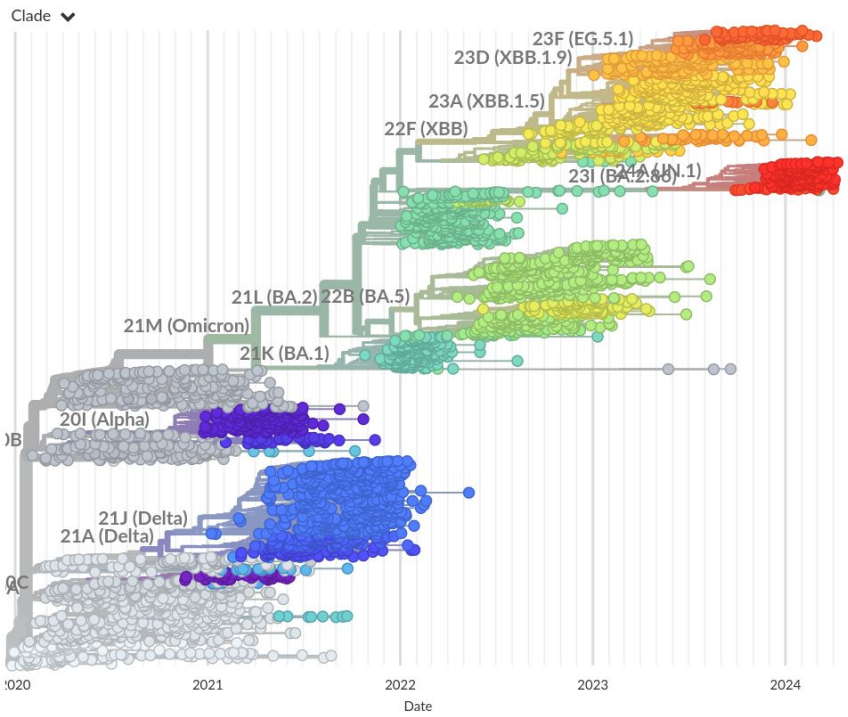
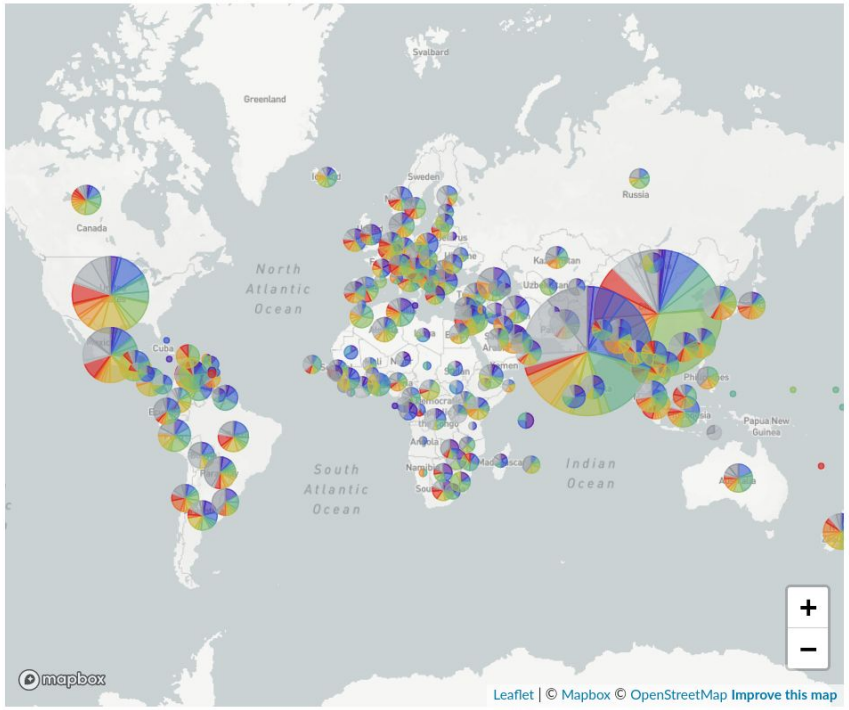
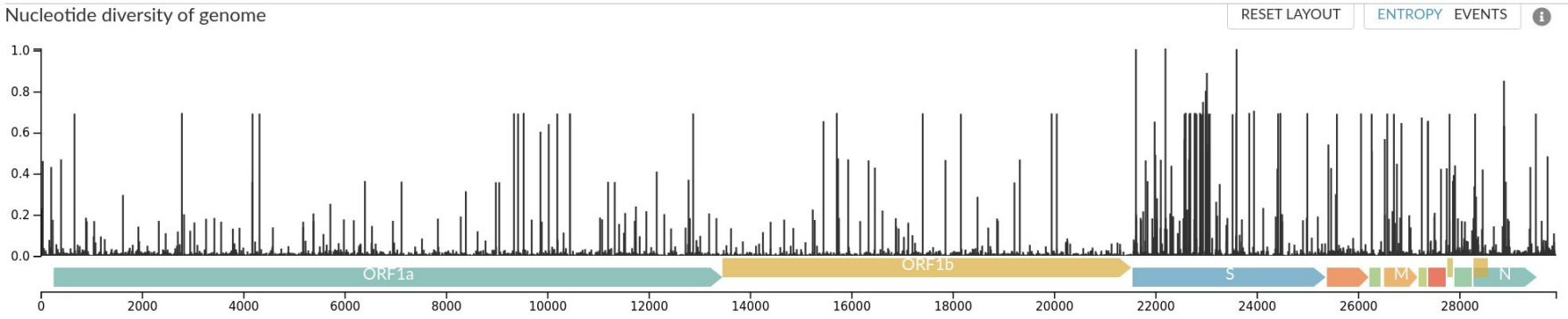
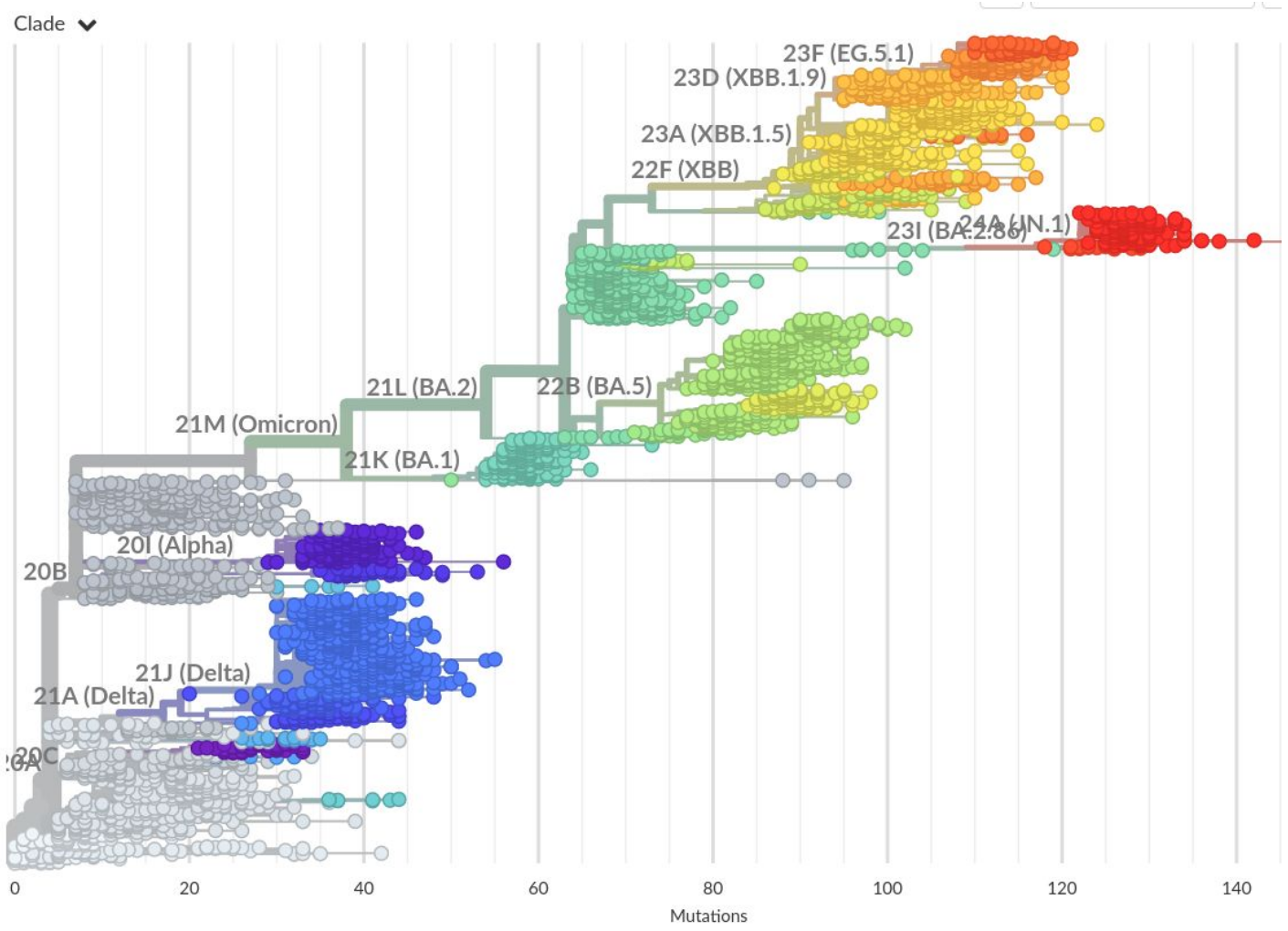
Evolution of SARS-CoV-2, immune escape, and the emergence variants of concern

Richard Neher, University of Basel

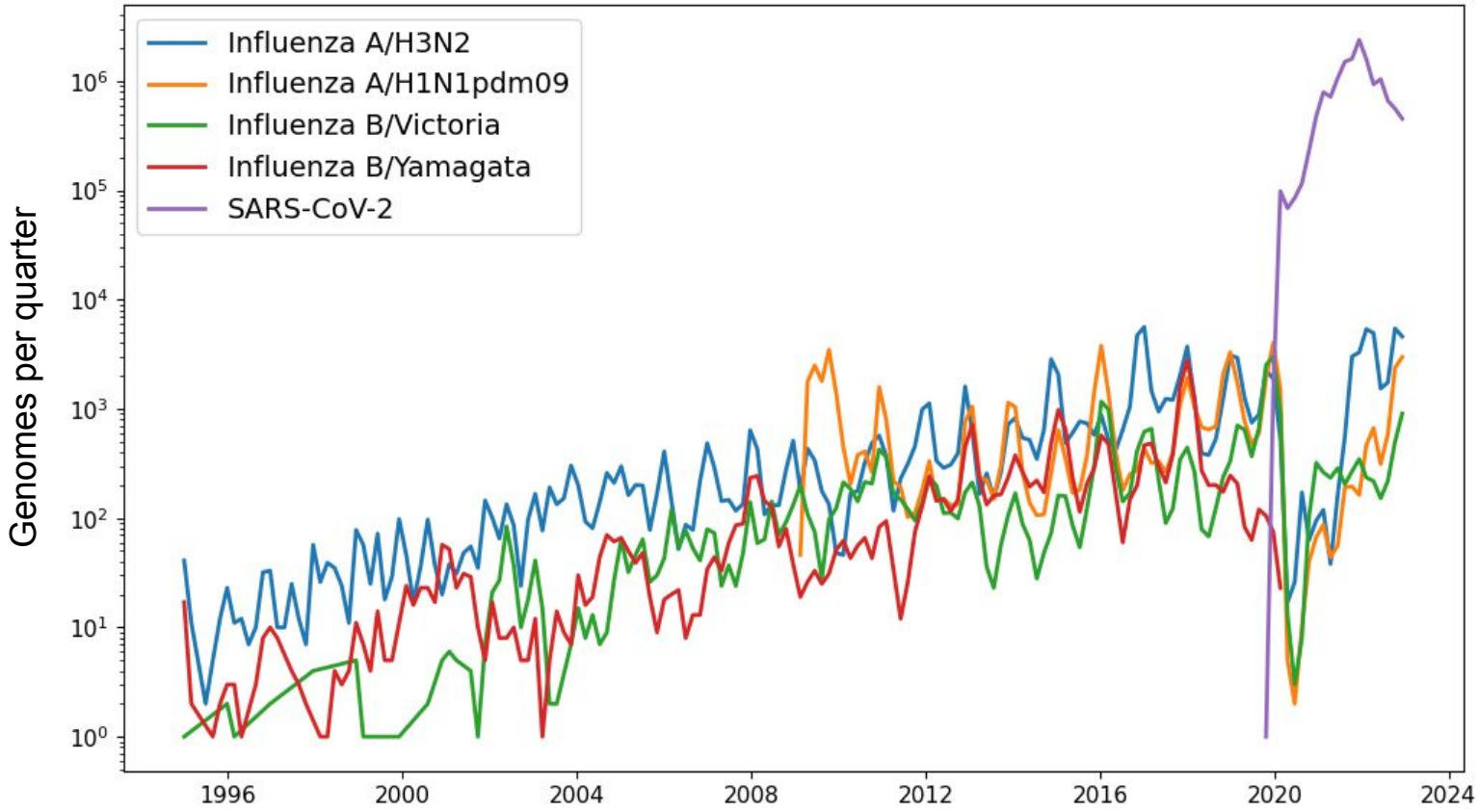

Clade
23F (EG.5.1)
23D (XBB.1.9)
23A (XBB.1.5)
22F (XBB)
23I (BA.2.86)
21L (BA.2)
22B (BA.5)
21M (Omicron)
21K (BA.1)
20I (Alpha)
21J (Delta)
21A (Delta)
2020
2021
2022
2023
2024
Date


Greenland
Svalbard
Canada
Russia
Sweden
North Atlantic Ocean
South Atlantic Ocean
Indian Ocean
Papua New Guinea
mapbox
Leaflet | © Mapbox © OpenStreetMap Improve this map


Nucleotide diversity of genome
RESET LAYOUT
ENTROPY
EVENTS
1.0
0.8
0.6
0.4
0.2
0.0
ORF1a
ORF1b
S
M
N
0
2000
4000
6000
8000
10000
12000
14000
16000
18000
20000
22000
24000
26000
28000


Clade
23F (EG.5.1)
23D (XBB.1.9)
23A (XBB.1.5)
22F (XBB)
24A (JN.1)
23I (BA.2.86)
22B (BA.5)
21L (BA.2)
21M (Omicron)
21K (BA.1)
20I (Alpha)
20B
21J (Delta)
21A (Delta)
20C
0
20
40
60
80
100
120
140
Mutations

Millions of publicly available viral genomes



What can we learn from so much data?

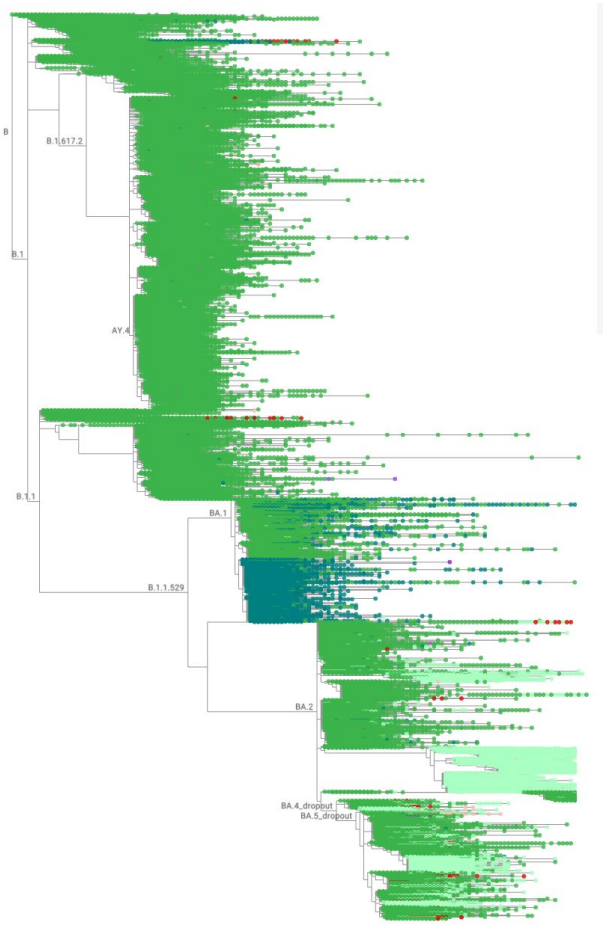
Pandemic scale trees – UShER and Taxonium

- Now >16M SARS-CoV-2 sequences
- Pre-pandemic tools can't deal with such volumes
- New tools: UShER, CoV-spectrum, Taxonium

  → Using UShER (by UCSC), Angie Hinrichs has continuously updated a tree of most available data

Rough estimate:

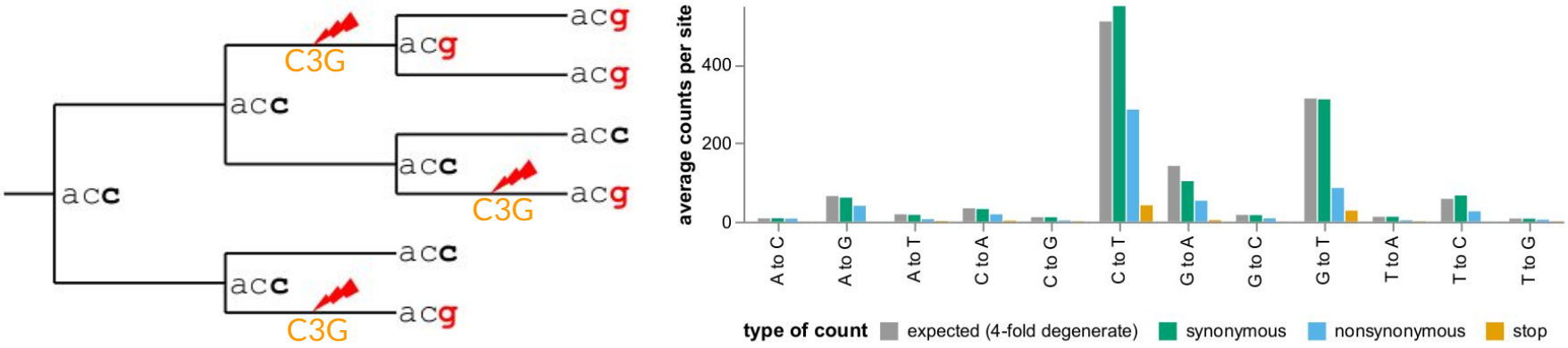
- Each leaf contributes 1 month of evolution
  → more than 1M years of evolution
- Every position mutated 100s of times somewhere on the tree.
  → should allow inference of position specific properties



Taxonium by Theo Sanderson

Mutation spectra are dominated by C->T and G->T mutations

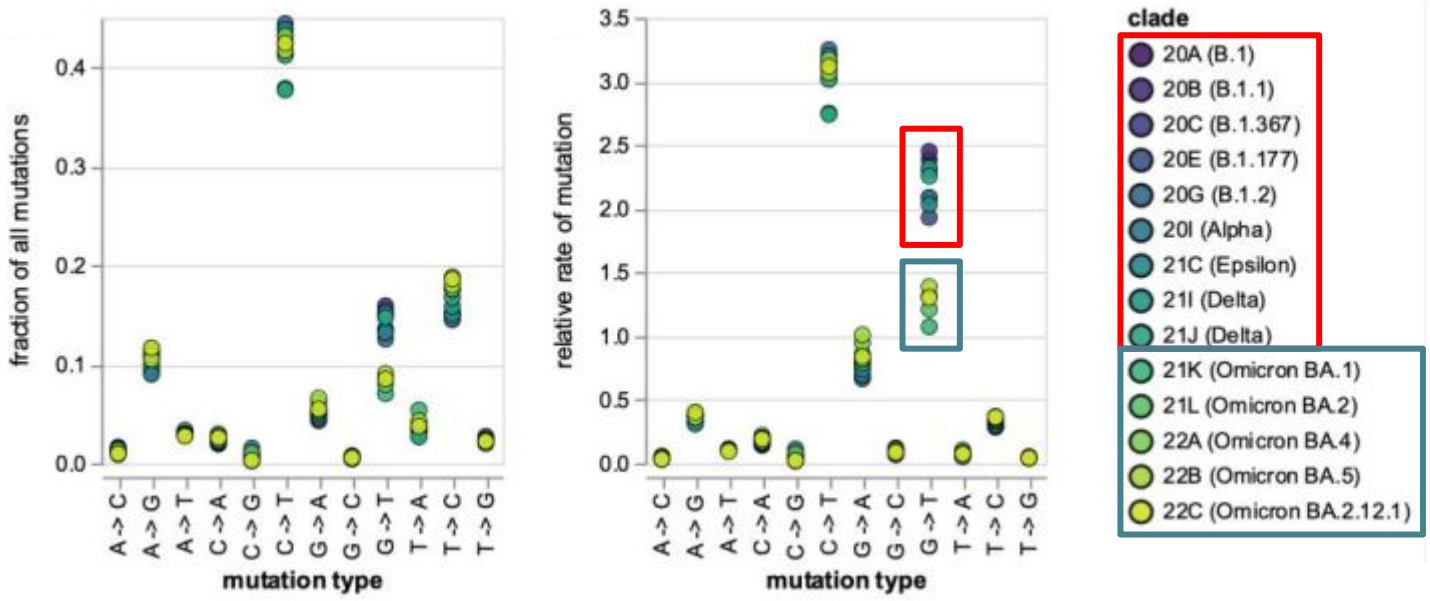
UShER tree is annotated with mutations:



- Expected patterns of purifying selection (more later)
- Fixed or high frequency mutations behave differently (more later)

Bloom and Neher, 2023

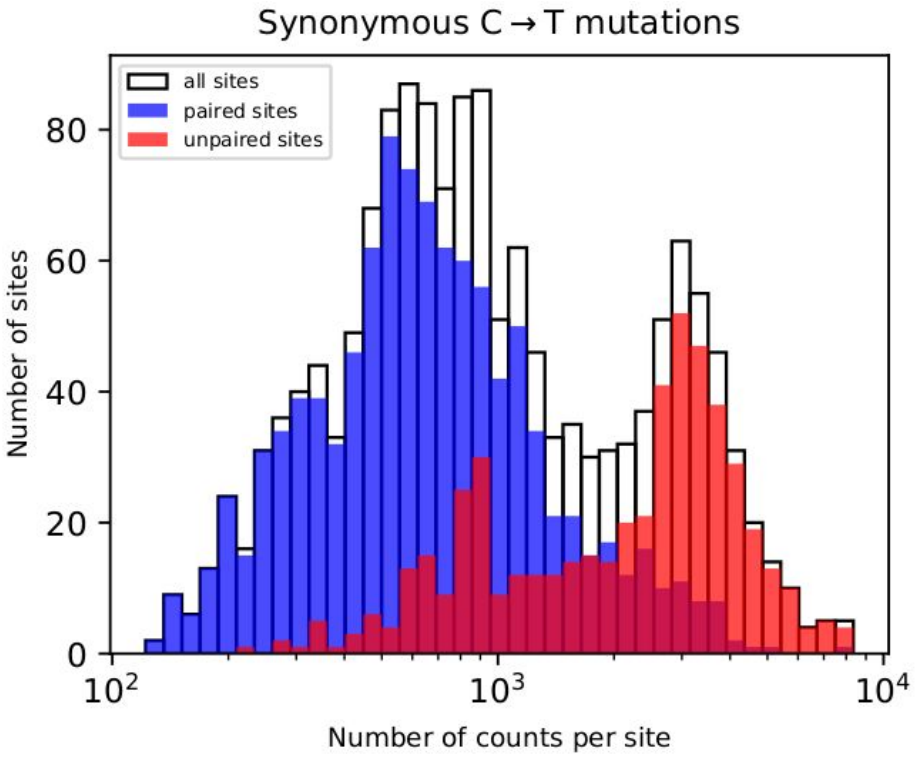
Mutation spectra differ between pre-Omicron and Omicron variants

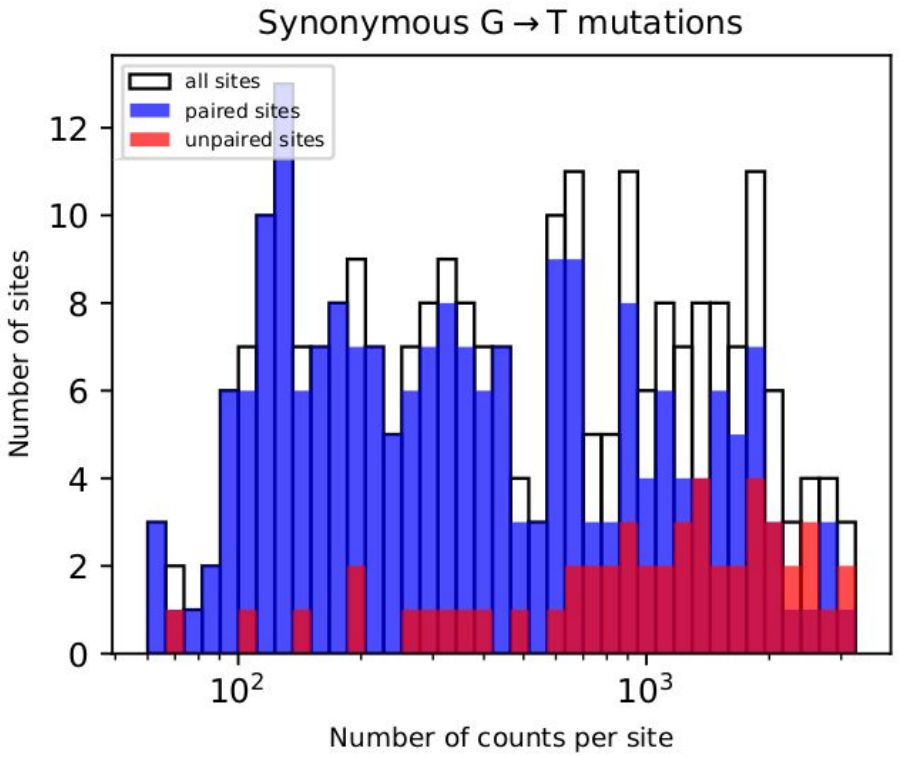


- C->T and G->T mutations are likely driven by processes that vary with the degree of immune activation
- C->T mutations depend strongly on secondary structure (Hensel, 2024)
- See also Lamb et al, 2024, Ruis et al, 2022

Bloom et al, 2023

Mutation rates vary from site to site and depend on context





Zach Hensel. 10.1101/2024.02.27.581995

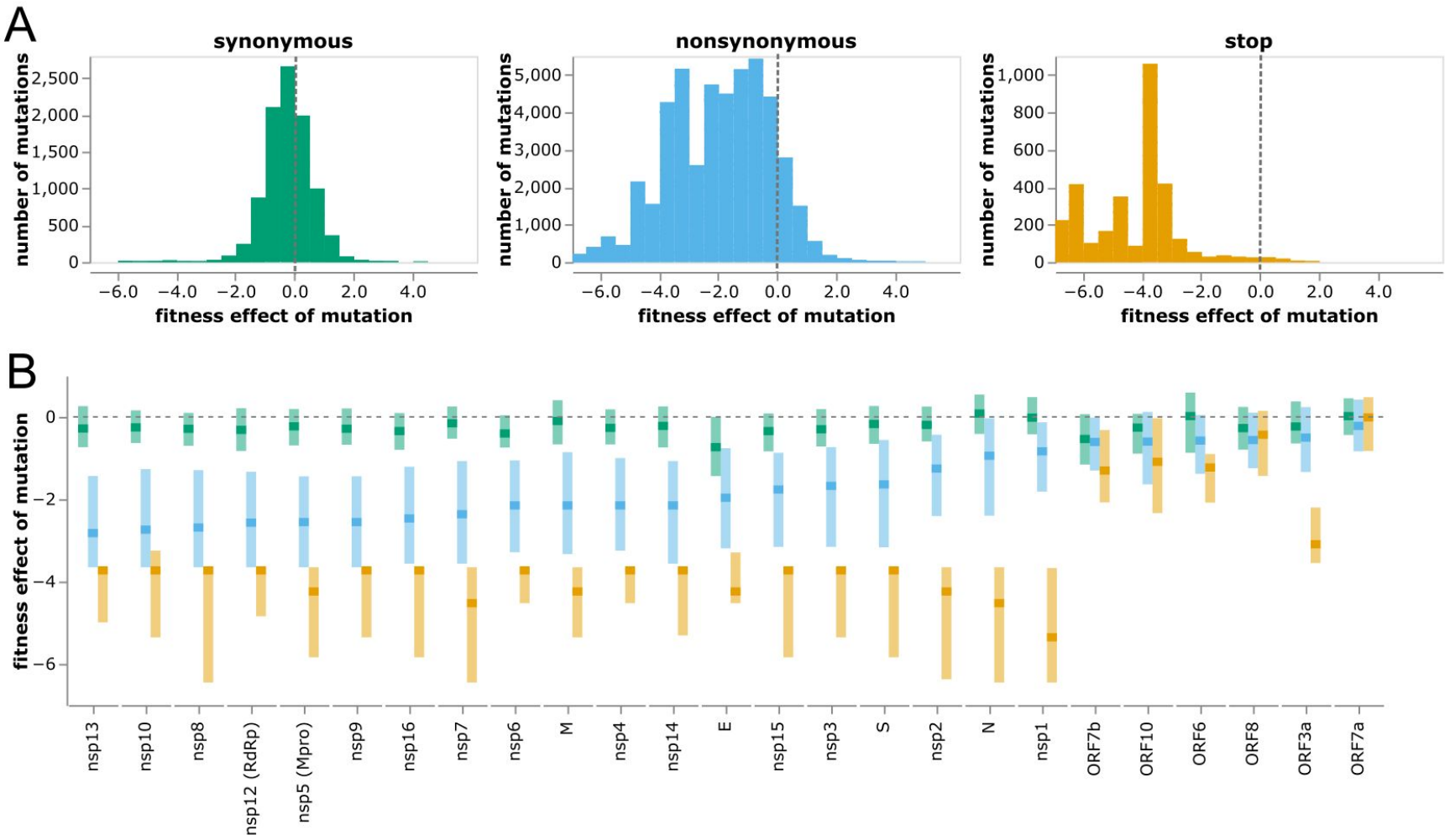
Site specific fitness effect estimates across most of the SARS-CoV-2 genome

1. Use four-fold synonymous sites as proxy for expected counts for each mutation (we are in the process of modeling expected counts more carefully)
2. Compare observed counts to expected counts.
3. Aggregate mutations for each amino acid substitution
4. Compute fitness proxy as

$$\log(n_{obs}/n_{exp})$$

5. Branching process models: deleterious mutations with selection coefficient s (epsilon is the sampling rate)

$$\log(1 + s/\epsilon) = \log(n_{obs}/n_{exp})$$

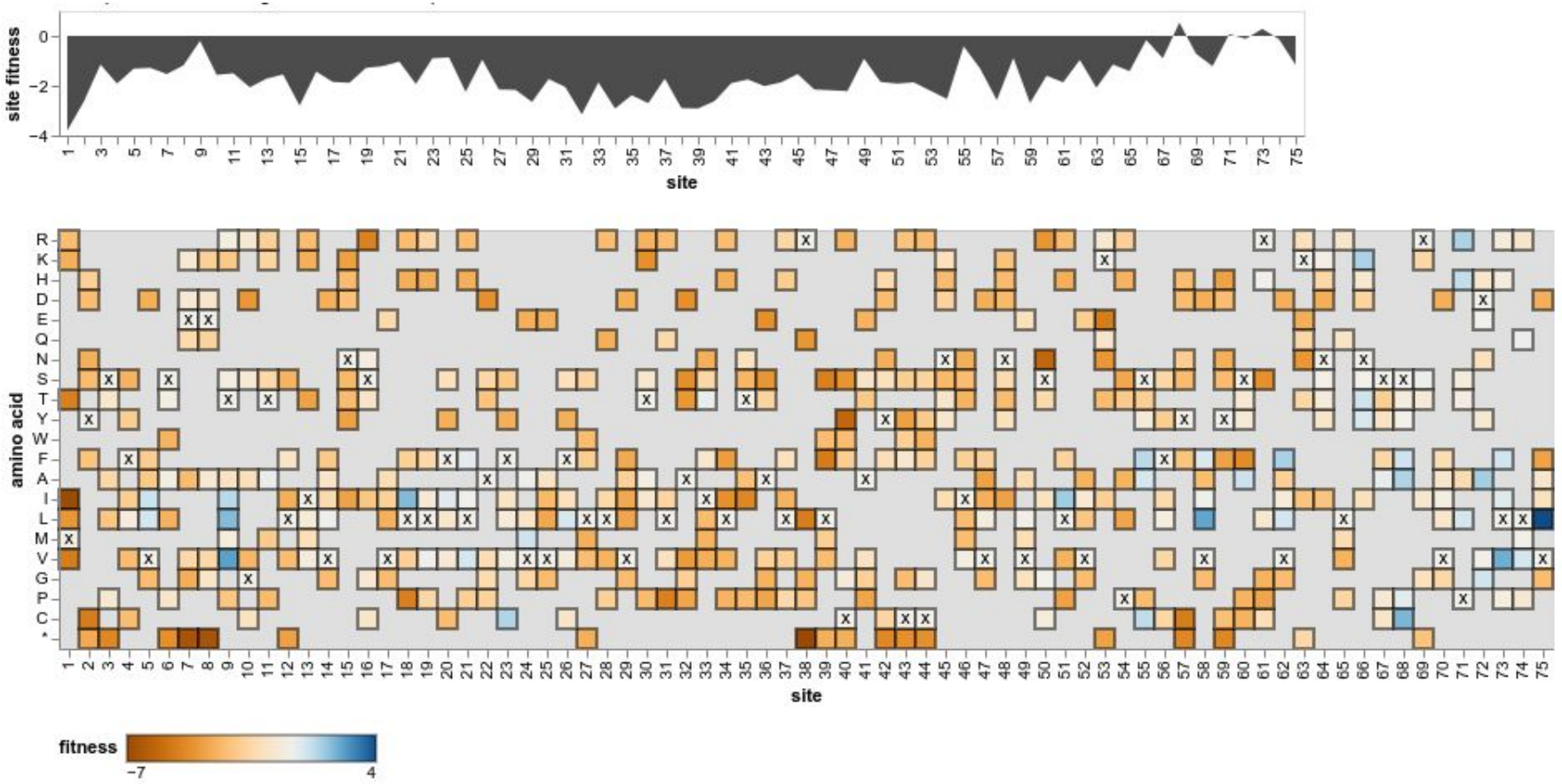


Bloom and Neher, 2023

Interactive plots at

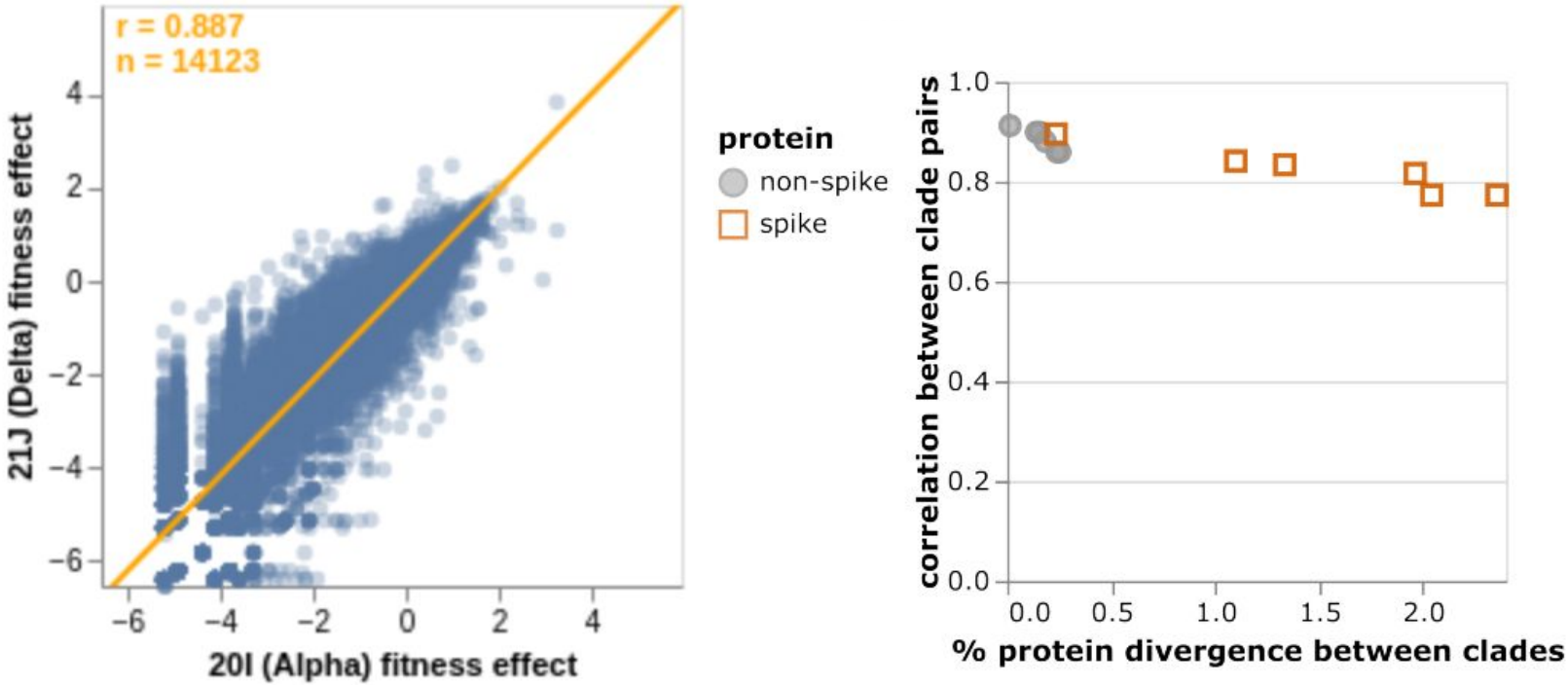
jbloomlab.github.io/SARS2-mut-fitness/

Fitness costs of mutations in the E protein



Bloom and Neher, 2023

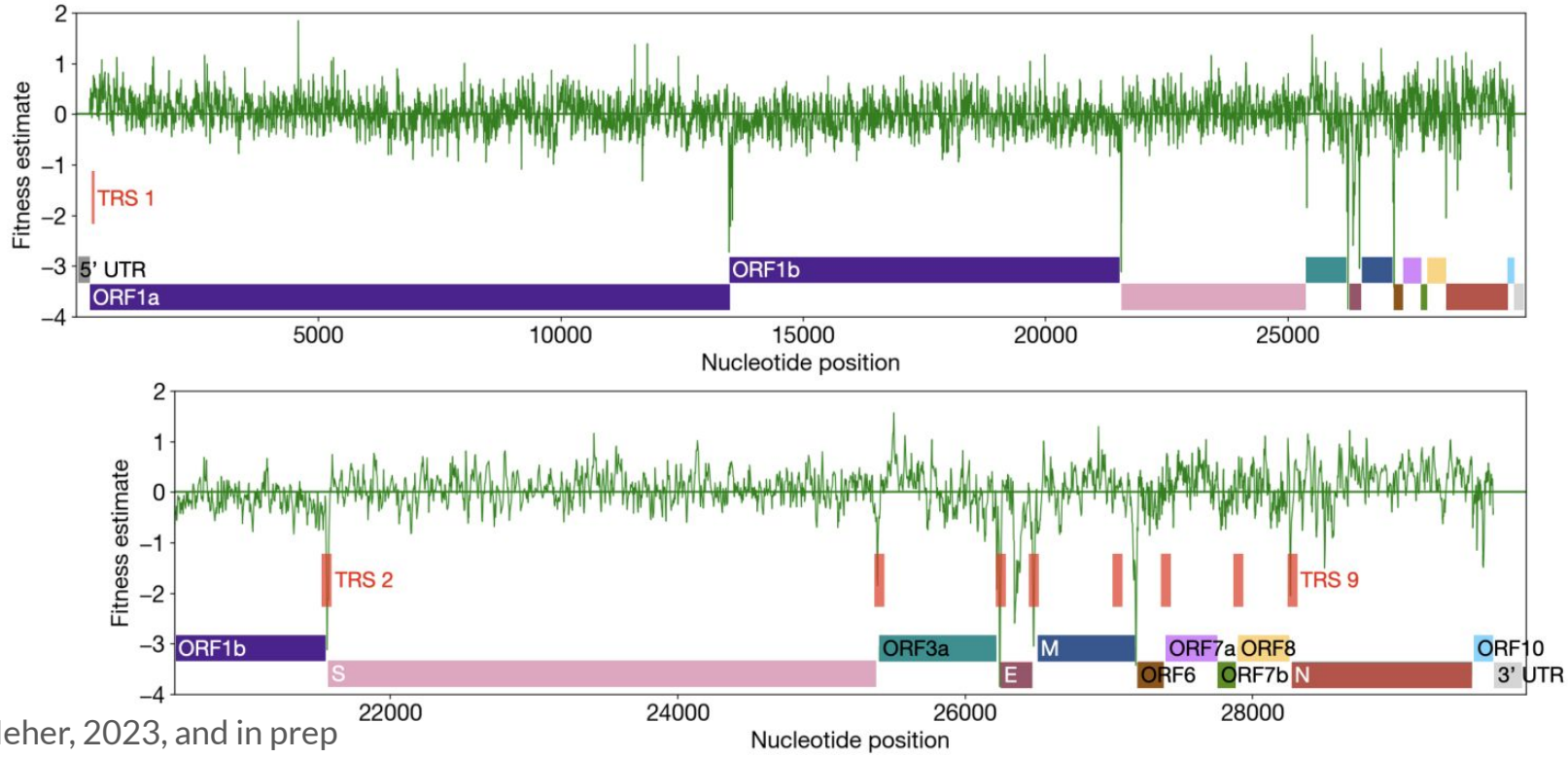
Concordance decreases with divergence – epistasis



Bloom and Neher, 2023

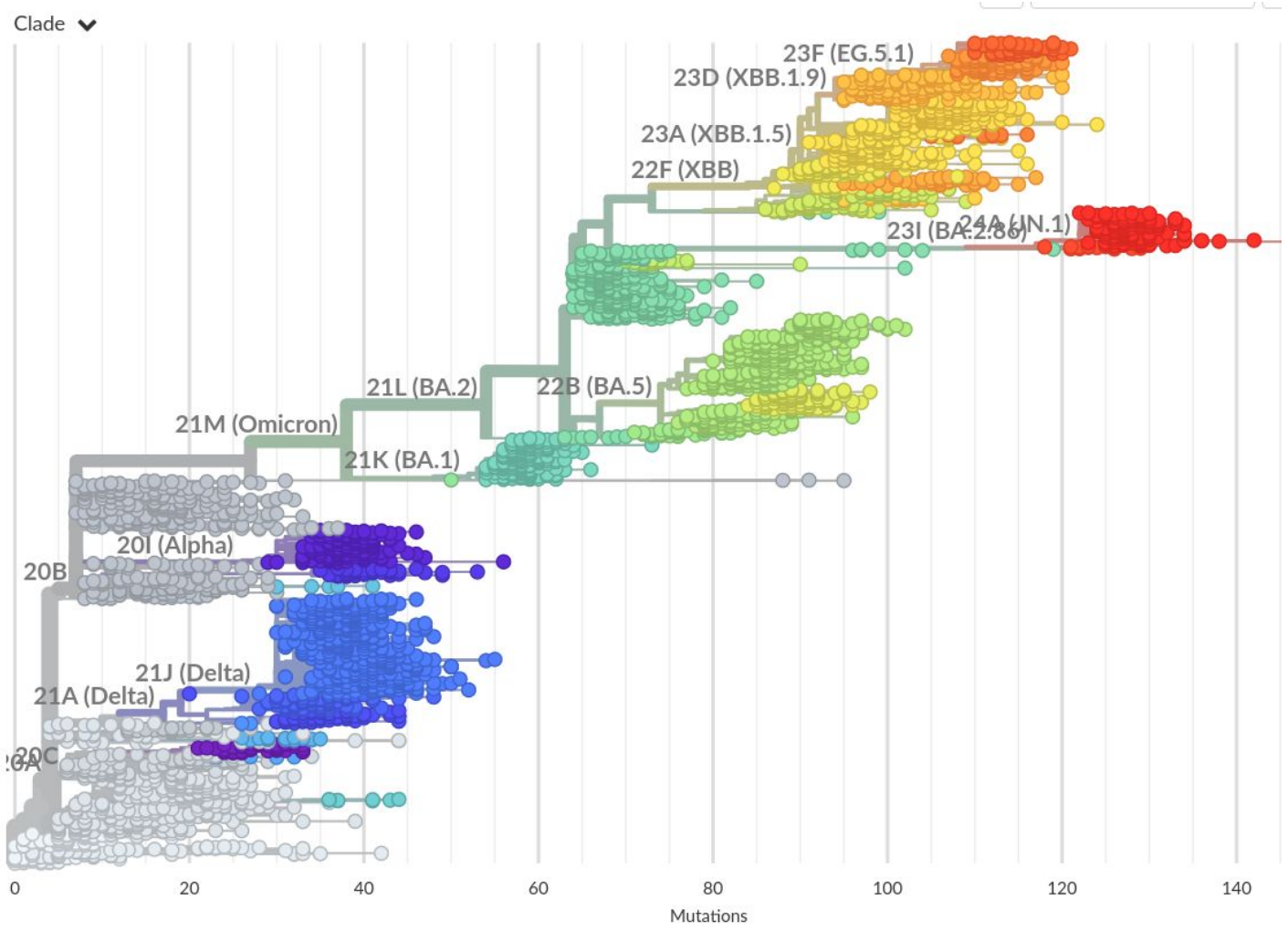
Selection beyond the coding sequence

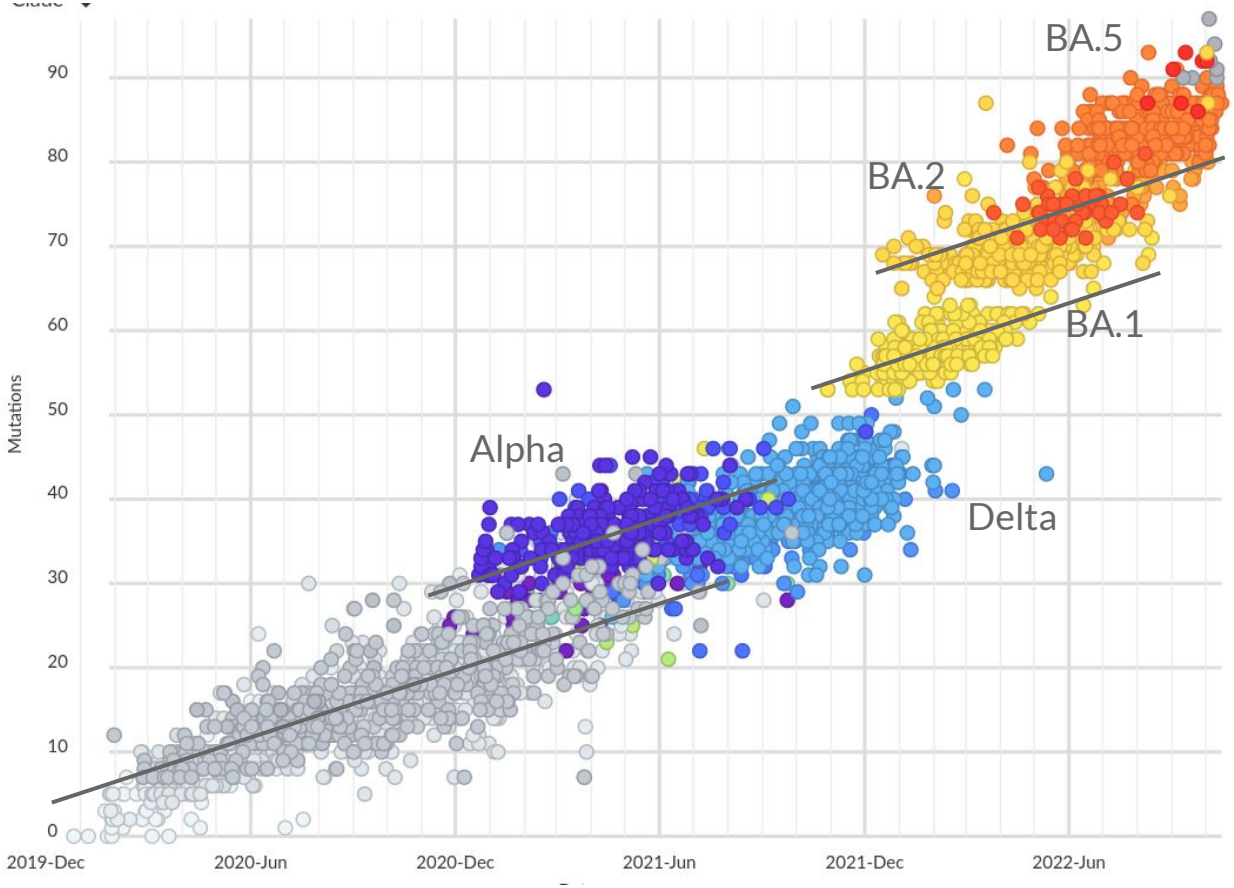
- In coding regions, constraints on nucleotide sequence can only be seen at synonymous sites
- But there are sufficiently many to detect conserved regions
  → only a minority of the genome is strongly constrained



Bloom and Neher, 2023, and in prep

From mutations and purifying selection to divergence…

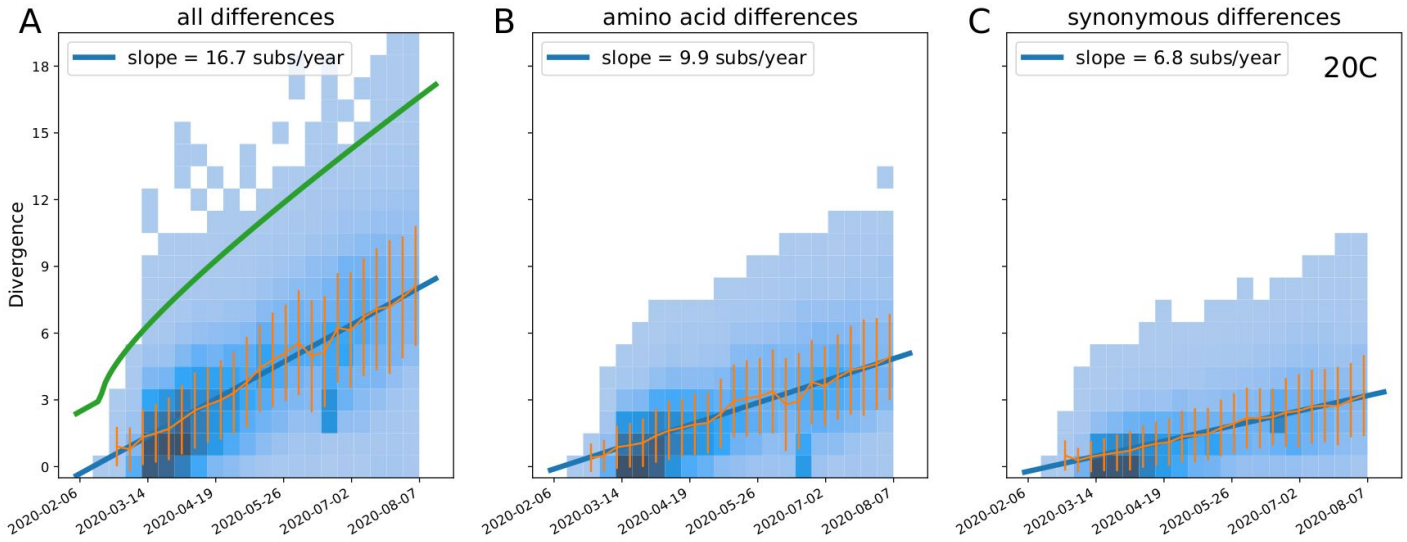

Clade
23F (EG.5.1)
23D (XBB.1.9)
23A (XBB.1.5)
22F (XBB)
24A (JN.1)
23I (BA.2.86)
22B (BA.5)
21L (BA.2)
21M (Omicron)
21K (BA.1)
20I (Alpha)
20B
21J (Delta)
21A (Delta)
20C
0
20
40
60
80
100
120
140
Mutations



- Rapid evolution
  (~30 changes per year)
  Coronaviruses were traditionally thought of as rather stable.
- Stepwise dynamics:
  - Slow within variants
  - Rapid jumps in between
- Rapid jumps possibly due to chronic infections; many hallmarks of adaptation

See also Duchene et al, Hill et al.

Robust determination of within-Clade evolutionary rates



- Use sequences that have all lineage defining mutations (removes problematic sequences)
- Linear regression on the number of **additional** synonymous or amino acid mutations
  (shared ancestry is a minor problem since most clades have approximately star like phylogenies)
- Straightforward to do for different proteins, regions etc.

→ Amino-acid and synonymous rate estimates for each clade

Neher, 2022

Amino acid rates within clades declined with time

Within vs Backbone rates:

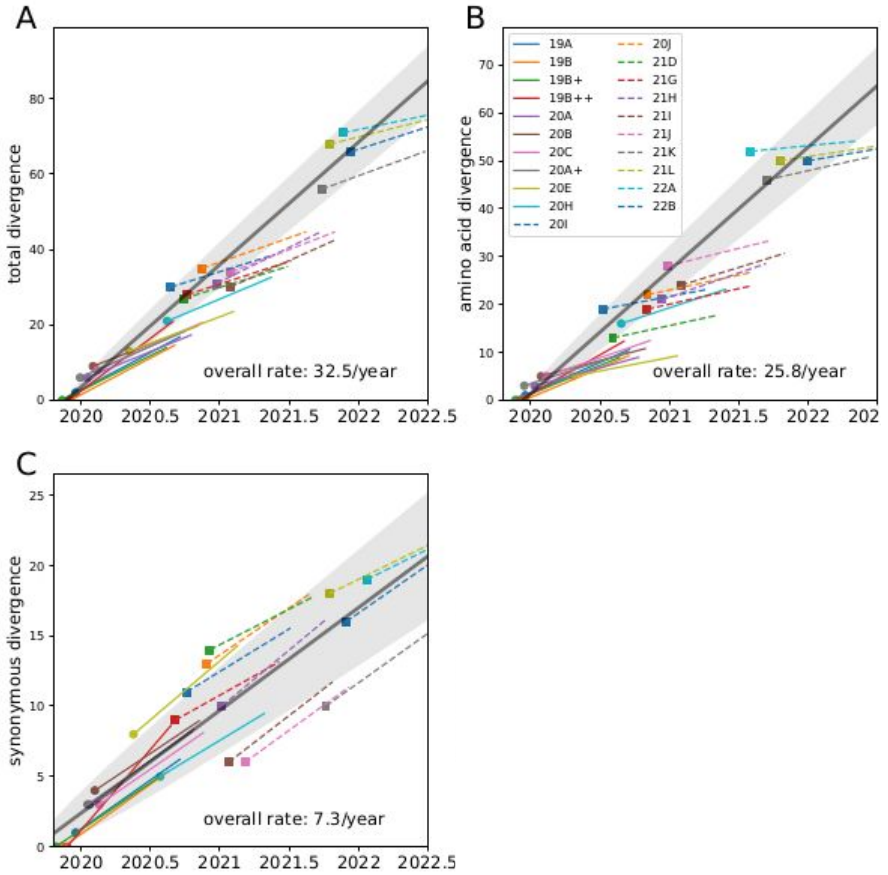
- All clades compatible with a common backbone rate
- Within clade rates are systematically lower

Synonymous rate:

- All variants roughly 6 changes per year
- Very little variation
- Overall rates similar, around 7 changes/year

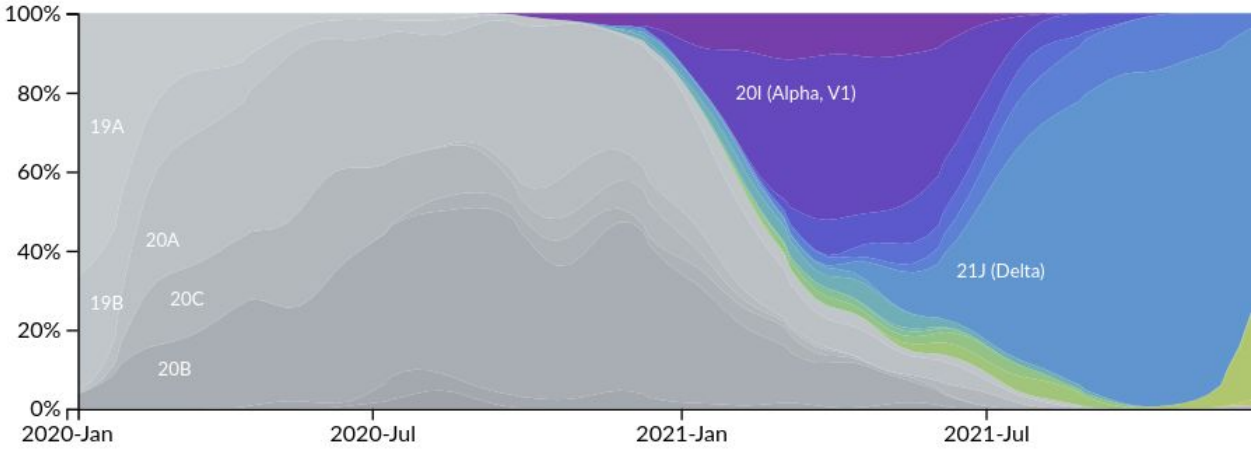
Amino acid rate:

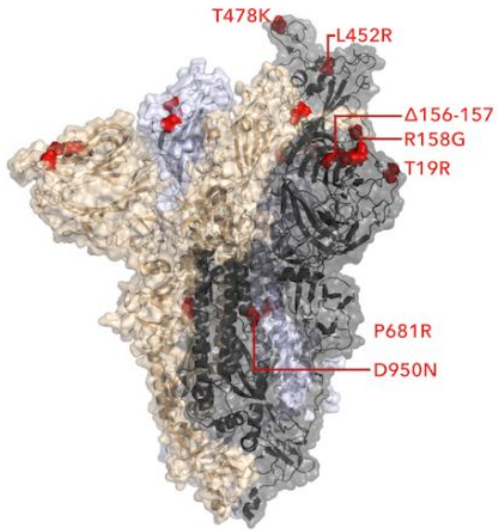
- The overall rate from clade to clade is much higher than the within clade rate



Neher, 2022

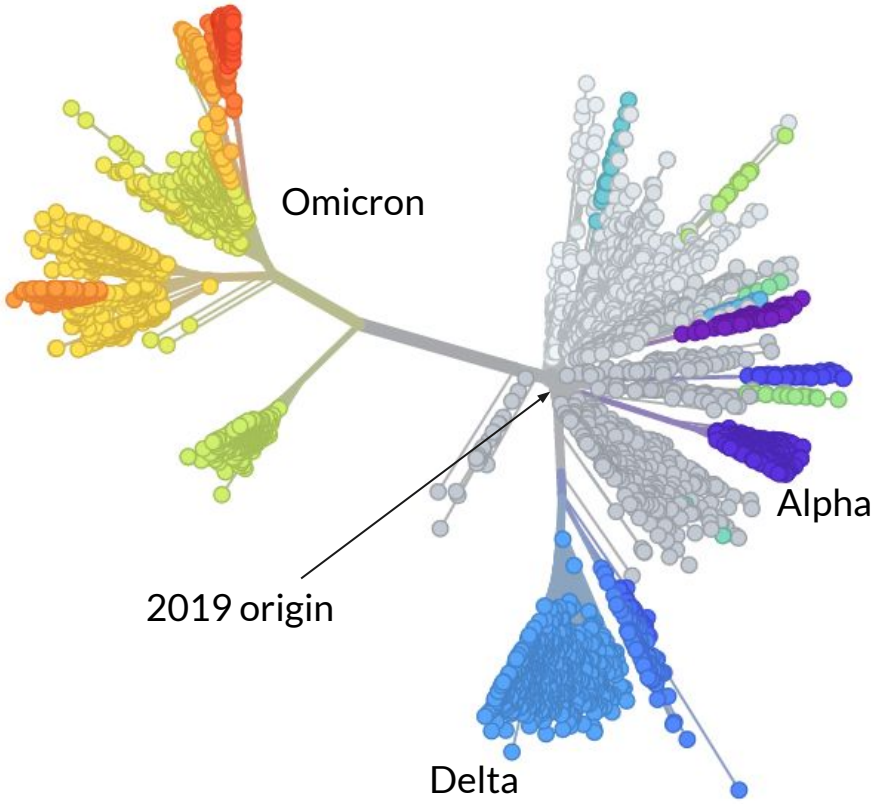
2021: Delta – global dominance



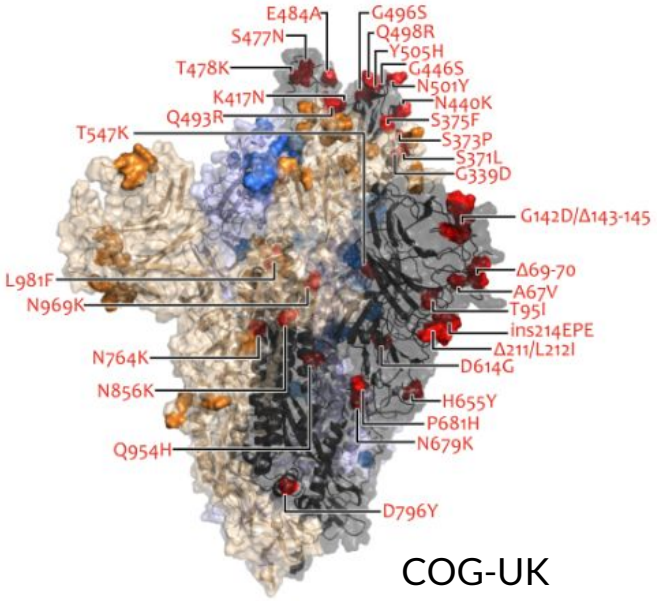


- Several mutations that increase transmissibility
- Increased severity of disease
- Moderately reduce immune recognition (less than Beta)

November 2021: Omicron



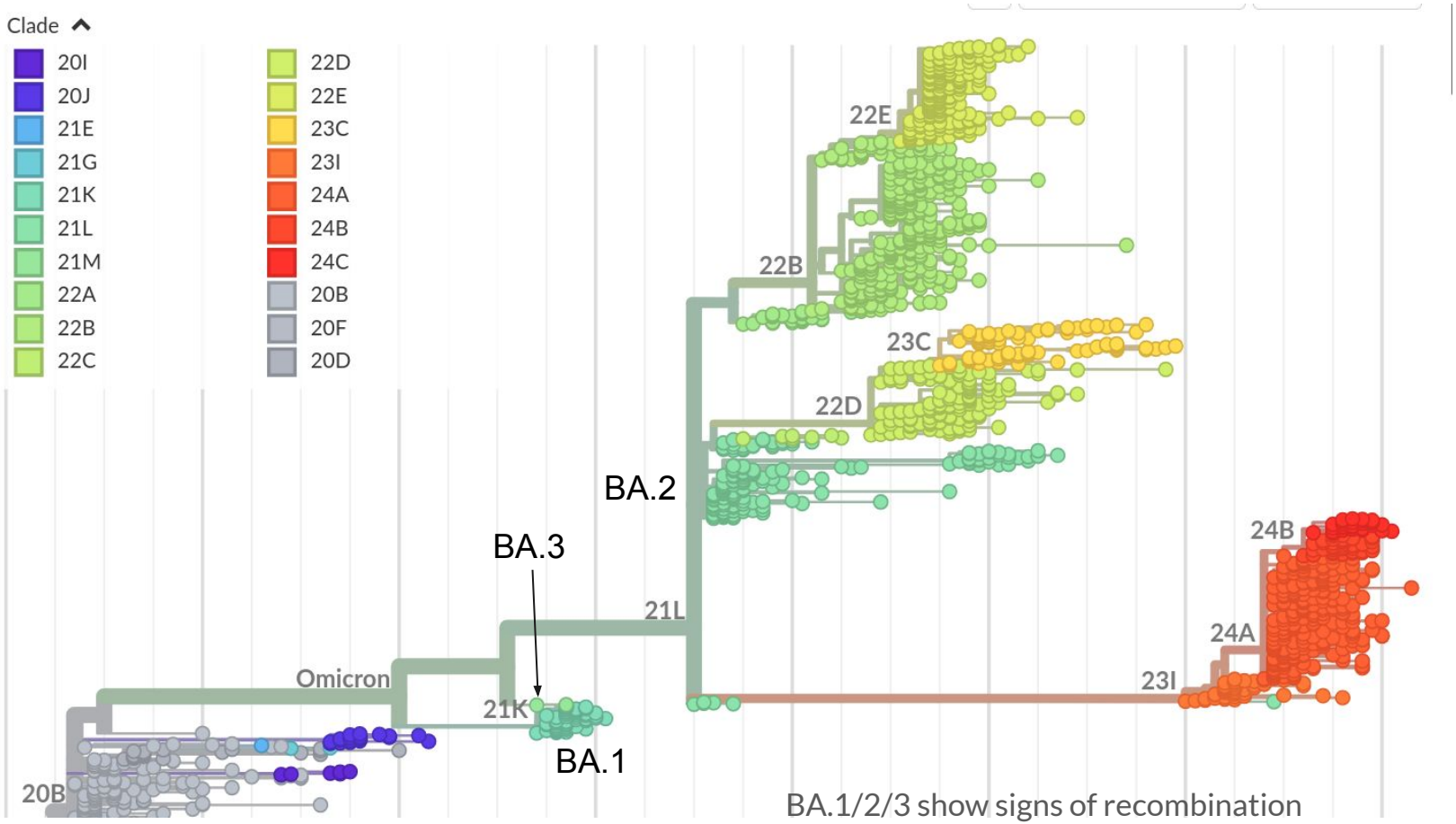
- Heavily mutated sister variant of previous VOCs
- Several distinct variants
- High rate of reinfections
- Very rapid spread



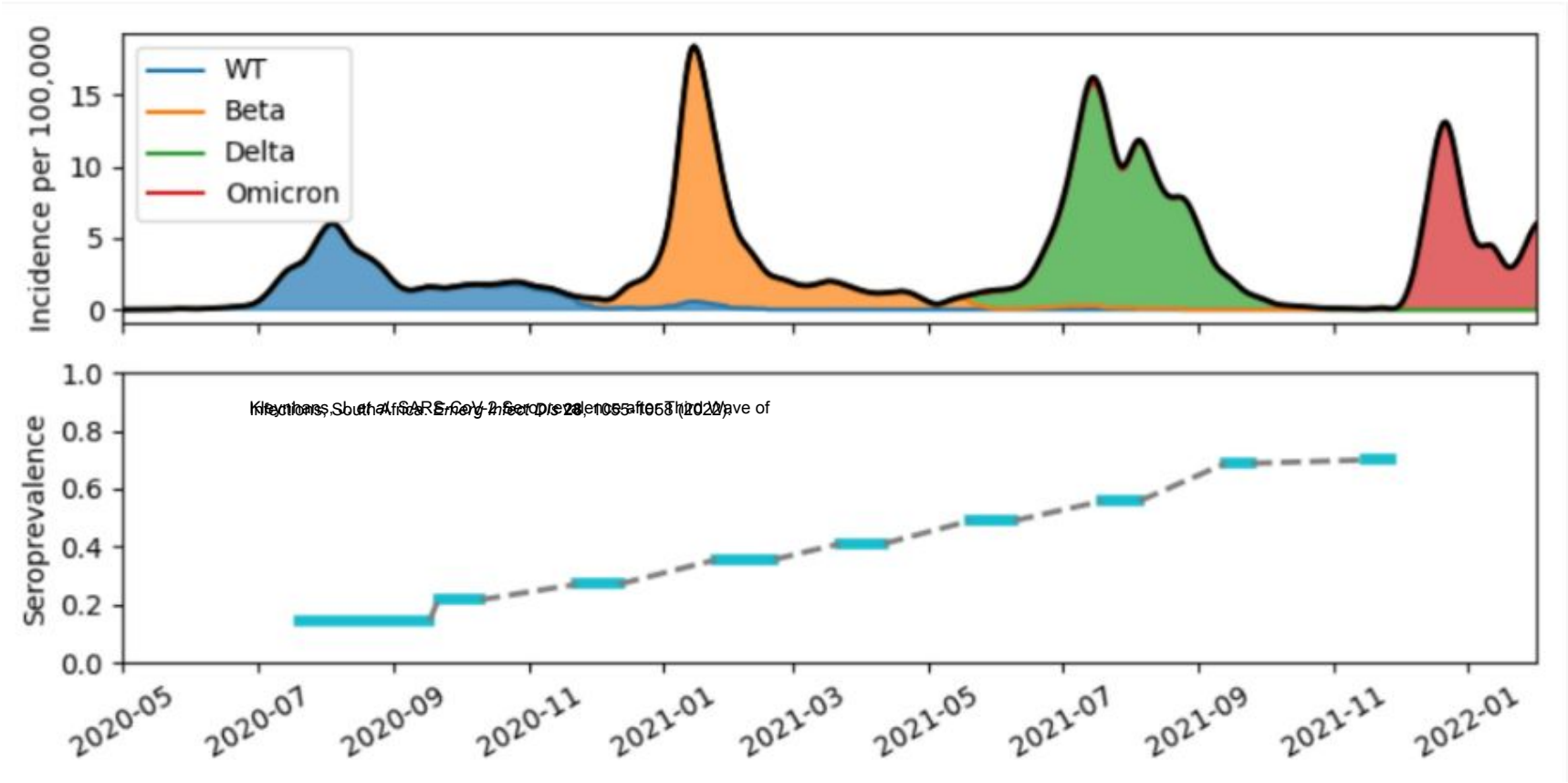
COG-UK

→ Michael Desai's talk later this week!!

Early Omicron diversity

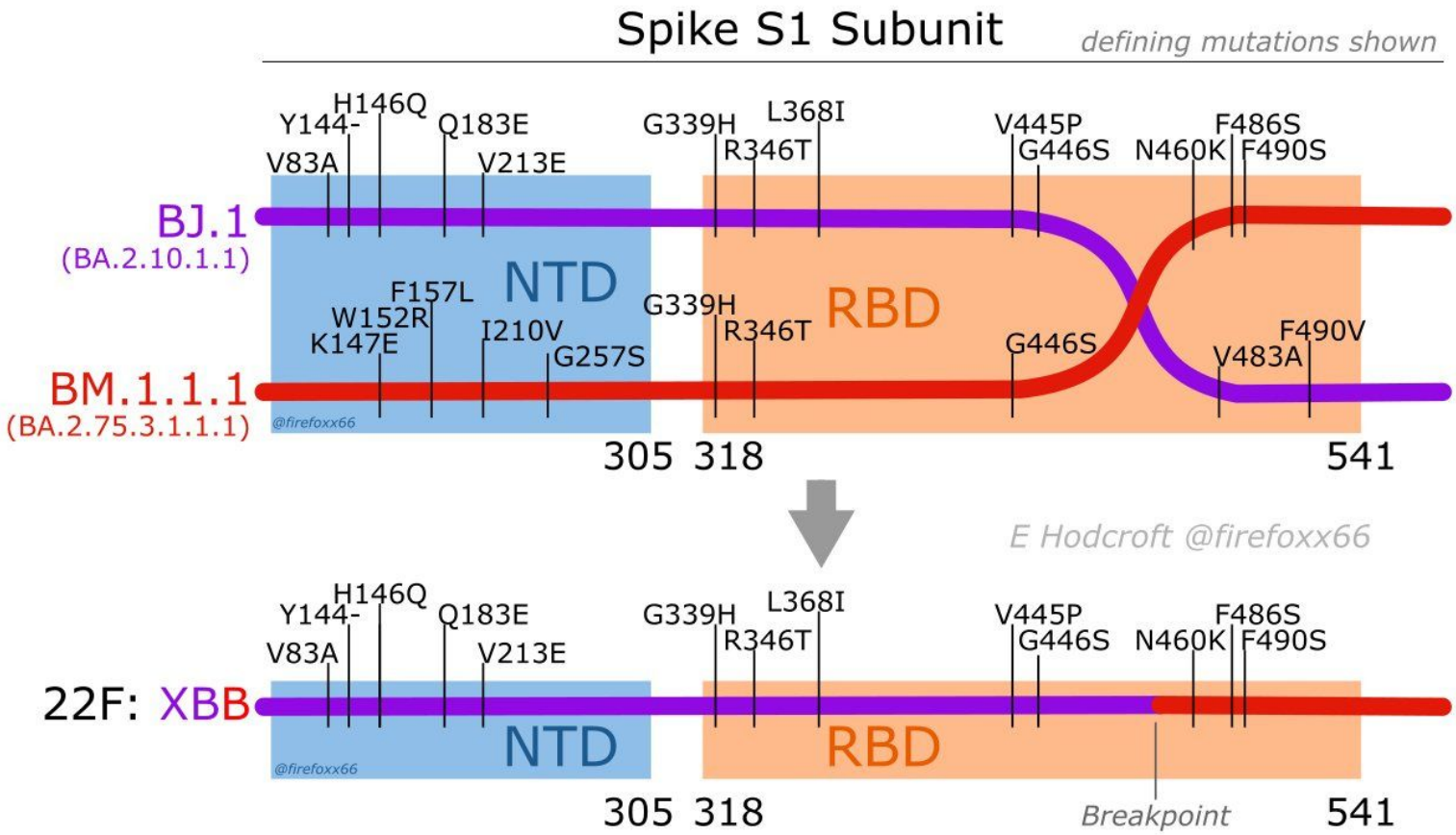


Seroprevalence in South-Africa

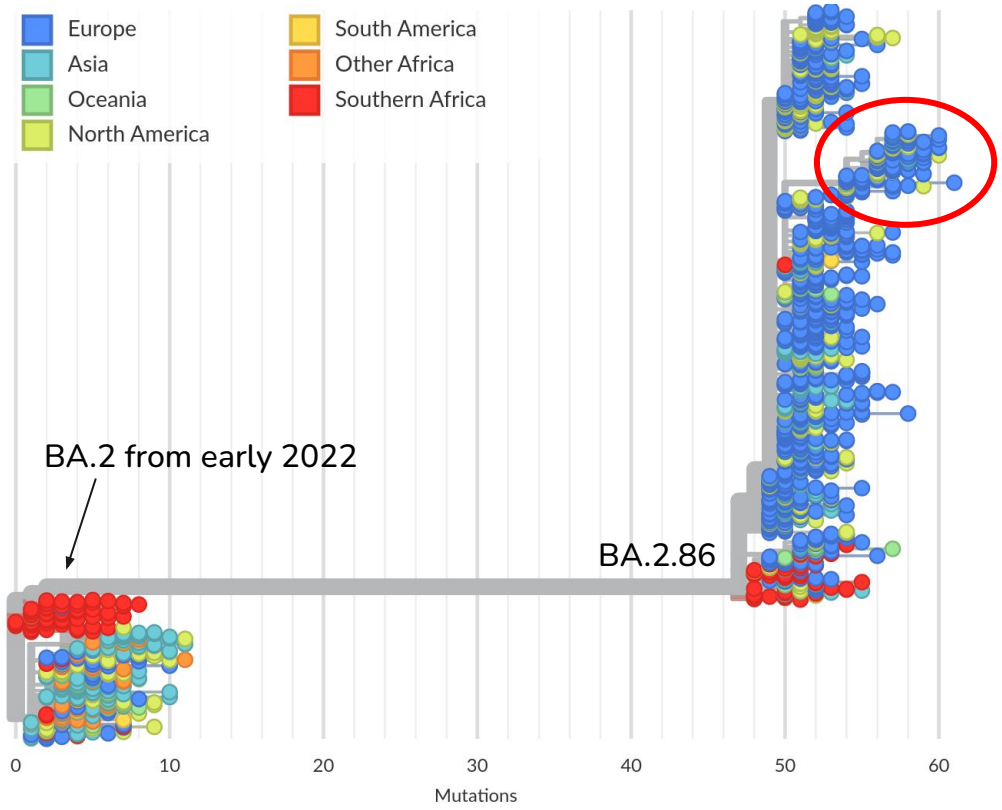


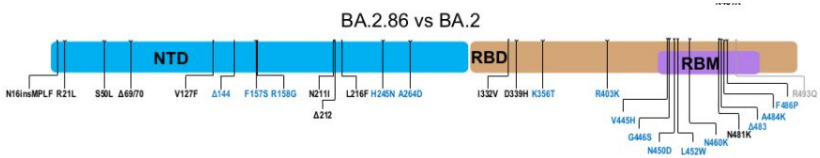
→ by end of 2021, most people were infected or vaccinated (outside of some parts of East-Asia)

XBB is likely a recombinant between two BA.2 descendents



Latest successful highly divergent variant: BA.2.86

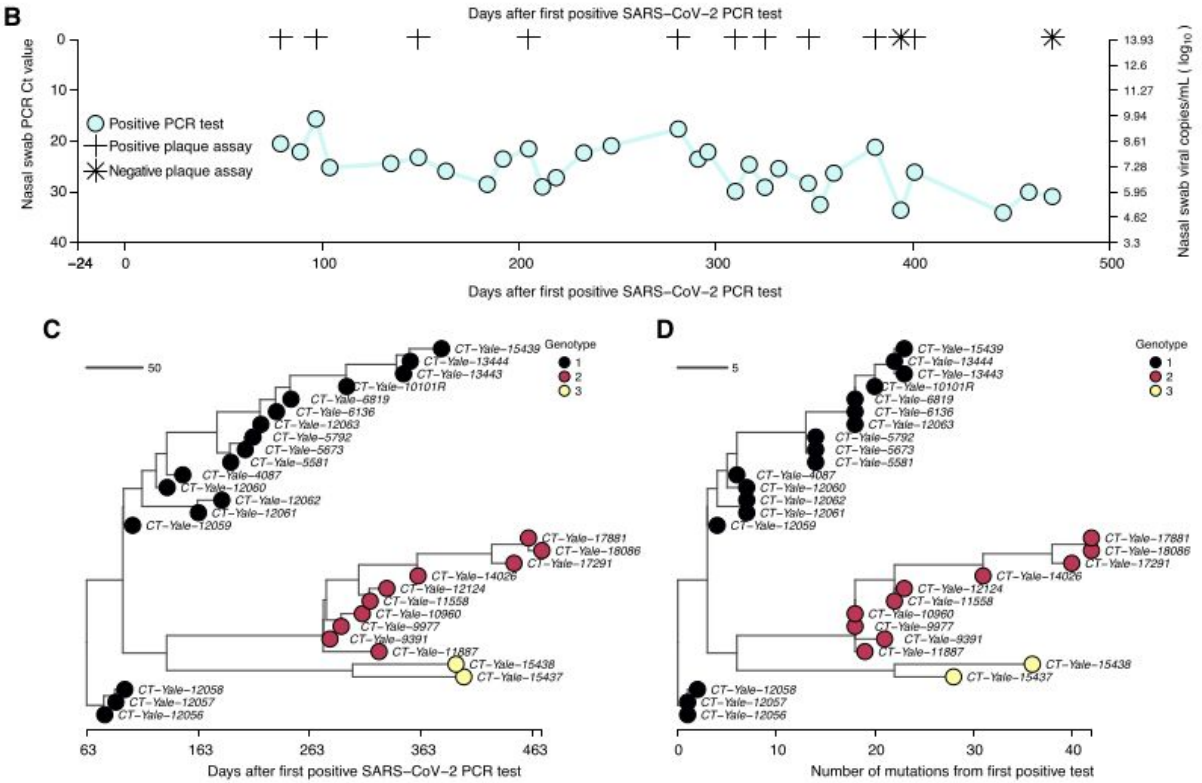




- Chronic infections are common in people living with unsuppressed HIV
- Rapid humoral immune escape of SARS-CoV-2 in many such individuals
- Emerged in April 2023, initially slow growth
- Sub-variant JN.1 rapidly took over in late 2023

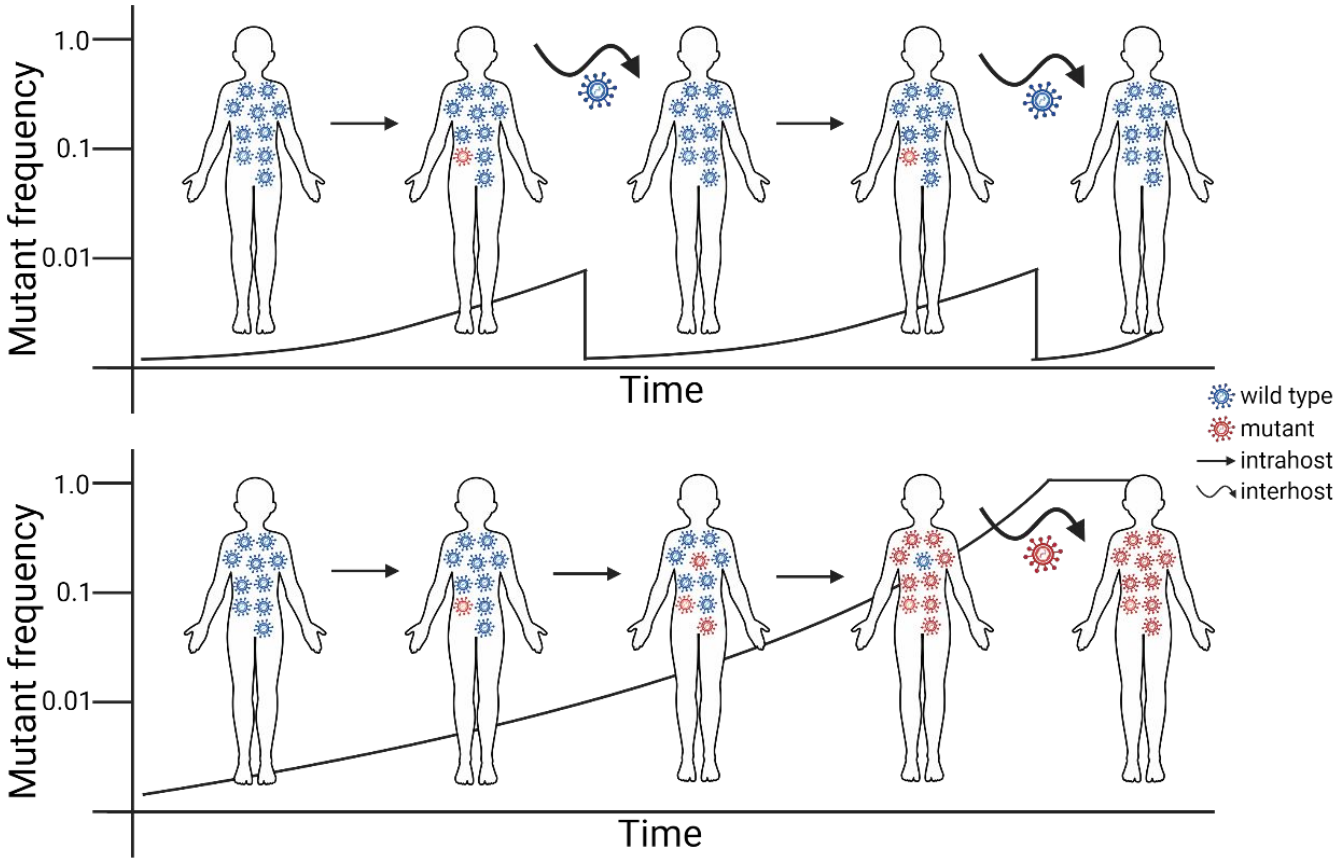
Khan et al, 2023

Emergence of VOCs: probably chronic infections



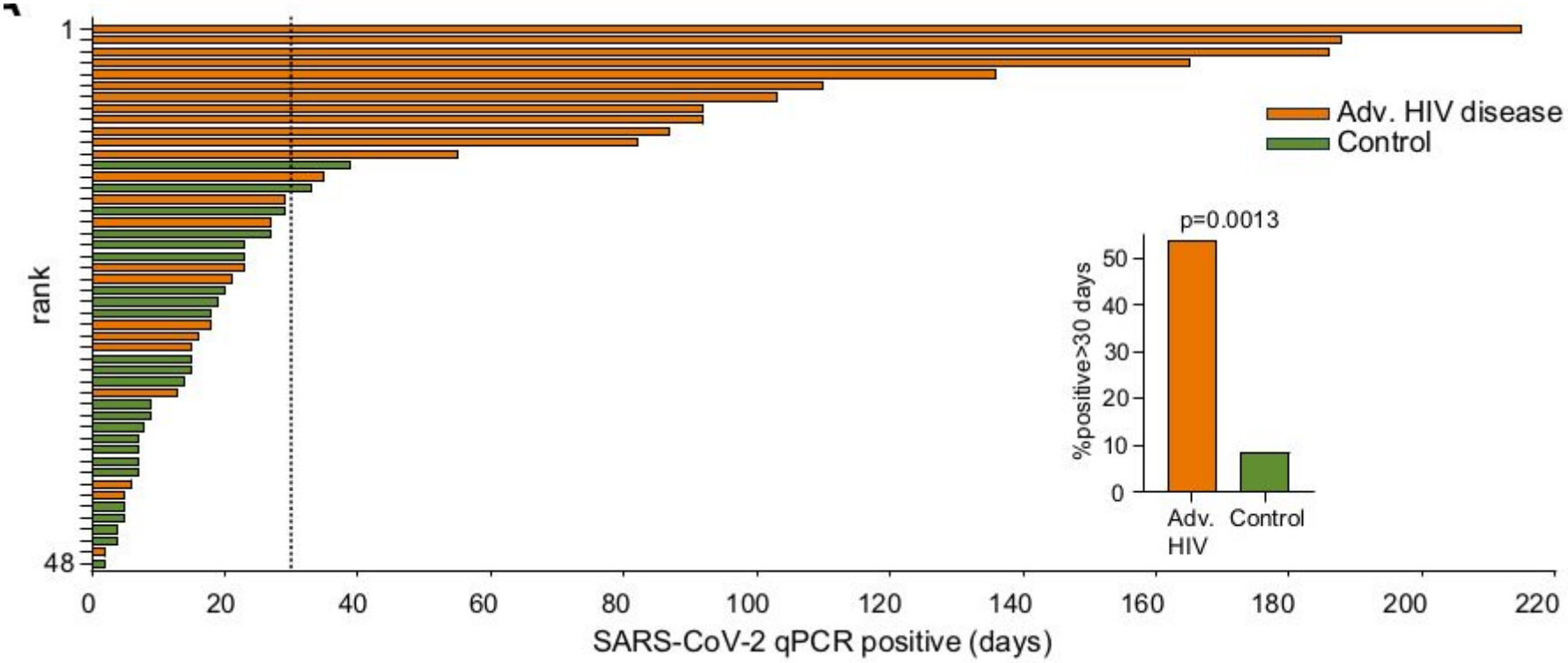
Chaguza et al, 2023

Adaptation is more efficient in chronic vs acute infection



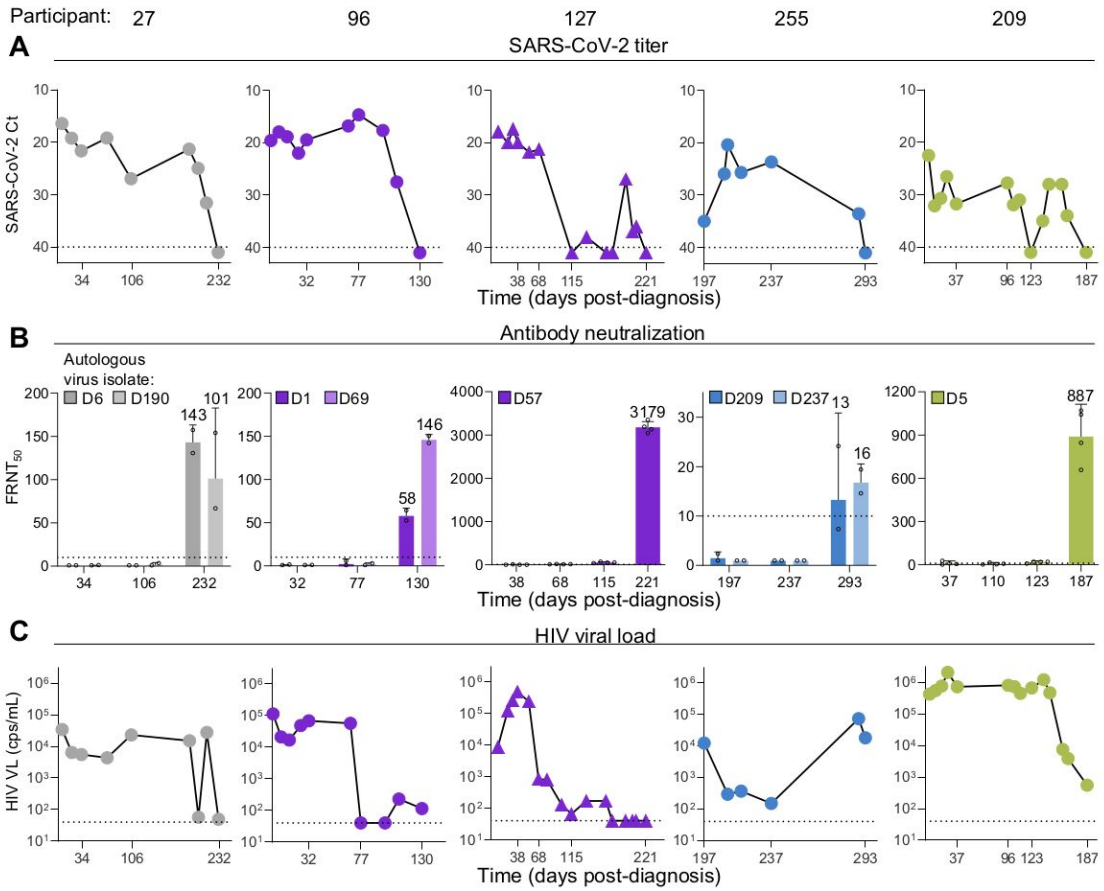
Sigal et al, in prep.

Persistent infections are common in people with advanced HIV

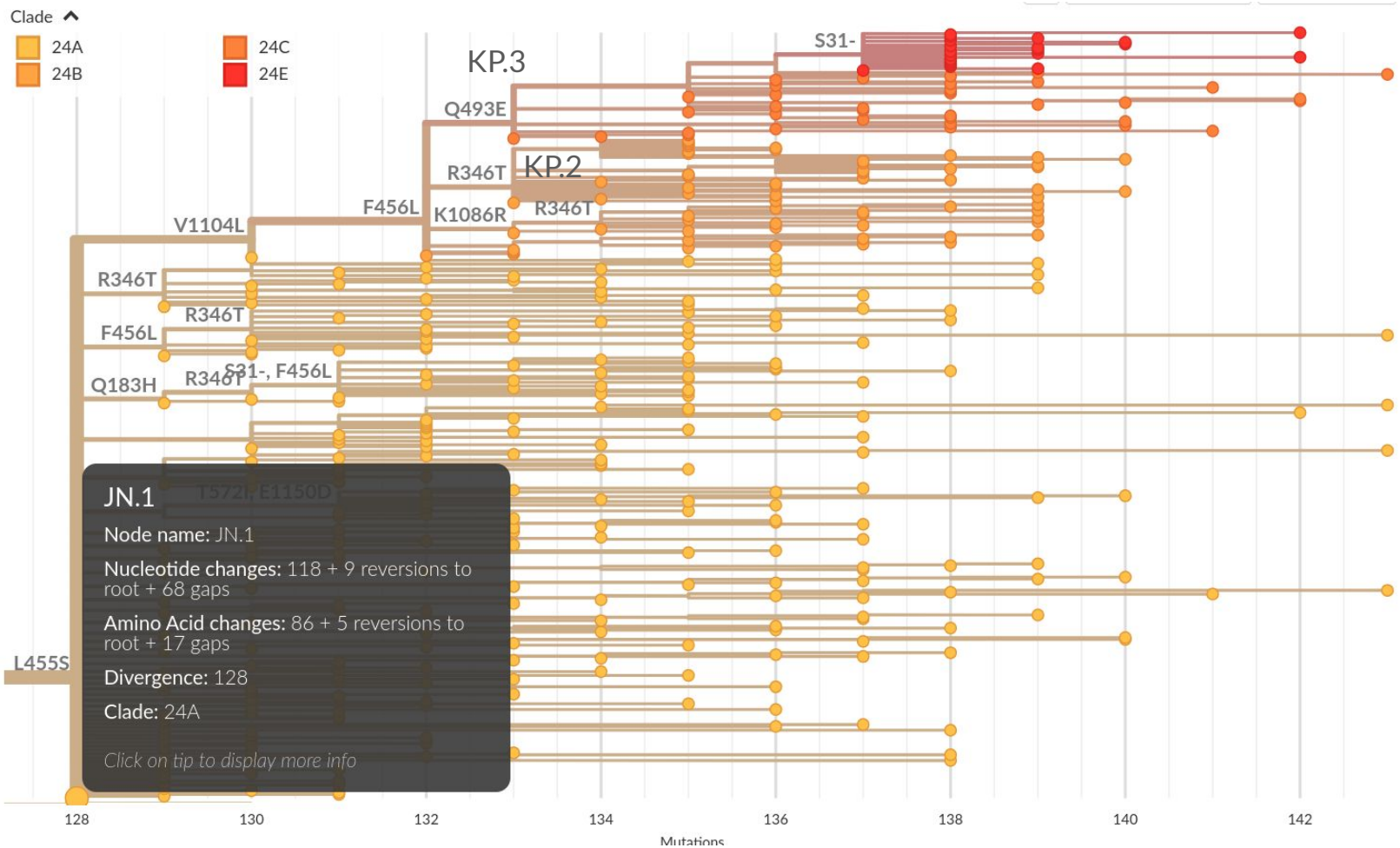


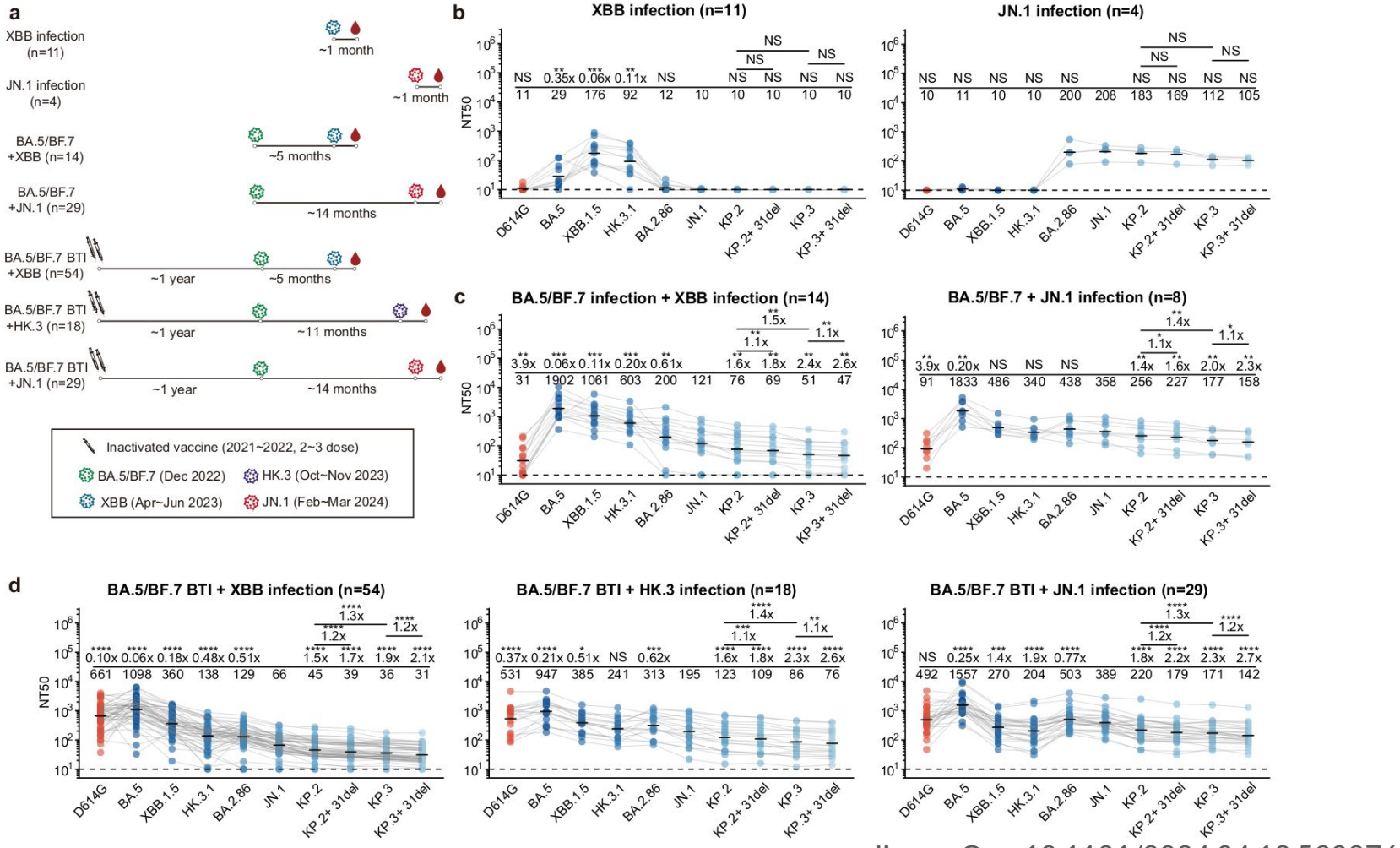
Karim, … Sigal. Nature Comm 2024

SARS-CoV-2 is cleared with HIV suppressed, AB titers go up.



Karim, … Sigal. Nature Comm 2024


Clade
24A
24B
24C
24E
KP.3
KP.2
S31-
Q493E
R346T
F456L
K1086R
R346T
V1104L
R346T
R346T
F456L
S31-, F456L
Q183H
R346T
T572I, E1150D
L455S
JN.1
Node name: JN.1
Nucleotide changes: 118 + 9 reversions to root + 68 gaps
Amino Acid changes: 86 + 5 reversions to root + 17 gaps
Divergence: 128
Clade: 24A
Click on tip to display more info
128
130
132
134
136
138
140
142
Mutations



Jian, … Cao. 10.1101/2024.04.19.590276

Acknowledgements

- Jesse Bloom, Erick Matsen, Hugh Haddox, Georg Angehrn
- Cornelius Roemer & Ivan Aksametov
- Alex Sigal and team!
- Sequence data contributors around the world (shared via GISAID or INSDC)


Trevor Bedford,

Richard Neher,

Ivan Aksamentov,

Jennifer Chang,

James Hadfield,

Emma Hodcroft,

John Huddleston,

Jover Lee,

Victor Lin,

Cornelius Roemer,

Thomas Sibley

Nextstrain