

# DateLife Workflows

Luna L. Sanchez Reyes

2019-04-30

## Taxon Anolis

### I. Query source data

There are 458 species in the Open Tree of Life Taxonomy for the taxon Anolis. Information on time of divergence is available for 302 of these species across 6 published and peer-reviewed chronograms. Original study citations as well as proportion of Anolis species found across those source chronograms is shown in Table 1.

All source chronograms are fully ultrametric.

Table 1: Anolis source chronogram studies information.

	<i>Citation</i>	<i>Source N</i>	<i>Taxon N</i>
1.	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	1	192/458
2.	Mahler, D. L., T. Ingram, L. J. Revell, J. B. Losos. 2013. Exceptional Convergence on the Macroevolutionary Landscape in Island Lizard Radiations. Science 341 (6143): 292-295.	1	98/458
3.	Pyron, R. Alexander, Frank T. Burbrink. 2013. Early origin of viviparity and multiple reversions to oviparity in squamate reptiles. Ecology Letters 17 (1): 13-21	1	205/458
4.	Steven Poe, Adrián Nieto-montes de oca, Omar Torres-carvajal, Kevin De Queiroz, Julián A. Velasco, Brad Truett, Levi N. Gray, Mason J. Ryan, Gunther Köhler, Fernando Ayala-varela, Ian Latella, 2017, 'A Phylogenetic, Biogeographic, and Taxonomic study of all Extant Species of Anolis (Squamata; Iguanidae)', Systematic Biology, vol. 66, no. 5, pp. 663-697	1	289/458
5.	Wright, April M., Kathleen M. Lyons, Matthew C. Brandley, David M. Hillis. 2015. Which came first: The lizard or the egg? Robustness in phylogenetic reconstruction of ancestral states. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution 324 (6): 504-516	1	203/458
6.	Yuchi Zheng, John J. Wiens, 2016, 'Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species', Molecular Phylogenetics and Evolution, vol. 94, pp. 537-547	1	202/458

*Source N*: Number of source chronograms reported in study.

*Taxon N*: Number of queried taxa found in source chronograms.

Source chronograms maximum age range from 47.843 to 72.578 million years ago (MYA). As a means for comparison, lineage through time plots of all source chronograms available in data base are shown in Fig. 1

## II. Summarize results.

LTT plots are a nice way to visually compare several trees. But what if you want to summarize information from all source chronograms into a single summary chronogram?

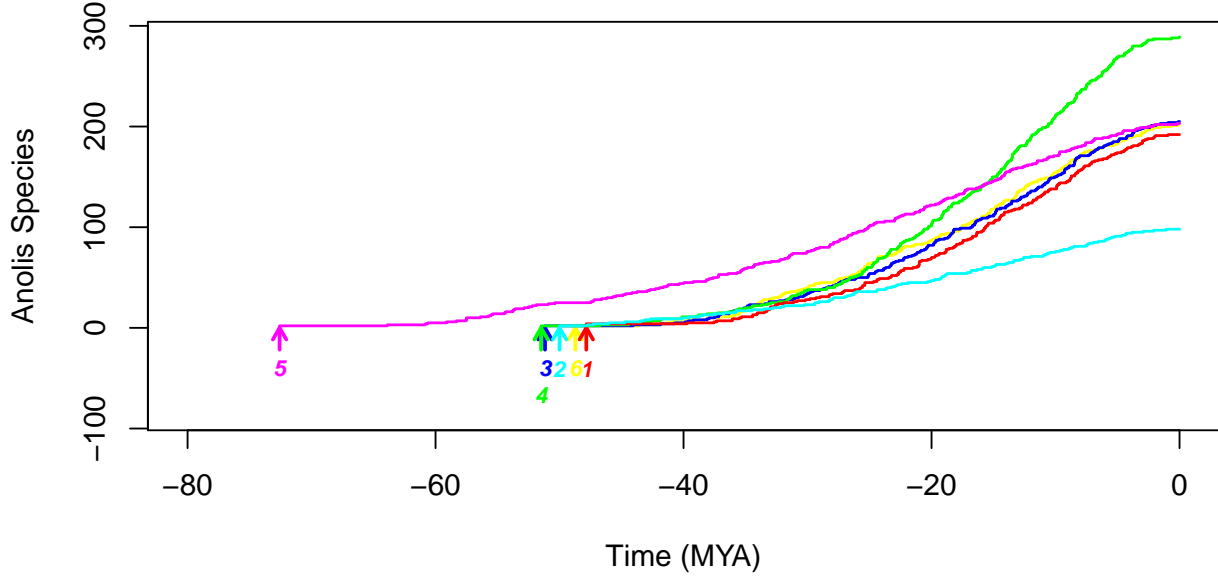


Figure 1: Lineage through time (LTT) plots of source chronograms available in data base for species in the Anolis. Numbers correspond to original studies in Table 1. Arrows indicate maximum age of each chronogram.

The first step is to identify the degree of species overlap among your source chronograms: if each source chronogram has a unique sample of species, it will not be possible to combine them into a single summary chronogram. To identify the set of trees or *grove* with the most source chronograms that have at least two overlapping taxa, we followed Ané et al. 2016. In this case, not all source chronograms found for the Anolis have at least two overlapping species. The largest grove has 2 chronograms (out of 6 total source chronograms). Now that we have identified a suitable grove we can go on to summarize it by translating the source chronograms into patristic distance matrices and then averaging them into a single summary matrix; yes, this first step is *that* straightforward. We can average the source matrices by simply using the mean or median distances, or we can use methods that involve transforming the original distance matrices –such as the super distance matrix (SDM) approach of Criscuolo et al. 2006– by minimizing the distances across source matrices.

Because our summary matrix is basically a distance matrix, a distance-based clustering algorithm could be used to reconstruct the tree. Algorithms such as neighbour joining (NJ) and unweighted pair group method with arithmetic mean (UPGMA) are fast and work well when there are no missing values in the matrices. However, summary matrices coming from source chronograms usually have several NAs and missing rows. When this happens, clustering algorithms that have been developed to deal with missing values do not work well, as shown in the following section. This is probably because these methods are usually applied to distance matrices that represent evolutionary distance in terms of substitution rate and not absolute time, as is the case in here.

### II.A. Detecting clustering issues.

We tested several clustering algorithms on summary distance matrices coming from median and SDM. UPGMA returns ultrametric trees that are considerably older than source chronograms. Even scaling the distance matrix down by a factor of 0.5 would not produce trees with ages that are coherent with the

source chronograms. NJ returned trees with reasonable ages, but trees are way non ultrametric, as you can see in Fig. S1 and Fig. 2.

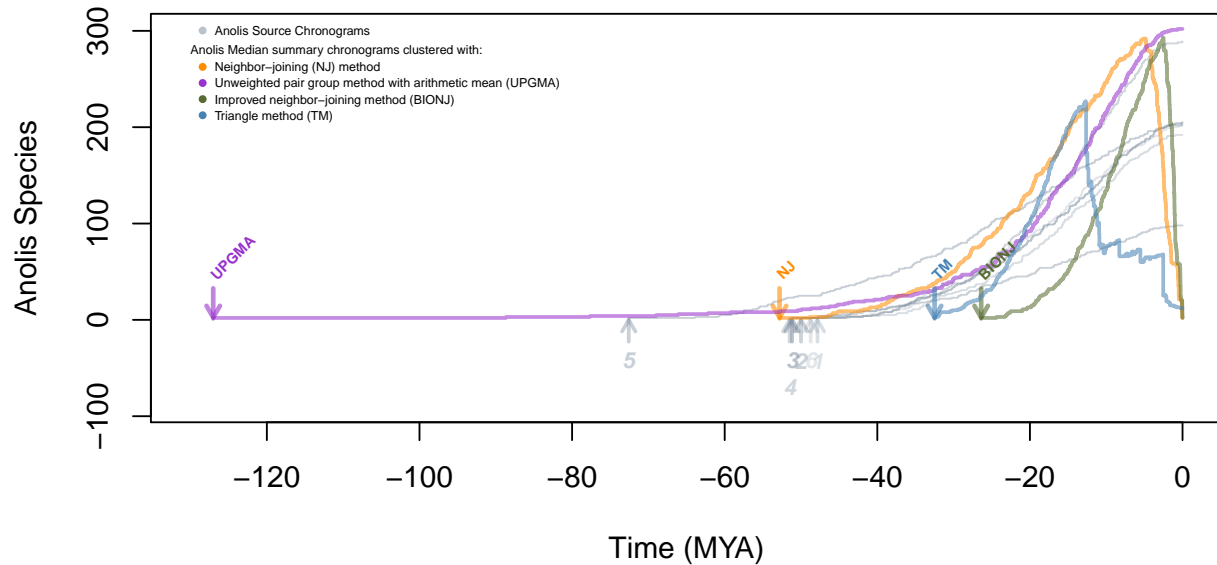


Figure 2: Lineage Through Time plots of Anolis summary chronograms from median (upper) and SDM (lower) summary matrices obtained with various clustering algorithms.

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

Comparison of summary chronograms reconstructed with min and max ages.

```
#> Error in figcap_lttplot_summ[[i]] <- paste(taxon, "lineage through time (LTT) plots from source chron
#> Error in paste0("\n!", figcap_lttplot_summ[[2]], ") (plots/", taxon, "_LTTplot_summtrees_Median.pdf)
#> Error in cat(lttplot_median): object 'lttplot_median' not found

#> Error in paste0("\n!", figcap_lttplot_summ[[1]], ") (plots/", taxon, "_LTTplot_summtrees_SDM.pdf)\n")
#> Error in cat(lttplot_sdm): object 'lttplot_sdm' not found
```

### 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 Anolis 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
Calibrations1	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Calibrations2	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Calibrations3	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Calibrations4	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Calibrations5	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Calibrations6	TRUE	TRUE	TRUE	TRUE	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:

## Appendix

The following species were completely absent from the chronogram data base: *Anolis alocomyos*, *Anolis altavelensis*, *Anolis anfiloquioi*, *Anolis anisolepis*, *Anolis attenuatus*, *Anolis bellipeniculus*, *Anolis birama*, *Anolis breedlovei*, *Anolis caquetae*, *Anolis carlliebi*, *Anolis carlostoddi*, *Anolis cf. alocomyos GK-2015*, *Anolis cf. humilis JJK-2013*, *Anolis cf. polylepis*, *Anolis charlesmyersi*, *Anolis chlorocyaneus*, *Anolis chrysops*, *Anolis concolor*, *Anolis cuscoensis*, *Anolis damulus*, *Anolis delafuentei*, *Anolis deltae*, *Anolis desiradei*, *Anolis dissimilis*, *Anolis eewi*, *Anolis fairchildi*, *Anolis forbesorum*, *Anolis fugitivus*, *Anolis gibbiceps*, *Anolis haguei*, *Anolis ibague*, *Anolis immaculogularis*, *Anolis impetigosus*, *Anolis incredulus*, *Anolis juangundlachi*, *Anolis kreutzi*, *Anolis laevis*, *Anolis lamari*, *Anolis landestoyi*, *Anolis leditzigorum*, *Anolis lemniscatus*, *Anolis limon*, *Anolis menta*, *Anolis microlepis*, *Anolis mirus*, *Anolis morazani*, *Anolis muralla*, *Anolis nasofrontalis*, *Anolis nietoi*, *Anolis nigrolineatus*, *Anolis osa*, *Anolis paravertebralis*, *Anolis philopunctatus*, *Anolis phyllorhinus*, *Anolis pigmaequestrus*, *Anolis pijolense*, *Anolis pinchoti*, *Anolis propinquus*, *Anolis pseudotigrinus*, *Anolis purpurescens*, *Anolis radulinus*, *Anolis rhombifer*, *Anolis rimarum*, *Anolis rivalis*, *Anolis roosevelti*, *Anolis ruibali*, *Anolis ruizi*, *Anolis rupinae*, *Anolis sacamecatensis*, *Anolis santamartae*, *Anolis scapularis*, *Anolis schiedei*, *Anolis schmidtii*, *Anolis sierramaestrae*, *Anolis simmonsii*, *Anolis solitarius*, *Anolis spectrum*, *Anolis squamulatus*, *Anolis stevepoei*, *Anolis tenorioensis*, *Anolis terueli*, *Anolis tetarii*, *Anolis toldo*, *Anolis umbrivagus*, *Anolis utowanae*, *Anolis vaupesianus*, *Anolis vescus*, *Anolis vicarius*, *Anolis villai*, *Anolis wampuensis*, *Anolis wattsii*, *Anolis wermuthi*, *Anolis williamsii*, *Anolis zapotecorum*, *Ctenonotus cybotes*, *Ctenonotus ferreus*, *Ctenonotus pulchellus*, *Norops alvarezdeltoroi*, *Norops anisolepis*, *Norops antonii*, *Norops baccatus*, *Norops birama*, *Norops bocourti*, *Norops bouvierii*, *Norops breedlovei*, *Norops chrysolepis*, *Norops cobanensis*, *Norops concolor*, *Norops conspersus*, *Norops cumingii*, *Norops cuprinus*, *Norops cymbops*, *Norops damulus*, *Norops delafuentei*, *Norops dollfusianus*, *Norops exsul*, *Norops forbesi*, *Norops fungosus*, *Norops gibbiceps*, *Norops granuliceps*, *Norops haguei*, *Norops hobartsmithi*, *Norops ibague*, *Norops imias*, *Norops johnmeyerii*, *Norops lemniscatus*, *Norops lineatus*, *Norops lynchi*, *Norops macrolepis*, *Norops macrophallus*, *Norops maculiventris*, *Norops mariarum*, *Norops matudai*, *Norops microlepis*, *Norops milleri*, *Norops notopholis*, *Norops opalinus*, *Norops parvicirculatus*, *Norops pentaprion*, *Norops petersii*, *Norops pinchoti*, *Norops pygmaeus*, *Norops rhombifer*, *Norops rivalis*, *Norops salvini*, *Norops scapularis*, *Norops schiedei*, *Norops schmidtii*, *Norops simmonsii*, *Norops sulcifrons*, *Norops tolimensis*, *Norops utowanae*, *Norops vicarius*, *Norops villai*, *Norops vittigerus*, *Norops vociferans*

This taxon's SDM matrix has some negative values in the following taxa: *Anolis porcus*, *Anolis guamuhaya*, *Chamaeleolis chameleontides*. This taxon's Median matrix has NO negative values.

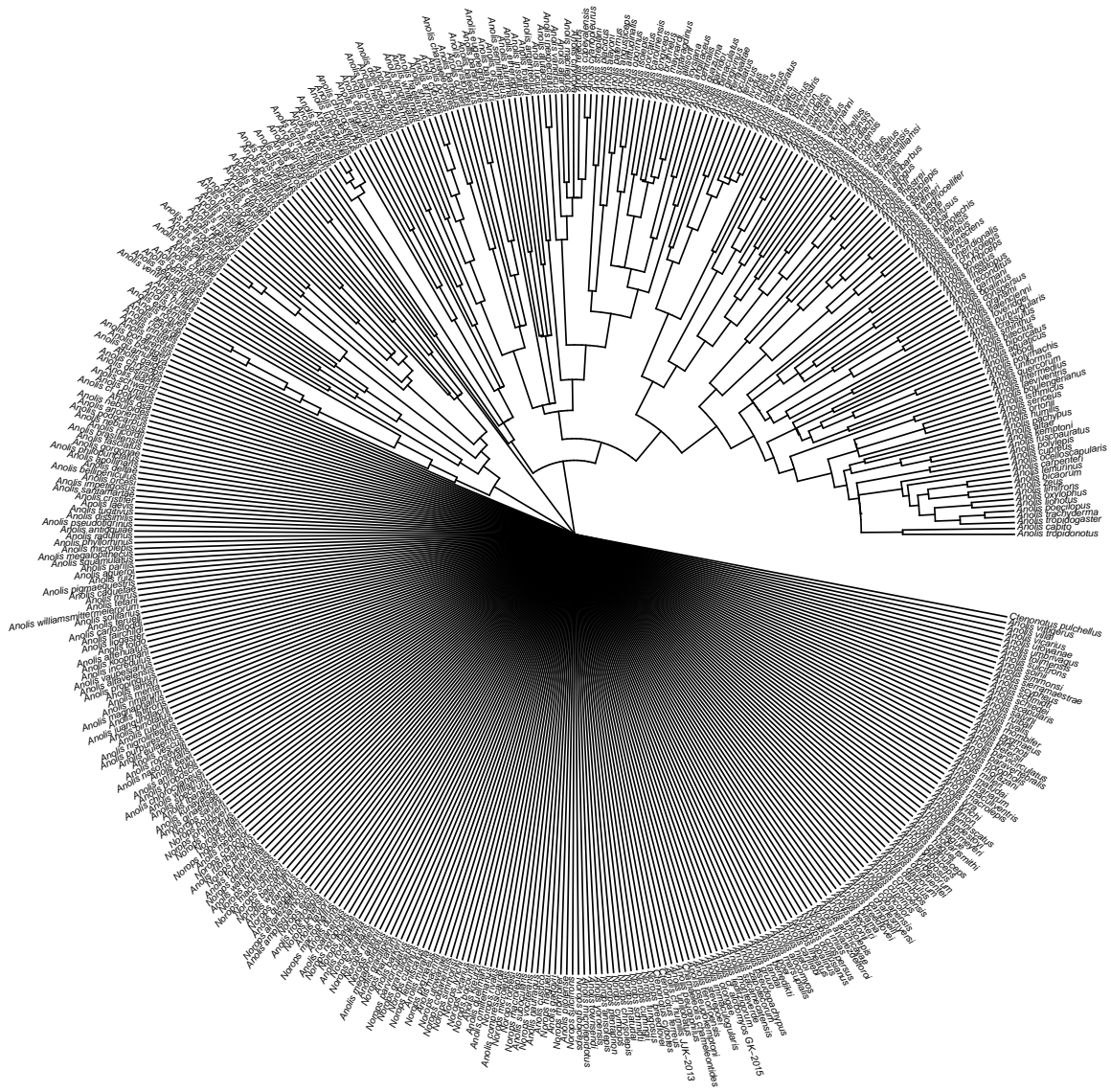


Figure 3: Anolis Species Dated Open Tree of Life Induced Subtree. This chronogram was obtained with `get_dated_otol_induced_subtree()` function.

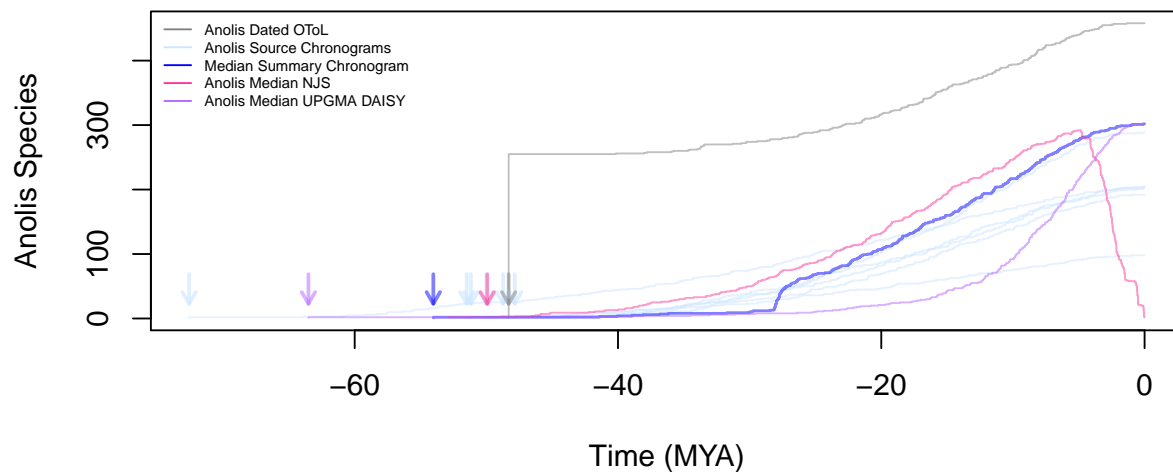


Figure 4: Anolis lineage through time (LTT) plots from source chronograms and Median summary matrix converted to phylo with different methods (NJ and UPGMA). Clustering algorithms used often are returning non-ultrametric trees or with maximum ages that are just off (too old or too young). So we developed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

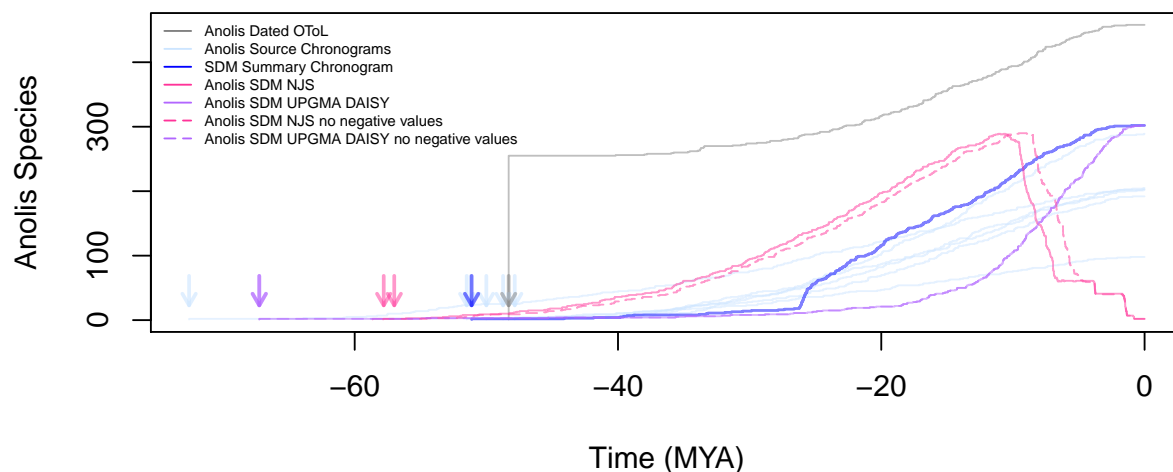


Figure 5: Anolis lineage through time (LTT) plots from source chronograms and SDM summary matrix converted to phylo with different methods (NJ and UPGMA). As you can note, dashed lines and solid lines from trees coming out from both types of clustering algorithms implemented are mostly overlapping. This means that removing negative values does not change results from clustering algorithms much. Clustering algorithms used often are returning non-ultrametric trees or with maximum ages that are just off (too old or too young). So we developed an alternative algorithm in `datelife` to go from a summary matrix to a fully ultrametric tree.