1 Team

Alexandra Gavryushkina, sasha.gavryushkina@auckland.ac.nz, Department of Computer Science, University of Auckland, Auckland, New Zealand.

David Welch, david.welch@auckland.ac.nz, Department of Computer Science, University of Auckland, Auckland, New Zealand

Alexei Drummond, alexei@cs.auckland.ac.nz, Department of Computer Science, University of Auckland, Auckland, New Zealand

2 Summary

We have looked at the Village simulation scenarios (October data) and used all three gene regions to estimate phylogenies and epidemiological parameters. We plan to run the same analysis only for the pol sequences and compare the estimates.

3 Methods

We used Bayesian MCMC methods to estimate phylogeny and epidemiological parameters from sequence data. We applied two different prior models for the tree branching process (transmission process): Bayesian skyline model and birth-death skyline model with sampled ancestors. There is a parameter unidentifiably for this model meaning that one of the parameters should be known to enable effective inference. Because of this, we placed a strong prior on the death rate parameter, μ, with mean rate of 0.1 per year, that is, we expect that an infected individuals dies in 10 years after being infected. The model also considers a probability, r, that an individual stops causing further infectious after sampling due to treatment or behaviour change.

4 Primary results

We have analysed three samples from epidemic 1 with sample scheme 1 and epidemic 3 with sample scheme 1 (6 analyses in total). At the time we have not finished the analyses. We present here preliminary results from the analysis with Bayesian skyline tree prior. We estimated that the epidemic

was growing at the time of sampling for all six analyses. Although the rate of growth differs from sample to sample and the growth was slowing down for all samples. Figure 1 shows the skyline estimates of the infected population size through time. From the estimates of the tree heights:

sample	mean	95%HPD
scA sample1 epidemic 1	43.7	[40, 48.2]
scB sample1 epidemic 1	41.3	[36.3, 47]
scC sample1 epidemic 1	38.1	[34.5, 41.9]
scG sample1 epidemic 3	50.1	[43.7, 57]
scH sample1 epidemic 3	45.2	[41.5, 49.1]
scI sample1 epidemic 3	36	[33.8, 38.5]

we can conclude that samples were taken in the order: scC sample1, scB sample1, and then scA sample1 for epidemic 1. ScI sample 1, scH sample 1, and scG for epidemic 3.

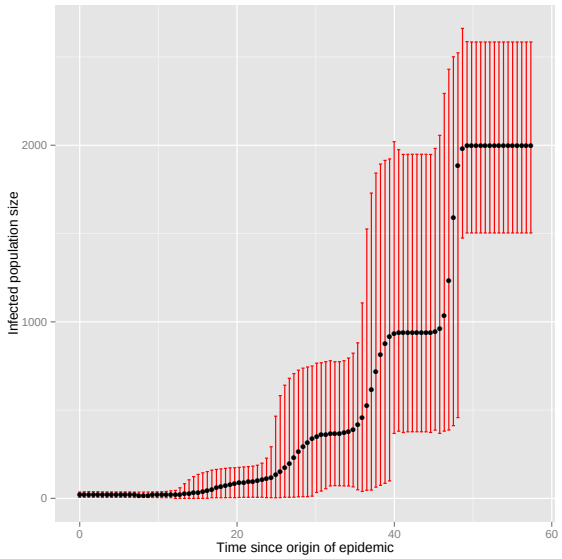

Figure 1: **The means and 95 % HPD intervals for posterior estimates of the infected population size through time.** The graph shows how the infected population size grows through time for scG sample 1 epidemic 3 (the most recent sample from epidemic 3).

Due to the time frame we cannot yet present the results from the analysis with the birth-death skyline model with sampled ancestors. We anticipate to estimate parameters:

- δ the total rate of becoming uninfectious due to death or sampling. It can be derived as $\mu + \psi r$;
- R_0 the effective reproductive number defined as $\frac{\lambda}{\delta}$, where λ is a transmission rate;
- s the sampling proportion and can be derived as $\frac{\psi}{\mu+\psi}$; and
- r the probability that a sampled individual will not cause further infectious;

given that we know one of the parameters, μ for example.

5 Interpretation

We can estimate the past epidemiological dynamics. The computation time is quite large although is feasible for datasets with approximately 300 samples.

6 Outlook

To use the birth-death model, we need information about one of the parameters, such as the total rate of becoming uninfectious, i.e., the expected time from the time of infection until the time when the person stops causing other infectious due to treatment, behaviour change or death.

7 Supplement

Our model allows to estimate samples that are direct ancestors of other samples in a transmission chain and removal (from epidemic) at sampling probability. Figure 2 shows preliminary posterior estimates of sampled ancestors for one of the samples. The mean estimate and 95% HPD for r was 0.9 [0.77, 0.98] for scA sample 1 epidemic 1.

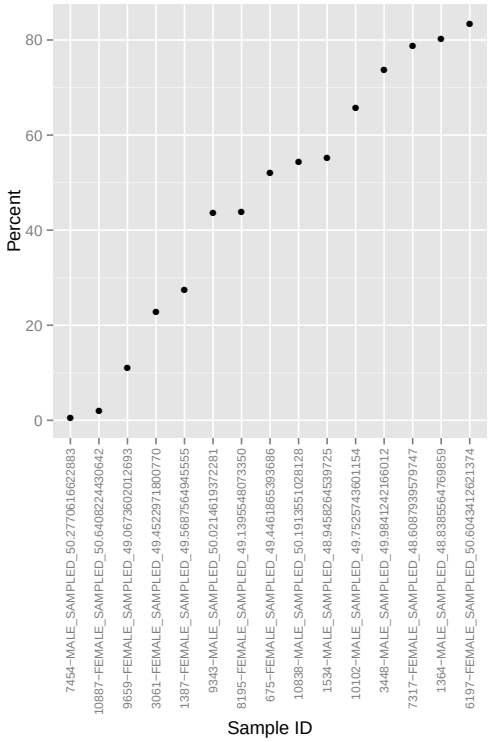

Figure 2: **Posterior probabilities of samples to be sampled ancestors.** The graph shows the posterior probabilities of samples to be sampled ancestors for scA sample 1 epidemic 1 (the most recent sample from epidemic 1). Only samples with non-zero probabilities are shown