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

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

3   [1] University of California, Merced

4   [2] University of Tennessee, Knoxville

5   Author Note

⁶   School of Natural Sciences, University of California, Merced, Science and Engineering

⁷ Building 1.

⁸   Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville,

⁹ 425 Hesler Biology Building, Knoxville, TN 37996, USA.

¹⁵   Correspondence concerning this article should be addressed to Luna L. Sánchez Reyes, .

¹⁶ E-mail: sanchez.reyes.luna@gmail.com

Abstract

One or two sentences providing a **basic introduction** to the field, comprehensible to a scientist in any discipline.

Two to three sentences of **more detailed background**, comprehensible to scientists in related disciplines.

One sentence clearly stating the **general problem** being addressed by this particular study.

One sentence summarizing the main result (with the words "**here we show**" or their equivalent).

Two or three sentences explaining what the **main result** reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge.

One or two sentences to put the results into a more **general context**.

Two or three sentences to provide a **broader perspective**, readily comprehensible to a scientist in any discipline.

The combination of new analytical techniques, availability of more fossil and molecular data, and better practices in data sharing in biology 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 few decades. However, getting a tree with branch lengths proportional to time remains difficult 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 published data on

₄₀ lineage divergence times. The main workflow starts with at least two taxon names as input,

₄₁ either as tip labels on a tree, or as a simple comma separated character string. A name

₄₂ search is then performed across the chronogram database and positively identified source

₄₃ trees are pruned to maintain queried taxa only and stored as a named list of patristic

₄₄ distance matrices. Source chronogram data can be summarised using branch length

₄₅ summary statistics or variance minimizing approaches to generate a single summary

₄₆ chronogram. Source chronogram data can also be used as calibration points to date a tree

₄₇ containing some or all names from the initial query. If there is no information available for

₄₈ any queried taxa, data can be simulated. All source and summary chronograms can be saved

₄₉ in formats that permit easy reuse and reanalysis. Summary and newly generated trees are

₅₀ potentially useful to evaluate evolutionary hypothesis in different areas of research in biology.

₅₁ How well this trees work for this purpose still needs to be tested. `datelife` will be useful to

₅₂ increase awereness on the existing variation in expert time of divergence data, and might

₅₃ foster exploration of the effect of alternative divergence time hypothesis on the results of

₅₄ analyses, nurturing a culture of more cautious interpretation of evolutionary results.

₅₅ *Keywords:* keywords

₅₆ Word count: X

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

58   **Methods**

59   We report how we determined our sample size, all data exclusions (if any), all

60 manipulations, and all measures in the study.

61 **Participants**

62 **Material**

63 **Procedure**

64 **Data analysis**

65   We used R (Version 4.1.0; R Core Team, 2021) and the R-package *papaja* (Version

66 0.1.0.9997; Aust & Barth, 2020) for all our analyses.

67   **Results**

68   **Discussion**

References

Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown.*
    Retrieved from https://github.com/crsh/papaja

R Core Team. (2021). *R: A language and environment for statistical computing.* Vienna,
    Austria: R Foundation for Statistical Computing. Retrieved from
    https://www.R-project.org/