Slide 1:

🡪Present the study:

-We are analysing the rates of mutation of Ebola over the 2013-2016 epidemic and the most recent 2018 epidemic, which is still ongoing, using BEAST.

-These are the 2 largest Ebola outbreaks that have been reported and they are significant as they are the first outbreaks were real time genomic sequencing was applied.

-This means that many full sequences are available for us to study for which we have information about the temporal signal.

Slide 2:

🡪Temporal signal:

-In phylogenetic methods we can use sequencing dates to calibrate the molecular clock, which helps estimate substitution rates of fast evolving viruses

-To do this we assign dates to the tips of the tree, which is possible when there is sufficient spread in the age of the samples analysed

-testing for temporal signal is useful for verifying whether the clock can be calibrated using sampling times

-A dataset is considered to have strong temporal signal if a sufficient amount of molecular evolution has occurred over the sampling time window. In this case the data can be treated as being from a measurably evolving population

Slide 3:

🡪 Models used

-To get the most accurate rate estimates we set up 4 different models in BEAST for phylogenetic reconstruction, from which we chose the best fitting one, that was then used to make further inferences and comparisons.

-These models included

* strict clock isochronous model
* strict clock heterochronous model
* relaxed clock isochronous model
* relaxed clock heterochronous model

Slide 4:

🡪Isochronous:

-In these models the molecular clock was not calibrated with the use of tip dates, so the data is constrained to be contemporaneous

-This can be a good model for datasets with samples collected over short periods of time, where there is no significant temporal signal, as the dates are very similar.

Slide 5:

🡪Heterochronous:

-In this case we take into account information about the age of the samples to calibrate the molecular clock

-This is only possible when we have sufficient information about the temporal signal, and it is also suited for datasets with fast evolving samples

Advantages:

-Apart from reflecting divergence events the age estimates can provide information about when the most recent common ancestor appeared, the timing of emergence of a novel specific variant and when host species jumps occur

-When combined with a relaxed molecular clock it can provide information about the rate at which mutations accumulate on their connecting lineages.

Disadvantages:

-The substitution process occurs stochastically, so a sequence sampled earlier might exhibit more divergence from the outgroup than one sampled later, and this could inflate the variance of the rate estimates and lead to uncertainty

Slide 6:

🡪 Strict clock

-The strict clock is the simplest model that can be used

-It assumes homogenous rates among branches (evolutionary change occurs at a predictable rate over time) and its only parameter is the rate of evolution (substitutions/site)

-Many rapidly evolving viruses follow clock like behaviour

Advantage

-Permits for greater flexibility in accounting for uncertainty in tree topology and in calibrations

Slide 7:

🡪Relaxed clock

-It accounts for variation among branch specific rates

-We used an uncorrelated relaxed clock model, that assumes no correlation between rates of neighbouring branches

-The branch specific rates were sampled from a lognormal distribution, that suggests that the they tend to cluster around a mean value

- The parameters of the distribution can provide a measure of rate heterogeneity

Slide 8:

🡪Why is it important to select the best fitting model?

-Choosing the best fitting model gives us the most accurate rate estimates

-In BEAST, parameter-rich models are not automatically preferred, because introducing a new parameter adds a new dimension that needs to be averaged over. This can complicate the computational process and create higher uncertainty in the results. It is important to find a balance between good fit and complexity

Slide 9:

🡪How to select the best fitting model?

-The best fitting model was selected based on the marginal likelihood estimator (MLE) score, which is the average fit of a model to a dataset.

-To determine the MLE we chose a general stepping-stone (GSS) path in BEAST, with a beta path step distribution (default setting, can’t be changed).

Slide 10:

🡪Graph of results (shows the rates from the sc+het, icld+het models and the best fitting model for each dataset)

Slide 11:

🡪Congo data