*Virus Evolution -Reflections*

**Molecular evolutionary rates of the 2013-2016 and 2018 African Ebola virus epidemics: implications for genome surveillance**

Melina Haritopoulou-Sinanidou1, Sebastian Duchene1

1Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, VIC, Australia.

**\*Corresponding author:** Pending.

**Abstract**

Pending.

**Keywords**

Ebola virus, Phylodynamics, Molecular clock, Genome surveillance

*(Total article should have 5,000 words max)*

*INTRODUCTION*

*-Describe these two epidemics and molecular work done on them*

*- Real-time genomics and surveillance. Mention Quick et al. and other papers on near-real time analyses. These two epidemics are the first time that we can use sequencing in real time.*

*- Info that has been obtained with real-time sequencing in Ebola (refer to Eddie’s paper). E.g. time of origin, geographic movement, potentially reservoir.*

*- These inferences depend on how accurate phylogenetic reconstructions, which in turn depend on information content in the data. Some phylodynamic inferences don’t just use sequence info, but sampling. E.g. Tanja’s and Veronika’s papers on the birth-death tree prior.*

*- Understanding the range of rates in Ebola (and any virus) is critical to assessing the range of inferences that can be obtained in real time. For example, some ebola virus rate estimates are too low for a sufficient amount of evolution to occur over a short amount of time and with limited spread. Here discuss the expected number of variable sites. Moreover, understanding the range of ebola rates is also important to understand when estimates are misleading due to methodological or statistical artefacts.*

*METHODS SUMMARY*

* *We collected publicly available genomes of ebola and subset them to represent different resolutions of temporal and geographic sampling*
* *We analysed the data using Bayesian phylogenetic methods and rapid phylogenetic approaches, which although use less realistic models, offer a practical solution to analyse large genomic data sets in a short amount of time.*
* *We also assessed temporal structure using cutting edge model selection approaches.*

*RESULTS AND DISCUSSION*

* *Pending.*
* *What is the range of estimates and does it depend on the methods (Fig with LSD and rates estimates using best model in beast and temporal structure)? What clock models are most commonly selected in BEAST? (models selected in beast in a fig and refer to table for p values from lsd and the tempest stuff etc..).*
* *If they vary substantially, are there any potential causes (selection, time dependency)? Probably not, but worth touching on. Our emphasis is what the rates are and potential biases (think sampling in the Congo data), rather than biological causes, which may not differ much between epidemics/geography.*
* *How does the rate compare to similar viruses (e.g. those with similar genome sizes)?*
* *What does this tell us about sampling for surveillance studies. For example, if we collect samples over a few months, can we trust our estimates of rates and times? Maybe if the data are uninformative we use rate calibrations from previous data (if they evolve within a small range) or include older samples (with care not to include very distant ones, as this can cause a bias).*