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 13, 2022

- Dynamic model of Sars-CoV-2 evolution, representing antigenic diversity on a lattice (as in e.g. [**?**, **?**, **?**])
- Antigenically distinct variants of the virus are mapped to 2D grid, distance between variants corresponds to the proportional reduction in maximum serum viral titre [**?**, **?**]

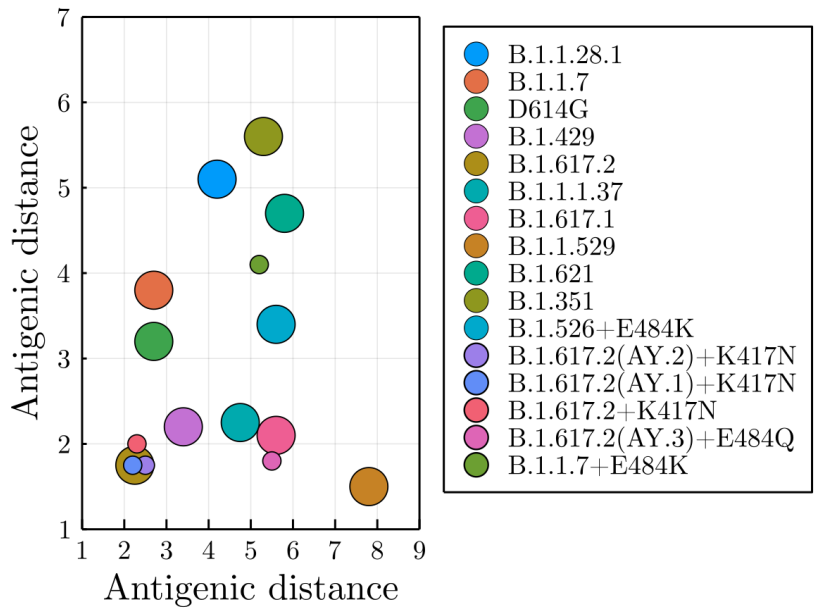


Figure: Antigenic cartography of Sars-CoV-2, reproduced from [**?**], Fig. 2

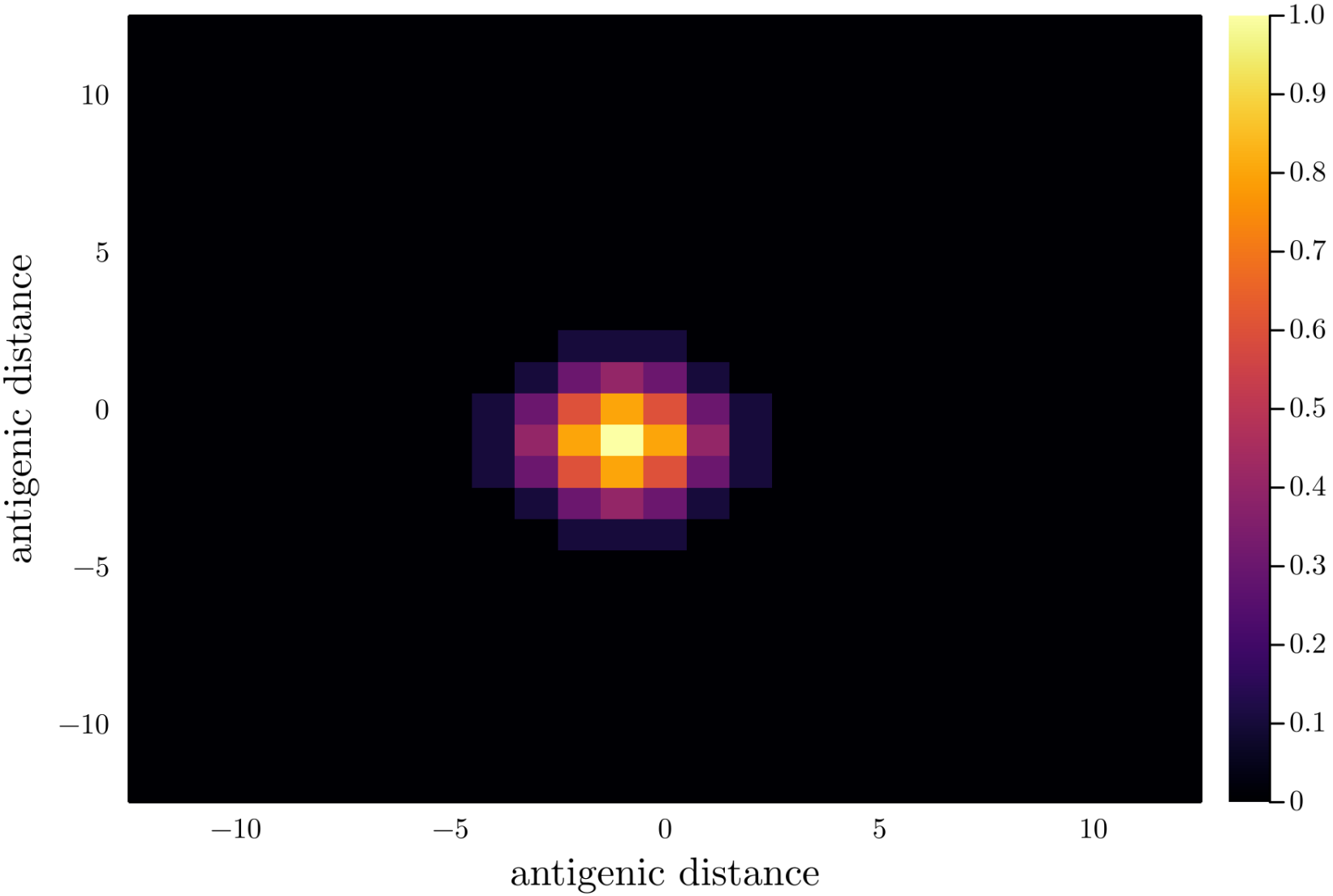
Model parameters/variables

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: 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).



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

To incorporate more realistic mutation rates, we can go to continuous strain-space and use nonlocal reaction-diffusion dynamics as in [**?**, **?**]

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

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

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

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

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, none of them seem to work very well
- Project to 2-d (hopefully) space with multidimensional scaling

Homoplasy in global tree

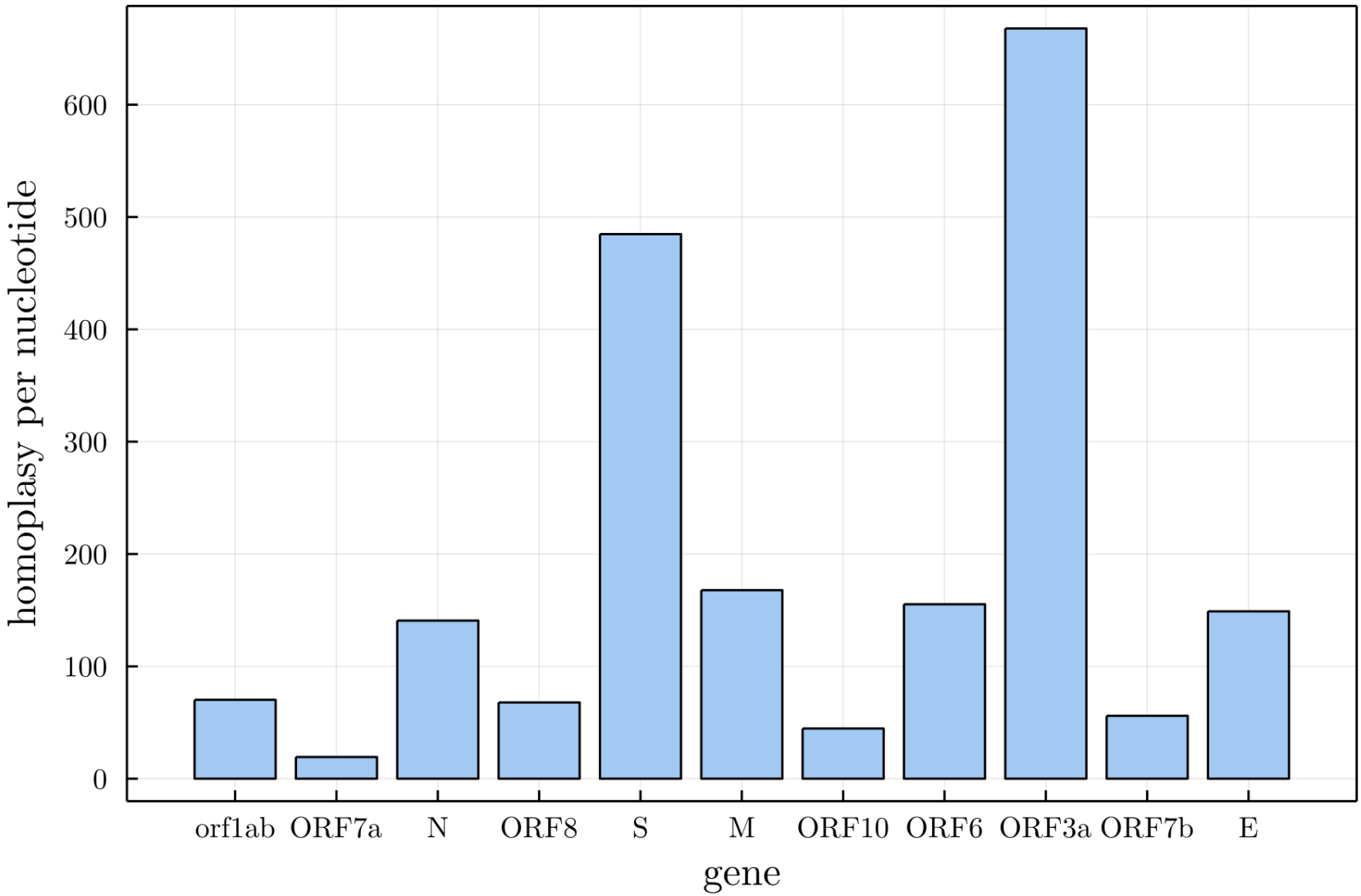
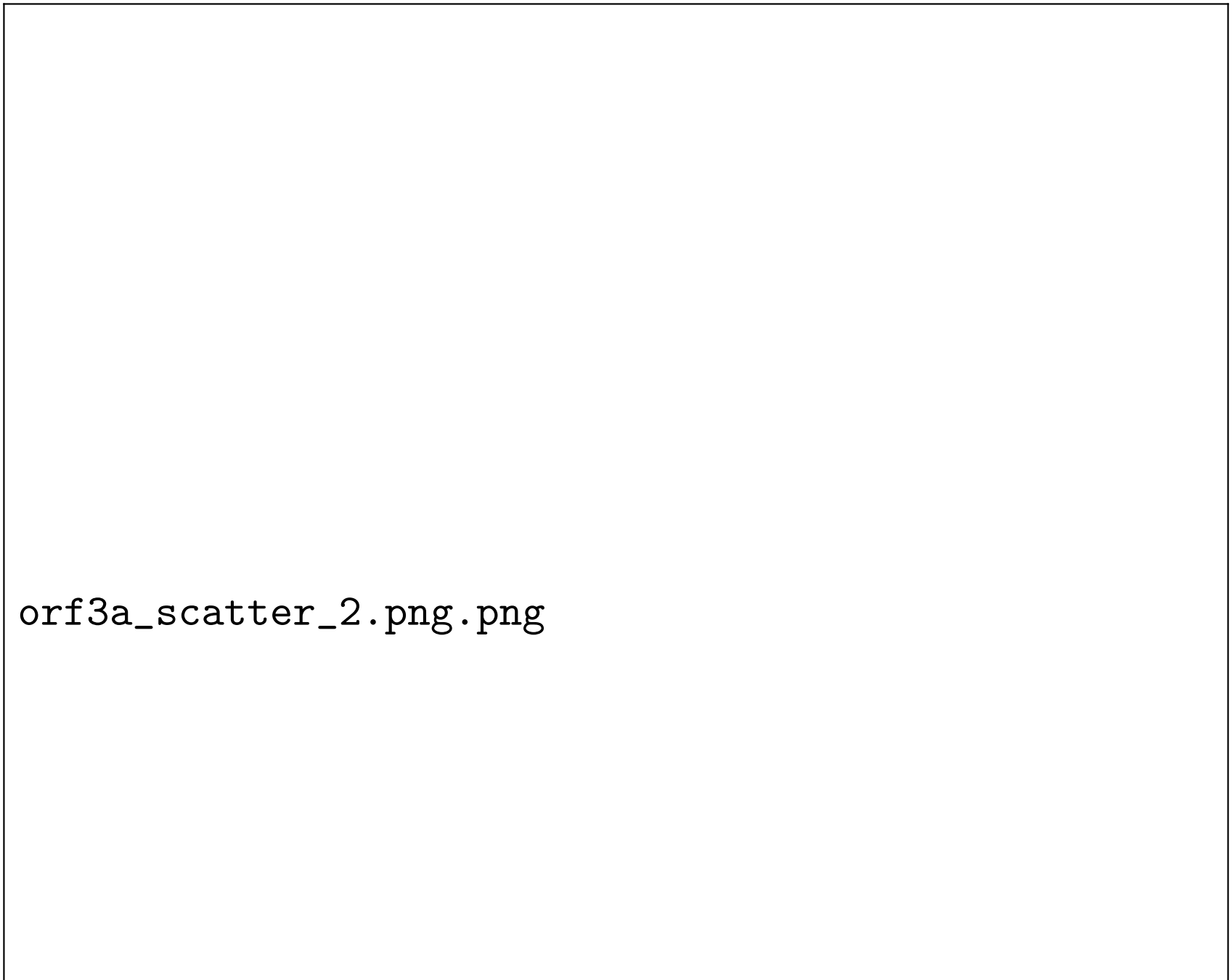


Figure: Number of recurrent (homoplasic) mutations per base by gene, (normalized by gene length)

Homoplasy in orf3a



Genome distance

Assume:

- a, b are SARS-CoV-2 genomes aligned with the reference
- a_i the ith nucleotide base in a
- $$\chi(a_i, b_i) = \begin{cases} 1 & \text{if } a_i = b_i \\ 0 & \text{otherwise} \end{cases}$$
- 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 [**?**]

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.