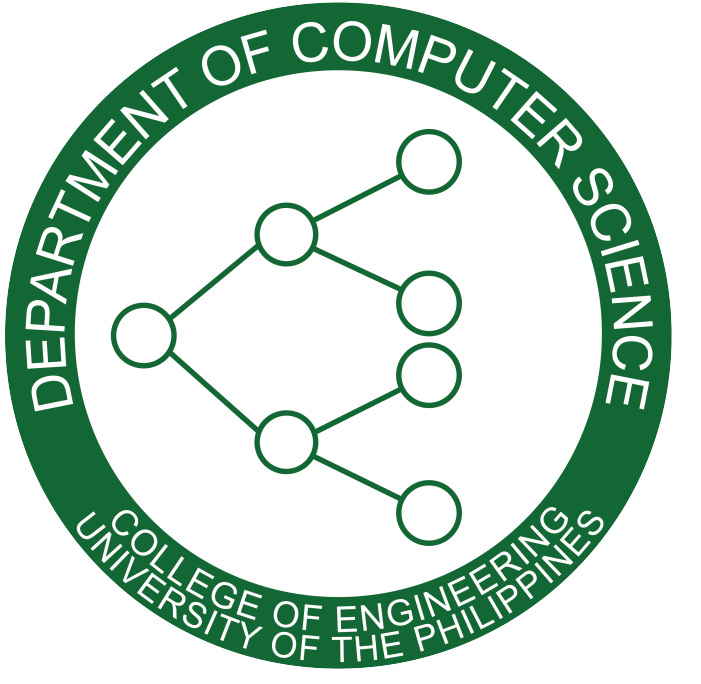




A scalable and parametrizable pipeline for the Bayesian phylodynamic inference of SARS-CoV-2



John Michael C. Egana^{1,2}, Francis A. Tablizo¹, John Justine S. Villar²

¹Philippine Genome Center, University of the Philippines System

²Department of Computer Science, University of the Philippines Diliman

Introduction

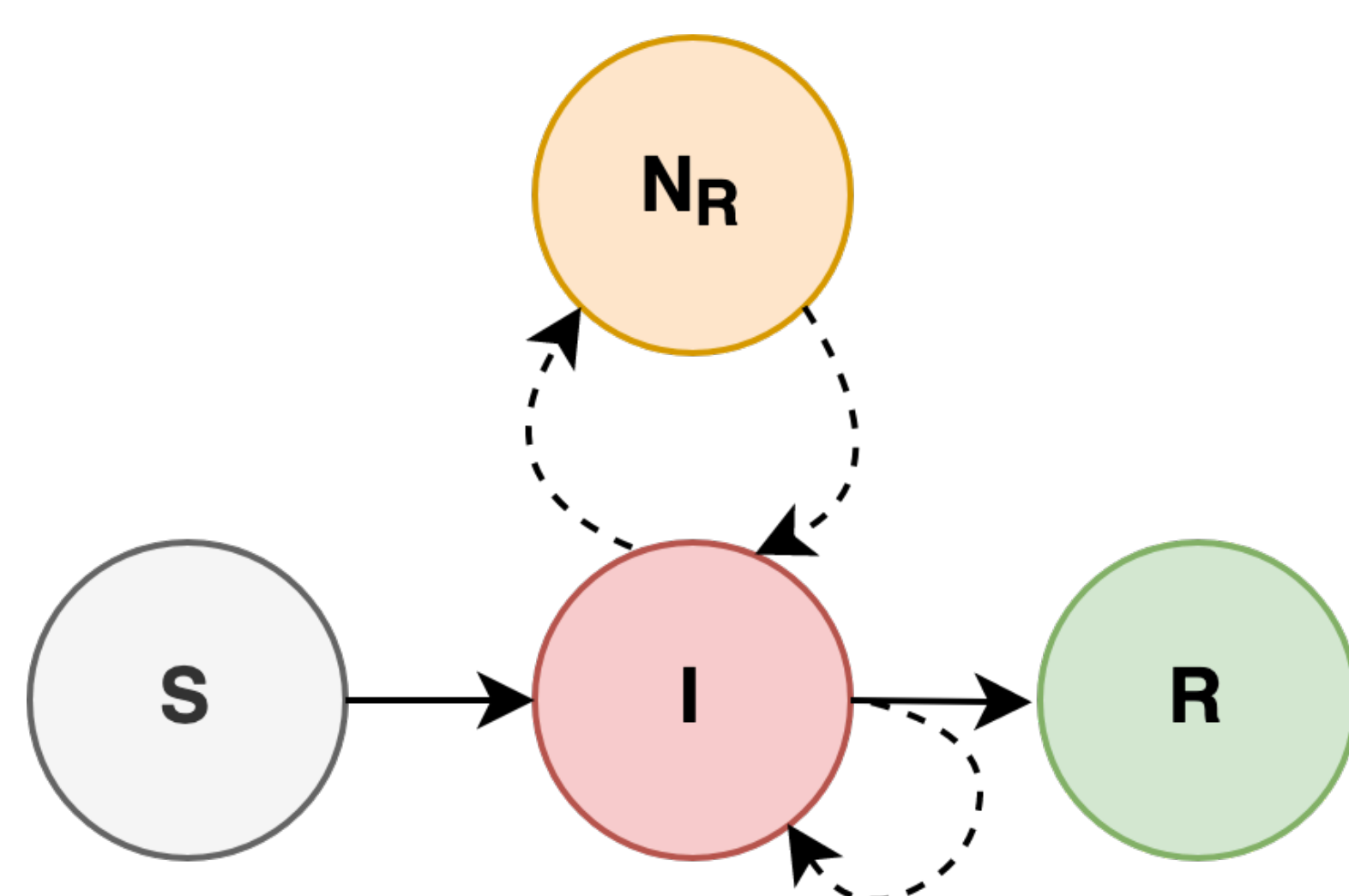
- The recent COVID-19 pandemic led to the **adoption of genomic biosurveillance** to identify variants that arise due to genome variations in the process of evolution.
- The **integration of the model of evolution and epidemiology** gives rise to the field of phylodynamics which is normally analyzed using Bayesian approaches.
- These approaches albeit come with advantages such as **summarizing phylogenetic uncertainty** among others, also come with various challenges:
 - scale of a global pandemic
 - complexity of the bioinformatics pipeline,
 - inherent computational intractability of Bayesian inference.

Solution

- To develop a scalable and parametrizable pipeline based on the principles of nf-core to streamline the process of:
 - compiling **sequence data** from public databases
 - setting up **prior distributions**
 - performing various **preliminary genomic analyses**
 - calculating the **posterior distributions** of the epidemiologic parameters being inferred using MCMC

Model

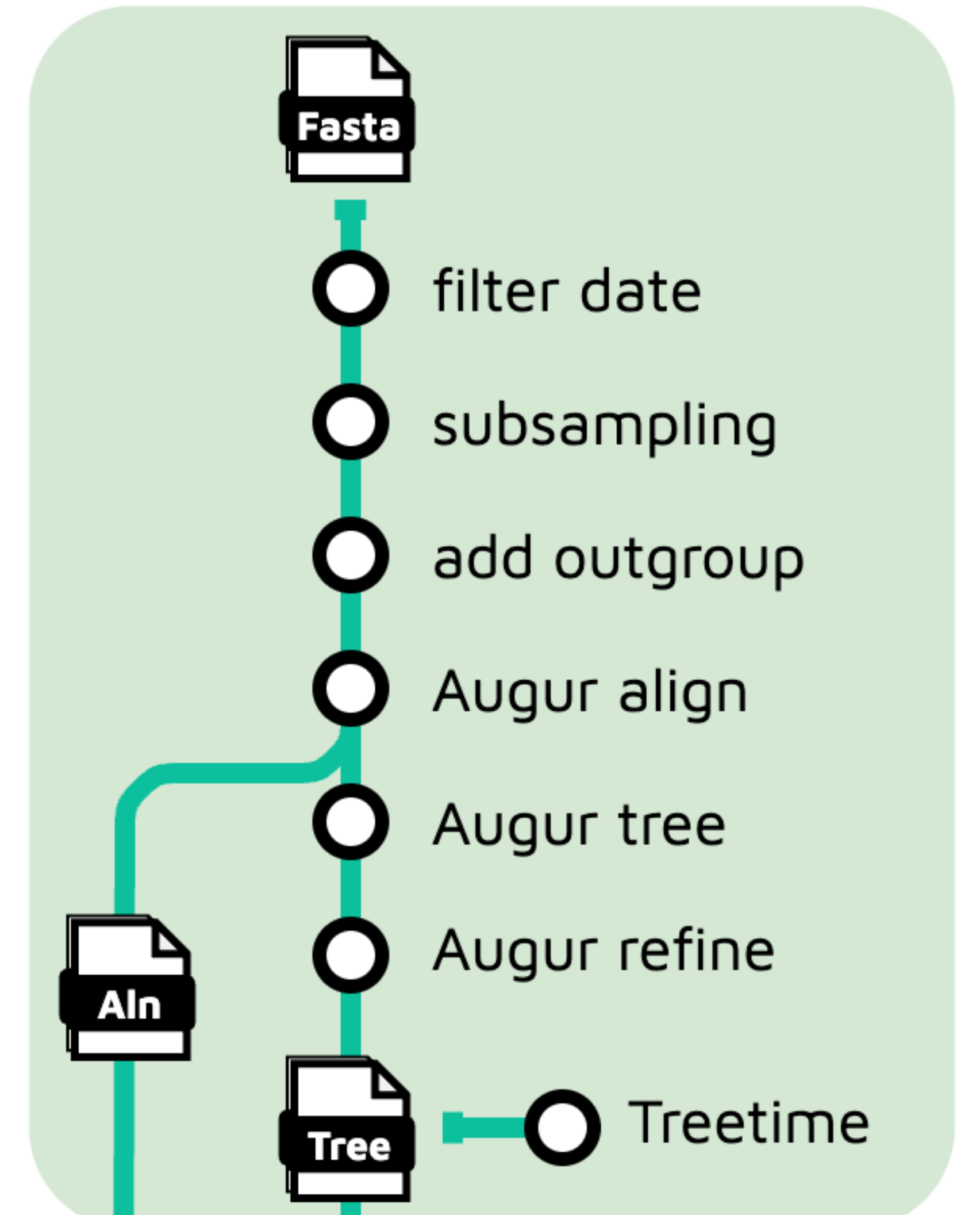
- Model is based on the structured coalescent model implemented in **PhyDyn**
- compartmental models are fitted into viral sequence data



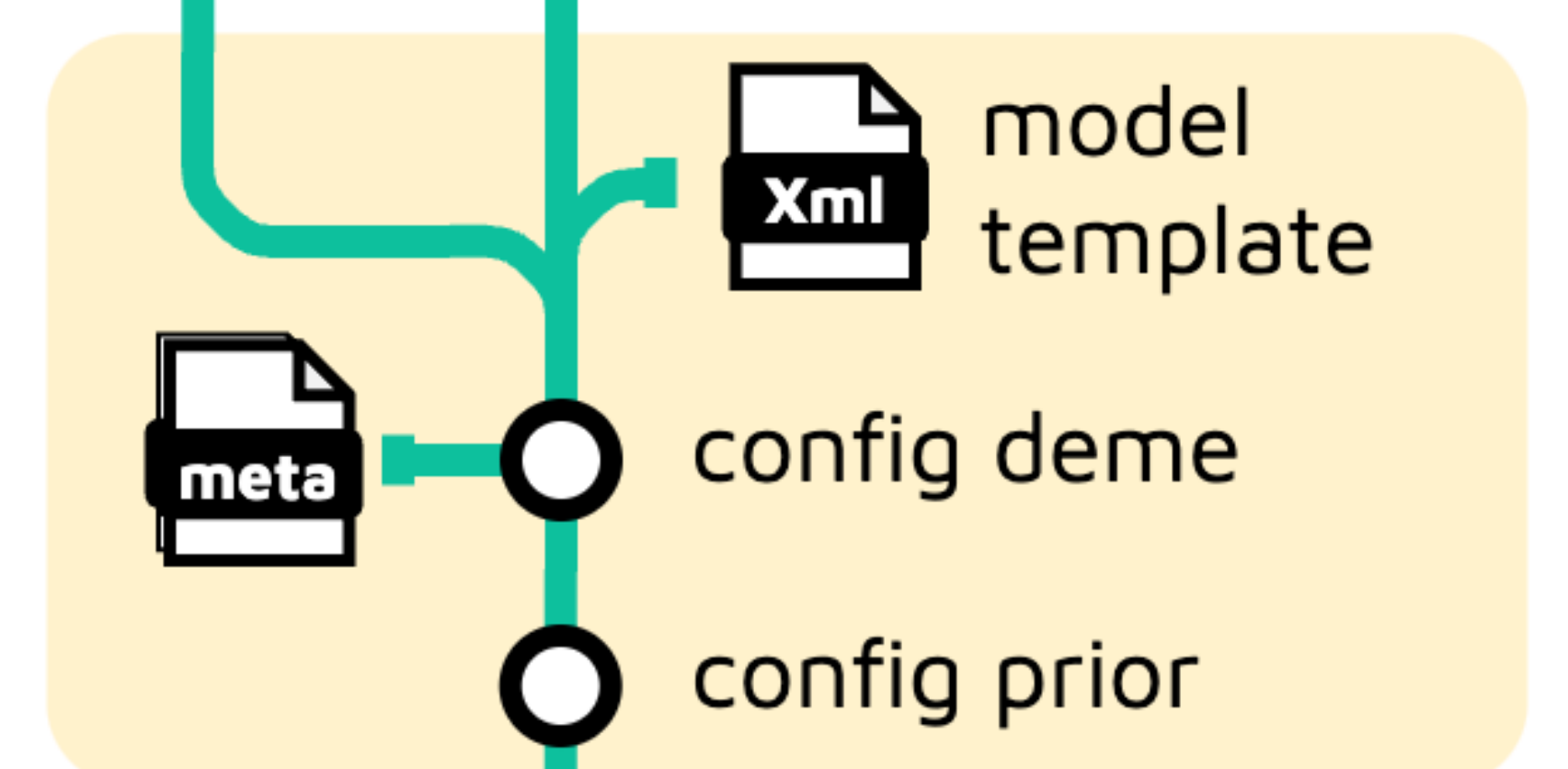
A sample extension SIR model* used in the pipeline

Workflow

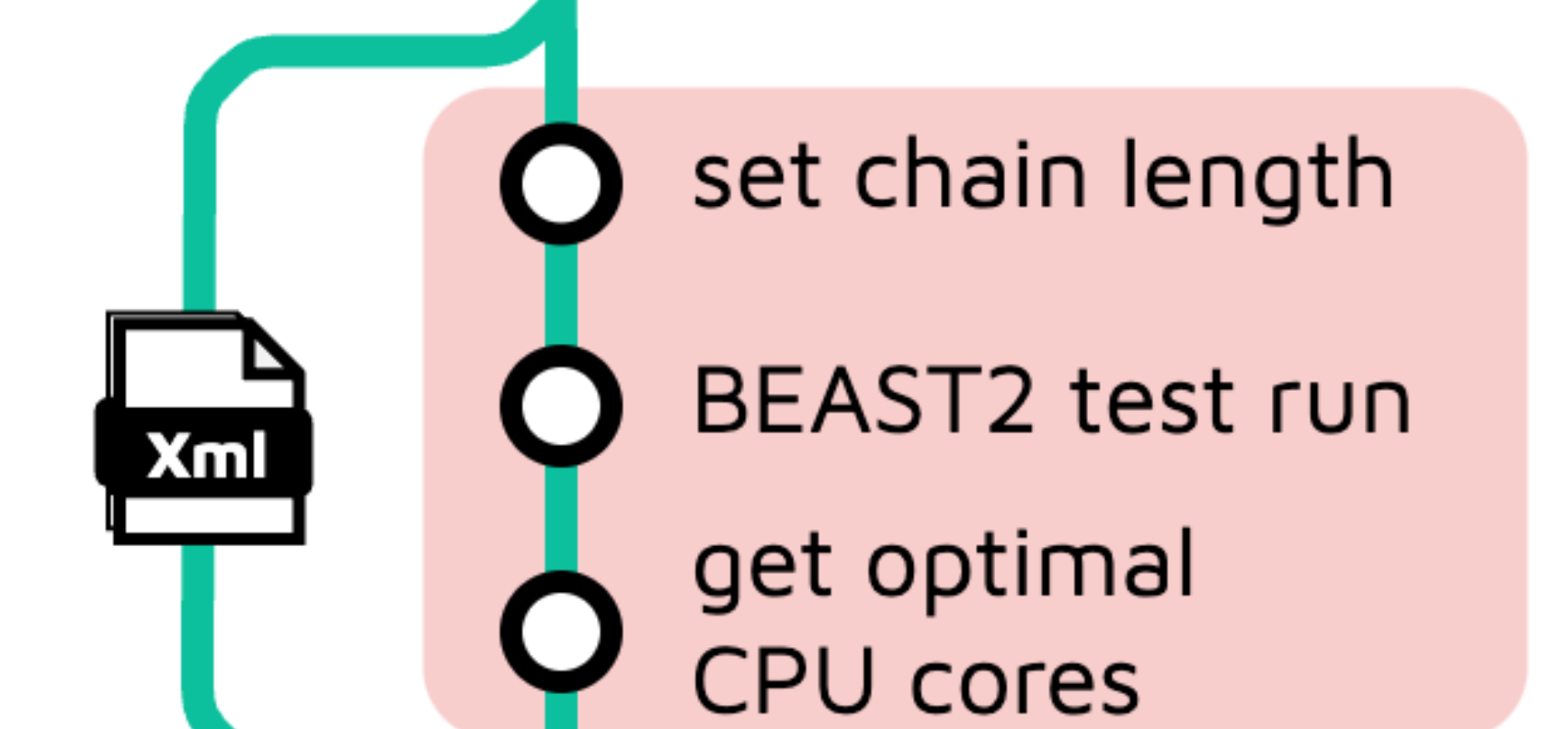
MAXIMUM LIKELIHOOD PHYLO TREE CONSTRUCTION



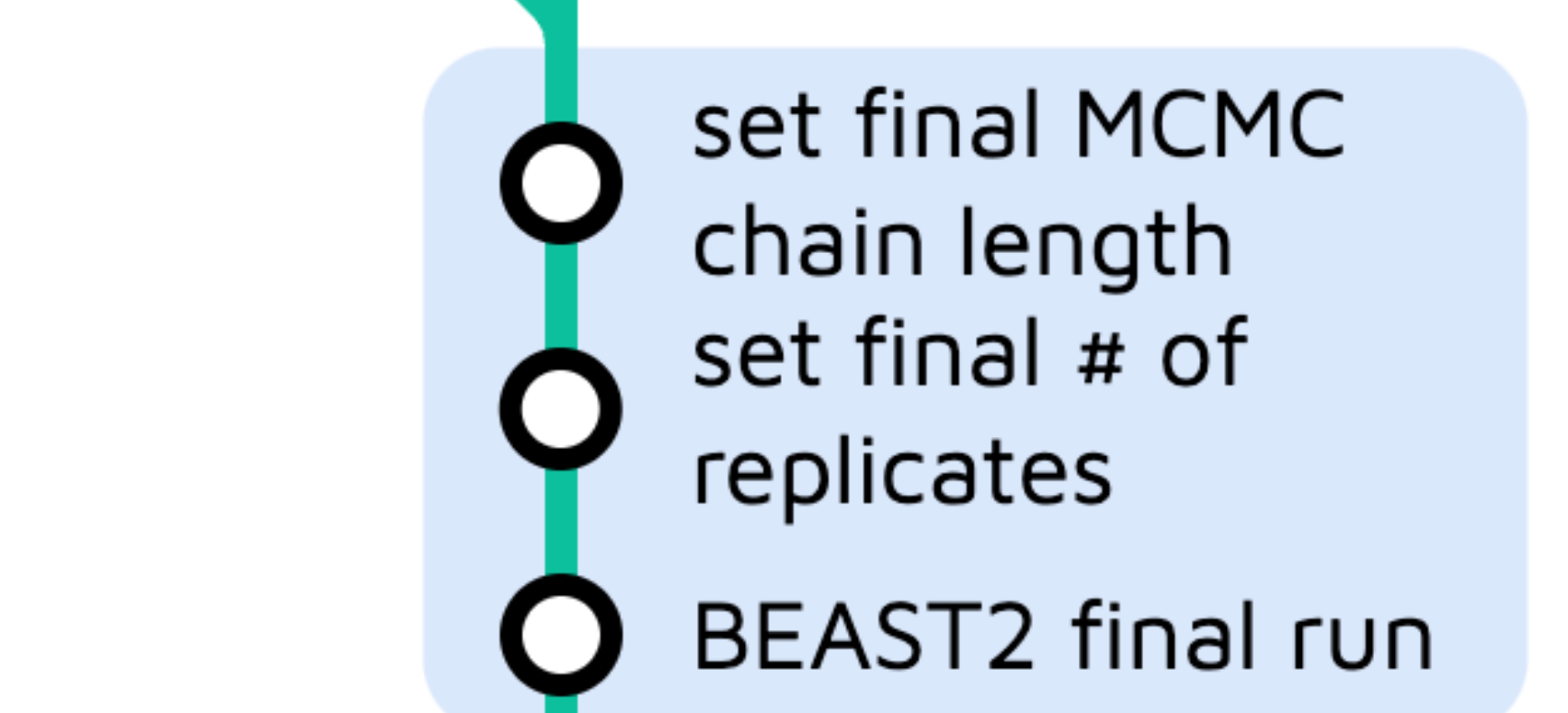
MODEL SPECIFICATION



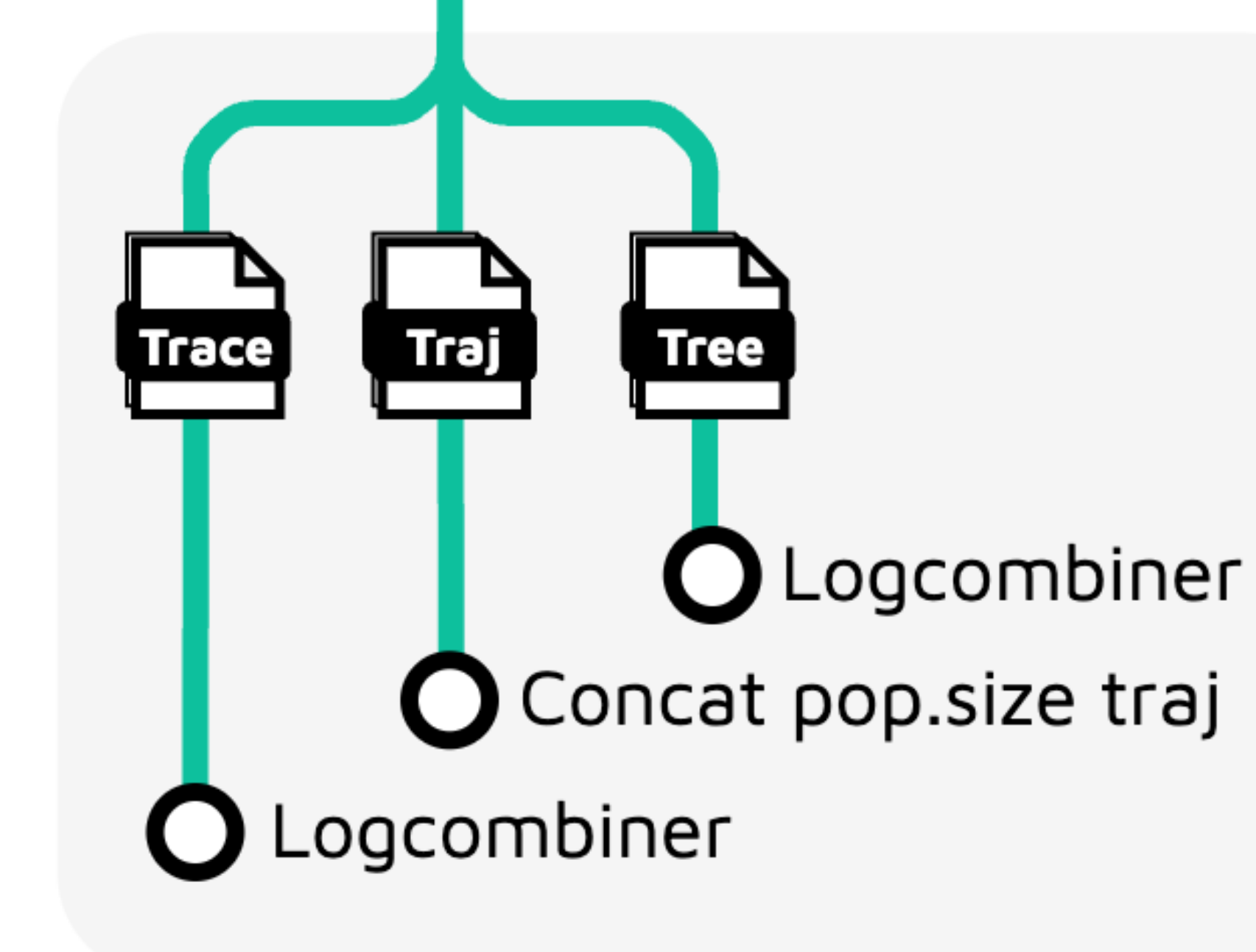
OPTIMIZATION



BAYESIAN INFERENCE



RESULTS CURATION



References

- Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., du Plessis, L., Poppinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., ... Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology*, 15(4), e1006650. <https://doi.org/10.1371/journal.pcbi.1006650>
- Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology*, 35(4), 316–319. <https://doi.org/10.1038/nbt.3820>
- Huddleston, J., Hadfield, J., Sibley, T., Lee, J., Fay, K., Ilcisin, M., Harkins, E., Bedford, T., Neher, R., & Hodcroft, E. (2021). Augur: A bioinformatics toolkit for phylogenetic analyses of human pathogens. *Journal of Open Source Software*, 6(57), 2906. <https://doi.org/10.21105/joss.02906>
- Volz, E. M., & Siveroni, I. (2018). Bayesian phylodynamic inference with complex models. *PLOS Computational Biology*, 14(11), e1006546. <https://doi.org/10.1371/journal.pcbi.1006546>

Acknowledgements

The authors express their gratitude to the Philippine Department of Health - Epidemiology Bureau for leading the genomic surveillance efforts in the Philippines, which provided the sequence data in this study.