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

Luna L. Sánchez Reyes[1,2], Emily Jane McTavish[1], & Brian O'Meara[2]

[1] University of California, Merced

[2] University of Tennessee, Knoxville

## Author Note

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

Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, 425 Hesler Biology Building, Knoxville, TN 37996, USA.

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

Correspondence concerning this article should be addressed to Luna L. Sánchez Reyes, . E-mail: sanchez.reyes.luna@gmail.com

## Abstract

Date estimates for times of evolutionary divergences are key data for research in the natural sciences. These estimates also provide valuable information for for education, science communication and policy decisions. Although achieving a high-quality reconstruction of a phylogenetic tree with branch lengths proportional to absolute time (chronogram), is a difficult and time-consuming task, the increased availability of fossil and molecular data, and time-efficient analytical techniques has resulted in many recent publications of large chronograms for a large number and wide diversity of organisms. When these estimates are shared in public, open databases this wealth of expertly-curated and peer-reviewed data on time of evolutionary origin is exposed in a programatic and reusable way. Intensive and localized efforts have improved data sharing practices, as well as incentivizited open science in biology. Here we present DateLife, a service implemented as an R package and an Rshiny website application available at www.datelife.org/query/, that provides functionalities for efficient and easy finding, summary, reuse, and reanalysis of expert, peer-reviewed, public data on time of evolutionary origin. The main DateLife workflow constructs a chronogram for any given combination of taxon names, by searching a local chronogram database constructed and curated from the Open Tree of Life Phylesystem phylogenetic database, which incorporates phylogenetic data from TreeBASE database as well. We implement and test methods for summarizing time data from multiple source chronograms using supertree and congruification algorithms, and using age data extracted from source chronograms as secondary calibration points to add branch lengths proportional to absolute time to a tree topology. DateLife will be useful to increase awereness on the existing variation in expert time of divergence data, and can foster exploration of the effect of alternative divergence time hypothesis on the results of analyses, providing a framework for a more informed interpretation of evolutionary results.


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

Congruification; Supertree; Calibrations; Secondary calibrations

Word count: 4362

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

## Introduction

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

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

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

## Description

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

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

## Processing an input

DateLife starts by processing an input consisting of at least two taxon names, which 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 below.

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

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 (**boyle2013?**). DateLife implements TNRS using OpenTree's unified taxonomy as standard (Open Tree Of Life et al., 2016; Rees & Cranston, 2017), storing OpenTree's

Taxonomy identification numbers for further processing.

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

## Searching the database

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

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 withing 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. Last but not least, single summary chronograms resulting from summarizing age data from available source chronograms, as follows:

***Choosing a topology.***– DateLife requires a tree topology to summarize age data upon. We recommend that users provide a tree topology as input from the literature, or one of their own making. If no topology is provided, DateLife automatically subsets one from the OpenTree synthetic tree (Open Tree Of Life et al., 2019). Alternatively, DateLife can reconstruct a tree with branch lengths proportional to substitution rates from a starting tree topology using genetic data from the Barcode of Life Data System, BOLD (Ratnasingham & Hebert, 2007), or combine topologies from source chronograms using a supertree approach.

***Reconstructing branch lengths.***– DateLife starts by mining the BOLD database to obtain genetic markers for the input taxa, and aligning them with MUSCLE (Edgar, 2004; or MAFFT, Katoh, Asimenos, & Toh, 2009). Currently, branch length reconstruction is performed with parsimony and the likelihood of the phylogenetic tree given a sequence

alignment is computed (Schliep, 2011).

***Combining source chronograms.***– To combine topologies from source chronograms into a single summary (or supertree) topology, the DateLife algorithm starts by identifying 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). 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. If source chronograms do not form a grove, the supertree reconstruction will fail.

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

***Summarizing congruified ages.***– For each congruent node, the pairwise distance that traverse that node are summarized into a single summary matrix using classic statistics (mean, median, minimum and maximum ages). These single summary ages per taxon pair are then used as secondary calibrations to date the tree topology.

***Dating the tree topology.***– By default, DateLife implements the Branch Length Adjuster (BLADJ) algorithm to obtain a fully dated topology. BLADJ fixes node ages that have calibration data, and distributes time between nodes with no data evenly between nodes with calibration data. This minimizes age variance in the resulting chronogram (Campbell O. Webb, Ackerly, & Kembel, 2008). When there is conflict in ages between nodes with calibration data, the algorithm ignores ages that are older than ages of parent nodes

and/or younger than ages from descendant nodes.

If there is no information on the age of the root in the chronogram database, users can provide an estimate from the literature. If none is provided, DateLife assigns an arbitrary age to the root as 10% older than the oldest age available within the group.

Summarized calibrations can be applied as secondary calibrations with different dating methods currently supported within DateLife: MrBayes (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003), PATHd8 (Britton, Anderson, Jacquet, Lundqvist, & Bremer, 2007), BLADJ (Campbell O. Webb et al., 2008; Campbell O. Webb & Donoghue, 2005), and treePL (Stephen A. Smith & O'Meara, 2012).

***Visualizing results.***– Finally, users can save all source and summary chronograms in formats that permit reuse and reanalyses (newick and R "phylo" format), as well as visualize and compare results graphically, or construct their own graphs using `datelife`'s chronogram plot generation functions.

## Benchmark

`datelife`'s 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 through the database relative to number of queried taxon names. Query processing time increases roughly linearly with number of input taxon names, and increases considerably if Taxonomic Name Resolution Service (TNRS) is activated. Up to ten thousand names can be processed and searched in less than 30 minutes with the most time consuming settings. Once names have been processed as described in methods, a name search through the chronogram database can be performed in less than a minute, even with a very large number of taxon names (Fig. 2). `datelife`'s code performance was evaluated with a set of unit tests designed and implemented with the R package testthat (R Core Team, 2018) that were run both locally with the devtools package (R Core Team, 2018), and on a public server –via GitHub,

using the continuous integration tool Travis CI (https://travis-ci.org). At present, unit tests cover more than 40% of `datelife`'s code (https://codecov.io/gh/phylotastic/datelife).

## Case studies

We illustrate the DateLife algorithm using a group within the Passeriform birds encompassing the family of true finches, Fringillidae and allies as case study. The first example analyses 6 bird species and shows all steps of the algorithm. The second example is a real life application

### Small example

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

Processing input names found that *Emberiza elegans* is synonym for *Schoeniclus elegans* in the default reference taxonomy [Open Tree of Life Taxonomy v3.3, June 1, 2021]. For a detailed discussion on the state of the synonym refer to Avibase (Avibase, 2022; Lepage, 2004; Lepage, Vaidya, & Guralnick, 2014). DateLife used the processed input names to search the local chronogram database and found 9 matching chronograms in 6 different studies. Three studies matched five input names (Barker, Burns, Klicka, Lanyon, & Lovette, 2015; Hedges, Marin, Suleski, Paymer, & Kumar, 2015; Jetz, Thomas, Joy, Hartmann, & Mooers, 2012), one study matched four input names (Hooper & Price, 2017) and two studies matched two input names (Barker, Burns, Klicka, Lanyon, & Lovette, 2013; Burns et al., 2014). No studies matched all input names. Together, matching chronograms have 28 unique age data points. All nodes have age data. As fixed tree topology, DateLife used OpenTree's

synthetic tree as default and mapped age data to nodes in the tree. As expected, more inclusive nodes (e.g., node "n1") have more age data than less inclusive nodes (e.g., node "n5"). The processing step allowed discovering five data points for node "n4" that would not have had any data otherwise. Age summary statistics per node were calculated and tested as secondary calibrations to date the tree topology using the BLADJ algorithm. Age data for node "n2" was excluded as final calibration because it is older than age data of a more inclusive node.

## Real life application

A college educator wishes to obtain state-of-the-art data on time of evolutionary origin of species belonging to the true finches for their class. They decide to use `datelife` because they are teaching best practices for reproducibility. Students have the option to go to the website at www.datelife.org and perform an interactive run. However, the educator also wants the students to practice their R skills. The first step is to run a `datelife` query using the "get species from taxon" flag. This will get all recognised species names within their chosen inclusive taxon. The Fringillidae has 289 species, according to the Open Tree of Life taxonomy. Once with a curated set of species taxon names, the next step is to run a `datelife` search that will find all chronograms that contain at least two species names. The algorithm proceeds to prune the trees to keep matching species names on tips only, and transform the pruned trees to pairwise distance matrices. There are 13 chronograms containing at least two Fringillidae species, published in 9 different studies (Barker et al., 2013, 2015; Burns et al., 2014; Claramunt & Cracraft, 2015; Gibb et al., 2015; Hedges et al., 2015; Hooper & Price, 2017; Jetz et al., 2012; Price et al., 2014). The final step is to summarize the available information using two alternative types of summary chronograms, median and SDM. As explained in the "Description" section, data from source chronograms is first summarised into a single distance matrix and then the available node ages are used as fixed node calibrations over a consensus tree topology, to obtain a fully dated tree with the

program BLADJ (Fig. 5). Median summary chronograms are older and have wider variation in maximum ages than chronograms obtained with SDM.

## Cross-validation test

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

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

## Discussion

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

At the time of writing of this manuscript (May 05, 2022), `datelife`'s database has 253 chronograms, pulled entirely from OpenTree's database, the Phylesystem (McTavish et al.,

2015). A unique feature of OpenTree's Phylesystem is that the community can add new state-of-the-art chronograms any time. As chronograms are added to Phylesystem, they are incorporated into an updated `datelife`'s database that is assigned a new version number, followed by a package release on CRAN. `datelife`'s chronogram database is updated as new chronogram data is added to Phylesystem, at a minimum of once a month and a maximum of every 6 months. Users can also upload new chronograms to OpenTree themselves, and trigger an update of their local `datelife` database to incorporate the new chronograms, to have them immediately available for analysis.

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

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

By default, `datelife` currently summarizes all source chronograms that overlap with at least two species names. Users can exclude source chronograms if they have reasons to do

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

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

Furthermore, chronograms can be obtained with primary fossil data or with molecular substitution rates obtained experimentally, which can deepen the already substantial

variation in time estimates between lineages, as observed from the comparison of source chronograms in the Fringillidae example. This observation is often encountered in the literature (see, for example, the ongoing debate about crown group age of angiosperms (Barba-Montoya, Reis, Schneider, Donoghue, & Yang, 2018; Magallón et al., 2015; Ramshaw et al., 1972; Sanderson & Doyle, 2001; Sauquet, Ramírez-Barahona, & Magallón, 2021). For some studies, especially ones based on branch lengths (e.g., studies of species diversification, timing of evolutionary events, phenotypic trait evolution), using a different chronogram may return different results (Title & Rabosky, 2016). Stitching together these chronograms can create a larger tree that uses information from multiple studies, but the effect of uncertainties and errors at this level on downstream analyses is still largely unknown.

Summarizing chronograms might also imply summarizing fundamentally distinct evolutionary hypotheses. For example, two different researchers working on the same clade both carefully select and argument their choices of fossil calibrations. Still, if one researcher decides a fossil will calibrate the ingroup of a clade, while another researcher uses the same one to calibrate outside the clade, the resulting age estimates will often differ substantially, as the placement of calibrations as stem or crown group is proved to deeply affect estimated times of lineage divergence (Sauquet, 2013). Trying to summarize the resulting chronograms into a single one using simple summary statistics can erase many types of relevant information from the source chronograms. Accordingly, the prevailing view is that we should favor time of lineage divergence estimates obtained from a single analysis, using fossil data as primary sources of calibrations, and using fossils that have been widely discussed and curated as calibrations to date other trees, making sure that all data used in the analysis reflect a coherent evolutionary history (Antonelli et al., 2017). However, the exercise of summarizing different chronograms has the potential to help getting a single global evolutionary history for a lineage by putting together evidence from different hypothesis. Choosing the elements of the chronograms that we are going to keep and the ones that we are going to discard is key, since we are potentially loosing important parts of the

evolutionary history of a lineage that might only be reflected in source chronograms and not on the summary chronogram (Sauquet et al., 2021).

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

## Conclusions

Divergence time information is key to many areas of evolutionary studies: trait evolution, diversification, biogeography, macroecology and more. It is also crucial for science communication and education, but generating chronograms is difficult, especially for those who want to use phylogenies but who are not systematists, or do not have the time to acquire and develop the necessary knowledge and data curation skills. Moreover, years of primarily public funded research have resulted in vast amounts of chronograms that are

already available on scientific publications, but hidden to the public and scientific community for reuse.

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

## Availability

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

## Supplementary Material

Code used to generate all versions of this manuscript, the biological examples, as well as the benchmark of functionalities are available at datelifeMS1, datelife_examples, and

datelife_benchmark repositories in LLSR's GitHub account.

## Funding

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

## Acknowledgements

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

## References


Ané, C., Eulenstein, O., Piaggio-Talice, R., & Sanderson, M. J. (2009). Groves of phylogenetic trees. *Annals of Combinatorics*, *13*(2), 139–167.

Antonelli, A., Hettling, H., Condamine, F. L., Vos, K., Nilsson, R. H., Sanderson, M. J., … Vos, R. A. (2017). Toward a self-updating platform for estimating rates of speciation and migration, ages, and relationships of Taxa. *Systematic Biology*, *66*(2), 153–166. https://doi.org/10.1093/sysbio/syw066

Archie, J., Day, W. H., Felsenstein, J., Maddison, W., Meacham, C., Rohlf, F. J., & Swofford, D. (1986). The Newick tree format. Retrieved from %7Bhttps://evolution.genetics.washington.edu/phylip/newicktree.html%7D

Avibase. (2022). Yellow-throated Bunting. *Avibase - The World Bird Database*, (Online Resource). Retrieved from %7Bhttps://avibase.bsc-eoc.org/species.jsp?lang=EN&avibaseid=82D1EE0049D8D927%7D

Bapst, D. W. (2012). Paleotree: An R package for paleontological and phylogenetic analyses of evolution. *Methods in Ecology and Evolution*, *3*(5), 803–807. https://doi.org/10.1111/j.2041-210X.2012.00223.x

Barba-Montoya, J., Reis, M. dos, Schneider, H., Donoghue, P. C., & Yang, Z. (2018). Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a cretaceous terrestrial revolution. *New Phytologist*, *218*(2), 819–834.

Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2013). Going to extremes: Contrasting rates of diversification in a recent radiation of new world passerine birds. *Systematic Biology*, *62*(2), 298–320.

Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2015). New insights into new world biogeography: An integrated view from the phylogeny of blackbirds, cardinals, sparrows, tanagers, warblers, and allies. *The Auk: Ornithological Advances*, *132*(2), 333–348.

Britton, T., Anderson, C. L., Jacquet, D., Lundqvist, S., & Bremer, K. (2007).

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

*and Evolution*, *7*(11), 2983–2995.

Harmon, L., Weir, J., Brock, C., Glor, R., & Challenger, W. (2008). GEIGER: investigating evolutionary radiations. *Bioinformatics*, *24*, 129–131.

Hedges, S. B., Marin, J., Suleski, M., Paymer, M., & Kumar, S. (2015). Tree of life reveals clock-like speciation and diversification. *Molecular Biology and Evolution*, *32*(4), 835–845. https://doi.org/10.1093/molbev/msv037

Heibl, C. (2008). *PHYLOCH: R language tree plotting tools and interfaces to diverse phylogenetic software packages.* Retrieved from http://www.christophheibl.de/Rpackages.html

Hooper, D. M., & Price, T. D. (2017). Chromosomal inversion differences correlate with range overlap in passerine birds. *Nature Ecology & Evolution*, *1*(10), 1526.

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, *17*(8), 754–755. https://doi.org/10.1093/bioinformatics/17.8.754

Jetz, W., Thomas, G., Joy, J. J. B., Hartmann, K., & Mooers, A. (2012). The global diversity of birds in space and time. *Nature*, *491*(7424), 444–448. https://doi.org/10.1038/nature11631

Katoh, K., Asimenos, G., & Toh, H. (2009). Multiple alignment of DNA sequences with MAFFT. In *Bioinformatics for DNA sequence analysis* (pp. 39–64). Springer.

Laubichler, M. D., & Maienschein, J. (2009). *Form and function in developmental evolution*. Cambridge University Press.

Lepage, D. (2004). *Avibase: The world bird database*. Bird Studies Canada.

Lepage, D., Vaidya, G., & Guralnick, R. (2014). Avibase–a database system for managing and organizing taxonomic concepts. *ZooKeys*, (420), 117.

Magallon, S., & Sanderson, M. J. (2001). Absolute diversification rates in angiosperm clades. *Evolution*, *55*(9), 1762–1780.

Magallón, S. (2010). Using fossils to break long branches in molecular dating: A comparison of relaxed clocks applied to the origin of angiosperms. *Systematic Biology*, *59*(4), 384–399.

Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., & Hernández-Hernández, T. (2015). A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist*, *207*(2), 437–453.

McTavish, E. J., Hinchliff, C. E., Allman, J. F., Brown, J. W., Cranston, K. A., Holder, M. T., … Smith, S. A. (2015). Phylesystem: A git-based data store for community-curated phylogenetic estimates. *Bioinformatics*, *31*(17), 2794–2800.

Michonneau, F., Brown, J. W., & Winter, D. J. (2016). rotl: an R package to interact with the Open Tree of Life data. *Methods in Ecology and Evolution*, *7*(12), 1476–1481. https://doi.org/10.1111/2041-210X.12593

Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecology Letters*, *17*(4), 508–525. https://doi.org/10.1111/ele.12251

Ooms, J., & Chamberlain, S. (2018). *Phylocomr: Interface to 'phylocom'*. Retrieved from https://CRAN.R-project.org/package=phylocomr

Open Tree Of Life, Redelings, B., Cranston, K. A., Allman, J., Holder, M. T., & McTavish, E. J. (2016). Open Tree of Life APIs v3.0. *Open Tree of Life Project*, (Online Resources). Retrieved from %7Bhttps://github.com/OpenTreeOfLife/germinator/wiki/Open-Tree-of-Life-Web-APIs%7D

Open Tree Of Life, Redelings, B., Sánchez Reyes, L. L., Cranston, K. A., Allman, J., Holder, M. T., & McTavish, E. J. (2019). Open tree of life synthetic tree v12.3. *Zenodo*. Retrieved from https://doi.org/10.5281/zenodo.3937742

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*(2), 289–290.

Posadas, P., Crisci, J. V., & Katinas, L. (2006). Historical biogeography: A review of

its basic concepts and critical issues. *Journal of Arid Environments*, *66*(3), 389–403.

Price, T. D., Hooper, D. M., Buchanan, C. D., Johansson, U. S., Tietze, D. T., Alström, P., ... others. (2014). Niche filling slows the diversification of himalayan songbirds. *Nature*, *509*(7499), 222.

R Core Team. (2018). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., ... others. (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*, *559*(7714), 392.

Ramshaw, J., Richardson, D., Meatyard, B., Brown, R., Richardson, M., Thompson, E., & Boulter, D. (1972). The time of origin of the flowering plants determined by using amino acid sequence data of cytochrome c. *New Phytologist*, *71*(5), 773–779.

Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The barcode of life data system (http://www. Barcodinglife. org). *Molecular Ecology Notes*, *7*(3), 355–364.

Rees, J. A., & Cranston, K. (2017). Automated assembly of a reference taxonomy for phylogenetic data synthesis. *Biodiversity Data Journal*, (5).

Revell, L. J. (2012). Phytools: An r package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*, 217–223.

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*(12), 1572–1574. https://doi.org/10.1093/bioinformatics/btg180

Sanchez-Reyes, L. L., O'Meara, B., Eastman, J., Heath, T., Wright, A., Schliep, K., ... Alfaro, M. (2022). datelife: Scientific Data on Time of Lineage Divergence for Your Taxa. *R Package Version 0.6.2*. Retrieved from https://doi.org/10.5281/zenodo.593938

Sanderson, M. J. (2003). r8s: Inferring absolute rates of molecular evolution and

divergence times in the absence of a molecular clock. *Bioinformatics*, *19*(2), 301–302.

Sanderson, M. J., & Doyle, J. A. (2001). Sources of error and confidence intervals in estimating the age of angiosperms from rbcL and 18S rDNA data. *American Journal of Botany*, *88*(8), 1499–1516.

Sauquet, H. (2013). A practical guide to molecular dating. *Comptes Rendus Palevol*, *12*(6), 355–367.

Sauquet, H., Ramírez-Barahona, S., & Magallón, S. (2021). *The age of flowering plants is unknown*.

Schenk, J. J. (2016). Consequences of secondary calibrations on divergence time estimates. *PLoS ONE*, *11*(1). https://doi.org/10.1371/journal.pone.0148228

Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in r. *Bioinformatics*, *27*(4), 592–593.

Smith, Stephen A., & Brown, J. W. (2018). Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany*, *105*(3), 302–314.

Smith, Stephen A., & O'Meara, B. C. (2012). TreePL: Divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics*, *28*(20), 2689–2690. https://doi.org/10.1093/bioinformatics/bts492

Stoltzfus, A., Lapp, H., Matasci, N., Deus, H., Sidlauskas, B., Zmasek, C. M., … Jordan, G. (2013). Phylotastic! Making tree-of-life knowledge accessible, reusable and convenient. *BMC Bioinformatics*, *14*. https://doi.org/10.1186/1471-2105-14-158

Title, P. O., & Rabosky, D. L. (2016). Do Macrophylogenies Yield Stable Macroevolutionary Inferences? An Example from Squamate Reptiles. *Systematic Biology*, syw102. https://doi.org/10.1093/sysbio/syw102

Vos, R. A., Balhoff, J. P., Caravas, J. A., Holder, M. T., Lapp, H., Maddison, W. P., … others. (2012). NeXML: Rich, extensible, and verifiable representation of

583 comparative data and metadata. *Systematic Biology*, *61*(4), 675–689.

584 Webb, C. (2000). Exploring the Phylogenetic Structure of Ecological Communities :

585 An Example for Rain Forest Trees. *The American Naturalist*, *156*(2), 145–155.

586 Webb, Campbell O., Ackerly, D. D., & Kembel, S. W. (2008). Phylocom: Software for

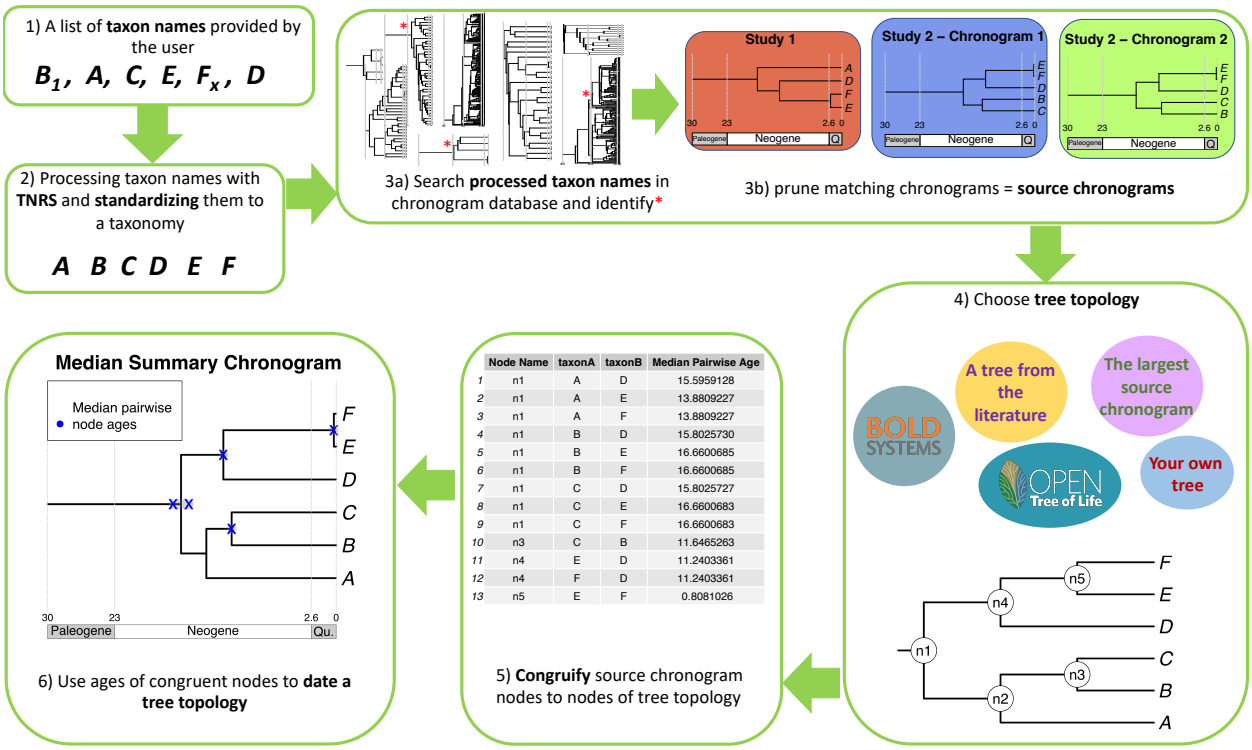
587 the analysis of phylogenetic community structure and trait evolution.

588 *Bioinformatics*, *24*(18), 2098–2100.

589 https://doi.org/10.1093/bioinformatics/btn358

590 Webb, Campbell O., & Donoghue, M. J. (2005). Phylomatic: Tree assembly for

591 applied phylogenetics. *Molecular Ecology Notes*, *5*(1), 181–183.




| | Node Name | taxonA | taxonB | Median Pairwise Age |
|---|---|---|---|---|
| 1 | n1 | A | D | 15.5959128 |
| 2 | n1 | A | E | 13.8809227 |
| 3 | n1 | A | F | 13.8809227 |
| 4 | n1 | B | D | 15.8025730 |
| 5 | n1 | B | E | 16.6600685 |
| 6 | n1 | B | F | 16.6600685 |
| 7 | n1 | C | D | 15.8025727 |
| 8 | n1 | C | E | 16.6600683 |
| 9 | n1 | C | F | 16.6600683 |
| 10 | n3 | C | B | 11.6465263 |
| 11 | n4 | E | D | 11.2403361 |
| 12 | n4 | F | D | 11.2403361 |
| 13 | n5 | E | F | 0.8081026 |

FIGURE 1. Stylized DateLife workflow. This shows the general worflows 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.

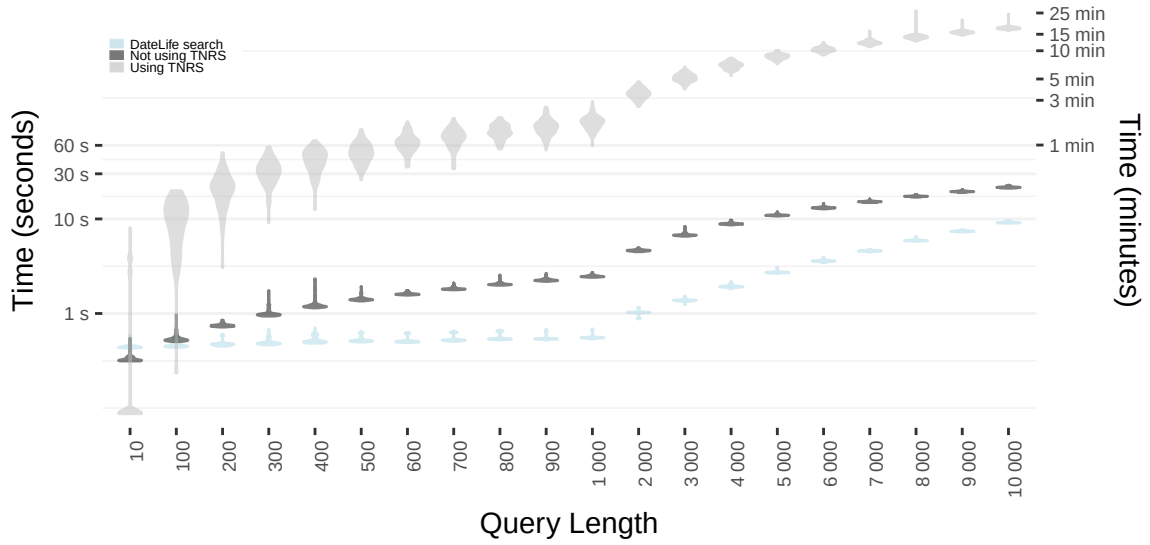
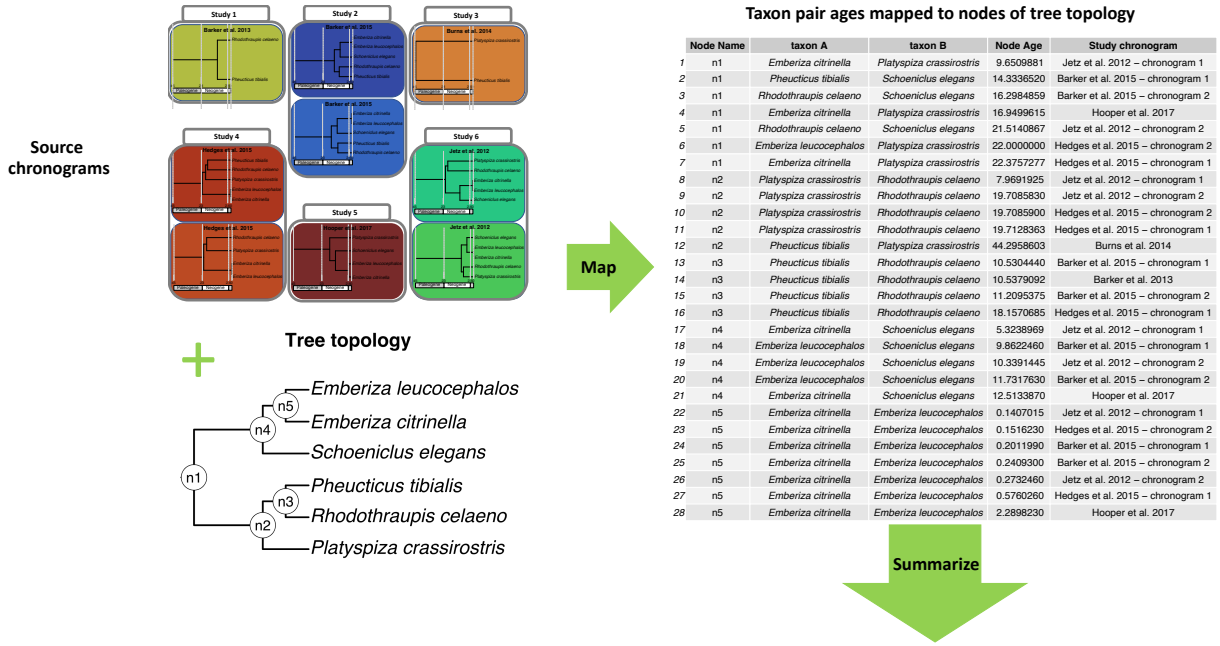



FIGURE 2. Computation time of query processing and search across `datelife`'s chronogram database relative to number of input taxon names. We sampled N names from the class Aves for each cohort 100 times and then performed a search with query processing not using the Taxon Names Resoultion Service (TNRS;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




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

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 | | *Pheucticus tibialis* | *Emberiza citrinella* | 16.298486 | |
| 2 | | *Pheucticus tibialis* | *Emberiza leucocephalos* | 16.298486 | |
| 3 | | *Platyspiza crassirostris* | *Emberiza citrinella* | 21.514085 | |
| 4 | | *Platyspiza crassirostris* | *Emberiza leucocephalos* | 21.514085 | |
| 5 | n1 | *Rhodothraupis celaeno* | *Emberiza citrinella* | 20.408031 | 19.301977 |
| 6 | | *Rhodothraupis celaeno* | *Emberiza leucocephalos* | 20.408031 | |
| 7 | | *Schoeniclus elegans* | *Pheucticus tibialis* | 15.316069 | |
| 8 | | *Schoeniclus elegans* | *Platyspiza crassirostris* | 19.301977 | |
| 9 | | *Schoeniclus elegans* | *Rhodothraupis celaeno* | 17.800231 | |
| 10 | n2 | *Platyspiza crassirostris* | *Pheucticus tibialis* | 32.004348 | 25.856467327225 |
| 11 | | *Rhodothraupis celaeno* | *Platyspiza crassirostris* | 19.708587 | |
| 12 | n3 | *Rhodothraupis celaeno* | *Pheucticus tibialis* | 10.873723 | 10.87372335475 |
| 13 | n4 | *Schoeniclus elegans* | *Emberiza citrinella* | 10.647794 | 10.6477935 |
| 14 | | *Schoeniclus elegans* | *Emberiza leucocephalos* | 10.647794 | |
| 15 | n5 | *Emberiza leucocephalos* | *Emberiza citrinella* | 0.273246 | 0.273246 |

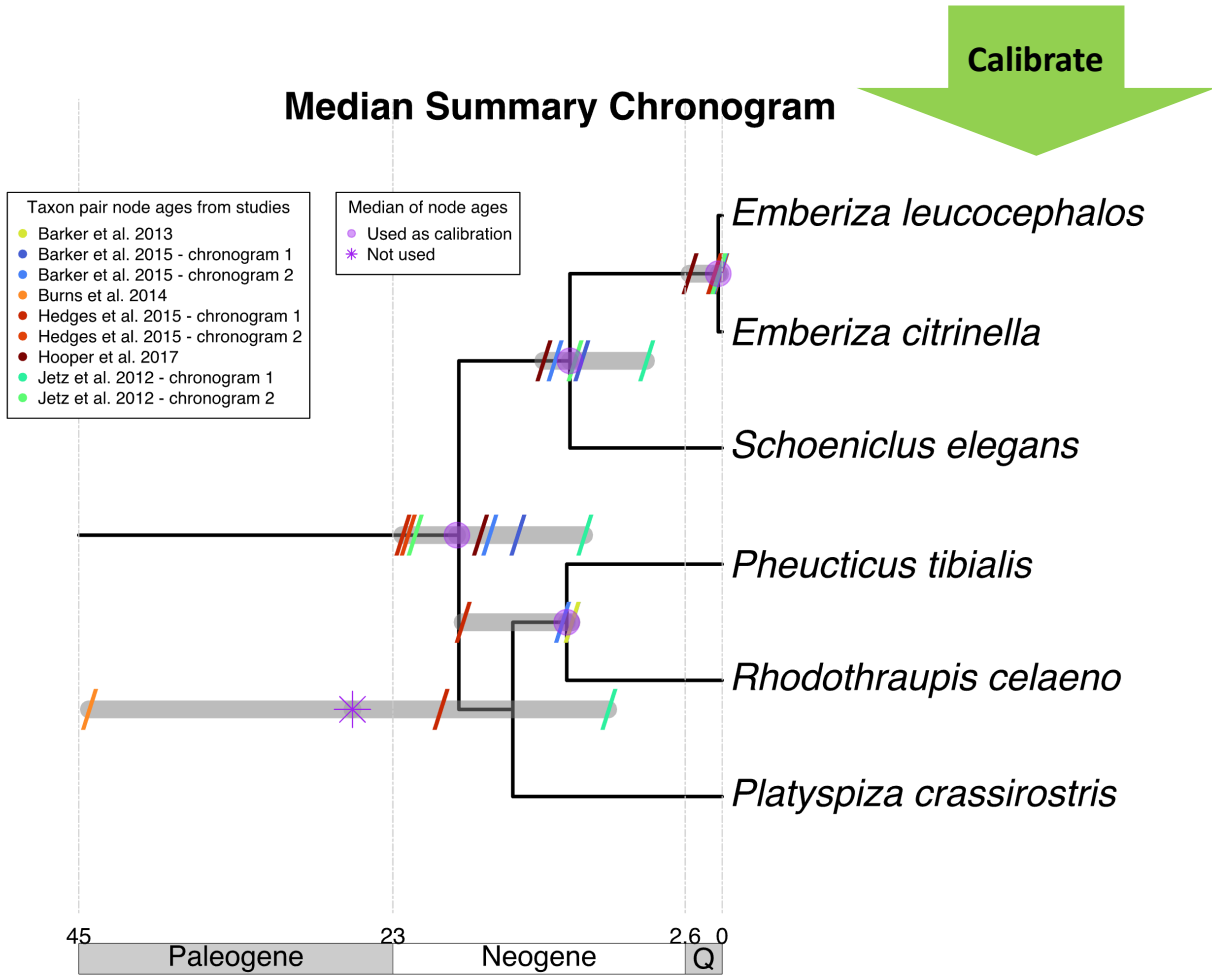



FIGURE 4. Summarized age data is used as secondary calibrations to date a tree topology as a summary chronogram.

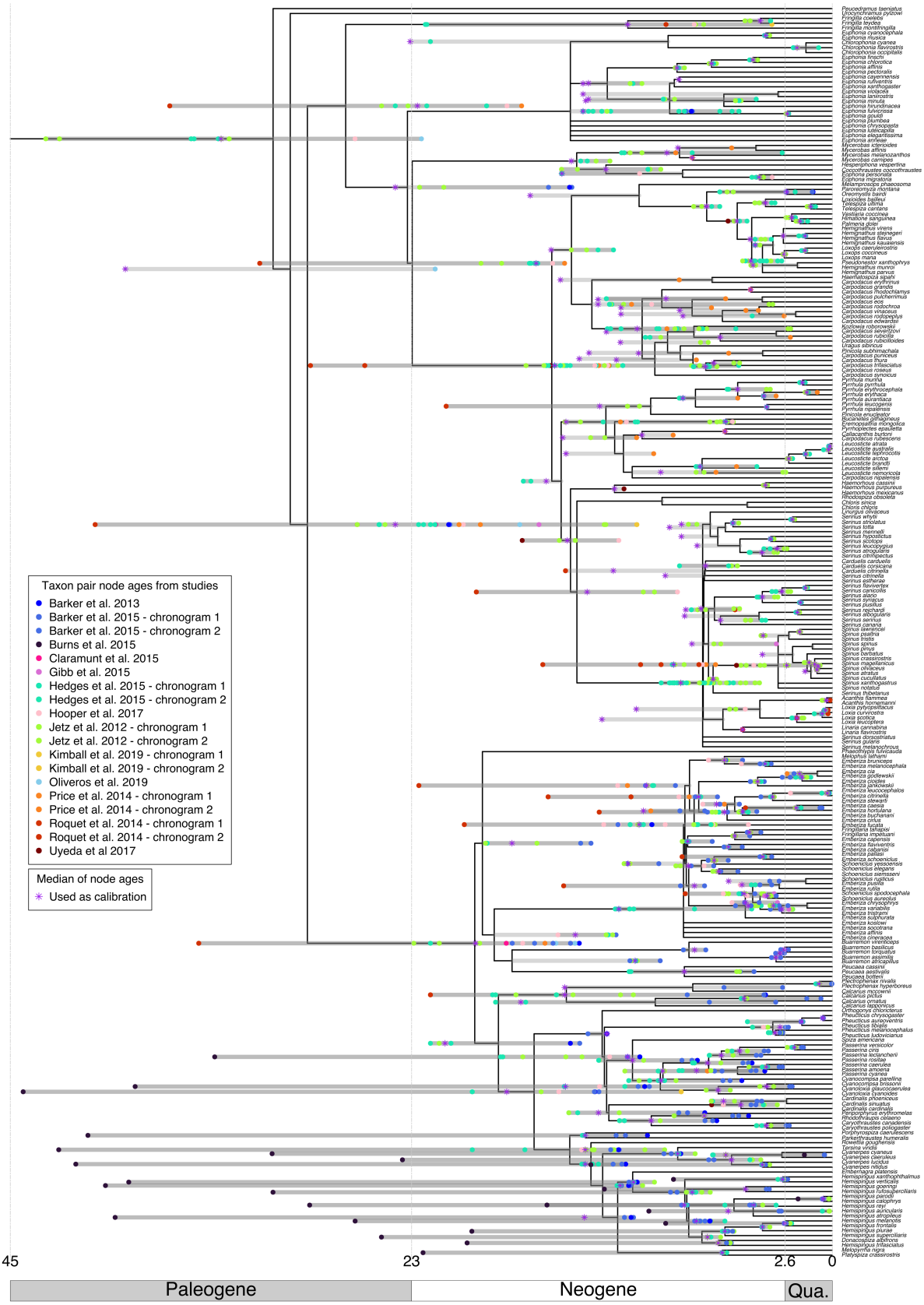



FIGURE 5. Fringillidae median summary chronogram genertaed with DateLife. It has 256 tips and 233 nodes.

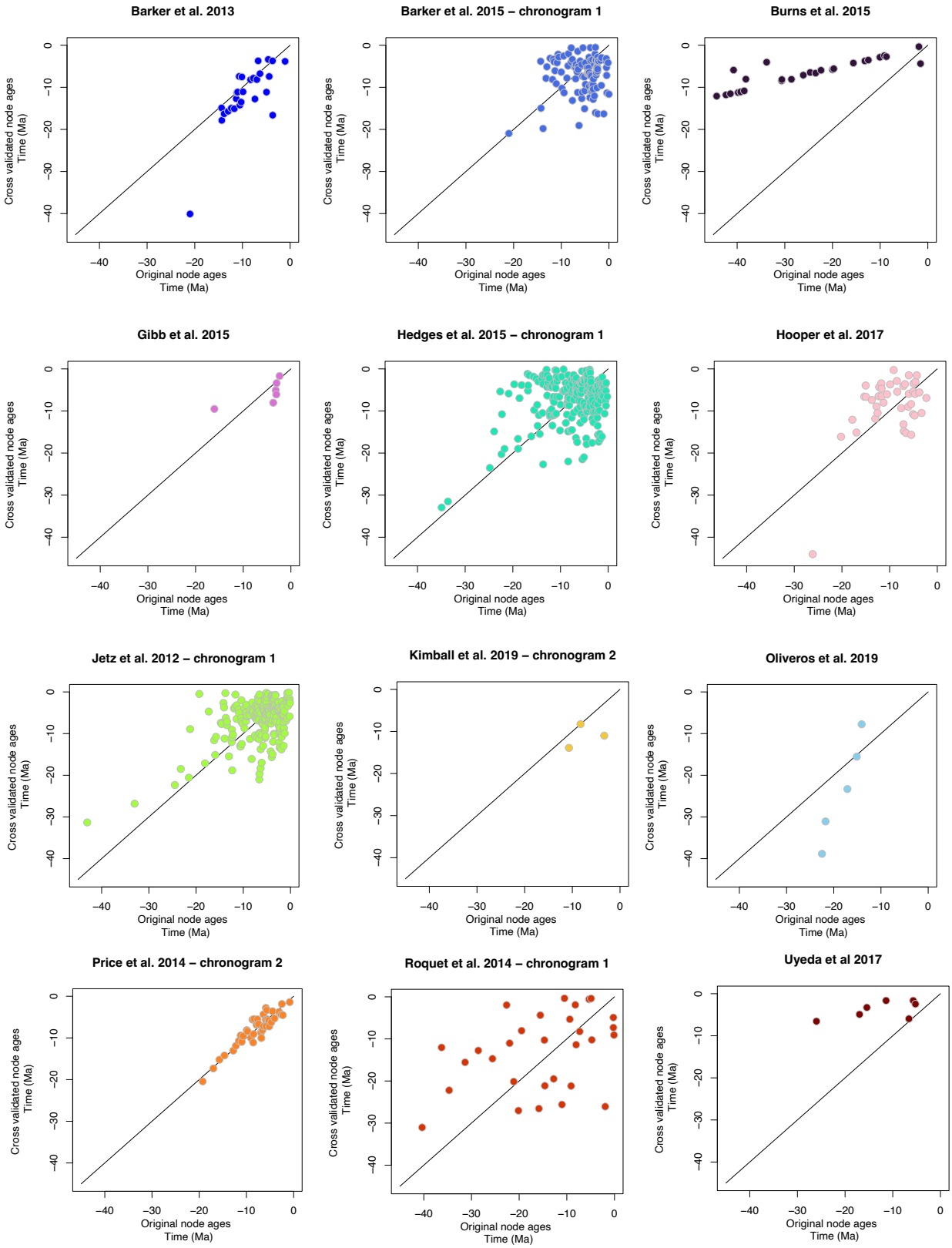



FIGURE 6. Results from cross validation analysis.

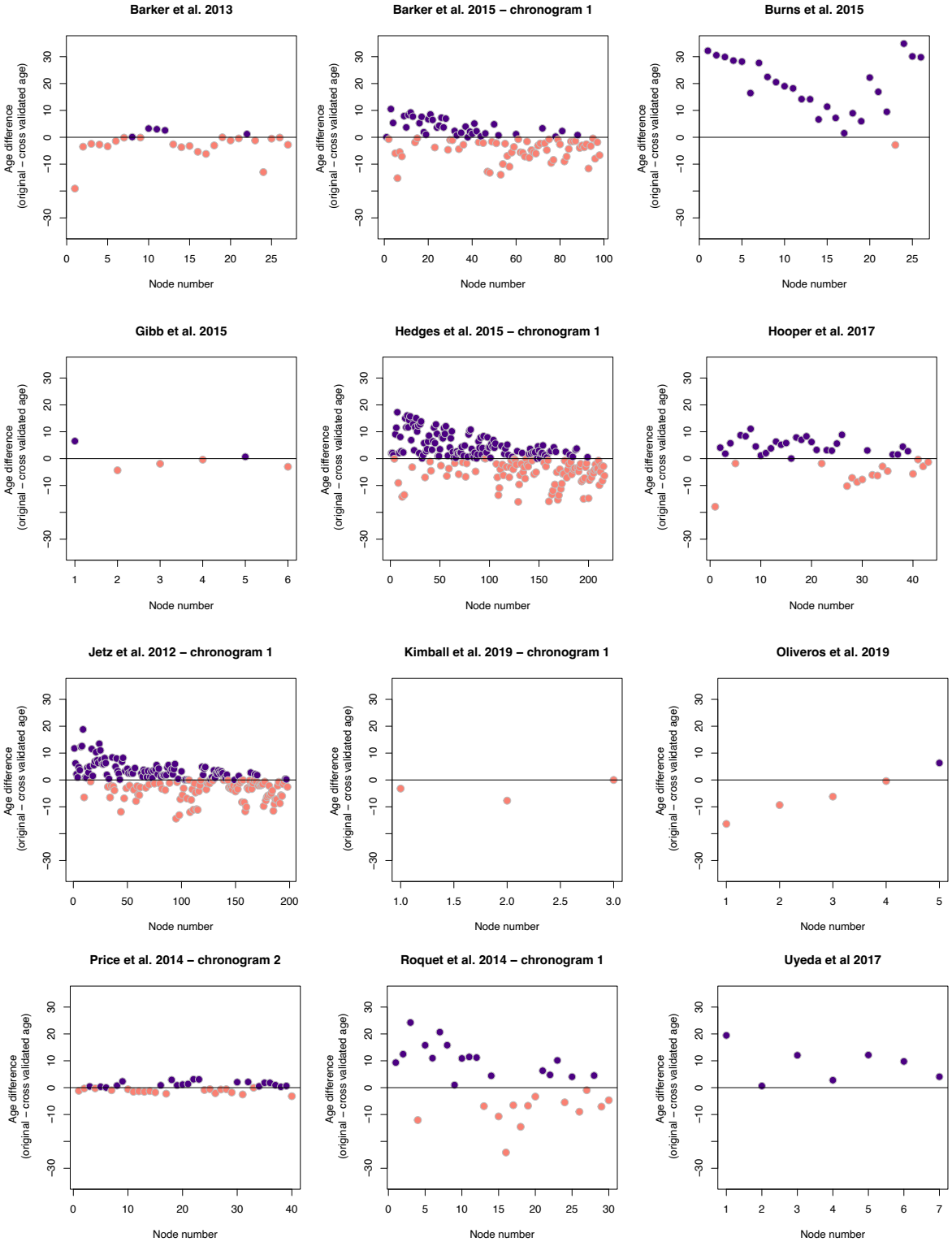



FIGURE 7. Results from cross validation analysis.

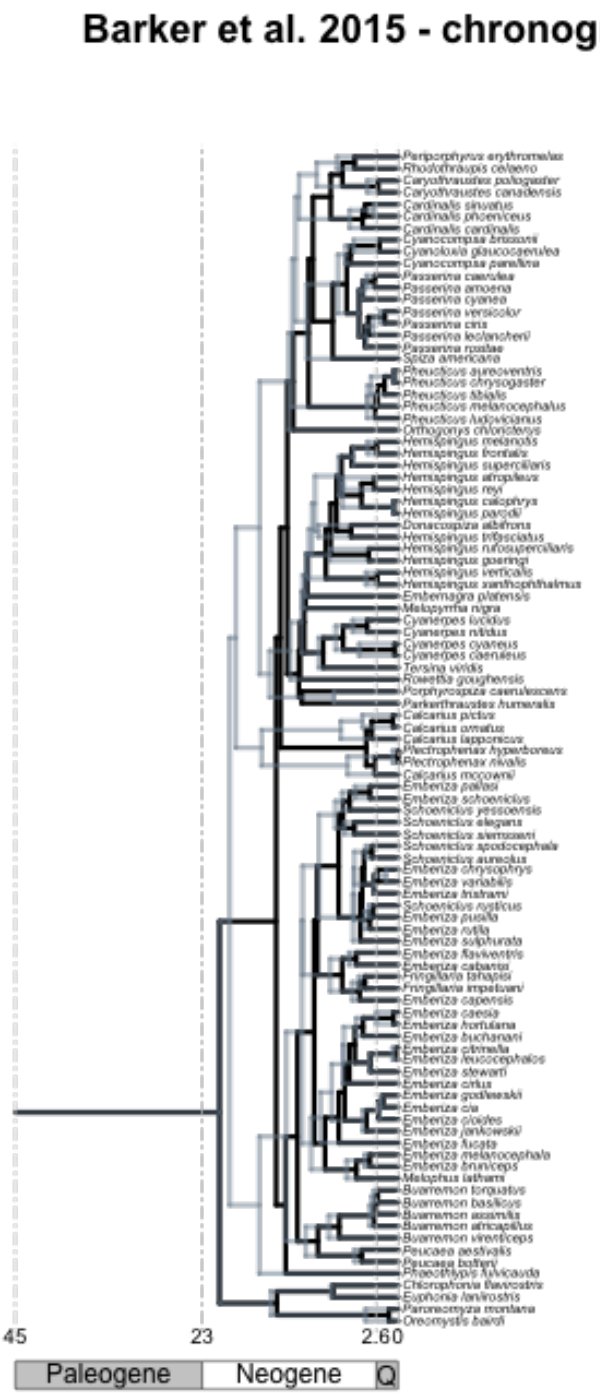



FIGURE 8. Cross validation of second source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to


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

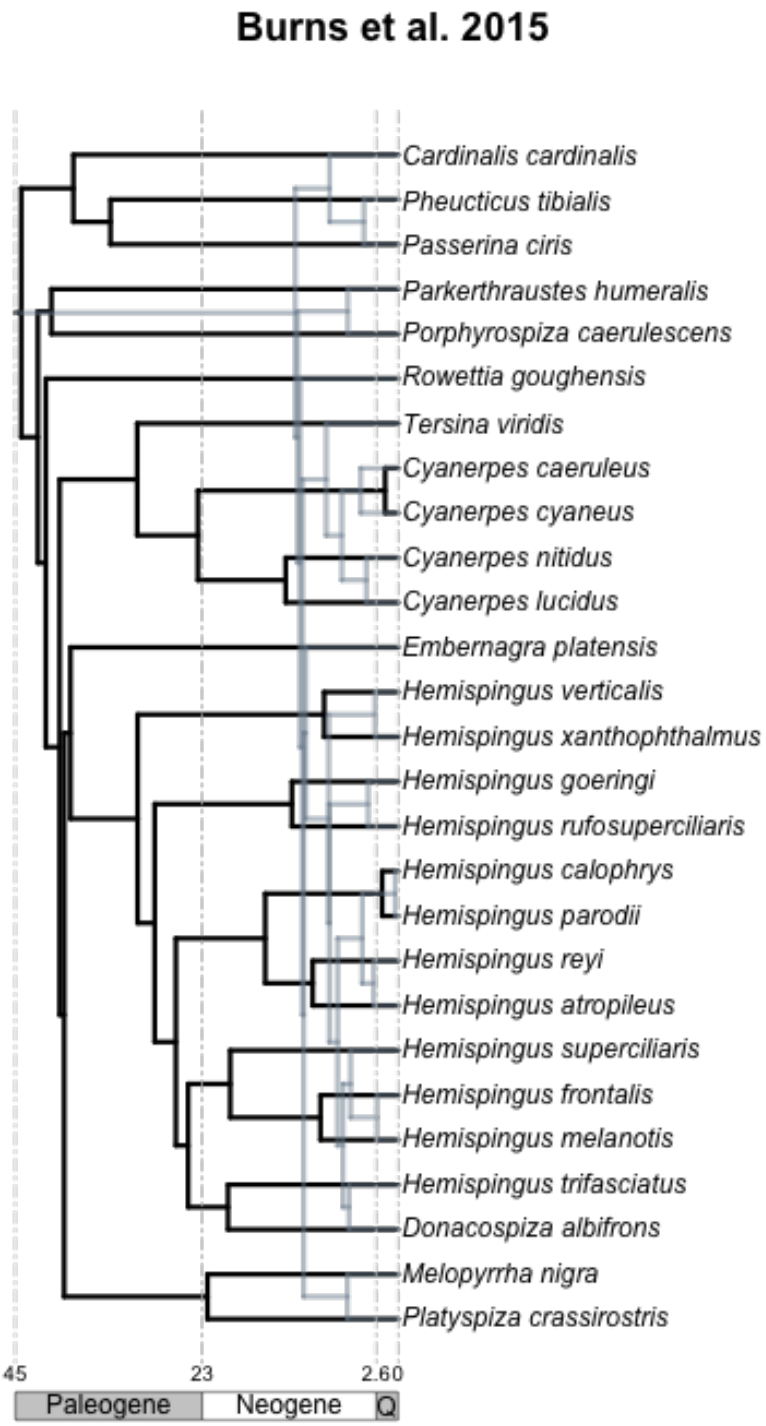



FIGURE 10. Cross validation of fourth source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to

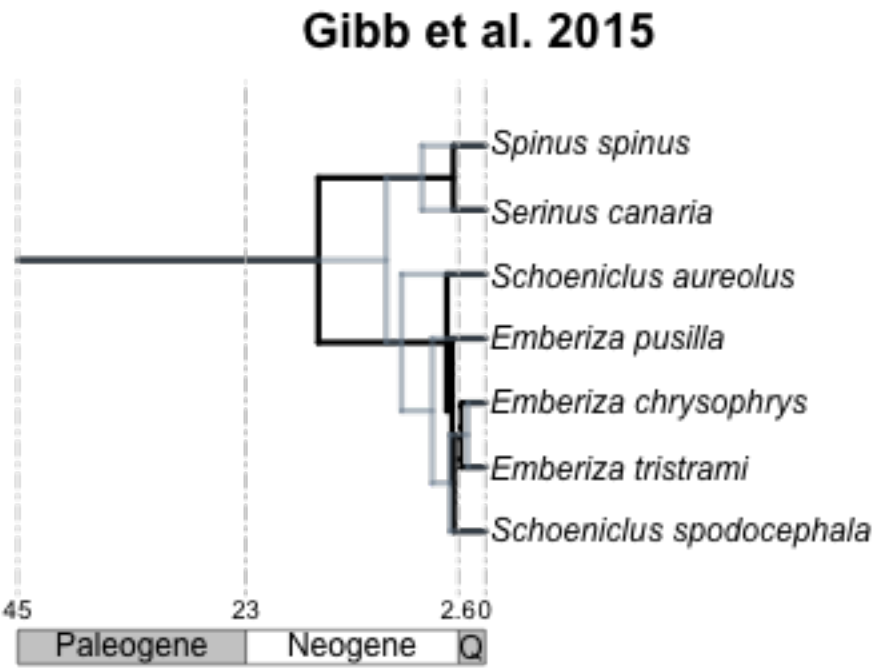



FIGURE 11. Cross validation of sixth source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to the same tree topology dated with BLADJ using node ages from all other source chronograms as secondary calibrations.

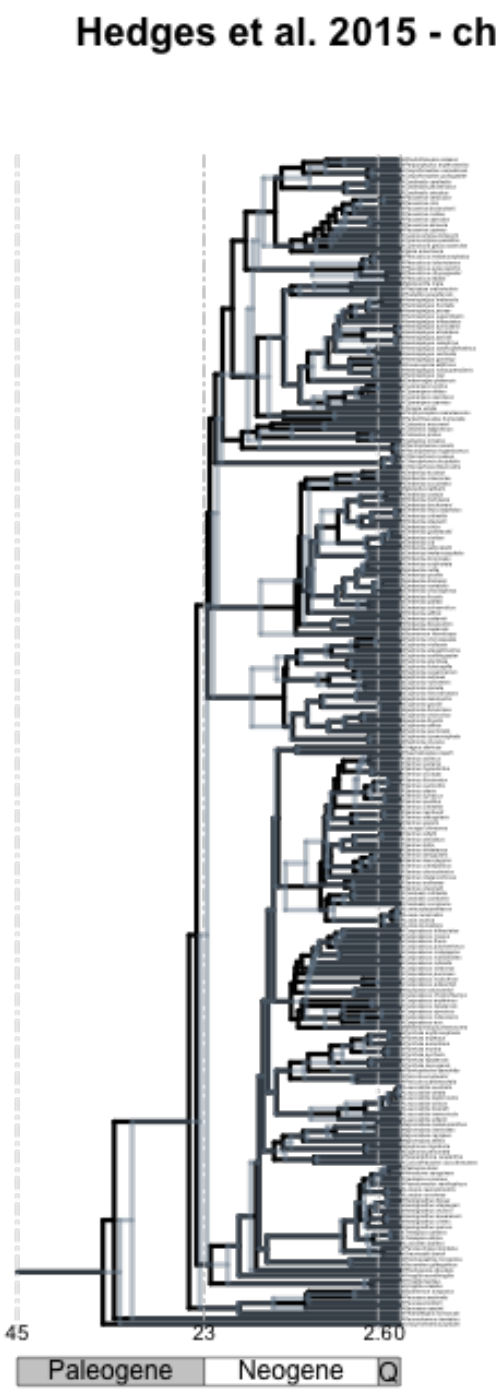



FIGURE 12. Cross validation of seventh source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to

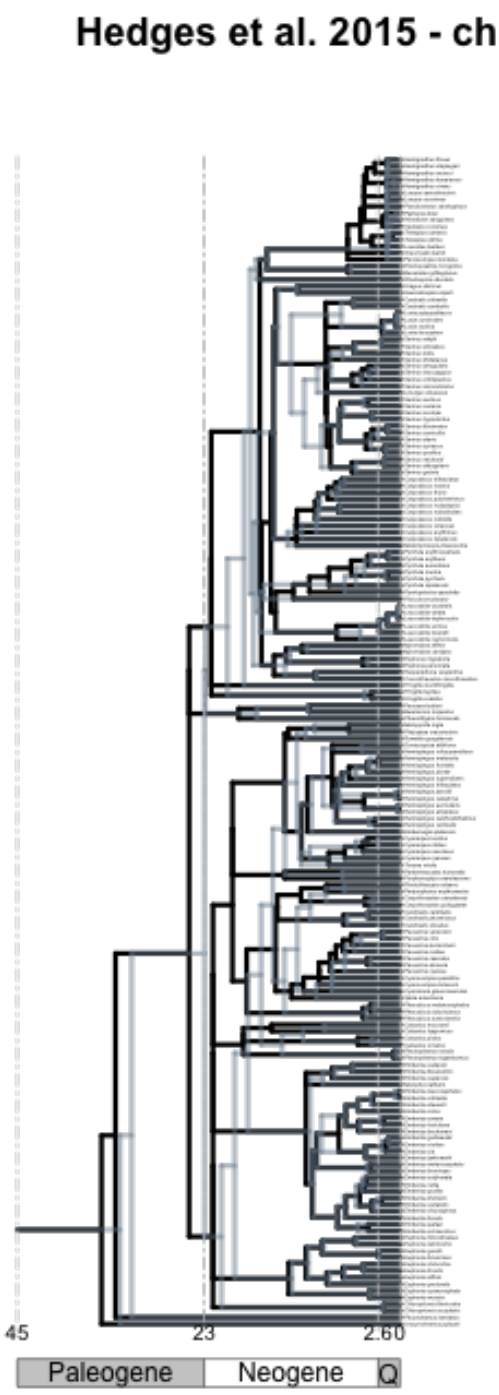



FIGURE 13. Cross validation of eight source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to

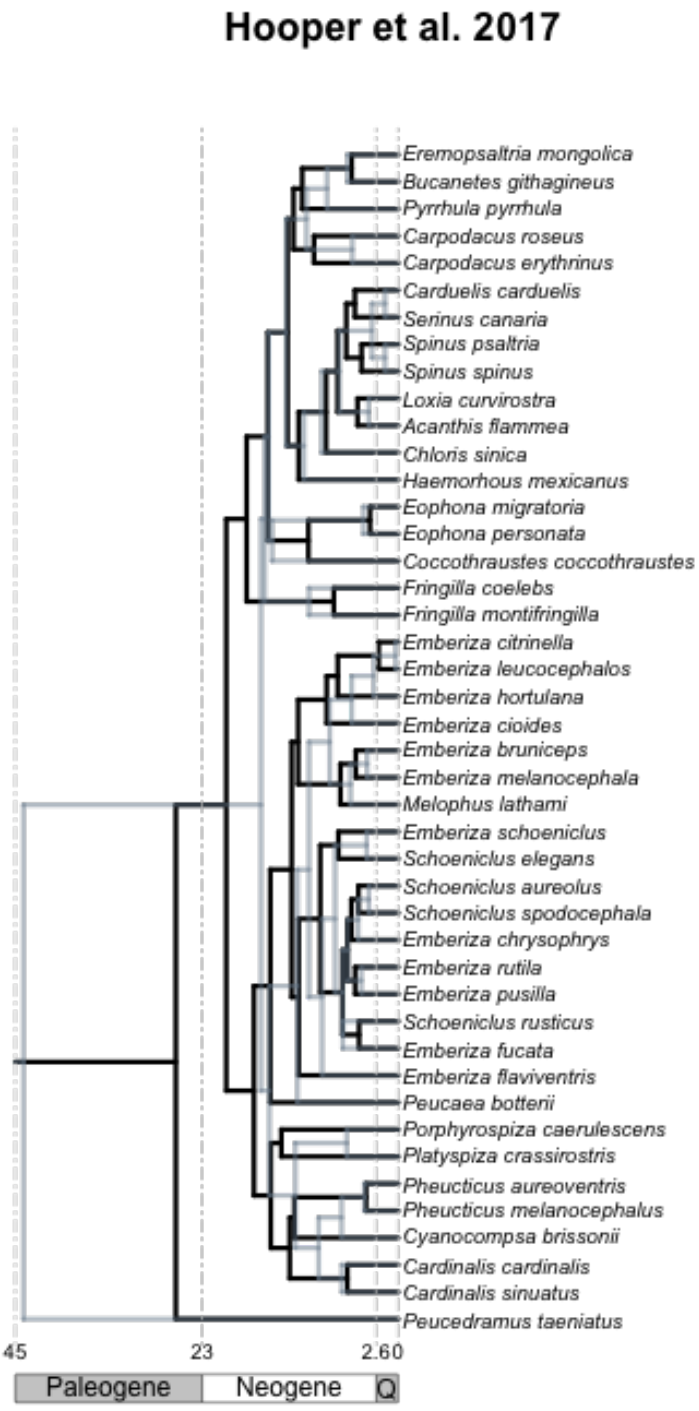



FIGURE 14. Cross validation of ninth source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to

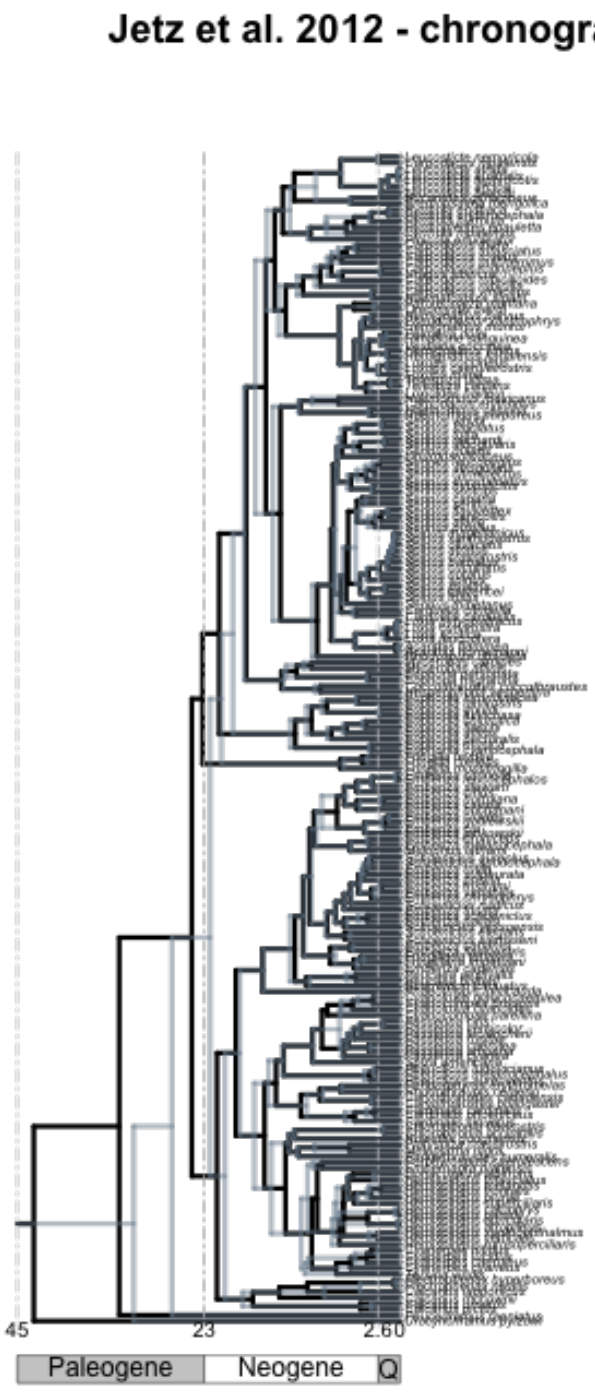



FIGURE 15. Cross validation of tenth source chronogram. The chronogram shown in black corresponds to the dates published in the original study. The gray chronogram corresponds to