Viral Bioinformatics 2022
Topic: Beast
Instructors: Marta Giovanetti and Carolina Torres
<u>Hands-on exercises</u>

1. *Loading the sequence data file*
   Input file: ZIKV.fasta
   LAC-2022/Typing-tool_Data-Visualization/1.Viral_typing_tool/2.Hands-on/ZIKV.fasta

To load the alignment, simply select the Import Data... option from the File menu and browse to the 'YFV_sequences.fasta' file to load it. This file contains an alignment of 65 YFV genomes 10,236 nucleotides in length. Once loaded, the sequence data will be listed under Partitions as shown in the figure:

2. *Specifying the sampling date information*

Select the box labelled Use tip dates. The actual sampling time in years is encoded in the name of each taxon and we could simply edit the value in the Date column of the table to reflect these. However, if the taxa names contain the calibration information, then a convenient way to specify the dates of the sequences in BEAUti is to use the Parse Dates button at the top left of the Data panel. Clicking this will make a dialog box appear:

3. *Setting the evolutionary model.*

For this tutorial HKY + Gamma and select the 3 partitions: positions 1, 2, 3 option so that each codon position has its own HKY substitution model, rate of evolution, Estimated base frequencies, and Gamma-distributed rate variation among sites:

4. *Setting the clock model*

Click on the Clocks tab at the top of the main window. We will perform our initial run using the (default) strict molecular clock model:

5. *Setting the starting tree and tree prior*

Click on the Trees tab at the top of the main window. We keep a default random starting tree and a skyline demographic model.

6. *Setting the MCMC options*

The MCMC tab in BEAUti provides settings to control the MCMC chain. Firstly, we have the Length of chain. This is the number of steps the MCMC will make in the chain before finishing. How long this should depend on the size of the dataset, the complexity of the model and the precision of the answer required. The default value of 10,000,000 is entirely arbitrary and should be adjusted according to the size of your dataset. We will see later how the resulting log file can be analyzed using Tracer to examine whether a particular chain length is adequate.

7. *Saving and Loading BEAUti files*

8. *Generating the BEAST XML file*

9. *Repeat the same steps by setting the Uncorrelated Relaxed Clock in the Clocks tab*

10. *Running the 2-beast xml files*

11. *Analyzing the BEAST outputs*

12. *Lets try to discuss "preliminary" results togheter!*