



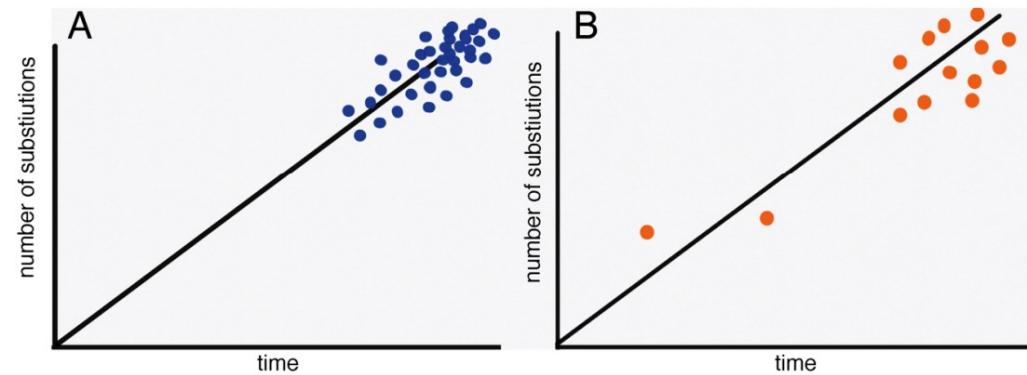
# Tutorial: TempEst and Temporal signal

Ash Porter

Phylogenetics workshop, University of Melbourne, 2021

# Temporal signal

- We often have genetic sequences that have been sampled over timescales of months to years: “temporally sampled sequence data”.
- Two sampling times are “evolutionary distinct” when the sequences have a **measurable** amount of substitution change within the population (1).
  - Can happen in RNA virus populations within months (seen in A).
  - Can use ancient DNA and contemporary sequences to measure evolutionary change over hundreds of years (seen in B).



# TempEst



When using sequence data that has been sampled through time:

Sampling times can be used to calibrate the molecular clock.

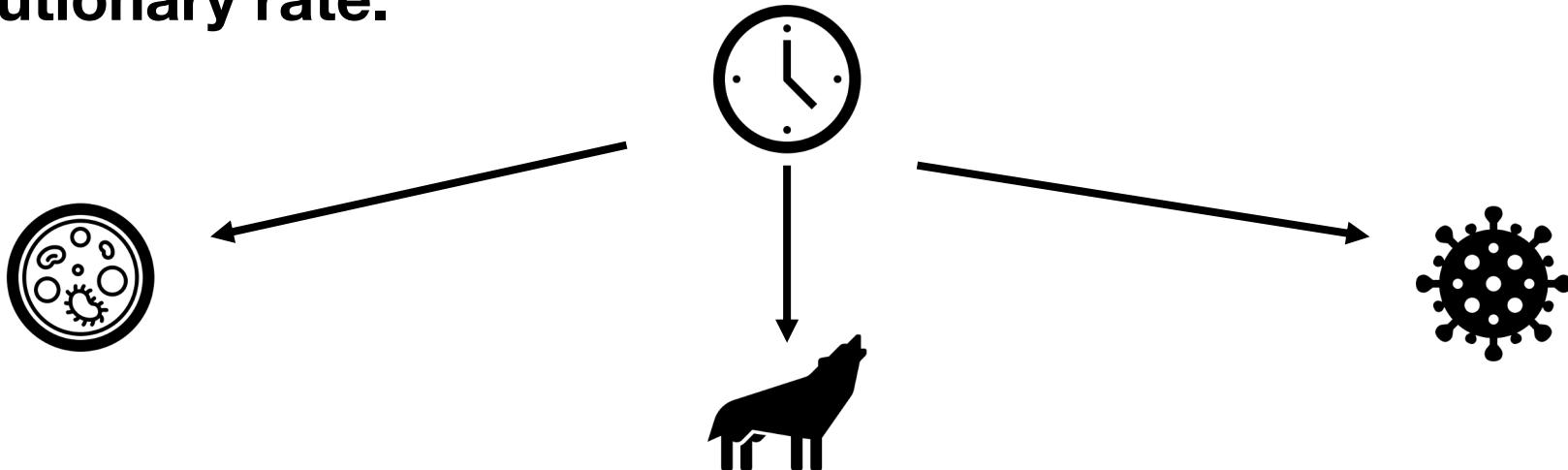
Used to infer the evolutionary rate and time-scale of the population.



## What is temporal signal?

We assume molecular evolution occurs over time at a predictable rate, and that our sampling period is sufficiently large enough to capture measurable evolutionary change.

- TempEst software (1) can be used to examine the temporal signal of sequences:
  - To identify **outliers**, and to get a rough estimate of the **TMCRA** and **evolutionary rate**.



# TempEst



When using sequence data that has been sampled through time:

Sampling times can be used to calibrate the molecular clock.

Used to infer the evolutionary rate and time-scale of the population.

## What is temporal signal?

We assume molecular evolution occurs over time at a predictable rate, and that our sampling period is sufficiently large enough to capture measurable evolutionary change

- TempEst software (2) can be used to examine the temporal signal of sequences:
  - To identify **outliers**, and to get a rough estimate of the **TMCRA** and **evolutionary rate**.

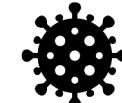
An organism that rapidly evolves, but only has been sampled very briefly.



An organism that has been sampled over a long period of time, but evolves very slowly.

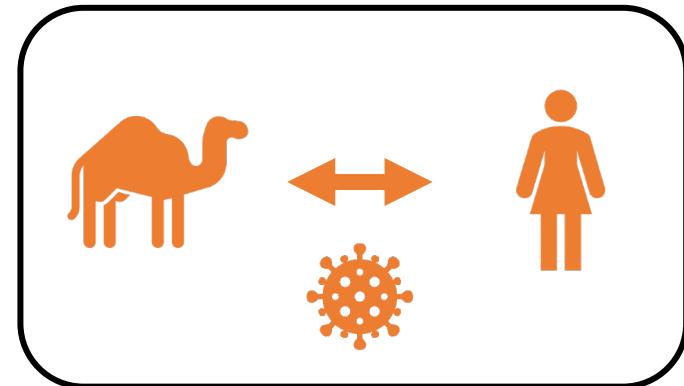


An organism that rapidly evolves, and has been sampled from a sufficient window of time.  
☐ measurably evolving population!



# TempEst

- Using a dataset of MERS coronaviruses.  
→ (n=301)
- Whole genome sequences (29,530 bp).
- Collected from human and camel hosts.



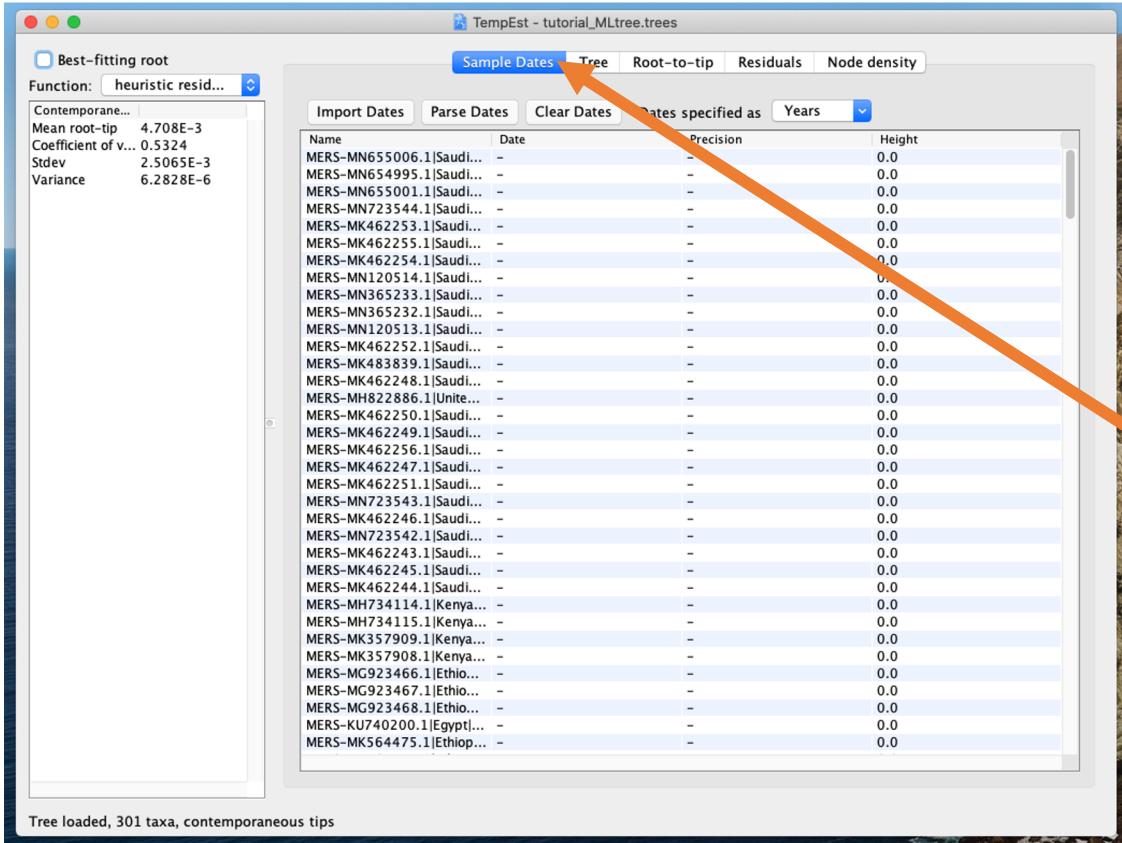
# TempEst: Materials

```
iqtree2 -s MERSTutorial.fasta -st DNA -pre tutorial_MLtree -nt AUTO -m HKY+G -bb 1000
```

- You will need a “non-clock” phylogenetic tree.
- Maximum likelihood methods are commonly used such as IQTree (3).

We will be using “tutorial\_MLtree.trees”.

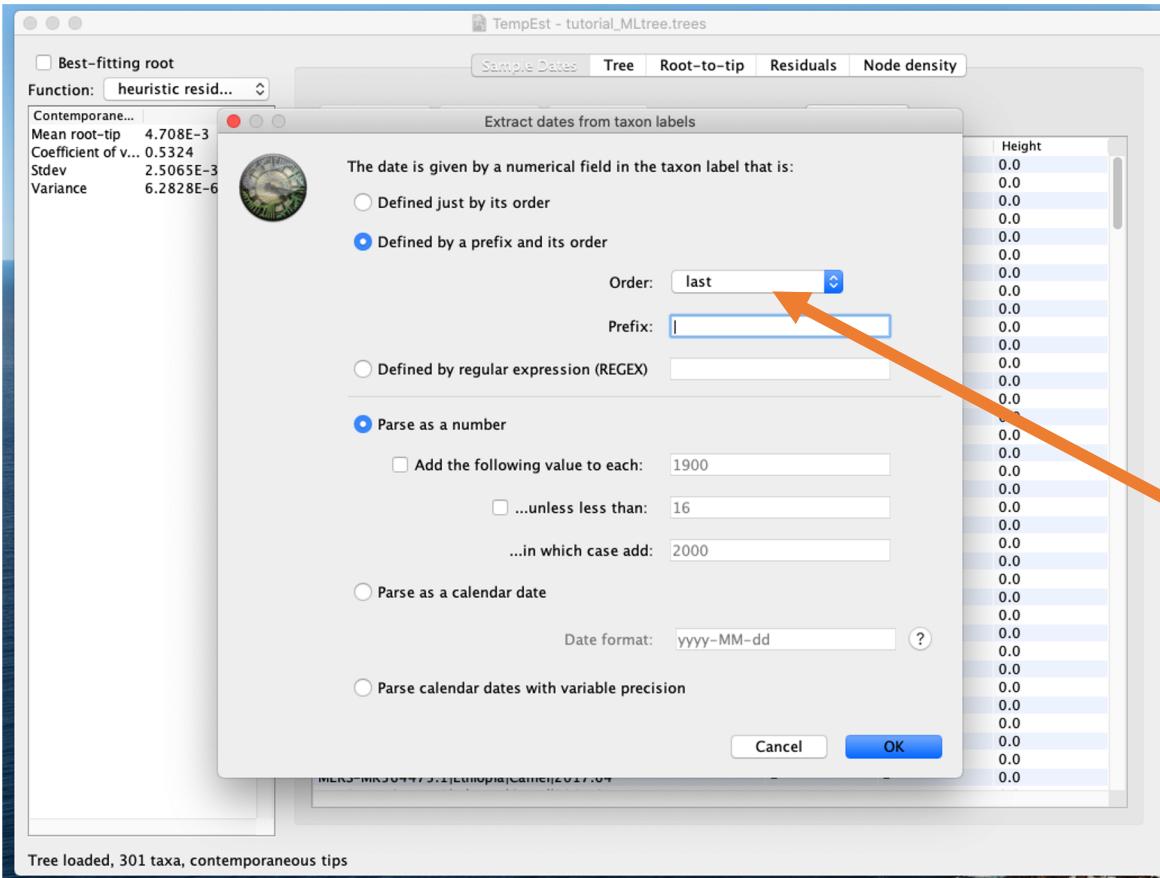
# TempEst: Importing tree



Click on the TempEst icon and load the tree file.

We first need to parse the sample dates from the label.

# TempEst: Setting dates



**Want to make this easier?**

When setting up your data, make sure that the taxon label has a delimiter before the date!

The date follows the last “|” delimiter in the label.

# TempEst: Setting dates



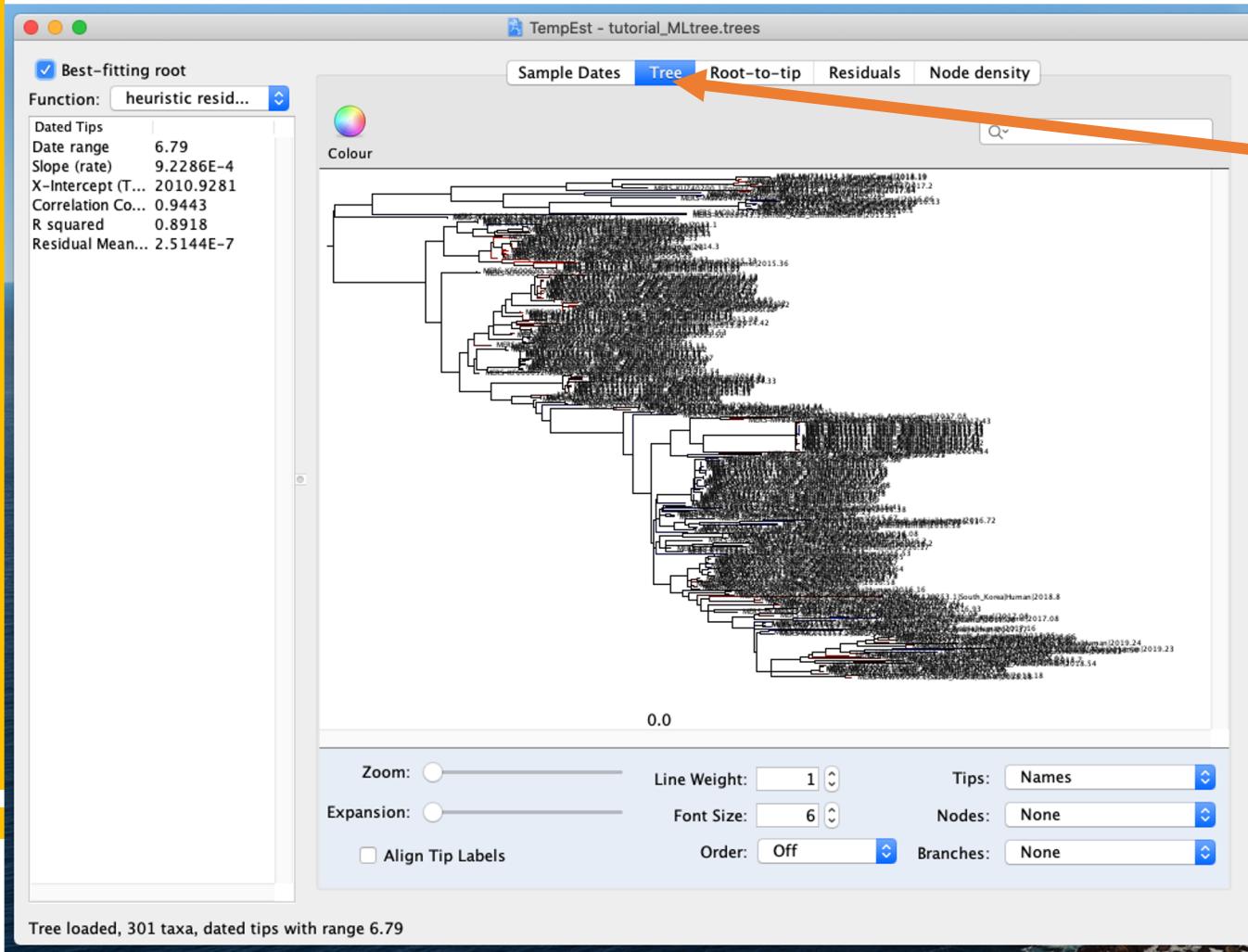
The screenshot shows the TempEst software interface with a dated tree titled "tutorial\_MLtree.trees". The tree has 301 taxa and a date range of 6.79. The main window displays a table of node details, including Name, Date, Precision, and Height. The table lists numerous entries, mostly MERS strains from Saudi Arabia and Ethiopia, with dates ranging from 2014.96 to 2018.64. The software interface includes tabs for Sample Dates, Tree, Root-to-tip, Residuals, and Node density, and a sidebar with various parameters like Best-fitting root, Function (heuristic resid...), Dated Tips (6.79), Date range (-0.0005), X-Intercept (2025.9081), Correlation Co. (-0.2808), R squared (7.8842E-2), and Residual Mean (5.8068E-6).

Tree loaded, 301 taxa, dated tips with range 6.79

## Tip!

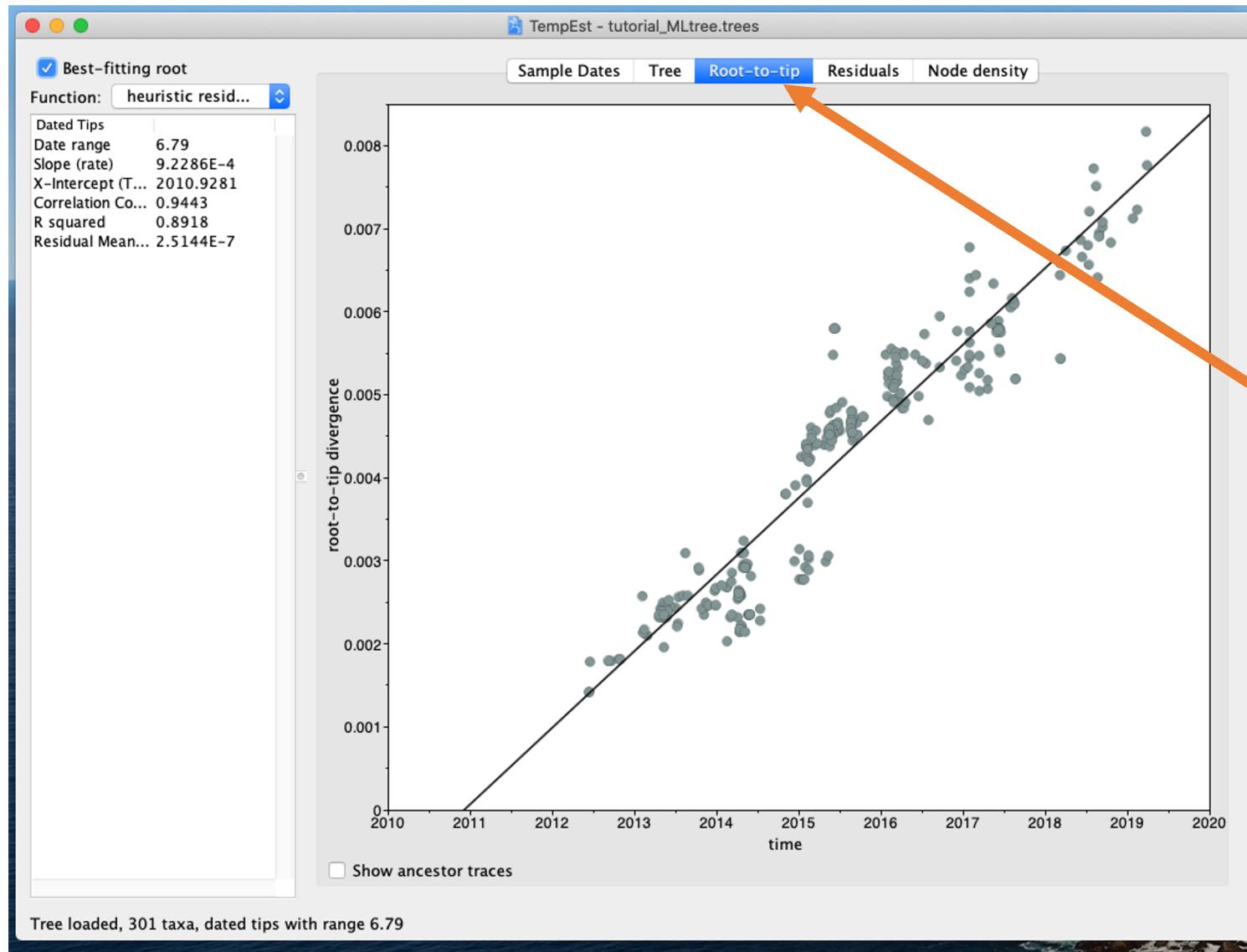
You don't have to use decimal dates— you can use many different formats, including day/month/year format.

# TempEst: Best fitting root



Now we can look at the tree panel and ask TempEst to find the “best-fitting root” of the tree.

The tree branches will have a colour associated with the residual: **blue** for tips with positive residuals (above the regression line), **red** for negative



### What is root-to-tip divergence?

“Root-to-tip” is a plot of the divergence from the root of the tree against time of sampling.

Now we can look at the root-to-tip divergence and see how our data fits along the regression.

# TempEst: TMRCA and rate



- Date range:
  - The span of dates from the earliest sample to latest sample.
- Slope (rate):
  - The slope of the regression line, which is a rough estimate of the rate of evolution (substitutions/site/year).
- X-Intercept (TMRCA):
  - The intercept on the x-axis at which the regression line crosses, which is a rough estimate of the date of the root of the tree.

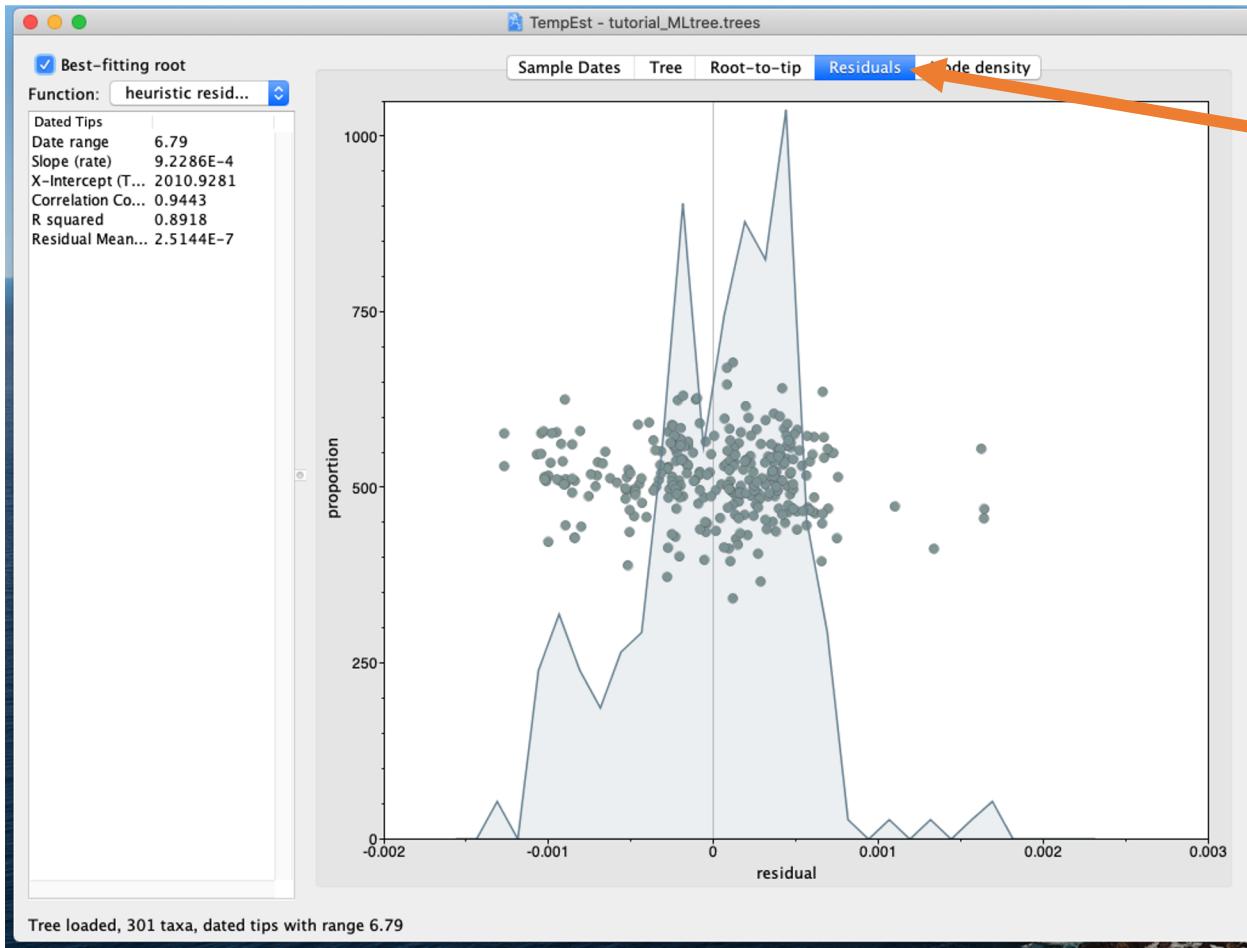
## Tip!

If the rate does not appear "normal" for your dataset, you might need to have a closer look at your data.

## Dated Tips

Date range	6.79
Slope (rate)	9.2286E-4
X-Intercept (T...	2010.9281
Correlation Co...	0.9443
R squared	0.8918
Residual Mean...	2.5144E-7

# TempEst: Residuals



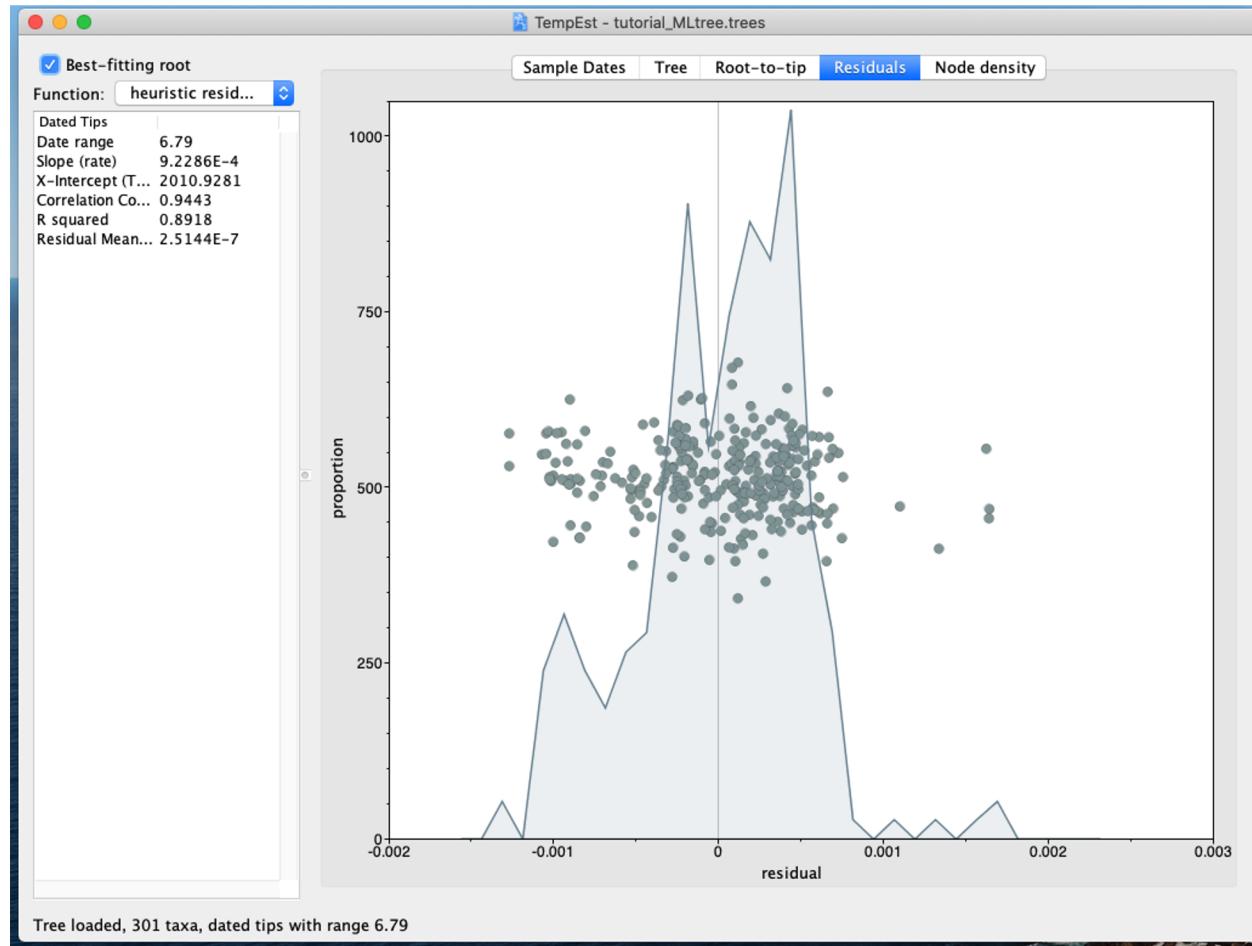
## Problematic data?

The residual plot is a handy way to spot any potential outliers. If you click on the dot, it highlights the sequence in the tree panel.

Now we can look at the residuals panel.

- The residual plot is the deviation from the regression line in the root-to-tip plot.
- If you click on the dot, it highlights the sequence/s in the tree.

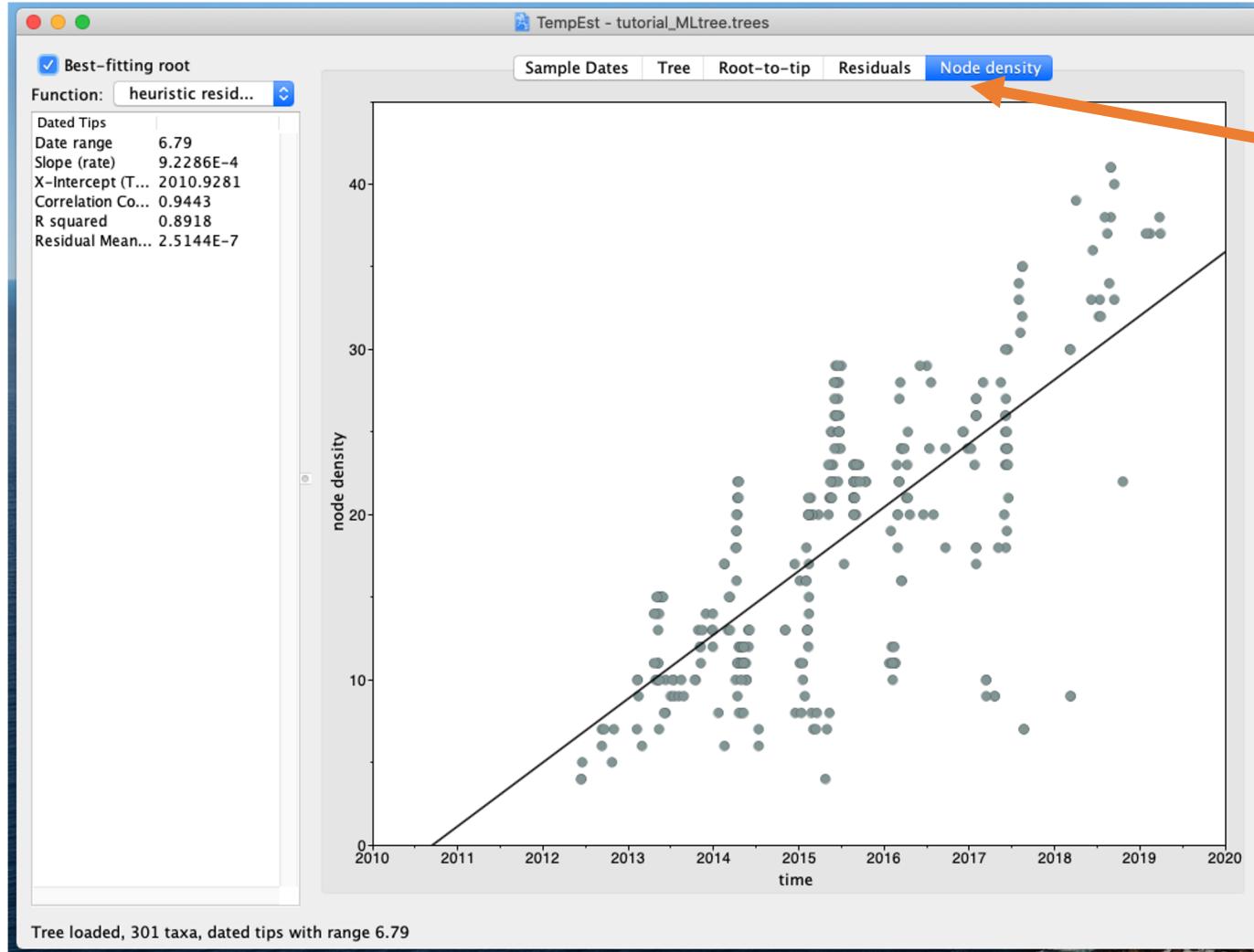
# TempEst: Residuals



## Data quality matters!

If you have an obvious outlier, it could be labelled incorrectly. Public databases often have sampling dates that might not be accurate.

# TempEst: Node density



Lastly, we can check the node density panel.

# References and further reading:

1. Drummond, A.J. et al. (2003) Measurably Evolving Populations. Trends in Ecology and Evolution, 18: 481–8
2. Rambaut, L. et al (2016) Exploring the temporal structure of heterochronous sequences using TempEst. Virus Evolution, 2: vew007
3. Available at <http://www.iqtree.org/>.