

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

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

³ ¹ University of California, Merced, USA

⁴ ² University of Tennessee, Knoxville, USA

⁵ Author Note

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

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

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

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

17

Abstract

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

42

Keywords: Tree; Phylogeny; Scaling; Dating; Ages; Divergence times; Open Science;

⁴³ Congruification; Supertree; Calibrations; Secondary calibrations

⁴⁴ Word count: 4960

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

46 **Abstract**

47 **Introduction**

48 Achieving a high-quality reconstruction of a phylogenetic tree with branch lengths
49 proportional to absolute time (chronogram) is a difficult and time-consuming task. But the
50 increased availability of fossil and molecular data, and time-efficient analytical techniques
51 has resulted in many recent publications of large chronograms for a large number and wide
52 diversity of organisms. Knowledge of the evolutionary time frame of organisms is key for
53 research in the natural sciences. It also represents valuable information for education,
54 science communication, and policy decisions. When chronograms are shared in public and
55 open databases, this wealth of expertly curated and peer-reviewed data on evolutionary
56 timeframe is exposed in a programmatic and reusable way, as intensive and localized efforts
57 have improved data sharing practices, as well as incentivized open science in biology.
58 Here we present DateLife, a service implemented as an R package and an R Shiny website
59 application available at www.datelife.org, that provides functionalities for efficient and easy
60 finding, summary, reuse, and reanalysis of expert, peer-reviewed, public data on time frame
61 of evolution. The main DateLife workflow constructs a chronogram for any given
62 combination of taxon names by searching a local chronogram database constructed and
63 curated from the Open Tree of Life Phylesystem phylogenetic database, which incorporates
64 phylogenetic data from the TreeBASE database as well. We implement and test methods
65 for summarizing time data from multiple source chronograms using supertree and
66 congruification algorithms, and using age data extracted from source chronograms as
67 secondary calibration points to add branch lengths proportional to absolute time to a tree
68 topology. DateLife will be useful to increase awareness of the existing variation in
69 alternative hypotheses of evolutionary time for the same organisms, and can foster
70 exploration of the effect of alternative evolutionary timing hypotheses on the results of

71 downstream analyses, providing a framework for a more informed interpretation of
72 evolutionary results.

73 *Keywords:* Tree; Phylogeny; Scaling; Dating; Ages; Divergence times; Open Science;
74 Congruification; Supertree; Calibrations; Secondary calibrations

75 Word count: 5393

76 Chronograms –phylogenies with branch lengths proportional to time– provide key data
77 on evolutionary time frame for the study of natural processes in many areas of biological
78 research, such as developmental biology (Delsuc et al., 2018; Laubichler & Maienschein,
79 2009), conservation biology (Felsenstein, 1985; Webb, 2000), historical biogeography
80 (Posadas, Crisci, & Katinas, 2006), and species diversification (Magallon & Sanderson, 2001;
81 Morlon, 2014).

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

88 Here we present the DateLife project which has the main goal of capturing age data
89 from published chronograms, and making these data readily accessible to the community for
90 reuse and reanalysis, for research, teaching, and science communication and policy.

91 DateLife’s core software application is available as an R package ([Sanchez-Reyes et al.,
92 2022???](#)), and as an online Rshiny interactive website at www.datelife.org. It features key
93 elements for scientific reproducibility, such as a versioned, open and fully public source
94 database (McTavish et al., 2015), data stored and available in a computer readable format
95 (Vos et al., 2012), automated and programmatic ways of accessing the data (Stoltzfus et al.,

96 2013) and methods to summarize and compare the data.

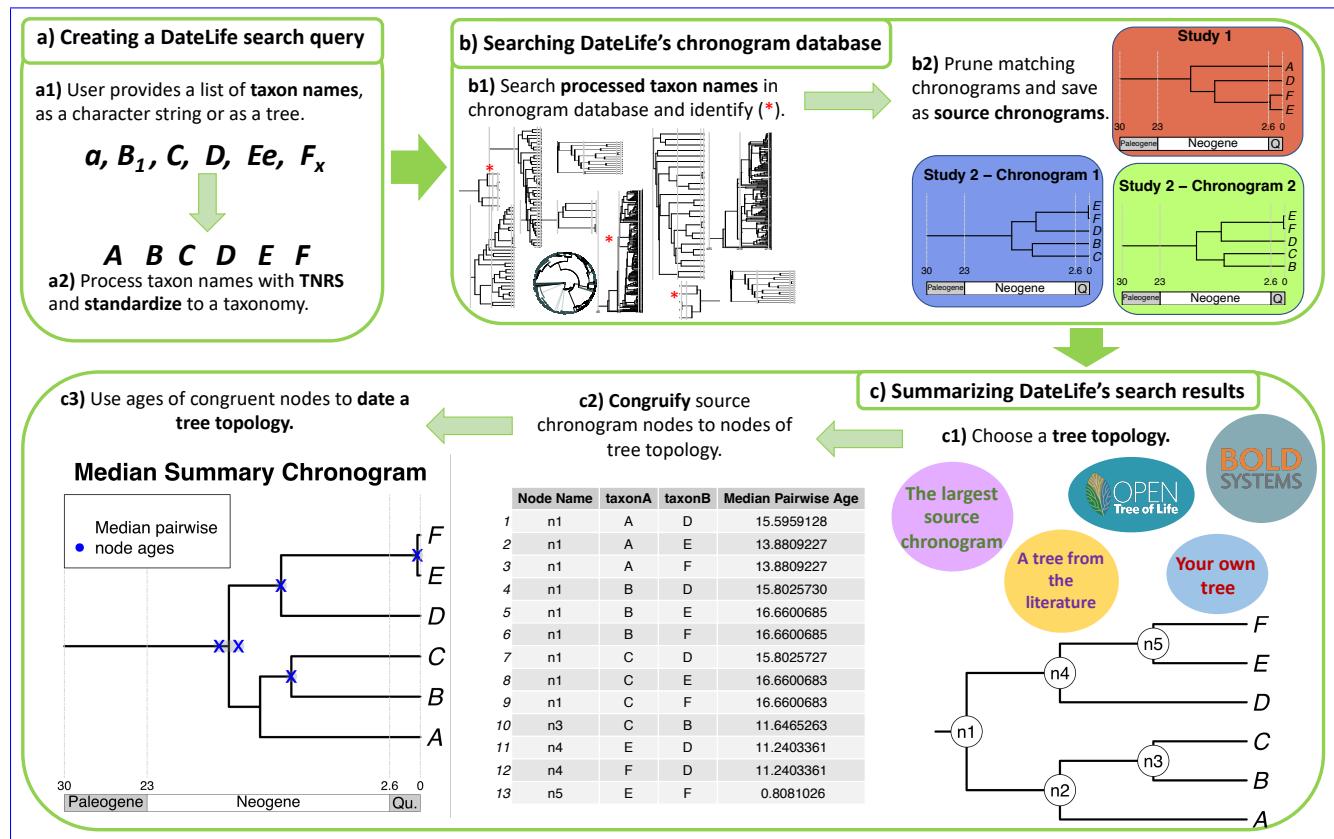
97 **DESCRIPTION**

98 **Description**

99 DateLife's core software application consists of the R package `datelife`. Its current
100 stable version – v0.6.6.8, is available from [the](#) The Comprehensive R Archive Network
101 (CRAN) repository ([Sanchez-Reyes et al., 2022???](#)), and relies on functionalities from
102 various biological R packages: ape (Paradis, Claude, & Strimmer, 2004), bold (Chamberlain,
103 2018), geiger (Pennell et al., 2014), paleotree (Bapst, 2012), phyloch (Heibl, 2008),
104 phylocomr (Ooms & Chamberlain, 2018), phytools (Revell, 2012), rotl (Michonneau, Brown,
105 & Winter, 2016), and taxize (Chamberlain, 2018; Chamberlain & Szöcs, 2013). Figure 1
106 provides a graphical summary of the three main steps of the DateLife workflow: creating a
107 search query, searching a database, and summarizing results from the search.

111 *Creating a Search Query*

112



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

113 Creating a search query

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

119 DateLife accepts scientific names that can belong to any inclusive taxonomic group
 120 (e.g., genus, family, tribe, etc.) or a binomial species name. Subspecies and variants are
 121 ignored. If an input taxon name belongs to an inclusive taxonomic group, DateLife has two

122 alternative behaviors defined by the “get species from taxon” flag. If the flag is active,
123 DateLife retrieves all species names within the inclusive taxonomic group following a
124 standard taxonomy of choice, and adds them to the input string. Taxonomies currently
125 supported by DateLife are Open Tree of Life (OpenTree) unified Taxonomy (OTT, Rees &
126 Cranston, 2017), the National Center of Biotechnology Information (NCBI) taxonomic
127 database (Schoch et al., 2020), the Global Biodiversity Information Facility (GBIF)
128 taxonomic backbone (GBIF Secretariat, 2022), and the Interim Register of Marine and
129 Nonmarine Genera (IRMNG) database (Rees et al., 2017). If the flag is inactive, DateLife
130 excludes any taxon names above the species level from the search query.

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

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

142 *Searching a Chronogram Database*

143 **Searching a chronogram database**

144 At the time of writing of this manuscript (Jun 22, 2022), DateLife’s chronogram
145 database latest version consist of 253 chronograms published in 187 different studies. It is
146 curated from OpenTree’s phylogenetic database, the Phylesystem, which constitutes an open

147 source of expert and peer-reviewed phylogenetic knowledge with rich metadata (McTavish et
148 al., 2015), which allows automatic and reproducible assembly of our chronogram database.
149 Datelife's chronogram database is navigable as an R data object within the **datelife** R
150 package.

151 A unique feature of the Phylesystem is that any user can add new published,
152 state-of-the-art chronograms any time, through their curator application
153 (<https://tree.opentreeoflife.org/curator>). As chronograms are added to Phylesystem, they are
154 incorporated into the chronogram database of the **datelife** package. The updated database
155 is assigned a new version number, followed by a package release on CRAN. **datelife**'s
156 chronogram database is updated as new chronogram data is added to Phylesystem, at a
157 minimum of once a month and a maximum of every 6 months. Users can also implement
158 functions from the **datelife** R package to trigger an update of the local chronogram
159 database, to incorporate any new chronograms to the user's DateLife analysis before an
160 official database update is released on CRAN.

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

171 *Summarizing Search Results*

172 **Summarizing search results**

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

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

186 *Choosing a Topology*[Choosing a topology.](#)

187 DateLife requires a tree topology to summarize age data upon. We recommend that
188 users provide ~~as input~~ a tree topology [as input](#) from the literature, or one of their own
189 making. If no topology is provided, DateLife automatically extracts one from the OpenTree
190 synthetic tree, a phylogeny encompassing 2.3 million taxa across all life, assembled from 1,
191 239 published phylogenetic trees and OpenTree’s unified Taxonomy, OTT (Open Tree Of
192 Life et al., 2019). Alternatively, DateLife can combine topologies from source chronograms
193 using a supertree approach. To combine topologies from source chronograms into a single
194 summary (or supertree) topology, the DateLife workflow identifies the source chronograms
195 that form a grove, roughly, a sufficiently overlapping set of taxa between trees, by
196 implementing definition 2.8 for n-overlap from Ané et al. (2009). In rare cases, a group of

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

200 *Dating the Topology* Dating the topology.–

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

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

220 For each congruent node, the pairwise distances that traverse that node are
221 summarized into a single summary matrix using classic summary statistics (i.e., mean,

median, minimum and maximum ages), and the Supermatrix Distance Method (SDM; Criscuolo, Berry, Douzery, & Gascuel, 2006), which deforms patristic distance matrices by minimizing variance and then averaging them. These single summary taxon pair age matrices (Summarized calibrations) can be applied as calibrations to date a tree topology, using different dating methods currently supported within DateLife: MrBayes (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003), PATHd8 (Britton, Anderson, Jacquet, Lundqvist, & Bremer, 2007), BLADJ (Webb, Ackerly, & Kembel, 2008; Webb & Donoghue, 2005), and treePL (Smith & O'Meara, 2012).

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

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

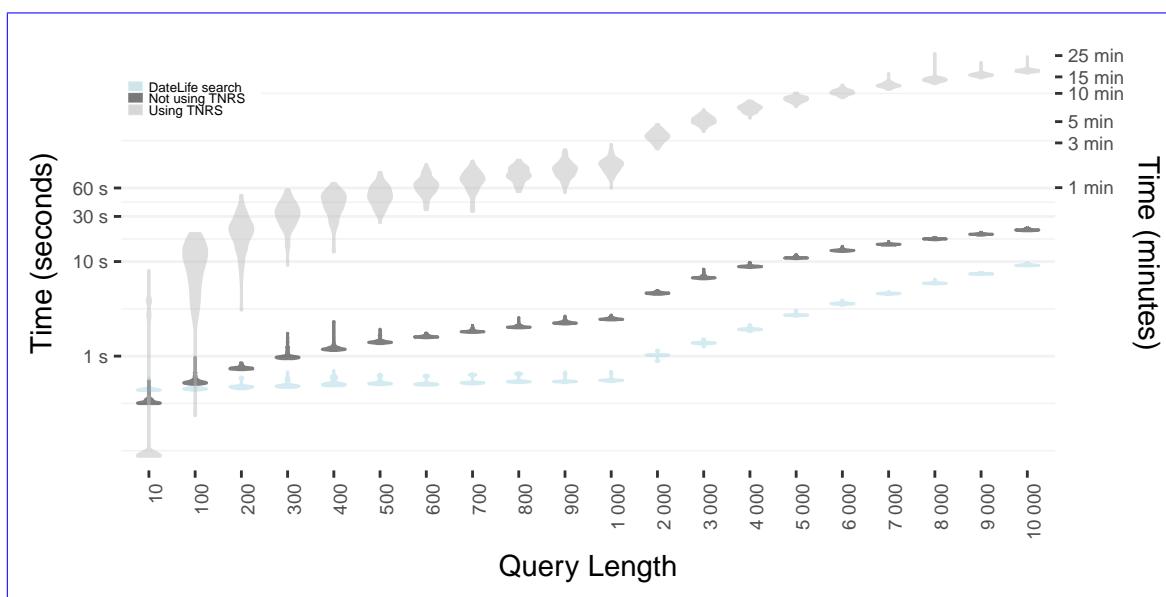
Visualizing Results Visualizing results.—

Finally, users can save all source and summary chronograms in 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).

252 BENCHMARK

253 Benchmark

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



263

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

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

278 **CASE STUDIES**

279 Case studies

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

286 **A Small Example**

287 **A small example**

288 **Creating a search query.**

289 Creating a search query. We chose 6 bird species within the Passeriformes. The sample
290 includes two species of cardinals: the black-thighed grosbeak – *Pheucticus tibialis* and the
291 crimson-collared grosbeak – *Rhodothraupis celaeno*; three species of buntings: the
292 yellowhammer – *Emberiza citrinella*, the pine bunting – *Emberiza leucocephalos* and the
293 yellow-throated bunting – *Emberiza elegans*; and one species of tanager, the vegetarian finch –
294 *Platyspiza crassirostris*. Processing of input names found that *Emberiza elegans* is synonym
295 for *Schoeniclus elegans* in the default reference taxonomy (OTT v3.3, June 1, 2021). For a
296 detailed discussion on the state of the synonym, refer to Avibase (Avibase, 2022; Lepage,
297 2004; Lepage, Vaidya, & Guralnick, 2014). Discovering this synonym allowed assigning five
298 age data points for the parent node of *Emberiza elegans*, shown as *Schoeniclus elegans* in
299 figure 3A, which would not have had any data otherwise.

300 ***Searching the database.***

301 **Searching the database.** DateLife used the processed input names to search the local
302 chronogram database and found 9 matching chronograms in 6 different studies (Fig. 3B).
303 Three studies matched five input names (Barker, Burns, Klicka, Lanyon, & Lovette, 2015;
304 Hedges, Marin, Suleski, Paymer, & Kumar, 2015; Jetz, Thomas, Joy, Hartmann, & Mooers,
305 2012), one study matched four input names (Hooper & Price, 2017) and two studies matched
306 two input names (Barker, Burns, Klicka, Lanyon, & Lovette, 2013; Burns et al., 2014). No
307 studies matched all input names. Together, source chronograms provide 28 unique age data
308 points, covering all nodes on our chosen tree topology to date (Table 1).

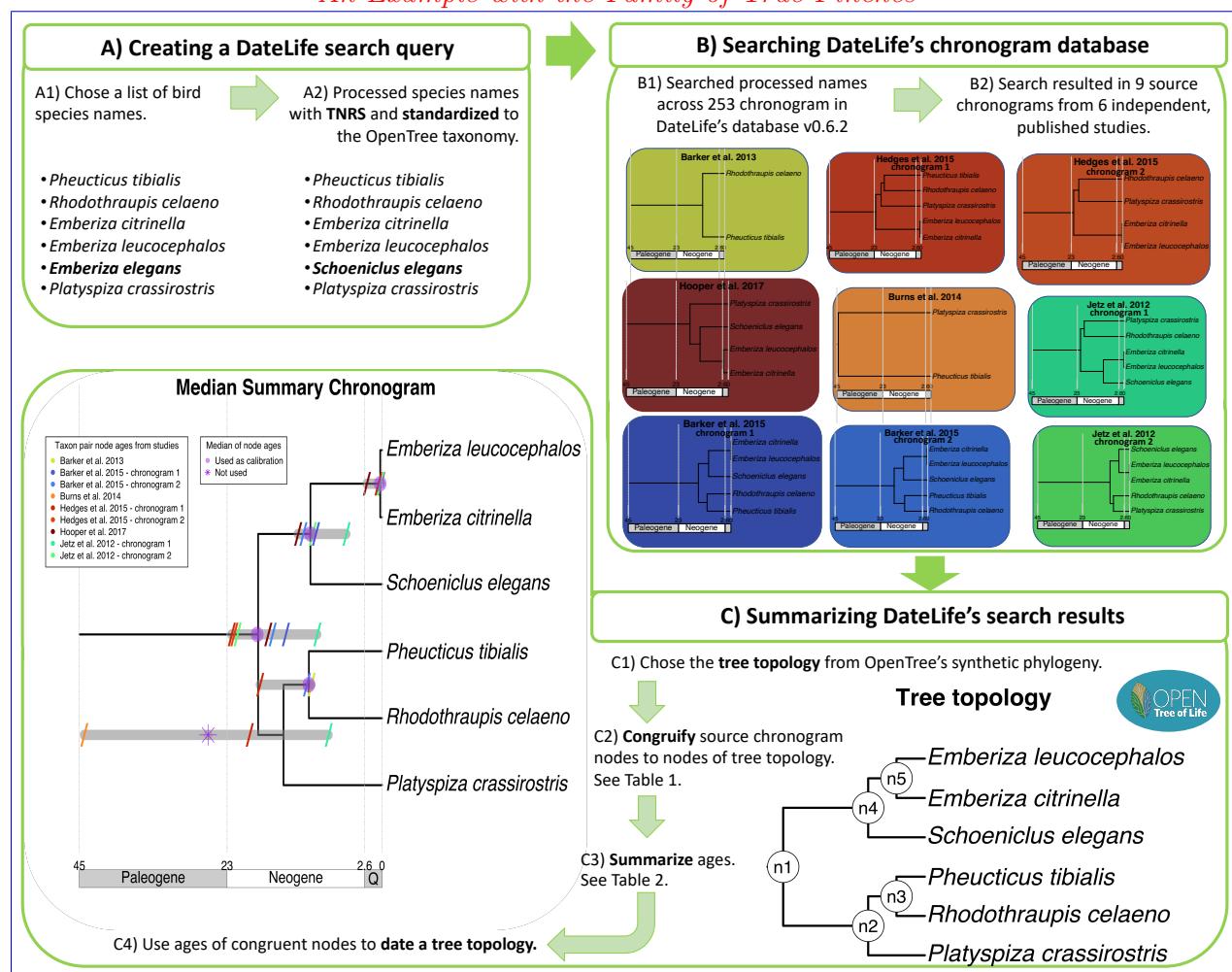
309 ***Summarizing search results.***

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

319

An Example with the Family of True Finches

320



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

328 ***Creating a query.***

329 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

330

331 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

332

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

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

342 ***Searching the database.***

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

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

352 ***Summarizing search results.***

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

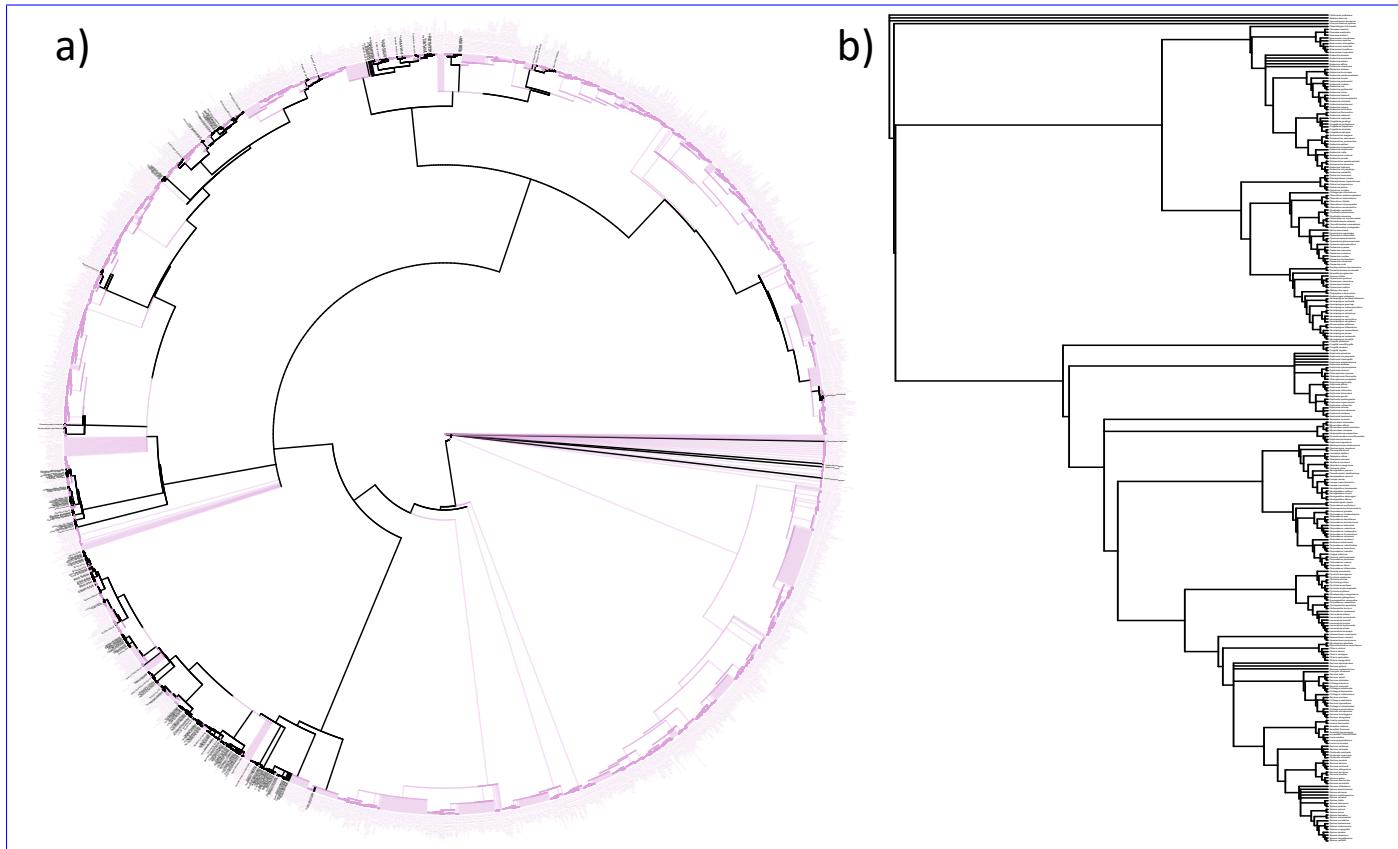


FIGURE 4. Tree topologies extracted from Open Tree of Life's (OpenTree) synthetic phylogenetic tree.
 360 A) Topology of 2,333 tips and 1,305 internal nodes, encompassing bird species within the family
 361 Fringillidae following the NCBI taxonomy (black), as well as all other bird species that share the same
 362 Most Recent Common Ancestor (MRCA) node in OpenTree's synthetic tree (purple). B) Subtree
 363 topology of 289 tips and 253 internal nodes, resulting from pruning species that do not belong to the
 364 family Fringillidae according to the NCBI taxonomy (purple branches in topology A). Bird species
 365 within the Fringillidae are paraphyletic (Alström et al. 2014, Barker, Cibois, Schikler, Feinstein, &
 366 Cracraft 2004, Barker et al. 2013, Barker 2014, Barker et al. 2015, Beresford, Barker, Ryan, &
 367 Crowe 2005, Bryson Jr et al. 2014, Burleigh, Kimball, & Braun 2015, Burns et al. 2014, Chaves,
 368 Hidalgo, & Klicka 2013, Claramunt & Cracraft 2015, Gibb et al. 2015, Hackett et al. 2008, Jetz et al.
 369 2012, Johansson, Fjeldså, & Bowi 200, Kimball et al. 2019, Klicka et al. 2014, Lamichhaney et al.
 370 2015, Lerner, Meyer, James, Hofreiter, & Fleischer 2011, Lovette et al. 2010, Moyle et al. 2016,
 371 Ödeen, Håstad, & Alström 2011, Oliveros et al. 2019, Päckert et al. 2012, Parchman, Benkman, &
 372 Mezquida 2007, Powell et al. 2014, Price et al. 2014, Pulgarín-R, Smith, Bryson Jr, Spellman, &
 373 Klicka 2013, Selvatti, Gonzaga, & Moraes Russo 2015, Tietze, Päckert, Martens, Lehmann, & Sun
 374 2013, Treplin et al. 2008, Zuccon, Prŷs-Jones, Rasmussen, & Ericson 2012).

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

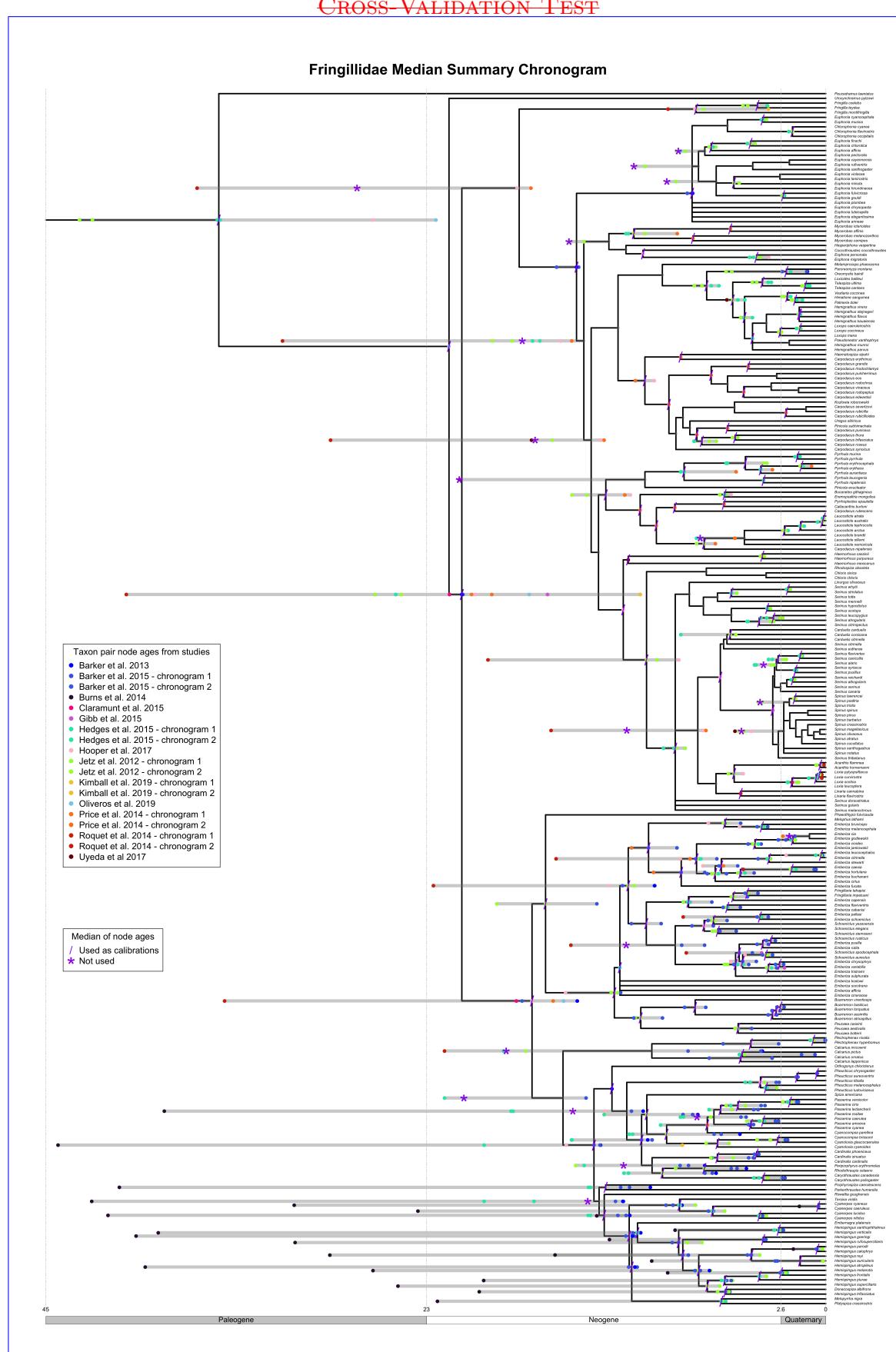


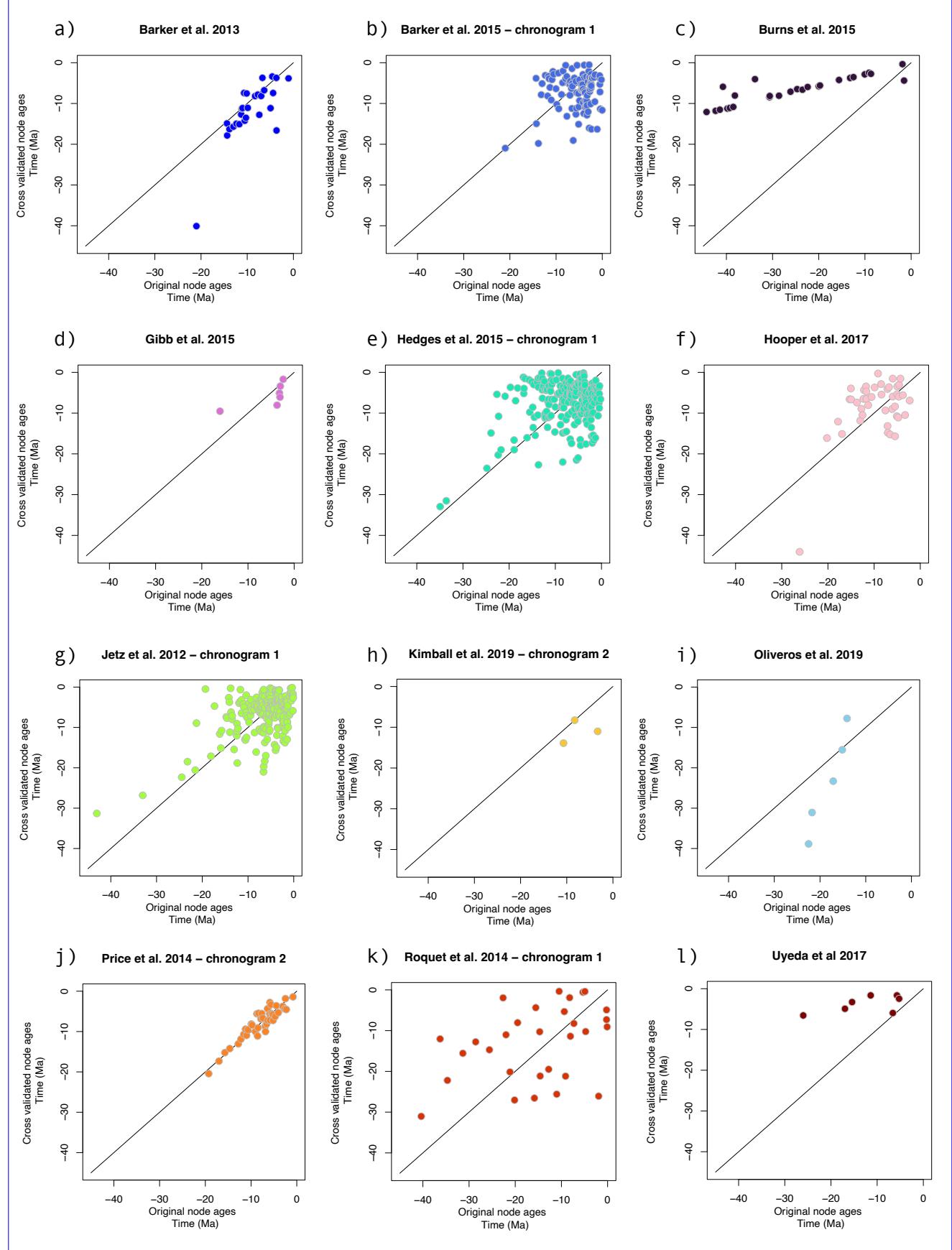
FIGURE 5. *Fringillidae* median summary chronogram generated with DateLife. It has 289 tips and 253 nodes, from which 212 have age data from at least one published chronogram. These nodes have

388

Cross-validation test

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

DISCUSSION



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

402

Discussion

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

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

417 *Age Variation in Source Chronograms*

418 **Age variation in source chronograms**

419 Conflict in estimated ages among alternative studies is common in the literature. See, for
420 example, the robust ongoing debate about crown group age of angiosperms (Barba-Montoya,
421 Reis, Schneider, Donoghue, & Yang, 2018; Magallón, Gómez-Acevedo, Sánchez-Reyes, &
422 Hernández-Hernández, 2015; Ramshaw et al., 1972; Sanderson & Doyle, 2001; Sauquet,
423 Ramírez-Barahona, & Magallón, 2021). Source chronograms available for the same
424 organisms have potentially been estimated implementing calibrations very differently. For
425 example, the chronograms from Burns et al. (2014) were inferred using molecular
426 substitution rate estimates across birds (Weir & Schluter, 2008), and have much older age
427 estimates for the same nodes than chronograms that were inferred using fossils as

428 calibrations (Figs. 5, 6; Supplementary Figs. S1, S5).
429 Different calibration implementations might also imply fundamentally distinct evolutionary
430 hypotheses (Antonelli et al., 2017). For example, two independent researchers working on
431 the same clade should both carefully select and justify their choices of fossil calibration
432 placement. Yet, if one researcher concludes that a fossil should calibrate the ingroup of a
433 clade, while another researcher concludes that the same fossil should calibrate the outgroup
434 of the clade, the resulting age estimates will differ, as the placement of calibrations as stem
435 or crown group has been proven to significantly affect time of lineage divergence estimates
436 (Sauquet, 2013).

437 *Primary vs Secondary Calibrations*

438 **Primary vs Secondary calibrations.** While most chronograms in DateLife's database
439 are constructed using primary calibrations (molecular substitution rates or ages obtained
440 from the fossil record or geological events), DateLife summarizes chronograms using
441 secondary calibrations (ages coming from other chronograms). Graur and Martin (2004)
442 cautioned on the increased error and uncertainty in estimated ages when using secondary
443 calibrations in dating analyses. Schenk (2016) showed that, in simulations, divergence times
444 inferred using secondary calibrations are significantly younger than those inferred with
445 primary calibrations, when obtained with Bayesian inference methods, and when priors are
446 implemented in similar ways in both analyses. Accordingly, the scientific community seems
447 to have more confidence in chronograms obtained from a single analysis, using fossil data as
448 primary sources of calibrations (Schenk, 2016), and using fossils that have been widely
449 discussed and curated as calibrations to date other trees, making sure that all data reflect a
450 coherent evolutionary history (Sauquet, 2013), as for example done by Antonelli et al.
451 (2017). There have been attempts to create fossil calibration databases (Ksepka et al., 2015),
452 though these still have room to grow.
453 It seems that using several (as opposed to just a few) secondary calibrations can provide
454 sufficient information to alleviate or even neutralize potential biases (Sauquet, 2013).

455 Certainly, further studies are required to fully understand the effect of secondary calibrations
456 on outputs from different tree dating methods, and on downstream analyses. It is possible
457 that secondary calibrations can be safely used with dating methods that do not require
458 setting priors, such as penalized likelihood (Sanderson, 2003), with methods that do not
459 make any assumptions on the ages and fix them to a node on a tree topology, such as
460 BLADJ (Webb et al., 2008; Webb & Donoghue, 2005), or methods that summarize age data
461 unto a tree topology.

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

470 *Summarizing Chronograms*

471 **Summarizing chronograms**

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

479 Some measures that have been proposed are the proportion of lineage sampling and the
480 number of calibrations used (Magallón, 2010; Magallón et al., 2015,

481 [@magallon2015metacalibrated](#)). Some characteristics that are often cited in published
482 studies as a measure of improved age estimates as compared to previously published
483 estimates are: quality of alignment (missing data, GC content), lineage sampling (strategy
484 and proportion), phylogenetic and dating inference method, number of fossils used as
485 calibrations, support for nodes and ages, and magnitude of confidence intervals.

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

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

500 *Effects on Downstream Analyses*

501 **Effects on downstream analyses**

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

505 In ecology and conservation biology, incorporating at least some data on lineage divergence
506 times represents a relevant improvement for testing alternative hypothesis using phylogenetic

507 distance (Webb et al., 2008). Hence, DateLife's workflow features different ways of
508 estimating node ages in the absence of calibrations and branch length information for certain
509 taxa. "Making up" branch lengths is a common practice in scientific publications: Jetz et al.
510 (2012), created a chronogram of all 9, 993 bird species, where 67% had molecular data and
511 the rest was simulated; Rabosky et al. (2018) created a chronogram of 31, 536 ray-finned
512 fishes, of which only 37% had molecular data; Smith and Brown (2018) constructed a
513 chronogram of 353, 185 seed plants where only 23% had molecular data.
514 Notably, risks come with this practice. Taken to the extreme, one could make a fully
515 resolved, calibrated tree of all modern and extinct taxa using a single taxonomy and a single
516 calibration, using polytomy resolution and branch estimation methods. There has yet to be
517 a thorough analysis of what can go wrong when one extends inferences beyond the data in
518 this way, so we urge caution; we also urge readers to follow the example of the large tree
519 papers cited above, by carefully considering the statistical assumptions being made, and
520 assessing the consistency of the results with prior work.

521 **CONCLUSIONS**

522 **Conclusions**

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

531 The DateLife project allows for easy and fast summary of public and state-of-the-art data on
532 time of lineage divergence. It provides a straightforward way to get an informed idea on the

533 state of knowledge of the time frame of evolution of different regions of the tree of life, and
534 allows identification of regions that require more research, or that have conflicting
535 information. It is available as an R package, and as a web-based R shiny application at
536 www.datelife.org Both summary and newly generated trees are useful to evaluate
537 evolutionary hypotheses in different areas of research. The DateLife project helps with
538 awareness of the existing variation in expert time of divergence data, and will foster
539 exploration of the effect of alternative divergence time hypothesis on the results of analyses,
540 nurturing a culture of more cautious interpretation of evolutionary results.

541 **AVAILABILITY**

542 **Availability**

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

553 **SUPPLEMENTARY MATERIAL**

554 **Supplementary Material**

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

558 **FUNDING**

559

Funding

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

563 **ACKNOWLEDGEMENTS**

564

Acknowledgements

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

References

- 577
- 578 Alström, P., Hooper, D. M., Liu, Y., Olsson, U., Mohan, D., Gelang, M., ... Price, T. D.
579 (2014). Discovery of a relict lineage and monotypic family of passerine birds. *Biology
580 Letters*, 10(3), 20131067.
- 581 Ané, C., Eulenstein, O., Piaggio-Talice, R., & Sanderson, M. J. (2009). Groves of
582 phylogenetic trees. *Annals of Combinatorics*, 13(2), 139–167.
- 583 Antonelli, A., Hettling, H., Condamine, F. L., Vos, K., Nilsson, R. H., Sanderson, M. J., ...
584 Vos, R. A. (2017). Toward a self-updating platform for estimating rates of speciation
585 and migration, ages, and relationships of Taxa. *Systematic Biology*, 66(2), 153–166.
586 <https://doi.org/10.1093/sysbio/syw066>
- 587 Archie, J., Day, W. H., Felsenstein, J., Maddison, W., Meacham, C., Rohlf, F. J., &
588 Swofford, D. (1986). The Newick tree format. Retrieved from
589 {<https://evolution.genetics.washington.edu/phylip/newicktree.html>}
- 590 Avibase. (2022). Yellow-throated Bunting. *Avibase - the World Bird Database*, (Online
591 Resource). Retrieved from
592 {<https://avibase.bsc-eoc.org/species.jsp?lang=EN&avibaseid=82D1EE0049D8D927>}
- 593 Bapst, D. W. (2012). Paleotree: An R package for paleontological and phylogenetic analyses
594 of evolution. *Methods in Ecology and Evolution*, 3(5), 803–807.
595 <https://doi.org/10.1111/j.2041-210X.2012.00223.x>
- 596 Barba-Montoya, J., Reis, M. dos, Schneider, H., Donoghue, P. C., & Yang, Z. (2018).
597 Constraining uncertainty in the timescale of angiosperm evolution and the veracity of
598 a cretaceous terrestrial revolution. *New Phytologist*, 218(2), 819–834.
- 599 Barker, F. K. (2014). Mitogenomic data resolve basal relationships among passeriform and
600 passeridan birds. *Molecular Phylogenetics and Evolution*, 79, 313–324.
- 601 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2013). Going to
602 extremes: Contrasting rates of diversification in a recent radiation of new world
603 passerine birds. *Systematic Biology*, 62(2), 298–320.

- 604 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2015). New insights
605 into new world biogeography: An integrated view from the phylogeny of blackbirds,
606 cardinals, sparrows, tanagers, warblers, and allies. *The Auk: Ornithological Advances*,
607 132(2), 333–348.
- 608 Barker, F. K., Cibois, A., Schikler, P., Feinstein, J., & Cracraft, J. (2004). Phylogeny and
609 diversification of the largest avian radiation. *Proceedings of the National Academy of
610 Sciences*, 101(30), 11040–11045.
- 611 Beresford, P., Barker, F., Ryan, P., & Crowe, T. (2005). African endemics span the tree of
612 songbirds (passeri): Molecular systematics of several evolutionary “enigmas”.
613 *Proceedings of the Royal Society B: Biological Sciences*, 272(1565), 849–858.
- 614 Boyle, B., Hopkins, N., Lu, Z., Raygoza Garay, J. A., Mozzherin, D., Rees, T., ... Enquist,
615 B. J. (2013). The taxonomic name resolution service: An online tool for automated
616 standardization of plant names. *BMC Bioinformatics*, 14(1).
617 <https://doi.org/10.1186/1471-2105-14-16>
- 618 Britton, T., Anderson, C. L., Jacquet, D., Lundqvist, S., & Bremer, K. (2007). Estimating
619 Divergence Times in Large Phylogenetic Trees. *Systematic Biology*, 56(788777878),
620 741–752. <https://doi.org/10.1080/10635150701613783>
- 621 Bryson Jr, R. W., Chaves, J., Smith, B. T., Miller, M. J., Winker, K., Pérez-Emán, J. L., &
622 Klicka, J. (2014). Diversification across the new world within the ‘blue’cardinalids
623 (aves: Cardinalidae). *Journal of Biogeography*, 41(3), 587–599.
- 624 Burleigh, J. G., Kimball, R. T., & Braun, E. L. (2015). Building the avian tree of life using a
625 large-scale, sparse supermatrix. *Molecular Phylogenetics and Evolution*, 84, 53–63.
- 626 Burns, K. J., Shultz, A. J., Title, P. O., Mason, N. A., Barker, F. K., Klicka, J., ... Lovette,
627 I. J. (2014). Phylogenetics and diversification of tanagers (passeriformes:
628 Thraupidae), the largest radiation of neotropical songbirds. *Molecular Phylogenetics
629 and Evolution*, 75, 41–77.
- 630 Chamberlain, S. (2018). *bold: Interface to Bold Systems API*. Retrieved from

- 631 <https://CRAN.R-project.org/package=bold>
- 632 Chamberlain, S. A., & Szöcs, E. (2013). taxize : taxonomic search and retrieval in R [version
633 2; referees: 3 approved]. *F1000Research*, 2(191), 1–29.
634 <https://doi.org/10.12688/f1000research.2-191.v2>
- 635 Chaves, J. A., Hidalgo, J. R., & Klicka, J. (2013). Biogeography and evolutionary history of
636 the neotropical genus *Saltator* (aves: Thraupini). *Journal of Biogeography*, 40(11),
637 2180–2190.
- 638 Claramunt, S., & Cracraft, J. (2015). A new time tree reveals earth history's imprint on the
639 evolution of modern birds. *Science Advances*, 1(11), e1501005.
- 640 Criscuolo, A., Berry, V., Douzery, E. J., & Gascuel, O. (2006). SDM: A fast distance-based
641 approach for (super)tree building in phylogenomics. *Systematic Biology*, 55(5),
642 740–755. <https://doi.org/10.1080/10635150600969872>
- 643 Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M.-K., Turon, X., ... Douzery,
644 E. J. (2018). A phylogenomic framework and timescale for comparative studies of
645 tunicates. *BMC Biology*, 16(1), 1–14.
- 646 Eastman, J. M., Harmon, L. J., & Tank, D. C. (2013). Congruification: Support for time
647 scaling large phylogenetic trees. *Methods in Ecology and Evolution*, 4(7), 688–691.
648 <https://doi.org/10.1111/2041-210X.12051>
- 649 Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high
650 throughput. *Nucleic Acids Research*, 32(5), 1792–1797.
- 651 Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American Naturalist*,
652 125(1), 1–15. Retrieved from <http://www.jstor.org/stable/2461605>
- 653 GBIF Secretariat. (2022). GBIF Backbone Taxonomy. *Checklist dataset*, (Online Resource
654 accessed via GBIF.org). Retrieved from {<https://doi.org/10.15468/39omei> }
- 655 Gibb, G. C., England, R., Hartig, G., McLenaghan, P. A., Taylor Smith, B. L., McComish,
656 B. J., ... Penny, D. (2015). New Zealand passerines help clarify the diversification of
657 major songbird lineages during the oligocene. *Genome Biology and Evolution*, 7(11),

- 658 2983–2995.
- 659 Graur, D., & Martin, W. (2004). Reading the entrails of chickens: Molecular timescales of
660 evolution and the illusion of precision. *TRENDS in Genetics*, 20(2), 80–86.
- 661 Hackett, S. J., Kimball, R. T., Reddy, S., Bowie, R. C., Braun, E. L., Braun, M. J., ...
662 others. (2008). A phylogenomic study of birds reveals their evolutionary history.
663 *Science*, 320(5884), 1763–1768.
- 664 Hedges, S. B., Marin, J., Suleski, M., Paymer, M., & Kumar, S. (2015). Tree of life reveals
665 clock-like speciation and diversification. *Molecular Biology and Evolution*, 32(4),
666 835–845. <https://doi.org/10.1093/molbev/msv037>
- 667 Heibl, C. (2008). *PHYLOCH: R language tree plotting tools and interfaces to diverse*
668 *phylogenetic software packages*. Retrieved from
669 <http://www.christophheibl.de/Rpackages.html>
- 670 Hooper, D. M., & Price, T. D. (2017). Chromosomal inversion differences correlate with
671 range overlap in passerine birds. *Nature Ecology & Evolution*, 1(10), 1526.
- 672 Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic
673 trees. *Bioinformatics*, 17(8), 754–755.
674 <https://doi.org/10.1093/bioinformatics/17.8.754>
- 675 Jetz, W., Thomas, G., Joy, J. J., Hartmann, K., & Mooers, A. (2012). The global diversity
676 of birds in space and time. *Nature*, 491(7424), 444–448.
677 <https://doi.org/10.1038/nature11631>
- 678 Johansson, U. S., Fjeldså, J., & Bowie, R. C. (2008). Phylogenetic relationships within
679 passerida (aves: Passeriformes): A review and a new molecular phylogeny based on
680 three nuclear intron markers. *Molecular Phylogenetics and Evolution*, 48(3), 858–876.
- 681 Katoh, K., Asimenos, G., & Toh, H. (2009). Multiple alignment of dna sequences with mafft.
682 In *Bioinformatics for dna sequence analysis* (pp. 39–64). Springer.
- 683 Kimball, R. T., Oliveros, C. H., Wang, N., White, N. D., Barker, F. K., Field, D. J., ...
684 others. (2019). A phylogenomic supertree of birds. *Diversity*, 11(7), 109.

- 685 Klicka, J., Barker, F. K., Burns, K. J., Lanyon, S. M., Lovette, I. J., Chaves, J. A., & Bryson
686 Jr, R. W. (2014). A comprehensive multilocus assessment of sparrow (aves:
687 Passerellidae) relationships. *Molecular Phylogenetics and Evolution*, 77, 177–182.
- 688 Ksepka, D. T., Parham, J. F., Allman, J. F., Benton, M. J., Carrano, M. T., Cranston, K.
689 A., . . . others. (2015). The fossil calibration database—a new resource for divergence
690 dating. *Systematic Biology*, 64(5), 853–859.
- 691 Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio,
692 A., . . . others. (2015). Evolution of darwin’s finches and their beaks revealed by
693 genome sequencing. *Nature*, 518(7539), 371–375.
- 694 Laubichler, M. D., & Maienschein, J. (2009). *Form and function in developmental evolution*.
695 Cambridge University Press.
- 696 Lepage, D. (2004). *Avibase: The world bird database*. Bird Studies Canada.
- 697 Lepage, D., Vaidya, G., & Guralnick, R. (2014). Avibase—a database system for managing
698 and organizing taxonomic concepts. *ZooKeys*, (420), 117.
- 699 Lerner, H. R., Meyer, M., James, H. F., Hofreiter, M., & Fleischer, R. C. (2011). Multilocus
700 resolution of phylogeny and timescale in the extant adaptive radiation of hawaiian
701 honeycreepers. *Current Biology*, 21(21), 1838–1844.
- 702 Lovette, I. J., Pérez-Emán, J. L., Sullivan, J. P., Banks, R. C., Fiorentino, I.,
703 Córdoba-Córdoba, S., . . . others. (2010). A comprehensive multilocus phylogeny for
704 the wood-warblers and a revised classification of the parulidae (aves). *Molecular
705 Phylogenetics and Evolution*, 57(2), 753–770.
- 706 Magallon, S., & Sanderson, M. (2001). Absolute diversification rates in angiosperm clades.
707 *Evolution*, 55(9), 1762–1780.
- 708 Magallón, S. (2010). Using fossils to break long branches in molecular dating: A comparison
709 of relaxed clocks applied to the origin of angiosperms. *Systematic Biology*, 59(4),
710 384–399.
- 711 Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., & Hernández-Hernández, T. (2015).

- 712 A metacalibrated time-tree documents the early rise of flowering plant phylogenetic
713 diversity. *New Phytologist*, 207(2), 437–453.
- 714 McTavish, E. J., Hinchliff, C. E., Allman, J. F., Brown, J. W., Cranston, K. A., Holder, M.
715 T., ... Smith, S. (2015). Phylesystem: A git-based data store for community-curated
716 phylogenetic estimates. *Bioinformatics*, 31(17), 2794–2800.
- 717 Michonneau, F., Brown, J. W., & Winter, D. J. (2016). rotl: an R package to interact with
718 the Open Tree of Life data. *Methods in Ecology and Evolution*, 7(12), 1476–1481.
719 <https://doi.org/10.1111/2041-210X.12593>
- 720 Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecology Letters*,
721 17(4), 508–525. <https://doi.org/10.1111/ele.12251>
- 722 Moyle, R. G., Oliveros, C. H., Andersen, M. J., Hosner, P. A., Benz, B. W., Manthey, J. D.,
723 ... Faircloth, B. C. (2016). Tectonic collision and uplift of Wallacea triggered the
724 global songbird radiation. *Nature Communications*, 7(1), 1–7.
- 725 Oliveros, C. H., Field, D. J., Ksepka, D. T., Barker, F. K., Aleixo, A., Andersen, M. J., ...
726 others. (2019). Earth history and the passerine superradiation. *Proceedings of the
727 National Academy of Sciences*, 116(16), 7916–7925.
- 728 Ooms, J., & Chamberlain, S. (2018). *Phylocomr: Interface to 'phylocom'*. Retrieved from
729 <https://CRAN.R-project.org/package=phylocomr>
- 730 Open Tree Of Life, Redelings, B., Cranston, K. A., Allman, J., Holder, M. T., & McTavish,
731 E. J. (2016). Open Tree of Life APIs v3.0. *Open Tree of Life Project*, (Online
732 Resources). Retrieved from
733 {735 Open Tree Of Life, Redelings, B., Sánchez Reyes, L. L., Cranston, K. A., Allman, J., Holder,
736 M. T., & McTavish, E. J. (2019). Open tree of life synthetic tree v12.3. *Zenodo*.
737 Retrieved from <https://doi.org/10.5281/zenodo.3937742>
738 Ödeen, A., Håstad, O., & Alström, P. (2011). Evolution of ultraviolet vision in the largest

- 739 avian radiation-the passerines. *BMC Evolutionary Biology*, 11(1), 1–8.
- 740 741 Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20(2), 289–290.
- 742 743 Parchman, T. L., Benkman, C. W., & Mezquida, E. T. (2007). Coevolution between hispaniolan crossbills and pine: Does more time allow for greater phenotypic 744 escalation at lower latitude? *Evolution*, 61(9), 2142–2153.
- 745 746 Päckert, M., Martens, J., Sun, Y.-H., Severinghaus, L. L., Nazarenko, A. A., Ting, J., ... Tietze, D. T. (2012). Horizontal and elevational phylogeographic patterns of 747 himalayan and southeast asian forest passerines (aves: Passeriformes). *Journal of 748 Biogeography*, 39(3), 556–573.
- 749 750 Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., FitzJohn, R. G., ... Harmon, L. J. (2014). Geiger v2. 0: An expanded suite of methods for fitting 751 macroevolutionary models to phylogenetic trees. *Bioinformatics*, 30(15), 2216–2218.
- 752 753 Posadas, P., Crisci, J. V., & Katinas, L. (2006). Historical biogeography: A review of its basic concepts and critical issues. *Journal of Arid Environments*, 66(3), 389–403.
- 754 755 Powell, A. F., Barker, F. K., Lanyon, S. M., Burns, K. J., Klicka, J., & Lovette, I. J. (2014). A comprehensive species-level molecular phylogeny of the new world blackbirds 756 (icteridae). *Molecular Phylogenetics and Evolution*, 71, 94–112.
- 757 758 Price, T. D., Hooper, D. M., Buchanan, C. D., Johansson, U. S., Tietze, D. T., Alström, P., ... others. (2014). Niche filling slows the diversification of himalayan songbirds. 759 *Nature*, 509(7499), 222.
- 760 761 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: 762 Cardinalidae). *Molecular Phylogenetics and Evolution*, 69(3), 1222–1227.
- 763 764 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*, 765 559(7714), 392.

- 766 Ramshaw, J., Richardson, D., Meatyard, B., Brown, R., Richardson, M., Thompson, E., &
767 Boulter, D. (1972). The time of origin of the flowering plants determined by using
768 amino acid sequence data of cytochrome c. *New Phytologist*, 71(5), 773–779.
- 769 Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The barcode of life data system
770 (<http://www.Barcodinglife.Org>). *Molecular Ecology Notes*, 7(3), 355–364.
- 771 R Core Team. (2018). *R: a language and environment for statistical computing*. Vienna,
772 Austria: R Foundation for Statistical Computing.
- 773 Rees, & Cranston, K. (2017). Automated assembly of a reference taxonomy for phylogenetic
774 data synthesis. *Biodiversity Data Journal*, (5).
- 775 Rees, Vandepitte, L., Decock, W., & Vanhoorne, B. (2017). IRMNG 2006–2016: 10 Years of
776 a Global Taxonomic Database. *Biodiversity Informatics*, 12.
- 777 Revell, L. J. (2012). Phytools: An r package for phylogenetic comparative biology (and other
778 things). *Methods in Ecology and Evolution*, 3, 217–223.
- 779 Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference
780 under mixed models. *Bioinformatics*, 19(12), 1572–1574.
781 <https://doi.org/10.1093/bioinformatics/btg180>
- 782 Sanchez-Reyes, L. L., & O'Meara, B. (2022). `datelifeplot`: Methods to plot chronograms
783 and outputs of the `datelife` package. *R Package Release V0.2.2*. Retrieved from
784 <https://zenodo.org/badge/latestdoi/381501451>
- 785 ~~Sanchez-Reyes, L. L., O'Meara, B., Eastman, J., Heath, T., Wright, A., Schliep, K.,...,~~
786 ~~Alfaro, M. (2022). `datelife`: Scientific Data on Time of Lineage Divergence for~~
787 ~~Your Taxa. In *R package version 0.6.6*. Retrieved from~~
788 ~~<https://CRAN.R-project.org/package=datelife> and~~
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 Tietze, D. T., Päckert, M., Martens, J., Lehmann, H., & Sun, Y.-H. (2013). Complete

820 phylogeny and historical biogeography of true rosefinches (aves: *Carpodacus*).

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

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

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

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

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

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

834 Webb, C., Ackerly, D., & Kembel, S. (2008). Phylocom: Software for the analysis of
835 phylogenetic community structure and trait evolution. *Bioinformatics*, 24(18),
836 2098–2100. <https://doi.org/10.1093/bioinformatics/btn358>

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

839 Weir, J., & Schluter, D. (2008). Calibrating the avian molecular clock. *Molecular Ecology*,
840 17(10), 2321–2328.

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