

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: 4298

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** – v0.6.2, is available from the CRAN repository
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
132 own making. If no topology is provided, DateLife automatically subsets one from the
133 OpenTree synthetic tree (Open Tree Of Life et al., 2019). Alternatively, DateLife can
134 reconstruct a tree with branch lengths proportional to substitution rates from a starting tree
135 topology using genetic data from the Barcode of Life Data System, BOLD (Ratnasingham &
136 Hebert, 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 summary statistics.

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 studies

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.

Cross-validation test

To perform a cross validation analysis of the DateLife workflow, we used resulting data from the previous section (Casestudy: Real Life Application). We took individual tree topologies from each of the 19 source chronograms found (Supplementary data XX). Then we congruified age data of source chronograms from studies, and used this ages to date the tree topology with the program BLADJ.

We found that node ages from original study, and ages estimated using all other age data available are generally correlated (Supplementary Fig. 6). In 5 studies, more inclusive nodes have older original ages, and less inclusive nodes have younger original ages than their cross-validated age estimates. Accordingly, root ages are generally older in the original study than estimated using cross-validated ages. Root ages were similar in original and cross validated ages in three vases (Supplementary Fig. 7). Notably, chronograms have different species sampling, hence roots are not comparable across studies. Yet, chronograms with a higher sampling number can potentially inform the age of the root of chronograms with less sampling.

Discussion

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

At the time of writing of this manuscript (Apr 14, 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

271 incorporated into an updated `datelife`'s database that is assigned a new version number,
272 followed by a package release on CRAN. `datelife`'s chronogram database is updated as new
273 chronogram data is added to Phylesystem, at a minimum of once a month and a maximum
274 of every 6 months. Users can also upload new chronograms to OpenTree themselves, and
275 trigger an update of their local `datelife` database to incorporate the new chronograms, to
276 have them immediately available for analysis.

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

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

293 By default, `datelife` currently summarizes all source chronograms that overlap with
294 at least two species names. Users can exclude source chronograms if they have reasons to do
295 so. Strictly speaking, the best chronogram should reflect the real time of lineage divergence
296 accurately and precisely. To our knowledge, there are no good measures to determine

297 independently if a chronogram is better than another. Some measures that have been
298 proposed are the proportion of lineage sampling and the number of calibrations used
299 Magallón, Gómez-Acevedo, Sánchez-Reyes, & Hernández-Hernández (2015). Several
300 characteristics of the data used for dating analyses as well as from the output chronogram
301 itself, could be used to score quality of source chronograms. Some characteristics that are
302 often cited in published studies as a measure of improved age estimates as compared to
303 previously published estimates are: quality of alignment (missing data, GC content), lineage
304 sampling (strategy and proportion), phylogenetic and dating inference method, number of
305 fossils used as calibrations, support for nodes and ages, and magnitude of confidence
306 intervals. DateLife provides an opportunity to capture concordance and conflict among date
307 estimates, which can also be used as a metric for chronogram reliability.

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

319 Furthermore, even chronograms obtained with primary fossil data can vary
320 substantially in time estimates between lineages, as observed from the comparison of source
321 chronograms in the Fringillidae example. This observation is often encountered in the
322 literature (see, for example, the ongoing debate about crown group age of angiosperms

323 (Barba-Montoya, Reis, Schneider, Donoghue, & Yang, 2018; Magallón et al., 2015; Ramshaw
324 et al., 1972; Sanderson & Doyle, 2001; Sauquet, Ramírez-Barahona, & Magallón, 2021). For
325 some studies, especially ones based on branch lengths (e.g., studies of species diversification,
326 timing of evolutionary events, phenotypic trait evolution), using a different chronogram may
327 return different results (Title & Rabosky, 2016). Stitching together these chronograms can
328 create a larger tree that uses information from multiple studies, but the effect of
329 uncertainties and errors at this level on downstream analyses is still largely unknown.

330 Summarizing chronograms might also imply summarizing fundamentally distinct
331 evolutionary hypotheses. For example, two different researchers working on the same clade
332 both carefully select and argument their choices of fossil calibrations. Still, if one researcher
333 decides a fossil will calibrate the ingroup of a clade, while another researcher uses the same
334 one to calibrate outside the clade, the resulting age estimates will often differ substantially,
335 as the placement of calibrations as stem or crown group is proved to deeply affect estimated
336 times of lineage divergence (Sauquet, 2013). Trying to summarize the resulting chronograms
337 into a single one using simple summary statistics can erase many types of relevant
338 information from the source chronograms. Accordingly, the prevailing view is that we should
339 favor time of lineage divergence estimates obtained from a single analysis, using fossil data as
340 primary sources of calibrations, and using fossils that have been widely discussed and
341 curated as calibrations to date other trees, making sure that all data used in the analysis
342 reflect a coherent evolutionary history (Antonelli et al., 2017). However, the exercise of
343 summarizing different chronograms has the potential to help getting a single global
344 evolutionary history for a lineage by putting together evidence from different hypothesis.
345 Choosing the elements of the chronograms that we are going to keep and the ones that we
346 are going to discard is key, since we are potentially loosing important parts of the
347 evolutionary history of a lineage that might only be reflected in source chronograms and not
348 on the summary chronogram (Sauquet et al., 2021).

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

366 Conclusions

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

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

385 Availability

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

395 Supplementary Material

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

399

Funding

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

404

Acknowledgements

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

417

References

- 418 Ané, C., Eulenstein, O., Piaggio-Talice, R., & Sanderson, M. J. (2009). Groves of
419 phylogenetic trees. *Annals of Combinatorics*, 13(2), 139–167.
- 420 Antonelli, A., Hettling, H., Condamine, F. L., Vos, K., Nilsson, R. H., Sanderson, M.
421 J., . . . Vos, R. A. (2017). Toward a self-updating platform for estimating rates of
422 speciation and migration, ages, and relationships of Taxa. *Systematic Biology*,
423 66(2), 153–166. <https://doi.org/10.1093/sysbio/syw066>
- 424 Archie, J., Day, W. H., Felsenstein, J., Maddison, W., Meacham, C., Rohlf, F. J., &
425 Swofford, D. (1986). The Newick tree format. Retrieved from
426 %7B<https://evolution.genetics.washington.edu/phylip/newicktree.html%7D>
- 427 Avibase. (2022). Yellow-throated Bunting. *Avibase - The World Bird Database*,
428 (Online Resource). Retrieved from %7B<https://avibase.bsc-eoc.org/species.jsp?lang=EN&avibaseid=82D1EE0049D8D927%7D>
- 429 Bapst, D. W. (2012). Paleotree: An R package for paleontological and phylogenetic
430 analyses of evolution. *Methods in Ecology and Evolution*, 3(5), 803–807.
431 <https://doi.org/10.1111/j.2041-210X.2012.00223.x>
- 432 Barba-Montoya, J., Reis, M. dos, Schneider, H., Donoghue, P. C., & Yang, Z. (2018).
433 Constraining uncertainty in the timescale of angiosperm evolution and the
434 veracity of a cretaceous terrestrial revolution. *New Phytologist*, 218(2), 819–834.
- 435 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2013). Going
436 to extremes: Contrasting rates of diversification in a recent radiation of new world
437 passerine birds. *Systematic Biology*, 62(2), 298–320.
- 438 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2015). New
439 insights into new world biogeography: An integrated view from the phylogeny of
440 blackbirds, cardinals, sparrows, tanagers, warblers, and allies. *The Auk: Ornithological Advances*, 132(2), 333–348.
- 441 Britton, T., Anderson, C. L., Jacquet, D., Lundqvist, S., & Bremer, K. (2007).
442

- 444 Estimating Divergence Times in Large Phylogenetic Trees. *Systematic Biology*,
445 56(788777878), 741–752. <https://doi.org/10.1080/10635150701613783>
- 446 Burns, K. J., Shultz, A. J., Title, P. O., Mason, N. A., Barker, F. K., Klicka, J., ...
447 Lovette, I. J. (2014). Phylogenetics and diversification of tanagers (passeriformes:
448 Thraupidae), the largest radiation of neotropical songbirds. *Molecular
449 Phylogenetics and Evolution*, 75, 41–77.
- 450 Chamberlain, S. A., & Szöcs, E. (2013). taxize : taxonomic search and retrieval in R
451 [version 2; referees: 3 approved]. *F1000Research*, 2(191), 1–29.
452 <https://doi.org/10.12688/f1000research.2-191.v2>
- 453 Chamberlain, S. A., Szöcs, E., Foster, Z., Arendsee, Z., Boettiger, C., Ram, K., ...
454 Li, G. (2019). *taxize: Taxonomic information from around the web*. Retrieved
455 from <https://github.com/ropensci/taxize>
- 456 Claramunt, S., & Cracraft, J. (2015). A new time tree reveals earth history's imprint
457 on the evolution of modern birds. *Science Advances*, 1(11), e1501005.
- 458 Criscuolo, A., Berry, V., Douzery, E. J. P., & Gascuel, O. (2006). SDM: A fast
459 distance-based approach for (super)tree building in phylogenomics. *Systematic
460 Biology*, 55(5), 740–755. <https://doi.org/10.1080/10635150600969872>
- 461 Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M.-K., Turon, X., ...
462 Douzery, E. J. (2018). A phylogenomic framework and timescale for comparative
463 studies of tunicates. *BMC Biology*, 16(1), 1–14.
- 464 Eastman, J. M., Harmon, L. J., & Tank, D. C. (2013). Congruification: Support for
465 time scaling large phylogenetic trees. *Methods in Ecology and Evolution*, 4(7),
466 688–691. <https://doi.org/10.1111/2041-210X.12051>
- 467 Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and
468 high throughput. *Nucleic Acids Research*, 32(5), 1792–1797.
- 469 Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American
470 Naturalist*, 125(1), 1–15. Retrieved from <http://www.jstor.org/stable/2461605>

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

- 498 managing and organizing taxonomic concepts. *ZooKeys*, (420), 117.
- 499 Magallon, S., & Sanderson, M. J. (2001). Absolute diversification rates in angiosperm
500 clades. *Evolution*, 55(9), 1762–1780.
- 501 Magallón, S. (2010). Using fossils to break long branches in molecular dating: A
502 comparison of relaxed clocks applied to the origin of angiosperms. *Systematic
503 Biology*, 59(4), 384–399.
- 504 Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., & Hernández-Hernández, T.
505 (2015). A metacalibrated time-tree documents the early rise of flowering plant
506 phylogenetic diversity. *New Phytologist*, 207(2), 437–453.
- 507 McTavish, E. J., Hinchliff, C. E., Allman, J. F., Brown, J. W., Cranston, K. A.,
508 Holder, M. T., . . . Smith, S. A. (2015). Phylesystem: A git-based data store for
509 community-curated phylogenetic estimates. *Bioinformatics*, 31(17), 2794–2800.
- 510 Michonneau, F., Brown, J. W., & Winter, D. J. (2016). rotl: an R package to interact
511 with the Open Tree of Life data. *Methods in Ecology and Evolution*, 7(12),
512 1476–1481. <https://doi.org/10.1111/2041-210X.12593>
- 513 Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecology
514 Letters*, 17(4), 508–525. <https://doi.org/10.1111/ele.12251>
- 515 Ooms, J., & Chamberlain, S. (2018). *Phylocomr: Interface to 'phylocom'*. Retrieved
516 from <https://CRAN.R-project.org/package=phylocomr>
- 517 Open Tree Of Life, Redelings, B., Cranston, K. A., Allman, J., Holder, M. T., &
518 McTavish, E. J. (2016). Open Tree of Life APIs v3.0. *Open Tree of Life Project*,
519 (Online Resources). Retrieved from
520 [https://github.com/OpenTreeOfLife/germinator/wiki/Open-Tree-of-Life-
Web-APIs%7D](https://github.com/OpenTreeOfLife/germinator/wiki/Open-Tree-of-Life-
521 Web-APIs%7D)
- 522 Open Tree Of Life, Redelings, B., Sánchez Reyes, L. L., Cranston, K. A., Allman, J.,
523 Holder, M. T., & McTavish, E. J. (2019). Open tree of life synthetic tree v12.3.
524 *Zenodo*. Retrieved from <https://doi.org/10.5281/zenodo.3937742>

- 525 Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and
526 evolution in R language. *Bioinformatics*, 20(2), 289–290.
- 527 Posadas, P., Crisci, J. V., & Katinas, L. (2006). Historical biogeography: A review of
528 its basic concepts and critical issues. *Journal of Arid Environments*, 66(3),
529 389–403.
- 530 Price, T. D., Hooper, D. M., Buchanan, C. D., Johansson, U. S., Tietze, D. T.,
531 Alström, P., ... others. (2014). Niche filling slows the diversification of himalayan
532 songbirds. *Nature*, 509(7499), 222.
- 533 R Core Team. (2018). *R: a language and environment for statistical computing*.
534 Vienna, Austria: R Foundation for Statistical Computing.
- 535 Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., ...
536 others. (2018). An inverse latitudinal gradient in speciation rate for marine fishes.
537 *Nature*, 559(7714), 392.
- 538 Ramshaw, J., Richardson, D., Mealyard, B., Brown, R., Richardson, M., Thompson,
539 E., & Boulter, D. (1972). The time of origin of the flowering plants determined by
540 using amino acid sequence data of cytochrome c. *New Phytologist*, 71(5), 773–779.
- 541 Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The barcode of life data system
542 (<http://www.Barcodinglife.org>). *Molecular Ecology Notes*, 7(3), 355–364.
- 543 Rees, J. A., & Cranston, K. (2017). Automated assembly of a reference taxonomy for
544 phylogenetic data synthesis. *Biodiversity Data Journal*, (5).
- 545 Revell, L. J. (2012). Phytools: An r package for phylogenetic comparative biology
546 (and other things). *Methods in Ecology and Evolution*, 3, 217–223.
- 547 Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic
548 inference under mixed models. *Bioinformatics*, 19(12), 1572–1574.
549 <https://doi.org/10.1093/bioinformatics/btg180>
- 550 Sanchez-Reyes, L. L., O'Meara, B., Eastman, J., Heath, T., Wright, A., Schliep, K.,
551 ... Alfaro, M. (2022). datelife: Scientific Data on Time of Lineage Divergence for

- 552 Your Taxa. *R Package Version 0.6.2*. Retrieved from
553 <https://doi.org/10.5281/zenodo.593938>
- 554 Sanderson, M. J. (2003). r8s: Inferring absolute rates of molecular evolution and
555 divergence times in the absence of a molecular clock. *Bioinformatics*, 19(2),
556 301–302.
- 557 Sanderson, M. J., & Doyle, J. A. (2001). Sources of error and confidence intervals in
558 estimating the age of angiosperms from rbcL and 18S rDNA data. *American*
559 *Journal of Botany*, 88(8), 1499–1516.
- 560 Sauquet, H. (2013). A practical guide to molecular dating. *Comptes Rendus Palevol*,
561 12(6), 355–367.
- 562 Sauquet, H., Ramírez-Barahona, S., & Magallón, S. (2021). The age of flowering
563 plants is unknown.
- 564 Schenk, J. J. (2016). Consequences of secondary calibrations on divergence time
565 estimates. *PLoS ONE*, 11(1). <https://doi.org/10.1371/journal.pone.0148228>
- 566 Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in r. *Bioinformatics*, 27(4),
567 592–593.
- 568 Smith, Stephen A., & Brown, J. W. (2018). Constructing a broadly inclusive seed
569 plant phylogeny. *American Journal of Botany*, 105(3), 302–314.
- 570 Smith, Stephen A., & O'Meara, B. C. (2012). TreePL: Divergence time estimation
571 using penalized likelihood for large phylogenies. *Bioinformatics*, 28(20),
572 2689–2690. <https://doi.org/10.1093/bioinformatics/bts492>
- 573 Stoltzfus, A., Lapp, H., Matasci, N., Deus, H., Sidlauskas, B., Zmasek, C. M., ...
574 Jordan, G. (2013). Phylotastic! Making tree-of-life knowledge accessible, reusable
575 and convenient. *BMC Bioinformatics*, 14.
576 <https://doi.org/10.1186/1471-2105-14-158>
- 577 Title, P. O., & Rabosky, D. L. (2016). Do Macrophylogenies Yield Stable
578 Macroevolutionary Inferences? An Example from Squamate Reptiles. *Systematic*

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

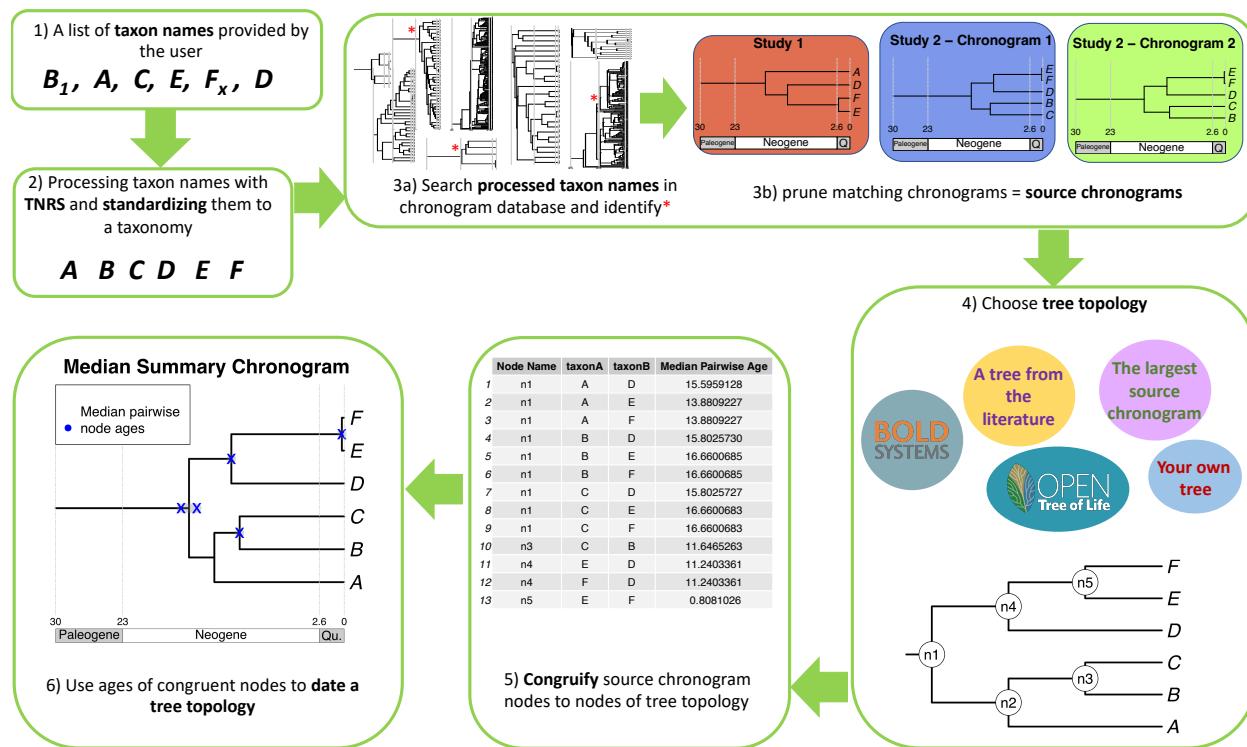


FIGURE 1. Stylized DateLife workflow. This shows the general workflows and analyses that can be performed with `datelife`, via the R package or through the website at <http://www.datelife.org/>. Details on the functions involved on each workflow are shown in `datelife`'s R package vignette.

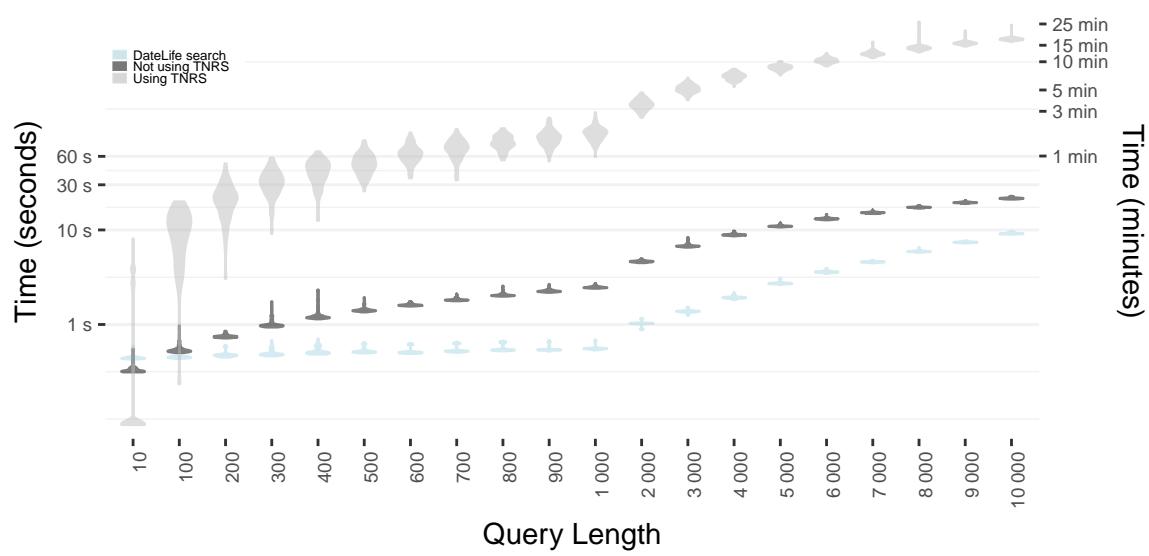


FIGURE 2. Computation time of query processing and search across **datelife**'s chronogram database relative to number of input taxon names. We sampled N names from the class Aves for each cohort 100 times and then performed a search with query processing not using the Taxon Names Resolution Service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

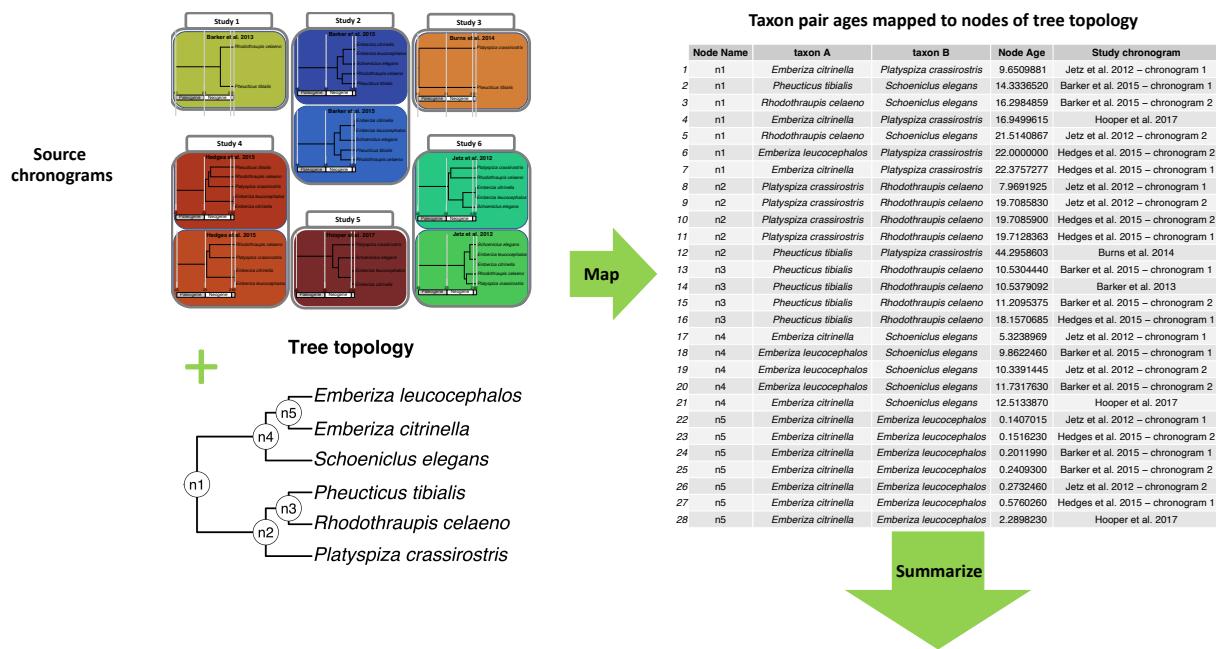


FIGURE 3. Age data results of a DateLife search of a small sample of 6 bird species within the Passeriformes. Input names were found across 9 chronograms within 6 independent studies (Barker et al. (2012), Barker et al. (2015), Burns et al. (2014), Hedges et al. (2015), Hooper and Price (2017), Jetz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

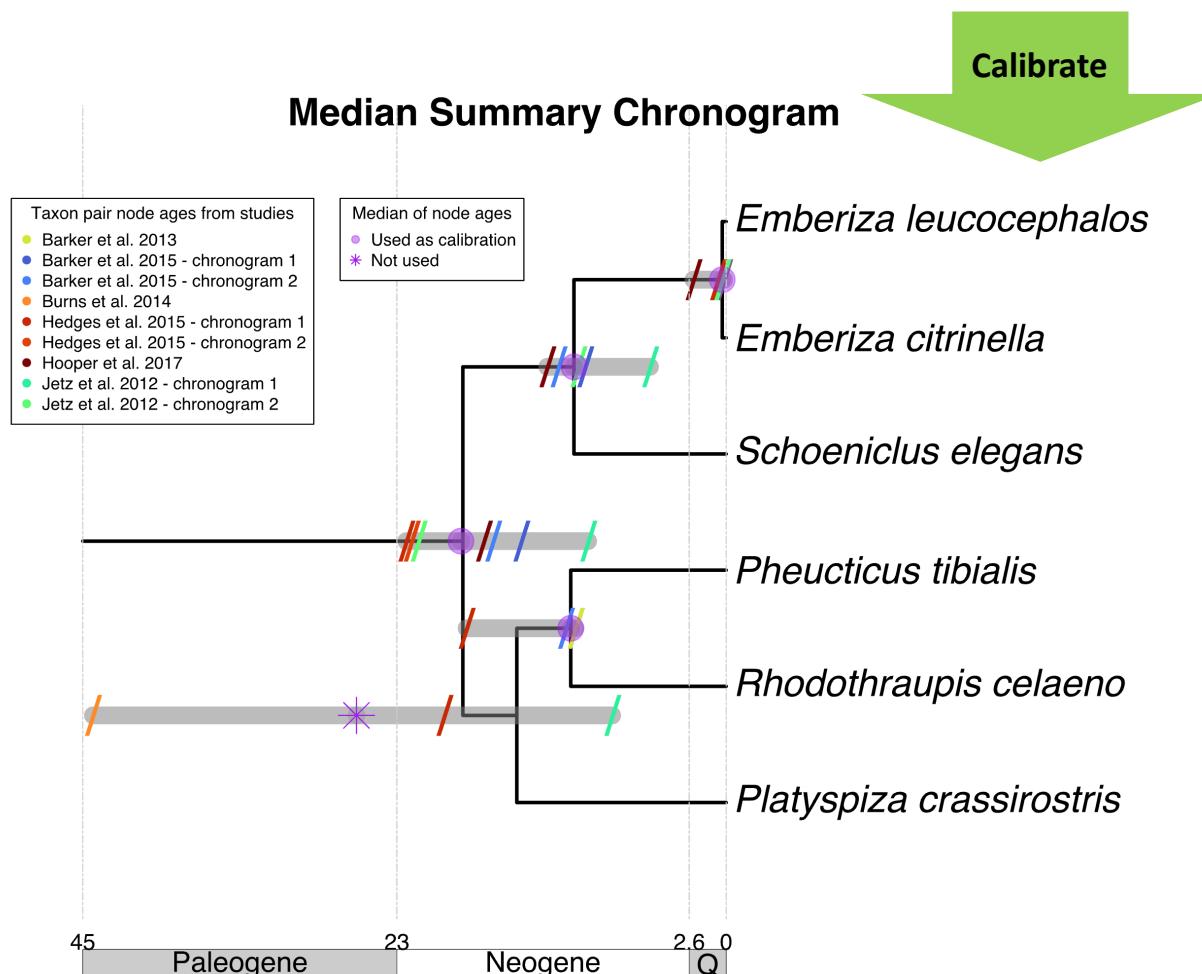


FIGURE 4. Summarized age data is used as secondary calibrations to date a tree topology as a summary chronogram.

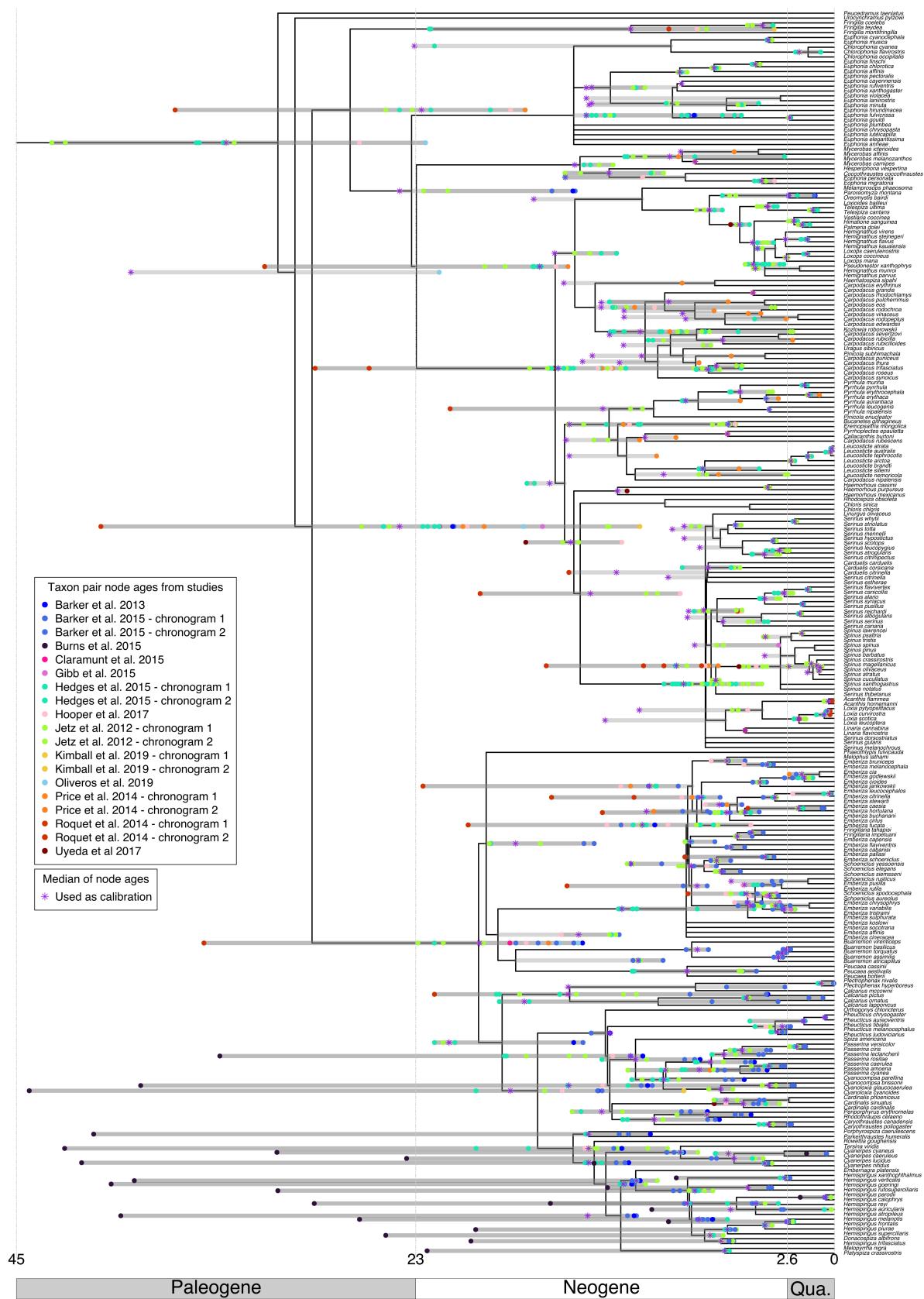


FIGURE 5. Fringillidae median summary chronogram generated with DateLife. It has 256 tips and 233 nodes.

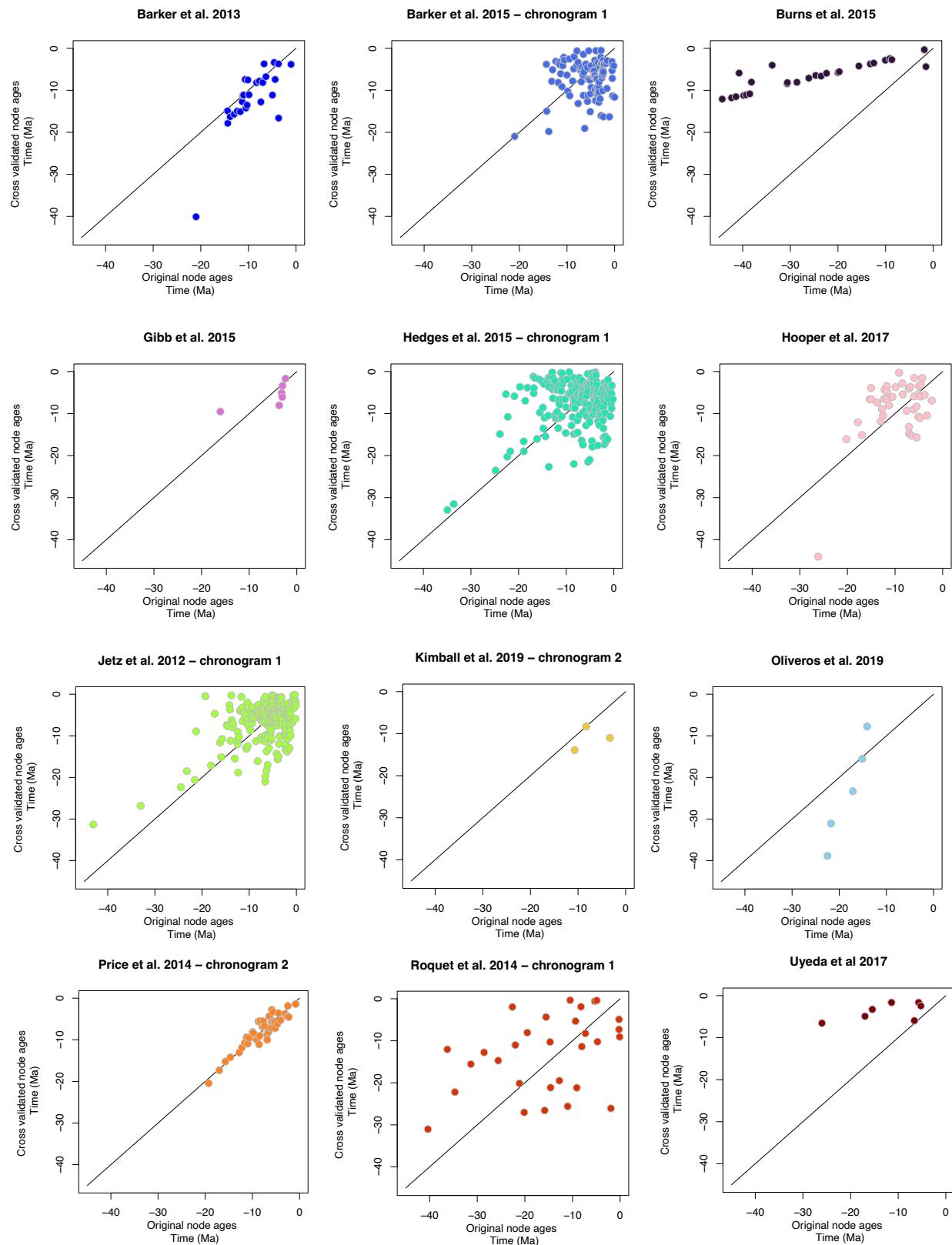


FIGURE 6. Results from cross validation analysis.

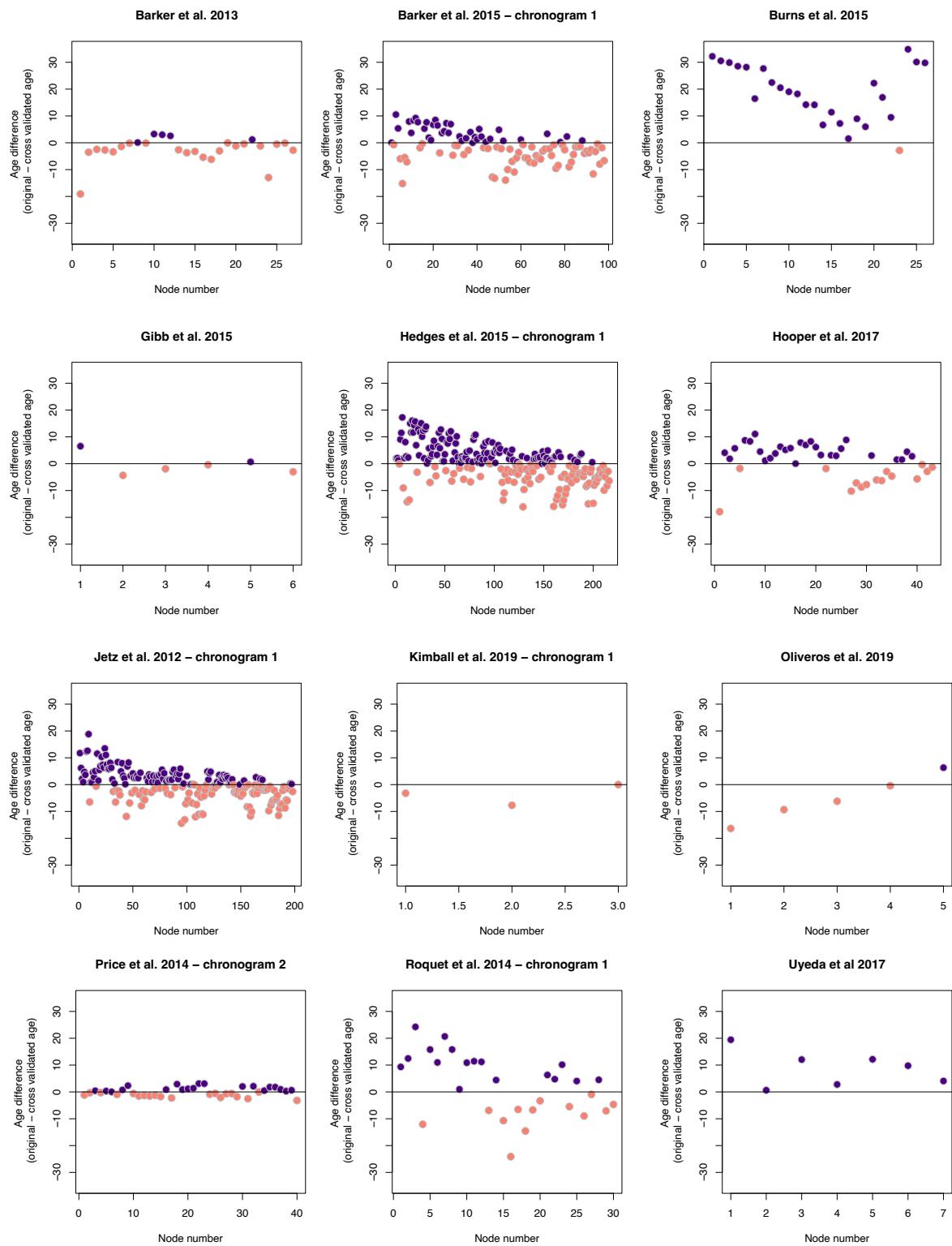


FIGURE 7. Results from cross validation analysis.

Barker et al. 2015 - chronogram 1

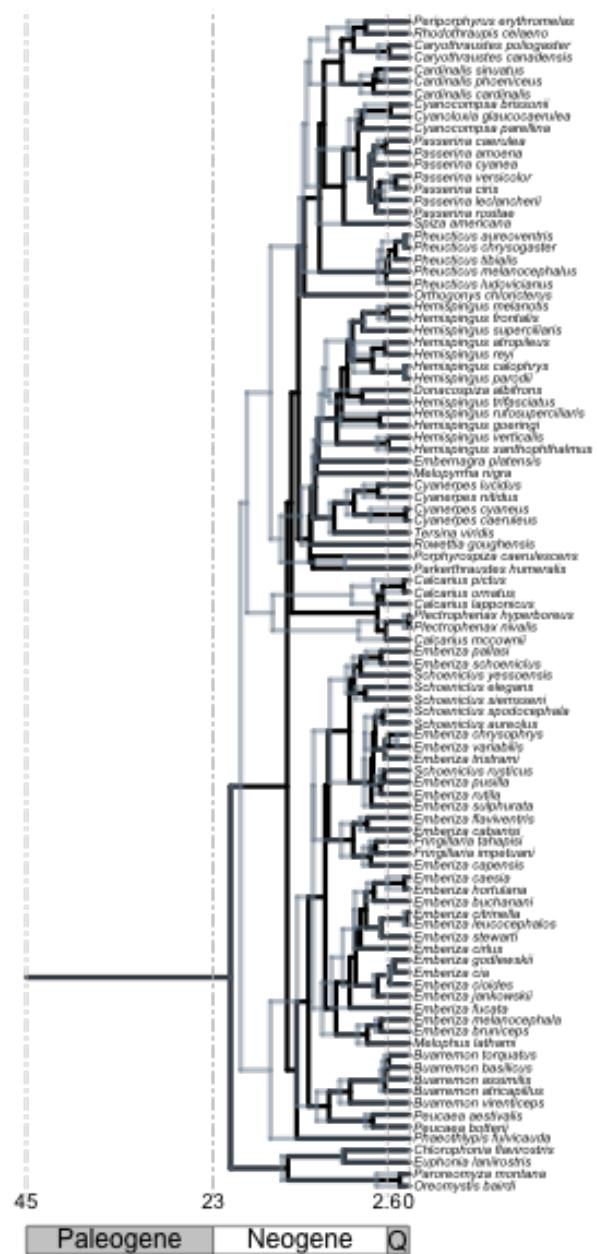


FIGURE 8. Cross validation of second source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to

Barker et al. 2015 - chronogram 2

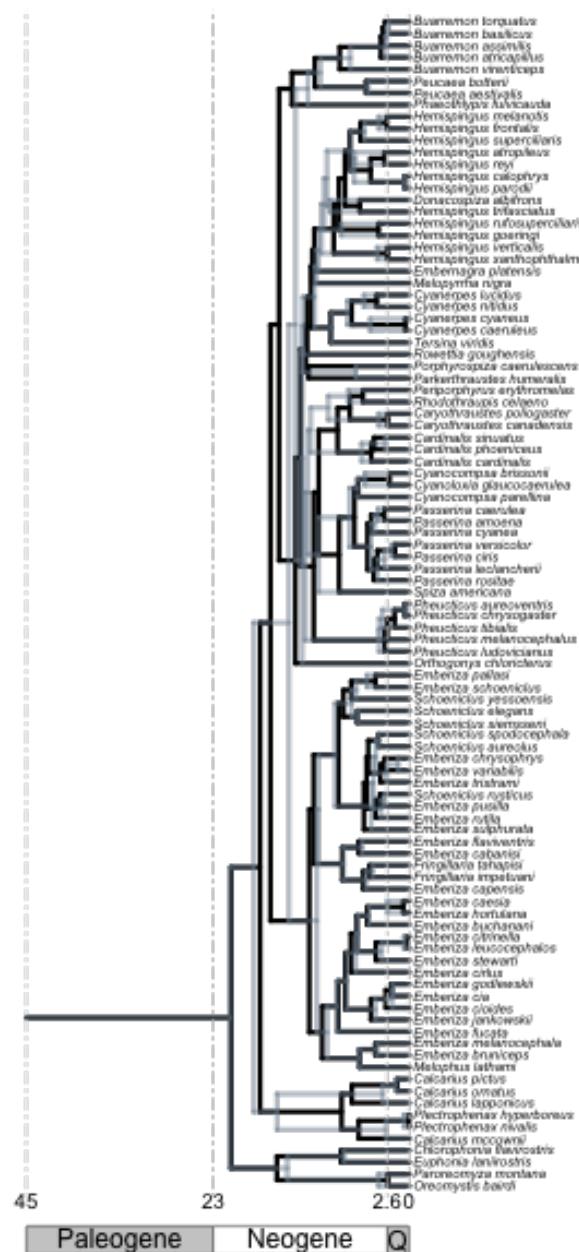


FIGURE 9. Cross validation of third source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the dates calculated with BLADJ, i.e., the same for all the nodes.

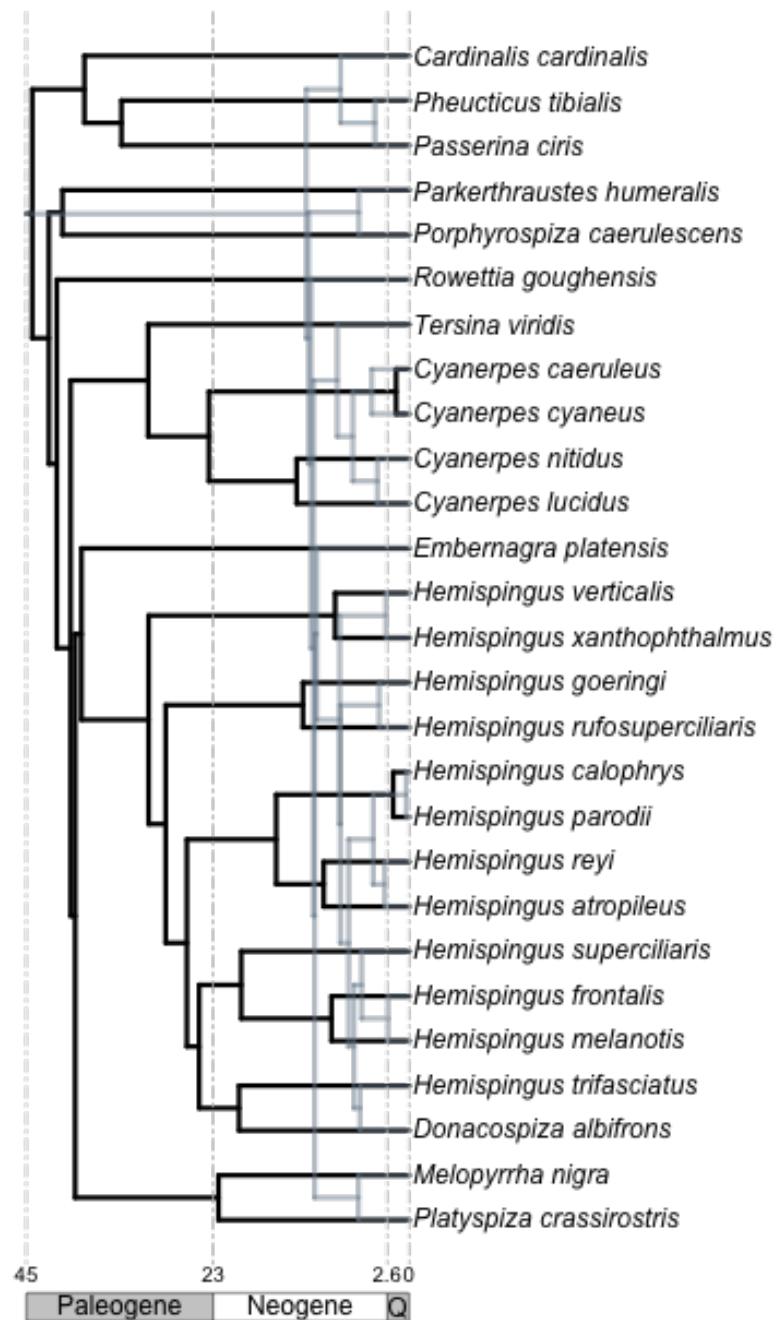
Burns et al. 2015

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, i.e., the same for all the nodes.

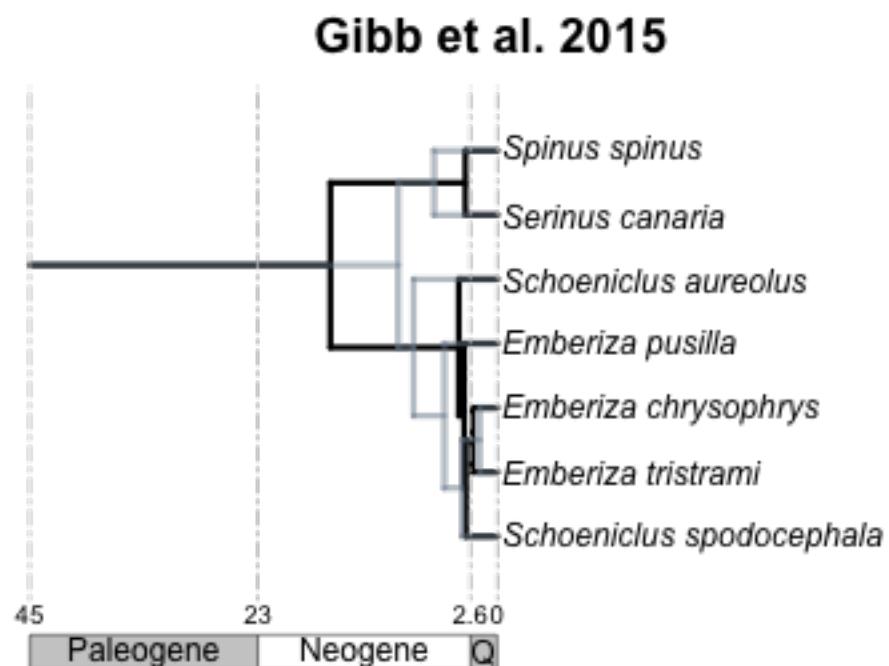


FIGURE 11. Cross validation of sixth source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the same tree topology dated with BLADJ using node ages from all other source chronograms as secondary calibrations.

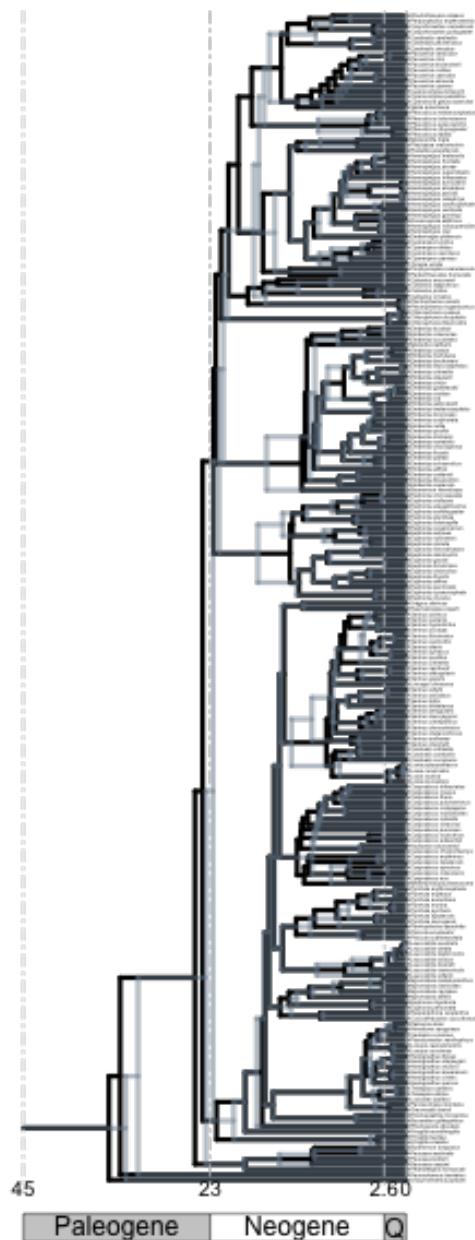
Hedges et al. 2015 - chronogram 1

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

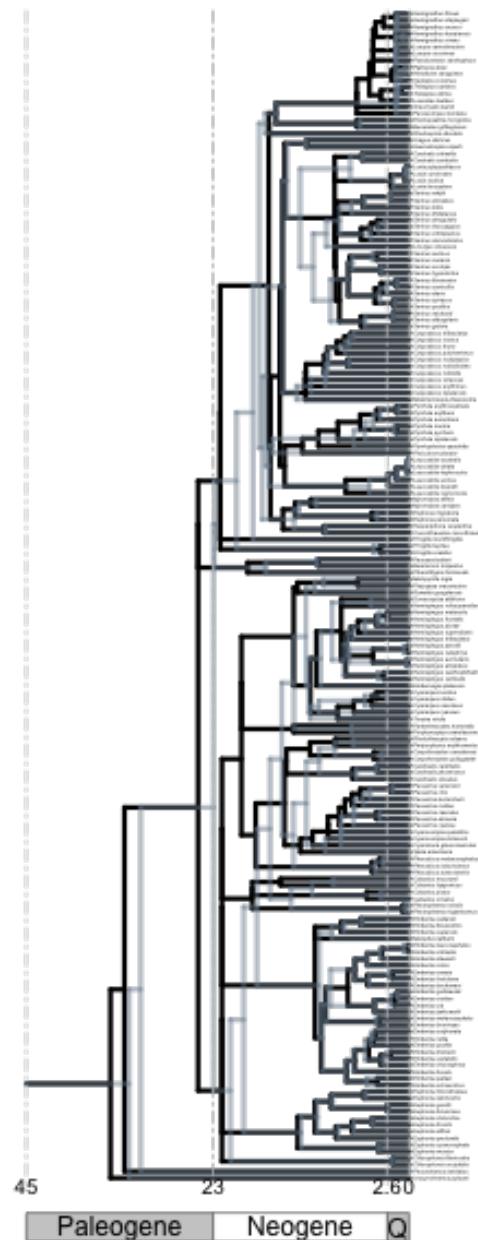
Hedges et al. 2015 - chronogram 2

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 BLADJ, i.e., the cross-validation procedure.

Hooper et al. 2017

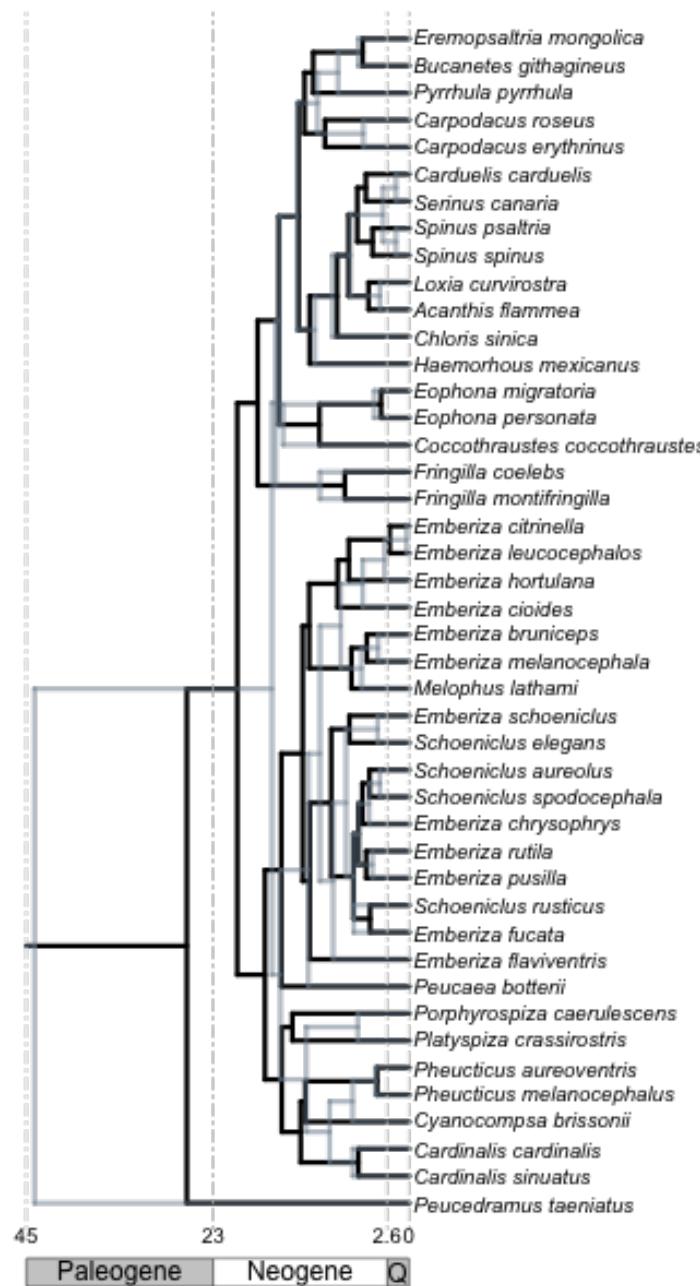


FIGURE 14. Cross validation of ninth source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the dates calculated with BLADJ in our analysis. The tree is rooted on the left and branches to the right.

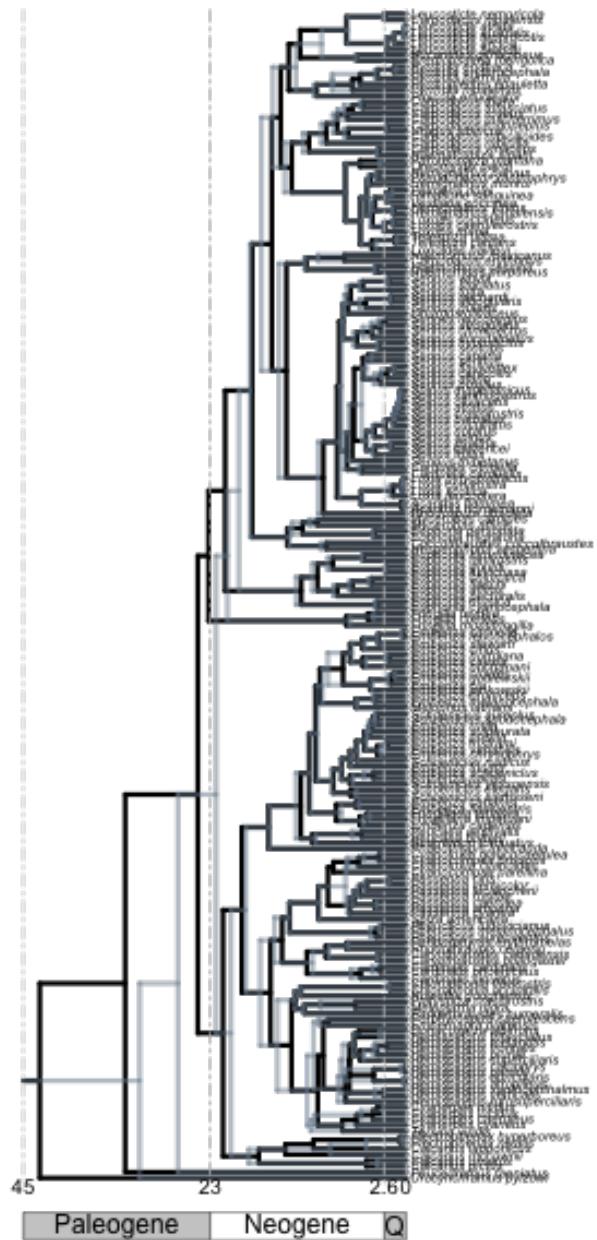
Jetz et al. 2012 - chronogram 1

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 order to facilitate the comparison, the two chronograms are plotted side-by-side.