Phylodynamics in BEAST2 using the birth-death SIR model

In this practical session we will use a sequence data set from the 2009 H1N1 influenza pandemic collected from March to November 2009. As such, these samples cover a large portion of the timescale of the pandemic where compartmental models might be useful. In particular, we will use the birth-death susceptible-infected-recovered model (Kühnert et al. 2004), which is an approximation of the epidemiological SIR model.

The birth-death SIR model

The birth-death SIR model is a reparameterisation of the birth-death skyline (Stadler et al. 2013). It assumes that there exist that the population of hosts can be divided into three compartments; susceptible, infected, and recovered. In this model, the recovered class includes two types of individuals; those that are recovered and observed (i.e. sampled) and those that are unobserved. These dynamics can be expressed in terms of reaction kinetics:

Infections:	$I + S \rightarrow 2I$	mass-infection rate: β	(a)
Unobserved recovery:	$I \rightarrow R_h$	unobserved recovery rate: $\gamma^*(1-s)$	(b)
Sampling:	$I \rightarrow R_s$	sampled recovery rate: $s^*\gamma$	(c)

Where S stands for susceptible individuals, I for infected individuals, Rh for hidden (or unobserved) recoveries, and Rs for sampled individuals. These reactions occur different rates. Infections, equation (a), occur at an mass-action transmission rate of β. The rate of unobserved recoveries is the product of the recovery rate, γ, and the probability that an individual is not sampled, $1 - s$, where s is the sampling rate. The rate of observed recoveries is simply the product of the sampling rate and the recovery rate, $s\gamma$. We also assume that the population size, N = S + I + R is constant over time.

In the birth-death model we assume that the susceptible population size is infinite, but in the birth-death SIR the susceptible population, n_s, can be depleted over time, leading to a decrease in the number of infections, i.e. n_s is a function of time, $n_s(t)$, where t stands for time. Thus, the reproductive number can change over time, and we will refer to it as R_e to distinguish it from the basic reproductive number, R_0. In terms of the reaction rates above $R_e = \beta^* n_s(t) / \gamma$. Importantly, R_e is equivalent to R_0 at the start of the infectious spread, when the population is fully susceptible; $R_0 = \beta^* n_s(0) / \gamma$. For further details of the implementation of this model see Kühnert et al. (2014).

Setting up the birth-death SIR model in BEAUTI

Data set

- Sequence alignment in fasta format of samples collected until November 2009 of the H1N1 pandemic in North America: `NorthAm.Nov.fasta`

Software

- BEAST 2.4 (beast2.org)
- Phylodynamics addon
- R (www.r-project.org)

1. Open BEAUTI and drag and drop the alignment file. You might get a pop-up window asking you what kind of data this is (Fig 1). Select 'import alignment' and click 'OK'. You might get an other prompt asking you to select aminoacid or nucleotide sequences. Select nucleotide and click 'OK'.

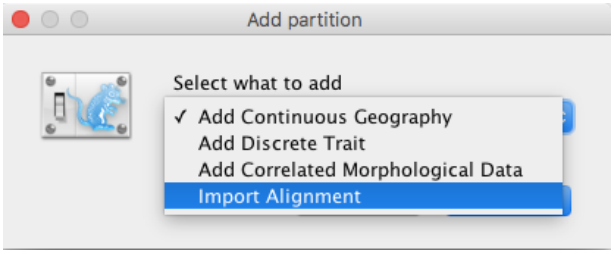
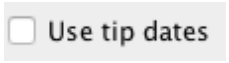


Fig 1.

2. The data will appear in the 'Partitions' tab. It should correspond to 100 taxa and 13,154 sites. Click on the 'Tip Dates' tab. You should see a blank window with a 

box. Check the box, and you should see the table in Fig 2 appear. Click on 'Auto-configure'. You should a window to set the dates, as shown in Fig 3. Select 'use everything', 'after last' and type in a | symbol. Click OK. The table from Fig 2 should be updated to list the sampling times for all the samples.

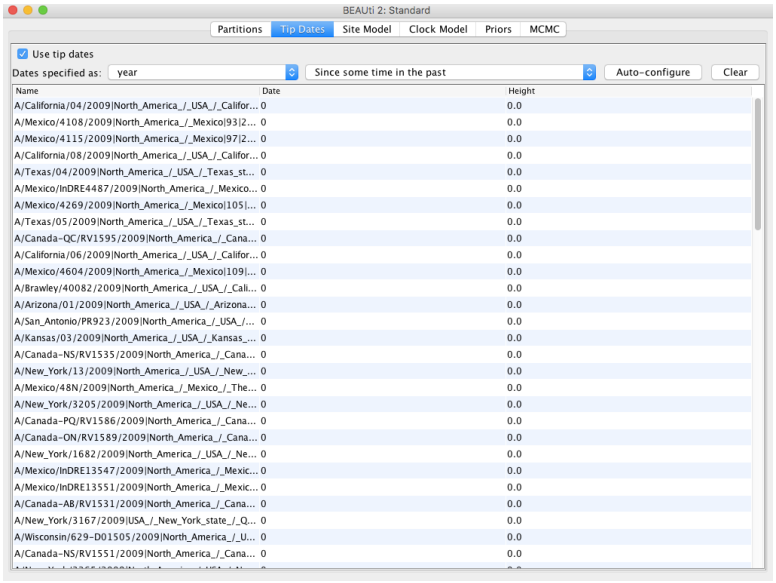


Fig 2. Tip dates tab in BEAUTI.

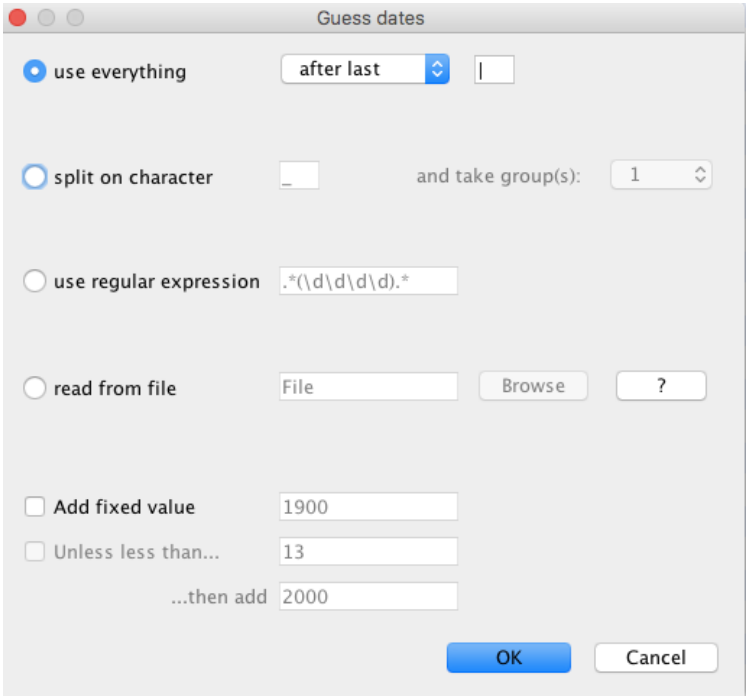


Fig 3. Specifying tip dates in BEAUTI.

3. Click on the ‘site model’ tab. Select the HKY+G model as shown in Fig 4. This model will account for rate heterogeneity among sites and for the transition to tranversion bias.

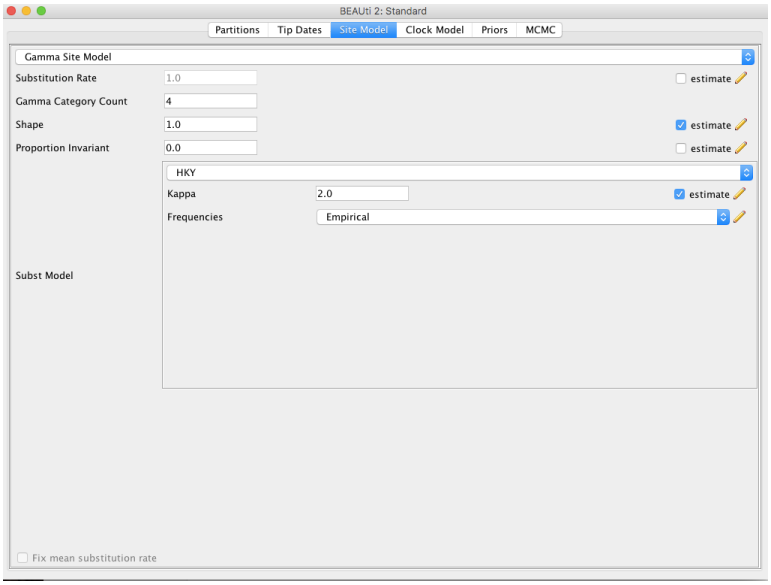


Fig 4. Setting up the HKY+G model in BEAUTI.

4. Next, select the 'Clock model' tab and ensure that the strict clock model is selected. This is the default, so there is probably no need to change anything.

5. In the 'Priors' tab go to the tree prior and select 'Phylodynamics: Birth Death SIR (serial) (Fig 1).

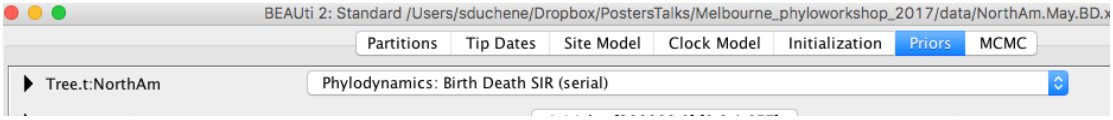


Fig 1. Selecting the Birth Death SIR (serial) model.

6. Click on 'View' and select 'Show Initialization panel' (Fig 2).

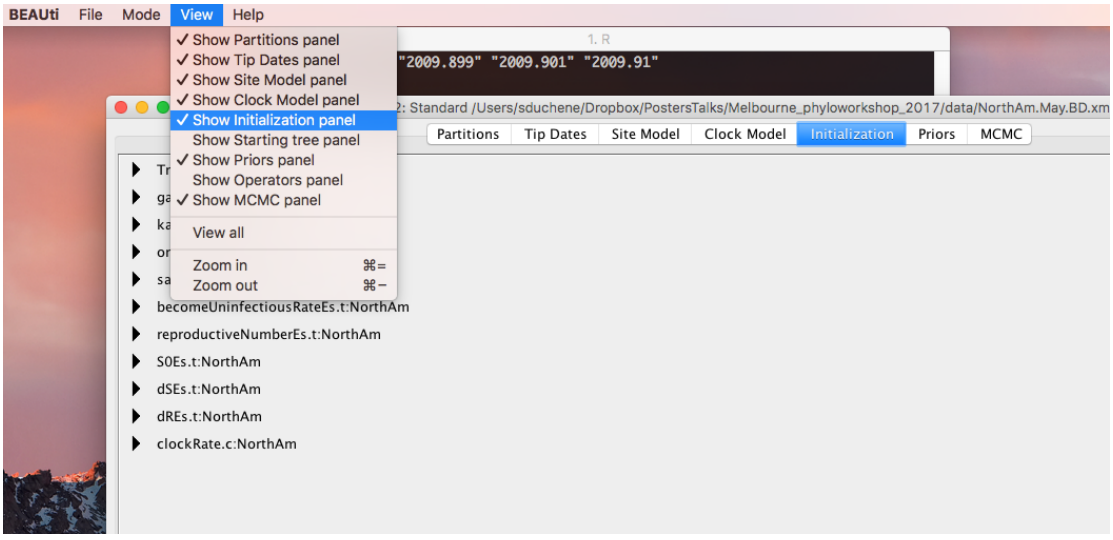


Fig 2. Initialisation panel selection in BEAUTI.

7. At the ‘Initialization’ tab, click on the arrow next to ‘becomeUninfectiousRateEs.t:NorthAm’ and change the value for ‘Upper’ to 1000 (Fig 3). This step ensures that the become uninfectious rate can take large values (about 140 in our analyses of the constant birth-death model).

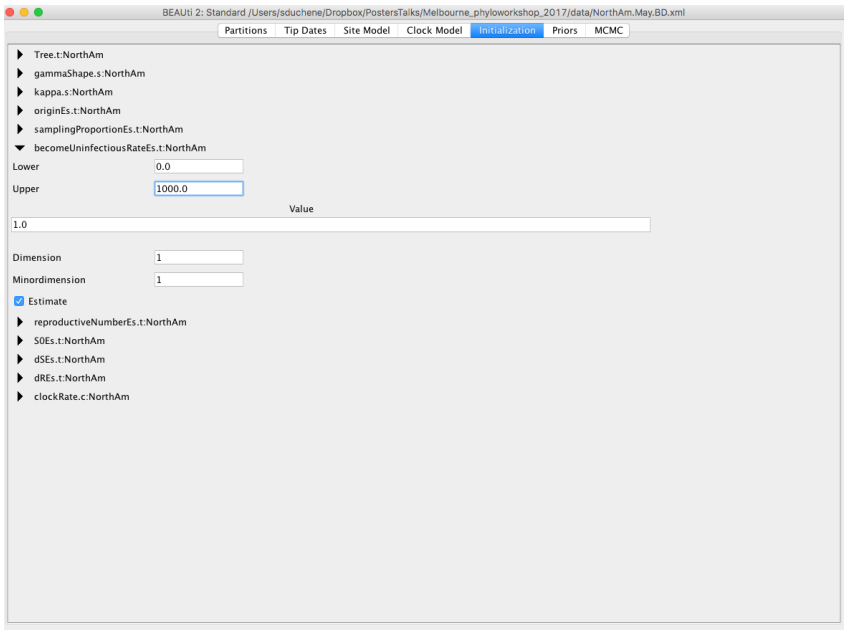


Fig 3. Changing the upper threshold for the become uninfectious rate.

8. Go back to the ‘Priors section. Below is some information about the priors for the parameters of the model and how to set them up (Fig 4):

S0Es.t:NorthAm
This is the population size of susceptible individuals, $n_s(0)$, at t=0. Set this to a gamma distribution with an alpha parameter of 100 and beta of 40000. The ‘mode’ should be set to ‘ShapeScale’

becomeUninfectiousRateEs.t:NorthAm
This corresponds to γ, which is analogous to δ in the birth-death model (in this model we also assume that the duration of infection is constant over time). We will use the same prior as in the constant birth-death model. Set this to a normal distribution with mean of 140 and standard deviation of 1.3.

The remaining default prior distributions are fine for this analyses, but it is always useful to inspect them to check that they represent our knowledge of these parameters.

9. Go to the ‘MCMC’ tab. This analysis requires more computation than the birth-death and the exponential growth coalescent. Set the chain length to $5x10^7$. Under ‘tracelog’ and ‘treelog.t:NorthAm’ change the ‘Log Every’ value to 5000. For the File Name use NorthAm.Nov.BDSIR.log and NorthAm.Nov.BDSIR.trees (Fig 5). Go to ‘File’, ‘Save’ and save the file as NorthAm.Nov.BDSIR.xml in a convenient directory.

10. Run this analysis in BEAST.

Fig 4. Setting up prior distributions for the birth-death SIR model.

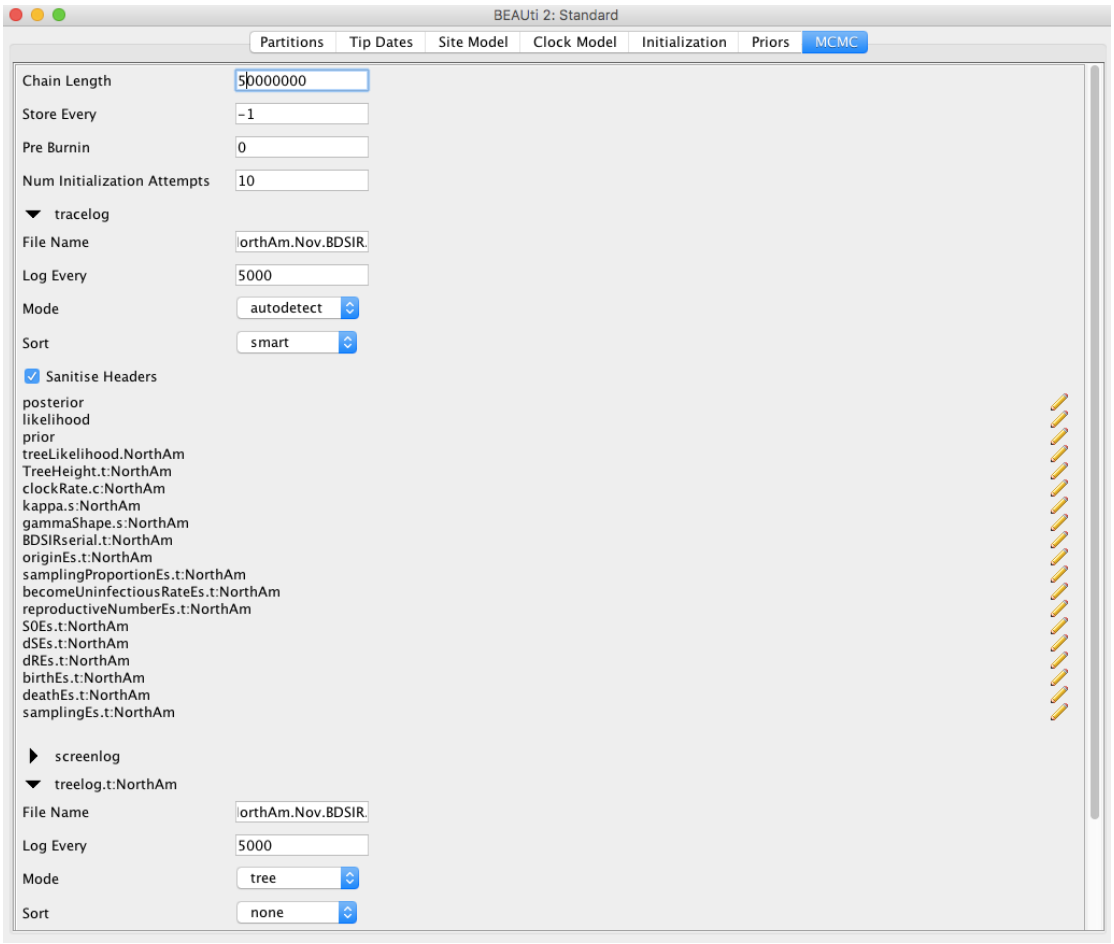


Fig 5. MCMC settings for the birth-death SIR model in BEAUTI.

11. This analysis will be very slow, and it might not complete by the end of this session. We have uploaded a complete run NorthAm.Nov.BDSIR.log where you can inspect the results (remember you can also open log files in Tracer before the analysis is complete). Load this file (or your own if the analysis is complete) in Tracer.

Question 1: What is the basic reproductive number for this data set? Is it similar to what we estimated for the data collected until May 2009?

12. The log file from the birth-death SIR model has the estimates for some important epidemiological parameters, but it doe not inform us about the population dynamics over time. To do this, we will use R to do some calculations using the log file.

Open Rstudio and load the script ‘plot_BDSIR.R’. The first line reads the log file, make sure that it points to the correct path. For example :

```
data <- read.table('~/Desktop/NorthAm.Nov.BDSIR.log',
head = T)
```

The script will make four plots; the trajectories of susceptible, infected, and recovered individuals over time; the prevalence (current number of infections) and incidence (number of new infections); the reproductive number over time, R_e; and the posterior distribution of R_0 (Fig 6).

Question 2: Does there appear to be evidence of a depletion of susceptible individuals?

Question 3: At around what time did the number of infections over time (incidence) start decreasing?

Question 4: At what time did the pandemic appear to stop spreading (R_e<=1)? Does this coincide with the peak prevalence and incidence?

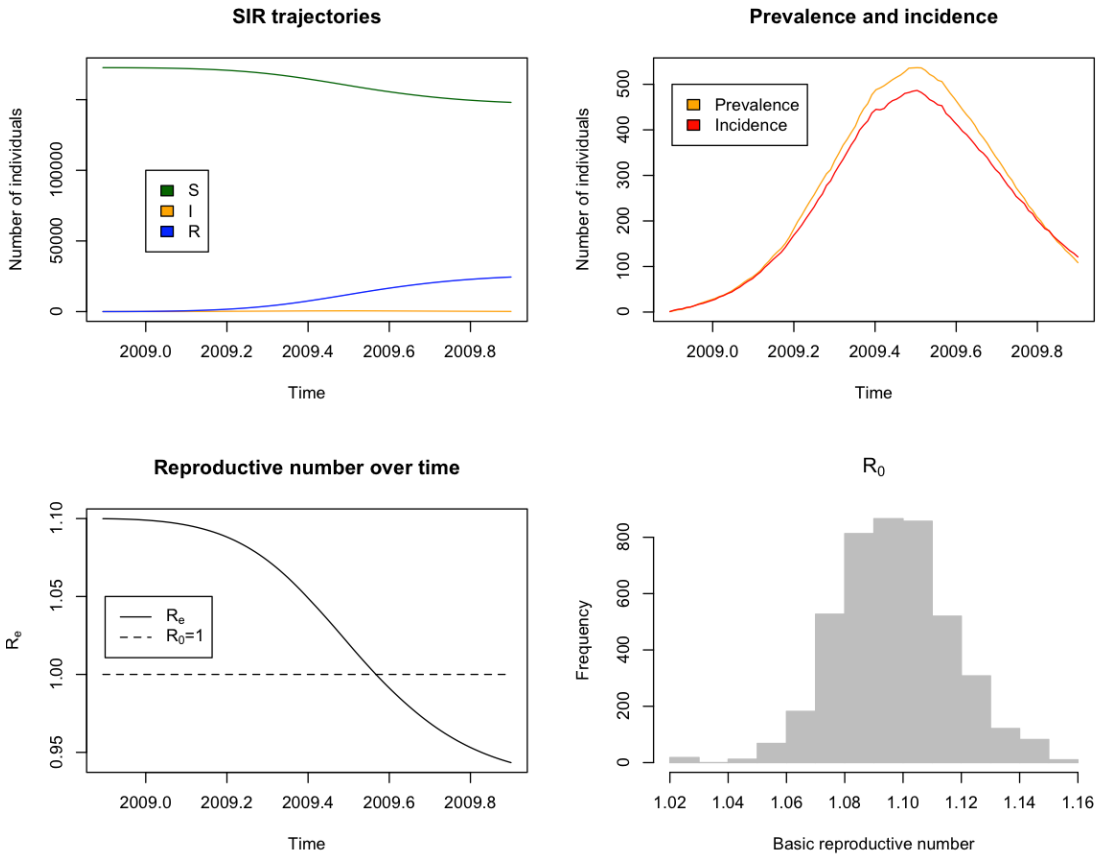


Fig 6. Epidemiological information from the birth-death SIR model.

References

Kühnert, D., Stadler, T., Vaughan, T. G., & Drummond, A. J. (2014). Simultaneous reconstruction of evolutionary history and epidemiological dynamics from viral sequences with the birth–death SIR model. *Journal of the Royal Society Interface*, *11*(94), 20131106.

Stadler, T., Kühnert, D., Bonhoeffer, S., & Drummond, A. J. (2013). Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proceedings of the National Academy of Sciences*, *110*(1), 228-233.