Strain-space model for Sars-CoV-2

Peter C. Jentsch, PhD [1,4] Finlay Maguire, PhD [3,5]
Samira Mubareka, MD, FRCPC [1,2]

[1]Sunnybrook Research Institute, Toronto, Canada

[2]University of Toronto, Toronto, Canada

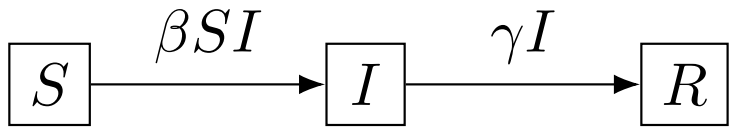
[3]Dalhousie University, Halifax, Canada

[4]Simon Fraser University, Burnaby, Canada

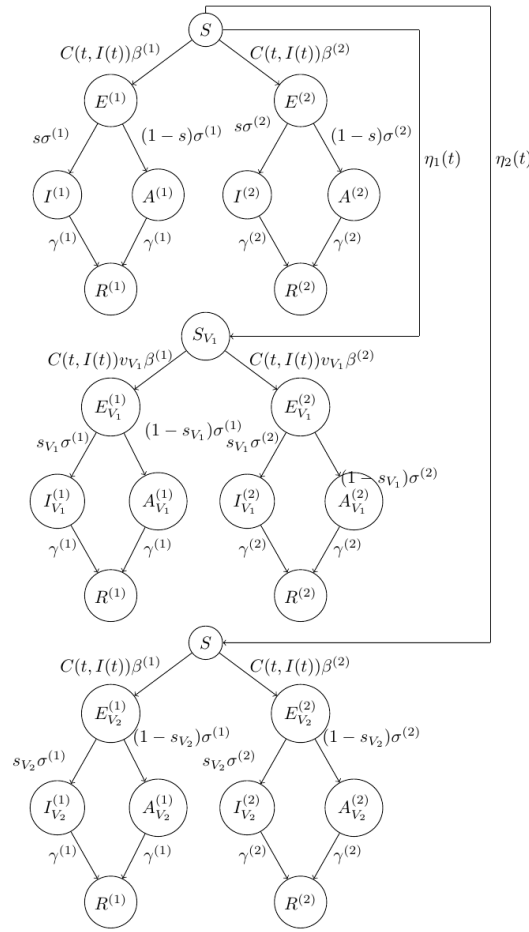
[5]Shared Hospital Laboratory, Toronto, Canada

June 29, 2022

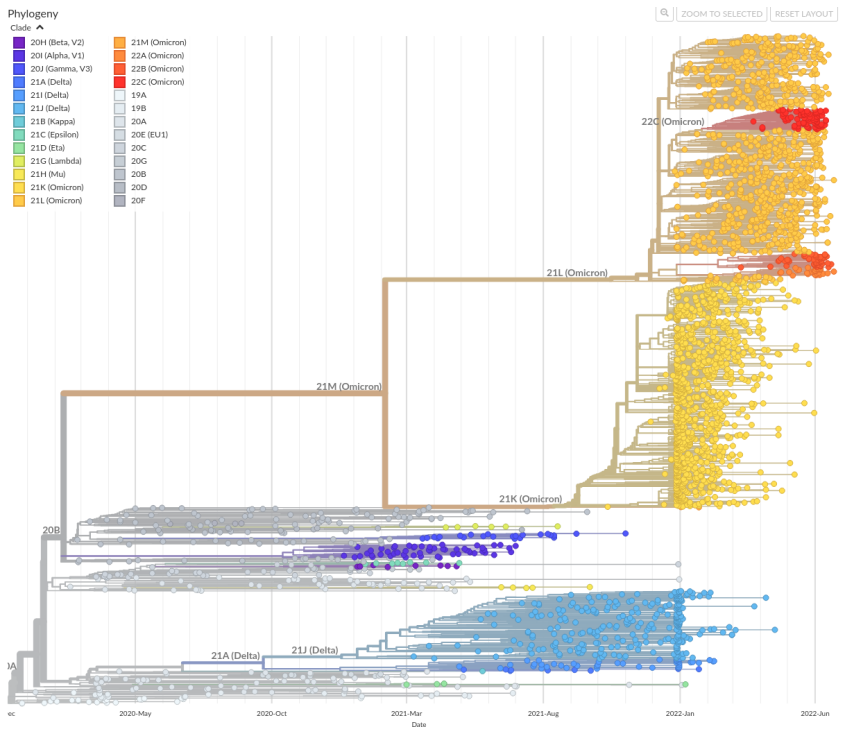
- Infection spread is often modelled using compartmental models
- Represent subsets of a host population and rates of movement between them



- Multiple infections (e.g. competing VoCs) can be represented as more compartments
- Work on multiple infections is usually here due to lack of data, increasing complexity

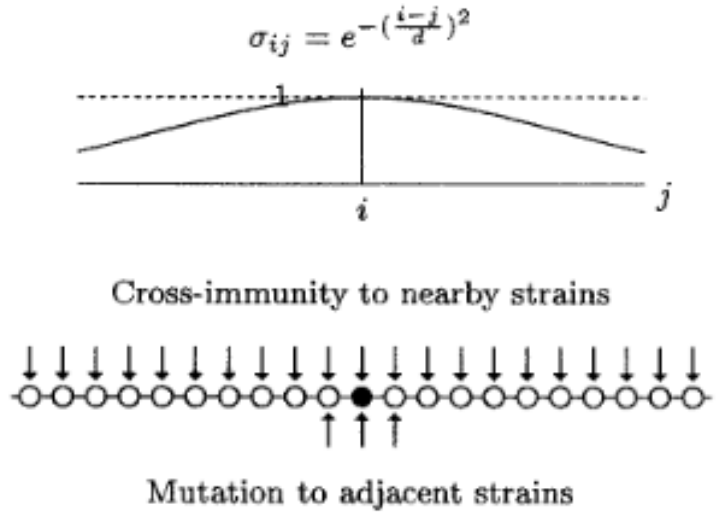


This only represents a tiny amount of the genomic data we have for Sars-CoV-2!

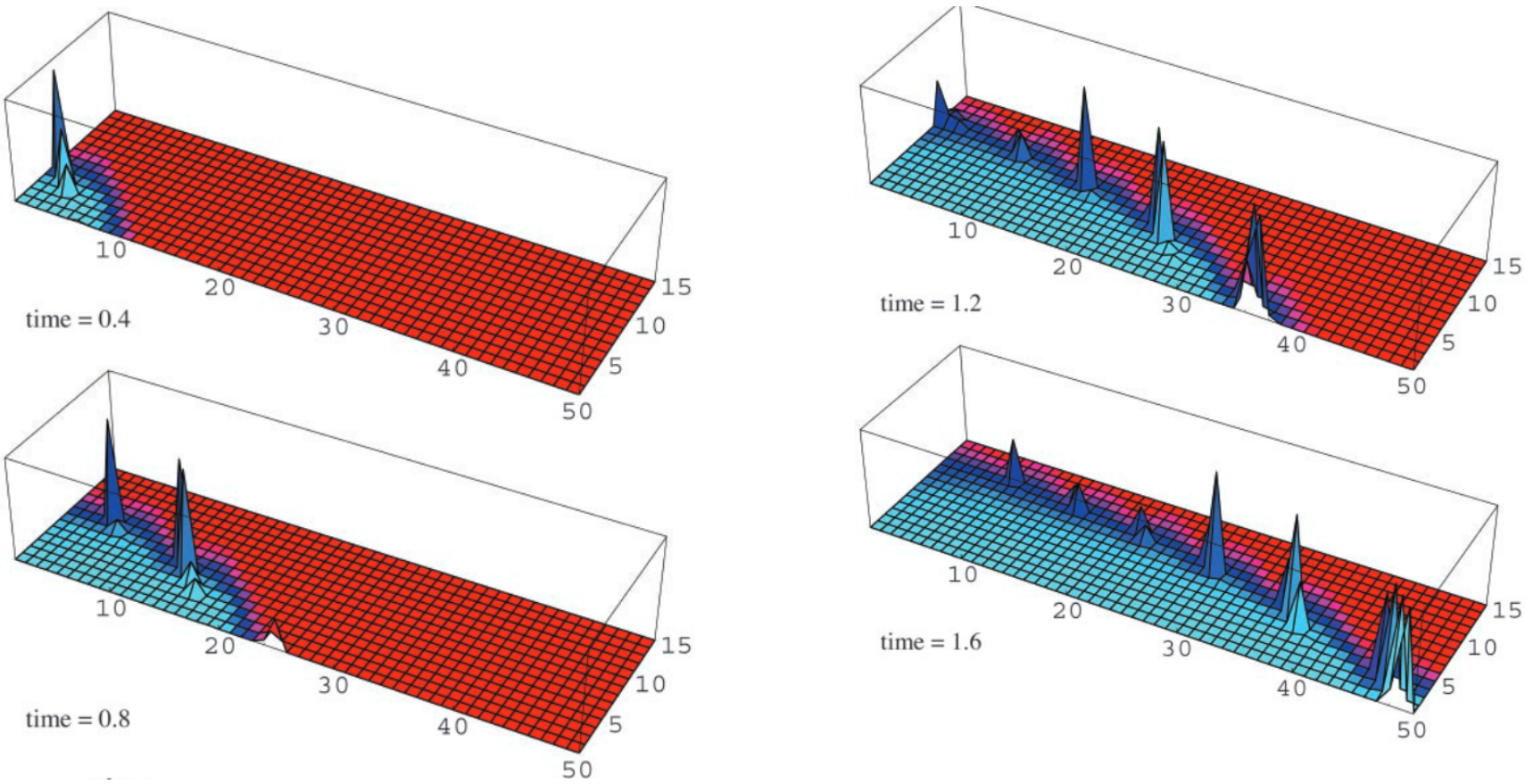


- Can extend these models to a sequence of variants
- Assume each variant is indexed by i
- The dynamics at each variant i are determined by a simple compartmental model

- Variants are related by a function $\sigma(i, j)$ that determines how much an infection by variant i reduces probability of infection to variant j.
- A variant i mutates to neighbouring indices $i + 1, i - 1$ proportional to the population of variant i



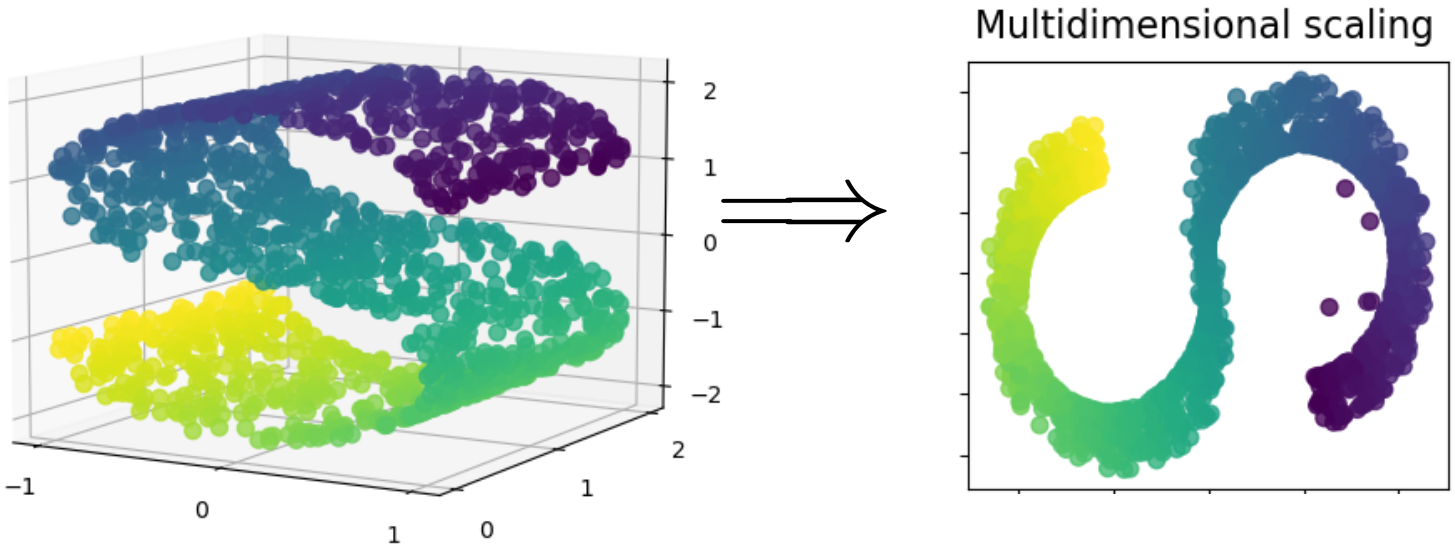
[Gog and Grenfell, 2002]



[Gog and Grenfell, 2002]

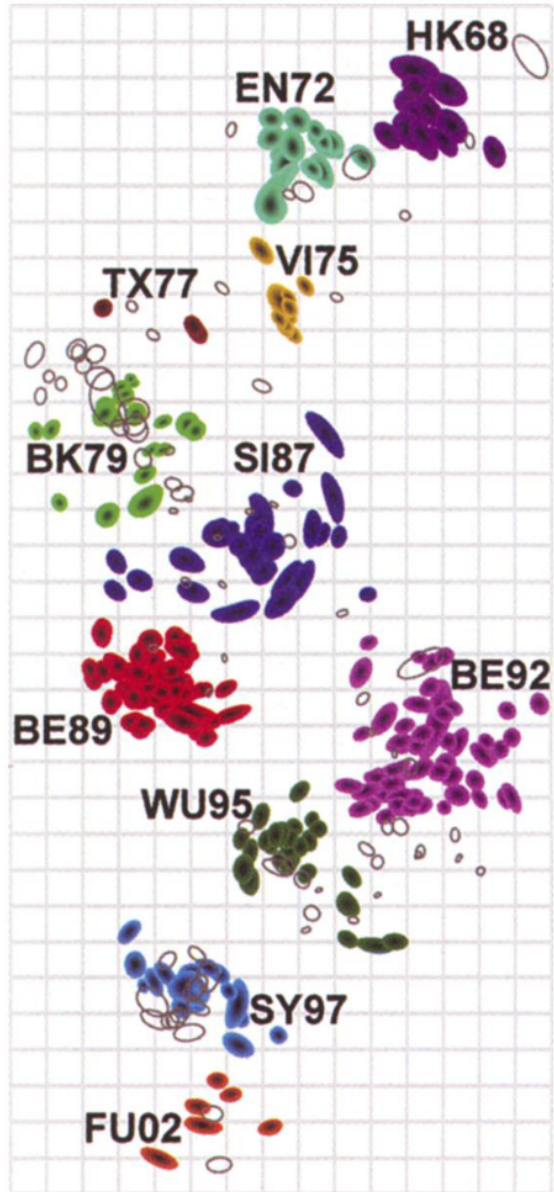
Antigenic cartography

- Practice of mapping out immune responses to related pathogens
- Distance between serums and pathogen is quantified, these points are visualized using multidimensional scaling

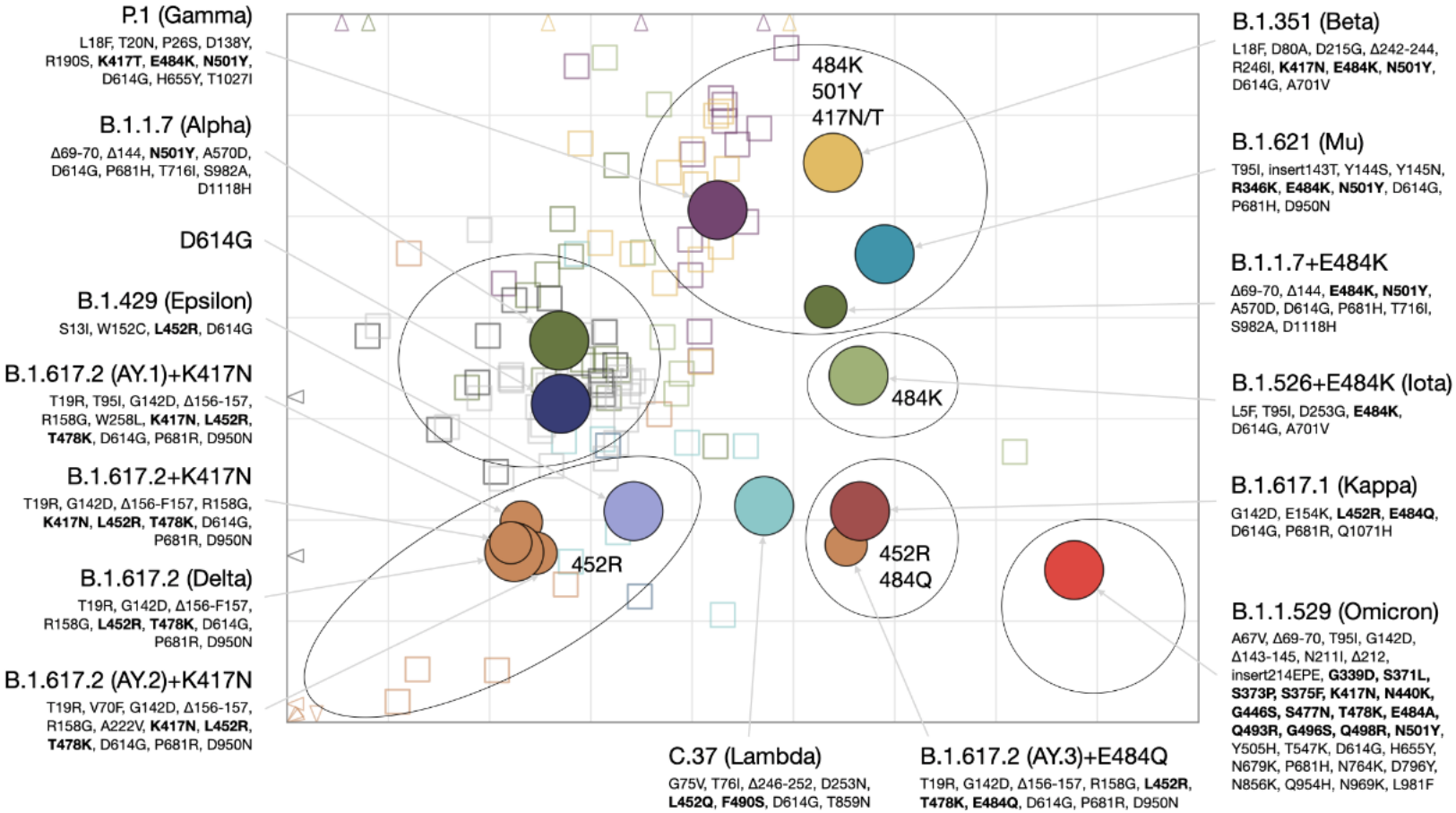


[Pedregosa et al., 2011]

This technique was developed to visualize the antigenic drift of H3N2



[Lapedes and Farber, 2001, Smith et al., 2004]



[Wilks et al., 2022]

These results suggest that 2 dimensions might be an adequate approximation to the full space!

Model Equations

$$\frac{S_{ij}}{dt} = -\sum_{kl} \beta_{kl}\sigma_{ijkl}S_{ij}I_{kl} + \gamma R_{ij} \tag{1}$$

$$\frac{I_{ij}(t)}{dt} = \beta_{ij}S_{ij}I_{ij} - \xi I_{ij} + M\left(-4I_{ij} + I_{i-1,j} + I_{i+1,j} + I_{i,j-1} + I_{i,j+1}\right) \tag{2}$$

$$\frac{R_{ij}(t)}{dt} = \xi I_{ij} - \gamma R_{ij} \tag{3}$$

Boundary conditions: $I_{0,j} = 0, I_{j,0} = 0, I_{N,j} = 0, I_{j,N} = 0$
Initial conditions computed from genomic data in GISAID

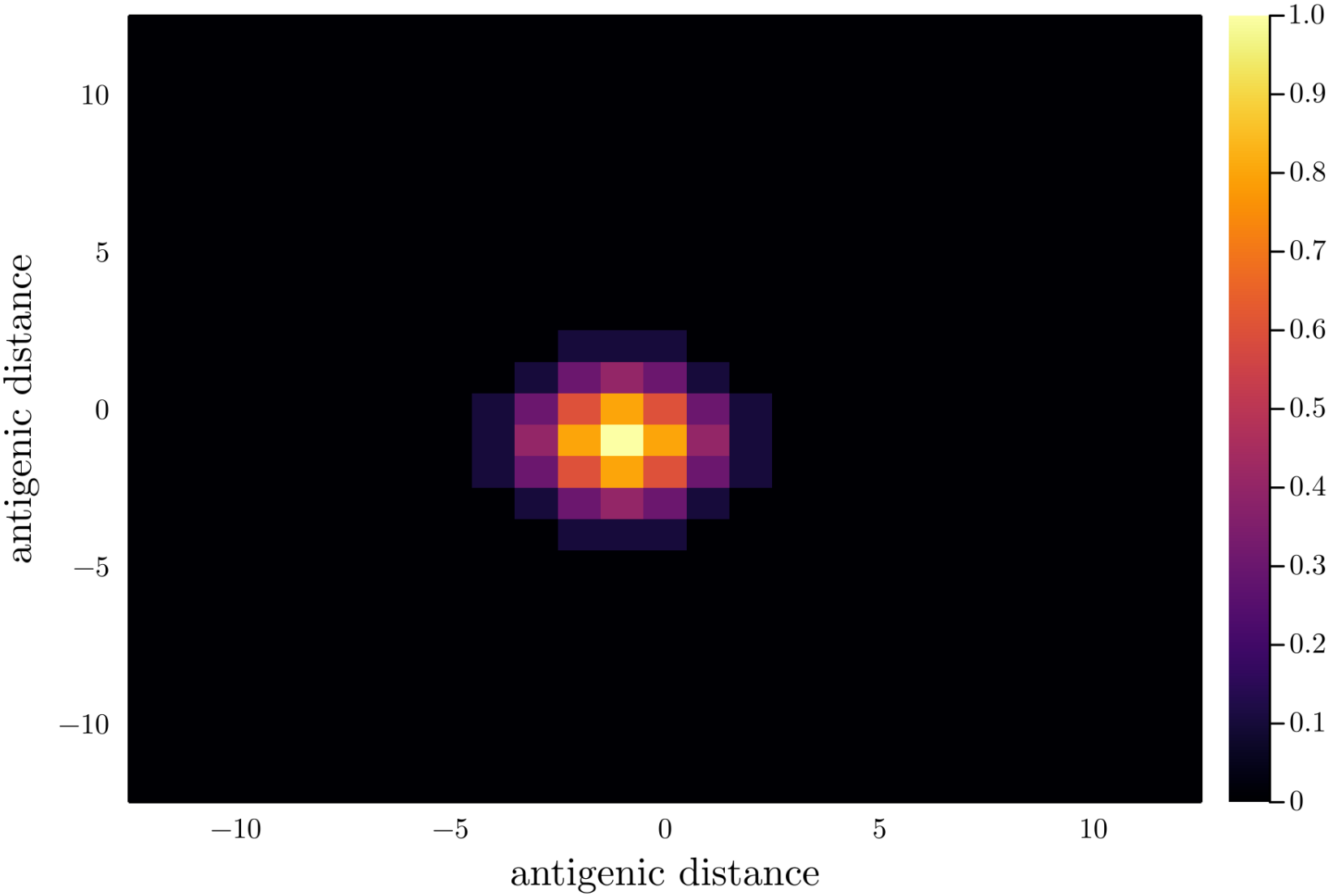
Model parameters/variables

Symbol	Description
N	Size of variant grid
S_{ij}	Population susceptible to variant $(i, j) \in [0, N]^2$
I_{ij}	Population infected by variant $(i, j) \in [0, N]^2$
R_{ij}	Recovered/Immune to variant $(i, j) \in [0, N]^2$
σ_{ijkl}	Probability that exposure to variant (i, j) causes immunity to variant (k, l)
β_{ij}	Transmission rate of variant (i, j)
ξ	Recovery rate of all strains
γ	Rate of immunity loss of all strains

Table of symbols for Model 2

σ matrix

In practice, we assume σ_{ijkl} is just a 2-D gaussian distribution parameterized by the distance between (i, j) and (k, l).



To incorporate more realistic mutation rates, we can go to continuous strain-space and use nonlocal reaction-diffusion dynamics as in [Rouzine and Rozhnova, 2018, Bessonov et al., 2021]

$$S_t(x,y,t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \beta(x',y')\sigma(x,y,x',y')S(x,y,t)I(x',y',t)dx'dy' + \gamma R_{ij} - \eta(t)v(x,y)S(x,y,t) \tag{4}$$

$$I_t(x,y,t) = \beta(x,y)S(x,y,t)I(x,y,t) - \xi I(x,y,t) + M\left(I_x(x,y,t) + I_y(x,y,t)\right) \tag{5}$$

$$R_t(x,y,t) = \xi I(x,y,t)I(x,y,t) - \gamma R(x,y,t) + \eta(t)v(x,y)S(x,y,t) \tag{6}$$

where β, σ, v have been generalized to their continuous counterparts. Given a dispersion kernel $K(x,y) \in L_2 : \mathbb{R}^2 \to \mathbb{R}$ this can be generalised to non-local diffusion as follows

$$I_t(x,y,t) = \beta(x,y)S(x,y,t)I(x,y,t) - \xi I(x,y,t) + M\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K(x-x',y-y')I(x',y',t)dx'dy'\right) \tag{7}$$

Developing an antigenic distance map

- We would like an approximate measure of antigenic distance for every sample genome
- Using all samples, we compute pairwise distances between each unique genome in some way that encodes antigenic response
- Many possible ways to do this, so far none of them seem to work very well
- Project to 2-d (hopefully) space with multidimensional scaling

Genome distance

Assume:

- a, b are SARS-CoV-2 genomes aligned with the reference
- a_i the ith nucleotide base in a and

$$\chi(a_i, b_i) = \begin{cases} 1 & \text{if } a_i = b_i \\ 0 & \text{otherwise} \end{cases}$$

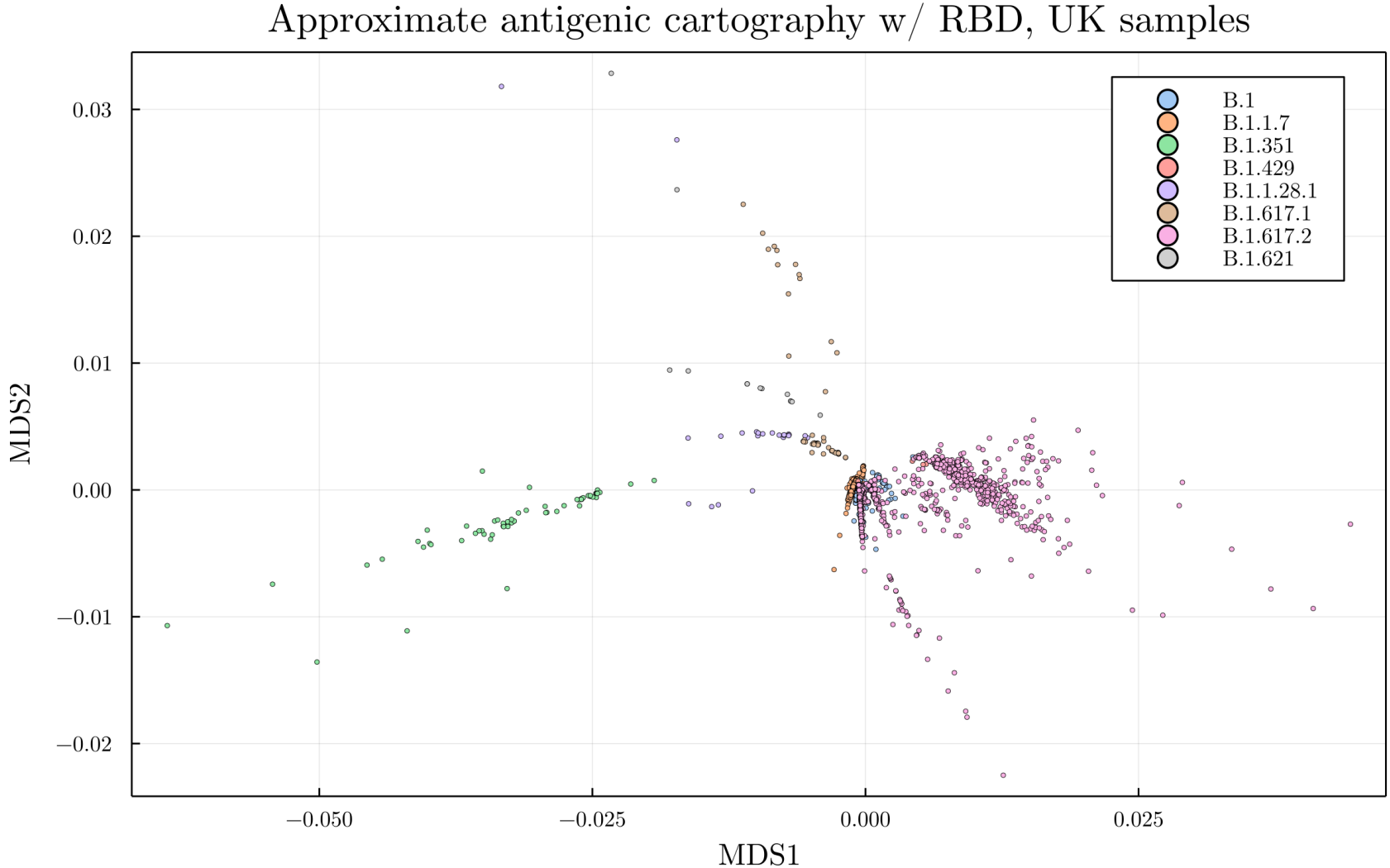
- h_i is a vector containing the number of homoplasic mutations at site i in the global tree
- $\mathfrak{B}(a)$ computes the polyclonal binding affinity of genome a as per [Starr et al., 2020]

One option for a distance measure is something like

$$d(a, b) = \frac{\mathfrak{B}(a) + \mathfrak{B}(b)}{2} + \sum_i \chi(a_i, b_i) h_i \tag{8}$$

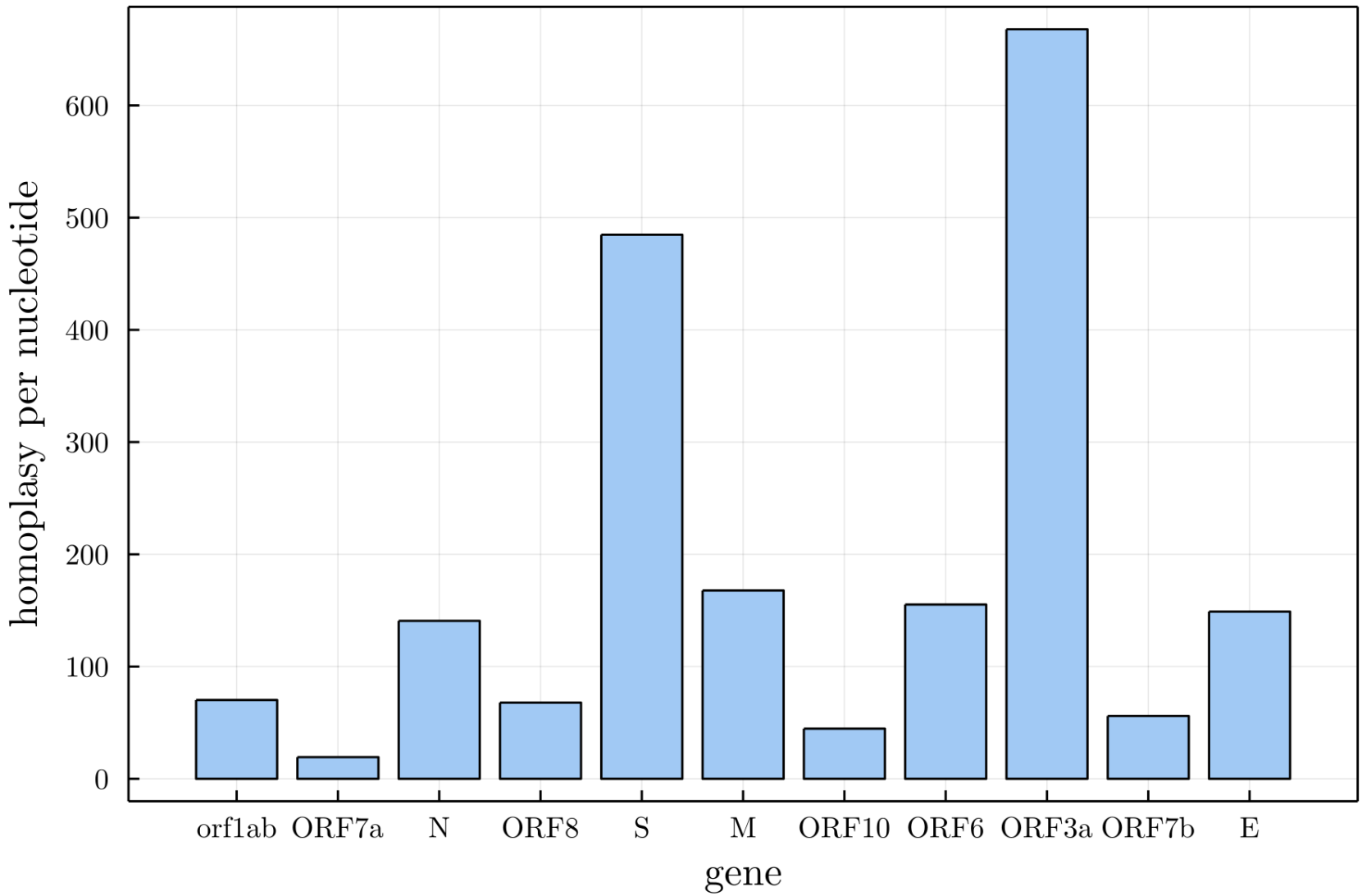
That is, the average binding between two genomes plus the SNP distance weighted by the relative homoplasy of each mutation.

Example antigenic distance map



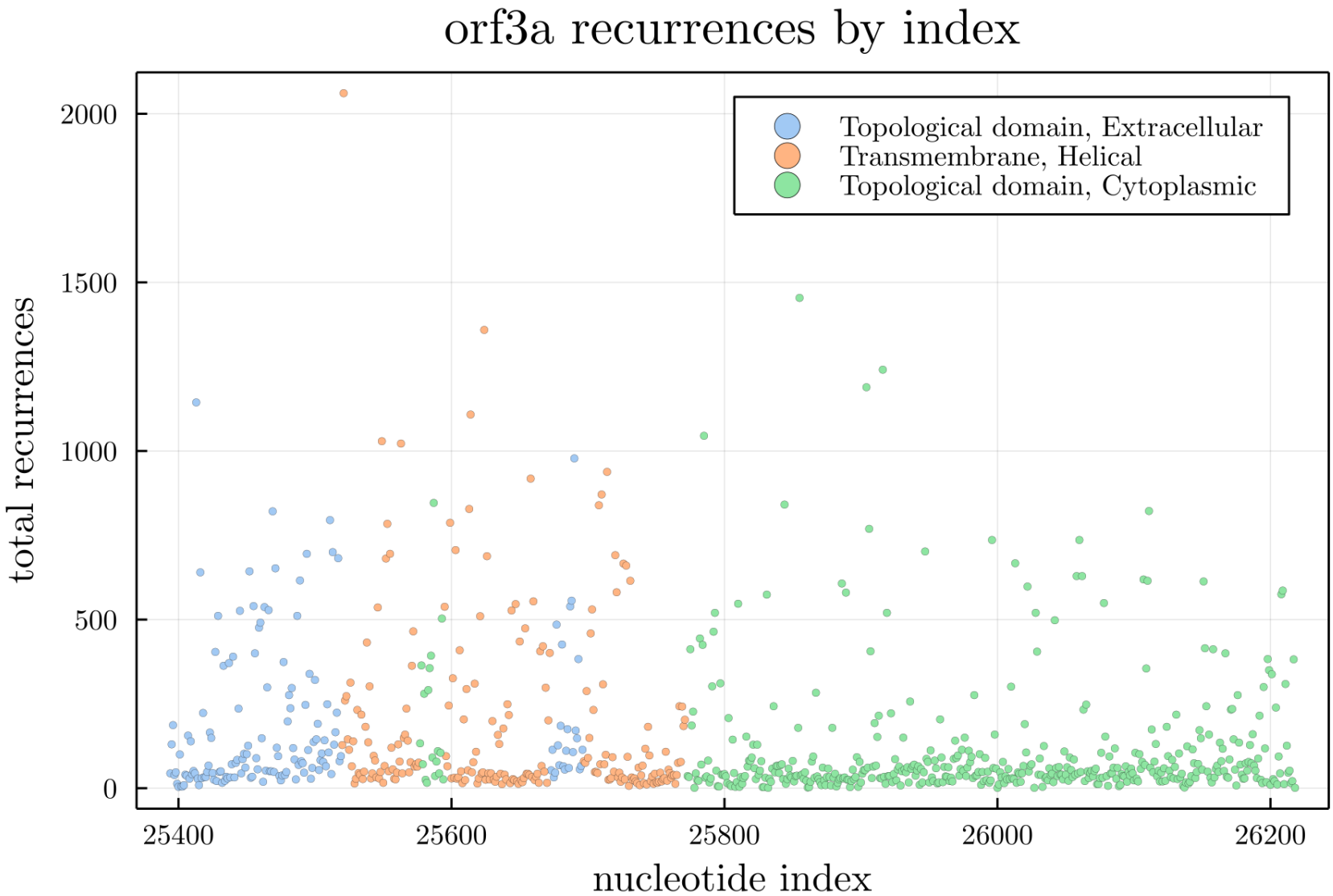
Multidimensional scaling plot using samples from the UK up to mid November

Homoplasy in global tree



Number of recurrent (homoplasic) mutations per base by gene, (normalized by gene length)

Homoplasy in orf3a



Bessonov, N., Bocharov, G., Meyerhans, A., Popov, V., and Volpert, V. (2021).
Existence and dynamics of strains in a nonlocal reaction-diffusion model of viral evolution.
SIAM Journal on Applied Mathematics, 81(1):107–128.

Gog, J. R. and Grenfell, B. T. (2002).
Dynamics and selection of many-strain pathogens.
Proceedings of the National Academy of Sciences, 99(26):17209–17214.

Lapedes, A. and Farber, R. (2001).
The Geometry of Shape Space: Application to Influenza.
Journal of Theoretical Biology, 212(1):57–69.


Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011).
Scikit-learn: Machine learning in Python.
Journal of Machine Learning Research, 12:2825–2830.


Rouzine, I. M. and Rozhnova, G. (2018).
Antigenic evolution of viruses in host populations.
PLoS Pathogens, 14(9):e1007291.


Smith, D. J., Lapedes, A. S., de Jong, J. C., Bestebroer, T. M., Rimmelzwaan, G. F., Osterhaus, A. D. M. E., and Fouchier, R. A. M. (2004).
Mapping the Antigenic and Genetic Evolution of Influenza Virus.
Science, 305(5682):371–376.

Starr, T. N., Greaney, A. J., Hilton, S. K., Ellis, D., Crawford, K. H., Dingens, A. S., Navarro, M. J., Bowen, J. E., Tortorici, M. A., Walls, A. C., et al. (2020). Deep mutational scanning of sars-cov-2 receptor binding domain reveals constraints on folding and ace2 binding. *Cell*, 182(5):1295–1310.

Wilks, S. H., Mühlemann, B., Shen, X., Türeli, S., LeGresley, E. B., Netzl, A., Caniza, M. A., Chacaltana-Huarcaya, J. N., Daniell, X., Datto, M. B., Denny, T. N., Drosten, C., Fouchier, R. A. M., Garcia, P. J., Halfmann, P. J., Jassem, A., Jones, T. C., Kawaoka, Y., Krammer, F., McDanal, C., Pajon, R., Simon, V., Stockwell, M., Tang, H., van Bakel, H., Webby, R., Montefiori, D. C., and Smith, D. J. (2022). Mapping SARS-CoV-2 antigenic relationships and serological responses. Preprint, Immunology.