Phylodynamics of Ebola in West Africa

Luiz Max de Carvalho[1,a] & Andrew Rambaut[1]

[a]lm.carvalho@ed.ac.uk [1] Institute of Evolutionary Biology, University of Edinburgh, UK.

THE UNIVERSITY *of* EDINBURGH

Background

Mode and tempo of the epidemic

- The 2013-2016 West African Ebola virus disease (EVD) epidemic was the largest in history;
- A massive international collaboration produced the most comprehensive data set for an acute virus to date (over 5% sampling);
- Dudas et al. (2016) produced a rich data set with detailed information about viral movement in West Africa using 1610 viral genomes;
- Our challenge is to combine different sources of information (epidemiological, genetic, climatic, etc) to trace the epidemic and explain its mode and tempo.

▶ Which climatic and socio-economic factors predict outbreak sizes?

Association between a particular mutation and disease severity

- There is experimental evidence that an A→V mutation in the glycoprotein (GP) confers increased infectivity in human cells;

▶ Investigate the association between GP82-AV and disease severity using 236 genomes for which clinical metadata was available;

Methods

We take a Bayesian approach and estimate time-calibrated phylogenies with **BEAST**.

∮ Case counts (**Y**) along climatic and socio-economic covariates (**X**) for 56 locations → negative binomial GLM + SSVS:

$$Y_i \sim \text{NegBin}(p_i, r)$$
$$p_i = \frac{r}{(r + \lambda_i)}$$
$$\log(\lambda_i) = \alpha + \beta_1\delta_1 x_{i1} + \ldots + \beta_P\delta_P x_{iP}$$

This approach has the advantage of allowing for the calculation of Bayes factors (BF) analytically. Let (prior) w_0 the probability that no predictors are included. Then the Bayes factor for predictor x_j is

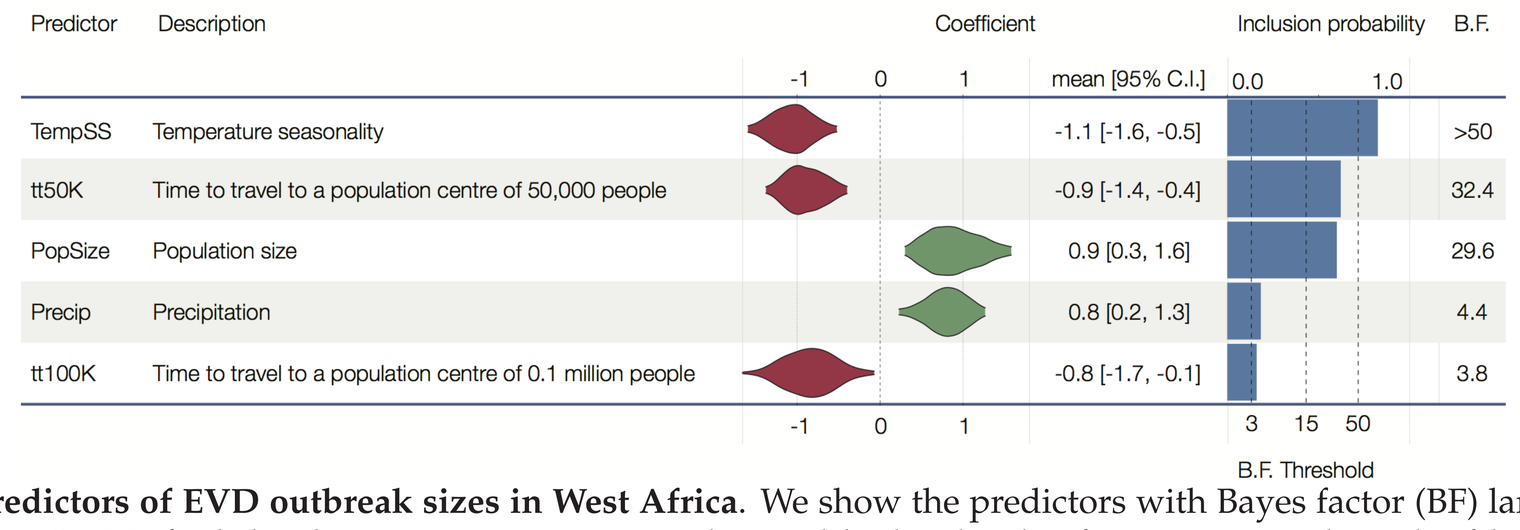
$$\text{BF}(x_j) = \frac{\hat{\delta}_j w_0^{1/P}}{(1-\hat{\delta}_j)(1-w_0^{1/P})}$$

where P is the number of predictors.

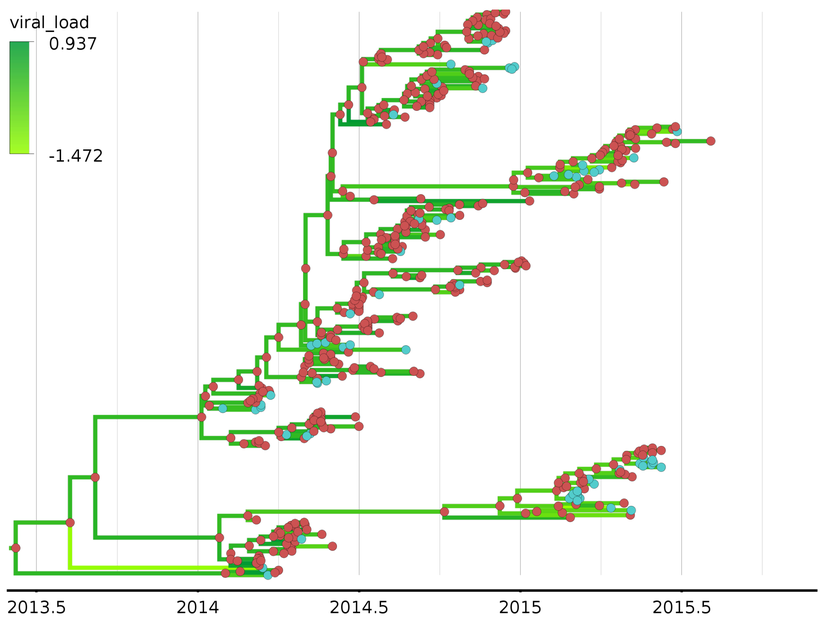
∮ We had viral load ($C(t)$ values) and outcome (death/survival) information for 236 patients → binomial GLM;

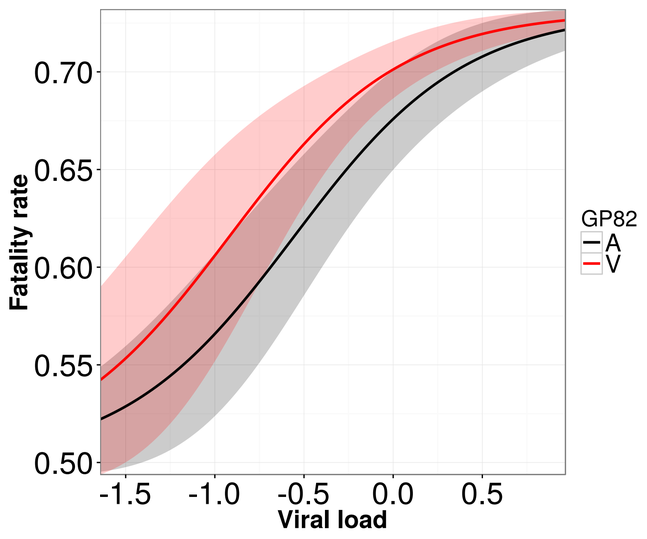
∮ Constructing outbreak clusters: every time a location receives a viral introduction that leads to >20 sampled cases, split the series before and after the introduction.

Results

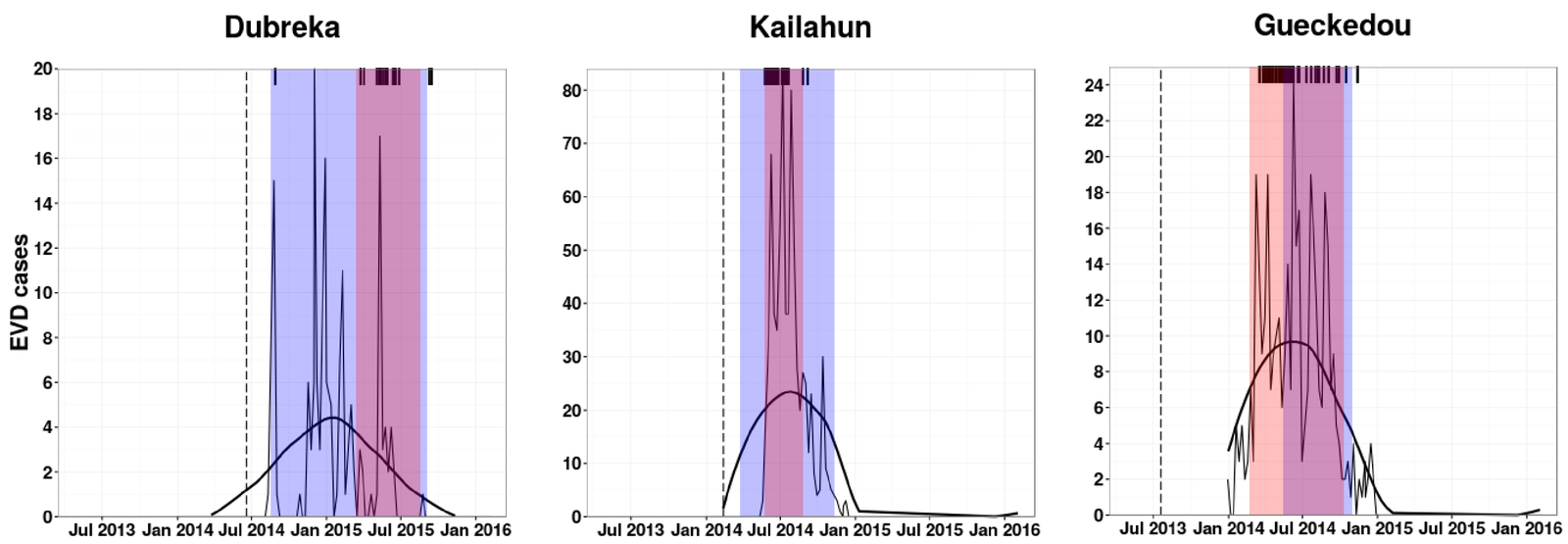


Predictors of EVD outbreak sizes in West Africa. We show the predictors with Bayes factor (BF) larger than 3. We find that low temperature seasonality and higher levels of rain increase the risk of larger outbreaks. Similarly, urbanicity seems to play a role, with locations closer to urban centres being at higher risk.





Association between the GP82-AV and disease severity. In the left panel we show a time-calibrated phylogeny with 236 sequences from patients whose outcome information was available. Branches with lower viral loads (transformed $C(t)$) have higher probability of leading to surviving tips, suggesting some degree of heritability of infectivity. The right panel shows the predicted fatality rates conditional on viral load for each genotype (wild type: **A**, mutant: **V**). Notice that although fatality rates seem be higher for **V** the confidence bands overlap considerably, indicating substantial uncertainty.



Combining epidemiological and genetic data. Here we overlay case notifications data with inferred introductions to gain insight into the dynamics of EVD the epidemic. Dashed vertical line marks the earliest introduction in the posterior. Shaded areas show the 95% credibility interval for introductions (blue) and exports (red). Rugplots show dates of sampling of the sequences in that location.

Conclusions

- Both climatic and social economic factors predict outbreak sizes. We have employed this approach to help explain why some regions did not experience cases despite being in close proximity with affected areas.
- The A→V mutation in the EBOV glycoprotein is weakly associated with higher fatality rates;
- By combining epidemiological information and viral movement inferred from sequenced genomes one can gain insight into the dynamics of Ebola in West Africa. Our challenge now is to integrate all of this information in a coherent statistical model.

Funding: I thank the School of Biological Sciences for topping up my stipend.