

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 Chronograms –phylogenies with branch lengths proportional to time– represent key data on
19 evolutionary time frame for the study of natural processes in many areas of biological
20 research. Chronograms also provide valuable information that can be used for education,
21 science communication, and policy decisions. Yet, achieving a high-quality reconstruction of
22 a chronogram is a difficult and resource-consuming task. Here we present DateLife, a service
23 implemented as an R package and an R Shiny web application available at www.datelife.org,
24 that provides services for efficient and easy discovery, summary, reuse, and reanalysis of node
25 age data mined from a curated database of expert, peer-reviewed, and openly available
26 chronograms. The main DateLife workflow starts with one or more scientific taxon names
27 provided by a user. Names are processed and standardized to a unified taxonomy. Then,
28 DateLife runs a name match across its local chronogram database that is curated from Open
29 Tree of Life’s phylogenetic repository, and extracts all chronograms that contain at least two
30 queried taxon names, along with their metadata. Finally, node ages from matching
31 chronograms are mapped upon a chosen tree topology using the congruification algorithm.
32 Congruified node ages are used as secondary calibrations to date the chosen topology, with
33 or without initial branch lengths, using different phylogenetic dating methods such as
34 BLADJ, treePL, PATHd8 and MrBayes. We performed a cross-validation test to compare
35 node ages resulting from a DateLife analysis to those from the original chronograms (i.e.,
36 obtained with primary calibrations), and found that DateLife’s node age estimates appear
37 generally correlated to those from the original chronograms, with the largest variation in
38 ages occurring around topologically deeper nodes. Software such as DateLife can help
39 increase awareness of the existing variation in alternative hypothesis of evolutionary time
40 frame for the same organisms, and can support exploration of the effect of alternative
41 chronogram hypotheses on downstream analyses, providing a framework for a more informed
42 interpretation of evolutionary results.

⁴³ *Keywords:* Tree; Phylogeny; Scaling; Dating; Ages; Divergence times; Open Science;

⁴⁴ Congruification; Supertree; Calibrations; Secondary calibrations

⁴⁵ Word count: 5605

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

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-a~~
55 homology hypothesis to construct a phylogeny, selecting and placing appropriate calibrations
56 on the phylogeny using independent age data points from the fossil record or other dated
57 events, and inferring ~~the-a~~ full dated tree; ~~it also generally requires~~. All of this entails
58 specialized biological training, taxonomic domain knowledge, and a non-negligible amount of
59 research time, computational resources and funding.

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

71 Description

72 DateLife's core software ~~application consists of applications are implemented in~~ the R
73 package `datelife`. Its current stable version – v0.6.6, is available from The Comprehensive
74 R Archive Network (CRAN) repository (Sanchez-Reyes et al., 2022), and relies on
75 functionalities from ~~various other~~ biological R packages: ape (Paradis, Claude, & Strimmer,
76 2004), bold (Chamberlain, 2018), geiger (Pennell et al., 2014), `msa` ([Bodenhofer, Bonatesta, Horejš-Kainrath, & Hochreiter, 2015](#)), paleotree (Bapst, 2012), phyloch (Heibl, 2008),
77 `phylocomr` (Ooms & Chamberlain, 2018), `phytools` (Revell, 2012), `rotl` (Michonneau, Brown,
78 & Winter, 2016), and `taxize` (Chamberlain, 2018; Chamberlain & Szöcs, 2013). Figure 1
79 provides a graphical summary of the three main steps of the DateLife workflow: creating a
80 search query, searching a database, and summarizing results from the search.

82 Creating a search query

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

89 DateLife ~~accepts scientific names that can belong to any inclusive taxonomic group~~
90 ~~(e.g., genus, family, tribe, etc.) or a binomial species name. Subspecies and variants are~~
91 ~~ignored. If an input taxon name belongs to an inclusive taxonomic group, DateLife has two~~
92 ~~alternative behaviors defined by~~ processes input scientific names using a Taxonomic Name
93 Resolution Service (TNRS), which increases the probability of correctly finding the queried
94 taxon names in the chronogram database. TNRS detects, corrects and standardizes name
95 misspellings and typos, variant spellings and authorities, and nomenclatural synonyms to a
96 single taxonomic standard (Boyle et al., 2013). TNRS also allows to correctly choose

97 between homonyms, by considering other taxa provided as input to infer the taxonomic
98 context of the flag. If the flag is active, DateLife retrieves all species names within the
99 inclusive taxonomic group following a standard taxonomy of choice, and adds them to the
100 input string. Taxonomies currently supported by DateLife are homonym. DateLife
101 implements TNRS using the Open Tree of Life (OpenTree) unified Taxonomy (OTT, Open
102 Tree Of Life et al., 2016; Rees & Cranston, 2017), as standard, storing taxonomic
103 identification numbers (OTT ids) for further processing and analysis. Other taxonomies
104 currently supported by DateLife are the National Center of Biotechnology Information
105 (NCBI) taxonomic database (Schoch et al., 2020), the Global Biodiversity Information
106 Facility (GBIF) taxonomic backbone (GBIF Secretariat, 2022), and the Interim Register of
107 Marine and Nonmarine Genera (IRMNG) database (Rees et al., 2017). If the
108 flag is

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

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

123 numbers for further processing.

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

130 Searching a chronogram database

131 At the time of writing of this manuscript (Jun 22, 2022Feb 20, 2023), DateLife's
132 chronogram database latest version consist of 253 chronograms published in 187 different
133 studies. It is curated from OpenTree's phylogenetic database, the Phylesystem, which
134 constitutes an open source of expert and peer-reviewed phylogenetic knowledge with rich
135 metadata (McTavish et al., 2015), which allows automatic and reproducible assembly of our
136 chronogram database. Datelife's chronogram database is navigable as an R data object
137 within the `datelife` R package.

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

¹⁴⁸ chronograms to ~~the user's~~ their DateLife analysis before an official database update is
¹⁴⁹ 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.***

175 **Choosing a topology**

176 DateLife requires a tree topology to summarize age data upon. We recommend that

177 users provide as input a tree topology from the literature, or one of their own making. If no

178 topology is provided, DateLife automatically extracts one from the OpenTree synthetic tree,

179 a phylogeny currently encompassing 2.3 million taxa across all life, assembled from 1, 239

180 published phylogenetic trees and OpenTree's unified Taxonomy, OTT (Open Tree Of Life et

181 al., 2019). Alternatively, DateLife can combine topologies from source chronograms using a

182 supertree approach ~~. To combine topologies from source chronograms into a single summary~~

183 ~~(or supertree) topology, the DateLife workflow~~ (Criscuolo, Berry, Douzery, & Gascuel, 2006).

184 To do this, DateLife first identifies the source chronograms that form a grove, roughly, a

185 sufficiently overlapping set of taxa between trees, by implementing definition 2.8 for

186 n-overlap from Ané et al. (2009). If the source chronograms do not form a grove, the

187 supertree reconstruction will fail. In rare cases, a group of trees can have multiple groves.

188 By default, DateLife chooses the grove with the most taxa, however, the “criterion = trees”

189 flag allows the user to choose the grove with the most trees instead. ~~If source chronograms~~

190 ~~do not form a grove, the supertree reconstruction will fail. The result is a single summary~~

191 ~~(or supertree) topology, that combines topologies from source chronograms in a grove.~~

192 ***Dating the topology.*** Input topologies from OpenTree or the supertree approach

193 ~~described above do not include branch length estimates of any kind. Optionally, to~~

194 ~~estimate branch lengths proportional to substitution rates on these topologies, DateLife~~

195 ~~can mine the Barcode of Life Data System, BOLD (Ratnasingham & Hebert, 2007) to~~

196 ~~obtain genetic markers for the input taxa. These markers are aligned with MUSCLE~~

197 ~~(Edgar, 2004) (by default) or MAFFT (Katoh, Asimenos, & Toh, 2009). This alignment~~

198 can be used to estimate branch lengths on input topologies that lack branch lengths.
 199 Currently, branch length reconstruction in DateLife is performed using parsimony and the
 200 likelihood of the phylogenetic tree given a sequence alignment is computed (Sehliep, 2011).
 201 While relative branch length information provides additional data for nodes without
 202 secondary date calibrations, topologies without branch lengths can also be dated.

203 Applying secondary calibrations

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

214 For each congruent node, the pairwise distances that traverse that node are
 215 summarized into a single summary matrix using classic summary statistics (i.e., mean,
 216 median, minimum and maximum ages), and the Supermatrix Distance Method (SDM;
 217 Criscuolo, Berry, Douzery, & Gascuel, et al., 2006), which deforms patristic distance
 218 matrices by minimizing variance and then averaging them. These single summary taxon pair
 219 age matrices (Summarized calibrations) can be applied as are stored as summarized
 220 calibrations that can be used as secondary calibrations to date a tree topology, using
 221 different - with or without initial branch lengths, using phylogenetic dating methods
 222 currently supported within DateLife: BLADJ (Webb, Ackerly, & Kembel, 2008; Webb &
 223 Donoghue, 2005), MrBayes (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003),

²²⁴ PATHd8 (Britton, Anderson, Jacquet, Lundqvist, & Bremer, 2007), **BLADJ** (Webb, Ackerly,
²²⁵ & Kembel, 2008; Webb & Donoghue, 2005), and treePL (Smith & O'Meara, 2012).

²²⁶ By default, DateLife implements the

²²⁷ **Dating a tree topology**

²²⁸ **With no branch lengths.** When producing or obtaining a tree with branch lengths
²²⁹ for a group of interest is not possible, DateLife can date a topology without branch lengths,
²³⁰ obtained from OpenTree or by implementing the supertree approach described above, by
²³¹ implementing the Branch Length Adjuster (BLADJ) algorithm to obtain a fully dated
²³² topology. BLADJ fixes node ages that have calibration data, and distributes time between
²³³ ; Webb et al., 2008; Webb & Donoghue, 2005) algorithm, which requires no initial branch
²³⁴ lengths. The algorithm starts by fixing ages for nodes with calibration data upon the given
²³⁵ topology. Then, it distributes time for nodes with no data evenly between nodes with
²³⁶ calibration data. This minimizes calibrated nodes, minimizing age variance in the resulting
²³⁷ chronogram (Webb et al., 2008), which has proven useful for ecological analyses that require
²³⁸ age data (Webb et al., 2008). BLADJ does not use branch lengths even when they are
²³⁹ present in the input tree or summarizing topology. When there is conflict in ages between
²⁴⁰ nodes with calibration data between ages of calibrated nodes, BLADJ ignores node ages that
²⁴¹ are older than the age of a parent node. BLADJ The BLADJ algorithm requires a root age
²⁴² estimate to run. If there is no information on the age of the root in the chronogram database,
²⁴³ users can provide an estimate from the literature. If none is provided, DateLife assigns an
²⁴⁴ arbitrary age to the root as 10% older than the oldest age available within the group will
²⁴⁵ not return a dated topology and will provide a warning message along with suggestions on
²⁴⁶ how the user can provide an age for the root so that a fully dated tree can be returned.

²⁴⁷ In the absence of genetic data, we consider that BLADJ is the most agnostic way to
²⁴⁸ assign ages to nodes with no available data, as it does not require any assumptions on the

underlying model of branch length distribution. It is however common practice in the literature to use a birth-death model to assign ages to nodes with no genetic data (Jetz, Thomas, Joy, Hartmann, & Mooers, 2012; Rabosky et al., 2018; Smith & Brown, 2018). To do so, DateLife implements MrBayes (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003), using nodes with published age data as calibration priors on a fixed topology, a simple birth-death model with parameters that can be determined by the user, and no genetic data.

Alternative phylogenetic dating options supported in DateLife (MrBayes, PATH8, TreePL)

Dating a tree with branch lengths. Relative branch lengths provide key information for phylogenetic dating, specifically for nodes without any calibration data available. While, using initial branch lengths data is the golden standard for phylogenetic dating analyses, producing such requires assembling and curating a homology hypothesis, and choosing and implementing a method for phylogenetic inference. DateLife implements a workflow to streamline this process by leveraging on genetic data from the Barcode of Life Data System, BOLD (Ratnasingham & Hebert, 2007) to obtain markers for input taxa. By default, mined genetic sequences are aligned with MUSCLE (Edgar, 2004) using functions from the msa R package (Bodenhofer et al., 2015). Alternatively, sequences can be aligned with MAFFT (Katoh, Asimenos, & Toh, 2009), using functions from the ape R package (Paradis et al., 2004). The BOLD sequence alignment is then used to obtain initial branch lengths with the accelerated transformation (ACCTRAN) parsimony algorithm, which resolves ambiguous character optimization by assigning changes along branches of the tree as close to the root as possible (Agnarsson & Miller, 2008), resulting in older nodes (Forest et al., 2005). The parsimony branch lengths are then optimized using Maximum Likelihood, given the alignment, the topology and a simple Jukes-Cantor model, producing a BOLD tree with branch lengths proportional to expected number of substitutions per site. Both parsimony and ML optimizations are done with functions from the phangorn

276 package (Schliep, 2011). The BOLD workflow is currently only available on DateLife's R
277 package and not on the web application.

278 Phylogenetic dating methods supported in DateLife that incorporate branch length
279 information from the input topology in combination with the calibrations. secondary
280 calibrations include: PATHd8is, a non-clock, rate-smoothing method to date trees (Britton
281 et al., 2007)to date trees; treePL (Smith & O'Meara, 2012), is a semi-parametric,
282 rate-smoothing, penalized likelihood dating method (Sanderson, 2002). The; and MrBayes
283 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003) approach in DateLife uses
284 the calibrations as priors on node ages, a Bayesian inference program implementing Markov
285 chain Monte Carlo (MCMC) methods to estimate a posterior distribution of model
286 parameters.

287 ***Visualizing results.***

288 **Visualizing results**

289 Finally, users can save all source and summary chronograms in formats that permit
290 reuse and reanalyses (newick and allowing for reuse and reanalysis, such as newick and the
291 R "phylo" format), as well as visualize and compare results graphically, or. Input and
292 summary chronograms can be visualized and compared graphically, and users can construct
293 their own graphs using DateLife's chronogram plot generation functions available from the R
294 package `datelifeplot` (Sanchez-Reyes & O'Meara, 2022).

295 **Benchmark**

296 R package `datelife` 's R package' code speed was tested on an Apple iMac with one
297 3.4 GHz Intel Core i5 processor. We registered variation in computing time of query
298 processing and search through the database relative to number of queried taxon names.
299 Query processing time increases roughly linearly with number of input taxon names, and
300 increases considerably if Taxonomic Name Resolution Service (TNRS) is activated. Up to

ten thousand names can be processed and searched in less than 30 minutes with the most time consuming settings. Once names have been processed as described in methods, a name search through the chronogram database can be performed in less than a minute, even with a very large number of taxon names (Fig. 2).

`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](#) 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

326 for *Schoeniclus elegans* in the default reference taxonomy (OTT v3.3, June 1, 2021). For a
327 detailed discussion on the state of the synonym, refer to Avibase (Avibase, 2022; Lepage,
328 2004; Lepage, Vaidya, & Guralnick, 2014). Discovering this synonym allowed assigning five
329 age data points for the parent node of *Emberiza elegans*, shown as *Schoeniclus elegans* in
330 figure 3A, which would not have had any data otherwise.

331 **Searching the database.** DateLife used the processed input names to search the
332 local chronogram database and found 9 matching chronograms in 6 different studies (Fig.
333 3B). Three studies matched five input names (Barker, Burns, Klicka, Lanyon, & Lovette,
334 2015; Hedges, Marin, Suleski, Paymer, & Kumar, 2015; Jetz, ~~Thomas, Joy, Hartmann, &~~
335 ~~Mooers, et al.~~, 2012), one study matched four input names (Hooper & Price, 2017) and two
336 studies matched two input names (Barker, Burns, Klicka, Lanyon, & Lovette, 2013; Burns et
337 al., 2014). No studies matched all input names. Together, source chronograms provide 28
338 unique age data points, covering all nodes on our chosen tree topology to date (Table 1).

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

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

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

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

372 **Summarizing search results.** The final step ~~is to congruify and summarize~~
373 entailed congruifying and summarizing the age data available for the Fringillidae species
374 into two single summary chronograms, using two different types of summary ages, median

375 and SDM. As explained in the “Description” section, a tree topology to summarize age data
376 upon is required. By default, DateLife uses the topology from OpenTree’s synthetic tree that
377 contains ~~the species in the search query to summarize age data upon all taxa from the search~~
378 query. According to OpenTree’s synthetic tree, species belonging to the family Fringillidae
379 do not form a monophyletic group (Fig. 4). Hence, a topology containing only the 289
380 species from the original query was extracted from Open Tree of Life’s synthetic tree v12.3
381 (Open Tree Of Life et al., 2019).

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

387 Cross-validation test

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

398

Discussion

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

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

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

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

419 **Age variation in source chronograms**

420 Conflict in estimated ages among alternative studies is common in the literature. See,
421 for example, the robust ongoing debate about crown group age of angiosperms
422 (Barba-Montoya, Reis, Schneider, Donoghue, & Yang, 2018; Magallón, Gómez-Acevedo,

423 Sánchez-Reyes, & Hernández-Hernández, 2015; Ramshaw et al., 1972; Sanderson & Doyle,
424 2001; Sauquet, Ramírez-Barahona, & Magallón, 2021). Source chronograms available for the
425 same organisms have potentially been estimated implementing calibrations very differently.
426 ~~For example~~ In the Fringillidae example shown here, the chronograms from Burns et al.
427 (2014) were inferred using molecular substitution rate estimates across birds (Weir &
428 Schlüter, 2008), and have much older age estimates for the same nodes than chronograms
429 that were inferred using fossils as calibrations (Figs. 5, 6; Supplementary Figs. S1, S5).

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

438 Primary vs Secondary calibrations

439 While most chronograms in DateLife's database are constructed using primary
440 calibrations (molecular substitution rates or ages obtained from the fossil record or geological
441 events), DateLife summarizes chronograms using secondary calibrations (ages coming from
442 other chronograms). Graur and Martin (2004) cautioned on the increased error and
443 uncertainty in estimated ages when using secondary calibrations in dating analyses. Schenk
444 (2016) showed that, in simulations, divergence times inferred using secondary calibrations are
445 significantly younger than those inferred with primary calibrations, when obtained with
446 Bayesian inference methods, and when priors are implemented in similar ways in both
447 analyses. Accordingly, the scientific community seems to have more confidence in
448 chronograms obtained from a single analysis, using fossil data as primary sources of

449 calibrations (Schenk, 2016), and using fossils that have been widely discussed and curated as
450 calibrations to date other trees, making sure that all data reflect a coherent evolutionary
451 history (Sauquet, 2013), as for example done by Antonelli et al. (2017). There have been
452 attempts to create fossil calibration databases (Ksepka et al., 2015), though these still have
453 room to grow.

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

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

472 **Summarizing chronograms**

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

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

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

492 The exercise of summarizing age data from across multiple studies provides the
493 opportunity to work with a **more-inclusive chronogram**, chronogram that reflects a unified
494 evolutionary history for a lineage, by putting together evidence from different hypotheses.
495 The largest, and taxonomically broadest chronogram currently available from OpenTree was
496 constructed summarizing age data from 2,274 published chronograms using NCBI's

497 taxonomic tree as backbone (Hedges et al., 2015). A summarizing exercise may also amplify
498 the effect of uncertainty and errors in source data, and blur parts of the evolutionary history
499 of a lineage that might only be reflected in source chronograms and lost on the summary
500 chronogram (Sauquet et al., 2021).

501 **Effects of phylogenetic sampling on downstream analyses**

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

505 ~~In ecology and conservation biology, incorporating at least some data on lineage~~
506 ~~divergence times represents a relevant improvement for testing alternative hypothesis using~~
507 ~~phylogenetic distance (Webb Analysis of species diversification using simulated and~~
508 ~~empirical phylogenies suggest that using a more completely sampled phylogeny provides~~
509 ~~estimates that are closer to the true diversification history than when analysing~~
510 ~~incompletely sampled phylogenies (Chang, Rabosky, & Alfaro, 2020; Cusimano, Stadler, &~~
511 ~~Renner, 2012; Sun et al., 2008)–2020). Ideally phylogenies should be completed using~~
512 ~~genetic data, but this is a very hard task to achieve for many biological groups.~~ Hence,
513 DateLife’s workflow features different ways of ~~estimating assigning~~ node ages in the absence
514 of calibrations and branch length information for certain taxa. ~~branch lengths Completing a~~
515 ~~phylogeny using a stochastic birth-death politomy resolver and a backbone taxonomy~~ is a
516 common practice in scientific publications: Jetz et al. (2012), created a chronogram of all 9,
517 993 bird species, where 67% had molecular data and the rest was simulated; Rabosky et al.
518 (2018) created a chronogram of 31, 536 ray-finned fishes, of which only 37% had molecular
519 data; Smith and Brown (2018) constructed a chronogram of 353, 185 seed plants where only
520 23% had molecular data. ~~These stochastically resolved chronograms provide less biased~~
521 ~~diversification rates estimated with methods that account for sampling fractions (Chang et~~
522 ~~al., 2020; Cusimano et al., 2012), but can also introduce spurious patterns of early bursts of~~

523 diversification (Cusimano & Renner, 2010; Sun et al., 2020).

524 Notably, Taxonomy based polity resolvers also introduce topological differences.

525 The study of macroevolutionary processes largely depends on an understanding of the
526 timing of species diversification events, and different phylogenetic and chronogram
527 hypothesis can provide very different overviews of the macroevolutionary history of a
528 biological group. For example, alternative topologies in chronograms from the same
529 biological group can infer very different species diversification patterns (Rabosky, 2015;
530 Title & Rabosky, 2016). Similarly there are worries that patterns of morphological
531 evolution cannot be accurately inferred with these types of phylogenies, as any patterns
532 would be erased by randomization (Rabosky, 2015). We note that the same applies for
533 geographical and morphological dependent diversification analysis. Hence, we suggest that
534 these types of randomly resolved phylogenies can be used as null or neutral models,
535 representing the case of diversification independent of traits and geographical scenario.

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

546

Conclusions

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

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

565

Availability

566 The DateLife software is free and open source~~and it~~. It can be used online through its
567 R shiny web application hosted at <http://www.datelife.org>, and locally through the
568 datelife R package, ~~and through Phylotastic's project web portal, downloadable from~~
569 Zenodo (<https://doi.org/10.5281/zenodo.593938>) and the CRAN repository (Sanchez-Reyes
570 et al., 2022). DateLife's web application is maintained using RStudio's shiny server and the
571 shiny package open infrastructure, as well as Docker and OpenTree's infrastructure

572 (dates.opentreeoflife.org/datelife). `datelife`'s R package stable version is available for
573 installation from the CRAN repository () CRAN repository using the command
574 `install.packages(pkgs = "datelife")` from within R. Development versions are
575 available from the DateLife's GitHub repository (<https://github.com/phylotastic/datelife>)
576 and can be installed using the command
577 `devtools::install_github("phylotastic/datelife")`.

578 Supplementary Material

579 ~~Code used to generate all versions of this manuscript, the Supplementary material,~~
580 ~~including code, biological examples, benchmark results, data files and online-only~~
581 ~~appendices, can be found in the Dryad data repository~~
582 ~~(<https://doi.org/10.5061/dryad.cnp5hqc6w>)~~, as well as ~~the benchmark of functionalities are~~
583 ~~available at datelifeMS1, datelife_examples, and datelife_benchmark repositories in~~
584 ~~LCSR's GitHub account in the Zenodo stable repositories that host the reproducible~~
585 ~~manuscript~~ (<https://doi.org/10.5281/zenodo.7435094>), ~~the biological examples~~
586 (<https://doi.org/10.5281/zenodo.7435101>), ~~and the software benchmark~~
587 (<https://doi.org/10.5281/zenodo.7435106>). ~~The corresponding development versions hosted~~
588 ~~on GitHub can be found at~~ <https://github.com/LunaSare/datelifeMS1#readme>
589 https://github.com/LunaSare/datelife_examples#readme, ~~and~~
590 https://github.com/LunaSare/datelife_benchmark#readme.

591 Funding

592 Funding was provided by the US National Science Foundation (NSF) grants
593 ABI-1458603 to the Datelife DateLife project; DBI-0905606 to the National Evolutionary
594 Synthesis Center (NESCenT), ABI-1458572 to the Phylotastic project, and ABI-1759846
595 to the Open Tree of Life project.

596

Acknowledgements

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

References

- 610
- 611 612 Agnarsson, I., & Miller, J. A. (2008). Is acctrans better than deltran? *Cladistics*, 24(6), 1032–1038.
- 613 614 615 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.
- 616 617 Ané, C., Eulenstein, O., Piaggio-Talice, R., & Sanderson, M. J. (2009). Groves of phylogenetic trees. *Annals of Combinatorics*, 13(2), 139–167.
- 618 619 620 621 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>
- 622 623 624 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>}
- 625 626 627 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>}
- 628 629 630 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>
- 631 632 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

- 633 a cretaceous terrestrial revolution. *New Phytologist*, 218(2), 819–834.
- 634 Barker, F. K. (2014). Mitogenomic data resolve basal relationships among passeriform and
635 passeridan birds. *Molecular Phylogenetics and Evolution*, 79, 313–324.
- 636 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2013). Going to
637 extremes: Contrasting rates of diversification in a recent radiation of new world
638 passerine birds. *Systematic Biology*, 62(2), 298–320.
- 639 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2015). New insights
640 into new world biogeography: An integrated view from the phylogeny of blackbirds,
641 cardinals, sparrows, tanagers, warblers, and allies. *The Auk: Ornithological Advances*,
642 132(2), 333–348.
- 643 Barker, F. K., Cibois, A., Schikler, P., Feinstein, J., & Cracraft, J. (2004). Phylogeny and
644 diversification of the largest avian radiation. *Proceedings of the National Academy of
645 Sciences*, 101(30), 11040–11045.
- 646 Beresford, P., Barker, F., Ryan, P., & Crowe, T. (2005). African endemics span the tree of
647 songbirds (passeri): Molecular systematics of several evolutionary “enigmas”.
648 *Proceedings of the Royal Society B: Biological Sciences*, 272(1565), 849–858.
- 649 [Bodenhofer, U., Bonatesta, E., Horejš-Kainrath, C., & Hochreiter, S. \(2015\). Msa: An r
650 package for multiple sequence alignment.](#) *Bioinformatics*, 31(24), 3997–3999.
- 651 Boyle, B., Hopkins, N., Lu, Z., Raygoza Garay, J. A., Mozzherin, D., Rees, T., ... Enquist,
652 B. J. (2013). The taxonomic name resolution service: An online tool for automated
653 standardization of plant names. *BMC Bioinformatics*, 14(1).
654 <https://doi.org/10.1186/1471-2105-14-16>
- 655 Britton, T., Anderson, C. L., Jacquet, D., Lundqvist, S., & Bremer, K. (2007). Estimating

- 656 Divergence Times in Large Phylogenetic Trees. *Systematic Biology*, 56(788777878),
657 741–752. <https://doi.org/10.1080/10635150701613783>
- 658 Bryson Jr, R. W., Chaves, J., Smith, B. T., Miller, M. J., Winker, K., Pérez-Emán, J. L., &
659 Klicka, J. (2014). Diversification across the new world within the ‘blue’cardinalids
660 (aves: Cardinalidae). *Journal of Biogeography*, 41(3), 587–599.
- 661 Burleigh, J. G., Kimball, R. T., & Braun, E. L. (2015). Building the avian tree of life using a
662 large-scale, sparse supermatrix. *Molecular Phylogenetics and Evolution*, 84, 53–63.
- 663 Burns, K. J., Shultz, A. J., Title, P. O., Mason, N. A., Barker, F. K., Klicka, J., ... Lovette,
664 I. J. (2014). Phylogenetics and diversification of tanagers (passeriformes:
665 Thraupidae), the largest radiation of neotropical songbirds. *Molecular Phylogenetics
666 and Evolution*, 75, 41–77.
- 667 Chamberlain, S. (2018). *bold: Interface to Bold Systems API*. Retrieved from
668 <https://CRAN.R-project.org/package=bold>
- 669 Chamberlain, S. A., & Szöcs, E. (2013). taxize : taxonomic search and retrieval in R [version
670 2; referees: 3 approved]. *F1000Research*, 2(191), 1–29.
671 <https://doi.org/10.12688/f1000research.2-191.v2>
- 672 Chang, J., Rabosky, D. L., & Alfaro, M. E. (2020). Estimating diversification rates on
673 incompletely sampled phylogenies: Theoretical concerns and practical solutions.
674 *Systematic Biology*, 69(3), 602–611.
- 675 Chaves, J. A., Hidalgo, J. R., & Klicka, J. (2013). Biogeography and evolutionary history of
676 the neotropical genus Saltator (aves: Thraupini). *Journal of Biogeography*, 40(11),
677 2180–2190.
- 678 Claramunt, S., & Cracraft, J. (2015). A new time tree reveals earth history’s imprint on the

- 679 evolution of modern birds. *Science Advances*, 1(11), e1501005.
- 680 Criscuolo, A., Berry, V., Douzery, E. J., & Gascuel, O. (2006). SDM: A fast distance-based
681 approach for (super)tree building in phylogenomics. *Systematic Biology*, 55(5),
682 740–755. <https://doi.org/10.1080/10635150600969872>
- 683 [Cusimano, N., & Renner, S. S. \(2010\). Slowdowns in diversification rates from real](#)
684 [phylogenies may not be real. *Systematic Biology*, 59\(4\), 458–464.](#)
- 685 [Cusimano, N., Stadler, T., & Renner, S. S. \(2012\). A new method for handling missing](#)
686 [species in diversification analysis applicable to randomly or nonrandomly sampled](#)
687 [phylogenies. *Systematic Biology*, 61\(5\), 785–792.](#)
- 688 Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M.-K., Turon, X., ... Douzery,
689 E. J. (2018). A phylogenomic framework and timescale for comparative studies of
690 tunicates. *BMC Biology*, 16(1), 1–14.
- 691 Eastman, J. M., Harmon, L. J., & Tank, D. C. (2013). Congruification: Support for time
692 scaling large phylogenetic trees. *Methods in Ecology and Evolution*, 4(7), 688–691.
693 <https://doi.org/10.1111/2041-210X.12051>
- 694 Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high
695 throughput. *Nucleic Acids Research*, 32(5), 1792–1797.
- 696 Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American Naturalist*,
697 125(1), 1–15. Retrieved from <http://www.jstor.org/stable/2461605>
- 698 [Forest, F., Savolainen, V., Chase, M. W., Lupia, R., Bruneau, A., & Crane, P. R. \(2005\).](#)
699 [Teasing apart molecular-versus fossil-based error estimates when dating](#)
700 [phylogenetic trees: A case study in the birch family \(betulaceae\). *Systematic*](#)
701 [Botany, 30\(1\), 118–133.](#)

- 702 Freckleton, R. P., Harvey, P. H., & Pagel, M. (2002). Phylogenetic analysis and
703 comparative data: A test and review of evidence. *The American Naturalist*.
- 704 GBIF Secretariat. (2022). GBIF Backbone Taxonomy. *Checklist dataset*, (Online Resource
705 accessed via GBIF.org). Retrieved from {<https://doi.org/10.15468/39omei> }
- 706 Gibb, G. C., England, R., Hartig, G., McLenachan, P. A., Taylor Smith, B. L., McComish,
707 B. J., ... Penny, D. (2015). New zealand passerines help clarify the diversification of
708 major songbird lineages during the oligocene. *Genome Biology and Evolution*, 7(11),
709 2983–2995.
- 710 Graur, D., & Martin, W. (2004). Reading the entrails of chickens: Molecular timescales of
711 evolution and the illusion of precision. *TRENDS in Genetics*, 20(2), 80–86.
- 712 Hackett, S. J., Kimball, R. T., Reddy, S., Bowie, R. C., Braun, E. L., Braun, M. J., ...
713 others. (2008). A phylogenomic study of birds reveals their evolutionary history.
714 *Science*, 320(5884), 1763–1768.
- 715 Harvey, P. H., Pagel, M. D., & others. (1991). *The comparative method in evolutionary
716 biology* (Vol. 239). Oxford university press Oxford.
- 717 Hedges, S. B., Marin, J., Suleski, M., Paymer, M., & Kumar, S. (2015). Tree of life reveals
718 clock-like speciation and diversification. *Molecular Biology and Evolution*, 32(4),
719 835–845. <https://doi.org/10.1093/molbev/msv037>
- 720 Heibl, C. (2008). *PHYLOCH: R language tree plotting tools and interfaces to diverse
721 phylogenetic software packages*. Retrieved from
722 <http://www.christophheibl.de/Rpackages.html>
- 723 Hooper, D. M., & Price, T. D. (2017). Chromosomal inversion differences correlate with
724 range overlap in passerine birds. *Nature Ecology & Evolution*, 1(10), 1526.

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

748 Cambridge University Press.

- 749 Lepage, D. (2004). *Avibase: The world bird database*. Bird Studies Canada.
- 750 Lepage, D., Vaidya, G., & Guralnick, R. (2014). Avibase—a database system for managing
751 and organizing taxonomic concepts. *ZooKeys*, (420), 117.
- 752 Lerner, H. R., Meyer, M., James, H. F., Hofreiter, M., & Fleischer, R. C. (2011). Multilocus
753 resolution of phylogeny and timescale in the extant adaptive radiation of hawaiian
754 honeycreepers. *Current Biology*, 21(21), 1838–1844.
- 755 Lovette, I. J., Pérez-Emán, J. L., Sullivan, J. P., Banks, R. C., Fiorentino, I.,
756 Córdoba-Córdoba, S., ... others. (2010). A comprehensive multilocus phylogeny for
757 the wood-warblers and a revised classification of the parulidae (aves). *Molecular
758 Phylogenetics and Evolution*, 57(2), 753–770.
- 759 Magallon, S., & Sanderson, M. (2001). Absolute diversification rates in angiosperm clades.
760 *Evolution*, 55(9), 1762–1780.
- 761 Magallón, S. (2010). Using fossils to break long branches in molecular dating: A comparison
762 of relaxed clocks applied to the origin of angiosperms. *Systematic Biology*, 59(4),
763 384–399.
- 764 Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., & Hernández-Hernández, T. (2015).
765 A metacalibrated time-tree documents the early rise of flowering plant phylogenetic
766 diversity. *New Phytologist*, 207(2), 437–453.
- 767 McTavish, E. J., Hinchliff, C. E., Allman, J. F., Brown, J. W., Cranston, K. A., Holder, M.
768 T., ... Smith, S. (2015). Phylesystem: A git-based data store for community-curated
769 phylogenetic estimates. *Bioinformatics*, 31(17), 2794–2800.
- 770 Michonneau, F., Brown, J. W., & Winter, D. J. (2016). rotl: an R package to interact with

- 771 the Open Tree of Life data. *Methods in Ecology and Evolution*, 7(12), 1476–1481.
- 772 <https://doi.org/10.1111/2041-210X.12593>
- 773 Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecology Letters*,
- 774 17(4), 508–525. <https://doi.org/10.1111/ele.12251>
- 775 Moyle, R. G., Oliveros, C. H., Andersen, M. J., Hosner, P. A., Benz, B. W., Manthey, J. D.,
- 776 ... Faircloth, B. C. (2016). Tectonic collision and uplift of Wallacea triggered the
- 777 global songbird radiation. *Nature Communications*, 7(1), 1–7.
- 778 Oliveros, C. H., Field, D. J., Ksepka, D. T., Barker, F. K., Aleixo, A., Andersen, M. J., ...
- 779 others. (2019). Earth history and the passerine superradiation. *Proceedings of the*
- 780 *National Academy of Sciences*, 116(16), 7916–7925.
- 781 Ooms, J., & Chamberlain, S. (2018). *Phylocomr: Interface to 'phylocom'*. Retrieved from
- 782 <https://CRAN.R-project.org/package=phylocomr>
- 783 Open Tree Of Life, Redelings, B., Cranston, K. A., Allman, J., Holder, M. T., & McTavish,
- 784 E. J. (2016). Open Tree of Life APIs v3.0. *Open Tree of Life Project*, (Online
- 785 Resources). Retrieved from
- 786 {<https://github.com/OpenTreeOfLife/germinator/wiki/Open-Tree-of-Life-Web->
- 787 APIs}
- 788 Open Tree Of Life, Redelings, B., Sánchez Reyes, L. L., Cranston, K. A., Allman, J., Holder,
- 789 M. T., & McTavish, E. J. (2019). Open tree of life synthetic tree v12.3. *Zenodo*.
- 790 Retrieved from <https://doi.org/10.5281/zenodo.3937742>
- 791 Ödeen, A., Håstad, O., & Alström, P. (2011). Evolution of ultraviolet vision in the largest
- 792 avian radiation—the passerines. *BMC Evolutionary Biology*, 11(1), 1–8.
- 793 Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and

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

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

819 Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., ... others.
820 (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*,
821 559(7714), 392.

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

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

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

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

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

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

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

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

- 840 Sanchez-Reyes, L. L., & O'Meara, B. (2022). *datelifeplot*: Methods to plot chronograms
841 and outputs of the *datelife* package. *R Package Release V0.2.2*. Retrieved from
842 <https://zenodo.org/badge/latestdoi/381501451>
- 843 Sanchez-Reyes, L. L., O'Meara, B., Eastman, J., Heath, T., Wright, A., Schliep, K., ...
844 Alfaro, M. (2022). *datelife*: Scientific Data on Time of Lineage Divergence for Your
845 Taxa. In *R Package Release V0 package version 0.6.6.2*. Retrieved from
846 <https://CRAN.R-project.org/package=datelife> and
847 <https://doi.org/10.5281/zenodo.593938>
- 848 Sanderson, M. (2002). Estimating Absolute Rates of Molecular Evolution and Divergence
849 Times: A Penalized Likelihood Approach. *Molecular Biology and Evolution*, 19(1),
850 101–109. <https://doi.org/10.1093/oxfordjournals.molbev.a003974>
- 851 Sanderson, M. (2003). r8s: Inferring Absolute Rates of Molecular Evolution and Divergence
852 Times in the Absence of a Molecular Clock. *Bioinformatics*, 19(2), 301–302.
- 853 Sanderson, M., & Doyle, J. (2001). Sources of error and confidence intervals in estimating
854 the age of angiosperms from *rbcL* and 18S rDNA data. *American Journal of Botany*,
855 88(8), 1499–1516.
- 856 Sauquet, H. (2013). A practical guide to molecular dating. *Comptes Rendus Palevol*, 12(6),
857 355–367.
- 858 Sauquet, H., Ramírez-Barahona, S., & Magallón, S. (2021). *The age of flowering plants is
859 unknown*.
- 860 Schenk, J. J. (2016). Consequences of secondary calibrations on divergence time estimates.
861 *PLoS ONE*, 11(1). <https://doi.org/10.1371/journal.pone.0148228>
- 862 Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in r. *Bioinformatics*, 27(4), 592–593.

- 863 Schoch, C. L., Ciufo, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R., ...
864 others. (2020). NCBI Taxonomy: a Comprehensive Update on Curation, Resources
865 and Tools. *Database, 2020*.
- 866 Selvatti, A. P., Gonzaga, L. P., & Moraes Russo, C. A. de. (2015). A paleogene origin for
867 crown passerines and the diversification of the oscines in the new world. *Molecular
868 Phylogenetics and Evolution, 88*, 1–15.
- 869 Smith, S., & Brown, J. (2018). Constructing a broadly inclusive seed plant phylogeny.
870 *American Journal of Botany, 105*(3), 302–314.
- 871 Smith, S., & O'Meara, B. (2012). TreePL: Divergence time estimation using penalized
872 likelihood for large phylogenies. *Bioinformatics, 28*(20), 2689–2690.
873 <https://doi.org/10.1093/bioinformatics/bts492>
- 874 Stoltzfus, A., Lapp, H., Matasci, N., Deus, H., Sidlauskas, B., Zmasek, C. M., ... Jordan, G.
875 (2013). Phylotastic! Making tree-of-life knowledge accessible, reusable and convenient.
876 *BMC Bioinformatics, 14*. <https://doi.org/10.1186/1471-2105-14-158>
- 877 Sun, M., Folk, R. A., Gitzendanner, M. A., Soltis, P. S., Chen, Z., Soltis, D. E., &
878 Guralnick, R. P. (2020). Estimating rates and patterns of diversification with
879 incomplete sampling: A case study in the rosids. *American Journal of Botany,
880 107*(6), 895–909.
- 881 Tietze, D. T., Päckert, M., Martens, J., Lehmann, H., & Sun, Y.-H. (2013). Complete
882 phylogeny and historical biogeography of true rosefinches (aves: Carpodacus).
883 *Zoological Journal of the Linnean Society, 169*(1), 215–234.
- 884 Title, P. O., & Rabosky, D. L. (2016). Do Macrophylogenies Yield Stable Macroevolutionary
885 Inferences? An Example from Squamate Reptiles. *Systematic Biology, syw102*.
886 <https://doi.org/10.1093/sysbio/syw102>

- 887 Treplin, S., Siegert, R., Bleidorn, C., Thompson, H. S., Fotso, R., & Tiedemann, R. (2008).
888 Molecular phylogeny of songbirds (aves: Passeriformes) and the relative utility of
889 common nuclear marker loci. *Cladistics*, 24(3), 328–349.
- 890 Uyeda, J. C., Pennell, M. W., Miller, E. T., Maia, R., & McClain, C. R. (2017). The
891 evolution of energetic scaling across the vertebrate tree of life. *The American*
892 Naturalist, 190(2), 185–199.
- 893 Vos, R. A., Balhoff, J. P., Caravas, J. A., Holder, M. T., Lapp, H., Maddison, W. P., ...
894 others. (2012). NeXML: Rich, extensible, and verifiable representation of
895 comparative data and metadata. *Systematic Biology*, 61(4), 675–689.
896 <https://doi.org/10.1093/sysbio/sys025>
- 897 Webb, C. (2000). Exploring the Phylogenetic Structure of Ecological Communities : An
898 Example for Rain Forest Trees. *The American Naturalist*, 156(2), 145–155.
- 899 Webb, C., Ackerly, D., & Kembel, S. (2008). Phylocom: Software for the analysis of
900 phylogenetic community structure and trait evolution. *Bioinformatics*, 24(18),
901 2098–2100. <https://doi.org/10.1093/bioinformatics/btn358>
- 902 Webb, C., & Donoghue, M. (2005). Phylomatic: Tree assembly for applied phylogenetics.
903 *Molecular Ecology Notes*, 5(1), 181–183.
- 904 Weir, J., & Schluter, D. (2008). Calibrating the avian molecular clock. *Molecular Ecology*,
905 17(10), 2321–2328.
- 906 Zuccon, D., Prŷs-Jones, R., Rasmussen, P. C., & Ericson, P. G. (2012). The phylogenetic
907 relationships and generic limits of finches (fringillidae). *Molecular Phylogenetics and*
908 *Evolution*, 62(2), 581–596.

909 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

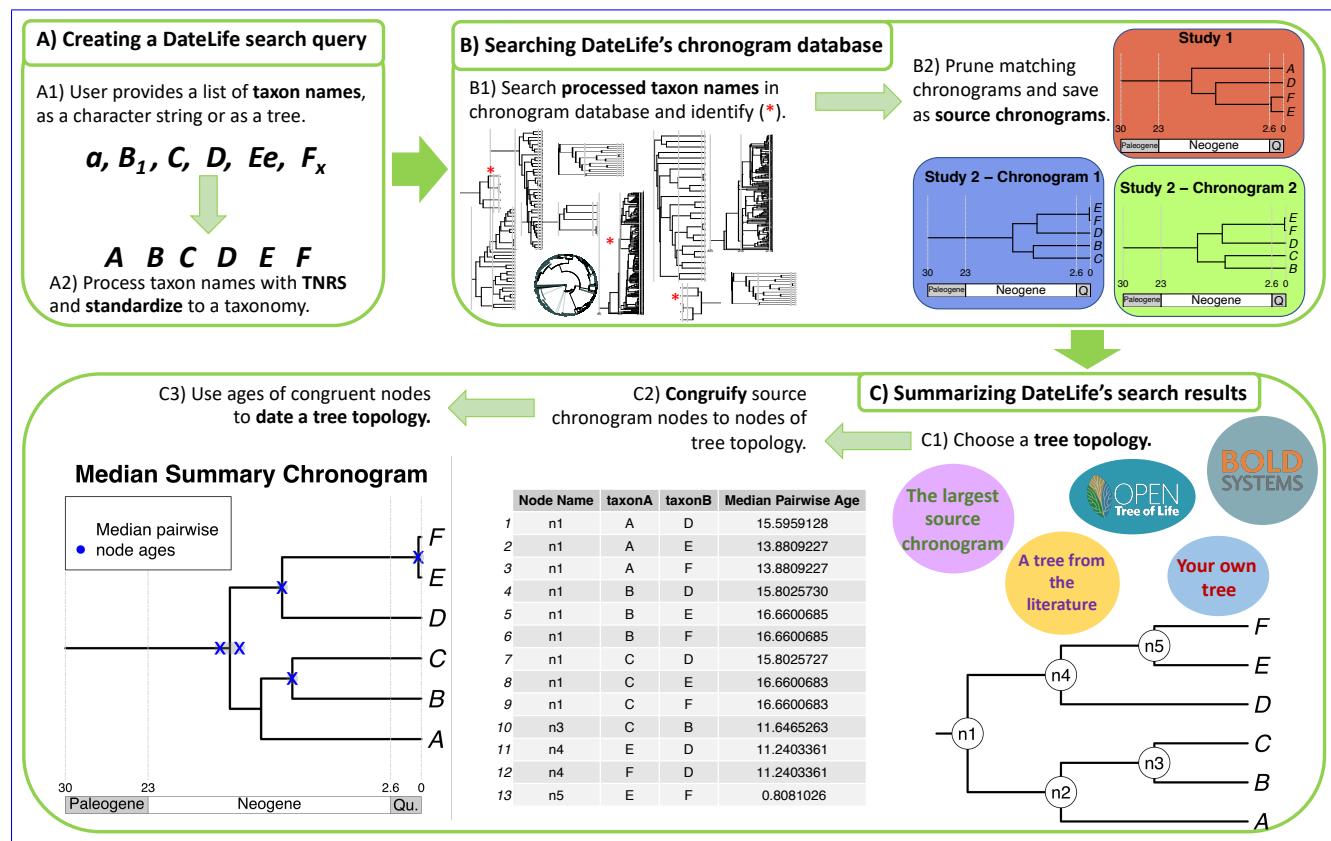
910

911 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

912

917



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

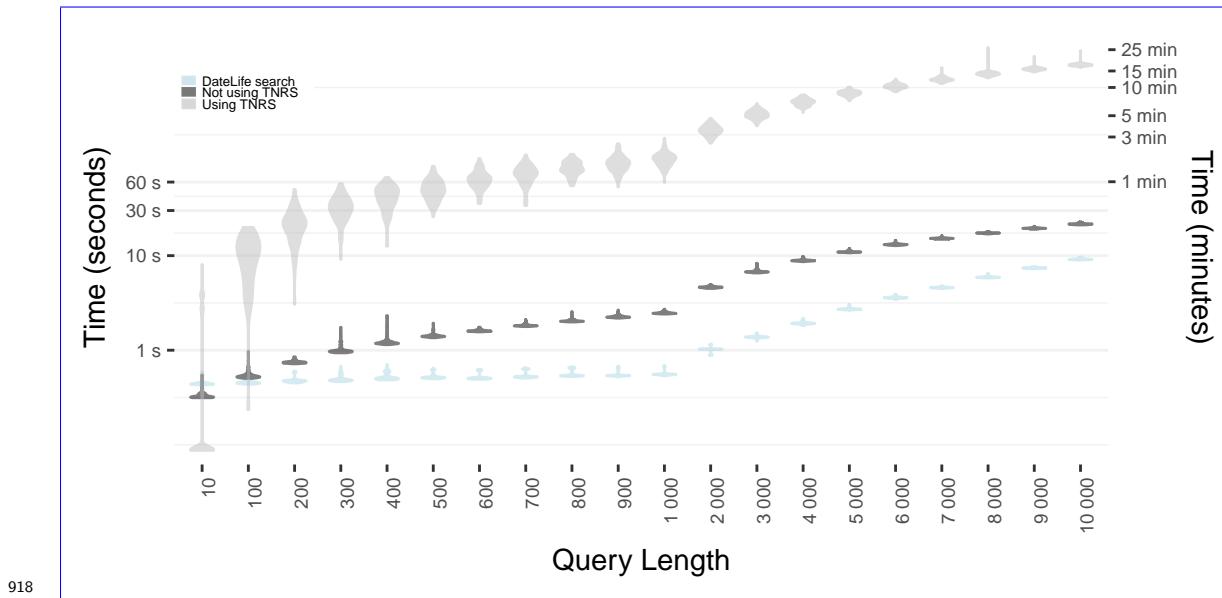
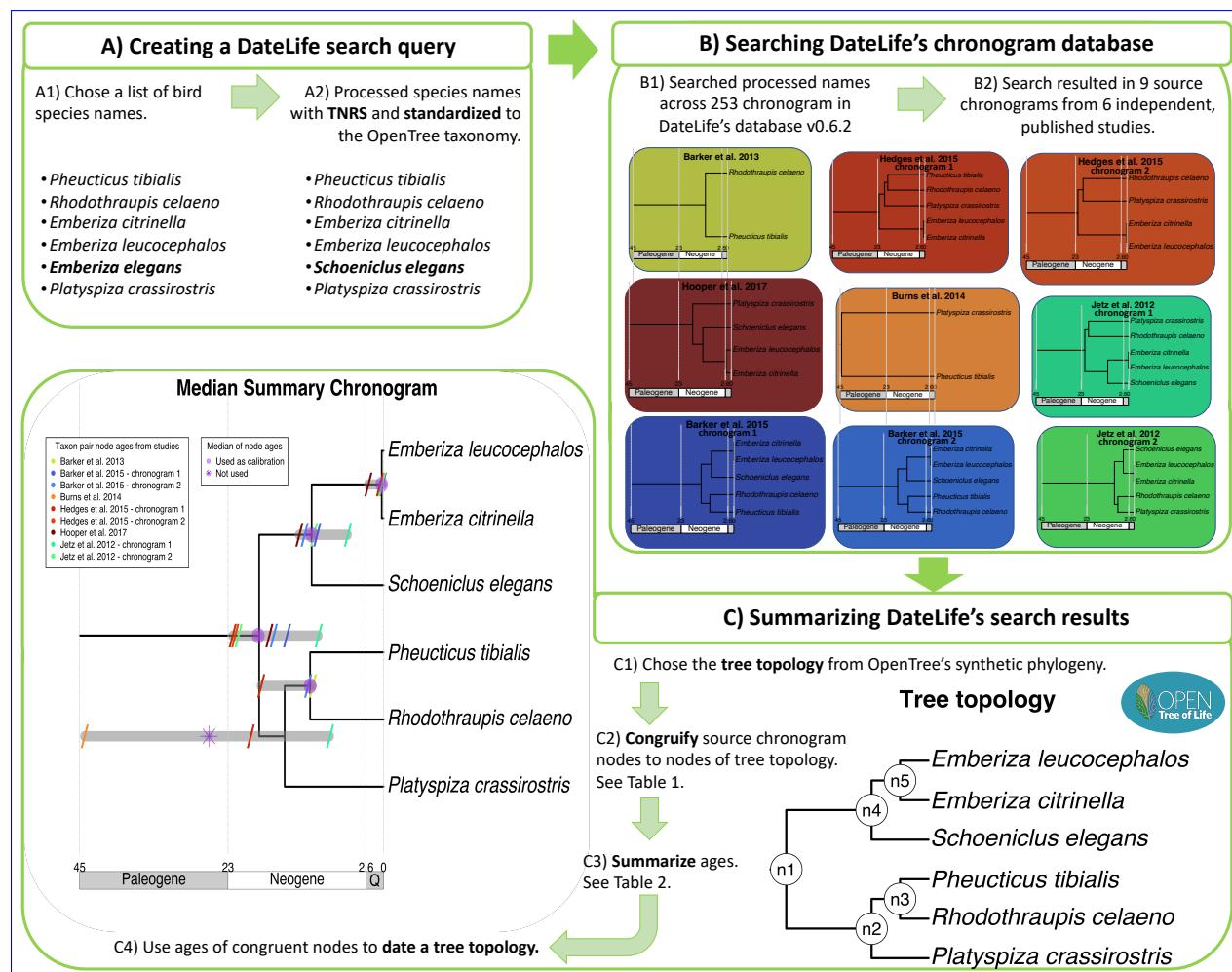


FIGURE 2. *DateLife's* benchmarking results. Computation time used to process a query and a search across *datelife*'s chronogram database, relative to number of input taxon names. For each $N = \{10, 100, 200, \dots, 1\,000, \dots, 9\,000, 10\,000\}$, we sampled N species names from the class Aves a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

926



927

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

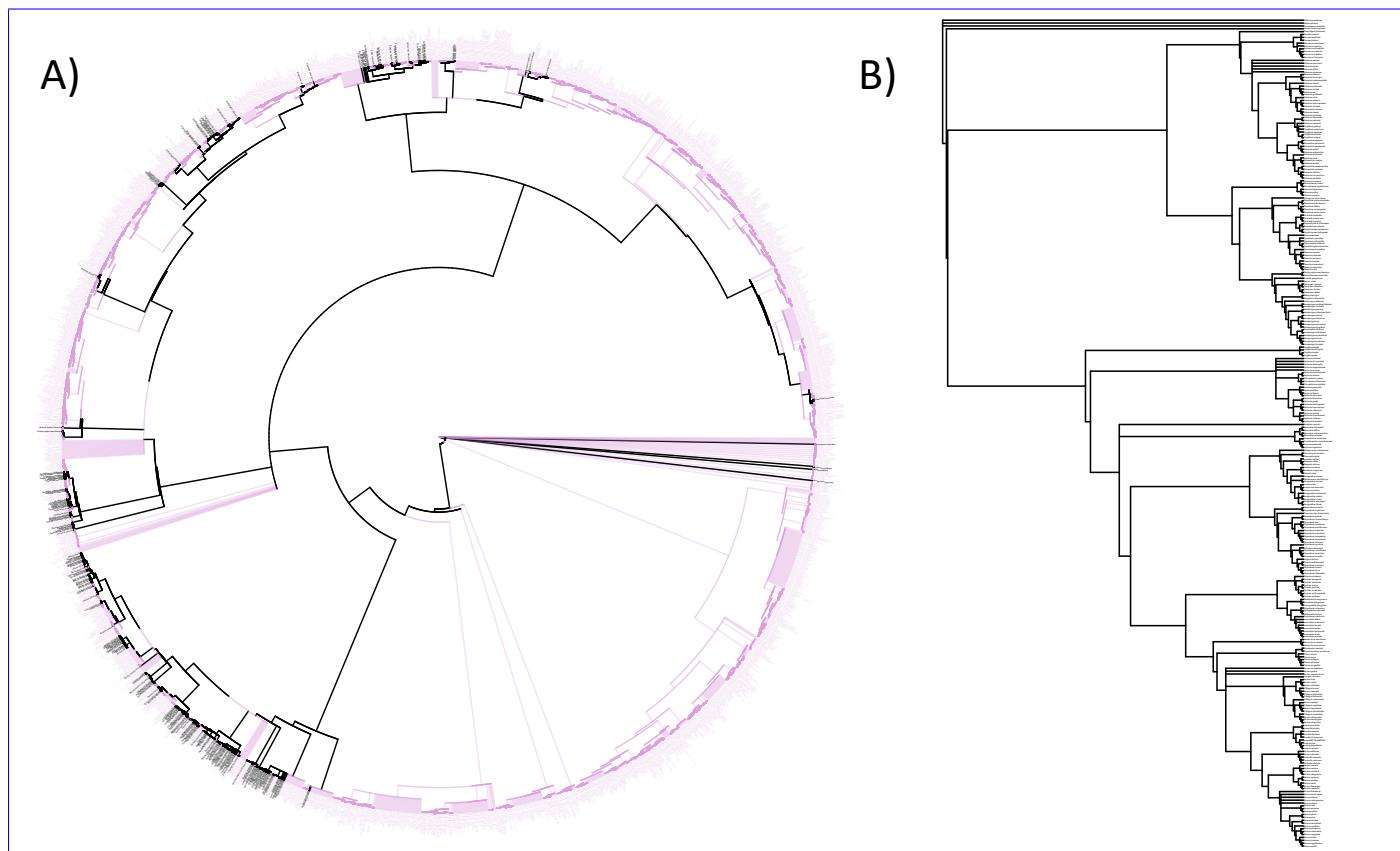
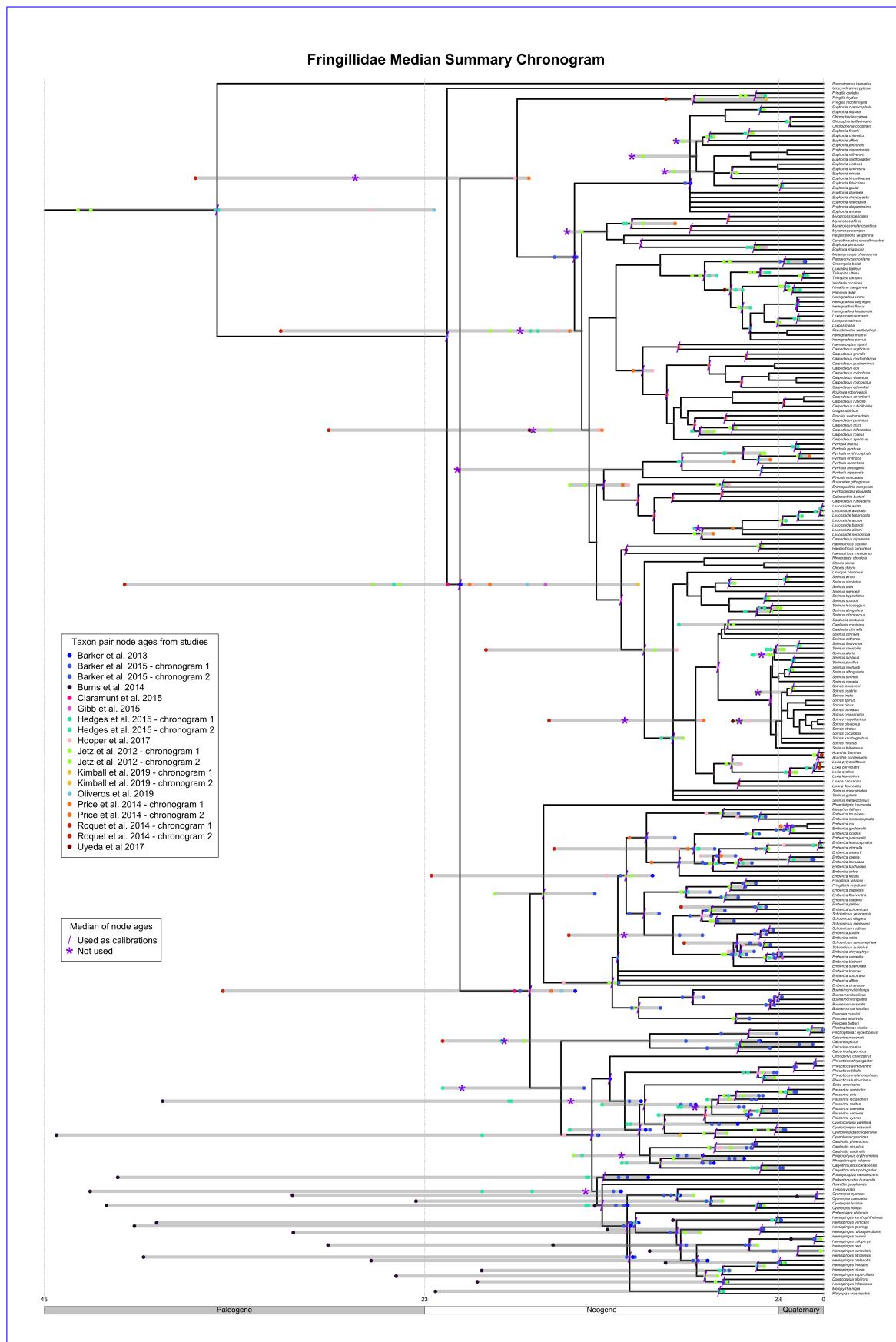
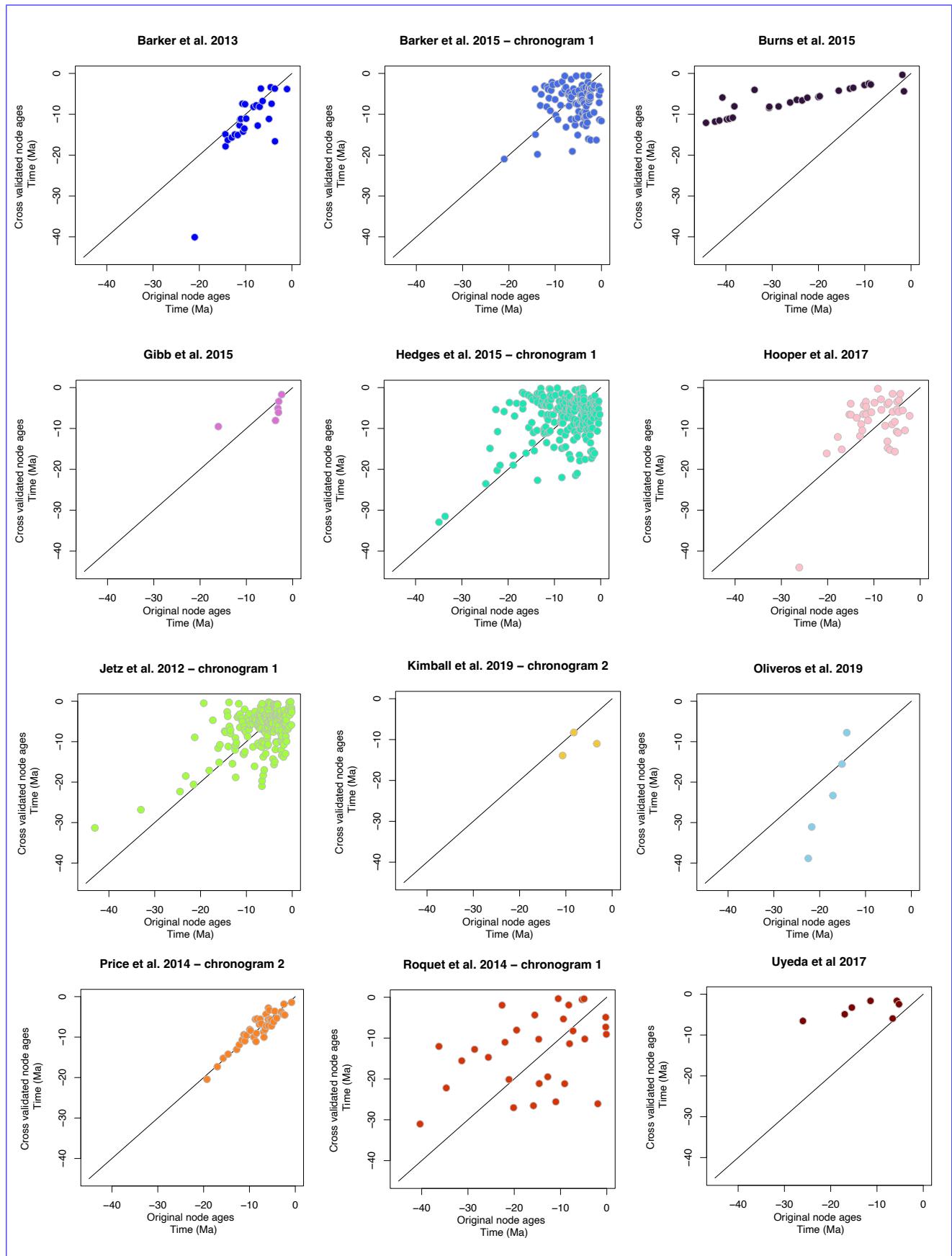


FIGURE 4. Tree topologies extracted from Open Tree of Life's (OpenTree) synthetic phylogenetic tree.
 A) Topology of 2,333 tips and 1,305 internal nodes, encompassing bird species within the family Fringillidae following the NCBI taxonomy (black), as well as all other bird species that share the same 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).



952 FIGURE 5. *Fringillidae* median summary chronogram generated with DateLife. It has 289 tips and
953 253 nodes, from which 212 have age data from at least one published chronogram. These nodes have



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