

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, that provides functionalities for efficient
30 and easy finding, summary, reuse, and reanalysis of expert, peer-reviewed, public data on
31 time of evolutionary origin. The main DateLife workflow constructs a chronogram for any
32 given combination of taxon names, by searching a local chronogram database constructed
33 and curated from the Open Tree of Life Phylesystem phylogenetic database, which
34 incorporates phylogenetic data from TreeBASE database as well. We implement and test
35 methods for summarizing time data from multiple source chronograms using supertree and
36 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 of the existing variation in expert
39 time of divergence data, and can foster exploration of the effect of alternative divergence
40 time hypotheses 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: 4493

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; it also generally requires specialized biological training, taxonomic domain
56 knowledge, and a non-negligible amount of research time, computational resources and
57 funding.

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

66 **Description**

67 DateLife’s core software application consists of the R package `datelife`. Its latest
68 stable version – v0.6.2, is available from the CRAN repository (Sanchez-Reyes et al., 2022),
69 and relies on functionalities from various biological R packages: `ape` (Paradis, Claude, &

70 Strimmer, 2004), bold (Chamberlain et al., 2019), geiger (Harmon, Weir, Brock, Glor, &
71 Challenger, 2008), paleotree (Bapst, 2012), phyloch (Heibl, 2008), phylocomr (Ooms &
72 Chamberlain, 2018), phytools (Revell, 2012), rotl (Michonneau, Brown, & Winter, 2016),
73 and taxize (Chamberlain & Szöcs, 2013; Chamberlain et al., 2019). Figure 1 provides a
74 graphical summary of the three main steps of the DateLife algorithm: providing an input,
75 searching a chronogram database, and summarizing results from the search.

76 Processing an input

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

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

89 DateLife processes input scientific names using a Taxonomic Name Resolution Service
90 (TNRS), which increases the probability of correctly finding the queried taxon names in the
91 chronogram database. TNRS detects, corrects and standardizes name misspellings and typos,
92 variant spellings and authorities, and nomenclatural synonyms to a single taxonomic
93 standard (Boyle et al., 2013). DateLife implements TNRS using OpenTree’s unified
94 taxonomy as standard (Open Tree Of Life et al., 2016; Rees & Cranston, 2017), storing

95 OpenTree's Taxonomy identification numbers for further processing.

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

100 **Searching the database**

101 DateLife's chronogram database latest version consist of 253 chronograms published in
102 187 different studies. It is curated from OpenTree's phylogenetic database, the Phylesystem,
103 which constitutes an open source of expert and peer-reviewed phylogenetic knowledge with
104 rich metadata (McTavish et al., 2015), which allows automatic and reproducible assembly of
105 our chronogram database. Datelife's chronogram database is navigable as an R data object
106 within the `datelife` R package. Published chronograms can be added to Phylesystem by
107 any user, at any time, and are immediately publicly available
108 (<https://tree.opentreeoflife.org/curator>). This facilitates an immediate update of DateLife's
109 chronogram database to include new chronogram data on a following search.

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

120 **Summarizing search results**

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

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

134 ***Choosing a topology.***— DateLife requires a tree topology to summarize age data
135 upon. We recommend that users provide a tree topology as input from the literature, or one
136 of their own making. If no topology is provided, DateLife automatically subsets one from the
137 OpenTree synthetic tree (Open Tree Of Life et al., 2019). Alternatively, DateLife can
138 combine topologies from source chronograms using a supertree approach. To combine
139 topologies from source chronograms into a single summary (or supertree) topology, the
140 DateLife algorithm starts by identifying the source chronograms that form a grove, roughly,
141 a sufficiently overlapping set of taxa between trees, by implementing definition 2.8 for
142 n-overlap from Ané et al. (2009). In rare cases, a group of trees can have multiple groves. By
143 default, DateLife chooses the grove with the most taxa, however, the “criterion = trees” flag
144 allows the user to choose the grove with the most trees instead. If source chronograms do
145 not form a grove, the supertree reconstruction will fail.

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

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

For each congruent node, the pairwise distances that traverse that node are summarized into a single summary matrix using classic summary statistics (i.e., mean, median, minimum and maximum ages), and the Supermatrix Distance Method [SDM; Criscuolo, Berry, Douzery, and Gascuel (2006)], which deforms patristic distance matrices by minimizing variance and then averaging them. These single summary taxon pair age matrices (Summarized calibrations) can be applied as secondary calibrations to date a tree topology, using different dating methods currently supported within DateLife: MrBayes (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003), PATHd8 (Britton, Anderson, Jacquet,

- ¹⁷² Lundqvist, & Bremer, 2007), BLADJ (Campbell O. Webb, Ackerly, & Kembel, 2008);
¹⁷³ Campbell O. Webb & Donoghue, 2005), and treePL (Stephen A. Smith & O'Meara, 2012).

¹⁷⁴ By default, DateLife implements the Branch Length Adjuster (BLADJ) algorithm to
¹⁷⁵ obtain a fully dated topology. BLADJ fixes node ages that have calibration data, and
¹⁷⁶ distributes time between nodes with no data evenly between nodes with calibration data.
¹⁷⁷ This minimizes age variance in the resulting chronogram (Campbell O. Webb et al., 2008).
¹⁷⁸ BLADJ does not incorporate branch lengths even when they are present. When there is
¹⁷⁹ conflict in ages between nodes with calibration data, the algorithm ignores node ages that
¹⁸⁰ are older than the age of a parent node. BLADJ requires a root age estimate. If there is no
¹⁸¹ information on the age of the root in the chronogram database, users can provide an
¹⁸² estimate from the literature. If none is provided, DateLife assigns an arbitrary age to the
¹⁸³ root as 10% older than the oldest age available within the group.

¹⁸⁴ Alternative options supported in DateLife (MrBayes, PATHd8, TreePL) incorporate
¹⁸⁵ branch length information from the input topology in combination with the secondary
¹⁸⁶ calibrations. PATHd8 is a non-clock, rate-smoothing method (Britton et al., 2007) to date
¹⁸⁷ trees. treePL, is a semi-parametric, rate-smoothing, penalized likelihood dating method
¹⁸⁸ (Stephen A. Smith & O'Meara, 2012). The MrBayes (Huelsenbeck & Ronquist, 2001;
¹⁸⁹ Ronquist & Huelsenbeck, 2003) approach in datelife uses the secondary calibrations as priors
¹⁹⁰ on node ages.

¹⁹¹ **Visualizing results.**— Finally, users can save all source and summary chronograms in
¹⁹² formats that permit reuse and reanalyses (newick and R “phylo” format), as well as visualize
¹⁹³ and compare results graphically, or construct their own graphs using `datelife`'s chronogram
¹⁹⁴ plot generation functions available from the R package `datelifeplot` (Sanchez-Reyes &
¹⁹⁵ O'Meara, 2022).

196

Benchmark

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

210

Case studies

211 We illustrate the DateLife algorithm using a group within the Passeriform birds
212 encompassing the family of true finches, Fringillidae as case study. The first example analyses
213 6 bird species and shows all steps of the algorithm. The second example is an application
214 analysing all 289 species in the family Fringillidae for which we have date informtaion.

215 Small example

216 We randomly chose 6 bird species in the family Fringillidae, the true finches. The
217 sample includes two species of cardinals: the black-thighed grosbeak – *Pheucticus tibialis*
218 and the crimson-collared grosbeak – *Rhodothraupis celaeno*; three species of buntings: the
219 yellowhammer – *Emberiza citrinella*, the pine bunting – *Emberiza leucocephalos* and the
220 yellow-throated bunting – *Emberiza elegans*; and one species of tanager, the vegetarian finch –

221 *Platyspiza crassirostris.*

222 Processing input names found that *Emberiza elegans* is synonym for *Schoeniclus*
223 *elegans* in the default reference taxonomy [Open Tree of Life Taxonomy v3.3, June 1, 2021].
224 For a detailed discussion on the state of the synonym refer to Avibase (Avibase, 2022;
225 Lepage, 2004; Lepage, Vaidya, & Guralnick, 2014). DateLife used the processed input names
226 to search the local chronogram database and found 9 matching chronograms in 6 different
227 studies. Three studies matched five input names (Barker, Burns, Klicka, Lanyon, & Lovette,
228 2015; Hedges, Marin, Suleski, Paymer, & Kumar, 2015; Jetz, Thomas, Joy, Hartmann, &
229 Mooers, 2012), one study matched four input names (Hooper & Price, 2017) and two studies
230 matched two input names (Barker, Burns, Klicka, Lanyon, & Lovette, 2013; Burns et al.,
231 2014). No studies matched all input names. Together, matching chronograms have 28 unique
232 age data points. All nodes have age data. DateLife used OpenTree's synthetic tree topology
233 for these taxa and mapped age data to nodes in the tree. As expected, more inclusive nodes
234 (e.g., node "n1") have more age data than less inclusive nodes (e.g., node "n5"). The
235 processing step allowed discovering five data points for node "n4" that would not have had
236 any data otherwise. **I don't understand what this sentence means** Age summary statistics
237 per node were calculated and used as secondary calibrations to date the tree topology using
238 the BLADJ algorithm. Age data for node "n2" was excluded as final calibration because it is
239 older than age data of a more inclusive node (n1).

240 **Fringillidae**

241 To estimate ages for the taxa in the family Fringillidae, we ran a `datelife` query using
242 the "get species from taxon" flag. This will get all recognised species names within the
243 named taxon. The Fringillidae has species, according to the Open Tree of Life taxonomy. ???
244 We have date estimates for 289 of those species.

245 ??? There is an option to include all, right? According to the OpenTree of life

²⁴⁶ synthetic phylogeny, Fringillidae is not a monophyletic family, so this taxon-constrained
²⁴⁷ approach returns a topology and ages for the species in that named group.

²⁴⁸ The next step is to run a `datelife` search on this curated set of species taxon names
²⁴⁹ that will find all chronograms that contain at least two species. The algorithm proceeds to
²⁵⁰ prune the trees to keep matching species names on tips only, and transform the pruned trees
²⁵¹ to pairwise distance matrices. There are 13 chronograms containing at least two Fringillidae
²⁵² species, published in 9 different studies (Barker et al., 2013, 2015; Burns et al., 2014;
²⁵³ Claramunt & Cracraft, 2015; Gibb et al., 2015; Hedges et al., 2015; Hooper & Price, 2017;
²⁵⁴ Jetz et al., 2012; Price et al., 2014). The final step is to summarize the available information
²⁵⁵ using two alternative types of summary chronograms, median and SDM. As explained in the
²⁵⁶ “Description” section, data from source chronograms is first summarised into a single
²⁵⁷ distance matrix and then the available node ages are used as fixed node calibrations over a
²⁵⁸ consensus tree topology, to obtain a fully dated tree with the program BLADJ (Fig. 5).
²⁵⁹ Median summary chronograms are older and have wider variation in maximum ages than
²⁶⁰ chronograms obtained with SDM.

²⁶¹ **Cross-validation test**

²⁶² We performed a cross validation analysis of the DateLife workflow using the
²⁶³ Fringillidae chronograms. We used the individual tree topologies from each of the ???
²⁶⁴ TODO CHECK NUMBER 19 source chronograms from ?? studies as inputs, treating their
²⁶⁵ node ages as unknown. We estimated dates for these topologies using the node ages from the
²⁶⁶ chronograms from the other studies as calibrations and smoothing using BLADJ.

²⁶⁷ We found that node ages from original study, and ages estimated using all other age
²⁶⁸ data available are correlated **R-square value?** (Supplementary Fig. 6).

²⁶⁹ In 5 studies, more inclusive nodes have older original ages, and less inclusive nodes have
²⁷⁰ younger original ages than their cross-validated age estimates. **more inclusive nodes have to**

271 be older! :P I think there is something differnet you are trying to say here - maybe "Datelife
272 tended to underestimate ages for more inclusive nodes relative to the orginal estimate, and
273 overestimate ages for nodes closer to the tips" Accordingly, root ages are generally older in
274 the original study than estimated using cross-validated ages (Supplementary Fig. 7).

275 **Discussion**

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

281 At the time of writing of this manuscript (May 07, 2022), **datelife**'s database has 253
282 chronograms, pulled entirely from OpenTree's database, the Phylesystem (McTavish et al.,
283 2015). A unique feature of OpenTree's Phylesystem is that the community can add new
284 state-of-the-art chronograms any time. As chronograms are added to Phylesystem, they are
285 incorporated into an updated **datelife**'s database that is assigned a new version number,
286 followed by a package release on CRAN. **datelife**'s chronogram database is updated as new
287 chronogram data is added to Phylesystem, at a minimum of once a month and a maximum
288 of every 6 months. Users can also upload new chronograms to OpenTree themselves, and
289 trigger an update of their local **datelife** database to incorporate the new chronograms, to
290 have them immediately available for analysis.

291 Incorporation of more chronograms into **datelife**'s database is crucial to improve its
292 services. One option to increase chronogram number in the database is the Dryad data
293 repository. Methods to automatically mine chronograms from Dryad could be designed and
294 implemented. However, Dryad's metadata system has no information to automatically detect
295 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).

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

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

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

345 Summarizing chronograms might also imply summarizing fundamentally distinct
346 evolutionary hypotheses. For example, two different researchers working on the same clade
347 both carefully select and argument their choices of fossil calibrations. Still, if one researcher

348 decides a fossil will calibrate the ingroup of a clade, while another researcher uses the same
349 one to calibrate outside the clade, the resulting age estimates will often differ substantially,
350 as the placement of calibrations as stem or crown group is proved to deeply affect estimated
351 times of lineage divergence (Sauquet, 2013). Trying to summarize the resulting chronograms
352 into a single one using simple summary statistics can erase many types of relevant
353 information from the source chronograms. Accordingly, the prevailing view is that we should
354 favor time of lineage divergence estimates obtained from a single analysis, using fossil data as
355 primary sources of calibrations, and using fossils that have been widely discussed and
356 curated as calibrations to date other trees, making sure that all data used in the analysis
357 reflect a coherent evolutionary history (Antonelli et al., 2017). However, the exercise of
358 summarizing different chronograms has the potential to help getting a single global
359 evolutionary history for a lineage by putting together evidence from different hypothesis.
360 Choosing the elements of the chronograms that we are going to keep and the ones that we
361 are going to discard is key, since we are potentially loosing important parts of the
362 evolutionary history of a lineage that might only be reflected in source chronograms and not
363 on the summary chronogram (Sauquet et al., 2021).

364 Nonetheless, in ecology and conservation biology, incorporating at least some data on
365 lineage divergence times represents a relevant improvement for testing alternative hypothesis
366 using phylogenetic distance (Campbell O. Webb et al., 2008). Hence, we integrated into
367 datelife's workflow different ways of estimating node ages in the absence of calibrations and
368 branch length information for taxa lacking this information. "Making up" branch lengths is
369 an accepted practice in scientific publications: Jetz et al. (2012), created a time-calibrated
370 tree of all 9,993 bird species, where 67% had molecular data and the rest was simulated;
371 Rabosky et al. (2018) created a time-calibrated tree of 31,536 ray-finned fishes, of which only
372 37% had molecular data; Stephen A. Smith and Brown (2018) constructed a tree of 353,185
373 seed plants where only 23% had molecular data. Obviously, there are risks in this practice!
374 Taken to the extreme, one could make a fully resolved, calibrated tree of all modern and

375 extinct taxa using a single taxonomy and a single calibration with the polytomy resolution
376 and branch estimation methods. There has yet to be a thorough analysis of what can go
377 wrong when one extends inferences beyond the data in this way, so we urge caution; we also
378 urge readers to follow the example of many of the large tree papers cited above and make
379 carefully consider the statistical assumptions being made, and assess the consistency of the
380 results with prior work.

381 Conclusions

382 Divergence time information is key to many areas of evolutionary studies: trait
383 evolution, diversification, biogeography, macroecology and more. It is also crucial for science
384 communication and education, but generating chronograms is difficult, especially for those
385 who want to use phylogenies but who are not systematists, or do not have the time to
386 acquire and develop the necessary knowledge and data curation skills. Moreover, years of
387 primarily public funded research have resulted in vast amounts of chronograms that are
388 already available on scientific publications, but hidden to the public and scientific community
389 for reuse.

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

400

Availability

401 `datelife` is free and open source and it can be used through its current website
402 <http://www.datelife.org>, through the `datelife` R package, and through Phylotastic's
403 project web portal <https://phylo.cs.nmsu.edu/>. `datelife`'s website is maintained using
404 RStudio's shiny server and the shiny package open infrastructure, as well as Docker.
405 `datelife`'s R package stable version is available for installation from the CRAN repository
406 (<https://cran.r-project.org/package=datelife>) using the command `install.packages(pkgs`
407 `= "datelife"`) from within R. Development versions are available from the GitHub
408 repository (<https://github.com/phylotastic/datelife>) and can be installed using the
409 command `devtools::install_github("phylotastic/datelife")`.

410

Supplementary Material

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

414

Funding

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

419

Acknowledgements

420 The DateLife project was born as a prototype tool aiming to provide these services,
421 and was developed over a series of hackathons at the National Evolutionary Synthesis
422 Center, NC, USA (Stoltzfus et al., 2013). We thank colleagues from the O'Meara Lab at the
423 University of Tennessee Knoxville for suggestions, discussions and software testing. The late

424 National Evolutionary Synthesis Center (NESCent), which sponsored hackathons that led to
425 initial work on this project. The team that assembled `datelife`'s first proof of concept:
426 Tracy Heath, Jonathan Eastman, Peter Midford, Joseph Brown, Matt Pennell, Mike Alfaro,
427 and Luke Harmon. The Open Tree of Life project that provides the open, metadata rich
428 repository of trees used for `datelife`. The many scientists who publish their chronograms in
429 an open, reusable form, and the scientists who curate them for deposition in the Open Tree
430 of Life repository. The NSF for funding nearly all the above, in addition to the ABI grant
431 that funded this project itself.

432

References

- 433 Ané, C., Eulenstein, O., Piaggio-Talice, R., & Sanderson, M. J. (2009). Groves of
434 phylogenetic trees. *Annals of Combinatorics*, 13(2), 139–167.
- 435 Antonelli, A., Hettling, H., Condamine, F. L., Vos, K., Nilsson, R. H., Sanderson, M.
436 J., ... Vos, R. A. (2017). Toward a self-updating platform for estimating rates of
437 speciation and migration, ages, and relationships of Taxa. *Systematic Biology*,
438 66(2), 153–166. <https://doi.org/10.1093/sysbio/syw066>
- 439 Archie, J., Day, W. H., Felsenstein, J., Maddison, W., Meacham, C., Rohlf, F. J., &
440 Swofford, D. (1986). The Newick tree format. Retrieved from
441 %7B<https://evolution.genetics.washington.edu/phylip/newicktree.html%7D>
- 442 Avibase. (2022). Yellow-throated Bunting. *Avibase - The World Bird Database*,
443 (Online Resource). Retrieved from %7B<https://avibase.bsc-eoc.org/species.jsp?lang=EN&avibaseid=82D1EE0049D8D927%7D>
- 444 Bapst, D. W. (2012). Paleotree: An R package for paleontological and phylogenetic
445 analyses of evolution. *Methods in Ecology and Evolution*, 3(5), 803–807.
446 <https://doi.org/10.1111/j.2041-210X.2012.00223.x>
- 447 Barba-Montoya, J., Reis, M. dos, Schneider, H., Donoghue, P. C., & Yang, Z. (2018).
448 Constraining uncertainty in the timescale of angiosperm evolution and the
449 veracity of a cretaceous terrestrial revolution. *New Phytologist*, 218(2), 819–834.
- 450 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2013). Going
451 to extremes: Contrasting rates of diversification in a recent radiation of new world
452 passerine birds. *Systematic Biology*, 62(2), 298–320.
- 453 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2015). New
454 insights into new world biogeography: An integrated view from the phylogeny of
455 blackbirds, cardinals, sparrows, tanagers, warblers, and allies. *The Auk: Ornithological Advances*, 132(2), 333–348.
- 456 Boyle, B., Hopkins, N., Lu, Z., Raygoza Garay, J. A., Mozzherin, D., Rees, T., ...
457

- 459 Enquist, B. J. (2013). The taxonomic name resolution service: An online tool for
460 automated standardization of plant names. *BMC Bioinformatics*, 14(1).
461 <https://doi.org/10.1186/1471-2105-14-16>
- 462 Britton, T., Anderson, C. L., Jacquet, D., Lundqvist, S., & Bremer, K. (2007).
463 Estimating Divergence Times in Large Phylogenetic Trees. *Systematic Biology*,
464 56(788777878), 741–752. <https://doi.org/10.1080/10635150701613783>
- 465 Burns, K. J., Shultz, A. J., Title, P. O., Mason, N. A., Barker, F. K., Klicka, J., ...
466 Lovette, I. J. (2014). Phylogenetics and diversification of tanagers (passeriformes:
467 Thraupidae), the largest radiation of neotropical songbirds. *Molecular
468 Phylogenetics and Evolution*, 75, 41–77.
- 469 Chamberlain, S. A., & Szöcs, E. (2013). taxize : taxonomic search and retrieval in R
470 [version 2; referees: 3 approved]. *F1000Research*, 2(191), 1–29.
471 <https://doi.org/10.12688/f1000research.2-191.v2>
- 472 Chamberlain, S. A., Szöcs, E., Foster, Z., Arendsee, Z., Boettiger, C., Ram, K., ...
473 Li, G. (2019). *taxize: Taxonomic information from around the web*. Retrieved
474 from <https://github.com/ropensci/taxize>
- 475 Claramunt, S., & Cracraft, J. (2015). A new time tree reveals earth history's imprint
476 on the evolution of modern birds. *Science Advances*, 1(11), e1501005.
- 477 Criscuolo, A., Berry, V., Douzery, E. J. P., & Gascuel, O. (2006). SDM: A fast
478 distance-based approach for (super)tree building in phylogenomics. *Systematic
479 Biology*, 55(5), 740–755. <https://doi.org/10.1080/10635150600969872>
- 480 Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M.-K., Turon, X., ...
481 Douzery, E. J. (2018). A phylogenomic framework and timescale for comparative
482 studies of tunicates. *BMC Biology*, 16(1), 1–14.
- 483 Eastman, J. M., Harmon, L. J., & Tank, D. C. (2013). Congruification: Support for
484 time scaling large phylogenetic trees. *Methods in Ecology and Evolution*, 4(7),
485 688–691. <https://doi.org/10.1111/2041-210X.12051>

- 486 Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and
487 high throughput. *Nucleic Acids Research*, 32(5), 1792–1797.
- 488 Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American
489 Naturalist*, 125(1), 1–15. Retrieved from <http://www.jstor.org/stable/2461605>
- 490 Gibb, G. C., England, R., Hartig, G., McLenachan, P. A., Taylor Smith, B. L.,
491 McComish, B. J., ... Penny, D. (2015). New zealand passerines help clarify the
492 diversification of major songbird lineages during the oligocene. *Genome Biology
493 and Evolution*, 7(11), 2983–2995.
- 494 Harmon, L., Weir, J., Brock, C., Glor, R., & Challenger, W. (2008). GEIGER:
495 investigating evolutionary radiations. *Bioinformatics*, 24, 129–131.
- 496 Hedges, S. B., Marin, J., Suleski, M., Paymer, M., & Kumar, S. (2015). Tree of life
497 reveals clock-like speciation and diversification. *Molecular Biology and Evolution*,
498 32(4), 835–845. <https://doi.org/10.1093/molbev/msv037>
- 499 Heibl, C. (2008). *PHYLOCH: R language tree plotting tools and interfaces to diverse
500 phylogenetic software packages*. Retrieved from
501 <http://www.christophheibl.de/Rpackages.html>
- 502 Hooper, D. M., & Price, T. D. (2017). Chromosomal inversion differences correlate
503 with range overlap in passerine birds. *Nature Ecology & Evolution*, 1(10), 1526.
- 504 Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of
505 phylogenetic trees. *Bioinformatics*, 17(8), 754–755.
506 <https://doi.org/10.1093/bioinformatics/17.8.754>
- 507 Jetz, W., Thomas, G., Joy, J. J. B., Hartmann, K., & Mooers, A. (2012). The global
508 diversity of birds in space and time. *Nature*, 491(7424), 444–448.
509 <https://doi.org/10.1038/nature11631>
- 510 Katoh, K., Asimenos, G., & Toh, H. (2009). Multiple alignment of DNA sequences
511 with MAFFT. In *Bioinformatics for DNA sequence analysis* (pp. 39–64).
512 Springer.

- 513 Laubichler, M. D., & Maienschein, J. (2009). *Form and function in developmental*
514 *evolution*. Cambridge University Press.
- 515 Lepage, D. (2004). *Avibase: The world bird database*. Bird Studies Canada.
- 516 Lepage, D., Vaidya, G., & Guralnick, R. (2014). Avibase—a database system for
517 managing and organizing taxonomic concepts. *ZooKeys*, (420), 117.
- 518 Magallon, S., & Sanderson, M. J. (2001). Absolute diversification rates in angiosperm
519 clades. *Evolution*, 55(9), 1762–1780.
- 520 Magallón, S. (2010). Using fossils to break long branches in molecular dating: A
521 comparison of relaxed clocks applied to the origin of angiosperms. *Systematic*
522 *Biology*, 59(4), 384–399.
- 523 Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., & Hernández-Hernández, T.
524 (2015). A metacalibrated time-tree documents the early rise of flowering plant
525 phylogenetic diversity. *New Phytologist*, 207(2), 437–453.
- 526 McTavish, E. J., Hinchliff, C. E., Allman, J. F., Brown, J. W., Cranston, K. A.,
527 Holder, M. T., . . . Smith, S. A. (2015). Phylesystem: A git-based data store for
528 community-curated phylogenetic estimates. *Bioinformatics*, 31(17), 2794–2800.
- 529 Michonneau, F., Brown, J. W., & Winter, D. J. (2016). rotl: an R package to interact
530 with the Open Tree of Life data. *Methods in Ecology and Evolution*, 7(12),
531 1476–1481. <https://doi.org/10.1111/2041-210X.12593>
- 532 Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecology*
533 *Letters*, 17(4), 508–525. <https://doi.org/10.1111/ele.12251>
- 534 Ooms, J., & Chamberlain, S. (2018). *Phylocomr: Interface to 'phylocom'*. Retrieved
535 from <https://CRAN.R-project.org/package=phylocomr>
- 536 Open Tree Of Life, Redelings, B., Cranston, K. A., Allman, J., Holder, M. T., &
537 McTavish, E. J. (2016). Open Tree of Life APIs v3.0. *Open Tree of Life Project*,
538 (Online Resources). Retrieved from
539 <https://github.com/OpenTreeOfLife/germinator/wiki/Open-Tree-of-Life->

- 540 Web-APIs%7D
- 541 Open Tree Of Life, Redelings, B., Sánchez Reyes, L. L., Cranston, K. A., Allman, J.,
542 Holder, M. T., & McTavish, E. J. (2019). Open tree of life synthetic tree v12.3.
543 *Zenodo*. Retrieved from <https://doi.org/10.5281/zenodo.3937742>
- 544 Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and
545 evolution in R language. *Bioinformatics*, 20(2), 289–290.
- 546 Posadas, P., Crisci, J. V., & Katinas, L. (2006). Historical biogeography: A review of
547 its basic concepts and critical issues. *Journal of Arid Environments*, 66(3),
548 389–403.
- 549 Price, T. D., Hooper, D. M., Buchanan, C. D., Johansson, U. S., Tietze, D. T.,
550 Alström, P., ... others. (2014). Niche filling slows the diversification of himalayan
551 songbirds. *Nature*, 509(7499), 222.
- 552 R Core Team. (2018). *R: a language and environment for statistical computing*.
553 Vienna, Austria: R Foundation for Statistical Computing.
- 554 Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., ...
555 others. (2018). An inverse latitudinal gradient in speciation rate for marine fishes.
556 *Nature*, 559(7714), 392.
- 557 Ramshaw, J., Richardson, D., Mealyard, B., Brown, R., Richardson, M., Thompson,
558 E., & Boulter, D. (1972). The time of origin of the flowering plants determined by
559 using amino acid sequence data of cytochrome c. *New Phytologist*, 71(5), 773–779.
- 560 Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The barcode of life data system
561 (<http://www.Barcodinglife.org>). *Molecular Ecology Notes*, 7(3), 355–364.
- 562 Rees, J. A., & Cranston, K. (2017). Automated assembly of a reference taxonomy for
563 phylogenetic data synthesis. *Biodiversity Data Journal*, (5).
- 564 Revell, L. J. (2012). Phytools: An r package for phylogenetic comparative biology
565 (and other things). *Methods in Ecology and Evolution*, 3, 217–223.
- 566 Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic

- 567 inference under mixed models. *Bioinformatics*, 19(12), 1572–1574.
- 568 <https://doi.org/10.1093/bioinformatics/btg180>
- 569 Sanchez-Reyes, L. L., & O'Meara, B. (2022). *datelifeplot*: Methods to plot
570 chronograms and outputs of the *datelife* package. *R Package Release V0.2.2*.
571 Retrieved from <https://zenodo.org/badge/latestdoi/381501451>
- 572 Sanchez-Reyes, L. L., O'Meara, B., Eastman, J., Heath, T., Wright, A., Schliep, K.,
573 ... Alfaro, M. (2022). *datelife*: Scientific Data on Time of Lineage Divergence
574 for Your Taxa. *R Package Release V0.6.2*. Retrieved from
575 <https://doi.org/10.5281/zenodo.593938>
- 576 Sanderson, M. J. (2003). r8s: Inferring absolute rates of molecular evolution and
577 divergence times in the absence of a molecular clock. *Bioinformatics*, 19(2),
578 301–302.
- 579 Sanderson, M. J., & Doyle, J. A. (2001). Sources of error and confidence intervals in
580 estimating the age of angiosperms from rbcL and 18S rDNA data. *American
581 Journal of Botany*, 88(8), 1499–1516.
- 582 Sauquet, H. (2013). A practical guide to molecular dating. *Comptes Rendus Palevol*,
583 12(6), 355–367.
- 584 Sauquet, H., Ramírez-Barahona, S., & Magallón, S. (2021). *The age of flowering
585 plants is unknown*.
- 586 Schenk, J. J. (2016). Consequences of secondary calibrations on divergence time
587 estimates. *PLoS ONE*, 11(1). <https://doi.org/10.1371/journal.pone.0148228>
- 588 Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in r. *Bioinformatics*, 27(4),
589 592–593.
- 590 Smith, Stephen A., & Brown, J. W. (2018). Constructing a broadly inclusive seed
591 plant phylogeny. *American Journal of Botany*, 105(3), 302–314.
- 592 Smith, Stephen A., & O'Meara, B. C. (2012). TreePL: Divergence time estimation
593 using penalized likelihood for large phylogenies. *Bioinformatics*, 28(20),

- 594 2689–2690. <https://doi.org/10.1093/bioinformatics/bts492>
- 595 Stoltzfus, A., Lapp, H., Matasci, N., Deus, H., Sidlauskas, B., Zmasek, C. M., ...
- 596 Jordan, G. (2013). Phylotastic! Making tree-of-life knowledge accessible, reusable
- 597 and convenient. *BMC Bioinformatics*, 14.
- 598 <https://doi.org/10.1186/1471-2105-14-158>
- 599 Title, P. O., & Rabosky, D. L. (2016). Do Macrophylogenies Yield Stable
- 600 Macroevolutionary Inferences? An Example from Squamate Reptiles. *Systematic*
- 601 *Biology*, syw102. <https://doi.org/10.1093/sysbio/syw102>
- 602 Vos, R. A., Balhoff, J. P., Caravas, J. A., Holder, M. T., Lapp, H., Maddison, W. P.,
- 603 ... others. (2012). NeXML: Rich, extensible, and verifiable representation of
- 604 comparative data and metadata. *Systematic Biology*, 61(4), 675–689.
- 605 Webb, C. (2000). Exploring the Phylogenetic Structure of Ecological Communities :
- 606 An Example for Rain Forest Trees. *The American Naturalist*, 156(2), 145–155.
- 607 Webb, Campbell O., Ackerly, D. D., & Kembel, S. W. (2008). Phylocom: Software for
- 608 the analysis of phylogenetic community structure and trait evolution.
- 609 *Bioinformatics*, 24(18), 2098–2100.
- 610 <https://doi.org/10.1093/bioinformatics/btn358>
- 611 Webb, Campbell O., & Donoghue, M. J. (2005). Phylomatic: Tree assembly for
- 612 applied phylogenetics. *Molecular Ecology Notes*, 5(1), 181–183.

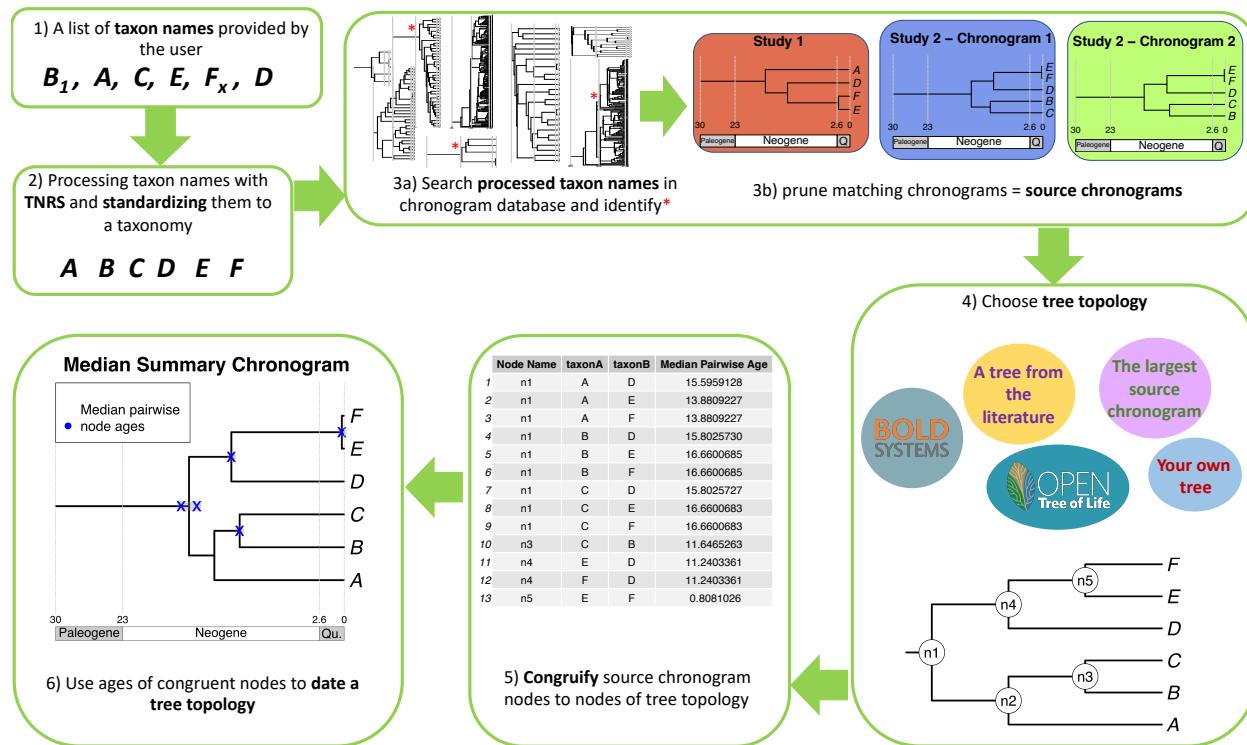


FIGURE 1. Stylized DateLife workflow. This shows the general workflows and analyses that can be performed with `datelife`, via the R package or through the website at <http://www.datelife.org/>. Details on the functions involved on each workflow are shown in `datelife`'s R package vignette.

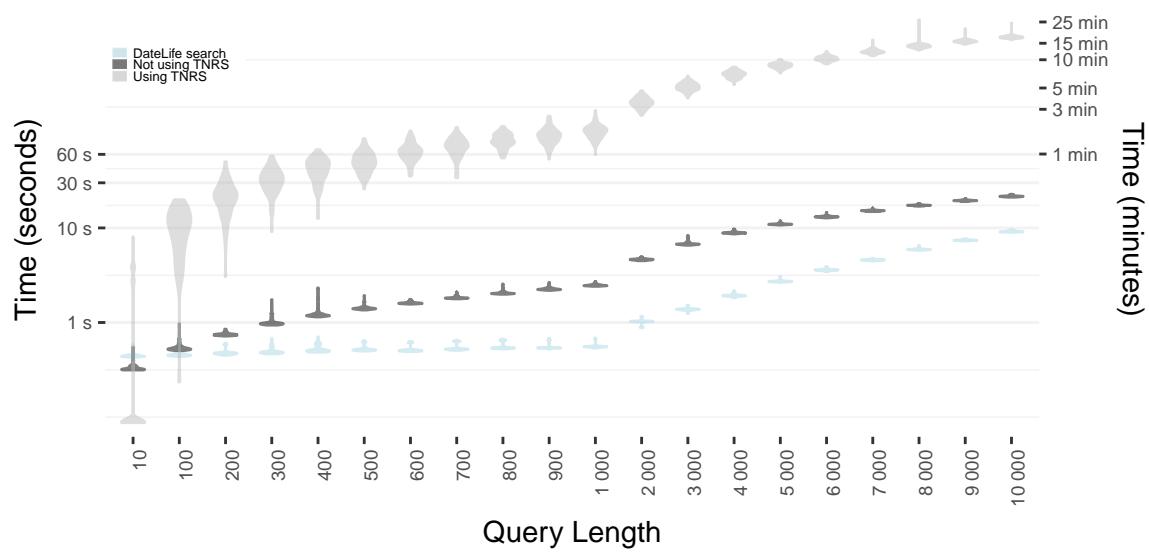


FIGURE 2. Computation time of query processing and search across **datelife**'s chronogram database relative to number of input taxon names. We sampled N names from the class Aves for each cohort 100 times and then performed a search with query processing not using the Taxon Names Resolution Service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

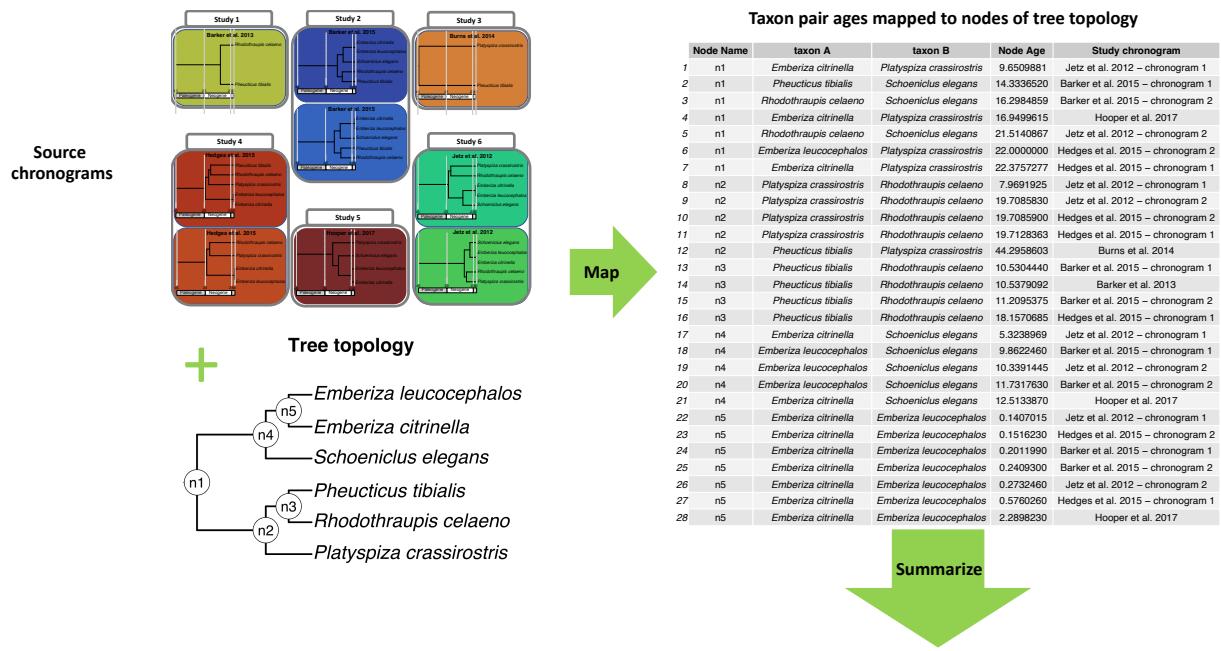


FIGURE 3. Age data results of a DateLife search of a small sample of 6 bird species within the Passeriformes. Input names were found across 9 chronograms within 6 independent studies (Barker et al. (2012), Barker et al. (2015), Burns et al. (2014), Hedges et al. (2015), Hooper and Price (2017), Jetz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

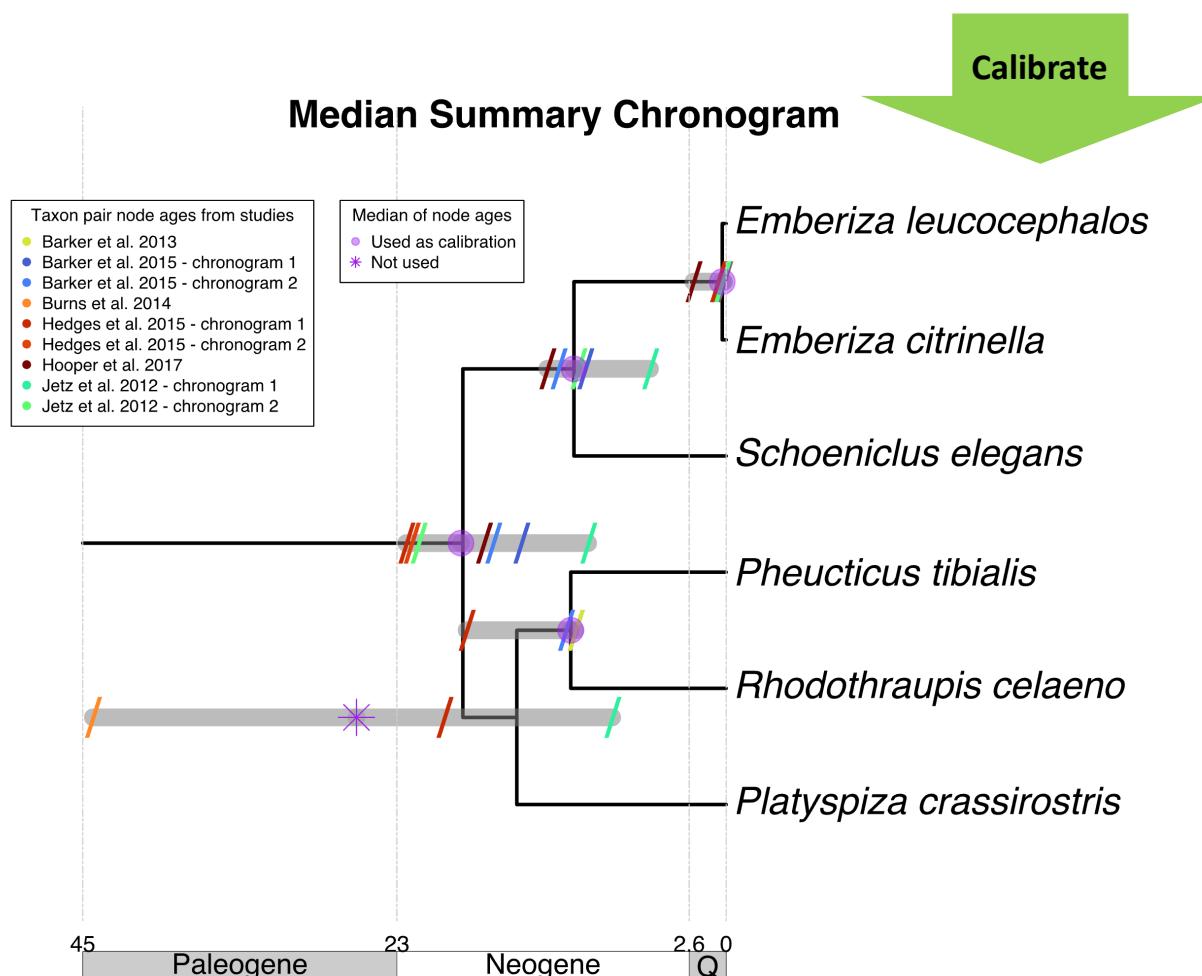


FIGURE 4. Summarized age data is used as secondary calibrations to date a tree topology as a summary chronogram.

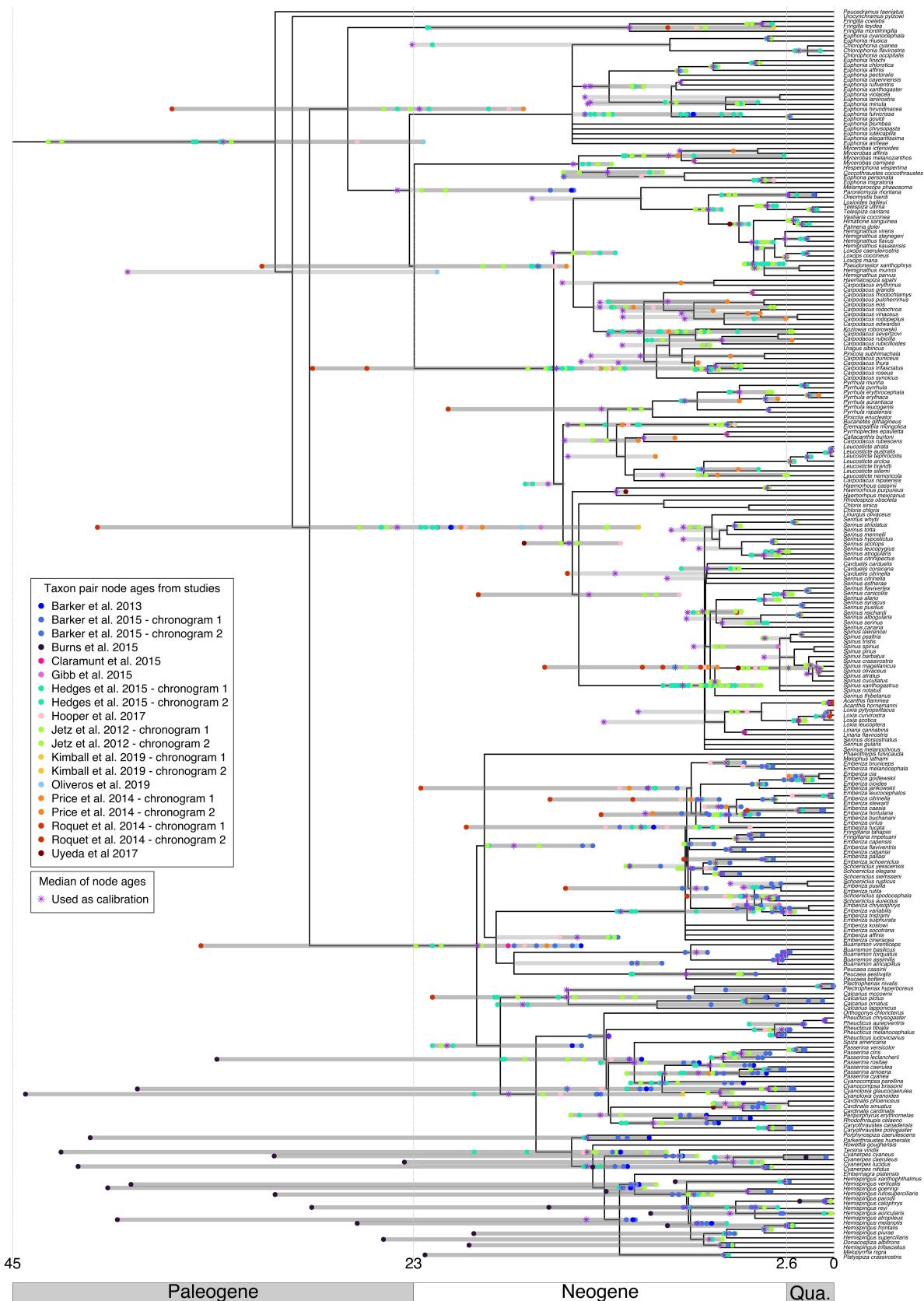


FIGURE 5. Fringillidae median summary chronogram generated with DateLife. It has 256 tips and 233 nodes.

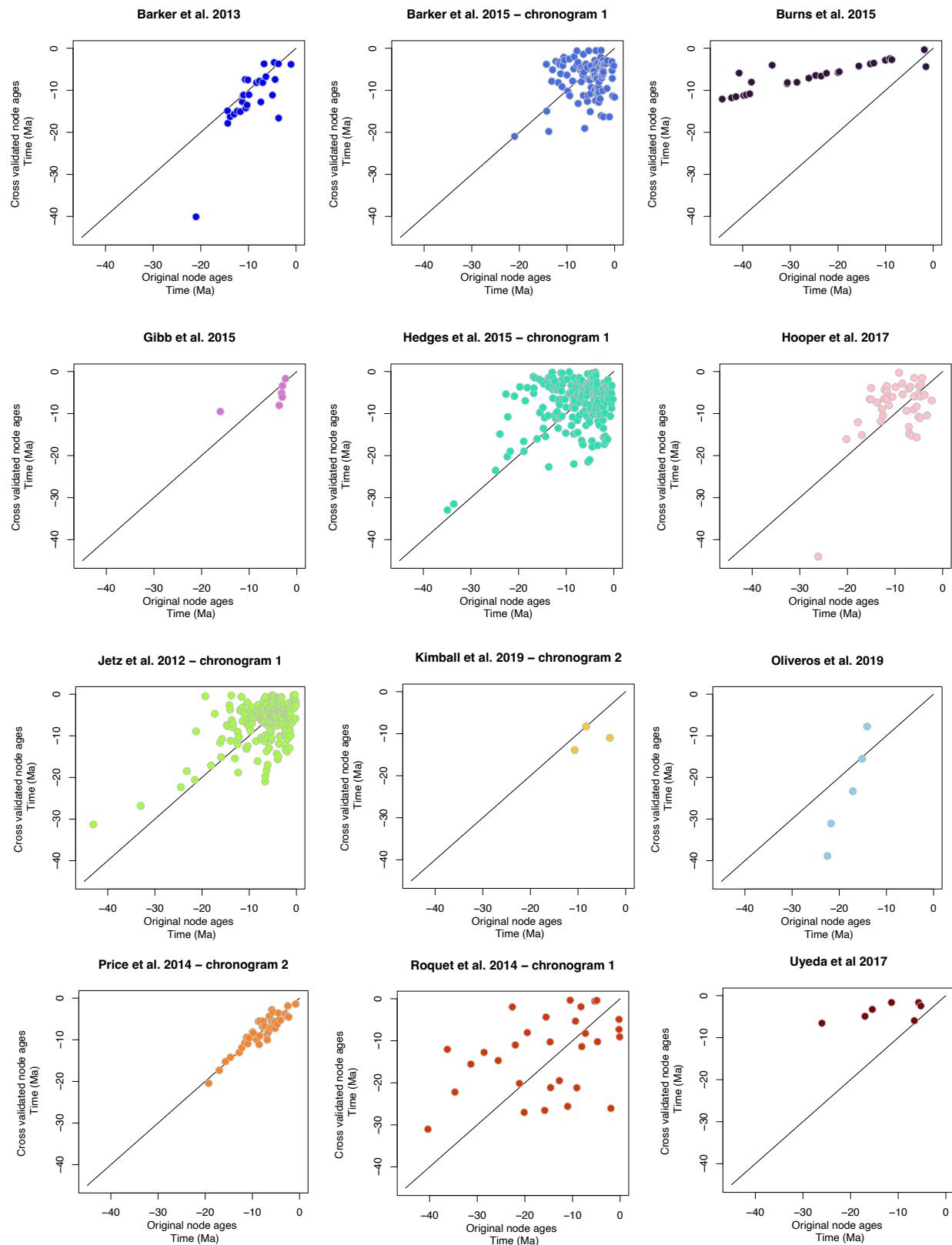


FIGURE 6. Results from cross validation analysis.

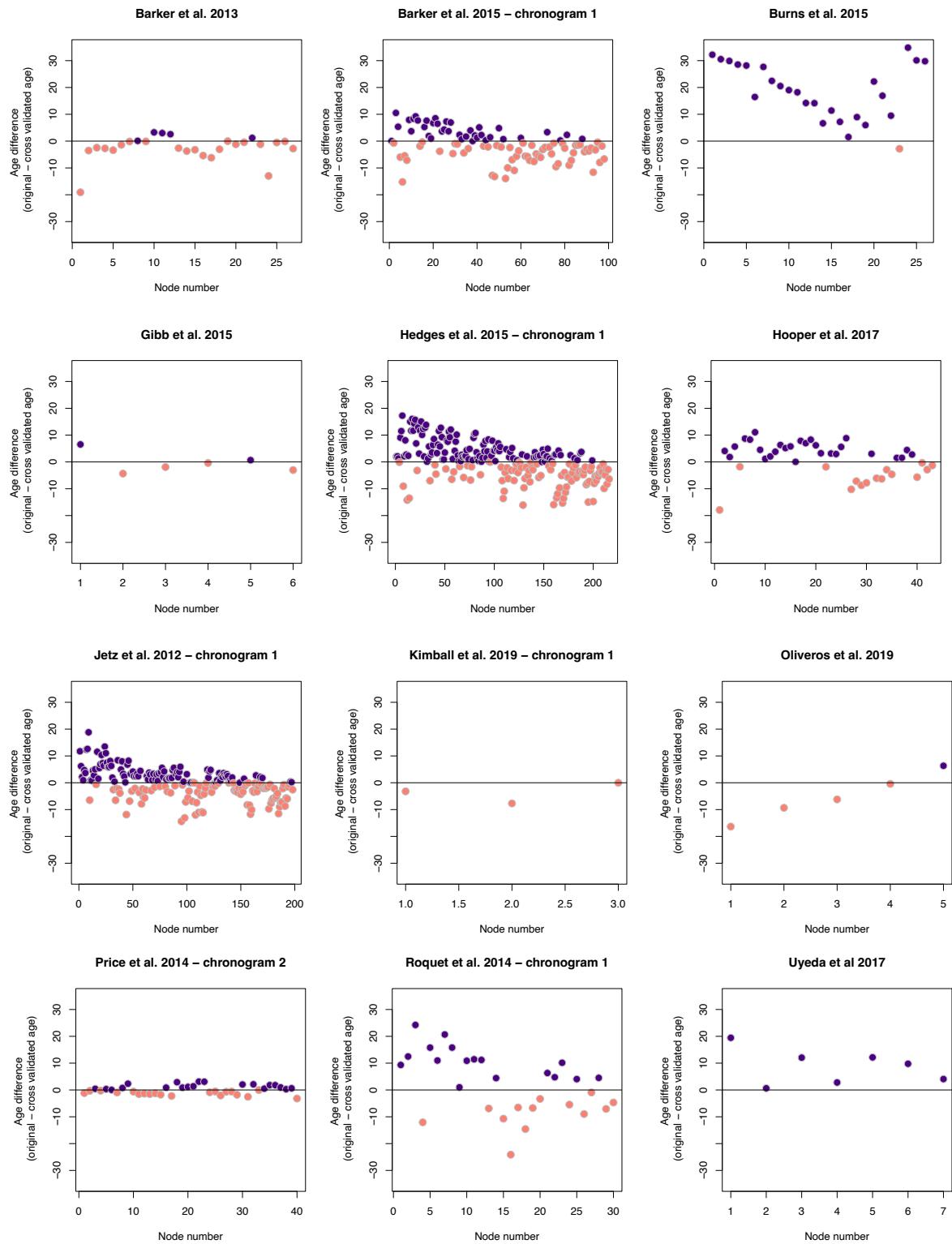


FIGURE 7. Results from cross validation analysis.

Barker et al. 2015 - chronogram 1

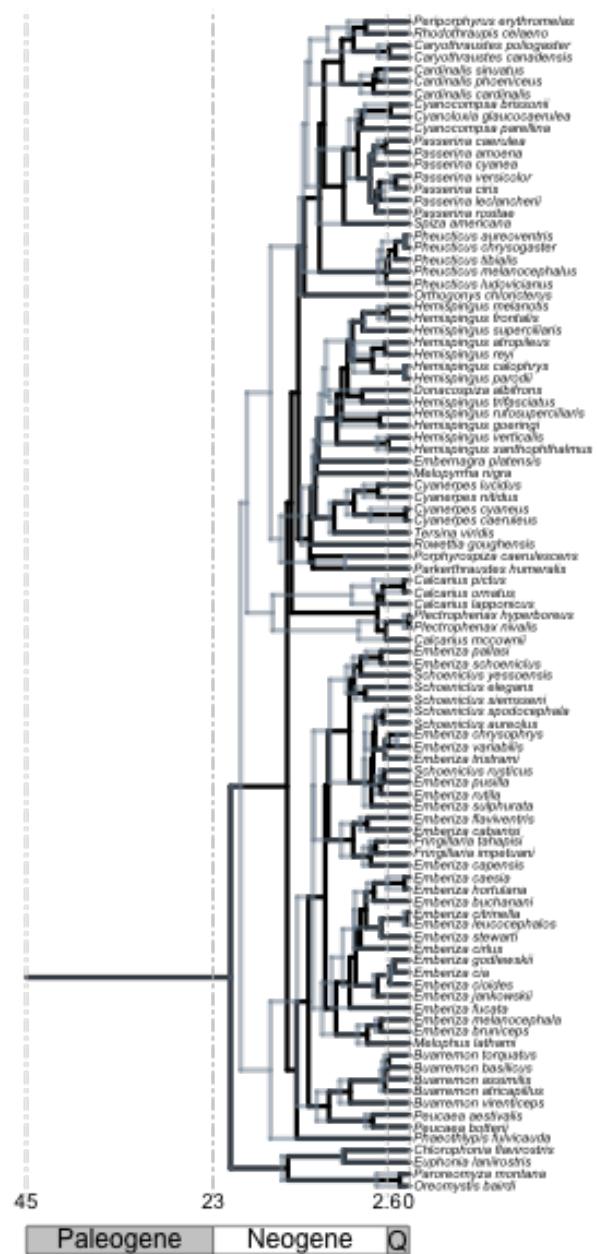


FIGURE 8. Cross validation of second source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to

Barker et al. 2015 - chronogram 2

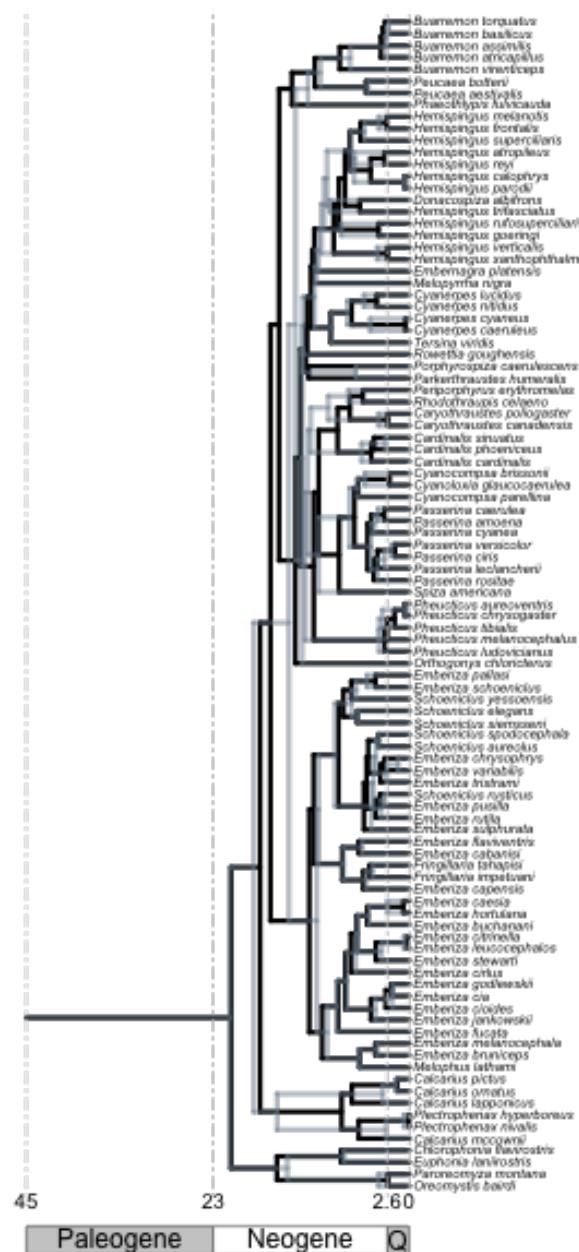


FIGURE 9. Cross validation of third source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the dates calculated with BLADJ, i.e., the same for all the nodes.

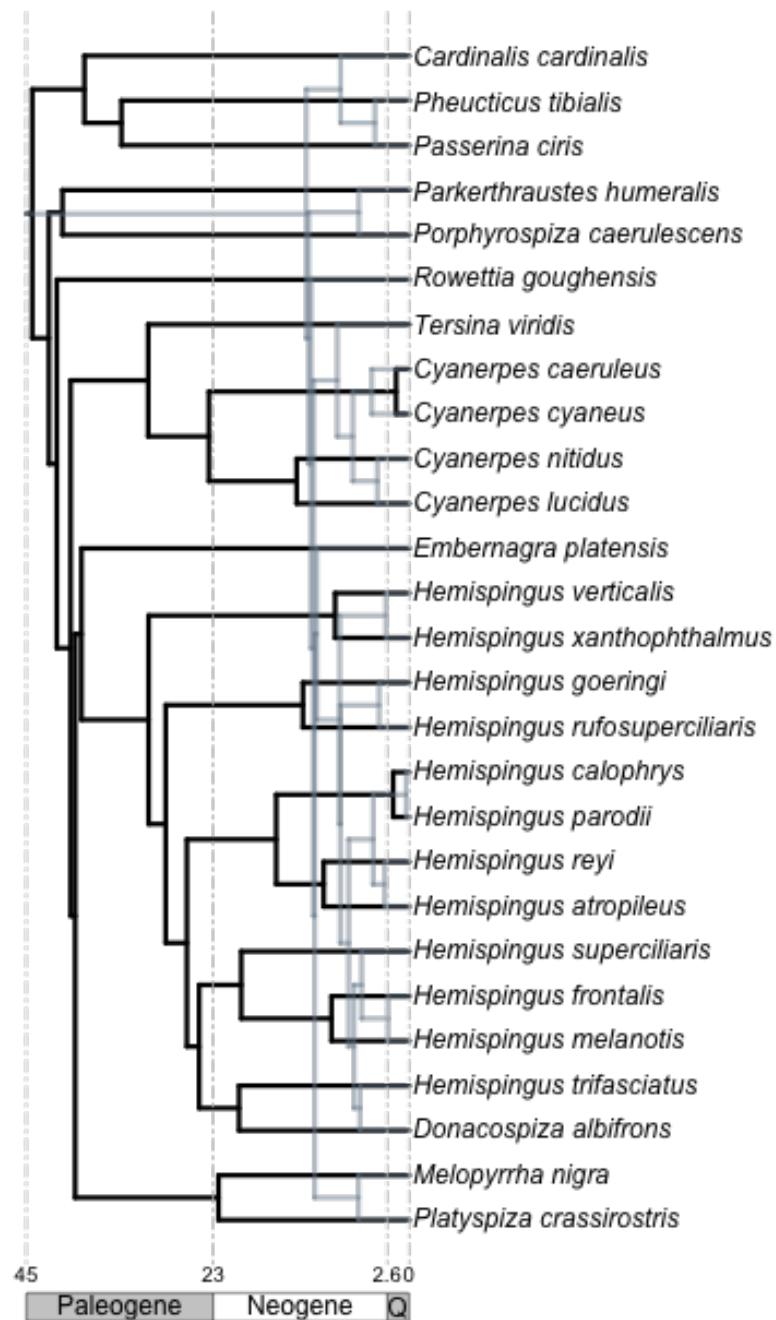
Burns et al. 2015

FIGURE 10. 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 in each run of the cross-validation.

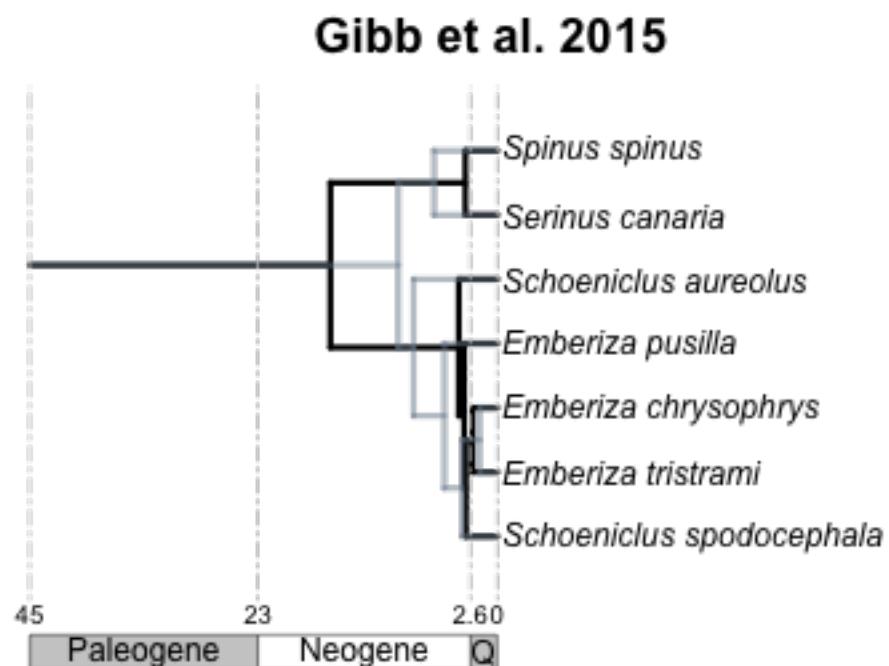


FIGURE 11. Cross validation of sixth source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the same tree topology dated with BLADJ using node ages from all other source chronograms as secondary calibrations.

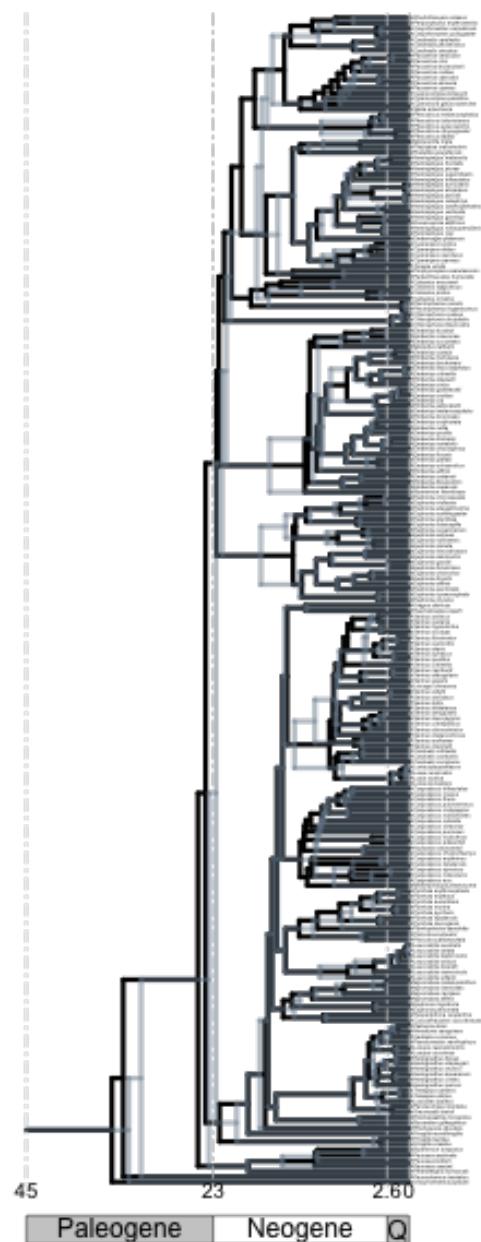
Hedges et al. 2015 - chronogram 1

FIGURE 12. 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 order to facilitate the comparison, the two chronograms are plotted side-by-side. The x-axis represents time in millions of years ago (Mya).

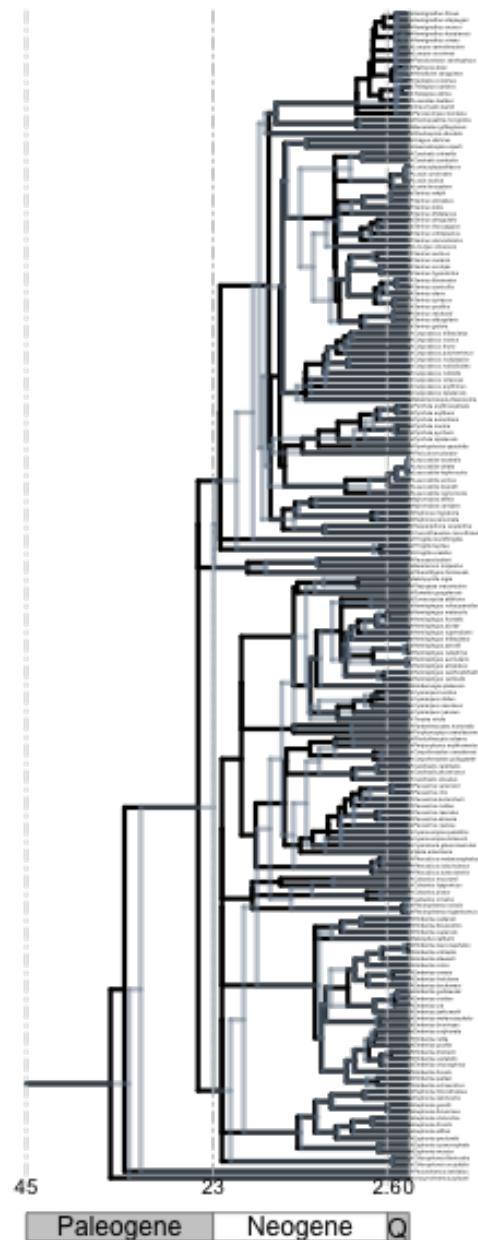
Hedges et al. 2015 - chronogram 2

FIGURE 13. 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

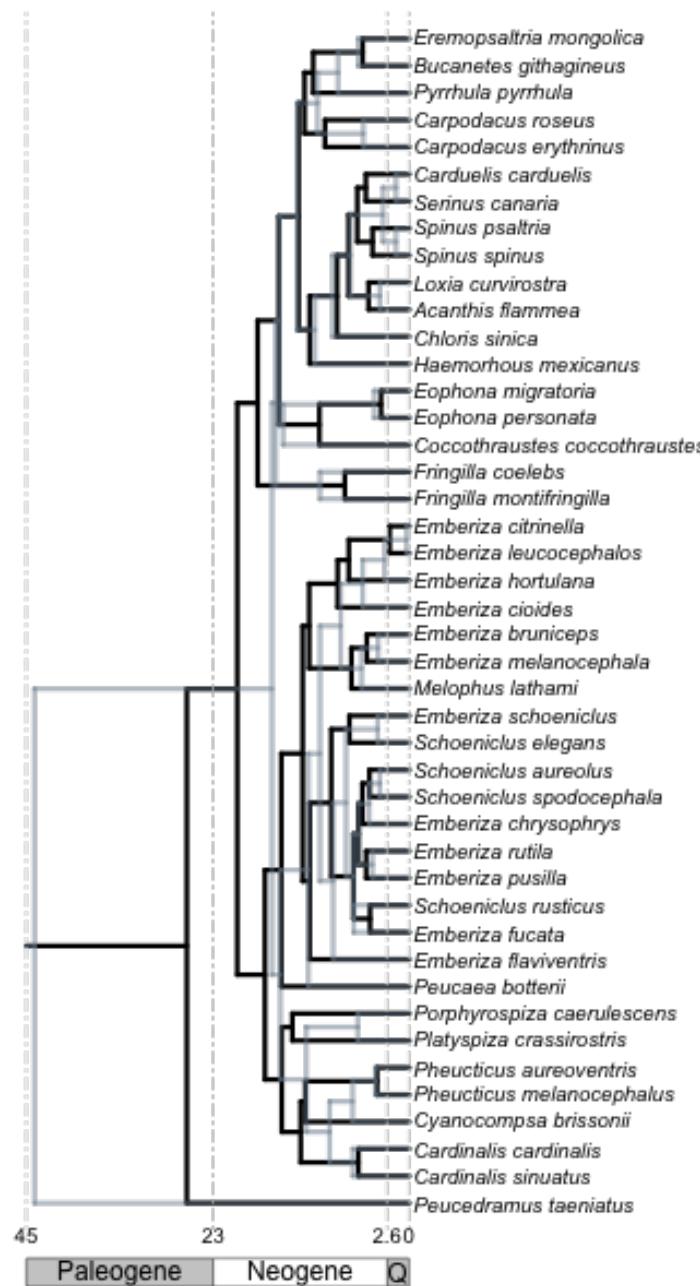


FIGURE 14. Cross validation of ninth source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the dates calculated with BLADJ.

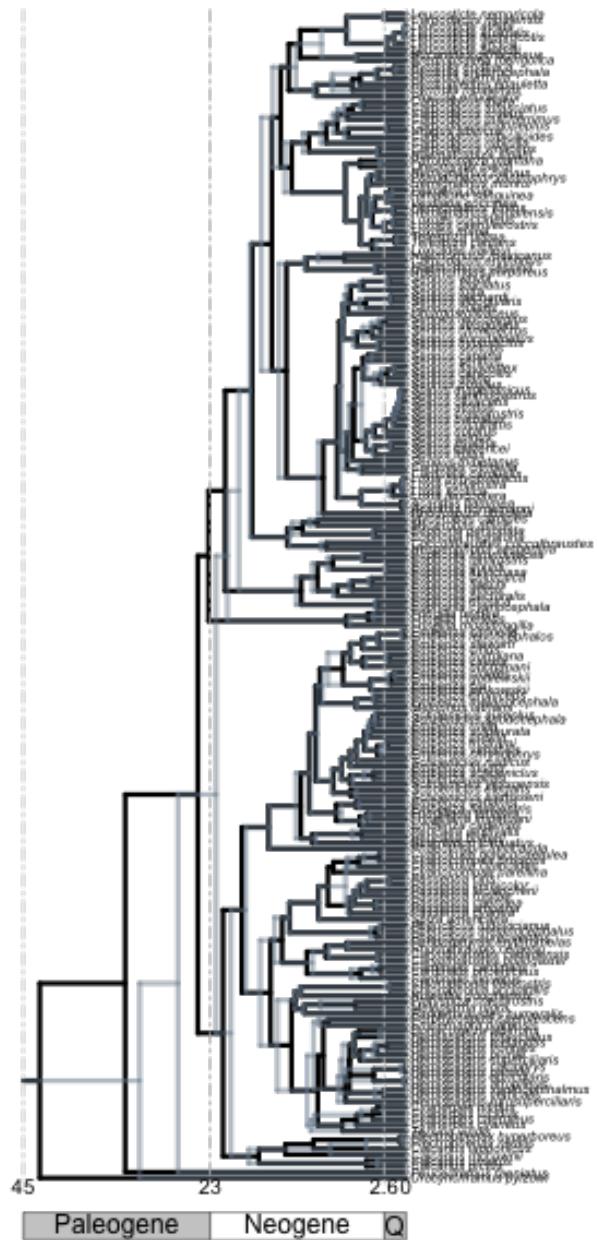
Jetz et al. 2012 - chronogram 1

FIGURE 15. 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.