

¹ DateLife: leveraging databases and analytical tools to reveal the dated Tree of Life

² Luna L. Sánchez Reyes^{1,2}, Emily Jane McTavish¹, & Brian O'Meara²

³ ¹ University of California, Merced, USA

⁴ ² University of Tennessee, Knoxville, USA

⁵ Author Note

⁶ School of Natural Sciences, University of California, Merced, 258 Science and
⁷ Engineering Building 1, Merced, CA 95340, USA.

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

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 and 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, 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;

⁴³ Congruification; Supertree; Calibrations; Secondary calibrations

⁴⁴ Word count: 5629

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

46 Chronograms –phylogenies with branch lengths proportional to time– provide key data
47 on evolutionary time frame for the study of natural processes in many areas of biological
48 research, such as comparative analysis (Freckleton, Harvey, & Pagel, 2002; Harvey, Pagel, &
49 others, 1991), developmental biology (Delsuc et al., 2018; Laubichler & Maienschein, 2009),
50 conservation biology and ecology (Felsenstein, 1985; 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 a
54 homology hypothesis to construct a phylogeny, selecting and placing appropriate calibrations
55 on the phylogeny using independent age data points from the fossil record or other dated
56 events, and inferring a full dated tree. All of this entails specialized biological training,
57 taxonomic domain knowledge, and a non-negligible amount of research time, computational
58 resources and funding.

59 Here we present the DateLife project which has the main goal of extracting and
60 exposing age data from published chronograms, making age data readily accessible to the
61 wider community for reuse and reanalysis, research, teaching, and science communication
62 and policy. DateLife’s core software application is available as an R package (Sanchez-Reyes
63 et al., 2022), and as an online Rshiny interactive website at www.datelife.org. It features key
64 elements for scientific reproducibility, such as a curated, versioned, open and fully public
65 chronogram database (McTavish et al., 2015) that stores data in a computer-readable format
66 (Vos et al., 2012); automated and programmatic ways of accessing and downloading the data
67 in a computer-readable format also (Stoltzfus et al., 2013); and methods to summarize and
68 compare the data.

Description

DateLife's core software application consists of the R package `datelife`. Its current stable version – v0.6.6, is available from The Comprehensive R Archive Network (CRAN) repository (Sanchez-Reyes et al., 2022), and relies on functionalities from various biological R packages: `ape` (Paradis, Claude, & Strimmer, 2004), `bold` (Chamberlain, 2018), `geiger` (Pennell et al., 2014), `msa` (Bodenhofer, Bonatesta, Horejš-Kainrath, & Hochreiter, 2015), `paleotree` (Bapst, 2012), `phyloch` (Heibl, 2008), `phylocomr` (Ooms & Chamberlain, 2018), `phytools` (Revell, 2012), `rotl` (Michonneau, Brown, & Winter, 2016), and `taxize` (Chamberlain, 2018; Chamberlain & Szöcs, 2013). Figure 1 provides a graphical summary of the three main steps of the DateLife workflow: creating a search query, searching a database, and summarizing results from the search.

Creating a search query

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

DateLife processes input scientific names using a Taxonomic Name Resolution Service (TNRS), which increases the probability of correctly finding the queried taxon names in the chronogram database. TNRS detects, corrects and standardizes name misspellings and typos, variant spellings and authorities, and nomenclatural synonyms to a single taxonomic standard (Boyle et al., 2013). TNRS also allows to correctly choose between homonyms, by considering other taxa provided as input to infer the taxonomic context of the homonym. DateLife implements TNRS using the Open Tree of Life (OpenTree) unified Taxonomy (OTT, Open Tree Of Life et al., 2016; Rees & Cranston, 2017) as standard, storing

95 taxonomic identification numbers (OTT ids) for further processing and analysis. Other
96 taxonomies currently supported by DateLife are the National Center of Biotechnology
97 Information (NCBI) taxonomic database (Schoch et al., 2020), the Global Biodiversity
98 Information Facility (GBIF) taxonomic backbone (GBIF Secretariat, 2022), and the Interim
99 Register of Marine and Nonmarine Genera (IRMNG) database (Rees et al., 2017).

100 Besides binomial species names, DateLife accepts scientific names from any inclusive
101 taxonomic group (e.g., genus, family, tribe), as well as subspecific taxonomic variants (e.g.,
102 subspecies, variants, strains). If a taxon name belongs to an inclusive taxonomic group,
103 DateLife has two alternative behaviors defined by the “get species from taxon” flag. If the
104 flag is active, DateLife retrieves all species names within the taxonomic group from the
105 standard taxonomy of choice, and adds them to the search query. In this case, subspecific
106 variants are excluded. If the flag is inactive, DateLife excludes any taxon names above the
107 species level from the search query. Species and subspecific variant names are processed and
108 searched as provided by the user. The processed taxon names are saved as an R object of a
109 newly defined class, `datelifeQuery`, that is used in the following steps. This object contains
110 the input names standardized to a taxonomy of choice (OTT by default), the corresponding
111 OTT id numbers, and the topology of an input tree, if one was provided.

112 **Searching a chronogram database**

113 At the time of writing of this manuscript (Feb 20, 2023), DateLife’s chronogram
114 database latest version consist of 253 chronograms published in 187 different studies. It is
115 curated from OpenTree’s phylogenetic database, the Phylesystem, which constitutes an open
116 source of expert and peer-reviewed phylogenetic knowledge with rich metadata (McTavish et
117 al., 2015), which allows automatic and reproducible assembly of our chronogram database.
118 Datelife’s chronogram database is navigable as an R data object within the `datelife` R
119 package.

A unique feature of the Phylesystem is that any user can add new published, state-of-the-art chronograms any time, through their curator application (<https://tree.opentreeoflife.org/curator>). As chronograms are added to Phylesystem, they can be incorporated into the chronogram database of the `datelife` package. `datelife`'s chronogram database is currently manually updated as new chronogram data is added to Phylesystem. The updated database is assigned a new version number, followed by a package release on CRAN. Users can also implement functions from the `datelife` R package to trigger an update of the local chronogram database, to incorporate any new chronograms to the user's DateLife analysis before an official database update is released on CRAN.

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

Summarizing search results

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

1. The matching pruned chronograms as newick strings or "phylo" objects.
2. The ages of the root of all source chronograms. These ages can correspond to the age

of the most recent common ancestor (mrca) of the user's group of interest if the source chronograms have all taxa belonging to the group. If not, the root corresponds to the mrca of a subgroup within the group of interest.

3. Study citations where original chronograms were published.
4. A report of input taxon names matches across source chronograms.
5. The source chronogram(s) with the most input taxon names.
6. Various single summary chronograms resulting from summarizing age data, generated using the methodology described next.

Choosing a topology

DateLife requires a tree topology to summarize age data upon. We recommend that users provide as input a tree topology from the literature, or one of their own making. If no topology is provided, DateLife automatically extracts one from the OpenTree synthetic tree, a phylogeny currently encompassing 2.3 million taxa across all life, assembled from 1, 239 published phylogenetic trees and OpenTree's unified Taxonomy, OTT (Open Tree Of Life et al., 2019). Alternatively, DateLife can combine topologies from source chronograms using a supertree approach (Criscuolo, Berry, Douzery, & Gascuel, 2006). To do this, DateLife first identifies the source chronograms that form a grove, roughly, a sufficiently overlapping set of taxa between trees, by implementing definition 2.8 for n-overlap from Ané et al. (2009). If the source chronograms do not form a grove, the supertree reconstruction will fail. In rare cases, a group of trees can have multiple groves. By default, DateLife chooses the grove with the most taxa, however, the "criterion = trees" flag allows the user to choose the grove with the most trees instead. The result is a single summary (or supertree) topology, that combines topologies from source chronograms in a grove.

Applying secondary calibrations

Once a topology is chosen, DateLife applies the congruification method (Eastman, Harmon, & Tank, 2013) that finds nodes belonging to the same clade across source

171 chronograms, and then extracts the corresponding node ages from patristic distance matrices
172 stored as a `datelifeResult` object. Note that by definition, these matrices store total
173 distance (time from tip to tip), assuming that the terminal taxa are coeval and occur at the
174 present. Hence, node ages correspond to half the values stored in the `datelifeResult`
175 matrices. A table of congruified node ages that can be used as calibrations for a dating
176 analysis is stored as a `congruifiedCalibrations` object.

177 For each congruent node, the pairwise distances that traverse that node are summarized
178 into a single summary matrix using classic summary statistics (i.e., mean, median, minimum
179 and maximum ages), and the Supermatrix Distance Method (SDM; Criscuolo et al., 2006),
180 which deforms patristic distance matrices by minimizing variance and then averaging them.
181 These single summary taxon pair age matrices are stored as summarized calibrations that
182 can be used as secondary calibrations to date a tree topology - with or without initial branch
183 lengths, using phylogenetic dating methods currently supported within DateLife: BLADJ
184 (Webb, Ackerly, & Kembel, 2008; Webb & Donoghue, 2005), MrBayes (Huelsenbeck &
185 Ronquist, 2001; Ronquist & Huelsenbeck, 2003), PATHd8 (Britton, Anderson, Jacquet,
186 Lundqvist, & Bremer, 2007), and treePL (Smith & O'Meara, 2012).

187 **Dating a tree topology**

188 **With no branch lengths.** When producing or obtaining a tree with branch lengths
189 for a group of interest is not possible, DateLife can date a topology without branch lengths,
190 obtained from OpenTree or by implementing the supertree approach described above, by
191 implementing the Branch Length Adjuster (BLADJ; Webb et al., 2008; Webb & Donoghue,
192 2005) algorithm, which requires no initial branch lengths. The algorithm starts by fixing
193 ages for nodes with calibration data upon the given topology. Then, it distributes time for
194 nodes with no data evenly between calibrated nodes, minimizing age variance in the
195 resulting chronogram (Webb et al., 2008), which has proven useful for ecological analyses
196 that require age data (Webb et al., 2008). When there is conflict between ages of calibrated

197 nodes, BLADJ ignores node ages that are older than the age of a parent node. The BLADJ
198 algorithm requires a root age estimate to run. If there is no information on the age of the
199 root in the chronogram database, users can provide an estimate from the literature. If none
200 is provided, DateLife will not return a dated topology and will provide a warning message
201 along with suggestions on how the user can provide an age for the root so that a fully dated
202 tree can be returned.

203 In the absence of genetic data, we consider that BLADJ is the most agnostic way to
204 assign ages to nodes with no available data, as it does not require any assumptions on the
205 underlying model of branch length distribution. It is however common practice in the
206 literature to use a birth-death model to assign ages to nodes with no genetic data (Jetz,
207 Thomas, Joy, Hartmann, & Mooers, 2012; Rabosky et al., 2018; Smith & Brown, 2018). To
208 do so, DateLife implements MrBayes (Huelsenbeck & Ronquist, 2001; Ronquist &
209 Huelsenbeck, 2003), using nodes with published age data as calibration priors on a fixed
210 topology, a simple birth-death model with parameters that can be determined by the user,
211 and no genetic data.

212 **Dating a tree with branch lengths.** Relative branch lengths provide key
213 information for phylogenetic dating, specifically for nodes without any calibration data
214 available. While, using initial branch lengths data is the golden standard for phylogenetic
215 dating analyses, producing such requires assembling and curating a homology hypothesis,
216 and choosing and implementing a method for phylogenetic inference. DateLife implements a
217 workflow to streamline this process by leveraging on genetic data from the Barcode of Life
218 Data System, BOLD (Ratnasingham & Hebert, 2007) to obtain markers for input taxa. By
219 default, mined genetic sequences are aligned with MUSCLE (Edgar, 2004) using functions
220 from the msa R package (Bodenhofer et al., 2015). Alternatively, sequences can be aligned
221 with MAFFT (Katoh, Asimenos, & Toh, 2009), using functions from the ape R package
222 (Paradis et al., 2004). The BOLD sequence alignment is then used to obtain initial branch
223 lengths with the accelerated transformation (ACCTRAN) parsimony algorithm, which

resolves ambiguous character optimization by assigning changes along branches of the tree as close to the root as possible (Agnarsson & Miller, 2008), resulting in older nodes (Forest et al., 2005). The parsimony branch lengths are then optimized using Maximum Likelihood, given the alignment, the topology and a simple Jukes-Cantor model, producing a BOLD tree with branch lengths proportional to expected number of substitutions per site. Both parsimony and ML optimizations are done with functions from the `phangorn` package (Schliep, 2011). The BOLD workflow is currently only available on DateLife's R package and not on the web application.

BOLD trees generated by DateLife, and any tree with branch lengths provided by a user can be analysed with phylogenetic dating methods supported in DateLife that incorporate branch length information from the input topology in combination with the secondary calibrations: PATHd8, a non-clock, rate-smoothing method to date trees (Britton et al., 2007); treePL (Smith & O'Meara, 2012), a semi-parametric, rate-smoothing, penalized likelihood dating method (Sanderson, 2002); and MRBayes (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003), a Bayesian inference program implementing Markov chain Monte Carlo (MCMC) methods to estimate a posterior distribution of model parameters.

Visualizing results

Finally, users can save all source and summary chronograms in formats allowing for reuse and reanalysis, such as newick and the R "phylo" format. Input and summary chronograms can be visualized and compared graphically, and users can construct their own graphs using DateLife's chronogram plot generation functions available from the R package `datelifeplot` (Sanchez-Reyes & O'Meara, 2022).

Benchmark

`datelife`'s R package code speed was tested on an Apple iMac with one 3.4 GHz Intel Core i5 processor. We registered variation in computing time of query processing and search

249 through the database relative to number of queried taxon names. Query processing time
250 increases roughly linearly with number of input taxon names, and increases considerably if
251 Taxonomic Name Resolution Service (TNRS) is activated. Up to ten thousand names can be
252 processed and searched in less than 30 minutes with the most time consuming settings. Once
253 names have been processed as described in methods, a name search through the chronogram
254 database can be performed in less than a minute, even with a very large number of taxon
255 names (Fig. 2).

256 `datelife`'s code performance was evaluated with a set of unit tests designed and
257 implemented with the R package `testthat` (R Core Team, 2018) that were run both locally
258 with the `devtools` package (R Core Team, 2018), and on a public server using the continuous
259 integration tool of GitHub actions (<https://docs.github.com/en/actions>). At present, unit
260 tests cover more than 40% of `datelife`'s code (<https://codecov.io/gh/phylotastic/datelife>).
261 Unit testing helps identify potential issues as code is updated or, more critically, as services
262 code relies upon may change.

263 Case studies

264 We illustrate the DateLife workflow using a family within the Passeriform birds
265 encompassing the true finches, Fringillidae, as case study. On a small example, we analysed 6
266 bird species, and results from each step of the workflow are shown in Fig. 3. As a second
267 example, we analysed 289 bird species in the family Fringillidae that are included in the
268 NCBI taxonomy. The resulting summary chronogram is shown in Fig. 5, and results from
269 previous steps of the workflow are available as Supplementary Figures.

270 A small example

271 **Creating a search query.** We chose 6 bird species within the Passeriformes. The
272 sample includes two species of cardinals: the black-thighed grosbeak – *Pheucticus tibialis*
273 and the crimson-collared grosbeak – *Rhodothraupis celaeno*; three species of buntings: the

274 yellowhammer – *Emberiza citrinella*, the pine bunting – *Emberiza leucocephalos* and the
275 yellow-throated bunting – *Emberiza elegans*; and one species of tanager, the vegetarian finch –
276 *Platyspiza crassirostris*. Processing of input names found that *Emberiza elegans* is synonym
277 for *Schoeniclus elegans* in the default reference taxonomy (OTT v3.3, June 1, 2021). For a
278 detailed discussion on the state of the synonym, refer to Avibase (Avibase, 2022; Lepage,
279 2004; Lepage, Vaidya, & Guralnick, 2014). Discovering this synonym allowed assigning five
280 age data points for the parent node of *Emberiza elegans*, shown as *Schoeniclus elegans* in
281 figure 3A, which would not have had any data otherwise.

282 **Searching the database.** DateLife used the processed input names to search the
283 local chronogram database and found 9 matching chronograms in 6 different studies (Fig.
284 3B). Three studies matched five input names (Barker, Burns, Klicka, Lanyon, & Lovette,
285 2015; Hedges, Marin, Suleski, Paymer, & Kumar, 2015; Jetz et al., 2012), one study matched
286 four input names (Hooper & Price, 2017) and two studies matched two input names (Barker,
287 Burns, Klicka, Lanyon, & Lovette, 2013; Burns et al., 2014). No studies matched all input
288 names. Together, source chronograms provide 28 unique age data points, covering all nodes
289 on our chosen tree topology to date (Table 1).

290 **Summarizing search results.** DateLife obtained OpenTree’s synthetic tree
291 topology for these taxa (Fig. 3C), and congruified and mapped age data to nodes in this
292 chosen topology (Table 1). The name processing step allowed including five data points for
293 node “n4” (parent of *Schoeniclus elegans*; Fig. 3A) that would not have had any data
294 otherwise due to name mismatch. Age summary statistics per node were calculated (Table 2)
295 and used as calibrations to date the tree topology using the BLADJ algorithm. As expected,
296 more inclusive nodes (e.g., node “n1”) have more variance in age data than less inclusive
297 nodes (e.g., node “n5”). Summary age data for node “n2” were excluded as final calibration
298 because they are older than age data of the more inclusive node, “n1” (Fig. 3C4).

299 **An example with the family of true finches**

300 **Creating a query.** To obtain ages for all species within the family of true finches,
301 Fringillidae, we ran a DateLife query using the “get species from taxon” flag, which gets all
302 recognized species names within a named group from a taxonomy of choice. Following the
303 NCBI taxonomy, our DateLife query has 289 Fringillidae species names. This
304 taxon-constrained approach implies that the full DateLife analysis will be performed using a
305 tree topology and ages available for species names from a given taxonomic group, which do
306 not necessarily correspond to a monophyletic group. Users can change this behaviour by
307 providing all species names corresponding to a monophyletic group as input for a DateLife
308 search, or a monophyletic tree to construct a DateLife summary.

309 **Searching the database.** Next, we used the processed species names in our
310 DateLife query to identify chronograms with at least two Fringillidae species as tip taxa.
311 The DateLife search identified 19 chronograms matching this criteria, published in 13
312 different studies (Barker et al., 2013, 2015; Burns et al., 2014; Claramunt & Cracraft, 2015;
313 Gibb et al., 2015; Hedges et al., 2015; Hooper & Price, 2017; Jetz et al., 2012; Kimball et al.,
314 2019; Oliveros et al., 2019; Price et al., 2014; Roquet, Lavergne, & Thuiller, 2014; Uyeda,
315 Pennell, Miller, Maia, & McClain, 2017). Once identified, DateLife pruned these matching
316 chronograms to remove tips that do not belong to the queried taxon names, and transformed
317 these pruned chronograms to pairwise distance matrices, revealing 1, 206 different age data
318 points available for species within the Fringillidae (Supplementray Table S1).

319 **Summarizing search results.** The final step entailed congruifying and
320 summarizing the age data available for the Fringillidae species into two single summary
321 chronograms, using two different types of summary ages, median and SDM. As explained in
322 the “Description” section, a tree topology to summarize age data upon is required. By
323 default, DateLife uses the topology from OpenTree’s synthetic tree that contains all taxa
324 from the search query. According to OpenTree’s synthetic tree, species belonging to the
325 family Fringillidae do not form a monophyletic group (Fig. 4). Hence, a topology containing

326 only the 289 species from the original query was extracted from Open Tree of Life's synthetic
327 tree v12.3 (Open Tree Of Life et al., 2019).

328 Age data from source chronograms was congruified to OpenTree's topology (Fig. 4B),
329 reducing the age data set to 818 different data points (Supplementray Table S2). For each
330 congruent node, age summary statistics were calculated and used as fixed secondary
331 calibrations over the chosen tree topology, to obtain a fully dated phylogeny with the
332 program BLADJ (Fig. 5).

333 **Cross-validation test**

334 We performed a cross validation analysis of the DateLife workflow using the
335 Fringillidae chronograms. We used the individual tree topologies from each of the 19 source
336 chronograms from 13 studies as inputs, treating their node ages as unknown. We then
337 estimated dates for these topologies using the node ages from the chronograms from the
338 other studies as calibrations and smoothing using BLADJ. We found that node ages from
339 original study, and ages estimated using all other age data available are correlated (Fig. 6).
340 For five studies, Datelife tended to underestimate ages for topologically deeper nodes (those
341 with many descendant taxa, aka “closer to the root”) relative to the original estimate, and
342 overestimate ages for nodes closer to the tips. Accordingly, root ages are generally older in
343 the original study than estimated using cross-validated ages (Supplementary Fig. S1).

344 **Discussion**

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

350 A total of 99,474 unique terminal taxa are represented in DateLife's database.

351 Incorporation of more chronograms into the database will continue to improve DateLife's
352 services. One option to increase the number of chronograms in the DateLife database is the
353 Dryad data repository. Methods to automatically mine chronograms from Dryad could be
354 designed and implemented. However, Dryad's metadata system has no information to
355 automatically detect branch length units, and those would still need to be determined
356 manually by a human curator. We would like to emphasize on the importance of sharing
357 chronogram data, including systematically curated metadata, into open repositories, such as
358 OpenTree's Phylesystem (McTavish et al., 2015) for the benefit of the scientific community
359 as a whole.

360 While it is known that incorporating at least some data on lineage divergence times
361 represents a relevant improvement for testing alternative hypothesis using phylogenetic
362 distance in ecological and conservation biology studies (Webb et al., 2008). Next, we discuss
363 some particularities of using summarized node age data from a DateLife analysis that users
364 should be aware of.

365 **Age variation in source chronograms**

366 Conflict in estimated ages among alternative studies is common in the literature. See,
367 for example, the robust ongoing debate about crown group age of angiosperms
368 (Barba-Montoya, Reis, Schneider, Donoghue, & Yang, 2018; Magallón, Gómez-Acevedo,
369 Sánchez-Reyes, & Hernández-Hernández, 2015; Ramshaw et al., 1972; Sanderson & Doyle,
370 2001; Sauquet, Ramírez-Barahona, & Magallón, 2021). Source chronograms available for the
371 same organisms have potentially been estimated implementing calibrations very differently.
372 For example, the chronograms from Burns et al. (2014) were inferred using molecular
373 substitution rate estimates across birds (Weir & Schlüter, 2008), and have much older age
374 estimates for the same nodes than chronograms that were inferred using fossils as
375 calibrations (Figs. 5, 6; Supplementary Figs. S1, S5).

Different calibration implementations might also imply fundamentally distinct evolutionary hypotheses (Antonelli et al., 2017). For example, two independent researchers working on the same clade should both carefully select and justify their choices of fossil calibration placement. Yet, if one researcher concludes that a fossil should calibrate the ingroup of a clade, while another researcher concludes that the same fossil should calibrate the outgroup of the clade, the resulting age estimates will differ, as the placement of calibrations as stem or crown group has been proven to significantly affect time of lineage divergence estimates (Sauquet, 2013).

Primary vs Secondary calibrations

While most chronograms in DateLife's database are constructed using primary calibrations (molecular substitution rates or ages obtained from the fossil record or geological events), DateLife summarizes chronograms using secondary calibrations (ages coming from other chronograms). Graur and Martin (2004) cautioned on the increased error and uncertainty in estimated ages when using secondary calibrations in dating analyses. Schenk (2016) showed that, in simulations, divergence times inferred using secondary calibrations are significantly younger than those inferred with primary calibrations, when obtained with Bayesian inference methods, and when priors are implemented in similar ways in both analyses. Accordingly, the scientific community seems to have more confidence in chronograms obtained from a single analysis, using fossil data as primary sources of calibrations (Schenk, 2016), and using fossils that have been widely discussed and curated as calibrations to date other trees, making sure that all data reflect a coherent evolutionary history (Sauquet, 2013), as for example done by Antonelli et al. (2017). There have been attempts to create fossil calibration databases (Ksepka et al., 2015), though these still have room to grow.

It seems that using several (as opposed to just a few) secondary calibrations can provide sufficient information to alleviate or even neutralize potential biases (Sauquet, 2013).

402 Certainly, further studies are required to fully understand the effect of secondary calibrations
403 on outputs from different tree dating methods, and on downstream analyses. It is possible
404 that secondary calibrations can be safely used with dating methods that do not require
405 setting priors, such as penalized likelihood (Sanderson, 2003), with methods that do not
406 make any assumptions on the ages and fix them to a node on a tree topology, such as
407 BLADJ (Webb et al., 2008; Webb & Donoghue, 2005), or methods that summarize age data
408 unto a tree topology.

409 Our cross validation analysis might provide some insight in this regard. When ages are
410 estimated with secondary calibrations, nodes closer to the root do tend to be slightly
411 younger than ages estimated with primary calibrations. However, nodes closer to the tip
412 tend to be older when estimated using secondary calibrations with a dating method that
413 does not make any prior assumptions on the nature of the calibrations themselves
414 (Supplementary Figures S2-S20). The only exception to this was observed on cross validation
415 results of the Burns et al. (2014) chronogram, which displays much younger node ages when
416 estimated using secondary calibrations (Supplementary Figs. S1, S5).

417 **Summarizing chronograms**

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

425 Some measures that have been proposed are the proportion of lineage sampling and the
426 number of calibrations used (Magallón, 2010; Magallón et al., 2015). Some characteristics

427 that are often cited in published studies as a measure of improved age estimates as compared
428 to previously published estimates are: quality of alignment (missing data, GC content),
429 lineage sampling (strategy and proportion), phylogenetic and dating inference method,
430 number of fossils used as calibrations, support for nodes and ages, and magnitude of
431 confidence intervals.

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

437 The exercise of summarizing age data from across multiple studies provides the
438 opportunity to work with a chronogram that reflects a unified evolutionary history for a
439 lineage, by putting together evidence from different hypotheses. The largest, and
440 taxonomically broadest chronogram currently available from OpenTree was constructed
441 summarizing age data from 2,274 published chronograms using NCBI's taxonomic tree as
442 backbone (Hedges et al., 2015). A summarizing exercise may also amplify the effect of
443 uncertainty and errors in source data, and blur parts of the evolutionary history of a lineage
444 that might only be reflected in source chronograms and lost on the summary chronogram
445 (Sauquet et al., 2021).

446 Effects of phylogenetic sampling on downstream analyses

447 Analysis of species diversification using simulated and empirical phylogenies suggest
448 that using a more completely sampled phylogeny provides estimates that are closer to the
449 true diversification history than when analysing incompletely sampled phylogenies (Chang,
450 Rabosky, & Alfaro, 2020; Cusimano, Stadler, & Renner, 2012; Sun et al., 2020). Ideally
451 phylogenies should be completed using genetic data, but this is a very hard task to achieve

452 for many biological groups. Hence, DateLife's workflow features different ways of assigning
453 node ages in the absence of calibrations and branch length information for certain taxa.
454 Completing a phylogeny using a stochastic birth-death politomy resolver and a backbone
455 taxonomy is a common practice in scientific publications: Jetz et al. (2012), created a
456 chronogram of all 9, 993 bird species, where 67% had molecular data and the rest was
457 simulated; Rabosky et al. (2018) created a chronogram of 31, 536 ray-finned fishes, of which
458 only 37% had molecular data; Smith and Brown (2018) constructed a chronogram of 353, 185
459 seed plants where only 23% had molecular data. These stochastically resolved chronograms
460 provide less biased diversification rates estimated with methods that account for sampling
461 fractions (Chang et al., 2020; Cusimano et al., 2012), but can also introduce spurious
462 patterns of early bursts of diversification (Cusimano & Renner, 2010; Sun et al., 2020).

463 Taxonomy based politomy resolvers also introduce topological differences. The study of
464 macroevolutionary processes largely depends on an understanding of the timing of species
465 diversification events, and different phylogenetic and chronogram hypothesis can provide
466 very different overviews of the macroevolutionary history of a biological group. For example,
467 alternative topologies in chronograms from the same biological group can infer very different
468 species diversification patterns (Rabosky, 2015; Title & Rabosky, 2016). Similarly there are
469 worries that patterns of morphological evolution cannot be accurately inferred with these
470 types of phylogenies, as any patterns would be erased by randomization (Rabosky, 2015).
471 We note that the same applies for geographical and morphological dependent diversification
472 analysis. Hence, we suggest that these types of randomly resolved phylogenies can be used as
473 null or neutral models, representing the case of diversification independent of traits and
474 geographical scenario.

475 While stochastic politomy resolvers have been useful to advance research, notably, risks
476 come with this practice. Taken to the extreme, one could generate a fully resolved,
477 calibrated tree of all modern and extinct taxa using a single taxonomy, a single calibration,

478 and assigning branch lengths following a birth-death diversification model. Clearly, this can
479 lead to a misrepresentation of the true evolutionary history. We urge DateLife users to follow
480 the example of the large tree papers cited above, by carefully considering the statistical
481 assumptions being made, potential biases, and assessing the consistency of the results with
482 prior work.

483 **Conclusions**

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

492 The DateLife project allows for easy and fast summary of public and state-of-the-art
493 data on time of lineage divergence. It provides a straightforward way to get an informed idea
494 on the state of knowledge of the time frame of evolution of different regions of the tree of life,
495 and allows identifying regions that require more research, or that have conflicting
496 information. It is available as an R package, and as a web-based R shiny application at
497 www.datelife.org Both summary and newly generated trees are useful to evaluate
498 evolutionary hypotheses in different areas of research. The DateLife project should improve
499 awareness of the existing variation in expert time of divergence data, and foster exploration
500 of the effect of alternative divergence time hypothesis on the results of analyses, nurturing a
501 culture of more cautious interpretation of evolutionary results.

Availability

The DateLife software is free and open source. It can be used online through its R shiny web application hosted at <http://www.datelife.org>, and locally through the `datelife` R package, downloadable from Zenodo (<https://doi.org/10.5281/zenodo.593938>) and the CRAN repository (Sanchez-Reyes et al., 2022). DateLife’s web application is maintained using RStudio’s shiny server and the shiny package open infrastructure, as well as Docker and OpenTree’s infrastructure (dates.opentreeoflife.org/datelife). `datelife`’s R package stable version is available for installation from CRAN repository using the command `install.packages(pkgs = "datelife")` from within R. Development versions are available from DateLife’s GitHub repository (<https://github.com/phylotastic/datelife>) and can be installed using the command `devtools::install_github("phylotastic/datelife")`.

Supplementary Material

Supplementary material, including code, biological examples, benchmark results, data files and online-only appendices, can be found in the Dryad data repository (<https://doi.org/10.5061/dryad.cnp5hqc6w>), as well as in the Zenodo stable repositories that host the reproducible manuscript (<https://doi.org/10.5281/zenodo.7435094>), the biological examples (<https://doi.org/10.5281/zenodo.7435101>), and the software benchmark (<https://doi.org/10.5281/zenodo.7435106>). The corresponding development versions hosted on GitHub can be found at <https://github.com/LunaSare/datelifeMS1#readme>, https://github.com/LunaSare/datelife_examples#readme, and https://github.com/LunaSare/datelife_benchmark#readme.

Funding

Funding was provided by the US National Science Foundation (NSF) grants ABI-1458603 to the Datelife project; DBI-0905606 to the National Evolutionary Synthesis

527 Center (NESCent); ABI-1458572 to the Phylotastic project; and ABI-1759846 to the Open
528 Tree of Life project.

529

Acknowledgements

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

542

References

- 543 Agnarsson, I., & Miller, J. A. (2008). Is acctran better than deltran? *Cladistics*, 24(6),
544 1032–1038.
- 545 Alström, P., Hooper, D. M., Liu, Y., Olsson, U., Mohan, D., Gelang, M., … Price, T. D.
546 (2014). Discovery of a relict lineage and monotypic family of passerine birds. *Biology
547 Letters*, 10(3), 20131067.
- 548 Ané, C., Eulenstein, O., Piaggio-Talice, R., & Sanderson, M. J. (2009). Groves of
549 phylogenetic trees. *Annals of Combinatorics*, 13(2), 139–167.
- 550 Antonelli, A., Hettling, H., Condamine, F. L., Vos, K., Nilsson, R. H., Sanderson, M. J., …
551 Vos, R. A. (2017). Toward a self-updating platform for estimating rates of speciation
552 and migration, ages, and relationships of Taxa. *Systematic Biology*, 66(2), 153–166.
553 <https://doi.org/10.1093/sysbio/syw066>
- 554 Archie, J., Day, W. H., Felsenstein, J., Maddison, W., Meacham, C., Rohlf, F. J., &
555 Swofford, D. (1986). The Newick tree format. Retrieved from
556 {<https://evolution.genetics.washington.edu/phylip/newicktree.html>}
- 557 Avibase. (2022). Yellow-throated Bunting. *Avibase - the World Bird Database*, (Online
558 Resource). Retrieved from
559 {<https://avibase.bsc-eoc.org/species.jsp?lang=EN&avibaseid=82D1EE0049D8D927>}
- 560 Bapst, D. W. (2012). Paleotree: An R package for paleontological and phylogenetic analyses
561 of evolution. *Methods in Ecology and Evolution*, 3(5), 803–807.
562 <https://doi.org/10.1111/j.2041-210X.2012.00223.x>
- 563 Barba-Montoya, J., Reis, M. dos, Schneider, H., Donoghue, P. C., & Yang, Z. (2018).
564 Constraining uncertainty in the timescale of angiosperm evolution and the veracity of

- 565 a cretaceous terrestrial revolution. *New Phytologist*, 218(2), 819–834.
- 566 Barker, F. K. (2014). Mitogenomic data resolve basal relationships among passeriform and
567 passeridan birds. *Molecular Phylogenetics and Evolution*, 79, 313–324.
- 568 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2013). Going to
569 extremes: Contrasting rates of diversification in a recent radiation of new world
570 passerine birds. *Systematic Biology*, 62(2), 298–320.
- 571 Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2015). New insights
572 into new world biogeography: An integrated view from the phylogeny of blackbirds,
573 cardinals, sparrows, tanagers, warblers, and allies. *The Auk: Ornithological Advances*,
574 132(2), 333–348.
- 575 Barker, F. K., Cibois, A., Schikler, P., Feinstein, J., & Cracraft, J. (2004). Phylogeny and
576 diversification of the largest avian radiation. *Proceedings of the National Academy of
577 Sciences*, 101(30), 11040–11045.
- 578 Beresford, P., Barker, F., Ryan, P., & Crowe, T. (2005). African endemics span the tree of
579 songbirds (passeri): Molecular systematics of several evolutionary “enigmas”.
580 *Proceedings of the Royal Society B: Biological Sciences*, 272(1565), 849–858.
- 581 Bodenhofer, U., Bonatesta, E., Horejš-Kainrath, C., & Hochreiter, S. (2015). Msa: An r
582 package for multiple sequence alignment. *Bioinformatics*, 31(24), 3997–3999.
- 583 Boyle, B., Hopkins, N., Lu, Z., Raygoza Garay, J. A., Mozzherin, D., Rees, T., ... Enquist,
584 B. J. (2013). The taxonomic name resolution service: An online tool for automated
585 standardization of plant names. *BMC Bioinformatics*, 14(1).
586 <https://doi.org/10.1186/1471-2105-14-16>
- 587 Britton, T., Anderson, C. L., Jacquet, D., Lundqvist, S., & Bremer, K. (2007). Estimating

- 588 Divergence Times in Large Phylogenetic Trees. *Systematic Biology*, 56(788777878),
589 741–752. <https://doi.org/10.1080/10635150701613783>
- 590 Bryson Jr, R. W., Chaves, J., Smith, B. T., Miller, M. J., Winker, K., Pérez-Emán, J. L., &
591 Klicka, J. (2014). Diversification across the new world within the ‘blue’cardinalids
592 (aves: Cardinalidae). *Journal of Biogeography*, 41(3), 587–599.
- 593 Burleigh, J. G., Kimball, R. T., & Braun, E. L. (2015). Building the avian tree of life using a
594 large-scale, sparse supermatrix. *Molecular Phylogenetics and Evolution*, 84, 53–63.
- 595 Burns, K. J., Shultz, A. J., Title, P. O., Mason, N. A., Barker, F. K., Klicka, J., ... Lovette,
596 I. J. (2014). Phylogenetics and diversification of tanagers (passeriformes:
597 Thraupidae), the largest radiation of neotropical songbirds. *Molecular Phylogenetics
598 and Evolution*, 75, 41–77.
- 599 Chamberlain, S. (2018). *bold: Interface to Bold Systems API*. Retrieved from
600 <https://CRAN.R-project.org/package=bold>
- 601 Chamberlain, S. A., & Szöcs, E. (2013). taxize : taxonomic search and retrieval in R [version
602 2; referees: 3 approved]. *F1000Research*, 2(191), 1–29.
603 <https://doi.org/10.12688/f1000research.2-191.v2>
- 604 Chang, J., Rabosky, D. L., & Alfaro, M. E. (2020). Estimating diversification rates on
605 incompletely sampled phylogenies: Theoretical concerns and practical solutions.
606 *Systematic Biology*, 69(3), 602–611.
- 607 Chaves, J. A., Hidalgo, J. R., & Klicka, J. (2013). Biogeography and evolutionary history of
608 the neotropical genus Saltator (aves: Thraupini). *Journal of Biogeography*, 40(11),
609 2180–2190.
- 610 Claramunt, S., & Cracraft, J. (2015). A new time tree reveals earth history’s imprint on the

- 611 evolution of modern birds. *Science Advances*, 1(11), e1501005.
- 612 Criscuolo, A., Berry, V., Douzery, E. J., & Gascuel, O. (2006). SDM: A fast distance-based
613 approach for (super)tree building in phylogenomics. *Systematic Biology*, 55(5),
614 740–755. <https://doi.org/10.1080/10635150600969872>
- 615 Cusimano, N., & Renner, S. S. (2010). Slowdowns in diversification rates from real
616 phylogenies may not be real. *Systematic Biology*, 59(4), 458–464.
- 617 Cusimano, N., Stadler, T., & Renner, S. S. (2012). A new method for handling missing
618 species in diversification analysis applicable to randomly or nonrandomly sampled
619 phylogenies. *Systematic Biology*, 61(5), 785–792.
- 620 Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M.-K., Turon, X., ... Douzery,
621 E. J. (2018). A phylogenomic framework and timescale for comparative studies of
622 tunicates. *BMC Biology*, 16(1), 1–14.
- 623 Eastman, J. M., Harmon, L. J., & Tank, D. C. (2013). Congruification: Support for time
624 scaling large phylogenetic trees. *Methods in Ecology and Evolution*, 4(7), 688–691.
625 <https://doi.org/10.1111/2041-210X.12051>
- 626 Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high
627 throughput. *Nucleic Acids Research*, 32(5), 1792–1797.
- 628 Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American Naturalist*,
629 125(1), 1–15. Retrieved from <http://www.jstor.org/stable/2461605>
- 630 Forest, F., Savolainen, V., Chase, M. W., Lupia, R., Bruneau, A., & Crane, P. R. (2005).
631 Teasing apart molecular-versus fossil-based error estimates when dating phylogenetic
632 trees: A case study in the birch family (betulaceae). *Systematic Botany*, 30(1),
633 118–133.

- 634 Freckleton, R. P., Harvey, P. H., & Pagel, M. (2002). Phylogenetic analysis and comparative
635 data: A test and review of evidence. *The American Naturalist*.
- 636 GBIF Secretariat. (2022). GBIF Backbone Taxonomy. *Checklist dataset*, (Online Resource
637 accessed via GBIF.org). Retrieved from {<https://doi.org/10.15468/39omei> }
- 638 Gibb, G. C., England, R., Hartig, G., McLenaghan, P. A., Taylor Smith, B. L., McComish,
639 B. J., ... Penny, D. (2015). New zealand passerines help clarify the diversification of
640 major songbird lineages during the oligocene. *Genome Biology and Evolution*, 7(11),
641 2983–2995.
- 642 Graur, D., & Martin, W. (2004). Reading the entrails of chickens: Molecular timescales of
643 evolution and the illusion of precision. *TRENDS in Genetics*, 20(2), 80–86.
- 644 Hackett, S. J., Kimball, R. T., Reddy, S., Bowie, R. C., Braun, E. L., Braun, M. J., ...
645 others. (2008). A phylogenomic study of birds reveals their evolutionary history.
646 *Science*, 320(5884), 1763–1768.
- 647 Harvey, P. H., Pagel, M. D., & others. (1991). *The comparative method in evolutionary
648 biology* (Vol. 239). Oxford university press Oxford.
- 649 Hedges, S. B., Marin, J., Suleski, M., Paymer, M., & Kumar, S. (2015). Tree of life reveals
650 clock-like speciation and diversification. *Molecular Biology and Evolution*, 32(4),
651 835–845. <https://doi.org/10.1093/molbev/msv037>
- 652 Heibl, C. (2008). *PHYLOCH: R language tree plotting tools and interfaces to diverse
653 phylogenetic software packages*. Retrieved from
654 <http://www.christophheibl.de/Rpackages.html>
- 655 Hooper, D. M., & Price, T. D. (2017). Chromosomal inversion differences correlate with
656 range overlap in passerine birds. *Nature Ecology & Evolution*, 1(10), 1526.

- 657 Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic
658 trees. *Bioinformatics*, 17(8), 754–755.
659 <https://doi.org/10.1093/bioinformatics/17.8.754>
- 660 Jetz, W., Thomas, G., Joy, J. J., Hartmann, K., & Mooers, A. (2012). The global diversity
661 of birds in space and time. *Nature*, 491(7424), 444–448.
662 <https://doi.org/10.1038/nature11631>
- 663 Johansson, U. S., Fjeldså, J., & Bowie, R. C. (2008). Phylogenetic relationships within
664 passerida (aves: Passeriformes): A review and a new molecular phylogeny based on
665 three nuclear intron markers. *Molecular Phylogenetics and Evolution*, 48(3), 858–876.
- 666 Katoh, K., Asimenos, G., & Toh, H. (2009). Multiple alignment of dna sequences with mafft.
667 In *Bioinformatics for dna sequence analysis* (pp. 39–64). Springer.
- 668 Kimball, R. T., Oliveros, C. H., Wang, N., White, N. D., Barker, F. K., Field, D. J., ...
669 others. (2019). A phylogenomic supertree of birds. *Diversity*, 11(7), 109.
- 670 Klicka, J., Barker, F. K., Burns, K. J., Lanyon, S. M., Lovette, I. J., Chaves, J. A., & Bryson
671 Jr, R. W. (2014). A comprehensive multilocus assessment of sparrow (aves:
672 Passerellidae) relationships. *Molecular Phylogenetics and Evolution*, 77, 177–182.
- 673 Ksepka, D. T., Parham, J. F., Allman, J. F., Benton, M. J., Carrano, M. T., Cranston, K.
674 A., ... others. (2015). The fossil calibration database—a new resource for divergence
675 dating. *Systematic Biology*, 64(5), 853–859.
- 676 Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio,
677 A., ... others. (2015). Evolution of darwin's finches and their beaks revealed by
678 genome sequencing. *Nature*, 518(7539), 371–375.
- 679 Laubichler, M. D., & Maienschein, J. (2009). *Form and function in developmental evolution*.

- 680 Cambridge University Press.
- 681 Lepage, D. (2004). *Avibase: The world bird database*. Bird Studies Canada.
- 682 Lepage, D., Vaidya, G., & Guralnick, R. (2014). Avibase—a database system for managing
683 and organizing taxonomic concepts. *ZooKeys*, (420), 117.
- 684 Lerner, H. R., Meyer, M., James, H. F., Hofreiter, M., & Fleischer, R. C. (2011). Multilocus
685 resolution of phylogeny and timescale in the extant adaptive radiation of hawaiian
686 honeycreepers. *Current Biology*, 21(21), 1838–1844.
- 687 Lovette, I. J., Pérez-Emán, J. L., Sullivan, J. P., Banks, R. C., Fiorentino, I.,
688 Córdoba-Córdoba, S., ... others. (2010). A comprehensive multilocus phylogeny for
689 the wood-warblers and a revised classification of the parulidae (aves). *Molecular
690 Phylogenetics and Evolution*, 57(2), 753–770.
- 691 Magallon, S., & Sanderson, M. (2001). Absolute diversification rates in angiosperm clades.
692 *Evolution*, 55(9), 1762–1780.
- 693 Magallón, S. (2010). Using fossils to break long branches in molecular dating: A comparison
694 of relaxed clocks applied to the origin of angiosperms. *Systematic Biology*, 59(4),
695 384–399.
- 696 Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., & Hernández-Hernández, T. (2015).
697 A metacalibrated time-tree documents the early rise of flowering plant phylogenetic
698 diversity. *New Phytologist*, 207(2), 437–453.
- 699 McTavish, E. J., Hinchliff, C. E., Allman, J. F., Brown, J. W., Cranston, K. A., Holder, M.
700 T., ... Smith, S. (2015). Phylesystem: A git-based data store for community-curated
701 phylogenetic estimates. *Bioinformatics*, 31(17), 2794–2800.
- 702 Michonneau, F., Brown, J. W., & Winter, D. J. (2016). rotl: an R package to interact with

- 703 the Open Tree of Life data. *Methods in Ecology and Evolution*, 7(12), 1476–1481.
- 704 <https://doi.org/10.1111/2041-210X.12593>
- 705 Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecology Letters*,
706 17(4), 508–525. <https://doi.org/10.1111/ele.12251>
- 707 Moyle, R. G., Oliveros, C. H., Andersen, M. J., Hosner, P. A., Benz, B. W., Manthey, J. D.,
708 ... Faircloth, B. C. (2016). Tectonic collision and uplift of Wallacea triggered the
709 global songbird radiation. *Nature Communications*, 7(1), 1–7.
- 710 Oliveros, C. H., Field, D. J., Ksepka, D. T., Barker, F. K., Aleixo, A., Andersen, M. J., ...
711 others. (2019). Earth history and the passerine superradiation. *Proceedings of the
712 National Academy of Sciences*, 116(16), 7916–7925.
- 713 Ooms, J., & Chamberlain, S. (2018). *Phylocomr: Interface to 'phylocom'*. Retrieved from
714 <https://CRAN.R-project.org/package=phylocomr>
- 715 Open Tree Of Life, Redelings, B., Cranston, K. A., Allman, J., Holder, M. T., & McTavish,
716 E. J. (2016). Open Tree of Life APIs v3.0. *Open Tree of Life Project*, (Online
717 Resources). Retrieved from
718 {<https://github.com/OpenTreeOfLife/germinator/wiki/Open-Tree-of-Life-Web->
719 APIs}
- 720 Open Tree Of Life, Redelings, B., Sánchez Reyes, L. L., Cranston, K. A., Allman, J., Holder,
721 M. T., & McTavish, E. J. (2019). Open tree of life synthetic tree v12.3. *Zenodo*.
722 Retrieved from <https://doi.org/10.5281/zenodo.3937742>
- 723 Ödeen, A., Håstad, O., & Alström, P. (2011). Evolution of ultraviolet vision in the largest
724 avian radiation—the passerines. *BMC Evolutionary Biology*, 11(1), 1–8.
- 725 Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and

- 726 evolution in R language. *Bioinformatics*, 20(2), 289–290.
- 727 Parchman, T. L., Benkman, C. W., & Mezquida, E. T. (2007). Coevolution between
728 hispaniolan crossbills and pine: Does more time allow for greater phenotypic
729 escalation at lower latitude? *Evolution*, 61(9), 2142–2153.
- 730 Päckert, M., Martens, J., Sun, Y.-H., Severinghaus, L. L., Nazarenko, A. A., Ting, J., ...
731 Tietze, D. T. (2012). Horizontal and elevational phylogeographic patterns of
732 himalayan and southeast asian forest passerines (aves: Passeriformes). *Journal of
733 Biogeography*, 39(3), 556–573.
- 734 Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., FitzJohn, R. G.,
735 ... Harmon, L. J. (2014). Geiger v2. 0: An expanded suite of methods for fitting
736 macroevolutionary models to phylogenetic trees. *Bioinformatics*, 30(15), 2216–2218.
- 737 Posadas, P., Crisci, J. V., & Katinas, L. (2006). Historical biogeography: A review of its
738 basic concepts and critical issues. *Journal of Arid Environments*, 66(3), 389–403.
- 739 Powell, A. F., Barker, F. K., Lanyon, S. M., Burns, K. J., Klicka, J., & Lovette, I. J. (2014).
740 A comprehensive species-level molecular phylogeny of the new world blackbirds
741 (icteridae). *Molecular Phylogenetics and Evolution*, 71, 94–112.
- 742 Price, T. D., Hooper, D. M., Buchanan, C. D., Johansson, U. S., Tietze, D. T., Alström, P.,
743 ... others. (2014). Niche filling slows the diversification of himalayan songbirds.
744 *Nature*, 509(7499), 222.
- 745 Pulgarín-R, P. C., Smith, B. T., Bryson Jr, R. W., Spellman, G. M., & Klicka, J. (2013).
746 Multilocus phylogeny and biogeography of the new world pheucticus grosbeaks (aves:
747 Cardinalidae). *Molecular Phylogenetics and Evolution*, 69(3), 1222–1227.
- 748 Rabosky, D. L. (2015). No substitute for real data: A cautionary note on the use of

- 749 phylogenies from birth–death polytomy resolvers for downstream comparative
750 analyses. *Evolution*, 69(12), 3207–3216.
- 751 Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., ... others.
752 (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*,
753 559(7714), 392.
- 754 Ramshaw, J., Richardson, D., Meatyard, B., Brown, R., Richardson, M., Thompson, E., &
755 Boulter, D. (1972). The time of origin of the flowering plants determined by using
756 amino acid sequence data of cytochrome c. *New Phytologist*, 71(5), 773–779.
- 757 Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The barcode of life data system
758 (<http://www.Barcodinglife.Org>). *Molecular Ecology Notes*, 7(3), 355–364.
- 759 R Core Team. (2018). *R: a language and environment for statistical computing*. Vienna,
760 Austria: R Foundation for Statistical Computing.
- 761 Rees, & Cranston, K. (2017). Automated assembly of a reference taxonomy for phylogenetic
762 data synthesis. *Biodiversity Data Journal*, (5).
- 763 Rees, Vandepitte, L., Decock, W., & Vanhoorne, B. (2017). IRMNG 2006–2016: 10 Years of
764 a Global Taxonomic Database. *Biodiversity Informatics*, 12.
- 765 Revell, L. J. (2012). Phytools: An r package for phylogenetic comparative biology (and other
766 things). *Methods in Ecology and Evolution*, 3, 217–223.
- 767 Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference
768 under mixed models. *Bioinformatics*, 19(12), 1572–1574.
769 <https://doi.org/10.1093/bioinformatics/btg180>
- 770 Roquet, C., Lavergne, S., & Thuiller, W. (2014). One tree to link them all: A phylogenetic
771 dataset for the european tetrapoda. *PLoS Currents*, 6.

- 772 Sanchez-Reyes, L. L., & O'Meara, B. (2022). *datelifeplot*: Methods to plot chronograms
773 and outputs of the *datelife* package. *R Package Release V0.2.2*. Retrieved from
774 <https://zenodo.org/badge/latestdoi/381501451>
- 775 Sanchez-Reyes, L. L., O'Meara, B., Eastman, J., Heath, T., Wright, A., Schliep, K., ...
776 Alfaro, M. (2022). *datelife*: Scientific Data on Time of Lineage Divergence for Your
777 Taxa. In *R package version 0.6.6*. Retrieved from
778 <https://CRAN.R-project.org/package=datelife> and
779 <https://doi.org/10.5281/zenodo.593938>
- 780 Sanderson, M. (2002). Estimating Absolute Rates of Molecular Evolution and Divergence
781 Times: A Penalized Likelihood Approach. *Molecular Biology and Evolution*, 19(1),
782 101–109. <https://doi.org/10.1093/oxfordjournals.molbev.a003974>
- 783 Sanderson, M. (2003). r8s: Inferring Absolute Rates of Molecular Evolution and Divergence
784 Times in the Absence of a Molecular Clock. *Bioinformatics*, 19(2), 301–302.
- 785 Sanderson, M., & Doyle, J. (2001). Sources of error and confidence intervals in estimating
786 the age of angiosperms from *rbcL* and 18S rDNA data. *American Journal of Botany*,
787 88(8), 1499–1516.
- 788 Sauquet, H. (2013). A practical guide to molecular dating. *Comptes Rendus Palevol*, 12(6),
789 355–367.
- 790 Sauquet, H., Ramírez-Barahona, S., & Magallón, S. (2021). *The age of flowering plants is
791 unknown*.
- 792 Schenk, J. J. (2016). Consequences of secondary calibrations on divergence time estimates.
793 *PLoS ONE*, 11(1). <https://doi.org/10.1371/journal.pone.0148228>
- 794 Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in r. *Bioinformatics*, 27(4), 592–593.

- 795 Schoch, C. L., Ciufo, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R., ...
796 others. (2020). NCBI Taxonomy: a Comprehensive Update on Curation, Resources
797 and Tools. *Database, 2020*.
- 798 Selvatti, A. P., Gonzaga, L. P., & Moraes Russo, C. A. de. (2015). A paleogene origin for
799 crown passerines and the diversification of the oscines in the new world. *Molecular
800 Phylogenetics and Evolution, 88*, 1–15.
- 801 Smith, S., & Brown, J. (2018). Constructing a broadly inclusive seed plant phylogeny.
802 *American Journal of Botany, 105*(3), 302–314.
- 803 Smith, S., & O'Meara, B. (2012). TreePL: Divergence time estimation using penalized
804 likelihood for large phylogenies. *Bioinformatics, 28*(20), 2689–2690.
805 <https://doi.org/10.1093/bioinformatics/bts492>
- 806 Stoltzfus, A., Lapp, H., Matasci, N., Deus, H., Sidlauskas, B., Zmasek, C. M., ... Jordan, G.
807 (2013). Phylotastic! Making tree-of-life knowledge accessible, reusable and convenient.
808 *BMC Bioinformatics, 14*. <https://doi.org/10.1186/1471-2105-14-158>
- 809 Sun, M., Folk, R. A., Gitzendanner, M. A., Soltis, P. S., Chen, Z., Soltis, D. E., & Guralnick,
810 R. P. (2020). Estimating rates and patterns of diversification with incomplete
811 sampling: A case study in the rosids. *American Journal of Botany, 107*(6), 895–909.
- 812 Tietze, D. T., Päckert, M., Martens, J., Lehmann, H., & Sun, Y.-H. (2013). Complete
813 phylogeny and historical biogeography of true rosefinches (aves: Carpodacus).
814 *Zoological Journal of the Linnean Society, 169*(1), 215–234.
- 815 Title, P. O., & Rabosky, D. L. (2016). Do Macrophylogenies Yield Stable Macroevolutionary
816 Inferences? An Example from Squamate Reptiles. *Systematic Biology, syw102*.
817 <https://doi.org/10.1093/sysbio/syw102>

- 818 Treplin, S., Siegert, R., Bleidorn, C., Thompson, H. S., Fotso, R., & Tiedemann, R. (2008).
819 Molecular phylogeny of songbirds (aves: Passeriformes) and the relative utility of
820 common nuclear marker loci. *Cladistics*, 24(3), 328–349.
- 821 Uyeda, J. C., Pennell, M. W., Miller, E. T., Maia, R., & McClain, C. R. (2017). The
822 evolution of energetic scaling across the vertebrate tree of life. *The American
823 Naturalist*, 190(2), 185–199.
- 824 Vos, R. A., Balhoff, J. P., Caravas, J. A., Holder, M. T., Lapp, H., Maddison, W. P., ...
825 others. (2012). NeXML: Rich, extensible, and verifiable representation of
826 comparative data and metadata. *Systematic Biology*, 61(4), 675–689.
827 <https://doi.org/10.1093/sysbio/sys025>
- 828 Webb, C. (2000). Exploring the Phylogenetic Structure of Ecological Communities : An
829 Example for Rain Forest Trees. *The American Naturalist*, 156(2), 145–155.
- 830 Webb, C., Ackerly, D., & Kembel, S. (2008). Phylocom: Software for the analysis of
831 phylogenetic community structure and trait evolution. *Bioinformatics*, 24(18),
832 2098–2100. <https://doi.org/10.1093/bioinformatics/btn358>
- 833 Webb, C., & Donoghue, M. (2005). Phylomatic: Tree assembly for applied phylogenetics.
834 *Molecular Ecology Notes*, 5(1), 181–183.
- 835 Weir, J., & Schluter, D. (2008). Calibrating the avian molecular clock. *Molecular Ecology*,
836 17(10), 2321–2328.
- 837 Zuccon, D., Prŷs-Jones, R., Rasmussen, P. C., & Ericson, P. G. (2012). The phylogenetic
838 relationships and generic limits of finches (fringillidae). *Molecular Phylogenetics and
839 Evolution*, 62(2), 581–596.

840 TABLE 1. Ages of congruified nodes. See Figure 3, step C2.

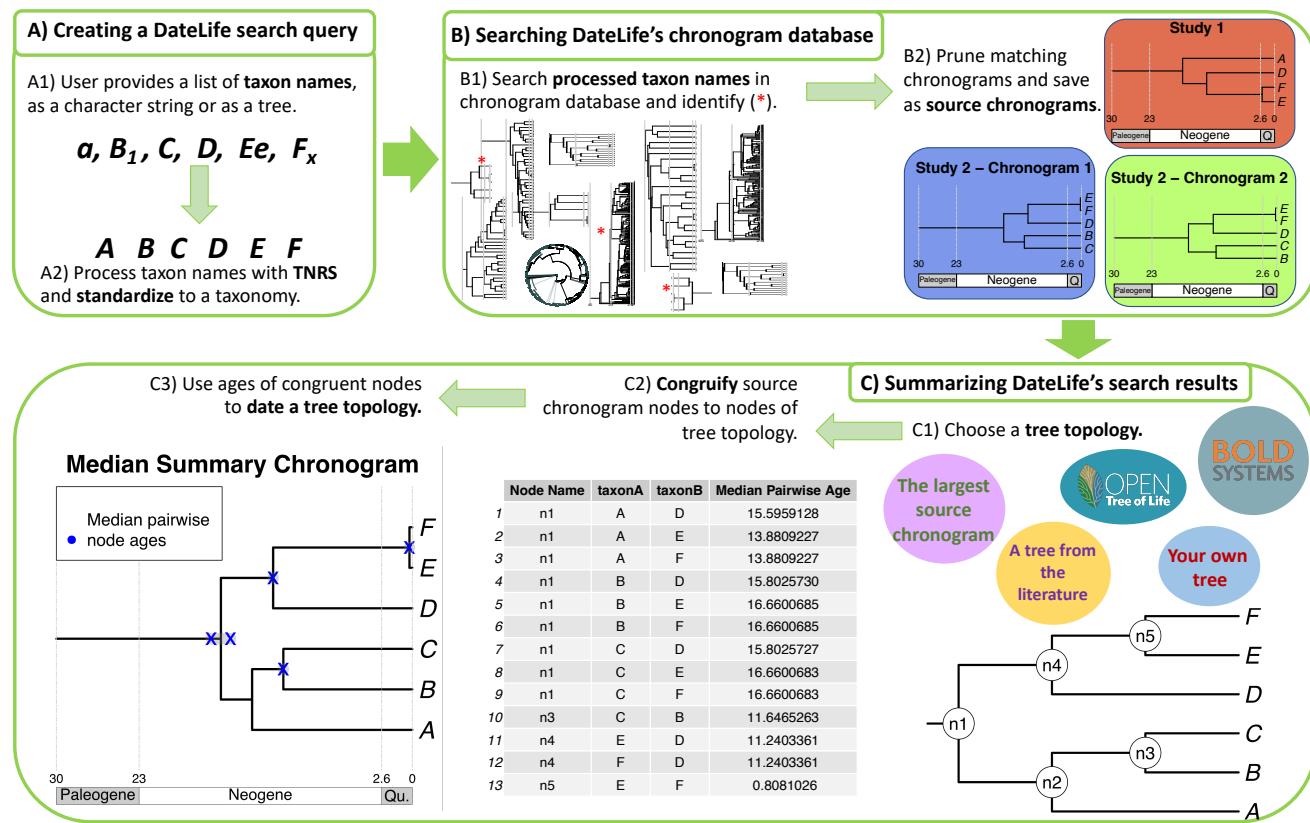
	Node Name	taxon A	taxon B	Node Age	Study chronogram
1	n1	<i>Emberiza citrinella</i>	<i>Platyspiza crassirostris</i>	9.6509881	Jetz et al. 2012 – chronogram 1
2	n1	<i>Pheucticus tibialis</i>	<i>Schoeniclus elegans</i>	14.3336520	Barker et al. 2015 – chronogram 1
3	n1	<i>Rhodothraupis celaeno</i>	<i>Schoeniclus elegans</i>	16.2984859	Barker et al. 2015 – chronogram 2
4	n1	<i>Emberiza citrinella</i>	<i>Platyspiza crassirostris</i>	16.9499615	Hooper et al. 2017
5	n1	<i>Rhodothraupis celaeno</i>	<i>Schoeniclus elegans</i>	21.5140867	Jetz et al. 2012 – chronogram 2
6	n1	<i>Emberiza leucocephalos</i>	<i>Platyspiza crassirostris</i>	22.0000000	Hedges et al. 2015 – chronogram 2
7	n1	<i>Emberiza citrinella</i>	<i>Platyspiza crassirostris</i>	22.3757277	Hedges et al. 2015 – chronogram 1
8	n2	<i>Platyspiza crassirostris</i>	<i>Rhodothraupis celaeno</i>	7.9691925	Jetz et al. 2012 – chronogram 1
9	n2	<i>Platyspiza crassirostris</i>	<i>Rhodothraupis celaeno</i>	19.7085830	Jetz et al. 2012 – chronogram 2
10	n2	<i>Platyspiza crassirostris</i>	<i>Rhodothraupis celaeno</i>	19.7085900	Hedges et al. 2015 – chronogram 2
11	n2	<i>Platyspiza crassirostris</i>	<i>Rhodothraupis celaeno</i>	19.7128363	Hedges et al. 2015 – chronogram 1
12	n2	<i>Pheucticus tibialis</i>	<i>Platyspiza crassirostris</i>	44.2958603	Burns et al. 2014
13	n3	<i>Pheucticus tibialis</i>	<i>Rhodothraupis celaeno</i>	10.5304440	Barker et al. 2015 – chronogram 1
14	n3	<i>Pheucticus tibialis</i>	<i>Rhodothraupis celaeno</i>	10.5379092	Barker et al. 2013
15	n3	<i>Pheucticus tibialis</i>	<i>Rhodothraupis celaeno</i>	11.2095375	Barker et al. 2015 – chronogram 2
16	n3	<i>Pheucticus tibialis</i>	<i>Rhodothraupis celaeno</i>	18.1570685	Hedges et al. 2015 – chronogram 1
17	n4	<i>Emberiza citrinella</i>	<i>Schoeniclus elegans</i>	5.3238969	Jetz et al. 2012 – chronogram 1
18	n4	<i>Emberiza leucocephalos</i>	<i>Schoeniclus elegans</i>	9.8622460	Barker et al. 2015 – chronogram 1
19	n4	<i>Emberiza leucocephalos</i>	<i>Schoeniclus elegans</i>	10.3391445	Jetz et al. 2012 – chronogram 2
20	n4	<i>Emberiza leucocephalos</i>	<i>Schoeniclus elegans</i>	11.7317630	Barker et al. 2015 – chronogram 2
21	n4	<i>Emberiza citrinella</i>	<i>Schoeniclus elegans</i>	12.5133870	Hooper et al. 2017
22	n5	<i>Emberiza citrinella</i>	<i>Emberiza leucocephalos</i>	0.1407015	Jetz et al. 2012 – chronogram 1
23	n5	<i>Emberiza citrinella</i>	<i>Emberiza leucocephalos</i>	0.1516230	Hedges et al. 2015 – chronogram 2
24	n5	<i>Emberiza citrinella</i>	<i>Emberiza leucocephalos</i>	0.2011990	Barker et al. 2015 – chronogram 1
25	n5	<i>Emberiza citrinella</i>	<i>Emberiza leucocephalos</i>	0.2409300	Barker et al. 2015 – chronogram 2
26	n5	<i>Emberiza citrinella</i>	<i>Emberiza leucocephalos</i>	0.2732460	Jetz et al. 2012 – chronogram 2
27	n5	<i>Emberiza citrinella</i>	<i>Emberiza leucocephalos</i>	0.5760260	Hedges et al. 2015 – chronogram 1
28	n5	<i>Emberiza citrinella</i>	<i>Emberiza leucocephalos</i>	2.2898230	Hooper et al. 2017

841

842 TABLE 2. Summary of congruified nodes ages. See Figure 3, step C3.

Node Name	Min Age	Q1	Median Age	Mean Age	Q3	Max Age	Variance	SD
n1	9.6509881	15.316069	16.94996	17.5889860	21.757043	22.375728	22.2431847	4.7162681
n2	7.9691925	19.708583	19.70859	22.2790124	19.712836	44.295860	177.3279940	13.3164558
n3	10.5304440	10.536043	10.87372	12.6087398	12.946420	18.157069	13.7831237	3.7125630
n4	5.3238969	9.862246	10.33914	9.9540875	11.731763	12.513387	7.8263782	2.7975665
n5	0.1407015	0.176411	0.24093	0.5533641	0.424636	2.289823	0.6079318	0.7796998

843



844 FIGURE 1. *Main DateLife workflow. Analyses can be performed via DateLife's interactive*
845 *website at www.datelife.org, or using the `datelife` R package. Details on the R functions used to*
846 *perform the analyses are available from `datelife`'s R package vignettes at*
847 *<https://phylotastic.org/datelife>.*

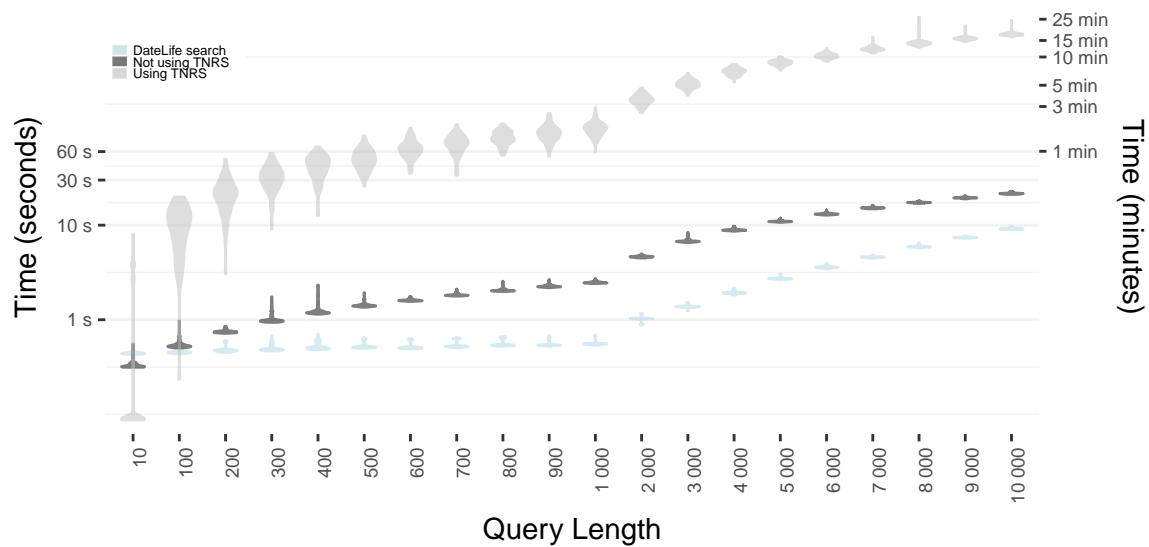
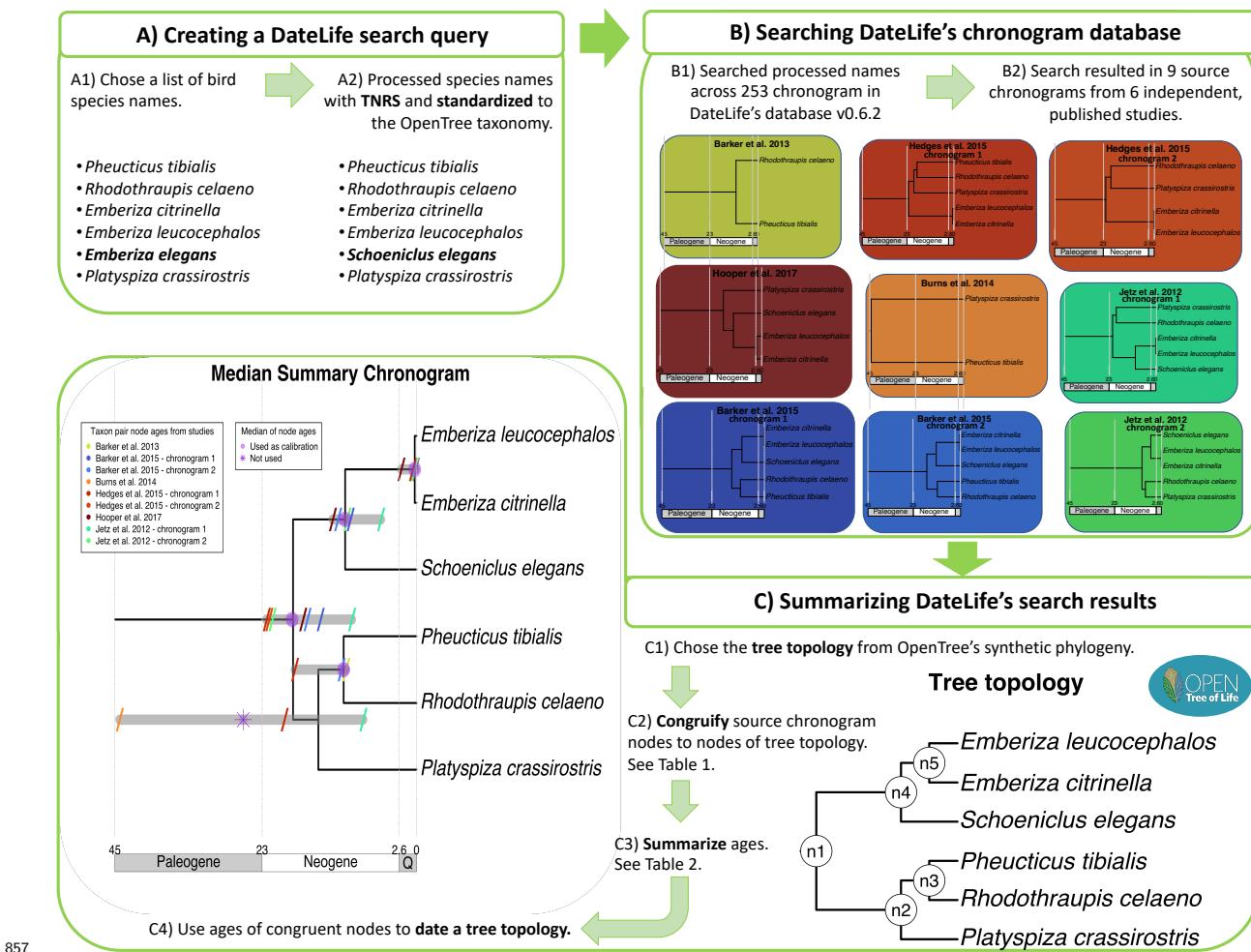


FIGURE 2. *DateLife's* benchmarking results. Computation time used to process a query and a search across *datelife*'s chronogram database, relative to number of input taxon names. For each $N = \{10, 100, 200, \dots, 1\,000, \dots, 9\,000, 10\,000\}$, we sampled N species names from the class Aves a hundred times, and then performed a *datelife* search processing the input names with Taxon Names Resolution Service (TNRS; light gray), and without processing names (dark gray). For comparison, we performed a search using an input that had been pre-processed with TNRS (light blue).



858 FIGURE 3. *DateLife analysis results for a small sample of A) 6 bird species within the*
859 *Passeriformes. B) Processed species names were found across 9 chronograms within 6*
860 *independent studies (Barker et al. (2012), Barker et al. (2015), Burns et al. (2014), Hedges*
861 *et al. (2015), Hooper and Price (2017), Jetz et al. (2012).) C) This revealed 28 source age*
862 *data points for the queried species names. Summarized age data is used as secondary*
863 *calibrations to date a tree topology obtained from OpenTree's synthetic tree, resulting in a*
864 *summary chronogram of source ages.*

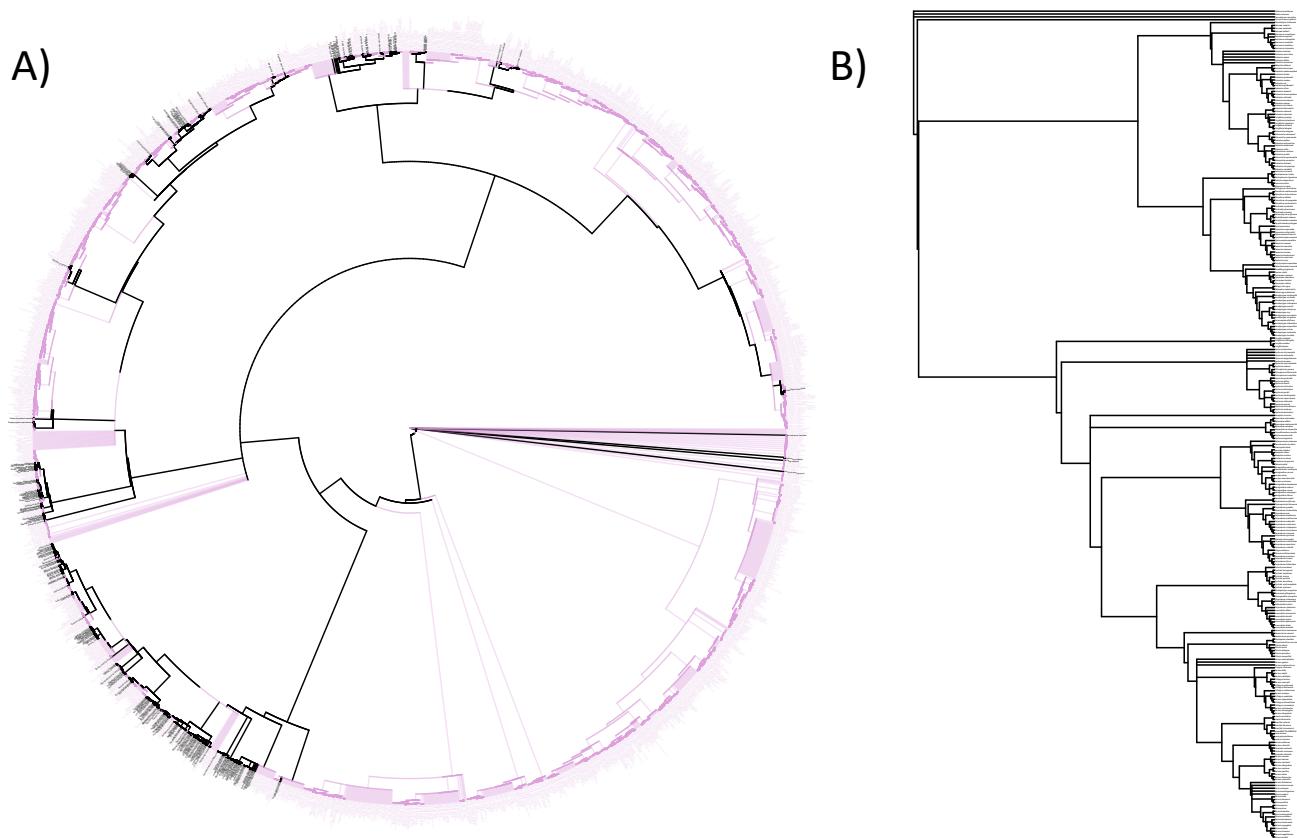
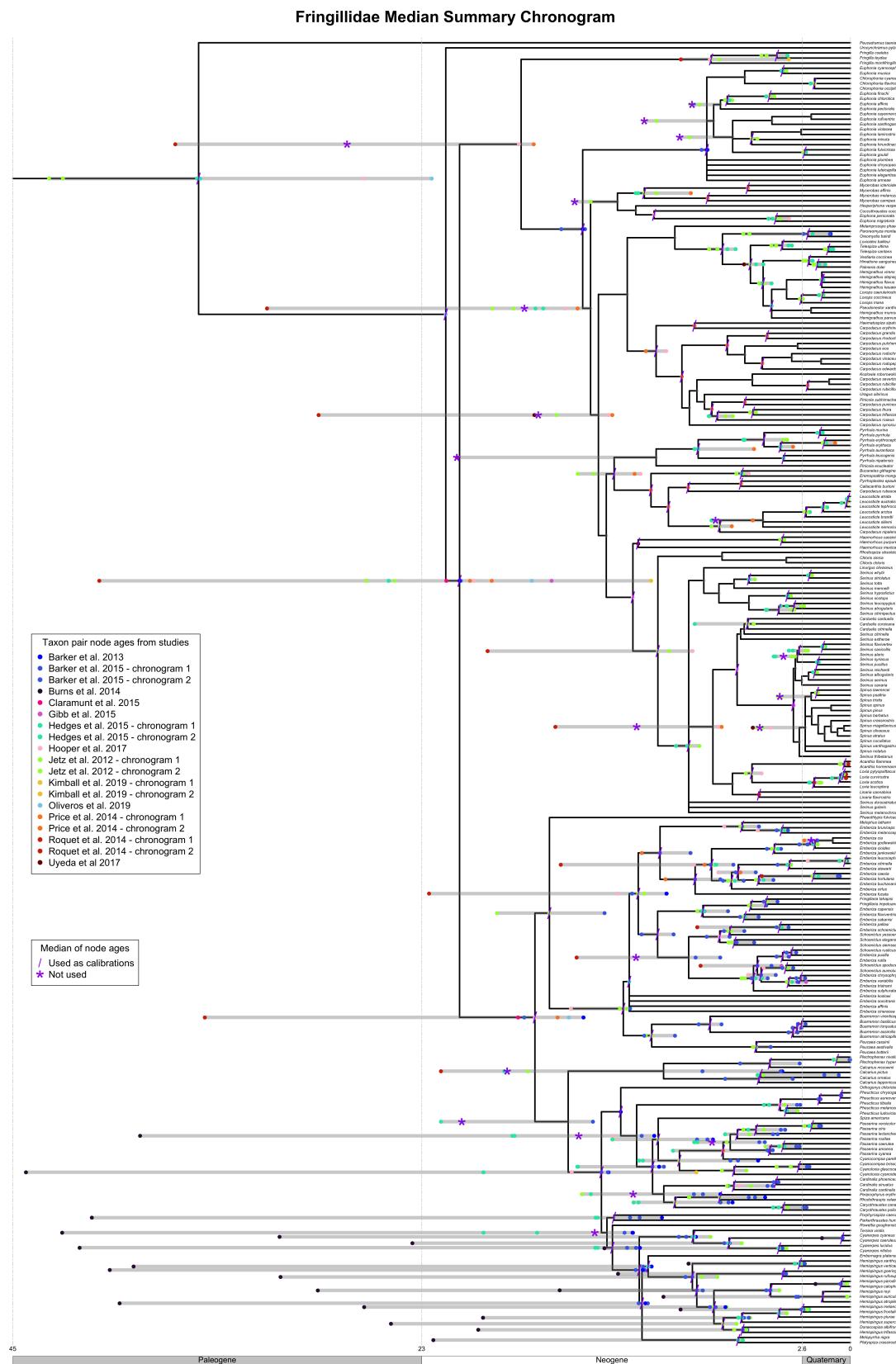


FIGURE 4. Tree topologies extracted from Open Tree of Life's (OpenTree) synthetic phylogenetic tree.
 865 A) Topology of 2,333 tips and 1,305 internal nodes, encompassing bird species within the family
 866 Fringillidae following the NCBI taxonomy (black), as well as all other bird species that share the same
 867 Most Recent Common Ancestor (MRCA) node in OpenTree's synthetic tree (purple). B) Subtree
 868 topology of 289 tips and 253 internal nodes, resulting from pruning species that do not belong to the
 869 family Fringillidae according to the NCBI taxonomy (purple branches in topology A). Bird species
 870 within the Fringillidae are paraphyletic (Alström et al. 2014, Barker, Cibois, Schikler, Feinstein, &
 871 Cracraft 2004, Barker et al. 2013, Barker 2014, Barker et al. 2015, Beresford, Barker, Ryan, &
 872 Crowe 2005, Bryson Jr et al. 2014, Burleigh, Kimball, & Braun 2015, Burns et al. 2014, Chaves,
 873 Hidalgo, & Klicka 2013, Claramunt & Cracraft 2015, Gibb et al. 2015, Hackett et al. 2008, Jetz et al.
 874 2012, Johansson, Fjeldså, & Bowi 200, Kimball et al. 2019, Klicka et al. 2014, Lamichhaney et al.
 875 2015, Lerner, Meyer, James, Hofreiter, & Fleischer 2011, Lovette et al. 2010, Moyle et al. 2016,
 876 Ödeen, Håstad, & Alström 2011, Oliveros et al. 2019, Päckert et al. 2012, Parchman, Benkman, &
 877 Mezquida 2007, Powell et al. 2014, Price et al. 2014, Pulgarín-R, Smith, Bryson Jr, Spellman, &
 878 Klicka 2013, Selvatti, Gonzaga, & Moraes Russo 2015, Tietze, Päckert, Martens, Lehmann, & Sun
 879 2013, Treplin et al. 2008, Zuccon, Prŷs-Jones, Rasmussen, & Ericson 2012).



887

882 FIGURE 5. *Fringillidae median summary chronogram generated with DateLife. It has 289 tips and*
 883 *253 nodes, from which 212 have age data from at least one published chronogram. These nodes have*

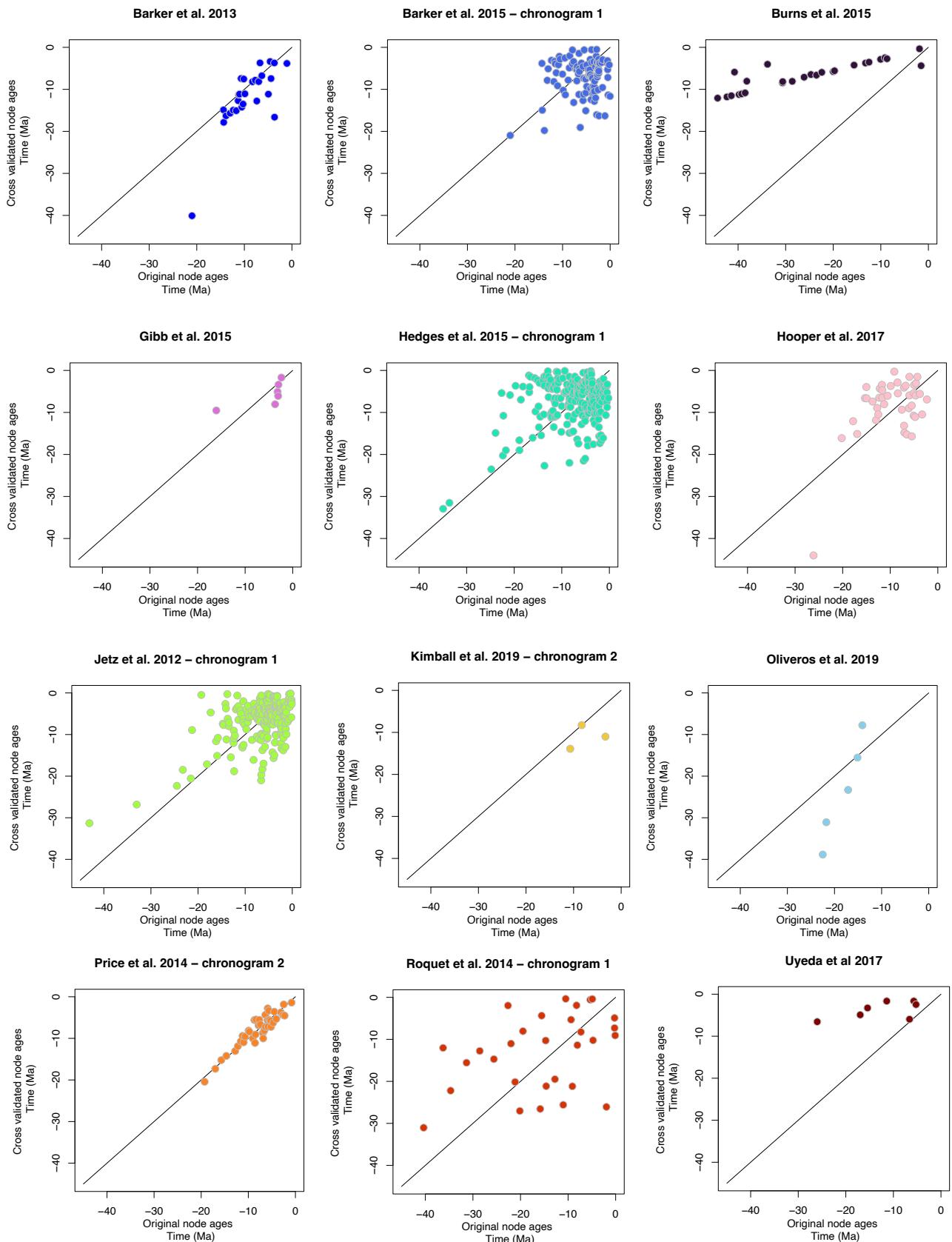


FIGURE 6. Results from cross validation analysis. Each plot compares the original age estimate (x axis) with the age obtained with a DateLife analysis (y axis), per node.