

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

⁴ ² University of Tennessee, Knoxville

⁵ Author Note

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

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

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

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

17

Abstract

18 Achieving a high-quality reconstruction of a phylogenetic tree with branch lengths
19 proportional to absolute time (chronogram) is a difficult and time-consuming task. But the
20 increased availability of fossil and molecular data, and time-efficient analytical techniques
21 has resulted in many recent publications of large chronograms for a large number and wide
22 diversity of organisms. Knowledge of the evolutionary time frame of organisms is key for
23 research in the natural sciences. It also represent valuable information for education, science
24 communication, and policy decisions. When chronograms are shared in public, open
25 databases this wealth of expertly-curated and peer-reviewed data on evolutionary timeframe
26 is exposed in a programmatic and reusable way, as intensive and localized efforts have
27 improved data sharing practices, as well as incentivized open science in biology. Here we
28 present DateLife, a service implemented as an R package and an R Shiny website application
29 available at www.datelife.org, 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: 4912

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

46 **Introduction**

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

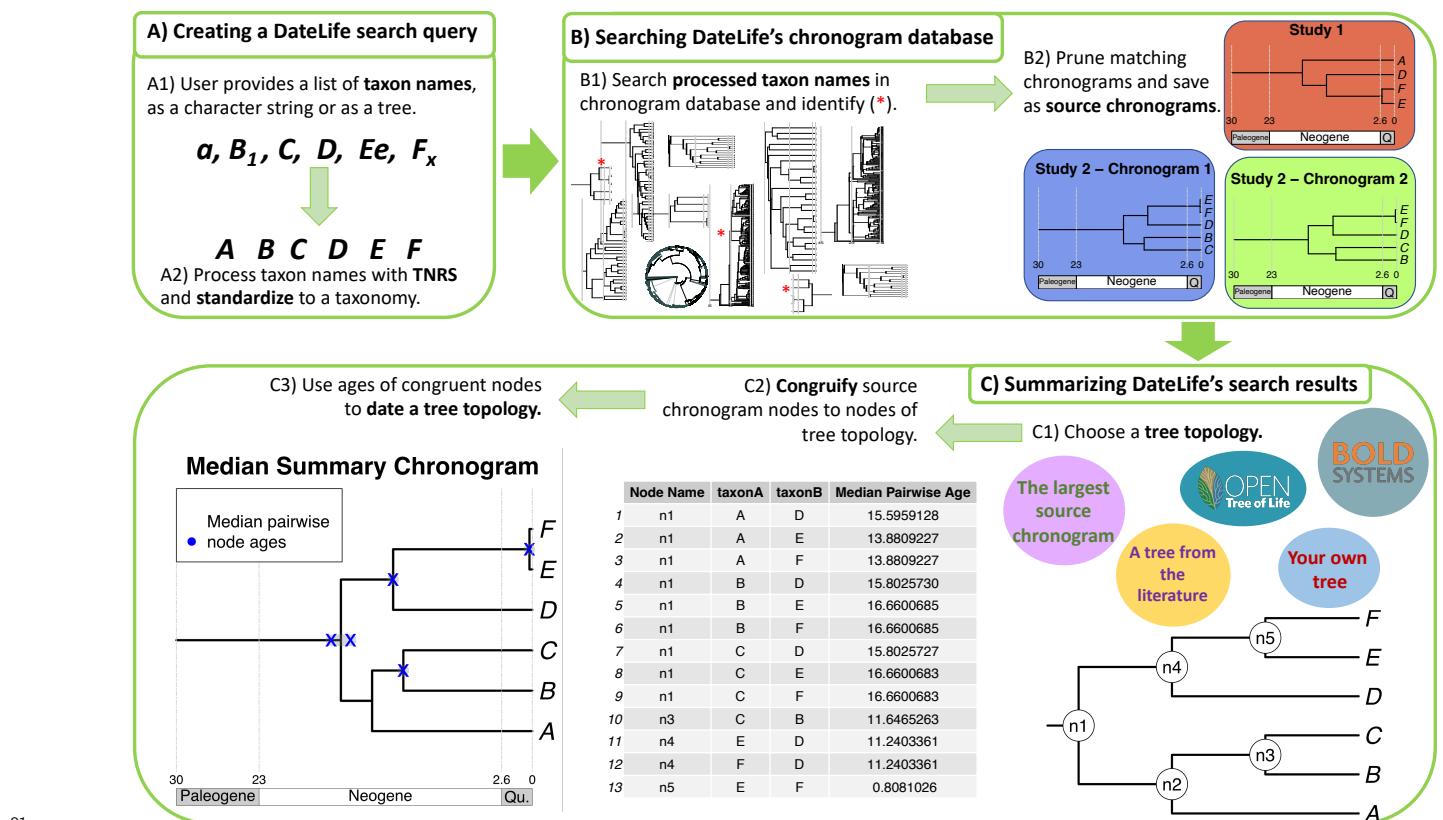
53 Building a chronogram is not an easy task. It requires obtaining and curating data to
54 construct a phylogeny, selecting and placing appropriate calibrations on the phylogeny using
55 independent age data points from the fossil record or other dated events, and inferring the
56 full dated tree; it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

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

68

Description

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



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 lengths,
87 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, the algorithm has
91 two 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 taxonomy of choice; current options are Open Tree of Life, NCBI, GBIF, or IRMNG
94 taxonomies) and adds them to the input string. If the flag is inactive, DateLife excludes any
95 taxon names above the species level from the search query.

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

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

107 **Searching a chronogram database**

108 At the time of writing of this manuscript (Jun 21, 2022), DateLife's chronogram
109 database latest version consist of 253 chronograms published in 187 different studies. It is
110 curated from OpenTree's phylogenetic database, the Phylesystem, which constitutes an open
111 source of expert and peer-reviewed phylogenetic knowledge with rich metadata (McTavish et
112 al., 2015), which allows automatic and reproducible assembly of our chronogram database.
113 Datelife's chronogram database is navigable as an R data object within the `datelife` R
114 package.

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

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

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

134 **Summarizing search results**

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

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

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

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

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

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

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

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

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

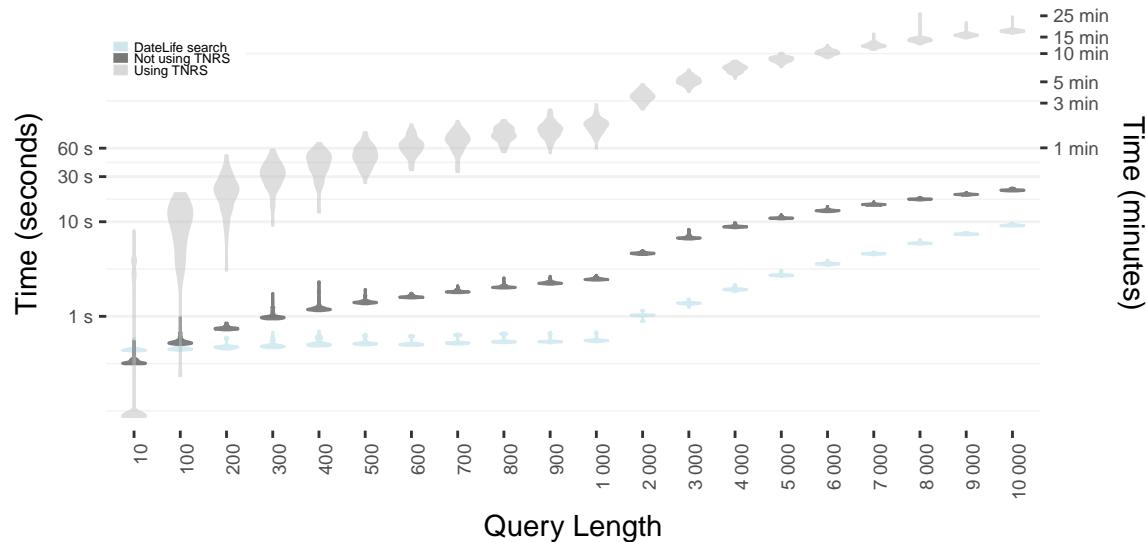
201 Alternative options supported in DateLife (MrBayes, PATHD8, TreePL) incorporate
202 branch length information from the input topology in combination with the calibrations.
203 PATHd8 is a non-clock, rate-smoothing method (Britton et al., 2007) to date trees. treePL
204 (Stephen A. Smith & O'Meara,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

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

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

Benchmark

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



223

FIGURE 2. DateLife's benchmarking results. Computation time used to process a query and a search across `datelife`'s chronogram database, relative to number of input taxon names.

226 *Caption continues.*

227 *Figure caption continued.* For each $N = \{10, 100, 200, \dots, 1\ 000, \dots, 9\ 000, 10\ 000\}$, we
228 sampled N species names from the class Aves a hundred times, and then performed a datelife
229 search processing the input names with Taxon Names Resolution Service (TNRS; light gray),
230 and without processing names (dark gray). For comparison, we performed a search using an
231 input that had been pre-processed with TNRS (light blue).

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

239

Case studies

240 We illustrate the DateLife algorithm using a family within the Passeriform birds
241 encompassing the true finches, Fringillidae, as case study. The first example analyses 6 bird
242 species and shows all steps of the algorithm. The second example is an application analysing
243 289 species in the family Fringillidae that are included in the NCBI taxonomy.

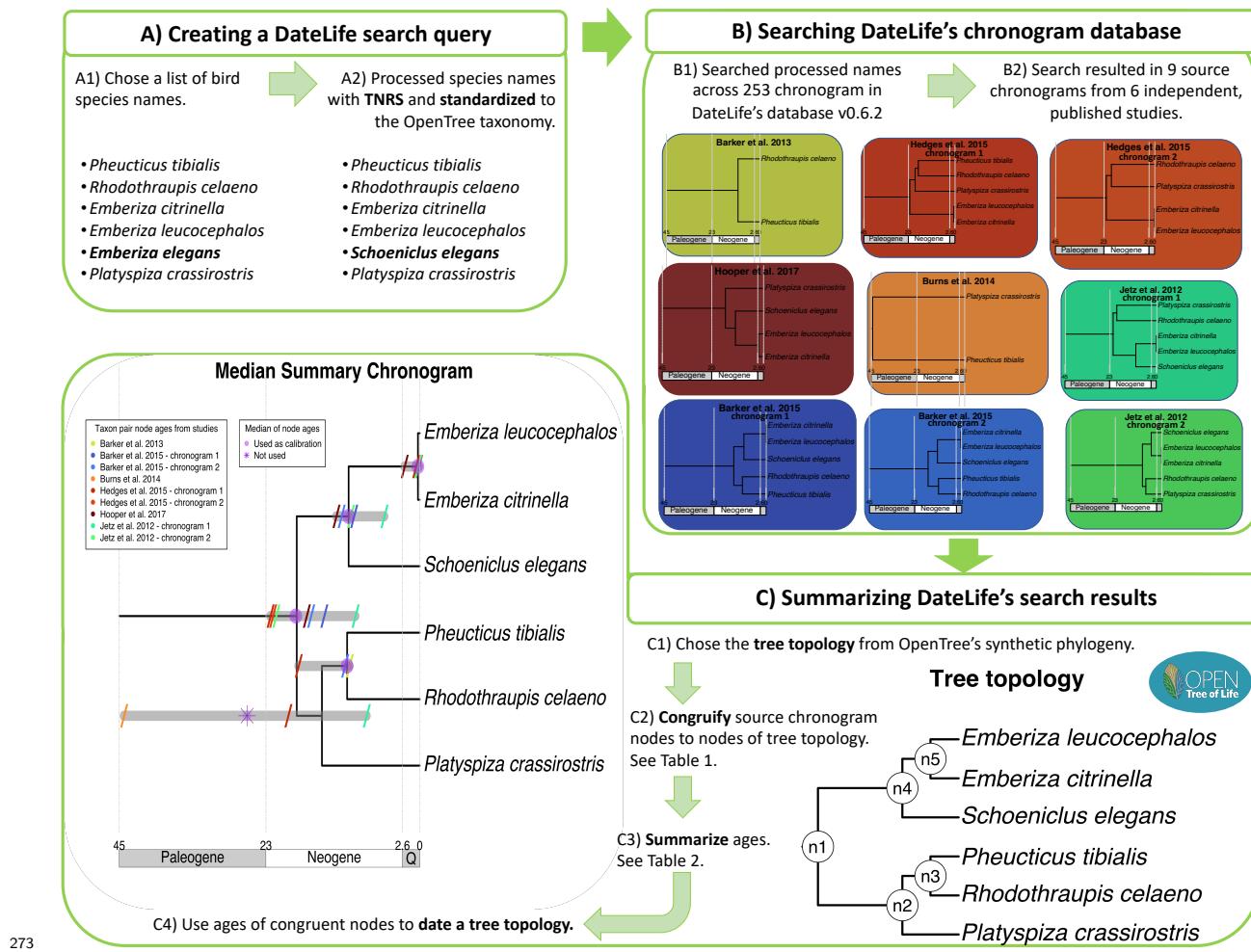
244 **A small example**

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

250 *Platyspiza crassirostris*. Processing of input names found that *Emberiza elegans* is synonym
251 for *Schoeniclus elegans* in the default reference taxonomy (Open Tree of Life Taxonomy v3.3,
252 June 1, 2021). For a detailed discussion on the state of the synonym, refer to Avibase
253 (Avibase, 2022; Lepage, 2004; Lepage, Vaidya, & Guralnick, 2014). Discovering this synonym
254 allowed assigning five age data points for the parent node of *Emberiza elegans*, shown as
255 *Schoeniclus elegans* in figure 3A, which would not have had any data otherwise.

256 **Searching the database.**- DateLife used the processed input names to search the local
257 chronogram database and found 9 matching chronograms in 6 different studies (Fig. 3B).
258 Three studies matched five input names (Barker, Burns, Klicka, Lanyon, & Lovette, 2015;
259 Hedges, Marin, Suleski, Paymer, & Kumar, 2015; Jetz, Thomas, Joy, Hartmann, & Mooers,
260 2012), one study matched four input names (Hooper & Price, 2017) and two studies matched
261 two input names (Barker, Burns, Klicka, Lanyon, & Lovette, 2013; Burns et al., 2014). No
262 studies matched all input names. Together, source chronograms provide 28 unique age data
263 points, covering all nodes on our chosen tree topology to date (Table 1).

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



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

	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

281 TABLE 1. Ages of congruified nodes obtained from a datelife search for the Small Example
 282 presented. See Figure 3, step C2.

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

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

²⁸⁴ **An example with the family of true finches**

²⁸⁵ ***Creating a query.***- To obtain ages for all species within the family of true finches,
²⁸⁶ Fringillidae, we ran a DateLife query using the “get species from taxon” flag, which gets all
²⁸⁷ recognized species names within a named group from a taxonomy of choice. Following the
²⁸⁸ NCBI taxonomy, our DateLife query has 289 Fringillidae species. This taxon-constrained
²⁸⁹ approach implies that the final results of a full DateLife analysis will be done using a tree
²⁹⁰ topology and ages for the species in a named group, which do not necessarily correspond to a
²⁹¹ monophyletic group. Users can change this behaviour by providing a monophyletic tree as
²⁹² input for a DateLife search, or as a tree topology for a DateLife summary.

²⁹³ ***Searching the database.***- Next, we used the processed species names in our DateLife
²⁹⁴ query to identify chronograms with at least two Fringillidae species. The DateLife search
²⁹⁵ identified 13 chronograms containing at least two Fringillidae species, published in 9 different
²⁹⁶ studies (Barker et al., 2013, 2015; Burns et al., 2014; Claramunt & Cracraft, 2015; Gibb et
²⁹⁷ al., 2015; Hedges et al., 2015; Hooper & Price, 2017; Jetz et al., 2012; Price et al., 2014).
²⁹⁸ Once identified, DateLife pruned matching chronograms to keep Fringillidae species names
²⁹⁹ on tips only, and transformed these pruned chronograms to pairwise distance matrices,
³⁰⁰ revealing 1206 different age data points available for species within the Fringillidae
³⁰¹ (supplementray table 1).

³⁰² ***Summarizing search results.***- The final step is to congruify and summarize the age data
³⁰³ available for the Fringillidae species into single summary chronograms, using different types
³⁰⁴ of summary ages, median and SDM. As explained in the “Description” section, a tree
³⁰⁵ topology to summarize age data upon is required. By default, DateLife uses as summarizing
³⁰⁶ topology the portion of OpenTree synthetic phylogeny that contains the species provided as
³⁰⁷ search query. According to this phylogeny, species belonging to the family Fringillidae do not
³⁰⁸ form a monophyletic group (Fig. 4).

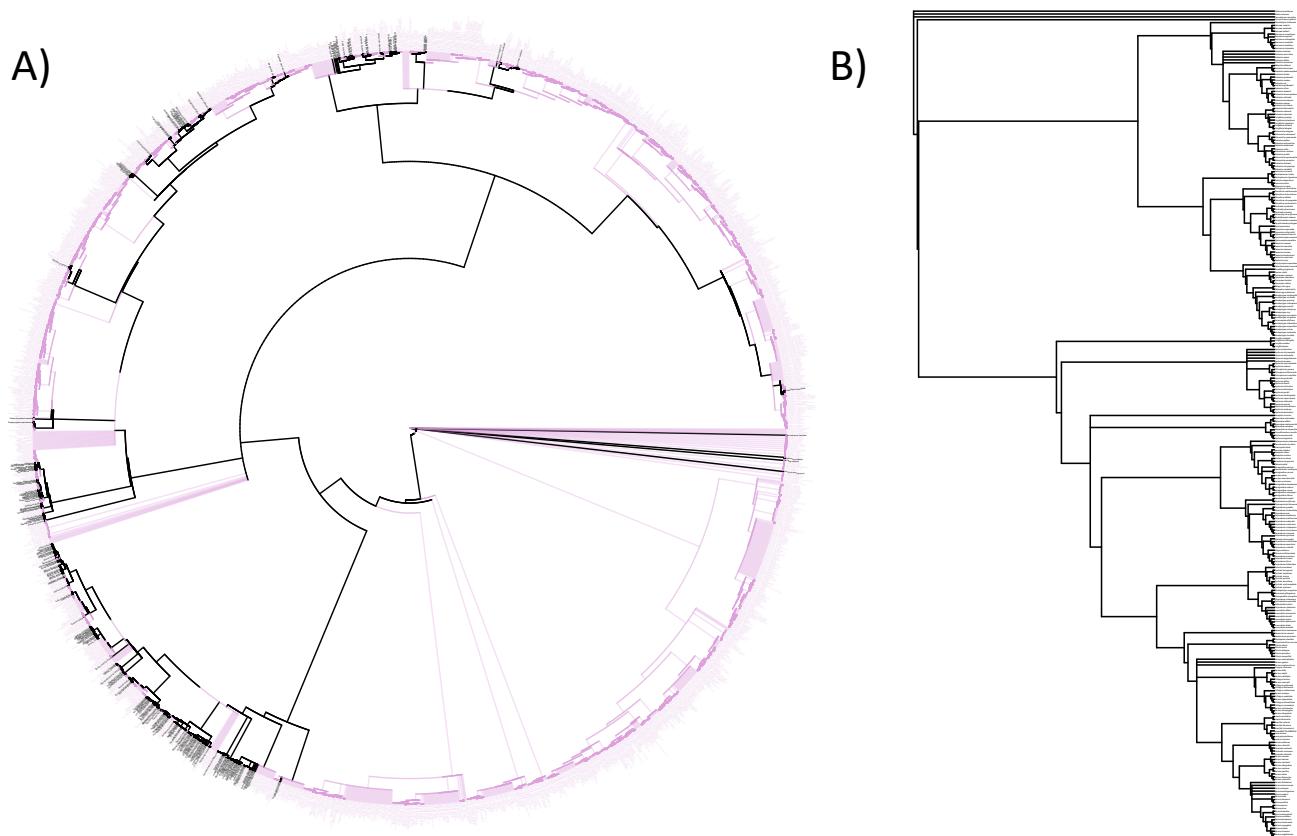
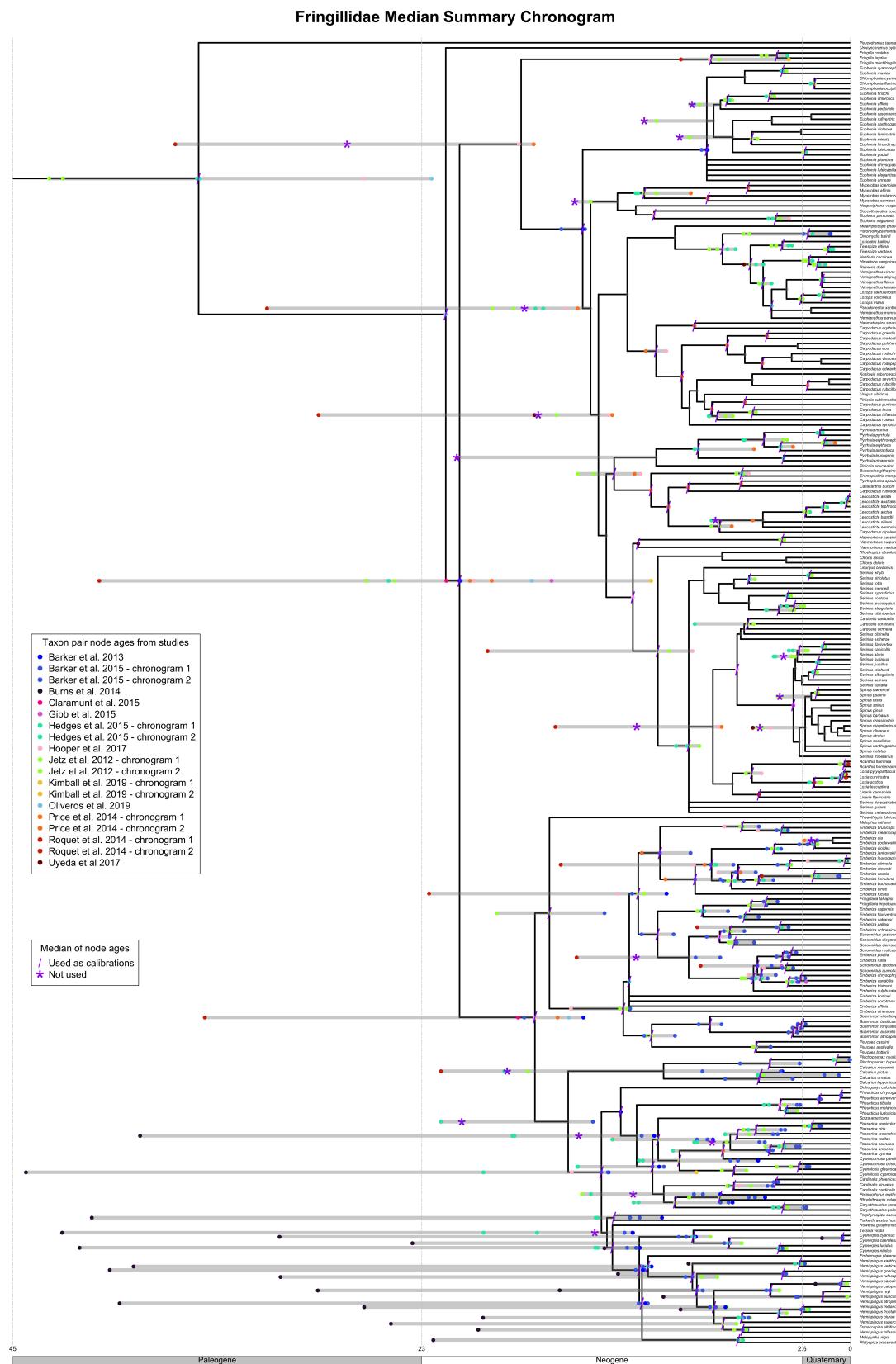


FIGURE 4. Tree topologies obtained from Open Tree of Life's (OpenTree) synthetic phylogenetic tree. A) Topology of 2,333 tips and 1,305 internal nodes, encompassing bird species within the family Fringillidae following the NCBI taxonomy (black), as well as all other bird species that share the same mrca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

326 Age data from source chronograms was congruified to the chosen tree topology (OpenTree's
327 synthetic subtree from figure 4B), reducing the age data set to 818 different data points
328 (supplementray table 2). Age summary statistics per congruified node were calculated and
329 used as fixed secondary calibrations over the chosen tree topology, to obtain a fully dated
330 phylogeny with the program BLADJ (Fig. 5).



333

331 FIGURE 5. Fringillidae median summary chronogram generated with DateLife. It has 256 tips and

332 233 nodes, from which 212 have age data from at least one published chronogram.

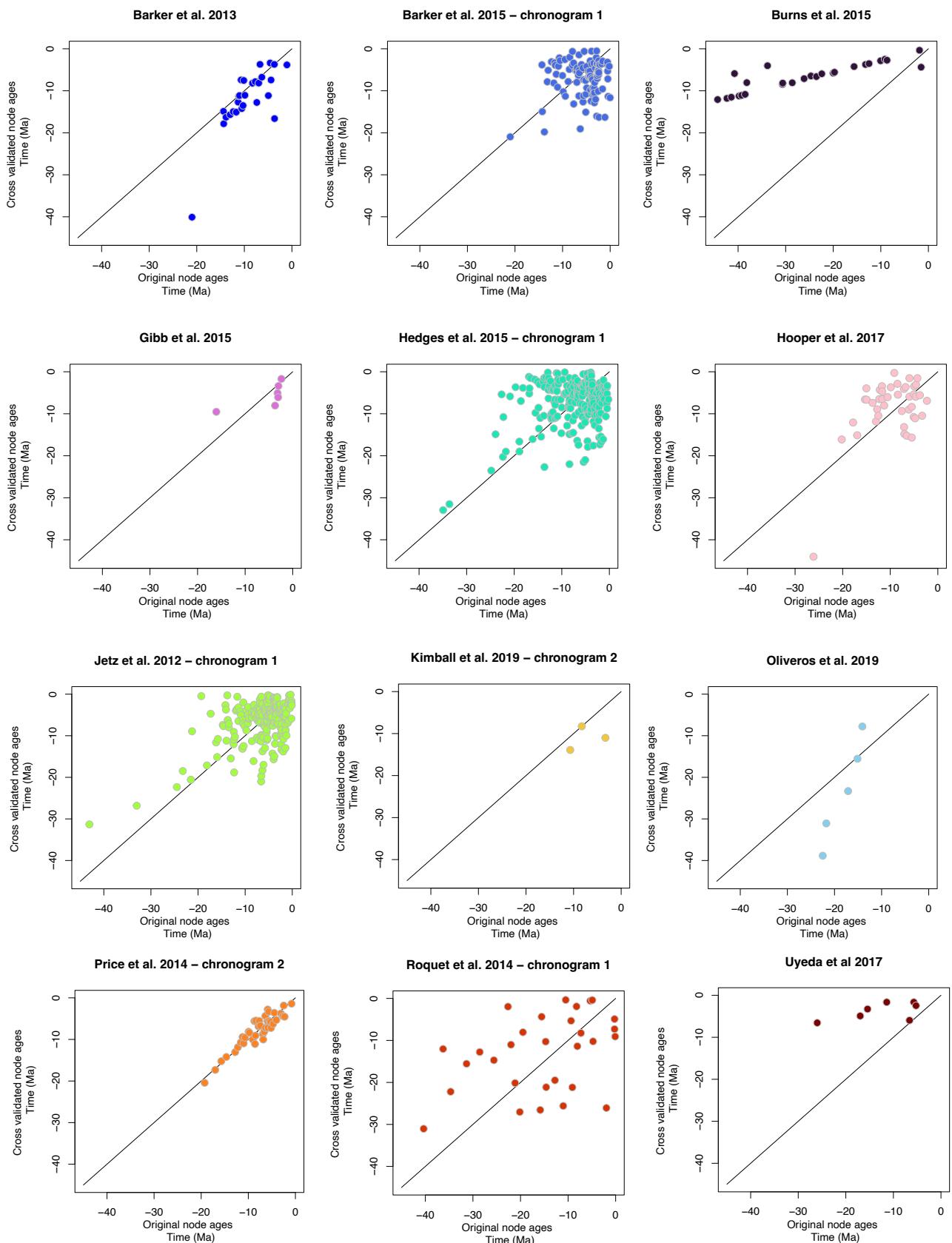
334

Cross-validation test

335 We performed a cross validation analysis of the DateLife workflow using the Fringillidae
336 chronograms. We used the individual tree topologies from each of the 19 source chronograms
337 from 13 studies as inputs, treating their node ages as unknown. We then estimated dates for
338 these topologies using the node ages from the chronograms from the other studies as
339 calibrations and smoothing using BLADJ.

340 We found that node ages from original study, and ages estimated using all other age data
341 available are correlated (Fig. 6).

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



346 FIGURE 6. Results from cross validation analysis. Each plot compares the original age estimate (x
 axis) with the age obtained with a DateLife analysis (y axis), per node.
 347

349

Discussion

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

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

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

365 While most chronograms in DateLife's database are constructed using primary calibrations
366 (molecular substitution rates or ages obtained from the fossil record or geological events),
367 DateLife summarizes chronograms using secondary calibrations only (ages coming from other
368 chronograms).

369 Graur and Martin (2004) cautioned on the increased error and uncertainty in estimated ages
370 when using secondary calibrations in dating analyses. Schenk (2016) showed that, in
371 simulations, divergence times inferred using secondary calibrations are significantly younger
372 than those inferred with primary calibrations, when obtained with Bayesian inference
373 methods, and when priors are implemented in similar ways in both analyses. Accordingly,
374 the scientific community seems to have more confidence in chronograms obtained from a

375 single analysis, using fossil data as primary sources of calibrations (Schenk, 2016), and using
376 fossils that have been widely discussed and curated as calibrations to date other trees,
377 making sure that all data reflect a coherent evolutionary history (Sauquet, 2013), as for
378 example done by Antonelli et al. (2017). There have been attempts to create fossil
379 calibration databases (Ksepka et al., 2015), though these still have room to grow.

380 It seems that using several (as opposed to just a few) secondary calibrations can provide
381 sufficient information to alleviate or even neutralize potential biases (Sauquet, 2013).

382 Certainly, further studies are required to fully understand the effect of secondary calibrations
383 on outputs from different tree dating methods, and on downstream analyses. It is possible
384 that secondary calibrations can be safely used with dating methods that do not require
385 setting priors, such as penalized likelihood (Michael J. Sanderson, 2003), with methods that
386 do not make any assumptions on the ages and fix them to a node on a tree topology, such as
387 BLADJ (Campbell O. Webb et al., 2008; Campbell O. Webb & Donoghue, 2005), or
388 methods that summarize age data unto a tree topology.

389 Our cross validation analysis might provide some insight in this regard. When ages are
390 estimated with secondary calibrations, nodes closer to the root do tend to be slightly
391 younger than ages estimated with primary calibrations. However, nodes closer to the tip
392 tend to be older when estimated using secondary calibrations with a dating method that
393 does not make any prior assumptions on the nature of the calibrations themselves
394 (Supplementary Figures 1-9), with a notable exception of cross validation of Burns et al.
395 (2014) chronogram, which results in much younger node ages when estimated using
396 secondary calibrations (Supplementary figs. 1 and 4).

397 Variation in source chronograms

398 Conflict among alternative dating studies is common in the literature (see, for example, the
399 robust ongoing debate about crown group age of angiosperms (Barba-Montoya, Reis,
400 Schneider, Donoghue, & Yang, 2018; Magallón, Gómez-Acevedo, Sánchez-Reyes, &

401 Hernández-Hernández, 2015; Ramshaw et al., 1972; Michael J. Sanderson & Doyle, 2001;
402 Sauquet, Ramírez-Barahona, & Magallón, 2021)).
403 Source chronograms may have been estimated using different types of primary calibrations
404 (fossil record, geologic or biogeographic events, molecular substitution rates). These
405 differences in approach can deepen the already substantial variation in time estimates
406 between lineages, as observed from the comparison of source chronograms in the Fringillidae
407 example. For example, the chronograms from Burns et al. (2014) were inferred using
408 molecular substitution rate estimates across birds (Weir & Schlüter, 2008), and have much
409 older date estimates for the same nodes than chronograms that were inferred using fossils as
410 primary calibrations (figs. fig:fringillidages and fig:cvXY; supplementary figs. 1 and 4).
411 Different chronograms might also imply fundamentally distinct evolutionary hypotheses. For
412 example, two different researchers working on the same clade could both carefully select and
413 argue for their choices of fossil calibrations. Still, if one researcher decides a fossil will
414 calibrate the ingroup of a clade, while another researcher uses the same one to calibrate
415 outside the clade, the resulting age estimates will often differ substantially, as the placement
416 of calibrations as stem or crown group is proved to deeply affect estimated times of lineage
417 divergence (Sauquet, 2013).

418 Summarizing chronograms

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

427 number of calibrations used Magallón et al. (2015). Some characteristics that are often cited
428 in published studies as a measure of improved age estimates as compared to previously
429 published estimates are: quality of alignment (missing data, GC content), lineage sampling
430 (strategy and proportion), phylogenetic and dating inference method, number of fossils used
431 as calibrations, support for nodes and ages, and magnitude of confidence intervals.

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

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

446 Effects on downstream analyses

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

450 In ecology and conservation biology, incorporating at least some data on lineage divergence
451 times represents a relevant improvement for testing alternative hypothesis using phylogenetic
452 distance (Campbell O. Webb et al., 2008). Hence, DateLife's workflow features different

ways of estimating node ages in the absence of calibrations and branch length information for certain taxa. “Making up” branch lengths is a common practice in scientific publications: Jetz et al. (2012), created a time-calibrated tree of all 9,993 bird species, where 67% had molecular data and the rest was simulated; Rabosky et al. (2018) created a time-calibrated tree of 31,536 ray-finned fishes, of which only 37% had molecular data; Stephen A. Smith and Brown (2018) constructed a tree of 353,185 seed plants where only 23% had molecular data. Notably, risks come with this practice. Taken to the extreme, one could make a fully resolved, calibrated tree of all modern and extinct taxa using a single taxonomy and a single calibration, using polytomy resolution and branch estimation methods. There has yet to be a thorough analysis of what can go wrong when one extends inferences beyond the data in this way, so we urge caution; we also urge readers to follow the example of the large tree papers cited above, by carefully considering the statistical assumptions being made, and assessing the consistency of the results with prior work.

Conclusions

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

The DateLife project allows for easy and fast summary of public and state-of-the-art data on time of lineage divergence. It provides a straightforward way to get an informed idea on the state of knowledge of the time frame of evolution of different regions of the tree of life, and allows identification of regions that require more research, or that have conflicting

479 information. It is available as an R package, and as a web-based R shiny application at
480 www.datelife.org Both summary and newly generated trees are useful to evaluate
481 evolutionary hypotheses in different areas of research. The DateLife project helps with
482 awareness of the existing variation in expert time of divergence data, and will foster
483 exploration of the effect of alternative divergence time hypothesis on the results of analyses,
484 nurturing a culture of more cautious interpretation of evolutionary results.

485 Availability

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

496 Supplementary Material

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

500 Funding

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

504

Acknowledgements

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

References

- 517
- 518 Alström, P., Hooper, D. M., Liu, Y., Olsson, U., Mohan, D., Gelang, M., ... Price, T.
519 D. (2014). Discovery of a relict lineage and monotypic family of passerine birds.
520 *Biology Letters*, 10(3), 20131067.
- 521 Ané, C., Eulenstein, O., Piaggio-Talice, R., & Sanderson, M. J. (2009). Groves of
522 phylogenetic trees. *Annals of Combinatorics*, 13(2), 139–167.
- 523 Antonelli, A., Hettling, H., Condamine, F. L., Vos, K., Nilsson, R. H., Sanderson, M.
524 J., ... Vos, R. A. (2017). Toward a self-updating platform for estimating rates of
525 speciation and migration, ages, and relationships of Taxa. *Systematic Biology*,
526 66(2), 153–166. <https://doi.org/10.1093/sysbio/syw066>
- 527 Archie, J., Day, W. H., Felsenstein, J., Maddison, W., Meacham, C., Rohlf, F. J., &
528 Swofford, D. (1986). The Newick tree format. Retrieved from
529 %7B<https://evolution.genetics.washington.edu/phylip/newicktree.html>%7D
- 530 Avibase. (2022). Yellow-throated Bunting. *Avibase - The World Bird Database*,
531 (Online Resource). Retrieved from %7B<https://avibase.bsc-eoc.org/species.jsp?lang=EN&avibaseid=82D1EE0049D8D927%7D>
- 532 Bapst, D. W. (2012). Paleotree: An R package for paleontological and phylogenetic
533 analyses of evolution. *Methods in Ecology and Evolution*, 3(5), 803–807.
534 <https://doi.org/10.1111/j.2041-210X.2012.00223.x>
- 535 Barba-Montoya, J., Reis, M. dos, Schneider, H., Donoghue, P. C., & Yang, Z. (2018).
536 Constraining uncertainty in the timescale of angiosperm evolution and the
537 veracity of a cretaceous terrestrial revolution. *New Phytologist*, 218(2), 819–834.
- 538 Barker, F. K. (2014). Mitogenomic data resolve basal relationships among passeriform
539 and passeridan birds. *Molecular Phylogenetics and Evolution*, 79, 313–324.
- 540 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2013). Going
541 to extremes: Contrasting rates of diversification in a recent radiation of new world
542 passerine birds. *Systematic Biology*, 62(2), 298–320.
- 543

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

- 571 *Phylogenetics and Evolution*, 75, 41–77.
- 572 Chamberlain, S. A., & Szöcs, E. (2013). taxize : taxonomic search and retrieval in R
573 [version 2; referees: 3 approved]. *F1000Research*, 2(191), 1–29.
574 <https://doi.org/10.12688/f1000research.2-191.v2>
- 575 Chamberlain, S. A., Szöcs, E., Foster, Z., Arendsee, Z., Boettiger, C., Ram, K., ...
576 Li, G. (2019). *taxize: Taxonomic information from around the web*. Retrieved
577 from <https://github.com/ropensci/taxize>
- 578 Chaves, J. A., Hidalgo, J. R., & Klicka, J. (2013). Biogeography and evolutionary
579 history of the n eotropical genus s altator (a ves: T hraupini). *Journal of
580 Biogeography*, 40(11), 2180–2190.
- 581 Claramunt, S., & Cracraft, J. (2015). A new time tree reveals earth history's imprint
582 on the evolution of modern birds. *Science Advances*, 1(11), e1501005.
- 583 Criscuolo, A., Berry, V., Douzery, E. J. P., & Gascuel, O. (2006). SDM: A fast
584 distance-based approach for (super)tree building in phylogenomics. *Systematic
585 Biology*, 55(5), 740–755. <https://doi.org/10.1080/10635150600969872>
- 586 Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M.-K., Turon, X., ...
587 Douzery, E. J. (2018). A phylogenomic framework and timescale for comparative
588 studies of tunicates. *BMC Biology*, 16(1), 1–14.
- 589 Eastman, J. M., Harmon, L. J., & Tank, D. C. (2013). Congruification: Support for
590 time scaling large phylogenetic trees. *Methods in Ecology and Evolution*, 4(7),
591 688–691. <https://doi.org/10.1111/2041-210X.12051>
- 592 Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and
593 high throughput. *Nucleic Acids Research*, 32(5), 1792–1797.
- 594 Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American
595 Naturalist*, 125(1), 1–15. Retrieved from <http://www.jstor.org/stable/2461605>
- 596 Gibb, G. C., England, R., Hartig, G., McLenaghan, P. A., Taylor Smith, B. L.,
597 McComish, B. J., ... Penny, D. (2015). New zealand passerines help clarify the

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

- 625 with MAFFT. In *Bioinformatics for DNA sequence analysis* (pp. 39–64).
626 Springer.
- 627 Kimball, R. T., Oliveros, C. H., Wang, N., White, N. D., Barker, F. K., Field, D. J.,
628 ... others. (2019). A phylogenomic supertree of birds. *Diversity*, 11(7), 109.
- 629 Klicka, J., Barker, F. K., Burns, K. J., Lanyon, S. M., Lovette, I. J., Chaves, J. A., &
630 Bryson Jr, R. W. (2014). A comprehensive multilocus assessment of sparrow (aves:
631 Passerellidae) relationships. *Molecular Phylogenetics and Evolution*, 77, 177–182.
- 632 Ksepka, D. T., Parham, J. F., Allman, J. F., Benton, M. J., Carrano, M. T.,
633 Cranston, K. A., ... others. (2015). The fossil calibration database—a new
634 resource for divergence dating. *Systematic Biology*, 64(5), 853–859.
- 635 Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M.,
636 Martinez-Barrio, A., ... others. (2015). Evolution of darwin's finches and their
637 beaks revealed by genome sequencing. *Nature*, 518(7539), 371–375.
- 638 Laubichler, M. D., & Maienschein, J. (2009). *Form and function in developmental
639 evolution*. Cambridge University Press.
- 640 Lepage, D. (2004). *Avibase: The world bird database*. Bird Studies Canada.
- 641 Lepage, D., Vaidya, G., & Guralnick, R. (2014). Avibase—a database system for
642 managing and organizing taxonomic concepts. *ZooKeys*, (420), 117.
- 643 Lerner, H. R., Meyer, M., James, H. F., Hofreiter, M., & Fleischer, R. C. (2011).
644 Multilocus resolution of phylogeny and timescale in the extant adaptive radiation
645 of hawaiian honeycreepers. *Current Biology*, 21(21), 1838–1844.
- 646 Lovette, I. J., Pérez-Emán, J. L., Sullivan, J. P., Banks, R. C., Fiorentino, I.,
647 Córdoba-Córdoba, S., ... others. (2010). A comprehensive multilocus phylogeny
648 for the wood-warblers and a revised classification of the parulidae (aves).
649 *Molecular Phylogenetics and Evolution*, 57(2), 753–770.
- 650 Magallon, S., & Sanderson, M. J. (2001). Absolute diversification rates in angiosperm
651 clades. *Evolution*, 55(9), 1762–1780.

- 652 Magallón, S. (2010). Using fossils to break long branches in molecular dating: A
653 comparison of relaxed clocks applied to the origin of angiosperms. *Systematic*
654 *Biology*, 59(4), 384–399.
- 655 Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., & Hernández-Hernández, T.
656 (2015). A metacalibrated time-tree documents the early rise of flowering plant
657 phylogenetic diversity. *New Phytologist*, 207(2), 437–453.
- 658 McTavish, E. J., Hinchliff, C. E., Allman, J. F., Brown, J. W., Cranston, K. A.,
659 Holder, M. T., . . . Smith, S. A. (2015). Phylesystem: A git-based data store for
660 community-curated phylogenetic estimates. *Bioinformatics*, 31(17), 2794–2800.
- 661 Michonneau, F., Brown, J. W., & Winter, D. J. (2016). rotl: an R package to interact
662 with the Open Tree of Life data. *Methods in Ecology and Evolution*, 7(12),
663 1476–1481. <https://doi.org/10.1111/2041-210X.12593>
- 664 Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecology*
665 *Letters*, 17(4), 508–525. <https://doi.org/10.1111/ele.12251>
- 666 Moyle, R. G., Oliveros, C. H., Andersen, M. J., Hosner, P. A., Benz, B. W., Manthey,
667 J. D., . . . Faircloth, B. C. (2016). Tectonic collision and uplift of wallacea
668 triggered the global songbird radiation. *Nature Communications*, 7(1), 1–7.
- 669 Ödeen, A., Håstad, O., & Alström, P. (2011). Evolution of ultraviolet vision in the
670 largest avian radiation—the passerines. *BMC Evolutionary Biology*, 11(1), 1–8.
- 671 Oliveros, C. H., Field, D. J., Ksepka, D. T., Barker, F. K., Aleixo, A., Andersen, M.
672 J., . . . others. (2019). Earth history and the passerine superradiation.
673 *Proceedings of the National Academy of Sciences*, 116(16), 7916–7925.
- 674 Ooms, J., & Chamberlain, S. (2018). *Phylocomr: Interface to 'phylocom'*. Retrieved
675 from <https://CRAN.R-project.org/package=phylocomr>
- 676 Open Tree Of Life, Redelings, B., Cranston, K. A., Allman, J., Holder, M. T., &
677 McTavish, E. J. (2016). Open Tree of Life APIs v3.0. *Open Tree of Life Project*,
678 (Online Resources). Retrieved from

- 679 %7Bhttps://github.com/OpenTreeOfLife/germinator/wiki/Open-Tree-of-Life-
680 Web-APIs%7D
- 681 Open Tree Of Life, Redelings, B., Sánchez Reyes, L. L., Cranston, K. A., Allman, J.,
682 Holder, M. T., & McTavish, E. J. (2019). Open tree of life synthetic tree v12.3.
683 *Zenodo*. Retrieved from <https://doi.org/10.5281/zenodo.3937742>
- 684 Päckert, M., Martens, J., Sun, Y.-H., Severinghaus, L. L., Nazarenko, A. A., Ting, J.,
685 ... Tietze, D. T. (2012). Horizontal and elevational phylogeographic patterns of
686 himalayan and southeast asian forest passerines (aves: passeriformes). *Journal of*
687 *Biogeography*, 39(3), 556–573.
- 688 Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and
689 evolution in R language. *Bioinformatics*, 20(2), 289–290.
- 690 Parchman, T. L., Benkman, C. W., & Mezquida, E. T. (2007). Coevolution between
691 hispaniolan crossbills and pine: Does more time allow for greater phenotypic
692 escalation at lower latitude? *Evolution*, 61(9), 2142–2153.
- 693 Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., FitzJohn,
694 R. G., ... Harmon, L. J. (2014). Geiger v2. 0: An expanded suite of methods for
695 fitting macroevolutionary models to phylogenetic trees. *Bioinformatics*, 30(15),
696 2216–2218.
- 697 Posadas, P., Crisci, J. V., & Katinas, L. (2006). Historical biogeography: A review of
698 its basic concepts and critical issues. *Journal of Arid Environments*, 66(3),
699 389–403.
- 700 Powell, A. F., Barker, F. K., Lanyon, S. M., Burns, K. J., Klicka, J., & Lovette, I. J.
701 (2014). A comprehensive species-level molecular phylogeny of the new world
702 blackbirds (icteridae). *Molecular Phylogenetics and Evolution*, 71, 94–112.
- 703 Price, T. D., Hooper, D. M., Buchanan, C. D., Johansson, U. S., Tietze, D. T.,
704 Alström, P., ... others. (2014). Niche filling slows the diversification of himalayan
705 songbirds. *Nature*, 509(7499), 222.

- 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: cardinalidae). *Molecular Phylogenetics and Evolution*, 69(3), 1222–1227.
- R Core Team. (2018). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- 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*, 559(7714), 392.
- Ramshaw, J., Richardson, D., Mealyard, B., Brown, R., Richardson, M., Thompson, E., & Boulter, D. (1972). The time of origin of the flowering plants determined by using amino acid sequence data of cytochrome c. *New Phytologist*, 71(5), 773–779.
- Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The barcode of life data system (<http://www.Barcodinglife.org>). *Molecular Ecology Notes*, 7(3), 355–364.
- Rees, J. A., & Cranston, K. (2017). Automated assembly of a reference taxonomy for phylogenetic data synthesis. *Biodiversity Data Journal*, (5).
- Revell, L. J. (2012). Phytools: An r package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3, 217–223.
- Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19(12), 1572–1574.
<https://doi.org/10.1093/bioinformatics/btg180>
- Sanchez-Reyes, L. L., & O'Meara, B. (2022). **datelifeplot**: Methods to plot chronograms and outputs of the **datelife** package. *R Package Release V0.2.2*. Retrieved from <https://zenodo.org/badge/latestdoi/381501451>
- Sanchez-Reyes, L. L., O'Meara, B., Eastman, J., Heath, T., Wright, A., Schliep, K., ... Alfaro, M. (2022). **datelife**: Scientific Data on Time of Lineage Divergence for Your Taxa. *R Package Release V0.6.2*. Retrieved from

- 733 <https://doi.org/10.5281/zenodo.593938>
- 734 Sanderson, Michael J. (2002). Estimating Absolute Rates of Molecular Evolution and
735 Divergence Times: A Penalized Likelihood Approach. *Molecular Biology and*
736 *Evolution*, 19(1), 101–109.
737 <https://doi.org/10.1093/oxfordjournals.molbev.a003974>
- 738 Sanderson, Michael J. (2003). r8s: Inferring absolute rates of molecular evolution and
739 divergence times in the absence of a molecular clock. *Bioinformatics*, 19(2),
740 301–302.
- 741 Sanderson, Michael J., & Doyle, J. A. (2001). Sources of error and confidence
742 intervals in estimating the age of angiosperms from rbcL and 18S rDNA data.
743 *American Journal of Botany*, 88(8), 1499–1516.
- 744 Sauquet, H. (2013). A practical guide to molecular dating. *Comptes Rendus Palevol*,
745 12(6), 355–367.
- 746 Sauquet, H., Ramírez-Barahona, S., & Magallón, S. (2021). *The age of flowering*
747 *plants is unknown*.
- 748 Schenk, J. J. (2016). Consequences of secondary calibrations on divergence time
749 estimates. *PLoS ONE*, 11(1). <https://doi.org/10.1371/journal.pone.0148228>
- 750 Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in r. *Bioinformatics*, 27(4),
751 592–593.
- 752 Selvatti, A. P., Gonzaga, L. P., & Moraes Russo, C. A. de. (2015). A paleogene origin
753 for crown passerines and the diversification of the oscines in the new world.
754 *Molecular Phylogenetics and Evolution*, 88, 1–15.
- 755 Smith, Stephen A., & Brown, J. W. (2018). Constructing a broadly inclusive seed
756 plant phylogeny. *American Journal of Botany*, 105(3), 302–314.
- 757 Smith, Stephen A., & O'Meara, B. C. (2012). TreePL: Divergence time estimation
758 using penalized likelihood for large phylogenies. *Bioinformatics*, 28(20),
759 2689–2690. <https://doi.org/10.1093/bioinformatics/bts492>

- 760 Stoltzfus, A., Lapp, H., Matasci, N., Deus, H., Sidlauskas, B., Zmasek, C. M., ...
761 Jordan, G. (2013). Phylotastic! Making tree-of-life knowledge accessible, reusable
762 and convenient. *BMC Bioinformatics*, 14.
763 <https://doi.org/10.1186/1471-2105-14-158>
- 764 Tietze, D. T., Päckert, M., Martens, J., Lehmann, H., & Sun, Y.-H. (2013). Complete
765 phylogeny and historical biogeography of true rosefinches (aves: carpodacus).
766 *Zoological Journal of the Linnean Society*, 169(1), 215–234.
- 767 Title, P. O., & Rabosky, D. L. (2016). Do Macrophylogenies Yield Stable
768 Macroevolutionary Inferences? An Example from Squamate Reptiles. *Systematic
769 Biology*, syw102. <https://doi.org/10.1093/sysbio/syw102>
- 770 Treplin, S., Siegert, R., Bleidorn, C., Thompson, H. S., Fotso, R., & Tiedemann, R.
771 (2008). Molecular phylogeny of songbirds (aves: Passeriformes) and the relative
772 utility of common nuclear marker loci. *Cladistics*, 24(3), 328–349.
- 773 Vos, R. A., Balhoff, J. P., Caravas, J. A., Holder, M. T., Lapp, H., Maddison, W. P.,
774 ... others. (2012). NeXML: Rich, extensible, and verifiable representation of
775 comparative data and metadata. *Systematic Biology*, 61(4), 675–689.
- 776 Webb, C. (2000). Exploring the Phylogenetic Structure of Ecological Communities :
777 An Example for Rain Forest Trees. *The American Naturalist*, 156(2), 145–155.
- 778 Webb, Campbell O., Ackerly, D. D., & Kembel, S. W. (2008). Phylocom: Software for
779 the analysis of phylogenetic community structure and trait evolution.
780 *Bioinformatics*, 24(18), 2098–2100.
781 <https://doi.org/10.1093/bioinformatics/btn358>
- 782 Webb, Campbell O., & Donoghue, M. J. (2005). Phylomatic: Tree assembly for
783 applied phylogenetics. *Molecular Ecology Notes*, 5(1), 181–183.
- 784 Weir, J., & Schluter, D. (2008). Calibrating the avian molecular clock. *Molecular
785 Ecology*, 17(10), 2321–2328.
- 786 Zuccon, D., Prŷs-Jones, R., Rasmussen, P. C., & Ericson, P. G. (2012). The

787 phylogenetic relationships and generic limits of finches (fringillidae). *Molecular*
788 *Phylogenetics and Evolution*, 62(2), 581–596.