

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,
9 425 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 Date estimates for times of evolutionary divergences are key data for research in the natural
19 sciences. These estimates also provide valuable information for education, science
20 communication and policy decisions. Although achieving a high-quality reconstruction of a
21 phylogenetic tree with branch lengths proportional to absolute time (chronogram), is a
22 difficult and time-consuming task, the increased availability of fossil and molecular data, and
23 time-efficient analytical techniques has resulted in many recent publications of large
24 chronograms for a large number and wide diversity of organisms. When these estimates are
25 shared in public, open databases this wealth of expertly-curated and peer-reviewed data on
26 time of evolutionary origin is exposed in a programmatic and reusable way. Intensive and
27 localized efforts have improved data sharing practices, as well as incentivized open science
28 in biology. Here we present DateLife, a service implemented as an R package and an Rshiny
29 website application available at www.datelife.org/query/, that provides functionalities for
30 efficient and easy finding, summary, reuse, and reanalysis of expert, peer-reviewed, public
31 data on time of evolutionary origin. The main DateLife workflow constructs a chronogram
32 for any given combination of taxon names, by searching a local chronogram database
33 constructed and curated from the Open Tree of Life Phylesystem phylogenetic database,
34 which incorporates phylogenetic data from TreeBASE database as well. We implement and
35 test methods for summarizing time data from multiple source chronograms using supertree
36 and congruification algorithms, and using age data extracted from source chronograms as
37 secondary calibration points to add branch lengths proportional to absolute time to a tree
38 topology. DateLife will be useful to increase awareness on the existing variation in expert
39 time of divergence data, and can foster exploration of the effect of alternative divergence
40 time hypothesis on the results of analyses, providing a framework for a more informed
41 interpretation of evolutionary results.

42

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

⁴³ Congruification; Supertree; Calibrations; Secondary calibrations

⁴⁴ Word count: 4379

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 for the study of natural processes in many areas of biological research, such as developmental
49 biology (Delsuc et al., 2018; Laubichler & Maienschein, 2009), conservation biology
50 (Felsenstein, 1985; C. Webb, 2000), historical biogeography (Posadas, Crisci, & Katinas,
51 2006), and species diversification (Magallon & Sanderson, 2001; Morlon, 2014).

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

58 Here we present the DateLife software application, available as an R package and as an
59 online Rshiny interactive website at www.datelife.org/query/, which captures data from
60 published chronograms, and make these data readily accessible to users. DateLife features a
61 versioned, open and fully public chronogram database (McTavish et al., 2015) storing age
62 information in a computer readable format (Vos et al., 2012), an automated and
63 programmatic way of accessing the data (Stoltzfus et al., 2013) and methods to summarize
64 and compare age data.

65 **Description**

66 The DateLife algorithm is fully implemented using the R language. The latest stable
67 version of the R package **datelife** is available from the CRAN repository (v0.6.2;
68 Sanchez-Reyes et al. (2022)), and relies on functionalities from various biological R packages:
69 **ape** (Paradis, Claude, & Strimmer, 2004), **bold** (Chamberlain et al., 2019), **geiger** (Harmon,

70 Weir, Brock, Glor, & Challenger, 2008), paleotree (Bapst, 2012), phyloch (Heibl, 2008),
71 phylocomr (Ooms & Chamberlain, 2018), phytools (Revell, 2012), rotl (Michonneau, Brown,
72 & Winter, 2016), and taxize (Chamberlain & Szöcs, 2013; Chamberlain et al., 2019). Figure
73 1 provides a graphical summary of the three main steps of the DateLife algorithm: providing
74 an input, searching a chronogram database, and summarizing results from the search.

75 **Providing an input**

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

81 DateLife accepts scientific names as input. These names can belong to any inclusive
82 taxonomic group (e.g., genus, family, tribe, etc.) or binomial specific. Subspecies and
83 variants are ignored. If an input taxon name belongs to an inclusive taxonomic group the
84 algorithm has two alternative behaviors defined by the “get species from taxon” flag. If the
85 flag is active, the DateLife algorithm retrieves all species names within the inclusive
86 taxonomic group and adds them to the input. If the flag is inactive, DateLife ignores the
87 inclusive taxon names from the input.

88 Input scientific names are processed using a Taxonomic Name Resolution Service
89 (TNRS), which increases the probability of correctly finding the queried taxon names in the
90 chronogram database. TNRS detects, corrects and standardizes name misspellings and typos,
91 variant spellings and authorities, and nomenclatural synonyms to a single taxonomic
92 standard. DateLife implements TNRS using OpenTree’s taxonomy as standard (Open Tree
93 Of Life et al., 2016; Rees & Cranston, 2017).

94 The processed input taxon names are saved as an R object of a newly defined class

95 `datelifeQuery` that is used in the following steps. This object contains the processed
96 names, the corresponding OpenTree taxonomic id numbers, and the topology of the input
97 tree if any was provided.

98 Searching the database

99 A DateLife search consists of matching processed taxon names to tip labels in a
100 chronogram database. Chronograms with at least two matching tip labels are identified and
101 pruned down to preserve only the matched tips. These matching pruned chronograms are
102 now referred to as source chronograms. Total distance (in units of millions of years) between
103 taxon pairs within each source chronogram are stored as a patristic distance matrix (Figure
104 1). The matrix format speeds up extraction of pairwise taxon ages of any queried taxa, as
105 opposed to searching the ancestor node of a pair of taxa in a “phylo” object or newick string.
106 Finally, the patristic matrices are associated to the study citation where the original
107 chronogram was published, and stored as an R object of the newly defined class
108 `datelifeResult`.

109 DateLife’s chronogram database latest version consist of 253 chronograms published in
110 187 different studies. It is constructed from OpenTree’s phylogenetic database, the
111 Phylesystem, which constitutes an open source of expert phylogenetic knowledge with rich
112 metadata (McTavish et al., 2015) that allows automatic and reproducible construction of a
113 chronogram database. New chronograms can be added to Phylesystem by any user and are
114 immediately publicly available. This permits an immediate update of DateLife’s chronogram
115 database to include those new data within a run.

116 Summarizing search results

117 Summary information is extracted from the `datelifeResult` object to inform
118 decisions for the subsequent steps in the user workflow. Basic summary information available
119 to the user is:

- 120 1. The matching pruned chronograms as newick strings or “phylo” objects.
- 121 2. The ages of the root of all source chronograms. This can correspond to the age of the
- 122 most recent common ancestor (mrca) of your group of interest if the source
- 123 chronograms have all taxa belonging to the group. If not, the root corresponds to the
- 124 mrca of a subgroup within your group of interest.
- 125 3. Study citations where original chronograms were published.
- 126 4. A report of input taxon names matches across source chronograms.
- 127 5. The source chronogram(s) with the most input taxon names.
- 128 6. Last but not least, age data from source chronograms are summarized and used to
- 129 generate a single summary chronogram:

130 ***Choosing a topology.***— DateLife requires a tree topology to summarize age data

131 upon. We recommend that users provide one as input from the literature, or one of their own
132 making. If no topology is provided, DateLife automatically subsets one from the OpenTree
133 synthetic tree (Open Tree Of Life et al., 2019). Alternatively, DateLife can reconstruct a tree
134 with branch lengths proportional to substitution rates from a starting tree topology using
135 genetic data from the Barcode of Life Data System [BOLD; Ratnasingham and Hebert
136 (2007)], or combine topologies from source chronograms using a supertree approach.

137 ***Reconstructing branch lengths.***— DateLife starts by mining the BOLD database to

138 obtain genetic markers for the input taxa, and aligning them with MUSCLE (Edgar, 2004;
139 or MAFFT Katoh, Asimenos, & Toh, 2009). Currently, branch length reconstruction is
140 performed with parsimony and the likelihood of the phylogenetic tree given a sequence
141 alignment is computed (Schliep, 2011).

142 ***Combining source chronograms.***— To combine topologies from source chronograms

143 into a single summary topology, the DateLife algorithm starts by identifying the source
144 chronograms that form a grove, roughly, a sufficiently overlapping set of taxa between trees,
145 by implementing definition 2.8 for n-overlap from Ané et al. (2009). In rare cases, a group of

¹⁴⁶ trees can have multiple groves. By default, DateLife chooses the grove with the most taxa,
¹⁴⁷ however, the “criterion = trees” flag allows the user to choose the grove with the most trees
¹⁴⁸ instead. If source chronograms do not form a grove, the supertree reconstruction will fail.

¹⁴⁹ **Congruifying nodes.**— Once with a chosen topology, DateLife applies the
¹⁵⁰ congruification method (Eastman, Harmon, & Tank, 2013) to find nodes belonging to the
¹⁵¹ same clade across source chronograms, and extract the corresponding node ages from the
¹⁵² patristic distance matrices stored as `datelifeResult`. Since by definition the matrices store
¹⁵³ total distance (time from tip to tip), node ages correspond to half the values stored in the
¹⁵⁴ patristic distance matrices. A table of congruified node ages that can be used as secondary
¹⁵⁵ calibrations is stored as a `congruifiedCalibrations` object.

¹⁵⁶ **Summarizing congruified ages.**— Congruified pairwise node ages are used to
¹⁵⁷ calculate summary statistics (mean, median, minimum and maximum ages) per taxon pair.
¹⁵⁸ into a single summary matrix using different methods. Summarizing options implemented
¹⁵⁹ include Super Distance Matrix method (SDM, Criscuolo, Berry, Douzery, & Gascuel, 2006)
¹⁶⁰ and .

¹⁶¹ Ages per taxon pair can be used as calibrations to date the chosen tree topology.

¹⁶² **Dating the tree topology.**— By default, DateLife implements the Branch Length
¹⁶³ Adjuster (BLADJ) algorithm to obtain a fully dated topology. BLADJ fixes node ages that
¹⁶⁴ have calibration data, and distributes time between nodes with no data evenly between
¹⁶⁵ nodes with calibration data. This minimizes age variance in the resulting chronogram
¹⁶⁶ (Campbell O. Webb, Ackerly, & Kembel, 2008). When there is conflict in ages between nodes
¹⁶⁷ with calibration data, the algorithm ignores ages that are older than ages of parent nodes
¹⁶⁸ and/or younger than ages from descendant nodes.

¹⁶⁹ If there is no information on the age of the root in the chronogram database, users can
¹⁷⁰ provide an estimate from the literature. If none is provided, DateLife assigns an arbitrary

171 age to the root as 10% older than the oldest age available within the group.

172 Summarized calibrations can be applied as secondary calibrations with different dating
173 methods currently supported within DateLife: MrBayes (Huelsenbeck & Ronquist, 2001;
174 Ronquist & Huelsenbeck, 2003), PATHd8 (Britton, Anderson, Jacquet, Lundqvist, &
175 Bremer, 2007), BLADJ (Campbell O. Webb et al., 2008; Campbell O. Webb & Donoghue,
176 2005), and treePL (Stephen A. Smith & O'Meara, 2012).

177 **Visualizing results.**— Finally, users can save all source and summary chronograms in
178 formats that permit reuse and reanalyses (newick and R “phylo” format), as well as view
179 and compare results graphically, or construct their own graphs using `datelife`’s chronogram
180 plot generation functions.

181 Benchmark

182 `datelife`’s code speed was tested on an Apple iMac with one 3.4 GHz Intel Core i5
183 processor. We registered variation in computing time of query processing and search through
184 the database relative to number of queried taxon names. Query processing time increases
185 roughly linearly with number of input taxon names, and increases considerably if Taxonomic
186 Name Resolution Service (TNRS) is activated. Up to ten thousand names can be processed
187 and searched in less than 30 minutes with the most time consuming settings. Once names
188 have been processed as described in methods, a name search through the chronogram
189 database can be performed in less than a minute, even with a very large number of taxon
190 names (Fig. 2). `datelife`’s code performance was evaluated with a set of unit tests designed
191 and implemented with the R package `testthat` (R Core Team, 2018) that were run both
192 locally with the `devtools` package (R Core Team, 2018), and on a public server –via GitHub,
193 using the continuous integration tool Travis CI (<https://travis-ci.org>). At present, unit tests
194 cover more than 40% of `datelife`’s code (<https://codecov.io/gh/phylotastic/datelife>).

195

Case study

196 We illustrate the DateLife algorithm using a group within the Passeriform birds
197 encompassing the family of true finches, Fringillidae and allies as case study. The first
198 example analyses 6 bird species and shows all steps of the algorithm. The second example is
199 a real life application

200 **Small example**

201 We randomly chose 6 bird species related to the family Fringillidae of true finches. The
202 sample includes two species of cardinals: the black-thighed grosbeak – *Pheucticus tibialis*
203 and the crimson-collared grosbeak – *Rhodothraupis celaeno*; three species of buntings: the
204 yellowhammer – *Emberiza citrinella*, the pine bunting – *Emberiza leucocephalos* and the
205 yellow-throated bunting – *Emberiza elegans*; and one species of tanager, the vegetarian finch –
206 *Platyspiza crassirostris*.

207 Processing input names found that *Emberiza elegans* is synonym for *Schoeniclus*
208 *elegans* in the default reference taxonomy [Open Tree of Life Taxonomy v3.3, June 1, 2021].
209 For a detailed discussion on the state of the synonym refer to Avibase (Avibase, 2022;
210 Lepage, 2004; Lepage, Vaidya, & Guralnick, 2014). DateLife used the processed input names
211 to search the local chronogram database and found 9 matching chronograms in 6 different
212 studies. Three studies matched five input names (Barker, Burns, Klicka, Lanyon, & Lovette,
213 2015; Hedges, Marin, Suleski, Paymer, & Kumar, 2015; Jetz, Thomas, Joy, Hartmann, &
214 Mooers, 2012), one study matched four input names (Hooper & Price, 2017) and two studies
215 matched two input names (Barker, Burns, Klicka, Lanyon, & Lovette, 2013; Burns et al.,
216 2014). No studies matched all input names. Together, matching chronograms have 28 unique
217 age data points. All nodes have age data. As fixed tree topology, DateLife used OpenTree's
218 synthetic tree as default and mapped age data to nodes in the tree. As expected, more
219 inclusive nodes (e.g., node "n1") have more age data than less inclusive nodes (e.g., node
220 "n5"). The processing step allowed discovering five data points for node "n4" that would not

221 have had any data otherwise. Age summary statistics per node were calculated and tested as
222 secondary calibrations to date the tree topology using the BLADJ algorithm. Age data for
223 node “n2” was excluded as final calibration because it is older than age data of a more
224 inclusive node.

225 **Real life application**

226 A college educator wishes to obtain state-of-the-art data on time of evolutionary origin
227 of species belonging to the true finches for their class. They decide to use **datelife** because
228 they are teaching best practices for reproducibility. Students have the option to go to the
229 website at www.datelife.org and perform an interactive run. However, the educator also
230 wants the students to practice their R skills. The first step is to run a **datelife** query using
231 the “get species from taxon” flag. This will get all recognised species names within their
232 chosen inclusive taxon. The Fringillidae has 289 species, according to the Open Tree of Life
233 taxonomy. Once with a curated set of species taxon names, the next step is to run a
234 **datelife** search that will find all chronograms that contain at least two species names. The
235 algorithm proceeds to prune the trees to keep matching species names on tips only, and
236 transform the pruned trees to pairwise distance matrices. There are 13 chronograms
237 containing at least two Fringillidae species, published in 9 different studies (Barker et al.,
238 2013, 2015; Burns et al., 2014; Claramunt & Cracraft, 2015; Gibb et al., 2015; Hedges et al.,
239 2015; Hooper & Price, 2017; Jetz et al., 2012; Price et al., 2014). The final step is to
240 summarize the available information using two alternative types of summary chronograms,
241 median and SDM. As explained in the “Description” section, data from source chronograms
242 is first summarised into a single distance matrix and then the available node ages are used as
243 fixed node calibrations over a consensus tree topology, to obtain a fully dated tree with the
244 program BLADJ (Fig. 5). Median summary chronograms are older and have wider variation
245 in maximum ages than chronograms obtained with SDM. ?????????????????? Say some
246 things about the results!

247

Cross-validation test

248 As a form of cross validation, we took the tree topologies from each study and used
249 time of lineage divergence data from all other source chronograms as fixed calibrations with
250 the program BLADJ (Figs. 6 to 14).

251 The ages of internal nodes were recovered with a high precision in almost all cases
252 (except for studies 3, and 5; Fig. ??). Maximum tree ages were only recovered in one case
253 (study 2; Fig. 7). We also demonstrate the usage of PATHd8 (Britton et al., 2007) as a
254 dating method alternative to BLADJ. PATHd8 requires a tree with initial branch lengths
255 which we obtained using **datelife**'s branch length reconstruction algorithm, that searches
256 for DNA sequence data from the Barcode of Life Data System [BOLD; Ratnasingham and
257 Hebert (2007)] to generate branch lengths. We were able to successfully generate a tree with
258 BOLD branch lengths for all of the Fringillidae source chronograms. However, dating with
259 PATHd8 using congruified calibrations, was only successful in three cases (studies 3, 5, and 9,
260 shown in Fig. ??). From these, two trees have a different sampling than the original source
261 chronogram, mainly because DNA BOLD data for some species is absent from the database.
262 Maximum ages are quite different from source chronograms, but this might be explained also
263 by the differences in sampling between source chronograms and BOLD trees. More examples
264 and code used to generate these trees were developed on an open repository that is available
265 for consultation and reuse at https://github.com/LunaSare/datelife_examples.

266

Discussion

267 The main goal of **datelife** is to make state-of-the-art information on time of lineage
268 divergence easily accessible for comparison, reuse, and reanalysis, to researchers in all areas
269 of science and with all levels of expertise in the matter. It is an open service that does not
270 require any expert biological knowledge from users –besides the names of the organisms they
271 want to work with, for any of its functionality.

At the time of writing of this manuscript (Apr 07, 2022), `datelife`'s database has 253 chronograms, pulled entirely from OpenTree's database, the Phylesystem (McTavish et al., 2015). A unique feature of OpenTree's Phylesystem is that the community can add new state-of-the-art chronograms any time. As chronograms are added to Phylesystem, they are incorporated into an updated `datelife`'s database that is assigned a new version number, followed by a package release on CRAN. `datelife`'s chronogram database is updated as new chronogram data is added to Phylesystem, at a minimum of once a month and a maximum of every 6 months. Users can also upload new chronograms to OpenTree themselves, and trigger an update of their local `datelife` database to incorporate the new chronograms, to have them immediately available for analysis.

Incorporation of more chronograms into `datelife`'s database is crucial to improve its services. One option to increase chronogram number in the database is the Dryad data repository. Methods to automatically mine chronograms from Dryad could be designed and implemented. However, Dryad's metadata system has no information to automatically detect branch length units, and those would still need to be determined manually by a curator.

The largest, and taxonomically broadest, summary chronogram currently available from OpenTree was constructed using age data from 2,274 published chronograms (Hedges et al., 2015). However the source chronograms used as input data for this tree are not available in computer readable format for reuse or reanalysis. As this tree is part of `datelife`'s database, the amount of lineages that can be queried using `datelife` (99474 unique terminal taxa) is substantial. Access to the input chronograms used to generate the Hedges et al. summary tree would improve measures of uncertainty in DateLife, but they are available only as image files and not as usable data (timetree.org). We would like to emphasize on the importance of sharing chronogram data for the benefit of the scientific community as a whole, into repositories that require expert input and manual curation, such as OpenTree's Phylesystem (McTavish et al., 2015).

298 By default, `datelife` currently summarizes all source chronograms that overlap with
299 at least two species names. Users can exclude source chronograms if they have reasons to do
300 so. Strictly speaking, the best chronogram should reflect the real time of lineage divergence
301 accurately and precisely. To our knowledge, there are no good measures to determine
302 independently if a chronogram is better than another. Some measures that have been
303 proposed are the proportion of lineage sampling and the number of calibrations used
304 Magallón, Gómez-Acevedo, Sánchez-Reyes, & Hernández-Hernández (2015). Several
305 characteristics of the data used for dating analyses as well as from the output chronogram
306 itself, could be used to score quality of source chronograms. Some characteristics that are
307 often cited in published studies as a measure of improved age estimates as compared to
308 previously published estimates are: quality of alignment (missing data, GC content), lineage
309 sampling (strategy and proportion), phylogenetic and dating inference method, number of
310 fossils used as calibrations, support for nodes and ages, and magnitude of confidence
311 intervals. DateLife provides an opportunity to capture concordance and conflict among date
312 estimates, which can also be used as a metric for chronogram reliability.

313 Scientists usually also favor chronograms constructed using primary calibrations (ages
314 obtained from the fossil or geological record) to ones constructed with secondary calibrations
315 (ages coming from other chronograms)(Schenk, 2016). It has been observed with simulations
316 that divergence times inferred with secondary calibrations are significantly younger than
317 those inferred with primary calibrations in analyses performed with Bayesian inference
318 methods when priors are implemented in similar ways in both analyses (Schenk, 2016).
319 However, secondary calibrations can be applied using other dating methods that do not
320 require setting priors, such as penalized likelihood (Sanderson, 2003), or as fixed ages,
321 potentially mitigating the bias reported with Bayesian methods. Certainly, further studies
322 are required to fully understand the effect of using secondary calibrations on time estimates
323 and downstream analyses.

Furthermore, even chronograms obtained with primary fossil data can vary substantially in time estimates between lineages, as observed from the comparison of source chronograms in the Fringillidae example. This observation is often encountered in the literature (see, for example, the ongoing debate about crown group age of angiosperms (Barba-Montoya, Reis, Schneider, Donoghue, & Yang, 2018; Magallón et al., 2015; Ramshaw et al., 1972; Sanderson & Doyle, 2001; Sauquet, Ramírez-Barahona, & Magallón, 2021). For some studies, especially ones based on branch lengths (e.g., studies of species diversification, timing of evolutionary events, phenotypic trait evolution), using a different chronogram may return different results (Title & Rabosky, 2016). Stitching together these chronograms can create a larger tree that uses information from multiple studies, but the effect of uncertainties and errors at this level on downstream analyses is still largely unknown.

Summarizing chronograms might also imply summarizing fundamentally distinct evolutionary hypotheses. For example, two different researchers working on the same clade both carefully select and argument their choices of fossil calibrations. Still, if one researcher decides a fossil will calibrate the ingroup of a clade, while another researcher uses the same one to calibrate outside the clade, the resulting age estimates will often differ substantially, as the placement of calibrations as stem or crown group is proved to deeply affect estimated times of lineage divergence (Sauquet, 2013). Trying to summarize the resulting chronograms into a single one using simple summary statistics can erase many types of relevant information from the source chronograms. Accordingly, the prevailing view is that we should favor time of lineage divergence estimates obtained from a single analysis, using fossil data as primary sources of calibrations, and using fossils that have been widely discussed and curated as calibrations to date other trees, making sure that all data used in the analysis reflect a coherent evolutionary history (Antonelli et al., 2017). However, the exercise of summarizing different chronograms has the potential to help getting a single global evolutionary history for a lineage by putting together evidence from different hypothesis. Choosing the elements of the chronograms that we are going to keep and the ones that we

351 are going to discard is key, since we are potentially loosing important parts of the
352 evolutionary history of a lineage that might only be reflected in source chronograms and not
353 on the summary chronogram (Sauquet et al., 2021).

354 Nonetheless, in ecology and conservation biology, incorporating at least some data on
355 lineage divergence times represents a relevant improvement for testing alternative hypothesis
356 using phylogenetic distance (Campbell O. Webb et al., 2008). Hence, we integrated into
357 datelife's workflow different ways of estimating node ages in the absence of calibrations and
358 branch length information for taxa lacking this information. "Making up" branch lengths is
359 an accepted practice in scientific publications: Jetz et al. (2012), created a time-calibrated
360 tree of all 9,993 bird species, where 67% had molecular data and the rest was simulated;
361 Rabosky et al. (2018) created a time-calibrated tree of 31,536 ray-finned fishes, of which only
362 37% had molecular data; Stephen A. Smith and Brown (2018) constructed a tree of 353,185
363 seed plants where only 23% had molecular data. Obviously, there are risks in this practice!
364 Taken to the extreme, one could make a fully resolved, calibrated tree of all modern and
365 extinct taxa using a single taxonomy and a single calibration with the polytomy resolution
366 and branch estimation methods. There has yet to be a thorough analysis of what can go
367 wrong when one extends inferences beyond the data in this way, so we urge caution; we also
368 urge readers to follow the example of many of the large tree papers cited above and make
369 carefully consider the statistical assumptions being made, and assess the consistency of the
370 results with prior work.

371 Conclusions

372 Divergence time information is key to many areas of evolutionary studies: trait
373 evolution, diversification, biogeography, macroecology and more. It is also crucial for science
374 communication and education, but generating chronograms is difficult, especially for those
375 who want to use phylogenies but who are not systematists, or do not have the time to
376 acquire and develop the necessary knowledge and data curation skills. Moreover, years of

377 primarily public funded research have resulted in vast amounts of chronograms that are
378 already available on scientific publications, but hidden to the public and scientific community
379 for reuse.

380 The `datelife` R package allows easy and fast summarization of publicly available
381 information on time of lineage divergence. This provides a straightforward way to get an
382 informed idea on the state of knowledge of the time frame of evolution of different regions of
383 the tree of life, and allows identification of regions that require more research or that have
384 conflicting information. It is available as an R package, or a web-based R shiny app at
385 dates.opentreeloflife.org/datelife. Both summary and newly generated trees are useful to
386 evaluate evolutionary hypotheses in different areas of research. The DateLife project helps
387 with awareness of the existing variation in expert time of divergence data, and will foster
388 exploration of the effect of alternative divergence time hypothesis on the results of analyses,
389 nurturing a culture of more cautious interpretation of evolutionary results.

390 Availability

391 `datelife` is free and open source and it can be used through its current website
392 <http://www.datelife.org/query/>, through its R package, and through Phylotastic's project
393 web portal <http://phylo.cs.nmsu.edu:3000/>. `datelife`'s website is maintained using
394 RStudio's shiny server and the shiny package open infrastructure, as well as Docker.
395 `datelife`'s R package stable version is available for installation from the CRAN repository
396 (<https://cran.r-project.org/package=datelife>) using the command `install.packages(pkgs`
397 `= "datelife")` from within R. Development versions are available from the GitHub
398 repository (<https://github.com/phylotastic/datelife>) and can be installed using the
399 command `devtools::install_github("phylotastic/datelife")`.

Supplementary Material

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

Funding

405 Funding was provided by the US National Science Foundation (NSF) grants
406 ABI-1458603 to Datelife project and DBI-0905606 to the National Evolutionary Synthesis
407 Center (NESCent), and the Phylotastic project Grant ABI-1458572, and the OpenTree grant
408 ABI-1759846.

Acknowledgements

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

References

- 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., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2013). Going to extremes: Contrasting rates of diversification in a recent radiation of new world passerine birds. *Systematic Biology*, 62(2), 298–320.
- Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2015). New insights into new world biogeography: An integrated view from the phylogeny of blackbirds, cardinals, sparrows, tanagers, warblers, and allies. *The Auk: Ornithological Advances*, 132(2), 333–348.
- Britton, T., Anderson, C. L., Jacquet, D., Lundqvist, S., & Bremer, K. (2007).

- 449 Estimating Divergence Times in Large Phylogenetic Trees. *Systematic Biology*,
450 56(788777878), 741–752. <https://doi.org/10.1080/10635150701613783>
- 451 Burns, K. J., Shultz, A. J., Title, P. O., Mason, N. A., Barker, F. K., Klicka, J., ...
452 Lovette, I. J. (2014). Phylogenetics and diversification of tanagers (passeriformes:
453 Thraupidae), the largest radiation of neotropical songbirds. *Molecular
454 Phylogenetics and Evolution*, 75, 41–77.
- 455 Chamberlain, S. A., & Szöcs, E. (2013). taxize : taxonomic search and retrieval in R
456 [version 2; referees: 3 approved]. *F1000Research*, 2(191), 1–29.
457 <https://doi.org/10.12688/f1000research.2-191.v2>
- 458 Chamberlain, S. A., Szöcs, E., Foster, Z., Arendsee, Z., Boettiger, C., Ram, K., ...
459 Li, G. (2019). *taxize: Taxonomic information from around the web*. Retrieved
460 from <https://github.com/ropensci/taxize>
- 461 Claramunt, S., & Cracraft, J. (2015). A new time tree reveals earth history's imprint
462 on the evolution of modern birds. *Science Advances*, 1(11), e1501005.
- 463 Criscuolo, A., Berry, V., Douzery, E. J. P., & Gascuel, O. (2006). SDM: A fast
464 distance-based approach for (super)tree building in phylogenomics. *Systematic
465 Biology*, 55(5), 740–755. <https://doi.org/10.1080/10635150600969872>
- 466 Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M.-K., Turon, X., ...
467 Douzery, E. J. (2018). A phylogenomic framework and timescale for comparative
468 studies of tunicates. *BMC Biology*, 16(1), 1–14.
- 469 Eastman, J. M., Harmon, L. J., & Tank, D. C. (2013). Congruification: Support for
470 time scaling large phylogenetic trees. *Methods in Ecology and Evolution*, 4(7),
471 688–691. <https://doi.org/10.1111/2041-210X.12051>
- 472 Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and
473 high throughput. *Nucleic Acids Research*, 32(5), 1792–1797.
- 474 Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American
475 Naturalist*, 125(1), 1–15. Retrieved from <http://www.jstor.org/stable/2461605>

- 476 Gibb, G. C., England, R., Hartig, G., McLenachan, P. A., Taylor Smith, B. L.,
477 McComish, B. J., ... Penny, D. (2015). New zealand passerines help clarify the
478 diversification of major songbird lineages during the oligocene. *Genome Biology
479 and Evolution*, 7(11), 2983–2995.
- 480 Harmon, L., Weir, J., Brock, C., Glor, R., & Challenger, W. (2008). GEIGER:
481 investigating evolutionary radiations. *Bioinformatics*, 24, 129–131.
- 482 Hedges, S. B., Marin, J., Suleski, M., Paymer, M., & Kumar, S. (2015). Tree of life
483 reveals clock-like speciation and diversification. *Molecular Biology and Evolution*,
484 32(4), 835–845. <https://doi.org/10.1093/molbev/msv037>
- 485 Heibl, C. (2008). *PHYLOCH: R language tree plotting tools and interfaces to diverse
486 phylogenetic software packages*. Retrieved from
487 <http://www.christophheibl.de/Rpackages.html>
- 488 Hooper, D. M., & Price, T. D. (2017). Chromosomal inversion differences correlate
489 with range overlap in passerine birds. *Nature Ecology & Evolution*, 1(10), 1526.
- 490 Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of
491 phylogenetic trees. *Bioinformatics*, 17(8), 754–755.
492 <https://doi.org/10.1093/bioinformatics/17.8.754>
- 493 Jetz, W., Thomas, G., Joy, J. J. B., Hartmann, K., & Mooers, A. (2012). The global
494 diversity of birds in space and time. *Nature*, 491(7424), 444–448.
495 <https://doi.org/10.1038/nature11631>
- 496 Katoh, K., Asimenos, G., & Toh, H. (2009). Multiple alignment of DNA sequences
497 with MAFFT. In *Bioinformatics for DNA sequence analysis* (pp. 39–64).
498 Springer.
- 499 Laubichler, M. D., & Maienschein, J. (2009). *Form and function in developmental
500 evolution*. Cambridge University Press.
- 501 Lepage, D. (2004). *Avibase: The world bird database*. Bird Studies Canada.
- 502 Lepage, D., Vaidya, G., & Guralnick, R. (2014). Avibase—a database system for

- managing and organizing taxonomic concepts. *ZooKeys*, (420), 117.
- Magallon, S., & Sanderson, M. J. (2001). Absolute diversification rates in angiosperm clades. *Evolution*, 55(9), 1762–1780.
- Magallón, S. (2010). Using fossils to break long branches in molecular dating: A comparison of relaxed clocks applied to the origin of angiosperms. *Systematic Biology*, 59(4), 384–399.
- Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., & Hernández-Hernández, T. (2015). A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist*, 207(2), 437–453.
- McTavish, E. J., Hinchliff, C. E., Allman, J. F., Brown, J. W., Cranston, K. A., Holder, M. T., . . . Smith, S. A. (2015). Phylesystem: A git-based data store for community-curated phylogenetic estimates. *Bioinformatics*, 31(17), 2794–2800.
- Michonneau, F., Brown, J. W., & Winter, D. J. (2016). rotl: an R package to interact with the Open Tree of Life data. *Methods in Ecology and Evolution*, 7(12), 1476–1481. <https://doi.org/10.1111/2041-210X.12593>
- Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecology Letters*, 17(4), 508–525. <https://doi.org/10.1111/ele.12251>
- Ooms, J., & Chamberlain, S. (2018). *Phylocomr: Interface to 'phylocom'*. Retrieved from <https://CRAN.R-project.org/package=phylocomr>
- Open Tree Of Life, Redelings, B., Cranston, K. A., Allman, J., Holder, M. T., & McTavish, E. J. (2016). Open Tree of Life APIs v3.0. *Open Tree of Life Project*, (Online Resources). Retrieved from
%7Bhttps://github.com/OpenTreeOfLife/germinator/wiki/Open-Tree-of-Life-Web-APIs%7D
- Open Tree Of Life, Redelings, B., Sánchez Reyes, L. L., Cranston, K. A., Allman, J., Holder, M. T., & McTavish, E. J. (2019). Open tree of life synthetic tree v12.3. *Zenodo*. Retrieved from <https://doi.org/10.5281/zenodo.3937742>

- 530 Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and
531 evolution in R language. *Bioinformatics*, 20(2), 289–290.
- 532 Posadas, P., Crisci, J. V., & Katinas, L. (2006). Historical biogeography: A review of
533 its basic concepts and critical issues. *Journal of Arid Environments*, 66(3),
534 389–403.
- 535 Price, T. D., Hooper, D. M., Buchanan, C. D., Johansson, U. S., Tietze, D. T.,
536 Alström, P., ... others. (2014). Niche filling slows the diversification of himalayan
537 songbirds. *Nature*, 509(7499), 222.
- 538 R Core Team. (2018). *R: a language and environment for statistical computing*.
539 Vienna, Austria: R Foundation for Statistical Computing.
- 540 Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., ...
541 others. (2018). An inverse latitudinal gradient in speciation rate for marine fishes.
542 *Nature*, 559(7714), 392.
- 543 Ramshaw, J., Richardson, D., Mealyard, B., Brown, R., Richardson, M., Thompson,
544 E., & Boulter, D. (1972). The time of origin of the flowering plants determined by
545 using amino acid sequence data of cytochrome c. *New Phytologist*, 71(5), 773–779.
- 546 Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The barcode of life data system
547 (<http://www.Barcodinglife.org>). *Molecular Ecology Notes*, 7(3), 355–364.
- 548 Rees, J. A., & Cranston, K. (2017). Automated assembly of a reference taxonomy for
549 phylogenetic data synthesis. *Biodiversity Data Journal*, (5).
- 550 Revell, L. J. (2012). Phytools: An r package for phylogenetic comparative biology
551 (and other things). *Methods in Ecology and Evolution*, 3, 217–223.
- 552 Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic
553 inference under mixed models. *Bioinformatics*, 19(12), 1572–1574.
554 <https://doi.org/10.1093/bioinformatics/btg180>
- 555 Sanchez-Reyes, L. L., O'Meara, B., Eastman, J., Heath, T., Wright, A., Schliep, K.,
556 ... Alfaro, M. (2022). datelife: Scientific Data on Time of Lineage Divergence for

- 557 Your Taxa. *R Package Version 0.6.2*. Retrieved from
558 <https://doi.org/10.5281/zenodo.593938>
- 559 Sanderson, M. J. (2003). r8s: Inferring absolute rates of molecular evolution and
560 divergence times in the absence of a molecular clock. *Bioinformatics*, 19(2),
561 301–302.
- 562 Sanderson, M. J., & Doyle, J. A. (2001). Sources of error and confidence intervals in
563 estimating the age of angiosperms from rbcL and 18S rDNA data. *American*
564 *Journal of Botany*, 88(8), 1499–1516.
- 565 Sauquet, H. (2013). A practical guide to molecular dating. *Comptes Rendus Palevol*,
566 12(6), 355–367.
- 567 Sauquet, H., Ramírez-Barahona, S., & Magallón, S. (2021). The age of flowering
568 plants is unknown.
- 569 Schenk, J. J. (2016). Consequences of secondary calibrations on divergence time
570 estimates. *PLoS ONE*, 11(1). <https://doi.org/10.1371/journal.pone.0148228>
- 571 Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in r. *Bioinformatics*, 27(4),
572 592–593.
- 573 Smith, Stephen A., & Brown, J. W. (2018). Constructing a broadly inclusive seed
574 plant phylogeny. *American Journal of Botany*, 105(3), 302–314.
- 575 Smith, Stephen A., & O'Meara, B. C. (2012). TreePL: Divergence time estimation
576 using penalized likelihood for large phylogenies. *Bioinformatics*, 28(20),
577 2689–2690. <https://doi.org/10.1093/bioinformatics/bts492>
- 578 Stoltzfus, A., Lapp, H., Matasci, N., Deus, H., Sidlauskas, B., Zmasek, C. M., ...
579 Jordan, G. (2013). Phylotastic! Making tree-of-life knowledge accessible, reusable
580 and convenient. *BMC Bioinformatics*, 14.
581 <https://doi.org/10.1186/1471-2105-14-158>
- 582 Title, P. O., & Rabosky, D. L. (2016). Do Macrophylogenies Yield Stable
583 Macroevolutionary Inferences? An Example from Squamate Reptiles. *Systematic*

- 584 *Biology*, syw102. <https://doi.org/10.1093/sysbio/syw102>
- 585 Vos, R. A., Balhoff, J. P., Caravas, J. A., Holder, M. T., Lapp, H., Maddison, W. P.,
586 ... others. (2012). NeXML: Rich, extensible, and verifiable representation of
587 comparative data and metadata. *Systematic Biology*, 61(4), 675–689.
- 588 Webb, C. (2000). Exploring the Phylogenetic Structure of Ecological Communities :
589 An Example for Rain Forest Trees. *The American Naturalist*, 156(2), 145–155.
- 590 Webb, Campbell O., Ackerly, D. D., & Kembel, S. W. (2008). Phylocom: Software for
591 the analysis of phylogenetic community structure and trait evolution.
592 *Bioinformatics*, 24(18), 2098–2100.
593 <https://doi.org/10.1093/bioinformatics/btn358>
- 594 Webb, Campbell O., & Donoghue, M. J. (2005). Phylomatic: Tree assembly for
595 applied phylogenetics. *Molecular Ecology Notes*, 5(1), 181–183.

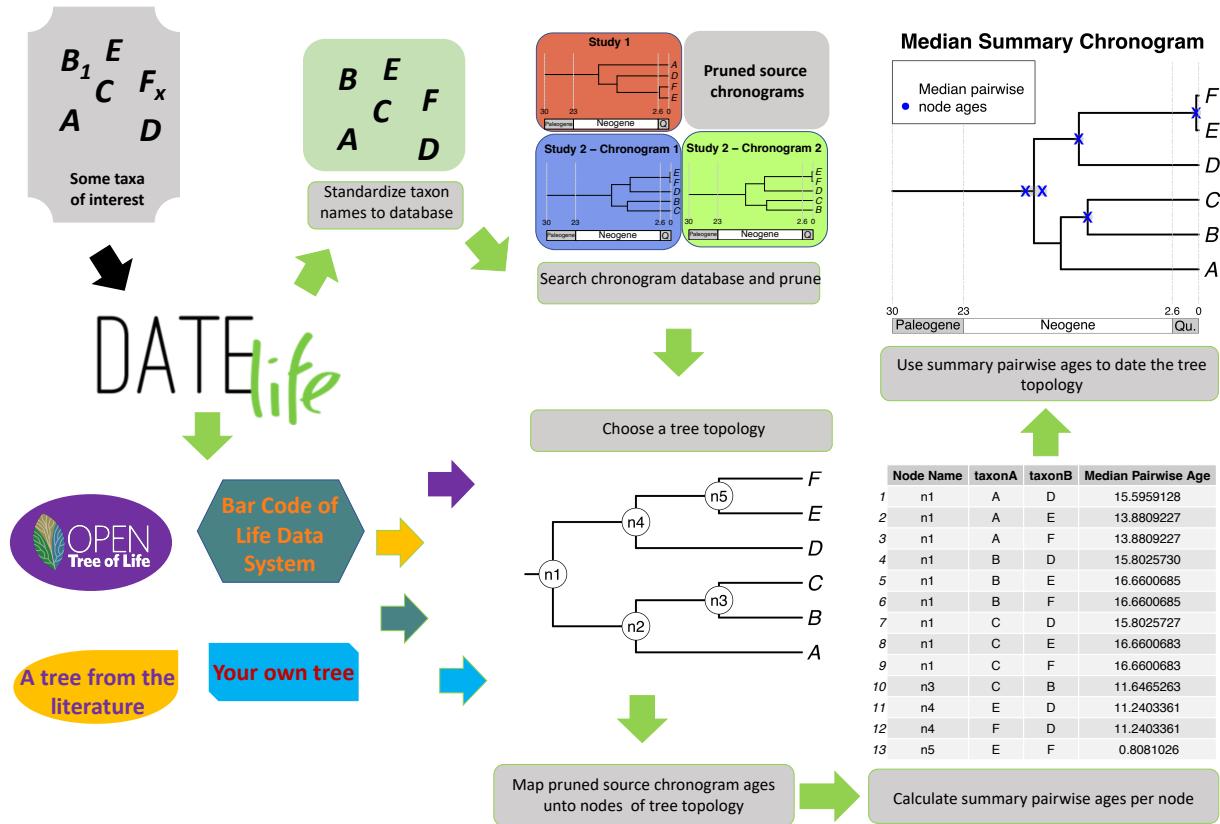


FIGURE 1. Stylized DateLife workflow. This shows the general workflows and analyses that can be performed with `datelife`, via the R package or through the website at <http://www.datelife.org/>. Details on the functions involved on each workflow are shown in `datelife`'s R package vignette.

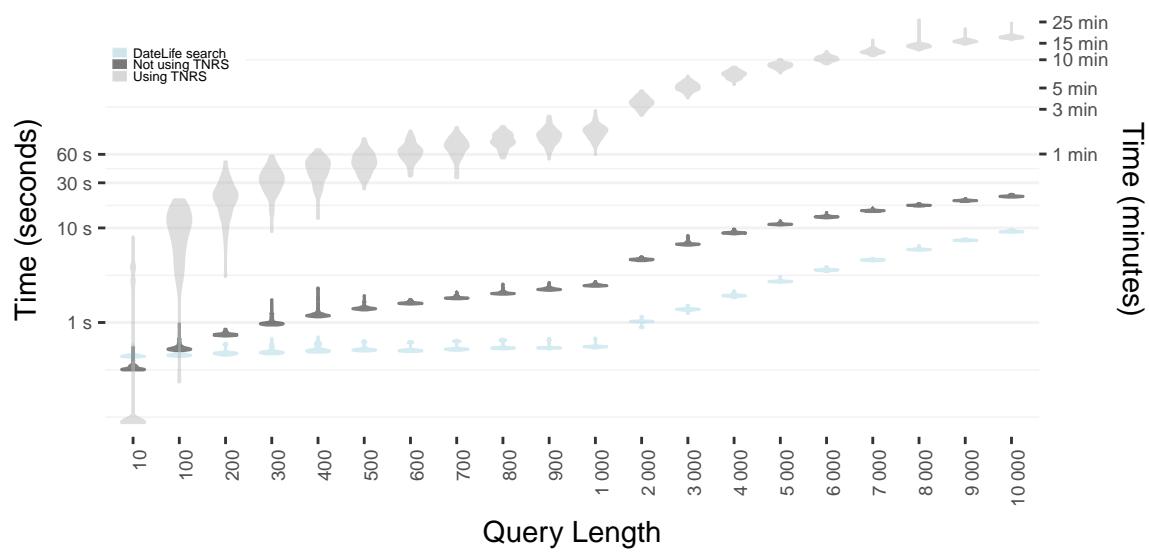


FIGURE 2. Computation time of query processing and search across **datelife**'s chronogram database relative to number of input taxon names. We sampled N names from the class Aves for each cohort 100 times and then performed a search with query processing not using the Taxon Names Resolution Service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

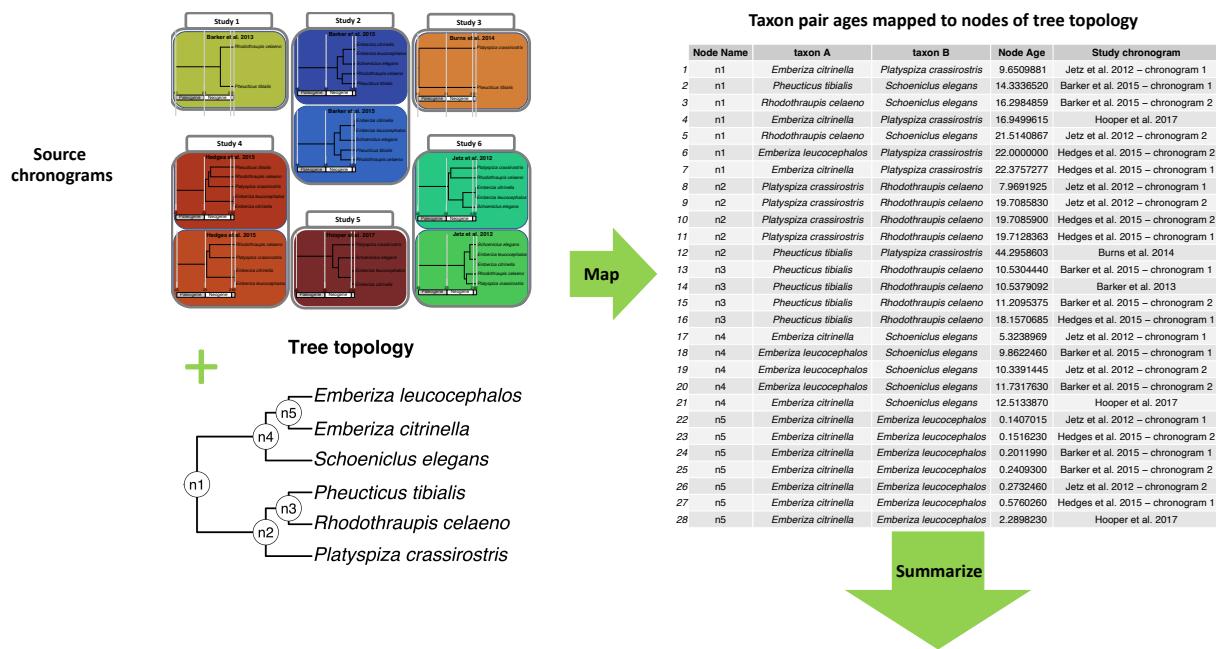


FIGURE 3. Age data results of a DateLife search of a small sample of 6 bird species within the Passeriformes. Input names were found across 9 chronograms within 6 independent studies (Barker et al. (2012), Barker et al. (2015), Burns et al. (2014), Hedges et al. (2015), Hooper and Price (2017), Jetz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

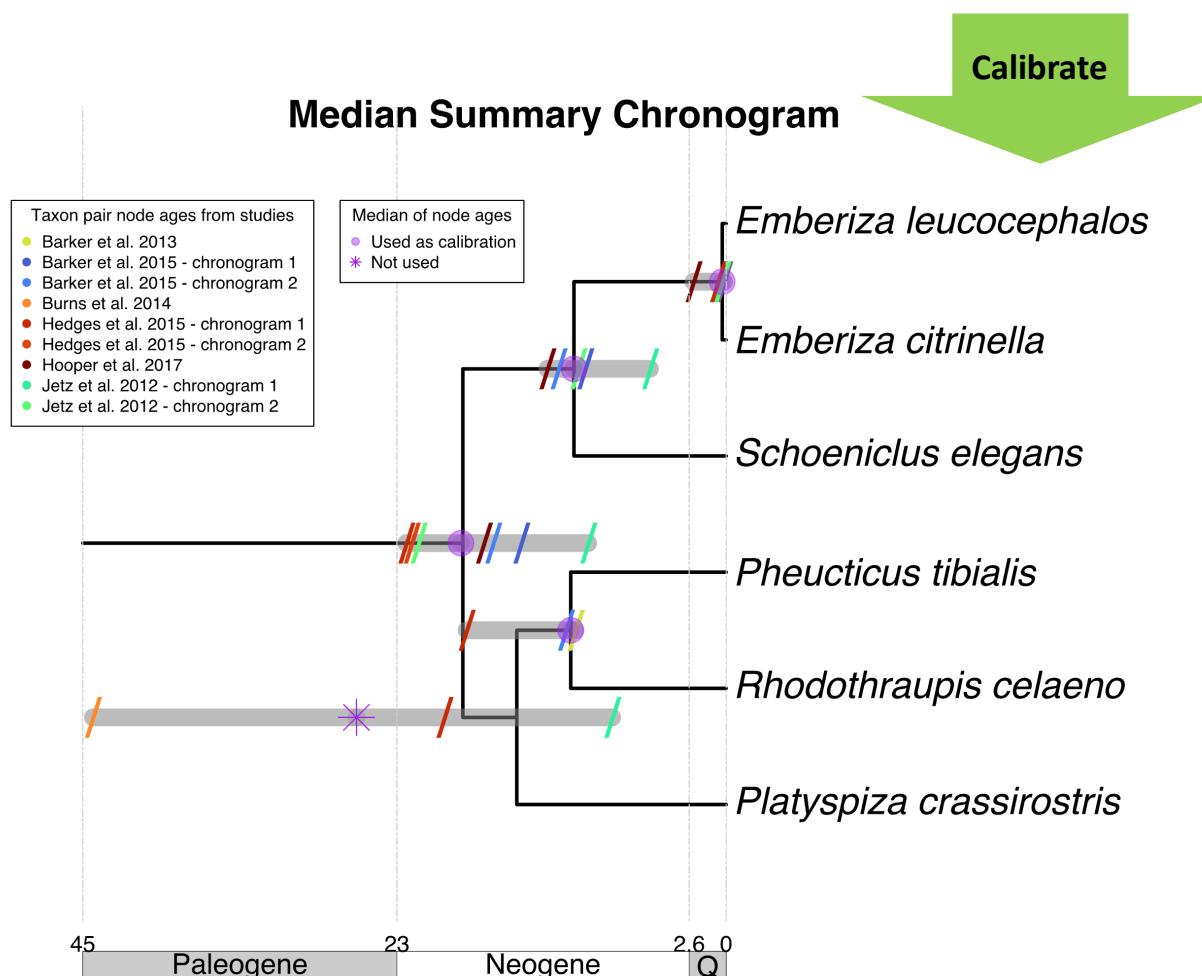


FIGURE 4. Summarized age data is used as secondary calibrations to date a tree topology as a summary chronogram.

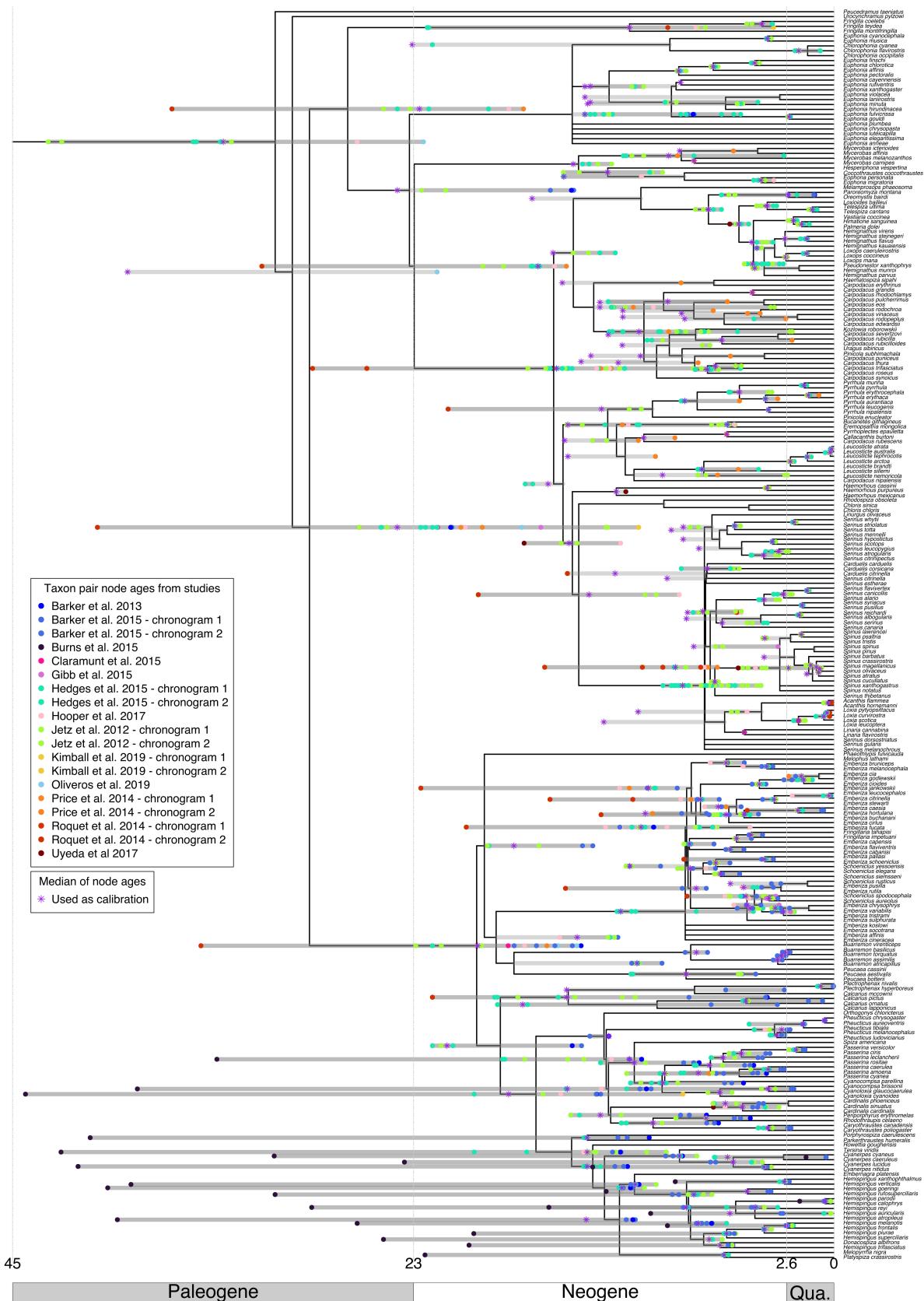


FIGURE 5. Fringillidae median summary chronogram generated with DateLife. It has 256 tips and 233 nodes.

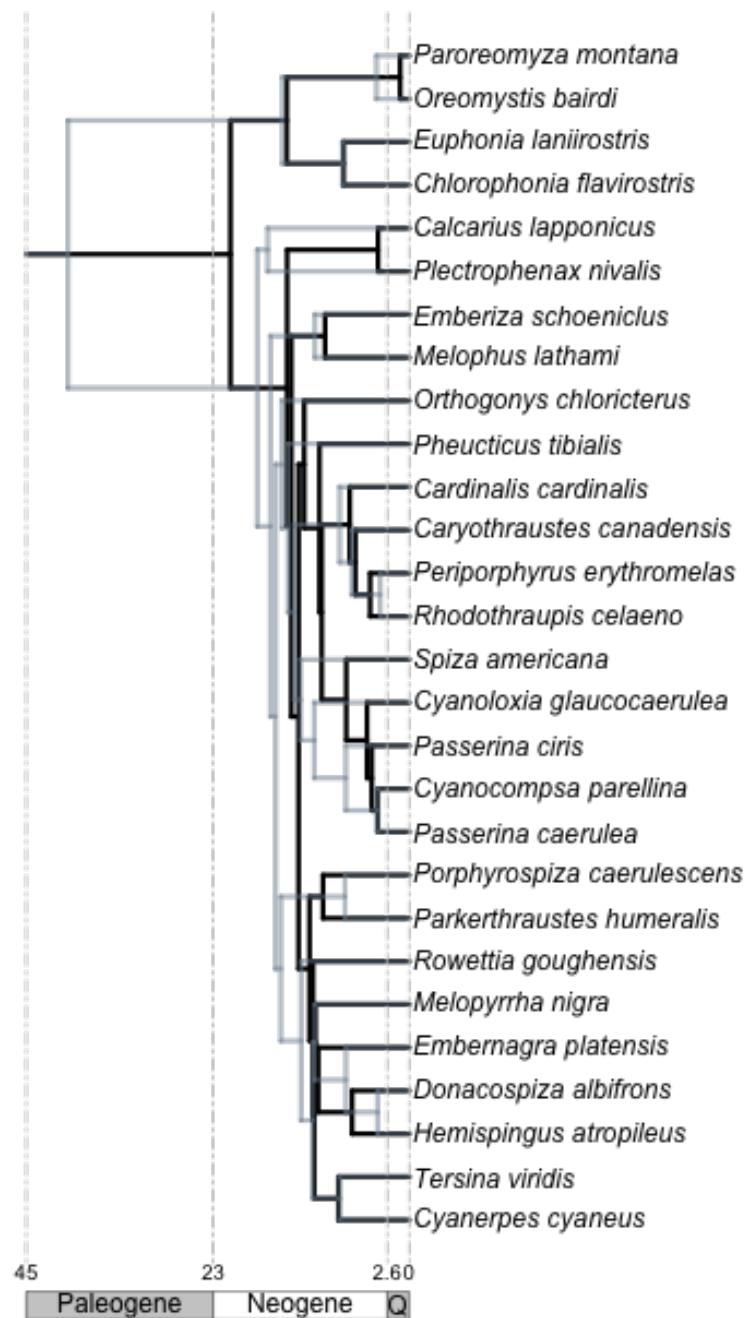
Barker et al. 2013

FIGURE 6. Cross validation of first source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the ones calculated with BLADJ in order to force all the nodes to fall within the same time interval.

Barker et al. 2015 - chronogram 1

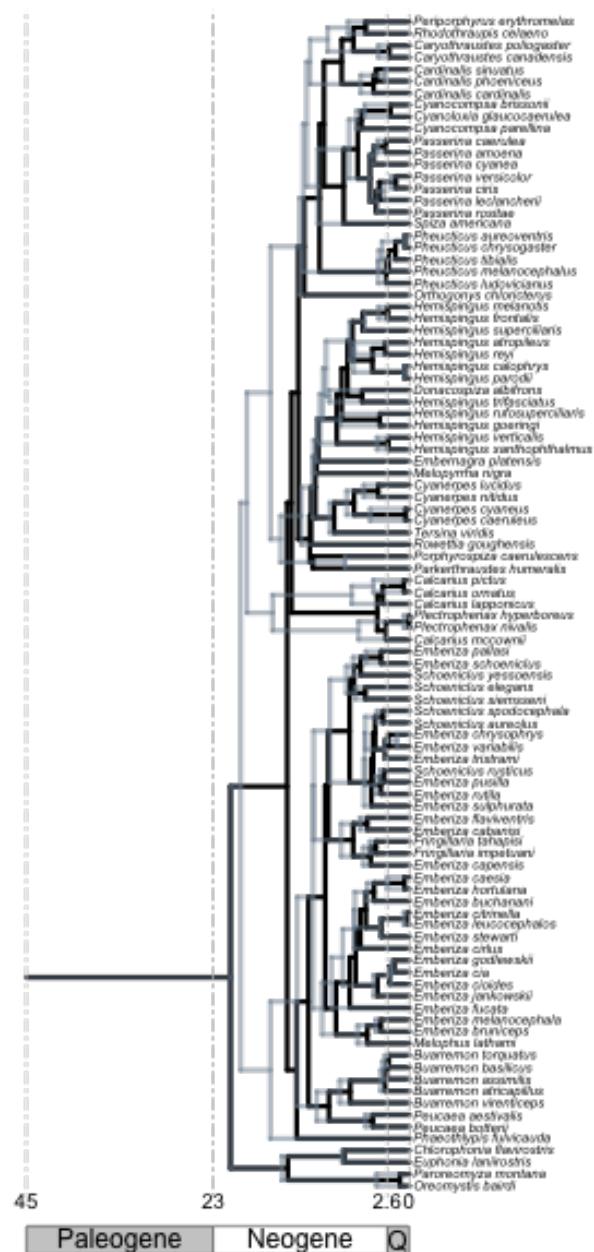


FIGURE 7. Cross validation of second source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to

Barker et al. 2015 - chronogram 2



FIGURE 8. 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 the dates calculated with BLADJ, i.e., the same for all the nodes.

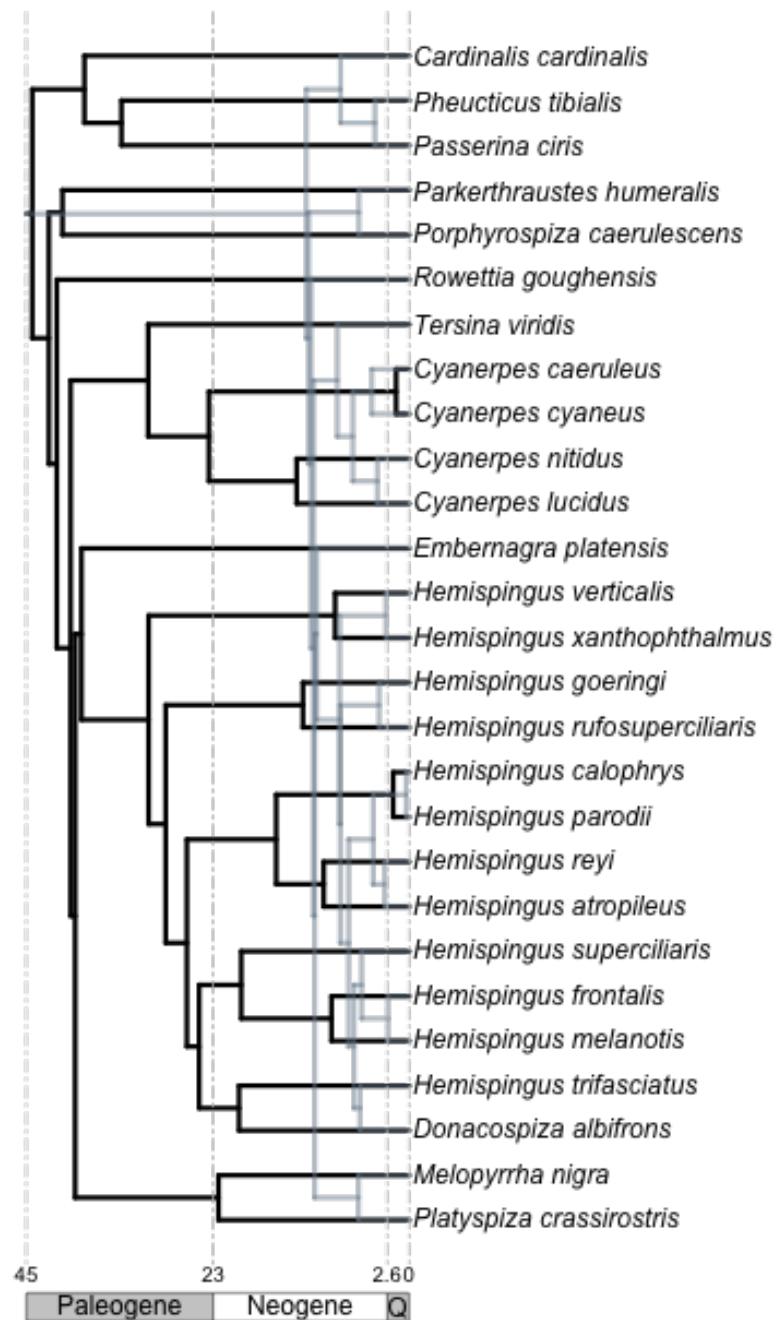
Burns et al. 2015

FIGURE 9. 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 mean of all the samples.

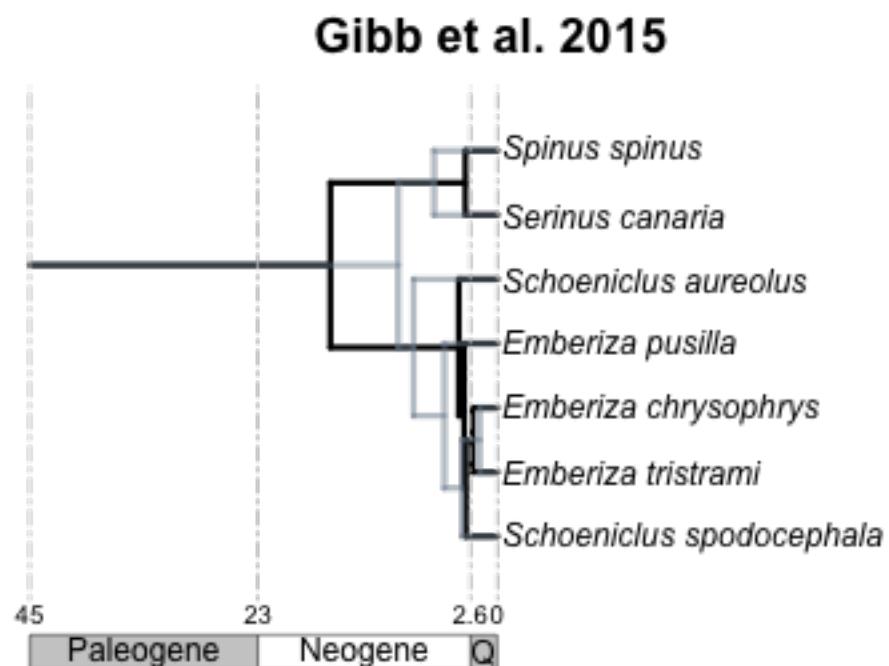


FIGURE 10. Cross validation of sixth source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the same tree topology dated with BLADJ using node ages from all other source chronograms as secondary calibrations.

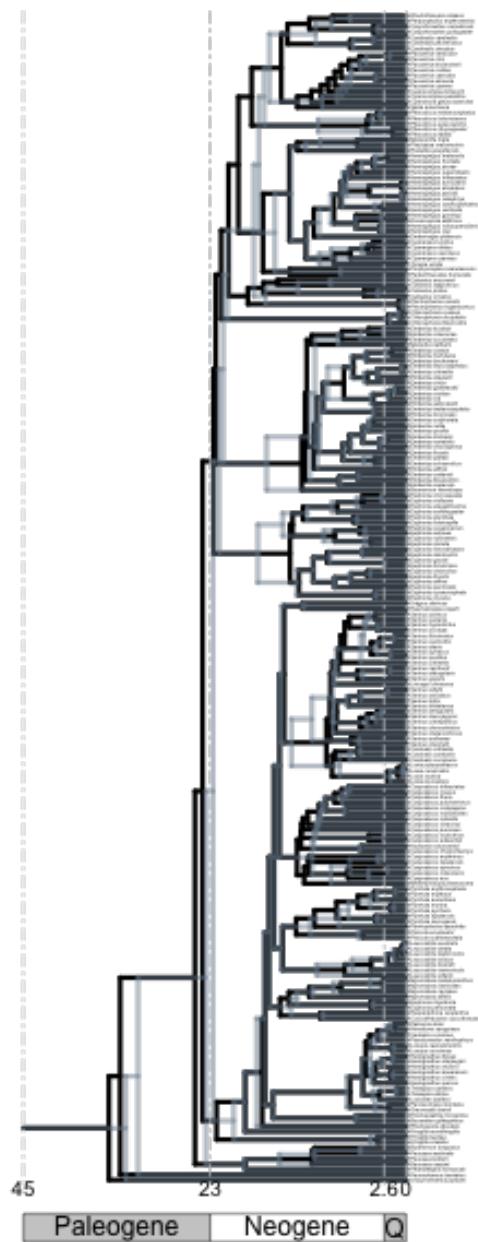
Hedges et al. 2015 - chronogram 1

FIGURE 11. 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 BLADJ, i.e., the cross-validation chronogram.

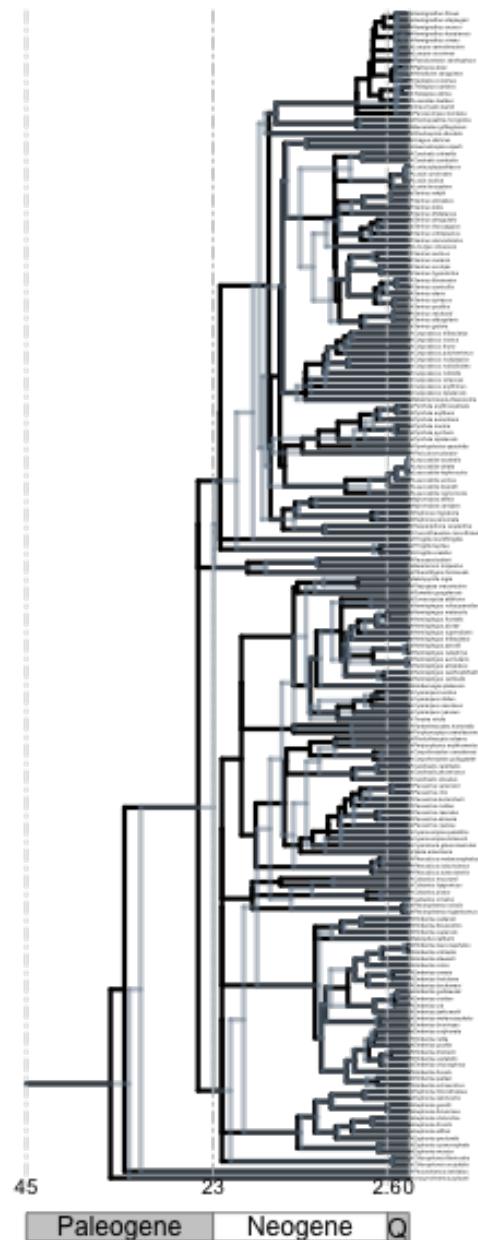
Hedges et al. 2015 - chronogram 2

FIGURE 12. 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 rooted at the same point.

Hooper et al. 2017

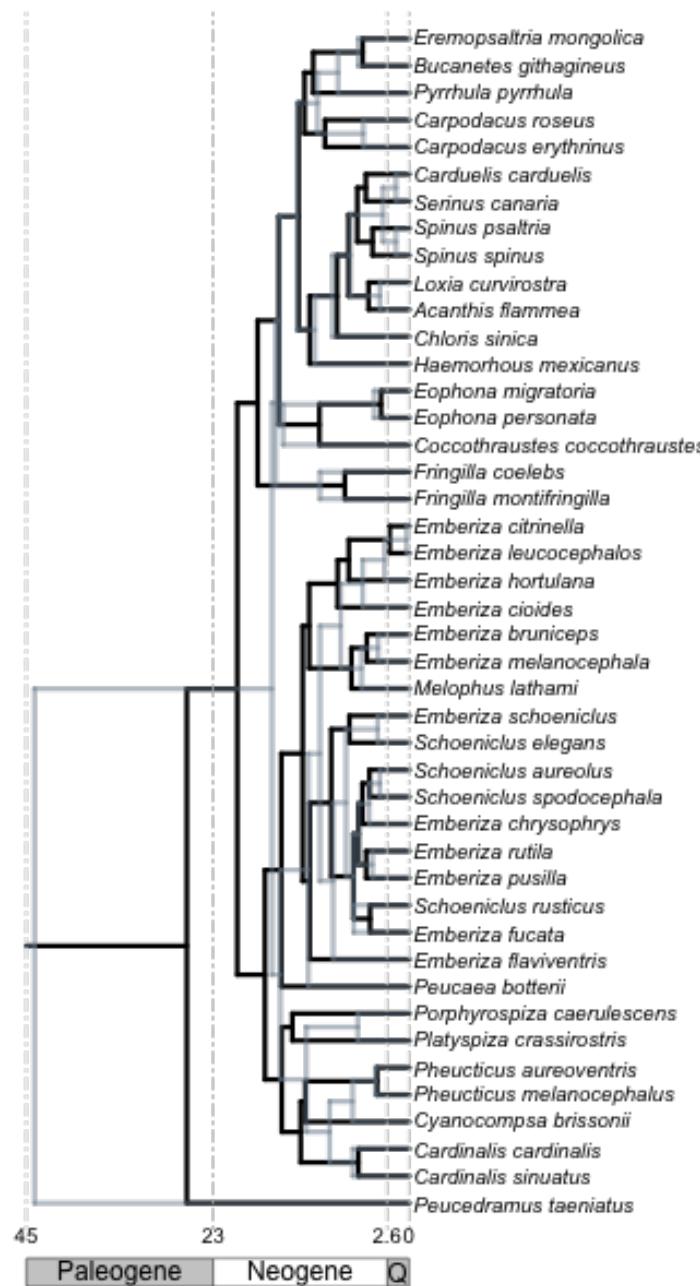


FIGURE 13. Cross validation of ninth source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the dates calculated with BLADJ, i.e., the ones from the tenth cross-validation.

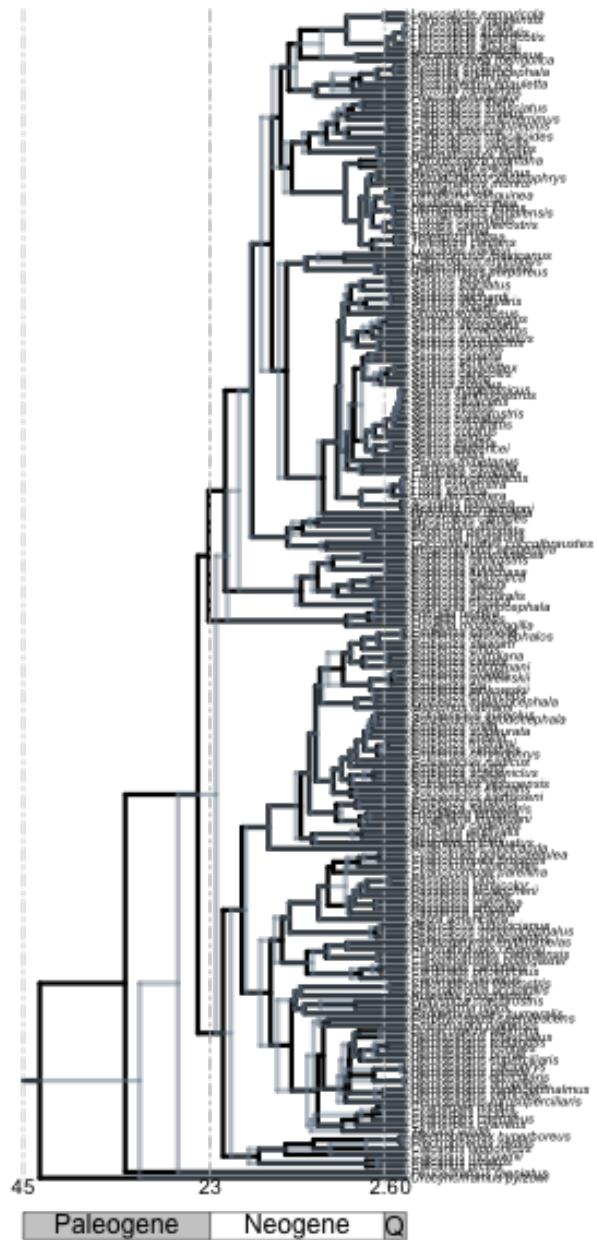
Jetz et al. 2012 - chronogram 1

FIGURE 14. 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.