

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

18 Abstract

19 Achieving a high-quality reconstruction of a phylogenetic tree Chronograms

20 phylogenies with branch lengths proportional to absolute time (chronogram) is a difficult

21 and time-consuming task. But the increased availability of fossil and molecular data, and

22 time-efficient analytical techniques has resulted in many recent publications of large

23 chronograms for a large number and wide diversity of organisms. Knowledge of the time-

24 represent key data on evolutionary time frame of organisms is key for research in the

25 natural sciences. It also represent valuable information for the study of natural processes in

26 many areas of biological research. Chronograms also provide valuable information that can

27 be used for education, science communication, and policy decisions. When chronograms are

28 shared in public and open databases, this wealth of expertly-curated and peer-reviewed

29 data on evolutionary timeframe is exposed in a programmatic and reusable way, as intensive

30 and localized efforts have improved data sharing practices, as well as incentivized open

31 science in biology Yet, achieving a high-quality reconstruction of a chronogram is a difficult

32 and resource-consuming task. Here we present DateLife, a service implemented as an R

33 package and an R Shiny website web application available at www.datelife.org, that provides

34 functionalities services for efficient and easy finding discovery, summary, reuse, and

35 reanalysis of node age data mined from a curated database of expert, peer-reviewed, public

36 data on time frame of evolution and openly available chronograms. The main DateLife

37 workflow constructs a chronogram for any given combination of taxon names by searching a

38 starts with one or more scientific taxon names provided by a user. Names are processed

39 and standardized to a unified taxonomy. Then, DateLife runs a name match across its local

40 chronogram database constructed and curated from the that is curated from Open Tree of

41 LifePhylesystem phylogenetic database, which incorporates phylogenetic data from the

42 TreeBASE database as well. We implement and test methods for summarizing time data

43 from multiple source chronograms using supertree and congruification algorithms, and
44 using age data extracted from source chronograms as secondary calibration points to add
45 branch lengths proportional to absolute time to a tree topology. DateLife will be useful to 's
46 phylogenetic repository, and extracts all chronograms that contain at least two queried
47 taxon names, along with their metadata. Finally, node ages from matching chronograms
48 are mapped upon a chosen tree topology using the congruification algorithm. Congruified
49 node ages are used as secondary calibrations to date the chosen topology, with or without
50 initial branch lengths, using different phylogenetic dating methods such as BLADJ, treePL,
51 PATHd8 and MrBayes. We performed a cross-validation test to compare node ages
52 resulting from a DateLife analysis (i.e. phylogenetic dating using secondary calibrations) to
53 those from the original chronograms (i.e. obtained with primary calibrations), and found
54 that DateLife's node age estimates appear generally correlated to those from the original
55 chronograms, with the largest variation in ages occurring around topologically deeper
56 nodes. Software such as DateLife can help increase awareness of the existing variation in
57 alternative hypothesis of evolutionary time frame for the same organisms, and can foster
58 support exploration of the effect of alternative evolutionary timing hypotheses on the
59 results of chronogram hypotheses on downstream analyses, providing a framework for a
60 more informed interpretation of evolutionary results.

61 *Keywords:* Tree; Phylogeny; Scaling; Dating; Ages; Divergence times; Open Science;
62 Congruification; Supertree; Calibrations; Secondary calibrations.

63 Word count: 5393-6111

64 Chronograms –phylogenies with branch lengths proportional to time– provide key data
65 on evolutionary time frame for the study of natural processes in many areas of biological
66 research, such as comparative analysis (Freckleton, Harvey, & Pagel, 2002; Harvey, Pagel,
67 & others, 1991), developmental biology (Delsuc et al., 2018; Laubichler & Maienschein,
68 2009), conservation biology and ecology (Felsenstein, 1985; Webb, 2000), historical
69 biogeography (Posadas, Crisci, & Katinas, 2006), and species diversification (Magallon &
70 Sanderson, 2001; Morlon, 2014).

71 Building a chronogram is not an easy task. It requires obtaining and curating ~~data-a~~
72 homology hypothesis to construct a phylogeny, selecting and placing appropriate calibrations
73 on the phylogeny using independent age data points from the fossil record or other dated
74 events, and inferring ~~the-a~~ full dated tree; ~~it also generally requires~~. All of this entails
75 specialized biological training, taxonomic domain knowledge, and a non-negligible amount of
76 research time, computational resources and funding.

77 Here we present the DateLife project which has the main goal of ~~eapturing-extracting~~
78 and exposing age data from published chronograms, ~~and making these-making age~~ data
79 readily accessible to the wider community for reuse and reanalysis, ~~for~~ research, teaching,
80 and science communication and policy. DateLife’s core software application is available as an
81 R package (Sanchez-Reyes et al., 2022), and as an online Rshiny interactive website at
82 www.datelife.org. It features key elements for scientific reproducibility, such as a curated,
83 versioned, open and fully public ~~source-chronogram~~ database (McTavish et al., 2015) ~~,-data~~
84 ~~stored-and-available-in-a-computer-readable~~-that stores data in a computer-readable format
85 (Vos et al., 2012)~~,-;~~ automated and programmatic ways of accessing ~~the-data-and~~
86 downloading the data in a computer-readable format also (Stoltzfus et al., 2013); and
87 methods to summarize and compare the data.

89 DateLife's core software application consists of applications are implemented in the R
90 package `datelife`. Its current stable version – v0.6.6, is available from The Comprehensive
91 R Archive Network (CRAN) repository (Sanchez-Reyes et al., 2022), and relies on
92 functionalities from various other biological R packages: ape (Paradis, Claude, & Strimmer,
93 2004), bold (Chamberlain, 2018), geiger (Pennell et al., 2014), `msa` (Bodenhofer, Bonatesta,
94 Horejš-Kainrath, & Hochreiter, 2015), paleotree (Bapst, 2012), phyloch (Heibl, 2008),
95 phylocomr (Ooms & Chamberlain, 2018), phytools (Revell, 2012), rotl (Michonneau, Brown,
96 & Winter, 2016), and taxize (Chamberlain, 2018; Chamberlain & Szöcs, 2013). Figure 1
97 provides a graphical summary of the three main steps of the DateLife workflow: creating a
98 search query, searching a database, and summarizing results from the search.

99 *Creating a Search Query*

100 DateLife starts by processing an input consisting of at least two taxon names, which
101 one taxon scientific name. Multiple input names can be provided as a comma separated
102 character string or as tip labels on a tree. If the input is a tree, it can be provided as a
103 classic newick character string (Archie et al., 1986), or as a “phylo” R object (Paradis et al.,
104 2004). The input tree is not required to have branch lengths, and its topology is used in the
105 summary steps described in the next section.

106 DateLife accepts scientific names that can belong to any inclusive taxonomic group
107 (e.g., genus, family, tribe, etc.) or a binomial species name. Subspecies and variants are
108 ignored. If an input taxon name belongs to an inclusive taxonomic group, DateLife has two
109 alternative behaviors defined by processes input scientific names using a Taxonomic Name
110 Resolution Service (TNRS), which increases the probability of correctly finding the queried
111 taxon names in the chronogram database. TNRS detects, corrects and standardizes name
112 misspellings and typos, variant spellings and authorities, and nomenclatural synonyms to a
113 single taxonomic standard (Boyle et al., 2013). TNRS also allows to correctly choose
114 between homonyms, by considering other taxa provided as input to infer the taxonomic

115 context of the flag. If the flag is active, DateLife retrieves all species names within the
116 inclusive taxonomic group following a standard taxonomy of choice, and adds them to the
117 input string. Taxonomies currently supported by DateLife are homonym. DateLife
118 implements TNRS using the Open Tree of Life (OpenTree) unified Taxonomy (OTT, Open
119 Tree Of Life et al., 2016; Rees & Cranston, 2017), as standard, storing taxonomic
120 identification numbers (OTT ids) for further processing and analysis. Other taxonomies
121 currently supported by DateLife are the National Center of Biotechnology Information
122 (NCBI) taxonomic database (Schoch et al., 2020), the Global Biodiversity Information
123 Facility (GBIF) taxonomic backbone (GBIF Secretariat, 2022), and the Interim Register of
124 Marine and Nonmarine Genera (IRMNG) database (Rees et al., 2017). If the
125 flag is

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

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

141 The processed input Species and subspecific variant names are processed and
142 searched as provided by the user. The processed taxon names are saved as an R object of a
143 newly defined class, `datelifeQuery`, that is used in the following steps. This object contains
144 the standardized names input names standardized to a taxonomy of choice (OTT by
145 default), the corresponding OTT identification id numbers, and the topology of the input
146 tree if any an input tree, if one was provided.

147 *Searching a Chronogram Database*

148 At the time of writing of this manuscript (~~Jun 22, 2022~~Feb 21, 2023), DateLife's
149 chronogram database latest version consist of 253 chronograms published in 187 different
150 studies. It is curated from OpenTree's phylogenetic database, the Phylesystem, which
151 constitutes an open source of expert and peer-reviewed phylogenetic knowledge with rich
152 metadata (McTavish et al., 2015), which allows automatic and reproducible assembly of our
153 chronogram database. Datelife's chronogram database is navigable as an R data object
154 within the `datelife` R package.

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

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

177 *Summarizing Search Results*

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

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

191 *Choosing a Topology*

DateLife requires a tree topology to summarize age data upon. We recommend that users provide as input a tree topology from the literature, or one of their own making. If no topology is provided, DateLife automatically extracts one from the OpenTree synthetic tree, a phylogeny currently encompassing 2.3 million taxa across all life, assembled from 1, 239 published phylogenetic trees and OpenTree's unified Taxonomy, OTT (Open Tree Of Life et al., 2019). Alternatively, DateLife can combine topologies from source chronograms using a supertree approach ~~To combine topologies from source chronograms into a single summary (or supertree) topology, the DateLife workflow (Criscuolo, Berry, Douzery, & Gascuel, 2006).~~

To do this, DateLife first identifies the source chronograms that form a grove, roughly, a sufficiently overlapping set of taxa between trees, by implementing definition 2.8 for n-overlap from Ané et al. (2009). If the source chronograms do not form a grove, the supertree reconstruction will fail. In rare cases, a group of trees can have multiple groves. By default, DateLife chooses the grove with the most taxa, however, the “criterion = trees” flag allows the user to choose the grove with the most trees instead. 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 Applying Secondary Calibrations

~~Input topologies from OpenTree or the supertreeapproach 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~~

218 branch length information provides additional data for nodes without secondary date
 219 calibrations, topologies without branch lengths can also be dated.

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

230 For each congruent node, the pairwise distances that traverse that node are
 231 summarized into a single summary matrix using classic summary statistics (i.e., mean,
 232 median, minimum and maximum ages), and the Supermatrix Distance Method (SDM;
 233 Criscuolo, Berry, Douzery, & Gascuel, et al., 2006), which deforms patristic distance
 234 matrices by minimizing variance and then averaging them. These single summary taxon pair
 235 age matrices (Summarized calibrations) can be applied as are stored as summarized
 236 calibrations that can be used as secondary calibrations to date a tree topology, using
 237 different - with or without initial branch lengths, using phylogenetic dating methods
 238 currently supported within DateLife: BLADJ (Webb, Ackerly, & Kembel, 2008; Webb &
 239 Donoghue, 2005), MrBayes (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003),
 240 PATHd8 (Britton, Anderson, Jacquet, Lundqvist, & Bremer, 2007), BLADJ (Webb, Ackerly,
 241 & Kembel, 2008; Webb & Donoghue, 2005), and treePL (Smith & O'Meara, 2012).

242 By default, DateLife implements the

243

Dating a Tree Topology

244 ***With no branch lengths.***— When producing or obtaining a tree with branch
245 lengths for a group of interest is not possible, DateLife can date a topology without branch
246 lengths, obtained from OpenTree or by implementing the supertree approach described
247 above, by implementing the Branch Length Adjuster (BLADJ) algorithm to obtain a fully
248 dated topology. BLADJ fixes node ages that have calibration data, and distributes time
249 between ; Webb et al., 2008; Webb & Donoghue, 2005) algorithm, which requires no initial
250 branch lengths. The algorithm starts by fixing ages for nodes with calibration data upon
251 the given topology. Then, it distributes time for nodes with no data evenly between nodes
252 with calibration data. This minimizes calibrated nodes, minimizing age variance in the
253 resulting chronogram (Webb et al., 2008), which has proven useful for ecological analyses
254 that require age data (Webb et al., 2008). BLADJ does not use branch lengths even when
255 they are present in the input tree or summarizing topology. When there is conflict in ages
256 between nodes with calibration data between ages of calibrated nodes, BLADJ ignores node
257 ages that are older than the age of a parent node. BLADJ The BLADJ algorithm requires a
258 root age estimate to run. If there is no information on the age of the root in the chronogram
259 database, users can provide an estimate from the literature. If none is provided, DateLife
260 assigns an arbitrary age to the root as 10% older than the oldest age available within the
261 group will not return a dated topology and will provide a warning message along with
262 suggestions on how the user can provide an age for the root so that a fully dated tree can
263 be returned.

264

In the absence of genetic data, we consider that BLADJ is the most agnostic way to
265 assign ages to nodes with no available data, as it does not require any assumptions on the
266 underlying model of branch length distribution. It is however common practice in the
267 literature to use a birth-death model to assign ages to nodes with no genetic data (Jetz,
268 Thomas, Joy, Hartmann, & Mooers, 2012; Rabosky et al., 2018; Smith & Brown, 2018). To

269 do so, DateLife implements MrBayes (Huelsenbeck & Ronquist, 2001; Ronquist &
270 Huelsenbeck, 2003), using nodes with published age data as calibration priors on a fixed
271 topology, a simple birth-death model with parameters that can be determined by the user,
272 and no genetic data.

273 Alternative phylogenetic dating options supported in DateLife (MrBayes, PATH8,
274 TreePL) **Dating a tree with branch lengths.**— Relative branch lengths provide key
275 information for phylogenetic dating, specifically for nodes without any calibration data
276 available. While, using initial branch lengths data is the golden standard for phylogenetic
277 dating analyses, producing such requires assembling and curating a homology hypothesis,
278 and choosing and implementing a method for phylogenetic inference. DateLife implements
279 a workflow to streamline this process by leveraging on genetic data from the Barcode of
280 Life Data System, BOLD (Ratnasingham & Hebert, 2007) to obtain markers for input taxa.
281 By default, mined genetic sequences are aligned with MUSCLE (Edgar, 2004) using
282 functions from the msa R package (Bodenhofer et al., 2015). Alternatively, sequences can
283 be aligned with MAFFT (Katoh, Asimenos, & Toh, 2009), using functions from the ape R
284 package (Paradis et al., 2004). The BOLD sequence alignment is then used to obtain initial
285 branch lengths with the accelerated transformation (ACCTRAN) parsimony algorithm,
286 which resolves ambiguous character optimization by assigning changes along branches of
287 the tree as close to the root as possible (Agnarsson & Miller, 2008), resulting in older nodes
288 (Forest et al., 2005). The parsimony branch lengths are then optimized using Maximum
289 Likelihood, given the alignment, the topology and a simple Jukes-Cantor model, producing
290 a BOLD tree with branch lengths proportional to expected number of substitutions per
291 site. Both parsimony and ML optimizations are done with functions from the phangorn
292 package (Schliep, 2011). The BOLD workflow is currently only available on DateLife's R
293 package and not on the web application.

294 Phylogenetic dating methods supported in DateLife that incorporate branch length

295 information from the input topology in combination with the ~~calibrations~~. ~~secondary~~
296 ~~calibrations include:~~ PATHd8~~is~~, a non-clock, rate-smoothing method ~~to date trees~~ (Britton
297 et al., 2007)~~to date trees~~; treePL (Smith & O'Meara, 2012), ~~is~~a semi-parametric,
298 rate-smoothing, penalized likelihood dating method (Sanderson, 2002). ~~The~~; ~~and~~ MrBayes
299 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003)~~approach in DateLife uses~~
300 ~~the calibrations as priors on node ages~~, a ~~Bayesian inference program implementing Markov~~
301 ~~chain Monte Carlo (MCMC) methods to estimate a posterior distribution of model~~
302 ~~parameters~~.

303 *Visualizing Results*

304 Finally, users can save all source and summary chronograms in formats ~~that permit~~
305 ~~reuse and reanalyses (newick and allowing for reuse and reanalysis, such as newick and the~~
306 R “phylo” format), as well as visualize and compare results graphically, or ~~Input and~~
307 ~~summary chronograms can be visualized and compared graphically, and users can~~ construct
308 their own graphs using DateLife’s chronogram plot generation functions available from the R
309 package `datelifeplot` (Sanchez-Reyes & O’Meara, 2022).

310 BENCHMARK

311 ~~R package~~ `datelife` ~~'s R package~~ code speed was tested on an Apple iMac with one
312 3.4 GHz Intel Core i5 processor. We registered variation in computing time of query
313 processing and search through the database relative to number of queried taxon names.
314 Query processing time increases roughly linearly with number of input taxon names, and
315 increases considerably if Taxonomic Name Resolution Service (TNRS) is activated. Up to
316 ten thousand names can be processed and searched in less than 30 minutes with the most
317 time consuming settings. Once names have been processed as described in methods, a name
318 search through the chronogram database can be performed in less than a minute, even with
319 a very large number of taxon names (Fig. 2).

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

CASE STUDIES

We illustrate the DateLife workflow using a family within the [Passeriforme](#)—[passeriform](#)
birds encompassing the true finches, Fringillidae, as case study. On a small example, we
analysed 6 bird species, and results from each step of the workflow are shown in Fig. 3. As a
second example, we analysed 289 bird species in the family Fringillidae that are included in
the NCBI taxonomy. The resulting summary chronogram is shown in Fig. 5, and results
from previous steps of the workflow are available as Supplementary Figures.

A Small Example

Creating a search query.— We chose 6 bird species within the Passeriformes. The sample includes two species of cardinals: the black-thighed grosbeak – *Pheucticus tibialis* and the crimson-collared grosbeak – *Rhodothraupis celaeno*; three species of buntings: the yellowhammer – *Emberiza citrinella*, the pine bunting – *Emberiza leucocephalos* and the yellow-throated bunting – *Emberiza elegans*; and one species of tanager, the vegetarian finch – *Platyspiza crassirostris*. Processing of input names found that *Emberiza elegans* is synonym for *Schoeniclus elegans* in the default reference taxonomy (OTT v3.3, June 1, 2021). For a detailed discussion on the state of the synonym, refer to Avibase (Avibase, 2022; Lepage, 2004; Lepage, Vaidya, & Guralnick, 2014). Discovering this synonym allowed assigning five age data points for the parent node of *Emberiza elegans*, shown as *Schoeniclus elegans* in

345 figure 3A, which would not have had any data otherwise.

346 ***Searching the database.***— DateLife used the processed input names to search the
347 local chronogram database and found 9 matching chronograms in 6 different studies (Fig.
348 3B). Three studies matched five input names (Barker, Burns, Klicka, Lanyon, & Lovette,
349 2015; Hedges, Marin, Suleski, Paymer, & Kumar, 2015; Jetz, Thomas, Joy, Hartmann, &
350 Mooers, et al., 2012), one study matched four input names (Hooper & Price, 2017) and two
351 studies matched two input names (Barker, Burns, Klicka, Lanyon, & Lovette, 2013; Burns et
352 al., 2014). No studies matched all input names. Together, source chronograms provide 28
353 unique age data points, covering all nodes on our chosen tree topology to date (Table 1).

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

363

An Example with the Family of True Finches

364

Creating a query.— To obtain ages for all species within the family of true finches,

365 Fringillidae, we ran a DateLife query using the “get species from taxon” flag, which gets all
366 recognized species names within a named group from a taxonomy of choice. Following the
367 NCBI taxonomy, our DateLife query has 289 Fringillidae species names. This
368 taxon-constrained approach implies that the ~~final results of a~~ full DateLife analysis will be
369 ~~done performed~~ using a tree topology and ages ~~for the species in a named available for~~
370 ~~species names from a given taxonomic~~ group, which do not necessarily correspond to a
371 monophyletic group. Users can change this ~~behaviour by providing a monophyletic tree~~
372 ~~behavior by providing all species names corresponding to a monophyletic group~~ as input for
373 a DateLife search, or ~~as a tree topology for a monophyletic tree to construct a~~ DateLife
374 summary.

375

Searching the database.— Next, we used the processed species names in our

376 DateLife query to identify chronograms with at least two Fringillidae species as tip taxa.
377 The DateLife search identified ~~13 chronograms containing at least two Fringillidae species~~¹⁹
378 chronograms matching this criteria, published in ~~9–13~~ different studies (Barker et al., 2013,
379 2015; Burns et al., 2014; Claramunt & Cracraft, 2015; Gibb et al., 2015; Hedges et al., 2015;
380 Hooper & Price, 2017; Jetz et al., 2012; Kimball et al., 2019; Oliveros et al., 2019; Price et
381 al., 2014; Roquet, Lavergne, & Thuiller, 2014; Uyeda, Pennell, Miller, Maia, & McClain,
382 2017). Once identified, DateLife pruned these matching chronograms to ~~keep Fringillidae~~
383 ~~species names on tips only remove tips that do not belong to the queried taxon names~~, and
384 transformed these pruned chronograms to pairwise distance matrices, revealing 1, 206
385 different age data points available for species within the Fringillidae (~~Supplementray~~
386 Supplementary Table S1).

387

Summarizing search results.— The final step ~~is to congrify and summarize~~

388 entailed congrifying and summarizing the age data available for the Fringillidae species

389 into two single summary chronograms, using two different types of summary ages, median
390 and SDM. As explained in the “Description” section, a tree topology to summarize age data
391 upon is required. By default, DateLife uses the topology from OpenTree’s synthetic tree that
392 contains ~~the species in the search query to summarize age data upon all taxa from the search~~
393 query. According to OpenTree’s synthetic tree, species belonging to the family Fringillidae
394 do not form a monophyletic group (Fig. 4). Hence, a topology containing only the 289
395 species from the original query was extracted from Open Tree of Life’s synthetic tree v12.3
396 (Open Tree Of Life et al., 2019).

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

402 CROSS-VALIDATION TEST

403 We performed a cross validation ~~analysis of the DateLife workflow test of a DateLife~~
404 analysis using the Fringillidae ~~ehronograms~~. ~~We used the source chronograms obtained~~
405 above (Supplementary Figs. S2-S20). We used as inputs for a DateLife analysis all
406 individual tree topologies from each of the 19 source chronograms from 13 studies ~~as inputs~~,
407 treating their node ages as unknown. We ~~then estimated dates for these topologies using~~
408 ~~the node ages~~ congruified node ages extracted from chronograms from ~~the chronograms~~
409 ~~from the other studies as calibrations and smoothing using BLADJ all other studies upon~~
410 ~~the individual topologies, effectively excluding original ages from each topology. Average~~
411 ~~node ages per node were applied as secondary calibrations, smoothed with the BLADJ~~
412 ~~algorithm~~. We found that node ages from ~~original study~~ the original studies, and ages
413 estimated using all other age data available are largely correlated (Fig. 6). For five studies,
414 ~~Dateli~~ DateLife tended to underestimate ages for topologically deeper nodes (those with

415 many descendant taxa, aka “closer to the root”) relative to the original estimate, and
416 overestimate ages for nodes closer to the tips. Accordingly, root ages are generally older in
417 the original study than estimated using cross-validated ages (Supplementary Fig. S1). In
418 general, topologically deeper nodes display the largest age variation between node ages
419 from the original chronograms and ages summarized with DateLife.

420 DISCUSSION

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

426 A total of 99,474 unique terminal taxa are represented in DateLife’s database.

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

436 While it is known that incorporating at least some data on lineage divergence times
437 represents a relevant improvement for testing alternative hypothesis using phylogenetic
438 distance in ecological and conservation biology studies (Webb et al., 2008). Next, we
439 discuss some particularities of using summarized node age data from a DateLife analysis

440 that users should be aware of.

441 *Age Variation in Source Chronograms*

442 Conflict in estimated ages among alternative studies is common in the literature. See,
443 for example, the robust ongoing debate about crown group age of angiosperms
444 (Barba-Montoya, Reis, Schneider, Donoghue, & Yang, 2018; Magallón, Gómez-Acevedo,
445 Sánchez-Reyes, & Hernández-Hernández, 2015; Ramshaw et al., 1972; Sanderson & Doyle,
446 2001; Sauquet, Ramírez-Barahona, & Magallón, 2021). Source chronograms available for the
447 same organisms have potentially been estimated implementing calibrations very differently.

448 ~~For example~~ In the Fringillidae example above, the chronograms from Burns et al. (2014)
449 were inferred using molecular substitution rate estimates across birds (Weir & Schlüter,
450 2008), and have much older age estimates for the same nodes than chronograms that were
451 inferred using fossils as calibrations (Figs. 5, 6; Supplementary Figs. S1, S5).

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

460 *Primary vs Secondary Calibrations*

461 While most chronograms in DateLife's database are constructed using primary
462 calibrations (molecular substitution rates or ages obtained from the fossil record or geological
463 events), DateLife summarizes chronograms using secondary calibrations (ages coming from
464 other chronograms). Graur and Martin (2004) cautioned on the increased error and

465 uncertainty in estimated ages when using secondary calibrations in dating analyses. Schenk
466 (2016) showed that, in simulations, divergence times inferred using secondary calibrations are
467 significantly younger than those inferred with primary calibrations, when obtained with
468 Bayesian inference methods, and when priors are implemented in similar ways in both
469 analyses. Accordingly, the scientific community seems to have more confidence in
470 chronograms obtained from a single analysis, using fossil data as primary sources of
471 calibrations (Schenk, 2016), and using fossils that have been widely discussed and curated as
472 calibrations to date other trees, making sure that all data reflect a coherent evolutionary
473 history (Sauquet, 2013), as for example done by Antonelli et al. (2017). There have been
474 attempts to create fossil calibration databases (Ksepka et al., 2015), though these still have
475 room to grow.

476 It seems that using several (as opposed to just a few) secondary calibrations can
477 provide sufficient information to alleviate or even neutralize potential biases (Sauquet, 2013).
478 Certainly, further studies are required to fully understand the effect of secondary calibrations
479 on outputs from different tree dating methods, and on downstream analyses. It is possible
480 that secondary calibrations can be safely used with dating methods that do not require
481 setting priors, such as penalized likelihood (Sanderson, 2003), with methods that do not
482 make any assumptions on the ages and fix them to a node on a tree topology, such as
483 BLADJ (Webb et al., 2008; Webb & Donoghue, 2005), or methods that summarize age data
484 unto a tree topology.

485 Our cross validation analysis might provide some insight in this regard. When ages are
486 estimated with secondary calibrations, nodes closer to the root do tend to be slightly
487 younger than ages estimated with primary calibrations. However, nodes closer to the tip
488 tend to be older when estimated using secondary calibrations with a dating method that
489 does not make any prior assumptions on the nature of the calibrations themselves
490 (Supplementary Figures S2-S20). The only ~~exepetion to tjis exception to this~~ was observed

491 on ~~results of the cross validation analysis~~ cross validation results of the Burns et al. (2014)
492 chronogram, which ~~results in displays~~ much younger node ages when estimated using
493 secondary calibrations (Supplementary Figs. S1, S5).

494 *Summarizing Chronograms*

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

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

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

514 The exercise of summarizing age data from across multiple studies provides the
515 opportunity to work with a ~~more inclusive chronogram~~, chronogram that reflects a unified

516 evolutionary history for a lineage, by putting together evidence from different hypotheses.
517 The largest, and taxonomically broadest chronogram currently available from OpenTree was
518 constructed summarizing age data from 2,274 published chronograms using NCBI's
519 taxonomic tree as backbone (Hedges et al., 2015). A summarizing exercise may also amplify
520 the effect of uncertainty and errors in source data, and blur parts of the evolutionary history
521 of a lineage that might only be reflected in source chronograms and lost on the summary
522 chronogram (Sauquet et al., 2021).

523 *Effects of Phylogenetic Sampling on Downstream Analyses*

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

527 ~~In ecology and conservation biology, incorporating at least some data on lineage~~
528 ~~divergence times represents a relevant improvement for testing alternative hypothesis using~~
529 ~~phylogenetic distance (Webb Analysis of species diversification using simulated and~~
530 ~~empirical phylogenies suggest that using a more completely sampled phylogeny provides~~
531 ~~estimates that are closer to the true diversification history than when analysing~~
532 ~~incompletely sampled phylogenies (Chang, Rabosky, & Alfaro, 2020; Cusimano, Stadler, &~~
533 ~~Renner, 2012; Sun et al., 2008–2020). Ideally, phylogenies should be completed using~~
534 ~~genetic data, but this is a difficult task to achieve for many biological groups.~~ Hence,
535 DateLife's workflow features different ways of ~~estimating assigning~~ node ages in the absence
536 of calibrations and branch length information for certain taxa. ~~branch lengths Completing a~~
537 ~~phylogeny using a stochastic birth-death politomy resolver and a backbone taxonomy~~ is a
538 common practice in scientific publications: Jetz et al. (2012), created a chronogram of all 9,
539 993 bird species, where 67% had molecular data and the rest was simulated; Rabosky et al.
540 (2018) created a chronogram of 31, 536 ray-finned fishes, of which only 37% had molecular
541 data; Smith and Brown (2018) constructed a chronogram of 353, 185 seed plants where only

542 23% had molecular data. These stochastically resolved chronograms provide less biased
543 diversification rates estimated with methods that account for sampling fractions (Chang et
544 al., 2020; Cusimano et al., 2012), but can also introduce spurious patterns of early bursts of
545 diversification (Cusimano & Renner, 2010; Sun et al., 2020).

546 Notably, Taxonomy based polity resolvers also introduce topological differences.
547 The study of macroevolutionary processes largely depends on an understanding of the
548 timing of species diversification events, and different phylogenetic and chronogram
549 hypothesis can provide very different overviews of the macroevolutionary history of a
550 biological group. For example, alternative topologies in chronograms from the same
551 biological group can infer very different species diversification patterns (Rabosky, 2015;
552 Title & Rabosky, 2016). Similarly there are worries that patterns of morphological
553 evolution cannot be accurately inferred with these types of phylogenies, as any patterns
554 would be erased by randomization (Rabosky, 2015). We note that the same applies for
555 geographical and morphological dependent diversification analysis. Hence, we suggest that
556 these types of randomly resolved phylogenies can be used as null or neutral models,
557 representing the case of diversification independent of traits and geographical scenario.

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

568

CONCLUSIONS

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

577 The DateLife project allows for easy and fast summary of public and state-of-the-art
578 data on time of lineage divergence. It provides a straightforward way to get an informed idea
579 on the state of knowledge of the time frame of evolution of different regions of the tree of life,
580 and allows ~~identification of identifying~~ regions that require more research, or that have
581 conflicting information. It is available as an R package, and as a web-based R shiny
582 application at www.datelife.org. Both summary and newly generated trees are useful to
583 evaluate evolutionary hypotheses in different areas of research. The DateLife project ~~helps~~
584 ~~with should improve~~ awareness of the existing variation in expert time of divergence data,
585 and ~~will~~ foster exploration of the effect of alternative divergence time hypothesis on the
586 results of analyses, nurturing a culture of more cautious interpretation of evolutionary results.

587

AVAILABILITY

588 The DateLife software is free and open source~~and it~~. It can be used online through its
589 R shiny web application hosted at <http://www.datelife.org>, and locally through the
590 `datelife` R package, ~~and through Phylotastic's project web portal~~ downloadable from
591 Zenodo (<https://doi.org/10.5281/zenodo.593938>) and the CRAN repository (Sanchez-Reyes
et al., 2022). DateLife's web application is maintained using RStudio's shiny server and the

593 shiny package open infrastructure, as well as Docker and OpenTree's infrastructure
594 (dates.opentreeoflife.org/datelife). datelife's ~~R package~~ stable version is available for
595 installation from ~~the CRAN repository ()~~ [CRAN repository](#) using the command
596 `install.packages(pkgs = "datelife")` from within R. Development versions are
597 available from ~~the DateLife's~~ GitHub repository (<https://github.com/phylotastic/datelife>)
598 and can be installed using the command
599 `devtools::install_github("phylotastic/datelife")`.

600 SUPPLEMENTARY MATERIAL

601 ~~Code used to generate all versions of this manuscript, the Supplementary material,~~
602 ~~including code, biological examples, benchmark results, data files and online-only~~
603 ~~appendices, can be found in the Dryad data repository~~
604 (<https://doi.org/10.5061/dryad.cnp5hqc6w>) as well as ~~the benchmark of functionalities are~~
605 ~~available at datelifeMS1, datelife_examples, and datelife_benchmark repositories in~~
606 ~~LCSR's GitHub account in the Zenodo stable repositories that host the reproducible~~
607 ~~manuscript~~ (<https://doi.org/10.5281/zenodo.7435094>), ~~the biological examples~~
608 (<https://doi.org/10.5281/zenodo.7435101>), ~~and the software benchmark~~
609 (<https://doi.org/10.5281/zenodo.7435106>). The corresponding development versions hosted
610 ~~on GitHub can be found at <https://github.com/LunaSare/datelifeMS1#readme>,~~
611 https://github.com/LunaSare/datelife_examples#readme, ~~and~~
612 https://github.com/LunaSare/datelife_benchmark#readme.

613 FUNDING

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

618

ACKNOWLEDGEMENTS

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

References

- 632
- 633 Agnarsson, I., & Miller, J. A. (2008). Is acctran better than deltran? *Cladistics*, 24(6),
634 1032–1038.
- 635 Alström, P., Hooper, D. M., Liu, Y., Olsson, U., Mohan, D., Gelang, M., ... Price, T. D.
636 (2014). Discovery of a relict lineage and monotypic family of passerine birds. *Biology
Letters*, 10(3), 20131067.
- 637
- 638 Ané, C., Eulenstein, O., Piaggio-Talice, R., & Sanderson, M. J. (2009). Groves of
639 phylogenetic trees. *Annals of Combinatorics*, 13(2), 139–167.
- 640 Antonelli, A., Hettling, H., Condamine, F. L., Vos, K., Nilsson, R. H., Sanderson, M. J., ...
641 Vos, R. A. (2017). Toward a self-updating platform for estimating rates of speciation
642 and migration, ages, and relationships of Taxa. *Systematic Biology*, 66(2), 153–166.
643 <https://doi.org/10.1093/sysbio/syw066>
- 644 Archie, J., Day, W. H., Felsenstein, J., Maddison, W., Meacham, C., Rohlf, F. J., &
645 Swofford, D. (1986). The Newick tree format. Retrieved from
646 {<https://evolution.genetics.washington.edu/phylip/newicktree.html>}
- 647 Avibase. (2022). Yellow-throated Bunting. *Avibase - the World Bird Database*, (Online
648 Resource). Retrieved from
649 {<https://avibase.bsc-eoc.org/species.jsp?lang=EN&avibaseid=82D1EE0049D8D927>}
- 650 Bapst, D. W. (2012). Paleotree: An R package for paleontological and phylogenetic analyses
651 of evolution. *Methods in Ecology and Evolution*, 3(5), 803–807.
652 <https://doi.org/10.1111/j.2041-210X.2012.00223.x>
- 653 Barba-Montoya, J., Reis, M. dos, Schneider, H., Donoghue, P. C., & Yang, Z. (2018).
654 Constraining uncertainty in the timescale of angiosperm evolution and the veracity of

- 655 a cretaceous terrestrial revolution. *New Phytologist*, 218(2), 819–834.
- 656 Barker, F. K. (2014). Mitogenomic data resolve basal relationships among passeriform and
657 passeridan birds. *Molecular Phylogenetics and Evolution*, 79, 313–324.
- 658 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2013). Going to
659 extremes: Contrasting rates of diversification in a recent radiation of new world
660 passerine birds. *Systematic Biology*, 62(2), 298–320.
- 661 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2015). New insights
662 into new world biogeography: An integrated view from the phylogeny of blackbirds,
663 cardinals, sparrows, tanagers, warblers, and allies. *The Auk: Ornithological Advances*,
664 132(2), 333–348.
- 665 Barker, F. K., Cibois, A., Schikler, P., Feinstein, J., & Cracraft, J. (2004). Phylogeny and
666 diversification of the largest avian radiation. *Proceedings of the National Academy of
667 Sciences*, 101(30), 11040–11045.
- 668 Beresford, P., Barker, F., Ryan, P., & Crowe, T. (2005). African endemics span the tree of
669 songbirds (passeri): Molecular systematics of several evolutionary “enigmas”.
670 *Proceedings of the Royal Society B: Biological Sciences*, 272(1565), 849–858.
- 671 [Bodenhofer, U., Bonatesta, E., Horejš-Kainrath, C., & Hochreiter, S. \(2015\). Msa: An r
672 package for multiple sequence alignment.](#) *Bioinformatics*, 31(24), 3997–3999.
- 673 Boyle, B., Hopkins, N., Lu, Z., Raygoza Garay, J. A., Mozzherin, D., Rees, T., ... Enquist,
674 B. J. (2013). The taxonomic name resolution service: An online tool for automated
675 standardization of plant names. *BMC Bioinformatics*, 14(1).
676 <https://doi.org/10.1186/1471-2105-14-16>
- 677 Britton, T., Anderson, C. L., Jacquet, D., Lundqvist, S., & Bremer, K. (2007). Estimating

- 678 Divergence Times in Large Phylogenetic Trees. *Systematic Biology*, 56(788777878),
679 741–752. <https://doi.org/10.1080/10635150701613783>
- 680 Bryson Jr, R. W., Chaves, J., Smith, B. T., Miller, M. J., Winker, K., Pérez-Emán, J. L., &
681 Klicka, J. (2014). Diversification across the new world within the ‘blue’cardinalids
682 (aves: Cardinalidae). *Journal of Biogeography*, 41(3), 587–599.
- 683 Burleigh, J. G., Kimball, R. T., & Braun, E. L. (2015). Building the avian tree of life using a
684 large-scale, sparse supermatrix. *Molecular Phylogenetics and Evolution*, 84, 53–63.
- 685 Burns, K. J., Shultz, A. J., Title, P. O., Mason, N. A., Barker, F. K., Klicka, J., ... Lovette,
686 I. J. (2014). Phylogenetics and diversification of tanagers (passeriformes:
687 Thraupidae), the largest radiation of neotropical songbirds. *Molecular Phylogenetics
688 and Evolution*, 75, 41–77.
- 689 Chamberlain, S. (2018). *bold: Interface to Bold Systems API*. Retrieved from
690 <https://CRAN.R-project.org/package=bold>
- 691 Chamberlain, S. A., & Szöcs, E. (2013). taxize : taxonomic search and retrieval in R [version
692 2; referees: 3 approved]. *F1000Research*, 2(191), 1–29.
693 <https://doi.org/10.12688/f1000research.2-191.v2>
- 694 Chang, J., Rabosky, D. L., & Alfaro, M. E. (2020). Estimating diversification rates on
695 incompletely sampled phylogenies: Theoretical concerns and practical solutions.
696 *Systematic Biology*, 69(3), 602–611.
- 697 Chaves, J. A., Hidalgo, J. R., & Klicka, J. (2013). Biogeography and evolutionary history of
698 the neotropical genus Saltator (aves: Thraupini). *Journal of Biogeography*, 40(11),
699 2180–2190.
- 700 Claramunt, S., & Cracraft, J. (2015). A new time tree reveals earth history’s imprint on the

- 701 evolution of modern birds. *Science Advances*, 1(11), e1501005.
- 702 Criscuolo, A., Berry, V., Douzery, E. J., & Gascuel, O. (2006). SDM: A fast distance-based
703 approach for (super)tree building in phylogenomics. *Systematic Biology*, 55(5),
704 740–755. <https://doi.org/10.1080/10635150600969872>
- 705 [Cusimano, N., & Renner, S. S. \(2010\). Slowdowns in diversification rates from real](#)
706 [phylogenies may not be real. *Systematic Biology*, 59\(4\), 458–464.](#)
- 707 [Cusimano, N., Stadler, T., & Renner, S. S. \(2012\). A new method for handling missing](#)
708 [species in diversification analysis applicable to randomly or nonrandomly sampled](#)
709 [phylogenies. *Systematic Biology*, 61\(5\), 785–792.](#)
- 710 Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M.-K., Turon, X., ... Douzery,
711 E. J. (2018). A phylogenomic framework and timescale for comparative studies of
712 tunicates. *BMC Biology*, 16(1), 1–14.
- 713 Eastman, J. M., Harmon, L. J., & Tank, D. C. (2013). Congruification: Support for time
714 scaling large phylogenetic trees. *Methods in Ecology and Evolution*, 4(7), 688–691.
715 <https://doi.org/10.1111/2041-210X.12051>
- 716 Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high
717 throughput. *Nucleic Acids Research*, 32(5), 1792–1797.
- 718 Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American Naturalist*,
719 125(1), 1–15. Retrieved from <http://www.jstor.org/stable/2461605>
- 720 [Forest, F., Savolainen, V., Chase, M. W., Lupia, R., Bruneau, A., & Crane, P. R. \(2005\).](#)
721 [Teasing apart molecular-versus fossil-based error estimates when dating](#)
722 [phylogenetic trees: A case study in the birch family \(betulaceae\). *Systematic*](#)
723 [*Botany*, 30\(1\), 118–133.](#)

- 724 Freckleton, R. P., Harvey, P. H., & Pagel, M. (2002). Phylogenetic analysis and
725 comparative data: A test and review of evidence. *The American Naturalist*.
- 726 GBIF Secretariat. (2022). GBIF Backbone Taxonomy. *Checklist dataset*, (Online Resource
727 accessed via GBIF.org). Retrieved from {<https://doi.org/10.15468/39omei> }
- 728 Gibb, G. C., England, R., Hartig, G., McLenaghan, P. A., Taylor Smith, B. L., McComish,
729 B. J., ... Penny, D. (2015). New zealand passerines help clarify the diversification of
730 major songbird lineages during the oligocene. *Genome Biology and Evolution*, 7(11),
731 2983–2995.
- 732 Graur, D., & Martin, W. (2004). Reading the entrails of chickens: Molecular timescales of
733 evolution and the illusion of precision. *TRENDS in Genetics*, 20(2), 80–86.
- 734 Hackett, S. J., Kimball, R. T., Reddy, S., Bowie, R. C., Braun, E. L., Braun, M. J., ...
735 others. (2008). A phylogenomic study of birds reveals their evolutionary history.
736 *Science*, 320(5884), 1763–1768.
- 737 Harvey, P. H., Pagel, M. D., & others. (1991). *The comparative method in evolutionary
738 biology* (Vol. 239). Oxford university press Oxford.
- 739 Hedges, S. B., Marin, J., Suleski, M., Paymer, M., & Kumar, S. (2015). Tree of life reveals
740 clock-like speciation and diversification. *Molecular Biology and Evolution*, 32(4),
741 835–845. <https://doi.org/10.1093/molbev/msv037>
- 742 Heibl, C. (2008). *PHYLOCH: R language tree plotting tools and interfaces to diverse
743 phylogenetic software packages*. Retrieved from
744 <http://www.christophheibl.de/Rpackages.html>
- 745 Hooper, D. M., & Price, T. D. (2017). Chromosomal inversion differences correlate with
746 range overlap in passerine birds. *Nature Ecology & Evolution*, 1(10), 1526.

- 747 Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic
748 trees. *Bioinformatics*, 17(8), 754–755.
749 <https://doi.org/10.1093/bioinformatics/17.8.754>
- 750 Jetz, W., Thomas, G., Joy, J. J., Hartmann, K., & Mooers, A. (2012). The global diversity
751 of birds in space and time. *Nature*, 491(7424), 444–448.
752 <https://doi.org/10.1038/nature11631>
- 753 Johansson, U. S., Fjeldså, J., & Bowie, R. C. (2008). Phylogenetic relationships within
754 passerida (aves: Passeriformes): A review and a new molecular phylogeny based on
755 three nuclear intron markers. *Molecular Phylogenetics and Evolution*, 48(3), 858–876.
- 756 Katoh, K., Asimenos, G., & Toh, H. (2009). Multiple alignment of dna sequences with mafft.
757 In *Bioinformatics for dna sequence analysis* (pp. 39–64). Springer.
- 758 Kimball, R. T., Oliveros, C. H., Wang, N., White, N. D., Barker, F. K., Field, D. J., ...
759 others. (2019). A phylogenomic supertree of birds. *Diversity*, 11(7), 109.
- 760 Klicka, J., Barker, F. K., Burns, K. J., Lanyon, S. M., Lovette, I. J., Chaves, J. A., & Bryson
761 Jr, R. W. (2014). A comprehensive multilocus assessment of sparrow (aves:
762 Passerellidae) relationships. *Molecular Phylogenetics and Evolution*, 77, 177–182.
- 763 Ksepka, D. T., Parham, J. F., Allman, J. F., Benton, M. J., Carrano, M. T., Cranston, K.
764 A., ... others. (2015). The fossil calibration database—a new resource for divergence
765 dating. *Systematic Biology*, 64(5), 853–859.
- 766 Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio,
767 A., ... others. (2015). Evolution of darwin's finches and their beaks revealed by
768 genome sequencing. *Nature*, 518(7539), 371–375.
- 769 Laubichler, M. D., & Maienschein, J. (2009). *Form and function in developmental evolution*.

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

- 793 the Open Tree of Life data. *Methods in Ecology and Evolution*, 7(12), 1476–1481.
- 794 <https://doi.org/10.1111/2041-210X.12593>
- 795 Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecology Letters*,
796 17(4), 508–525. <https://doi.org/10.1111/ele.12251>
- 797 Moyle, R. G., Oliveros, C. H., Andersen, M. J., Hosner, P. A., Benz, B. W., Manthey, J. D.,
798 ... Faircloth, B. C. (2016). Tectonic collision and uplift of Wallacea triggered the
799 global songbird radiation. *Nature Communications*, 7(1), 1–7.
- 800 Oliveros, C. H., Field, D. J., Ksepka, D. T., Barker, F. K., Aleixo, A., Andersen, M. J., ...
801 others. (2019). Earth history and the passerine superradiation. *Proceedings of the
802 National Academy of Sciences*, 116(16), 7916–7925.
- 803 Ooms, J., & Chamberlain, S. (2018). *Phylocomr: Interface to 'phylocom'*. Retrieved from
804 <https://CRAN.R-project.org/package=phylocomr>
- 805 Open Tree Of Life, Redelings, B., Cranston, K. A., Allman, J., Holder, M. T., & McTavish,
806 E. J. (2016). Open Tree of Life APIs v3.0. *Open Tree of Life Project*, (Online
807 Resources). Retrieved from
808 {<https://github.com/OpenTreeOfLife/germinator/wiki/Open-Tree-of-Life-Web->
809 APIs}
- 810 Open Tree Of Life, Redelings, B., Sánchez Reyes, L. L., Cranston, K. A., Allman, J., Holder,
811 M. T., & McTavish, E. J. (2019). Open tree of life synthetic tree v12.3. *Zenodo*.
812 Retrieved from <https://doi.org/10.5281/zenodo.3937742>
- 813 Ödeen, A., Håstad, O., & Alström, P. (2011). Evolution of ultraviolet vision in the largest
814 avian radiation—the passerines. *BMC Evolutionary Biology*, 11(1), 1–8.
- 815 Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and

- 816 evolution in R language. *Bioinformatics*, 20(2), 289–290.
- 817 Parchman, T. L., Benkman, C. W., & Mezquida, E. T. (2007). Coevolution between
818 hispaniolan crossbills and pine: Does more time allow for greater phenotypic
819 escalation at lower latitude? *Evolution*, 61(9), 2142–2153.
- 820 Päckert, M., Martens, J., Sun, Y.-H., Severinghaus, L. L., Nazarenko, A. A., Ting, J., ...
821 Tietze, D. T. (2012). Horizontal and elevational phylogeographic patterns of
822 himalayan and southeast asian forest passerines (aves: Passeriformes). *Journal of
823 Biogeography*, 39(3), 556–573.
- 824 Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., FitzJohn, R. G.,
825 ... Harmon, L. J. (2014). Geiger v2. 0: An expanded suite of methods for fitting
826 macroevolutionary models to phylogenetic trees. *Bioinformatics*, 30(15), 2216–2218.
- 827 Posadas, P., Crisci, J. V., & Katinas, L. (2006). Historical biogeography: A review of its
828 basic concepts and critical issues. *Journal of Arid Environments*, 66(3), 389–403.
- 829 Powell, A. F., Barker, F. K., Lanyon, S. M., Burns, K. J., Klicka, J., & Lovette, I. J. (2014).
830 A comprehensive species-level molecular phylogeny of the new world blackbirds
831 (icteridae). *Molecular Phylogenetics and Evolution*, 71, 94–112.
- 832 Price, T. D., Hooper, D. M., Buchanan, C. D., Johansson, U. S., Tietze, D. T., Alström, P.,
833 ... others. (2014). Niche filling slows the diversification of himalayan songbirds.
834 *Nature*, 509(7499), 222.
- 835 Pulgarín-R, P. C., Smith, B. T., Bryson Jr, R. W., Spellman, G. M., & Klicka, J. (2013).
836 Multilocus phylogeny and biogeography of the new world pheucticus grosbeaks (aves:
837 Cardinalidae). *Molecular Phylogenetics and Evolution*, 69(3), 1222–1227.
- 838 [Rabosky, D. L. \(2015\). No substitute for real data: A cautionary note on the use of](#)

839 phylogenies from birth-death polytomy resolvers for downstream comparative
840 analyses. *Evolution*, 69(12), 3207–3216.

841 Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., ... others.
842 (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*,
843 559(7714), 392.

844 Ramshaw, J., Richardson, D., Meatyard, B., Brown, R., Richardson, M., Thompson, E., &
845 Boulter, D. (1972). The time of origin of the flowering plants determined by using
846 amino acid sequence data of cytochrome c. *New Phytologist*, 71(5), 773–779.

847 Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The barcode of life data system
848 (<http://www.Barcodinglife.Org>). *Molecular Ecology Notes*, 7(3), 355–364.

849 R Core Team. (2018). *R: a language and environment for statistical computing*. Vienna,
850 Austria: R Foundation for Statistical Computing.

851 Rees, & Cranston, K. (2017). Automated assembly of a reference taxonomy for phylogenetic
852 data synthesis. *Biodiversity Data Journal*, (5).

853 Rees, Vandepitte, L., Decock, W., & Vanhoorne, B. (2017). IRMNG 2006–2016: 10 Years of
854 a Global Taxonomic Database. *Biodiversity Informatics*, 12.

855 Revell, L. J. (2012). Phytools: An r package for phylogenetic comparative biology (and other
856 things). *Methods in Ecology and Evolution*, 3, 217–223.

857 Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference
858 under mixed models. *Bioinformatics*, 19(12), 1572–1574.
859 <https://doi.org/10.1093/bioinformatics/btg180>

860 Roquet, C., Lavergne, S., & Thuiller, W. (2014). One tree to link them all: A phylogenetic
861 dataset for the european tetrapoda. *PLoS Currents*, 6.

- 862 Sanchez-Reyes, L. L., & O'Meara, B. (2022). *datelifeplot*: Methods to plot chronograms
863 and outputs of the *datelife* package. *R Package Release V0.2.2*. Retrieved from
864 <https://zenodo.org/badge/latestdoi/381501451>
- 865 Sanchez-Reyes, L. L., O'Meara, B., Eastman, J., Heath, T., Wright, A., Schliep, K., ...
866 Alfaro, M. (2022). *datelife*: Scientific Data on Time of Lineage Divergence for Your
867 Taxa. In *R package version 0.6.6*. Retrieved from
868 <https://CRAN.R-project.org/package=datelife> and
869 <https://doi.org/10.5281/zenodo.593938>
- 870 Sanderson, M. (2002). Estimating Absolute Rates of Molecular Evolution and Divergence
871 Times: A Penalized Likelihood Approach. *Molecular Biology and Evolution*, 19(1),
872 101–109. <https://doi.org/10.1093/oxfordjournals.molbev.a003974>
- 873 Sanderson, M. (2003). r8s: Inferring Absolute Rates of Molecular Evolution and Divergence
874 Times in the Absence of a Molecular Clock. *Bioinformatics*, 19(2), 301–302.
- 875 Sanderson, M., & Doyle, J. (2001). Sources of error and confidence intervals in estimating
876 the age of angiosperms from *rbcL* and 18S rDNA data. *American Journal of Botany*,
877 88(8), 1499–1516.
- 878 Sauquet, H. (2013). A practical guide to molecular dating. *Comptes Rendus Palevol*, 12(6),
879 355–367.
- 880 Sauquet, H., Ramírez-Barahona, S., & Magallón, S. (2021). *The age of flowering plants is
881 unknown*.
- 882 Schenk, J. J. (2016). Consequences of secondary calibrations on divergence time estimates.
883 *PLoS ONE*, 11(1). <https://doi.org/10.1371/journal.pone.0148228>
- 884 Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in r. *Bioinformatics*, 27(4), 592–593.

- 885 Schoch, C. L., Ciufo, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R., ...
886 others. (2020). NCBI Taxonomy: a Comprehensive Update on Curation, Resources
887 and Tools. *Database, 2020*.
- 888 Selvatti, A. P., Gonzaga, L. P., & Moraes Russo, C. A. de. (2015). A paleogene origin for
889 crown passerines and the diversification of the oscines in the new world. *Molecular
890 Phylogenetics and Evolution, 88*, 1–15.
- 891 Smith, S., & Brown, J. (2018). Constructing a broadly inclusive seed plant phylogeny.
892 *American Journal of Botany, 105*(3), 302–314.
- 893 Smith, S., & O'Meara, B. (2012). TreePL: Divergence time estimation using penalized
894 likelihood for large phylogenies. *Bioinformatics, 28*(20), 2689–2690.
895 <https://doi.org/10.1093/bioinformatics/bts492>
- 896 Stoltzfus, A., Lapp, H., Matasci, N., Deus, H., Sidlauskas, B., Zmasek, C. M., ... Jordan, G.
897 (2013). Phylotastic! Making tree-of-life knowledge accessible, reusable and convenient.
898 *BMC Bioinformatics, 14*. <https://doi.org/10.1186/1471-2105-14-158>
- 899 Sun, M., Folk, R. A., Gitzendanner, M. A., Soltis, P. S., Chen, Z., Soltis, D. E., &
900 Guralnick, R. P. (2020). Estimating rates and patterns of diversification with
901 incomplete sampling: A case study in the rosids. American Journal of Botany,
902 107(6), 895–909.
- 903 Tietze, D. T., Päckert, M., Martens, J., Lehmann, H., & Sun, Y.-H. (2013). Complete
904 phylogeny and historical biogeography of true rosefinches (aves: Carpodacus).
905 *Zoological Journal of the Linnean Society, 169*(1), 215–234.
- 906 Title, P. O., & Rabosky, D. L. (2016). Do Macrophylogenies Yield Stable Macroevolutionary
907 Inferences? An Example from Squamate Reptiles. *Systematic Biology, syw102.*
908 <https://doi.org/10.1093/sysbio/syw102>

- 909 Treplin, S., Siegert, R., Bleidorn, C., Thompson, H. S., Fotso, R., & Tiedemann, R. (2008).
910 Molecular phylogeny of songbirds (aves: Passeriformes) and the relative utility of
911 common nuclear marker loci. *Cladistics*, 24(3), 328–349.
- 912 Uyeda, J. C., Pennell, M. W., Miller, E. T., Maia, R., & McClain, C. R. (2017). The
913 evolution of energetic scaling across the vertebrate tree of life. *The American*
914 *Naturalist*, 190(2), 185–199.
- 915 Vos, R. A., Balhoff, J. P., Caravas, J. A., Holder, M. T., Lapp, H., Maddison, W. P., ...
916 others. (2012). NeXML: Rich, extensible, and verifiable representation of
917 comparative data and metadata. *Systematic Biology*, 61(4), 675–689.
918 <https://doi.org/10.1093/sysbio/sys025>
- 919 Webb, C. (2000). Exploring the Phylogenetic Structure of Ecological Communities : An
920 Example for Rain Forest Trees. *The American Naturalist*, 156(2), 145–155.
- 921 Webb, C., Ackerly, D., & Kembel, S. (2008). Phylocom: Software for the analysis of
922 phylogenetic community structure and trait evolution. *Bioinformatics*, 24(18),
923 2098–2100. <https://doi.org/10.1093/bioinformatics/btn358>
- 924 Webb, C., & Donoghue, M. (2005). Phylomatic: Tree assembly for applied phylogenetics.
925 *Molecular Ecology Notes*, 5(1), 181–183.
- 926 Weir, J., & Schluter, D. (2008). Calibrating the avian molecular clock. *Molecular Ecology*,
927 17(10), 2321–2328.
- 928 Zuccon, D., Prŷs-Jones, R., Rasmussen, P. C., & Ericson, P. G. (2012). The phylogenetic
929 relationships and generic limits of finches (fringillidae). *Molecular Phylogenetics and*
930 *Evolution*, 62(2), 581–596.

931 TABLE 1. Ages of congruified nodes. See Figure 3, step C2.

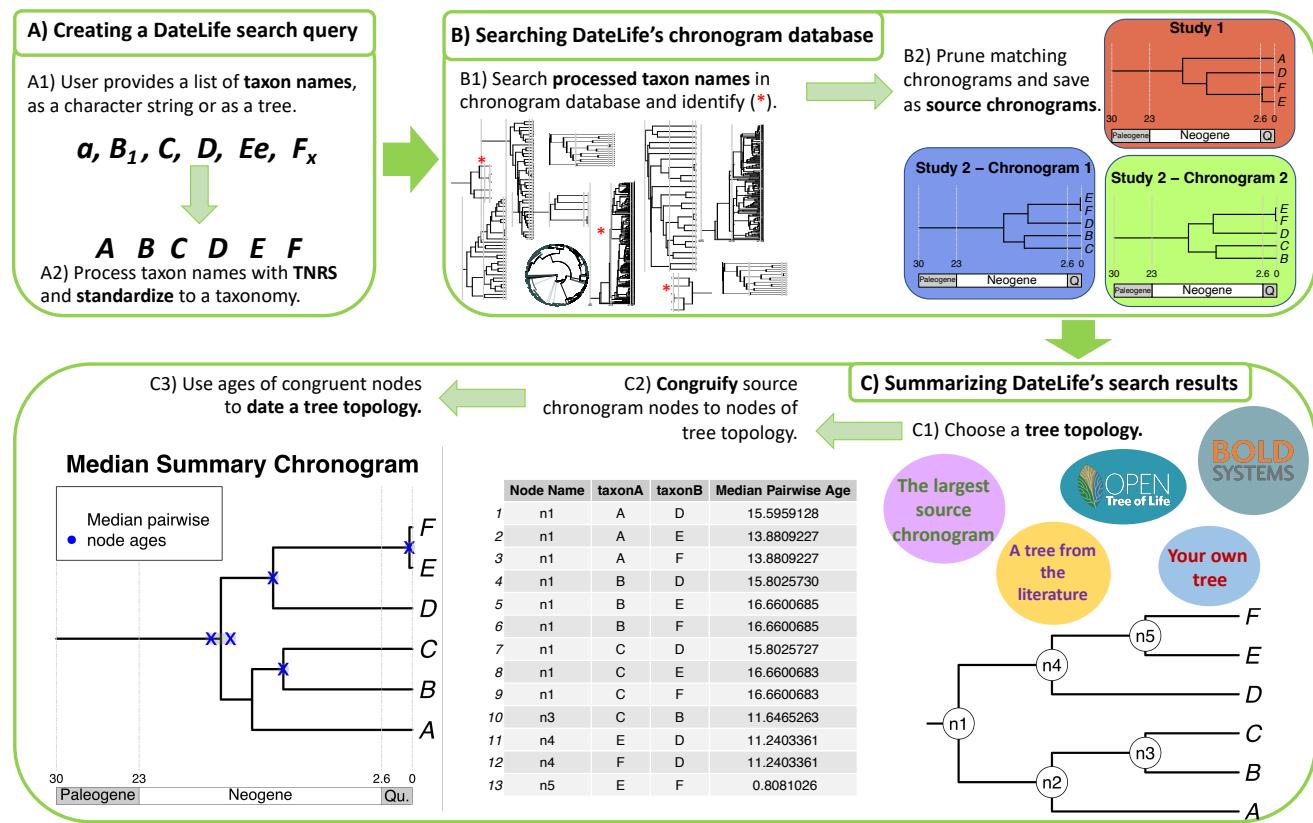
	Node Name	taxon A	taxon B	Node Age	Study chronogram
1	n1	<i>Emberiza citrinella</i>	<i>Platyspiza crassirostris</i>	9.6509881	Jetz et al. 2012 – chronogram 1
2	n1	<i>Pheucticus tibialis</i>	<i>Schoeniclus elegans</i>	14.3336520	Barker et al. 2015 – chronogram 1
3	n1	<i>Rhodothraupis celaeno</i>	<i>Schoeniclus 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

932

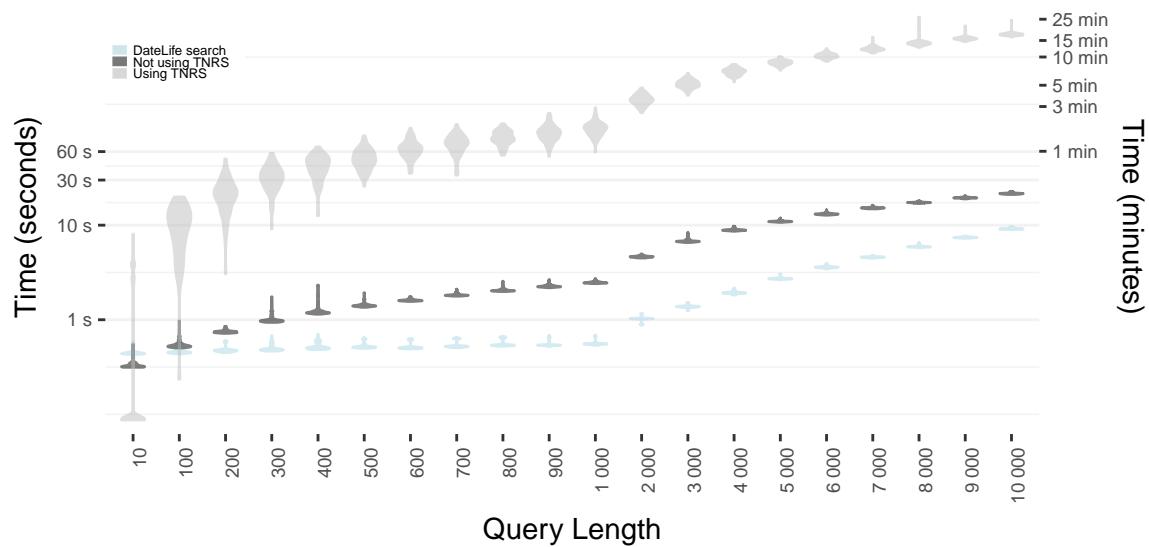
933 TABLE 2. Summary of congruified nodes ages. See Figure 3, step C3.

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

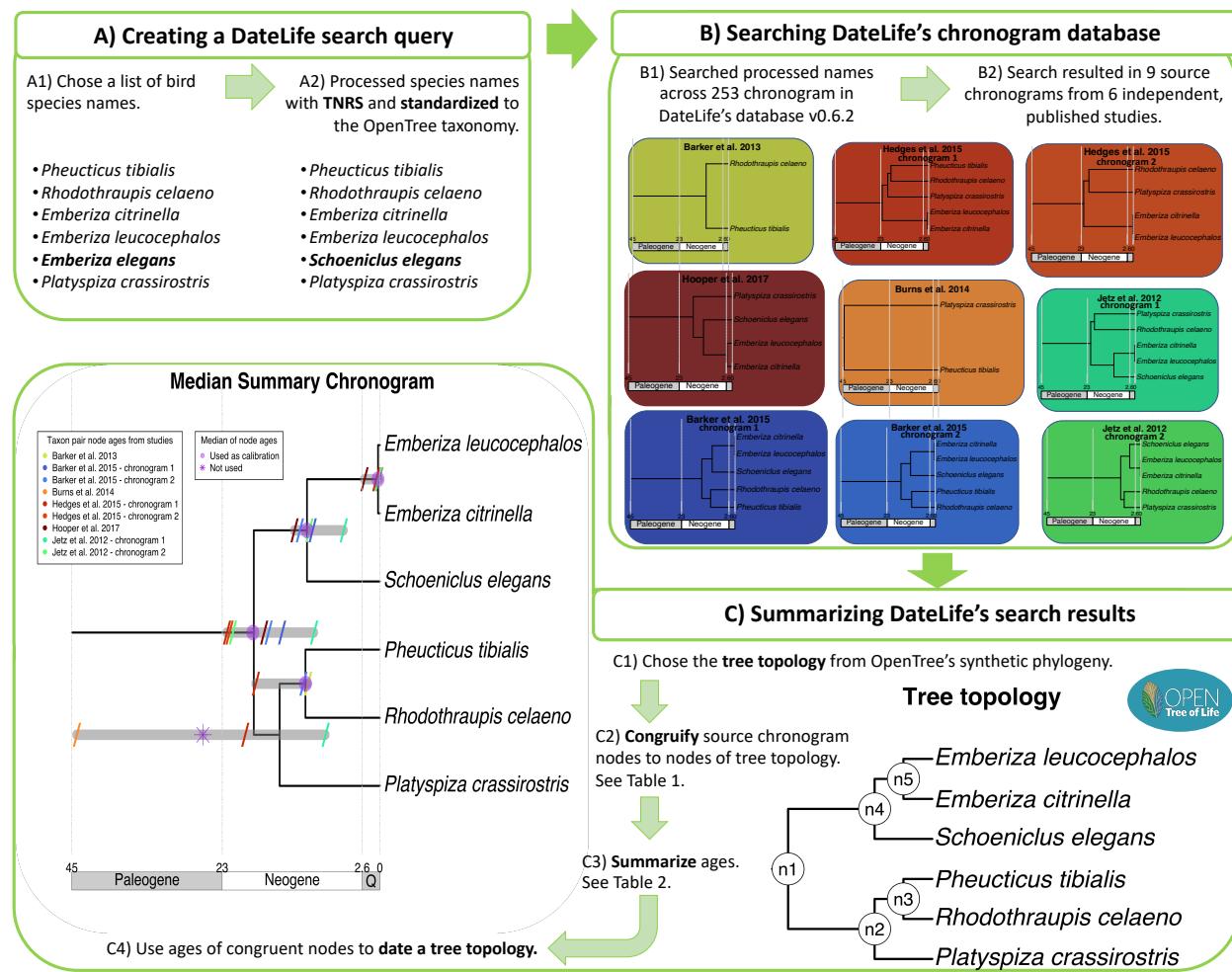
934



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



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



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

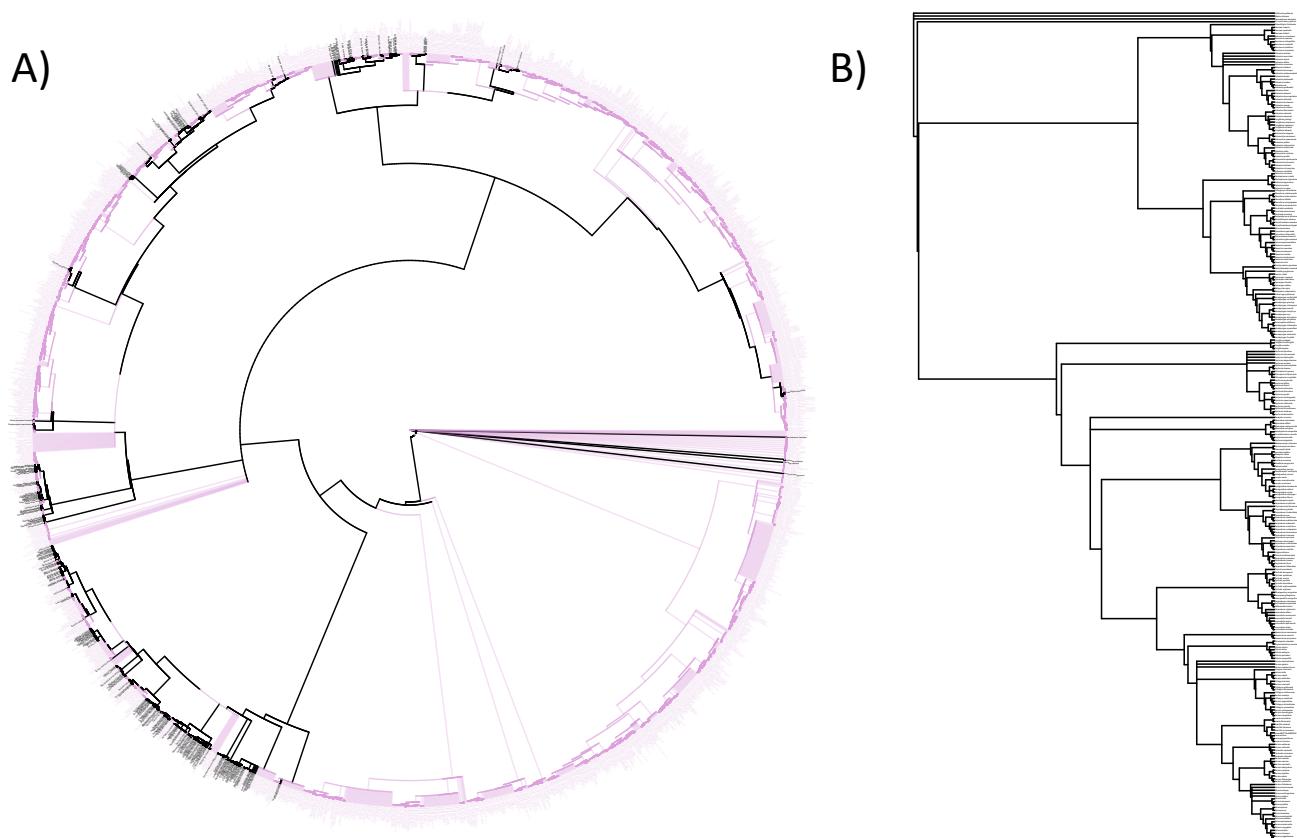


FIGURE 4. Tree topologies extracted from Open Tree of Life's (OpenTree) synthetic phylogenetic tree.
 A) Topology of 2,333 tips and 1,305 internal nodes, encompassing bird species within the family Fringillidae following the NCBI taxonomy (black), as well as all other bird species that share the same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

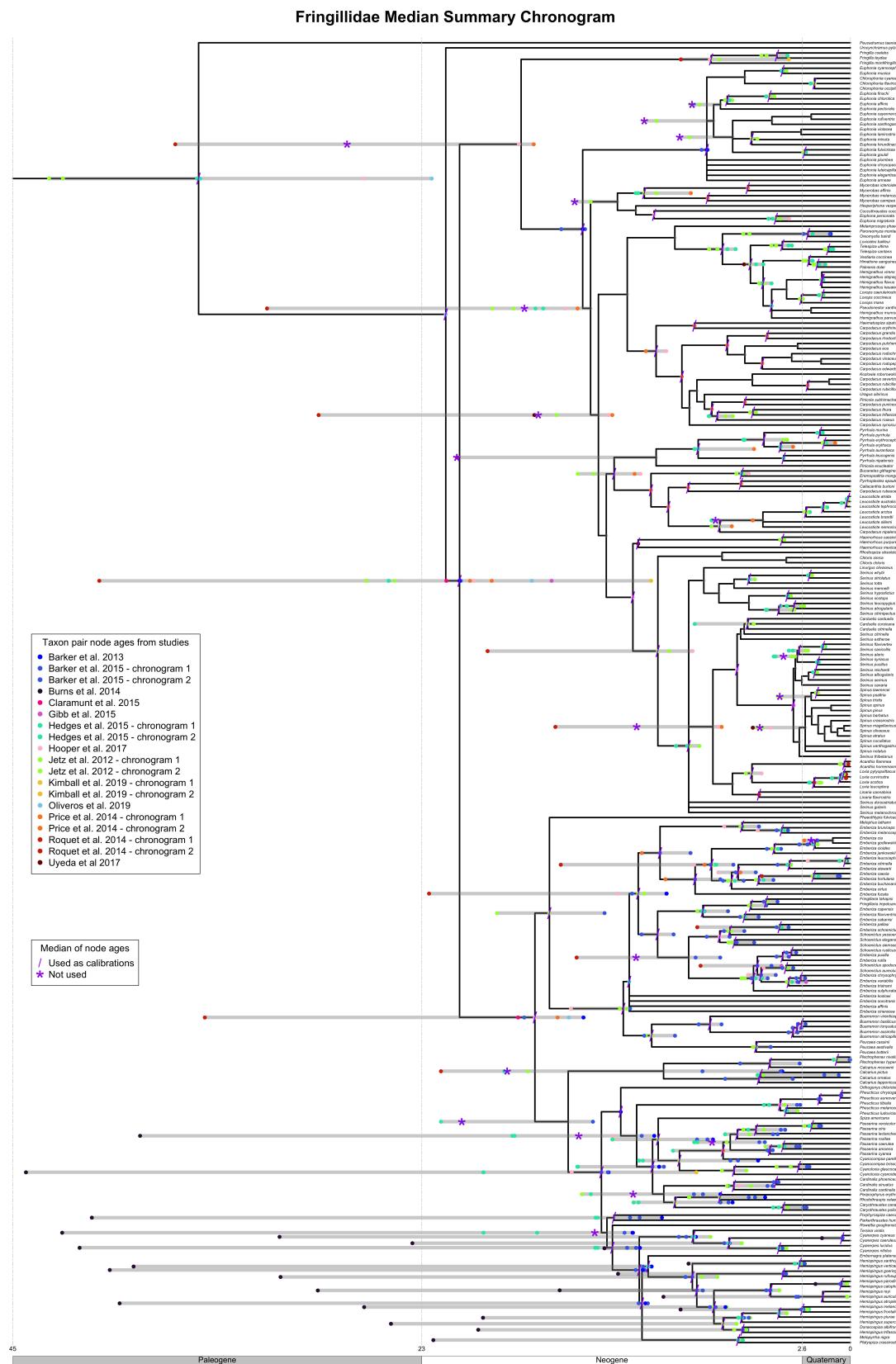
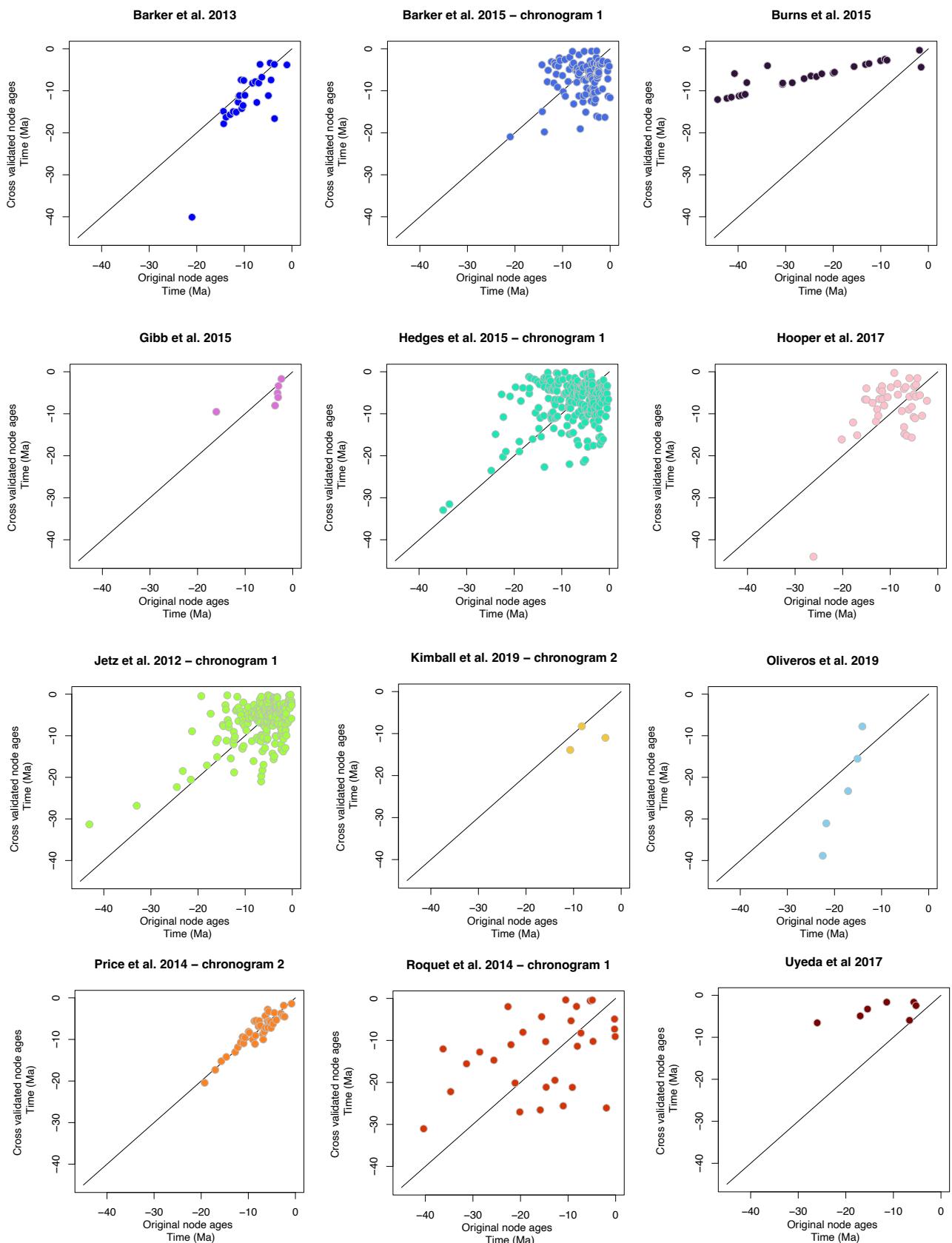


FIGURE 5. *Fringillidae* median summary chronogram generated with *DateLife*. It has 289 tips and 253 nodes, from which 212 have age data from at least one published chronogram. These nodes have



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