

Time of lineage divergence constitutes a fundamental piece of information for evolutionary understanding in many areas of research, from developmental to conservation biology (???; Webb 2000). This information can be obtained from fossils and from molecular dating studies. The developments in DNA sequencing and phylogenetic dating inferences have encouraged/spurred the application of molecular dating methods to a very large diversity of organisms, greatly increasing the quantity of data on taxon ages across the tree of life. Along with phylogenetic relationships, it allows testing alternative evolutionary hypothesis in many areas of research, from developmental to conservation biology (???; Webb 2000). It has been shown that only a bit of information on time of divergence can improve inferences of phylogenetic distance Webb et al. (2008) bladj Together with geographical distribution patterns, it is a crucial component of historical biogeography studies (Posadas et al. 2006). Coupled to species number, it allows studying the tempo and mode of speciation and extinction, central for understanding how biodiversity patterns are shaped across space, time, and clades (Morlon 2014). In the past two decades, the possibility to obtain good quality DNA sequences coupled to methodological developments in phylogenetic and dating inference, allowed the application of molecular dating methods to a very large diversity of organisms, greatly increasing the quantity and quality of data on taxon ages across the tree of life. To date, there is a large amount of data on taxon ages and phylogenetic relationships in public repositories such as Dryad, TreeBASE and Open Tree of Life (OToL). , ->

The TimeTree project (Hedges et al. 2006, 2015; Kumar et al. 2017) has aggregated chronograms from 3,163 studies, encompassing 97,085 species (Kumar et al. 2017), and continues to grow. However, even in this gold standard resource, the included taxa only encompass between 0.097 and 3.236% of total species diversity (following taxonomic expert opinion on the global, extant species numbers, which ranges from 3 to 100 million species [Mayr2010; Moran2011]). One advantage of TimeTree is that it includes taxa from across the tree of life, versus more specialized chronogram repositories focusing on land plants (Stevens and Davis 2001) and mammals (Webb and Donoghue 2005), birds (Jetz et al. 2012), and fish (Chang et al. 2019). Users can choose between a web interface or a mobile app to receive information on divergence times for the evolutionary history of a lineage, pairs of taxa, all lineages within a taxon, or a list of taxa. As a science communication tool, TimeTree project is very powerful: it has a friendly graphical interface, with informative and colorful outputs, that allows the general public to satisfy curiosity regarding a particular organism of

interest or group of them. It is of limited utility for scientific studies, however. The thousands of trees that have been entered are unavailable for examination or reuse; according to the creators (see TimeTree web FAQ), methods for allowing data downloading have been under discussion for the past several years yet the primary data remain closed. Moreover, there is no Application Programming Interface (API) allowing programmatic access to any data, greatly impairing the possibility of large-scale, automated data-mining, which is not allowed under TimeTree website's terms of use. The nearly hundred thousand taxon summary chronogram generated from TimeTree resources is not available with its publication (Kumar et al. 2017) or the TimeTree website, though the still substantial chronogram from a previous publication (Hedges et al. 2015) was made available at OToL.

Despite its great importance, analytical tools to summarize available information on taxon ages for the scientific community are still lacking. We identified several aspects that might have so far delayed the exploitation of existing data. First, original chronograms available publicly are scattered across repositories and journals supplementary data, usually with different formats too. Second, lineage names due to taxonomic idiosyncrasy can be different among studies and manual curation of that is usually necessary. Third, data curation

Platforms that contribute to knowledge in this areas:

Supersmart is an open tool but not easy to use, still requires a lot of curation and knowledge.

Phylogenerator (???)

Generate new dates using all available DNA sequence information; Perform one global analysis using all available information;

Problems or downsides: This might be time consuming for large groups and a lot of data curation and knowledge on the group of interest is still necessary. For example, choosing correct fossils for calibration requires a lot of expertise and knowledge on the group. An incorrect use of fossils can generate severe bias in dating results (Sauquet et al. 2012). Hence, data curation is still an important part of any biological

study. The research community considers it as an important or even crucial step before data analysis. Hence, automated processes for large data analysis are frequently received with skepticism.

datelife palliates this by only using information available from already published studies, which are ideally constructed using robust information, such as sequence data and thoughtfully curated fossil calibrations.

Rapidly increasing data on time of lineage divergence both from molecular and paleontological studies; the increasing importance of use of these data in distant areas of research, often not specialized enough to rapidly obtain data on their own; and the lack of an open (both the data sources and the code underlying the analyses) easy to use tool inspired the development of a prototype **datelife** service over a series of phylotastic hackathons (Stoltzfus et al. 2013) at the National Evolutionary Synthesis Center. In this paper we present the first formal description of **datelife**, featuring an improved database of chronograms, more methods to summarize trees, and new functions to visualize data, as well as comparisons of summary trees. **datelife** is the main service for scaling phylogenetic trees in Phylotastic! system (Stoltzfus et al. 2013) It can be used through its R package , and a web interface (<http://www.datelife.org/query/>).

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.

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.

Chang J., Rabosky D.L., Smith S.A., Alfaro M.E. 2019. An r package and online resource for macroevolutionary studies using the ray-finned fish tree of life. *Methods in Ecology and Evolution*. 0:1–7.

Claramunt S., Cracraft J. 2015. A new time tree reveals earth history’s imprint on the evolution of modern

76 birds. *Science advances*. 1:e1501005.

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

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

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

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

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

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

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

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

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

96 Sauquet H., Ho S.Y.W., Gandolfo M. a, Jordan G.J., Wilf P., Cantrill D.J., Bayly M.J., Bromham L., Brown
97 G.K., Carpenter R.J., Lee D.M., Murphy D.J., Sniderman J.M.K., Udovicic F. 2012. Testing the impact

98 of calibration on molecular divergence times using a fossil-rich group: the case of Nothofagus (Fagales).
99 Systematic Biology. 61:289–313.

100 Stevens P.F., Davis H. 2001. Angiosperm phylogeny website..

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

106 Webb C.O. 2000. Exploring the Phylogenetic Structure of Ecological Communities : An Example for Rain
107 Forest Trees. The American Naturalist. 156:145–155.

108 Webb C.O., Ackerly D.D., Kembel S.W. 2008. Phylocom: Software for the analysis of phylogenetic community
109 structure and trait evolution. Bioinformatics. 24:2098–2100.

110 Webb C.O., Donoghue M.J. 2005. Phylomatic: Tree assembly for applied phylogenetics. Molecular Ecology
111 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.

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

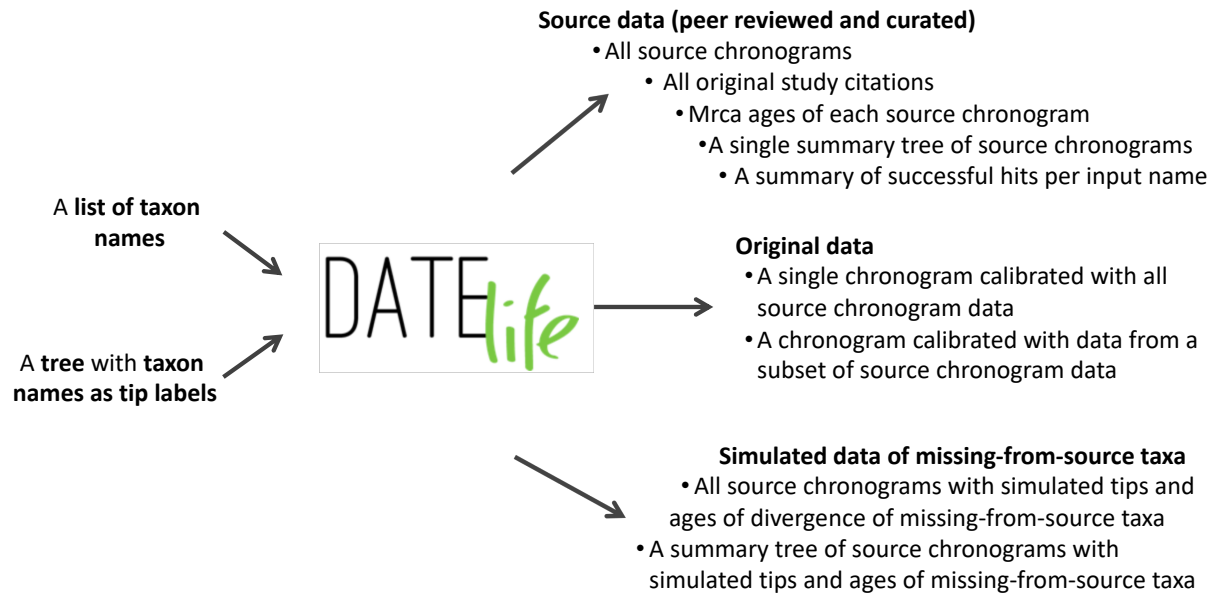


Figure 1:

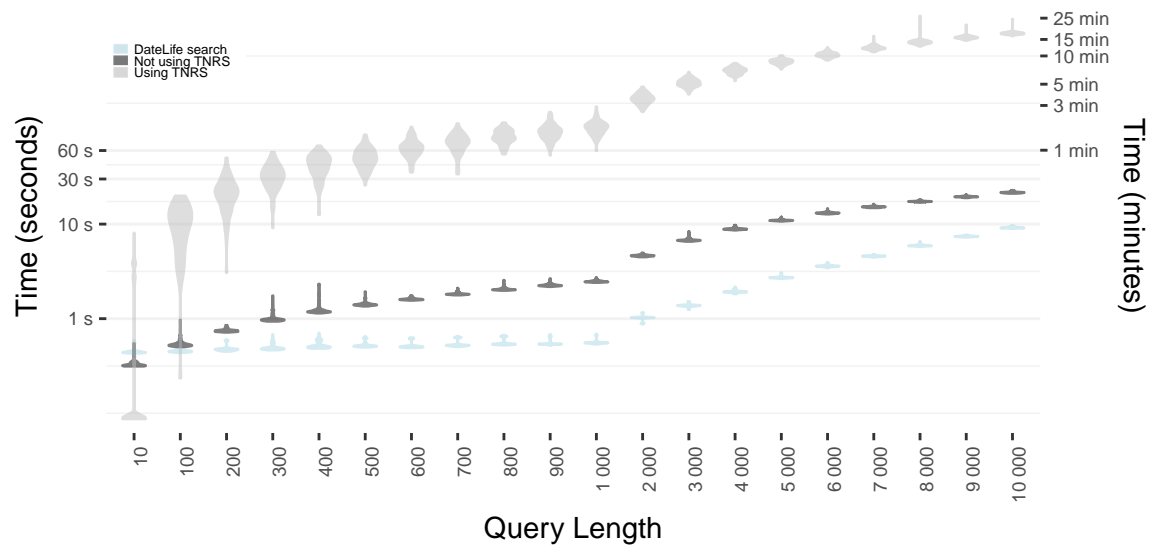


Figure 2:

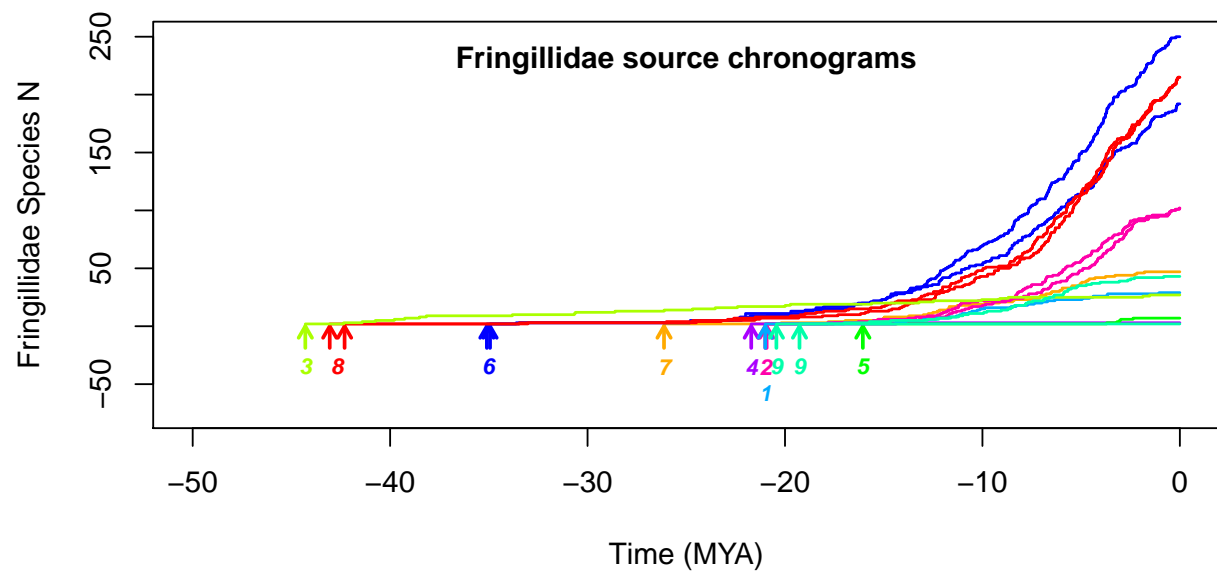


Figure 3:

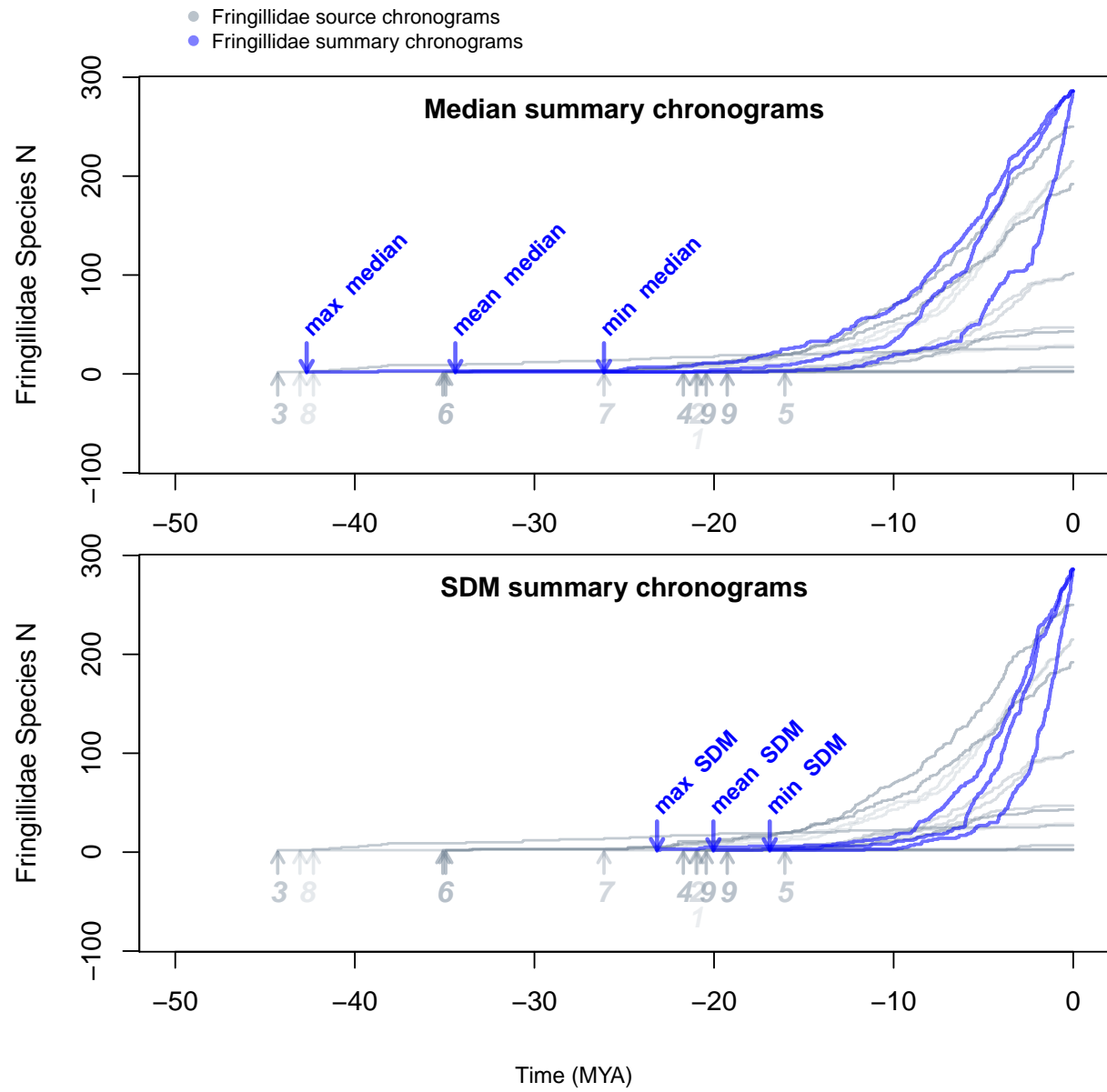


Figure 4:

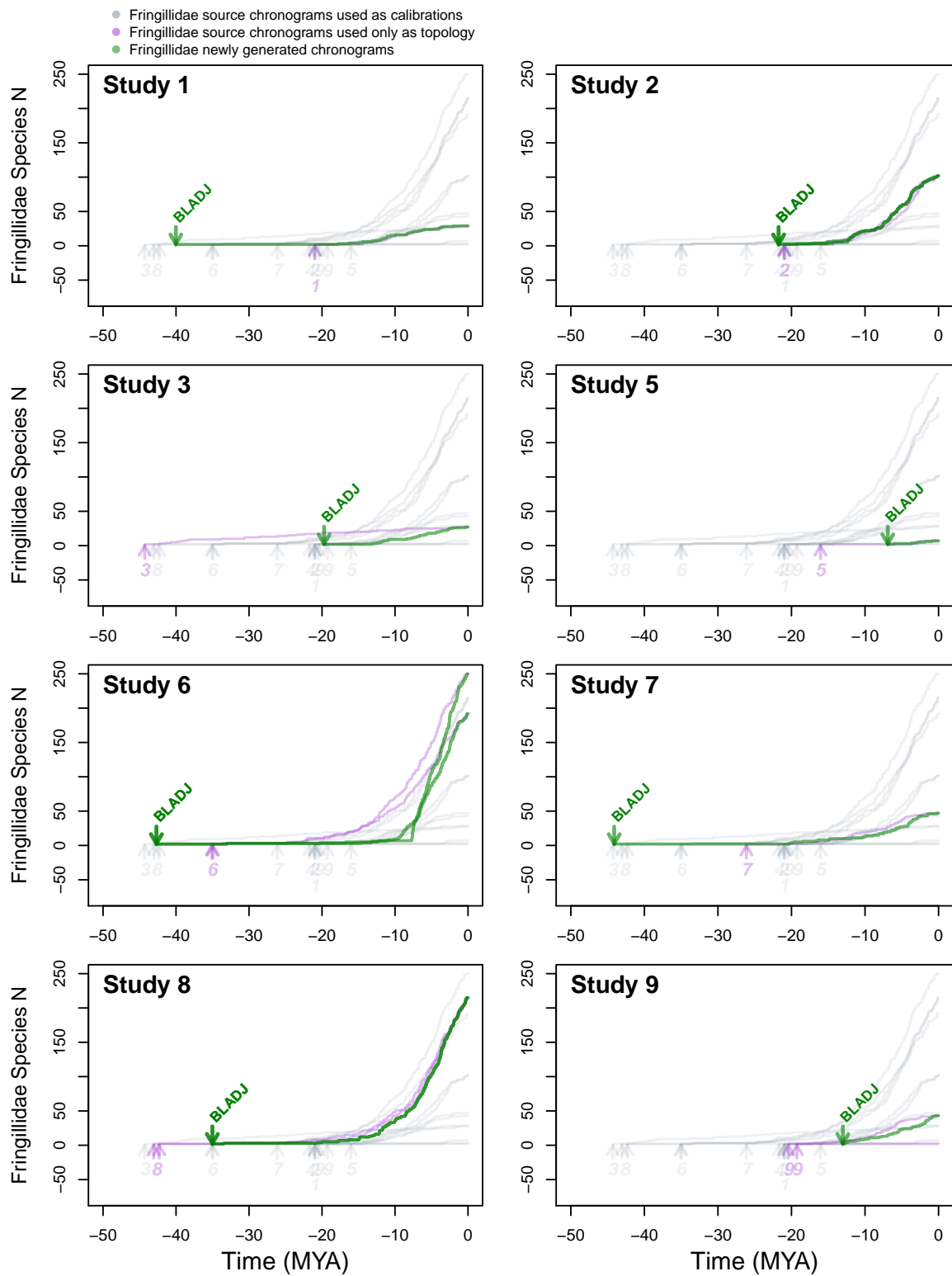


Figure 5:

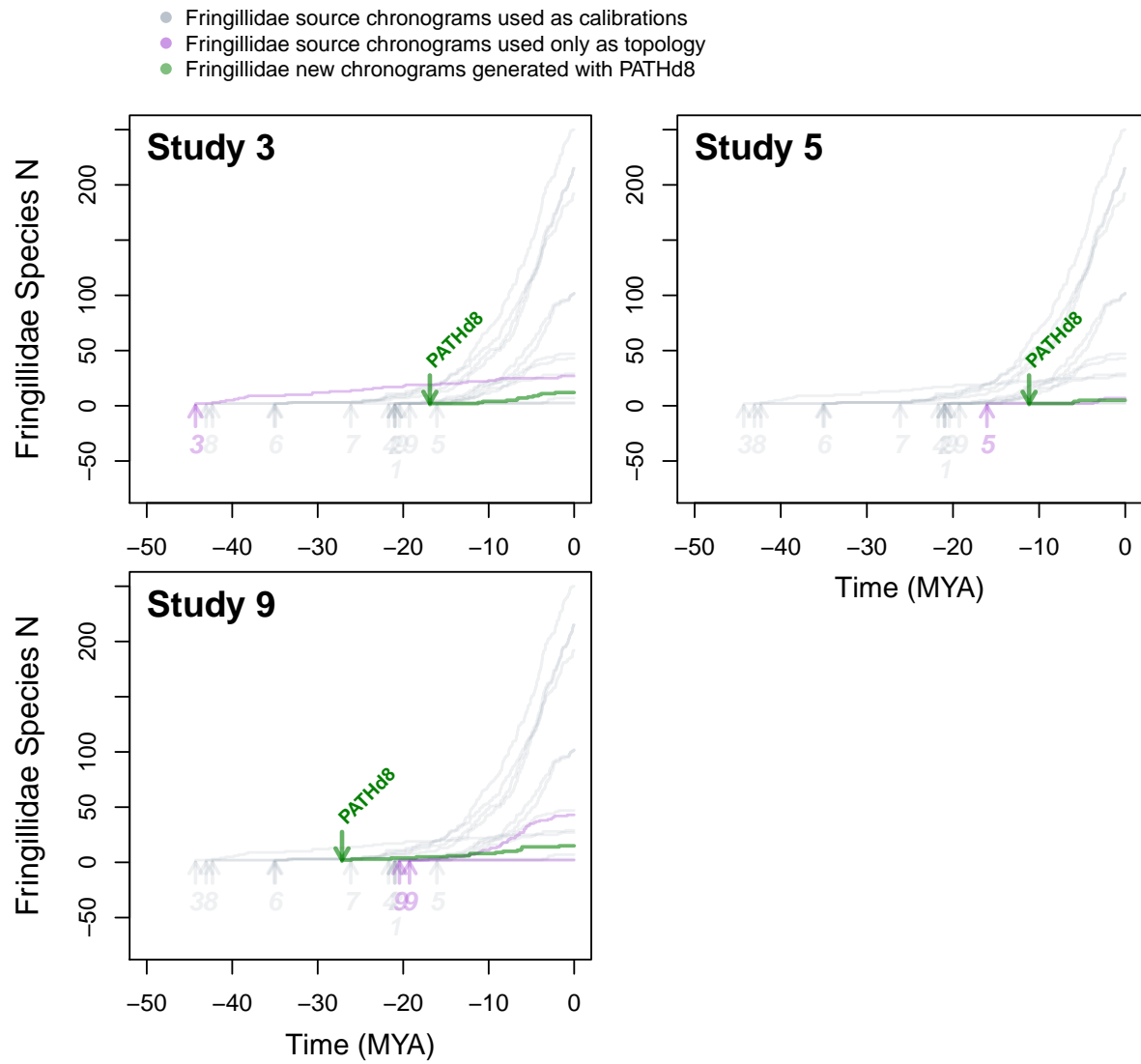


Figure 6: