University of Glasgow | Boyd Orr Centre for Population and Ecosystem Health

Using diversity-based methods to estimate true epidemic sizes from sampled outbreaks

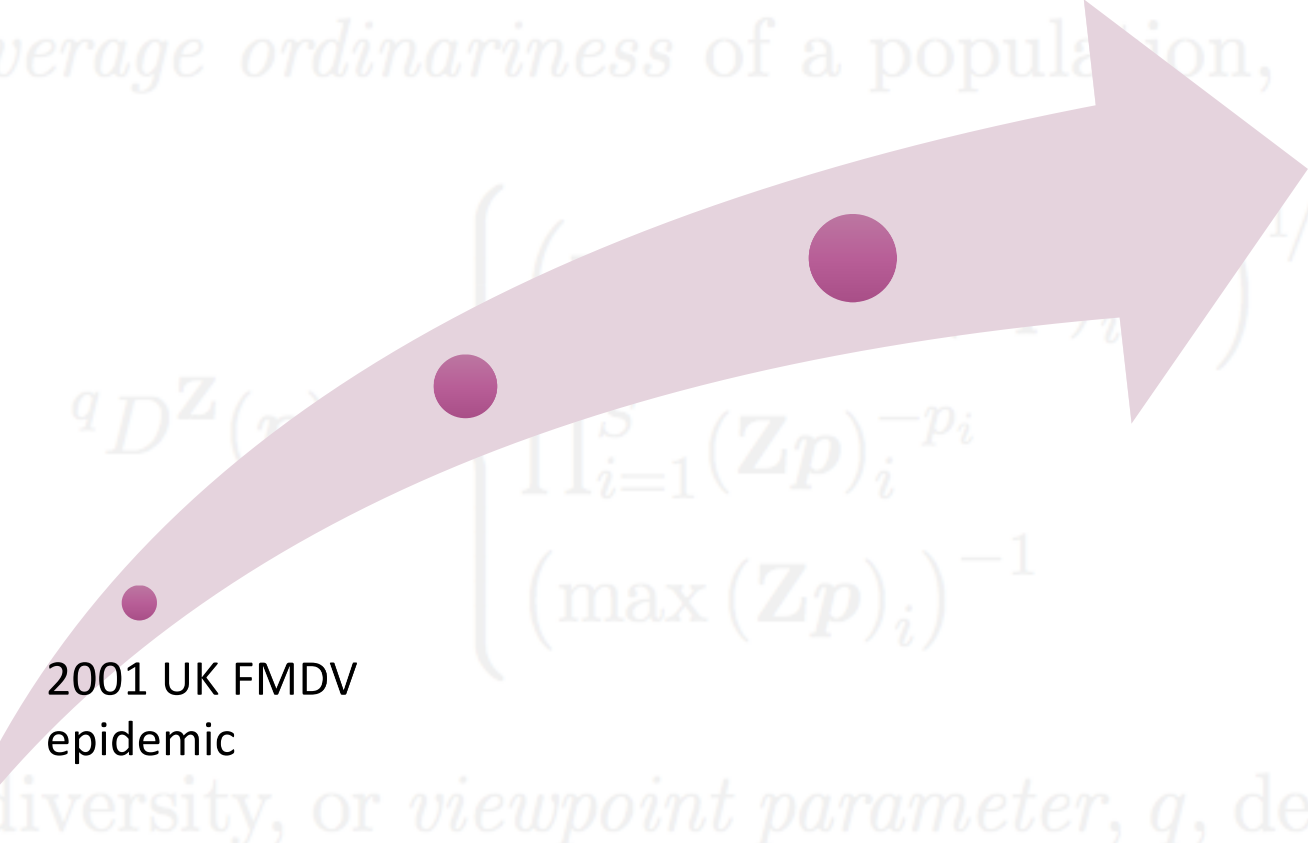
Sonia Mitchell

Richard Reeve
Louise Matthews
Christina Cobbold
Antonello Di Nardo

BBSRC

Previous work..

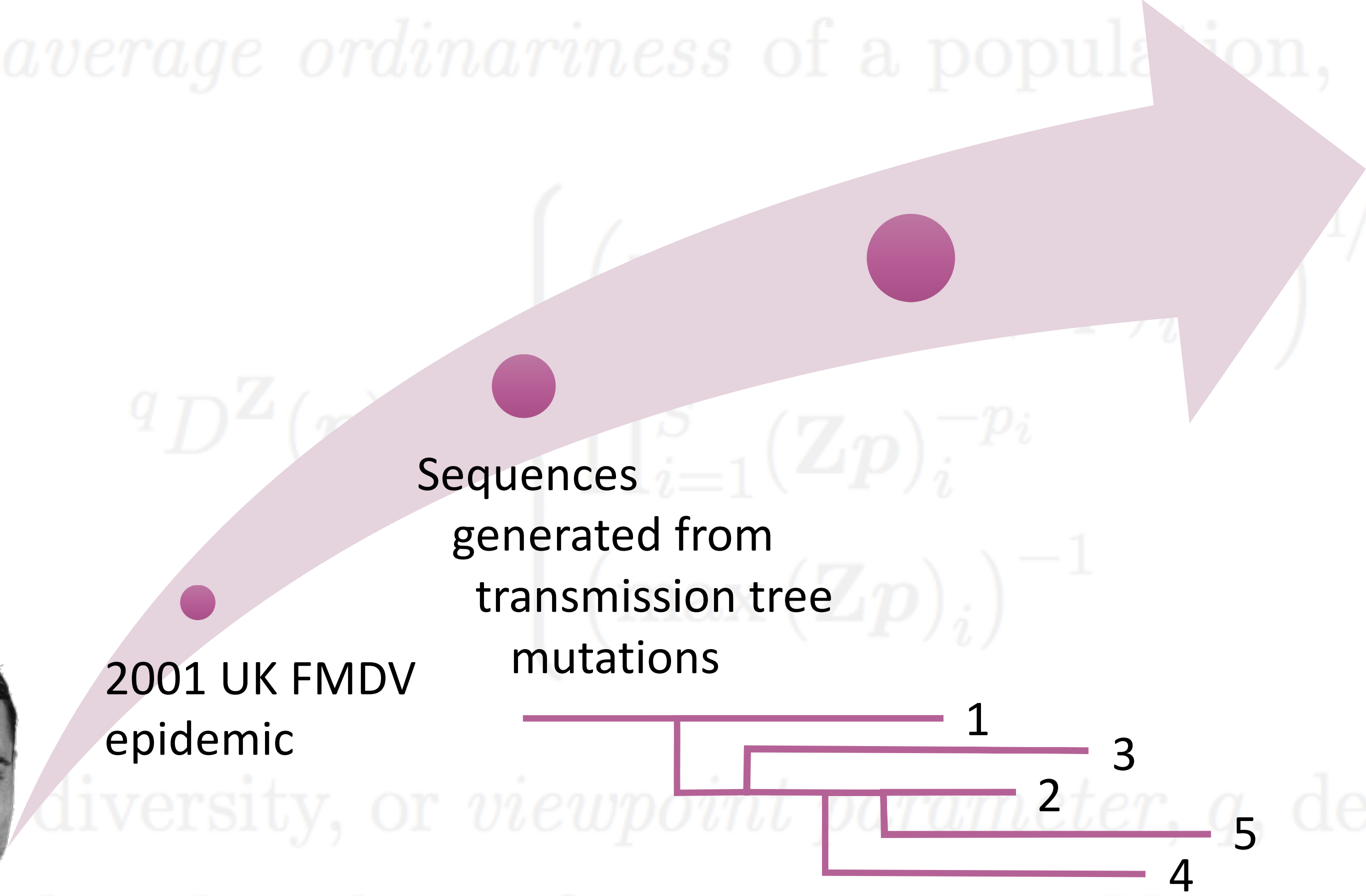
Antonello Di Nardo



2001 UK FMDV
epidemic

Di Nardo, A. 2016. *Phylodynamic modelling of foot-and-mouth disease virus sequence data*.

Antonello Di Nardo

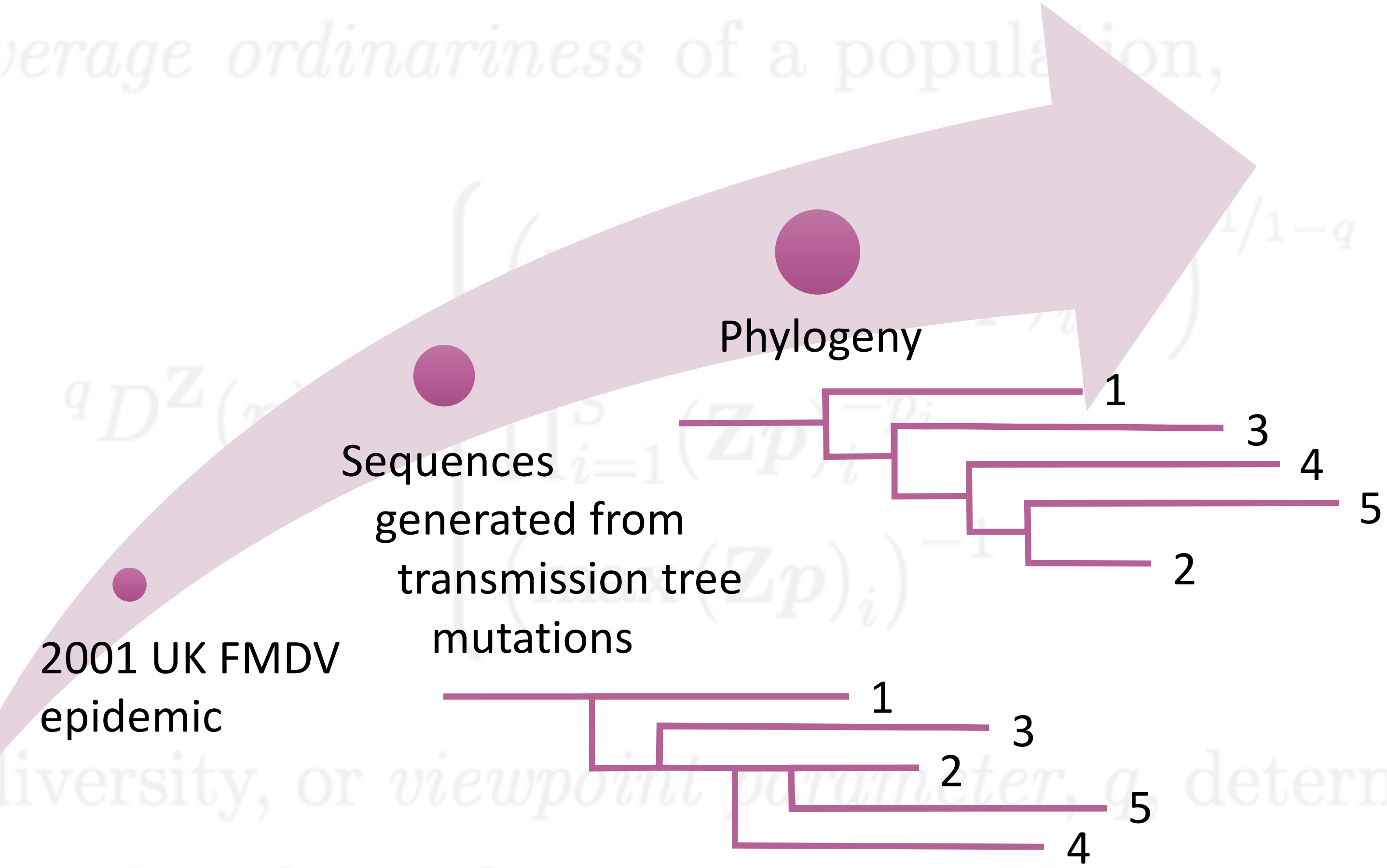


2001 UK FMDV epidemic

Sequences generated from transmission tree mutations

Di Nardo, A. 2016. *Phylodynamic modelling of foot-and-mouth disease virus sequence data*.

Antonello Di Nardo



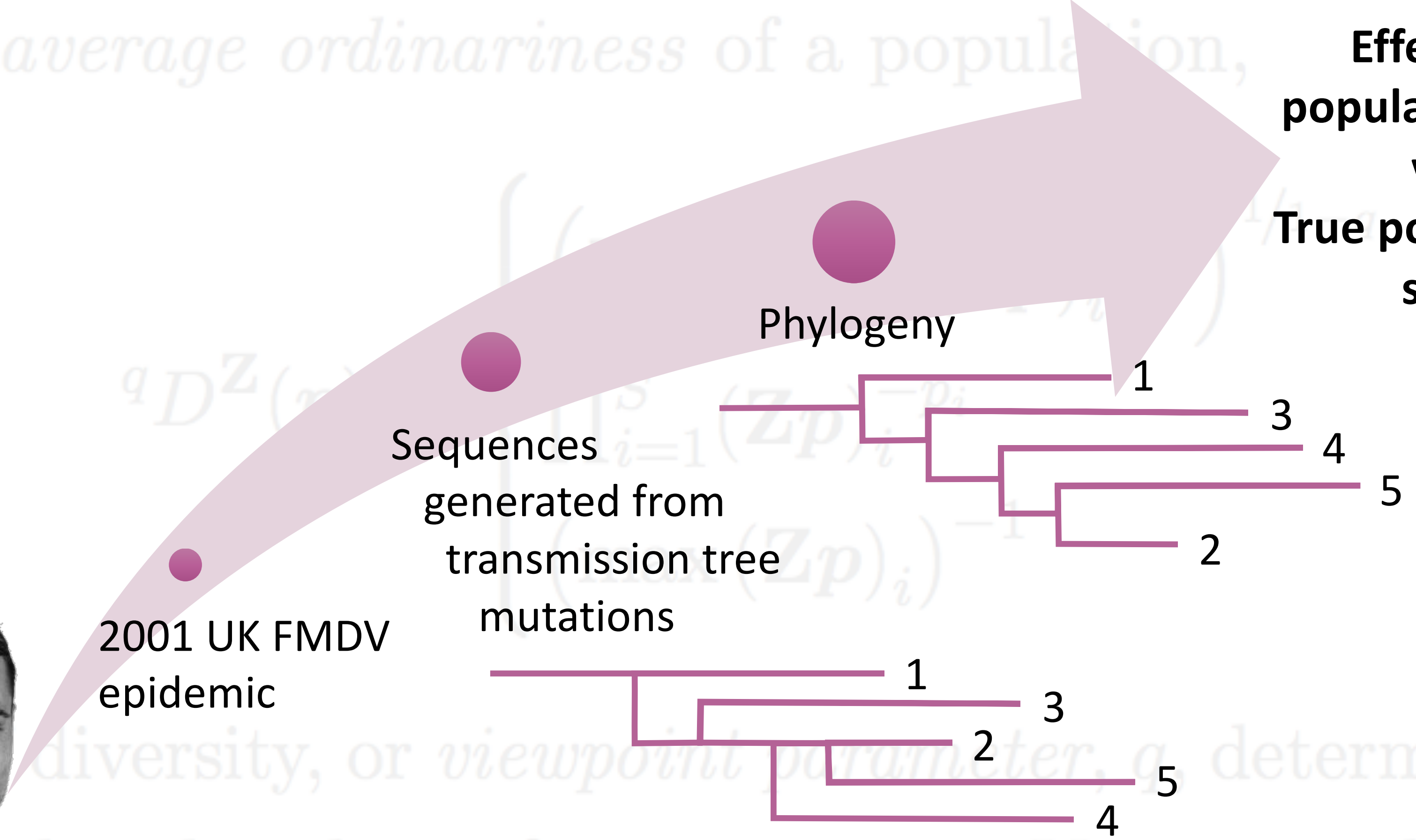
2001 UK FMDV epidemic

Sequences generated from transmission tree mutations

Phylogeny

Di Nardo, A. 2016. *Phylodynamic modelling of foot-and-mouth disease virus sequence data*.

Antonello Di Nardo



2001 UK FMDV epidemic

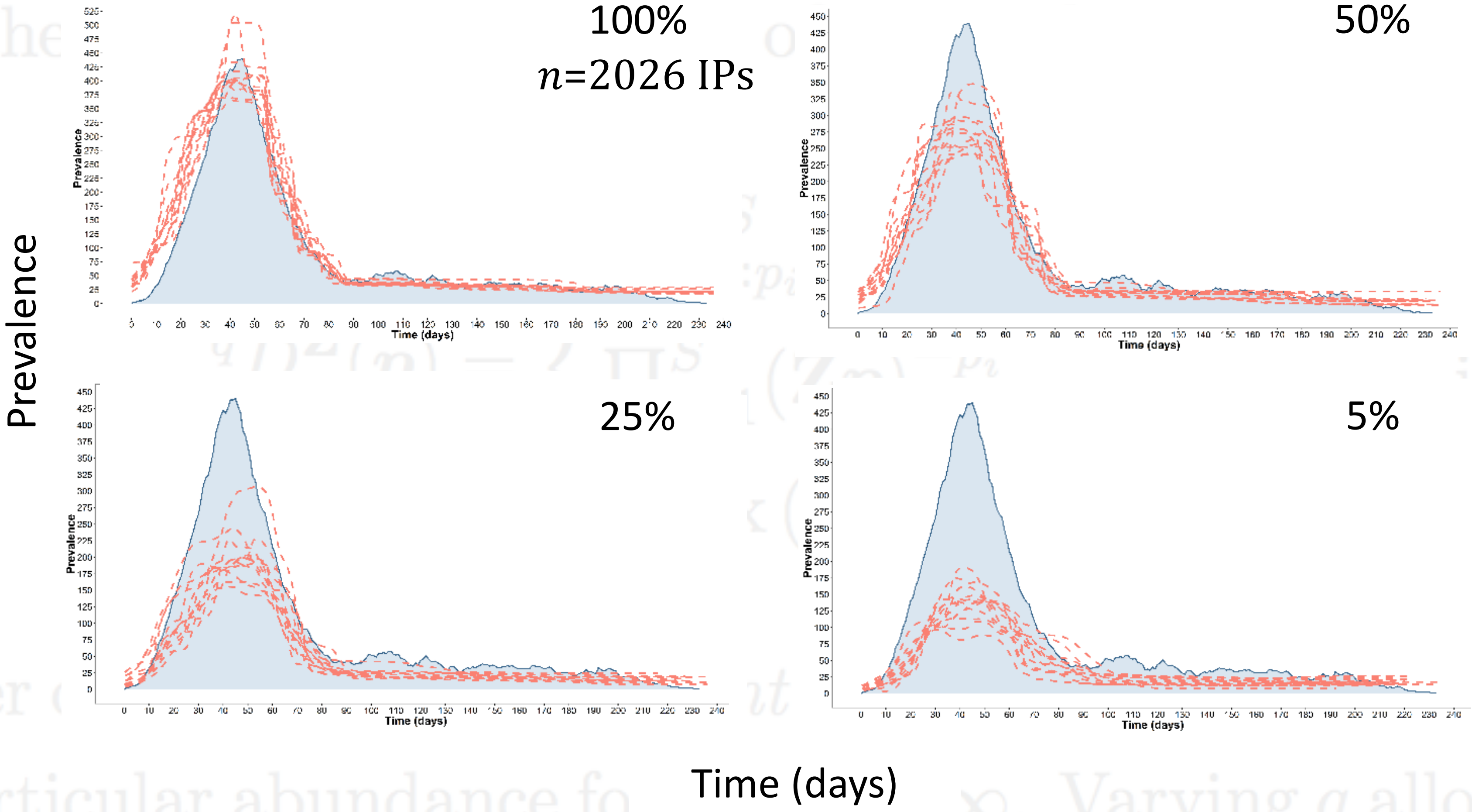
Sequences generated from transmission tree mutations

Phylogeny

Effective population size vs. True population size

Di Nardo, A. 2016. *Phylodynamic modelling of foot-and-mouth disease virus sequence data*.

Antonello Di Nardo



Di Nardo, A. 2016. *Phylodynamic modelling of foot-and-mouth disease virus sequence data*: Figure 5-2.

Antonello Di Nardo

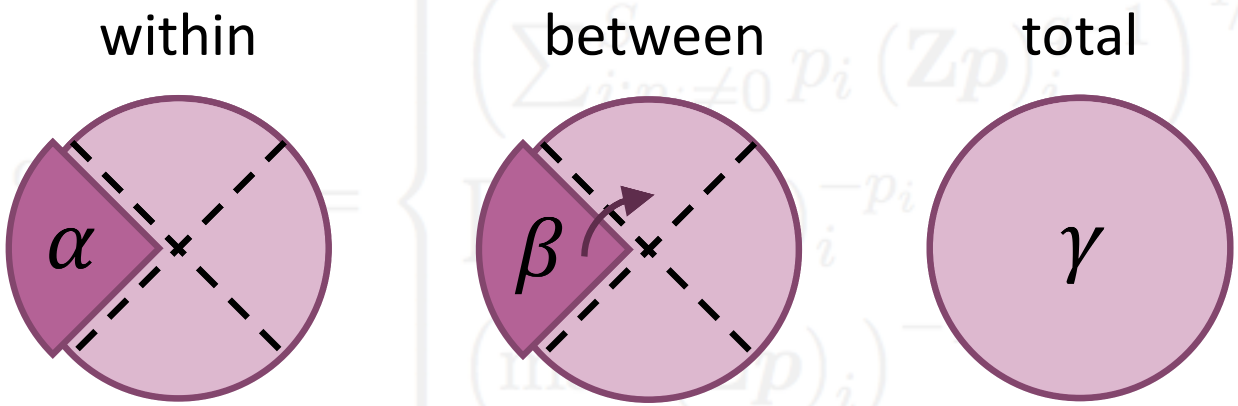
Incomplete surveillance data may cause a mismatch between N_e and N such that..

$$N_e \ll N$$

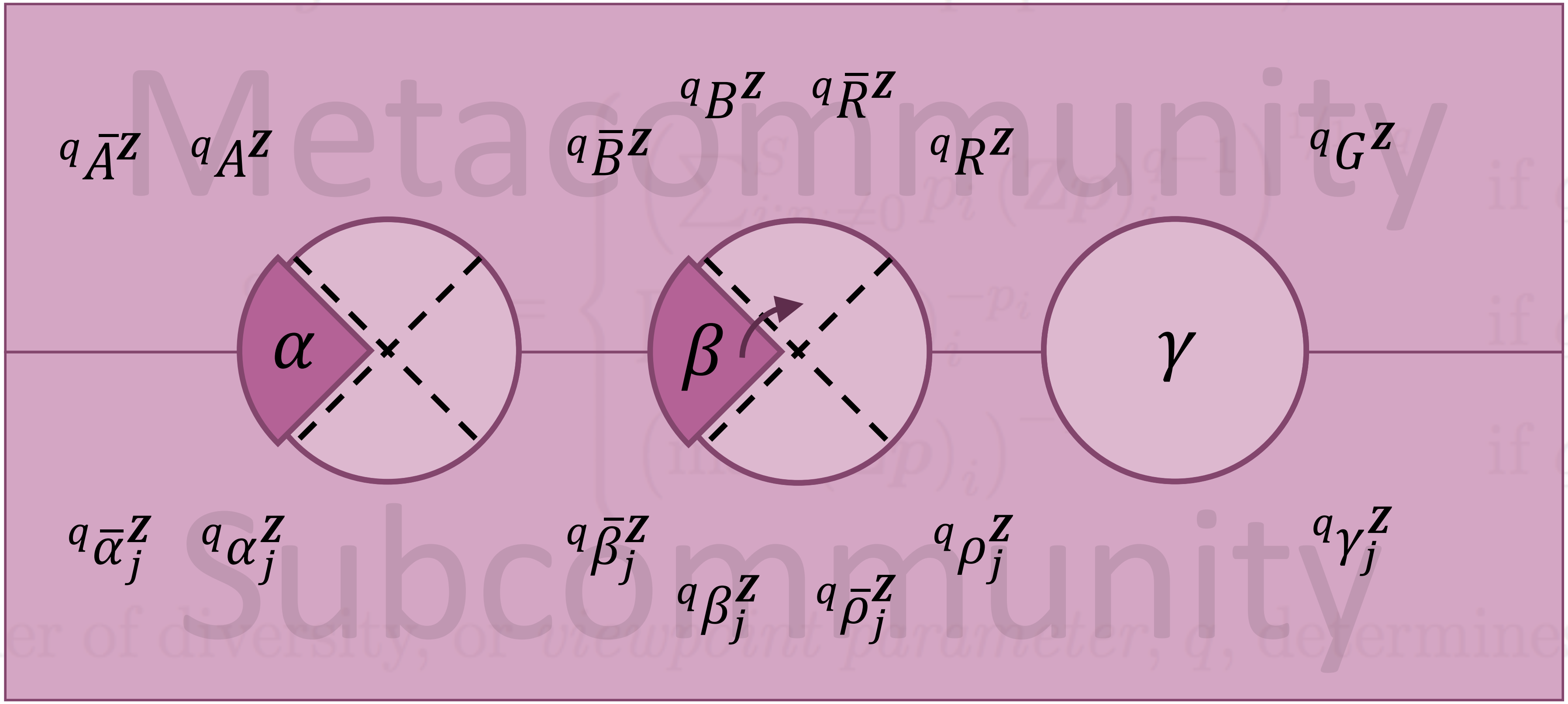
Di Nardo, A. 2016. *Phylodynamic modelling of foot-and-mouth disease virus sequence data*: Figure 5-2.

We propose a new method..

Framework for Partitioning Diversity



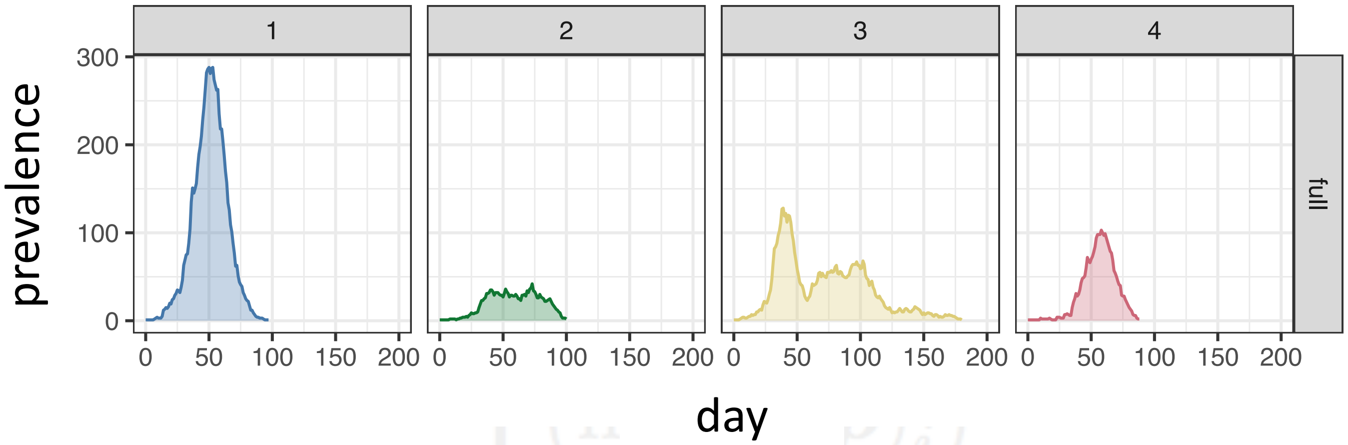
Framework for Partitioning Diversity



Reeve, R., et al. 2016. *How to partition diversity*. arXiv 1404.6520:1–9.

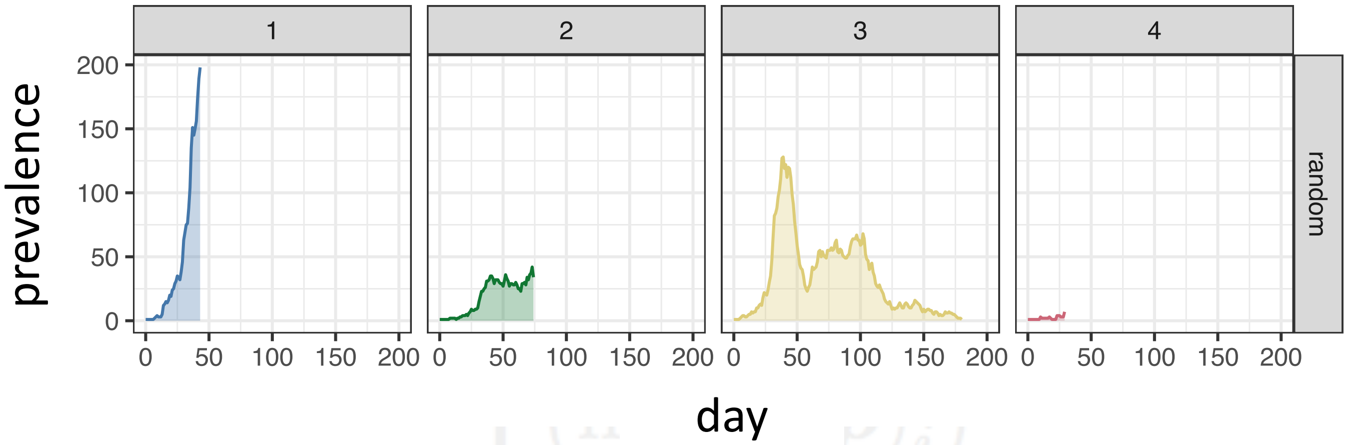
Methods

- Simulate outbreaks



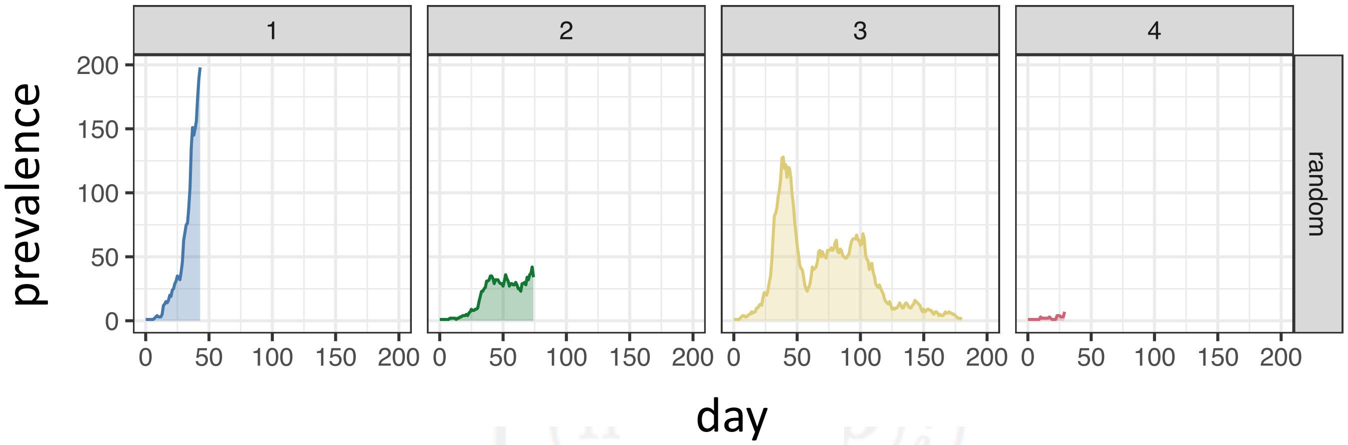
Methods

- Simulate outbreaks (truncated)



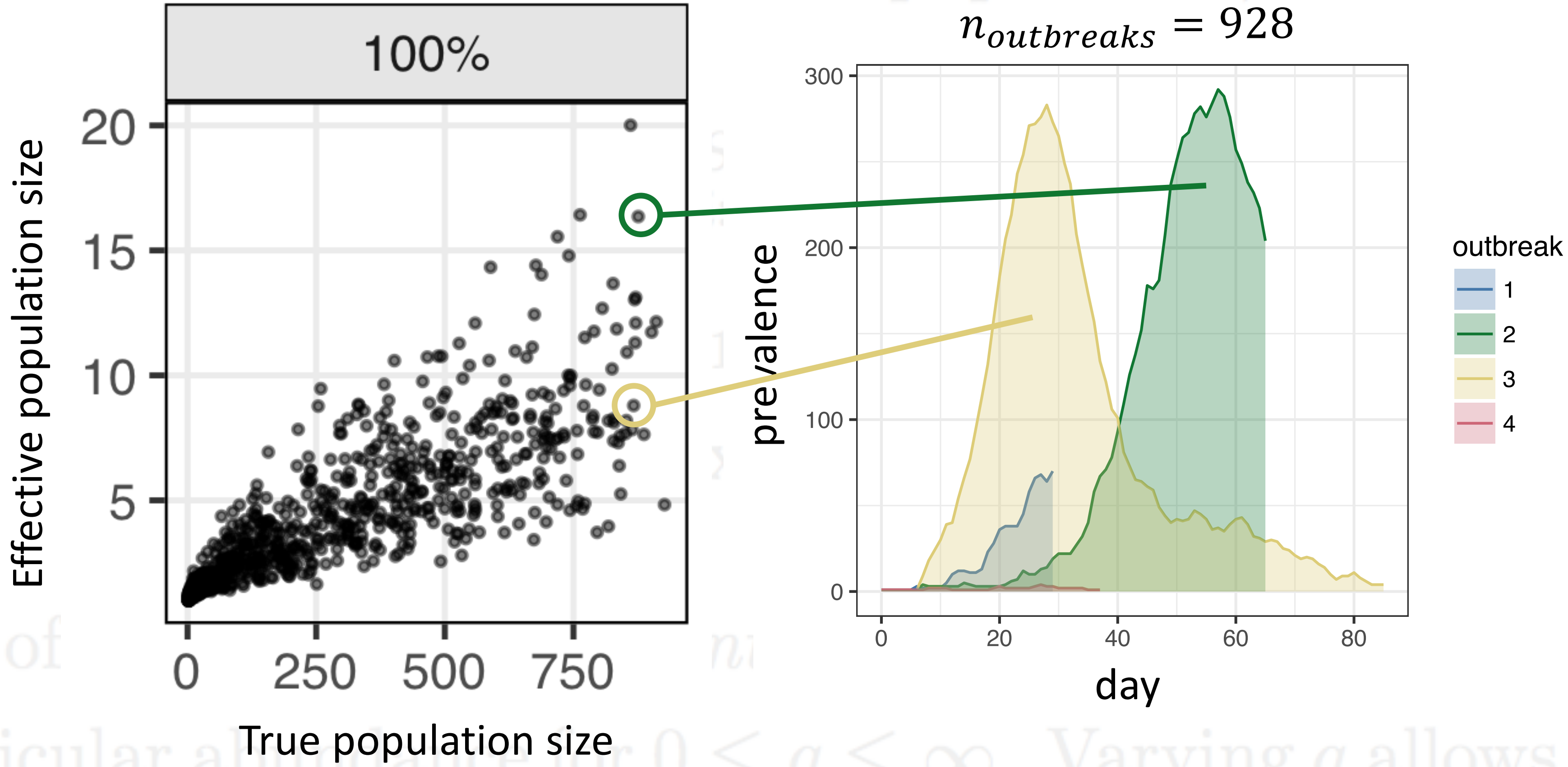
Methods

- Simulate outbreaks (truncated)

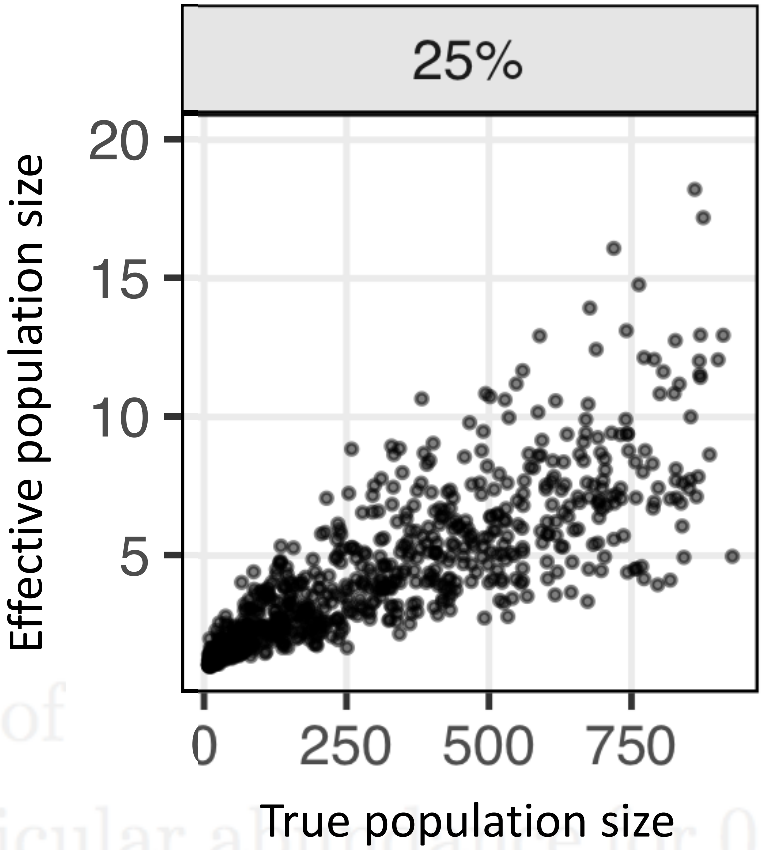


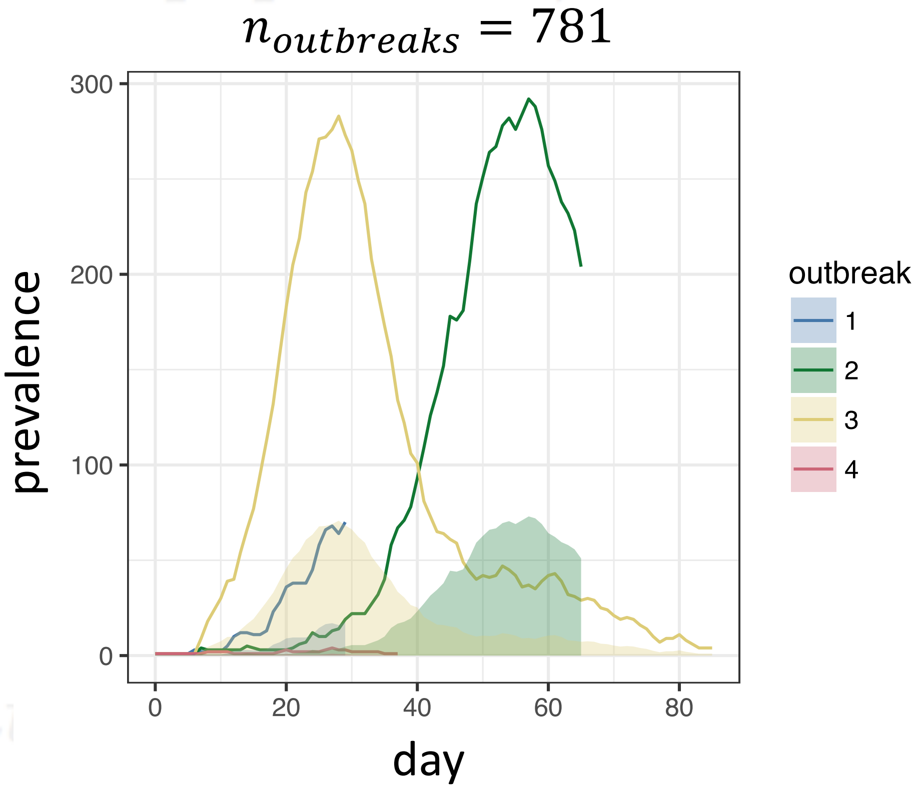
- Viral genetic diversity of the 2001 FMDV outbreak
- Calculate ${}^{q}G^{\boldsymbol{Z}}$ (Effective number of sequences / infected premises)

Effective number of sequences, $^{1}G^{\mathbf{Z}}$

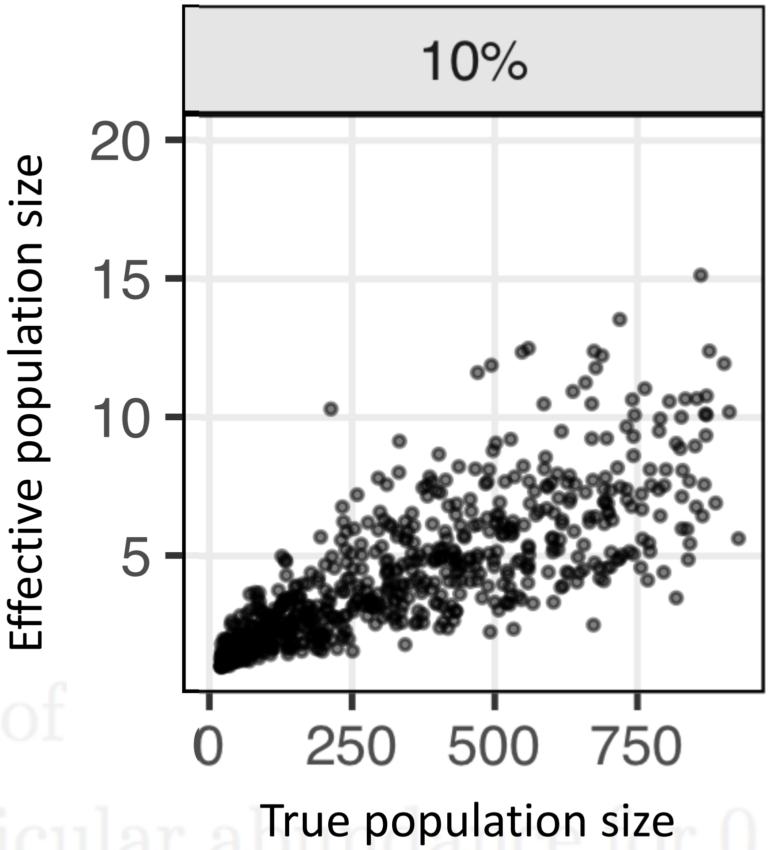


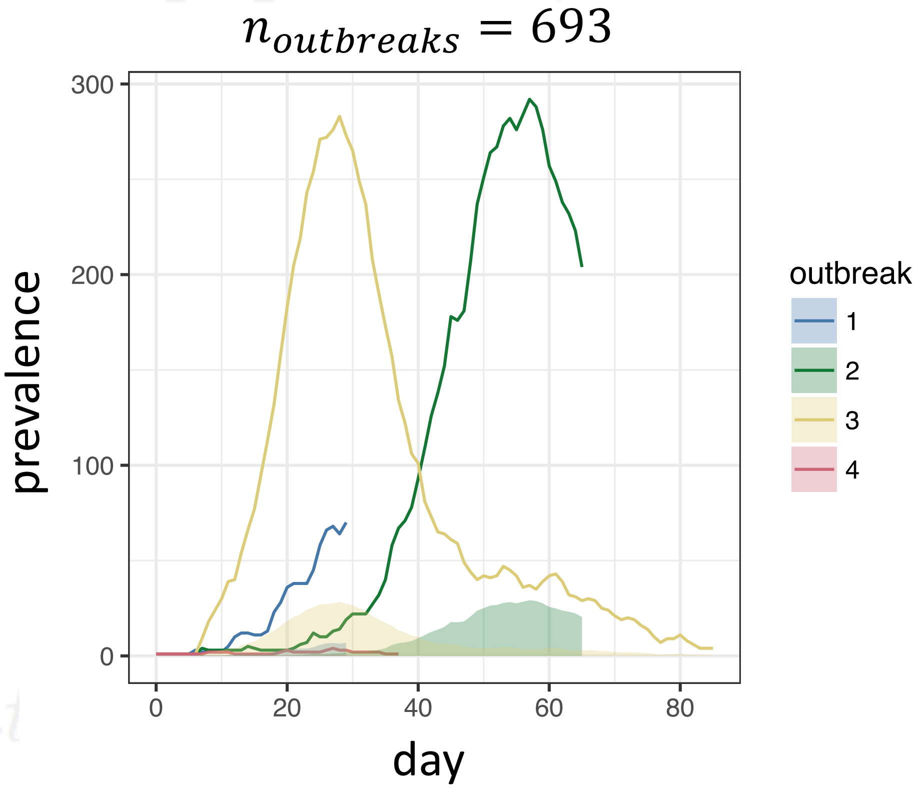
Effective number of sequences, ${}^{1}G^{\mathbf{Z}}$




$n_{outbreaks} = 781$

Effective number of sequences, $^1G^{\mathbf{Z}}$

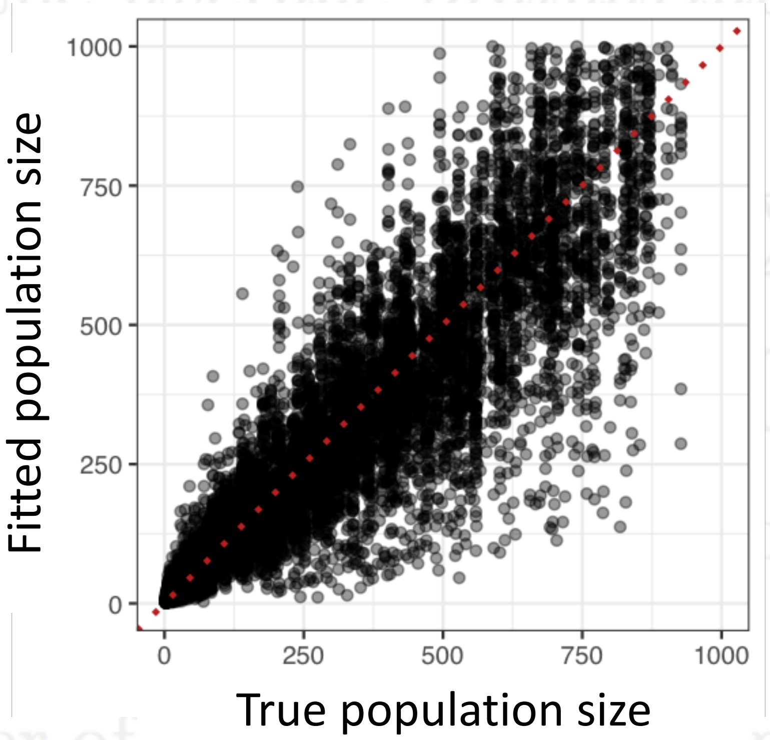




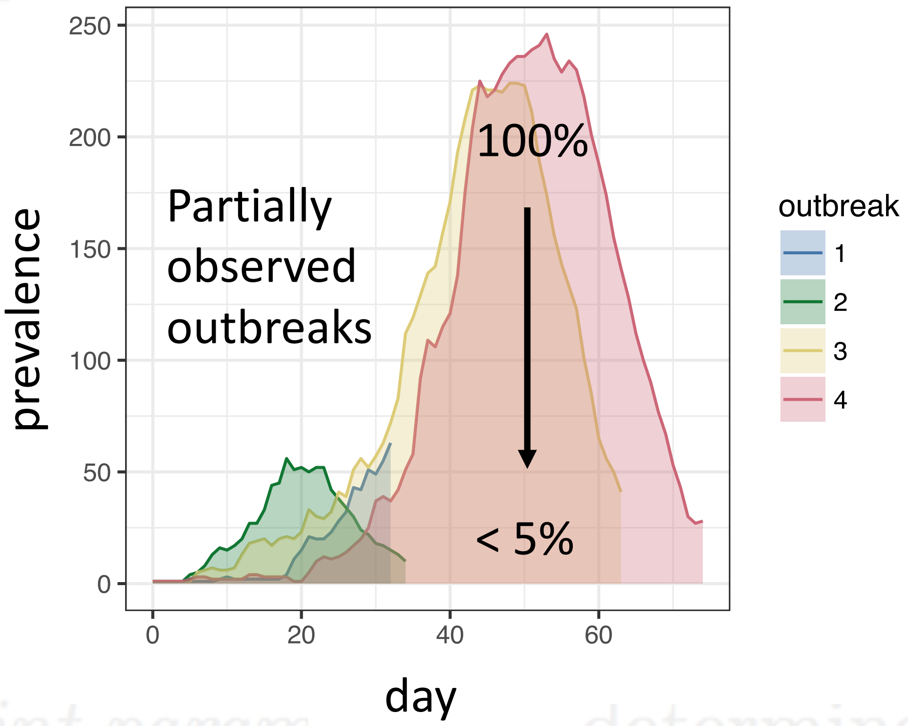
Next, we use ${}^{q}G^{\mathbf{Z}}$ with other simple summary statistics..

..to fit a linear model.

Model fit



Adjusted $R^2 = 0.9001$



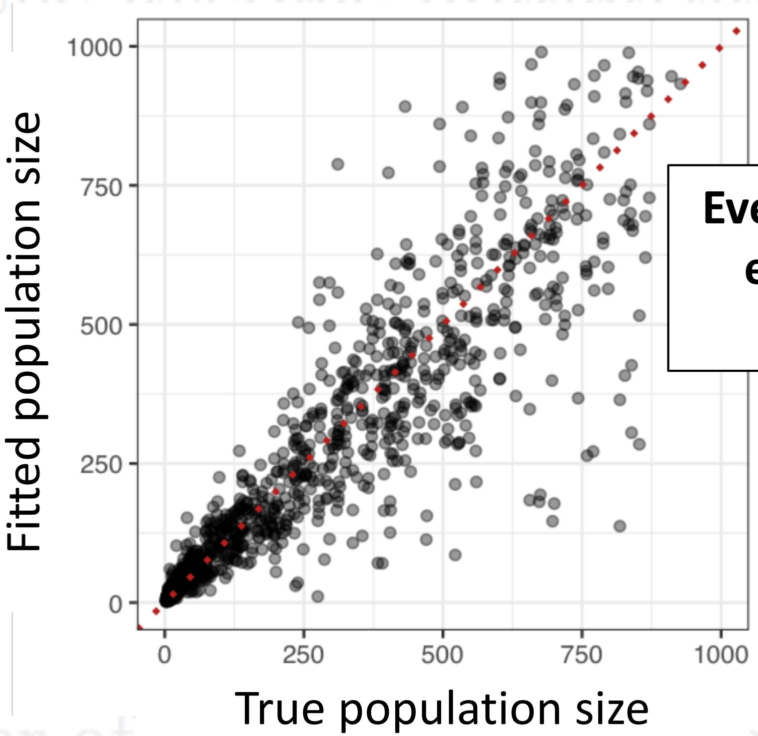
$n_{outbreaks} = 10759$

100%

Partially observed outbreaks

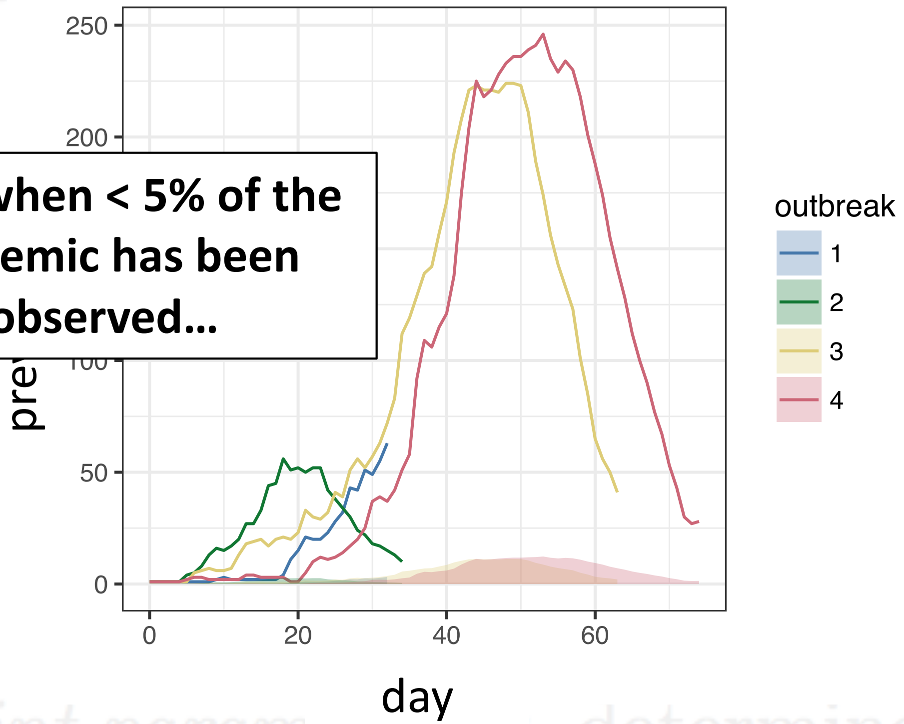
< 5%

Model fit



Even when < 5% of the epidemic has been observed…

$n_{outbreaks} = 4038$



Conclusion

- We can infer epidemic size from partially observed outbreaks
- These methods are more accurate than existing techniques

WHAT NEXT?

- Incorporate measures of β-diversity
- Apply more advanced machine learning techniques, *e.g.* Google's TensorFlow

https://www.tensorflow.org

University of Glasgow | Boyd Orr Centre for Population and Ecosystem Health

- From the R console: `install.packages('rdiversity')`
- Reeve, R. et al. (2016). *How to partition diversity*. arXiv:1404.6520 [q-bio.QM].
- Di Nardo, A. 2016. *Phylodynamic modelling of foot-and-mouth disease virus sequence data*. [PhD Thesis]

Richard Reeve
Antonello Di Nardo

Christina Cobbold
Louise Matthews

BBSRC