

¹ DateLife: leveraging databases and analytical tools to reveal the dated Tree of Life

² Luna L. Sánchez Reyes^{1,2}, Emily Jane McTavish¹, & Brian O'Meara²

³ ¹ University of California, Merced, USA

⁴ ² University of Tennessee, Knoxville, USA

⁵ Author Note

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

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

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 and 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, 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;

⁴³ Congruification; Supertree; Calibrations; Secondary calibrations

⁴⁴ Word count: 5355

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 comparative analysis (Freckleton, Harvey, & Pagel, 2002; Harvey, Pagel, &
50 others, 1991), developmental biology (Delsuc et al., 2018; Laubichler & Maienschein, 2009),
51 conservation and ecology (Felsenstein, 1985; Webb, 2000), historical biogeography (Posadas,
52 Crisci, & Katinas, 2006), and species diversification (Magallon & Sanderson, 2001; Morlon,
53 2014).

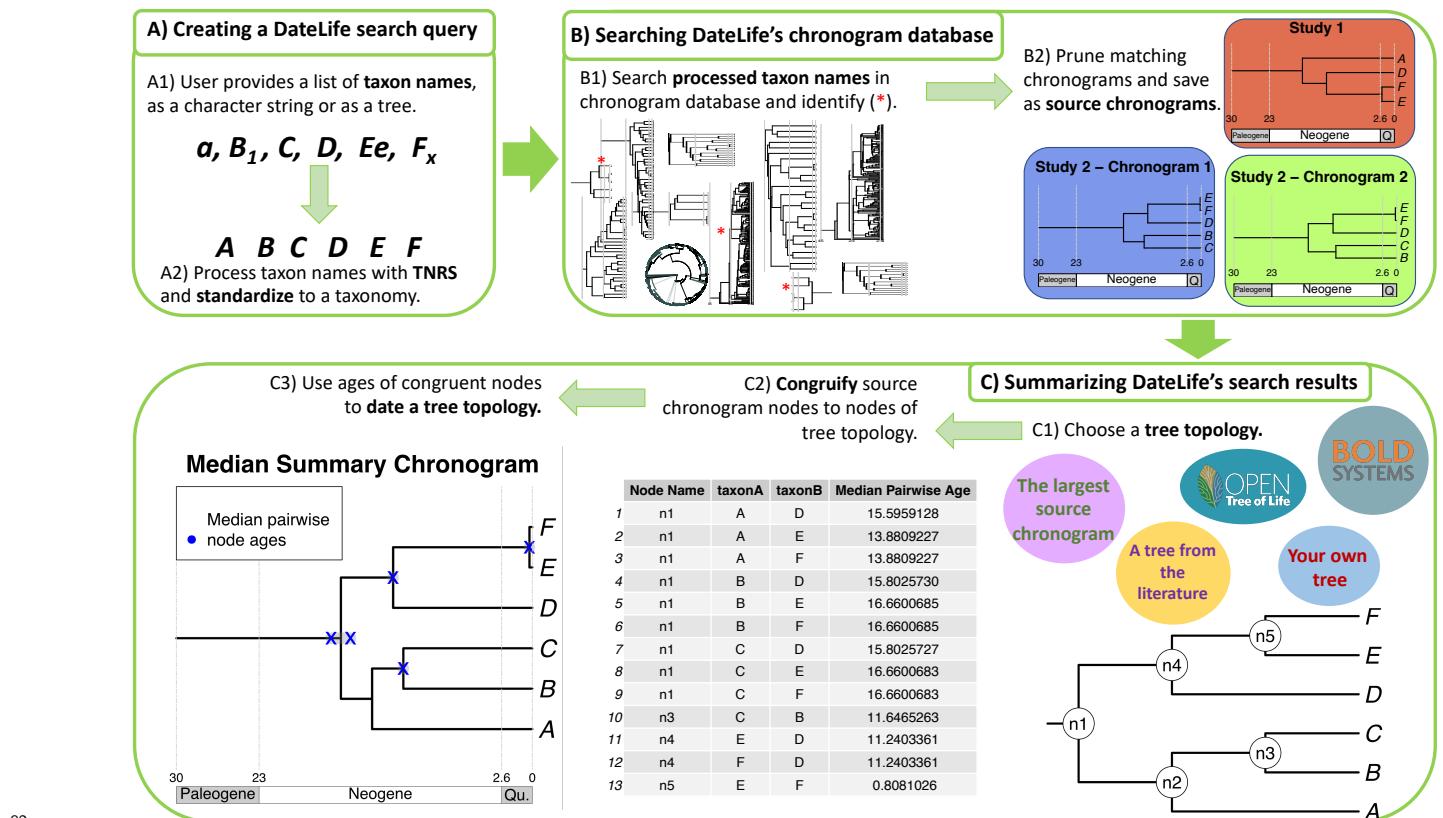
54 Building a chronogram is not an easy task. It requires obtaining and curating data to
55 construct a phylogeny, selecting and placing appropriate calibrations on the phylogeny using
56 independent age data points from the fossil record or other dated events, and inferring the
57 full dated tree; it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

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

69

Description

70 DateLife's core software application consists of the R package `datelife`. Its current
 71 stable version – v0.6.6, is available from the The Comprehensive R Archive Network (CRAN)
 72 repository (Sanchez-Reyes et al., 2022), and relies on functionalities from various biological
 73 R packages: `ape` (Paradis, Claude, & Strimmer, 2004), `bold` (Chamberlain, 2018), `geiger`
 74 (Pennell et al., 2014), `paleotree` (Bapst, 2012), `phyloch` (Heibl, 2008), `phylocomr` (Ooms &
 75 Chamberlain, 2018), `phytools` (Revell, 2012), `rotl` (Michonneau, Brown, & Winter, 2016),
 76 and `taxize` (Chamberlain, 2018; Chamberlain & Szöcs, 2013). Figure 1 provides a graphical
 77 summary of the three main steps of the DateLife workflow: creating a search query,
 78 searching a database, and summarizing results from the search.



82

79 FIGURE 1. Main DateLife workflow. Analyses can be performed via DateLife's interactive website
 80 at www.datelife.org, or using the `datelife` R package. Details on the R functions used to perform
 81 the analyses are available from `datelife`'s R package vignettes at phylotastic.org/datelife.

83 Creating a search query

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

90 DateLife processes input scientific names using a Taxonomic Name Resolution Service
91 (TNRS), which increases the probability of correctly finding the queried taxon names in the
92 chronogram database. TNRS detects, corrects and standardizes name misspellings and typos,
93 variant spellings and authorities, and nomenclatural synonyms to a single taxonomic
94 standard (Boyle et al., 2013). TNRS also allows to correctly choose between homonyms, by
95 considering other taxa provided as input to infer the taxonomic context of the homonym.

96 DateLife implements TNRS using the Open Tree of Life (OpenTree) unified Taxonomy
97 (OTT, Open Tree Of Life et al., 2016; Rees & Cranston, 2017) as standard, storing
98 taxonomic identification numbers (OTT ids) for further processing and analysis. Other
99 taxonomies currently supported by DateLife are the National Center of Biotechnology
100 Information (NCBI) taxonomic database (Schoch et al., 2020), the Global Biodiversity
101 Information Facility (GBIF) taxonomic backbone (GBIF Secretariat, 2022), and the Interim
102 Register of Marine and Nonmarine Genera (IRMNG) database (Rees et al., 2017).

103 Besides binomial species names, DateLife accepts scientific names from any inclusive
104 taxonomic group (e.g., genus, family, tribe), as well as subspecific taxonomic variants (e.g.,
105 subspecies, variants, strains). If a taxon name belongs to an inclusive taxonomic group,
106 DateLife has two alternative behaviors defined by the “get species from taxon” flag. If the
107 flag is active, DateLife retrieves all species names within the taxonomic group from the
108 standard taxonomy of choice, and adds them to the search query. In this case, subspecific

109 variants are excluded. If the flag is inactive, DateLife excludes any taxon names above the
110 species level from the search query. Species and subspecific variant names are processed and
111 searched as provided by the user. The processed taxon names are saved as an R object of a
112 newly defined class, `datelifeQuery`, that is used in the following steps. This object contains
113 the names standardized to the taxonomy of choice, the corresponding OTT id numbers, and
114 the topology of the input tree if one was provided.

115 **Searching a chronogram database**

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

123 A unique feature of the Phylesystem is that any user can add new published,
124 state-of-the-art chronograms any time, through their curator application
125 (<https://tree.opentreeoflife.org/curator>). As chronograms are added to Phylesystem, they
126 can be incorporated into the chronogram database of the `datelife` package. `datelife`'s
127 chronogram database is currently manually updated as new chronogram data is added to
128 Phylesystem. The updated database is assigned a new version number, followed by a
129 package release on CRAN. Users can also implement functions from the `datelife` R package
130 to trigger an update of the local chronogram database, to incorporate any new chronograms
131 to the user's DateLife analysis before an official database update is released on CRAN.

132 A DateLife search is implemented by matching processed taxon names provided by the
133 user to tip labels in the chronogram database. Chronograms with at least two matching

134 taxon names on their tip labels are identified and pruned down to preserve only the matched
135 taxa. These matching pruned chronograms are referred to as source chronograms. Total
136 distance (in units of millions of years) between taxon pairs within each source chronogram
137 are stored as a patristic distance matrix (Figure 1). The matrix format speeds up extraction
138 of pairwise taxon ages of any queried taxa, as opposed to searching the ancestor node of a
139 pair of taxa in a “phylo” object or newick string. Finally, the patristic matrices are
140 associated to the study citation where the original chronogram was published, and stored as
141 an R object of the newly defined class `datelifeResult`.

142 Summarizing search results

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

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

156 ***Choosing a topology.***— DateLife requires a tree topology to summarize age data
157 upon. We recommend that users provide a tree topology as input from the literature, or one
158 of their own making. If no topology is provided, DateLife automatically extracts one from

the OpenTree synthetic tree, a phylogeny currently encompassing 2.3 million taxa across all life, assembled from 1, 239 published phylogenetic trees and OpenTree’s unified Taxonomy, OTT (Open Tree Of Life et al., 2019). Alternatively, DateLife can combine topologies from source chronograms using a supertree approach. To do this DateLife first identifies the source chronograms that form a grove, roughly, a sufficiently overlapping set of taxa between trees, by implementing definition 2.8 for n-overlap from Ané et al. (2009). If the source chronograms do not form a grove, the supertree reconstruction will fail. In rare cases, a group of trees can have multiple groves. By default, DateLife chooses the grove with the most taxa, however, the “criterion = trees” flag allows the user to choose the grove with the most trees instead. The result is a single summary (or supertree) topology, that combines topologies from source chronograms in a grove.

Applying secondary calibrations.— Once a topology is chosen, DateLife applies the congruification method (Eastman, Harmon, & Tank, 2013) that find nodes belonging to the same clade across source chronograms, and then extracts the corresponding node ages from patristic distance matrices stored as a `datelifeResult` object. Note that by definition, these matrices store total distance (time from tip to tip), assuming that the terminal taxa are coeval and occur at the present. Hence, node ages correspond to half the values stored in the `datelifeResult` matrices. A table of congruified node ages that can be used as calibrations for a dating analysis is stored as a `congruifiedCalibrations` object.

For each congruent node, the pairwise distances that traverse that node are summarized into a single summary matrix using classic summary statistics (i.e., mean, median, minimum and maximum ages), and the Supermatrix Distance Method (SDM; Criscuolo, Berry, Douzery, & Gascuel, 2006), which deforms patristic distance matrices by minimizing variance and then averaging them. These single summary taxon pair age matrices are stored as summarized calibrations that can be used as secondary calibrations to date a tree topology - with or without initial branch lengths, using phylogenetic dating

185 methods currently supported within DateLife: BLADJ (Webb, Ackerly, & Kembel, 2008;
186 Webb & Donoghue, 2005), MrBayes (Huelsenbeck & Ronquist, 2001; Ronquist &
187 Huelsenbeck, 2003), PATHd8 (Britton, Anderson, Jacquet, Lundqvist, & Bremer, 2007), and
188 treePL (Smith & O'Meara, 2012).

189 **Dating a tree topology.**— By default, DateLife implements the Branch Length
190 Adjuster (BLADJ; Webb et al., 2008; Webb & Donoghue, 2005) algorithm to obtain a fully
191 dated topology. BLADJ is the only dating algorithm that can work with initial topologies
192 without any branch length data.

193 Alternatively, the user can chose phylogenetic dating options supported in DateLife
194 that incorporate branch length information from the input topology in combination with the
195 secondary calibrations: PATHd8 is a non-clock, rate-smoothing method (Britton et al., 2007)
196 to date trees; treePL (Smith & O'Meara, 2012), is a semi-parametric, rate-smoothing,
197 penalized likelihood dating method (Sanderson, 2002); the MrBayes (Huelsenbeck &
198 Ronquist, 2001; Ronquist & Huelsenbeck, 2003) approach in DateLife uses the calibrations as
199 priors on node ages.

200 The latter methods can assign dates using a birth-death model (all of them????), they
201 require initial branch lengths and thus require more time and expertise to run, which makes
202 BLADJ the fastest and most practical method to obtain a dated tree.

203 BLADJ fixes node ages that have calibration data, and distributes time between nodes
204 with no data evenly between calibrated nodes. This has proven effective minimizes age
205 variance in the resulting chronogram and useful for ecoogical analyses (Webb et al., 2008).
206 When there is conflict in ages between nodes with calibration data, BLADJ ignores node
207 ages that are older than the age of a parent node. BLADJ requires a root age estimate. If
208 there is no information on the age of the root in the chronogram database, users can provide
209 an estimate from the literature. If none is provided, DateLife will not return a dated

210 topology but provide a warning message along with suggestions on how the user can provide
211 an age for the root so DateLife can run.

212 **Dating a tree with branch lengths**.– Topologies obtained from OpenTree and
213 with the supertree approach described above lack branch length data. Yet, phylogenetic
214 dating using branch lengths data is the golden standard for phylogenetic dating analyses, but
215 it is costly and requires a lot of human/expert curation.

216 A fast solution implemented in DateLife is as follows. To estimate branch lengths
217 proportional to substitution rates for these topologies, DateLife currently implements a
218 simple algorithm. First, it mines the Barcode of Life Data System, BOLD (Ratnasingham &
219 Hebert, 2007) to obtain genetic markers for the input taxa. Mined genetic sequences are
220 aligned with MUSCLE (Edgar, 2004) (by default) or MAFFT (Katoh, Asimenos, & Toh,
221 2009).

222 The BOLD sequence alignment is then used to reconstruct branch lengths with the
223 accelerated transformation (ACCTRAN) parsimony algorithm, which resolves ambiguous
224 character optimization, by assigning changes along branches of the tree as close to the root as
225 possible (Agnarsson & Miller, 2008). This algorithm work rally fast and allows getting initial
226 branch lengths that ar ethen optimized using ML. Optionally, the likelihood of the tree
227 topology, the alignment and the reconstructed branch lengths given different evolutionary
228 models, is computed using functions from the `phangorn` package (Schliep, 2011).

229 Relative branch length information provides key data for phylogenetic dating,
230 especially for nodes without secondary calibrations available. Yet, topologies without branch
231 lengths can also be dated.

232 **Visualizing results**.– Finally, users can save all source and summary chronograms in
233 formats that permit reuse and reanalysis (such as newick and the R “phylo” format), as well
234 as visualize and compare results graphically, or construct their own graphs using DateLife’s

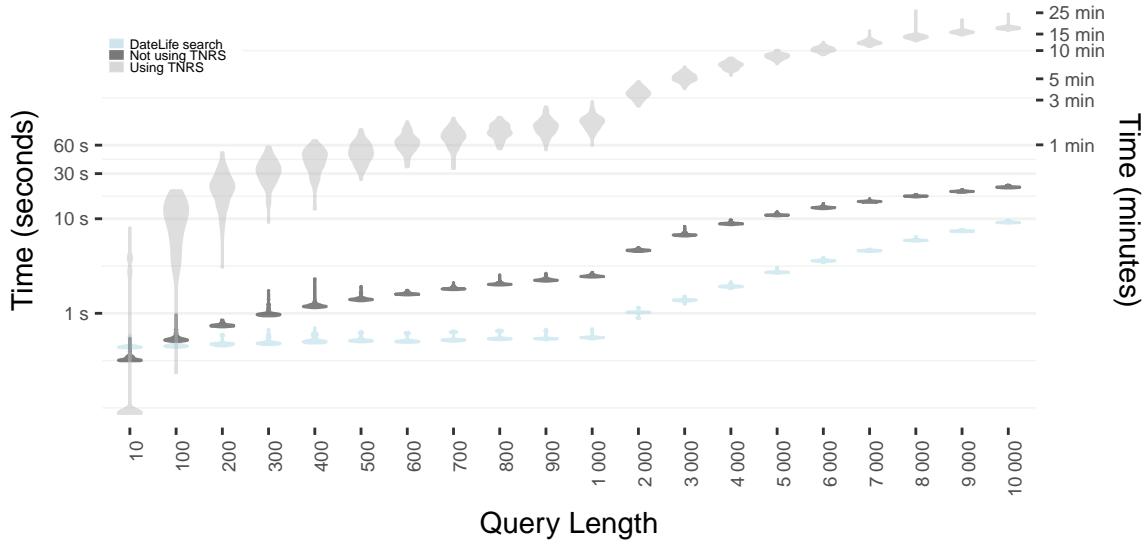
²³⁵ chronogram plot generation functions available from the R package **datelifeplot**

²³⁶ (Sanchez-Reyes & O'Meara, 2022).

237

Benchmark

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



247

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

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

262 **Case studies**

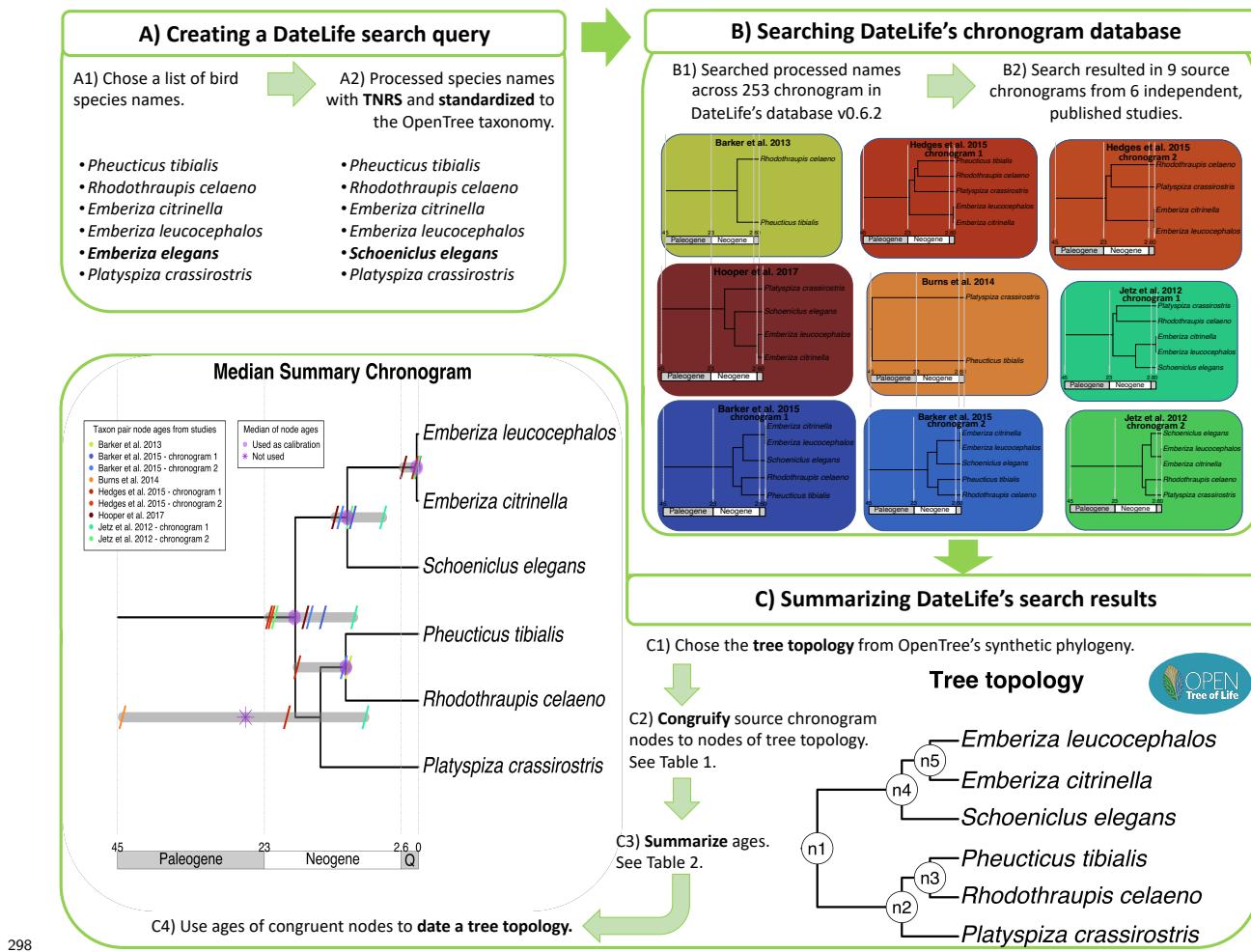
263 We illustrate the DateLife workflow using a family within the Passeriform birds
264 encompassing the true finches, Fringillidae, as case study. On a small example, we analysed 6
265 bird species, and results from each step of the workflow are shown in Fig. 3. As a second
266 example, we analysed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

269 **A small example**

270 **Creating a search query.** We chose 6 bird species within the Passeriformes. The sample
271 includes two species of cardinals: the black-thighed grosbeak – *Pheucticus tibialis* and the
272 crimson-collared grosbeak – *Rhodothraupis celaeno*; three species of buntings: the
273 yellowhammer – *Emberiza citrinella*, the pine bunting – *Emberiza leucocephalos* and the
274 yellow-throated bunting – *Emberiza elegans*; and one species of tanager, the vegetarian finch –
275 *Platyspiza crassirostris*. Processing of input names found that *Emberiza elegans* is synonym
276 for *Schoeniclus elegans* in the default reference taxonomy (OTT v3.3, June 1, 2021). For a
277 detailed discussion on the state of the synonym, refer to Avibase (Avibase, 2022; Lepage,
278 2004; Lepage, Vaidya, & Guralnick, 2014). Discovering this synonym allowed assigning five
279 age data points for the parent node of *Emberiza elegans*, shown as *Schoeniclus elegans* in
280 figure 3A, which would not have had any data otherwise.

281 **Searching the database.** DateLife used the processed input names to search the local
282 chronogram database and found 9 matching chronograms in 6 different studies (Fig. 3B).
283 Three studies matched five input names (Barker, Burns, Klicka, Lanyon, & Lovette, 2015;
284 Hedges, Marin, Suleski, Paymer, & Kumar, 2015; Jetz, Thomas, Joy, Hartmann, & Mooers,
285 2012), one study matched four input names (Hooper & Price, 2017) and two studies matched
286 two input names (Barker, Burns, Klicka, Lanyon, & Lovette, 2013; Burns et al., 2014). No
287 studies matched all input names. Together, source chronograms provide 28 unique age data
288 points, covering all nodes on our chosen tree topology to date (Table 1).

289 **Summarizing search results.** DateLife obtained OpenTree’s synthetic tree topology for
290 these taxa (Fig. 3C), and congruified and mapped age data to nodes in this chosen topology
291 (Table 1). The name processing step allowed including five data points for node “n4” (parent
292 of *Schoeniclus elegans*; Fig. 3A) that would not have had any data otherwise due to name
293 mismatch. Age summary statistics per node were calculated (Table 2) and used as
294 calibrations to date the tree topology using the BLADJ algorithm. As expected, more
295 inclusive nodes (e.g., node “n1”) have more variance in age data than less inclusive nodes
296 (e.g., node “n5”). Summary age data for node “n2” were excluded as final calibration
297 because they are older than age data of the more inclusive node, “n1” (Fig. 3C4).



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

306 TABLE 1. Ages of congruified nodes. 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

307

308 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

309

310 **An example with the family of true finches**

311 **Creating a query.** To obtain ages for all species within the family of true finches,
312 Fringillidae, we ran a DateLife query using the “get species from taxon” flag, which gets all
313 recognized species names within a named group from a taxonomy of choice. Following the
314 NCBI taxonomy, our DateLife query has 289 Fringillidae species names. This
315 taxon-constrained approach implies that the full DateLife analysis will be performed using a
316 tree topology and ages available for species names from a given taxonomic group, which do
317 not necessarily correspond to a monophyletic group. Users can change this behaviour by
318 providing all species names corresponding to a monophyletic group as input for a DateLife
319 search, or a monophyletic tree to construct a DateLife summary.

320 **Searching the database.** Next, we used the processed species names in our DateLife
321 query to identify chronograms with at least two Fringillidae species as tip taxa. The
322 DateLife search identified 19 chronograms matching this criteria, published in 13 different
323 studies (Barker et al., 2013, 2015; Burns et al., 2014; Claramunt & Cracraft, 2015; Gibb et
324 al., 2015; Hedges et al., 2015; Hooper & Price, 2017; Jetz et al., 2012; Kimball et al., 2019;
325 Oliveros et al., 2019; Price et al., 2014; Roquet, Lavergne, & Thuiller, 2014; Uyeda, Pennell,
326 Miller, Maia, & McClain, 2017). Once identified, DateLife pruned these matching
327 chronograms to remove tips that do not belong to the queried taxon names, and transformed
328 these pruned chronograms to pairwise distance matrices, revealing 1, 206 different age data
329 points available for species within the Fringillidae (Supplementray Table S1).

330 **Summarizing search results.** The final step entailed congruifying and summarizing the
331 age data available for the Fringillidae species into single summary chronograms, using two
332 different types of summary ages, median and SDM. As explained in the “Description”
333 section, a tree topology to summarize age data upon is required. By default, to do this,
334 DateLife uses the topology from OpenTree’s synthetic tree that contains species from the
335 search query. According to OpenTree’s synthetic tree, species belonging to the family
336 Fringillidae do not form a monophyletic group (Fig. 4).

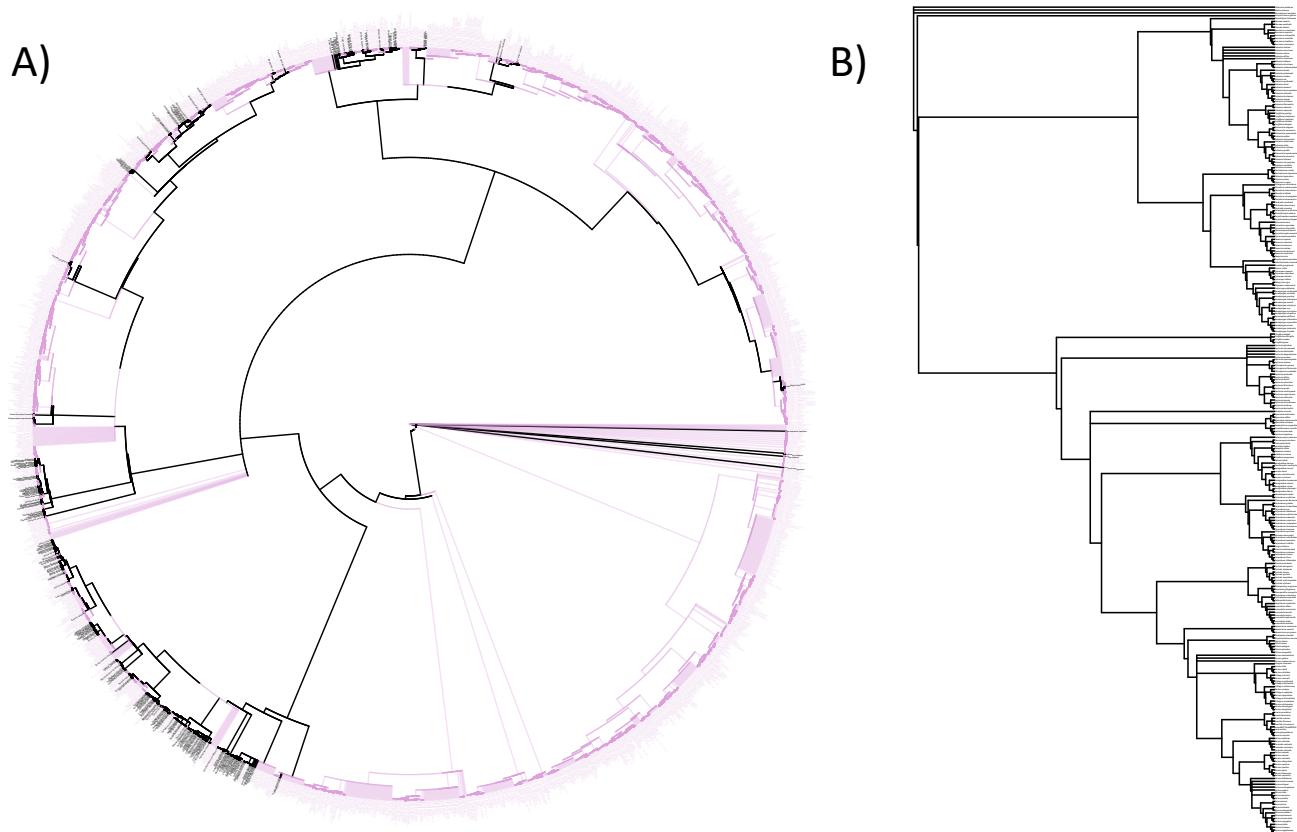
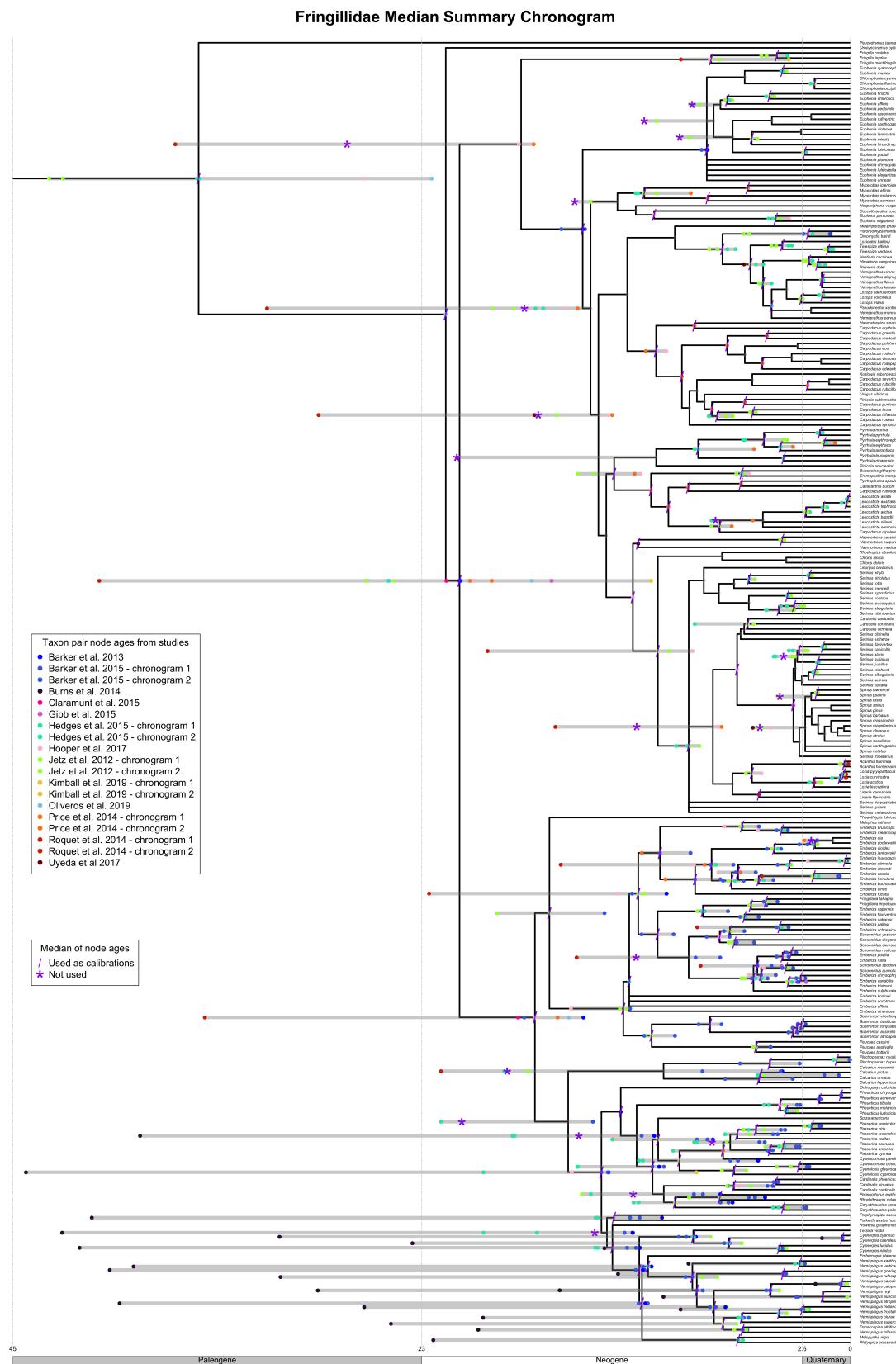


FIGURE 4. Tree topologies extracted from Open Tree of Life's (OpenTree) synthetic phylogenetic tree.
 337 A) Topology of 2,333 tips and 1,305 internal nodes, encompassing bird species within the family
 338 Fringillidae following the NCBI taxonomy (black), as well as all other bird species that share the same
 339 MRCA node in OpenTree's synthetic tree (purple). B) Subtree
 340 topology of 289 tips and 253 internal nodes, resulting from pruning species that do not belong to the
 341 family Fringillidae according to the NCBI taxonomy (purple branches in topology A). Bird species
 342 within the Fringillidae are paraphyletic (Alström et al. 2014, Barker, Cibois, Schikler, Feinstein, &
 343 Cracraft 2004, Barker et al. 2013, Barker 2014, Barker et al. 2015, Beresford, Barker, Ryan, &
 344 Crowe 2005, Bryson Jr et al. 2014, Burleigh, Kimball, & Braun 2015, Burns et al. 2014, Chaves,
 345 Hidalgo, & Klicka 2013, Claramunt & Cracraft 2015, Gibb et al. 2015, Hackett et al. 2008, Jetz et al.
 346 2012, Johansson, Fjeldså, & Bowi 200, Kimball et al. 2019, Klicka et al. 2014, Lamichhaney et al.
 347 2015, Lerner, Meyer, James, Hofreiter, & Fleischer 2011, Lovette et al. 2010, Moyle et al. 2016,
 348 Ödeen, Håstad, & Alström 2011, Oliveros et al. 2019, Päckert et al. 2012, Parchman, Benkman, &
 349 Mezquida 2007, Powell et al. 2014, Price et al. 2014, Pulgarín-R, Smith, Bryson Jr, Spellman, &
 350 Klicka 2013, Selvatti, Gonzaga, & Moraes Russo 2015, Tietze, Päckert, Martens, Lehmann, & Sun
 351 2013, Treplin et al. 2008, Zuccon, Prŷs-Jones, Rasmussen, & Ericson 2012).

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



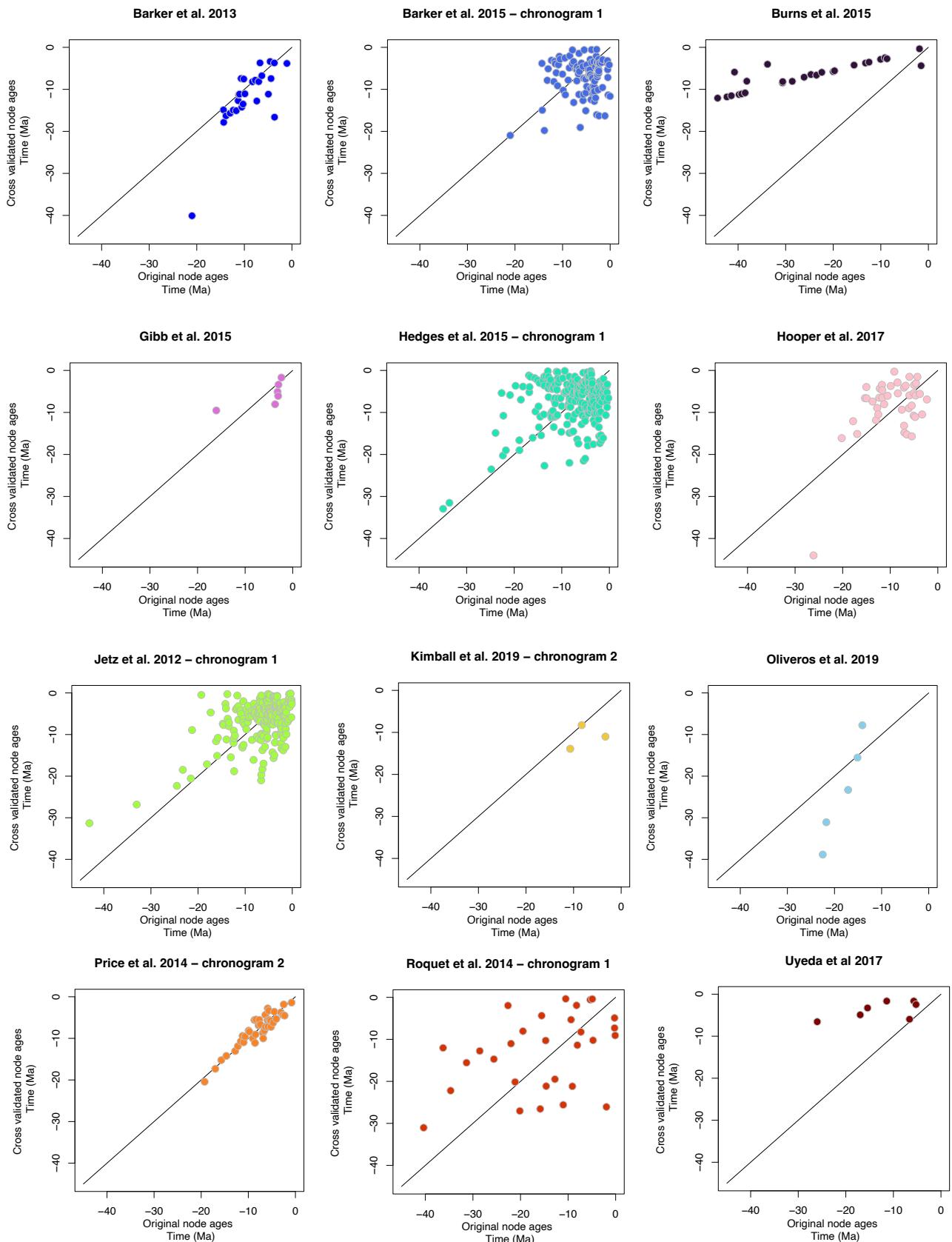
361

359 FIGURE 5. *Fringillidae* median summary chronogram generated with DateLife. It has 256 tips and
360 233 nodes, from which 212 have age data from at least one published chronogram.

362

Cross-validation test

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



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

376

Discussion

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

382 A total of 99,474 unique terminal taxa are represented in DateLife’s database. Incorporation
383 of more chronograms into the database will continue to improve DateLife’s services. One
384 option to increase the number of chronograms in the DateLife database is the Dryad data
385 repository. Methods to automatically mine chronograms from Dryad could be designed and
386 implemented. However, Dryad’s metadata system has no information to automatically detect
387 branch length units, and those would still need to be determined manually by a human
388 curator. We would like to emphasize on the importance of sharing chronogram data,
389 including systematically curated metadata, into open repositories, such as OpenTree’s
390 Phylesystem (McTavish et al., 2015) for the benefit of the scientific community as a whole.

391 **Age variation in source chronograms**

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

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

410 **Primary vs Secondary calibrations.** While most chronograms in DateLife’s database
411 are constructed using primary calibrations (molecular substitution rates or ages obtained
412 from the fossil record or geological events), DateLife summarizes chronograms using
413 secondary calibrations (ages coming from other chronograms). Graur and Martin (2004)
414 cautioned on the increased error and uncertainty in estimated ages when using secondary
415 calibrations in dating analyses. Schenk (2016) showed that, in simulations, divergence times
416 inferred using secondary calibrations are significantly younger than those inferred with
417 primary calibrations, when obtained with Bayesian inference methods, and when priors are
418 implemented in similar ways in both analyses. Accordingly, the scientific community seems
419 to have more confidence in chronograms obtained from a single analysis, using fossil data as
420 primary sources of calibrations (Schenk, 2016), and using fossils that have been widely
421 discussed and curated as calibrations to date other trees, making sure that all data reflect a
422 coherent evolutionary history (Sauquet, 2013), as for example done by Antonelli et al.
423 (2017). There have been attempts to create fossil calibration databases (Ksepka et al., 2015),
424 though these still have room to grow.

425 It seems that using several (as opposed to just a few) secondary calibrations can provide
426 sufficient information to alleviate or even neutralize potential biases (Sauquet, 2013).
427 Certainly, further studies are required to fully understand the effect of secondary calibrations
428 on outputs from different tree dating methods, and on downstream analyses. It is possible

429 that secondary calibrations can be safely used with dating methods that do not require
430 setting priors, such as penalized likelihood (Sanderson, 2003), with methods that do not
431 make any assumptions on the ages and fix them to a node on a tree topology, such as
432 BLADJ (Webb et al., 2008; Webb & Donoghue, 2005), or methods that summarize age data
433 unto a tree topology.

434 Our cross validation analysis might provide some insight in this regard. When ages are
435 estimated with secondary calibrations, nodes closer to the root do tend to be slightly
436 younger than ages estimated with primary calibrations. However, nodes closer to the tip
437 tend to be older when estimated using secondary calibrations with a dating method that
438 does not make any prior assumptions on the nature of the calibrations themselves
439 (Supplementary Figures S2-S20). The only exception to this was observed on cross validation
440 results of the Burns et al. (2014) chronogram, which displays much younger node ages when
441 estimated using secondary calibrations (Supplementary Figs. S1, S5).

442 Sumarizing chronograms

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

450 Some measures that have been proposed are the proportion of lineage sampling and the
451 number of calibrations used (Magallón, 2010, @magallon2015metacalibrated). Some
452 characteristics that are often cited in published studies as a measure of improved age
453 estimates as compared to previously published estimates are: quality of alignment (missing
454 data, GC content), lineage sampling (strategy and proportion), phylogenetic and dating

455 inference method, number of fossils used as calibrations, support for nodes and ages, and
456 magnitude of confidence intervals.

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

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

471 Effects on downstream analyses

472 The study of phenomena dependent on the timing of species diversification events, such as
473 macroevolutionary processes, is affected by the usage of alternative chronograms that vary in
474 topology (Rabosky, 2015; Title & Rabosky, 2016).

475 On the other hand, incorporating at least some data on lineage divergence times represents a
476 relevant improvement for testing alternative hypothesis using phylogenetic distance in
477 ecological and conservation biology studies (Webb et al., 2008). Hence, DateLife's workflow
478 features different ways of generating node ages in the absence of calibration and branch
479 length information for certain taxa.

480 Adding branch lengths sampled from a birth-death model in the absence of genetic data has

481 been found to improve insight in phylogeny-based analyses.

482 is a common practice in scientific publications: Jetz et al. (2012), created a chronogram of

483 all 9, 993 bird species, where 67% had molecular data and the rest was simulated; Rabosky

484 et al. (2018) created a chronogram of 31, 536 ray-finned fishes, of which only 37% had

485 molecular data; Smith and Brown (2018) constructed a chronogram of 353, 185 seed plants

486 where only 23% had molecular data.

487 Simulating branch lengths following a birth-death species diversification model for missing

488 taxa in chronograms with non random sampling, lowers type I error (false positive,

489 incorrectly accepting the null hypothesis of a constant-rate or temporally varying rate

490 birth-death model) when analyzing changes in diversification rate using the gamma statistic

491 (Cusimano, Stadler, & Renner, 2012).

492 Thomas et al. (2013) hypothesize that results of diversification analyses will be biased

493 towards the birth-death model used to simulate branch lengths; and note that “the effects of

494 missing species placement or polytomy resolution are less clear for other phylogeny-based

495 analyses (e.g. correlates of diversification, modelling trait evolution, community

496 phylogenetics), and future work should test how the treatment of missing species influences

497 both parameter estimation and type I and II errors”

498 Notably, risks come with this practice.

499 Taken to the extreme, one could generate a fully resolved, calibrated tree of all modern and

500 extinct taxa using a single taxonomy and a single calibration, using polytomy resolution and

501 branch length simulation methods. There has yet to be a thorough analysis of what can go

502 wrong when one extends inferences beyond the data in this way, so we urge caution; we also

503 urge readers to follow the example of the large tree papers cited above, by carefully

504 considering the statistical assumptions being made, and assessing the consistency of the

505 results with prior work.

506

Conclusions

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

515 The DateLife project allows for easy and fast summary of public and state-of-the-art data on
516 time of lineage divergence. It provides a straightforward way to get an informed idea on the
517 state of knowledge of the time frame of evolution of different regions of the tree of life, and
518 allows identifying regions that require more research, or that have conflicting information. It
519 is available as an R package, and as a web-based R shiny application at www.datelife.org

520 Both summary and newly generated trees are useful to evaluate evolutionary hypotheses in
521 different areas of research. The DateLife project should improve awareness of the existing
522 variation in expert time of divergence data, and foster exploration of the effect of alternative
523 divergence time hypothesis on the results of analyses, nurturing a culture of more cautious
524 interpretation of evolutionary results.

525

Availability

526 The DateLife software is free and open source and it can be used through its R shiny web
527 application at <http://www.datelife.org>, through the **datelife** R package, and through
528 Phylotastic's project web portal <https://phylo.cs.nmsu.edu/>. DateLife's web application is
529 maintained using RStudio's shiny server and the shiny package open infrastructure, as well
530 as Docker and OpenTree's infrastructure (dates.opentreeoflife.org/datelife). **datelife**'s R
531 package stable version is available for installation from the CRAN repository

532 (<https://cran.r-project.org/package=datelife>) using the command `install.packages(pkgs`
533 `= "datelife")` from within R. Development versions are available from the GitHub
534 repository (<https://github.com/phylotastic/datelife>) and can be installed using the
535 command `devtools::install_github("phylotastic/datelife")`.

536 **Supplementary Material**

537 Supplementary material, including code, biological examples and benchmark results data
538 files and online-only appendices, can be found in the Dryad data repository
539 (<https://doi.org/10.5061/dryad.cnp5hqc6w>), as well as in the GitHub repositories used to
540 develop the reproducible manuscript (<https://doi.org/10.5281/zenodo.7435094>), the
541 biological examples (<https://doi.org/10.5281/zenodo.7435101>), and the software benchmark
542 (<https://doi.org/10.5281/zenodo.7435106>).

543 **Funding**

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

547 **Acknowledgements**

548 The DateLife project was born as a prototype tool aiming to provide these services, and was
549 initially developed over a series of hackathons at the National Evolutionary Synthesis Center,
550 NC, USA (Stoltzfus et al., 2013). We thank colleagues from the O'Meara Lab at the
551 University of Tennessee Knoxville for suggestions, discussions and software testing. The late
552 National Evolutionary Synthesis Center (NESCenT), which sponsored hackathons that led to
553 initial work on this project. The team that assembled DateLife's first proof of concept:
554 Tracy Heath, Jonathan Eastman, Peter Midford, Joseph Brown, Matt Pennell, Mike Alfaro,
555 and Luke Harmon. The Open Tree of Life project that provides the open, metadata rich
556 repository of trees used to construct DateLife's chronogram database. The many scientists

557 who publish their chronograms in an open, reusable form, and the scientists who curate them
558 for deposition in the Open Tree of Life repository. The NSF for funding nearly all the above,
559 in addition to the ABI grant that funded this project itself.

560

References

- 561 562 Agnarsson, I., & Miller, J. A. (2008). Is acctran better than deltran? *Cladistics*, 24(6), 1032–1038.
- 563 564 Alström, P., Hooper, D. M., Liu, Y., Olsson, U., Mohan, D., Gelang, M., … Price, T. D. (2014). Discovery of a relict lineage and monotypic family of passerine birds. *Biology Letters*, 10(3), 20131067.
- 565 566 Ané, C., Eulenstein, O., Piaggio-Talice, R., & Sanderson, M. J. (2009). Groves of phylogenetic trees. *Annals of Combinatorics*, 13(2), 139–167.
- 567 568 Antonelli, A., Hettling, H., Condamine, F. L., Vos, K., Nilsson, R. H., Sanderson, M. J., … Vos, R. A. (2017). Toward a self-updating platform for estimating rates of speciation and migration, ages, and relationships of Taxa. *Systematic Biology*, 66(2), 153–166.
571 <https://doi.org/10.1093/sysbio/syw066>
- 572 573 Archie, J., Day, W. H., Felsenstein, J., Maddison, W., Meacham, C., Rohlf, F. J., & Swofford, D. (1986). The Newick tree format. Retrieved from
574 {<https://evolution.genetics.washington.edu/phylip/newicktree.html>}
- 575 576 Avibase. (2022). Yellow-throated Bunting. *Avibase - the World Bird Database*, (Online Resource). Retrieved from
577 {<https://avibase.bsc-eoc.org/species.jsp?lang=EN&avibaseid=82D1EE0049D8D927>}
- 578 579 Bapst, D. W. (2012). Paleotree: An R package for paleontological and phylogenetic analyses of evolution. *Methods in Ecology and Evolution*, 3(5), 803–807.
580 <https://doi.org/10.1111/j.2041-210X.2012.00223.x>
- 581 582 Barba-Montoya, J., Reis, M. dos, Schneider, H., Donoghue, P. C., & Yang, Z. (2018). Constraining uncertainty in the timescale of angiosperm evolution and the veracity of
583 a cretaceous terrestrial revolution. *New Phytologist*, 218(2), 819–834.
- 584 Barker, F. K. (2014). Mitogenomic data resolve basal relationships among passeriform and
585 passeridan birds. *Molecular Phylogenetics and Evolution*, 79, 313–324.
- 586 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2013). Going to

- extremes: Contrasting rates of diversification in a recent radiation of new world
passerine birds. *Systematic Biology*, 62(2), 298–320.
- Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2015). New insights
into new world biogeography: An integrated view from the phylogeny of blackbirds,
cardinals, sparrows, tanagers, warblers, and allies. *The Auk: Ornithological Advances*,
132(2), 333–348.
- Barker, F. K., Cibois, A., Schikler, P., Feinstein, J., & Cracraft, J. (2004). Phylogeny and
diversification of the largest avian radiation. *Proceedings of the National Academy of
Sciences*, 101(30), 11040–11045.
- Beresford, P., Barker, F., Ryan, P., & Crowe, T. (2005). African endemics span the tree of
songbirds (passeri): Molecular systematics of several evolutionary “enigmas”.
Proceedings of the Royal Society B: Biological Sciences, 272(1565), 849–858.
- Boyle, B., Hopkins, N., Lu, Z., Raygoza Garay, J. A., Mozzherin, D., Rees, T., ... Enquist,
B. J. (2013). The taxonomic name resolution service: An online tool for automated
standardization of plant names. *BMC Bioinformatics*, 14(1).
<https://doi.org/10.1186/1471-2105-14-16>
- Britton, T., Anderson, C. L., Jacquet, D., Lundqvist, S., & Bremer, K. (2007). Estimating
Divergence Times in Large Phylogenetic Trees. *Systematic Biology*, 56(788777878),
741–752. <https://doi.org/10.1080/10635150701613783>
- Bryson Jr, R. W., Chaves, J., Smith, B. T., Miller, M. J., Winker, K., Pérez-Emán, J. L., &
Klicka, J. (2014). Diversification across the new world within the ‘blue’cardinalids
(aves: Cardinalidae). *Journal of Biogeography*, 41(3), 587–599.
- Burleigh, J. G., Kimball, R. T., & Braun, E. L. (2015). Building the avian tree of life using a
large-scale, sparse supermatrix. *Molecular Phylogenetics and Evolution*, 84, 53–63.
- Burns, K. J., Shultz, A. J., Title, P. O., Mason, N. A., Barker, F. K., Klicka, J., ... Lovette,
I. J. (2014). Phylogenetics and diversification of tanagers (passeriformes:
Thraupidae), the largest radiation of neotropical songbirds. *Molecular Phylogenetics*

- 614 and *Evolution*, 75, 41–77.
- 615 Chamberlain, S. (2018). *bold: Interface to Bold Systems API*. Retrieved from
616 <https://CRAN.R-project.org/package=bold>
- 617 Chamberlain, S. A., & Szöcs, E. (2013). taxize : taxonomic search and retrieval in R [version
618 2; referees: 3 approved]. *F1000Research*, 2(191), 1–29.
619 <https://doi.org/10.12688/f1000research.2-191.v2>
- 620 Chaves, J. A., Hidalgo, J. R., & Klicka, J. (2013). Biogeography and evolutionary history of
621 the n eotropical genus s altator (a ves: T hraupini). *Journal of Biogeography*, 40(11),
622 2180–2190.
- 623 Claramunt, S., & Cracraft, J. (2015). A new time tree reveals earth history's imprint on the
624 evolution of modern birds. *Science Advances*, 1(11), e1501005.
- 625 Criscuolo, A., Berry, V., Douzery, E. J., & Gascuel, O. (2006). SDM: A fast distance-based
626 approach for (super)tree building in phylogenomics. *Systematic Biology*, 55(5),
627 740–755. <https://doi.org/10.1080/10635150600969872>
- 628 Cusimano, N., Stadler, T., & Renner, S. S. (2012). A new method for handling missing
629 species in diversification analysis applicable to randomly or nonrandomly sampled
630 phylogenies. *Systematic Biology*, 61(5), 785–792.
- 631 Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M.-K., Turon, X., ... Douzery,
632 E. J. (2018). A phylogenomic framework and timescale for comparative studies of
633 tunicates. *BMC Biology*, 16(1), 1–14.
- 634 Eastman, J. M., Harmon, L. J., & Tank, D. C. (2013). Congruification: Support for time
635 scaling large phylogenetic trees. *Methods in Ecology and Evolution*, 4(7), 688–691.
636 <https://doi.org/10.1111/2041-210X.12051>
- 637 Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high
638 throughput. *Nucleic Acids Research*, 32(5), 1792–1797.
- 639 Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American Naturalist*,
640 125(1), 1–15. Retrieved from <http://www.jstor.org/stable/2461605>

- 641 Freckleton, R. P., Harvey, P. H., & Pagel, M. (2002). Phylogenetic analysis and comparative
642 data: A test and review of evidence. *The American Naturalist*.
- 643 GBIF Secretariat. (2022). GBIF Backbone Taxonomy. *Checklist dataset*, (Online Resource
644 accessed via GBIF.org). Retrieved from {<https://doi.org/10.15468/39omei> }
- 645 Gibb, G. C., England, R., Hartig, G., McLenaghan, P. A., Taylor Smith, B. L., McComish,
646 B. J., ... Penny, D. (2015). New zealand passerines help clarify the diversification of
647 major songbird lineages during the oligocene. *Genome Biology and Evolution*, 7(11),
648 2983–2995.
- 649 Graur, D., & Martin, W. (2004). Reading the entrails of chickens: Molecular timescales of
650 evolution and the illusion of precision. *TRENDS in Genetics*, 20(2), 80–86.
- 651 Hackett, S. J., Kimball, R. T., Reddy, S., Bowie, R. C., Braun, E. L., Braun, M. J., ...
652 others. (2008). A phylogenomic study of birds reveals their evolutionary history.
653 *Science*, 320(5884), 1763–1768.
- 654 Harvey, P. H., Pagel, M. D., & others. (1991). *The comparative method in evolutionary
655 biology* (Vol. 239). Oxford university press Oxford.
- 656 Hedges, S. B., Marin, J., Suleski, M., Paymer, M., & Kumar, S. (2015). Tree of life reveals
657 clock-like speciation and diversification. *Molecular Biology and Evolution*, 32(4),
658 835–845. <https://doi.org/10.1093/molbev/msv037>
- 659 Heibl, C. (2008). *PHYLOCH: R language tree plotting tools and interfaces to diverse
660 phylogenetic software packages*. Retrieved from
661 <http://www.christophheibl.de/Rpackages.html>
- 662 Hooper, D. M., & Price, T. D. (2017). Chromosomal inversion differences correlate with
663 range overlap in passerine birds. *Nature Ecology & Evolution*, 1(10), 1526.
- 664 Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic
665 trees. *Bioinformatics*, 17(8), 754–755.
666 <https://doi.org/10.1093/bioinformatics/17.8.754>
- 667 Jetz, W., Thomas, G., Joy, J. J., Hartmann, K., & Mooers, A. (2012). The global diversity

- 668 of birds in space and time. *Nature*, 491(7424), 444–448.
- 669 <https://doi.org/10.1038/nature11631>
- 670 Johansson, U. S., Fjeldså, J., & Bowie, R. C. (2008). Phylogenetic relationships within
671 passerida (aves: Passeriformes): A review and a new molecular phylogeny based on
672 three nuclear intron markers. *Molecular Phylogenetics and Evolution*, 48(3), 858–876.
- 673 Katoh, K., Asimenos, G., & Toh, H. (2009). Multiple alignment of dna sequences with mafft.
674 In *Bioinformatics for dna sequence analysis* (pp. 39–64). Springer.
- 675 Kimball, R. T., Oliveros, C. H., Wang, N., White, N. D., Barker, F. K., Field, D. J., ...
676 others. (2019). A phylogenomic supertree of birds. *Diversity*, 11(7), 109.
- 677 Klicka, J., Barker, F. K., Burns, K. J., Lanyon, S. M., Lovette, I. J., Chaves, J. A., & Bryson
678 Jr, R. W. (2014). A comprehensive multilocus assessment of sparrow (aves:
679 Passerellidae) relationships. *Molecular Phylogenetics and Evolution*, 77, 177–182.
- 680 Ksepka, D. T., Parham, J. F., Allman, J. F., Benton, M. J., Carrano, M. T., Cranston, K.
681 A., ... others. (2015). The fossil calibration database—a new resource for divergence
682 dating. *Systematic Biology*, 64(5), 853–859.
- 683 Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio,
684 A., ... others. (2015). Evolution of darwin's finches and their beaks revealed by
685 genome sequencing. *Nature*, 518(7539), 371–375.
- 686 Laubichler, M. D., & Maienschein, J. (2009). *Form and function in developmental evolution*.
687 Cambridge University Press.
- 688 Lepage, D. (2004). *Avibase: The world bird database*. Bird Studies Canada.
- 689 Lepage, D., Vaidya, G., & Guralnick, R. (2014). Avibase—a database system for managing
690 and organizing taxonomic concepts. *ZooKeys*, (420), 117.
- 691 Lerner, H. R., Meyer, M., James, H. F., Hofreiter, M., & Fleischer, R. C. (2011). Multilocus
692 resolution of phylogeny and timescale in the extant adaptive radiation of hawaiian
693 honeycreepers. *Current Biology*, 21(21), 1838–1844.
- 694 Lovette, I. J., Pérez-Emán, J. L., Sullivan, J. P., Banks, R. C., Fiorentino, I.,

695 Córdoba-Córdoba, S., ... others. (2010). A comprehensive multilocus phylogeny for
696 the wood-warblers and a revised classification of the parulidae (aves). *Molecular*
697 *Phylogenetics and Evolution*, 57(2), 753–770.

698 Magallon, S., & Sanderson, M. (2001). Absolute diversification rates in angiosperm clades.
699 *Evolution*, 55(9), 1762–1780.

700 Magallón, S. (2010). Using fossils to break long branches in molecular dating: A comparison
701 of relaxed clocks applied to the origin of angiosperms. *Systematic Biology*, 59(4),
702 384–399.

703 Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., & Hernández-Hernández, T. (2015).
704 A metacalibrated time-tree documents the early rise of flowering plant phylogenetic
705 diversity. *New Phytologist*, 207(2), 437–453.

706 McTavish, E. J., Hinchliff, C. E., Allman, J. F., Brown, J. W., Cranston, K. A., Holder, M.
707 T., ... Smith, S. (2015). Phylesystem: A git-based data store for community-curated
708 phylogenetic estimates. *Bioinformatics*, 31(17), 2794–2800.

709 Michonneau, F., Brown, J. W., & Winter, D. J. (2016). rotl: an R package to interact with
710 the Open Tree of Life data. *Methods in Ecology and Evolution*, 7(12), 1476–1481.
711 <https://doi.org/10.1111/2041-210X.12593>

712 Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecology Letters*,
713 17(4), 508–525. <https://doi.org/10.1111/ele.12251>

714 Moyle, R. G., Oliveros, C. H., Andersen, M. J., Hosner, P. A., Benz, B. W., Manthey, J. D.,
715 ... Faircloth, B. C. (2016). Tectonic collision and uplift of Wallacea triggered the
716 global songbird radiation. *Nature Communications*, 7(1), 1–7.

717 Oliveros, C. H., Field, D. J., Ksepka, D. T., Barker, F. K., Aleixo, A., Andersen, M. J., ...
718 others. (2019). Earth history and the passerine superradiation. *Proceedings of the*
719 *National Academy of Sciences*, 116(16), 7916–7925.

720 Ooms, J., & Chamberlain, S. (2018). *Phylocomr: Interface to 'phylocom'*. Retrieved from
721 <https://CRAN.R-project.org/package=phylocomr>

- 722 Open Tree Of Life, Redelings, B., Cranston, K. A., Allman, J., Holder, M. T., & McTavish,
723 E. J. (2016). Open Tree of Life APIs v3.0. *Open Tree of Life Project*, (Online
724 Resources). Retrieved from
725 {<https://github.com/OpenTreeOfLife/germinator/wiki/Open-Tree-of-Life-Web->
726 APIs}
- 727 Open Tree Of Life, Redelings, B., Sánchez Reyes, L. L., Cranston, K. A., Allman, J., Holder,
728 M. T., & McTavish, E. J. (2019). Open tree of life synthetic tree v12.3. *Zenodo*.
729 Retrieved from <https://doi.org/10.5281/zenodo.3937742>
- 730 Ödeen, A., Håstad, O., & Alström, P. (2011). Evolution of ultraviolet vision in the largest
731 avian radiation-the passerines. *BMC Evolutionary Biology*, 11(1), 1–8.
- 732 Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and
733 evolution in R language. *Bioinformatics*, 20(2), 289–290.
- 734 Parchman, T. L., Benkman, C. W., & Mezquida, E. T. (2007). Coevolution between
735 hispaniolan crossbills and pine: Does more time allow for greater phenotypic
736 escalation at lower latitude? *Evolution*, 61(9), 2142–2153.
- 737 Päckert, M., Martens, J., Sun, Y.-H., Severinghaus, L. L., Nazarenko, A. A., Ting, J., ...
738 Tietze, D. T. (2012). Horizontal and elevational phylogeographic patterns of
739 himalayan and southeast asian forest passerines (aves: Passeriformes). *Journal of
740 Biogeography*, 39(3), 556–573.
- 741 Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., FitzJohn, R. G.,
742 ... Harmon, L. J. (2014). Geiger v2. 0: An expanded suite of methods for fitting
743 macroevolutionary models to phylogenetic trees. *Bioinformatics*, 30(15), 2216–2218.
- 744 Posadas, P., Crisci, J. V., & Katinas, L. (2006). Historical biogeography: A review of its
745 basic concepts and critical issues. *Journal of Arid Environments*, 66(3), 389–403.
- 746 Powell, A. F., Barker, F. K., Lanyon, S. M., Burns, K. J., Klicka, J., & Lovette, I. J. (2014).
747 A comprehensive species-level molecular phylogeny of the new world blackbirds
748 (icteridae). *Molecular Phylogenetics and Evolution*, 71, 94–112.

- 749 Price, T. D., Hooper, D. M., Buchanan, C. D., Johansson, U. S., Tietze, D. T., Alström, P.,
750 ... others. (2014). Niche filling slows the diversification of himalayan songbirds.
751 *Nature*, 509(7499), 222.
- 752 Pulgarín-R, P. C., Smith, B. T., Bryson Jr, R. W., Spellman, G. M., & Klicka, J. (2013).
753 Multilocus phylogeny and biogeography of the new world pheucticus grosbeaks (aves:
754 Cardinalidae). *Molecular Phylogenetics and Evolution*, 69(3), 1222–1227.
- 755 Rabosky, D. L. (2015). No substitute for real data: A cautionary note on the use of
756 phylogenies from birth–death polytomy resolvers for downstream comparative
757 analyses. *Evolution*, 69(12), 3207–3216.
- 758 Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., ... others.
759 (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*,
760 559(7714), 392.
- 761 Ramshaw, J., Richardson, D., Meatyard, B., Brown, R., Richardson, M., Thompson, E., &
762 Boulter, D. (1972). The time of origin of the flowering plants determined by using
763 amino acid sequence data of cytochrome c. *New Phytologist*, 71(5), 773–779.
- 764 Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The barcode of life data system
765 (<http://www.Barcodinglife.Org>). *Molecular Ecology Notes*, 7(3), 355–364.
- 766 R Core Team. (2018). *R: a language and environment for statistical computing*. Vienna,
767 Austria: R Foundation for Statistical Computing.
- 768 Rees, & Cranston, K. (2017). Automated assembly of a reference taxonomy for phylogenetic
769 data synthesis. *Biodiversity Data Journal*, (5).
- 770 Rees, Vandepitte, L., Decock, W., & Vanhoorne, B. (2017). IRMNG 2006–2016: 10 Years of
771 a Global Taxonomic Database. *Biodiversity Informatics*, 12.
- 772 Revell, L. J. (2012). Phytools: An r package for phylogenetic comparative biology (and other
773 things). *Methods in Ecology and Evolution*, 3, 217–223.
- 774 Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference
775 under mixed models. *Bioinformatics*, 19(12), 1572–1574.

- 776 https://doi.org/10.1093/bioinformatics/btg180
- 777 Roquet, C., Lavergne, S., & Thuiller, W. (2014). One tree to link them all: A phylogenetic
778 dataset for the european tetrapoda. *PLoS Currents*, 6.
- 779 Sanchez-Reyes, L. L., & O'Meara, B. (2022). *datelifeplot*: Methods to plot chronograms
780 and outputs of the *datelife* package. *R Package Release V0.2.2*. Retrieved from
781 <https://zenodo.org/badge/latestdoi/381501451>
- 782 Sanchez-Reyes, L. L., O'Meara, B., Eastman, J., Heath, T., Wright, A., Schliep, K., ...
783 Alfaro, M. (2022). *datelife*: Scientific Data on Time of Lineage Divergence for Your
784 Taxa. *R Package Release V0.6.2*. Retrieved from
785 <https://doi.org/10.5281/zenodo.593938>
- 786 Sanderson, M. (2002). Estimating Absolute Rates of Molecular Evolution and Divergence
787 Times: A Penalized Likelihood Approach. *Molecular Biology and Evolution*, 19(1),
788 101–109. <https://doi.org/10.1093/oxfordjournals.molbev.a003974>
- 789 Sanderson, M. (2003). r8s: Inferring Absolute Rates of Molecular Evolution and Divergence
790 Times in the Absence of a Molecular Clock. *Bioinformatics*, 19(2), 301–302.
- 791 Sanderson, M., & Doyle, J. (2001). Sources of error and confidence intervals in estimating
792 the age of angiosperms from rbcL and 18S rDNA data. *American Journal of Botany*,
793 88(8), 1499–1516.
- 794 Sauquet, H. (2013). A practical guide to molecular dating. *Comptes Rendus Palevol*, 12(6),
795 355–367.
- 796 Sauquet, H., Ramírez-Barahona, S., & Magallón, S. (2021). *The age of flowering plants is
797 unknown*.
- 798 Schenk, J. J. (2016). Consequences of secondary calibrations on divergence time estimates.
799 *PLoS ONE*, 11(1). <https://doi.org/10.1371/journal.pone.0148228>
- 800 Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in r. *Bioinformatics*, 27(4), 592–593.
- 801 Schoch, C. L., Ciufo, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R., ...
802 others. (2020). NCBI Taxonomy: a Comprehensive Update on Curation, Resources

- 803 and Tools. *Database*, 2020.
- 804 Selvatti, A. P., Gonzaga, L. P., & Moraes Russo, C. A. de. (2015). A paleogene origin for
805 crown passerines and the diversification of the oscines in the new world. *Molecular
806 Phylogenetics and Evolution*, 88, 1–15.
- 807 Smith, S., & Brown, J. (2018). Constructing a broadly inclusive seed plant phylogeny.
808 *American Journal of Botany*, 105(3), 302–314.
- 809 Smith, S., & O'Meara, B. (2012). TreePL: Divergence time estimation using penalized
810 likelihood for large phylogenies. *Bioinformatics*, 28(20), 2689–2690.
811 <https://doi.org/10.1093/bioinformatics/bts492>
- 812 Stoltzfus, A., Lapp, H., Matasci, N., Deus, H., Sidlauskas, B., Zmasek, C. M., ... Jordan, G.
813 (2013). Phylotastic! Making tree-of-life knowledge accessible, reusable and convenient.
814 *BMC Bioinformatics*, 14. <https://doi.org/10.1186/1471-2105-14-158>
- 815 Thomas, G. H., Hartmann, K., Jetz, W., Joy, J. B., Mimoto, A., & Mooers, A. O. (2013).
816 PASTIS: An r package to facilitate phylogenetic assembly with soft taxonomic
817 inferences. *Methods in Ecology and Evolution*, 4(11), 1011–1017.
- 818 Tietze, D. T., Päckert, M., Martens, J., Lehmann, H., & Sun, Y.-H. (2013). Complete
819 phylogeny and historical biogeography of true rosefinches (aves: Carpodacus).
820 *Zoological Journal of the Linnean Society*, 169(1), 215–234.
- 821 Title, P. O., & Rabosky, D. L. (2016). Do Macrophylogenies Yield Stable Macroevolutionary
822 Inferences? An Example from Squamate Reptiles. *Systematic Biology*, syw102.
823 <https://doi.org/10.1093/sysbio/syw102>
- 824 Treplin, S., Siegert, R., Bleidorn, C., Thompson, H. S., Fotso, R., & Tiedemann, R. (2008).
825 Molecular phylogeny of songbirds (aves: Passeriformes) and the relative utility of
826 common nuclear marker loci. *Cladistics*, 24(3), 328–349.
- 827 Uyeda, J. C., Pennell, M. W., Miller, E. T., Maia, R., & McClain, C. R. (2017). The
828 evolution of energetic scaling across the vertebrate tree of life. *The American
829 Naturalist*, 190(2), 185–199.

- 830 Vos, R. A., Balhoff, J. P., Caravas, J. A., Holder, M. T., Lapp, H., Maddison, W. P., ...
831 others. (2012). NeXML: Rich, extensible, and verifiable representation of
832 comparative data and metadata. *Systematic Biology*, 61(4), 675–689.
833 <https://doi.org/10.1093/sysbio/sys025>
- 834 Webb, C. (2000). Exploring the Phylogenetic Structure of Ecological Communities : An
835 Example for Rain Forest Trees. *The American Naturalist*, 156(2), 145–155.
- 836 Webb, C., Ackerly, D., & Kembel, S. (2008). Phylocom: Software for the analysis of
837 phylogenetic community structure and trait evolution. *Bioinformatics*, 24(18),
838 2098–2100. <https://doi.org/10.1093/bioinformatics/btn358>
- 839 Webb, C., & Donoghue, M. (2005). Phylomatic: Tree assembly for applied phylogenetics.
840 *Molecular Ecology Notes*, 5(1), 181–183.
- 841 Weir, J., & Schluter, D. (2008). Calibrating the avian molecular clock. *Molecular Ecology*,
842 17(10), 2321–2328.
- 843 Zuccon, D., Prŷs-Jones, R., Rasmussen, P. C., & Ericson, P. G. (2012). The phylogenetic
844 relationships and generic limits of finches (fringillidae). *Molecular Phylogenetics and*
845 *Evolution*, 62(2), 581–596.