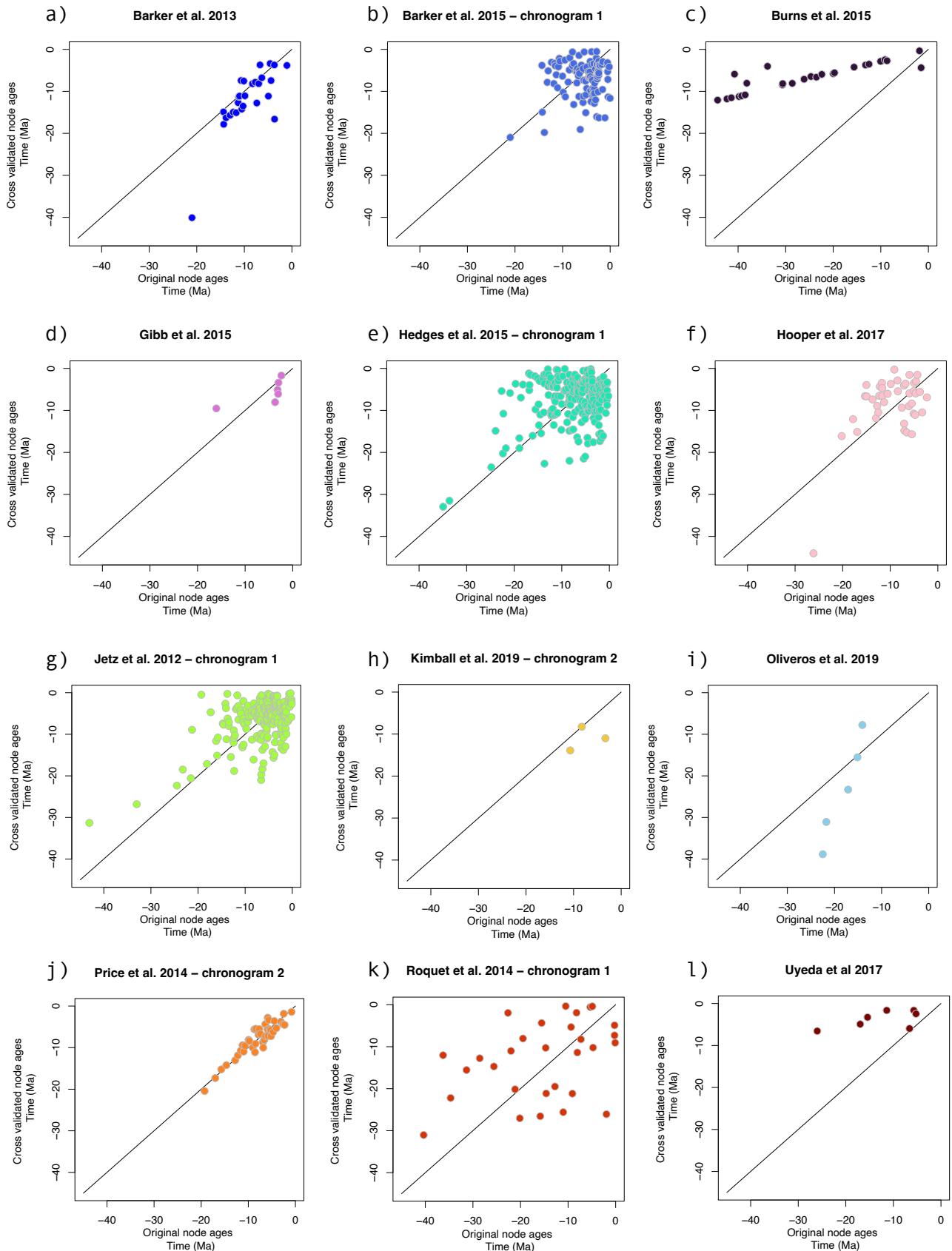


FIGURE 6. Results from cross validation analysis. Each plot compares the original age estimate (x axis) with the age obtained with a DateLife analysis (y axis), per node.