PHYLODYNAMICS OF EBOLA VIRUS IN WEST AFRICA

LUIZ MAX DE CARVALHO[1,a]

THE UNIVERSITY of EDINBURGH

[a] lm.carvalho@ed.ac.uk
[1] Institute of Evolutionary Biology, University of Edinburgh, UK.
Supervisors: Andrew Rambaut and Darren Obbard.

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). Detailed information about viral movement in West Africa using 1610 viral genomes was recently made available (Dudas et al. (2016)). A major challenge is to develop the statistical and computational tools 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. I investigate the association between GP82-AV and disease severity using 236 genomes for which clinical metadata was available.

METHODS

Weekly reported cases were available for 56 locations. For 236 cases complete metadata, including C_t values and outcome (death/survival) information was available. C_t stands for "threshold cycle" and records the number of qPCR cycles required to detect a real signal from the samples. These values are inversely correlated with viral load. Using 1610 genomes, Dudas et al. (2016) inferred how many viral introductions each location received, and how many samples resulted from each introduction.

§ Let Y_i be total number of cases reported for location i. Let also $\mathbf{X_i}$ be a set of climatic and socio-economic covariates measured at location i. I then employ a negative binomial generalised linear model (GLM) coupled with stochastic search variable selection:

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

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

§ To use C_t values as a proxy for viral load and also consistent with the mutation (binary) predictor we standardise C_t by subtracting from the mean and dividing by two standard deviations. I then fit a binomial GLM using outcome as response variable and the absence/presence of the mutation and C_t values as covariates.

§ 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

Predictor	Description	Coefficient -1 0 1	mean [95% C.I.]	Inclusion probability 0.0 1.0	B.F.
TempSS	Temperature seasonality		-1.1 [-1.6, -0.5]		>50
tt50K	Time to travel to a population centre of 50,000 people		-0.9 [-1.4, -0.4]		32.4
PopSize	Population size		0.9 [0.3, 1.6]		29.6
Precip	Precipitation		0.8 [0.2, 1.3]		4.4
tt100K	Time to travel to a population centre of 0.1 million people		-0.8 [-1.7, -0.1]		3.8

-1 0 1   3 15 50

B.F. Threshold

Predictors of EVD outbreak sizes in West Africa. We show the predictors with Bayes factor (BF) greater 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 having more cases (controlling for population sizes).
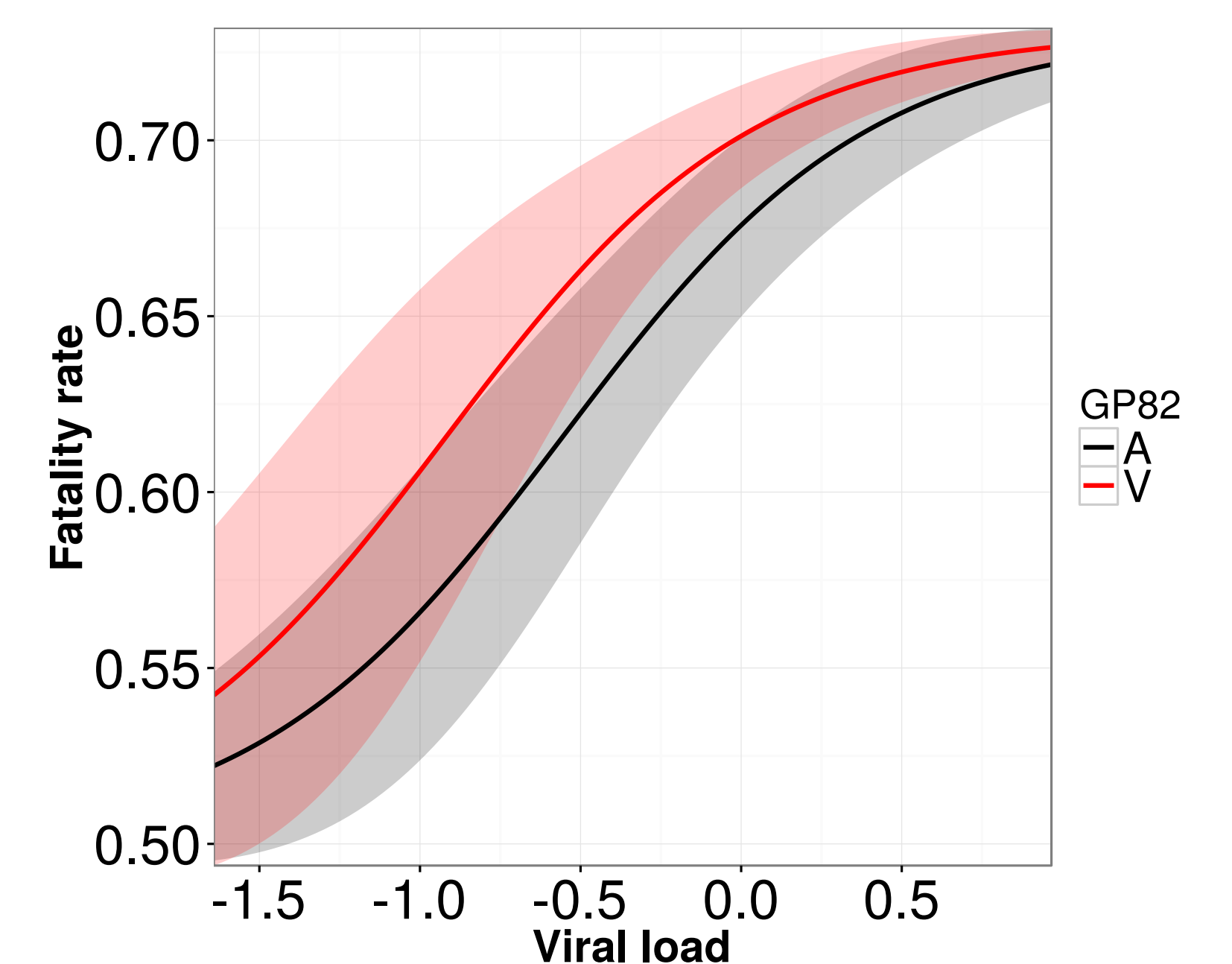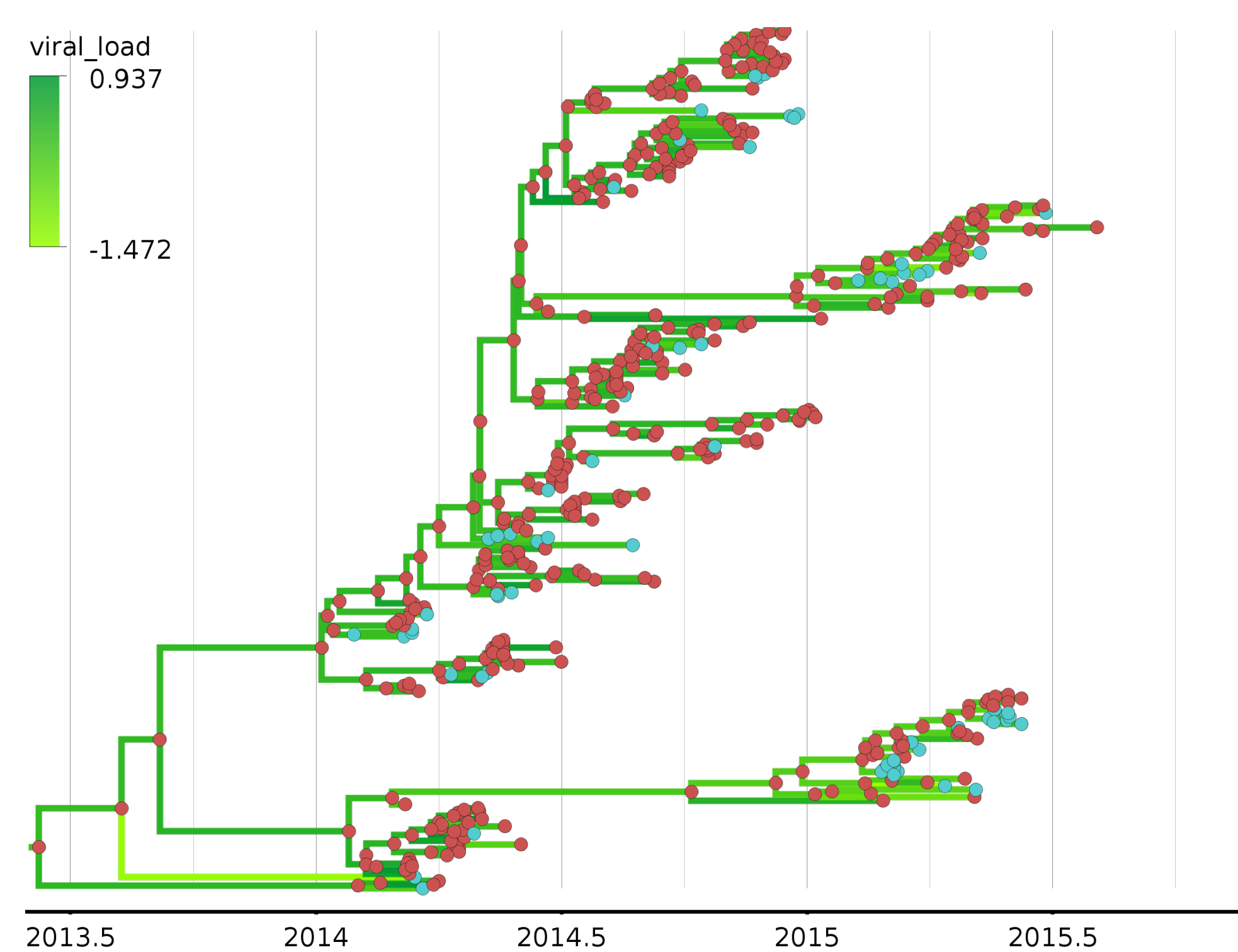
Association between the GP82-AV and disease severity. The left panel shows a time-calibrated phylogeny with 236 sequences from patients whose outcome information was available. Branches with lower viral loads (transformed C_t) have higher probability of leading to cases where the patient survived, suggesting 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.
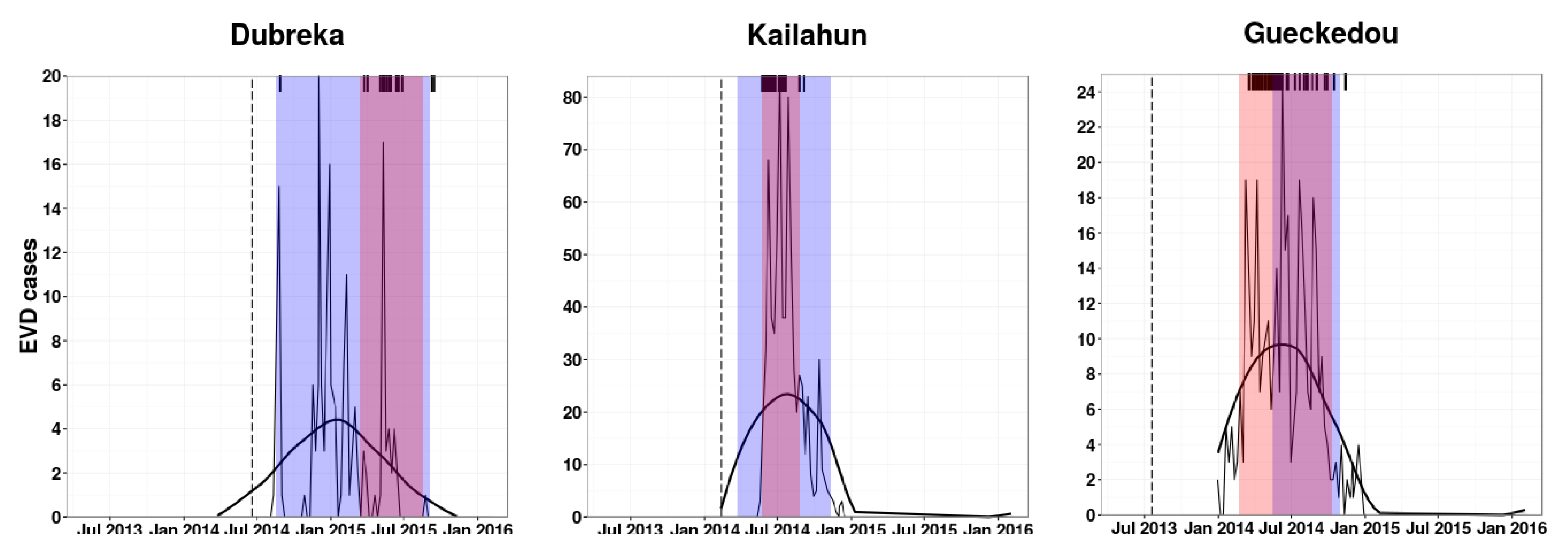
Combining epidemiological and genetic data. Here weekly recorded cases for three representative locations are overlaid with inferred introductions to gain insight into the dynamics of EVD the epidemic. Dashed vertical line marks the earliest introduction in the posterior. Tick marks at the top show the 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.