

Generating a README.txt file for the Dryad submission of data from study “DateLife: leveraging databases and analytical tools to reveal the dated Tree of Life”

Description of README file

```
line0 <- paste("This README file was generated on", Sys.Date(), "by Luna L. Sanchez-Reyes, <https://orcid.org/0000-0001-7668-2528>")
```

General Information on data set

```
line1 <- "
```

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 :

Sanchez-Reyes LL, McTavish EJ & Mearns 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

"

Original file names, rename

```
file.copy(from = "../tables/table-fringillidae-small-example.csv",
          to = "../dryad/Sanchez-Reyes_etal_2022_table_1.csv",
          overwrite = TRUE)
file.copy(from = "../tables/table-fringillidae-small-example-summary.csv",
          to = "../dryad/Sanchez-Reyes_etal_2022_table_2.csv",
          overwrite = TRUE)
file.copy(from = "../tables/table-fringillidae-all-congruified.csv",
          to = "../dryad/Sanchez-Reyes_etal_2022_supplementary_table_S1.csv",
          overwrite = TRUE)
file.copy(from = "../tables/table-fringillidae-all-congruified.pdf",
          to = "../dryad/Sanchez-Reyes_etal_2022_supplementary_table_S1.pdf",
          overwrite = TRUE)
file.copy(from = "../tables/table-fringillidae-all-summary.csv",
          to = "../dryad/Sanchez-Reyes_etal_2022_supplementary_table_S2.csv",
          overwrite = TRUE)
file.copy(from = ".../tables/table-fringillidae-all-summary.pdf",
          to = "../dryad/Sanchez-Reyes_etal_2022_supplementary_table_S2.pdf",
          overwrite = TRUE)
```

Description of data files

line2 <- "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/phylostatic/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 with node age results from datelife 's small example, presented in Table 1 of the manuscript. It contains node names, taxon names defining the nodes, the corresponding node ages (in Million years), and references of source chronograms that were congruified to the chosen tree topology to extract node ages.

File 2 Name: Sanchez-Reyes_etal_2022_table_2.csv

File 2 Description: CSV (comma separated values) file summarizing age data (in Million years) from table 1, per node. These data are presented in Table 2 of the manuscript.

File 3 Name: Sanchez-Reyes_etal_2022_supplementary_table_S1.csv

File 3 Description: CSV (comma separated values) file with node age results from datelife 's Fringillidae example. It contains node names, taxon names defining the nodes, the corresponding node ages (in Million years), and references of source chronograms that were congruified to the chosen tree topology (shown in Figure 4B of the manuscript) to extract node ages.

File 4 Name: Sanchez-Reyes_etal_2022_supplementary_table_S1.pdf

File 4 Description: PDF version of file Sanchez-Reyes_etal_2022_supplementary_table_S1.csv

File 5 Name: Sanchez-Reyes_etal_2022_supplementary_table_S2.csv

File 5 Description: CSV (comma separated values) file summarizing age data (in Million years) from table S1, per node.

File 6 Name: Sanchez-Reyes_etal_2022_supplementary_table_S2.pdf

File 6 Description: PDF file version of file Sanchez-Reyes_etal_2022_supplementary_table_S2.csv.

File 7 Name: Sanchez-Reyes_etal_2022_figure_1_chronogram_mock_example.tre

File 7 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 8 Name: Sanchez-Reyes_etal_2022_figure_3_chronogram_small_example.tre

File 8 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 9 Name: Sanchez-Reyes_etal_2022_figure_4A_topology_finches_mrca.tre

File 9 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 10 Name: Sanchez-Reyes_etal_2022_figure_4B_topology_finches_ncbi.tre

File 10 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_11_Name: Sanchez-Reyes_etal_2022_figure_5_chronogram_finches_example.tre
File_11_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_12_Name: Sanchez-Reyes_etal_2022_supplementary_figure_S1.pdf
File_12_Description: Pdf file showing results of cross validation analysis of
datelife 's chronogram generating functions.

File_13_Name: Sanchez-Reyes_etal_2022_supplementary_figure_S2.jpg
File_13_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_14_Name: Sanchez-Reyes_etal_2022_supplementary_figure_S2.tre
File_14_Description: Newick file of chronogram obtained with datelife, shown
in supplementary Figure S2 (gray).

File_15_Name: Sanchez-Reyes_etal_2022_supplementary_figure_S3.jpg
File_15_Description: JPG file of results of cross validation analysis of
datelife 's chronogram generating functions, using a tree topology from
Barker et al. 2015. Comparison of original chronogram (black)
and the chronogram obtained using datelife (gray).

File_16_Name: Sanchez-Reyes_etal_2022_supplementary_figure_S3.tre
File_16_Description: Newick file of chronogram obtained with datelife, shown
in supplementary Figure S3 (gray).

File_17_Name: Sanchez-Reyes_etal_2022_supplementary_figure_S4.jpg
File_17_Description: JPG file of results of cross validation analysis of
datelife 's chronogram generating functions, using a tree topology from
Barker et al. 2015. Comparison of original chronogram (black)
and the chronogram obtained using datelife (gray).

File_18_Name: Sanchez-Reyes_etal_2022_supplementary_figure_S4.tre
File_18_Description: Newick file of chronogram obtained with datelife, shown
in supplementary Figure S4 (gray).

File_19_Name: Sanchez-Reyes_etal_2022_supplementary_figure_S5.jpg
File_19_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_20_Name: Sanchez-Reyes_etal_2022_supplementary_figure_S5.tre
File_20_Description: Newick file of chronogram obtained with datelife, shown
in supplementary Figure S5 (gray).

File_21_Name: Sanchez-Reyes_etal_2022_supplementary_figure_S6.jpg
File_21_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 22 Name: Sanchez-Reyes_etal_2022_supplementary_figure_S6.tre

File 22 Description: Newick file of chronogram obtained with datelife, shown in supplementary Figure S6 (gray).

File 23 Name: Sanchez-Reyes_etal_2022_supplementary_figure_S7.jpg

File 23 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 24 Name: Sanchez-Reyes_etal_2022_supplementary_figure_S7.tre

File 24 Description: Newick file of chronogram obtained with datelife, shown in supplementary Figure S7 (gray).

File 25 Name: Sanchez-Reyes_etal_2022_supplementary_figure_S8.jpg

File 25 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 26 Name: Sanchez-Reyes_etal_2022_supplementary_figure_S8.tre

File 26 Description: Newick file of chronogram obtained with datelife, shown in supplementary Figure S8 (gray).

File 27 Name: Sanchez-Reyes_etal_2022_supplementary_figure_S9.jpg

File 27 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 28 Name: Sanchez-Reyes_etal_2022_supplementary_figure_S9.tre

File 28 Description: Newick file of chronogram obtained with datelife, shown in supplementary Figure S9 (gray).

File 29 Name: Sanchez-Reyes_etal_2022_supplementary_figure_S10.jpg

File 29 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 30 Name: Sanchez-Reyes_etal_2022_supplementary_figure_S10.tre

File 30 Description: Newick file of chronogram obtained with datelife, shown in supplementary Figure S10 (gray).

File 31 Name: Sanchez-Reyes_etal_2022_supplementary_figure_S11.jpg

File 31 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 32 Name: Sanchez-Reyes_etal_2022_supplementary_figure_S11.tre

File 32 Description: Newick file of chronogram obtained with datelife, shown

in supplementary Figure S11 (gray).

File 33 Name: Sanchez-Reyes_et al_2022_supplementary_figure_S12.jpg

File 33 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 34 Name: Sanchez-Reyes_et al_2022_supplementary_figure_S12.tre

File 34 Description: Newick file of chronogram obtained with datelife, shown in supplementary Figure S12 (gray).

File 35 Name: Sanchez-Reyes_et al_2022_supplementary_figure_S13.jpg

File 35 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 36 Name: Sanchez-Reyes_et al_2022_supplementary_figure_S13.tre

File 36 Description: Newick file of chronogram obtained with datelife, shown in supplementary Figure S13 (gray).

File 37 Name: Sanchez-Reyes_et al_2022_supplementary_figure_S14.jpg

File 37 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 38 Name: Sanchez-Reyes_et al_2022_supplementary_figure_S14.tre

File 38 Description: Newick file of chronogram obtained with datelife, shown in supplementary Figure S14 (gray).

File 39 Name: Sanchez-Reyes_et al_2022_supplementary_figure_S15.jpg

File 39 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 40 Name: Sanchez-Reyes_et al_2022_supplementary_figure_S15.tre

File 40 Description: Newick file of chronogram obtained with datelife, shown in supplementary Figure S15 (gray).

File 41 Name: Sanchez-Reyes_et al_2022_supplementary_figure_S16.jpg

File 41 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 42 Name: Sanchez-Reyes_et al_2022_supplementary_figure_S16.tre

File 42 Description: Newick file of chronogram obtained with datelife, shown in supplementary Figure S16 (gray).

File 43 Name: Sanchez-Reyes_et al_2022_supplementary_figure_S17.jpg

File 43 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 44 Name : Sanchez-Reyes_etal_2022_supplementary_figure_S17.tre

File 44 Description : Newick file of chronogram obtained with datelife , shown in supplementary Figure S17 (gray) .

File 45 Name : Sanchez-Reyes_etal_2022_supplementary_figure_S18.jpg

File 45 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 46 Name : Sanchez-Reyes_etal_2022_supplementary_figure_S18.tre

File 46 Description : Newick file of chronogram obtained with datelife , shown in supplementary Figure S18 (gray) .

File 47 Name : Sanchez-Reyes_etal_2022_supplementary_figure_S19.jpg

File 47 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 48 Name : Sanchez-Reyes_etal_2022_supplementary_figure_S19.tre

File 48 Description : Newick file of chronogram obtained with datelife , shown in supplementary Figure S19 (gray) .

File 49 Name : Sanchez-Reyes_etal_2022_supplementary_figure_S20.jpg

File 49 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 50 Name : Sanchez-Reyes_etal_2022_supplementary_figure_S20.tre

File 50 Description : Newick file of chronogram obtained with datelife , shown in supplementary Figure S20 (gray) .

"

Methods that generated the data files

```
line3 <- "
METHODOLOGICAL INFORMATION
```

All age data and chronograms were obtained using the R package datelife v0.6.5
<https://CRAN.R-project.org/package=datelife> and the chronogram database
 OpenTreeChronograms v2022.1.28

"

Individual descriptions of each data file

```
var_description_1 <- c( "Node names in tree topology , as defined by taxon A and  

  taxon B. " ,
```

```

      "Scientific name of the first element of a taxon pair
        that defines a node from a tree topology.",
      "Scientific name of the second element of a taxon pair
        that defines a node from a tree topology.",
      "Ages of nodes defined by taxon A and taxon B, in
        Million years.",
      "Reference of the study that published the chronogram
        from which the node ages were extracted.")

var_description_2 <- c("Names of nodes from the tree topology used for the
  dating analysis.",
  "Minimum node age, in Million years.",
  "First quartile of node age distributions, in Million
    years.",
  "Mean node age, in Million years.",
  "Median node age, in Million years.",
  "Third quartile of node age distributions, in Million
    years.",
  "Maximum node age, in Million years.",
  "Variance of node age, in Million years.",
  "Standard deviation of node age, in Million years.")

csv_file_names <- c("Sanchez-Reyes_etal_2022_table_1.csv",
  "Sanchez-Reyes_etal_2022_table_2.csv",
  "Sanchez-Reyes_etal_2022_supplementary_table_S1.csv",
  "Sanchez-Reyes_etal_2022_supplementary_table_S2.csv")
tre_file_names <- c("Sanchez-Reyes_etal_2022_figure_1_chronogram_mock_example.
  tre")

text <- c()
var_descriptions <- list(var_description_1, var_description_2, var_description
  _1, var_description_2)
names(var_descriptions) <- csv_file_names

for (file_name in csv_file_names) {
  # read the csv file in:
  csv_data <- utils::read.csv(file = paste0("../dryad/", file_name), header =
    TRUE, row.names = NULL)
  # remove the column with row numbers:
  csv_data <- csv_data[, -1]
  # create the description text
  var_list <- gsub("[.] ", "_", colnames(csv_data))
  var_list <- paste0("\t", var_list, ":", var_descriptions[[file_name]])
  text <- c(text,
    c(paste("\nDATA-SPECIFIC INFORMATION FOR:", file_name, "\n"),
      paste("1. Number of variables/columns:", ncol(csv_data), "\n"),
      paste("2. Number of cases/rows:", nrow(csv_data), "\n"),
      "3. Variable List:", var_list,
      "\n4. Missing data codes: None",
      "\n5. Abbreviations used: NA = not applicable\n"))
}

line4 <- text

```


Writing everything to TXT file

```
writeLines(text = c(line0 ,  
                    line1 ,  
                    line2 ,  
                    line3 ,  
                    line4) , con = "../dryad/README_Sanchez-Reyes_etal_2022_  
                    data.txt")
```