

Interpreting results and summarising trees: practical session

Data used in this prac:

BEAST output files of SARS-CoV-2 data collected up until Feb15

feb15.log

feb15.trees

Inspecting traces

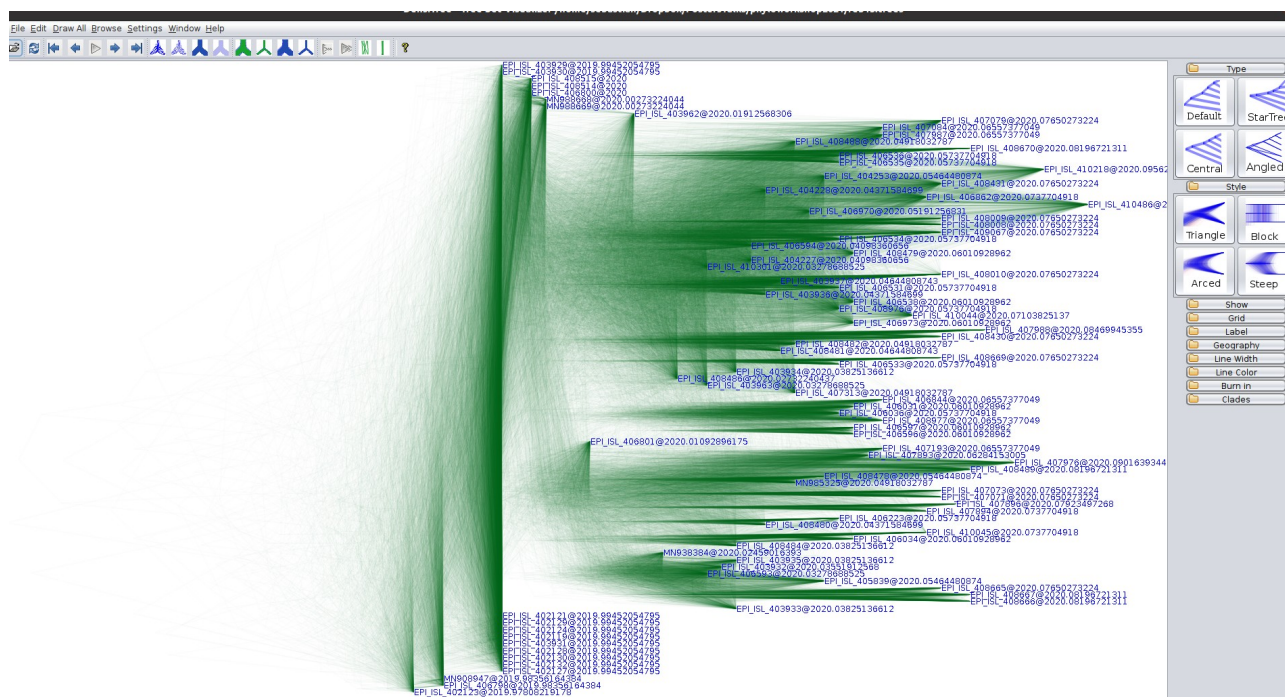
Step 1.1: Load the `feb15.log` file in Tracer or Beast and check the effective sample size for all traces.

Questions:

- Is the effective sample size (ESS) 'sufficiently high' for all traces?
- By looking at the parameters logged here, can you describe what model is being used here? (tip: try to identify the tree prior, the substitution model, and the clock model by using the trace names)
- Do the data seem to support a strict molecular clock model? Why?
- Is there evidence of population size growth in this data set? How much posterior support do you find?
- The last sample collected in this data set was taken in February 15 2020. Can you identify the date of emergence of the virus according to these data and model?

2) Inspecting trees

Step 2.1: Open the DensiTree package (usually within the BEAST2 folder, within the bin directory). Go to File, Load, and find the `feb15.trees` file and click Open. You should see the window below appear:



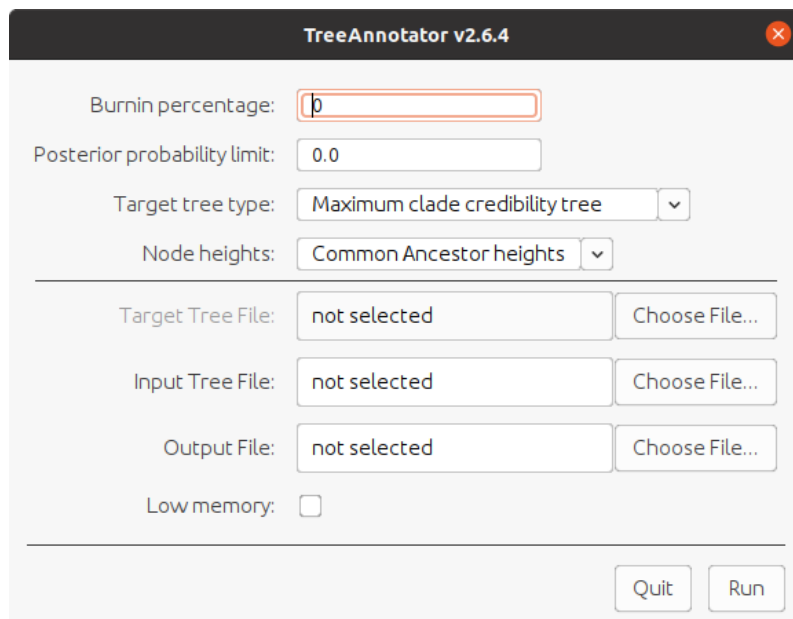
Step 2.2: Explore different visualisation options, such as the Blocked display.

Questions:

- Does the tree appear to be well supported?
- How much uncertainty (qualitatively) do think there is for some of the internal divergence events?

3. Summarising trees

Step 3.1: Open the treeannotator within your BEAST2 folder. A window will pop up:

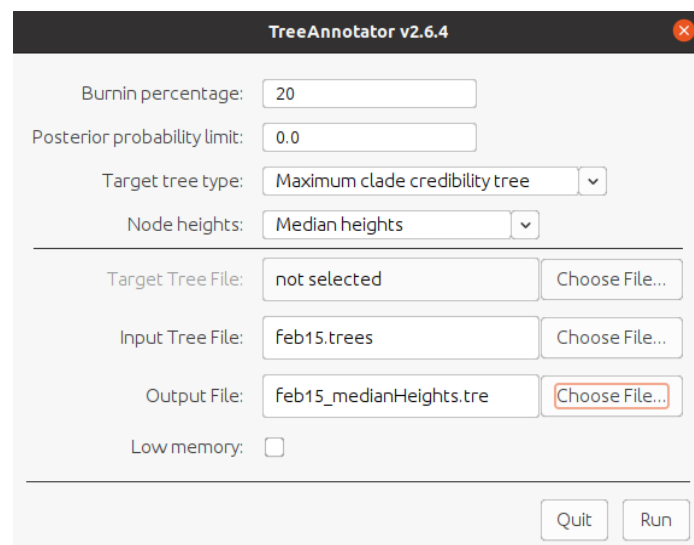


The screenshot shows the TreeAnnotator v2.6.4 window with the following settings:

- Burnin percentage: 0
- Posterior probability limit: 0.0
- Target tree type: Maximum clade credibility tree
- Node heights: Common Ancestor heights
- Target Tree File: not selected
- Input Tree File: not selected
- Output File: not selected
- Low memory: ☐

Buttons at the bottom: Quit, Run

Step 3.2: Select Burnin percentage to 20, for *Target tree type* select *Maximum clade credibility tree* and for *Node heights* select *Median heights*. For *Input Tree File* choose the `feb15.trees` and for *Output File*, choose an informative name (e.g. `feb15_medianHeights.tre`; note the `.tre` extension to distinguish it from `.trees` files). Click on *Run* when your window looks like the one below:



The screenshot shows the TreeAnnotator v2.6.4 window with the following settings:

- Burnin percentage: 20
- Posterior probability limit: 0.0
- Target tree type: Maximum clade credibility tree
- Node heights: Median heights
- Target Tree File: not selected
- Input Tree File: feb15.trees
- Output File: feb15_medianHeights.tre
- Low memory: ☐

Buttons at the bottom: Quit, Run

When the program has run it will produce the `feb15_medianHeights.tre` file in the location you chose.

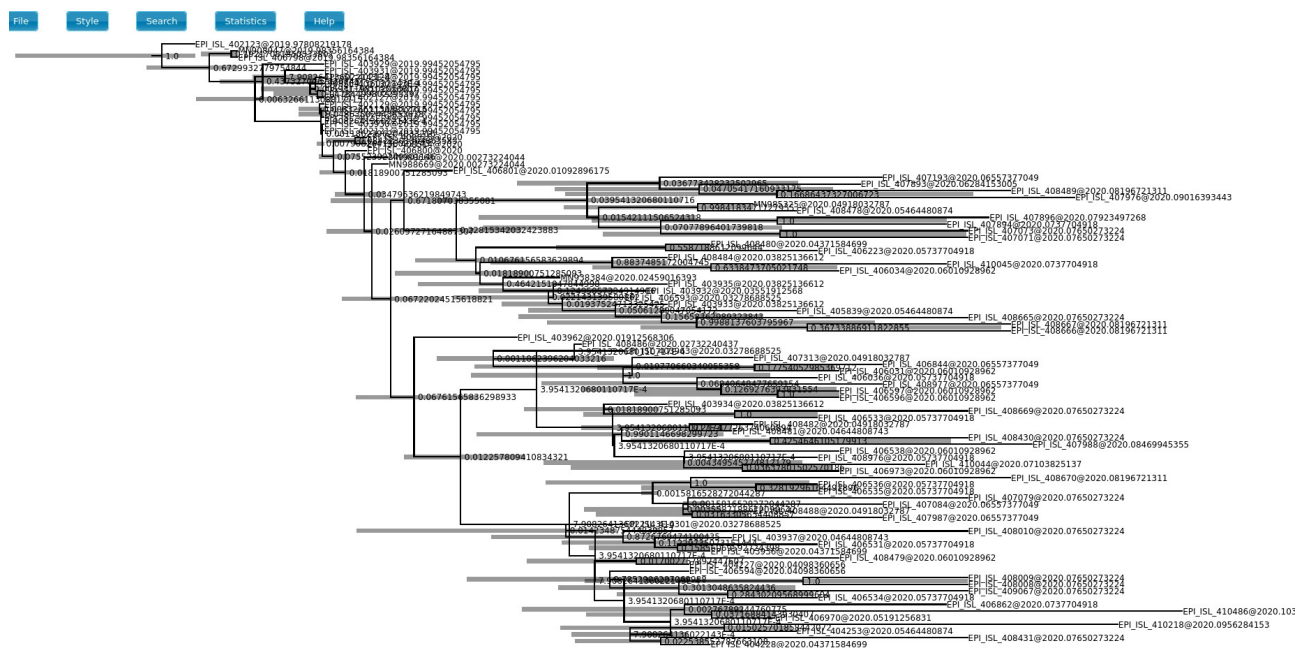
Step 3.3: Open a browser window and go to icytree.org. When the website loads drag and drop the `feb15_medianHeights.tre` file. If this does not work, load the tree via *File, Load from file*.

Step 3.4: When the tree loads, click on *Style, Internal node text*, and select *posterior*. You should see the posterior probability of nodes appear. Zoom by scrolling with your mouse to look at the numbers closely.

Question:

- Is our tree well supported overall? Does this match our expectation from the visualisation of the trees in [DensiTree](#)?

Step 4.5: Click *Style, Node height error bars*, and select *height_95%_HPD*. Your tree should look like the one I have below:



Questions:

- Do all internal nodes fall within the HPD intervals?
- Can you identify any negative branch lengths?

Step 4.6: Repeat this procedure by selecting *Keep target heights* and *Common Ancestor Heights* and inspect the trees.

Question:

- List three key differences between the summary trees.