

<sup>1</sup> DateLife: leveraging databases and analytical tools to reveal the dated Tree of Life

<sup>2</sup> Luna L. Sánchez Reyes<sup>1,2</sup>, Emily Jane McTavish<sup>1</sup>, & Brian O'Meara<sup>2</sup>

<sup>3</sup> <sup>1</sup> University of California, Merced

<sup>4</sup> <sup>2</sup> University of Tennessee, Knoxville

<sup>5</sup> Author Note

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

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

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

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

17

## Abstract

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

42

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

<sup>43</sup> Congruification; Supertree; Calibrations; Secondary calibrations

<sup>44</sup> Word count: 4765

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

46 **Introduction**

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

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

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

67 **Description**

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

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

## 77 Creating a search query

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

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

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

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

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

## 101 Searching a chronogram database

102 At the time of writing of this manuscript (May 27, 2022), DateLife's chronogram  
103 database latest version consist of 253 chronograms published in 187 different studies. It is  
104 curated from OpenTree's phylogenetic database, the Phylesystem, which constitutes an open  
105 source of expert and peer-reviewed phylogenetic knowledge with rich metadata (McTavish et  
106 al., 2015), which allows automatic and reproducible assembly of our chronogram database.  
107 Datelife's chronogram database is navigable as an R data object within the `datelife` R  
108 package.

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

118 A DateLife search is implemented by matching processed taxon names provided by the  
119 user to tip labels in the chronogram database. Chronograms with at least two matching

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

128 **Summarizing search results**

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

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

142 ***Choosing a topology.***— DateLife requires a tree topology to summarize age data  
143 upon. We recommend that users provide a tree topology as input from the literature, or one  
144 of their own making. If no topology is provided, DateLife automatically subsets one from the

<sup>145</sup> OpenTree synthetic tree (Open Tree Of Life et al., 2019). Alternatively, DateLife can  
<sup>146</sup> combine topologies from source chronograms using a supertree approach. To combine  
<sup>147</sup> topologies from source chronograms into a single summary (or supertree) topology, the  
<sup>148</sup> DateLife algorithm starts by identifying the source chronograms that form a grove, roughly,  
<sup>149</sup> a sufficiently overlapping set of taxa between trees, by implementing definition 2.8 for  
<sup>150</sup> n-overlap from Ané et al. (2009). In rare cases, a group of trees can have multiple groves. By  
<sup>151</sup> default, DateLife chooses the grove with the most taxa, however, the “criterion = trees” flag  
<sup>152</sup> allows the user to choose the grove with the most trees instead. If source chronograms do  
<sup>153</sup> not form a grove, the supertree reconstruction will fail.

<sup>154</sup> **Dating the topology.**— Input topologies from OpenTree or the supertree approach  
<sup>155</sup> described above do not include branch length estimates of any kind. Optionally, to estimate  
<sup>156</sup> branch lengths proportional to substitution rates on these topologies, DateLife can mine the  
<sup>157</sup> Barcode of Life Data System, BOLD (Ratnasingham & Hebert, 2007) to obtain genetic  
<sup>158</sup> markers for the input taxa. These markers are aligned with MUSCLE (Edgar, 2004) (by  
<sup>159</sup> default) or MAFFT (Katoh, Asimenos, & Toh, 2009). This alignment can be used to  
<sup>160</sup> estimate branch lengths on input topologies that lack branch lengths. Currently, branch  
<sup>161</sup> length reconstruction in DateLife is performed using parsimony and the likelihood of the  
<sup>162</sup> phylogenetic tree given a sequence alignment is computed (Schliep, 2011). While relative  
<sup>163</sup> branch length information provides additional data for nodes without secondary date  
<sup>164</sup> calibrations, topologies without branch lengths can also be dated.

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

171 congruified node ages that can be used as calibrations for a dating analysis is stored as a  
172 `congruifiedCalibrations` object.

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

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

194 Alternative options supported in DateLife (MrBayes, PATHD8, TreePL) incorporate  
195 branch length information from the input topology in combination with the calibrations.  
196 PATHd8 is a non-clock, rate-smoothing method (Britton et al., 2007) to date trees. treePL

197 (Stephen A. Smith & O'Meara, 2012), is a semi-parametric, rate-smoothing, penalized  
198 likelihood dating method (Michael J. Sanderson, 2002). The MrBayes (Huelsenbeck &  
199 Ronquist, 2001; Ronquist & Huelsenbeck, 2003) approach in datelife uses the calibrations as  
200 priors on node ages.

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

## 206 Benchmark

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

222

## Case studies

223 We illustrate the DateLife algorithm using a family within the Passeriform birds  
224 encompassing the true finches, Fringillidae, as case study. The first example analyses 6 bird  
225 species and shows all steps of the algorithm. The second example is an application analysing  
226 289 species in the family Fringillidae that are included in the NCBI taxonomy.

227 **Small example**

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

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

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

247 points. All nodes have age data.

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

256        **Fringillidae**

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

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

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

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

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

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

284 **Cross-validation test**

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

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

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

296

## Discussion

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

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

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

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

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

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

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

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

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

339 **Variation in source chronograms**

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

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

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

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

## 360 Sumarizing chronograms

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

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

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

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

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

### 391 Effects on downstream analyses

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

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

396 divergence times represents a relevant improvement for testing alternative hypothesis using  
397 phylogenetic distance (Campbell O. Webb et al., 2008). Hence, DateLife's workflow features  
398 different ways of estimating node ages in the absence of calibrations and branch length  
399 information for certain taxa. "Making up" branch lengths is a common practice in scientific  
400 publications: Jetz et al. (2012), created a time-calibrated tree of all 9,993 bird species, where  
401 67% had molecular data and the rest was simulated; Rabosky et al. (2018) created a  
402 time-calibrated tree of 31,536 ray-finned fishes, of which only 37% had molecular data;  
403 Stephen A. Smith and Brown (2018) constructed a tree of 353,185 seed plants where only  
404 23% had molecular data.

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

## 412 Conclusions

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

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

## 431 Availability

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

## 442 Supplementary Material

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

446

## Funding

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

451

## Acknowledgements

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

**References**

- 464
- 465 Alström, P., Hooper, D. M., Liu, Y., Olsson, U., Mohan, D., Gelang, M., ... Price, T.  
466 D. (2014). Discovery of a relict lineage and monotypic family of passerine birds.  
467 *Biology Letters*, 10(3), 20131067.
- 468 Ané, C., Eulenstein, O., Piaggio-Talice, R., & Sanderson, M. J. (2009). Groves of  
469 phylogenetic trees. *Annals of Combinatorics*, 13(2), 139–167.
- 470 Antonelli, A., Hettling, H., Condamine, F. L., Vos, K., Nilsson, R. H., Sanderson, M.  
471 J., ... Vos, R. A. (2017). Toward a self-updating platform for estimating rates of  
472 speciation and migration, ages, and relationships of Taxa. *Systematic Biology*,  
473 66(2), 153–166. <https://doi.org/10.1093/sysbio/syw066>
- 474 Archie, J., Day, W. H., Felsenstein, J., Maddison, W., Meacham, C., Rohlf, F. J., &  
475 Swofford, D. (1986). The Newick tree format. Retrieved from  
476 %7B<https://evolution.genetics.washington.edu/phylip/newicktree.html>%7D
- 477 Avibase. (2022). Yellow-throated Bunting. *Avibase - The World Bird Database*,  
478 (Online Resource). Retrieved from %7B<https://avibase.bsc-eoc.org/species.jsp?lang=EN&avibaseid=82D1EE0049D8D927%7D>
- 479 Bapst, D. W. (2012). Paleotree: An R package for paleontological and phylogenetic  
480 analyses of evolution. *Methods in Ecology and Evolution*, 3(5), 803–807.  
481 <https://doi.org/10.1111/j.2041-210X.2012.00223.x>
- 482 Barba-Montoya, J., Reis, M. dos, Schneider, H., Donoghue, P. C., & Yang, Z. (2018).  
483 Constraining uncertainty in the timescale of angiosperm evolution and the  
484 veracity of a cretaceous terrestrial revolution. *New Phytologist*, 218(2), 819–834.
- 485 Barker, F. K. (2014). Mitogenomic data resolve basal relationships among passeriform  
486 and passeridan birds. *Molecular Phylogenetics and Evolution*, 79, 313–324.
- 487 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2013). Going  
488 to extremes: Contrasting rates of diversification in a recent radiation of new world  
489 passerine birds. *Systematic Biology*, 62(2), 298–320.
- 490

- 491 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2015). New  
492 insights into new world biogeography: An integrated view from the phylogeny of  
493 blackbirds, cardinals, sparrows, tanagers, warblers, and allies. *The Auk: Ornithological Advances*, 132(2), 333–348.
- 494
- 495 Barker, F. K., Cibois, A., Schikler, P., Feinstein, J., & Cracraft, J. (2004). Phylogeny  
496 and diversification of the largest avian radiation. *Proceedings of the National  
497 Academy of Sciences*, 101(30), 11040–11045.
- 498 Beresford, P., Barker, F., Ryan, Pg., & Crowe, T. (2005). African endemics span the  
499 tree of songbirds (passeri): Molecular systematics of several evolutionary ‘enigmas.’  
500 *Proceedings of the Royal Society B: Biological Sciences*, 272(1565), 849–858.
- 501 Boyle, B., Hopkins, N., Lu, Z., Raygoza Garay, J. A., Mozzherin, D., Rees, T., ...  
502 Enquist, B. J. (2013). The taxonomic name resolution service: An online tool for  
503 automated standardization of plant names. *BMC Bioinformatics*, 14(1).  
504 <https://doi.org/10.1186/1471-2105-14-16>
- 505 Britton, T., Anderson, C. L., Jacquet, D., Lundqvist, S., & Bremer, K. (2007).  
506 Estimating Divergence Times in Large Phylogenetic Trees. *Systematic Biology*,  
507 56(788777878), 741–752. <https://doi.org/10.1080/10635150701613783>
- 508 Bryson Jr, R. W., Chaves, J., Smith, B. T., Miller, M. J., Winker, K., Pérez-Emán, J.  
509 L., & Klicka, J. (2014). Diversification across the new world within the  
510 ‘blue’cardinalids (aves: cardinalidae). *Journal of Biogeography*, 41(3), 587–599.
- 511 Burleigh, J. G., Kimball, R. T., & Braun, E. L. (2015). Building the avian tree of life  
512 using a large-scale, sparse supermatrix. *Molecular Phylogenetics and Evolution*,  
513 84, 53–63.
- 514 Burns, K. J., Shultz, A. J., Title, P. O., Mason, N. A., Barker, F. K., Klicka, J., ...  
515 Lovette, I. J. (2014). Phylogenetics and diversification of tanagers (passeriformes:  
516 Thraupidae), the largest radiation of neotropical songbirds. *Molecular Phylogenetics and Evolution*, 75, 41–77.
- 517

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

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

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

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

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

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

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

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

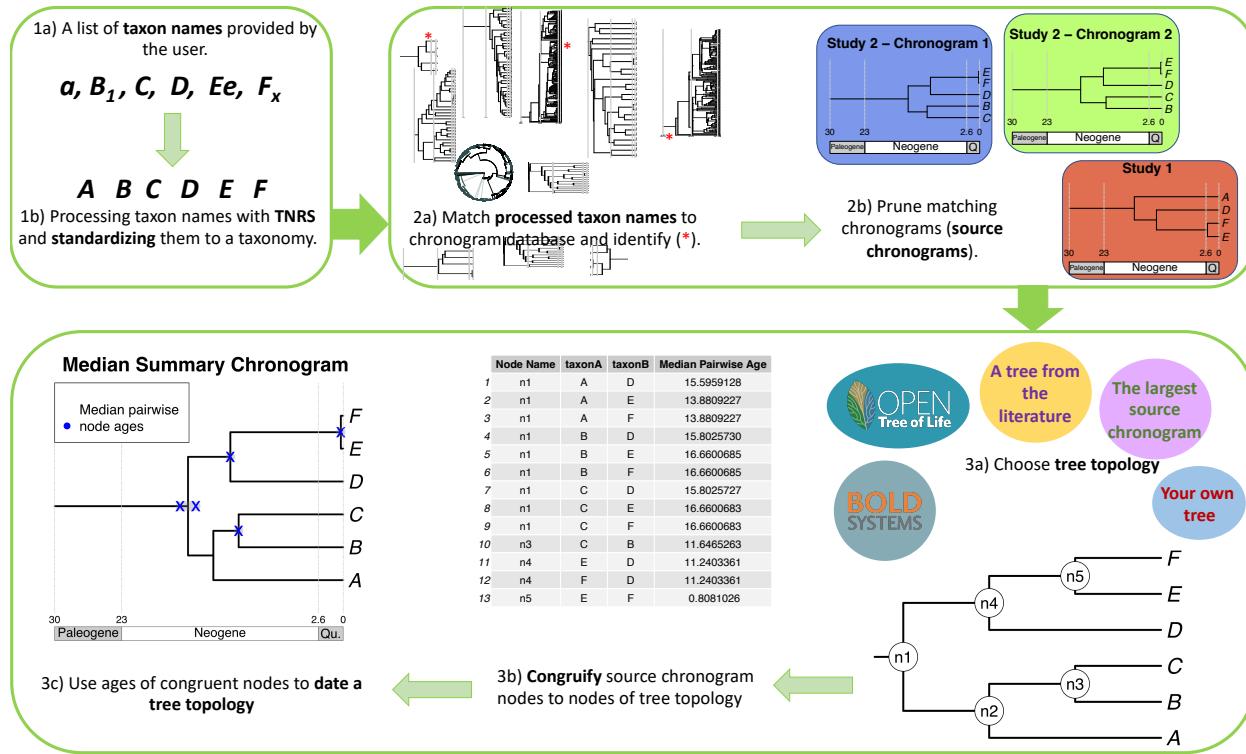


FIGURE 1. Stylized DateLife workflow. This shows the general workflows and analyses that can be performed with `datelife`, via the R package or through the website at <http://www.datelife.org/>. Details on the functions involved on each workflow are shown in `datelife`'s R package vignette.

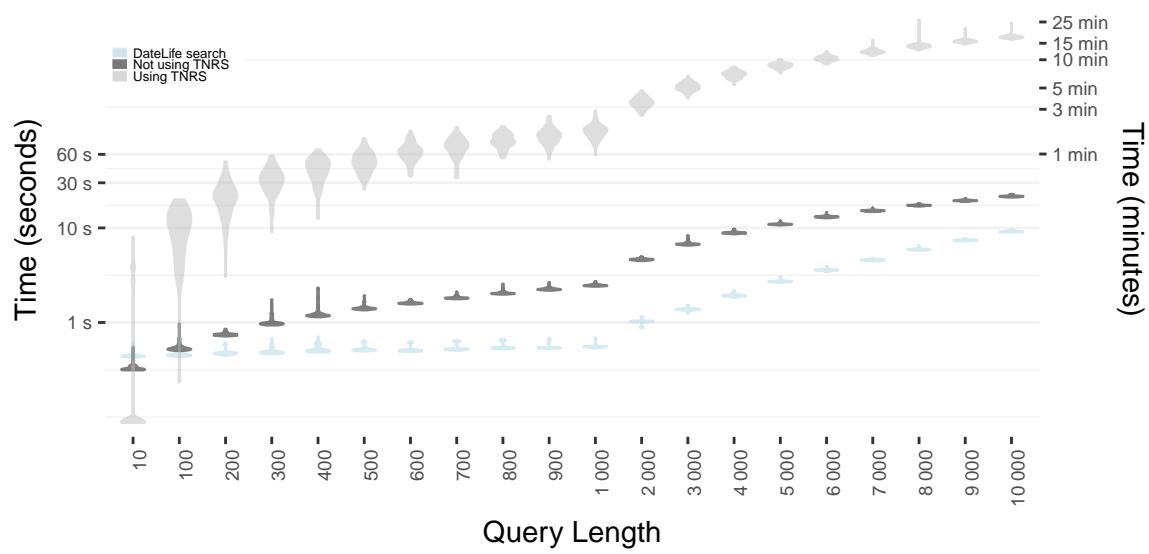


FIGURE 2. Computation time of query processing and search across **datelife**'s chronogram database relative to number of input taxon names. We sampled N names from the class Aves for each cohort 100 times and then performed a search with query processing not using the Taxon Names Resolution Service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

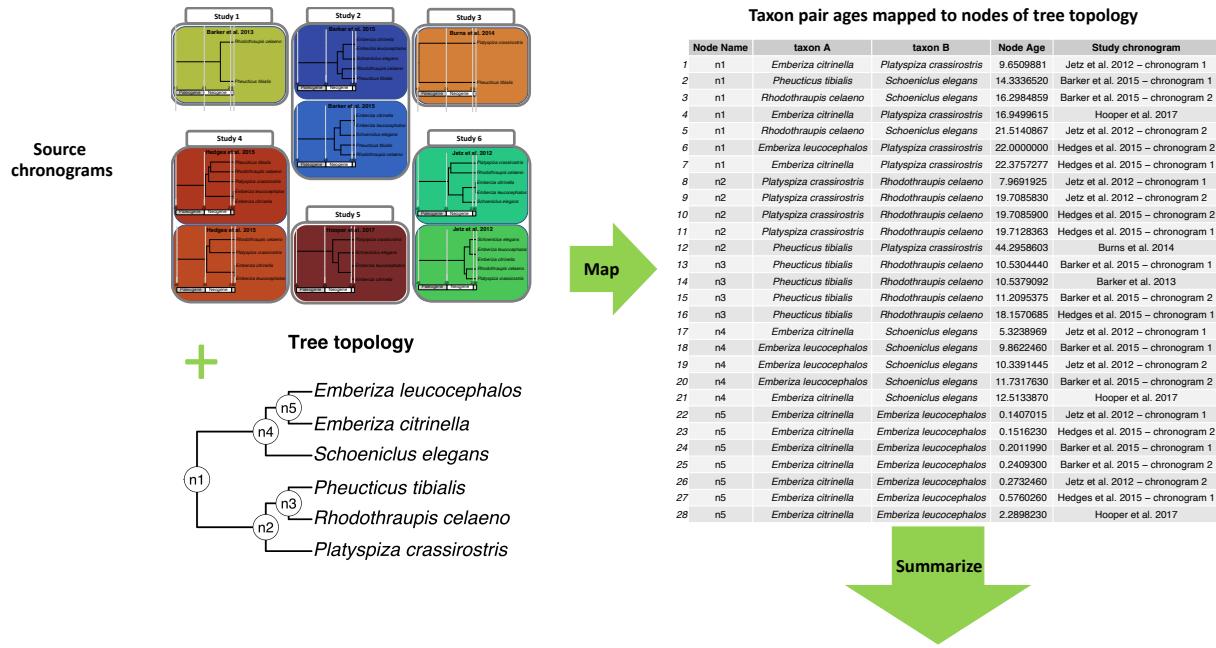


FIGURE 3. Age data results of a DateLife search of a small sample of 6 bird species within the Passeriformes. Input names were found across 9 chronograms within 6 independent studies (Barker et al. (2012), Barker et al. (2015), Burns et al. (2014), Hedges et al. (2015), Hooper and Price (2017), Jetz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

### Summary of mapped taxon pair age data

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

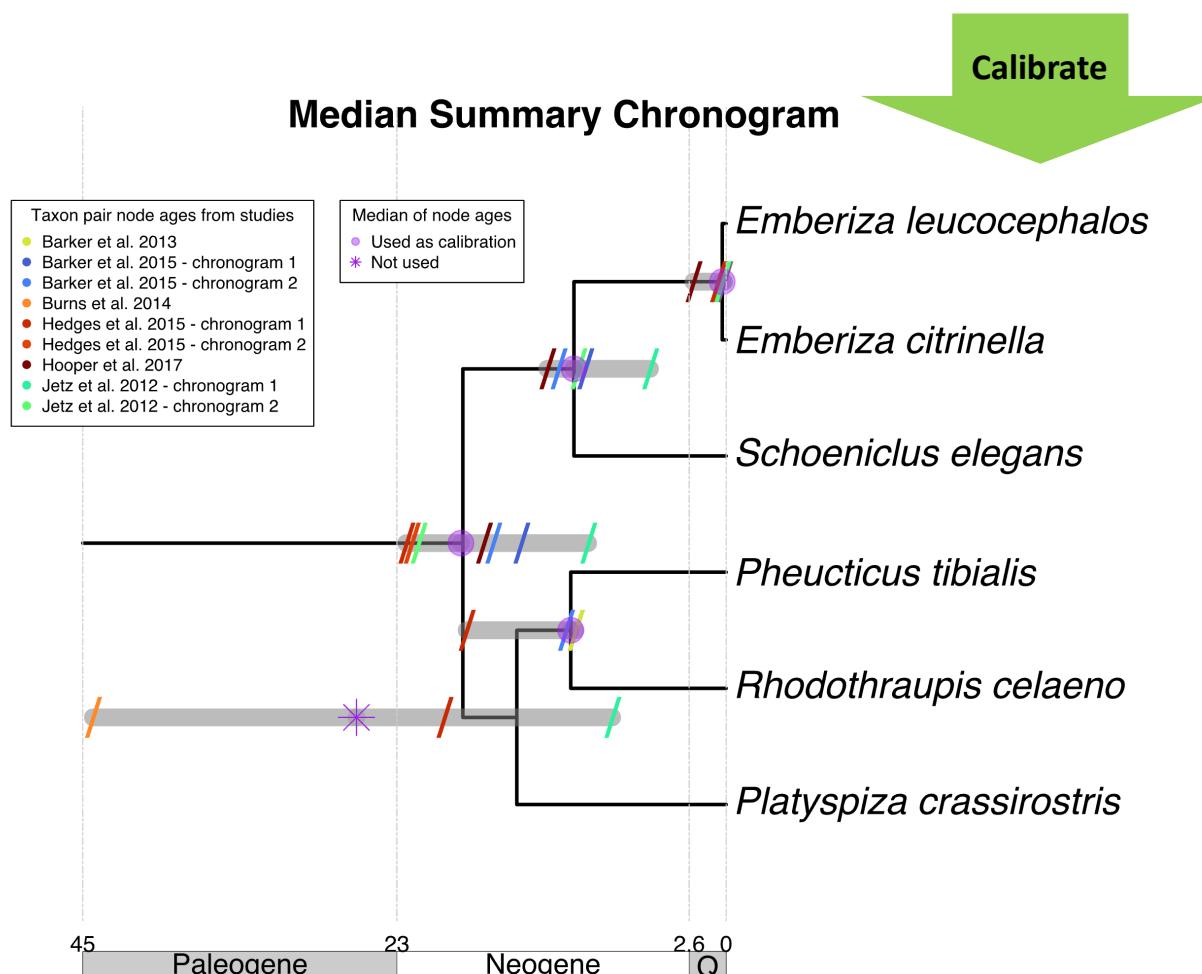


FIGURE 4. Summarized age data is used as secondary calibrations to date a tree topology as a summary chronogram.

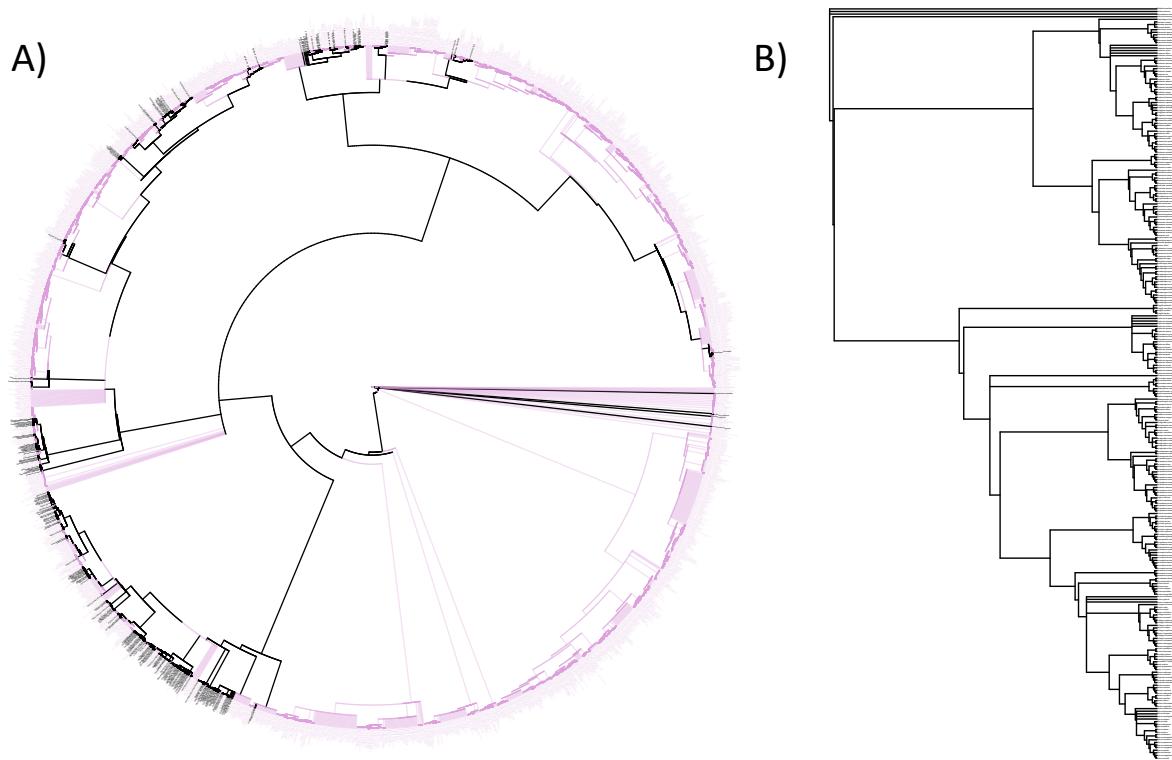


FIGURE 5. Tree topologies obtained from Open Tree of Life's (OpenTree) synthetic phylogenetic tree. A) Topology of 2333 tips and 1305 internal nodes, encompassing bird species within the family Fringillidae following the NCBI taxonomy (black), as well as all other bird species that share the same mrca node in OpenTree's synthetic tree (purple). B) Topology of 289 tips and 253 internal nodes, encompassing bird species within the Fringillidae only, resulting from pruning purple branches from topology A. Bird species within the Fringillidae do not form a monophyletic group. (Alström et al. 2014, Barker, Cibois, Schikler, Feinstein, & Cracraft 2004, Barker et al. 2013, Barker 2014, Barker et al. 2015, Beresford, Barker, Ryan, & Crowe 2005, Bryson Jr et al. 2014, Burleigh, Kimball, & Braun 2015, Burns et al. 2014, Chaves, Hidalgo, & Klicka 2013, Claramunt & Cracraft 2015, Gibb et al. 2015, Hackett et al. 2008, Jetz et al. 2012, Johansson, Fjeldså, & Bowi 200, Kimball et al. 2019, Klicka et al. 2014, Lamichhaney et al. 2015, Lerner, Meyer, James, Hofreiter, & Fleischer 2011, Lovette et al. 2010, Moyle et al. 2016, Ödeen, Håstad, & Alström 2011, Oliveros et al. 2019, Päckert et al. 2012, Parchman, Benkman, & Mezquida 2007, Powell et al. 2014, Price et al. 2014, Pulgarín-R, Smith, Bryson Jr, Spellman, & Klicka 2013, Selvatti, Gonzaga, & Moraes Russo 2015, Tietze, Päckert, Martens, Lehmann, & Sun 2013, Treplin et al. 2008, Zuccon, Prŷs-Jones, Rasmussen, & Ericson 2012)

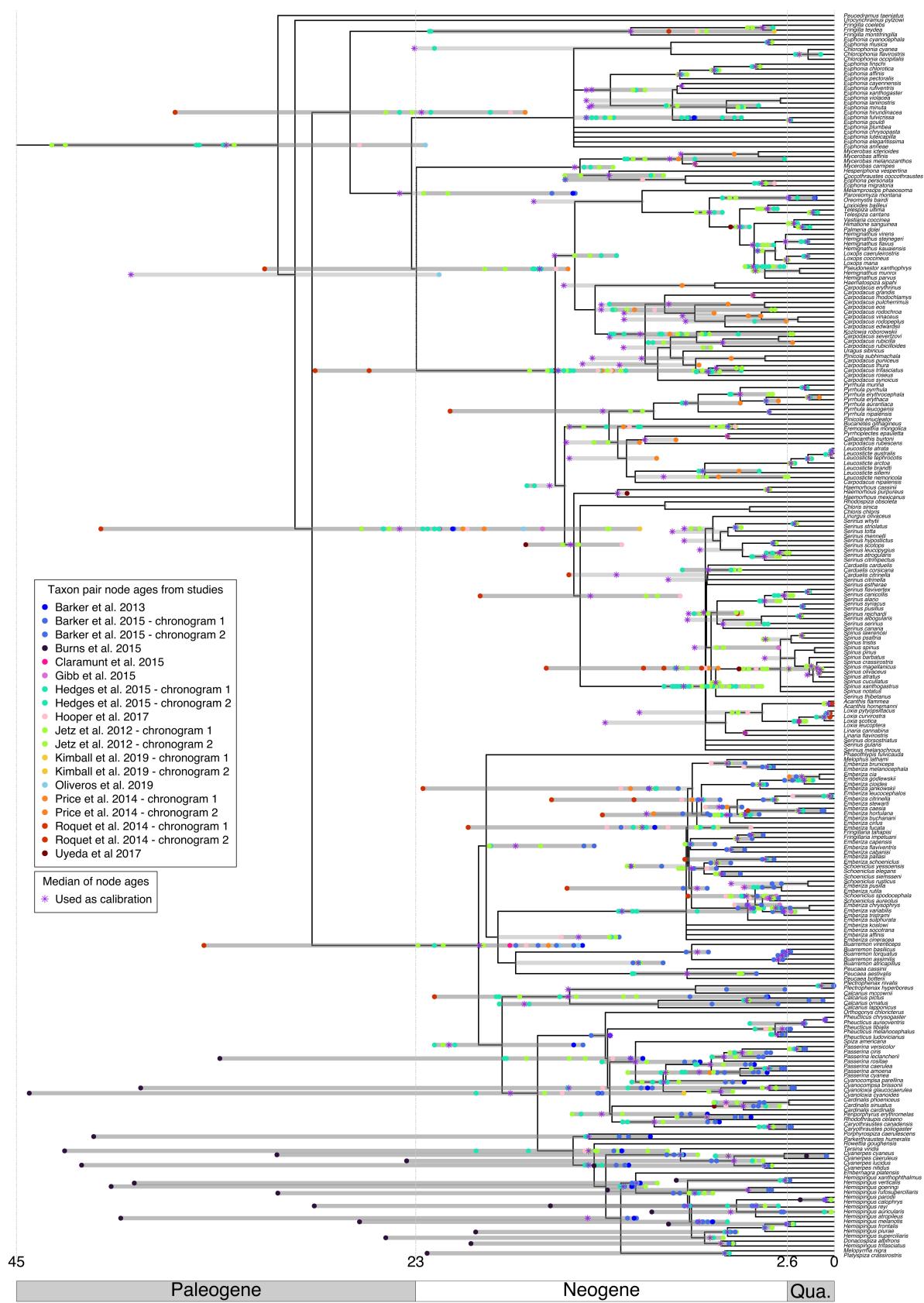


FIGURE 6. Fringillidae median summary chronogram generated with DateLife. It has 256 tips and 233 nodes.

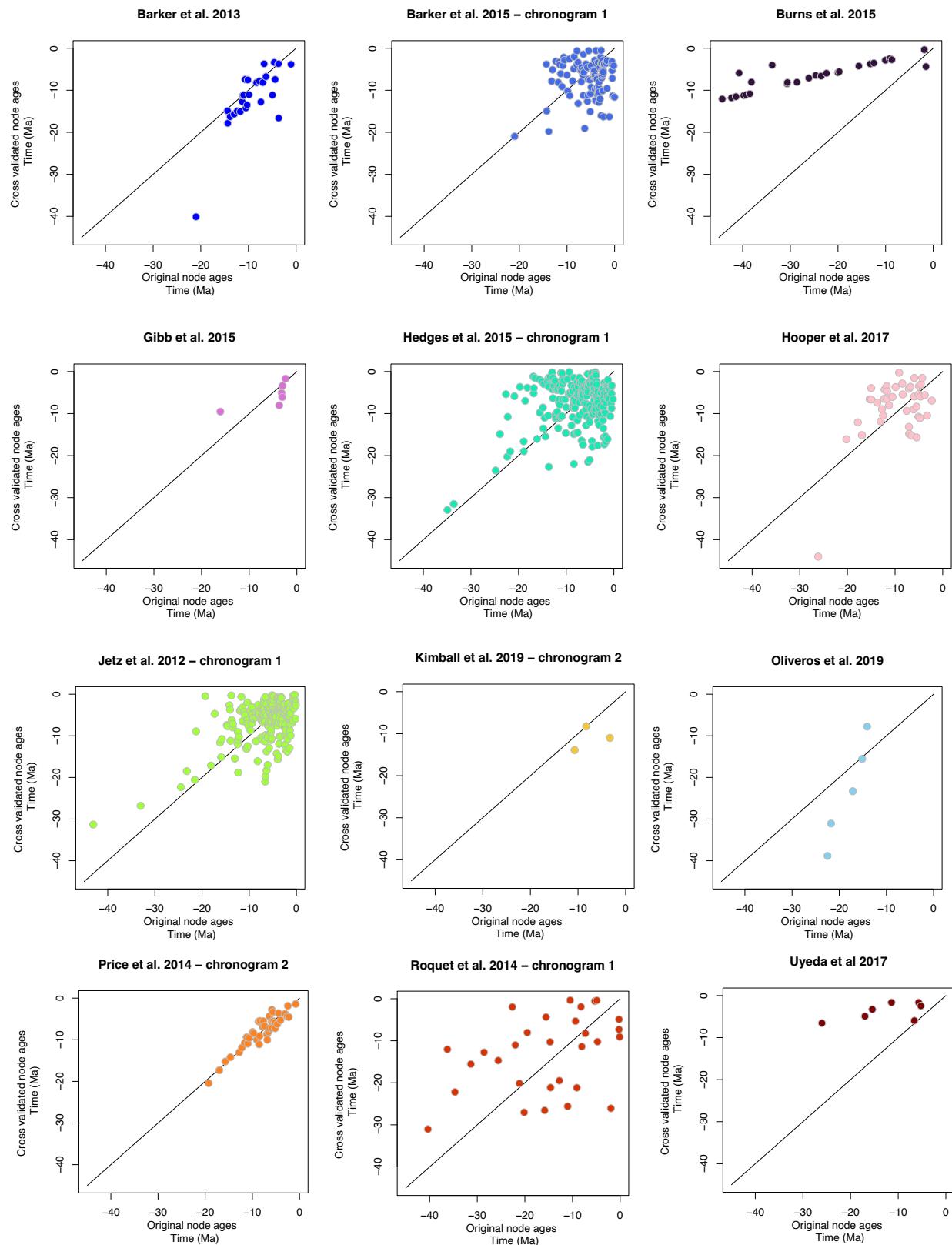


FIGURE 7. Results from cross validation analysis.

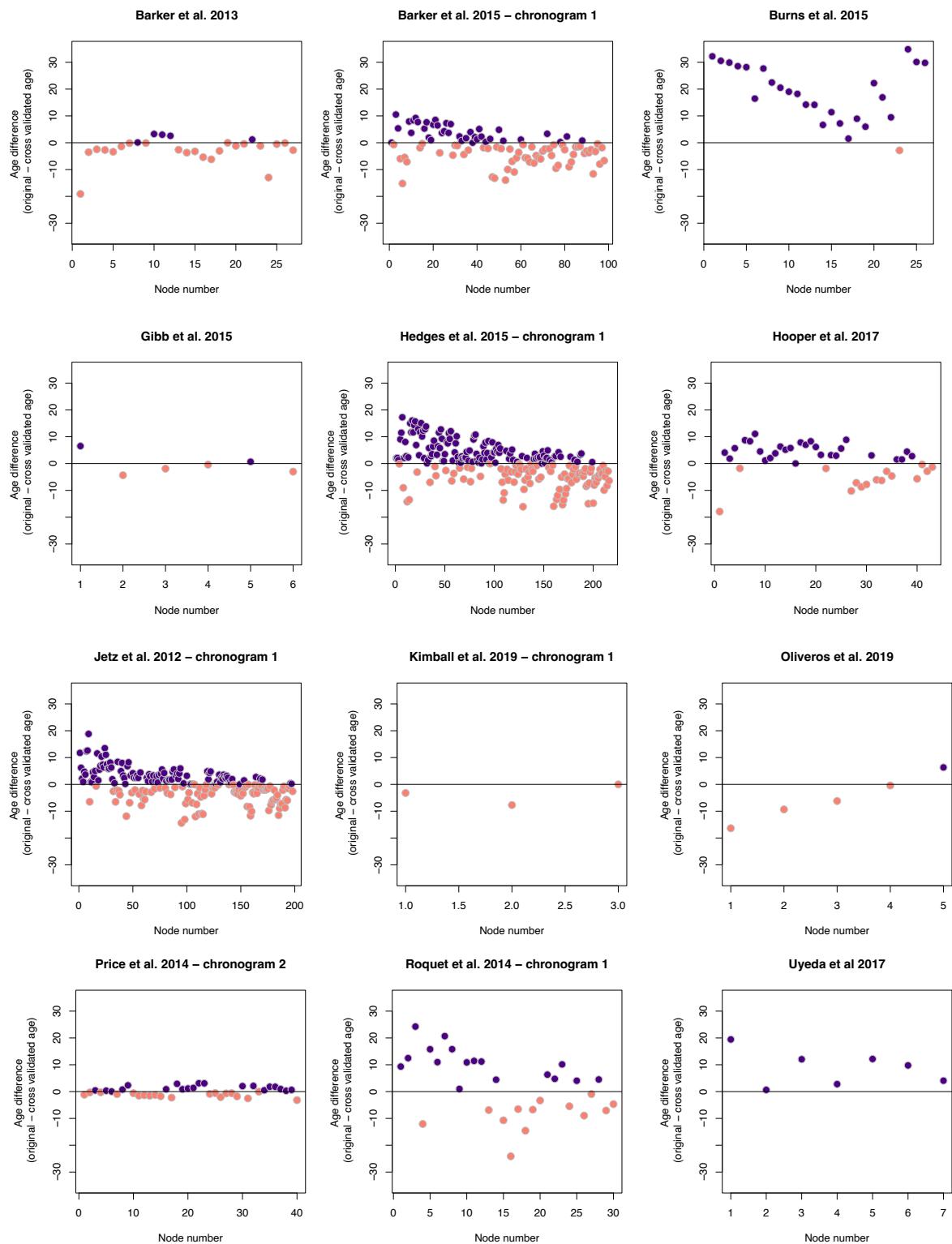


FIGURE 8. Results from cross validation analysis.

## Barker et al. 2015 - chronogram 1



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

## Barker et al. 2015 - chronogram 2



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

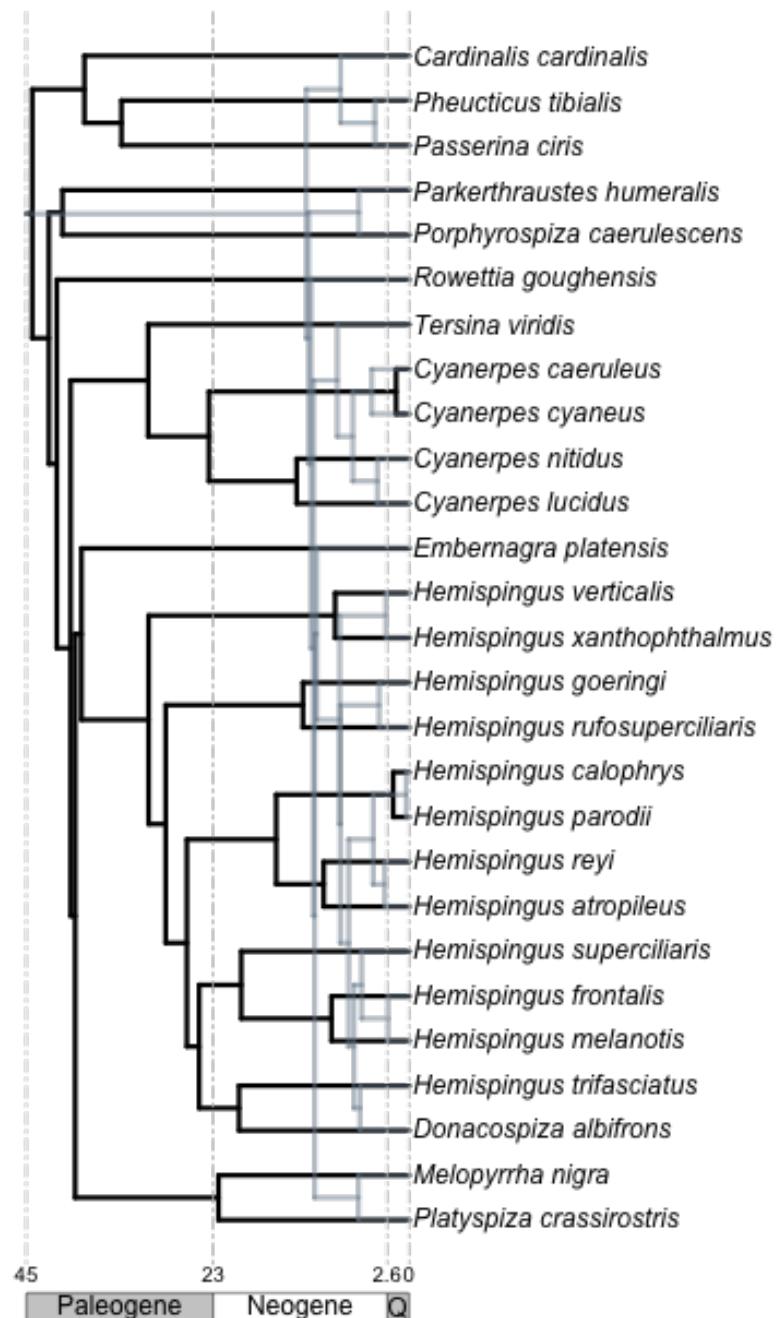
**Burns et al. 2015**

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

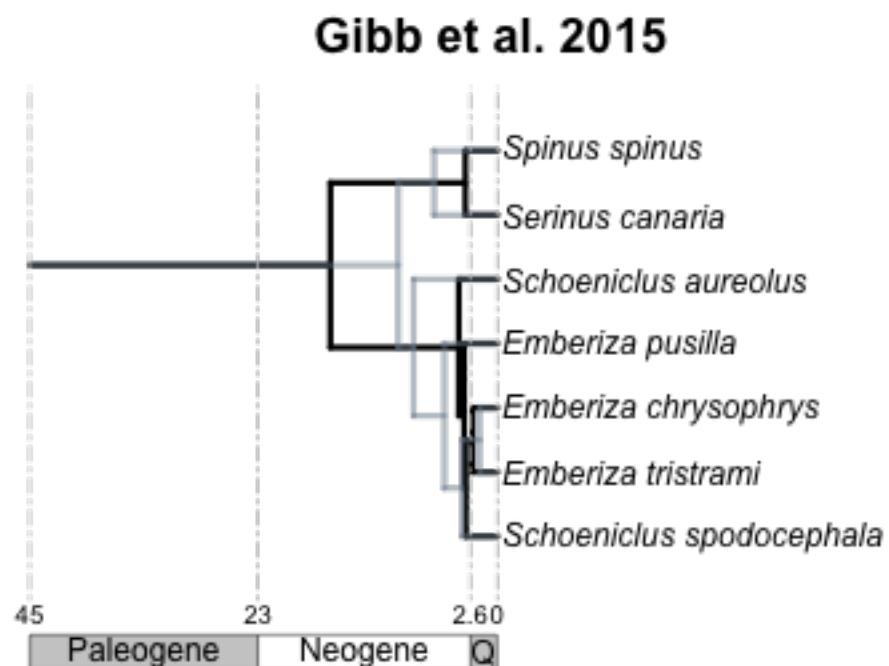


FIGURE 12. Cross validation of sixth source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the same tree topology dated with BLADJ using node ages from all other source chronograms as secondary calibrations.

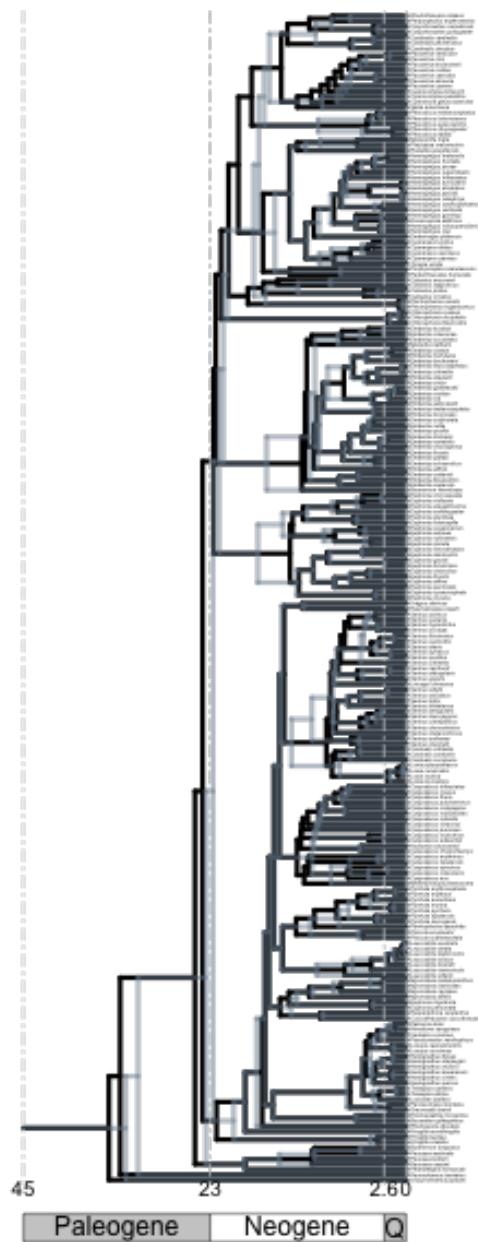
**Hedges et al. 2015 - chronogram 1**

FIGURE 13. Cross validation of seventh source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the dates calculated with BLADe. In order to facilitate the comparison, the two chronograms are plotted side-by-side.

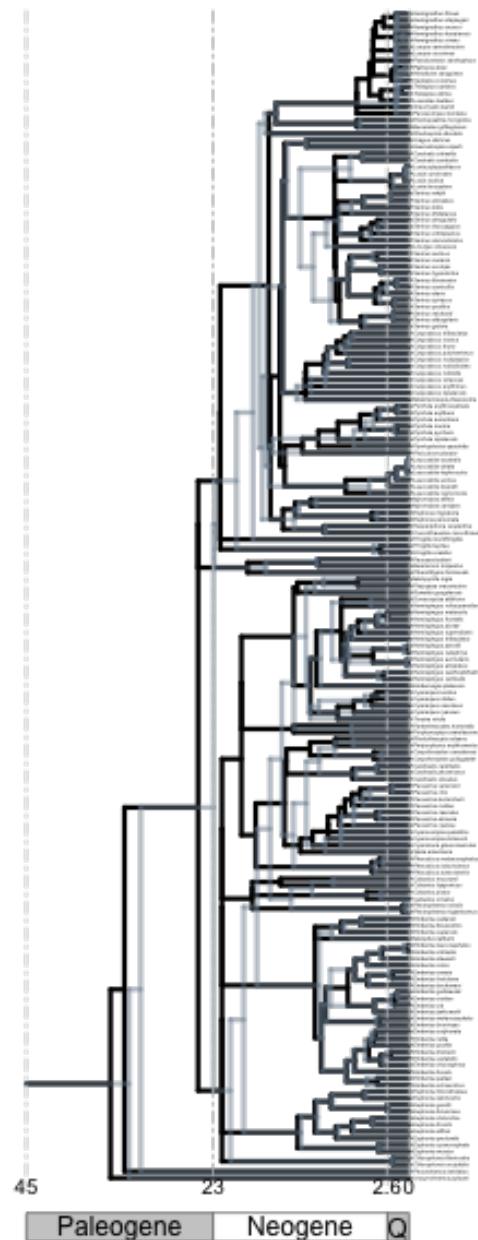
**Hedges et al. 2015 - chronogram 2**

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

**Hooper et al. 2017**

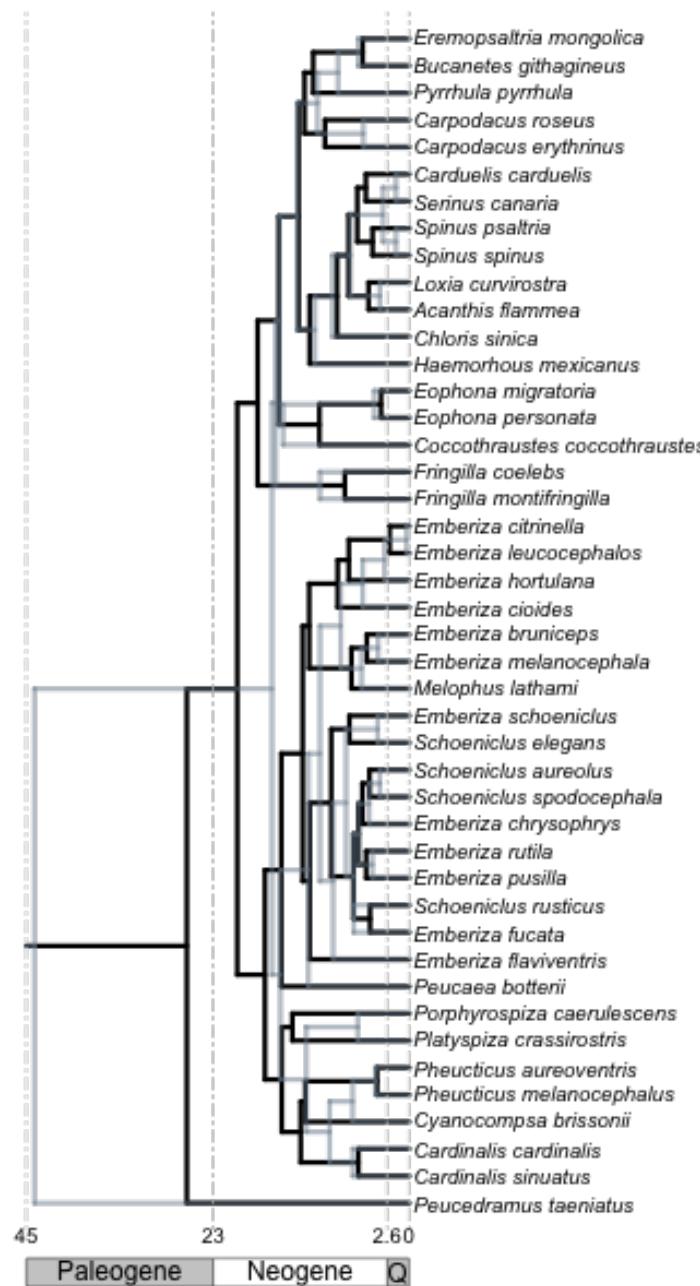


FIGURE 15. Cross validation of ninth source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the dates calculated with BLADJ.

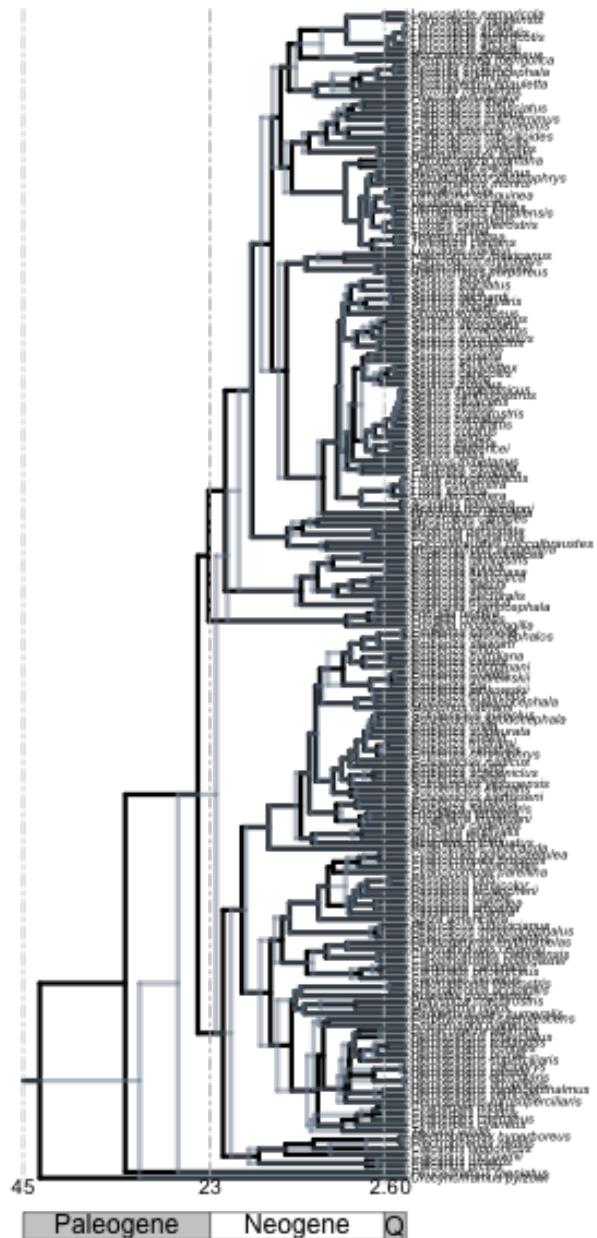
**Jetz et al. 2012 - chronogram 1**

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