

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: 5480

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,
51 [2009\), conservation biology and ecology](#) (Felsenstein, 1985; Webb, 2000), historical
52 biogeography (Posadas, Crisci, & Katinas, 2006), and species diversification (Magallon &
53 Sanderson, 2001; Morlon, 2014).

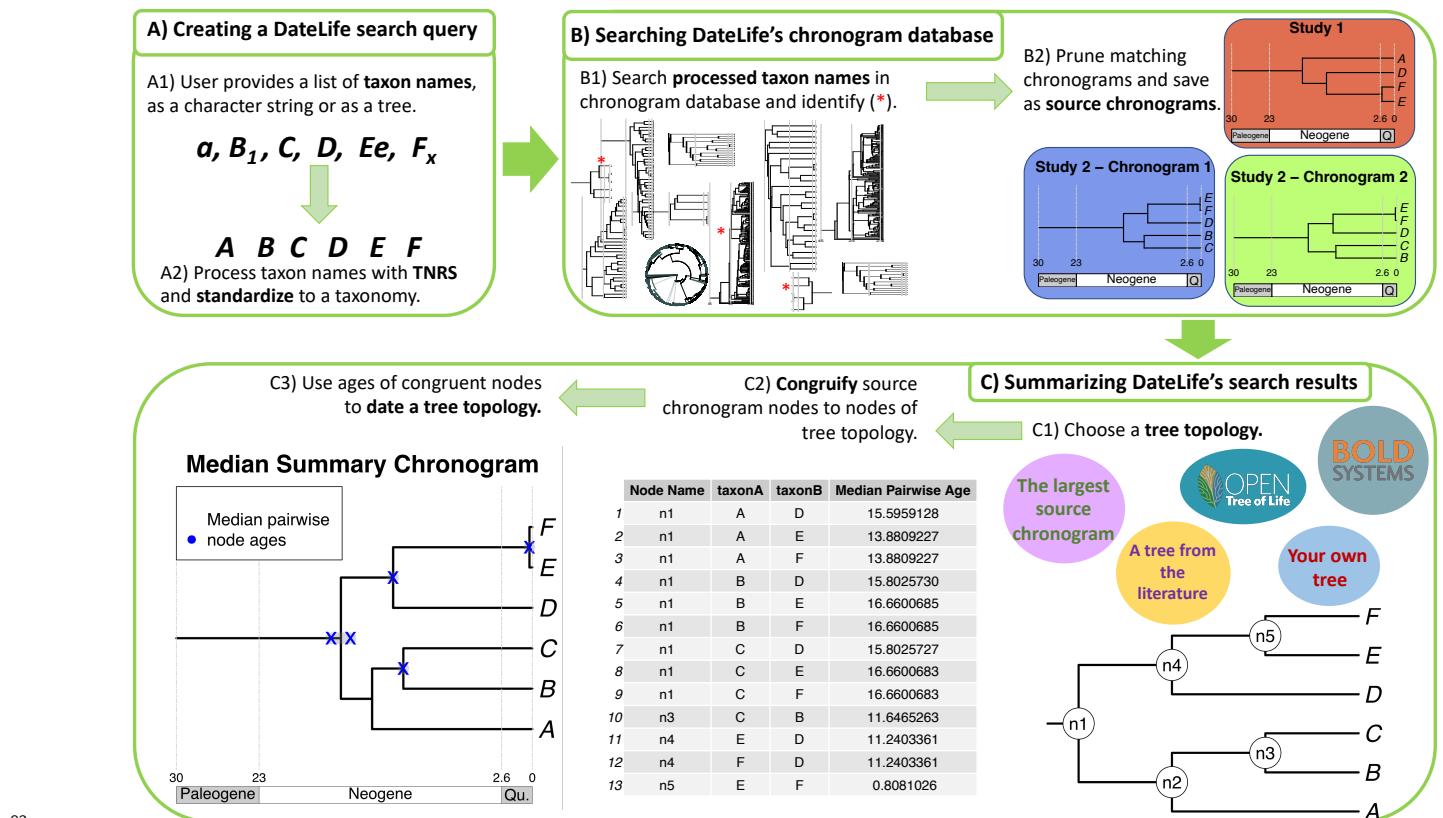
54 Building a chronogram is not an easy task. It requires obtaining and curating data to
55 construct a phylogeny, selecting and placing appropriate calibrations on the phylogeny using
56 independent age data points from the fossil record or other dated events, and inferring the
57 full dated tree; it 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) ~~, data stored and available in a computer readable that stores data~~
67 [in a computer-readable](#) format (Vos et al., 2012); automated and programmatic ways of
68 accessing ~~the data and downloading the data in a computer-readable format also~~ (Stoltzfus
69 et al., 2013); and methods to summarize and 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 ~~two taxon names, which~~
86 one taxon scientific name. Multiple input names can be provided as a comma separated
87 character string or as tip labels on a tree. If the input is a tree, it can be provided as a
88 classic newick character string (Archie et al., 1986), or as a “phylo” R object (Paradis et al.,
89 2004). The input tree is not required to have branch lengths, and its topology is used in the
90 summary steps described in the next section.

91 DateLife ~~accepts scientific names that can belong to any inclusive taxonomic group~~
92 ~~(e.g., genus, family, tribe, etc.) or a binomial species name. Subspecies and variants are~~
93 ~~ignored. If an input taxon name belongs to an inclusive taxonomic group, DateLife has two~~
94 ~~alternative behaviors defined by processes input scientific names using a Taxonomic Name~~
95 ~~Resolution Service (TNRS), which increases the probability of correctly finding the queried~~
96 ~~taxon names in the chronogram database. TNRS detects, corrects and standardizes name~~
97 ~~misspellings and typos, variant spellings and authorities, and nomenclatural synonyms to a~~
98 ~~single taxonomic standard (Boyle et al., 2013). TNRS also allows to correctly choose~~
99 ~~between homonyms, by considering other taxa provided as input to infer the taxonomic~~
100 ~~context of the flag. If the flag is active, DateLife retrieves all species names within the~~
101 ~~inclusive taxonomic group following a standard taxonomy of choice, and adds them to the~~
102 ~~input string. Taxonomies currently supported by DateLife are homonym. DateLife~~
103 ~~implements TNRS using the Open Tree of Life (OpenTree) unified Taxonomy (OTT, Open~~
104 ~~Tree Of Life et al., 2016; Rees & Cranston, 2017) as standard, storing taxonomic~~
105 ~~identification numbers (OTT ids) for further processing and analysis. Other taxonomies~~
106 ~~currently supported by DateLife are~~ the National Center of Biotechnology Information
107 (NCBI) taxonomic database (Schoch et al., 2020), the Global Biodiversity Information
108 Facility (GBIF) taxonomic backbone (GBIF Secretariat, 2022), and the Interim Register of
109 Marine and Nonmarine Genera (IRMNG) database (Rees et al., 2017). ~~If the flag is~~

110 Besides binomial species names, DateLife accepts scientific names from any inclusive
111 taxonomic group (e.g., genus, family, tribe), as well as subspecific taxonomic variants (e.g.,
112 subspecies, variants, strains). If a taxon name belongs to an inclusive taxonomic group,
113 DateLife has two alternative behaviors defined by the “get species from taxon” flag. If the
114 flag is active, DateLife retrieves all species names within the taxonomic group from the
115 standard taxonomy of choice, and adds them to the search query. In this case, subspecific
116 variants are excluded. If the flag is inactive, DateLife excludes any taxon names above the
117 species level from the search query.

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

125 The processed input Species and subspecific variant names are processed and
126 searched as provided by the user. The processed taxon names are saved as an R object of a
127 newly defined class, datelifeQuery, that is used in the following steps. This object contains
128 the standardized names input names standardized to a taxonomy of choice (OTT by
129 default), the corresponding OTT identification id numbers, and the topology of the input
130 treeif any an input tree, if one was provided.

131 Searching a chronogram database

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

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

139 A unique feature of the Phylesystem is that any user can add new published,
140 state-of-the-art chronograms any time, through their curator application
141 (<https://tree.opentreeoflife.org/curator>). As chronograms are added to Phylesystem, they ~~are~~
142 can be incorporated into the chronogram database of the `datelife` package. ~~datelife's~~
143 chronogram database is currently manually updated as new chronogram data is added to
144 Phylesystem. The updated database is assigned a new version number, followed by a
145 package release on CRAN. ~~datelife's chronogram database is updated as new chronogram~~
146 ~~data is added to Phylesystem, at a minimum of once a month and a maximum of every 6~~
147 ~~months.~~ Users can also implement functions from the `datelife` R package to trigger an
148 update of the local chronogram database, to incorporate any new chronograms to the user's
149 DateLife analysis before an official database update is released on CRAN.

150 A DateLife search is implemented by matching processed taxon names provided by the
151 user to tip labels in the chronogram database. Chronograms with at least two matching
152 taxon names on their tip labels are identified and pruned down to preserve only the matched
153 taxa. These matching pruned chronograms are referred to as source chronograms. Total
154 distance (in units of millions of years) between taxon pairs within each source chronogram
155 are stored as a patristic distance matrix (Figure 1). The matrix format speeds up extraction
156 of pairwise taxon ages of any queried taxa, as opposed to searching the ancestor node of a
157 pair of taxa in a “phylo” object or newick string. Finally, the patristic matrices are
158 associated to the study citation where the original chronogram was published, and stored as
159 an R object of the newly defined class `datelifeResult`.

160 **Summarizing search results**

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

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

174 ***Choosing a topology.***— DateLife requires a tree topology to summarize age data
175 upon. We recommend that users provide as input a tree topology from the literature, or one
176 of their own making. If no topology is provided, DateLife automatically extracts one from
177 the OpenTree synthetic tree, a phylogeny currently encompassing 2.3 million taxa across all
178 life, assembled from 1, 239 published phylogenetic trees and OpenTree’s unified Taxonomy,
179 OTT (Open Tree Of Life et al., 2019). Alternatively, DateLife can combine topologies from
180 source chronograms using a supertree approach ~~To combine topologies from source~~
181 ~~chronograms into a single summary (or supertree) topology, the DateLife workflow~~
182 ~~(Criscuolo, Berry, Douzery, & Gascuel, 2006).~~ To do this, DateLife first identifies the source
183 chronograms that form a grove, roughly, a sufficiently overlapping set of taxa between trees,
184 by implementing definition 2.8 for n-overlap from Ané et al. (2009). If the source
185 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. ~~If source chronograms do not form a grove, the supertree reconstruction will fail.~~ The result is a single summary (or supertree) topology, that combines topologies from source chronograms in a grove.

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

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

For each congruent node, the pairwise distances that traverse that node are summarized into a single summary matrix using classic summary statistics (i.e., mean, median, minimum and maximum ages), and the Supermatrix Distance Method (SDM; Criscuolo, ~~Berry, Douzery, & Gascuel, et al.~~, 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 ~~are stored as summarized calibrations that can be used as secondary~~ calibrations to date a tree topology, ~~using different~~ with or without initial branch lengths, using phylogenetic dating methods currently supported within DateLife: ~~BLADJ (Webb, Ackerly, & Kembel, 2008; Webb & Donoghue, 2005)~~, 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).

Dating a tree topology.— By default, DateLife implements the Branch Length Adjuster (BLADJ; ~~Webb et al., 2008; Webb & Donoghue, 2005~~) algorithm to obtain a fully dated topology. BLADJ ~~fixes node ages that have~~ takes an initial topology and ignores branch length data if present. Next, it fixes ages on nodes with calibration data, and distributes time ~~between for~~ nodes with no data evenly between ~~nodes with calibration data~~. This minimizes calibrated nodes, minimizing 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. This method is useful for ecological analyses (Webb et al., 2008)~~. When there is conflict ~~in ages between nodes with calibration data, BLADJ~~ between ages of calibrated nodes, 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~~ will not return a dated topology and will provide a warning message along with suggestions on how the user can provide an age for the root so the

239 dating analysis can be finished.

240 Alternatively, the user can choose to obtain a fully dated phylogeny MrBayes
241 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003) approach in DateLife uses
242 the calibrations as priors on node ages.

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

246 In the case the user can provide or obtain branch length data along with a tree
247 topology, phylogenetic dating options supported in DateLife (MrBayes, PATHd8, TreePL)
248 that incorporate branch length information from the input topology in combination with the
249 calibrations. secondary calibrations: PATHd8 is a non-clock, rate-smoothing method
250 (Britton et al., 2007) to date trees. treePL (Smith & O'Meara, 2012), is a semi-parametric,
251 rate-smoothing, penalized likelihood dating method (Sanderson, 2002). The MrBayes
252 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003) approach in DateLife uses
253 the calibrations as priors on node ages.

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

258 A fast solution implemented in DateLife is as follows. To estimate branch lengths
259 proportional to substitution rates for these topologies, DateLife currently implements a
260 simple algorithm. First, it mines the Barcode of Life Data System, BOLD (Ratnasingham
261 & Hebert, 2007) to obtain genetic markers for the input taxa. Mined genetic sequences are
262 aligned with MUSCLE (Edgar, 2004) (by default) or MAFFT (Katoh, Asimenos, & Toh,
263 2009).

264 The BOLD sequence alignment is then used to obtain initial branch lengths with the
265 accelerated transformation (ACCTRAN) parsimony algorithm, which resolves ambiguous
266 character optimization, by assigning changes along branches of the tree as close to the root
267 as possible (Agnarsson & Miller, 2008). This results in older nodes (Forest et al., 2005).
268 Optionally, the likelihood of the tree topology, the alignment and the reconstructed branch
269 lengths given different evolutionary models, is computed using functions from the
270 [phangorn](#) package (Schliep, 2011).

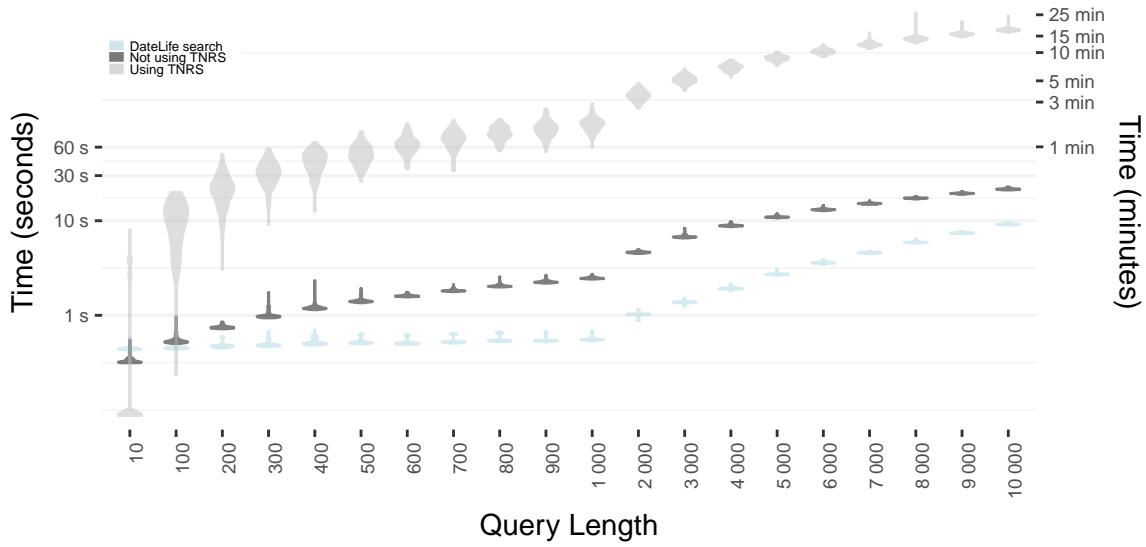
271 Relative branch length information provides key data for phylogenetic dating,
272 especially for nodes without secondary calibrations available. Yet, topologies without
273 branch lengths can also be dated.

274 **Visualizing results.**— Finally, users can save all source and summary chronograms in
275 formats that permit reuse and ~~reanalyses (newick and reanalysis (such as newick and the R~~
276 “phylo” format), as well as visualize and compare results graphically, or construct their own
277 graphs using DateLife’s chronogram plot generation functions available from the R package
278 [datelifeplot](#) (Sanchez-Reyes & O’Meara, 2022).

279

Benchmark

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



289

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

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

304 **Case studies**

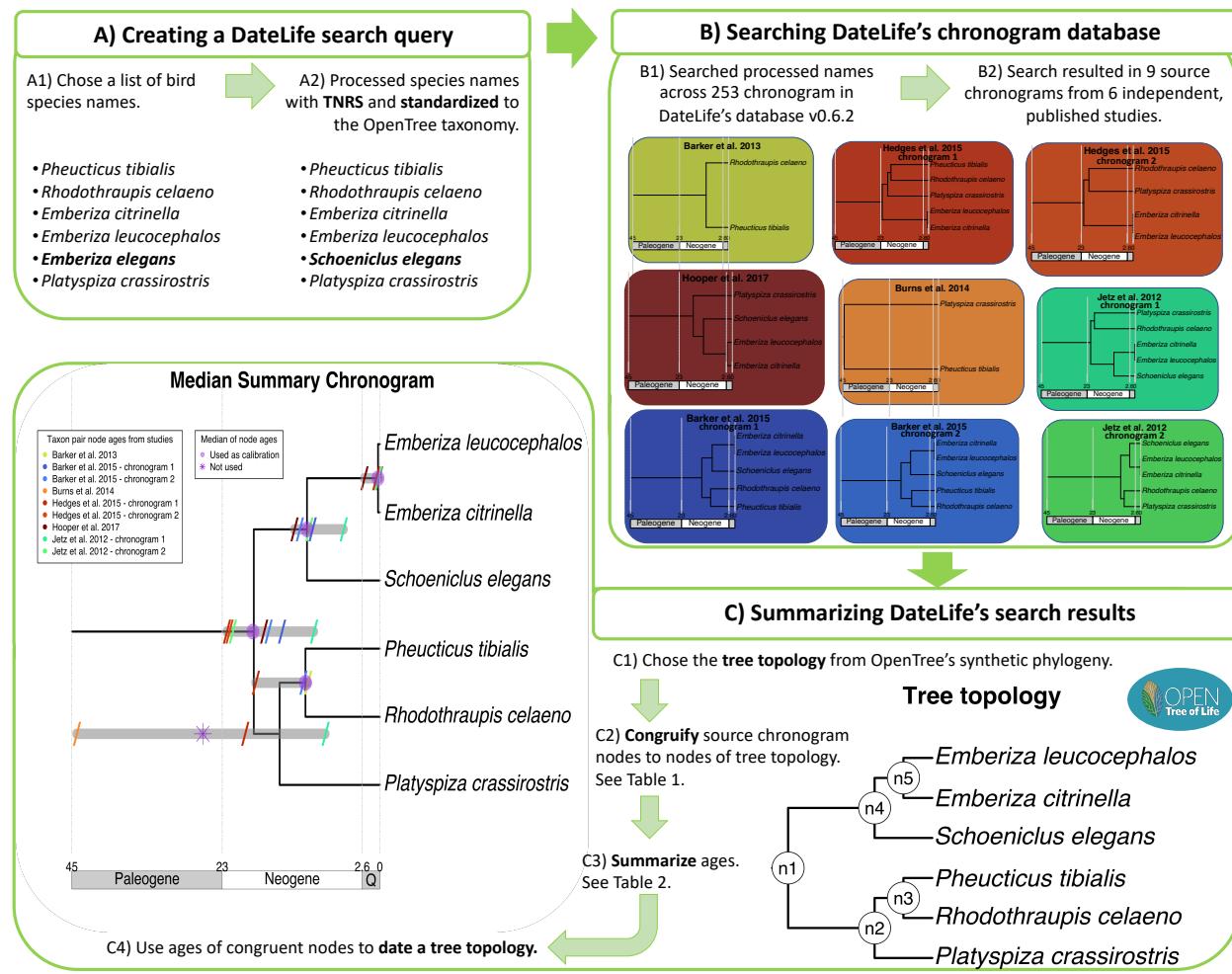
305 We illustrate the DateLife workflow using a family within the Passeriform birds
306 encompassing the true finches, Fringillidae, as case study. On a small example, we analysed 6
307 bird species, and results from each step of the workflow are shown in Fig. 3. As a second
308 example, we analysed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

311 **A small example**

312 **Creating a search query.** We chose 6 bird species within the Passeriformes. The sample
313 includes two species of cardinals: the black-thighed grosbeak – *Pheucticus tibialis* and the
314 crimson-collared grosbeak – *Rhodothraupis celaeno*; three species of buntings: the
315 yellowhammer – *Emberiza citrinella*, the pine bunting – *Emberiza leucocephalos* and the
316 yellow-throated bunting – *Emberiza elegans*; and one species of tanager, the vegetarian finch –
317 *Platyspiza crassirostris*. Processing of input names found that *Emberiza elegans* is synonym
318 for *Schoeniclus elegans* in the default reference taxonomy (OTT v3.3, June 1, 2021). For a
319 detailed discussion on the state of the synonym, refer to Avibase (Avibase, 2022; Lepage,
320 2004; Lepage, Vaidya, & Guralnick, 2014). Discovering this synonym allowed assigning five
321 age data points for the parent node of *Emberiza elegans*, shown as *Schoeniclus elegans* in
322 figure 3A, which would not have had any data otherwise.

323 **Searching the database.** DateLife used the processed input names to search the local
324 chronogram database and found 9 matching chronograms in 6 different studies (Fig. 3B).
325 Three studies matched five input names (Barker, Burns, Klicka, Lanyon, & Lovette, 2015;
326 Hedges, Marin, Suleski, Paymer, & Kumar, 2015; Jetz, Thomas, Joy, Hartmann, & Mooers,
327 2012), one study matched four input names (Hooper & Price, 2017) and two studies matched
328 two input names (Barker, Burns, Klicka, Lanyon, & Lovette, 2013; Burns et al., 2014). No
329 studies matched all input names. Together, source chronograms provide 28 unique age data
330 points, covering all nodes on our chosen tree topology to date (Table 1).

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



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

348 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

349

350 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

351

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

353 **Creating a query.** To obtain ages for all species within the family of true finches,
354 Fringillidae, we ran a DateLife query using the “get species from taxon” flag, which gets all
355 recognized species names within a named group from a taxonomy of choice. Following the
356 NCBI taxonomy, our DateLife query has 289 Fringillidae species names. This
357 taxon-constrained approach implies that the ~~final results of a~~ full DateLife analysis will be
358 ~~done performed~~ using a tree topology and ages ~~for the species in a named available for~~
359 ~~species names from a given taxonomic~~ group, which do not necessarily correspond to a
360 monophyletic group. Users can change this behaviour by providing ~~a monophyletic tree all~~
361 ~~species names corresponding to a monophyletic group~~ as input for a DateLife search, or ~~as a~~
362 ~~tree topology for a monophyletic tree to construct a DateLife summary~~.

363 **Searching the database.** Next, we used the processed species names in our DateLife
364 query to identify chronograms with at least two Fringillidae species as tip taxa. The DateLife
365 search identified ~~13 chronograms containing at least two Fringillidae species~~19 chronograms
366 matching this criteria, published in ~~9–13~~ different studies (Barker et al., 2013, 2015; Burns et
367 al., 2014; Claramunt & Cracraft, 2015; Gibb et al., 2015; Hedges et al., 2015; Hooper &
368 Price, 2017; Jetz et al., 2012; Kimball et al., 2019; Oliveros et al., 2019; Price et al., 2014;
369 Roquet, Lavergne, & Thuiller, 2014; Uyeda, Pennell, Miller, Maia, & McClain, 2017). Once
370 identified, DateLife pruned these matching chronograms to ~~keep Fringillidae species names~~
371 ~~on tips only remove tips that do not belong to the queried taxon names~~, and transformed
372 these pruned chronograms to pairwise distance matrices, revealing 1, 206 different age data
373 points available for species within the Fringillidae (Supplementray Table S1).

374 **Summarizing search results.** The final step ~~is to congrify and summarize entailed~~
375 congruifying and summarizing the age data available for the Fringillidae species into two
376 single summary chronograms, using two different types of summary ages, median and SDM.
377 As explained in the “Description” section, a tree topology to summarize age data upon is
378 required. By default, DateLife uses the topology from OpenTree’s synthetic tree that

379 contains the species in the search query to summarize age data upon all taxa from the search
380 query. According to OpenTree's synthetic tree, species belonging to the family Fringillidae
381 do not form a monophyletic group (Fig. 4). Hence, a topology containing only the 289
382 species from the original query was extracted from Open Tree of Life's synthetic tree v12.3
383 (Open Tree Of Life et al., 2019).

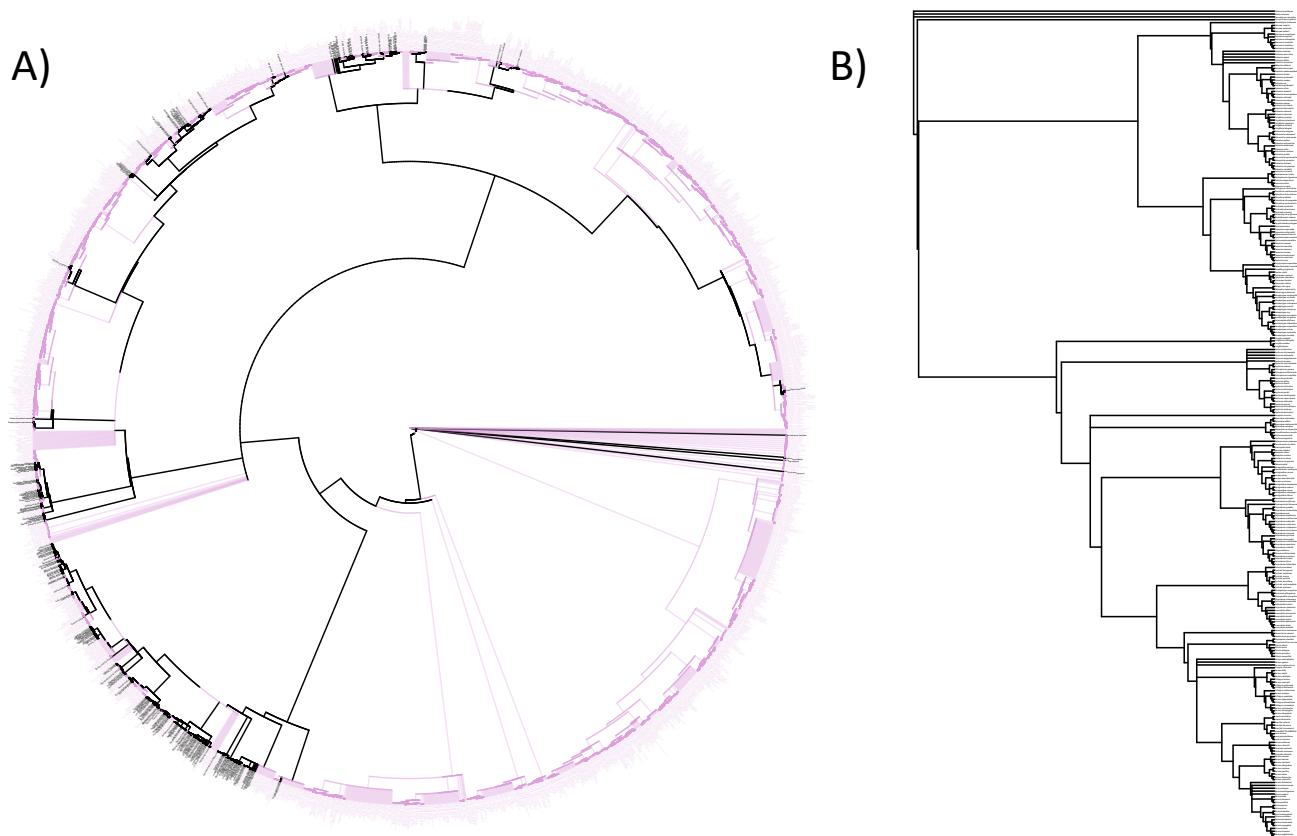
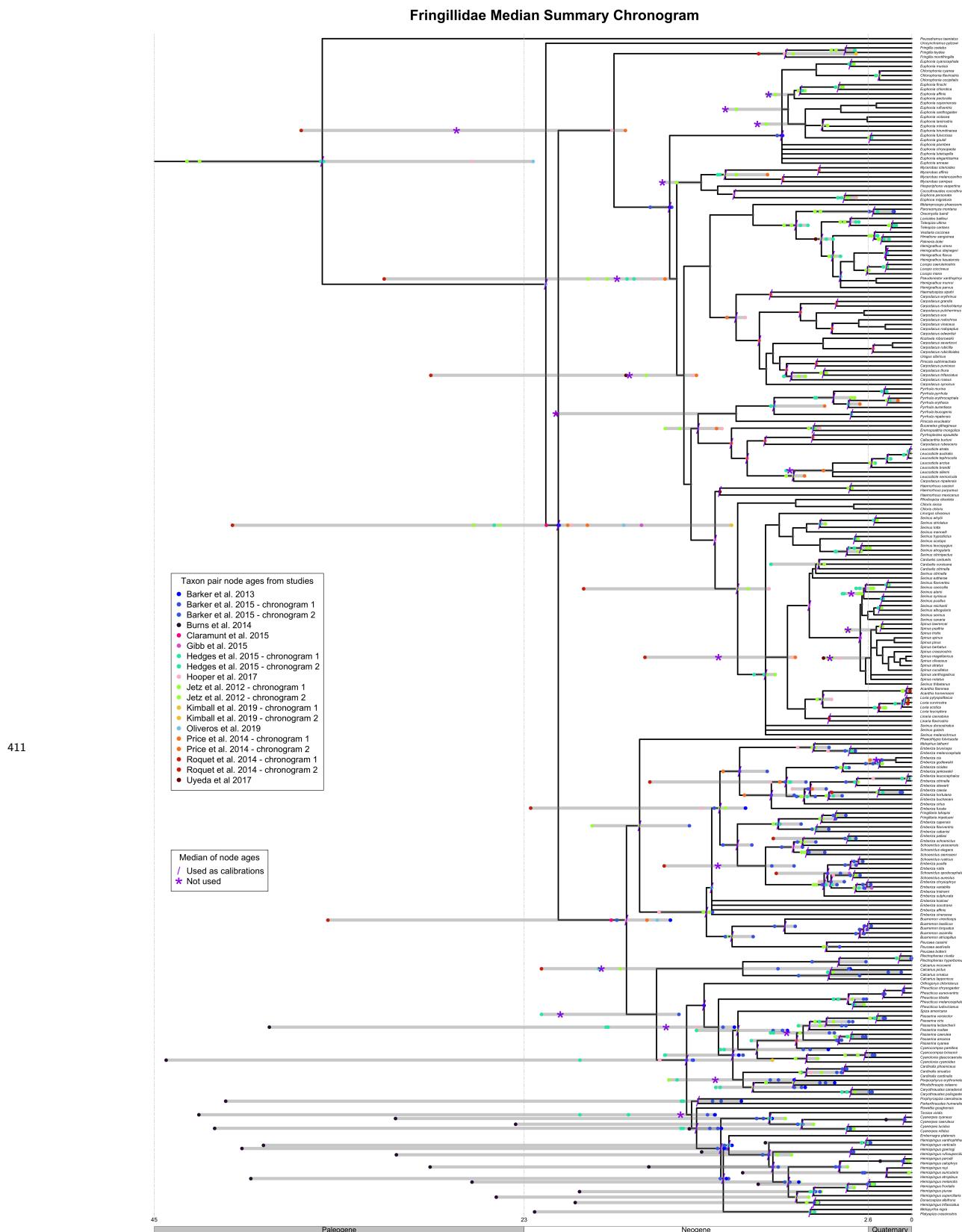


FIGURE 4. Tree topologies extracted from Open Tree of Life's (OpenTree) synthetic phylogenetic tree.
 384 A) Topology of 2,333 tips and 1,305 internal nodes, encompassing bird species within the family
 385 Fringillidae following the NCBI taxonomy (black), as well as all other bird species that share the same
 386 Most Recent Common Ancestor (MRCA) node in OpenTree's synthetic tree (purple). B) Subtree
 387 topology of 289 tips and 253 internal nodes, resulting from pruning species that do not belong to the
 388 family Fringillidae according to the NCBI taxonomy (purple branches in topology A). Bird species
 389 within the Fringillidae are paraphyletic (Alström et al. 2014, Barker, Cibois, Schikler, Feinstein, &
 390 Cracraft 2004, Barker et al. 2013, Barker 2014, Barker et al. 2015, Beresford, Barker, Ryan, &
 391 Crowe 2005, Bryson Jr et al. 2014, Burleigh, Kimball, & Braun 2015, Burns et al. 2014, Chaves,
 392 Hidalgo, & Klicka 2013, Claramunt & Cracraft 2015, Gibb et al. 2015, Hackett et al. 2008, Jetz et al.
 393 2012, Johansson, Fjeldså, & Bowi 200, Kimball et al. 2019, Klicka et al. 2014, Lamichhaney et al.
 394 2015, Lerner, Meyer, James, Hofreiter, & Fleischer 2011, Lovette et al. 2010, Moyle et al. 2016,
 395 Ödeen, Håstad, & Alström 2011, Oliveros et al. 2019, Päckert et al. 2012, Parchman, Benkman, &
 396 Mezquida 2007, Powell et al. 2014, Price et al. 2014, Pulgarín-R, Smith, Bryson Jr, Spellman, &
 397 Klicka 2013, Selvatti, Gonzaga, & Moraes Russo 2015, Tietze, Päckert, Martens, Lehmann, & Sun
 398 2013, Treplin et al. 2008, Zuccon, Prŷs-Jones, Rasmussen, & Ericson 2012).

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

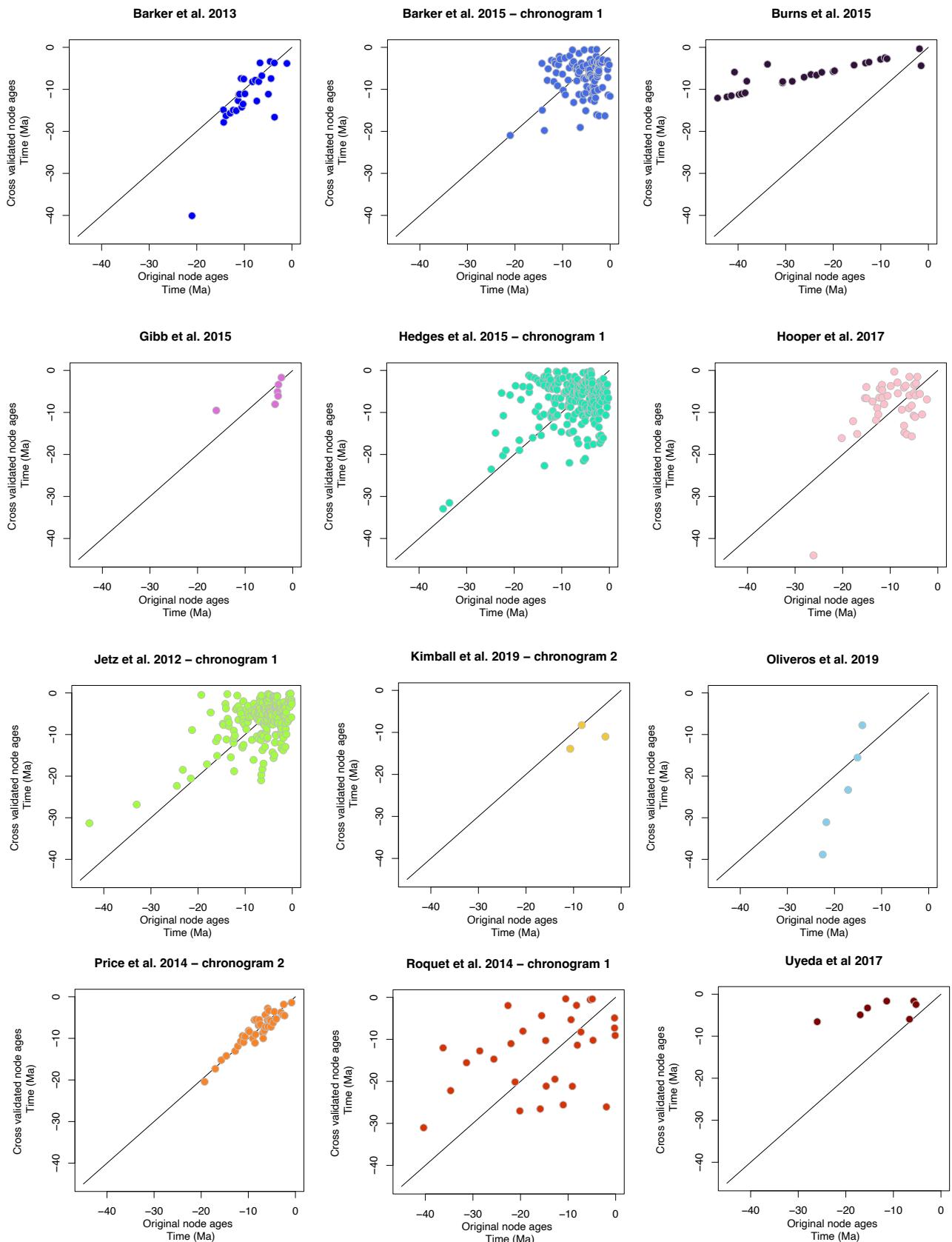


406 FIGURE 5. *Fringillidae median summary chronogram generated with DateLife*. It has 289 tips and
 407 253 nodes, from which 212 have age data from at least one published chronogram. These nodes have
 408 age data points from 19 different chronograms from 13 different studies (colored circles), which were

412

Cross-validation test

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



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

426

Discussion

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

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

441 While it is known that incorporating at least some data on lineage divergence times
442 represents a relevant improvement for testing alternative hypothesis using phylogenetic
443 distance in ecological and conservation biology studies (Webb et al., 2008). Next, we
444 discuss some particularities of using summarized node age data from a DateLife analysis
445 that users should be aware of.

446 Age variation in source chronograms

447 Conflict in estimated ages among alternative studies is common in the literature. See, for
448 example, the robust ongoing debate about crown group age of angiosperms (Barba-Montoya,
449 Reis, Schneider, Donoghue, & Yang, 2018; Magallón, Gómez-Acevedo, Sánchez-Reyes, &
450 Hernández-Hernández, 2015; Ramshaw et al., 1972; Sanderson & Doyle, 2001; Sauquet,
451 Ramírez-Barahona, & Magallón, 2021). Source chronograms available for the same

452 organisms have potentially been estimated implementing calibrations very differently. For
453 example, the chronograms from Burns et al. (2014) were inferred using molecular
454 substitution rate estimates across birds (Weir & Schluter, 2008), and have much older age
455 estimates for the same nodes than chronograms that were inferred using fossils as
456 calibrations (Figs. 5, 6; Supplementary Figs. S1, S5).

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

465 **Primary vs Secondary calibrations.** While most chronograms in DateLife's database
466 are constructed using primary calibrations (molecular substitution rates or ages obtained
467 from the fossil record or geological events), DateLife summarizes chronograms using
468 secondary calibrations (ages coming from other chronograms). Graur and Martin (2004)
469 cautioned on the increased error and uncertainty in estimated ages when using secondary
470 calibrations in dating analyses. Schenk (2016) showed that, in simulations, divergence times
471 inferred using secondary calibrations are significantly younger than those inferred with
472 primary calibrations, when obtained with Bayesian inference methods, and when priors are
473 implemented in similar ways in both analyses. Accordingly, the scientific community seems
474 to have more confidence in chronograms obtained from a single analysis, using fossil data as
475 primary sources of calibrations (Schenk, 2016), and using fossils that have been widely
476 discussed and curated as calibrations to date other trees, making sure that all data reflect a
477 coherent evolutionary history (Sauquet, 2013), as for example done by Antonelli et al.
478 (2017). There have been attempts to create fossil calibration databases (Ksepka et al., 2015),

479 though these still have room to grow.

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

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

489 Our cross validation analysis might provide some insight in this regard. When ages are
490 estimated with secondary calibrations, nodes closer to the root do tend to be slightly
491 younger than ages estimated with primary calibrations. However, nodes closer to the tip
492 tend to be older when estimated using secondary calibrations with a dating method that
493 does not make any prior assumptions on the nature of the calibrations themselves
494 (Supplementary Figures S2-S20). The only ~~exeption to tjis exception to this~~ was observed
495 on ~~results of the cross-validation analysis~~ cross validation results of the Burns et al. (2014)
496 chronogram, which ~~results in displays~~ much younger node ages when estimated using
497 secondary calibrations (Supplementary Figs. S1, S5).

498 Summarizing chronograms

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

505 score the quality of source chronograms.

506 Some measures that have been proposed are the proportion of lineage sampling and the
507 number of calibrations used (Magallón, 2010; Magallón et al., 2015). Some characteristics
508 that are often cited in published studies as a measure of improved age estimates as compared
509 to previously published estimates are: quality of alignment (missing data, GC content),
510 lineage sampling (strategy and proportion), phylogenetic and dating inference method,
511 number of fossils used as calibrations, support for nodes and ages, and magnitude of
512 confidence intervals.

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

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

527 Effects of phylogenetic sampling on downstream analyses

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

531 In ecology and conservation biology, incorporating at least some data on lineage divergence
532 times represents a relevant improvement for testing alternative hypothesis using
533 phylogenetic distance (Webb). Analysis of species diversification using simulated and
534 empirical phylogenies suggest that using a more completely sampled phylogeny provides
535 estimates that are closer to the true diversification history than when analysing
536 incompletely sampled phylogenies (Chang, Rabosky, & Alfaro, 2020; Cusimano, Stadler, &
537 Renner, 2012; Sun et al., 2008). Ideally phylogenies should be completed using
538 genetic data, but this is a very hard task to achieve for many biological groups. Hence,
539 DateLife's workflow features different ways of estimating assigning node ages in the absence
540 of calibrations and branch length information for certain taxa. "Making up" branch lengths
541 Completing a phylogeny using a stochastic birth-death politomy resolver and a backbone
542 taxonomy is a common practice in scientific publications: Jetz et al. (2012), created a
543 chronogram of all 9, 993 bird species, where 67% had molecular data and the rest was
544 simulated; Rabosky et al. (2018) created a chronogram of 31, 536 ray-finned fishes, of which
545 only 37% had molecular data; Smith and Brown (2018) constructed a chronogram of 353, 185
546 seed plants where only 23% had molecular data. These stochastically resolved chronograms
547 provide less biased diversification rates estimated with methods that account for sampling
548 fractions (Chang et al., 2020; Cusimano et al., 2012), but can also introduce spurious
549 patterns of early bursts of diversification (Cusimano & Renner, 2010; Sun et al., 2020).
550 Notably, Taxonomy based politomy resolvers also introduce topological differences. The
551 study of macroevolutionary processes largely depends on an understanding of the timing of
552 species diversification events, and different phylogenetic and chronogram hypothesis can
553 provide very different overviews of the macroevolutionary history of a biological group. For
554 example, alternative topologies in chronograms from the same biological group can infer
555 very different species diversification patterns (Rabosky, 2015; Title & Rabosky, 2016).
556 Similarly there are worries that patterns of morphological evolution cannot be accurately
557 inferred with these types of phylogenies, as any patterns would be erased by randomization

558 (Rabosky, 2015). We note that the same applies for geographical and morphological
559 dependent diversification analysis. Hence, we suggest that these types of randomly resolved
560 phylogenies can be used as null or neutral models, representing the case of diversification
561 independent of traits and geographical scenario.

562 While stochastic politomy resolvers have been useful to advance research, notably, risks
563 come with this practice. Taken to the extreme, one could make generate a fully resolved,
564 calibrated tree of all modern and extinct taxa using a single taxonomy and a single
565 calibration, using polyclad resolution and branch estimation methods. There has yet to be
566 a thorough analysis of what can go wrong when one extends inferences beyond the data in
567 this way, so we urge caution; we also urge readers and assigning branch lengths following a
568 birth-death diversification model. Clearly, this can lead to a misrepresentation of the true
569 evolutionary history. We urge DateLife users to follow the example of the large tree papers
570 cited above, by carefully considering the statistical assumptions being made, potential biases,
571 and assessing the consistency of the results with prior work.

572 Conclusions

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

581 The DateLife project allows for easy and fast summary of public and state-of-the-art data on
582 time of lineage divergence. It provides a straightforward way to get an informed idea on the
583 state of knowledge of the time frame of evolution of different regions of the tree of life, and

584 allows identification of identifying regions that require more research, or that have conflicting
585 information. It is available as an R package, and as a web-based R shiny application at
586 www.datelife.org Both summary and newly generated trees are useful to evaluate
587 evolutionary hypotheses in different areas of research. The DateLife project helps with
588 should improve awareness of the existing variation in expert time of divergence data, and
589 will-foster exploration of the effect of alternative divergence time hypothesis on the results of
590 analyses, nurturing a culture of more cautious interpretation of evolutionary results.

591 Availability

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

602 Supplementary Material

603 Code used to generate all versions of this manuscript, the biological examples,
604 Supplementary material, including code, biological examples and benchmark results data
605 files and online-only appendices, can be found in the Dryad data repository
606 (<https://doi.org/10.5061/dryad.cnp5hqc6w>), as well as the benchmark of functionalities are
607 available at datelifeMS1, datelife_examples, and datelife_benchmark repositories in
608 LLSR's GitHub accountin the GitHub repositories used to develop the reproducible
609 manuscript (<https://doi.org/10.5281/zenodo.7435094>), the biological examples

610 <https://doi.org/10.5281/zenodo.7435101>, and the software benchmark

611 <https://doi.org/10.5281/zenodo.7435106>.

612 **Funding**

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

617 **Acknowledgements**

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

630 **References**

- 631 632 Agnarsson, I., & Miller, J. A. (2008). Is acctran better than deltran? *Cladistics*, 24(6),
1032–1038.
- 633 634 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.
- 635 636 Ané, C., Eulenstein, O., Piaggio-Talice, R., & Sanderson, M. J. (2009). Groves of
phylogenetic trees. *Annals of Combinatorics*, 13(2), 139–167.
- 637 638 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.
<https://doi.org/10.1093/sysbio/syw066>
- 641 642 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>}
- 643 644 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>}
- 645 646 Bapst, D. W. (2012). Paleotree: An R package for paleontological and phylogenetic analyses
of evolution. *Methods in Ecology and Evolution*, 3(5), 803–807.
<https://doi.org/10.1111/j.2041-210X.2012.00223.x>
- 647 648 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
a cretaceous terrestrial revolution. *New Phytologist*, 218(2), 819–834.
- 649 650 Barker, F. K. (2014). Mitogenomic data resolve basal relationships among passeriform and
passeridan birds. *Molecular Phylogenetics and Evolution*, 79, 313–324.
- 651 652 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2013). Going to

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

- 684 and *Evolution*, 75, 41–77.
- 685 Chamberlain, S. (2018). *bold: Interface to Bold Systems API*. Retrieved from
686 <https://CRAN.R-project.org/package=bold>
- 687 Chamberlain, S. A., & Szöcs, E. (2013). taxize : taxonomic search and retrieval in R [version
688 2; referees: 3 approved]. *F1000Research*, 2(191), 1–29.
689 <https://doi.org/10.12688/f1000research.2-191.v2>
- 690 Chang, J., Rabosky, D. L., & Alfaro, M. E. (2020). Estimating diversification rates on
691 incompletely sampled phylogenies: Theoretical concerns and practical solutions.
692 *Systematic Biology*, 69(3), 602–611.
- 693 Chaves, J. A., Hidalgo, J. R., & Klicka, J. (2013). Biogeography and evolutionary history of
694 the n eotropical genus s altator (a ves: T hraupini). *Journal of Biogeography*, 40(11),
695 2180–2190.
- 696 Claramunt, S., & Cracraft, J. (2015). A new time tree reveals earth history's imprint on the
697 evolution of modern birds. *Science Advances*, 1(11), e1501005.
- 698 Criscuolo, A., Berry, V., Douzery, E. J., & Gascuel, O. (2006). SDM: A fast distance-based
699 approach for (super)tree building in phylogenomics. *Systematic Biology*, 55(5),
700 740–755. <https://doi.org/10.1080/10635150600969872>
- 701 Cusimano, N., & Renner, S. S. (2010). Slowdowns in diversification rates from real
702 phylogenies may not be real. *Systematic Biology*, 59(4), 458–464.
- 703 Cusimano, N., Stadler, T., & Renner, S. S. (2012). A new method for handling missing
704 species in diversification analysis applicable to randomly or nonrandomly sampled
705 phylogenies. *Systematic Biology*, 61(5), 785–792.
- 706 Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M.-K., Turon, X., ... Douzery,
707 E. J. (2018). A phylogenomic framework and timescale for comparative studies of
708 tunicates. *BMC Biology*, 16(1), 1–14.
- 709 Eastman, J. M., Harmon, L. J., & Tank, D. C. (2013). Congruification: Support for time
710 scaling large phylogenetic trees. *Methods in Ecology and Evolution*, 4(7), 688–691.

- 711 https://doi.org/10.1111/2041-210X.12051
- 712 Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high
713 throughput. *Nucleic Acids Research*, 32(5), 1792–1797.
- 714 Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American Naturalist*,
715 125(1), 1–15. Retrieved from http://www.jstor.org/stable/2461605
- 716 Forest, F., Savolainen, V., Chase, M. W., Lupia, R., Bruneau, A., & Crane, P. R. (2005).
717 [Teasing apart molecular-versus fossil-based error estimates when dating](#)
718 [phylogenetic trees: A case study in the birch family \(betulaceae\).](#) *Systematic*
719 [Botany](#), 30(1), 118–133.
- 720 Freckleton, R. P., Harvey, P. H., & Pagel, M. (2002). Phylogenetic analysis and
721 comparative data: A test and review of evidence. *The American Naturalist*.
- 722 GBIF Secretariat. (2022). GBIF Backbone Taxonomy. *Checklist dataset*, (Online Resource
723 accessed via GBIF.org). Retrieved from {https://doi.org/10.15468/39omei }
- 724 Gibb, G. C., England, R., Hartig, G., McLenaghan, P. A., Taylor Smith, B. L., McComish,
725 B. J., ... Penny, D. (2015). New zealand passerines help clarify the diversification of
726 major songbird lineages during the oligocene. *Genome Biology and Evolution*, 7(11),
727 2983–2995.
- 728 Graur, D., & Martin, W. (2004). Reading the entrails of chickens: Molecular timescales of
729 evolution and the illusion of precision. *TRENDS in Genetics*, 20(2), 80–86.
- 730 Hackett, S. J., Kimball, R. T., Reddy, S., Bowie, R. C., Braun, E. L., Braun, M. J., ...
731 others. (2008). A phylogenomic study of birds reveals their evolutionary history.
732 *Science*, 320(5884), 1763–1768.
- 733 Harvey, P. H., Pagel, M. D., & others. (1991). [The comparative method in evolutionary](#)
734 [biology \(Vol. 239\)](#). Oxford university press Oxford.
- 735 Hedges, S. B., Marin, J., Suleski, M., Paymer, M., & Kumar, S. (2015). Tree of life reveals
736 clock-like speciation and diversification. *Molecular Biology and Evolution*, 32(4),
737 835–845. https://doi.org/10.1093/molbev/msv037

- 738 Heibl, C. (2008). *PHYLOCH: R language tree plotting tools and interfaces to diverse*
739 *phylogenetic software packages*. Retrieved from
740 <http://www.christophheibl.de/Rpackages.html>
- 741 Hooper, D. M., & Price, T. D. (2017). Chromosomal inversion differences correlate with
742 range overlap in passerine birds. *Nature Ecology & Evolution*, 1(10), 1526.
- 743 Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic
744 trees. *Bioinformatics*, 17(8), 754–755.
745 <https://doi.org/10.1093/bioinformatics/17.8.754>
- 746 Jetz, W., Thomas, G., Joy, J. J., Hartmann, K., & Mooers, A. (2012). The global diversity
747 of birds in space and time. *Nature*, 491(7424), 444–448.
748 <https://doi.org/10.1038/nature11631>
- 749 Johansson, U. S., Fjeldså, J., & Bowie, R. C. (2008). Phylogenetic relationships within
750 passerida (aves: Passeriformes): A review and a new molecular phylogeny based on
751 three nuclear intron markers. *Molecular Phylogenetics and Evolution*, 48(3), 858–876.
- 752 Katoh, K., Asimenos, G., & Toh, H. (2009). Multiple alignment of dna sequences with mafft.
753 In *Bioinformatics for dna sequence analysis* (pp. 39–64). Springer.
- 754 Kimball, R. T., Oliveros, C. H., Wang, N., White, N. D., Barker, F. K., Field, D. J., ...
755 others. (2019). A phylogenomic supertree of birds. *Diversity*, 11(7), 109.
- 756 Klicka, J., Barker, F. K., Burns, K. J., Lanyon, S. M., Lovette, I. J., Chaves, J. A., & Bryson
757 Jr, R. W. (2014). A comprehensive multilocus assessment of sparrow (aves:
758 Passerellidae) relationships. *Molecular Phylogenetics and Evolution*, 77, 177–182.
- 759 Ksepka, D. T., Parham, J. F., Allman, J. F., Benton, M. J., Carrano, M. T., Cranston, K.
760 A., ... others. (2015). The fossil calibration database—a new resource for divergence
761 dating. *Systematic Biology*, 64(5), 853–859.
- 762 Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio,
763 A., ... others. (2015). Evolution of darwin's finches and their beaks revealed by
764 genome sequencing. *Nature*, 518(7539), 371–375.

- 765 Laubichler, M. D., & Maienschein, J. (2009). *Form and function in developmental evolution.*
766 Cambridge University Press.
- 767 Lepage, D. (2004). *Avibase: The world bird database.* Bird Studies Canada.
- 768 Lepage, D., Vaidya, G., & Guralnick, R. (2014). Avibase—a database system for managing
769 and organizing taxonomic concepts. *ZooKeys*, (420), 117.
- 770 Lerner, H. R., Meyer, M., James, H. F., Hofreiter, M., & Fleischer, R. C. (2011). Multilocus
771 resolution of phylogeny and timescale in the extant adaptive radiation of hawaiian
772 honeycreepers. *Current Biology*, 21(21), 1838–1844.
- 773 Lovette, I. J., Pérez-Emán, J. L., Sullivan, J. P., Banks, R. C., Fiorentino, I.,
774 Córdoba-Córdoba, S., ... others. (2010). A comprehensive multilocus phylogeny for
775 the wood-warblers and a revised classification of the parulidae (aves). *Molecular
776 Phylogenetics and Evolution*, 57(2), 753–770.
- 777 Magallon, S., & Sanderson, M. (2001). Absolute diversification rates in angiosperm clades.
778 *Evolution*, 55(9), 1762–1780.
- 779 Magallón, S. (2010). Using fossils to break long branches in molecular dating: A comparison
780 of relaxed clocks applied to the origin of angiosperms. *Systematic Biology*, 59(4),
781 384–399.
- 782 Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., & Hernández-Hernández, T. (2015).
783 A metacalibrated time-tree documents the early rise of flowering plant phylogenetic
784 diversity. *New Phytologist*, 207(2), 437–453.
- 785 McTavish, E. J., Hinchliff, C. E., Allman, J. F., Brown, J. W., Cranston, K. A., Holder, M.
786 T., ... Smith, S. (2015). Phylesystem: A git-based data store for community-curated
787 phylogenetic estimates. *Bioinformatics*, 31(17), 2794–2800.
- 788 Michonneau, F., Brown, J. W., & Winter, D. J. (2016). rotl: an R package to interact with
789 the Open Tree of Life data. *Methods in Ecology and Evolution*, 7(12), 1476–1481.
790 <https://doi.org/10.1111/2041-210X.12593>
- 791 Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecology Letters*,

- 792 17(4), 508–525. <https://doi.org/10.1111/ele.12251>
- 793 Moyle, R. G., Oliveros, C. H., Andersen, M. J., Hosner, P. A., Benz, B. W., Manthey, J. D.,
794 ... Faircloth, B. C. (2016). Tectonic collision and uplift of wallacea triggered the
795 global songbird radiation. *Nature Communications*, 7(1), 1–7.
- 796 Oliveros, C. H., Field, D. J., Ksepka, D. T., Barker, F. K., Aleixo, A., Andersen, M. J., ...
797 others. (2019). Earth history and the passerine superradiation. *Proceedings of the
798 National Academy of Sciences*, 116(16), 7916–7925.
- 799 Ooms, J., & Chamberlain, S. (2018). *Phylocomr: Interface to 'phylocom'*. Retrieved from
800 <https://CRAN.R-project.org/package=phylocomr>
- 801 Open Tree Of Life, Redelings, B., Cranston, K. A., Allman, J., Holder, M. T., & McTavish,
802 E. J. (2016). Open Tree of Life APIs v3.0. *Open Tree of Life Project*, (Online
803 Resources). Retrieved from
804 {[https://github.com/OpenTreeOfLife/germinator/wiki/Open-Tree-of-Life-Web-
APIs](https://github.com/OpenTreeOfLife/germinator/wiki/Open-Tree-of-Life-Web-
805 APIs)}
- 806 Open Tree Of Life, Redelings, B., Sánchez Reyes, L. L., Cranston, K. A., Allman, J., Holder,
807 M. T., & McTavish, E. J. (2019). Open tree of life synthetic tree v12.3. *Zenodo*.
808 Retrieved from <https://doi.org/10.5281/zenodo.3937742>
- 809 Ödeen, A., Håstad, O., & Alström, P. (2011). Evolution of ultraviolet vision in the largest
810 avian radiation-the passerines. *BMC Evolutionary Biology*, 11(1), 1–8.
- 811 Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and
812 evolution in R language. *Bioinformatics*, 20(2), 289–290.
- 813 Parchman, T. L., Benkman, C. W., & Mezquida, E. T. (2007). Coevolution between
814 hispaniolan crossbills and pine: Does more time allow for greater phenotypic
815 escalation at lower latitude? *Evolution*, 61(9), 2142–2153.
- 816 Päckert, M., Martens, J., Sun, Y.-H., Severinghaus, L. L., Nazarenko, A. A., Ting, J., ...
817 Tietze, D. T. (2012). Horizontal and elevational phylogeographic patterns of
818 himalayan and southeast asian forest passerines (aves: Passeriformes). *Journal of*

- 819 *Biogeography*, 39(3), 556–573.
- 820 Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., FitzJohn, R. G.,
821 ... Harmon, L. J. (2014). Geiger v2. 0: An expanded suite of methods for fitting
822 macroevolutionary models to phylogenetic trees. *Bioinformatics*, 30(15), 2216–2218.
- 823 Posadas, P., Crisci, J. V., & Katinas, L. (2006). Historical biogeography: A review of its
824 basic concepts and critical issues. *Journal of Arid Environments*, 66(3), 389–403.
- 825 Powell, A. F., Barker, F. K., Lanyon, S. M., Burns, K. J., Klicka, J., & Lovette, I. J. (2014).
826 A comprehensive species-level molecular phylogeny of the new world blackbirds
827 (icteridae). *Molecular Phylogenetics and Evolution*, 71, 94–112.
- 828 Price, T. D., Hooper, D. M., Buchanan, C. D., Johansson, U. S., Tietze, D. T., Alström, P.,
829 ... others. (2014). Niche filling slows the diversification of himalayan songbirds.
830 *Nature*, 509(7499), 222.
- 831 Pulgarín-R, P. C., Smith, B. T., Bryson Jr, R. W., Spellman, G. M., & Klicka, J. (2013).
832 Multilocus phylogeny and biogeography of the new world pheucticus grosbeaks (aves:
833 Cardinalidae). *Molecular Phylogenetics and Evolution*, 69(3), 1222–1227.
- 834 [Rabosky, D. L. \(2015\). No substitute for real data: A cautionary note on the use of](#)
835 [phylogenies from birth–death polytomy resolvers for downstream comparative](#)
836 [analyses. *Evolution*, 69\(12\), 3207–3216.](#)
- 837 Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., ... others.
838 (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*,
839 559(7714), 392.
- 840 Ramshaw, J., Richardson, D., Meatyard, B., Brown, R., Richardson, M., Thompson, E., &
841 Boulter, D. (1972). The time of origin of the flowering plants determined by using
842 amino acid sequence data of cytochrome c. *New Phytologist*, 71(5), 773–779.
- 843 Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The barcode of life data system
844 (<http://www.Barcodinglife.Org>). *Molecular Ecology Notes*, 7(3), 355–364.
- 845 R Core Team. (2018). *R: a language and environment for statistical computing*. Vienna,

- 846 Austria: R Foundation for Statistical Computing.
- 847 Rees, & Cranston, K. (2017). Automated assembly of a reference taxonomy for phylogenetic
848 data synthesis. *Biodiversity Data Journal*, (5).
- 849 Rees, Vandepitte, L., Decock, W., & Vanhoorne, B. (2017). IRMNG 2006–2016: 10 Years of
850 a Global Taxonomic Database. *Biodiversity Informatics*, 12.
- 851 Revell, L. J. (2012). Phytools: An r package for phylogenetic comparative biology (and other
852 things). *Methods in Ecology and Evolution*, 3, 217–223.
- 853 Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference
854 under mixed models. *Bioinformatics*, 19(12), 1572–1574.
855 <https://doi.org/10.1093/bioinformatics/btg180>
- 856 [Roquet, C., Lavergne, S., & Thuiller, W. \(2014\). One tree to link them all: A phylogenetic
857 dataset for the european tetrapoda. PLoS Currents, 6.](#)
- 858 Sanchez-Reyes, L. L., & O'Meara, B. (2022). datelifeplot: Methods to plot chronograms
859 and outputs of the datelife package. *R Package Release V0.2.2*. Retrieved from
860 <https://zenodo.org/badge/latestdoi/381501451>
- 861 Sanchez-Reyes, L. L., O'Meara, B., Eastman, J., Heath, T., Wright, A., Schliep, K., ...
862 Alfaro, M. (2022). datelife: Scientific Data on Time of Lineage Divergence for Your
863 Taxa. *R Package Release V0.6.2*. Retrieved from
864 <https://doi.org/10.5281/zenodo.593938>
- 865 Sanderson, M. (2002). Estimating Absolute Rates of Molecular Evolution and Divergence
866 Times: A Penalized Likelihood Approach. *Molecular Biology and Evolution*, 19(1),
867 101–109. <https://doi.org/10.1093/oxfordjournals.molbev.a003974>
- 868 Sanderson, M. (2003). r8s: Inferring Absolute Rates of Molecular Evolution and Divergence
869 Times in the Absence of a Molecular Clock. *Bioinformatics*, 19(2), 301–302.
- 870 Sanderson, M., & Doyle, J. (2001). Sources of error and confidence intervals in estimating
871 the age of angiosperms from rbcL and 18S rDNA data. *American Journal of Botany*,
872 88(8), 1499–1516.

- 873 Sauquet, H. (2013). A practical guide to molecular dating. *Comptes Rendus Palevol*, 12(6),
874 355–367.
- 875 Sauquet, H., Ramírez-Barahona, S., & Magallón, S. (2021). *The age of flowering plants is*
876 *unknown*.
- 877 Schenk, J. J. (2016). Consequences of secondary calibrations on divergence time estimates.
878 *PLoS ONE*, 11(1). <https://doi.org/10.1371/journal.pone.0148228>
- 879 Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in r. *Bioinformatics*, 27(4), 592–593.
- 880 Schoch, C. L., Ciufo, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R., ...
881 others. (2020). NCBI Taxonomy: a Comprehensive Update on Curation, Resources
882 and Tools. *Database*, 2020.
- 883 Selvatti, A. P., Gonzaga, L. P., & Moraes Russo, C. A. de. (2015). A paleogene origin for
884 crown passerines and the diversification of the oscines in the new world. *Molecular*
885 *Phylogenetics and Evolution*, 88, 1–15.
- 886 Smith, S., & Brown, J. (2018). Constructing a broadly inclusive seed plant phylogeny.
887 *American Journal of Botany*, 105(3), 302–314.
- 888 Smith, S., & O'Meara, B. (2012). TreePL: Divergence time estimation using penalized
889 likelihood for large phylogenies. *Bioinformatics*, 28(20), 2689–2690.
890 <https://doi.org/10.1093/bioinformatics/bts492>
- 891 Stoltzfus, A., Lapp, H., Matasci, N., Deus, H., Sidlauskas, B., Zmasek, C. M., ... Jordan, G.
892 (2013). Phylotastic! Making tree-of-life knowledge accessible, reusable and convenient.
893 *BMC Bioinformatics*, 14. <https://doi.org/10.1186/1471-2105-14-158>
- 894 Sun, M., Folk, R. A., Gitzendanner, M. A., Soltis, P. S., Chen, Z., Soltis, D. E., &
895 Guralnick, R. P. (2020). *Estimating rates and patterns of diversification with*
896 *incomplete sampling: A case study in the rosids*. *American Journal of Botany*
897 107(6), 895–909.
- 898 Tietze, D. T., Päckert, M., Martens, J., Lehmann, H., & Sun, Y.-H. (2013). Complete
899 phylogeny and historical biogeography of true rosefinches (aves: Carpodacus).

- 900 *Zoological Journal of the Linnean Society*, 169(1), 215–234.
- 901 Title, P. O., & Rabosky, D. L. (2016). Do Macrophylogenies Yield Stable Macroevolutionary
902 Inferences? An Example from Squamate Reptiles. *Systematic Biology*, syw102.
903 <https://doi.org/10.1093/sysbio/syw102>
- 904 Treplin, S., Siegert, R., Bleidorn, C., Thompson, H. S., Fotso, R., & Tiedemann, R. (2008).
905 Molecular phylogeny of songbirds (aves: Passeriformes) and the relative utility of
906 common nuclear marker loci. *Cladistics*, 24(3), 328–349.
907 Uyeda, J. C., Pennell, M. W., Miller, E. T., Maia, R., & McClain, C. R. (2017). The
908 evolution of energetic scaling across the vertebrate tree of life. *The American*
909 Naturalist, 190(2), 185–199.
- 910 Vos, R. A., Balhoff, J. P., Caravas, J. A., Holder, M. T., Lapp, H., Maddison, W. P., ...
911 others. (2012). NeXML: Rich, extensible, and verifiable representation of
912 comparative data and metadata. *Systematic Biology*, 61(4), 675–689.
913 <https://doi.org/10.1093/sysbio/sys025>
- 914 Webb, C. (2000). Exploring the Phylogenetic Structure of Ecological Communities : An
915 Example for Rain Forest Trees. *The American Naturalist*, 156(2), 145–155.
- 916 Webb, C., Ackerly, D., & Kembel, S. (2008). Phylocom: Software for the analysis of
917 phylogenetic community structure and trait evolution. *Bioinformatics*, 24(18),
918 2098–2100. <https://doi.org/10.1093/bioinformatics/btn358>
- 919 Webb, C., & Donoghue, M. (2005). Phylomatic: Tree assembly for applied phylogenetics.
920 *Molecular Ecology Notes*, 5(1), 181–183.
- 921 Weir, J., & Schluter, D. (2008). Calibrating the avian molecular clock. *Molecular Ecology*,
922 17(10), 2321–2328.
- 923 Zuccon, D., Prŷs-Jones, R., Rasmussen, P. C., & Ericson, P. G. (2012). The phylogenetic
924 relationships and generic limits of finches (fringillidae). *Molecular Phylogenetics and*
925 *Evolution*, 62(2), 581–596.