

1 Running head: DATELIFE: REVEALING THE DATED TREE OF LIFE

2 Title: DateLife: Leveraging databases and analytical tools to reveal the dated Tree of Life

3 Authors: Luna L. Sánchez-Reyes¹, Brian C. O'Meara¹

4 Correspondence address:

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

7 Corresponding authors: sanchez.reyes.luna@gmail.com, bomeara@utk.edu

abstract.- The combination of new analytical techniques, availability of more fossil and molecular data, and better practices in data sharing has resulted in a steady accumulation of chronograms in public and open databases such as TreeBASE, Dryad, and Open Tree of Life, for a large quantity and diversity of organisms in the last decades. However, time of lineage divergence data remains difficult to be obtained and synthesized for many biologists and the non-academic community, despite its importance in many areas of research, education and science communication. **datelife** is a service implemented via an R package and a web site (<http://www.datelife.org/>) for efficient reuse, summary and reanalysis of expert, published data on time of lineage divergence. The main workflow starts with at least two taxon names as input, either as tip labels on a tree, or as a simple comma separated character string. A name search is then performed across the chronogram database and positively identified source trees are pruned to maintain queried taxa only and stored as a named list of patristic distance matrices. Source chronogram data can be summarised using branch length summary statistics or variance minimising approaches to generate a single summary chronogram. Source chronogram data can also be used as calibration points to date a tree containing some or all names from the initial query. If there is no information available for any queried taxa, data can be simulated. All source and summary chronograms can be saved in formats that permit easy reuse and reanalysis. Summary and newly generated trees are potentially useful to evaluate evolutionary hypothesis in different areas of research. How well this trees work for this purpose still needs to be tested. **datelife** will be useful to increase awareness on the existing variation in expert time of divergence data, and might 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.

Keywords: Tree; Phylogeny; Scaling; Dating; Ages; Divergence times; Open Science; Congruification; Supertree; Calibrations

Time of lineage divergence constitutes a fundamental piece of information for evolutionary understanding in many areas of research, from developmental to conservation biology (Felsenstein 1985; Webb 2000), from historical biogeography to species diversification studies (Posadas et al. 2006; Morlon 2014). The primary information needed for time of lineage divergence estimation comes from the fossil record. Coupled to molecular phylogenies (and more recently to morphological data), the time of divergence of extant and extinct lineages is reconstructed with molecular dating methods. Probably encouraged by the great developments in DNA sequencing techniques, phylogenetic inference and molecular dating methods, the number of studies publishing phylogenies with branch lengths proportional to geologic time (hereafter chronograms) have constantly increased in number for the last two decades (Kumar et al. 2017). Still, generating a chronogram is not an easy task unless you have specialized training. That’s why there has been an urge for promoting and facilitating reuse of the vast amount of phylogenetic and time of lineage divergence data that has been generated and made available in publications, for the advantage of research relying on this information (Webb and Donoghue 2005; Stoltzfus et al. 2013).

Wide interest from the scientific community to make information from phylogenies in general and chronograms in particular available for consultation and reuse has spurred the creation of public platforms with various goals and characteristics. TreeBASE (Morell 1996; Piel et al. 2002), the Dryad repository (<http://datadryad.org/>), and the Open Tree of Life (OToL; Hinchliff et al. 2015) platforms store and make available published phylogenies and chronograms for easy scientific reuse. All of them can be queried using automatised web procedures, which permit personalised, large scale queries that are also very fast. OToL stores trees with branch length information from a wide range of living organisms, implementing a metadata structure that stores the branch length units (i.e., time or relative substitution rates). Treebase and Dryad repositories also contain trees from all groups of life, but the former does not store branch length information and Dryad stores many other types of biological data using metadata that does not allow automatic distinction of types of trees and branch length units, impairing the automatised access to time of lineage divergence information.

Besides keeping a repository to easily store and share expert phylogenetic and chronogram knowledge, OToL also has the primary goal of synthesising all trees in their repository to expose to the community a single tree

of all life depicting the phylogenetic relationships among known lineages. All or parts of this synthetic tree can be reused for any purpose. However, it currently only focus on synthesizing tree topology, meaning that it does not expose branch length data of any type. The Timetree of Life project focuses on the synthesis of a single chronogram of life (Hedges et al. 2006). However, the thousands of chronograms they have compiled for synthesis are only publicly available for visual examination in their website or for download as images, but not for scientific reuse or reanalysis. The latest version of their synthetic chronogram (Kumar et al. 2017) can be queried only through their website in a non-automatised fashion, and only subsets of it can be reused for analyses with the permission of the authors. Other platforms such as SuperSmart (Antonelli et al. 2017) and phylogenerator (Pearse and Purvis 2013) are focused in automatised *de novo* chronogram inference, by reusing DNA sequence data to reconstruct phylogenetic trees. However, expert fossil information necessary for subsequent molecular dating analyses still needs to be compiled and curated by the user, rendering them a challenging tool to obtain data on time of lineage divergence for the non-specialist. Moreover, these tools do not provide information from already existent expert chronograms.

A tool for efficient reuse of expert, published data on time of lineage divergence should have an open and fully public chronogram database storing data in a format suitable for scientific reuse, an automatised way of accessing the information, and straightforward means of comparing and summarizing chronogram information as needed by the user. A prototype service aiming to meet this characteristics was developed over a series of hackathons at the National Evolutionary Synthesis Center (Stoltzfus et al. 2013). In here we present the formal description and implementation of the **datelife** service, constituted by an R package and a web site (<http://www.datelife.org/>). There is still much room for improvement, and flaws and limitations are addressed below. We strived for the current implementation of **datelife** to perform the basic tasks described above, featuring a system for maintenance of an open database of chronograms pulled from public repositories, methods to summarize and compare source chronograms, and new functions to visualize and graphically compare source and summary chronograms. R packages for benchmarking of functionalities and demonstrating services were also developed and made available.

DESCRIPTION

The basic **datelife** workflow is shown in figure 1 and consists of:

1. A user providing at least two taxon names as input, either as tip labels on a tree, or as a simple comma separated character string. The tree can be in newick or phylo format, and can be with or without branch lengths.
2. A name search is then performed across the chronogram database; source trees with at least two matching input names are identified; all other taxa that do not match the original query are then dropped from the positively identified source trees –these pruned chronograms are hereafter referred as source chronograms; finally, each source chronogram is transformed to a patristic matrix named by the citation of the original study. This format facilitates and greatly speeds up all further analyses and summarising algorithms.
3. The user can obtain different summary information from the source chronograms including: a) all source chronogram ages, b) maximum ages of source chronograms, c) citations of studies where source chronograms were originally published, d) a summary table with all of the above, e) a single summary tree of all or a subset of source chronograms, and f) a report of succesful matches of input taxon names across source chronograms. Summary information can be used to make decisions on the next steps of the workflow.
4. Source chronogram data can be used as calibration points to date a tree with or without branch lengths containing some or all names from the initial query.
5. If there is no information available for any queried taxa, users can also simulate both age and phylogenetic data for this missing taxa with a variety of algorithms described below.
6. Finally, users can easily save all source and summary chronograms in formats that permit easy reuse and reanalyses (newick and R ‘phylo’ format), as well as view and compare results graphically, or construct their own graphs using inbuilt ‘datelife’ graphic generation functions.

To gather, process, and present information, **datelife** builds up from functions available in several R

packages including `rotl` (Michonneau et al. 2016), `ape` (Paradis et al. 2004), `geiger` (Harmon et al. 2008),
`paleotree` (Bapst 2012), `bold` (Chamberlain 2018), `phytools` (Revell 2012), `taxize` (Chamberlain and Szöcs
2013; Chamberlain 2018), `phyloch` (Heibl 2008), `phylcomr` (Ooms and Chamberlain 2018) and `rphylotastic`
(O'Meara et al. 2019).

A **`datelife`** search currently accepts scientific names only. It can be any named clade or binomial specific.
Chronogram search is performed at the species level, so when input names correspond to named clades,
`datelife` pulls all accepted species names within the clade from OTOL's reference taxonomy to perform the
search. Searches at the infraspecies level are not currently allowed, so input names belonging to subspecies or
any other infraspecific category are collapsed to the species level. **`datelife`** processes input names with the
taxon name resolution service (TNRS; Boyle et al. 2013), which corrects potentially misspelled names and
typos, and standardizes spelling variations and synonyms, increasing the probability to correctly find the
queried taxa in **`datelife`**'s chronogram database.

The chronogram search is performed across **`datelife`**'s chronogram database which is assembled from OTOL's
tree repository. Compared to other existing open tree repositories OTOL's metadata rich tree store is the
only one that supports search, identification, and handling of chronograms in an automatised fashion. Also,
all their chronograms come from peer-reviewed published studies generated by specialists in the targeted
lineages, arguably representing expert knowledge on time of lineage divergence.

Information from source chronograms can be summarised with a summary statistic of tree branch lengths
(such as median or mean) or using the Super Distance Matrix (SDM) approach for supertree reconstruction
(Criscuolo et al. 2006). The resulting summary patristic distance matrix could be clustered with classic
algorithms. However, we noticed that the resulting trees are often non-ultrameric and do not reflect the source
chronogram data (see `datelife_examples` package). Instead, we obtained a distribution of age data from the
summary matrix available for nodes on a consensus tree. The Branch Length Adjuster (BLADJ) algorithm
(Webb et al. 2008) was then used to distribute node ages evenly over the consensus tree, minimizing age
variance in the resulting chronogram.

For tree dating, the congruification algorithm described by Eastman et al. (2013) is implemented to find shared nodes between trees (congruent nodes). The ages of these nodes are then used as calibrations to date any given tree. Currently implemented methods for tree dating are BLADJ, MrBayes (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003) and PATHd8 (Britton et al. 2007), a non-clock, rate-smoothing dating method.

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 increases roughly linearly with number of input taxon names, and increases considerably if TNRS service is activated. Up to ten thousand names can be processed and searched in less than 30 minutes. A name search through the chronogram database with an already processed query 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 locally –using 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 50% of **datelife**'s code (<https://codecov.io/gh/phylostatic/datelife>).

EXAMPLE

In this section we demonstrate the types of outputs that can be obtained with **datelife**, using as an example the bird family Fringillidae of true finches. We performed a higher-taxon search to obtain all data on lineage divergence available in **datelife**'s database for all recognised species within the Fringillidae (475 spp. according to the Open Tree of Life taxonomy). There are 13 chronograms containing at least two Fringillidae species, published in 9 different studies (Fig. 3). Data from these source chronograms was used to generate two types of summary chronograms, median and SDM. As explained in the **Description**, data from source chronograms was first summarised into a single distance matrix (using either the median or the SDM method) and then the available node ages were used as calibrations points over a consensus tree topology, to

obtain a dated tree with the program BLADJ (Fig. 4). Median summary chronograms are older and have wider variation in maximum ages than chronograms obtained with SDM. In both cases, ages are coherent with source ages. It is not certain if these chronograms should be used to perform all types of downstream evolutionary analyses and more research is needed in that direction.

Data from source chronograms was also used to date tree topologies with no branch length information and trees with branch lengths in relative substitution rates (Figs. 5 and 6). As a form of cross validation, we used tree topologies from each study and calibrated them using information from all other source chronograms. In the absence of branch length data, the ages of internal nodes were approximately recovered in almost all cases (except for studies 3, and 5; Fig. 5). Maximum tree ages were only approximately recovered in one case (study 2; Fig. 5). Branch lengths were successfully generated using the BOLD database for all source chronograms. However, dating with PATHd8 (using congruified calibrations) was only successful in three cases (studies 3, 5, and 9; Fig. 6). From these, two trees have a different sampling than the original source chronogram, mainly because DNA data for some species is absent from the BOLD. Maximum ages are quite different from source chronograms, but this might be explained also by the differences in sampling between source chronograms and BOLD trees. More examples and details can be consulted in <https://github.com/LunaSare/datelifexamples>.

FLAWS, LIMITATIONS AND PROSPECTS

The main goal of **datelife** is to make expert 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 a very fast tool that fulfills the quality of openness and does not require any expert biological knowledge from users –besides the names of the organisms they want to work with– for any of its functionalities. However, it has many flaws. Some of them can be overcome, some of them might represent limitations.

At the moment, **datelife**'s chronogram database is not very large, storing 231 chronograms up to the time the manuscript was written. This represents 5.79% of the largest existing chronogram database, which is

181 unfortunately not open for scientific reuse nor automatised data mining (Kumar et al. 2017). OToL is the
182 only public tree repository from where **datelife** can currently pull chronograms to construct its database.
183 Other open repositories are not suitable for time of divergence data mining because they either do not
184 store that data (i.e., TreeBASE), or their metadata does not have enough information to allow automatised
185 chronogram searches (i.e., Dryad). A previous version of TimeTree’s synthetic chronogram (Hedges et al.
186 2015) was made available in OToL repository, hence the amount of lineages represented in datelife’s database
187 is at least as substantial as TimeTree’s. This ensures that some information will be available for any given
188 query, but it does not ensure that the full state of knowledge of time of divergence data will be available for
189 any given lineage. Thus, incorporation of published chronograms available in the Dryad data repository and
190 in supplementary material of journals to **datelife**’s database is crucial to improve its services. Methods
191 to automatically mine chronogram data from the Dryad repository could be designed and implemented.
192 However, the unit of branch lengths would still need to be determined by “hand”. Consequently, we would
193 like to emphasize on the importance of sharing chronogram data for the scientific community, in repositories
194 that require expert input and manual curation, such as OToL’s tree repository.

195 Another potential concern comes from summary chronograms. We currently summarize all source chronogram
196 data by default. Users can subset source data if they have reasons to favor some or one source chronogram
197 over others. Strictly speaking, a good chronogram should reflect the real time of lineage divergence accurately
198 and precisely. To our knowledge, there is no objective way to determine if an expert chronogram is better
199 than other. Some criteria that have been put forward are the level of lineage sampling and the number of
200 calibrations used. Scientists usually also favor chronograms coming from studies with primary calibrations to
201 ones from secondary calibrations. It has been observed with simulations that divergence times inferred with
202 secondary calibrations are significantly younger than those inferred with primary calibrations in analyses
203 performed with bayesian inference methods when priors are implemented in similar ways in both analyses
204 (Schenk 2016). Yet, there are different ways to use secondary calibrations and the bias might not be
205 encountered with other dating methods that do not require prior assumptions (such as ML methods). But
206 this remains to be tested.

Furthermore, even chronograms obtained with primary data can be very different, as observed from the comparison of source chronograms in the Fringillidae example. A large discrepancy in time of lineage divergence across expert knowledge is well known for different groups of organisms (e.g., angiosperms; Magallón et al. 2015). Comparison of available chronogram data for a wide range of organisms shown here suggest that this is a widespread phenomenon that requires further attention. Characteristics of the data used for dating analyses as well as from the output chronogram itself, such as 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 confidence intervals could be used to score quality of source chronograms. To facilitate subsetting of source chronograms following different criteria, this information should be included as metadata manually entered by curators (as is done in OToL) in the future. Still, even if all source chronograms have been generated by excellent standards and using similar methods, the evolutionary history they depict might be very different. Hence, summarizing chronograms might imply summarizing evolutionary hypothesis. This could be good from certain point of view, since it could help to get a single global evolutionary history for a lineage. But it could also be bad, since we could be losing part of the evolutionary history that is only being reflected in some chronograms and not from the summary chronogram. Ideally, we should still rely on time of lineage divergence data obtained from a single analysis using fossil data as primary sources of calibrations, and using fossils that have already been curated as calibrations to date other trees, which should reflect a more homogeneous evolutionary history (Antonelli et al. 2017). This will be implemented in future **datelife** versions.

Synthetic or summary chronograms are being used more and more frequently to study evolutionary hypothesis. We introduce different ways of summarizing time of lineage divergence data, but, are any of these summary chronograms reliable to study evolutionary patterns? The short answer is that we do not know to what extent the summary trees should be used for reanalyses and if there are areas of research where they can be used more reliably than others. There is evidence that different chronograms will result in very different macroevolutionary histories inferred from them (Title and Rabosky 2016). In other areas of biological research, such as ecology and conservation biology, it has been indicated that at least some data on lineage divergence represents an improvement for testing alternative hypothesis using phylogenetic distance. It is uncertain,

however, how using very different age data could affect the outcomes of these tests, and further research is necessary in this direction.

CONCLUSIONS

Time of lineage divergence information is key to many areas of evolutionary studies: trait evolution, diversification, biogeography, macroecology and more. Generating this information 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 to produce chronograms *de novo*. Knowledge on taxon ages is also crucial for science communication and education.

datelife allows an easy and fast obtention, as well as comparison of publicly available information on time of lineage divergence, providing 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, allowing identification of regions that require more research or that have conflicting information. Both summary and newly generated trees are potentially useful to evaluate evolutionary hypothesis in different areas of research. How well this trees work for this purpose still needs to be further tested. If all, **datelife** helps with awareness on the existing variation in expert time of divergence data, and might 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

259 `devtools::install_github("phylotastic/datelife").`

260 SUPPLEMENTARY MATERIAL

261 Code used to generate all versions of this manuscript, the biological examples, as well as the benchmark
262 of functionalities can be found in GitHub repositories at https://github.com/LunaSare/datelife_paper1,
263 https://github.com/LunaSare/datelife_examples, and https://github.com/LunaSare/datelife_benchmark,
264 respectively.

265 FUNDING

266 Funding was provided by the US National Science Foundation (NSF) grant 1458603

267 NESCent

268 Open Tree of Life

269 University of Tennessee, Knoxville

270 ACKNOWLEDGEMENTS

271 We thank colleagues from the O'Meara Lab at the University of Tennessee Knoxville for suggestions, discus-
272 sions and software testing. The late National Evolutionary Synthesis Center (NESCent), which sponsored
273 hackathons that led to initial work on this project. The Open Tree of Life project that provides the open,
274 metadata rich repository of trees used for **datelife**. The many scientists who publish their chronograms in
275 an open, reusable form, and the scientists who curate them for deposition in the Open Tree of Life repository.
276 The NSF for funding nearly all the above, in addition to the ABI grant that funded this project itself.

- 278 Antonelli A., Hettling H., Condamine F.L., Vos K., Nilsson R.H., Sanderson M.J., Sauquet H., Scharn
279 R., Silvestro D., Töpel M., Bacon C.D., Oxelman B., Vos R.A. 2017. Toward a self-updating platform for
280 estimating rates of speciation and migration, ages, and relationships of Taxa. *Systematic Biology*. 66:153–166.
- 281 Bapst D.W. 2012. Paleotree: An R package for paleontological and phylogenetic analyses of evolution.
282 *Methods in Ecology and Evolution*. 3:803–807.
- 283 Barker F.K., Burns K.J., Klicka J., Lanyon S.M., Lovette I.J. 2012. Going to extremes: Contrasting rates of
284 diversification in a recent radiation of new world passerine birds. *Systematic biology*. 62:298–320.
- 285 Barker F.K., Burns K.J., Klicka J., Lanyon S.M., Lovette I.J. 2015. New insights into new world biogeography:
286 An integrated view from the phylogeny of blackbirds, cardinals, sparrows, tanagers, warblers, and allies. *The*
287 *Auk: Ornithological Advances*. 132:333–348.
- 288 Boyle B., Hopkins N., Lu Z., Raygoza Garay J.A., Mozzherin D., Rees T., Matasci N., Narro M.L., Piel
289 W.H., McKay S.J., Lowry S., Freeland C., Peet R.K., Enquist B.J. 2013. The taxonomic name resolution
290 service: An online tool for automated standardization of plant names. *BMC Bioinformatics*. 14.
- 291 Britton T., Anderson C.L., Jacquet D., Lundqvist S., Bremer K. 2007. Estimating Divergence Times in Large
292 Phylogenetic Trees. *Systematic Biology*. 56:741–752.
- 293 Burns K.J., Shultz A.J., Title P.O., Mason N.A., Barker F.K., Klicka J., Lanyon S.M., Lovette I.J. 2014.
294 Phylogenetics and diversification of tanagers (passeriformes: Thraupidae), the largest radiation of neotropical
295 songbirds. *Molecular Phylogenetics and Evolution*. 75:41–77.
- 296 Chamberlain S. 2018. bold: Interface to Bold Systems API..
- 297 Chamberlain S.A., Szöcs E. 2013. taxize : taxonomic search and retrieval in R [version 2; referees: 3 approved].
298 F1000Research. 2:1–29.

299 Claramunt S., Cracraft J. 2015. A new time tree reveals earth history’s imprint on the evolution of modern
300 birds. *Science advances*. 1:e1501005.

301 Criscuolo A., Berry V., Douzery E.J., Gascuel O. 2006. SDM: A fast distance-based approach for (super)tree
302 building in phylogenomics. *Systematic Biology*. 55:740–755.

303 Eastman J.M., Harmon L.J., Tank D.C. 2013. Congruification: Support for time scaling large phylogenetic
304 trees. *Methods in Ecology and Evolution*. 4:688–691.

305 Felsenstein J. 1985. Phylogenies and the Comparative Method. *The American Naturalist*. 125:1–15.

306 Gibb G.C., England R., Hartig G., McLenachan P.A., Taylor Smith B.L., McComish B.J., Cooper A., Penny
307 D. 2015. New zealand passerines help clarify the diversification of major songbird lineages during the oligocene.
308 *Genome biology and evolution*. 7:2983–2995.

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

311 Hedges S.B., Dudley J., Kumar S. 2006. TimeTree: A public knowledge-base of divergence times among
312 organisms. *Bioinformatics*. 22:2971–2972.

313 Hedges S.B., Marin J., Suleski M., Paymer M., Kumar S. 2015. Tree of life reveals clock-like speciation and
314 diversification. *Molecular Biology and Evolution*. 32:835–845.

315 Heibl C. 2008. PHYLOCH: R language tree plotting tools and interfaces to diverse phylogenetic software
316 packages..

317 Hinchliff C.E., Smith S.A., Allman J.F., Burleigh J.G., Chaudhary R., Coghill L.M., Crandall K.A., Deng J.,
318 Drew B.T., Gazis R., Gude K., Hibbett D.S., Katz L.A., Laughinghouse H.D., McTavish E.J., Midford P.E.,
319 Owen C.L., Ree R.H., Rees J.A., Soltis D.E., Williams T., Cranston K.A. 2015. Synthesis of phylogeny and
320 taxonomy into a comprehensive tree of life. *Proceedings of the National Academy of Sciences*. 112:12764–12769.

321 Hooper D.M., Price T.D. 2017. Chromosomal inversion differences correlate with range overlap in passerine
322 birds. *Nature ecology & evolution*. 1:1526.

323 Huelsenbeck J.P., Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*.
324 17:754–755.

325 Jetz W., Thomas G., Joy J.J., Hartmann K., Mooers A. 2012. The global diversity of birds in space and
326 time. *Nature*. 491:444–448.

327 Kumar S., Stecher G., Suleski M., Hedges S.B. 2017. TimeTree: A Resource for Timelines, Timetrees, and
328 Divergence Times. *Molecular biology and evolution*. 34:1812–1819.

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

331 Michonneau F., Brown J.W., Winter D.J. 2016. rotl: an R package to interact with the Open Tree of Life
332 data. *Methods in Ecology and Evolution*. 7:1476–1481.

333 Morell V. 1996. The roots of phylogeny. *Science*. 273:569.

334 Morlon H. 2014. Phylogenetic approaches for studying diversification. *Ecology Letters*. 17:508–525.

335 O’Meara B., Md Tayeen A.S., Sanchez Reyes L.L. 2019. Rphylotastic: An r interface to ‘phylotastic’ web
336 services..

337 Ooms J., Chamberlain S. 2018. Phylocomr: Interface to ‘phylocom’..

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

340 Pearse W.D., Purvis A. 2013. PhyloGenerator: An automated phylogeny generation tool for ecologists.
341 *Methods in Ecology and Evolution*. 4:692–698.

342 Piel W.H., Donoghue M., Sanderson M. 2002. TreeBASE : A database of phylogenetic information. In:
343 Shimura J., Wilson K., Gordon D., editors. To the interoperable “catalog of life” with partners. Tsukuba,
344 Japan: National Institute for Environmental Studies. p. 41–47.

345 Posadas P., Crisci J.V., Katinas L. 2006. Historical biogeography: A review of its basic concepts and critical
346 issues. *Journal of Arid Environments*. 66:389–403.

347 Price T.D., Hooper D.M., Buchanan C.D., Johansson U.S., Tietze D.T., Alström P., Olsson U., Ghosh-Harihar
348 M., Ishtiaq F., Gupta S.K., others. 2014. Niche filling slows the diversification of himalayan songbirds.
349 *Nature*. 509:222.

350 R Core Team. 2018. R: a language and environment for statistical computing. Vienna, Austria: R Foundation
351 for Statistical Computing.

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

354 Ronquist F., Huelsenbeck J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models.
355 *Bioinformatics*. 19:1572–1574.

356 Schenk J.J. 2016. Consequences of secondary calibrations on divergence time estimates. *PLoS ONE*. 11.

357 Stoltzfus A., Lapp H., Matasci N., Deus H., Sidlauskas B., Zmasek C.M., Vaidya G., Pontelli E., Cranston
358 K., Vos R., Webb C.O., Harmon L.J., Pirrung M., O’Meara B., Pennell M.W., Mirarab S., Rosenberg M.S.,
359 Balhoff J.P., Bik H.M., Heath T.A., Midford P.E., Brown J.W., McTavish E.J., Sukumaran J., Westneat M.,
360 Alfaro M.E., Steele A., Jordan G. 2013. Phylotastic! Making tree-of-life knowledge accessible, reusable and
361 convenient. *BMC Bioinformatics*. 14.

362 Title P.O., Rabosky D.L. 2016. Do Macrophylogenies Yield Stable Macroevoolutionary Inferences? An
363 Example from Squamate Reptiles. *Systematic Biology*::syw102.

364 Webb C.O. 2000. Exploring the Phylogenetic Structure of Ecological Communities : An Example for Rain

- 365 Forest Trees. *The American Naturalist*. 156:145–155.
- 366 Webb C.O., Ackerly D.D., Kembel S.W. 2008. Phylocom: Software for the analysis of phylogenetic community
367 structure and trait evolution. *Bioinformatics*. 24:2098–2100.
- 368 Webb C.O., Donoghue M.J. 2005. Phylomatic: Tree assembly for applied phylogenetics. *Molecular Ecology*
369 *Notes*. 5:181–183.

FIGURE 1

Stylized DateLife workflow. This shows the general workflows and analyses that can be performed with **datelife**, via the R package or through the website. 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 Resoulution Service (TNRS; dark gray), and using TNRS (light gray). We also performed a search using the already processed query for comparison (light blue).

FIGURE 3

Lineage through time (LTT) plots of source chronograms containing all or a subset of species from the bird family Fringillidae of true finches. Arrows indicate maximum age of each chronogram. Numbers reference to chronograms' original publications 1: Barker et al. (2012), 2: Barker et al. (2015), 3: Burns et al. (2014), 4: Claramunt and Cracraft (2015), 5: Gibb et al. (2015), 6: Hedges et al. (2015), 7: Hooper and Price (2017), 8: Jetz et al. (2012), 9: Price et al. (2014).

FIGURE 4

LTT plots of median and Supermatrix Distance Method (SDM) chronograms summarising information from source chronograms found for the Fringillidae. Arrows indicate maximum age.

FIGURE 5

LTT plots showing results from the cross-validation analyses of trees without branch lengths dated using BLADJ. The dating analysis can only be performed in trees with more than 2 tips, thus excluding chronogram from study 4; its data was still used as calibration for the other source chronograms.

393 LTT plots showing results from the cross-validation analyses of trees with branch length reconstructed with
394 data from the Barcode of Life Database (BOLD) dated using PATHd8. We could construct a tree with
395 branch lengths for all source chronograms. However, dating with PATHd8 was only successful in three source
396 chronograms shown here.

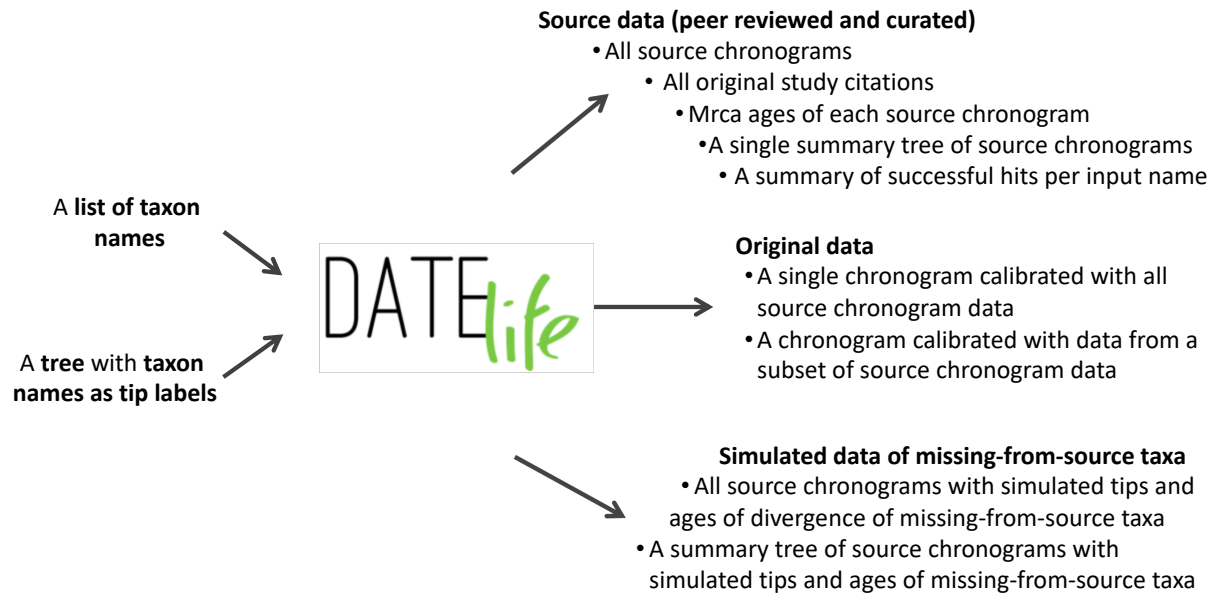


Figure 1:

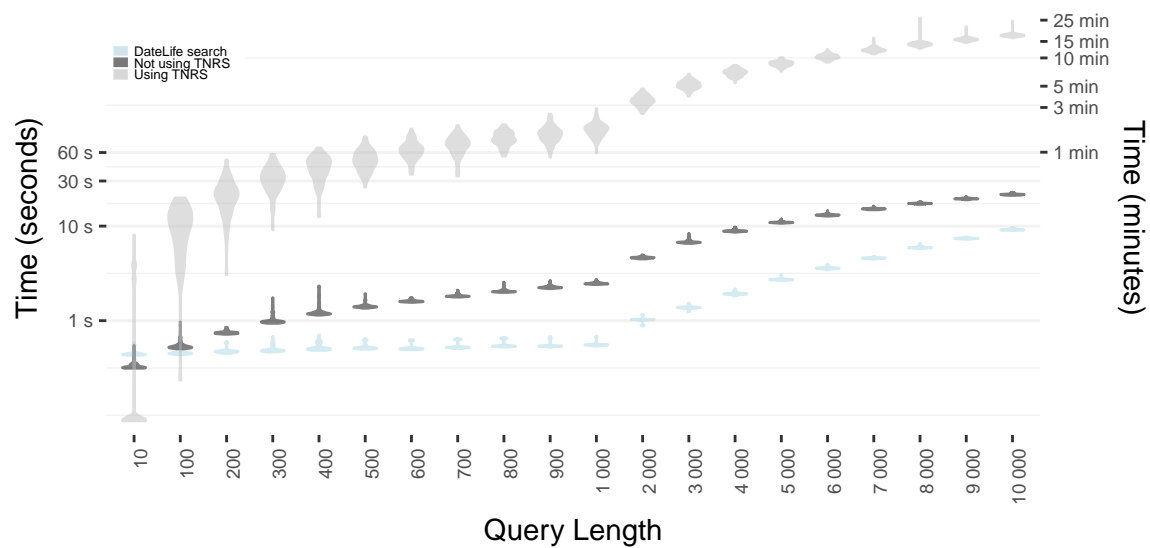


Figure 2:

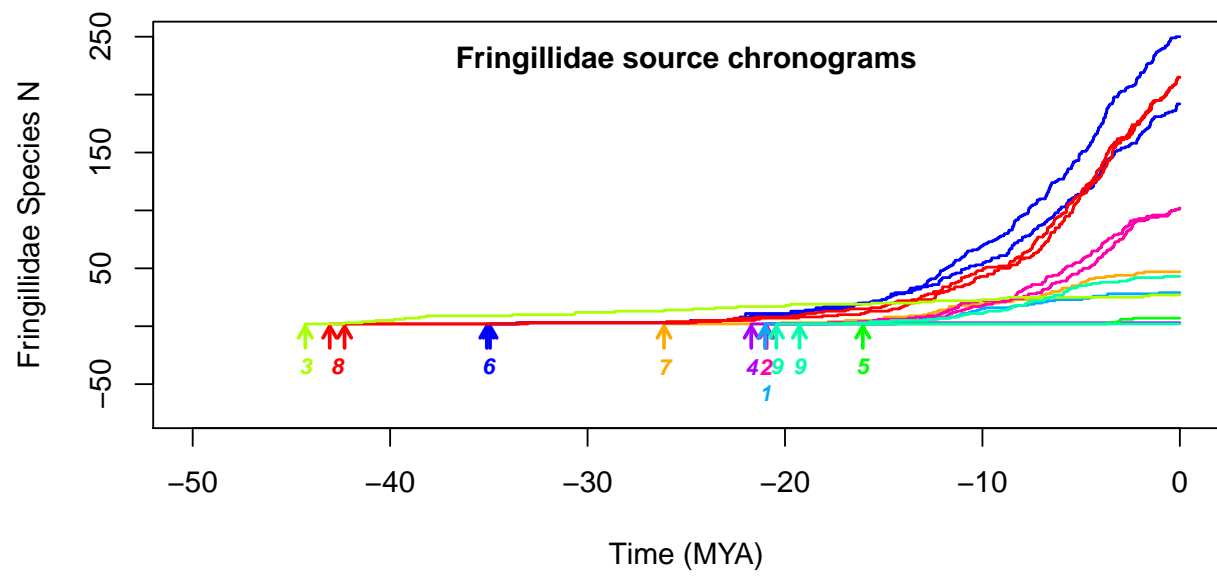


Figure 3:

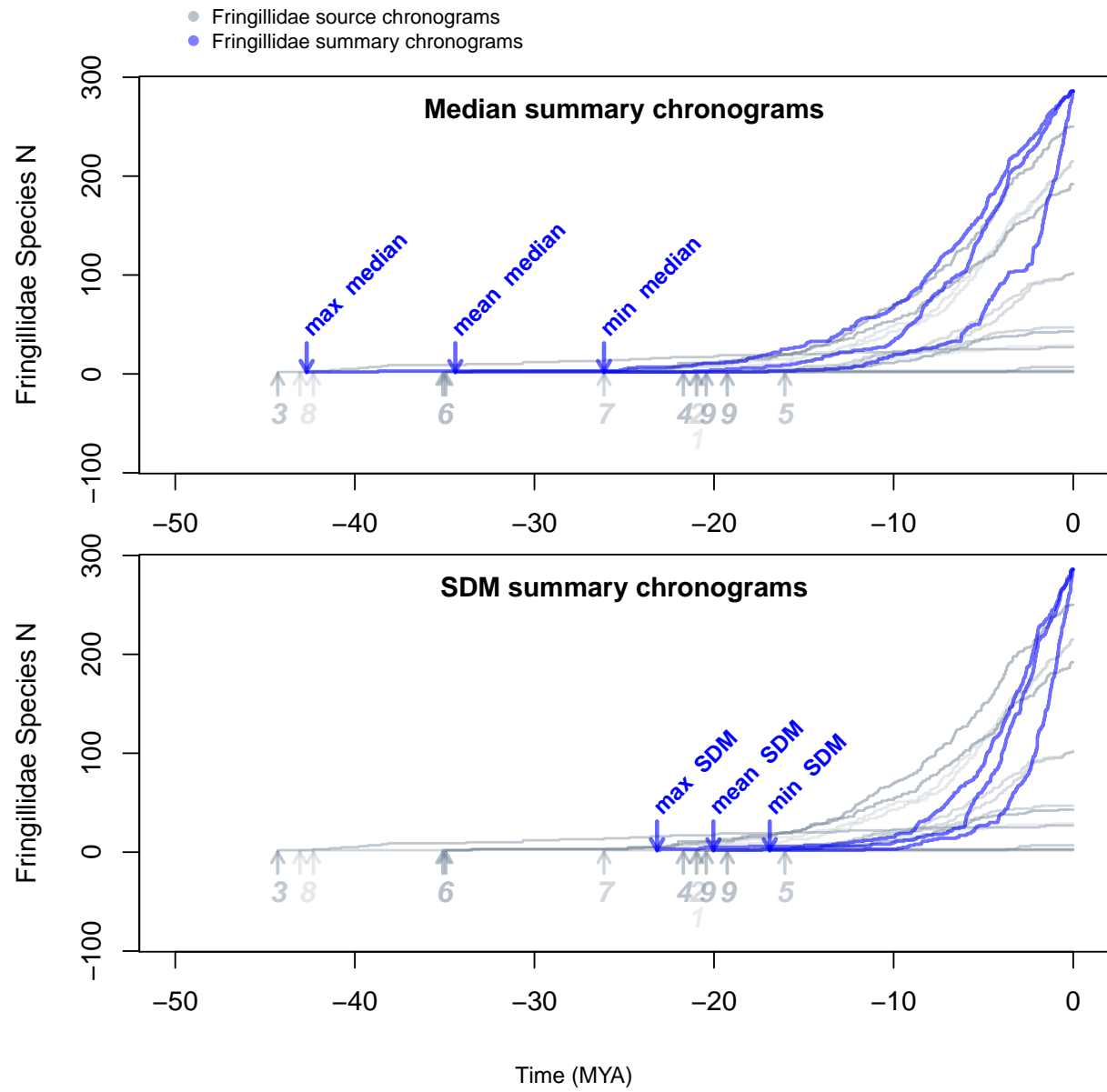


Figure 4:

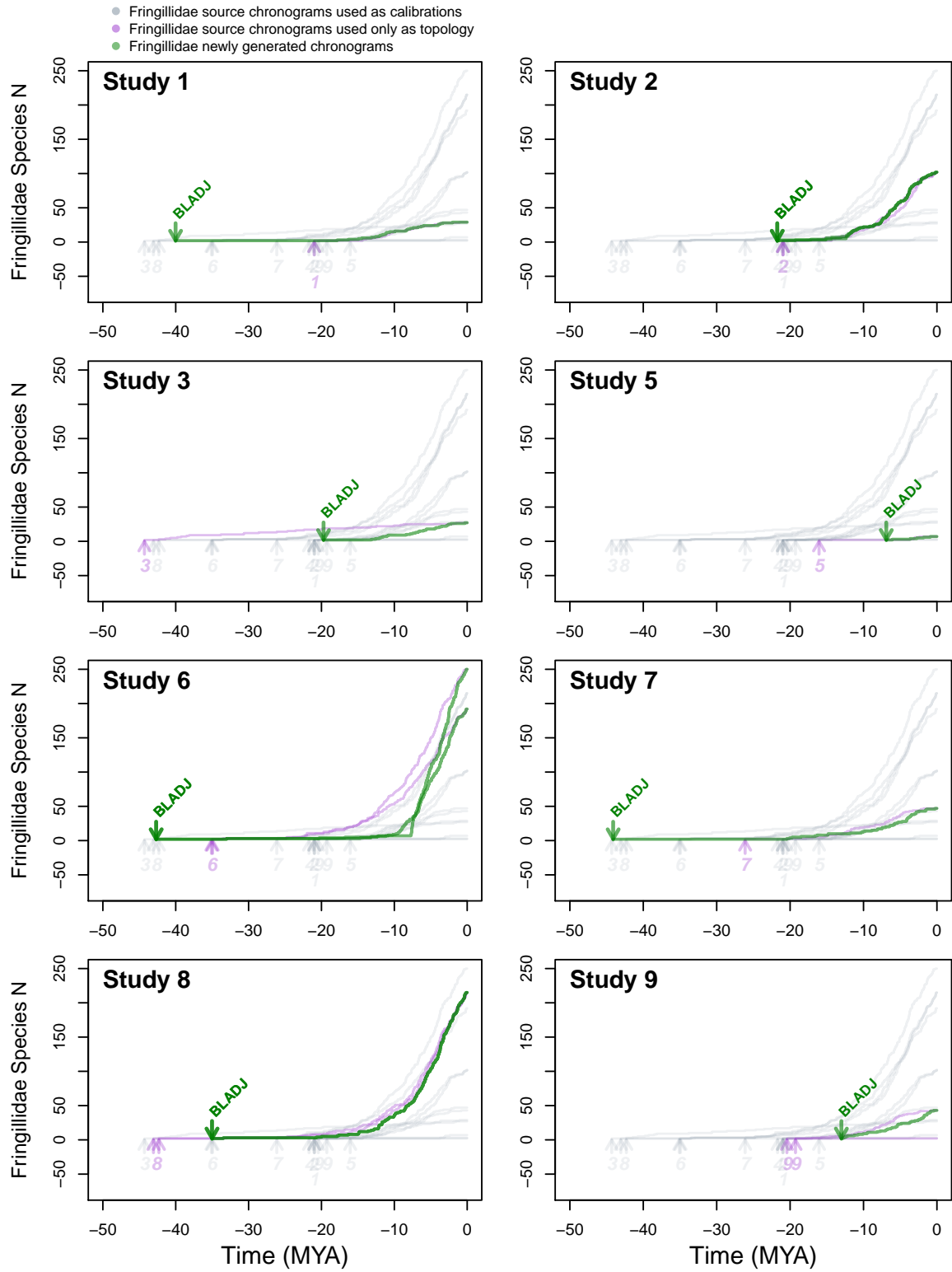


Figure 5:

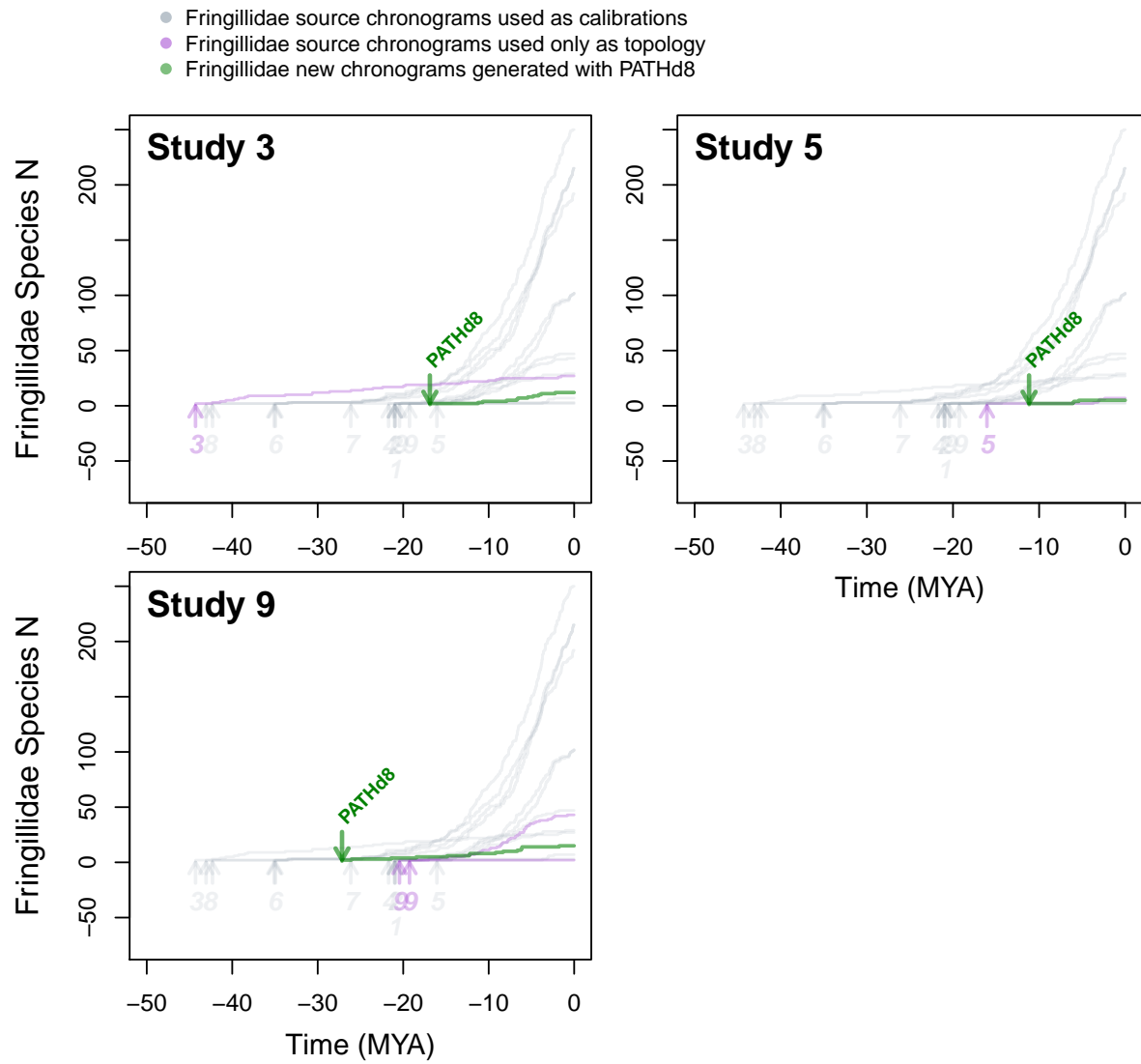


Figure 6: