Dryad Data Package for study "DateLife: leveraging databases and analytical tools to reveal the dated Tree of Life"

This README file was generated on 2022-07-16 by Luna L. Sánchez Reyes, https://orcid.org/0000-0001-7668-2528

GENERAL INFORMATION

1. Title of Dataset

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

2. Author Information

   Corresponding Researcher
   name: Luna L. Sanchez Reyes
   institution: University of California, Merced, USA
   email: sanchez.reyes.luna@gmail.com

   Co-researcher 1
   name: Emily Jane McTavish
   institution: University of California, Merced, USA

   Co-researcher 2
   name: Brian C. O'Meara
   institution: University of Tennessee, Knoxville, USA

3. Date of data collection: 2022-01-28

4. Geographic location of data collection: Online

5. Funding sources that supported the collection of the data: National Science Foundation, USA

6. Recommended citation for this dataset:

   Sánchez-Reyes LL, McTavish EJ & O'Meara BC. (2019). DateLife: leveraging databases and analytical tools to reveal the dated Tree of Life. BioRxiv, 782094.

   Data from: DateLife: leveraging databases and analytical tools to reveal the dated Tree of Life. Dryad, Dataset

DATA & FILE OVERVIEW

1. Description of dataset

These data were generated to investigate and showcase the performance of the datelife R package (https://github.com/phylotastic/datelife). We showcased the application of the package with one mock example and two different biological examples.

The first biological example uses datelife on a small sample of bird species. The second one uses datelife on bird species belonging to the family Fringillidae of "true finches", following the NCBI taxonomy. The mock example was created by replacing species names from the small example with letters, and reducing the number of source chronograms.

We evaluated the performance of the package datelife with two analysis: a benchmarking analysis to measure computing time of functions, and a cross validation analysis to test the accuracy and precision of the functions.

2. File List:

File 1 Name: Sanchez-Reyes_etal_2022_table_1.csv
File 1 Description: CSV (comma separated values) file containing results from the small example. Names and ages (in Million years) of nodes from source chronograms that have been congruified to nodes of the chosen tree topology.

File 2 Name: Sanchez−Reyes_etal_2022_table_2.csv
File 2 Description: CSV (comma separated values) file containing a summary of node age (in Million years) data from table 1.

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_table_S1.csv
File XXX Description: CSV (comma separated values) file containing results from the Fringillidae example. Names and ages (in Million years) of nodes from source chronograms that have been congruified to nodes of the chosen tree topology shown in Figure 4B of the manuscript.

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_table_S1.pdf
File XXX Description: PDF version of file Sanchez−Reyes_etal_2022_supplementary_table_S1.csv

File XXX Name: Sanchez-Reyes_etal_2022_supplementary_table_S2.csv
File XXX Description: CSV (comma separated values) file containing a summary of node age data (in Million years) from table S1.

File XXX Name: Sanchez-Reyes_etal_2022_supplementary_table_S2.pdf
File XXX Description: PDF file version of file Sanchez-Reyes_etal_2022_supplementary_table_S2.csv.

File 3 Name: Sanchez−Reyes_etal_2022_figure_1_chronogram_mock_example.tre
File 3 Description: Newick version of chronogram from the mock example shown in Figure 1 of the manuscript (https://www.biorxiv.org/content/10.1101/782094v2).

File 4 Name: Sanchez−Reyes_etal_2022_figure_3_chronogram_small_example.tre
File 4 Description: Newick version of chronogram from the small example shown in Figure 3 of the manuscript (https://www.biorxiv.org/content/10.1101/782094v2).

File 6 Name: Sanchez−Reyes_etal_2022_figure_4A_topology_finches_mrca.tre
File 6: Description: Newick version of topology extracted from the Open Tree of Life synthetic tree, shown in Figure 4A of the manuscript (https://www.biorxiv.org/content/10.1101/782094v2).

File 7 Name: Sanchez−Reyes_etal_2022_figure_4B_topology_finches_ncbi.tre
File 7 Description: Newick version of topology extracted from the Open Tree of Life synthetic tree, shown in Figure 4B of the manuscript (https://www.biorxiv.org/content/10.1101/782094v2).

File 5 Name: Sanchez−Reyes_etal_2022_figure_5_chronogram_finches_example.tre
File 5 Description: Newick version of chronogram from the finch example shown in Figure 5 of the manuscript (https://www.biorxiv.org/content/10.1101/782094v2).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S1.pdf
File XXX Description: PDf file showing results of cross validation analysis of datelife's chronogram generating functions.

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S2.jpg
File XXX Description: JPG file of results of cross validation analysis of datelife's chronogram generating functions, using a tree topology from Barker et al. 2013. Comparison of original chronogram (black) and the chronogram obtained using datelife (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S2.tre
File XXX Description: Newick file of chronogram obtained with datelife, shown in supplementary Figure S2 (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S3.jpg
File XXX Description: JPG file of results of cross validation analysis of datelife's chronogram generating functions, using a tree topology from Barker et al. 2015, chronogram 1. Comparison of original chronogram (black) and the chronogram obtained using datelife (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S3.tre
File XXX Description: Newick file of chronogram obtained with datelife, shown in supplementary Figure S3 (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S4.jpg
File XXX Description: JPG file of results of cross validation analysis of datelife's chronogram generating functions, using a tree topology from Barker et al. 2015, chronogram 2. Comparison of original chronogram (black) and the chronogram obtained using datelife (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S4.tre
File XXX Description: Newick file of chronogram obtained with datelife, shown in supplementary Figure S4 (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S5.jpg
File XXX Description: JPG file of results of cross validation analysis of datelife's chronogram generating functions, using a tree topology from Burns et al. 2015. Comparison of original chronogram (black) and the chronogram obtained using datelife (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S5.tre
File XXX Description: Newick file of chronogram obtained with datelife, shown in supplementary Figure S5 (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S6.jpg
File XXX Description: JPG file of results of cross validation analysis of datelife's chronogram generating functions, using a tree topology from Claramunt et al. 2015. Comparison of original chronogram (black) and the chronogram obtained using datelife (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S6.tre
File XXX Description: Newick file of chronogram obtained with datelife, shown in supplementary Figure S6 (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S7.jpg
File XXX Description: JPG file of results of cross validation analysis of datelife's chronogram generating functions, using a tree topology from Gibb et al. 2015. Comparison of original chronogram (black) and the chronogram obtained using datelife (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S7.tre
File XXX Description: Newick file of chronogram obtained with datelife, shown in supplementary Figure S7 (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S8.jpg
File XXX Description: JPG file of results of cross validation analysis of datelife's chronogram generating functions, using a tree topology from Hedges et al. 2015, chronogram 1. Comparison of original chronogram (black) and the chronogram obtained using datelife (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S8.tre
File XXX Description: Newick file of chronogram obtained with datelife, shown in supplementary Figure S8 (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S9.jpg
File XXX Description: JPG file of results of cross validation analysis of datelife's chronogram generating functions, using a tree topology from Hedges et al. 2015, chronogram 2. Comparison of original chronogram (black) and the chronogram obtained using datelife (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S9.tre
File XXX Description: Newick file of chronogram obtained with datelife, shown in supplementary Figure S9 (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S10.jpg
File XXX Description: JPG file of results of cross validation analysis of datelife's chronogram generating functions, using a tree topology from Hooper et al. 2017. Comparison of original chronogram (black) and the chronogram obtained using datelife (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S10.tre
File XXX Description: Newick file of chronogram obtained with datelife, shown in supplementary Figure S10 (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S11.jpg
File XXX Description: JPG file of results of cross validation analysis of datelife's chronogram generating functions, using a tree topology from Jetz et al. 2012, chronogram 1. Comparison of original chronogram (black) and the chronogram obtained using datelife (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S11.tre
File XXX Description: Newick file of chronogram obtained with datelife, shown in supplementary Figure S11 (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S12.jpg
File XXX Description: JPG file of results of cross validation analysis of datelife's chronogram generating functions, using a tree topology from

Jetz et al. 2012, chronogram 2. Comparison of original chronogram (black)
and the chronogram obtained using datelife (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S12.tre
File XXX Description: Newick file of chronogram obtained with datelife, shown
    in supplementary Figure S12 (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S13.jpg
File XXX Description: JPG file of results of cross validation analysis of
    datelife's chronogram generating functions, using a tree topology from
    Kimball et al. 2019, chronogram 1. Comparison of original chronogram (
    black) and the chronogram obtained using datelife (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S13.tre
File XXX Description: Newick file of chronogram obtained with datelife, shown
    in supplementary Figure S13 (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S14.jpg
File XXX Description: JPG file of results of cross validation analysis of
    datelife's chronogram generating functions, using a tree topology from
    Kimball et al. 2019, chronogram 2. Comparison of original chronogram (
    black) and the chronogram obtained using datelife (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S14.tre
File XXX Description: Newick file of chronogram obtained with datelife, shown
    in supplementary Figure S14 (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S15.jpg
File XXX Description: JPG file of results of cross validation analysis of
    datelife's chronogram generating functions, using a tree topology from
    Oliveros et al. 2019. Comparison of original chronogram (black) and the
    chronogram obtained using datelife (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S15.tre
File XXX Description: Newick file of chronogram obtained with datelife, shown
    in supplementary Figure S15 (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S16.jpg
File XXX Description: JPG file of results of cross validation analysis of
    datelife's chronogram generating functions, using a tree topology from
    Price et al. 2014, chronogram 1. Comparison of original chronogram (black)
     and the chronogram obtained using datelife (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S16.tre
File XXX Description: Newick file of chronogram obtained with datelife, shown
    in supplementary Figure S16 (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S17.jpg
File XXX Description: JPG file of results of cross validation analysis of
    datelife's chronogram generating functions, using a tree topology from
    Price et al. 2014, chronogram 2. Comparison of original chronogram (black)
     and the chronogram obtained using datelife (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S17.tre

File XXX Description: Newick file of chronogram obtained with datelife, shown
    in supplementary Figure S17 (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S18.jpg
File XXX Description: JPG file of results of cross validation analysis of
    datelife's chronogram generating functions, using a tree topology from
    Roquet et al. 2014, chronogram 1. Comparison of original chronogram (black
) and the chronogram obtained using datelife (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S18.tre
File XXX Description: Newick file of chronogram obtained with datelife, shown
    in supplementary Figure S18 (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S19.jpg
File XXX Description: JPG file of results of cross validation analysis of
    datelife's chronogram generating functions, using a tree topology from
    Roquet et al. 2014, chronogram 2. Comparison of original chronogram (black
) and the chronogram obtained using datelife (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S19.tre
File XXX Description: Newick file of chronogram obtained with datelife, shown
    in supplementary Figure S19 (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S20.jpg
File XXX Description: JPG file of results of cross validation analysis of
    datelife's chronogram generating functions, using a tree topology from
    Uyeda et al. 2017. Comparison of original chronogram (black) and the
    chronogram obtained using datelife (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_figure_S20.tre
File XXX Description: Newick file of chronogram obtained with datelife, shown
    in supplementary Figure S20 (gray).

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_
File XXX Description:

File XXX Name: Sanchez−Reyes_etal_2022_supplementary_
File XXX Description:

3. Original names and locations of files at https://github.com/LunaSare/datelifeMS1:

cp tables/table−fringillidae−all.pdf dryad/Supplementary_Table_S1.pdf
cp tables/table−fringillidae−all−summary.pdf dryad/Supplementary_Table_S2.pdf
cp figures/figure−cross−validation/fig−cross−validation−xy−plots−diffs.pdf
    dryad/Supplementary_Figure_S1.pdf

METHODOLOGICAL INFORMATION

Bacterial detection was performed on the ViiA7 Real-time PCR System (Thermo Fisher Scientific) using
PrimeTime® qPCR primers, probes and mastermix (IDT) according to the manufacturer's instructions.
Reactions were performed using 1X PrimeTime® Gene Expression Master Mix, 1X PrimeTime®qPCR Assay
and up to 10ng of DNA. Cycling conditions were 95°C for 3 minutes, 60 cycles of 95°C for 5 seconds and
60°C for 30 seconds. Amplification results were reviewed using QuantStudioTM Real-Time PCR Software

version 1.1 (Thermo Fisher Scientific). Amplification of beta-actin and prostaglandin transporter (PGT) was used to determine relative abundance.

Diversity profiling was performed by AGRF (Australian Genome Research Facility, Melbourne Australia). Samples were amplified with universal primers to the V1-V3 region of the bacterial 16S gene (forward AGAGTTTGATCMTGGCTCAG; reverse GWATTACCGCGGCKGCTG). Amplicons were indexed using the Nextera XT Index Kit (Illumina, San Diego, CA, USA) followed by Paired End sequencing on a MiSeq next generation sequencer (Illumina). Paired-end reads were assembled by aligning the forward and reverse reads using PEAR1 (version 0.9.5). Primers were identified and trimmed. Trimmed sequences were processed using Quantitative Insights into Microbial Ecology (QIIME 1.8) USEARCH (version 8.0.1623) and UPARSE software. Sequences were quality filtered and sorted by abundance after removal of full-length duplicate sequences. Singletons or unique reads were discarded. Sequences were clustered and then chimera filtered using "rdp_gold" database as reference. Reads were mapped back to Operational Taxonomic Units with a minimum identity of 97% and taxonomy was assigned using the QIIME 1 default classifier, pre-trained against Greengenes database5 (Version 13_8, Aug 2013).

DATA-SPECIFIC INFORMATION FOR: Sanchez-Reyes_etal_2022_table_1.csv

1. Number of variables/columns: 5

2. Number of cases/rows: 28

3. Variable List
    Node Name: Node names in tree topology, as defined by taxon A and taxon B
    taxon A: Scientific name of the first element of a taxon pair that defines a node from a tree topology
    taxon B: Scientific name of the second element of a taxon pair that defines a node from a tree topology
    Node Age: Ages of nodes defined by taxon A and taxon B, in Million years
    Study chronogram: Reference of the study that published the chronogram from which the node ages
were extracted

4. Missing data codes: None

5. Abbreviations used: NA = not applicable

DATA-SPECIFIC INFORMATION FOR: Sanchez-Reyes_etal_2022_table_2.csv

1. Number of variables/columns: 9

2. Number of cases/rows: 5

3. Variable List
    Node Name: Names of nodes from the tree topology used for the dating analysis
    Min Age: Minimum node age, in Million years
    Q1: First quartile of node age distributions, in Million years
    Median Age: Mean node age, in Million years
    Mean Age: Median node age, in Million years
    Q3: Third quartile of node age distributions, in Million years
    Max Age: Maximum node age, in Million years
    Variance: Variance of node age, in Million years
    SD: Standar deviation of node age, in Million years

4. Missing data codes: None

5. Abbreviations used: NA = not applicable

DATA-SPECIFIC INFORMATION FOR: Sanchez-Reyes_etal_2022_supplementary_table_S1.csv

1. Number of variables/columns: 5

2. Number of cases/rows: 818

3. Variable List
    Node Name: Node names in tree topology, as defined by taxon A and taxon B
    taxon A: Scientific name of the first element of a taxon pair that defines a node from a tree topology
    taxon B: Scientific name of the second element of a taxon pair that defines a node from a tree topology
    Node Age: Ages of nodes defined by taxon A and taxon B, in Million years
    Study reference: Reference of the study that published the chronogram from which the node ages were
extracted

4. Missing data codes: None

5. Abbreviations used: NA = not applicable

DATA-SPECIFIC INFORMATION FOR: Sanchez-Reyes_etal_2022_supplementary_table_S2.csv

1. Number of variables/columns: 9

2. Number of cases/rows: 194

3. Variable List
    Node Name: Names of nodes from the tree topology used for the dating analysis
    Min Age: Minimum node age, in Million years
    Q1: First quartile of node age distributions, in Million years
    Median Age: Mean node age, in Million years
    Mean Age: Median node age, in Million years
    Q3: Third quartile of node age distributions, in Million years
    Max Age: Maximum node age, in Million years
    Variance: Variance of node age, in Million years
    SD: Standar deviation of node age, in Million years

4. Missing data codes: None

5. Abbreviations used: NA = not applicable