

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

8 **abstract.-** The combination of new analytical techniques, availability of more fossil and molecular data, and
9 better practices in data sharing has resulted in a steady accumulation of chronograms in public and open
10 databases such as Dryad, TreeBASE and Open Tree of Life, for a large quantity and diversity of organisms.
11 However, this information remains difficult to synthesize for many biologists and the non-academic community.

12 **Keywords:** Tree; Phylogeny; Scaling; Dating; Ages; Divergence times; Open Science; Congruification;
13 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, the time of divergence of extant 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 and still accumulating information that is already available, for the advantage of researchers not specialized in this area (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 are committed to storing and making 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's stores trees with branch lengths from a wide range of living organisms, using a very good metadata structure. 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, and their metadata does not allow automatic distinction between types of trees, impairing the automatised access to time of lineage divergence information.

Besides storing public phylogenetic knowledge, OToL also has the primary goal of exposing a single tree of all life synthesised from trees in their repository, to show phylogenetic relationships of all known lineages.

All or parts of OToL’s synthetic tree can be reused for any purpose; however, it currently only focus on tree topology, meaning that it does not incorporate information on time of divergence from their stored chronograms. Phylomatic (Webb and Donoghue 2005) is a platform that synthesizes a tree containing taxa of interest from a user supplied set of trees or from its own database. It has the advantage that it allows grafting taxa of interest absent from their database, but it is limited to mostly seed plants and mammals. If your query does not match its records, it gives many polytomies back. The Timetree of Life project focuses on the synthesis of a single chronogram of life (Hedges et al. 2006). However, the thousands of chronograms 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. Also, 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 source database storing data in a format suitable for scientific reuse, an automatised way of accessing the information, and straightforward means of comparing and summarizing the source data as needed by the user. A prototype service aiming to meet this characteristics was initially developed over a series of hackathons at the National Evolutionary Synthesis Center (Stoltzfus et al. 2013). In here we present the first 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, new functions to visualize and graphically compare source and summary chronograms, as well as tools for comparison of source and summary trees. R packages for benchmarking of functionalities and exemplifying services were also developed.

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) **datelife** then performs a search across its database of peer reviewed and curated chronograms; identifies and gets source trees with at least two matching input names; drops unmatched taxa from positively identified source trees; and finally transforms each source tree to a patristic matrix named by the citation of the original study. This format facilitates and greatly speeds up all further analyses and summarization algorithms.
- 3) The user can obtain different types of summaries from the source data including: a) all source chronograms, b) mrca 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 source chronograms, and f) a report of successful matches per input taxon name across source chronograms.
- 4) At this point, users can choose to use all or some source data as calibration points to date a tree of their own making or choosing.
- 5) Users can also simulate age and/or phylogenetic data of input taxa not found in the database. A variety of algorithms are available for this purpose.
- 6) Finally, users can easily view results graphically as well as construct their own graphs using inbuilt **datelife** graphic generators.

datelife's chronogram database is currently built from OTOL's tree repository. Among currently existing repositories (i.e., TreeBASE, Dryad), OTOL's metadata rich tree store is the only one meeting the requirements

for automatized handling of source trees. TreeBASE trees are stored without branch lengths. Dryad metadata does not allow differentiating between different types of data sets, and trees would need to be manually curated anyways. **datelife** currently accepts scientific names only. It can be any named clade or binomial specific. Chronogram searches are 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. Currently, searches at the infraspecies level are not allowed, so input names belonging to subspecies or any other infraspecific category are collapsed to the species level. **datelife** also 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.

Source chronogram summary tree can be assembled using the Super Distance Matrix (SDM) supertree construction approach (Criscuolo et al. 2006) or using the median of branch lengths. Tree dating and simulation options are performed with various algorithms: Branch Length Adjuster (BLADJ) is a simple algorithm to distribute ages of undated nodes evenly, which minimizes age variance in the chronogram (Webb et al. 2008). PATHd8 is a non-clock, rate-smoothing method (Britton et al. 2007) to date trees. treePL, is a semi-parametric, rate-smoothing, penalized likelihood dating method (Smith and O’Meara 2012). MrBayes (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003) can be used when adding taxa at random, following a reference taxonomy or a topological constraint. It draws ages from a pure birth model, as implemented by Jetz and collaborators (2012). To apply calibrations to a tree, the congruification algorithm described by Eastman et al. (2013) is implemented to find shared nodes between trees (congruent nodes).

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), phylocomr (Ooms and Chamberlain 2018) and rphylotastic (O’Meara et al. 2019).

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 can be used to perform downstream evolutionary analyses. There is currently wide interest in determining this. However, we know that these chronograms are useful for...

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/dateliflife_examples.

FLAWS, LIMITATIONS AND PROSPECTS

The main goal of **dateliflife** 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.

There is only one public repository that has the metadata that **dateliflife** needs to perform its tasks and services. Other databases are not suitable for data mining because of their metadata structure. OTOL's tree database is very large for phylogenetic trees, but only stores a little bit more than 200 chronograms. This database is maintained by the scientific community, so the prospect is to keep inviting and doing work encouraging scientists to make their data sets available in this database.

We introduce different ways of summarizing this data, but are these summary chronograms reliable to study evolutionary patterns? The short answer is that we do not know to what extent the summary trees can be used for reanalyses and in which areas they can be used more reliably than others. Surely they might be

useful for ecological analyses, but further studies should be conducted.

As implemented now, we currently by default summarize all source data available. Users can subset source data if they have reasons to favor some or one source chronogram over others. To our knowledge there is no objective way to determine if an expert chronogram is better than other. Some criteria that have been put forward are the level of sampling and the number of calibrations used. Scientists usually also favor chronograms coming from studies with primary calibrations to ones from secondary calibrations. It would be important to evaluate how different source chronograms affect summary chronograms, in order to determine if any type of bias is introduced by certain types of chronograms in particular, so this chronograms are not included in the summary. A first step in that direction would be to add data on chronograms that come from primary or secondary calibrations, and maybe the number of calibrations used. This will need to be manually curated.

Even if all source chronograms have been generated by excellent standards and practices, the evolutionary context they depict might be very different. The ideal would be to use dates that were obtained from the same DNA context, which should reflect a more homogeneous evolutionary history or hypothesis (Antonelli et al. 2017). Summarizing trees implies summarizing evolutionary hypothesis.

However, the comparison of source trees in the example above shows that even chronograms obtained with primary data can be very different. There is some evidence that different chronograms will result in very different macroevolutionary histories inferred from them (Title and Rabosky 2016). In other areas of research, such as ecological and conservation biology, it has been demonstrated that the presence of data on lineage divergence represents an improvement for testing alternative hypothesis, compared to testing in a total absence of such time data. However, we do not know how using wrong or different kinds of age data will affect the outcomes of ecological tests. More work has to be performed in this area. The comparison of available data has made evident how much discrepancy in time of lineage divergence can be found across expert knowledge. This has been shown for angiosperms previously (Magallón et al. 2015), but our examples show that this is a widespread phenomenon that might require further attention. In the case of the example presented here, most chronograms are new, meaning that we expect that data was generated with high quality and software used was very advanced. However, this is just an assumption. Scores on source data

used (quality of alignment: missing data, GC content; tip sampling strategy; phylogenetic inference method; dating inference method; numer of fossils used) But we should also be able to choose on

Improvements: * fossils as calibrations: Using secondary calibrations can generate biased ages when using bayesian methods, mainly because we don't know what prior to give to secondary calibrations (Schenk 2016). * bayesian congruification * topological congruification

CONCLUSIONS

Time of lineage divergence data are key to many areas of evolutionary studies: trait evolution, diversification, biogeography, macroecology and more. Obtaining these ages 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 timeframe of different regions of the tree of life, allowing identification of regions that need more work or that have conflicting information. It can also summarize public data, or use it to generate new chronograms based on trees provided by the user. Both summary and newly generated trees have the potential to be useful to evaluate evolutionary hypothesis in different areas of research. How well this trees work for this purpose still needs to be tested. **datelife** helps with awereness on the variation of expert data, and the actual effect of the underlying data on the results of analyses, nurturing a culture of more cautious interpretation of evolutionary results.

datelife is also part of the reproducibility of science movement. It strives to facilitate usage of public data sets as well as encouraging the making of information publicly available. Acknowledging source information and describing workflows and analyses.

it allows exploration of alternative taxonomic and phylogenetic schemes; it allows rapid exploration of the

effect of alternative divergence time hypothesis; it allows rapid synthesis in a number of different formats;

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 with 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 can be found in GitHub repositories at https://github.com/LunaSare/datelife_paper1, https://github.com/LunaSare/datelife_examples, and https://github.com/LunaSare/datelife_benchmark, respectively.

FUNDING

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

NESCent

Open Tree of Life

University of Tennessee, Knoxville

ACKNOWLEDGEMENTS

We thank colleagues from the O'Meara Lab at the University of Tennessee Knoxville for suggestions, discussions and software testing. The late National Evolutionary Synthesis Center (NESCent), which sponsored

240 hackathons that led to initial work on this project. The Open Tree of Life project that provides the open,
241 metadata rich repository of trees used for **datelife**. The many scientists who publish their chronograms in
242 an open, reusable form, and the scientists who curate them for deposition in OpenTree. The NSF for funding
243 nearly all the above, in addition to the ABI grant that funded this project itself.

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

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

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

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

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

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

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

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

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

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

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

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

290 Huelsenbeck J.P., Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*.
291 17:754–755.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

324 Smith S.A., O’Meara B.C. 2012. TreePL: Divergence time estimation using penalized likelihood for large
325 phylogenies. *Bioinformatics*. 28:2689–2690.

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

331 Title P.O., Rabosky D.L. 2016. Do Macrophylogenies Yield Stable Macroevo-lutionary Inferences? An

- 332 Example from Squamate Reptiles. *Systematic Biology*:syw102.
- 333 Webb C.O. 2000. Exploring the Phylogenetic Structure of Ecological Communities : An Example for Rain
334 Forest Trees. *The American Naturalist*. 156:145–155.
- 335 Webb C.O., Ackerly D.D., Kembel S.W. 2008. Phylocom: Software for the analysis of phylogenetic community
336 structure and trait evolution. *Bioinformatics*. 24:2098–2100.
- 337 Webb C.O., Donoghue M.J. 2005. Phylomatic: Tree assembly for applied phylogenetics. *Molecular Ecology*
338 *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.

FIGURE 6

361

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

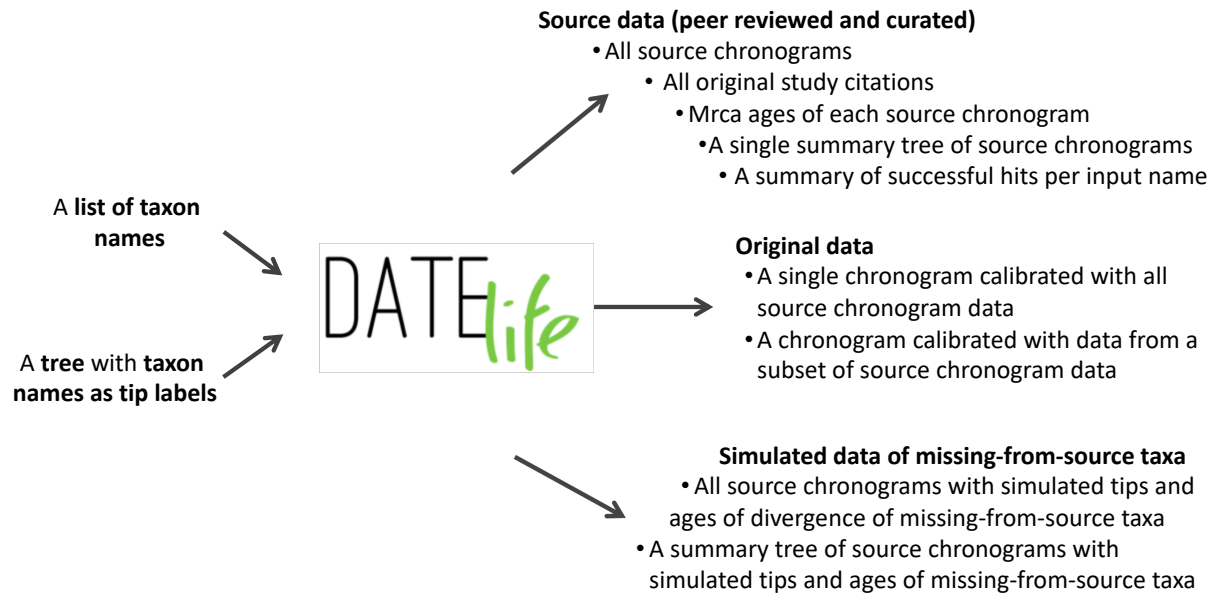


Figure 1:

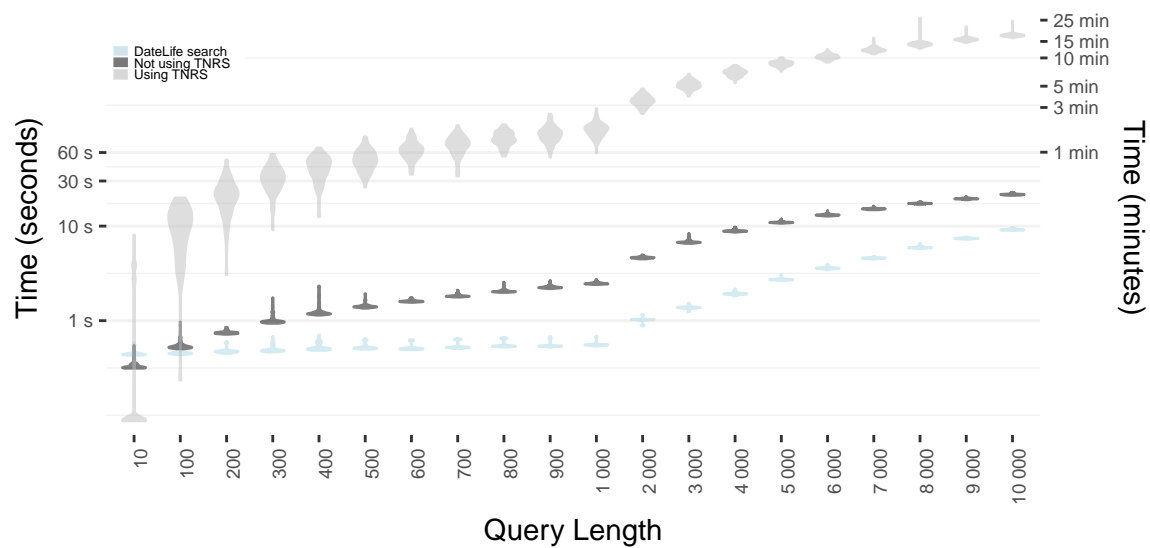


Figure 2:

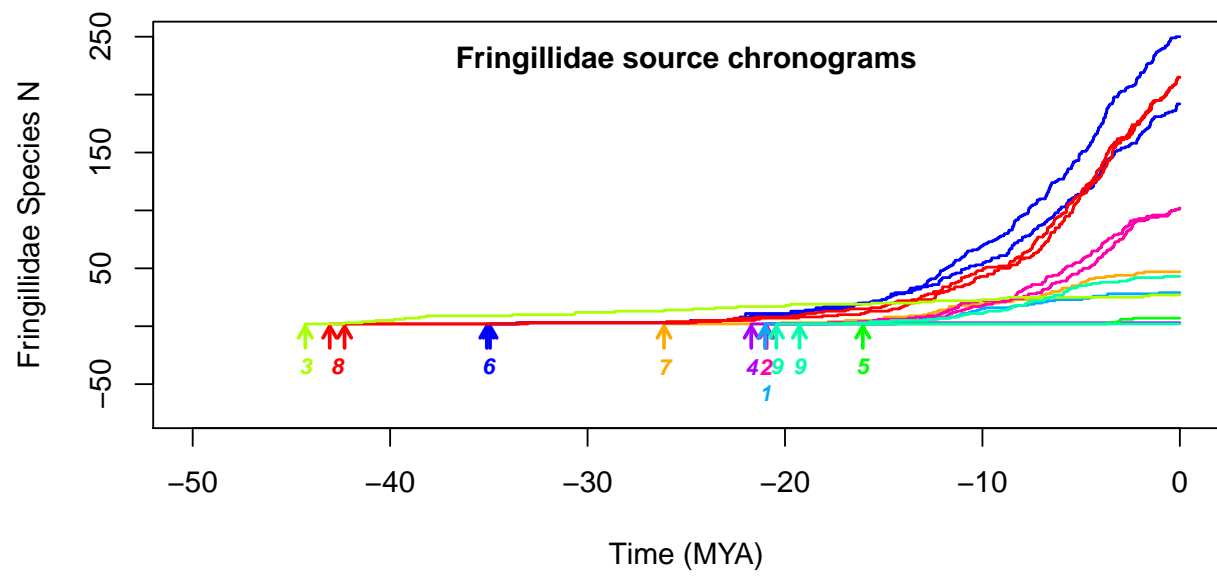


Figure 3:

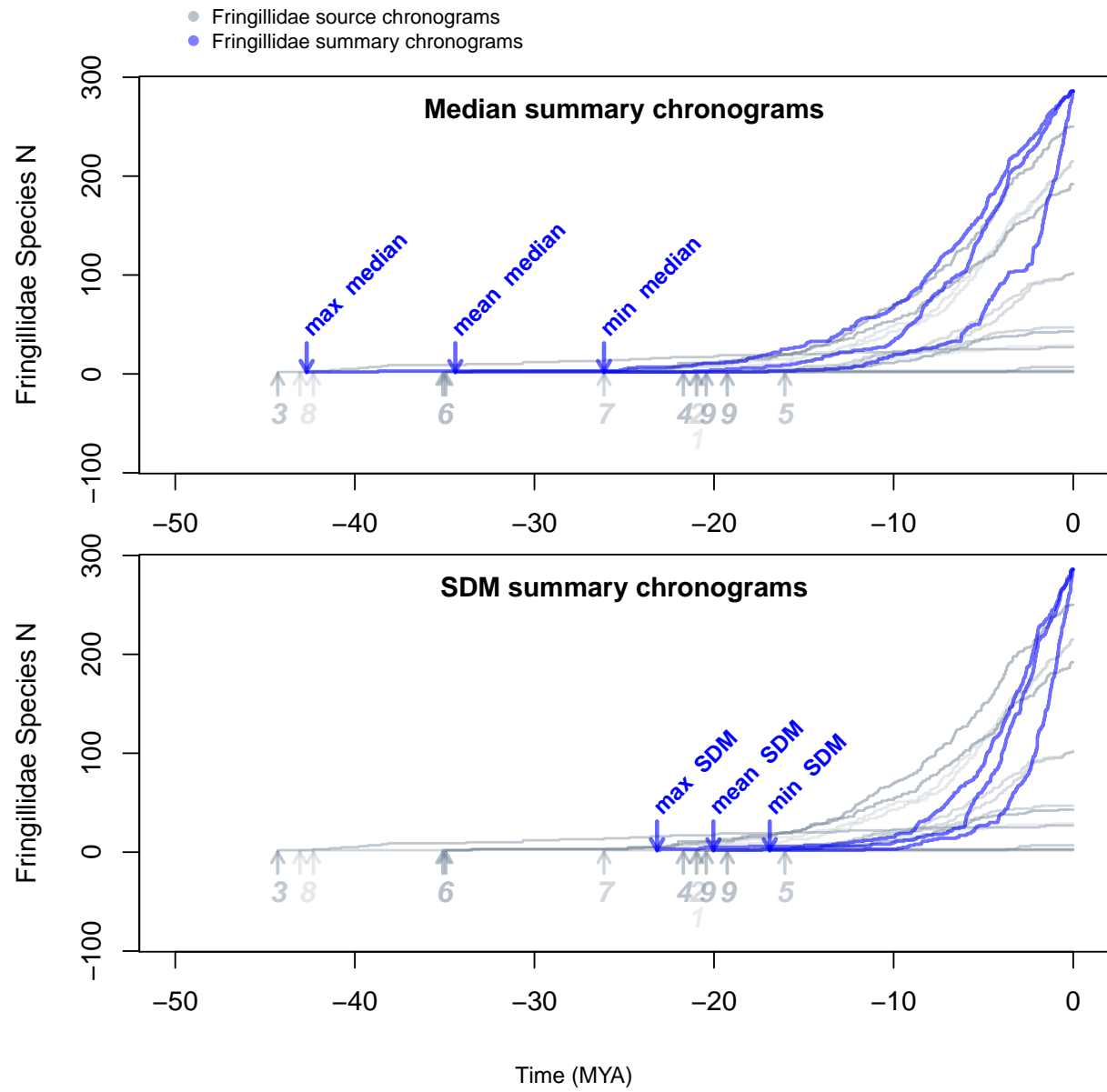


Figure 4:

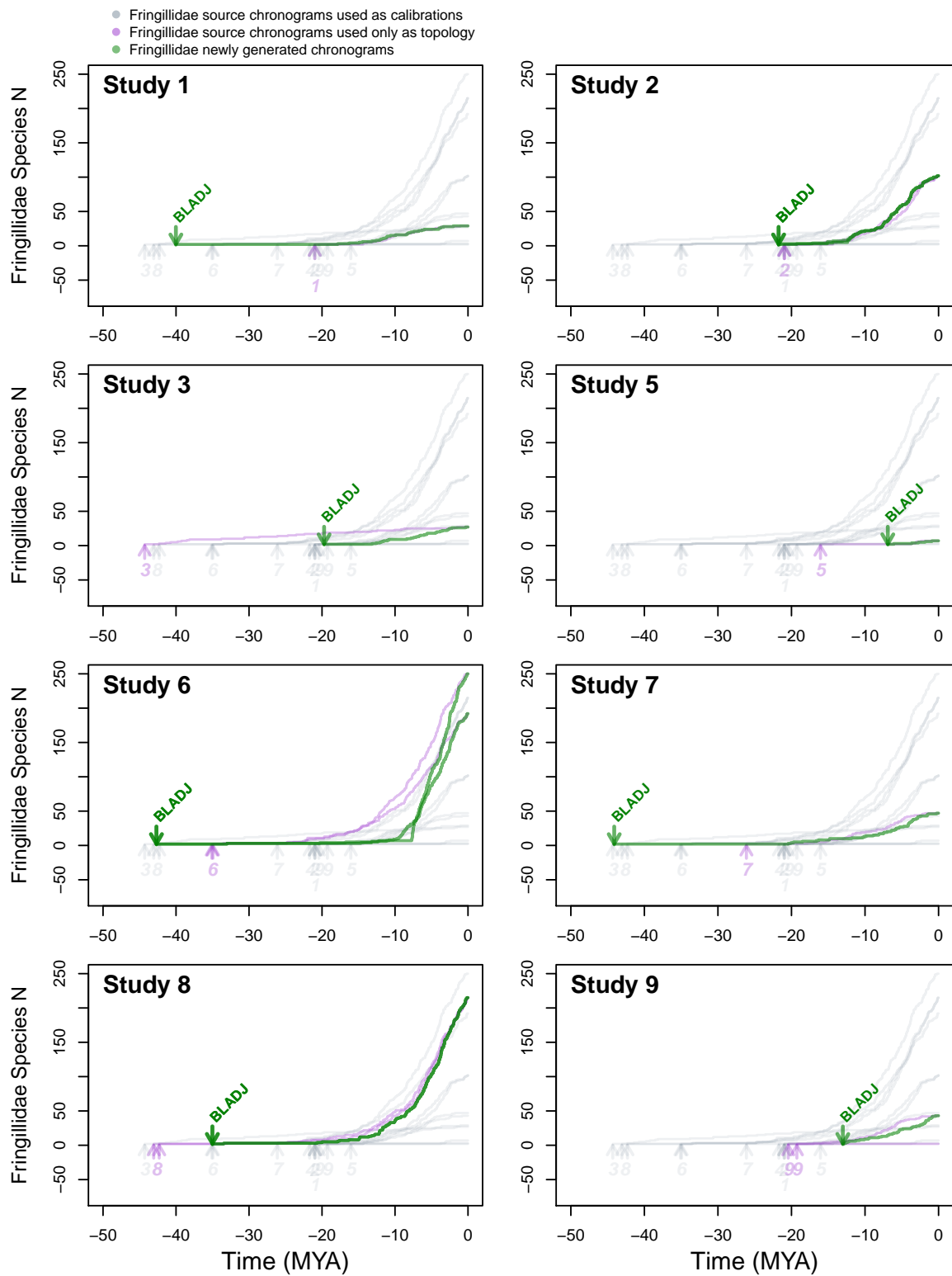


Figure 5:

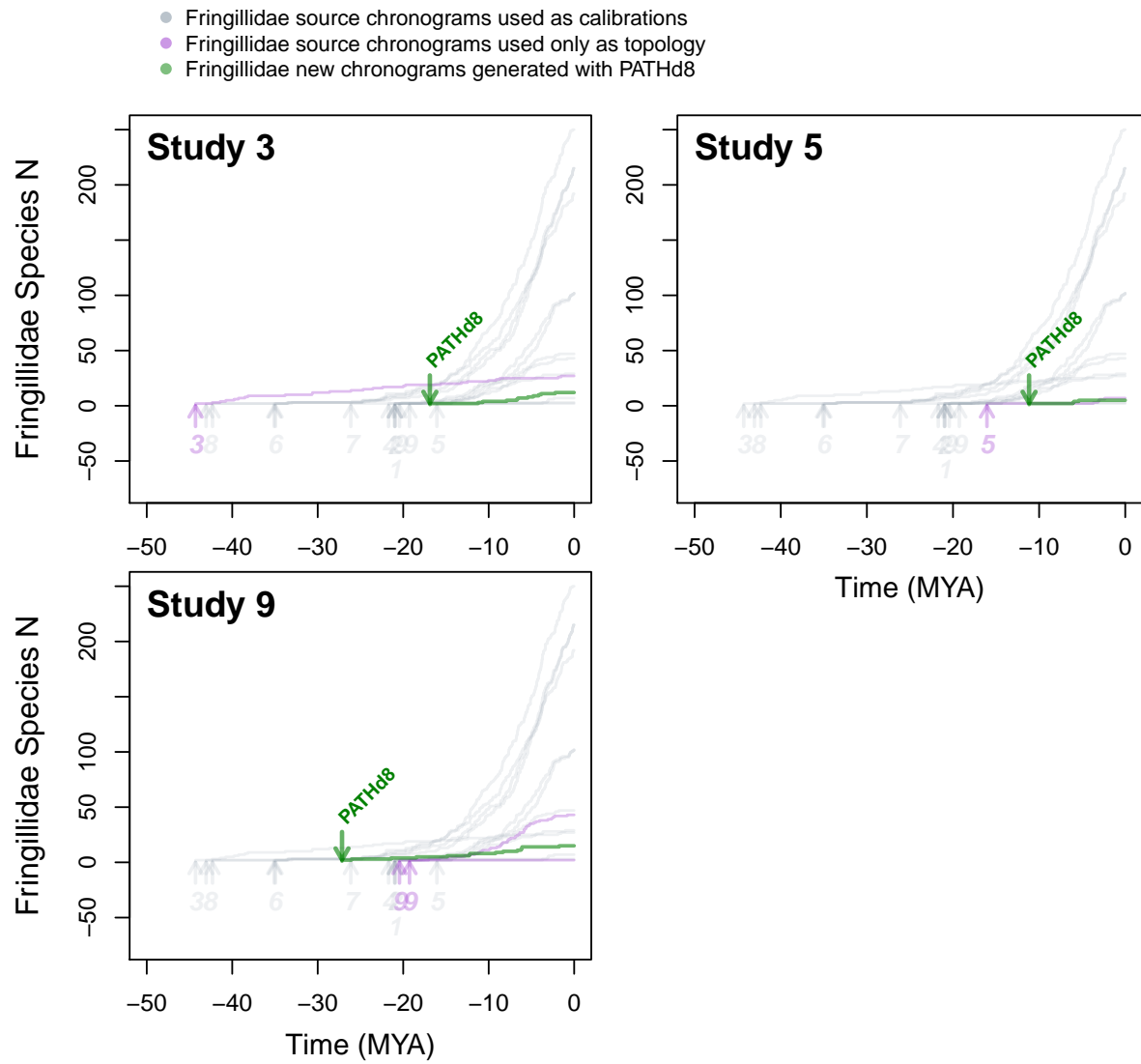


Figure 6: