

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: 4960

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

46 **Introduction**

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

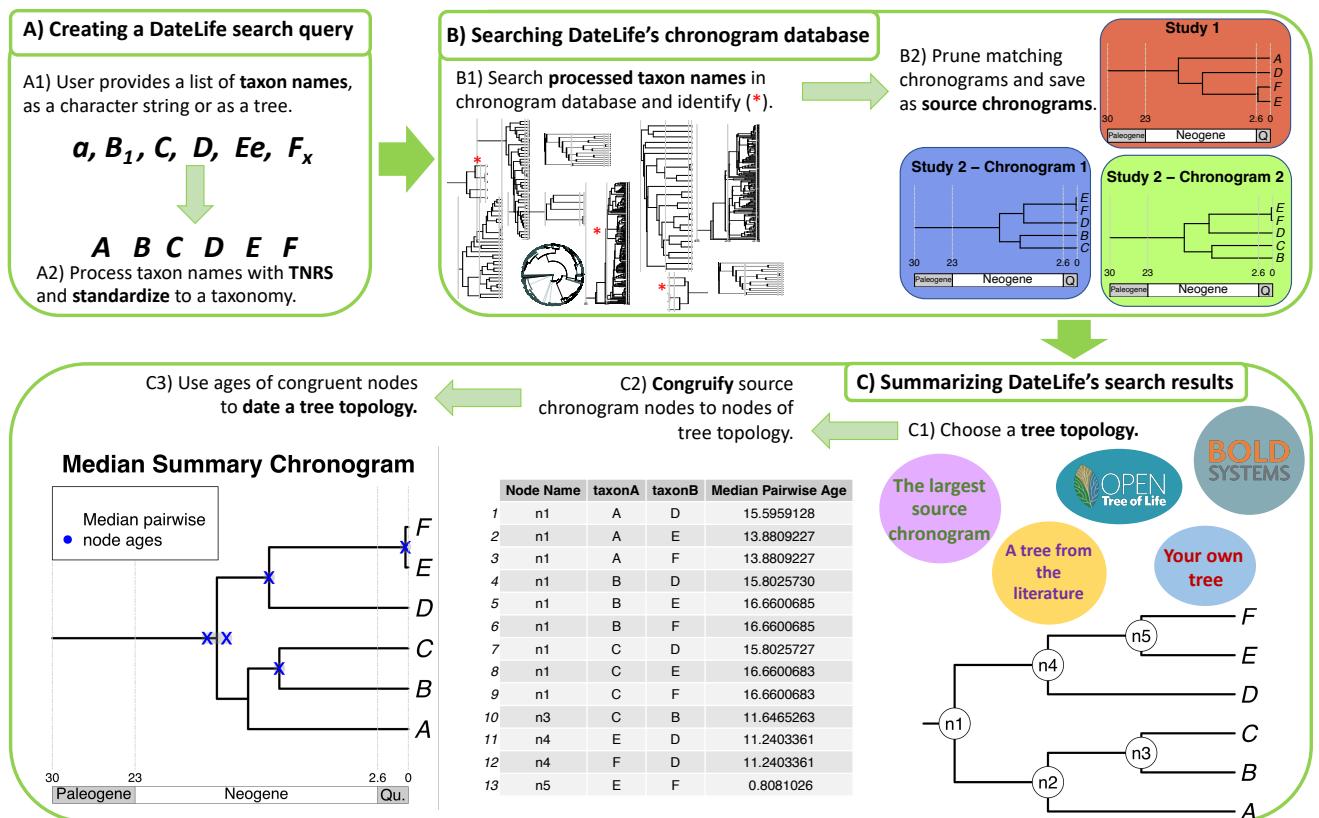
53 Building a chronogram is not an easy task. It requires obtaining and curating data to
54 construct a phylogeny, selecting and placing appropriate calibrations on the phylogeny using
55 independent age data points from the fossil record or other dated events, and inferring the
56 full dated tree; it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

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

68

Description

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



81

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

82 Creating a search query

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

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

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

107 The processed input taxon names are saved as an R object of a newly defined class,

108 `datelifeQuery`, that is used in the following steps. This object contains the standardized

109 names, the corresponding OTT identification numbers, and the topology of the input tree if

110 any was provided.

111 **Searching a chronogram database**

112 At the time of writing of this manuscript (Jun 22, 2022), DateLife's chronogram

113 database latest version consist of 253 chronograms published in 187 different studies. It is

114 curated from OpenTree's phylogenetic database, the Phylesystem, which constitutes an open

115 source of expert and peer-reviewed phylogenetic knowledge with rich metadata (McTavish et

116 al., 2015), which allows automatic and reproducible assembly of our chronogram database.

117 Datelife's chronogram database is navigable as an R data object within the `datelife` R

118 package.

119 A unique feature of the Phylesystem is that any user can add new published,

120 state-of-the-art chronograms any time, through their curator application

121 (<https://tree.opentreeoflife.org/curator>). As chronograms are added to Phylesystem, they are

122 incorporated into the chronogram database of the `datelife` package. The updated database

123 is assigned a new version number, followed by a package release on CRAN. `datelife`'s

124 chronogram database is updated as new chronogram data is added to Phylesystem, at a

125 minimum of once a month and a maximum of every 6 months. Users can also implement

126 functions from the `datelife` R package to trigger an update of the local chronogram

127 database, to incorporate any new chronograms to the user's DateLife analysis before an

128 official database update is released on CRAN.

129 A DateLife search is implemented by matching processed taxon names provided by the

130 user to tip labels in the chronogram database. Chronograms with at least two matching

131 taxon names on their tip labels are identified and pruned down to preserve only the matched

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

139 **Summarizing search results**

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

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

153 ***Choosing a topology.***— DateLife requires a tree topology to summarize age data
154 upon. We recommend that users provide a tree topology as input from the literature, or one
155 of their own making. If no topology is provided, DateLife automatically extracts one from
156 the OpenTree synthetic tree, a phylogeny 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 combine topologies from source chronograms into a single summary (or supertree) topology, the DateLife workflow 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). 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. If source chronograms do not form a grove, the supertree reconstruction will fail.

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

Once a topology is chosen, DateLife applies the congruification method (Eastman, Harmon, & Tank, 2013) to find nodes belonging to the same clade across source chronograms, and extract the corresponding node ages from the patristic distance matrices stored as `datelifeResult`. By definition, the matrices store total distance (time from tip to tip), hence, node ages correspond to half the values stored in the patristic distance matrices. This assumes that the terminal taxa are coeval and occur at the present. A table of

183 congruified node ages that can be used as calibrations for a dating analysis is stored as a
184 `congruifiedCalibrations` object.

185 For each congruent node, the pairwise distances that traverse that node are
186 summarized into a single summary matrix using classic summary statistics (i.e., mean,
187 median, minimum and maximum ages), and the Supermatrix Distance Method (SDM;
188 Criscuolo, Berry, Douzery, & Gascuel, 2006), which deforms patristic distance matrices by
189 minimizing variance and then averaging them. These single summary taxon pair age
190 matrices (Summarized calibrations) can be applied as calibrations to date a tree topology,
191 using different dating methods currently supported within DateLife: MrBayes (Huelsenbeck
192 & Ronquist, 2001; Ronquist & Huelsenbeck, 2003), PATHd8 (Britton, Anderson, Jacquet,
193 Lundqvist, & Bremer, 2007), BLADJ (Webb, Ackerly, & Kembel, 2008; Webb & Donoghue,
194 2005), and treePL (Smith & O'Meara, 2012).

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

205 Alternative phylogenetic dating options supported in DateLife (MrBayes, PATHD8,
206 TreePL) incorporate branch length information from the input topology in combination with
207 the calibrations. PATHd8 is a non-clock, rate-smoothing method (Britton et al., 2007) to
208 date trees. treePL (Smith & O'Meara, 2012), is a semi-parametric, rate-smoothing,

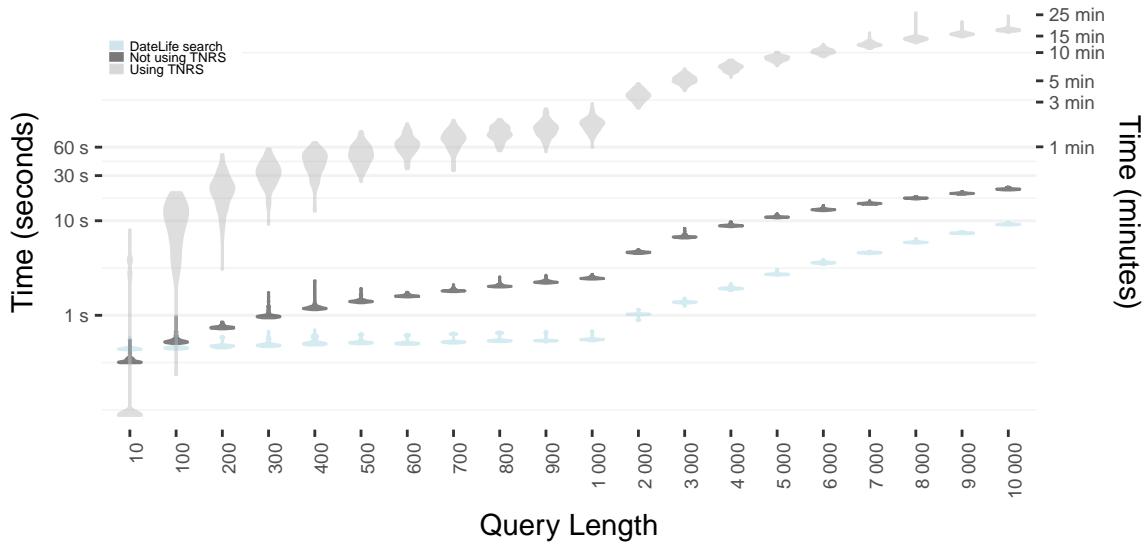
209 penalized likelihood dating method (Sanderson, 2002). The MrBayes (Huelsenbeck &
210 Ronquist, 2001; Ronquist & Huelsenbeck, 2003) approach in DateLife uses the calibrations as
211 priors on node ages.

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

217

Benchmark

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



227

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

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

242 **Case studies**

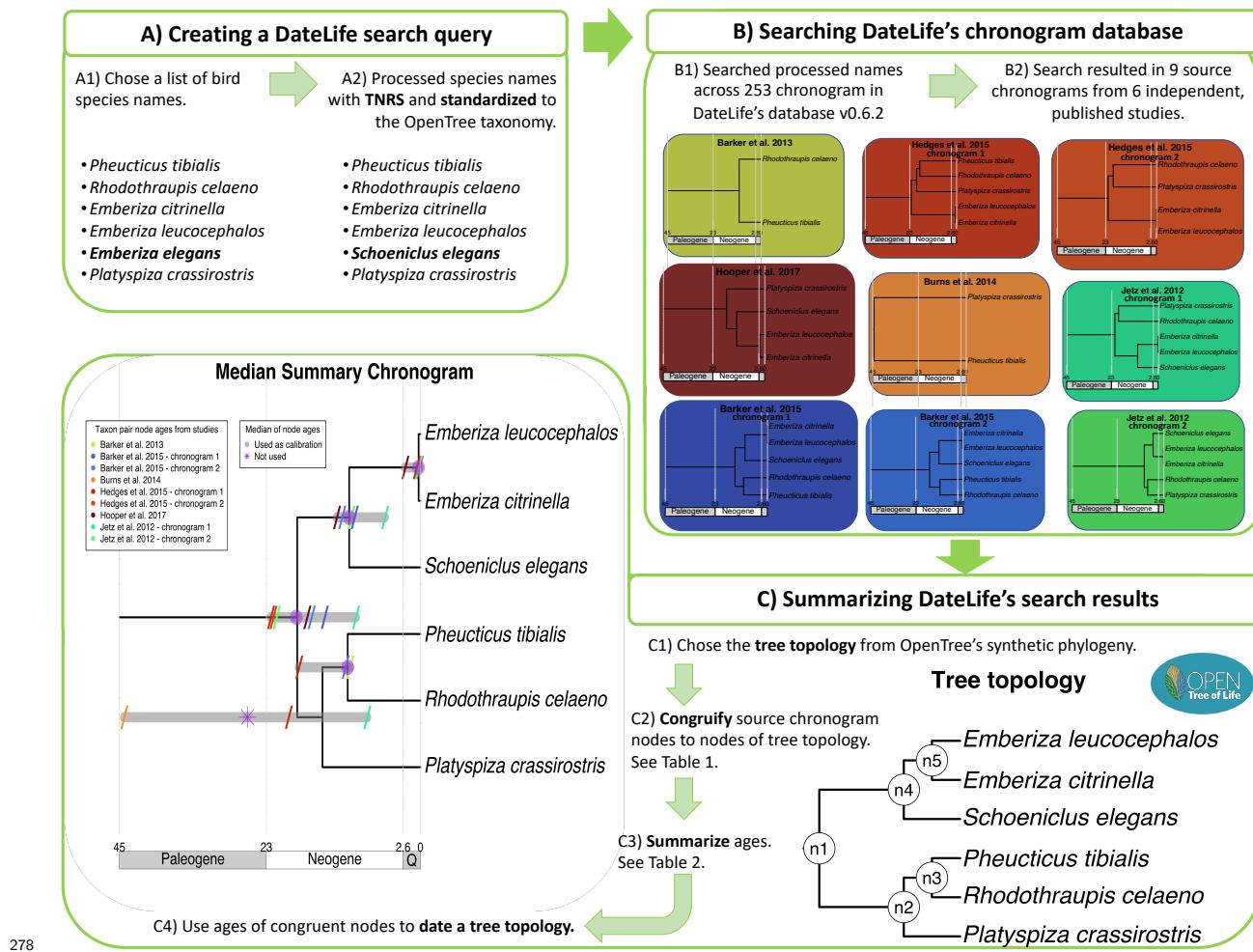
243 We illustrate the DateLife workflow using a family within the Passeriform birds
244 encompassing the true finches, Fringillidae, as case study. On a small example, we analysed 6
245 bird species, and results from each step of the workflow are shown in Fig. 3. As a second
246 example, we analysed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

249 **A small example**

250 **Creating a search query.** We chose 6 bird species within the Passeriformes. The sample
251 includes two species of cardinals: the black-thighed grosbeak – *Pheucticus tibialis* and the
252 crimson-collared grosbeak – *Rhodothraupis celaeno*; three species of buntings: the
253 yellowhammer – *Emberiza citrinella*, the pine bunting – *Emberiza leucocephalos* and the
254 yellow-throated bunting – *Emberiza elegans*; and one species of tanager, the vegetarian finch –
255 *Platyspiza crassirostris*. Processing of input names found that *Emberiza elegans* is synonym
256 for *Schoeniclus elegans* in the default reference taxonomy (OTT v3.3, June 1, 2021). For a
257 detailed discussion on the state of the synonym, refer to Avibase (Avibase, 2022; Lepage,
258 2004; Lepage, Vaidya, & Guralnick, 2014). Discovering this synonym allowed assigning five
259 age data points for the parent node of *Emberiza elegans*, shown as *Schoeniclus elegans* in
260 figure 3A, which would not have had any data otherwise.

261 **Searching the database.** DateLife used the processed input names to search the local
262 chronogram database and found 9 matching chronograms in 6 different studies (Fig. 3B).
263 Three studies matched five input names (Barker, Burns, Klicka, Lanyon, & Lovette, 2015;
264 Hedges, Marin, Suleski, Paymer, & Kumar, 2015; Jetz, Thomas, Joy, Hartmann, & Mooers,
265 2012), one study matched four input names (Hooper & Price, 2017) and two studies matched
266 two input names (Barker, Burns, Klicka, Lanyon, & Lovette, 2013; Burns et al., 2014). No
267 studies matched all input names. Together, source chronograms provide 28 unique age data
268 points, covering all nodes on our chosen tree topology to date (Table 1).

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



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

286 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

287

288 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

289

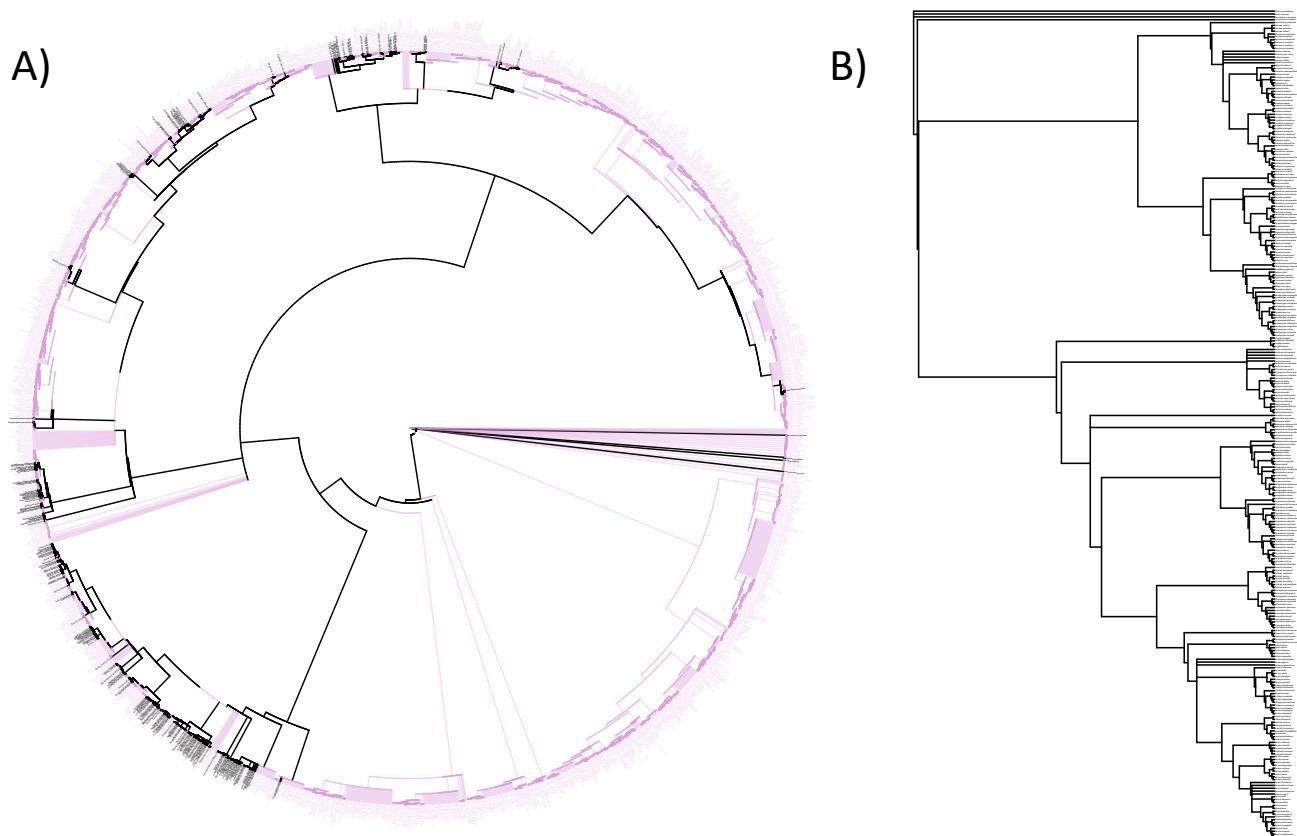
290 **An example with the family of true finches**

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

300 **Searching the database.** Next, we used the processed species names in our DateLife
301 query to identify chronograms with at least two Fringillidae species. The DateLife search
302 identified 13 chronograms containing at least two Fringillidae species, published in 9 different
303 studies (Barker et al., 2013, 2015; Burns et al., 2014; Claramunt & Cracraft, 2015; Gibb et
304 al., 2015; Hedges et al., 2015; Hooper & Price, 2017; Jetz et al., 2012; Price et al., 2014).

305 Once identified, DateLife pruned matching chronograms to keep Fringillidae species names
306 on tips only, and transformed these pruned chronograms to pairwise distance matrices,
307 revealing 1, 206 different age data points available for species within the Fringillidae
(Supplementray Table S1).

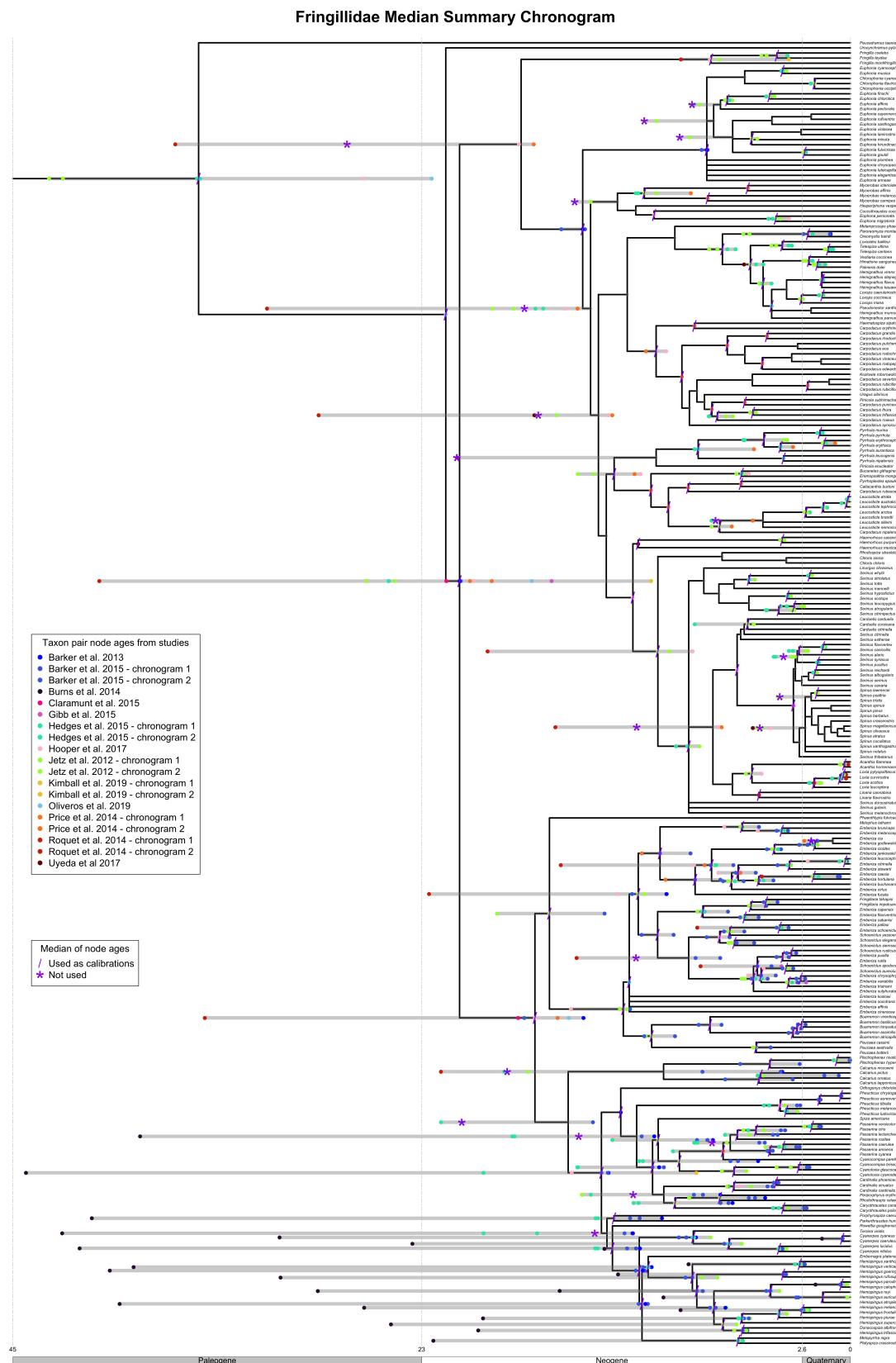
308 **Summarizing search results.** The final step is to congruify and summarize the age data
309 available for the Fringillidae species into single summary chronograms, using different types
310 of summary ages, median and SDM. As explained in the “Description” section, a tree
311 topology to summarize age data upon is required. By default, DateLife uses the topology
312 from OpenTree’s synthetic tree that contains the species in the search query to summarize
313 age data upon. According to OpenTree’s synthetic tree, species belonging to the family
314 Fringillidae do not form a monophyletic group (Fig. 4).



315 FIGURE 4. Tree topologies extracted from Open Tree of Life's (OpenTree) synthetic phylogenetic tree.

316 A) Topology of 2,333 tips and 1,305 internal nodes, encompassing bird species within the family
 317 Fringillidae following the NCBI taxonomy (black), as well as all other bird species that share the same
 318 Most Recent Common Ancestor (MRCA) node in OpenTree's synthetic tree (purple). B) Subtree
 319 topology of 289 tips and 253 internal nodes, resulting from pruning species that do not belong to the
 320 family Fringillidae according to the NCBI taxonomy (purple branches in topology A). Bird species
 321 within the Fringillidae are paraphyletic (Alström et al. 2014, Barker, Cibois, Schikler, Feinstein, &
 322 Cracraft 2004, Barker et al. 2013, Barker 2014, Barker et al. 2015, Beresford, Barker, Ryan, &
 323 Crowe 2005, Bryson Jr et al. 2014, Burleigh, Kimball, & Braun 2015, Burns et al. 2014, Chaves,
 324 Hidalgo, & Klicka 2013, Claramunt & Cracraft 2015, Gibb et al. 2015, Hackett et al. 2008, Jetz et al.
 325 2012, Johansson, Fjeldså, & Bowi 200, Kimball et al. 2019, Klicka et al. 2014, Lamichhaney et al.
 326 2015, Lerner, Meyer, James, Hofreiter, & Fleischer 2011, Lovette et al. 2010, Moyle et al. 2016,
 327 Ödeen, Håstad, & Alström 2011, Oliveros et al. 2019, Päckert et al. 2012, Parchman, Benkman, &
 328 Mezquida 2007, Powell et al. 2014, Price et al. 2014, Pulgarín-R, Smith, Bryson Jr, Spellman, &
 329 Klicka 2013, Selvatti, Gonzaga, & Moraes Russo 2015, Tietze, Päckert, Martens, Lehmann, & Sun
 330 2013, Treplin et al. 2008, Zuccon, Prŷs-Jones, Rasmussen, & Ericson 2012).

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

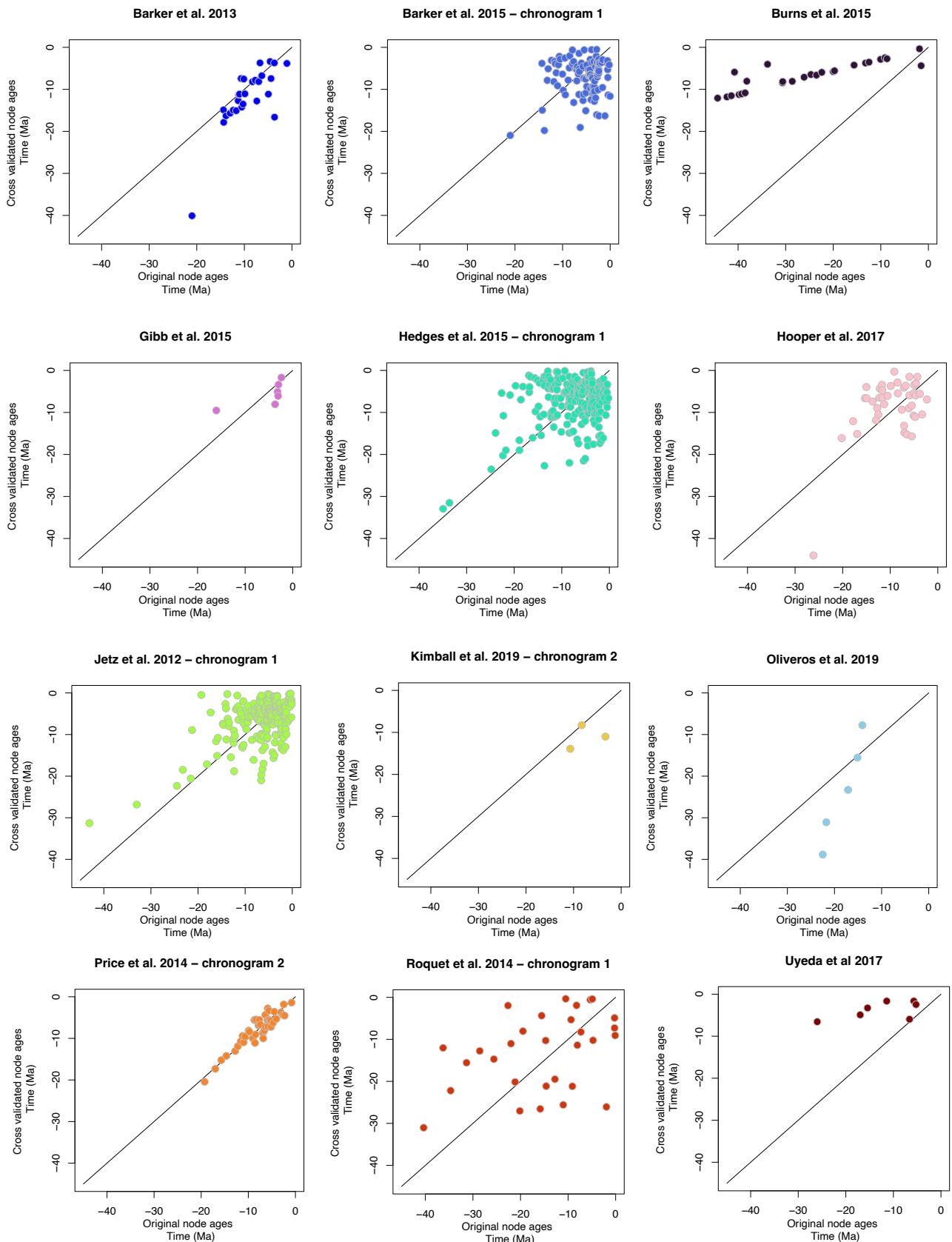


337 FIGURE 5. *Fringillidae* median summary chronogram generated with DateLife. It has 256 tips and
 338 233 nodes, from which 212 have age data from at least one published chronogram.

340

Cross-validation test

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



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

354

Discussion

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

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

369 **Age variation in source chronograms**

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

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

388 **Primary vs Secondary calibrations.** While most chronograms in DateLife's database
389 are constructed using primary calibrations (molecular substitution rates or ages obtained
390 from the fossil record or geological events), DateLife summarizes chronograms using
391 secondary calibrations (ages coming from other chronograms). Graur and Martin (2004)
392 cautioned on the increased error and uncertainty in estimated ages when using secondary
393 calibrations in dating analyses. Schenk (2016) showed that, in simulations, divergence times
394 inferred using secondary calibrations are significantly younger than those inferred with
395 primary calibrations, when obtained with Bayesian inference methods, and when priors are
396 implemented in similar ways in both analyses. Accordingly, the scientific community seems
397 to have more confidence in chronograms obtained from a single analysis, using fossil data as
398 primary sources of calibrations (Schenk, 2016), and using fossils that have been widely
399 discussed and curated as calibrations to date other trees, making sure that all data reflect a
400 coherent evolutionary history (Sauquet, 2013), as for example done by Antonelli et al.
401 (2017). There have been attempts to create fossil calibration databases (Ksepka et al., 2015),
402 though these still have room to grow.

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

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

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

420 Sumarizing chronograms

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

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

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

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

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

449 Effects on downstream analyses

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

453 In ecology and conservation biology, incorporating at least some data on lineage divergence
454 times represents a relevant improvement for testing alternative hypothesis using phylogenetic
455 distance (Webb et al., 2008). Hence, DateLife's workflow features different ways of
456 estimating node ages in the absence of calibrations and branch length information for certain
457 taxa. “Making up” branch lengths is a common practice in scientific publications: Jetz et al.
458 (2012), created a chronogram of all 9, 993 bird species, where 67% had molecular data and

459 the rest was simulated; Rabosky et al. (2018) created a chronogram of 31, 536 ray-finned
460 fishes, of which only 37% had molecular data; Smith and Brown (2018) constructed a
461 chronogram of 353, 185 seed plants where only 23% had molecular data.
462 Notably, risks come with this practice. Taken to the extreme, one could make a fully
463 resolved, calibrated tree of all modern and extinct taxa using a single taxonomy and a single
464 calibration, using polytomy resolution and branch estimation methods. There has yet to be
465 a thorough analysis of what can go wrong when one extends inferences beyond the data in
466 this way, so we urge caution; we also urge readers to follow the example of the large tree
467 papers cited above, by carefully considering the statistical assumptions being made, and
468 assessing the consistency of the results with prior work.

469 Conclusions

470 Knowledge of the evolutionary time frame of organisms is key to many research areas: trait
471 evolution, species diversification, biogeography, macroecology and more. It is also crucial for
472 education, science communication and policy, but generating chronograms is difficult,
473 especially for those who want to use phylogenies but who are not systematists, or do not
474 have the time to acquire and develop the necessary knowledge and skills to construct them
475 on their own. Importantly, years of primarily public funded research have resulted in vast
476 amounts of chronograms that are already available on scientific publications, but hidden to
477 the public and scientific community for reuse.
478 The DateLife project allows for easy and fast summary of public and state-of-the-art data on
479 time of lineage divergence. It provides a straightforward way to get an informed idea on the
480 state of knowledge of the time frame of evolution of different regions of the tree of life, and
481 allows identification of regions that require more research, or that have conflicting
482 information. It is available as an R package, and as a web-based R shiny application at
483 www.datelife.org Both summary and newly generated trees are useful to evaluate
484 evolutionary hypotheses in different areas of research. The DateLife project helps with

485 awareness of the existing variation in expert time of divergence data, and will foster
486 exploration of the effect of alternative divergence time hypothesis on the results of analyses,
487 nurturing a culture of more cautious interpretation of evolutionary results.

488 **Availability**

489 The DateLife software is free and open source and it can be used through its R shiny web
490 application at <http://www.datelife.org>, through the `datelife` R package, and through
491 Phylotastic's project web portal <https://phylo.cs.nmsu.edu/>. DateLife's web application is
492 maintained using RStudio's shiny server and the shiny package open infrastructure, as well
493 as Docker and OpenTree's infrastructure (dates.opentreeoflife.org/datelife). `datelife`'s R
494 package stable version is available for installation from the CRAN repository
495 (<https://cran.r-project.org/package=datelife>) using the command `install.packages(pkgs`
496 `= "datelife"`) from within R. Development versions are available from the GitHub
497 repository (<https://github.com/phylotastic/datelife>) and can be installed using the
498 command `devtools::install_github("phylotastic/datelife")`.

499 **Supplementary Material**

500 Code used to generate all versions of this manuscript, the biological examples, as well as the
501 benchmark of functionalities are available at `datelifeMS1`, `datelife_examples`, and
502 `datelife_benchmark` repositories in LLSR's GitHub account.

503 **Funding**

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

507

Acknowledgements

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

520

References

- 521 Alström, P., Hooper, D. M., Liu, Y., Olsson, U., Mohan, D., Gelang, M., ... Price, T. D.
522 (2014). Discovery of a relict lineage and monotypic family of passerine birds. *Biology
Letters*, 10(3), 20131067.
- 523
- 524 Ané, C., Eulenstein, O., Piaggio-Talice, R., & Sanderson, M. J. (2009). Groves of
525 phylogenetic trees. *Annals of Combinatorics*, 13(2), 139–167.
- 526 Antonelli, A., Hettling, H., Condamine, F. L., Vos, K., Nilsson, R. H., Sanderson, M. J., ...
527 Vos, R. A. (2017). Toward a self-updating platform for estimating rates of speciation
528 and migration, ages, and relationships of Taxa. *Systematic Biology*, 66(2), 153–166.
529 <https://doi.org/10.1093/sysbio/syw066>
- 530 Archie, J., Day, W. H., Felsenstein, J., Maddison, W., Meacham, C., Rohlf, F. J., &
531 Swofford, D. (1986). The Newick tree format. Retrieved from
532 {<https://evolution.genetics.washington.edu/phylip/newicktree.html>}
- 533 Avibase. (2022). Yellow-throated Bunting. *Avibase - the World Bird Database*, (Online
534 Resource). Retrieved from
535 {<https://avibase.bsc-eoc.org/species.jsp?lang=EN&avibaseid=82D1EE0049D8D927>}
- 536 Bapst, D. W. (2012). Paleotree: An R package for paleontological and phylogenetic analyses
537 of evolution. *Methods in Ecology and Evolution*, 3(5), 803–807.
538 <https://doi.org/10.1111/j.2041-210X.2012.00223.x>
- 539 Barba-Montoya, J., Reis, M. dos, Schneider, H., Donoghue, P. C., & Yang, Z. (2018).
540 Constraining uncertainty in the timescale of angiosperm evolution and the veracity of
541 a cretaceous terrestrial revolution. *New Phytologist*, 218(2), 819–834.
- 542 Barker, F. K. (2014). Mitogenomic data resolve basal relationships among passeriform and
543 passeridan birds. *Molecular Phylogenetics and Evolution*, 79, 313–324.
- 544 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2013). Going to
545 extremes: Contrasting rates of diversification in a recent radiation of new world
546 passerine birds. *Systematic Biology*, 62(2), 298–320.

- 547 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2015). New insights
548 into new world biogeography: An integrated view from the phylogeny of blackbirds,
549 cardinals, sparrows, tanagers, warblers, and allies. *The Auk: Ornithological Advances*,
550 132(2), 333–348.
- 551 Barker, F. K., Cibois, A., Schikler, P., Feinstein, J., & Cracraft, J. (2004). Phylogeny and
552 diversification of the largest avian radiation. *Proceedings of the National Academy of
553 Sciences*, 101(30), 11040–11045.
- 554 Beresford, P., Barker, F., Ryan, P., & Crowe, T. (2005). African endemics span the tree of
555 songbirds (passeri): Molecular systematics of several evolutionary “enigmas”.
556 *Proceedings of the Royal Society B: Biological Sciences*, 272(1565), 849–858.
- 557 Boyle, B., Hopkins, N., Lu, Z., Raygoza Garay, J. A., Mozzherin, D., Rees, T., ... Enquist,
558 B. J. (2013). The taxonomic name resolution service: An online tool for automated
559 standardization of plant names. *BMC Bioinformatics*, 14(1).
560 <https://doi.org/10.1186/1471-2105-14-16>
- 561 Britton, T., Anderson, C. L., Jacquet, D., Lundqvist, S., & Bremer, K. (2007). Estimating
562 Divergence Times in Large Phylogenetic Trees. *Systematic Biology*, 56(788777878),
563 741–752. <https://doi.org/10.1080/10635150701613783>
- 564 Bryson Jr, R. W., Chaves, J., Smith, B. T., Miller, M. J., Winker, K., Pérez-Emán, J. L., &
565 Klicka, J. (2014). Diversification across the new world within the ‘blue’cardinalids
566 (aves: Cardinalidae). *Journal of Biogeography*, 41(3), 587–599.
- 567 Burleigh, J. G., Kimball, R. T., & Braun, E. L. (2015). Building the avian tree of life using a
568 large-scale, sparse supermatrix. *Molecular Phylogenetics and Evolution*, 84, 53–63.
- 569 Burns, K. J., Shultz, A. J., Title, P. O., Mason, N. A., Barker, F. K., Klicka, J., ... Lovette,
570 I. J. (2014). Phylogenetics and diversification of tanagers (passeriformes:
571 Thraupidae), the largest radiation of neotropical songbirds. *Molecular Phylogenetics
572 and Evolution*, 75, 41–77.
- 573 Chamberlain, S. (2018). *bold: Interface to Bold Systems API*. Retrieved from

- 574 <https://CRAN.R-project.org/package=bold>
- 575 Chamberlain, S. A., & Szöcs, E. (2013). taxize : taxonomic search and retrieval in R [version
576 2; referees: 3 approved]. *F1000Research*, 2(191), 1–29.
577 <https://doi.org/10.12688/f1000research.2-191.v2>
- 578 Chaves, J. A., Hidalgo, J. R., & Klicka, J. (2013). Biogeography and evolutionary history of
579 the neotropical genus *saltator* (aves: Thraupini). *Journal of Biogeography*, 40(11),
580 2180–2190.
- 581 Claramunt, S., & Cracraft, J. (2015). A new time tree reveals earth history's imprint on the
582 evolution of modern birds. *Science Advances*, 1(11), e1501005.
- 583 Criscuolo, A., Berry, V., Douzery, E. J., & Gascuel, O. (2006). SDM: A fast distance-based
584 approach for (super)tree building in phylogenomics. *Systematic Biology*, 55(5),
585 740–755. <https://doi.org/10.1080/10635150600969872>
- 586 Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M.-K., Turon, X., ... Douzery,
587 E. J. (2018). A phylogenomic framework and timescale for comparative studies of
588 tunicates. *BMC Biology*, 16(1), 1–14.
- 589 Eastman, J. M., Harmon, L. J., & Tank, D. C. (2013). Congruification: Support for time
590 scaling large phylogenetic trees. *Methods in Ecology and Evolution*, 4(7), 688–691.
591 <https://doi.org/10.1111/2041-210X.12051>
- 592 Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high
593 throughput. *Nucleic Acids Research*, 32(5), 1792–1797.
- 594 Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American Naturalist*,
595 125(1), 1–15. Retrieved from <http://www.jstor.org/stable/2461605>
- 596 GBIF Secretariat. (2022). GBIF Backbone Taxonomy. *Checklist dataset*, (Online Resource
597 accessed via GBIF.org). Retrieved from {<https://doi.org/10.15468/39omei> }
- 598 Gibb, G. C., England, R., Hartig, G., McLenaghan, P. A., Taylor Smith, B. L., McComish,
599 B. J., ... Penny, D. (2015). New Zealand passerines help clarify the diversification of
600 major songbird lineages during the oligocene. *Genome Biology and Evolution*, 7(11),

- 601 2983–2995.
- 602 Graur, D., & Martin, W. (2004). Reading the entrails of chickens: Molecular timescales of
603 evolution and the illusion of precision. *TRENDS in Genetics*, 20(2), 80–86.
- 604 Hackett, S. J., Kimball, R. T., Reddy, S., Bowie, R. C., Braun, E. L., Braun, M. J., ...
605 others. (2008). A phylogenomic study of birds reveals their evolutionary history.
606 *Science*, 320(5884), 1763–1768.
- 607 Hedges, S. B., Marin, J., Suleski, M., Paymer, M., & Kumar, S. (2015). Tree of life reveals
608 clock-like speciation and diversification. *Molecular Biology and Evolution*, 32(4),
609 835–845. <https://doi.org/10.1093/molbev/msv037>
- 610 Heibl, C. (2008). *PHYLOCH: R language tree plotting tools and interfaces to diverse*
611 *phylogenetic software packages*. Retrieved from
612 <http://www.christophheibl.de/Rpackages.html>
- 613 Hooper, D. M., & Price, T. D. (2017). Chromosomal inversion differences correlate with
614 range overlap in passerine birds. *Nature Ecology & Evolution*, 1(10), 1526.
- 615 Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic
616 trees. *Bioinformatics*, 17(8), 754–755.
617 <https://doi.org/10.1093/bioinformatics/17.8.754>
- 618 Jetz, W., Thomas, G., Joy, J. J., Hartmann, K., & Mooers, A. (2012). The global diversity
619 of birds in space and time. *Nature*, 491(7424), 444–448.
620 <https://doi.org/10.1038/nature11631>
- 621 Johansson, U. S., Fjeldså, J., & Bowie, R. C. (2008). Phylogenetic relationships within
622 passerida (aves: Passeriformes): A review and a new molecular phylogeny based on
623 three nuclear intron markers. *Molecular Phylogenetics and Evolution*, 48(3), 858–876.
- 624 Katoh, K., Asimenos, G., & Toh, H. (2009). Multiple alignment of dna sequences with mafft.
625 In *Bioinformatics for dna sequence analysis* (pp. 39–64). Springer.
- 626 Kimball, R. T., Oliveros, C. H., Wang, N., White, N. D., Barker, F. K., Field, D. J., ...
627 others. (2019). A phylogenomic supertree of birds. *Diversity*, 11(7), 109.

- 628 Klicka, J., Barker, F. K., Burns, K. J., Lanyon, S. M., Lovette, I. J., Chaves, J. A., & Bryson
629 Jr, R. W. (2014). A comprehensive multilocus assessment of sparrow (aves:
630 Passerellidae) relationships. *Molecular Phylogenetics and Evolution*, 77, 177–182.
- 631 Ksepka, D. T., Parham, J. F., Allman, J. F., Benton, M. J., Carrano, M. T., Cranston, K.
632 A., . . . others. (2015). The fossil calibration database—a new resource for divergence
633 dating. *Systematic Biology*, 64(5), 853–859.
- 634 Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio,
635 A., . . . others. (2015). Evolution of darwin’s finches and their beaks revealed by
636 genome sequencing. *Nature*, 518(7539), 371–375.
- 637 Laubichler, M. D., & Maienschein, J. (2009). *Form and function in developmental evolution*.
638 Cambridge University Press.
- 639 Lepage, D. (2004). *Avibase: The world bird database*. Bird Studies Canada.
- 640 Lepage, D., Vaidya, G., & Guralnick, R. (2014). Avibase—a database system for managing
641 and organizing taxonomic concepts. *ZooKeys*, (420), 117.
- 642 Lerner, H. R., Meyer, M., James, H. F., Hofreiter, M., & Fleischer, R. C. (2011). Multilocus
643 resolution of phylogeny and timescale in the extant adaptive radiation of hawaiian
644 honeycreepers. *Current Biology*, 21(21), 1838–1844.
- 645 Lovette, I. J., Pérez-Emán, J. L., Sullivan, J. P., Banks, R. C., Fiorentino, I.,
646 Córdoba-Córdoba, S., . . . others. (2010). A comprehensive multilocus phylogeny for
647 the wood-warblers and a revised classification of the parulidae (aves). *Molecular
648 Phylogenetics and Evolution*, 57(2), 753–770.
- 649 Magallon, S., & Sanderson, M. (2001). Absolute diversification rates in angiosperm clades.
650 *Evolution*, 55(9), 1762–1780.
- 651 Magallón, S. (2010). Using fossils to break long branches in molecular dating: A comparison
652 of relaxed clocks applied to the origin of angiosperms. *Systematic Biology*, 59(4),
653 384–399.
- 654 Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., & Hernández-Hernández, T. (2015).

655 A metacalibrated time-tree documents the early rise of flowering plant phylogenetic
656 diversity. *New Phytologist*, 207(2), 437–453.

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

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

663 Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecology Letters*,
664 17(4), 508–525. <https://doi.org/10.1111/ele.12251>

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

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

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

673 Open Tree Of Life, Redelings, B., Cranston, K. A., Allman, J., Holder, M. T., & McTavish,
674 E. J. (2016). Open Tree of Life APIs v3.0. *Open Tree of Life Project*, (Online
675 Resources). Retrieved from
676 {<https://github.com/OpenTreeOfLife/germinator/wiki/Open-Tree-of-Life-Web->
677 APIs}

678 Open Tree Of Life, Redelings, B., Sánchez Reyes, L. L., Cranston, K. A., Allman, J., Holder,
679 M. T., & McTavish, E. J. (2019). Open tree of life synthetic tree v12.3. *Zenodo*.
680 Retrieved from <https://doi.org/10.5281/zenodo.3937742>

681 Ödeen, A., Håstad, O., & Alström, P. (2011). Evolution of ultraviolet vision in the largest

- 682 avian radiation-the passerines. *BMC Evolutionary Biology*, 11(1), 1–8.
- 683 684 Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20(2), 289–290.
- 685 686 Parchman, T. L., Benkman, C. W., & Mezquida, E. T. (2007). Coevolution between hispaniolan crossbills and pine: Does more time allow for greater phenotypic
- 687 escalation at lower latitude? *Evolution*, 61(9), 2142–2153.
- 688 689 Päckert, M., Martens, J., Sun, Y.-H., Severinghaus, L. L., Nazarenko, A. A., Ting, J., ... Tietze, D. T. (2012). Horizontal and elevational phylogeographic patterns of
- 690 himalayan and southeast asian forest passerines (aves: Passeriformes). *Journal of Biogeography*, 39(3), 556–573.
- 691 692 Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., FitzJohn, R. G., ... Harmon, L. J. (2014). Geiger v2. 0: An expanded suite of methods for fitting
- 693 694 macroevolutionary models to phylogenetic trees. *Bioinformatics*, 30(15), 2216–2218.
- 695 696 Posadas, P., Crisci, J. V., & Katinas, L. (2006). Historical biogeography: A review of its basic concepts and critical issues. *Journal of Arid Environments*, 66(3), 389–403.
- 697 698 Powell, A. F., Barker, F. K., Lanyon, S. M., Burns, K. J., Klicka, J., & Lovette, I. J. (2014). A comprehensive species-level molecular phylogeny of the new world blackbirds
- 699 (icteridae). *Molecular Phylogenetics and Evolution*, 71, 94–112.
- 700 701 Price, T. D., Hooper, D. M., Buchanan, C. D., Johansson, U. S., Tietze, D. T., Alström, P., ... others. (2014). Niche filling slows the diversification of himalayan songbirds.
- 702 *Nature*, 509(7499), 222.
- 703 704 Pulgarín-R, P. C., Smith, B. T., Bryson Jr, R. W., Spellman, G. M., & Klicka, J. (2013). Multilocus phylogeny and biogeography of the new world pheucticus grosbeaks (aves:
- 705 Cardinalidae). *Molecular Phylogenetics and Evolution*, 69(3), 1222–1227.
- 706 707 Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., ... others. (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*,
- 708 559(7714), 392.

- 709 Ramshaw, J., Richardson, D., Meatyard, B., Brown, R., Richardson, M., Thompson, E., &
710 Boulter, D. (1972). The time of origin of the flowering plants determined by using
711 amino acid sequence data of cytochrome c. *New Phytologist*, 71(5), 773–779.
- 712 Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The barcode of life data system
713 (<http://www.Barcodinglife.Org>). *Molecular Ecology Notes*, 7(3), 355–364.
- 714 R Core Team. (2018). *R: a language and environment for statistical computing*. Vienna,
715 Austria: R Foundation for Statistical Computing.
- 716 Rees, & Cranston, K. (2017). Automated assembly of a reference taxonomy for phylogenetic
717 data synthesis. *Biodiversity Data Journal*, (5).
- 718 Rees, Vandepitte, L., Decock, W., & Vanhoorne, B. (2017). IRMNG 2006–2016: 10 Years of
719 a Global Taxonomic Database. *Biodiversity Informatics*, 12.
- 720 Revell, L. J. (2012). Phytools: An r package for phylogenetic comparative biology (and other
721 things). *Methods in Ecology and Evolution*, 3, 217–223.
- 722 Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference
723 under mixed models. *Bioinformatics*, 19(12), 1572–1574.
724 <https://doi.org/10.1093/bioinformatics/btg180>
- 725 Sanchez-Reyes, L. L., & O'Meara, B. (2022). **datelifeplot**: Methods to plot chronograms
726 and outputs of the **datelife** package. *R Package Release V0.2.2*. Retrieved from
727 <https://zenodo.org/badge/latestdoi/381501451>
- 728 Sanchez-Reyes, L. L., O'Meara, B., Eastman, J., Heath, T., Wright, A., Schliep, K., ...
729 Alfaro, M. (2022). **datelife**: Scientific Data on Time of Lineage Divergence for Your
730 Taxa. *R Package Release V0.6.2*. Retrieved from
731 <https://doi.org/10.5281/zenodo.593938>
- 732 Sanderson, M. (2002). Estimating Absolute Rates of Molecular Evolution and Divergence
733 Times: A Penalized Likelihood Approach. *Molecular Biology and Evolution*, 19(1),
734 101–109. <https://doi.org/10.1093/oxfordjournals.molbev.a003974>
- 735 Sanderson, M. (2003). r8s: Inferring Absolute Rates of Molecular Evolution and Divergence

- 736 Times in the Absence of a Molecular Clock. *Bioinformatics*, 19(2), 301–302.
- 737 Sanderson, M., & Doyle, J. (2001). Sources of error and confidence intervals in estimating
738 the age of angiosperms from rbcL and 18S rDNA data. *American Journal of Botany*,
739 88(8), 1499–1516.
- 740 Sauquet, H. (2013). A practical guide to molecular dating. *Comptes Rendus Palevol*, 12(6),
741 355–367.
- 742 Sauquet, H., Ramírez-Barahona, S., & Magallón, S. (2021). *The age of flowering plants is
743 unknown*.
- 744 Schenk, J. J. (2016). Consequences of secondary calibrations on divergence time estimates.
745 *PLoS ONE*, 11(1). <https://doi.org/10.1371/journal.pone.0148228>
- 746 Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in r. *Bioinformatics*, 27(4), 592–593.
- 747 Schoch, C. L., Ciufo, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R., ...
748 others. (2020). NCBI Taxonomy: a Comprehensive Update on Curation, Resources
749 and Tools. *Database*, 2020.
- 750 Selvatti, A. P., Gonzaga, L. P., & Moraes Russo, C. A. de. (2015). A paleogene origin for
751 crown passerines and the diversification of the oscines in the new world. *Molecular
752 Phylogenetics and Evolution*, 88, 1–15.
- 753 Smith, S., & Brown, J. (2018). Constructing a broadly inclusive seed plant phylogeny.
754 *American Journal of Botany*, 105(3), 302–314.
- 755 Smith, S., & O'Meara, B. (2012). TreePL: Divergence time estimation using penalized
756 likelihood for large phylogenies. *Bioinformatics*, 28(20), 2689–2690.
757 <https://doi.org/10.1093/bioinformatics/bts492>
- 758 Stoltzfus, A., Lapp, H., Matasci, N., Deus, H., Sidlauskas, B., Zmasek, C. M., ... Jordan, G.
759 (2013). Phylotastic! Making tree-of-life knowledge accessible, reusable and convenient.
760 *BMC Bioinformatics*, 14. <https://doi.org/10.1186/1471-2105-14-158>
- 761 Tietze, D. T., Päckert, M., Martens, J., Lehmann, H., & Sun, Y.-H. (2013). Complete
762 phylogeny and historical biogeography of true rosefinches (aves: Carpodacus).

763 *Zoological Journal of the Linnean Society*, 169(1), 215–234.

764 Title, P. O., & Rabosky, D. L. (2016). Do Macrophylogenies Yield Stable Macroevolutionary
765 Inferences? An Example from Squamate Reptiles. *Systematic Biology*, syw102.

766 <https://doi.org/10.1093/sysbio/syw102>

767 Treplin, S., Siegert, R., Bleidorn, C., Thompson, H. S., Fotso, R., & Tiedemann, R. (2008).
768 Molecular phylogeny of songbirds (aves: Passeriformes) and the relative utility of
769 common nuclear marker loci. *Cladistics*, 24(3), 328–349.

770 Vos, R. A., Balhoff, J. P., Caravas, J. A., Holder, M. T., Lapp, H., Maddison, W. P., ...
771 others. (2012). NeXML: Rich, extensible, and verifiable representation of
772 comparative data and metadata. *Systematic Biology*, 61(4), 675–689.
773 <https://doi.org/10.1093/sysbio/sys025>

774 Webb, C. (2000). Exploring the Phylogenetic Structure of Ecological Communities : An
775 Example for Rain Forest Trees. *The American Naturalist*, 156(2), 145–155.

776 Webb, C., Ackerly, D., & Kembel, S. (2008). Phylocom: Software for the analysis of
777 phylogenetic community structure and trait evolution. *Bioinformatics*, 24(18),
778 2098–2100. <https://doi.org/10.1093/bioinformatics/btn358>

779 Webb, C., & Donoghue, M. (2005). Phylomatic: Tree assembly for applied phylogenetics.
780 *Molecular Ecology Notes*, 5(1), 181–183.

781 Weir, J., & Schluter, D. (2008). Calibrating the avian molecular clock. *Molecular Ecology*,
782 17(10), 2321–2328.

783 Zuccon, D., Prŷs-Jones, R., Rasmussen, P. C., & Ericson, P. G. (2012). The phylogenetic
784 relationships and generic limits of finches (fringillidae). *Molecular Phylogenetics and
785 Evolution*, 62(2), 581–596.