



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





& Cobbold (2012) derive a general expression for similarity numbers, is based on Rényi's generalised entropy and expands this framework reveals a coherent family of diversity measures. This framework reveals a coherent family of diversity measures of the *average ordinariness* of a population,

$${}^q D^{\mathbf{Z}}(\mathbf{p}) = \begin{cases} \left(\sum_{i:p_i \neq 0}^S p_i (\mathbf{Zp})_i^{q-1} \right)^{1/(1-q)} & \text{if } q \neq 1 \\ \prod_{i=1}^S (\mathbf{Zp})_i^{-p_i} & \text{if } q = 1 \\ \left(\max_i (\mathbf{Zp})_i \right)^{-1} & \text{if } q = \infty \end{cases}$$

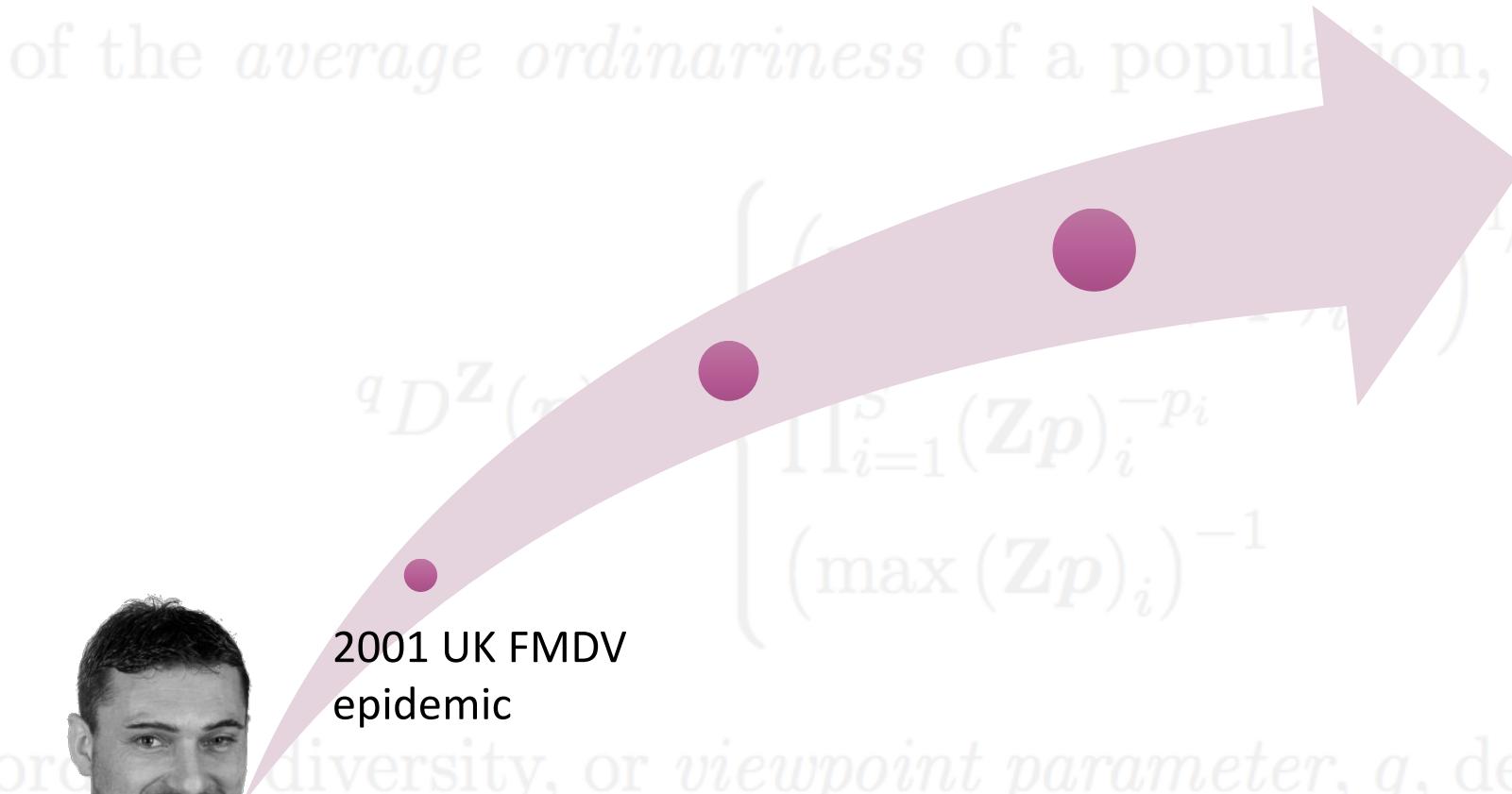
Previous work..

The order of diversity, or *viewpoint parameter*, q , determines the weight given to a particular abundance for $0 \leq q \leq \infty$. Varying q allows greater emphasis to be placed on species of differing ordinariness. Specifically,



Antonello Di Nardo

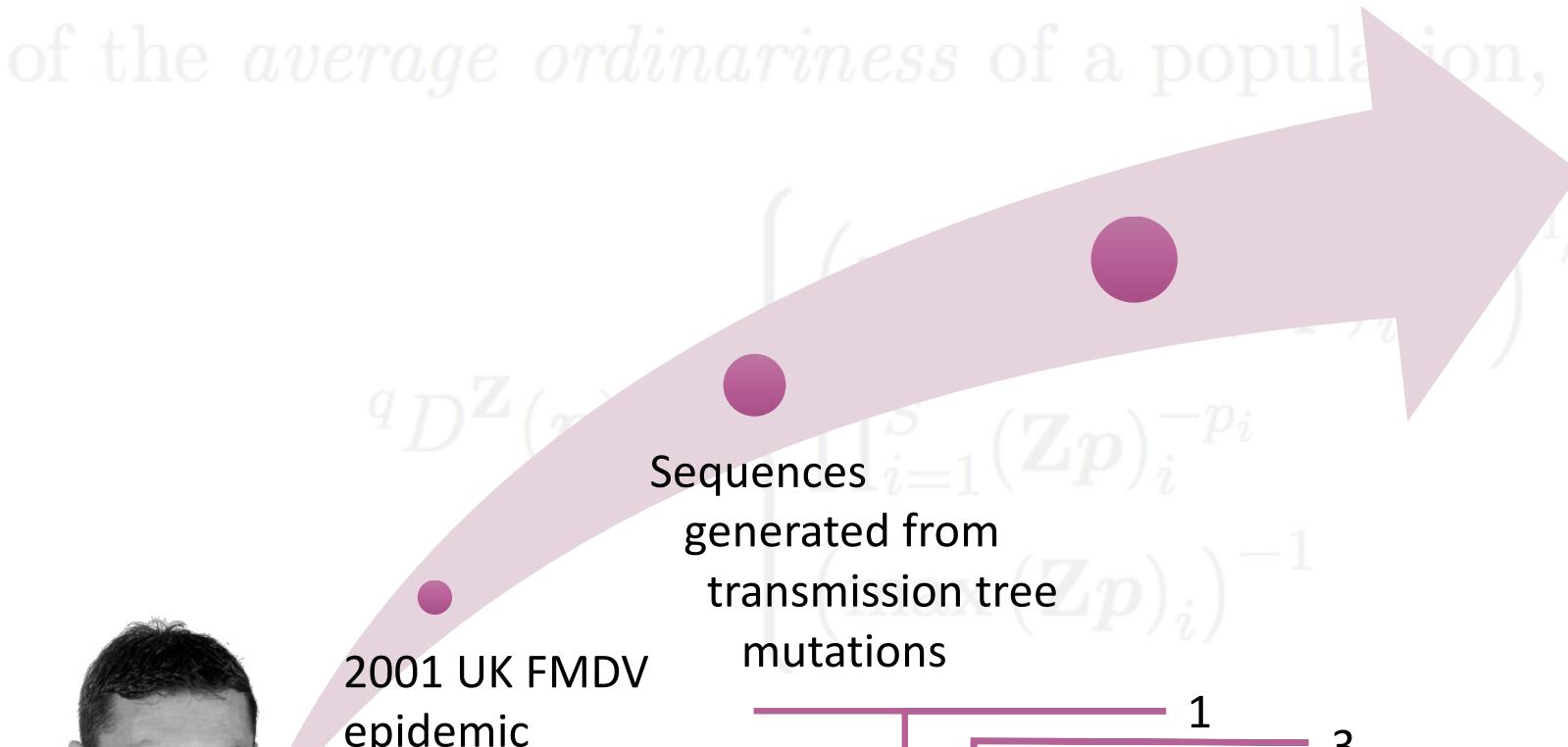
This framework reveals a coherent family of diversity measures of the *average ordinariness* of a population,





Antonello Di Nardo

This framework reveals a coherent family of diversity measures of the *average ordinariness* of a population,

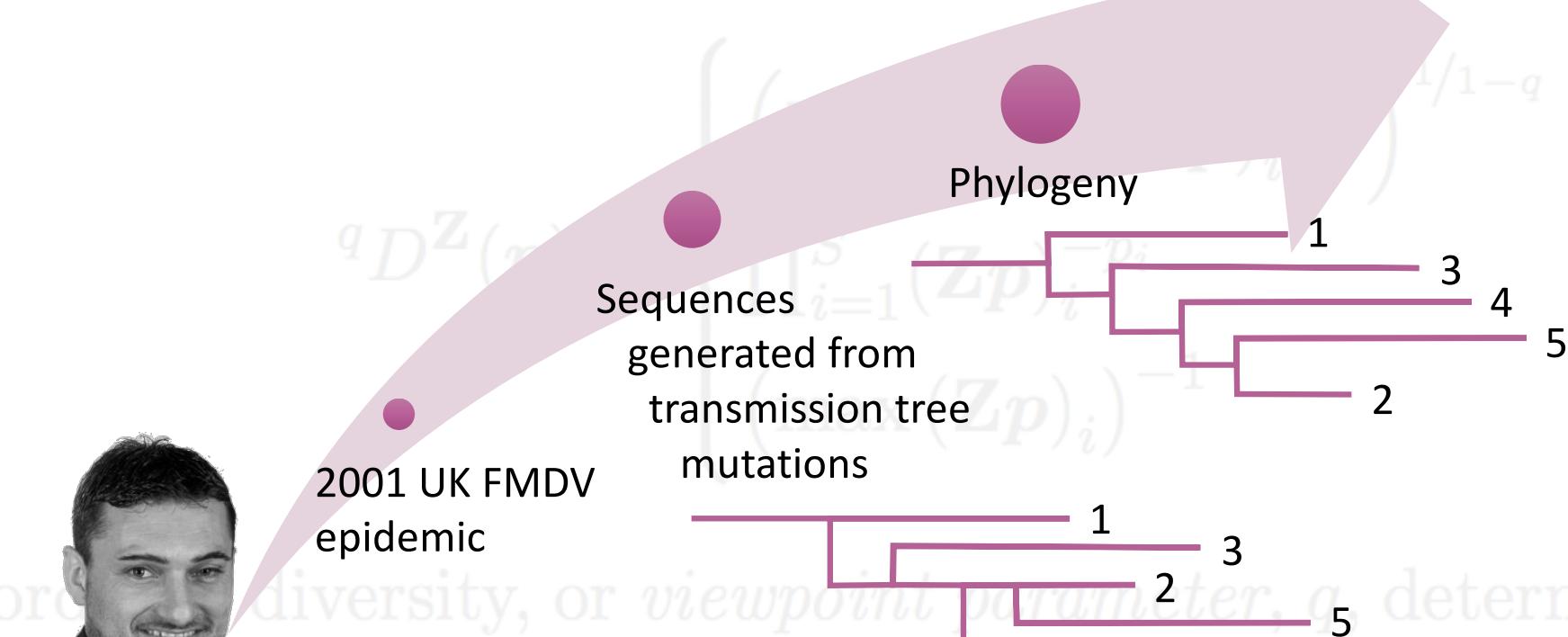




University
of Glasgow

Antonello Di Nardo

This framework reveals a coherent family of diversity measures of the *average ordinariness* of a population,



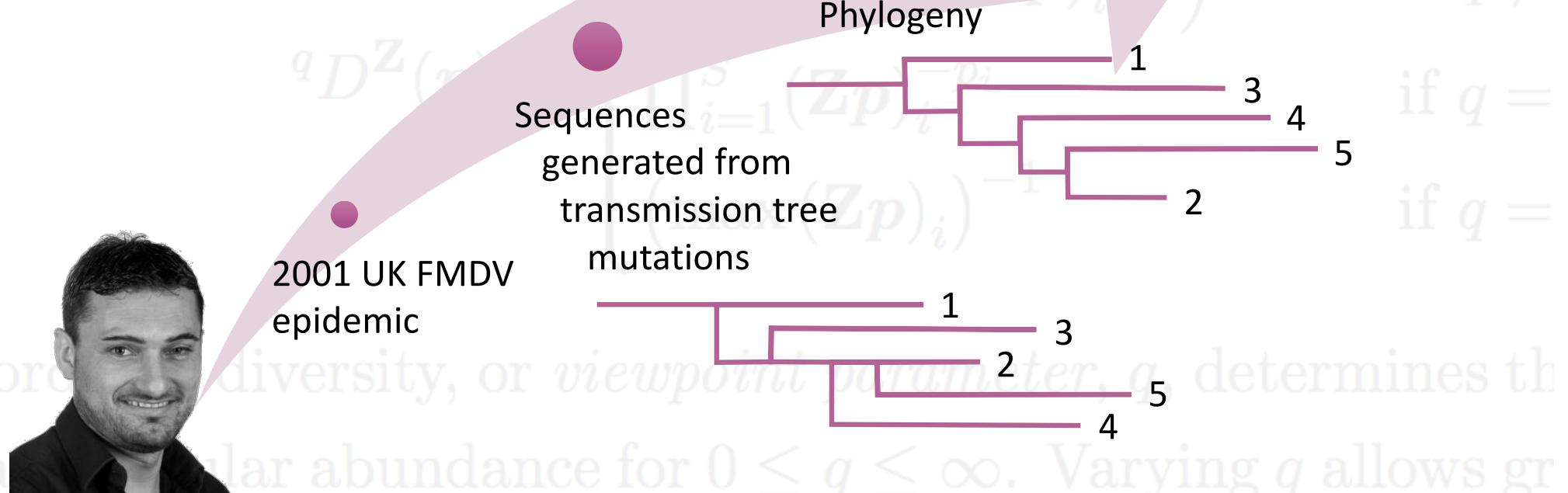


University
of Glasgow

Antonello Di Nardo

This framework reveals a coherent family of diversity measures of the *average ordinariness* of a population,

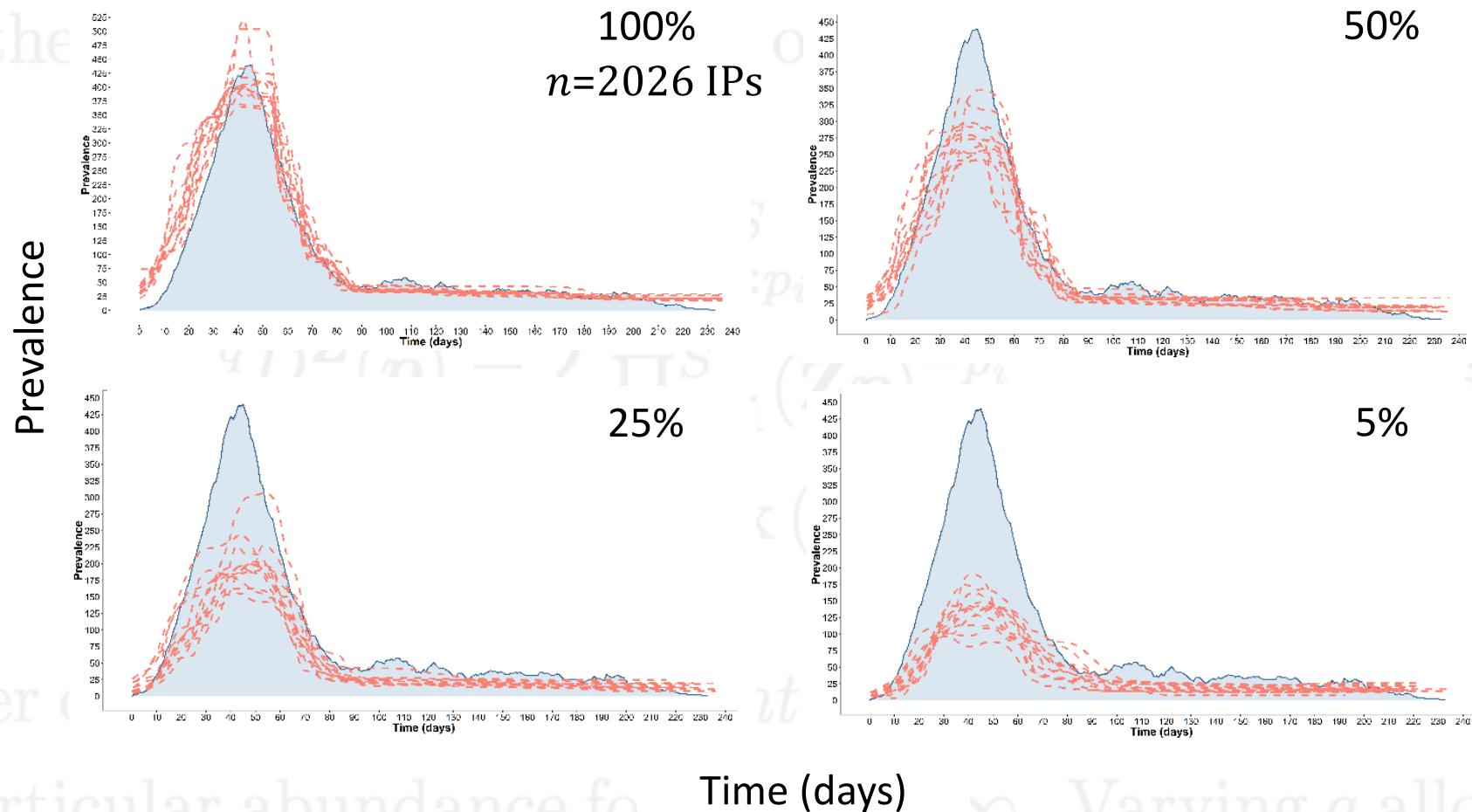
Effective population size vs. True population size



Antonello Di Nardo



University
of Glasgow



& Cobbold (2012) derive a general expression for similarity numbers is based on De Vries' generalised expression and exp

Antonello Di Nardo



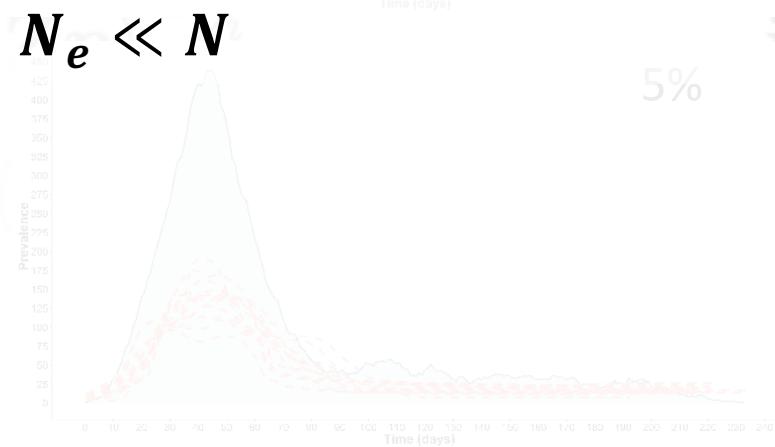
University
of Glasgow

This framework reveals a coherent family of diversity

l of the



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





& Cobbold (2012) derive a general expression for similarity numbers, is based on Rényi's generalised entropy and expands this framework reveals a coherent family of diversity measures. This framework reveals a coherent family of diversity measures of the *average ordinariness* of a population,

We propose a new method..

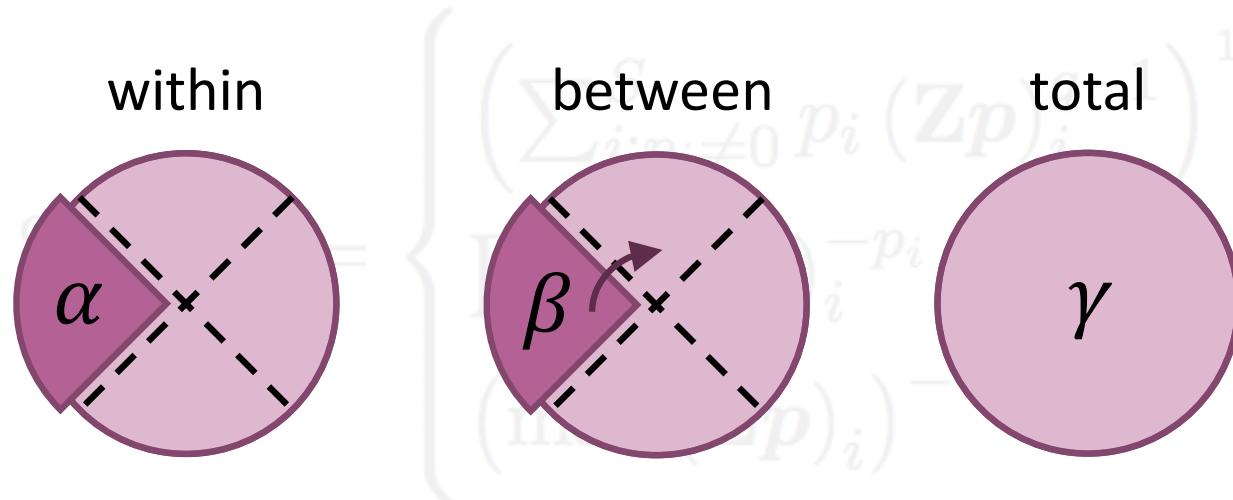
$$D^Z(p) = \begin{cases} \left(\sum_{i=1}^S p_i (\mathbf{Z}p)_i^{q-1} \right)^{1/(1-q)} & \text{if } q \neq 1 \\ \prod_{i=1}^S (\mathbf{Z}p)_i^{-p_i} & \text{if } q = 1 \\ \left(\max_i (\mathbf{Z}p)_i \right)^{-1} & \text{if } q = \infty \end{cases}$$

The order of diversity, or *viewpoint parameter*, q , determines the weight given to a particular abundance for $0 \leq q \leq \infty$. Varying q allows greater emphasis to be placed on different aspects of the distribution.



Framework for

Partitioning Diversity

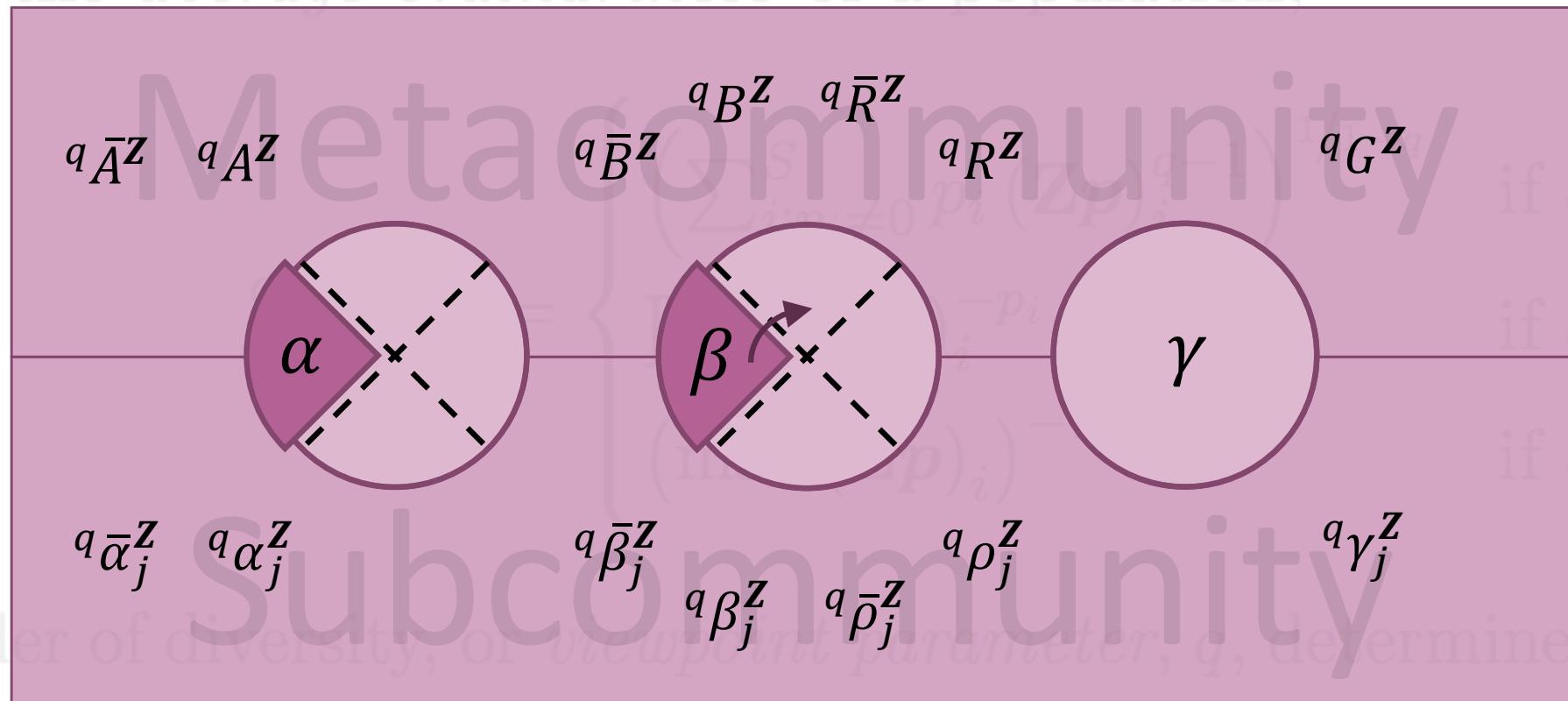


order of diversity, or *viewpoint parameter*, q , determines th
a particular abundance for $0 \leq q \leq \infty$. Varying q allows gr



University
of Glasgow

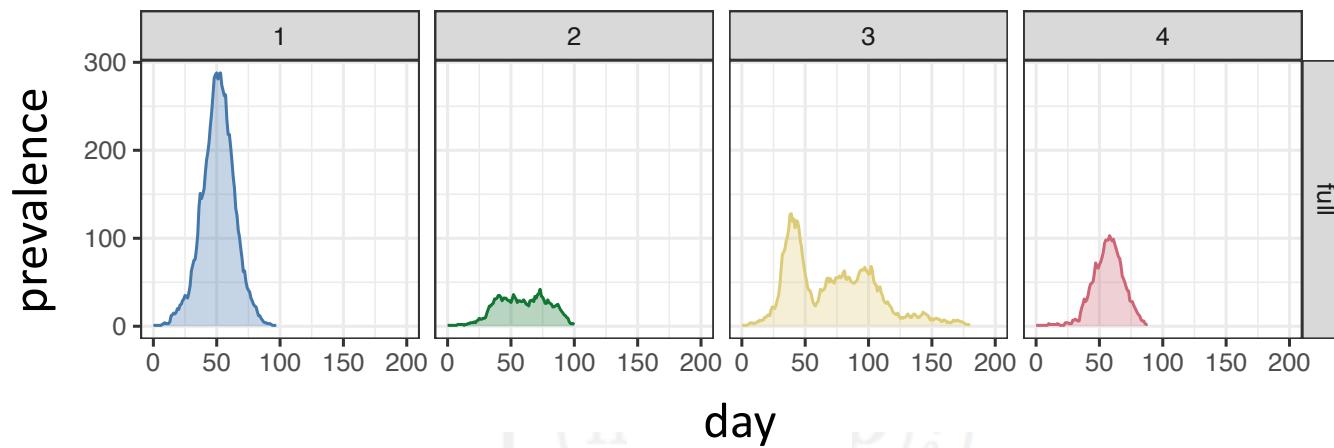
Framework for Partitioning Diversity





Methods

- This framework reveals a coherent family of diversity measures based on the *average ordinariness* of a population,
- Simulate outbreaks

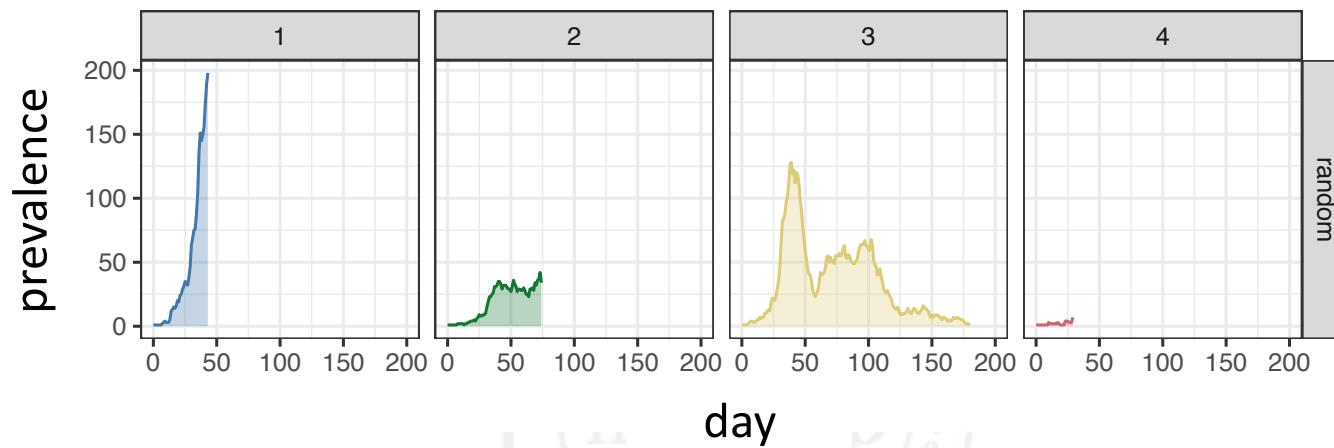


The order of diversity, or *viewpoint parameter*, q , determines the particular abundance for $0 \leq q \leq \infty$. Varying q allows greater



Methods

- This framework reveals a coherent family of diversity measures based on the *average ordinariness* of a population,
- Simulate outbreaks (truncated)

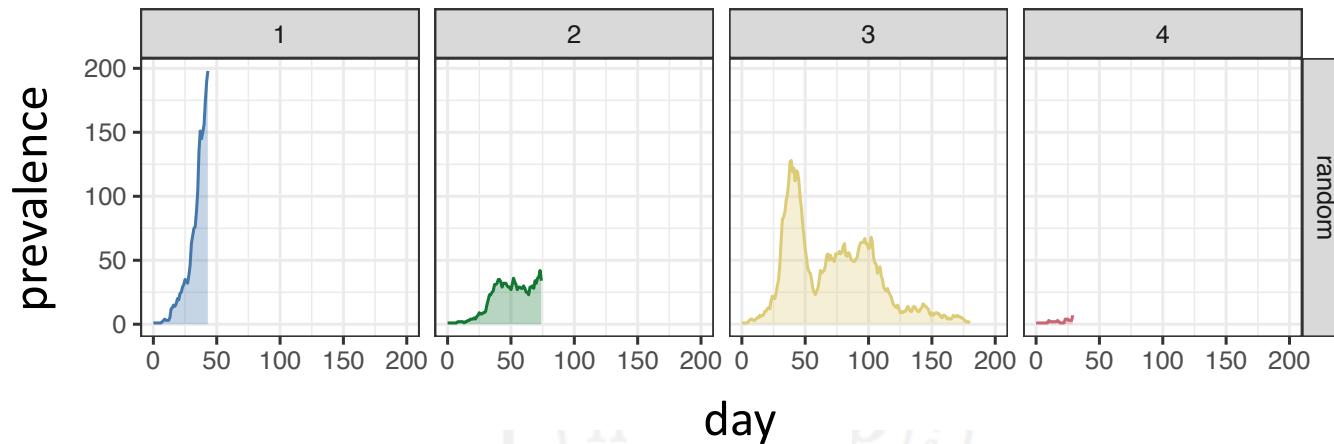


The order of diversity, or *viewpoint parameter*, q , determines the particular abundance for $0 \leq q \leq \infty$. Varying q allows generalisations from the extreme cases of the classical measures.



Methods

- Simulate outbreaks (truncated)

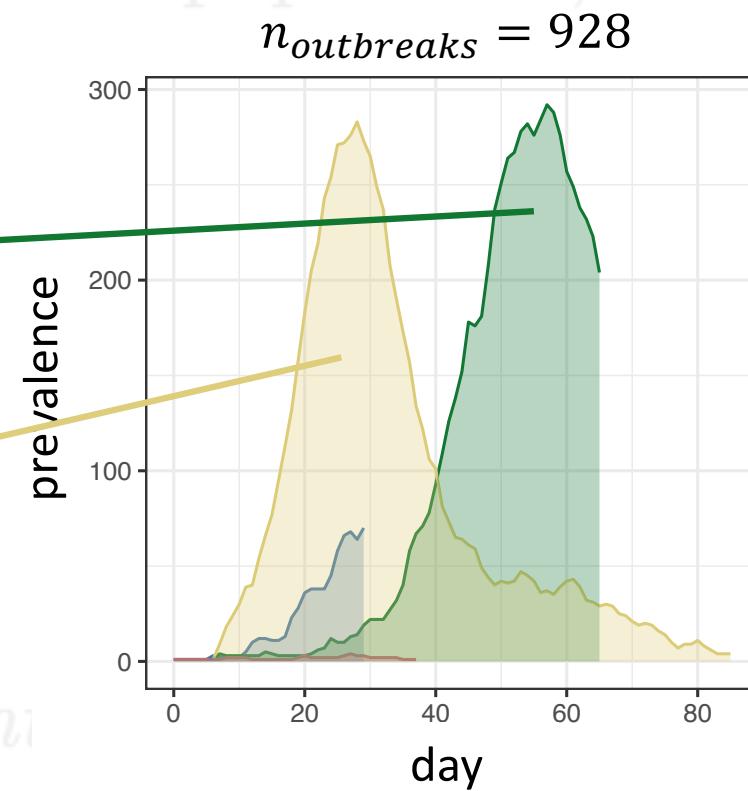
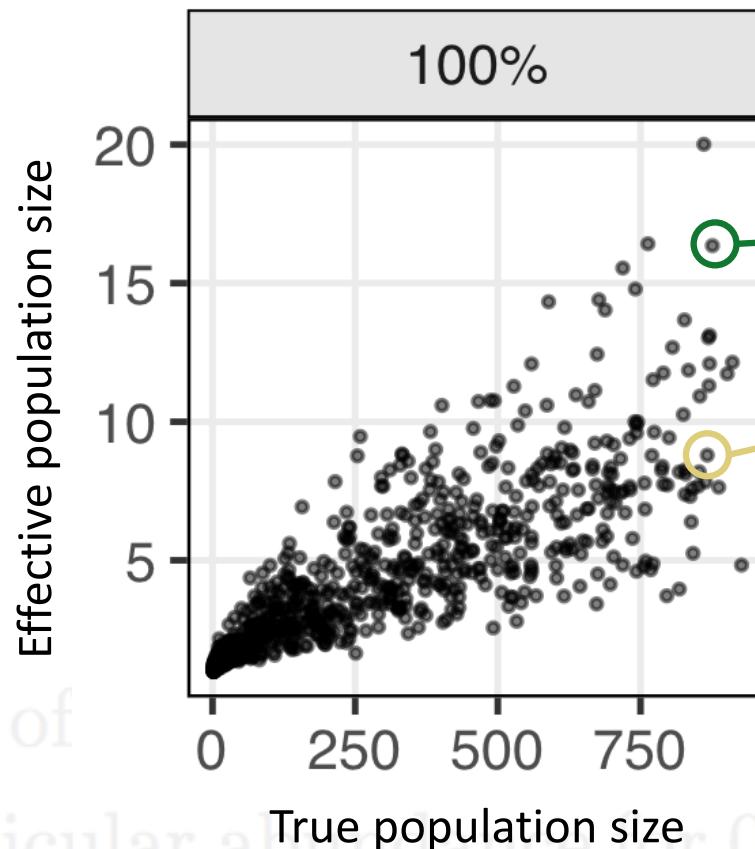


- Viral genetic diversity of the 2001 FMDV outbreak
- Calculate ${}^qG^Z$ (Effective number of sequences / infected premises)



University
of Glasgow

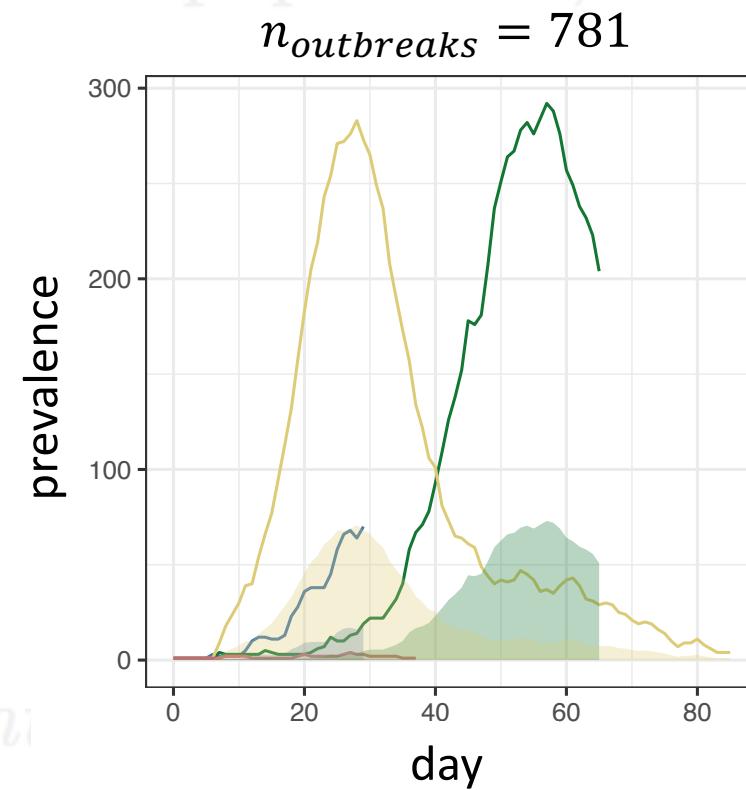
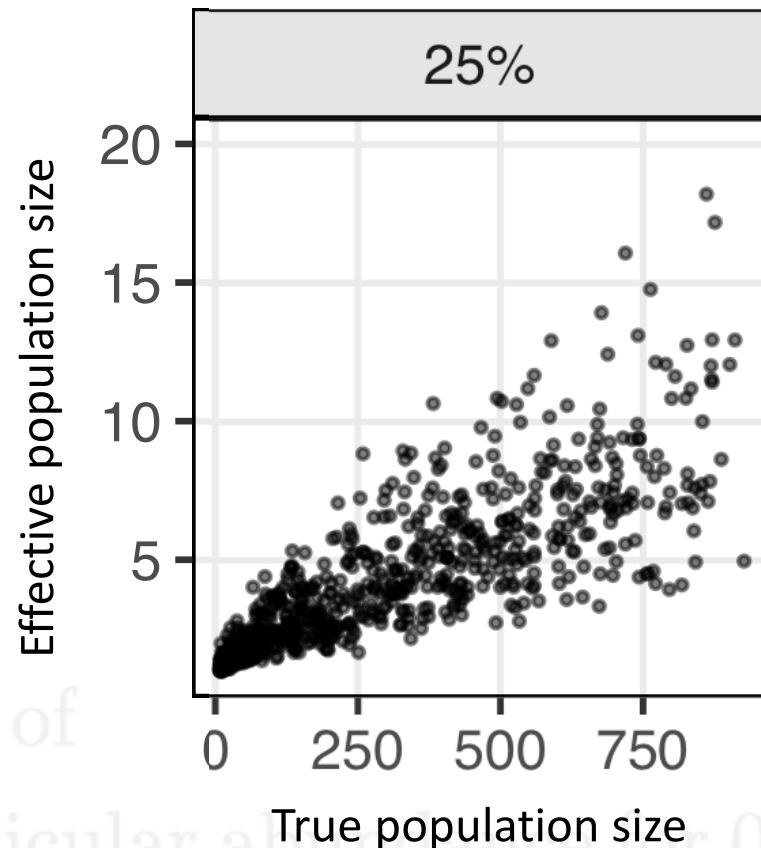
Effective number of sequences, ${}^1G^z$





University
of Glasgow

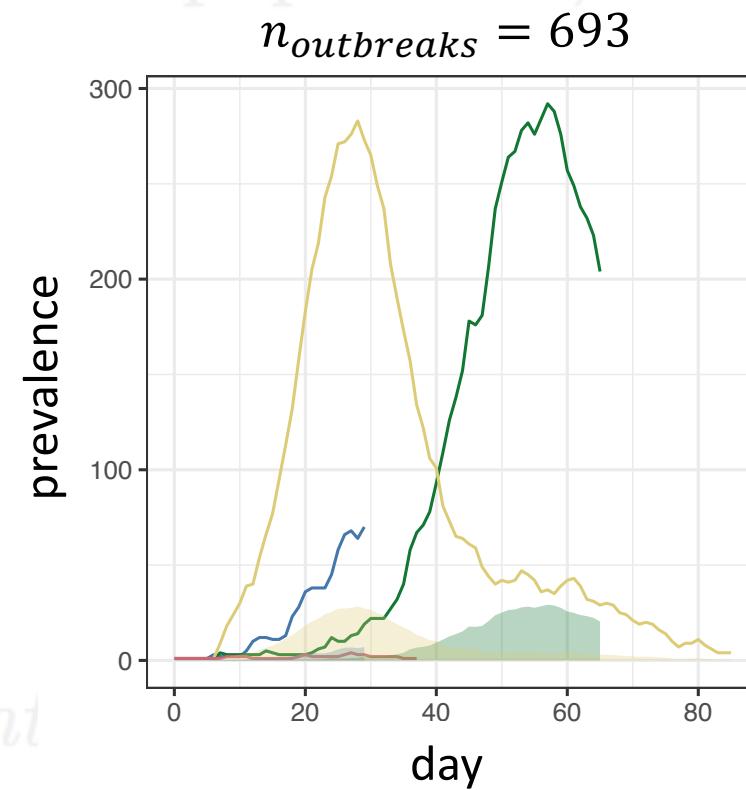
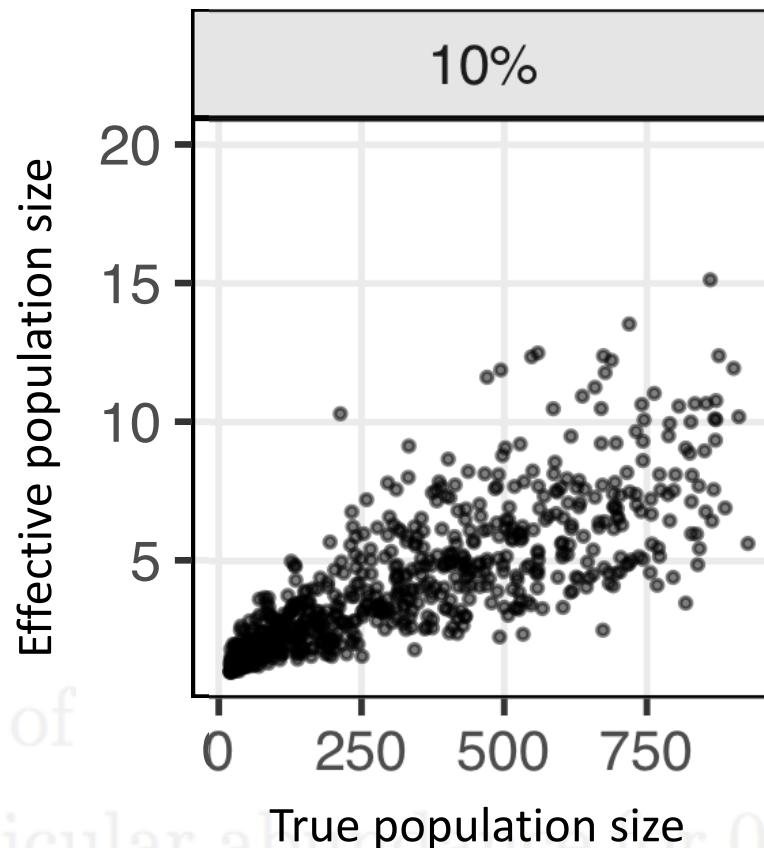
Effective number of sequences, ${}^1G^Z$





University
of Glasgow

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





Next, we use ${}^qG^Z$ with other simple summary statistics..

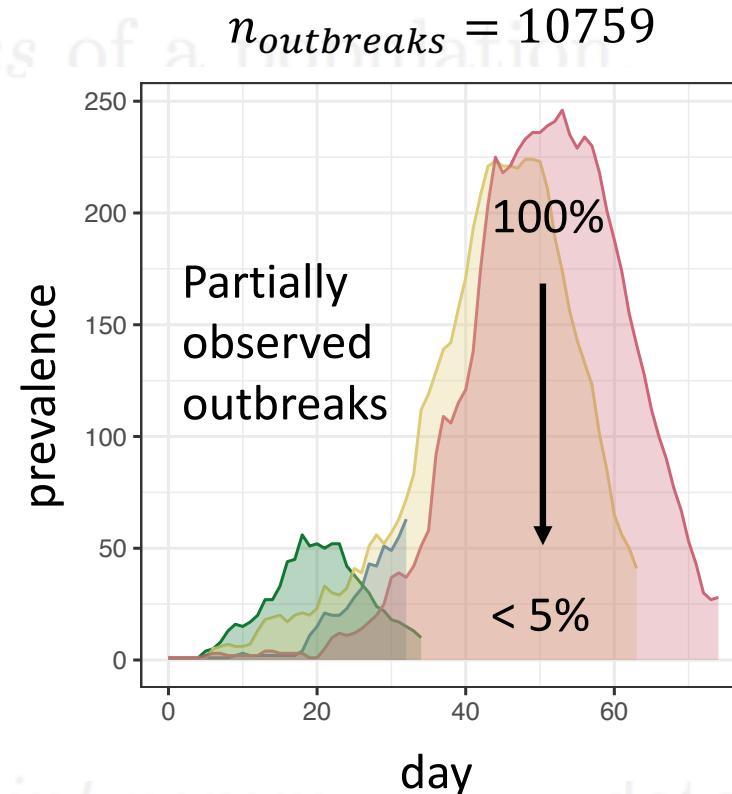
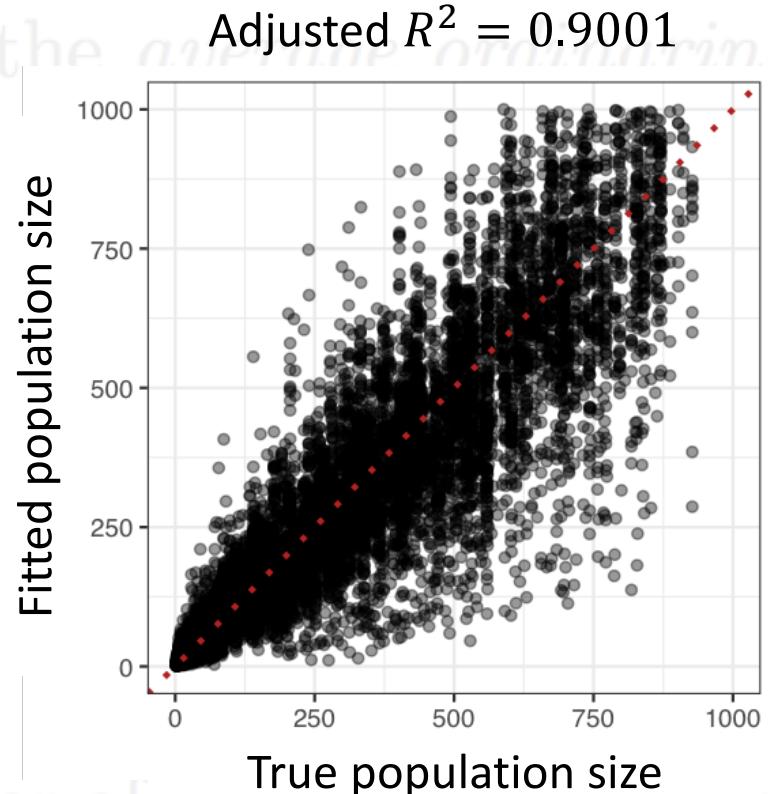
..to fit a linear model.



Model fit

This framework reveals a coherent family of diversity

l of the antimicrobial resistance of a illuminatin

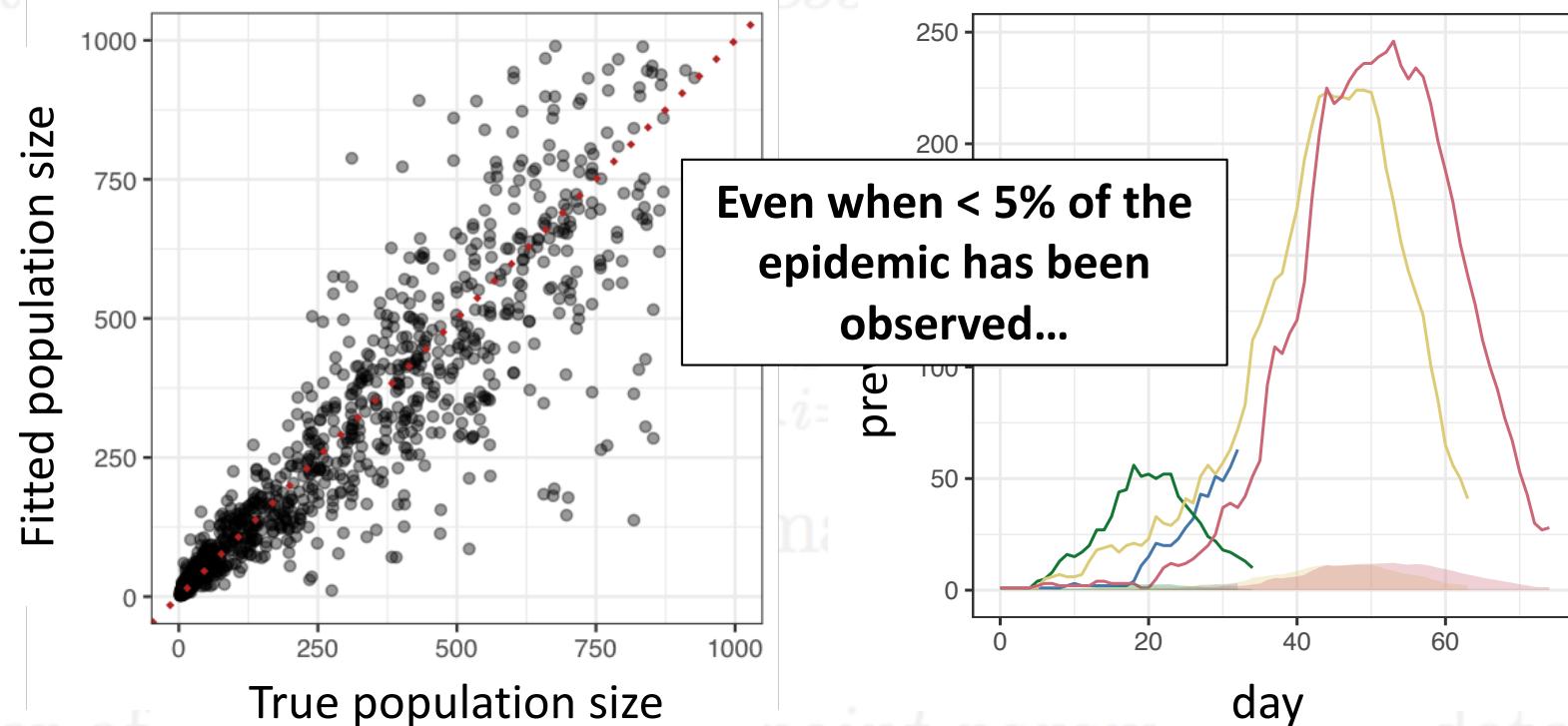




Model fit

This framework reveals a coherent family of diversity

l of the *anomalous ordinarity* of a random m





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



- 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. *Phylogenetic modelling of foot-and-mouth disease virus sequence data.* [PhD Thesis]



Richard Reeve
Antonello Di Nardo

Christina Cobbold
Louise Matthews