

<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 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 R Shiny  
29 website application available at [www.datelife.org](http://www.datelife.org), that provides functionalities for efficient  
30 and easy finding, summary, reuse, and reanalysis of expert, peer-reviewed, public data on  
31 time of evolutionary origin. The main DateLife workflow constructs a chronogram for any  
32 given combination of taxon names by searching a local chronogram database constructed and  
33 curated from the Open Tree of Life Phylesystem phylogenetic database, which incorporates  
34 phylogenetic data from the TreeBASE database as well. We implement and test methods for  
35 summarizing time data from multiple source chronograms using supertree and  
36 congruification algorithms, and using age data extracted from source chronograms as  
37 secondary calibration points to add branch lengths proportional to absolute time to a tree  
38 topology. DateLife will be useful to increase awareness of the existing variation in expert  
39 time of divergence data, and can foster exploration of the effect of alternative divergence  
40 time hypotheses on the results of analyses, providing a framework for a more informed  
41 interpretation of evolutionary results.

42

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

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

<sup>44</sup> Word count: 4607

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

46 **Introduction**

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

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

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

66 **Description**

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

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

76 **Creating a search query**

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

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

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

95 OpenTree's Taxonomy identification numbers for further processing.

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

100 **Searching a chronogram database**

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

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

120 **Summarizing search results**

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

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

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

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

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

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

<sup>172</sup> & Ronquist, 2001; Ronquist & Huelsenbeck, 2003), PATHd8 (Britton, Anderson, Jacquet,  
<sup>173</sup> Lundqvist, & Bremer, 2007), BLADJ (Campbell O. Webb, Ackerly, & Kembel, 2008;  
<sup>174</sup> Campbell O. Webb & Donoghue, 2005), and treePL (Stephen A. Smith & O'Meara, 2012).

<sup>175</sup> By default, DateLife implements the Branch Length Adjuster (BLADJ) algorithm to  
<sup>176</sup> obtain a fully dated topology. BLADJ fixes node ages that have calibration data, and  
<sup>177</sup> distributes time between nodes with no data evenly between nodes with calibration data.  
<sup>178</sup> This minimizes age variance in the resulting chronogram (Campbell O. Webb et al., 2008).  
<sup>179</sup> BLADJ does not incorporate input phylogeny branch lengths even when they are present.  
<sup>180</sup> When there is conflict in ages between nodes with calibration data, the algorithm ignores  
<sup>181</sup> node ages that are older than the age of a parent node. BLADJ requires a root age estimate.  
<sup>182</sup> If there is no information on the age of the root in the chronogram database, users can  
<sup>183</sup> provide an estimate from the literature. If none is provided, DateLife assigns an arbitrary  
<sup>184</sup> age to the root as 10% older than the oldest age available within the group.

<sup>185</sup> Alternative options supported in DateLife (MrBayes, PATHD8, TreePL) incorporate  
<sup>186</sup> branch length information from the input topology in combination with the secondary  
<sup>187</sup> calibrations. PATHd8 is a non-clock, rate-smoothing method (Britton et al., 2007) to date  
<sup>188</sup> trees. treePL (Stephen A. Smith & O'Meara, 2012), is a semi-parametric, rate-smoothing,  
<sup>189</sup> penalized likelihood dating method (Michael J. Sanderson, 2002). The MrBayes  
<sup>190</sup> (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003) approach in datelife uses the  
<sup>191</sup> secondary calibrations as priors on node ages.

<sup>192</sup> **Visualizing results.**— Finally, users can save all source and summary chronograms in  
<sup>193</sup> formats that permit reuse and reanalyses (newick and R “phylo” format), as well as visualize  
<sup>194</sup> and compare results graphically, or construct their own graphs using `datelife`'s chronogram  
<sup>195</sup> plot generation functions available from the R package `datelifeplot` (Sanchez-Reyes &  
<sup>196</sup> O'Meara, 2022).

197

## Benchmark

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

213

## Case studies

214        We illustrate the DateLife algorithm using a family within the Passeriform birds  
215 encompassing the true finches, Fringillidae, as case study. The first example analyses 6 bird  
216 species and shows all steps of the algorithm. The second example is an application analysing  
217 289 species in the family Fringillidae that are included in the NCBI taxonomy.

218        **Small example**

219        ***Creating a search query.-*** We chose 6 bird species in the family Fringillidae, known  
220 as the true finches. The sample includes two species of cardinals: the black-thighed grosbeak  
221 – *Pheucticus tibialis* and the crimson-collared grosbeak – *Rhodothraupis celaeno*; three species

222 of buntings: the yellowhammer – *Emberiza citrinella*, the pine bunting – *Emberiza*  
223 *leucocephalos* and the yellow-throated bunting – *Emberiza elegans*; and one species of  
224 tanager, the vegetarian finch – *Platyspiza crassirostris*.

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

231 **Searching the database.-** DateLife used the processed input names to search the  
232 local chronogram database and found 9 matching chronograms in 6 different studies. Three  
233 studies matched five input names (Barker, Burns, Klicka, Lanyon, & Lovette, 2015; Hedges,  
234 Marin, Suleski, Paymer, & Kumar, 2015; Jetz, Thomas, Joy, Hartmann, & Mooers, 2012),  
235 one study matched four input names (Hooper & Price, 2017) and two studies matched two  
236 input names (Barker, Burns, Klicka, Lanyon, & Lovette, 2013; Burns et al., 2014). No  
237 studies matched all input names. Together, matching chronograms have 28 unique age data  
238 points. All nodes have age data.

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

247 **Fringillidae**

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

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

260        ***Summarizing search results.***- The final step is to summarize the age data available  
261        for the Fringillidae species into single summary chronograms, using different types of  
262        summary ages, median and SDM. As explained in the “Description” section, a tree topology  
263        to summarize age data upon is required. By default, `datelife` uses the Open Tree of Life  
264        synthetic phylogeny as summarizing topology. According to this phylogeny, Fringillidae is  
265        not a monophyletic family (Alström et al., 2014; Barker, 2014; Barker et al., 2013, 2015;  
266        Barker, Cibois, Schikler, Feinstein, & Cracraft, 2004; Beresford, Barker, Ryan, & Crowe,  
267        2005; Bryson Jr et al., 2014; Burleigh, Kimball, & Braun, 2015; Burns et al., 2014; Chaves,  
268        Hidalgo, & Klicka, 2013; Claramunt & Cracraft, 2015; Gibb et al., 2015; Hackett et al., 2008;  
269        Jetz et al., 2012; Johansson, Fjeldså, & Bowie, 2008; Kimball et al., 2019; Klicka et al., 2014;  
270        Lamichhaney et al., 2015; Lerner, Meyer, James, Hofreiter, & Fleischer, 2011; Lovette et al.,  
271        2010; Moyle et al., 2016; Ödeen, Håstad, & Alström, 2011; Oliveros et al., 2019; Päckert et  
272        al., 2012; Parchman, Benkman, & Mezquida, 2007; Powell et al., 2014; Price et al., 2014;

<sup>273</sup> Pulgarín-R, Smith, Bryson Jr, Spellman, & Klicka, 2013; Selvatti, Gonzaga, & Moraes Russo,  
<sup>274</sup> 2015; Tietze, Päckert, Martens, Lehmann, & Sun, 2013; Treplin et al., 2008; Zuccon,  
<sup>275</sup> Prŷs-Jones, Rasmussen, & Ericson, 2012).

<sup>276</sup> Hence, DateLife's taxon-constrained approach returns a topology and ages for the  
<sup>277</sup> species in a named group.

<sup>278</sup> Age data from source chronograms is summarised into a single patristic distance  
<sup>279</sup> matrix and then the available node ages are used as fixed node calibrations over a consensus  
<sup>280</sup> tree topology, to obtain a fully dated phylogeny with the program BLADJ (Fig. 5).

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

### <sup>283</sup> Cross-validation test

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

<sup>289</sup> We found that node ages from original study, and ages estimated using all other age  
<sup>290</sup> data available are correlated (Supplementary Fig. 6).

<sup>291</sup> For five studies, Datelife tended to underestimate ages for topologically deeper nodes  
<sup>292</sup> (those with many descendant taxa, aka 'closer to the root') relative to the orginal estimate,  
<sup>293</sup> and overestimate ages for nodes closer to the tips. Accordingly, root ages are generally older  
<sup>294</sup> in the original study than estimated using cross-validated ages (Supplementary Fig. 7).

## Discussion

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

301        At the time of writing of this manuscript (May 22, 2022), `datelife`'s database has 253  
302        chronograms, pulled entirely from OpenTree's database, the Phylesystem (McTavish et al.,  
303        2015). A unique feature of OpenTree's Phylesystem is that the community can add new  
304        state-of-the-art chronograms any time. As chronograms are added to Phylesystem, they are  
305        incorporated into an updated `datelife`'s database that is assigned a new version number,  
306        followed by a package release on CRAN. `datelife`'s chronogram database is updated as new  
307        chronogram data is added to Phylesystem, at a minimum of once a month and a maximum  
308        of every 6 months. Users can also upload new chronograms to OpenTree themselves, and  
309        trigger an update of their local `datelife` database to incorporate the new chronograms, to  
310        have them immediately available for analysis.

311        A total of 99474 unique terminal taxa are represented in `datelife`'s database.  
312        Incorporation of more chronograms into `datelife`'s database will continue to improve its  
313        services. One option to increase the of chronograms in the DateLife database is the Dryad  
314        data repository, which now holds some chronograms. Methods to automatically mine  
315        chronograms from Dryad could be designed and implemented. However, Dryad's metadata  
316        system has no information to automatically detect branch length units, and those would still  
317        need to be determined manually by a curator.

318        The largest, and taxonomically broadest, summary chronogram currently available  
319        from OpenTree was constructed using age data from 2,274 published chronograms (Hedges et

320 al., 2015).

321 We would like to emphasize on the importance of sharing chronogram data, including  
322 systematically curated metadata, into open repositories, such as OpenTree's Phylesystem  
323 (McTavish et al., 2015) for the benefit of the scientific community as a whole.

324 By default, **datelife** currently summarizes all source chronograms that overlap with  
325 at least two species names. Users can exclude source chronograms if they have reasons to do  
326 so. Strictly speaking, the best chronogram should reflect the real time of lineage divergence  
327 accurately and precisely. To our knowledge, there are no good measures to determine  
328 independently if a chronogram is better than another. Some measures that have been  
329 proposed are the proportion of lineage sampling and the number of calibrations used  
330 Magallón, Gómez-Acevedo, Sánchez-Reyes, & Hernández-Hernández (2015). Several  
331 characteristics of the data used for dating analyses, as well as from the output chronogram  
332 itself, could be used to score quality of source chronograms. Some characteristics that are  
333 often cited in published studies as a measure of improved age estimates as compared to  
334 previously published estimates are: quality of alignment (missing data, GC content), lineage  
335 sampling (strategy and proportion), phylogenetic and dating inference method, number of  
336 fossils used as calibrations, support for nodes and ages, and magnitude of confidence  
337 intervals. DateLife provides an opportunity to capture concordance and conflict among date  
338 estimates, which can also be used as a metric for chronogram reliability. Its open database of  
339 trees allows other researchers to do such analyses themselves reproducibly and without  
340 needing permission, though of course they should follow proper citation practices, especially  
341 for the source trees.

342 Scientists have more confidence in chronograms constructed using primary calibrations  
343 (ages obtained from the fossil or geological record) to ones constructed with secondary  
344 calibrations (ages coming from other chronograms)(Schenk, 2016). Schenk (2016) showed  
345 that in simulations divergence times inferred with secondary calibrations are significantly

346 younger than those inferred with primary calibrations in analyses performed with Bayesian  
347 inference methods when priors are implemented in similar ways in both analyses. However,  
348 secondary calibrations are useful in dating methods that do not require setting priors, such  
349 as penalized likelihood (Michael J. Sanderson, 2003), or as fixed ages for nodes. Certainly,  
350 further studies are required to fully understand the effect of using secondary calibrations on  
351 time estimates and downstream analyses.

352 The input chronograms may have been estimated using primary fossil data or using  
353 estimates of molecular substitution rates. These differences in approach which can deepen  
354 the already substantial variation in time estimates between lineages, as observed from the  
355 comparison of source chronograms in the Fringillidae example. For example, the  
356 chronograms from Burns et al. (2014) tend to have much older date estimates for the same  
357 nodes than other inferences (Figure 5). This study did not use any primary calibrations, and  
358 inferred dates using an substitution rate estimates across birds (Weir & Schluter, 2008)

359 Conflict among alternative date estimates is common in the literature (see, for  
360 example, the robust ongoing debate about crown group age of angiosperms (Barba-Montoya,  
361 Reis, Schneider, Donoghue, & Yang, 2018; Magallón et al., 2015; Ramshaw et al., 1972;  
362 Michael J. Sanderson & Doyle, 2001; Sauquet, Ramírez-Barahona, & Magallón, 2021). Using  
363 a different chronogram may return different results (Title & Rabosky, 2016) particularly for  
364 studies of species diversification, or focussed on the timing of evolutionary events. Stitching  
365 together multiple chronograms creates a larger tree that uses information from across studies,  
366 but may amplify uncertainty and errors.

367 Summarizing chronograms might also imply summarizing fundamentally distinct  
368 evolutionary hypotheses. For example, two different researchers working on the same clade  
369 could both carefully select and argue for their choices of fossil calibrations. Still, if one  
370 researcher decides a fossil will calibrate the ingroup of a clade, while another researcher uses  
371 the same one to calibrate outside the clade, the resulting age estimates will often differ

372 substantially, as the placement of calibrations as stem or crown group is proved to deeply  
373 affect estimated times of lineage divergence (Sauquet, 2013). Trying to summarize the  
374 resulting chronograms into a single one using simple summary statistics can erase many  
375 types of relevant information from the source chronograms. Accordingly, the prevailing view  
376 is that we should favor lineage divergence estimates obtained from a single analysis, using  
377 fossil data as primary sources of calibrations, and using fossils that have been widely  
378 discussed and curated as calibrations to date other trees, making sure that all data used in  
379 the analysis reflect a coherent evolutionary history, as for example done by (Antonelli et al.,  
380 2017). There have been attempts to create fossil calibration databases , though these still  
381 have room to grow. However, the exercise of summarizing different chronograms has the  
382 potential to help getting a single global evolutionary history for a lineage by putting together  
383 evidence from different hypotheses. Choosing the elements of the chronograms that we are  
384 going to keep and the ones that we are going to discard is key, since we are potentially losing  
385 important parts of the evolutionary history of a lineage that might only be reflected in  
386 source chronograms and not on the summary chronogram (Sauquet et al., 2021).

387 Nonetheless, in ecology and conservation biology, incorporating at least some data on  
388 lineage divergence times represents a relevant improvement for testing alternative hypothesis  
389 using phylogenetic distance (Campbell O. Webb et al., 2008). Hence, we integrated into  
390 datelife’s workflow different ways of estimating node ages in the absence of calibrations and  
391 branch length information for taxa lacking this information. “Making up” branch lengths is  
392 an common practice in scientific publications: Jetz et al. (2012), created a time-calibrated  
393 tree of all 9,993 bird species, where 67% had molecular data and the rest was simulated;  
394 Rabosky et al. (2018) created a time-calibrated tree of 31,536 ray-finned fishes, of which only  
395 37% had molecular data; Stephen A. Smith and Brown (2018) constructed a tree of 353,185  
396 seed plants where only 23% had molecular data. Obviously, there are risks in this practice!  
397 Taken to the extreme, one could make a fully resolved, calibrated tree of all modern and  
398 extinct taxa using a single taxonomy and a single calibration with the polytomy resolution

399 and branch estimation methods. There has yet to be a thorough analysis of what can go  
400 wrong when one extends inferences beyond the data in this way, so we urge caution; we also  
401 urge readers to follow the example of many of the large tree papers cited above, by carefully  
402 considering the statistical assumptions being made, and assessing the consistency of the  
403 results with prior work.

404

## Conclusions

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

413 The `datelife` R package allows easy and fast summarization of publicly available  
414 information on time of lineage divergence. This provides a straightforward way to get an  
415 informed idea on the state of knowledge of the time frame of evolution of different regions of  
416 the tree of life, and allows identification of regions that require more research or that have  
417 conflicting information. It is available as an R package, or a web-based R shiny app at  
418 [dates.opentreeloflife.org/datelife](http://dates.opentreeloflife.org/datelife). Both summary and newly generated trees are useful to  
419 evaluate evolutionary hypotheses in different areas of research. The DateLife project helps  
420 with awareness of the existing variation in expert time of divergence data, and will foster  
421 exploration of the effect of alternative divergence time hypothesis on the results of analyses,  
422 nurturing a culture of more cautious interpretation of evolutionary results.

**Availability**

423       **Availability**

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

**Supplementary Material**

433       **Supplementary Material**

434       Code used to generate all versions of this manuscript, the biological examples, as well  
435       as the benchmark of functionalities are available at datelifeMS1, datelife\_examples, and  
436       datelife\_benchmark repositories in LLSR's GitHub account.

**Funding**

437       **Funding**

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

**Acknowledgements**

442       **Acknowledgements**

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

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

**References**

- 455 Alström, P., Hooper, D. M., Liu, Y., Olsson, U., Mohan, D., Gelang, M., ... Price, T.  
456 D. (2014). Discovery of a relict lineage and monotypic family of passerine birds.  
457 *Biology Letters*, 10(3), 20131067.
- 458 Ané, C., Eulensteiner, O., Piaggio-Talice, R., & Sanderson, M. J. (2009). Groves of  
459 phylogenetic trees. *Annals of Combinatorics*, 13(2), 139–167.
- 460 Antonelli, A., Hettling, H., Condamine, F. L., Vos, K., Nilsson, R. H., Sanderson, M.  
461 J., ... Vos, R. A. (2017). Toward a self-updating platform for estimating rates of  
462 speciation and migration, ages, and relationships of Taxa. *Systematic Biology*,  
463 66(2), 153–166. <https://doi.org/10.1093/sysbio/syw066>
- 464 Archie, J., Day, W. H., Felsenstein, J., Maddison, W., Meacham, C., Rohlf, F. J., &  
465 Swofford, D. (1986). The Newick tree format. Retrieved from  
466 %7B<https://evolution.genetics.washington.edu/phylip/newicktree.html>%7D
- 467 Avibase. (2022). Yellow-throated Bunting. *Avibase - The World Bird Database*,  
468 (Online Resource). Retrieved from %7B<https://avibase.bsc-eoc.org/species.jsp?lang=EN&avibaseid=82D1EE0049D8D927%7D>
- 469 Bapst, D. W. (2012). Paleotree: An R package for paleontological and phylogenetic  
470 analyses of evolution. *Methods in Ecology and Evolution*, 3(5), 803–807.  
471 <https://doi.org/10.1111/j.2041-210X.2012.00223.x>
- 472 Barba-Montoya, J., Reis, M. dos, Schneider, H., Donoghue, P. C., & Yang, Z. (2018).  
473 Constraining uncertainty in the timescale of angiosperm evolution and the  
474 veracity of a cretaceous terrestrial revolution. *New Phytologist*, 218(2), 819–834.
- 475 Barker, F. K. (2014). Mitogenomic data resolve basal relationships among passeriform  
476 and passeridan birds. *Molecular Phylogenetics and Evolution*, 79, 313–324.
- 477 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2013). Going  
478 to extremes: Contrasting rates of diversification in a recent radiation of new world  
479 passerine birds. *Systematic Biology*, 62(2), 298–320.
- 480
- 481

- 482 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2015). New  
483 insights into new world biogeography: An integrated view from the phylogeny of  
484 blackbirds, cardinals, sparrows, tanagers, warblers, and allies. *The Auk: Ornithological Advances*, 132(2), 333–348.
- 485
- 486 Barker, F. K., Cibois, A., Schikler, P., Feinstein, J., & Cracraft, J. (2004). Phylogeny  
487 and diversification of the largest avian radiation. *Proceedings of the National  
488 Academy of Sciences*, 101(30), 11040–11045.
- 489 Beresford, P., Barker, F., Ryan, Pg., & Crowe, T. (2005). African endemics span the  
490 tree of songbirds (passeri): Molecular systematics of several evolutionary ‘enigmas.’  
491 *Proceedings of the Royal Society B: Biological Sciences*, 272(1565), 849–858.
- 492 Boyle, B., Hopkins, N., Lu, Z., Raygoza Garay, J. A., Mozzherin, D., Rees, T., ...  
493 Enquist, B. J. (2013). The taxonomic name resolution service: An online tool for  
494 automated standardization of plant names. *BMC Bioinformatics*, 14(1).  
495 <https://doi.org/10.1186/1471-2105-14-16>
- 496 Britton, T., Anderson, C. L., Jacquet, D., Lundqvist, S., & Bremer, K. (2007).  
497 Estimating Divergence Times in Large Phylogenetic Trees. *Systematic Biology*,  
498 56(788777878), 741–752. <https://doi.org/10.1080/10635150701613783>
- 499 Bryson Jr, R. W., Chaves, J., Smith, B. T., Miller, M. J., Winker, K., Pérez-Emán, J.  
500 L., & Klicka, J. (2014). Diversification across the new world within the  
501 ‘blue’cardinalids (aves: cardinalidae). *Journal of Biogeography*, 41(3), 587–599.
- 502 Burleigh, J. G., Kimball, R. T., & Braun, E. L. (2015). Building the avian tree of life  
503 using a large-scale, sparse supermatrix. *Molecular Phylogenetics and Evolution*,  
504 84, 53–63.
- 505 Burns, K. J., Shultz, A. J., Title, P. O., Mason, N. A., Barker, F. K., Klicka, J., ...  
506 Lovette, I. J. (2014). Phylogenetics and diversification of tanagers (passeriformes:  
507 Thraupidae), the largest radiation of neotropical songbirds. *Molecular Phylogenetics and Evolution*, 75, 41–77.
- 508

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

- 536                   *and Evolution*, 7(11), 2983–2995.
- 537                   Hackett, S. J., Kimball, R. T., Reddy, S., Bowie, R. C., Braun, E. L., Braun, M. J.,  
538                   ... others. (2008). A phylogenomic study of birds reveals their evolutionary  
539                   history. *Science*, 320(5884), 1763–1768.
- 540                   Harmon, L., Weir, J., Brock, C., Glor, R., & Challenger, W. (2008). GEIGER:  
541                   investigating evolutionary radiations. *Bioinformatics*, 24, 129–131.
- 542                   Hedges, S. B., Marin, J., Suleski, M., Paymer, M., & Kumar, S. (2015). Tree of life  
543                   reveals clock-like speciation and diversification. *Molecular Biology and Evolution*,  
544                   32(4), 835–845. <https://doi.org/10.1093/molbev/msv037>
- 545                   Heibl, C. (2008). *PHYLOCH: R language tree plotting tools and interfaces to diverse*  
546                   *phylogenetic software packages*. Retrieved from  
547                   <http://www.christophheibl.de/Rpackages.html>
- 548                   Hooper, D. M., & Price, T. D. (2017). Chromosomal inversion differences correlate  
549                   with range overlap in passerine birds. *Nature Ecology & Evolution*, 1(10), 1526.
- 550                   Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of  
551                   phylogenetic trees. *Bioinformatics*, 17(8), 754–755.  
552                   <https://doi.org/10.1093/bioinformatics/17.8.754>
- 553                   Jetz, W., Thomas, G., Joy, J. J. B., Hartmann, K., & Mooers, A. (2012). The global  
554                   diversity of birds in space and time. *Nature*, 491(7424), 444–448.  
555                   <https://doi.org/10.1038/nature11631>
- 556                   Johansson, U. S., Fjeldså, J., & Bowie, R. C. (2008). Phylogenetic relationships  
557                   within passerida (aves: Passeriformes): A review and a new molecular phylogeny  
558                   based on three nuclear intron markers. *Molecular Phylogenetics and Evolution*,  
559                   48(3), 858–876.
- 560                   Katoh, K., Asimenos, G., & Toh, H. (2009). Multiple alignment of DNA sequences  
561                   with MAFFT. In *Bioinformatics for DNA sequence analysis* (pp. 39–64).  
562                   Springer.

- 563 Kimball, R. T., Oliveros, C. H., Wang, N., White, N. D., Barker, F. K., Field, D. J.,  
564 ... others. (2019). A phylogenomic supertree of birds. *Diversity*, 11(7), 109.
- 565 Klicka, J., Barker, F. K., Burns, K. J., Lanyon, S. M., Lovette, I. J., Chaves, J. A., &  
566 Bryson Jr, R. W. (2014). A comprehensive multilocus assessment of sparrow (aves:  
567 Passerellidae) relationships. *Molecular Phylogenetics and Evolution*, 77, 177–182.
- 568 Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M.,  
569 Martinez-Barrio, A., ... others. (2015). Evolution of darwin's finches and their  
570 beaks revealed by genome sequencing. *Nature*, 518(7539), 371–375.
- 571 Laubichler, M. D., & Maienschein, J. (2009). *Form and function in developmental  
572 evolution*. Cambridge University Press.
- 573 Lepage, D. (2004). *Avibase: The world bird database*. Bird Studies Canada.
- 574 Lepage, D., Vaidya, G., & Guralnick, R. (2014). Avibase—a database system for  
575 managing and organizing taxonomic concepts. *ZooKeys*, (420), 117.
- 576 Lerner, H. R., Meyer, M., James, H. F., Hofreiter, M., & Fleischer, R. C. (2011).  
577 Multilocus resolution of phylogeny and timescale in the extant adaptive radiation  
578 of hawaiian honeycreepers. *Current Biology*, 21(21), 1838–1844.
- 579 Lovette, I. J., Pérez-Emán, J. L., Sullivan, J. P., Banks, R. C., Fiorentino, I.,  
580 Córdoba-Córdoba, S., ... others. (2010). A comprehensive multilocus phylogeny  
581 for the wood-warblers and a revised classification of the parulidae (aves).  
582 *Molecular Phylogenetics and Evolution*, 57(2), 753–770.
- 583 Magallon, S., & Sanderson, M. J. (2001). Absolute diversification rates in angiosperm  
584 clades. *Evolution*, 55(9), 1762–1780.
- 585 Magallón, S. (2010). Using fossils to break long branches in molecular dating: A  
586 comparison of relaxed clocks applied to the origin of angiosperms. *Systematic  
587 Biology*, 59(4), 384–399.
- 588 Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., & Hernández-Hernández, T.  
589 (2015). A metacalibrated time-tree documents the early rise of flowering plant

- 590 phylogenetic diversity. *New Phytologist*, 207(2), 437–453.
- 591 McTavish, E. J., Hinchliff, C. E., Allman, J. F., Brown, J. W., Cranston, K. A.,  
592 Holder, M. T., ... Smith, S. A. (2015). Phylesystem: A git-based data store for  
593 community-curated phylogenetic estimates. *Bioinformatics*, 31(17), 2794–2800.
- 594 Michonneau, F., Brown, J. W., & Winter, D. J. (2016). rotl: an R package to interact  
595 with the Open Tree of Life data. *Methods in Ecology and Evolution*, 7(12),  
596 1476–1481. <https://doi.org/10.1111/2041-210X.12593>
- 597 Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecology*  
598 Letters, 17(4), 508–525. <https://doi.org/10.1111/ele.12251>
- 599 Moyle, R. G., Oliveros, C. H., Andersen, M. J., Hosner, P. A., Benz, B. W., Manthey,  
600 J. D., ... Faircloth, B. C. (2016). Tectonic collision and uplift of wallacea  
601 triggered the global songbird radiation. *Nature Communications*, 7(1), 1–7.
- 602 Ödeen, A., Håstad, O., & Alström, P. (2011). Evolution of ultraviolet vision in the  
603 largest avian radiation—the passerines. *BMC Evolutionary Biology*, 11(1), 1–8.
- 604 Oliveros, C. H., Field, D. J., Ksepka, D. T., Barker, F. K., Aleixo, A., Andersen, M.  
605 J., ... others. (2019). Earth history and the passerine superradiation.  
606 *Proceedings of the National Academy of Sciences*, 116(16), 7916–7925.
- 607 Ooms, J., & Chamberlain, S. (2018). *Phylocomr: Interface to 'phylocom'*. Retrieved  
608 from <https://CRAN.R-project.org/package=phylocomr>
- 609 Open Tree Of Life, Redelings, B., Cranston, K. A., Allman, J., Holder, M. T., &  
610 McTavish, E. J. (2016). Open Tree of Life APIs v3.0. *Open Tree of Life Project*,  
611 (Online Resources). Retrieved from  
612 %7B<https://github.com/OpenTreeOfLife/germinator/wiki/Open-Tree-of-Life->  
613 Web-APIs%7D
- 614 Open Tree Of Life, Redelings, B., Sánchez Reyes, L. L., Cranston, K. A., Allman, J.,  
615 Holder, M. T., & McTavish, E. J. (2019). Open tree of life synthetic tree v12.3.  
616 *Zenodo*. Retrieved from <https://doi.org/10.5281/zenodo.3937742>

- 617 Päckert, M., Martens, J., Sun, Y.-H., Severinghaus, L. L., Nazarenko, A. A., Ting, J.,  
618 ... Tietze, D. T. (2012). Horizontal and elevational phylogeographic patterns of  
619 himalayan and southeast asian forest passerines (aves: passeriformes). *Journal of*  
620 *Biogeography*, 39(3), 556–573.
- 621 Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and  
622 evolution in R language. *Bioinformatics*, 20(2), 289–290.
- 623 Parchman, T. L., Benkman, C. W., & Mezquida, E. T. (2007). Coevolution between  
624 hispaniolan crossbills and pine: Does more time allow for greater phenotypic  
625 escalation at lower latitude? *Evolution*, 61(9), 2142–2153.
- 626 Posadas, P., Crisci, J. V., & Katinas, L. (2006). Historical biogeography: A review of  
627 its basic concepts and critical issues. *Journal of Arid Environments*, 66(3),  
628 389–403.
- 629 Powell, A. F., Barker, F. K., Lanyon, S. M., Burns, K. J., Klicka, J., & Lovette, I. J.  
630 (2014). A comprehensive species-level molecular phylogeny of the new world  
631 blackbirds (icteridae). *Molecular Phylogenetics and Evolution*, 71, 94–112.
- 632 Price, T. D., Hooper, D. M., Buchanan, C. D., Johansson, U. S., Tietze, D. T.,  
633 Alström, P., ... others. (2014). Niche filling slows the diversification of himalayan  
634 songbirds. *Nature*, 509(7499), 222.
- 635 Pulgarín-R, P. C., Smith, B. T., Bryson Jr, R. W., Spellman, G. M., & Klicka, J.  
636 (2013). Multilocus phylogeny and biogeography of the new world pheucticus  
637 grosbeaks (aves: cardinalidae). *Molecular Phylogenetics and Evolution*, 69(3),  
638 1222–1227.
- 639 R Core Team. (2018). *R: a language and environment for statistical computing*.  
640 Vienna, Austria: R Foundation for Statistical Computing.
- 641 Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., ...  
642 others. (2018). An inverse latitudinal gradient in speciation rate for marine fishes.  
643 *Nature*, 559(7714), 392.

- 644 Ramshaw, J., Richardson, D., Mealyard, B., Brown, R., Richardson, M., Thompson,  
645 E., & Boulter, D. (1972). The time of origin of the flowering plants determined by  
646 using amino acid sequence data of cytochrome c. *New Phytologist*, 71(5), 773–779.
- 647 Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The barcode of life data system  
648 (<http://www.Barcodinglife.org>). *Molecular Ecology Notes*, 7(3), 355–364.
- 649 Rees, J. A., & Cranston, K. (2017). Automated assembly of a reference taxonomy for  
650 phylogenetic data synthesis. *Biodiversity Data Journal*, (5).
- 651 Revell, L. J. (2012). Phytools: An r package for phylogenetic comparative biology  
652 (and other things). *Methods in Ecology and Evolution*, 3, 217–223.
- 653 Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic  
654 inference under mixed models. *Bioinformatics*, 19(12), 1572–1574.  
<https://doi.org/10.1093/bioinformatics/btg180>
- 655 Sanchez-Reyes, L. L., & O'Meara, B. (2022). **datelifeplot**: Methods to plot  
656 chronograms and outputs of the **datelife** package. *R Package Release V0.2.2*.  
657 Retrieved from <https://zenodo.org/badge/latestdoi/381501451>
- 658 Sanchez-Reyes, L. L., O'Meara, B., Eastman, J., Heath, T., Wright, A., Schliep, K.,  
659 ... Alfaro, M. (2022). **datelife**: Scientific Data on Time of Lineage Divergence  
660 for Your Taxa. *R Package Release V0.6.2*. Retrieved from  
661 <https://doi.org/10.5281/zenodo.593938>
- 662 Sanderson, Michael J. (2002). Estimating Absolute Rates of Molecular Evolution and  
663 Divergence Times: A Penalized Likelihood Approach. *Molecular Biology and*  
664 *Evolution*, 19(1), 101–109.  
665 <https://doi.org/10.1093/oxfordjournals.molbev.a003974>
- 666 Sanderson, Michael J. (2003). r8s: Inferring absolute rates of molecular evolution and  
667 divergence times in the absence of a molecular clock. *Bioinformatics*, 19(2),  
668 301–302.
- 669 Sanderson, Michael J., & Doyle, J. A. (2001). Sources of error and confidence

- 671 intervals in estimating the age of angiosperms from *rbcL* and 18S rDNA data.
- 672     *American Journal of Botany*, 88(8), 1499–1516.
- 673     Sauquet, H. (2013). A practical guide to molecular dating. *Comptes Rendus Palevol*,
- 674       12(6), 355–367.
- 675     Sauquet, H., Ramírez-Barahona, S., & Magallón, S. (2021). *The age of flowering*
- 676       *plants is unknown*.
- 677     Schenk, J. J. (2016). Consequences of secondary calibrations on divergence time
- 678       estimates. *PLoS ONE*, 11(1). <https://doi.org/10.1371/journal.pone.0148228>
- 679     Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in r. *Bioinformatics*, 27(4),
- 680       592–593.
- 681     Selvatti, A. P., Gonzaga, L. P., & Moraes Russo, C. A. de. (2015). A paleogene origin
- 682       for crown passerines and the diversification of the oscines in the new world.
- 683       *Molecular Phylogenetics and Evolution*, 88, 1–15.
- 684     Smith, Stephen A., & Brown, J. W. (2018). Constructing a broadly inclusive seed
- 685       plant phylogeny. *American Journal of Botany*, 105(3), 302–314.
- 686     Smith, Stephen A., & O'Meara, B. C. (2012). TreePL: Divergence time estimation
- 687       using penalized likelihood for large phylogenies. *Bioinformatics*, 28(20),
- 688       2689–2690. <https://doi.org/10.1093/bioinformatics/bts492>
- 689     Stoltzfus, A., Lapp, H., Matasci, N., Deus, H., Sidlauskas, B., Zmasek, C. M., ...
- 690     Jordan, G. (2013). Phylotastic! Making tree-of-life knowledge accessible, reusable
- 691       and convenient. *BMC Bioinformatics*, 14.
- 692       <https://doi.org/10.1186/1471-2105-14-158>
- 693     Tietze, D. T., Päckert, M., Martens, J., Lehmann, H., & Sun, Y.-H. (2013). Complete
- 694       phylogeny and historical biogeography of true rosefinches (aves: *carpodacus*).
- 695       *Zoological Journal of the Linnean Society*, 169(1), 215–234.
- 696     Title, P. O., & Rabosky, D. L. (2016). Do Macrophylogenies Yield Stable
- 697       Macroevolutionary Inferences? An Example from Squamate Reptiles. *Systematic*

- 698        *Biology*, syw102. <https://doi.org/10.1093/sysbio/syw102>
- 699        Treplin, S., Siegert, R., Bleidorn, C., Thompson, H. S., Fotso, R., & Tiedemann, R.
- 700        (2008). Molecular phylogeny of songbirds (aves: Passeriformes) and the relative
- 701        utility of common nuclear marker loci. *Cladistics*, 24(3), 328–349.
- 702        Vos, R. A., Balhoff, J. P., Caravas, J. A., Holder, M. T., Lapp, H., Maddison, W. P.,
- 703        ... others. (2012). NeXML: Rich, extensible, and verifiable representation of
- 704        comparative data and metadata. *Systematic Biology*, 61(4), 675–689.
- 705        Webb, C. (2000). Exploring the Phylogenetic Structure of Ecological Communities :
- 706        An Example for Rain Forest Trees. *The American Naturalist*, 156(2), 145–155.
- 707        Webb, Campbell O., Ackerly, D. D., & Kembel, S. W. (2008). Phylocom: Software for
- 708        the analysis of phylogenetic community structure and trait evolution.
- 709        *Bioinformatics*, 24(18), 2098–2100.
- 710        <https://doi.org/10.1093/bioinformatics/btn358>
- 711        Webb, Campbell O., & Donoghue, M. J. (2005). Phylomatic: Tree assembly for
- 712        applied phylogenetics. *Molecular Ecology Notes*, 5(1), 181–183.
- 713        Weir, J., & Schluter, D. (2008). Calibrating the avian molecular clock. *Molecular*
- 714        *Ecology*, 17(10), 2321–2328.
- 715        Zuccon, D., Prŷs-Jones, R., Rasmussen, P. C., & Ericson, P. G. (2012). The
- 716        phylogenetic relationships and generic limits of finches (fringillidae). *Molecular*
- 717        *Phylogenetics and Evolution*, 62(2), 581–596.

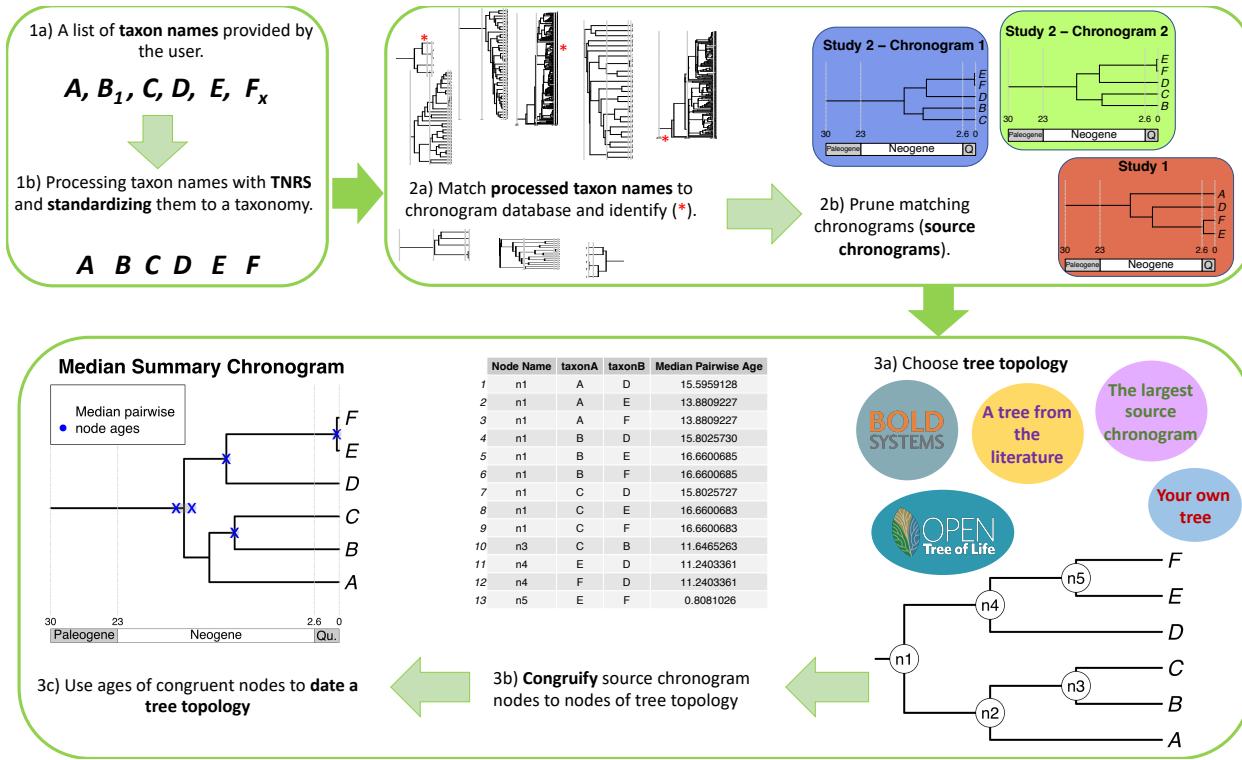


FIGURE 1. Stylized DateLife workflow. This shows the general workflows and analyses that can be performed with `datelife`, via the R package or through the website at <http://www.datelife.org/>. Details on the functions involved on each workflow are shown in `datelife`'s R package vignette.

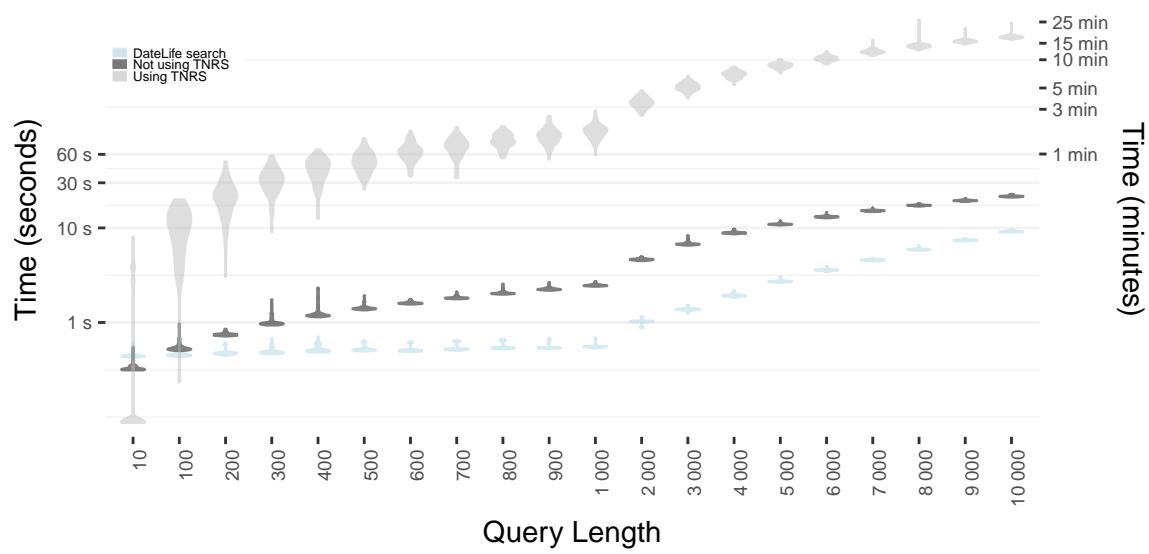


FIGURE 2. Computation time of query processing and search across **datelife**'s chronogram database relative to number of input taxon names. We sampled N names from the class Aves for each cohort 100 times and then performed a search with query processing not using the Taxon Names Resolution Service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

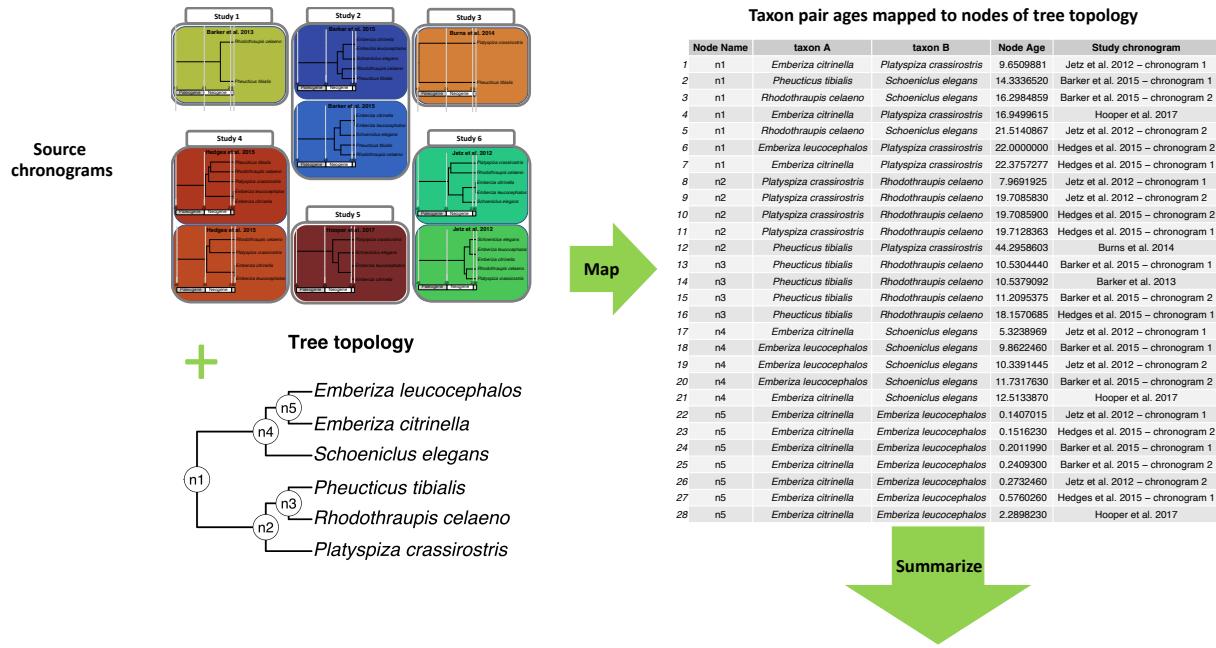


FIGURE 3. Age data results of a DateLife search of a small sample of 6 bird species within the Passeriformes. Input names were found across 9 chronograms within 6 independent studies (Barker et al. (2012), Barker et al. (2015), Burns et al. (2014), Hedges et al. (2015), Hooper and Price (2017), Jetz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

### Summary of mapped taxon pair age data

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

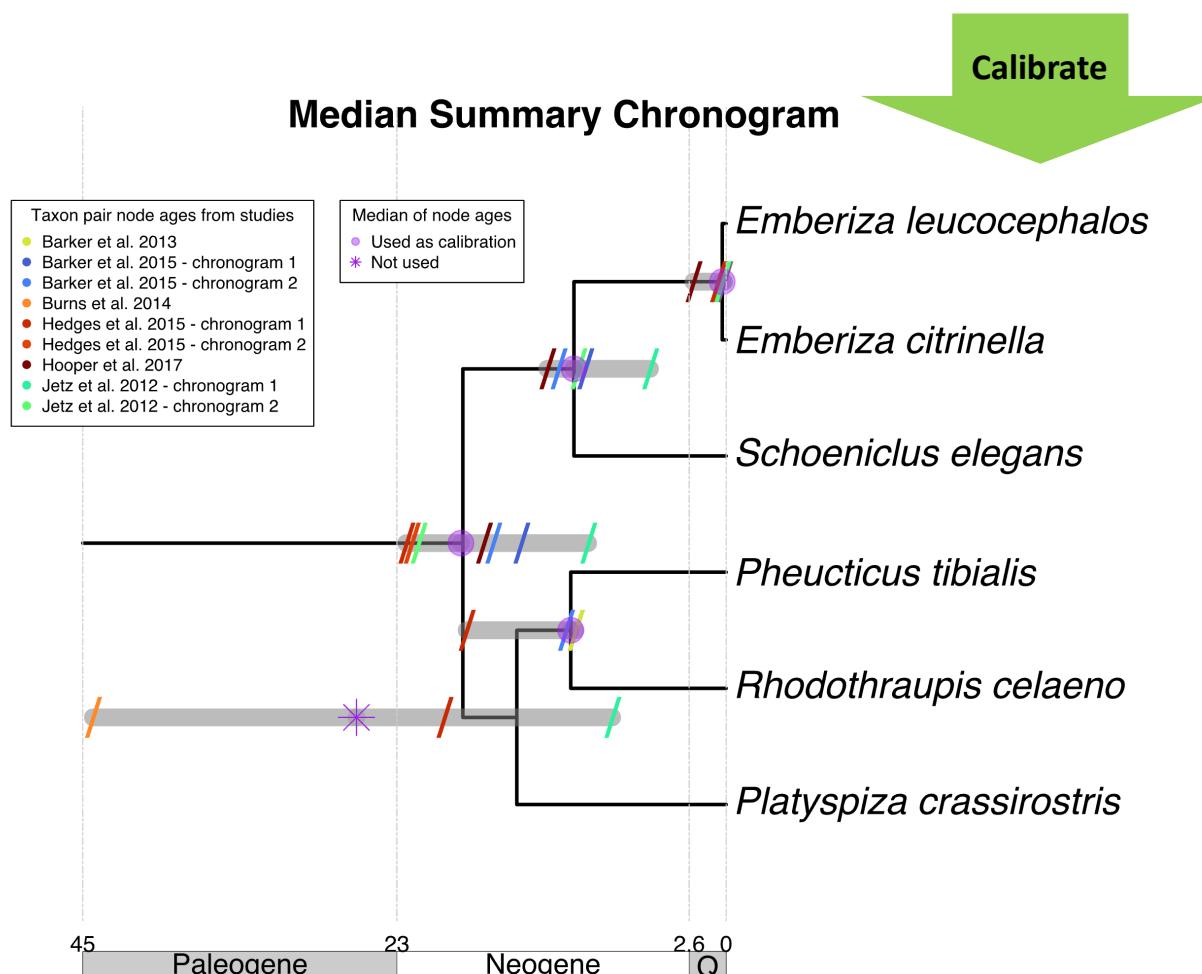


FIGURE 4. Summarized age data is used as secondary calibrations to date a tree topology as a summary chronogram.

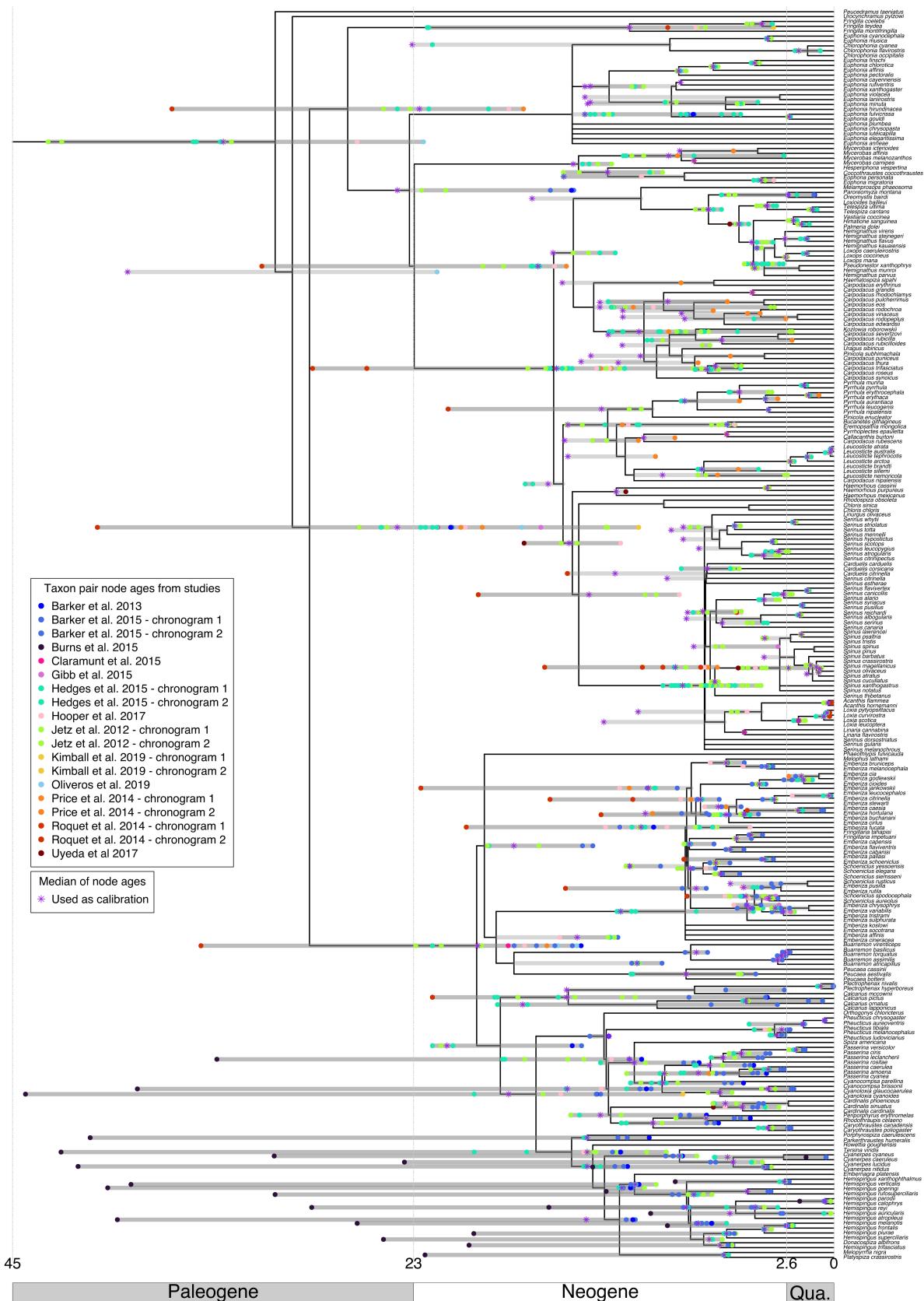


FIGURE 5. Fringillidae median summary chronogram generated with DateLife. It has 256 tips and 233 nodes.

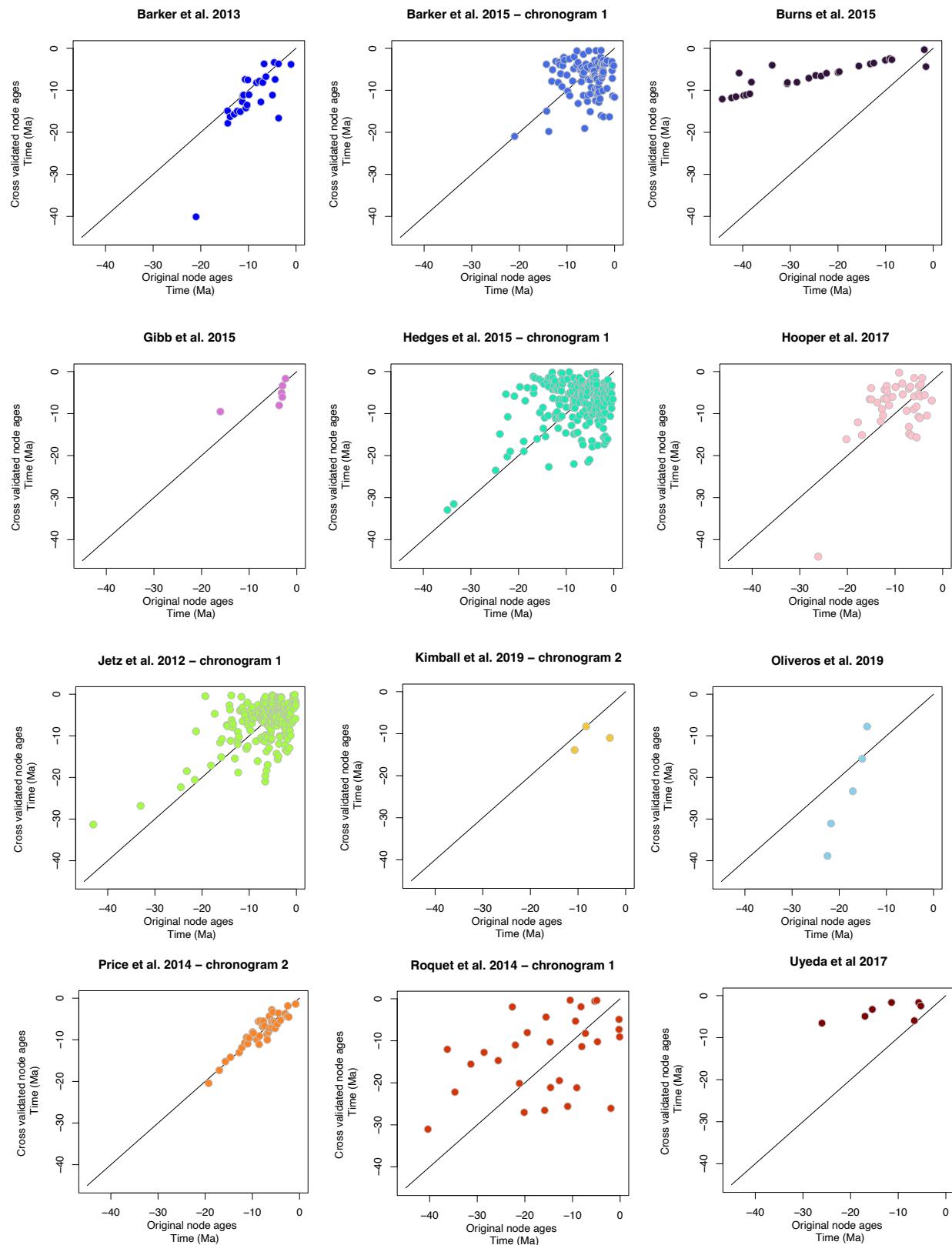


FIGURE 6. Results from cross validation analysis.

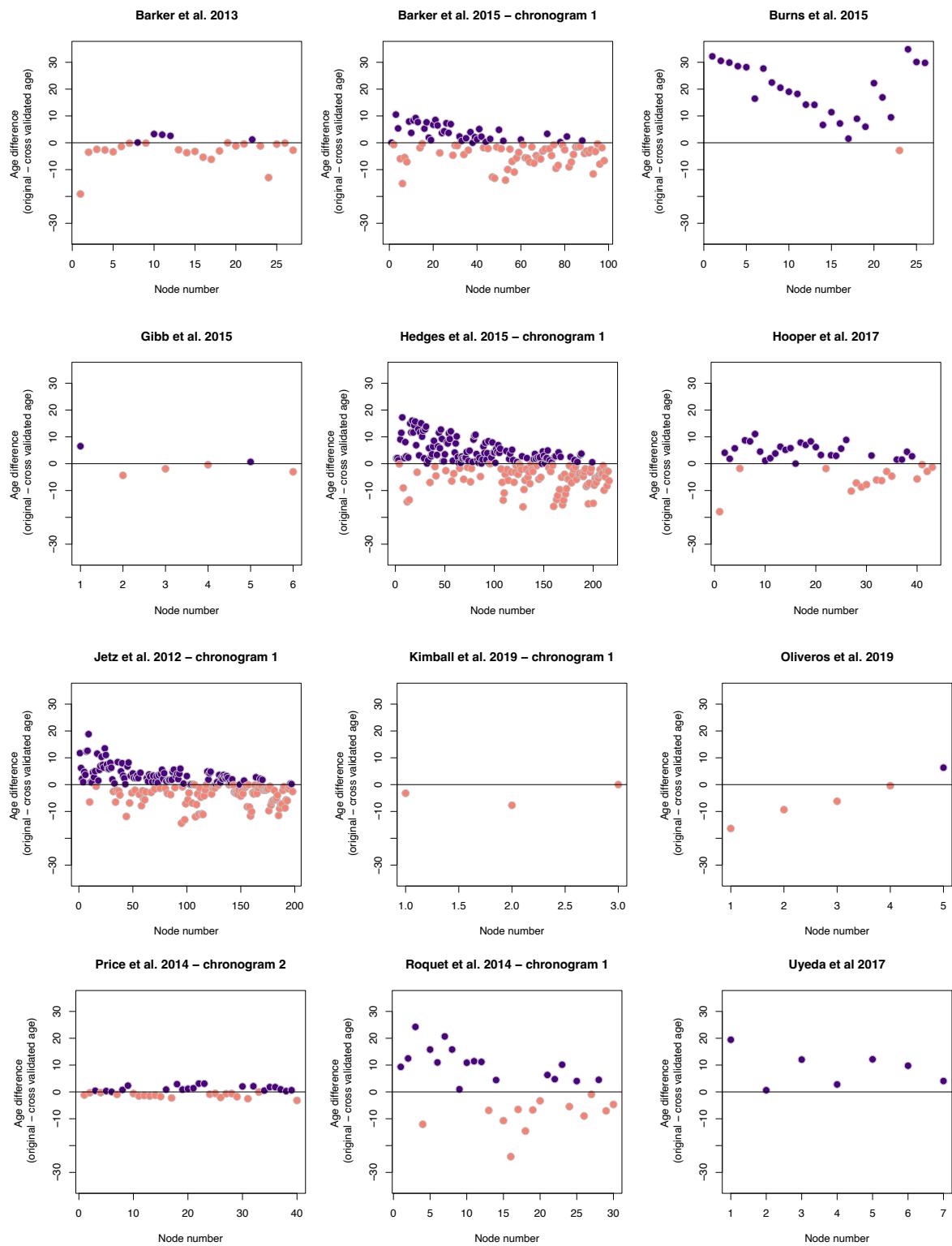


FIGURE 7. Results from cross validation analysis.

## Barker et al. 2015 - chronogram 1



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

## Barker et al. 2015 - chronogram 2

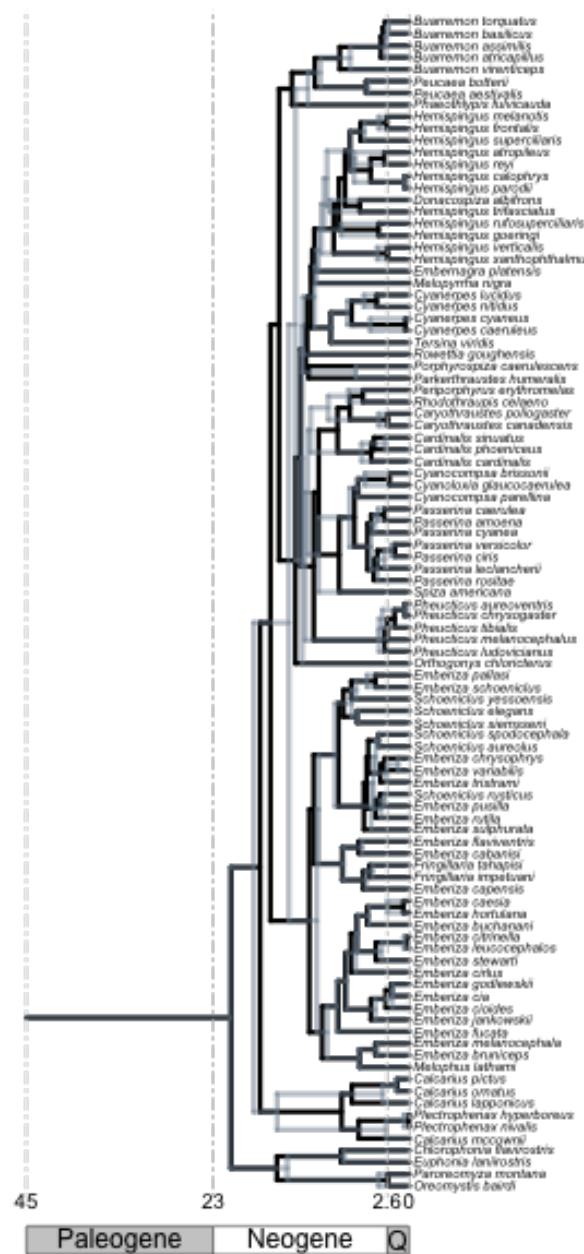


FIGURE 9. Cross validation of third source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the dates calculated with BLADJ, i.e., the same for all the nodes.

**Burns et al. 2015**

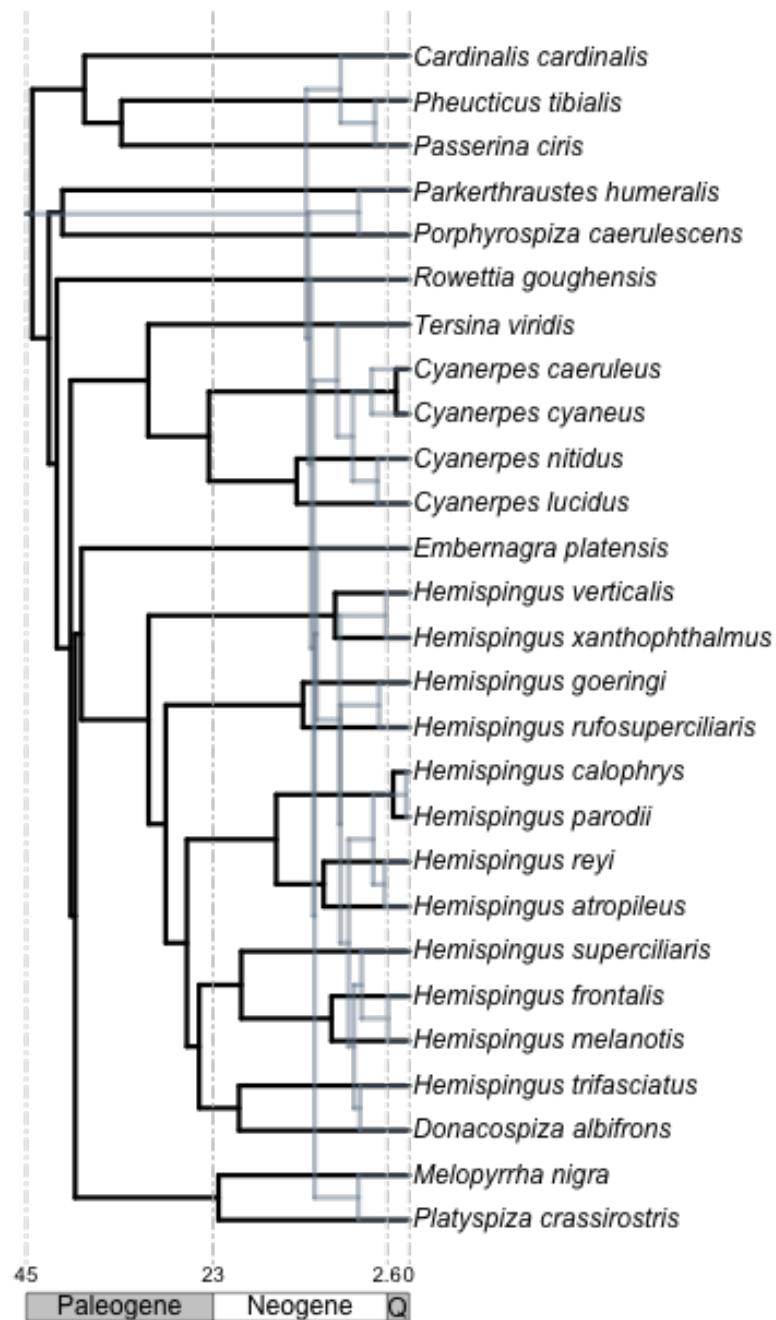


FIGURE 10. Cross validation of fourth source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the dates calculated with BLADJ in each run of the cross-validation.

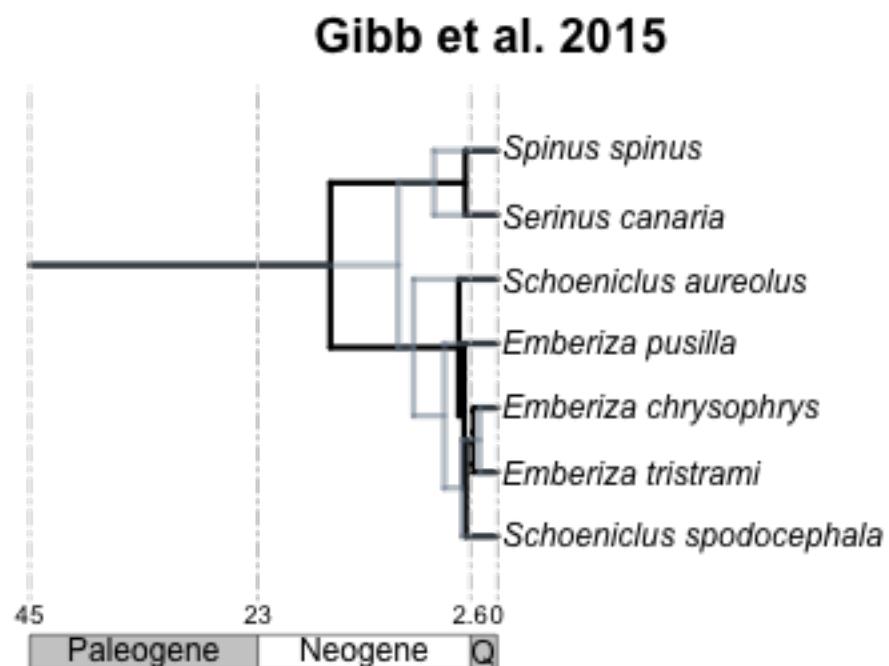


FIGURE 11. Cross validation of sixth source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the same tree topology dated with BLADJ using node ages from all other source chronograms as secondary calibrations.

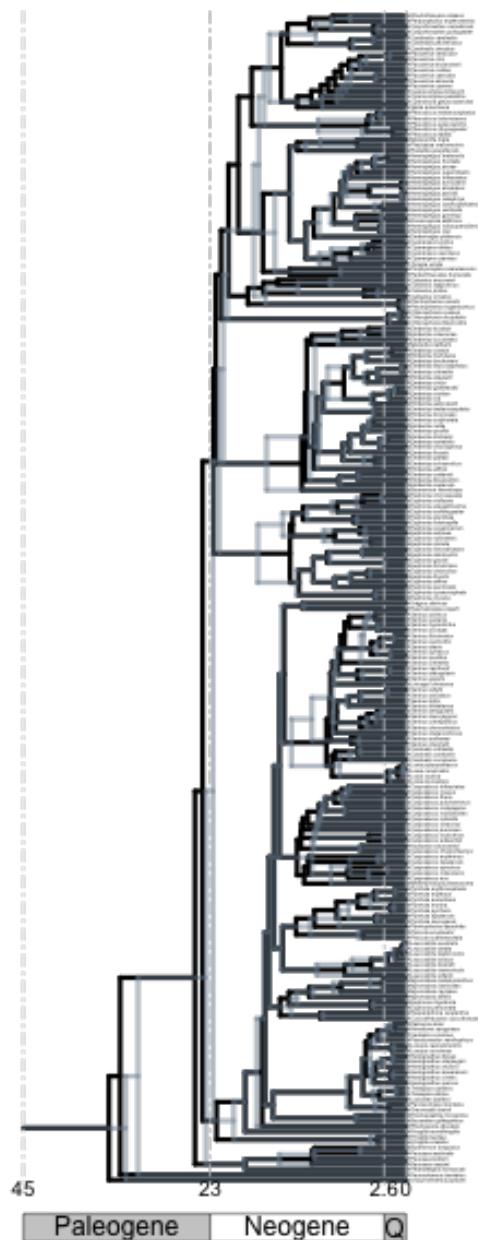
**Hedges et al. 2015 - chronogram 1**

FIGURE 12. Cross validation of seventh source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the dates calculated with BLADeL. In order to facilitate the comparison, the

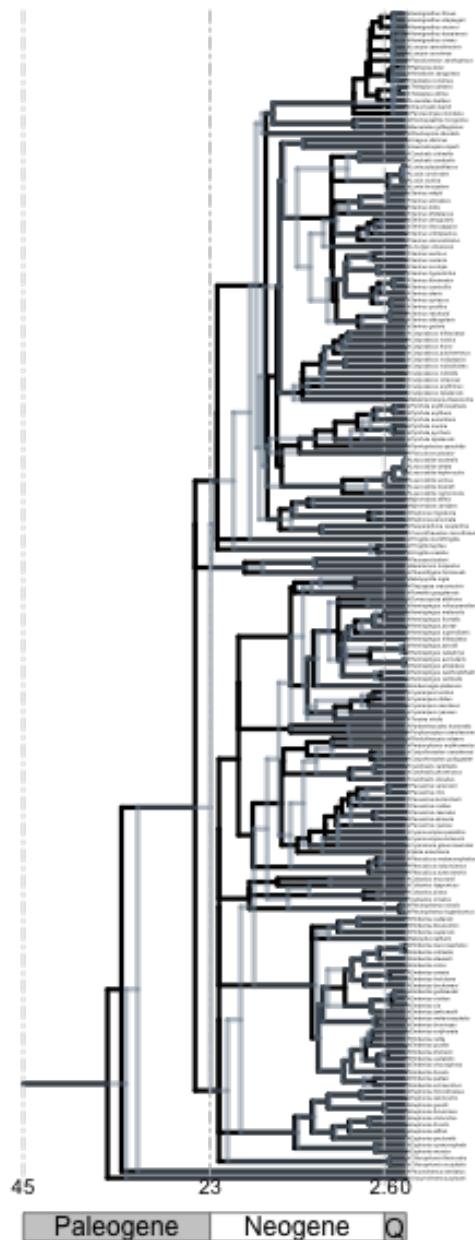
**Hedges et al. 2015 - chronogram 2**

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 DATELIFE in each of the eight source studies.

### Hooper et al. 2017

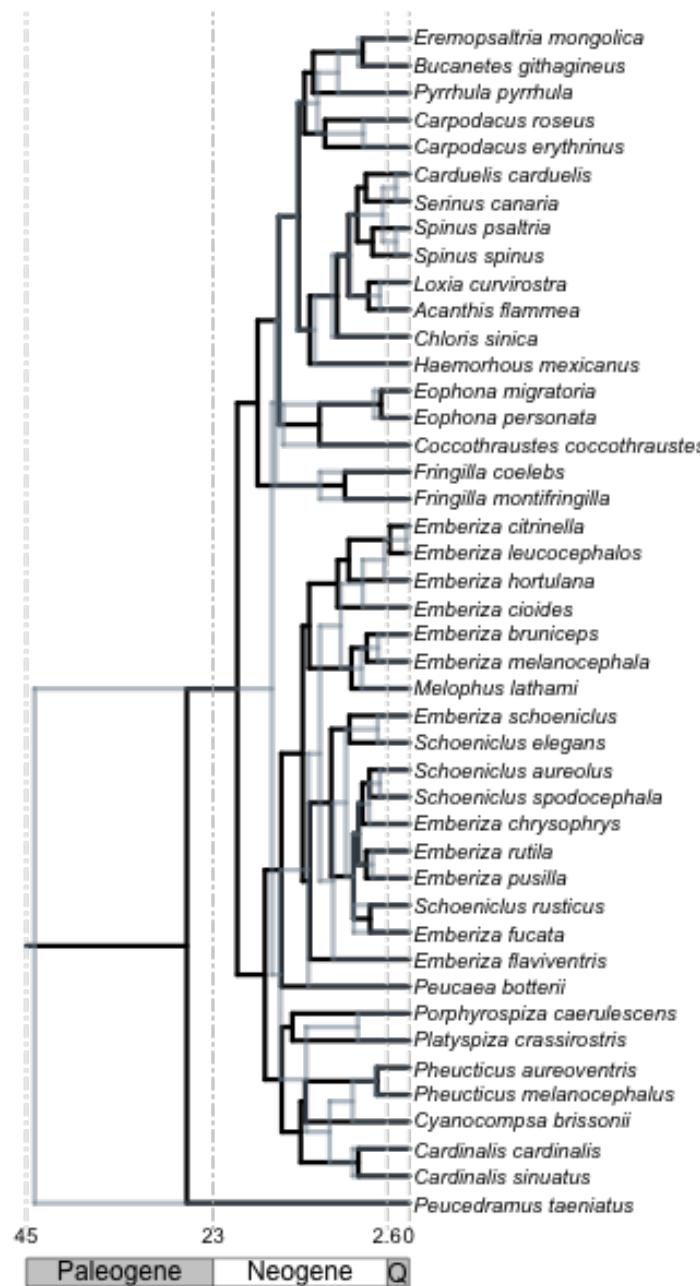


FIGURE 14. Cross validation of ninth source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the dates calculated with BLADJ.

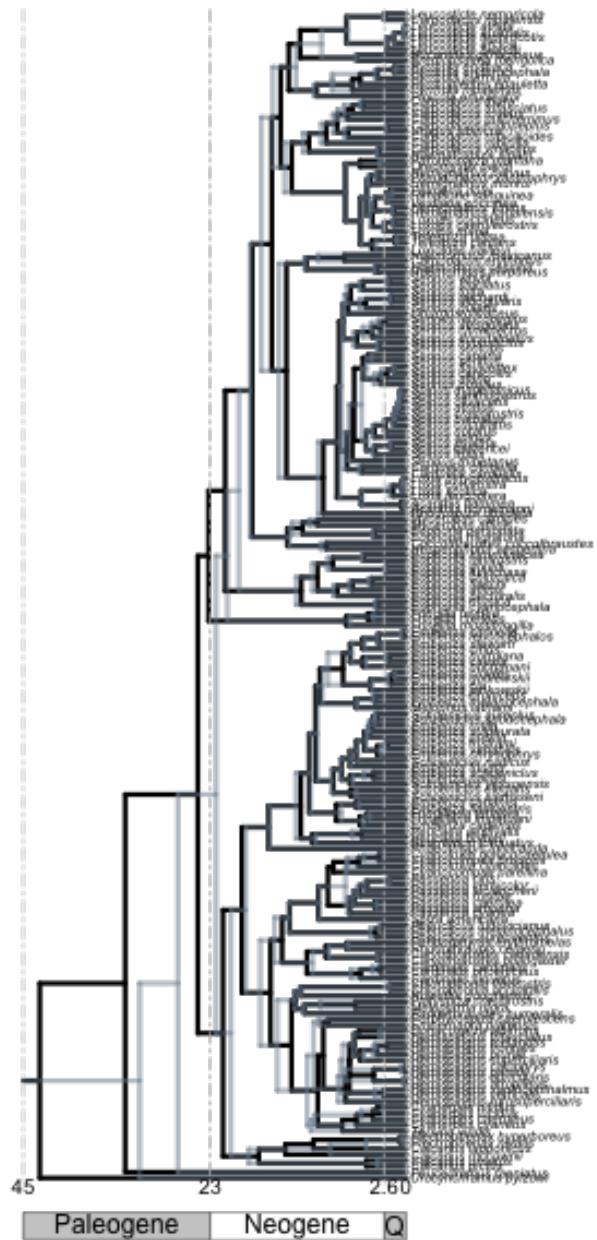
**Jetz et al. 2012 - chronogram 1**

FIGURE 15. Cross validation of tenth source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the dates calculated with BLADe. In each case, the tree is the same.