

# DateLife Workflows

Luna L. Sanchez Reyes

2019-03-01

## Taxon Phyllostomidae

### I. Query data

There are 223 species in the Open Tree of Life Taxonomy for the taxon Phyllostomidae. Information on time of divergence is available for 170 of these species across 7 published and peer-reviewed chronograms from the following studies:

1. Shi, Jeff J., Daniel L. Rabosky. 2015. Speciation dynamics during the global radiation of extant bats. *Evolution* 69 (6): 1528-1545
2. Bininda-Emonds, Olaf R. P., Marcel Cardillo, Kate E. Jones, Ross D. E. MacPhee, Robin M. D. Beck, Richard Grenyer, Samantha A. Price, Rutger A. Vos, John L. Gittleman, Andy Purvis. 2007. The delayed rise of present-day mammals. *Nature* 446 (7135): 507-512
3. Hedges, S. Blair, Julie Marin, Michael Suleski, Madeline Paymer, Sudhir Kumar. 2015. Tree of life reveals clock-like speciation and diversification. *Molecular Biology and Evolution* 32 (4): 835-845
4. Lack J.B., & Van den bussche R.A. 2010. Identifying the Confounding Factors in Resolving Phylogenetic Relationships in Vespertilionidae. *Journal of Mammalogy*, .
5. Dumont E.R., Davalos L.M., Goldberg A., Santana S.E., Rex K., & Voigt C.C. 2012. Morphological innovation, diversification and invasion of a new adaptive zone. *Proceedings of the Royal Society B: Biological Sciences*, 279: 1797-1805.

All source chronograms are fully ultrametric. The proportion of Phyllostomidae queried species found across source chronograms is as follows:

|   | Trees                           | Tips | Resolved |
|---|---------------------------------|------|----------|
| 1 | Open Tree of Life Subtree       | 220  | 100%     |
| 2 | Dated Open Tree of Life Subtree | 220  | 90%      |
| 3 | Median Summary Chronogram       | 170  | 99%      |
| 4 | SDM Summary Chronogram          | 170  | 99%      |

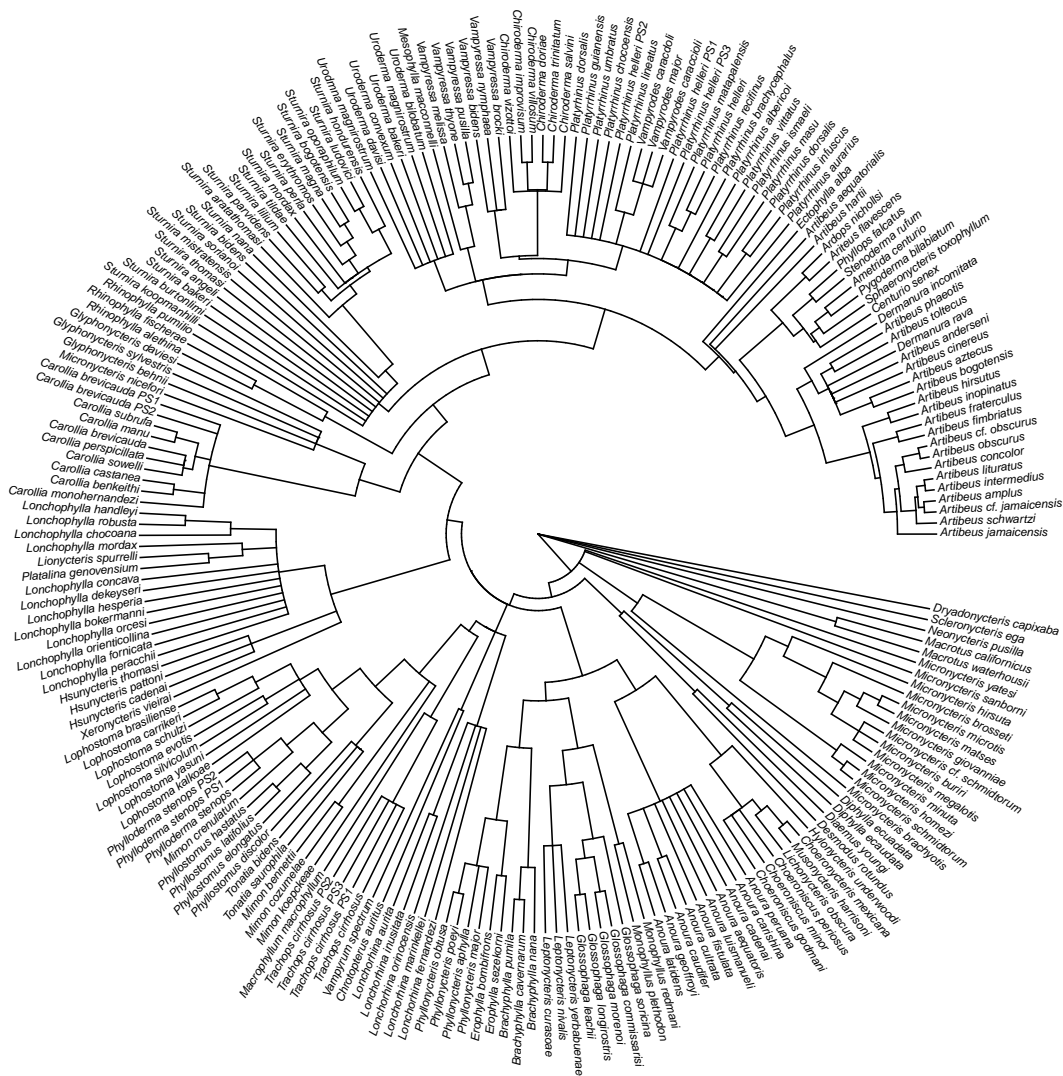


Figure 1: Phyllostomidae Species Dated Open Tree of Life Induced Subtree. This chronogram was obtained with `get_dated_oto1_induced_subtree()` function.

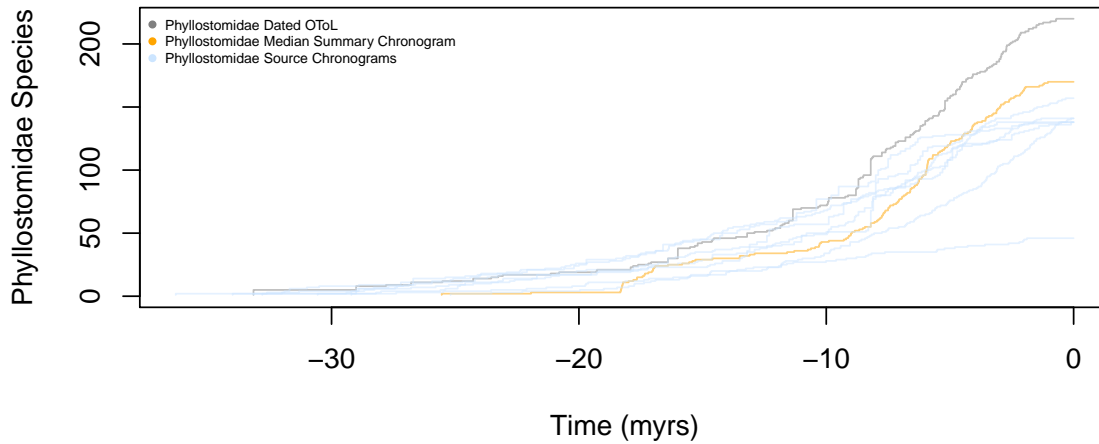


Figure 2: Phyllostomidae lineage through time (LTT) plots from source chronograms, summary median chronogram and dated Open Tree of Life chronogram.

## II. Summarize results.

### II.A. Diagnosing clustering issues.

We identified some issues with chronograms coming from SDM and Median summary matrices. First, clustering algorithms implemented to go from a summary distance matrix to a tree return trees that are too old (generally with UPGMA algorithms) or non-ultrametric (generally with Neighbour Joining algorithms). In most studied cases, UPGMA returns fully ultrametric trees but with very old ages (we had to multiply the matrix by 0.25 to get ages approximate to source chronograms ages, however this is a number chosen at random, it was just the number that worked well. NJ returned reasonable ages, but trees are way non ultrametric, as you can see in Fig. 3 and Fig. 4.

This taxon's SDM matrix has NO negative values. This taxon's Median matrix has NO negative values.

### II.B. Age distributions form Median and SDM summary trees.

Comparison of summary chronograms reconstructed with min and max ages.

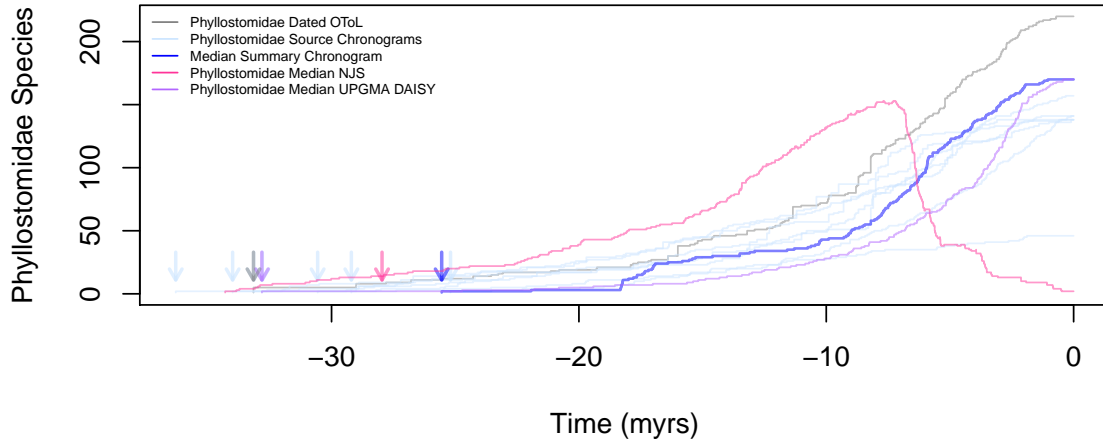


Figure 3: Phyllostomidae lineage through time (LTT) plots from source chronograms and Median summary matrix converted to phylo with different methods (NJ and UPGMA). Clustering algorithms used often are returning non-ultrametric trees or with maximum ages that are just off (too old or too young). So we developped an alternative algorithm in **datelife** to go from a summary matrix to a fully ultrametric tree.

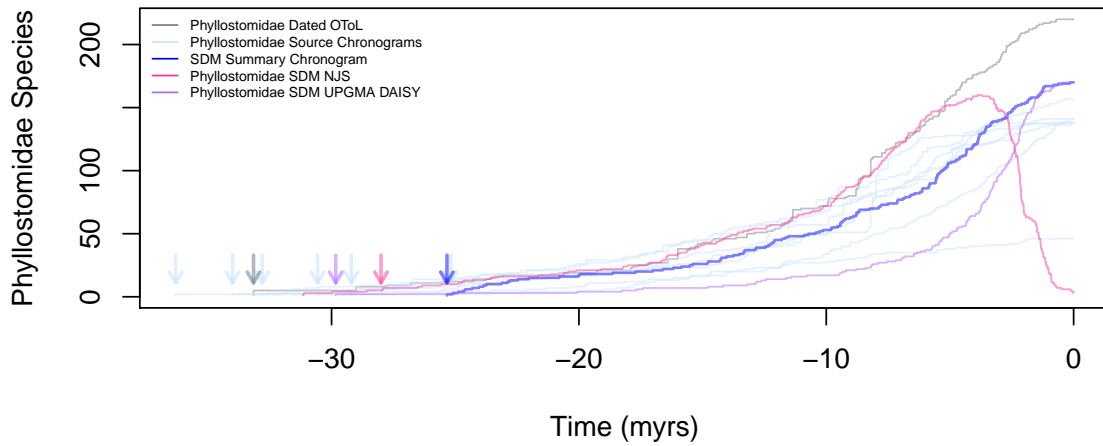


Figure 4: Phyllostomidae lineage through time (LTT) plots from source chronograms and SDM summary matrix converted to phylo with different methods (NJ and UPGMA). Clustering algorithms used often are returning non-ultrametric trees or with maximum ages that are just off (too old or too young). So we developped an alternative algorithm in **datelife** to go from a summary matrix to a fully ultrametric tree.

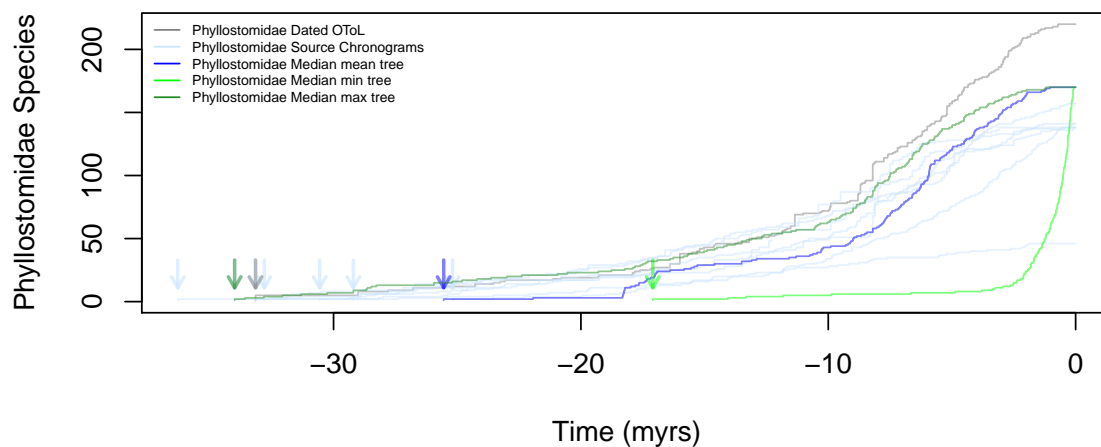


Figure 5: Phyllostomidae lineage through time (LTT) plots from source chronograms and Median summary matrix converted to phylo with `datelife` algorithm.

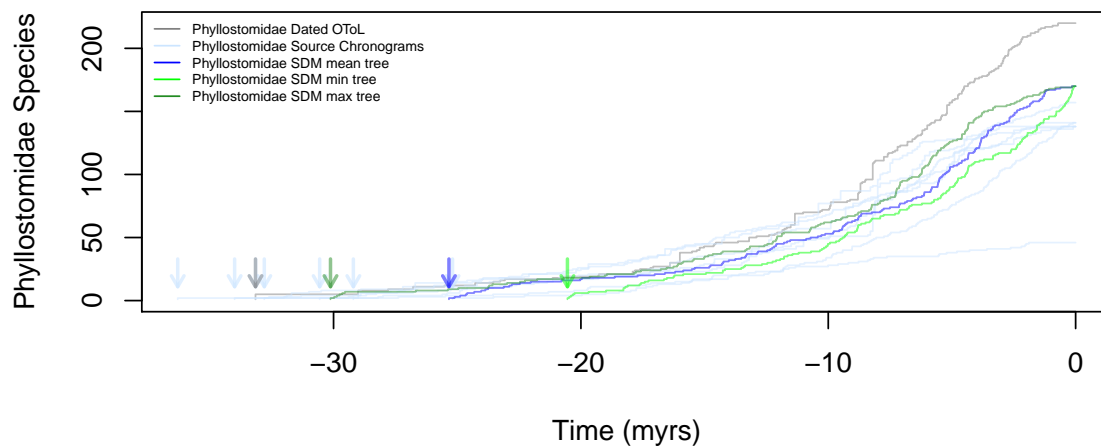


Figure 6: Phyllostomidae lineage through time (LTT) plots from source chronograms and SDM summary matrix converted to phylo with `datelife` algorithm.

### III. Create new data

As an example, we're gonna date the Open Tree Synthetic tree (mainly because the taxonomic tree is usually less well resolved.)

Now, let's say you like the Open Tree of Life Taxonomy and you want to stick to that tree. Dates from available studies were tested over the Open Tree of Life Synthetic tree of Phyllostomidae and a tree was constructed, but all branch lengths are NA. We also tried each source chronogram independently, with the Dated OToL and with each other, as a form of cross validation in Table 2. This is not working perfectly yet, but we are developping new ways to use all calibrations efficiently.

Table 2: Was it successful to use each source chronogram independently as calibration (CalibN) against the Dated Open Tree of Life (dOToL) and each other (ChronoN)?

|               | dOToL | Chrono1 | Chrono2 | Chrono3 | Chrono4 | Chrono5 | Chrono6 | Chrono7 |
|---------------|-------|---------|---------|---------|---------|---------|---------|---------|
| Calibrations1 | TRUE  | TRUE    | FALSE   | FALSE   | FALSE   | TRUE    | TRUE    | TRUE    |
| Calibrations2 | TRUE  | TRUE    | FALSE   | FALSE   | FALSE   | TRUE    | TRUE    | TRUE    |
| Calibrations3 | TRUE  | TRUE    | FALSE   | FALSE   | FALSE   | TRUE    | TRUE    | TRUE    |
| Calibrations4 | TRUE  | TRUE    | FALSE   | FALSE   | FALSE   | TRUE    | TRUE    | TRUE    |
| Calibrations5 | TRUE  | TRUE    | FALSE   | FALSE   | FALSE   | TRUE    | TRUE    | TRUE    |
| Calibrations6 | TRUE  | TRUE    | FALSE   | FALSE   | FALSE   | TRUE    | TRUE    | TRUE    |
| Calibrations7 | TRUE  | TRUE    | FALSE   | FALSE   | FALSE   | TRUE    | TRUE    | TRUE    |

### III. Simulate data

An alternative to generate a dated tree from a set of taxa is to take the available information and simulate into it the missing data. We will take the median and sdm summary chronograms to date the Synthetic tree of Life:

```
#> Error in paste0("\n![", figcap_lttplot_sdm, "](plots/", taxon, "_LTTplot_sdm.pdf)\n"): object 'fig'
#> Error in cat(lttplot): object 'lttplot' not found
```

## Appendix

The following species were completely absent from the chronogram data base: *Anoura aequatoris*, *Anoura cadenai*, *Anoura carishina*, *Anoura fistulata*, *Anoura luismanueli*, *Anoura peruana*, *Artibeus aequatorialis*, *Artibeus bogotensis*, *Artibeus cf. jamaicensis*, *Artibeus cf. obscurus*, *Carollia brevicauda* PS1, *Carollia brevicauda* PS2, *Carollia monohernandezi*, *Chiroderma vizottoi*, *Diphylla ecuadata*, *Dryadonycteris capixaba*, *Glyphonycteris behnii*, *Hsunnycteris cadenai*, *Hsunnycteris pattoni*, *Lonchophylla concava*, *Lonchophylla fornicata*, *Lonchophylla orcesi*, *Lonchophylla orienticollina*, *Lonchophylla peracchii*, *Lophostoma kalkoae*, *Lophostoma yasuni*, *Micronycteris sanborni*, *Micronycteris yatesi*, *Mimon koepckeae*, *Neonycteris pusilla*, *Phylloderma stenops* PS1, *Phylloderma stenops* PS2, *Phyllonycteris major*, *Platyrrhinus dorsalis*, *Platyrrhinus guianensis*, *Platyrrhinus helleri* PS1, *Platyrrhinus helleri* PS2, *Platyrrhinus helleri* PS3, *Sturnira angeli*, *Sturnira bakeri*, *Sturnira burtonlimi*, *Sturnira koopmanhilli*, *Sturnira mistratensis*, *Sturnira sorianoii*, *Trachops cirrhosus* PS1, *Trachops cirrhosus* PS2, *Trachops cirrhosus* PS3, *Uroderma bakeri*, *Uroderma convexum*, *Uroderma davisi*, *Uroderma magnirostrum*, *Vampyrodes caracdoli*, *Xeronycteris vieirai*