

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

⁴ ² University of Tennessee, Knoxville

⁵ Author Note

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

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

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

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

17

Abstract

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

42

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

⁴³ Congruification; Supertree; Calibrations; Secondary calibrations

⁴⁴ Word count: 4765

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

46 **Introduction**

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

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

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

67 **Description**

68 DateLife’s core software application consists of the R package `datelife`. Its latest
69 stable version – v0.6.2, is available from the CRAN repository (Sanchez-Reyes et al., 2022),

70 and relies on functionalities from various biological R packages: ape (Paradis, Claude, &
71 Strimmer, 2004), bold (Chamberlain et al., 2019), geiger (Pennell et al., 2014), paleotree
72 (Bapst, 2012), phyloch (Heibl, 2008), phylocomr (Ooms & Chamberlain, 2018), phytools
73 (Revell, 2012), rotl (Michonneau, Brown, & Winter, 2016), and taxize (Chamberlain & Szöcs,
74 2013; Chamberlain et al., 2019). Figure 1 provides a graphical summary of the three main
75 steps of the DateLife algorithm: creating a search query, searching a database, and
76 summarizing results from the search.

77 Creating a search query

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

83 DateLife accepts scientific names that can belong to any inclusive taxonomic group
84 (e.g., genus, family, tribe, etc.) or a binomial species name. Subspecies and variants are
85 ignored. If an input taxon name belongs to an inclusive taxonomic group, the algorithm has
86 two alternative behaviors defined by the “get species from taxon” flag. If the flag is active,
87 DateLife retrieves all species names within the inclusive taxonomic group (according to a
88 taxonomy of choice) and adds them to the input string. If the flag is inactive, DateLife
89 excludes the taxon names above the species level from the query.

90 DateLife processes input scientific names using a Taxonomic Name Resolution Service
91 (TNRS), which increases the probability of correctly finding the queried taxon names in the
92 chronogram database. TNRS detects, corrects and standardizes name misspellings and typos,
93 variant spellings and authorities, and nomenclatural synonyms to a single taxonomic
94 standard (Boyle et al., 2013). DateLife implements TNRS using OpenTree’s unified

95 taxonomy as standard (Open Tree Of Life et al., 2016; Rees & Cranston, 2017), storing
96 OpenTree's Taxonomy identification numbers for further processing.

97 The processed input taxon names are saved as an R object of a newly defined class,
98 `datelifeQuery`, that is used in the following steps. This object contains the standardized
99 names, the corresponding OpenTree taxonomic id numbers, and the topology of the input
100 tree if any was provided.

101 Searching a chronogram database

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

108 A unique feature of OpenTree's Phylesystem is that the community can add new
109 published, state-of-the-art chronograms any time through their curator application
110 (<https://tree.opentreeoflife.org/curator>). As chronograms are added to Phylesystem, they are
111 incorporated into an updated `datelife`'s database that is assigned a new version number,
112 followed by a package release on CRAN. `datelife`'s chronogram database is updated as new
113 chronogram data is added to Phylesystem, at a minimum of once a month and a maximum
114 of every 6 months. Users can also upload and curate new chronograms to OpenTree
115 themselves. Then, trigger an update of their local `datelife` database to incorporate the new
116 chronograms and have them available for their current analysis.

117 A DateLife search is implemented by matching processed taxon names provided by the
118 user to tip labels in the chronogram database. Chronograms with at least two matching
119 taxon names on their tip labels are identified and pruned down to preserve only the matched

120 taxa. These matching pruned chronograms are referred to as source chronograms. Total
121 distance (in units of millions of years) between taxon pairs within each source chronogram
122 are stored as a patristic distance matrix (Figure 1). The matrix format speeds up extraction
123 of pairwise taxon ages of any queried taxa, as opposed to searching the ancestor node of a
124 pair of taxa in a “phylo” object or newick string. Finally, the patristic matrices are
125 associated to the study citation where the original chronogram was published, and stored as
126 an R object of the newly defined class `datelifeResult`.

127 **Summarizing search results**

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

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

141 ***Choosing a topology.***— DateLife requires a tree topology to summarize age data
142 upon. We recommend that users provide a tree topology as input from the literature, or one
143 of their own making. If no topology is provided, DateLife automatically subsets one from the
144 OpenTree synthetic tree (Open Tree Of Life et al., 2019). Alternatively, DateLife can

145 combine topologies from source chronograms using a supertree approach. To combine
146 topologies from source chronograms into a single summary (or supertree) topology, the
147 DateLife algorithm starts by identifying the source chronograms that form a grove, roughly,
148 a sufficiently overlapping set of taxa between trees, by implementing definition 2.8 for
149 n-overlap from Ané et al. (2009). In rare cases, a group of trees can have multiple groves. By
150 default, DateLife chooses the grove with the most taxa, however, the “criterion = trees” flag
151 allows the user to choose the grove with the most trees instead. If source chronograms do
152 not form a grove, the supertree reconstruction will fail.

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

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

171 congruifiedCalibrations object.

172 For each congruent node, the pairwise distances that traverse that node are
173 summarized into a single summary matrix using classic summary statistics (i.e., mean,
174 median, minimum and maximum ages), and the Supermatrix Distance Method [SDM;
175 Criscuolo, Berry, Douzery, and Gascuel (2006)], which deforms patristic distance matrices by
176 minimizing variance and then averaging them. These single summary taxon pair age
177 matrices (Summarized calibrations) can be applied as calibrations to date a tree topology,
178 using different dating methods currently supported within DateLife: MrBayes (Huelsenbeck
179 & Ronquist, 2001; Ronquist & Huelsenbeck, 2003), PATHd8 (Britton, Anderson, Jacquet,
180 Lundqvist, & Bremer, 2007), BLADJ (Campbell O. Webb, Ackerly, & Kembel, 2008;
181 Campbell O. Webb & Donoghue, 2005), and treePL (Stephen A. Smith & O'Meara, 2012).

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

193 Alternative options supported in DateLife (MrBayes, PATHD8, TreePL) incorporate
194 branch length information from the input topology in combination with the calibrations.
195 PATHd8 is a non-clock, rate-smoothing method (Britton et al., 2007) to date trees. treePL
196 (Stephen A. Smith & O'Meara, 2012), is a semi-parametric, rate-smoothing, penalized

¹⁹⁷ likelihood dating method (Michael J. Sanderson, 2002). The MrBayes (Huelsenbeck &
¹⁹⁸ Ronquist, 2001; Ronquist & Huelsenbeck, 2003) approach in datelife uses the calibrations as
¹⁹⁹ priors on node ages.

²⁰⁰ **Visualizing results.**— Finally, users can save all source and summary chronograms in
²⁰¹ formats that permit reuse and reanalyses (newick and R “phylo” format), as well as visualize
²⁰² and compare results graphically, or construct their own graphs using **datelife**’s chronogram
²⁰³ plot generation functions available from the R package **datelifeplot** (Sanchez-Reyes &
²⁰⁴ O’Meara, 2022).

²⁰⁵ Benchmark

²⁰⁶ **datelife**’s code speed was tested on an Apple iMac with one 3.4 GHz Intel Core i5
²⁰⁷ processor. We registered variation in computing time of query processing and search through
²⁰⁸ the database relative to number of queried taxon names. Query processing time increases
²⁰⁹ roughly linearly with number of input taxon names, and increases considerably if Taxonomic
²¹⁰ Name Resolution Service (TNRS) is activated. Up to ten thousand names can be processed
²¹¹ and searched in less than 30 minutes with the most time consuming settings. Once names
²¹² have been processed as described in methods, a name search through the chronogram
²¹³ database can be performed in less than a minute, even with a very large number of taxon
²¹⁴ names (Fig. 2). **datelife**’s code performance was evaluated with a set of unit tests designed
²¹⁵ and implemented with the R package **testthat** (R Core Team, 2018) that were run both
²¹⁶ locally with the **devtools** package (R Core Team, 2018), and on a public server –via GitHub,
²¹⁷ using the continuous integration tool Travis CI (<https://travis-ci.org>). At present, unit tests
²¹⁸ cover more than 40% of **datelife**’s code (<https://codecov.io/gh/phylotastic/datelife>). Unit
²¹⁹ testing helps identify potential issues as code is updated or, more critically, as services code
²²⁰ relies upon may change.

221

Case studies

222 We illustrate the DateLife algorithm using a family within the Passeriform birds
223 encompassing the true finches, Fringillidae, as case study. The first example analyses 6 bird
224 species and shows all steps of the algorithm. The second example is an application analysing
225 289 species in the family Fringillidae that are included in the NCBI taxonomy.

226 **Small example**

227 ***Creating a search query.***- We chose 6 bird species in the family Fringillidae, known
228 as the true finches. The sample includes two species of cardinals: the black-thighed grosbeak
229 – *Pheucticus tibialis* and the crimson-collared grosbeak – *Rhodothraupis celaeno*; three species
230 of buntings: the yellowhammer – *Emberiza citrinella*, the pine bunting – *Emberiza*
231 *leucocephalos* and the yellow-throated bunting – *Emberiza elegans*; and one species of
232 tanager, the vegetarian finch – *Platyspiza crassirostris*.

233 Processing input names found that *Emberiza elegans* is synonym for *Schoeniclus*
234 *elegans* in the default reference taxonomy [Open Tree of Life Taxonomy v3.3, June 1, 2021].
235 For a detailed discussion on the state of the synonym refer to Avibase (Avibase, 2022;
236 Lepage, 2004; Lepage, Vaidya, & Guralnick, 2014). Discovering this synonym allowed
237 assigning five age data points for the parent node of *Emberiza elegans*, shown as *Schoeniclus*
238 *elegans* in figure 3, which would not have had any data otherwise.

239 ***Searching the database.***- DateLife used the processed input names to search the
240 local chronogram database and found 9 matching chronograms in 6 different studies. Three
241 studies matched five input names (Barker, Burns, Klicka, Lanyon, & Lovette, 2015; Hedges,
242 Marin, Suleski, Paymer, & Kumar, 2015; Jetz, Thomas, Joy, Hartmann, & Mooers, 2012),
243 one study matched four input names (Hooper & Price, 2017) and two studies matched two
244 input names (Barker, Burns, Klicka, Lanyon, & Lovette, 2013; Burns et al., 2014). No
245 studies matched all input names. Together, matching chronograms have 28 unique age data

246 points. All nodes have age data.

247 ***Summarizing search results.-*** DateLife used OpenTree's synthetic tree topology
248 for these taxa and mapped age data to nodes in the tree. As expected, more inclusive nodes
249 (e.g., node "n1") have more age data than less inclusive nodes (e.g., node "n5"). The name
250 resolution step allowed discovering five data points for node "n4" that would not have had
251 any data otherwise due to name mismatch. Age summary statistics per node were calculated
252 and used as calibrations to date the tree topology using the BLADJ algorithm. Age data for
253 node "n2" was excluded as final calibration because it is older than age data of the more
254 inclusive node "n1."

255 **Fringillidae**

256 ***Creating a query.-*** To estimate ages for species in the family Fringillidae, we ran a
257 `datelife` query using the "get species from taxon" flag, which gets all recognized species
258 names within a named group from a taxonomy of choice (options are Open Tree of Life,
259 NCBI, GBIF, or IRMNG). Following the NCBI taxonomy, the Fringillidae has 289 species.

260 ***Searching the database.-*** The next step is to use this curated set of species taxon
261 names to identify all chronograms that contain at least two Fringillidae species. Once
262 identified, the algorithm proceeds to prune matching chronograms to keep Fringillidae
263 species names on tips only, and transform these pruned chronograms to pairwise distance
264 matrices. The `datelife` search revealed 13 chronograms containing at least two Fringillidae
265 species, published in 9 different studies (Barker et al., 2013, 2015; Burns et al., 2014;
266 Claramunt & Cracraft, 2015; Gibb et al., 2015; Hedges et al., 2015; Hooper & Price, 2017;
267 Jetz et al., 2012; Price et al., 2014).

268 ***Summarizing search results.-*** The final step is to summarize the age data available
269 for the Fringillidae species into single summary chronograms, using different types of
270 summary ages, median and SDM. As explained in the "Description" section, a tree topology

271 to summarize age data upon is required. By default, `datelife` uses the Open Tree of Life
272 synthetic phylogeny as summarizing topology. According to this phylogeny, species belonging
273 to the family Fringillidae do not form a monophyletic group (Fig. 5). By default, DateLife's
274 taxon-constrained approach returns a topology and ages for the species in a named group,
275 which do not necessarily correspond to a monophyletic group. Users can change this
276 behaviour by providing a monophyletic tree as input for a DateLife search, or as a tree
277 topology for a DateLife summary.

278 Age data from source chronograms is summarised into a single patristic distance
279 matrix and then the available node ages are used as fixed node calibrations over a consensus
280 tree topology, to obtain a fully dated phylogeny with the program BLADJ (Fig. 6).

281 Median summary chronograms are older and have wider variation in maximum ages
282 than chronograms obtained with SDM.

283 **Cross-validation test**

284 We performed a cross validation analysis of the DateLife workflow using the
285 Fringillidae chronograms. We used the individual tree topologies from each of the 19 source
286 chronograms from 13 studies as inputs, treating their node ages as unknown. We then
287 estimated dates for these topologies using the node ages from the chronograms from the
288 other studies as calibrations and smoothing using BLADJ.

289 We found that node ages from original study, and ages estimated using all other age
290 data available are correlated (Supplementary Fig. 7).

291 For five studies, Datelife tended to underestimate ages for topologically deeper nodes
292 (those with many descendant taxa, aka 'closer to the root') relative to the orginal estimate,
293 and overestimate ages for nodes closer to the tips. Accordingly, root ages are generally older
294 in the original study than estimated using cross-validated ages (Supplementary Fig. 8).

295

Discussion

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

301 A total of 99474 unique terminal taxa are represented in `datelife`'s database.

302 Incorporation of more chronograms into `datelife`'s database will continue to improve its
303 services. One option to increase the number of chronograms in the DateLife database is the
304 Dryad data repository. Methods to automatically mine chronograms from Dryad could be
305 designed and implemented. However, Dryad's metadata system has no information to
306 automatically detect branch length units, and those would still need to be determined
307 manually by a human curator. We would like to emphasize on the importance of sharing
308 chronogram data, including systematically curated metadata, into open repositories, such as
309 OpenTree's Phylesystem (McTavish et al., 2015) for the benefit of the scientific community
310 as a whole.

311 **Primary vs Secondary calibrations in dating analyses**

312 While most chronograms in DateLife's database are constructed using primary
313 calibrations (molecular substitution rates or ages obtained from the fossil record or geological
314 events), DateLife summarizes chronograms using secondary calibrations only (ages coming
315 from other chronograms).

316 Graur and Martin (2004) cautioned on the increased error and uncertainty in estimated
317 ages when using secondary calibrations in dating analyses. Schenk (2016) showed that, in
318 simulations, divergence times inferred using secondary calibrations are significantly younger
319 than those inferred with primary calibrations, when obtained with Bayesian inference

320 methods, and when priors are implemented in similar ways in both analyses. Accordingly, the
321 scientific community seems to have more confidence in chronograms obtained from a single
322 analysis, using fossil data as primary sources of calibrations (Schenk, 2016), and using fossils
323 that have been widely discussed and curated as calibrations to date other trees, making sure
324 that all data used in the analysis reflect a coherent evolutionary history (Sauquet, 2013), as
325 for example done by (Antonelli et al., 2017). There have been attempts to create fossil
326 calibration databases (Ksepka et al., 2015), though these still have room to grow.

327 **LLSR: What does our cross validation analysis show in relation to diffs between**
328 **primary and secondary calibrations?**

329 It seems that using several (as opposed to just a few) secondary calibrations can
330 provide sufficient information to alleviate or even neutralize potential biases (Sauquet, 2013).
331 Certainly, further studies are required to fully understand the effect of secondary calibrations
332 on outputs from different tree dating methods, and on downstream analyses. It is possible
333 that secondary calibrations can be safely used with dating methods that do not require
334 setting priors, such as penalized likelihood (Michael J. Sanderson, 2003), methods that do
335 not make any assumptions on the ages and fix them to a node on a tree topology, such as
336 BLADJ (Campbell O. Webb et al., 2008; Campbell O. Webb & Donoghue, 2005), or
337 methods that summarize age data unto a tree topology.

338 **Variation in source chronograms**

339 Conflict among alternative dating studies is common in the literature (see, for example,
340 the robust ongoing debate about crown group age of angiosperms (Barba-Montoya, Reis,
341 Schneider, Donoghue, & Yang, 2018; Magallón, Gómez-Acevedo, Sánchez-Reyes, &
342 Hernández-Hernández, 2015; Ramshaw et al., 1972; Michael J. Sanderson & Doyle, 2001;
343 Sauquet, Ramírez-Barahona, & Magallón, 2021)).

344 Source chronograms may have been estimated using different types of primary

345 calibrations (fossil record, geologic or biogeographic events, molecular substitution rates).
346 These differences in approach can deepen the already substantial variation in time estimates
347 between lineages, as observed from the comparison of source chronograms in the Fringillidae
348 example. For example, the chronograms from Burns et al. (2014) tend to have much older
349 date estimates for the same nodes than other inferences (Figure 5). This study inferred dates
350 using molecular substitution rate estimates across birds (Weir & Schluter, 2008) **EJM:**
351 **MORE ON the FINCH examples and the cross validation here**

352 Different chronograms might also imply fundamentally distinct evolutionary
353 hypotheses. For example, two different researchers working on the same clade could both
354 carefully select and argue for their choices of fossil calibrations. Still, if one researcher
355 decides a fossil will calibrate the ingroup of a clade, while another researcher uses the same
356 one to calibrate outside the clade, the resulting age estimates will often differ substantially,
357 as the placement of calibrations as stem or crown group is proved to deeply affect estimated
358 times of lineage divergence (Sauquet, 2013).

359 Sumarizing chronograms

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

367 Some measures that have been proposed are the proportion of lineage sampling and the
368 number of calibrations used Magallón et al. (2015). Some characteristics that are often cited
369 in published studies as a measure of improved age estimates as compared to previously

370 published estimates are: quality of alignment (missing data, GC content), lineage sampling
371 (strategy and proportion), phylogenetic and dating inference method, number of fossils used
372 as calibrations, support for nodes and ages, and magnitude of confidence intervals.

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

378 The exercise of summarizing age data from across multiple studies provides the
379 opportunity to work with a more inclusive chronogram, that reflects a unified evolutionary
380 history for a lineage, by putting together evidence from different hypotheses. The largest,
381 and taxonomically broadest chronogram currently available from OpenTree was constructed
382 summarizing age data from 2,274 published chronograms using NCBI's taxonomic tree as
383 backbone (Hedges et al., 2015). A summarizing exercise may also amplify the effect of
384 uncertainty and errors in source data, and blur parts of the evolutionary history of a lineage
385 that might only be reflected in source chronograms and lost on the summary chronogram
386 (Sauquet et al., 2021). **EJM: ??? it's not clear to me what this last part means! What might**
387 **you be losing? LLSR: I changed loosing to blurring. But I will elaborate more on it**
388 Choosing source chronograms that we are going to keep and the ones that we are going to
389 discard might be key. **LLSR: Example**

390 Effects on downstream analyses

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

394 In ecology and conservation biology, incorporating at least some data on lineage

395 divergence times represents a relevant improvement for testing alternative hypothesis using
396 phylogenetic distance (Campbell O. Webb et al., 2008). Hence, DateLife's workflow features
397 different ways of estimating node ages in the absence of calibrations and branch length
398 information for certain taxa. "Making up" branch lengths is a common practice in scientific
399 publications: Jetz et al. (2012), created a time-calibrated tree of all 9,993 bird species, where
400 67% had molecular data and the rest was simulated; Rabosky et al. (2018) created a
401 time-calibrated tree of 31,536 ray-finned fishes, of which only 37% had molecular data;
402 Stephen A. Smith and Brown (2018) constructed a tree of 353,185 seed plants where only
403 23% had molecular data.

404 Notably, risks come with this practice. Taken to the extreme, one could make a fully
405 resolved, calibrated tree of all modern and extinct taxa using a single taxonomy and a single
406 calibration, using polytomy resolution and branch estimation methods. There has yet to be
407 a thorough analysis of what can go wrong when one extends inferences beyond the data in
408 this way, so we urge caution; we also urge readers to follow the example of the large tree
409 papers cited above, by carefully considering the statistical assumptions being made, and
410 assessing the consistency of the results with prior work.

411 Conclusions

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

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

Availability

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

Supplementary Material

⁴⁴² Code used to generate all versions of this manuscript, the biological examples, as well
⁴⁴³ as the benchmark of functionalities are available at datelifeMS1, datelife_examples, and
⁴⁴⁴ datelife_benchmark repositories in LLSR's GitHub account.

445

Funding

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

450

Acknowledgements

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

References

- 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.
- Ané, C., Eulenstein, O., Piaggio-Talice, R., & Sanderson, M. J. (2009). Groves of phylogenetic trees. *Annals of Combinatorics*, 13(2), 139–167.
- 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>
- Archie, J., Day, W. H., Felsenstein, J., Maddison, W., Meacham, C., Rohlf, F. J., & Swofford, D. (1986). The Newick tree format. Retrieved from %7B<https://evolution.genetics.washington.edu/phylip/newicktree.html>%7D
- Avibase. (2022). Yellow-throated Bunting. *Avibase - The World Bird Database*, (Online Resource). Retrieved from %7B<https://avibase.bsc-eoc.org/species.jsp?lang=EN&avibaseid=82D1EE0049D8D927>%7D
- 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>
- Barba-Montoya, J., Reis, M. dos, Schneider, H., Donoghue, P. C., & Yang, Z. (2018). Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a cretaceous terrestrial revolution. *New Phytologist*, 218(2), 819–834.
- Barker, F. K. (2014). Mitogenomic data resolve basal relationships among passeriform and passeridan birds. *Molecular Phylogenetics and Evolution*, 79, 313–324.
- Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2013). Going to extremes: Contrasting rates of diversification in a recent radiation of new world passerine birds. *Systematic Biology*, 62(2), 298–320.

- 490 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2015). New
491 insights into new world biogeography: An integrated view from the phylogeny of
492 blackbirds, cardinals, sparrows, tanagers, warblers, and allies. *The Auk: Ornithological Advances*, 132(2), 333–348.
- 493
- 494 Barker, F. K., Cibois, A., Schikler, P., Feinstein, J., & Cracraft, J. (2004). Phylogeny
495 and diversification of the largest avian radiation. *Proceedings of the National
496 Academy of Sciences*, 101(30), 11040–11045.
- 497 Beresford, P., Barker, F., Ryan, Pg., & Crowe, T. (2005). African endemics span the
498 tree of songbirds (passeri): Molecular systematics of several evolutionary ‘enigmas.’
499 *Proceedings of the Royal Society B: Biological Sciences*, 272(1565), 849–858.
- 500 Boyle, B., Hopkins, N., Lu, Z., Raygoza Garay, J. A., Mozzherin, D., Rees, T., ...
501 Enquist, B. J. (2013). The taxonomic name resolution service: An online tool for
502 automated standardization of plant names. *BMC Bioinformatics*, 14(1).
503 <https://doi.org/10.1186/1471-2105-14-16>
- 504 Britton, T., Anderson, C. L., Jacquet, D., Lundqvist, S., & Bremer, K. (2007).
505 Estimating Divergence Times in Large Phylogenetic Trees. *Systematic Biology*,
506 56(788777878), 741–752. <https://doi.org/10.1080/10635150701613783>
- 507 Bryson Jr, R. W., Chaves, J., Smith, B. T., Miller, M. J., Winker, K., Pérez-Emán, J.
508 L., & Klicka, J. (2014). Diversification across the new world within the
509 ‘blue’cardinalids (aves: cardinalidae). *Journal of Biogeography*, 41(3), 587–599.
- 510 Burleigh, J. G., Kimball, R. T., & Braun, E. L. (2015). Building the avian tree of life
511 using a large-scale, sparse supermatrix. *Molecular Phylogenetics and Evolution*,
512 84, 53–63.
- 513 Burns, K. J., Shultz, A. J., Title, P. O., Mason, N. A., Barker, F. K., Klicka, J., ...
514 Lovette, I. J. (2014). Phylogenetics and diversification of tanagers (passeriformes:
515 Thraupidae), the largest radiation of neotropical songbirds. *Molecular Phylogenetics and Evolution*, 75, 41–77.
- 516

- 517 Chamberlain, S. A., & Szöcs, E. (2013). taxize : taxonomic search and retrieval in R
518 [version 2; referees: 3 approved]. *F1000Research*, 2(191), 1–29.
519 <https://doi.org/10.12688/f1000research.2-191.v2>
- 520 Chamberlain, S. A., Szöcs, E., Foster, Z., Arendsee, Z., Boettiger, C., Ram, K., ...
521 Li, G. (2019). *taxize: Taxonomic information from around the web*. Retrieved
522 from <https://github.com/ropensci/taxize>
- 523 Chaves, J. A., Hidalgo, J. R., & Klicka, J. (2013). Biogeography and evolutionary
524 history of the n eotropical genus s altator (a ves: T hraupini). *Journal of*
525 *Biogeography*, 40(11), 2180–2190.
- 526 Claramunt, S., & Cracraft, J. (2015). A new time tree reveals earth history's imprint
527 on the evolution of modern birds. *Science Advances*, 1(11), e1501005.
- 528 Criscuolo, A., Berry, V., Douzery, E. J. P., & Gascuel, O. (2006). SDM: A fast
529 distance-based approach for (super)tree building in phylogenomics. *Systematic*
530 *Biology*, 55(5), 740–755. <https://doi.org/10.1080/10635150600969872>
- 531 Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M.-K., Turon, X., ...
532 Douzery, E. J. (2018). A phylogenomic framework and timescale for comparative
533 studies of tunicates. *BMC Biology*, 16(1), 1–14.
- 534 Eastman, J. M., Harmon, L. J., & Tank, D. C. (2013). Congruification: Support for
535 time scaling large phylogenetic trees. *Methods in Ecology and Evolution*, 4(7),
536 688–691. <https://doi.org/10.1111/2041-210X.12051>
- 537 Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and
538 high throughput. *Nucleic Acids Research*, 32(5), 1792–1797.
- 539 Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American*
540 *Naturalist*, 125(1), 1–15. Retrieved from <http://www.jstor.org/stable/2461605>
- 541 Gibb, G. C., England, R., Hartig, G., McLenachan, P. A., Taylor Smith, B. L.,
542 McComish, B. J., ... Penny, D. (2015). New zealand passerines help clarify the
543 diversification of major songbird lineages during the oligocene. *Genome Biology*

- 544 and *Evolution*, 7(11), 2983–2995.
- 545 Graur, D., & Martin, W. (2004). Reading the entrails of chickens: Molecular
546 timescales of evolution and the illusion of precision. *TRENDS in Genetics*, 20(2),
547 80–86.
- 548 Hackett, S. J., Kimball, R. T., Reddy, S., Bowie, R. C., Braun, E. L., Braun, M. J.,
549 ... others. (2008). A phylogenomic study of birds reveals their evolutionary
550 history. *Science*, 320(5884), 1763–1768.
- 551 Hedges, S. B., Marin, J., Suleski, M., Paymer, M., & Kumar, S. (2015). Tree of life
552 reveals clock-like speciation and diversification. *Molecular Biology and Evolution*,
553 32(4), 835–845. <https://doi.org/10.1093/molbev/msv037>
- 554 Heibl, C. (2008). *PHYLOCH: R language tree plotting tools and interfaces to diverse*
555 *phylogenetic software packages*. Retrieved from
556 <http://www.christophheibl.de/Rpackages.html>
- 557 Hooper, D. M., & Price, T. D. (2017). Chromosomal inversion differences correlate
558 with range overlap in passerine birds. *Nature Ecology & Evolution*, 1(10), 1526.
- 559 Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of
560 phylogenetic trees. *Bioinformatics*, 17(8), 754–755.
561 <https://doi.org/10.1093/bioinformatics/17.8.754>
- 562 Jetz, W., Thomas, G., Joy, J. J. B., Hartmann, K., & Mooers, A. (2012). The global
563 diversity of birds in space and time. *Nature*, 491(7424), 444–448.
564 <https://doi.org/10.1038/nature11631>
- 565 Johansson, U. S., Fjeldså, J., & Bowie, R. C. (2008). Phylogenetic relationships
566 within passerida (aves: Passeriformes): A review and a new molecular phylogeny
567 based on three nuclear intron markers. *Molecular Phylogenetics and Evolution*,
568 48(3), 858–876.
- 569 Katoh, K., Asimenos, G., & Toh, H. (2009). Multiple alignment of DNA sequences
570 with MAFFT. In *Bioinformatics for DNA sequence analysis* (pp. 39–64).

- 571 Springer.
- 572 Kimball, R. T., Oliveros, C. H., Wang, N., White, N. D., Barker, F. K., Field, D. J.,
573 . . . others. (2019). A phylogenomic supertree of birds. *Diversity*, 11(7), 109.
- 574 Klicka, J., Barker, F. K., Burns, K. J., Lanyon, S. M., Lovette, I. J., Chaves, J. A., &
575 Bryson Jr, R. W. (2014). A comprehensive multilocus assessment of sparrow (aves:
576 Passerellidae) relationships. *Molecular Phylogenetics and Evolution*, 77, 177–182.
- 577 Ksepka, D. T., Parham, J. F., Allman, J. F., Benton, M. J., Carrano, M. T.,
578 Cranston, K. A., . . . others. (2015). The fossil calibration database—a new
579 resource for divergence dating. *Systematic Biology*, 64(5), 853–859.
- 580 Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M.,
581 Martinez-Barrio, A., . . . others. (2015). Evolution of darwin’s finches and their
582 beaks revealed by genome sequencing. *Nature*, 518(7539), 371–375.
- 583 Laubichler, M. D., & Maienschein, J. (2009). *Form and function in developmental
584 evolution*. Cambridge University Press.
- 585 Lepage, D. (2004). *Avibase: The world bird database*. Bird Studies Canada.
- 586 Lepage, D., Vaidya, G., & Guralnick, R. (2014). Avibase—a database system for
587 managing and organizing taxonomic concepts. *ZooKeys*, (420), 117.
- 588 Lerner, H. R., Meyer, M., James, H. F., Hofreiter, M., & Fleischer, R. C. (2011).
589 Multilocus resolution of phylogeny and timescale in the extant adaptive radiation
590 of hawaiian honeycreepers. *Current Biology*, 21(21), 1838–1844.
- 591 Lovette, I. J., Pérez-Emán, J. L., Sullivan, J. P., Banks, R. C., Fiorentino, I.,
592 Córdoba-Córdoba, S., . . . others. (2010). A comprehensive multilocus phylogeny
593 for the wood-warblers and a revised classification of the parulidae (aves).
594 *Molecular Phylogenetics and Evolution*, 57(2), 753–770.
- 595 Magallon, S., & Sanderson, M. J. (2001). Absolute diversification rates in angiosperm
596 clades. *Evolution*, 55(9), 1762–1780.
- 597 Magallón, S. (2010). Using fossils to break long branches in molecular dating: A

- 598 comparison of relaxed clocks applied to the origin of angiosperms. *Systematic*
599 *Biology*, 59(4), 384–399.
- 600 Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., & Hernández-Hernández, T.
601 (2015). A metacalibrated time-tree documents the early rise of flowering plant
602 phylogenetic diversity. *New Phytologist*, 207(2), 437–453.
- 603 McTavish, E. J., Hinchliff, C. E., Allman, J. F., Brown, J. W., Cranston, K. A.,
604 Holder, M. T., ... Smith, S. A. (2015). Phylesystem: A git-based data store for
605 community-curated phylogenetic estimates. *Bioinformatics*, 31(17), 2794–2800.
- 606 Michonneau, F., Brown, J. W., & Winter, D. J. (2016). rotl: an R package to interact
607 with the Open Tree of Life data. *Methods in Ecology and Evolution*, 7(12),
608 1476–1481. <https://doi.org/10.1111/2041-210X.12593>
- 609 Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecology*
610 *Letters*, 17(4), 508–525. <https://doi.org/10.1111/ele.12251>
- 611 Moyle, R. G., Oliveros, C. H., Andersen, M. J., Hosner, P. A., Benz, B. W., Manthey,
612 J. D., ... Faircloth, B. C. (2016). Tectonic collision and uplift of wallacea
613 triggered the global songbird radiation. *Nature Communications*, 7(1), 1–7.
- 614 Ödeen, A., Håstad, O., & Alström, P. (2011). Evolution of ultraviolet vision in the
615 largest avian radiation—the passerines. *BMC Evolutionary Biology*, 11(1), 1–8.
- 616 Oliveros, C. H., Field, D. J., Ksepka, D. T., Barker, F. K., Aleixo, A., Andersen, M.
617 J., ... others. (2019). Earth history and the passerine superradiation.
618 *Proceedings of the National Academy of Sciences*, 116(16), 7916–7925.
- 619 Ooms, J., & Chamberlain, S. (2018). *Phylocomr: Interface to 'phylocom'*. Retrieved
620 from <https://CRAN.R-project.org/package=phylocomr>
- 621 Open Tree Of Life, Redelings, B., Cranston, K. A., Allman, J., Holder, M. T., &
622 McTavish, E. J. (2016). Open Tree of Life APIs v3.0. *Open Tree of Life Project*,
623 (Online Resources). Retrieved from
624 <https://github.com/OpenTreeOfLife/germinator/wiki/Open-Tree-of-Life->

- 625 Web-APIs%7D
- 626 Open Tree Of Life, Redelings, B., Sánchez Reyes, L. L., Cranston, K. A., Allman, J.,
627 Holder, M. T., & McTavish, E. J. (2019). Open tree of life synthetic tree v12.3.
628 *Zenodo*. Retrieved from <https://doi.org/10.5281/zenodo.3937742>
- 629 Päckert, M., Martens, J., Sun, Y.-H., Severinghaus, L. L., Nazarenko, A. A., Ting, J.,
630 ... Tietze, D. T. (2012). Horizontal and elevational phylogeographic patterns of
631 himalayan and southeast asian forest passerines (aves: passeriformes). *Journal of
632 Biogeography*, 39(3), 556–573.
- 633 Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and
634 evolution in R language. *Bioinformatics*, 20(2), 289–290.
- 635 Parchman, T. L., Benkman, C. W., & Mezquida, E. T. (2007). Coevolution between
636 hispaniolan crossbills and pine: Does more time allow for greater phenotypic
637 escalation at lower latitude? *Evolution*, 61(9), 2142–2153.
- 638 Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., FitzJohn,
639 R. G., ... Harmon, L. J. (2014). Geiger v2. 0: An expanded suite of methods for
640 fitting macroevolutionary models to phylogenetic trees. *Bioinformatics*, 30(15),
641 2216–2218.
- 642 Posadas, P., Crisci, J. V., & Katinas, L. (2006). Historical biogeography: A review of
643 its basic concepts and critical issues. *Journal of Arid Environments*, 66(3),
644 389–403.
- 645 Powell, A. F., Barker, F. K., Lanyon, S. M., Burns, K. J., Klicka, J., & Lovette, I. J.
646 (2014). A comprehensive species-level molecular phylogeny of the new world
647 blackbirds (icteridae). *Molecular Phylogenetics and Evolution*, 71, 94–112.
- 648 Price, T. D., Hooper, D. M., Buchanan, C. D., Johansson, U. S., Tietze, D. T.,
649 Alström, P., ... others. (2014). Niche filling slows the diversification of himalayan
650 songbirds. *Nature*, 509(7499), 222.
- 651 Pulgarín-R, P. C., Smith, B. T., Bryson Jr, R. W., Spellman, G. M., & Klicka, J.

- 652 (2013). Multilocus phylogeny and biogeography of the new world pheucticus
653 grosbeaks (aves: cardinalidae). *Molecular Phylogenetics and Evolution*, 69(3),
654 1222–1227.
- 655 R Core Team. (2018). *R: a language and environment for statistical computing*.
656 Vienna, Austria: R Foundation for Statistical Computing.
- 657 Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., ...
658 others. (2018). An inverse latitudinal gradient in speciation rate for marine fishes.
659 *Nature*, 559(7714), 392.
- 660 Ramshaw, J., Richardson, D., Meatyard, B., Brown, R., Richardson, M., Thompson,
661 E., & Boulter, D. (1972). The time of origin of the flowering plants determined by
662 using amino acid sequence data of cytochrome c. *New Phytologist*, 71(5), 773–779.
- 663 Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The barcode of life data system
664 (<http://www.Barcodinglife.org>). *Molecular Ecology Notes*, 7(3), 355–364.
- 665 Rees, J. A., & Cranston, K. (2017). Automated assembly of a reference taxonomy for
666 phylogenetic data synthesis. *Biodiversity Data Journal*, (5).
- 667 Revell, L. J. (2012). Phytools: An r package for phylogenetic comparative biology
668 (and other things). *Methods in Ecology and Evolution*, 3, 217–223.
- 669 Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic
670 inference under mixed models. *Bioinformatics*, 19(12), 1572–1574.
671 <https://doi.org/10.1093/bioinformatics/btg180>
- 672 Sanchez-Reyes, L. L., & O'Meara, B. (2022). datelifeplot: Methods to plot
673 chronograms and outputs of the **datelife** package. *R Package Release V0.2.2*.
674 Retrieved from <https://zenodo.org/badge/latestdoi/381501451>
- 675 Sanchez-Reyes, L. L., O'Meara, B., Eastman, J., Heath, T., Wright, A., Schliep, K.,
676 ... Alfaro, M. (2022). **datelife**: Scientific Data on Time of Lineage Divergence
677 for Your Taxa. *R Package Release V0.6.2*. Retrieved from
678 <https://doi.org/10.5281/zenodo.593938>

- 679 Sanderson, Michael J. (2002). Estimating Absolute Rates of Molecular Evolution and
680 Divergence Times: A Penalized Likelihood Approach. *Molecular Biology and*
681 *Evolution*, 19(1), 101–109.
682 <https://doi.org/10.1093/oxfordjournals.molbev.a003974>
- 683 Sanderson, Michael J. (2003). r8s: Inferring absolute rates of molecular evolution and
684 divergence times in the absence of a molecular clock. *Bioinformatics*, 19(2),
685 301–302.
- 686 Sanderson, Michael J., & Doyle, J. A. (2001). Sources of error and confidence
687 intervals in estimating the age of angiosperms from rbcL and 18S rDNA data.
688 *American Journal of Botany*, 88(8), 1499–1516.
- 689 Sauquet, H. (2013). A practical guide to molecular dating. *Comptes Rendus Palevol*,
690 12(6), 355–367.
- 691 Sauquet, H., Ramírez-Barahona, S., & Magallón, S. (2021). *The age of flowering*
692 *plants is unknown*.
- 693 Schenk, J. J. (2016). Consequences of secondary calibrations on divergence time
694 estimates. *PLoS ONE*, 11(1). <https://doi.org/10.1371/journal.pone.0148228>
- 695 Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in r. *Bioinformatics*, 27(4),
696 592–593.
- 697 Selvatti, A. P., Gonzaga, L. P., & Moraes Russo, C. A. de. (2015). A paleogene origin
698 for crown passerines and the diversification of the oscines in the new world.
699 *Molecular Phylogenetics and Evolution*, 88, 1–15.
- 700 Smith, Stephen A., & Brown, J. W. (2018). Constructing a broadly inclusive seed
701 plant phylogeny. *American Journal of Botany*, 105(3), 302–314.
- 702 Smith, Stephen A., & O'Meara, B. C. (2012). TreePL: Divergence time estimation
703 using penalized likelihood for large phylogenies. *Bioinformatics*, 28(20),
704 2689–2690. <https://doi.org/10.1093/bioinformatics/bts492>
- 705 Stoltzfus, A., Lapp, H., Matasci, N., Deus, H., Sidlauskas, B., Zmasek, C. M., ...

- 706 Jordan, G. (2013). Phylotastic! Making tree-of-life knowledge accessible, reusable
707 and convenient. *BMC Bioinformatics*, 14.
708 <https://doi.org/10.1186/1471-2105-14-158>
- 709 Tietze, D. T., Päckert, M., Martens, J., Lehmann, H., & Sun, Y.-H. (2013). Complete
710 phylogeny and historical biogeography of true rosefinches (aves: carpodacus).
711 *Zoological Journal of the Linnean Society*, 169(1), 215–234.
- 712 Title, P. O., & Rabosky, D. L. (2016). Do Macrophylogenies Yield Stable
713 Macroevolutionary Inferences? An Example from Squamate Reptiles. *Systematic
714 Biology*, syw102. <https://doi.org/10.1093/sysbio/syw102>
- 715 Treplin, S., Siegert, R., Bleidorn, C., Thompson, H. S., Fotso, R., & Tiedemann, R.
716 (2008). Molecular phylogeny of songbirds (aves: Passeriformes) and the relative
717 utility of common nuclear marker loci. *Cladistics*, 24(3), 328–349.
- 718 Vos, R. A., Balhoff, J. P., Caravas, J. A., Holder, M. T., Lapp, H., Maddison, W. P.,
719 ... others. (2012). NeXML: Rich, extensible, and verifiable representation of
720 comparative data and metadata. *Systematic Biology*, 61(4), 675–689.
- 721 Webb, C. (2000). Exploring the Phylogenetic Structure of Ecological Communities :
722 An Example for Rain Forest Trees. *The American Naturalist*, 156(2), 145–155.
- 723 Webb, Campbell O., Ackerly, D. D., & Kembel, S. W. (2008). Phylocom: Software for
724 the analysis of phylogenetic community structure and trait evolution.
725 *Bioinformatics*, 24(18), 2098–2100.
726 <https://doi.org/10.1093/bioinformatics/btn358>
- 727 Webb, Campbell O., & Donoghue, M. J. (2005). Phylomatic: Tree assembly for
728 applied phylogenetics. *Molecular Ecology Notes*, 5(1), 181–183.
- 729 Weir, J., & Schluter, D. (2008). Calibrating the avian molecular clock. *Molecular
730 Ecology*, 17(10), 2321–2328.

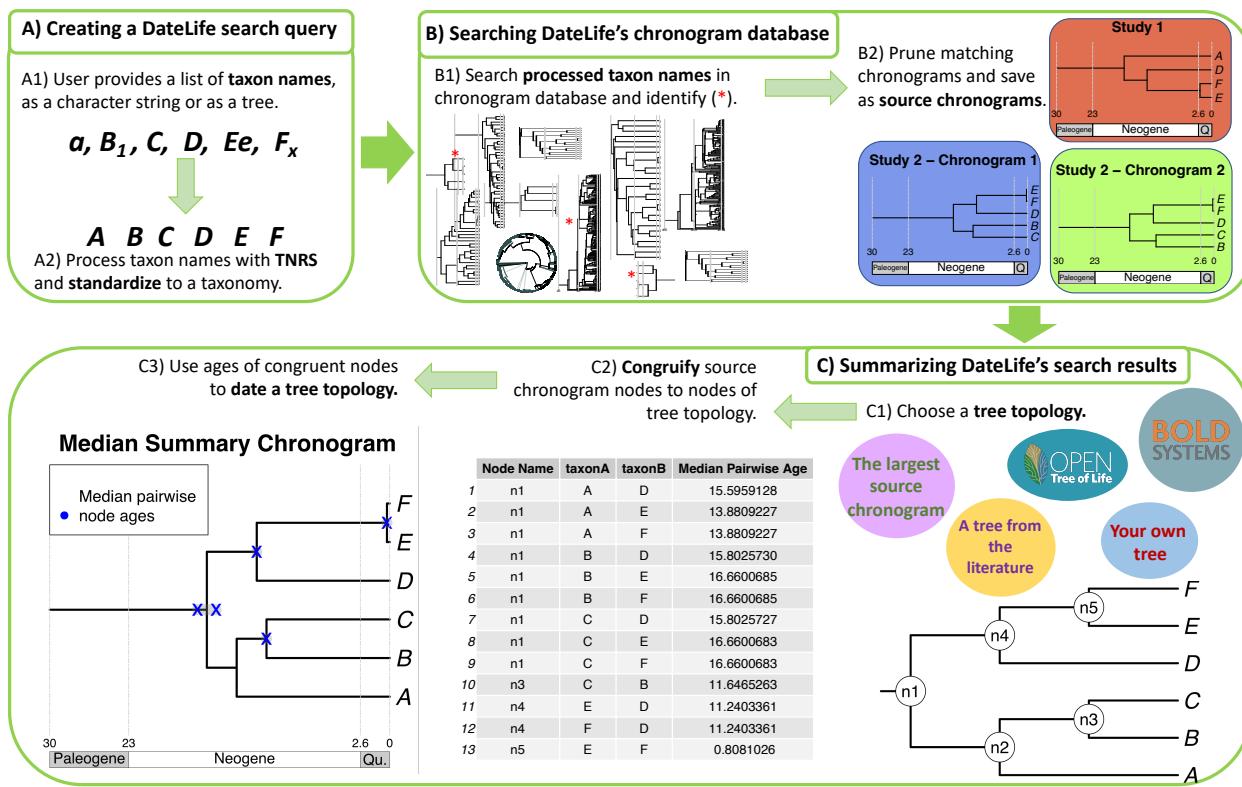


FIGURE 1. Stylized DateLife workflow. This shows the general workflows and analyses that can be performed with `datelife`, via the R package or through the website at <http://www.datelife.org/>. Details on the functions involved on each workflow are shown in `datelife`'s R package vignette.

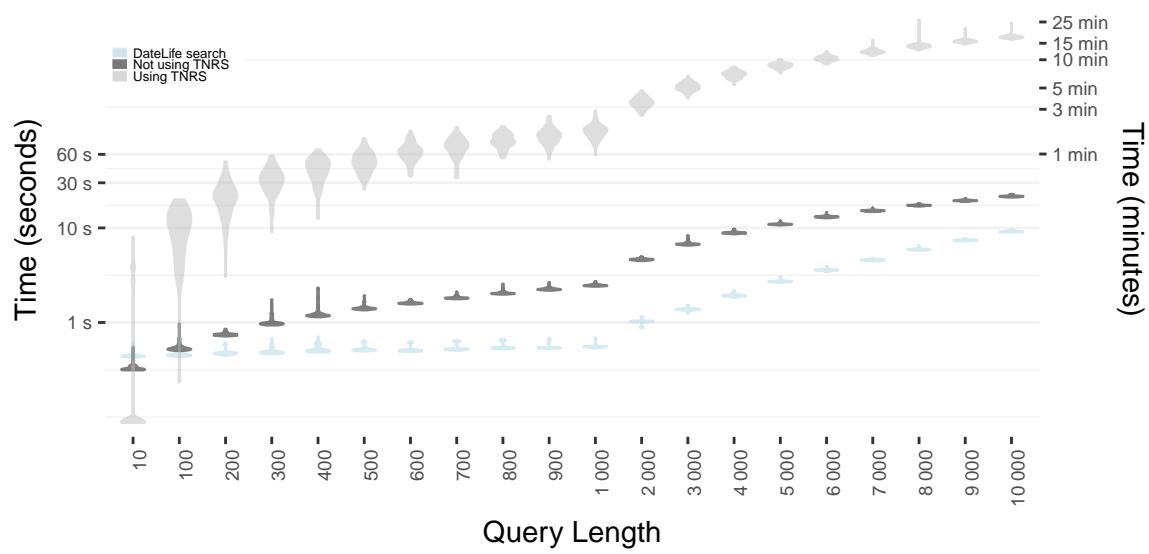


FIGURE 2. Computation time of query processing and search across **datelife**'s chronogram database relative to number of input taxon names. We sampled N names from the class Aves for each cohort 100 times and then performed a search with query processing not using the Taxon Names Resolution Service (TNRS; dark gray), and using TNRS (light gray). We also performed a search using the already processed query for comparison (light blue).

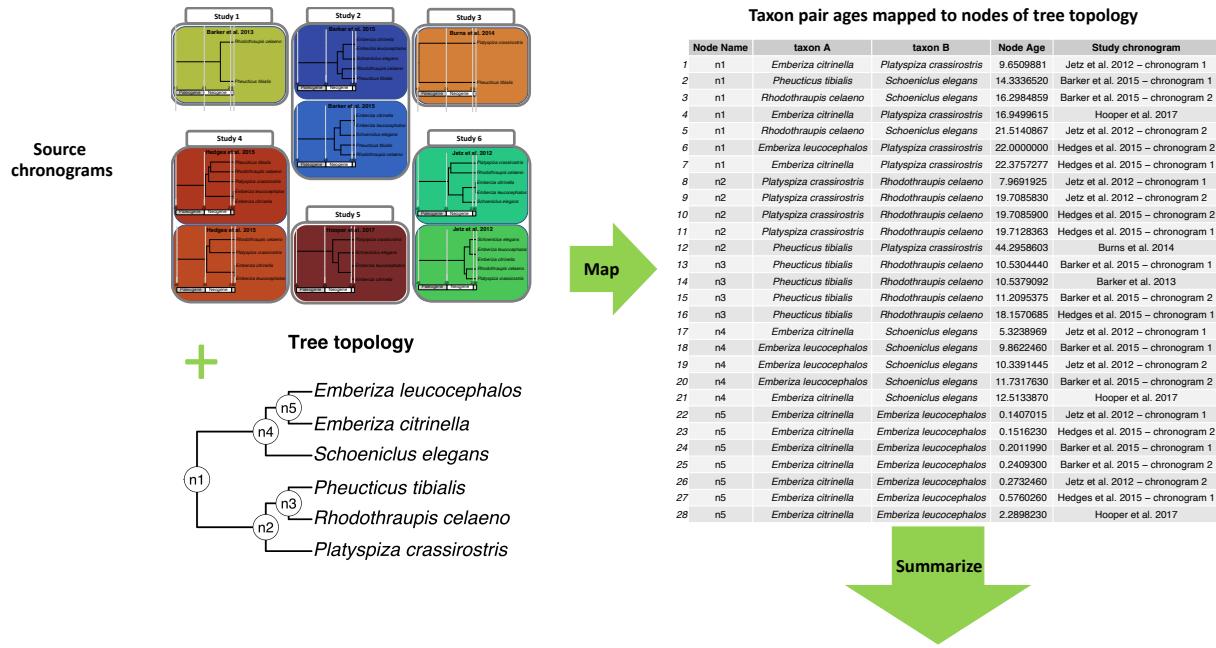


FIGURE 3. Age data results of a DateLife search of a small sample of 6 bird species within the Passeriformes. Input names were found across 9 chronograms within 6 independent studies (Barker et al. (2012), Barker et al. (2015), Burns et al. (2014), Hedges et al. (2015), Hooper and Price (2017), Jetz et al. (2012).) This revealed 28 age data points for the queried species names.

Summary of mapped taxon pair age data

Node Name	taxon A	taxon B	Pairwise Median Age	Node Median Age
1	<i>Pheucticus tibialis</i>	<i>Emberiza citrinella</i>	16.298486	
2	<i>Pheucticus tibialis</i>	<i>Emberiza leucocephalos</i>	16.298486	
3	<i>Platyspiza crassirostris</i>	<i>Emberiza citrinella</i>	21.514085	
4	<i>Platyspiza crassirostris</i>	<i>Emberiza leucocephalos</i>	21.514085	
5 n1	<i>Rhodothraupis celaeno</i>	<i>Emberiza citrinella</i>	20.408031	19.301977
6	<i>Rhodothraupis celaeno</i>	<i>Emberiza leucocephalos</i>	20.408031	
7	<i>Schoeniclus elegans</i>	<i>Pheucticus tibialis</i>	15.316069	
8	<i>Schoeniclus elegans</i>	<i>Platyspiza crassirostris</i>	19.301977	
9	<i>Schoeniclus elegans</i>	<i>Rhodothraupis celaeno</i>	17.800231	
10 n2	<i>Platyspiza crassirostris</i>	<i>Pheucticus tibialis</i>	32.004348	25.856467327225
11	<i>Rhodothraupis celaeno</i>	<i>Platyspiza crassirostris</i>	19.708587	
12 n3	<i>Rhodothraupis celaeno</i>	<i>Pheucticus tibialis</i>	10.873723	10.87372335475
13 n4	<i>Schoeniclus elegans</i>	<i>Emberiza citrinella</i>	10.647794	10.6477935
14	<i>Schoeniclus elegans</i>	<i>Emberiza leucocephalos</i>	10.647794	
15 n5	<i>Emberiza leucocephalos</i>	<i>Emberiza citrinella</i>	0.273246	0.273246

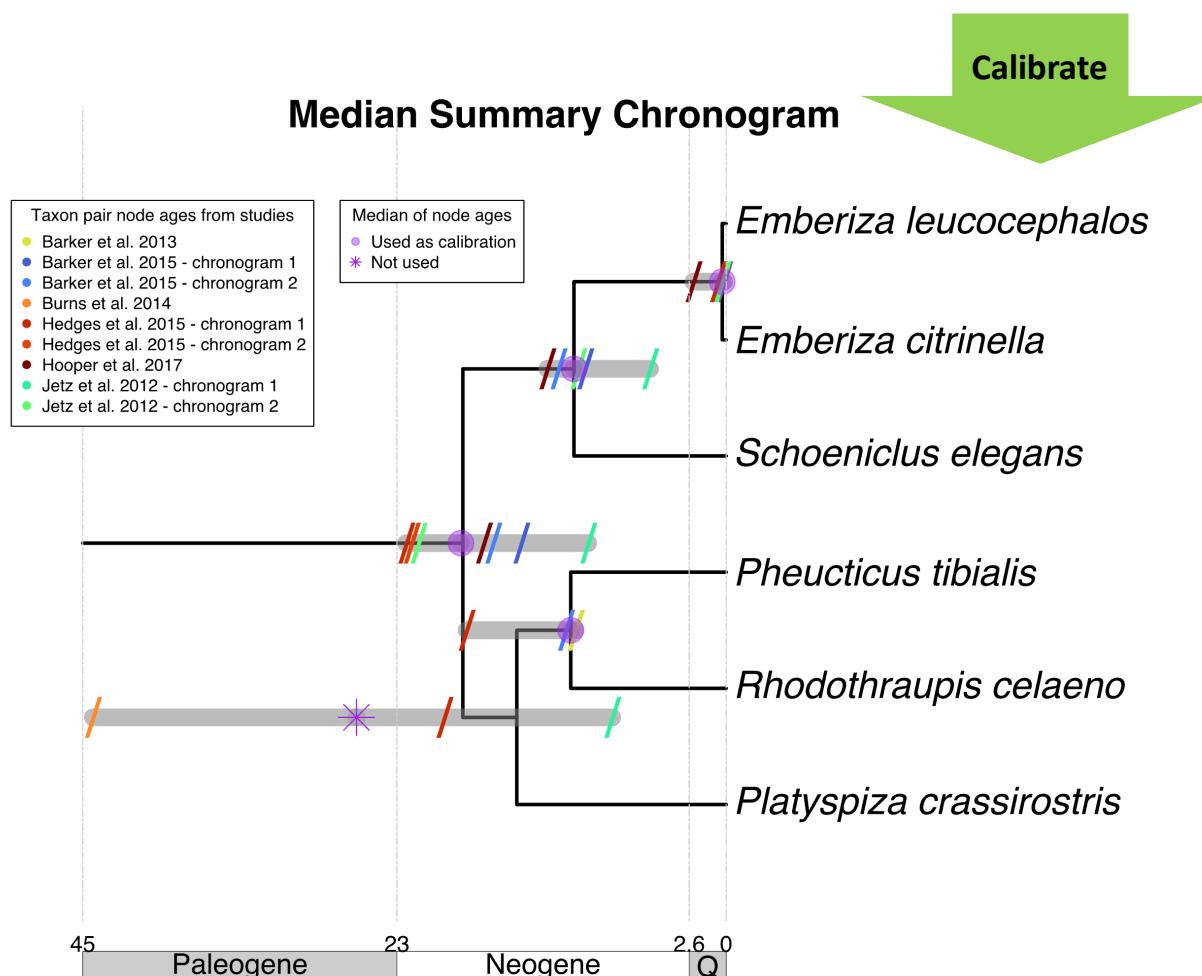


FIGURE 4. Summarized age data is used as secondary calibrations to date a tree topology as a summary chronogram.

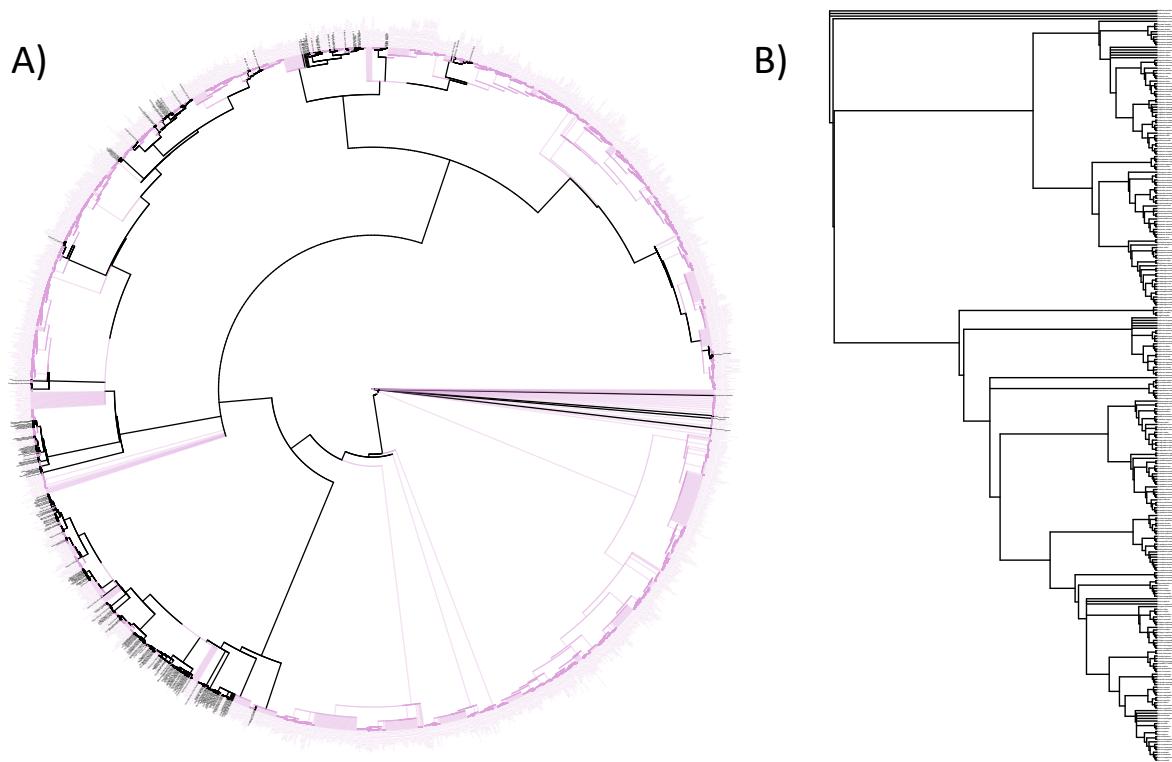


FIGURE 5. Tree topologies obtained from Open Tree of Life's (OpenTree) synthetic phylogenetic tree. A) Topology of 2333 tips and 1305 internal nodes, encompassing bird species within the family Fringillidae following the NCBI taxonomy (black), as well as all other bird species that share the same mrca node in OpenTree's synthetic tree (purple). B) Topology of 289 tips and 253 internal nodes, encompassing bird species within the Fringillidae only, resulting from pruning purple branches from topology A. Bird species within the Fringillidae do not form a monophyletic group. (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)

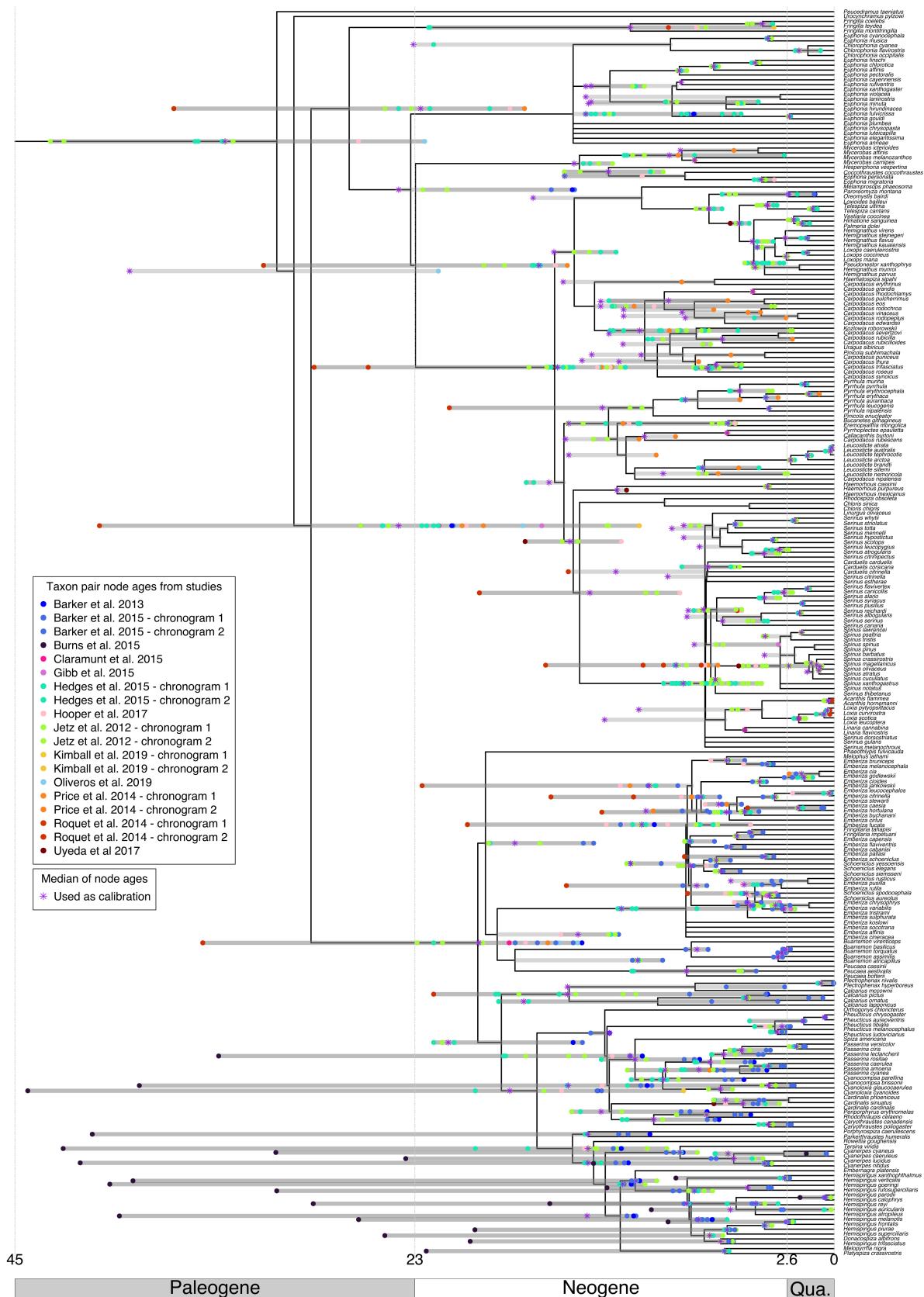


FIGURE 6. Fringillidae median summary chronogram generated with DateLife. It has 256 tips and 233 nodes.

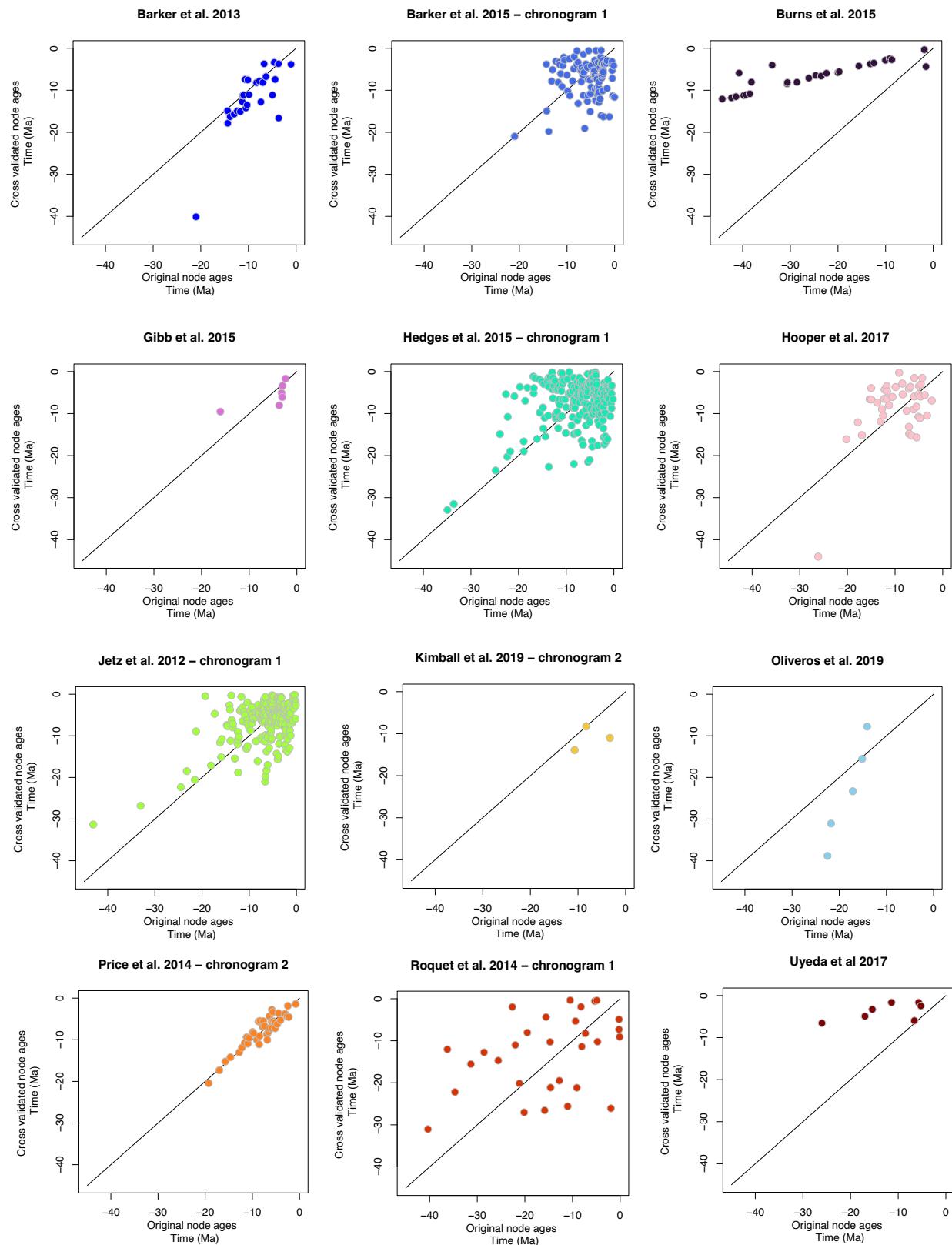


FIGURE 7. Results from cross validation analysis.

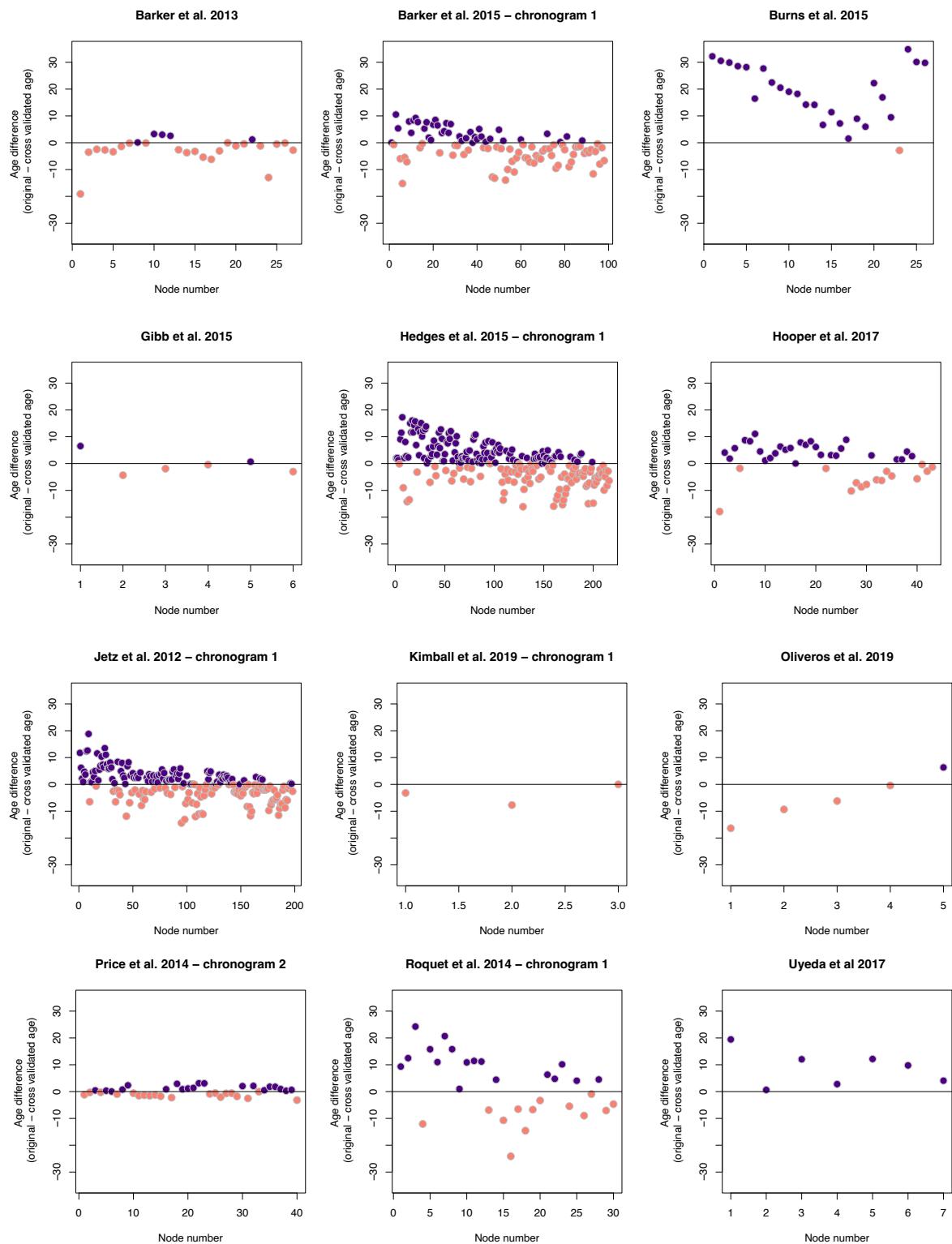


FIGURE 8. Results from cross validation analysis.

Barker et al. 2015 - chronogram 1



FIGURE 9. Cross validation of second source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to

Barker et al. 2015 - chronogram 2



FIGURE 10. Cross validation of third source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to

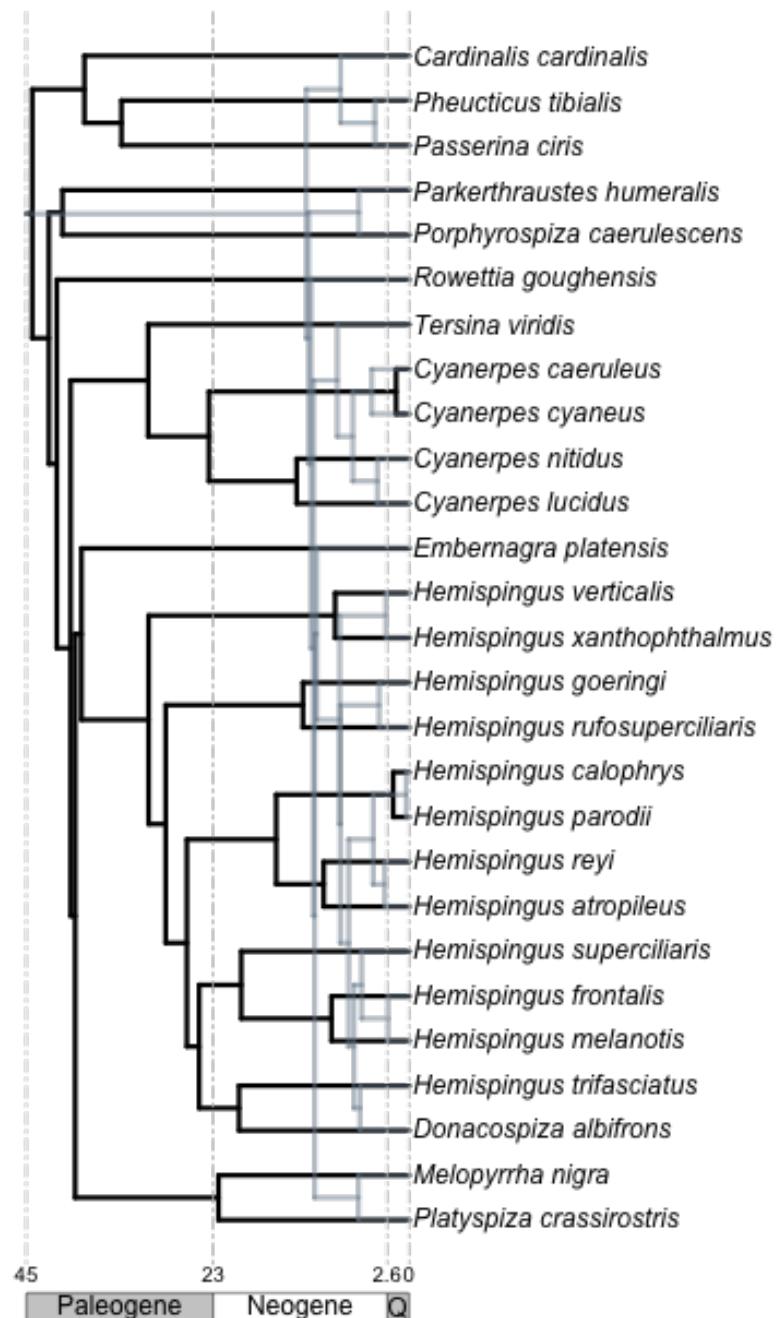
Burns et al. 2015

FIGURE 11. Cross validation of fourth source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the dates calculated with BLADJ, i.e., the same fossil dates used.

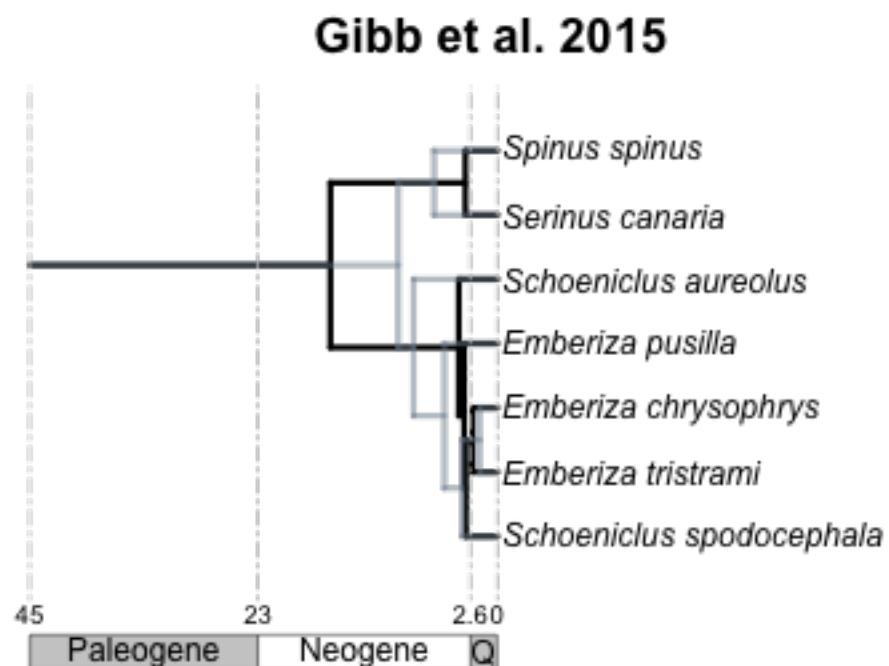


FIGURE 12. Cross validation of sixth source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the same tree topology dated with BLADJ using node ages from all other source chronograms as secondary calibrations.

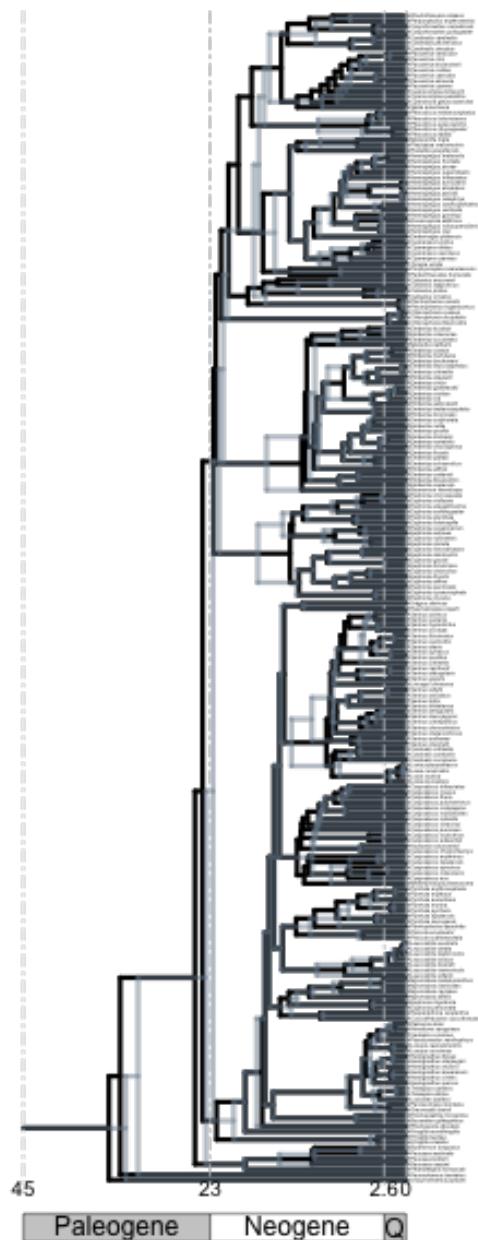
Hedges et al. 2015 - chronogram 1

FIGURE 13. Cross validation of seventh source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the dates calculated with BLADe. In each case, the tree is the same.

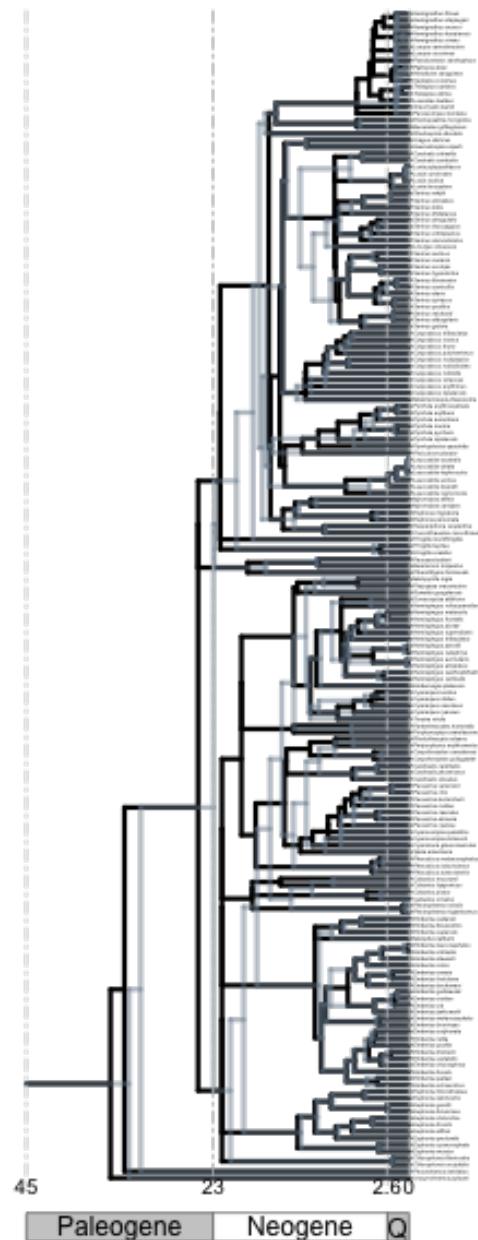
Hedges et al. 2015 - chronogram 2

FIGURE 14. Cross validation of eight source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the dates calculated with BLADe. In each case, the tree is the same.

Hooper et al. 2017

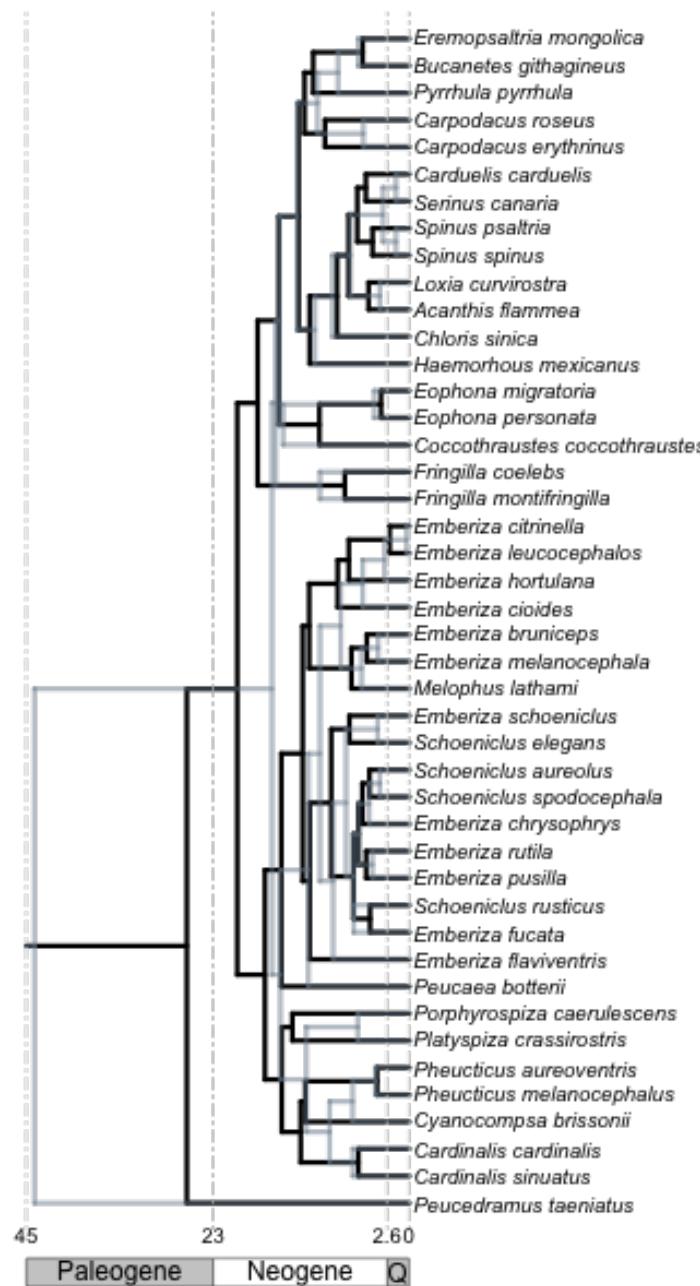


FIGURE 15. Cross validation of ninth source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the dates calculated with BLADJ, i.e., the ones from the ninth source chronogram.

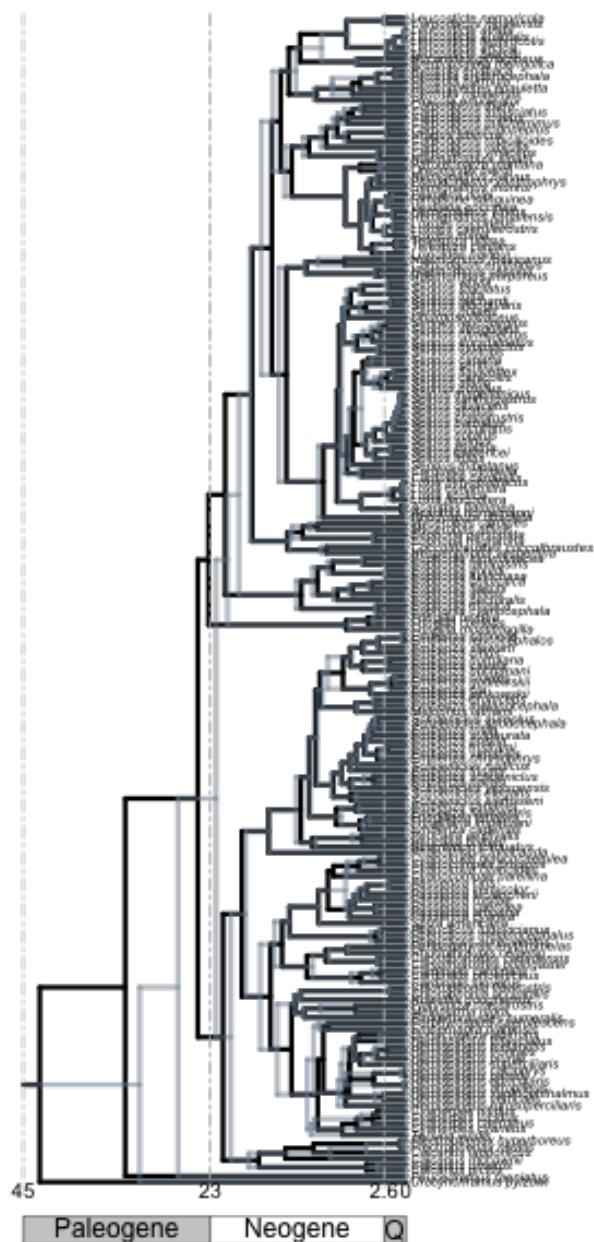
Jetz et al. 2012 - chronogram 1

FIGURE 16. Cross validation of tenth source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the dates calculated with BLADe. In each case, the tree is identical.