Methods:

Nucleotide sequences of the Ebola virus, between the years 2014 and 2019, were obtained from different databases including GenBank and Nextstrain and publications (reference). Each dataset included 72 to 2013 samples, with specified sampling times, in calendar dates, and sampling periods ranging from 0.05 to 3 years (Table 1: dataset information).

BEAST v1.10.4 was used for analysis of the datasets, were we selected a coalescent exponential growth model, which suggests that the gene variants in each originate from the same ancestor. The parameters for this model included the exponential population size, that was sampled under a uniform distribution with 0 and 0.5 lower and upper bounds and the growth rate parameter, drawn from a Laplace distribution. The settings also specified an HKY model for substitution rate with a gamma distribution accounting for among-site rate heterogeneity. For determination of the maximum likelihood estimator (MLE), a generalised stepping-stone (GSS) path was chosen to perform after the MCMC analysis.

The datasets were stimulated under multiple settings with different clock models that were either calibrated with the use of tip dates (heterochronous) or had no information about the tip dates and had to estimate the time of divergence (isochronous). The clock models included a strict clock model, that assumes homogenous rates among branches with the rate of evolution (substitution rate/ site) being its only parameter, and an uncorrelated relaxed clock model that was sampled under a lognormal distribution and suggests that the branch specific rates vary. Thus, there were four models created for each set of sequences: i) strict cock, isochronous, ii) strict clock, heterochronous, iii) relaxed clock, isochronous and iv) relaxed clock heterochronous.

Once the different models were created, the best fitting one for each dataset was chosen, based on the MLE score. The results were then analysed with Tracer and TempEst, in order to draw inferences about the substitution rates within and across datasets.