- Immune systems are the most complex of complex adaptive systems
- Viruses are (almost) pure replicating information
- Disease is an emergent phenomenon

Mapping the Antigenic and Genetic Evolution of Influenza Virus

Derek J. Smith,[1,2*†] Alan S. Lapedes,[3*] Jan C. de Jong,[2]
Theo M. Bestebroer,[2] Guus F. Rimmelzwaan,[2]
Albert D. M. E. Osterhaus,[2] Ron A. M. Fouchier[2*]

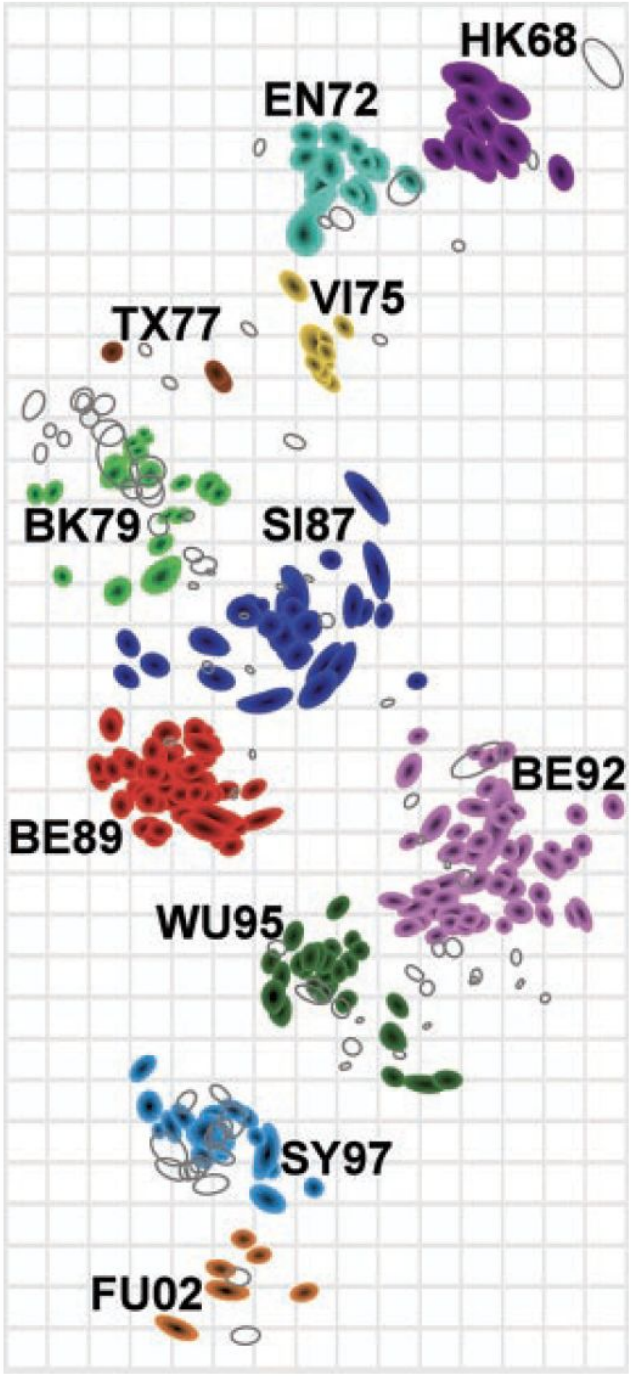
www.sciencemag.org SCIENCE VOL 305 16 JULY 2004

Main points

- Quantified and visualized antigenic evolution of influenza A (H3N2) virus 1968 to 2003.
  - Note (not H1N1 which caused both the 1918 and 2009 swine flu pandemic)
- Antigenic (phenotypic) and genetic evolution are similar but
  - Antigenic evolution was more punctuated than genetic evolution,
  - Genetic change sometimes had a disproportionately large antigenic effect
- This method allows
  - Estimating how effective vaccines will be against circulating strains
  - ? Predicting the relative success of emerging viral strains by
    - quantifying the combined effects of population level immune escape
    - and viral fitness on strain evolution.
- Derek Smith (UNM CS PhD 1997, WHO flu vaccine design task force)

Main points

- Flu names: location of emergence + year
- Colored shapes = flu
  - Open circles = antibodies that fight that flu
- Distance on the map represents “HI distance” (phenotypic distance)
- Each grid cell is 1 unit of antigenic (phenotypic distance)
  - 2 units = 4-fold distance
  - 3 units – 8 -fold distance

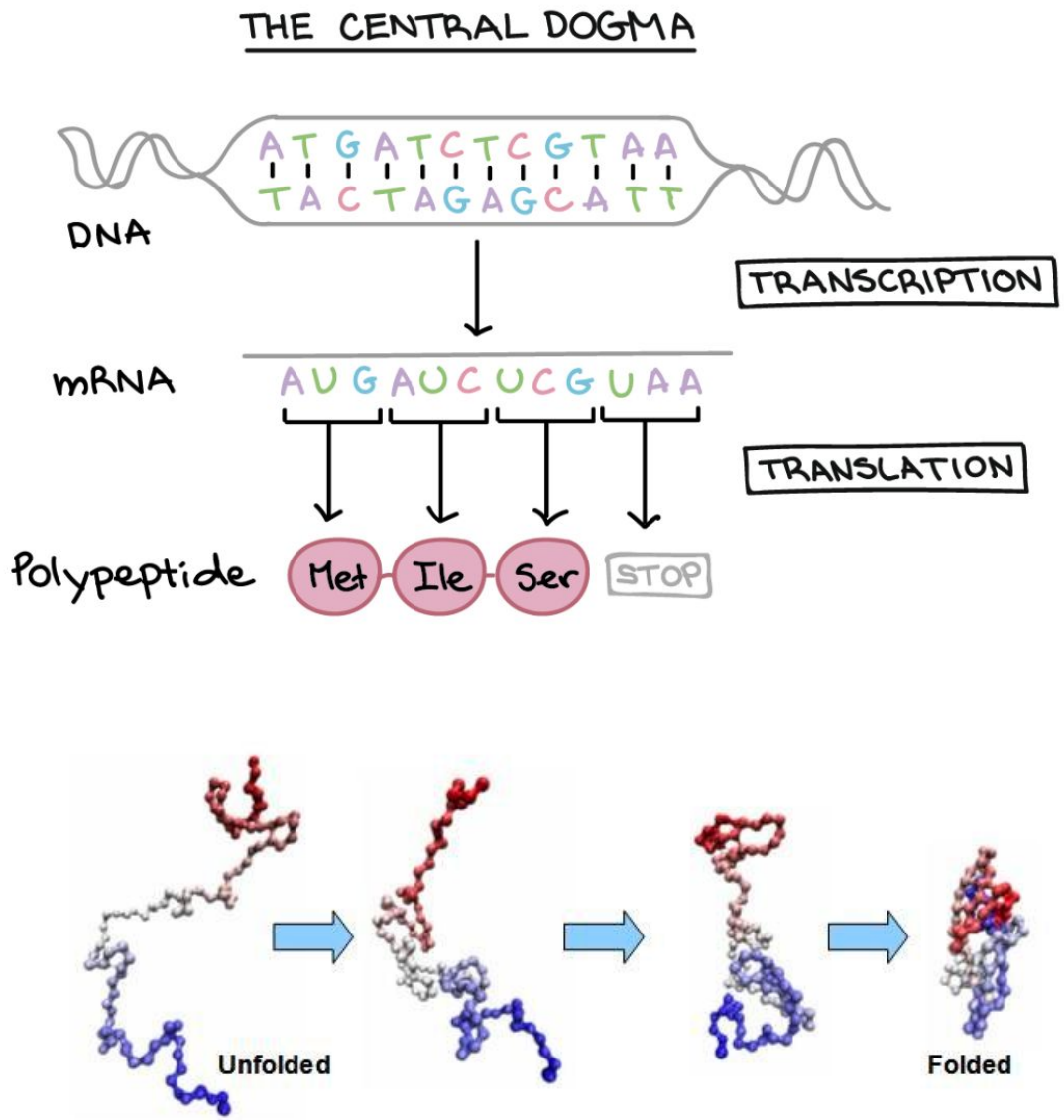


Background BIOLOGY

- Flu proteins (H1N1) and antibody response
- HI assay measures the virus phenotype based on what antibody binds to it

4 base pairs (nucleotides), 20 amino acids

First nucleotide	Second nucleotide: U	Second nucleotide: C	Second nucleotide: A	Second nucleotide: G	Third nucleotide
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA STOP	UGA STOP	A
	UUG Leu	UCG Ser	UAG STOP	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G



https://www.khanacademy.org/science/high-school-biology/hs-molecular-genetics/hs-rna-and-protein-synthesis/a/intro-to-gene-expression-central-dogma

Influenza Virus **hemagglutinin** (**HA**) surface protein

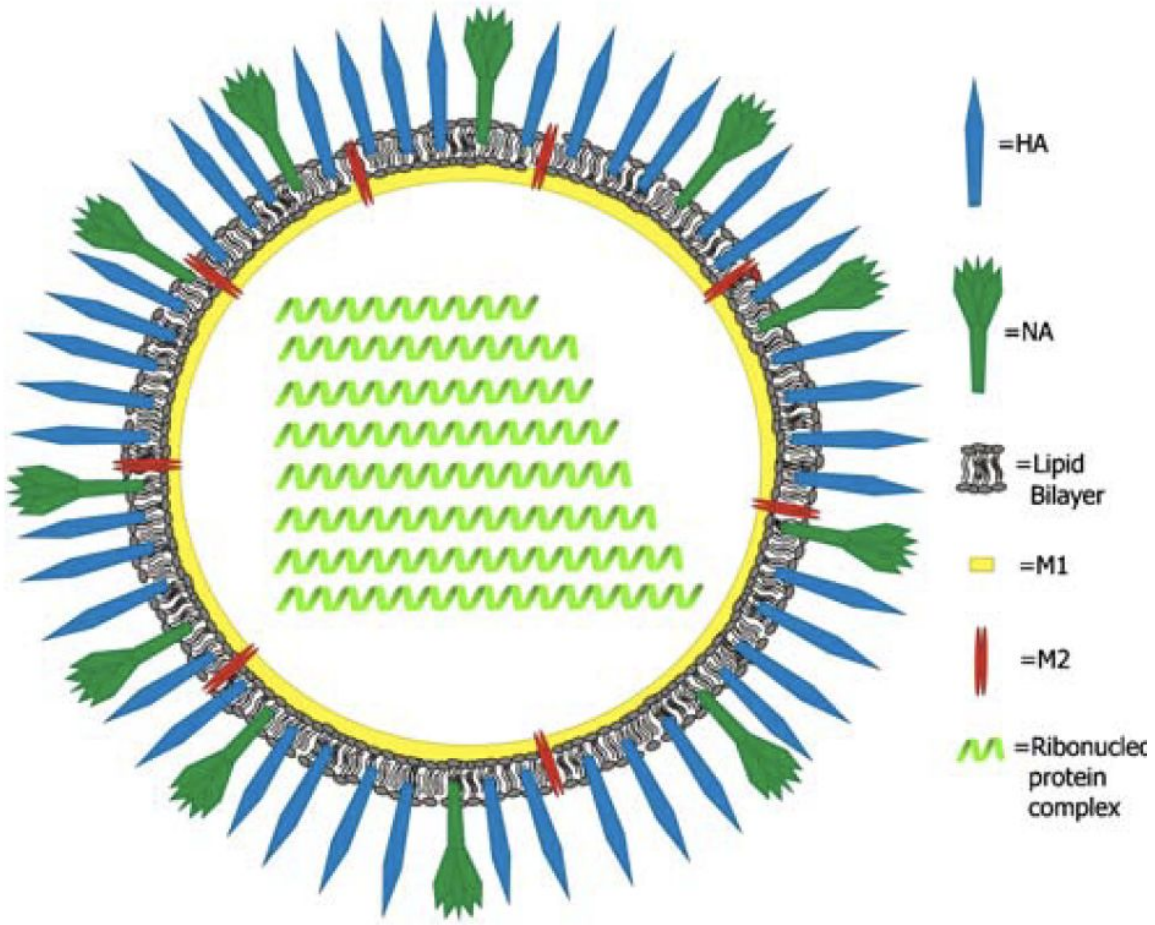
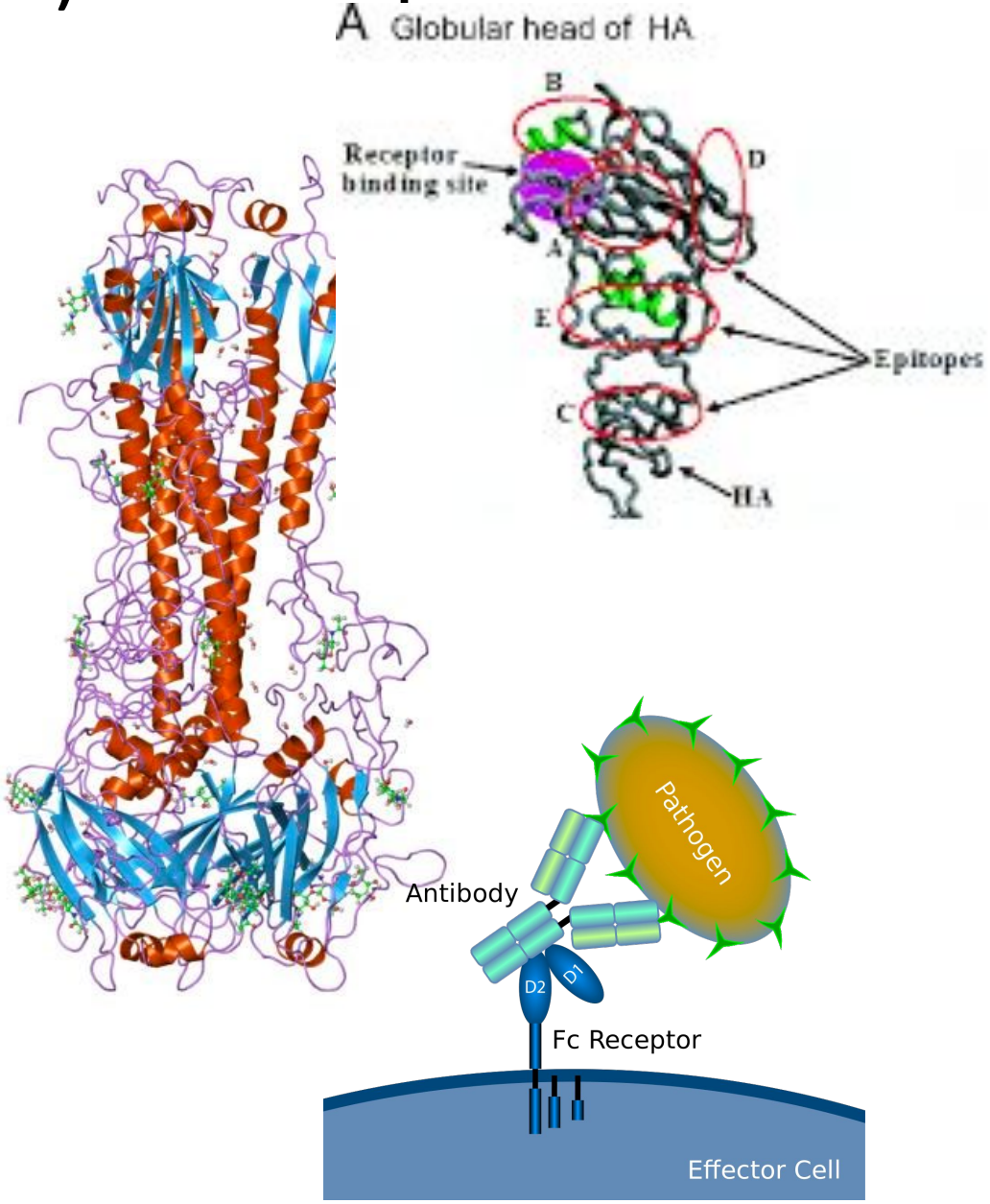
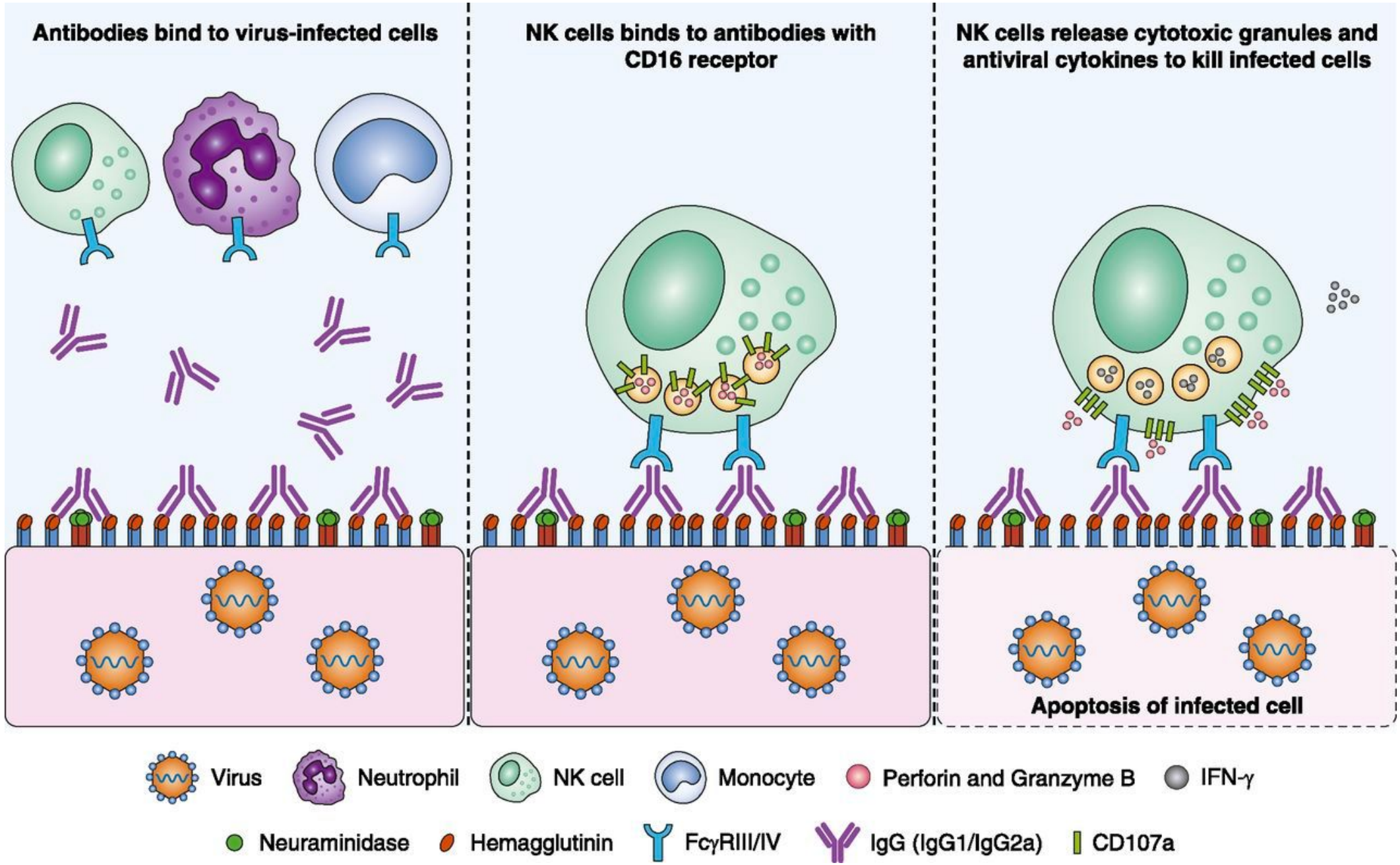


Fig. 1. A schematic drawing of the influenza virus.



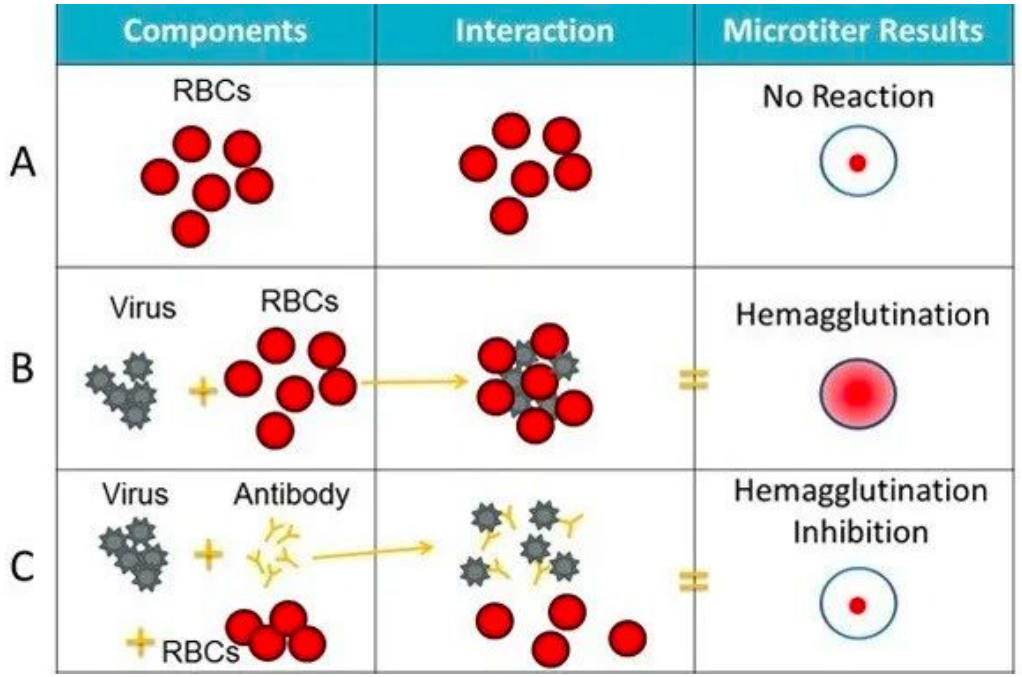
Neuraminidase (NA) is another protein that antibodies bind to
Strains identified as, eg H1NI or H3N2 or H5NI (pathogenic bird flu)

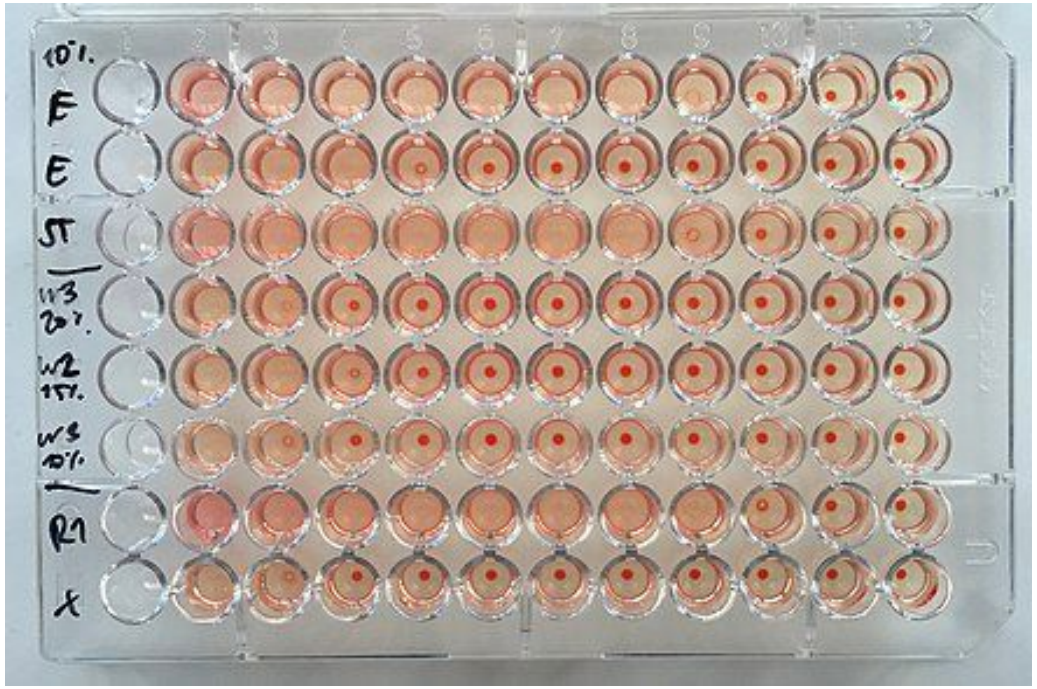

Antibodies bind to virus-infected cells
NK cells binds to antibodies with CD16 receptor
NK cells release cytotoxic granules and antiviral cytokines to kill infected cells
Apoptosis of infected cell
Virus
Neutrophil
NK cell
Monocyte
Perforin and Granzyme B
IFN-γ
Neuraminidase
Hemagglutinin
FcγRIII/IV
IgG (IgG1/IgG2a)
CD107a

Hemagglutination inhibition (HI) assay

- Hemagglutination: red blood cells bind to influenza virus → red blob suspended in solution

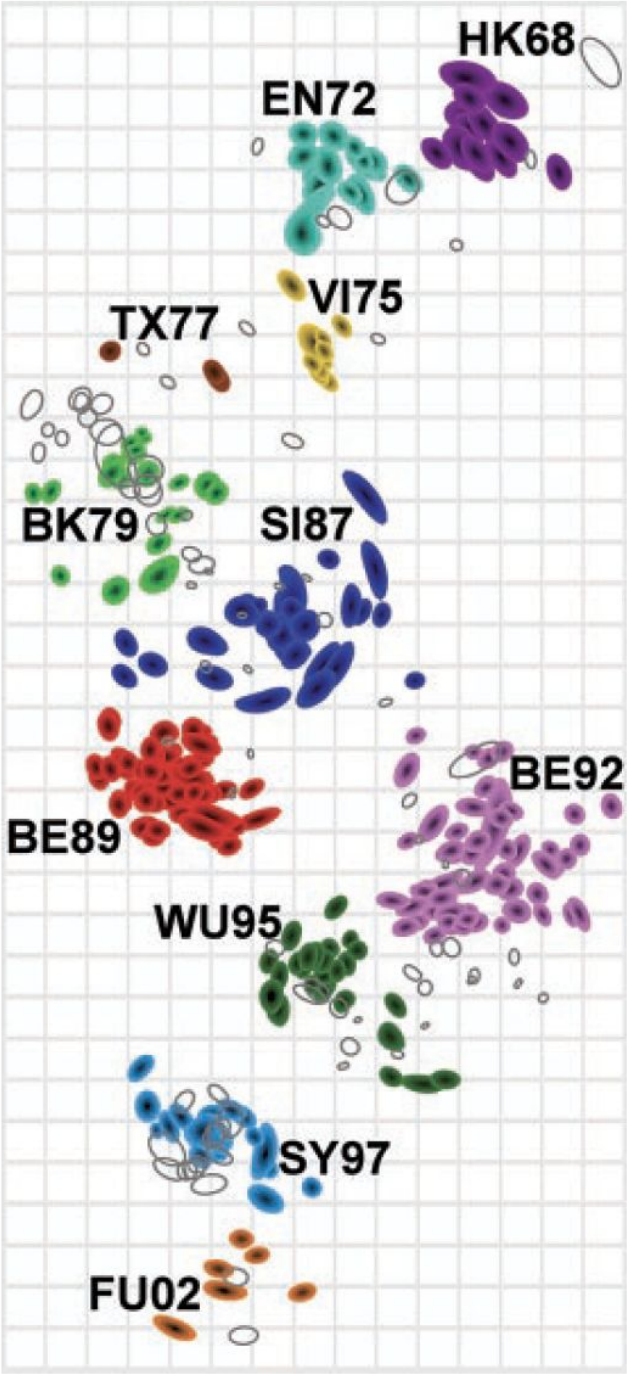
- The virus is diluted in each column
- Each row is a different anti-viral antibody inhibitor that disrupts virus-RBC interaction and dilutes the color
- Measure how far its diluted before the color disappears tells you how well the antibody binds to the virus



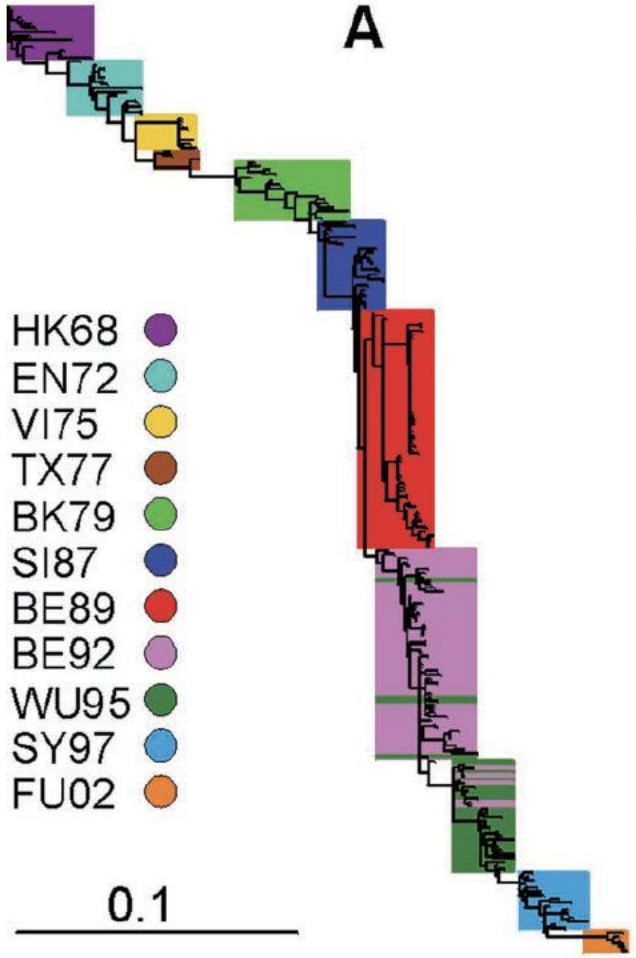


Detailed analysis

- Flu names: location of emergence + year
- Colored shapes = flu
  - Open circles = antibodies that bind to that flu, taken from people who survived that flu
- Distance on the map represents “HI distance” (phenotypic distance)
  - spaced to indicate how well each virus binds to all the different antibodies
- Each grid cell is 1 unit of antigenic (phenotypic distance)
  - 2 units = 4-fold distance
  - 3 units – 8 -fold distance
  - Rotation is irrelevant here, only distance matters

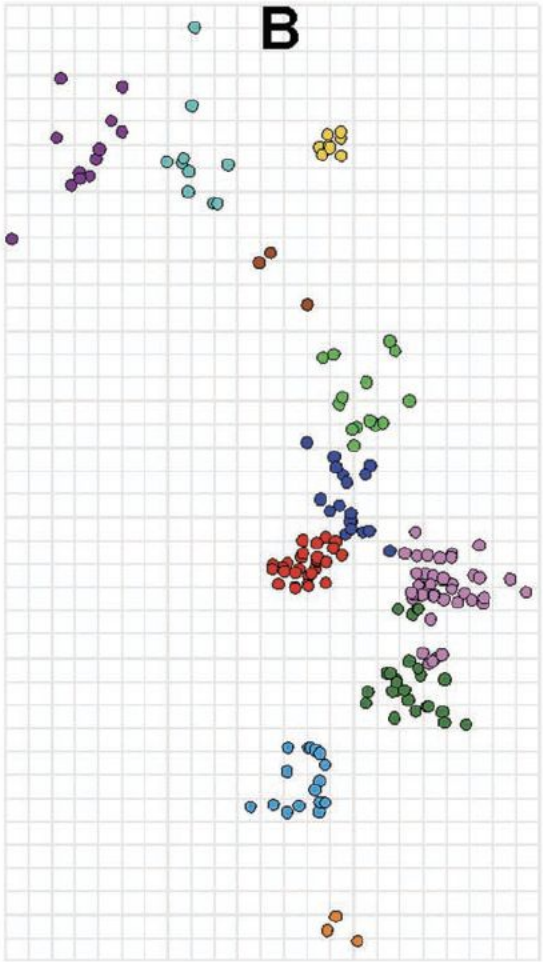


Phylogenetic Tree



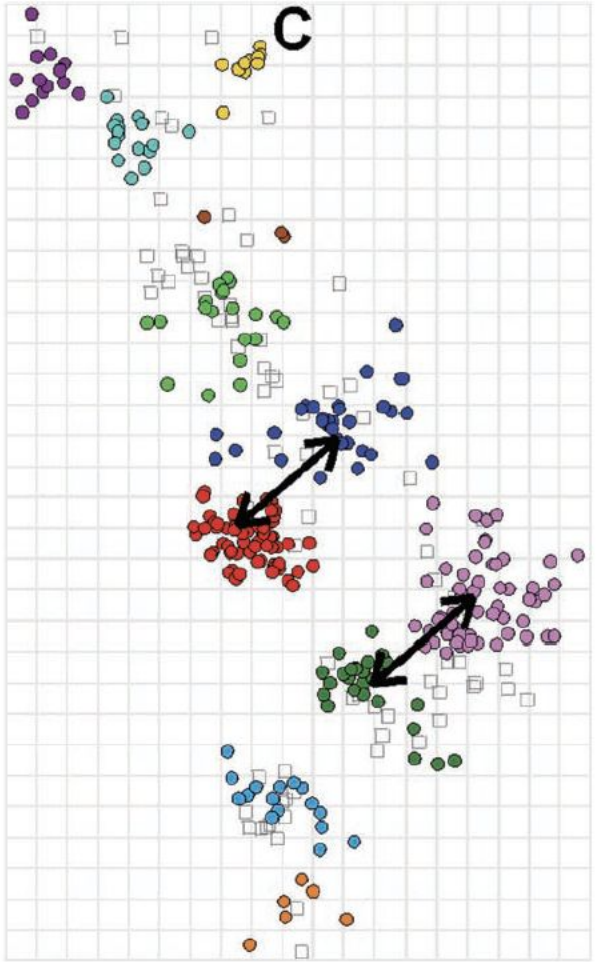
Genetic Map

1 cell = 2.5 Amino Acid substitutions



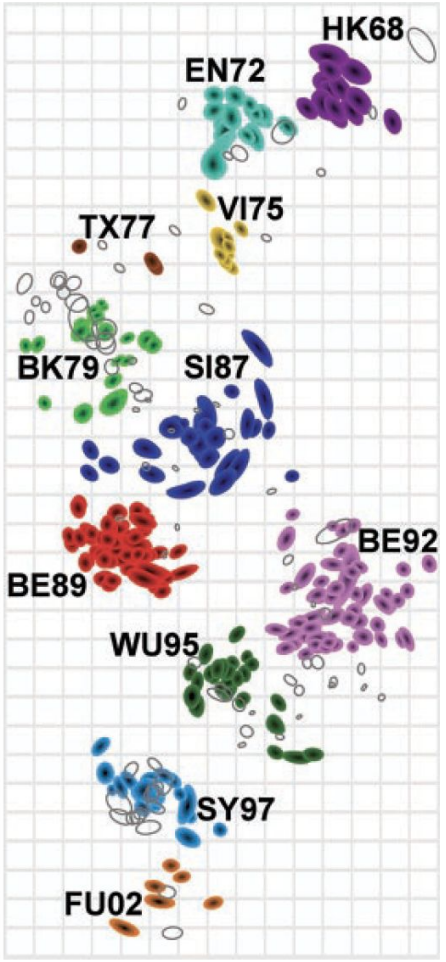
Antigenic Map

Rotated from original before TX77



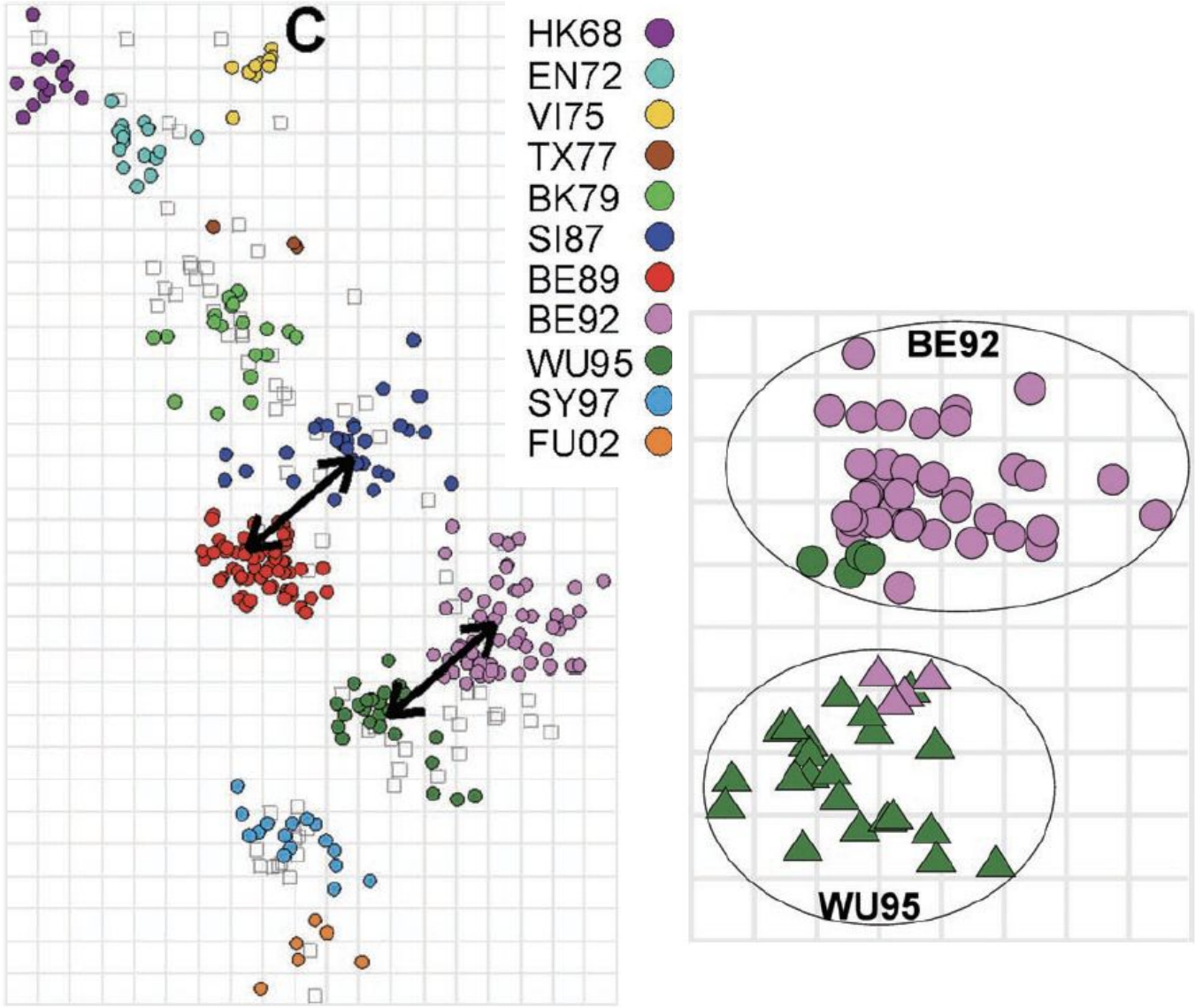
Antigenic Map

(Originally computed)



One amino acid change *can* shift to a new cluster (requiring a new vaccine needed)

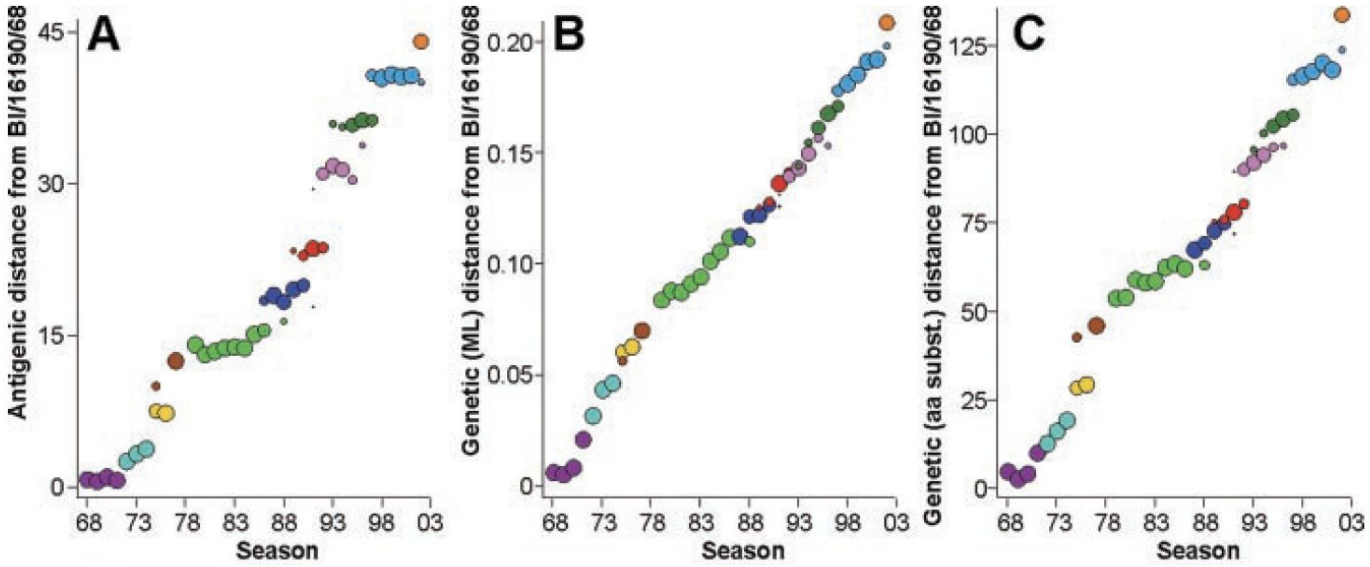
- Arrows indicate the two cluster transitions for which the amino acid substitution N145K is the only cluster-difference substitution
- Pink symbols have an N (asparagine)
- Green symbols replaced N with K (lysine) at position 145
- Circles in the upper BE92 oval are in the same genetic cluster
- Triangles in the WU95 are in the same genetic cluster
- Same N□K 145 shift happened in SI87 to BE89

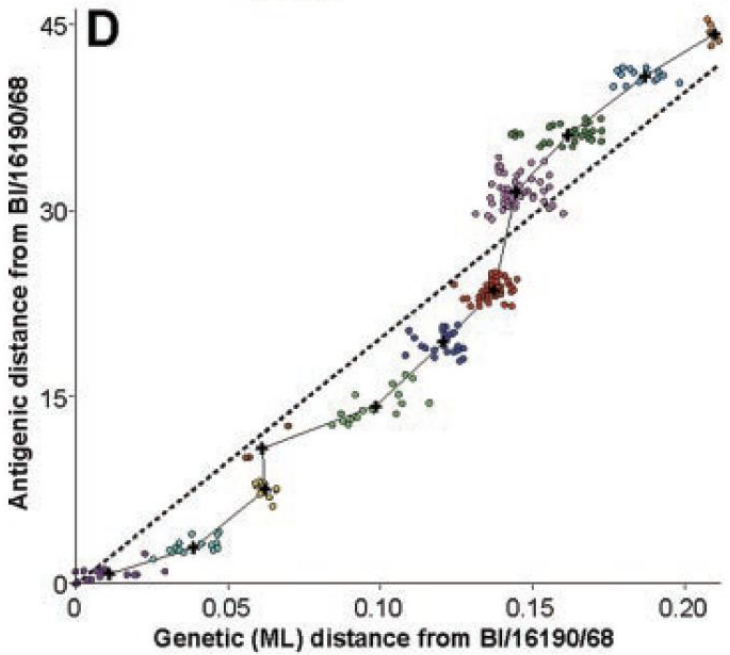


Antigenic (phenotype) Distance

Phylogenetic Distance

Amino Acid Distance





- Comparison of Antigenic and Genetic distance
  - (A vs B)
  - Most mutations are neutral
  - Occasionally small genetic change cause big phenotypic change
  - Indicated steepness in curved line compared to dashed straight line

Computational approach is scalable, easily interpretable, and readily applied to new data

“The method used in this manuscript is based on the fundamental ideas described by Lapedes and Farber (*4*) and, in particular, takes advantage of their observation that antigenic distance is linearly related to the logarithm of the HI measurement.

Exploiting this observation allowed us to create a new method that is parametric yet still handles HI measurements that are beyond the sensitivity of the HI assay (*9*).

We use a modification of metric MDS (*25*) to position the antigens and antisera in the map (*9*). This new approach offers computational advantages over the ordinal approach, including reduced running time and fewer local minima, making it tractable to run on datasets the size of the one used in this manuscript, and on larger datasets.”

“The resolution of the map can be greater than the resolution of the assay because the location of a point in the map is fixed by measurements to multiple other points, thereby averaging out errors.”

Some of these “cluster-difference” substitutions (*9*) will contribute to the antigenic difference between clusters, some may be compensatory mutations to retain function, and others may be hitchhikers carried along by chance. Of the 67 cluster-difference amino acid substitutions, 63 were in antigenic sites (*28*), 8 were in the receptor-binding site (*29*), and 21 were in codons previously identified as positively selected

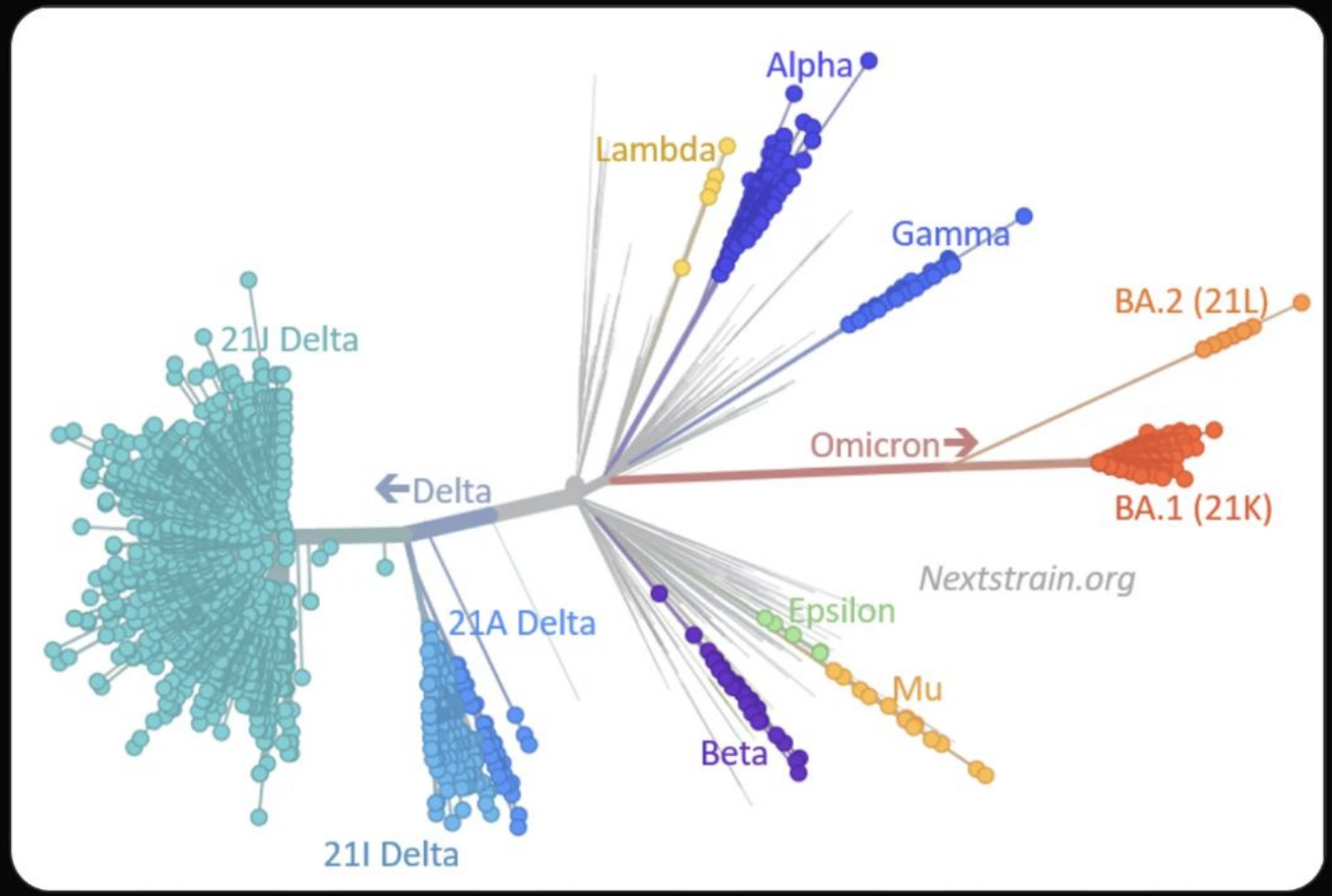
The correlation between antigenic distance and the number of amino acid substitutions between strains was 0.81, and on average, 2.9 amino acid substitutions resulted in one unit change in antigenic distance.

Surprisingly, a single amino acid substitution, N145K (*32*), is the only cluster-difference substitution between the SI87 and BE89 and between the Beijing 1992 (BE92) and Wuhan 1995 (WU95) clusters.

Dr Emma Hodcroft @firefoxx66 · Jan 27

This filtered @Nextstrain build gives a nice visual display of how distant the Omicron family is from everything else, & how different BA.1 (21K) & BA.2 (21L) are from each other. Distance is in mutations.

nextstrain.org/ncov/gisaid/gl...



Antigenic Map of SARS-CoV-2 Before Omicron

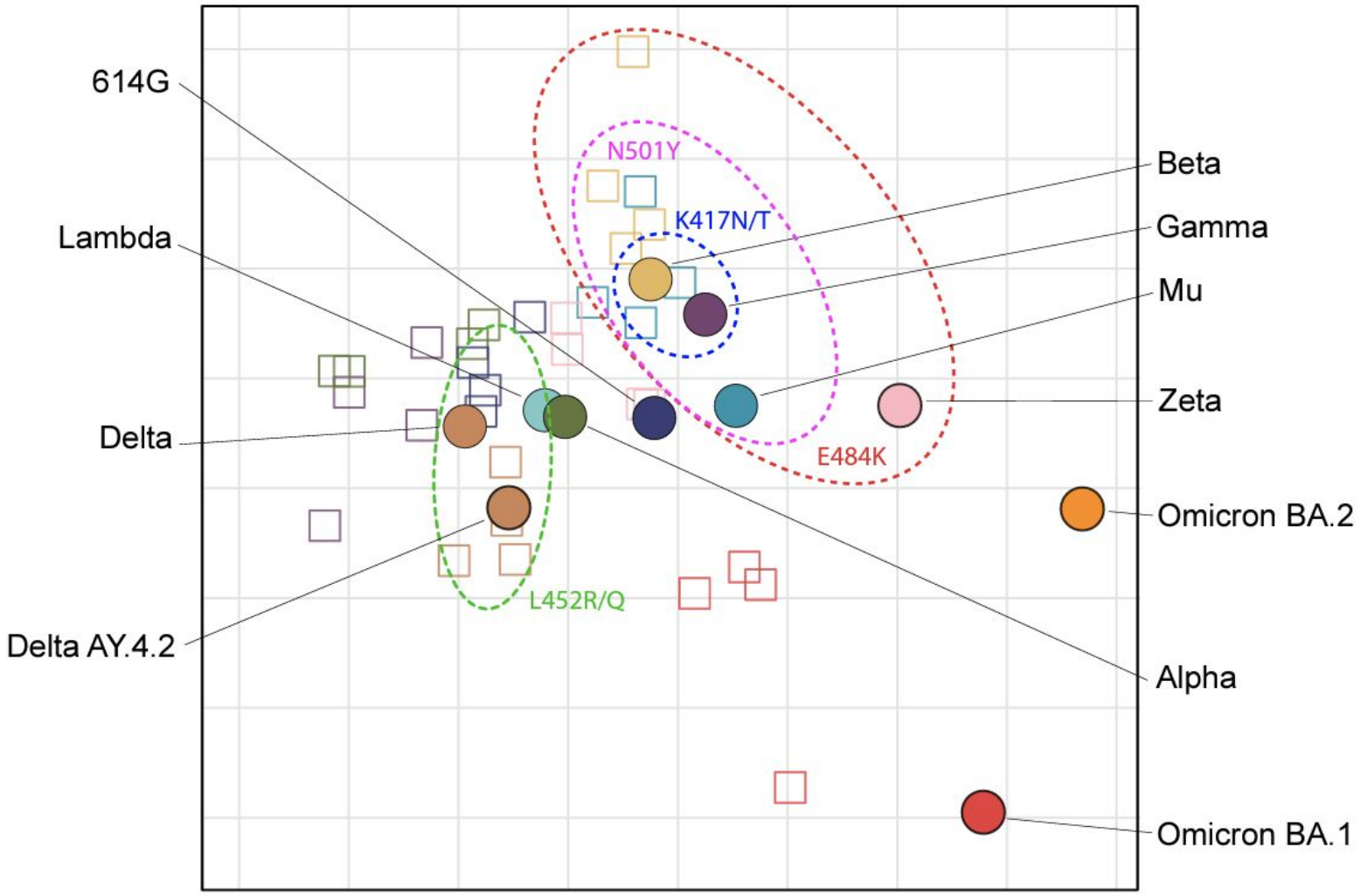
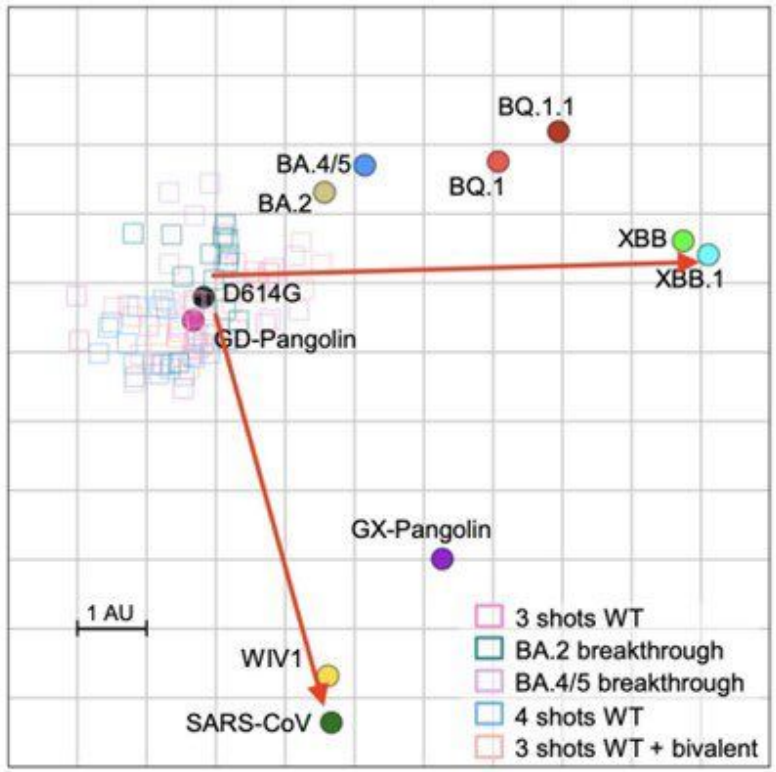


Fig. 4. Antigenic cartography using authentic SARS-CoV-2. a-h, Neutralizing titers of hamsters infected with either (**a**) 614G, (**b**) Alpha, (**c**) Beta, (**d**) Gamma, (**e**) Zeta, (**f**) Delta, (**g**) Mu or (**h**) Omicron BA.1 viruses. **i,** Multidimensional scaling was used to create an antigenic map utilizing PRNT50 titers generated from authentic SARS-CoV-2 on Calu-3 cells. See legend to Fig. 3 for details. Subdivided by dotted ellipses a variants possessing overlapping substitutions as indicated. Geometric mean is displayed above each graph. PRNT50: plaque reduction neutralization titers resulting in 50% plaque reduction. Dotted lines indicate limits of detection. Error bars indicate SEM.

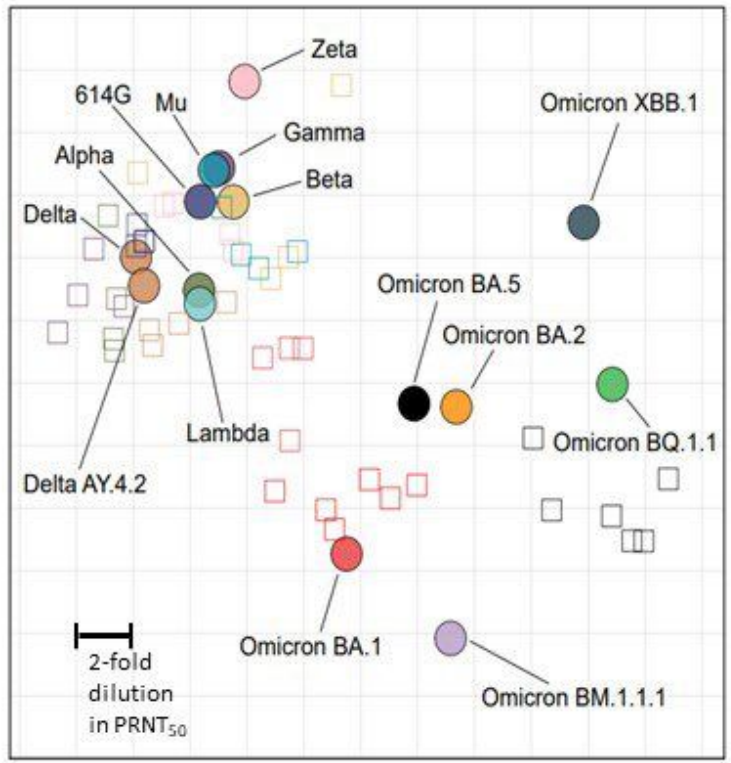
https://www.biorxiv.org/content/10.1101/2022.02.23.481644v1

Antigenic maps after Omicron

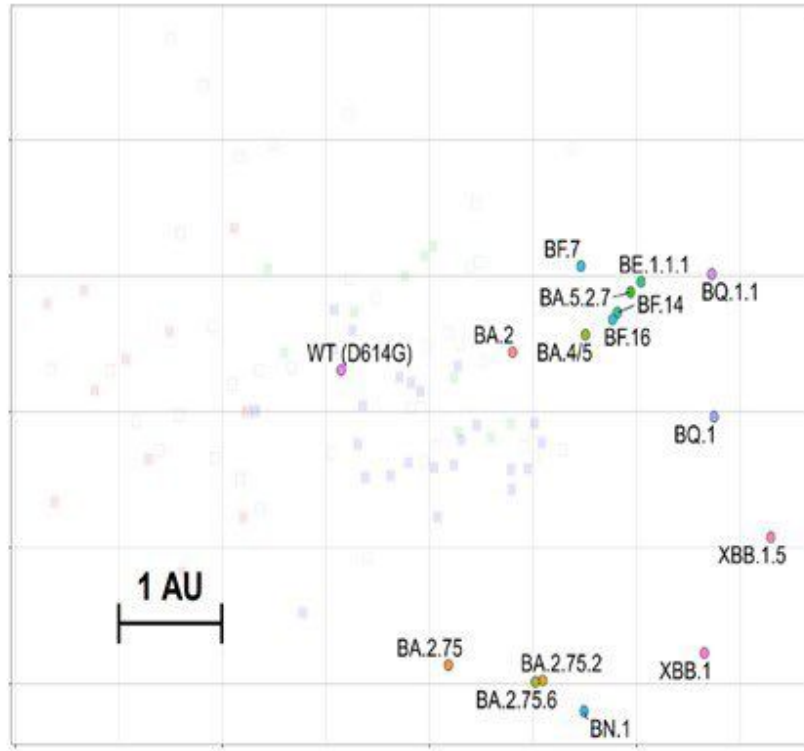
Not just genetic distance, but also serological distance



https://www.cell.com/cell/fulltext/S0092-8674(22)01531-8



https://www.thelancet.com/cms/10.1016/S2666-5247(22)00384-6/attachment/55ffc73f-eb19-4d55-b863-988d1c151589/mmc1.pdf



https://www.biorxiv.org/content/10.1101/2023.02.07.527406v1

Application:

Antigenic maps show that antibody escape can be measured in psuedoviruses (safer to use in lab) to estimate antibody escape in the real virus.

https://www.science.org/doi/full/10.1126/sciimmunol.abq4450

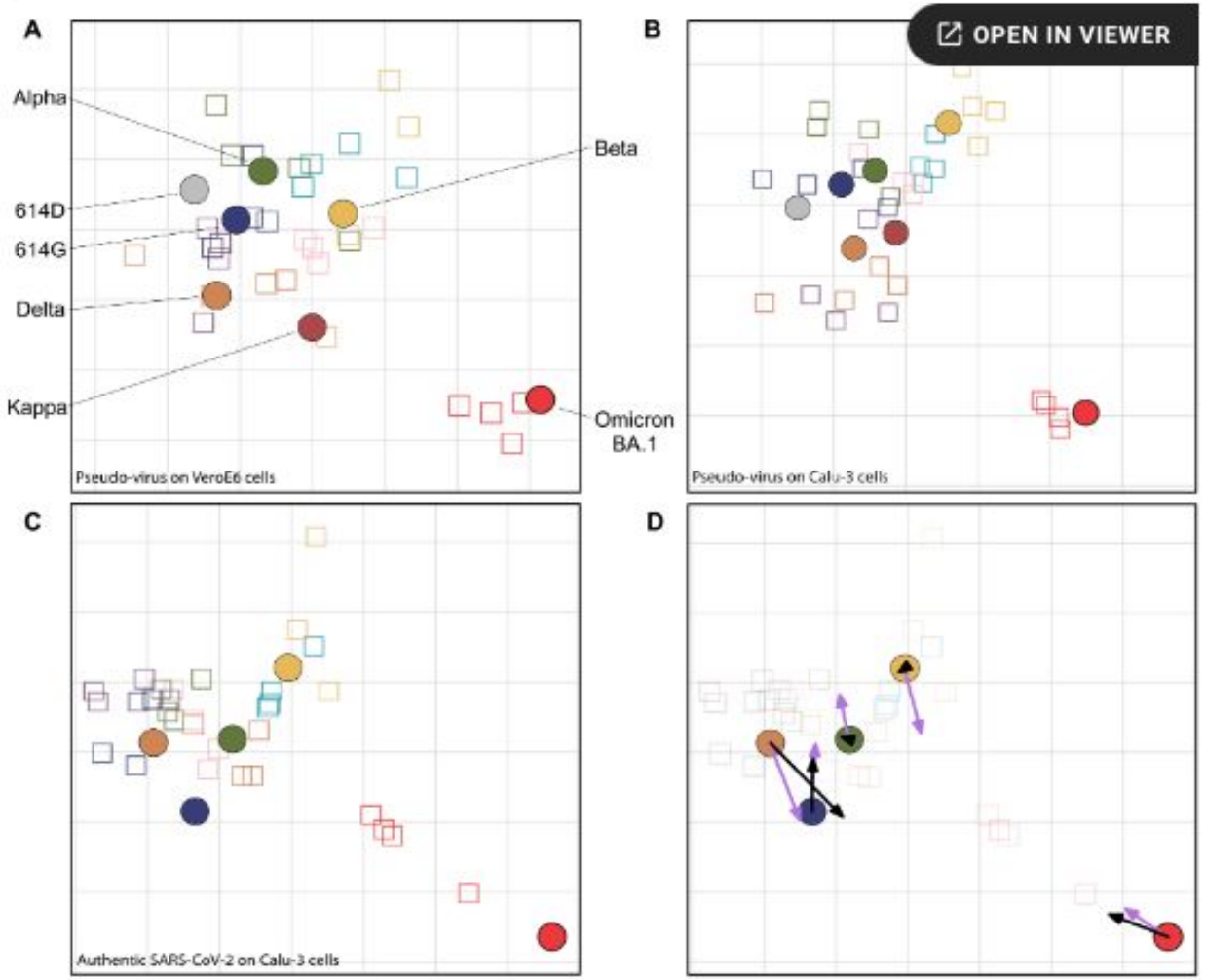
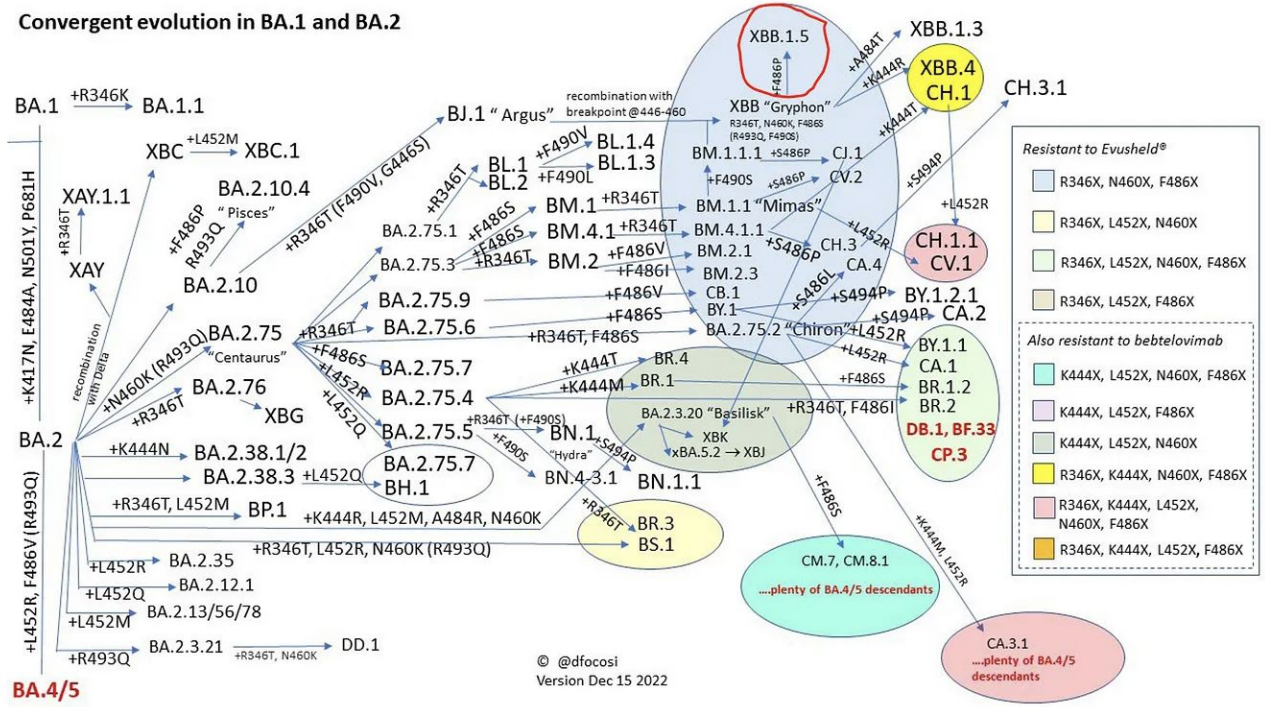
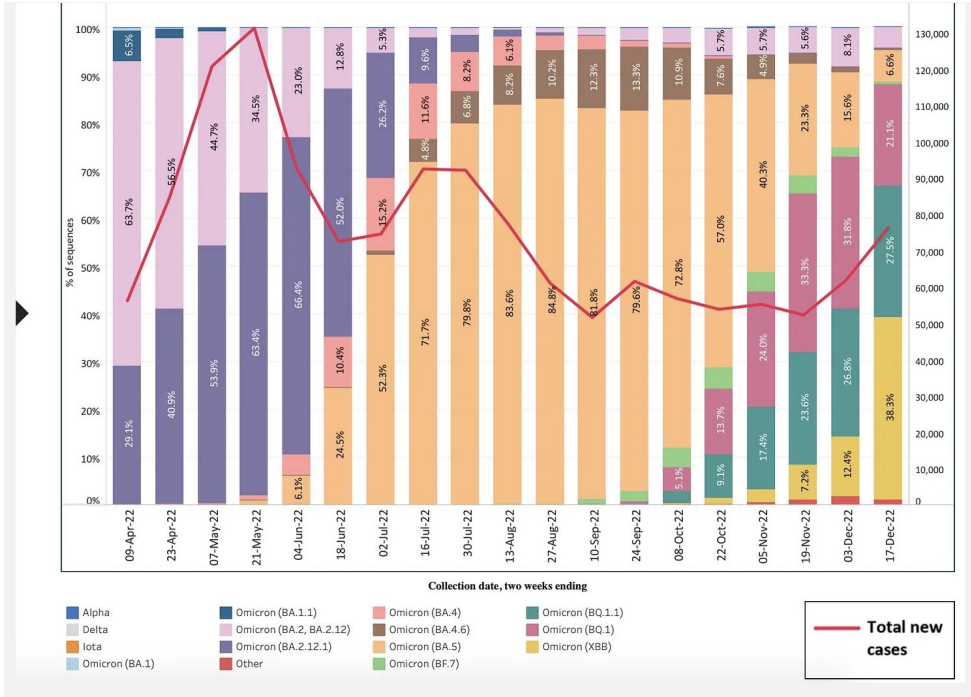


Fig. 3. Antigenic maps comparing neutralizations with SARS-CoV-2 pseudo-viruses and authentic SARS-CoV-2.

(**A** and **B**) MDS was used to create an antigenic map from the PRNT50 titers generated against 614D, 614G, Alpha, Beta, Delta, Kappa, and Omicron pseudo-viruses on either VeroE6 (A) or Calu-3 (B) cells. (**C**) MDS was used to create an antigenic map from the PRNT50 titers generated against 614G, Alpha, Beta, Delta, and Omicron authentic SARS-CoV-2. (**D**) Redisplay of antigenic map in (C) with lilac arrows indicating antigen positions in map (A) and black arrows indicating antigen positions in map (B). Viruses are shown as colored circles and antisera are shown as squares with the same outline color as the matching viruses. Viruses and antisera are positioned in the map so that the distances between them are inversely related to the antibody titers, with minimized error. The grid in the background scales to a twofold dilution of antisera in the titrations.

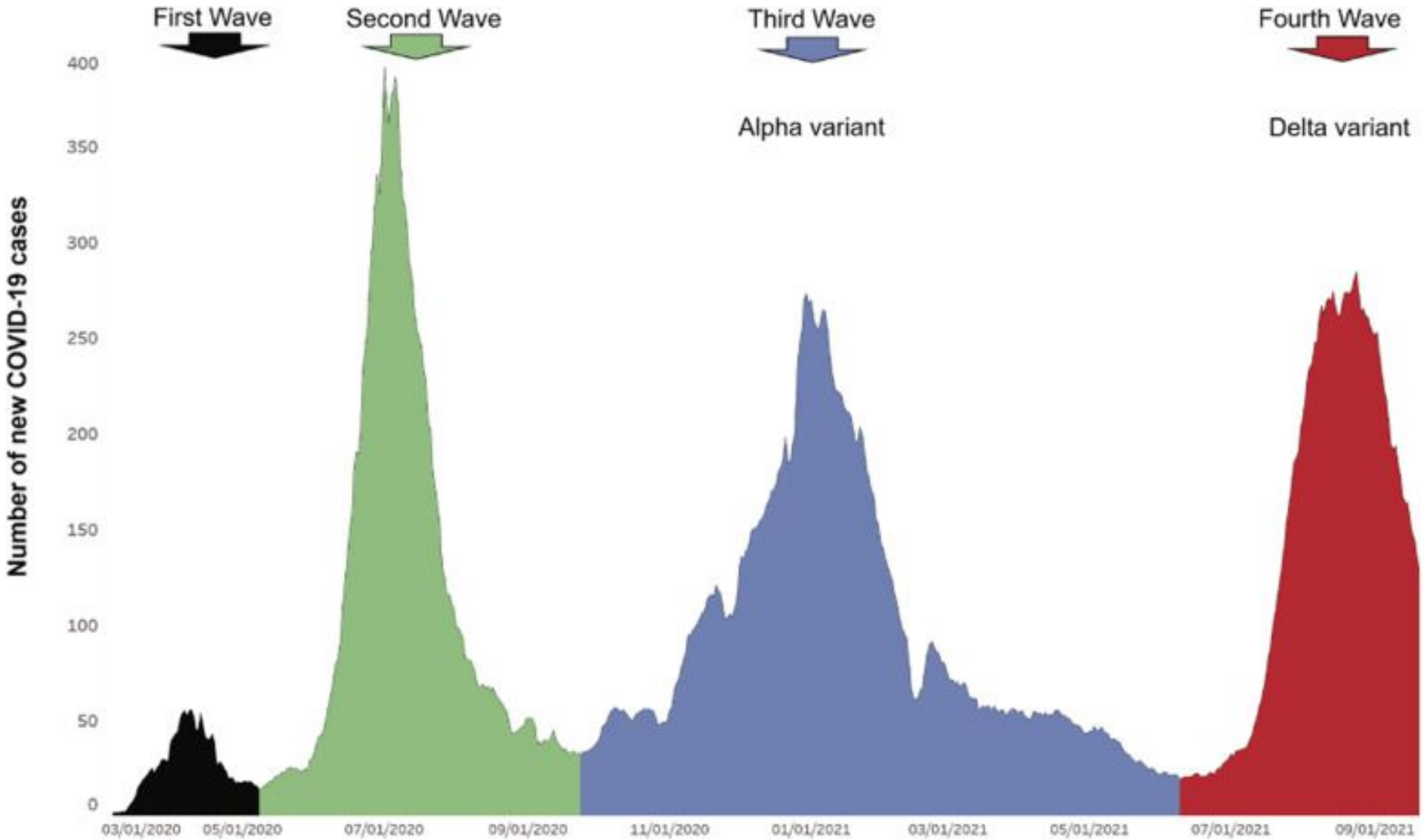




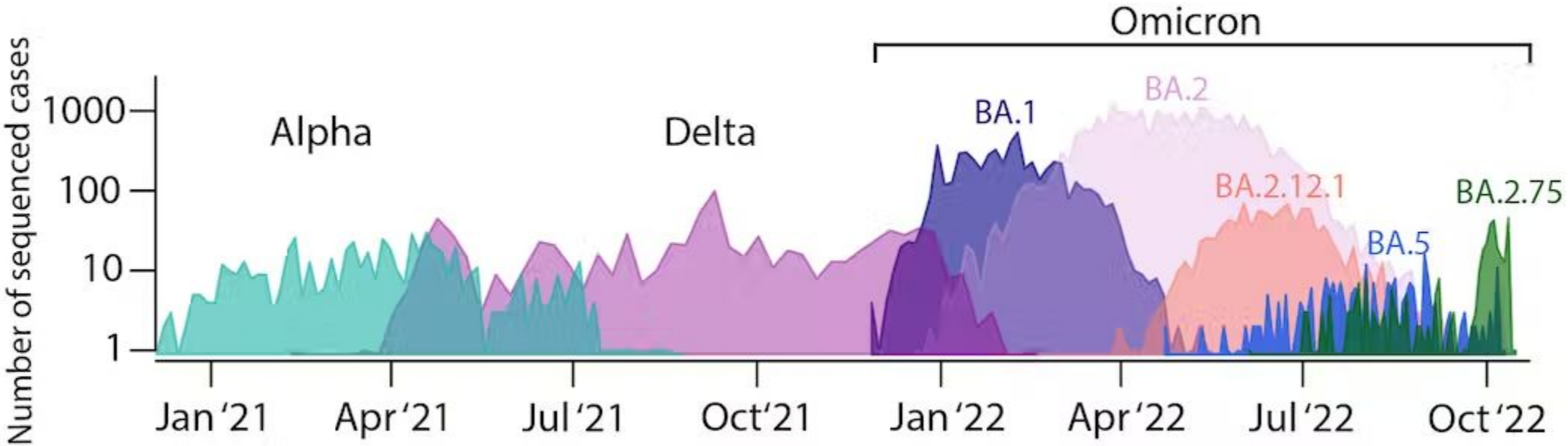
As Jesse Bloom warned us from their comprehensive screening of spike protein nucelotides, "Watch those sites for future mutations!" (see 486 below) As Ryan Hisner very recently posted: "F486P is a hell of a mutation, as Jesse Bloom's lab RBD heat map made clear many months ago. It's the #1 AA residue for RBD expression and ACE2 binding strength. It just took the virus a while to get over the 2-nuc-mutation barrier."

https://erictopol.substack.com/p/a-new-variant-alert

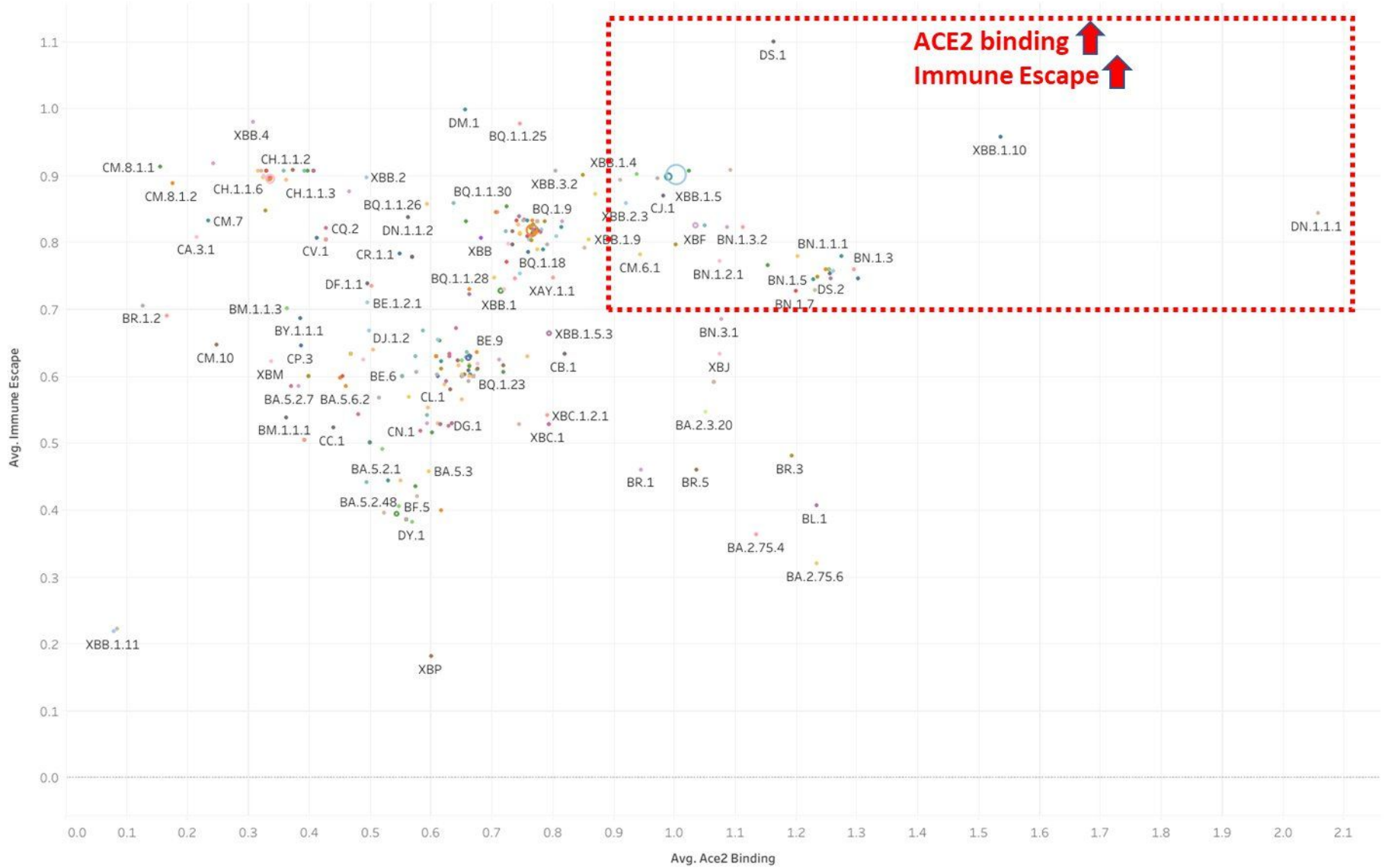
In the good old days



Variant soup



Predicted Immune Escape vs ACE2 Binding Scores (Relative to BA.2)



Specimen Collection Date
Non-Null Values Only

Division
All

Country
All

Pangolin Lineage
~~Multiple values~~

QC Status
- [ ] Null
- [ ] bad
- [x] good
- [ ] mediocre

Sequences
1
2,000
4,000
6,000
8,934

|Tracking SARSCoV2 Lineages - Global 20 Day Trends - Antigua and Barbuda, Australia, Austria and 46 more |NYITCOMResearchReport | In Silico Immune Escape and ACE2 binding scores were generated using nextclade tool by Cornelius Roemer et al

Predicted impact of RBD mutations on ACE2 binding, relative to a BA.2 baseline.
The score is calculated using the same data and logic as Bloom lab's ACE2 binding calculator (see https://doi.org/10.1101/2022.09.20.508745). Predicted neutralizing antibody escape relative to BA.2. The score is calculated using the same logic as Bloom lab's antibody escape calculator (see https://doi.org/10.1093/ve/veac021) using data generated by Yunlong Cao's group (see https://doi.org/10.1101/2022.09.15.507787).

Note: Select good QC sequences from the menu above

Evolutionary success of virus

Outcompete current strains (invasion)

Evade existing immunity

Produce lots of infectious virions