

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: 5400

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

46 **Introduction**

47 Chronograms –phylogenies with branch lengths proportional to time– provide key data
48 on evolutionary time frame for the study of natural processes in many areas of biological
49 research, such as comparative analysis (Freckleton, Harvey, & Pagel, 2002; Harvey, Pagel, &
50 others, 1991), developmental biology (Delsuc et al., 2018; Laubichler & Maienschein, 2009),
51 conservation biology and ecology (Felsenstein, 1985; Webb, 2000), historical biogeography
52 (Posadas, Crisci, & Katinas, 2006), and species diversification (Magallon & Sanderson, 2001;
53 Morlon, 2014).

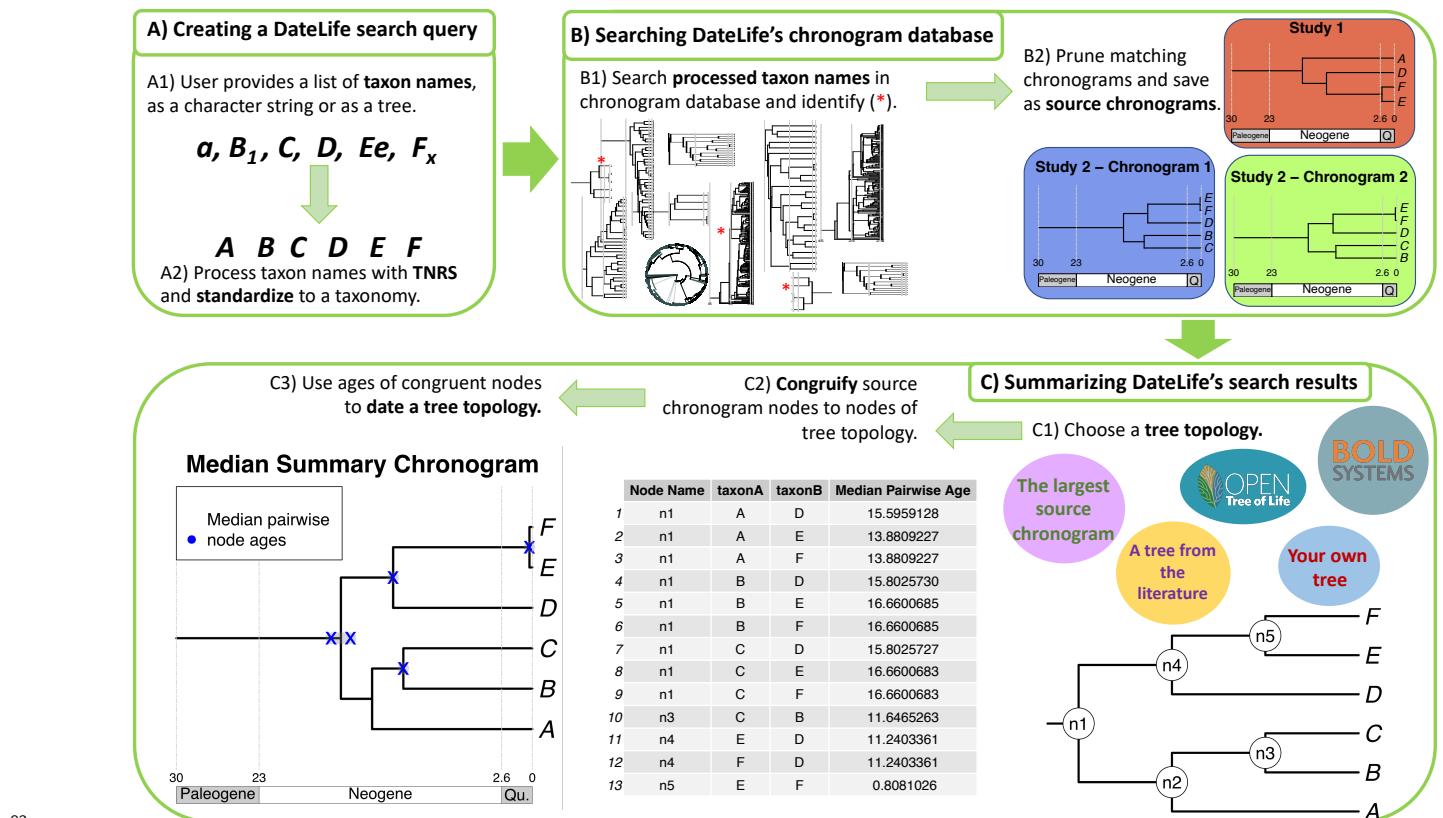
54 Building a chronogram is not an easy task. It requires obtaining and curating data to
55 construct a phylogeny, selecting and placing appropriate calibrations on the phylogeny using
56 independent age data points from the fossil record or other dated events, and inferring the
57 full dated tree; it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

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

70

Description

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



83

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

84 Creating a search query

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

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

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

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

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

116 **Searching a chronogram database**

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

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

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

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

143 **Summarizing search results**

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

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

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

the OpenTree synthetic tree, a phylogeny currently encompassing 2.3 million taxa across all life, assembled from 1, 239 published phylogenetic trees and OpenTree’s unified Taxonomy, OTT (Open Tree Of Life et al., 2019). Alternatively, DateLife can combine topologies from source chronograms using a supertree approach (Criscuolo, Berry, Douzery, & Gascuel, 2006). To do this DateLife first identifies the source chronograms that form a grove, roughly, a sufficiently overlapping set of taxa between trees, by implementing definition 2.8 for n-overlap from Ané et al. (2009). If the source chronograms do not form a grove, the supertree reconstruction will fail. In rare cases, a group of trees can have multiple groves. By default, DateLife chooses the grove with the most taxa, however, the “criterion = trees” flag allows the user to choose the grove with the most trees instead. The result is a single summary (or supertree) topology, that combines topologies from source chronograms in a grove.

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

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

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

¹⁸⁹ **Dating a tree topology.**— By default, DateLife implements the Branch Length
¹⁹⁰ Adjuster (BLADJ; Webb et al., 2008; Webb & Donoghue, 2005) algorithm to obtain a fully
¹⁹¹ dated topology. BLADJ takes an initial topologies without any branch length data, fixes
¹⁹² node ages that have calibration data, and distributes time between nodes with no data evenly
¹⁹³ between calibrated nodes. This has proven effective to minimize age variance in the resulting
¹⁹⁴ chronogram and useful for ecological analyses (Webb et al., 2008). When there is conflict in
¹⁹⁵ ages between nodes with calibration data, BLADJ ignores node ages that are older than the
¹⁹⁶ age of a parent node. BLADJ requires a root age estimate. If there is no information on the
¹⁹⁷ age of the root in the chronogram database, users can provide an estimate from the literature.
¹⁹⁸ If none is provided, DateLife will not return a dated topology but provide a warning message
¹⁹⁹ along with suggestions on how the user can provide an age for the root so DateLife can run.

²⁰⁰ Alternatively, the user can use the MrBayes (Huelsenbeck & Ronquist, 2001; Ronquist
²⁰¹ & Huelsenbeck, 2003) approach in DateLife uses the calibrations as priors on node ages.

²⁰² The latter methods can assign dates using a birth-death model (all of them????), they
²⁰³ require initial branch lengths and thus require more time and expertise to run, which makes
²⁰⁴ BLADJ the fastest and most practical method to obtain a dated tree.

²⁰⁵ In the case the user can provide or obtain branch length data along with a tree
²⁰⁶ topology, phylogenetic dating options supported in DateLife that incorporate branch length
²⁰⁷ information from the input topology in combination with the secondary calibrations:
²⁰⁸ PATHd8 is a non-clock, rate-smoothing method (Britton et al., 2007) to date trees; treePL
²⁰⁹ (Smith & O'Meara, 2012), is a semi-parametric, rate-smoothing, penalized likelihood dating
²¹⁰ method (Sanderson, 2002).

211 **Dating a tree with branch lengths**.– Topologies obtained from OpenTree and

212 with the supertree approach described above lack branch length data. Yet, phylogenetic

213 dating using branch lengths data is the golden standard for phylogenetic dating analyses, but

214 it is costly and requires a lot of human/expert curation.

215 A fast solution implemented in DateLife is as follows. To estimate branch lengths

216 proportional to substitution rates for these topologies, DateLife currently implements a

217 simple algorithm. First, it mines the Barcode of Life Data System, BOLD (Ratnasingham &

218 Hebert, 2007) to obtain genetic markers for the input taxa. Mined genetic sequences are

219 aligned with MUSCLE (Edgar, 2004) (by default) or MAFFT (Katoh, Asimenos, & Toh,

220 2009).

221 The BOLD sequence alignment is then used to reconstruct branch lengths with the

222 accelerated transformation (ACCTRAN) parsimony algorithm, which resolves ambiguous

223 character optimization, by assigning changes along branches of the tree as close to the root as

224 possible (Agnarsson & Miller, 2008). This algorithm work rally fast and allows getting initial

225 branch lengths that ar ethen optimized using ML. Optionally, the likelihood of the tree

226 topology, the alignment and the reconstructed branch lengths given different evolutionary

227 models, is computed using functions from the `phangorn` package (Schliep, 2011).

228 Relative branch length information provides key data for phylogenetic dating,

229 especially for nodes without secondary calibrations available. Yet, topologies without branch

230 lengths can also be dated.

231 **Visualizing results**.– Finally, users can save all source and summary chronograms in

232 formats that permit reuse and reanalysis (such as newick and the R “phylo” format), as well

233 as visualize and compare results graphically, or construct their own graphs using DateLife’s

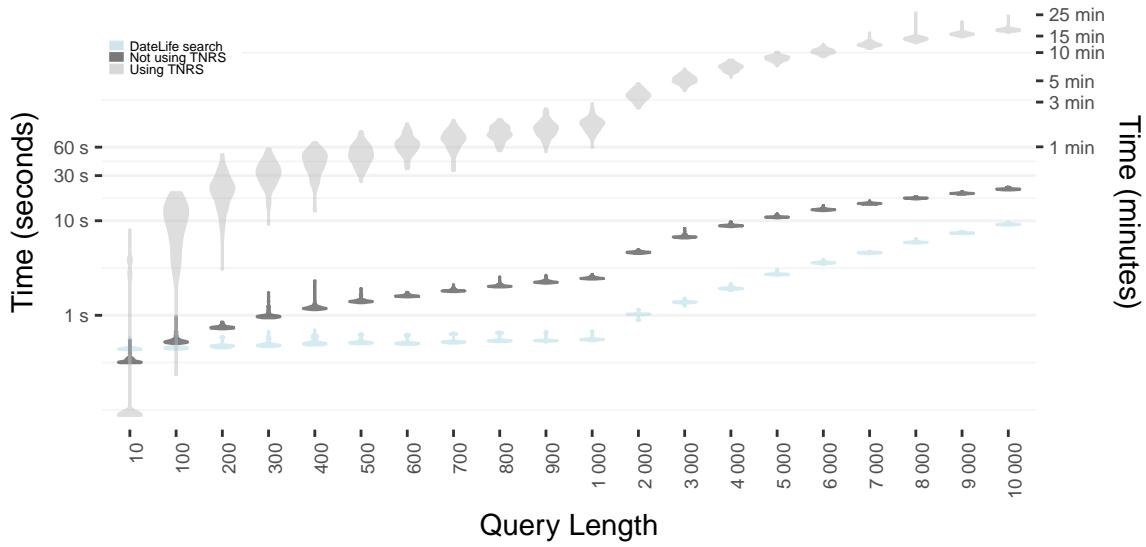
234 chronogram plot generation functions available from the R package `datelifeplot`

235 (Sanchez-Reyes & O’Meara, 2022).

236

Benchmark

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



246

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

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

261 Case studies

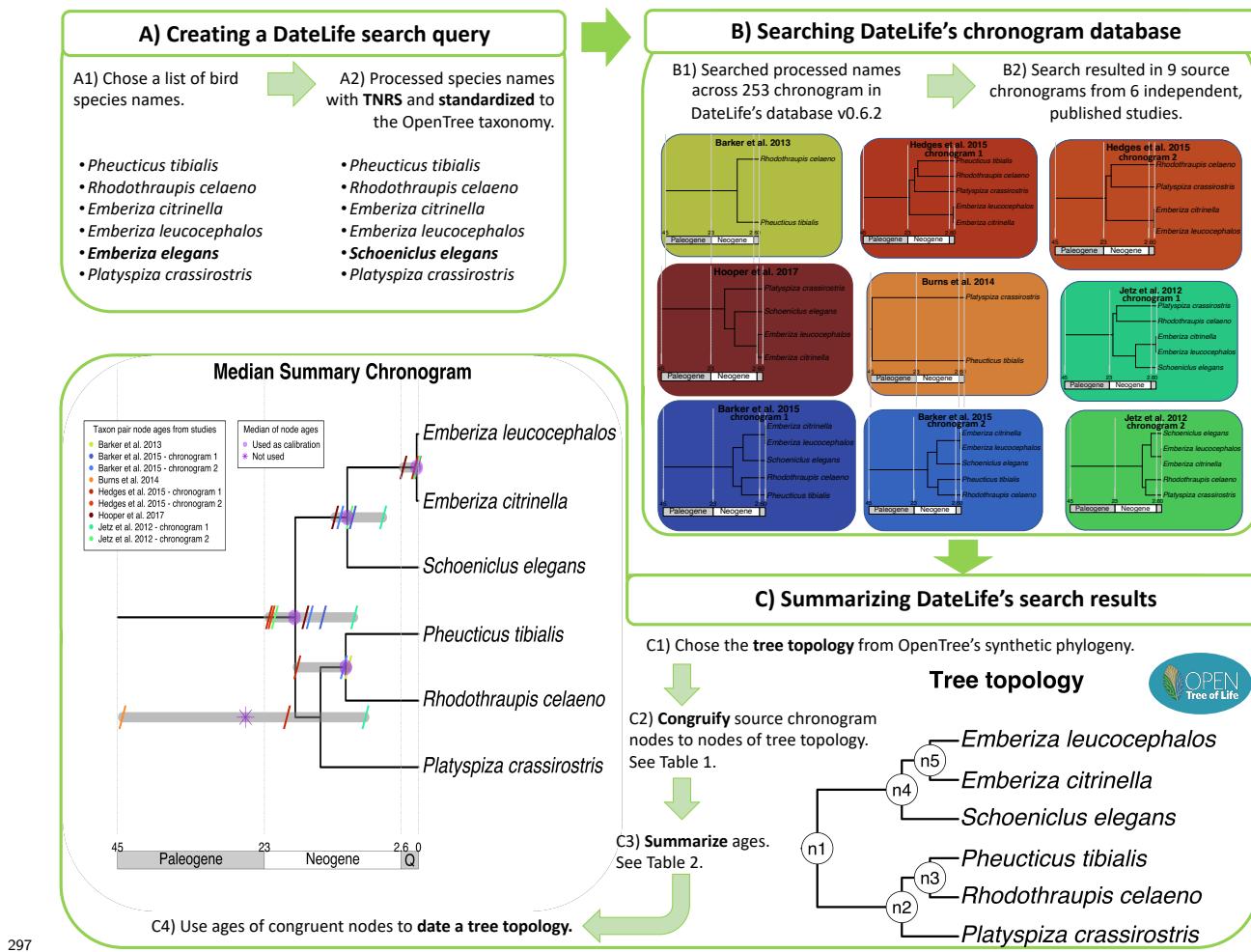
262 We illustrate the DateLife workflow using a family within the Passeriform birds
263 encompassing the true finches, Fringillidae, as case study. On a small example, we analysed 6
264 bird species, and results from each step of the workflow are shown in Fig. 3. As a second
265 example, we analysed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

268 A small example

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

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

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



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

305 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

306

307 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

308

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

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

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

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

³³⁶ the 289 species from the original query was extracted from Open Tree of Life's synthetic tree
³³⁷ v12.3 (Open Tree Of Life et al., 2019).

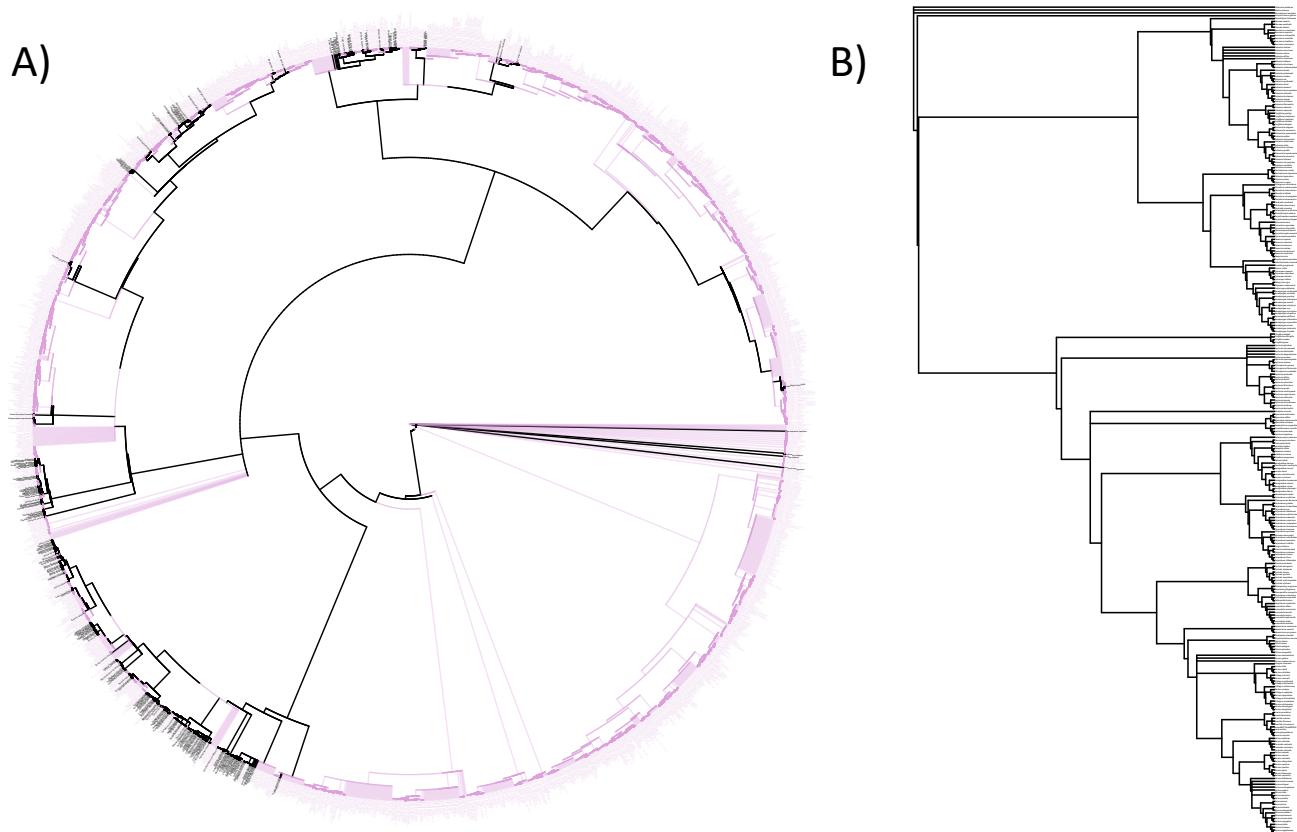
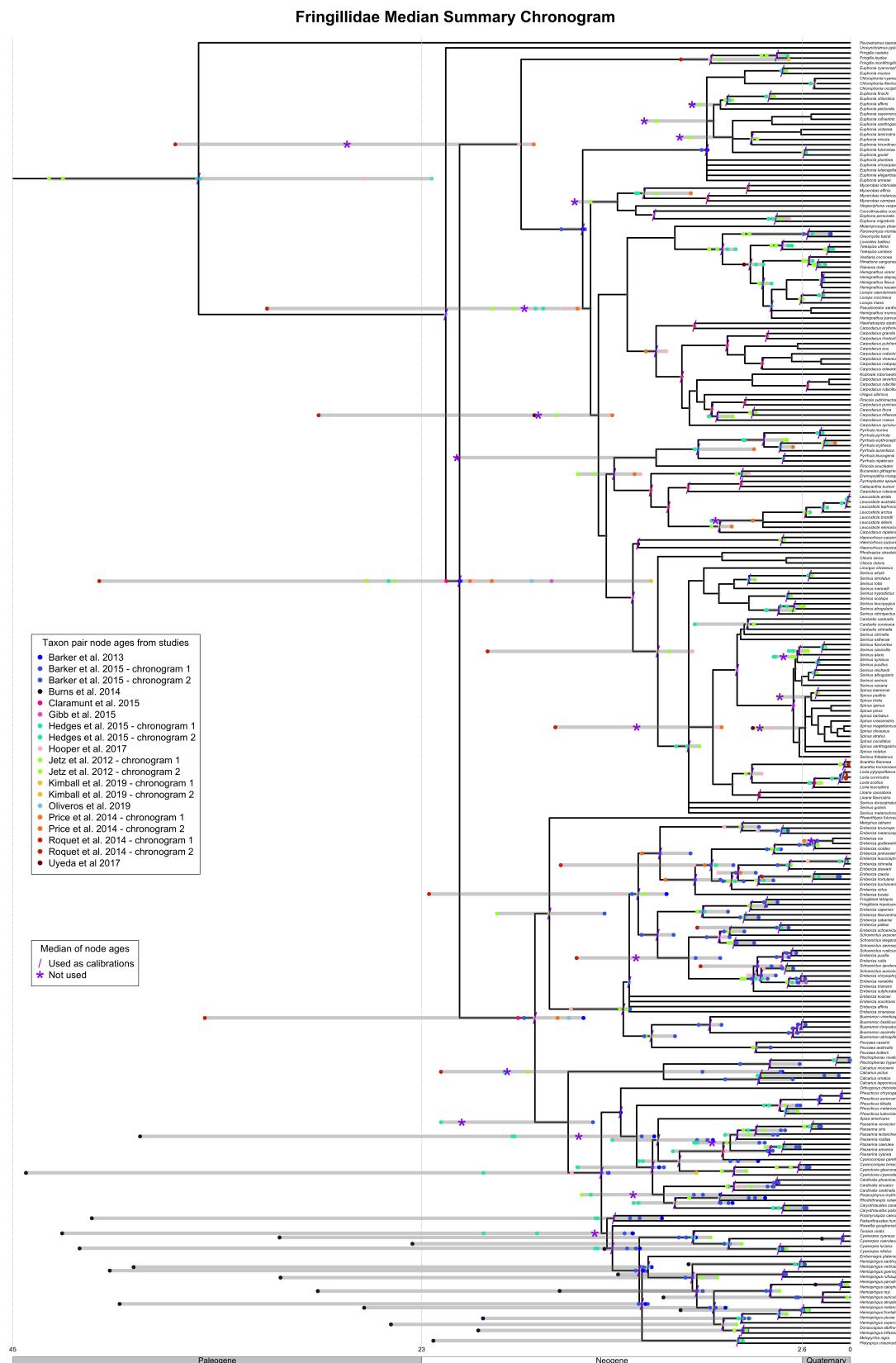


FIGURE 4. Tree topologies extracted from Open Tree of Life's (OpenTree) synthetic phylogenetic tree.
 338 A) Topology of 2,333 tips and 1,305 internal nodes, encompassing bird species within the family
 339 Fringillidae following the NCBI taxonomy (black), as well as all other bird species that share the same
 340 MRCA node in OpenTree's synthetic tree (purple). B) Subtree
 341 topology of 289 tips and 253 internal nodes, resulting from pruning species that do not belong to the
 342 family Fringillidae according to the NCBI taxonomy (purple branches in topology A). Bird species
 343 within the Fringillidae are paraphyletic (Alström et al. 2014, Barker, Cibois, Schikler, Feinstein, &
 344 Cracraft 2004, Barker et al. 2013, Barker 2014, Barker et al. 2015, Beresford, Barker, Ryan, &
 345 Crowe 2005, Bryson Jr et al. 2014, Burleigh, Kimball, & Braun 2015, Burns et al. 2014, Chaves,
 346 Hidalgo, & Klicka 2013, Claramunt & Cracraft 2015, Gibb et al. 2015, Hackett et al. 2008, Jetz et al.
 347 2012, Johansson, Fjeldså, & Bowi 200, Kimball et al. 2019, Klicka et al. 2014, Lamichhaney et al.
 348 2015, Lerner, Meyer, James, Hofreiter, & Fleischer 2011, Lovette et al. 2010, Moyle et al. 2016,
 349 Ödeen, Håstad, & Alström 2011, Oliveros et al. 2019, Päckert et al. 2012, Parchman, Benkman, &
 350 Mezquida 2007, Powell et al. 2014, Price et al. 2014, Pulgarín-R, Smith, Bryson Jr, Spellman, &
 351 Klicka 2013, Selvatti, Gonzaga, & Moraes Russo 2015, Tietze, Päckert, Martens, Lehmann, & Sun
 352 2013, Treplin et al. 2008, Zuccon, Prŷs-Jones, Rasmussen, & Ericson 2012).

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

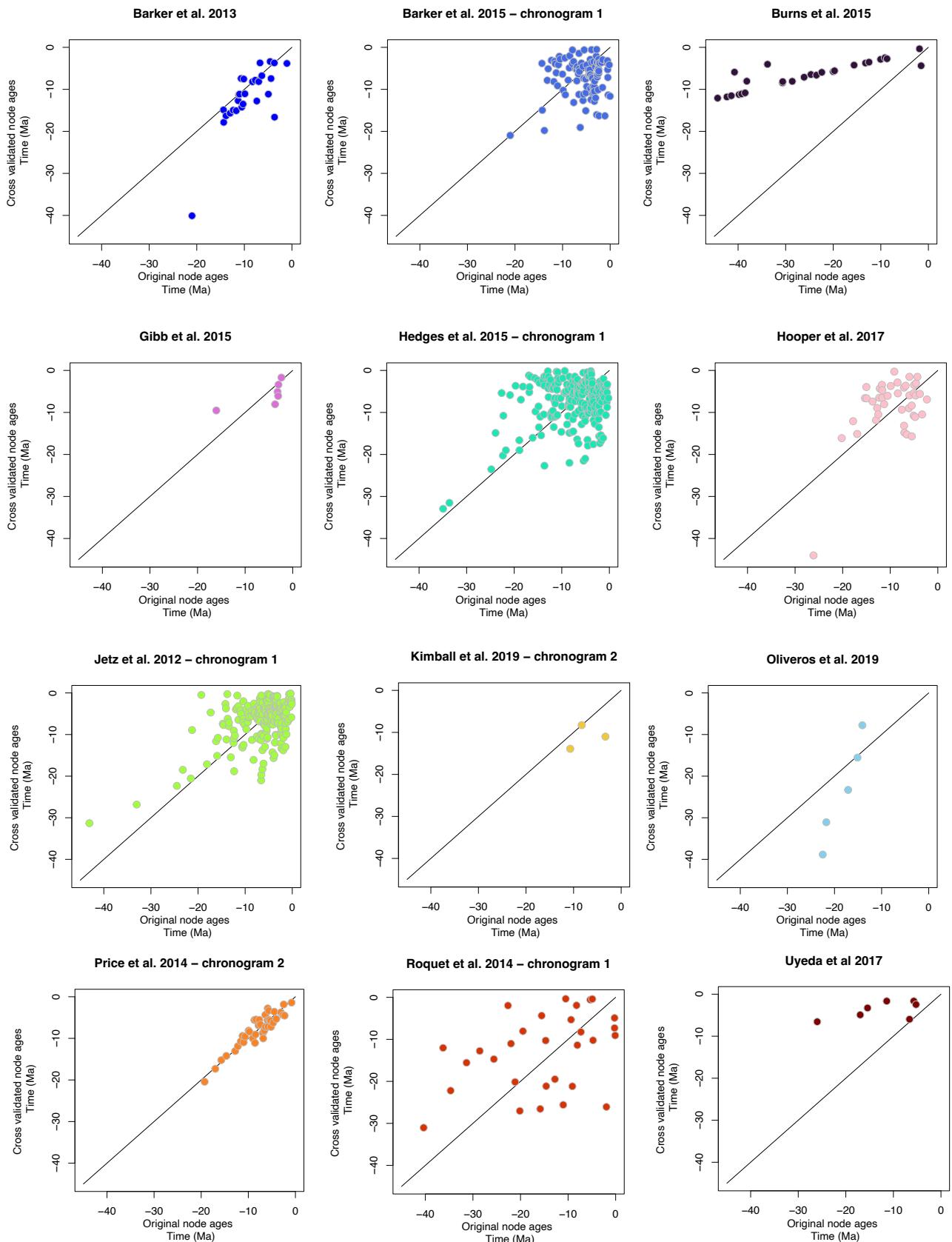


360 FIGURE 5. *Fringillidae median summary chronogram generated with DateLife*. It has 289 tips and
 361 253 nodes, from which 212 have age data from at least one published chronogram. These nodes have
 362 age data points from 19 different chronograms from 13 different studies (colored circles), which were

366

Cross-validation test

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



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

380

Discussion

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

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

395 **Age variation in source chronograms**

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

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

414 **Primary vs Secondary calibrations.** While most chronograms in DateLife’s database
415 are constructed using primary calibrations (molecular substitution rates or ages obtained
416 from the fossil record or geological events), DateLife summarizes chronograms using
417 secondary calibrations (ages coming from other chronograms). Graur and Martin (2004)
418 cautioned on the increased error and uncertainty in estimated ages when using secondary
419 calibrations in dating analyses. Schenk (2016) showed that, in simulations, divergence times
420 inferred using secondary calibrations are significantly younger than those inferred with
421 primary calibrations, when obtained with Bayesian inference methods, and when priors are
422 implemented in similar ways in both analyses. Accordingly, the scientific community seems
423 to have more confidence in chronograms obtained from a single analysis, using fossil data as
424 primary sources of calibrations (Schenk, 2016), and using fossils that have been widely
425 discussed and curated as calibrations to date other trees, making sure that all data reflect a
426 coherent evolutionary history (Sauquet, 2013), as for example done by Antonelli et al.
427 (2017). There have been attempts to create fossil calibration databases (Ksepka et al., 2015),
428 though these still have room to grow.

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

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

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

446 Sumarizing chronograms

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

454 Some measures that have been proposed are the proportion of lineage sampling and the
455 number of calibrations used (Magallón, 2010; Magallón et al., 2015). Some characteristics
456 that are often cited in published studies as a measure of improved age estimates as compared
457 to previously published estimates are: quality of alignment (missing data, GC content),
458 lineage sampling (strategy and proportion), phylogenetic and dating inference method,

459 number of fossils used as calibrations, support for nodes and ages, and magnitude of
460 confidence intervals.

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

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

475 Effects on downstream analyses

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

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

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

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

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

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

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

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

490 where only 23% had molecular data.

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

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

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

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

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

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

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

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

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

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

501 both parameter estimation and type I and II errors”

502 Notably, risks come with this practice.

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

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

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

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

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

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

509 results with prior work.

510

Conclusions

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

519 The DateLife project allows for easy and fast summary of public and state-of-the-art data on
520 time of lineage divergence. It provides a straightforward way to get an informed idea on the
521 state of knowledge of the time frame of evolution of different regions of the tree of life, and
522 allows identifying regions that require more research, or that have conflicting information. It
523 is available as an R package, and as a web-based R shiny application at www.datelife.org
524 Both summary and newly generated trees are useful to evaluate evolutionary hypotheses in
525 different areas of research. The DateLife project should improve awareness of the existing
526 variation in expert time of divergence data, and foster exploration of the effect of alternative
527 divergence time hypothesis on the results of analyses, nurturing a culture of more cautious
528 interpretation of evolutionary results.

529

Availability

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

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

540 **Supplementary Material**

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

547 **Funding**

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

551 **Acknowledgements**

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

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

References

- 564
- 565 566 Agnarsson, I., & Miller, J. A. (2008). Is acctran better than deltran? *Cladistics*, 24(6), 1032–1038.
- 567 568 Alström, P., Hooper, D. M., Liu, Y., Olsson, U., Mohan, D., Gelang, M., … Price, T. D. (2014). Discovery of a relict lineage and monotypic family of passerine birds. *Biology Letters*, 10(3), 20131067.
- 569
- 570 571 Ané, C., Eulenstein, O., Piaggio-Talice, R., & Sanderson, M. J. (2009). Groves of phylogenetic trees. *Annals of Combinatorics*, 13(2), 139–167.
- 572 573 Antonelli, A., Hettling, H., Condamine, F. L., Vos, K., Nilsson, R. H., Sanderson, M. J., … Vos, R. A. (2017). Toward a self-updating platform for estimating rates of speciation and migration, ages, and relationships of Taxa. *Systematic Biology*, 66(2), 153–166.
- 574
- 575 <https://doi.org/10.1093/sysbio/syw066>
- 576 577 Archie, J., Day, W. H., Felsenstein, J., Maddison, W., Meacham, C., Rohlf, F. J., & Swofford, D. (1986). The Newick tree format. Retrieved from {<https://evolution.genetics.washington.edu/phylip/newicktree.html>}
- 578
- 579 580 Avibase. (2022). Yellow-throated Bunting. *Avibase - the World Bird Database*, (Online Resource). Retrieved from {<https://avibase.bsc-eoc.org/species.jsp?lang=EN&avibaseid=82D1EE0049D8D927>}
- 581
- 582 583 Bapst, D. W. (2012). Paleotree: An R package for paleontological and phylogenetic analyses of evolution. *Methods in Ecology and Evolution*, 3(5), 803–807.
- 584 <https://doi.org/10.1111/j.2041-210X.2012.00223.x>
- 585 586 Barba-Montoya, J., Reis, M. dos, Schneider, H., Donoghue, P. C., & Yang, Z. (2018). Constraining uncertainty in the timescale of angiosperm evolution and the veracity of 587 a cretaceous terrestrial revolution. *New Phytologist*, 218(2), 819–834.
- 588 Barker, F. K. (2014). Mitogenomic data resolve basal relationships among passeriform and 589 passeridan birds. *Molecular Phylogenetics and Evolution*, 79, 313–324.
- 590 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2013). Going to

- 591 extremes: Contrasting rates of diversification in a recent radiation of new world
592 passerine birds. *Systematic Biology*, 62(2), 298–320.
- 593 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2015). New insights
594 into new world biogeography: An integrated view from the phylogeny of blackbirds,
595 cardinals, sparrows, tanagers, warblers, and allies. *The Auk: Ornithological Advances*,
596 132(2), 333–348.
- 597 Barker, F. K., Cibois, A., Schikler, P., Feinstein, J., & Cracraft, J. (2004). Phylogeny and
598 diversification of the largest avian radiation. *Proceedings of the National Academy of
599 Sciences*, 101(30), 11040–11045.
- 600 Beresford, P., Barker, F., Ryan, P., & Crowe, T. (2005). African endemics span the tree of
601 songbirds (passeri): Molecular systematics of several evolutionary “enigmas”.
602 *Proceedings of the Royal Society B: Biological Sciences*, 272(1565), 849–858.
- 603 Boyle, B., Hopkins, N., Lu, Z., Raygoza Garay, J. A., Mozzherin, D., Rees, T., ... Enquist,
604 B. J. (2013). The taxonomic name resolution service: An online tool for automated
605 standardization of plant names. *BMC Bioinformatics*, 14(1).
606 <https://doi.org/10.1186/1471-2105-14-16>
- 607 Britton, T., Anderson, C. L., Jacquet, D., Lundqvist, S., & Bremer, K. (2007). Estimating
608 Divergence Times in Large Phylogenetic Trees. *Systematic Biology*, 56(788777878),
609 741–752. <https://doi.org/10.1080/10635150701613783>
- 610 Bryson Jr, R. W., Chaves, J., Smith, B. T., Miller, M. J., Winker, K., Pérez-Emán, J. L., &
611 Klicka, J. (2014). Diversification across the new world within the ‘blue’cardinalids
612 (aves: Cardinalidae). *Journal of Biogeography*, 41(3), 587–599.
- 613 Burleigh, J. G., Kimball, R. T., & Braun, E. L. (2015). Building the avian tree of life using a
614 large-scale, sparse supermatrix. *Molecular Phylogenetics and Evolution*, 84, 53–63.
- 615 Burns, K. J., Shultz, A. J., Title, P. O., Mason, N. A., Barker, F. K., Klicka, J., ... Lovette,
616 I. J. (2014). Phylogenetics and diversification of tanagers (passeriformes:
617 Thraupidae), the largest radiation of neotropical songbirds. *Molecular Phylogenetics*

618 and *Evolution*, 75, 41–77.

619 Chamberlain, S. (2018). *bold: Interface to Bold Systems API*. Retrieved from
620 <https://CRAN.R-project.org/package=bold>

621 Chamberlain, S. A., & Szöcs, E. (2013). taxize : taxonomic search and retrieval in R [version
622 2; referees: 3 approved]. *F1000Research*, 2(191), 1–29.
623 <https://doi.org/10.12688/f1000research.2-191.v2>

624 Chaves, J. A., Hidalgo, J. R., & Klicka, J. (2013). Biogeography and evolutionary history of
625 the n eotropical genus s altator (a ves: T hraupini). *Journal of Biogeography*, 40(11),
626 2180–2190.

627 Claramunt, S., & Cracraft, J. (2015). A new time tree reveals earth history's imprint on the
628 evolution of modern birds. *Science Advances*, 1(11), e1501005.

629 Criscuolo, A., Berry, V., Douzery, E. J., & Gascuel, O. (2006). SDM: A fast distance-based
630 approach for (super)tree building in phylogenomics. *Systematic Biology*, 55(5),
631 740–755. <https://doi.org/10.1080/10635150600969872>

632 Cusimano, N., Stadler, T., & Renner, S. S. (2012). A new method for handling missing
633 species in diversification analysis applicable to randomly or nonrandomly sampled
634 phylogenies. *Systematic Biology*, 61(5), 785–792.

635 Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M.-K., Turon, X., ... Douzery,
636 E. J. (2018). A phylogenomic framework and timescale for comparative studies of
637 tunicates. *BMC Biology*, 16(1), 1–14.

638 Eastman, J. M., Harmon, L. J., & Tank, D. C. (2013). Congruification: Support for time
639 scaling large phylogenetic trees. *Methods in Ecology and Evolution*, 4(7), 688–691.
640 <https://doi.org/10.1111/2041-210X.12051>

641 Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high
642 throughput. *Nucleic Acids Research*, 32(5), 1792–1797.

643 Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American Naturalist*,
644 125(1), 1–15. Retrieved from <http://www.jstor.org/stable/2461605>

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

of birds in space and time. *Nature*, 491(7424), 444–448.

<https://doi.org/10.1038/nature11631>

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

- 699 Córdoba-Córdoba, S., ... others. (2010). A comprehensive multilocus phylogeny for
700 the wood-warblers and a revised classification of the parulidae (aves). *Molecular*
701 *Phylogenetics and Evolution*, 57(2), 753–770.
- 702 Magallon, S., & Sanderson, M. (2001). Absolute diversification rates in angiosperm clades.
703 *Evolution*, 55(9), 1762–1780.
- 704 Magallón, S. (2010). Using fossils to break long branches in molecular dating: A comparison
705 of relaxed clocks applied to the origin of angiosperms. *Systematic Biology*, 59(4),
706 384–399.
- 707 Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., & Hernández-Hernández, T. (2015).
708 A metacalibrated time-tree documents the early rise of flowering plant phylogenetic
709 diversity. *New Phytologist*, 207(2), 437–453.
- 710 McTavish, E. J., Hinchliff, C. E., Allman, J. F., Brown, J. W., Cranston, K. A., Holder, M.
711 T., ... Smith, S. (2015). Phylesystem: A git-based data store for community-curated
712 phylogenetic estimates. *Bioinformatics*, 31(17), 2794–2800.
- 713 Michonneau, F., Brown, J. W., & Winter, D. J. (2016). rotl: an R package to interact with
714 the Open Tree of Life data. *Methods in Ecology and Evolution*, 7(12), 1476–1481.
715 <https://doi.org/10.1111/2041-210X.12593>
- 716 Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecology Letters*,
717 17(4), 508–525. <https://doi.org/10.1111/ele.12251>
- 718 Moyle, R. G., Oliveros, C. H., Andersen, M. J., Hosner, P. A., Benz, B. W., Manthey, J. D.,
719 ... Faircloth, B. C. (2016). Tectonic collision and uplift of Wallacea triggered the
720 global songbird radiation. *Nature Communications*, 7(1), 1–7.
- 721 Oliveros, C. H., Field, D. J., Ksepka, D. T., Barker, F. K., Aleixo, A., Andersen, M. J., ...
722 others. (2019). Earth history and the passerine superradiation. *Proceedings of the*
723 *National Academy of Sciences*, 116(16), 7916–7925.
- 724 Ooms, J., & Chamberlain, S. (2018). *Phylocomr: Interface to 'phylocom'*. Retrieved from
725 <https://CRAN.R-project.org/package=phylocomr>

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

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

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

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

- 834 Vos, R. A., Balhoff, J. P., Caravas, J. A., Holder, M. T., Lapp, H., Maddison, W. P., ...
835 others. (2012). NeXML: Rich, extensible, and verifiable representation of
836 comparative data and metadata. *Systematic Biology*, 61(4), 675–689.
837 <https://doi.org/10.1093/sysbio/sys025>
- 838 Webb, C. (2000). Exploring the Phylogenetic Structure of Ecological Communities : An
839 Example for Rain Forest Trees. *The American Naturalist*, 156(2), 145–155.
- 840 Webb, C., Ackerly, D., & Kembel, S. (2008). Phylocom: Software for the analysis of
841 phylogenetic community structure and trait evolution. *Bioinformatics*, 24(18),
842 2098–2100. <https://doi.org/10.1093/bioinformatics/btn358>
- 843 Webb, C., & Donoghue, M. (2005). Phylomatic: Tree assembly for applied phylogenetics.
844 *Molecular Ecology Notes*, 5(1), 181–183.
- 845 Weir, J., & Schluter, D. (2008). Calibrating the avian molecular clock. *Molecular Ecology*,
846 17(10), 2321–2328.
- 847 Zuccon, D., Prŷs-Jones, R., Rasmussen, P. C., & Ericson, P. G. (2012). The phylogenetic
848 relationships and generic limits of finches (fringillidae). *Molecular Phylogenetics and*
849 *Evolution*, 62(2), 581–596.