**Viral Bioinformatics 2022**

**Topic: Phylogeny & Virus Typing**

**Instructors: Dr. Urmila Kulkarni-Kale & Ms. K. Sunitha Manjari**

**Hands-on exercises**

1. Perform multiple genome alignment of SARS-CoV-2 isolates (sars-cov-2.fas) using MAFFT. Use WIV04 isolate as the reference sequence for this purpose (reference\_sars-cov-.fas)s. What is the percentage of identical sites in the alignment? Save the genome alignment in fasta and aln formats.

Input dataset: phylogeny\_typing\1\_mafft\input\sars-cov-2.fas

Output files: phylogeny\_typing\1\_mafft\output\

1. Use the genome alignment of SARS-CoV-2 isolates and generate whole genome phylogenetic tree with the help of IQTREE. Select the best nucleotide substitution model that fits the data using ModelSelector. Reconstruct maximum likelihood-based phylogeny with 1000 bootstrap replicates using UltraFast method available in IQTREE. What are the number of invariant sites and parsimoniously informative sites? Which nucleotide model best fits the data provided? Using the consensus tree, find out the number of clusters in which Indian isolates are observed.

Input dataset: \phylogeny\_typing\2\_iqtree\input\sars-cov-2\_aln.fas

Output files: \phylogeny\_typing\2\_iqtree\output\

1. Using the consensus tree generated with IQTREE, check the presence or absence of temporal signal in the SARS-CoV-2 data with the help of TempEst. What can be the initial value for nucleotide substitution rate?

Input dataset: \phylogeny\_typing\3\_tempest\input\sars-cov-2.contree

Output files: \phylogeny\_typing\3\_tempest\output\sars-cov-2.pdf

1. Estimate the genome-wide nucleotide substitution rate of SARS-CoV-2 dataset using BEAST package. With the help of ‘BeauTi’ tool, choose GTR+I+G as the nucleotide substitution model and the value of slope obtained with TempEst as the initial value for ‘*meanrate*’ parameter. Use uniform distribution for ‘*treeprior*’ and normal distribution as prior for ‘*meanrate*’. Set molecular clock to ‘uncorrelated lognormal distribution’ with demographic model as ‘coalescent’. MCMC to be set to 10 million steps with log at every 10,000 steps. Generate an xml file with all the parameters set and use this as input to run ‘beast’(Takes 20-30 minutes on 8GB laptop or desktop). After the beast run, two files are obtained namely, log file and tree file. Check convergence of the log file using Tracer (Hint: ESS values to be greater than 200 for every parameter). If convergence is obtained, run the same in triplicate and combine the log files using the tool ‘logcombiner’. If convergence is not obtained, then increase the MCMC steps to 50 million and repeat the same in triplicate. Generate the maximum clade credibility tree using the tool ‘treeannotator’ with trees file as input and visualise the same using FigTree.
2. BeauTi

* Input dataset: \phylogeny\_typing\4\_beast\1\_beauti\input\sars-cov-2\_aln.fas
* Output files: \phylogeny\_typing\4\_beast\1\_beauti\output\sars-cov-2.xml

1. BEAST

* Input dataset: \phylogeny\_typing\4\_beast\2\_beast\input\sars-cov-2.xml
* Output files: \phylogeny\_typing\4\_beast\2\_beast\output\

1. Tracer

* Input dataset: \phylogeny\_typing\4\_beast\3\_tracer\\*.log

1. TreeAnnotator

* Input dataset: \phylogeny\_typing\4\_beast\4\_treeannotator\input\sars-cov-2.trees
* Output files: \phylogeny\_typing\4\_beast\4\_treeannotator\output\mcc\_sars.tree

1. FigTree

* Input dataset: \phylogeny\_typing\4\_beast\5\_figtree\input\mcc\_sars.tree
* Output files: \phylogeny\_typing\4\_beast\5\_figtree\output\annotated\_sars-cov-2.tree

1. Perform genotype assignment for the provided data set of *Dengue virus* sequences (denv.fas) using –

Input dataset: \phylogeny\_typing\5\_genotyping\rtd\denv.fasta

* 1. RTD server: http://bioinfo.unipune.ac.in/Dengue/Home.html
  2. Genome detective: https://www.genomedetective.com/app/typingtool/dengue/

1. Perform lineage assignment of two SARS-CoV-2 isolates (sars-cov-2\_genotype.fasta) using Genome Detective (https://www.genomedetective.com/app/typingtool/cov/).

Input dataset: \phylogeny\_typing\5\_genotyping\genomedetective\sars-cov-2\_genotype.fasta