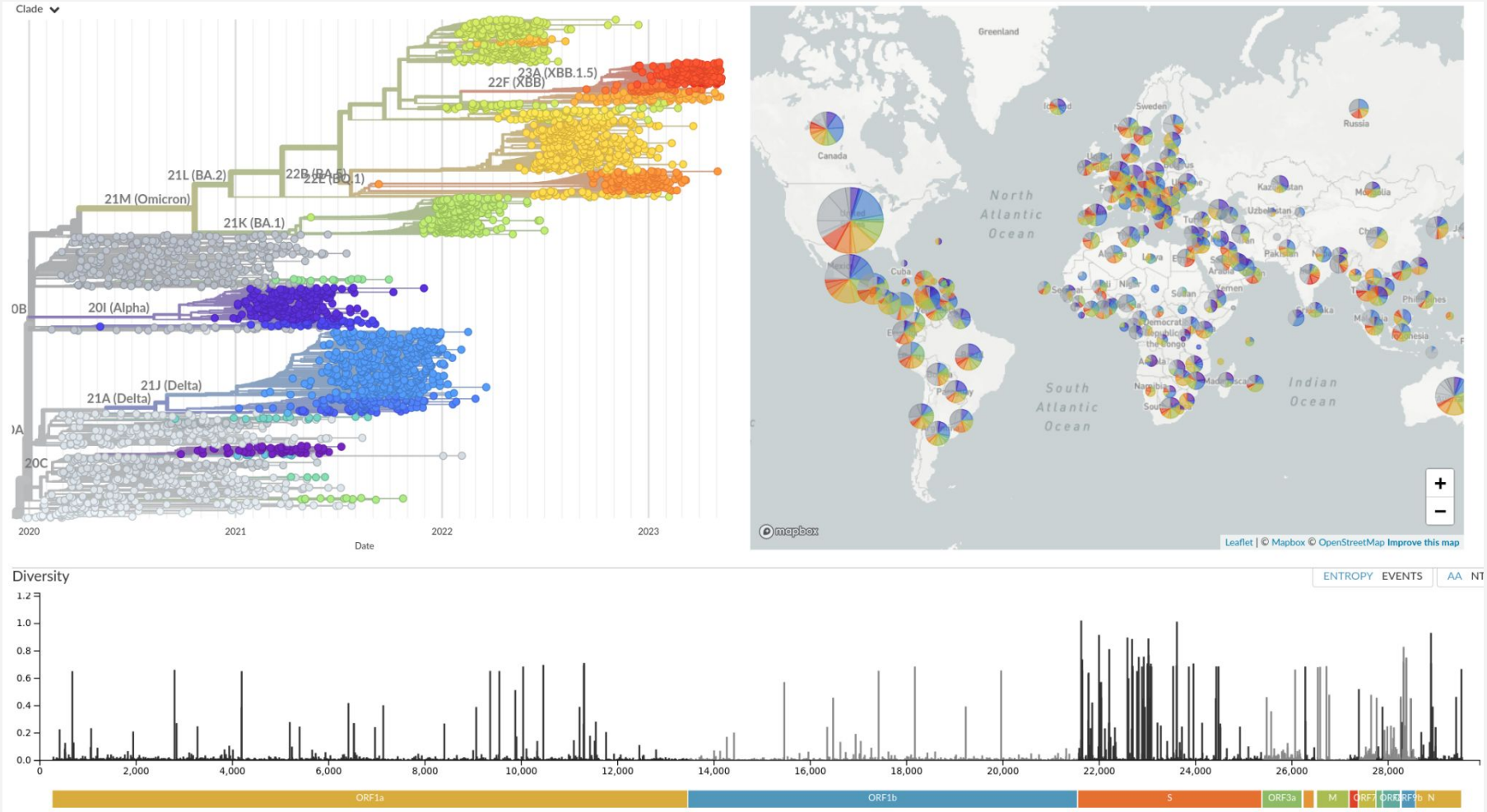
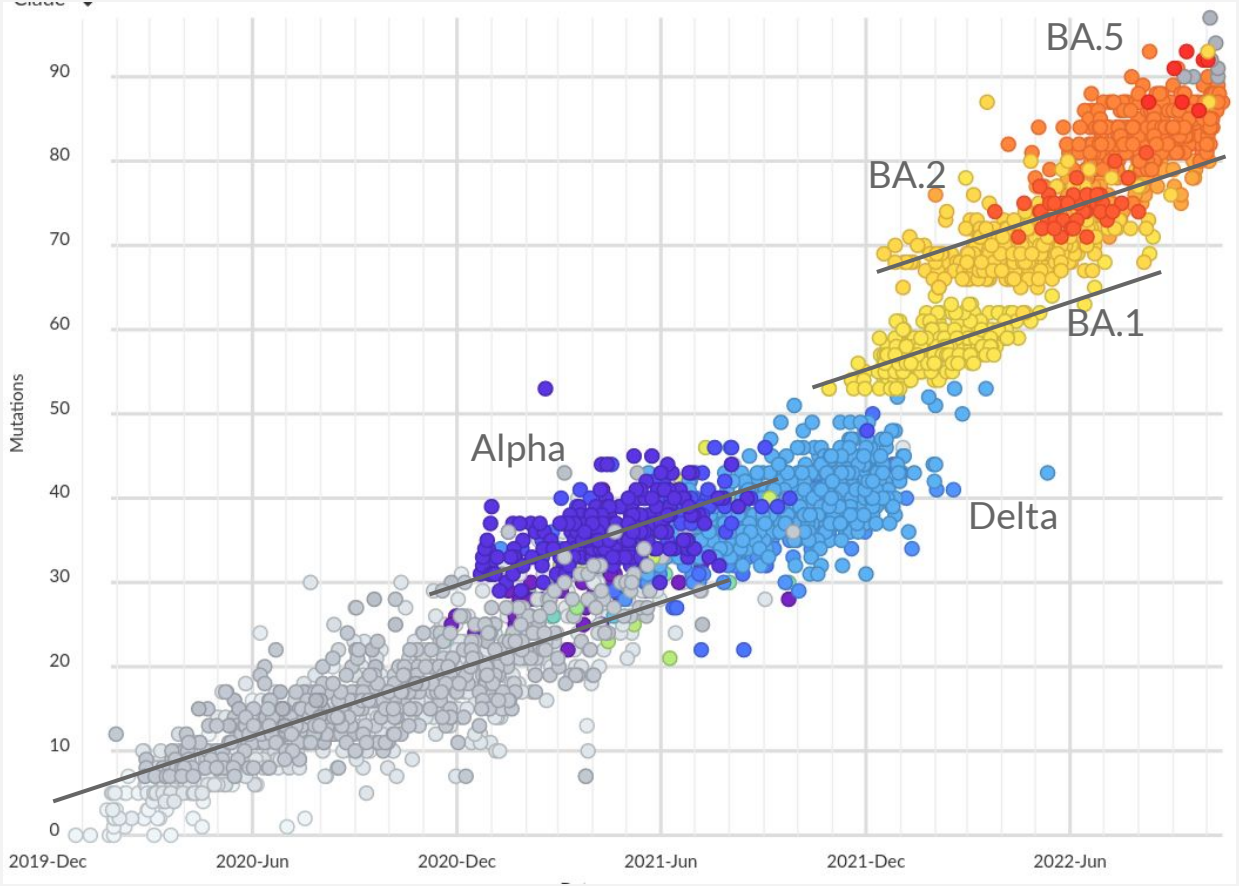
# Fitness landscapes and rates of SARS-CoV-2 evolution

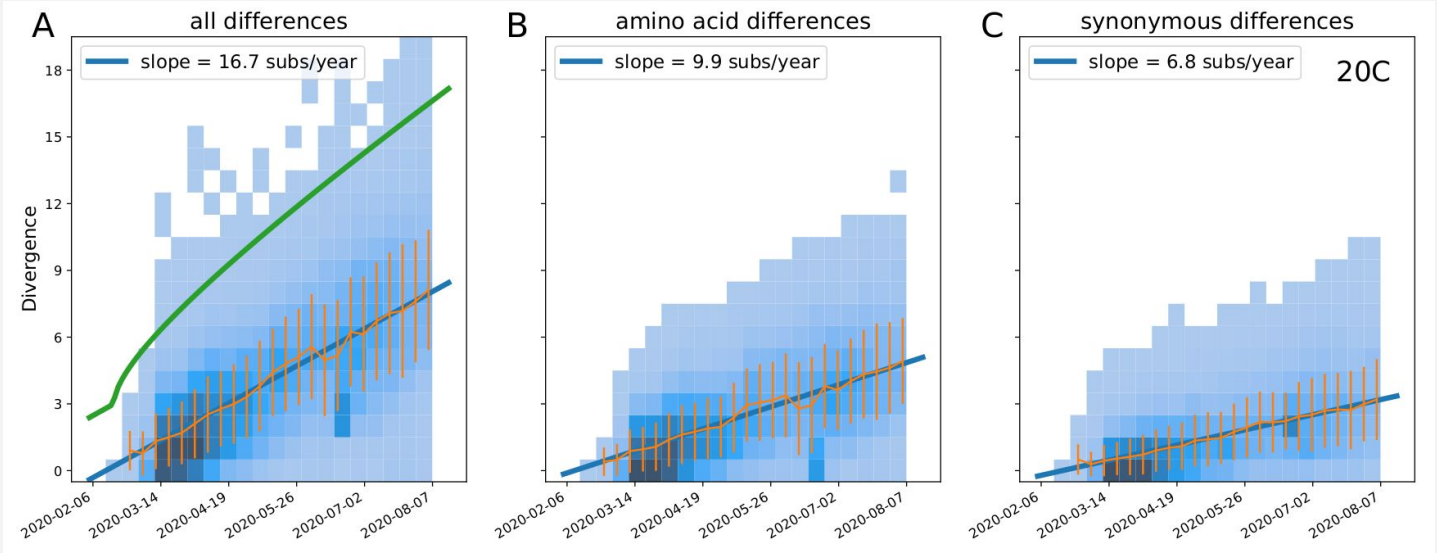

Clade
23A (XBB.1.5)
22F (XBB)
21L (BA.2)
22B (BA.2)
22E (BQ.1)
21M (Omicron)
21K (BA.1)
20I (Alpha)
21J (Delta)
21A (Delta)
20C
2020
2021
2022
2023
Date
Greenland
Canada
Russia
North Atlantic Ocean
South Atlantic Ocean
Indian Ocean
Leaflet | © Mapbox © OpenStreetMap Improve this map
Diversity
ENTROPY EVENTS
AA NT
ORF1a
ORF1b
S
ORF3a
M
N




- 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 possible due to chronic infections; many hallmarks of adaptation

See also Duchene et al, Hill et al.

# 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)

→ Amino-acid and synonymous rate estimates for each clade

# Amino acid rates within clades declined with time

**Within vs Backbone rates:**

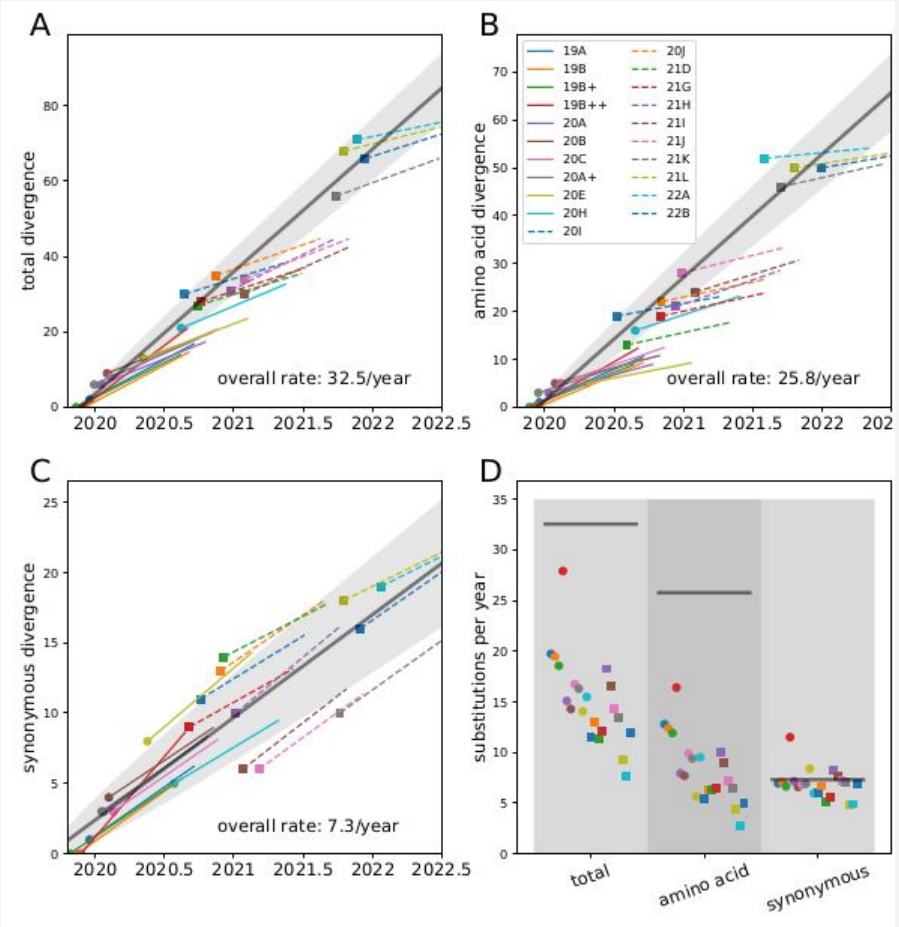
- All clades compatible with a common backbone rate
- Within clade rates are systematically lower

**Synonymous rate:**

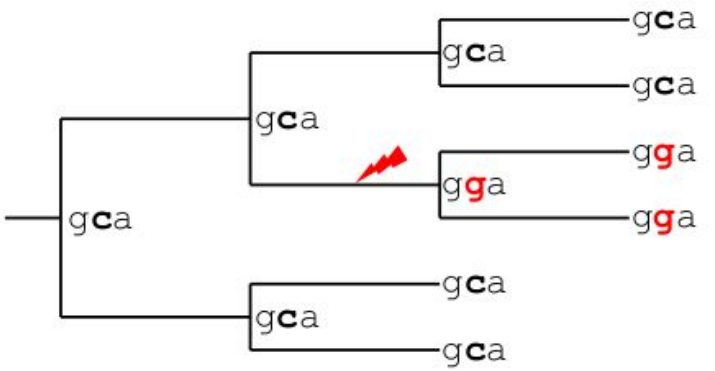
- All variants roughly 6 changes per year
- Very little variation
- Overall rates similar, around 7 changes/year

**Amino acid rate:**

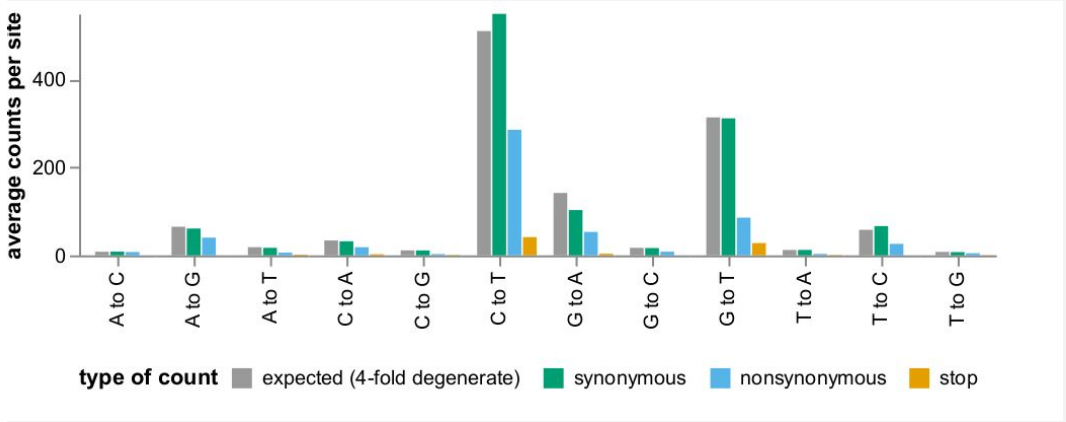
- Early variants evolved faster
- Large variation
- The overall rate from clade to clade is much higher than the within clade rate



# Site specific mutation rates and fitness landscapes

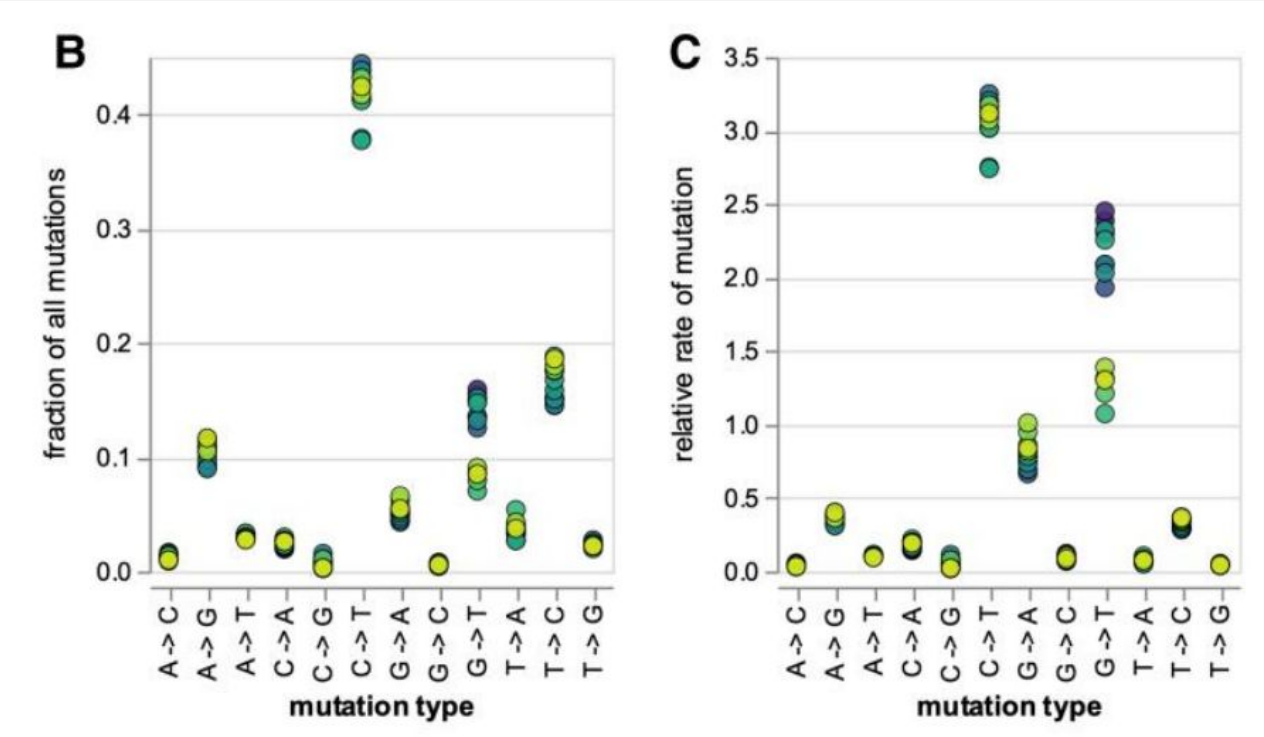



non-synonymous position




- Between 100 and 500 mutations per site! → allows quantitative estimation of site specific properties
- UShER (UC Santa Cruz) provides phylogenetic trees of millions of SC2 genomes

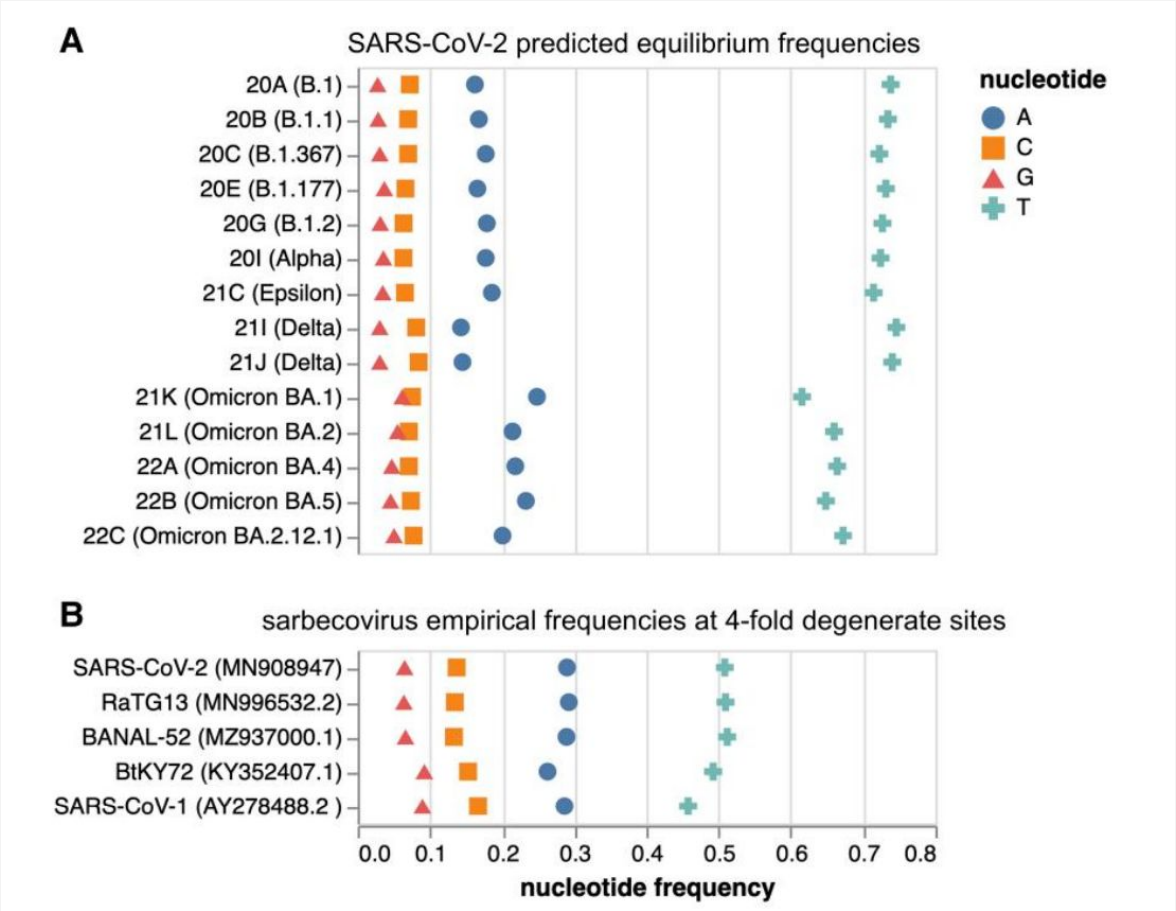
# Mutation rates and their clade dependence



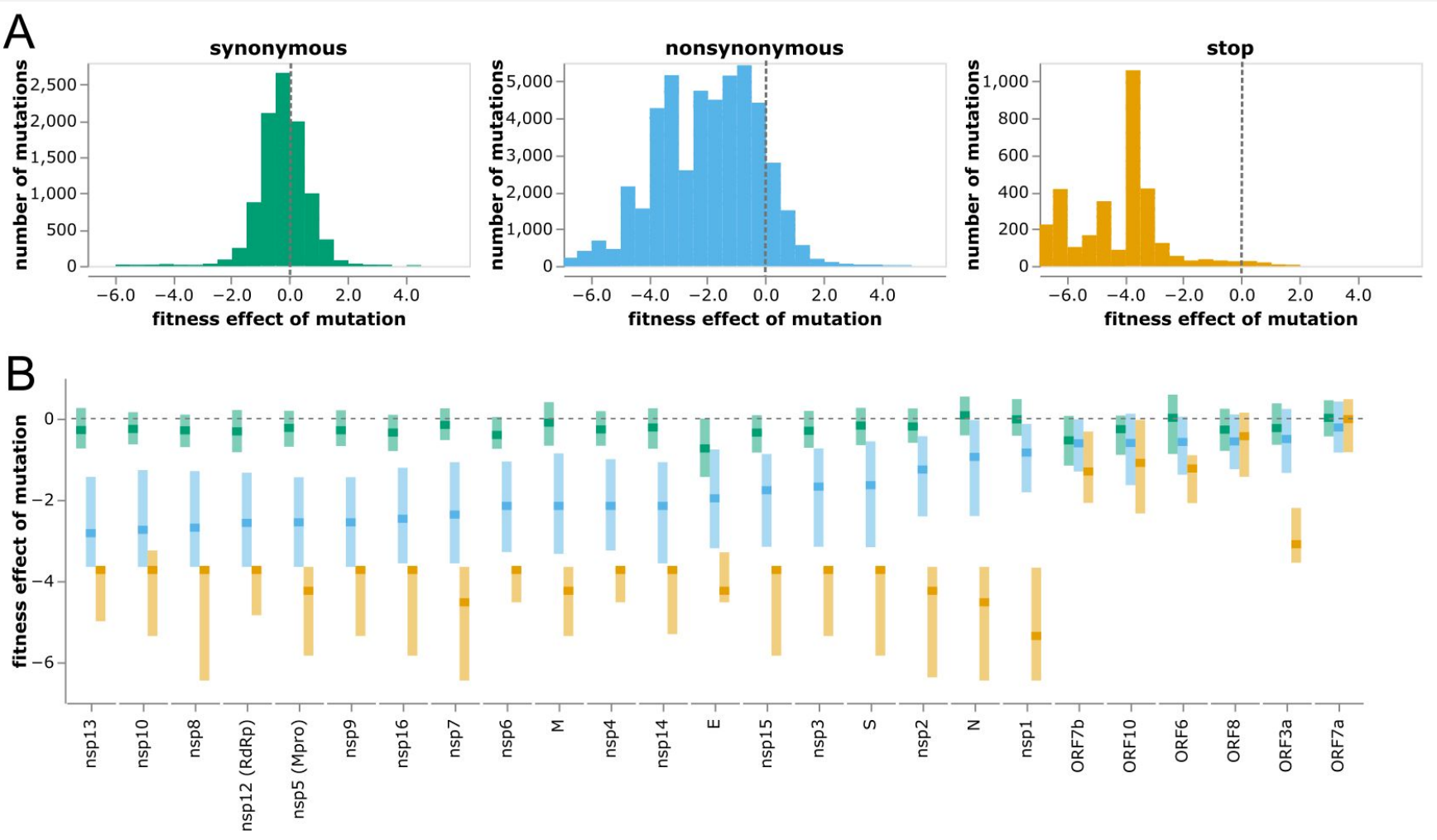


Bloom et al, 2023

# Mutation rates and their background dependence



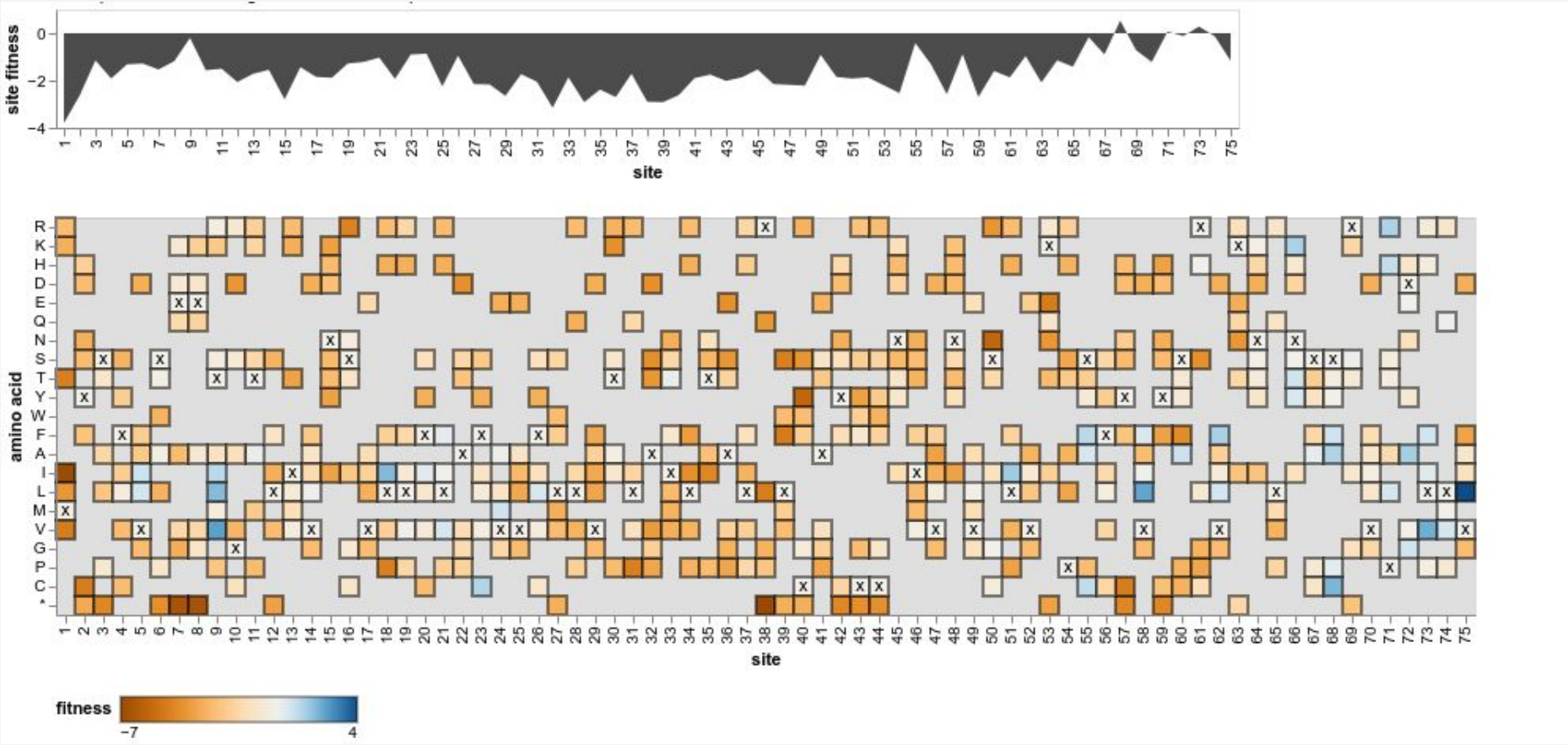

Bloom et al, 2023


A
synonymous
nonsynonymous
stop
number of mutations
fitness effect of mutation
B
fitness effect of mutation
nsp13
nsp10
nsp8
nsp12 (RdRp)
nsp5 (Mpro)
nsp9
nsp16
nsp7
nsp6
M
nsp4
nsp14
E
nsp15
nsp3
S
nsp2
N
nsp1
ORF7b
ORF10
ORF6
ORF8
ORF3a
ORF7a

Interactive plots:

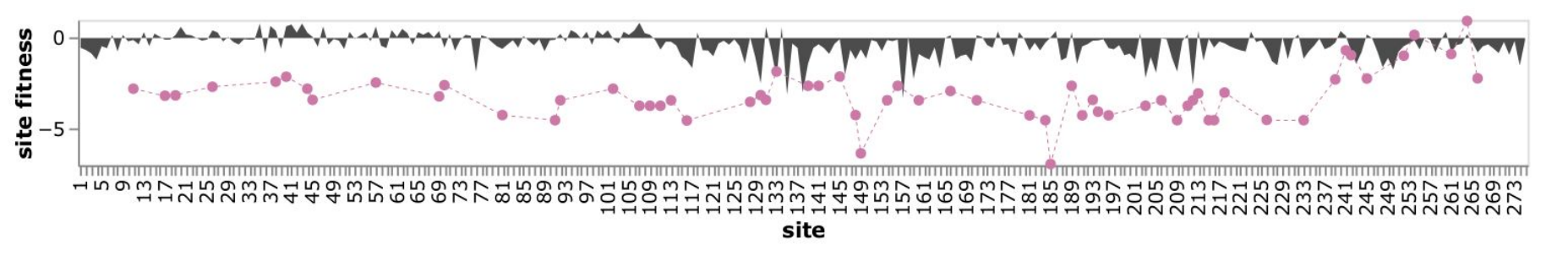
jbloomlab.github.io/SARS2-mut-fitness/

# Example: Fitness costs of mutations in the E protein

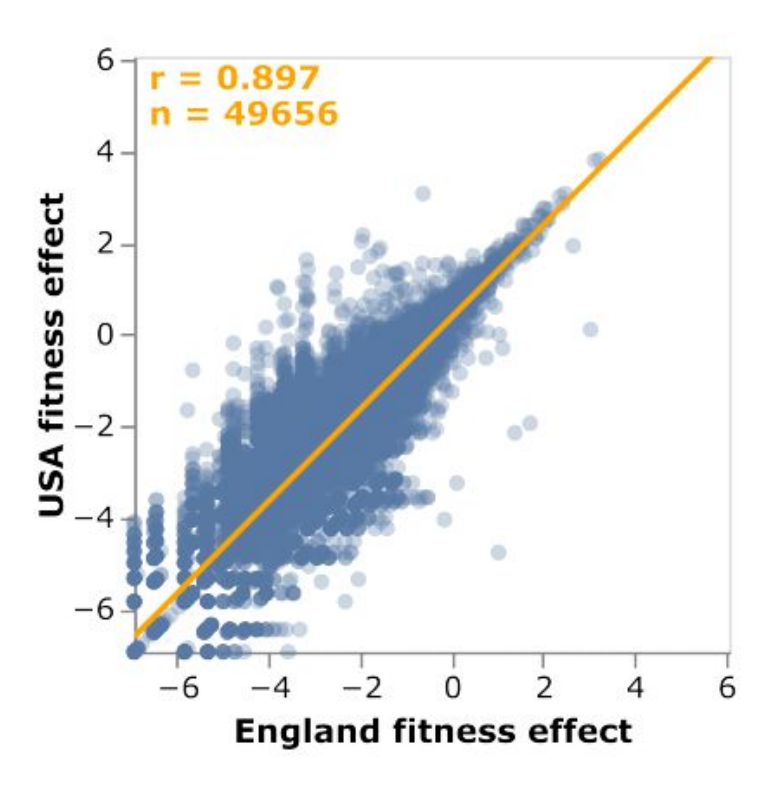


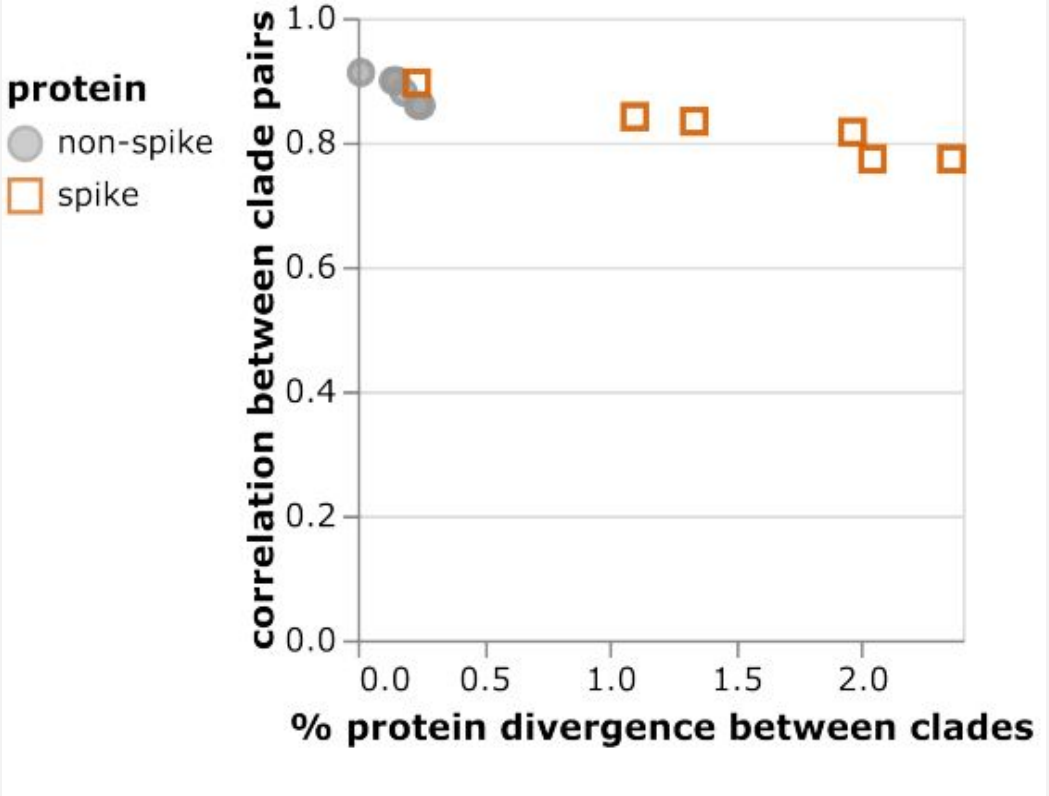
# Limited selection on amino acid sequences in accessory proteins

- Stop codons in ORF6, ORF7a/b, ORF8, and ORF10 don't seem to matter
- Circulating variants have stop codons in these genes
- ORF3a has little selection on the amino acid sequence,
  but stop codons are deleterious up to position ~240



# Estimates are consistent across geographies and clades



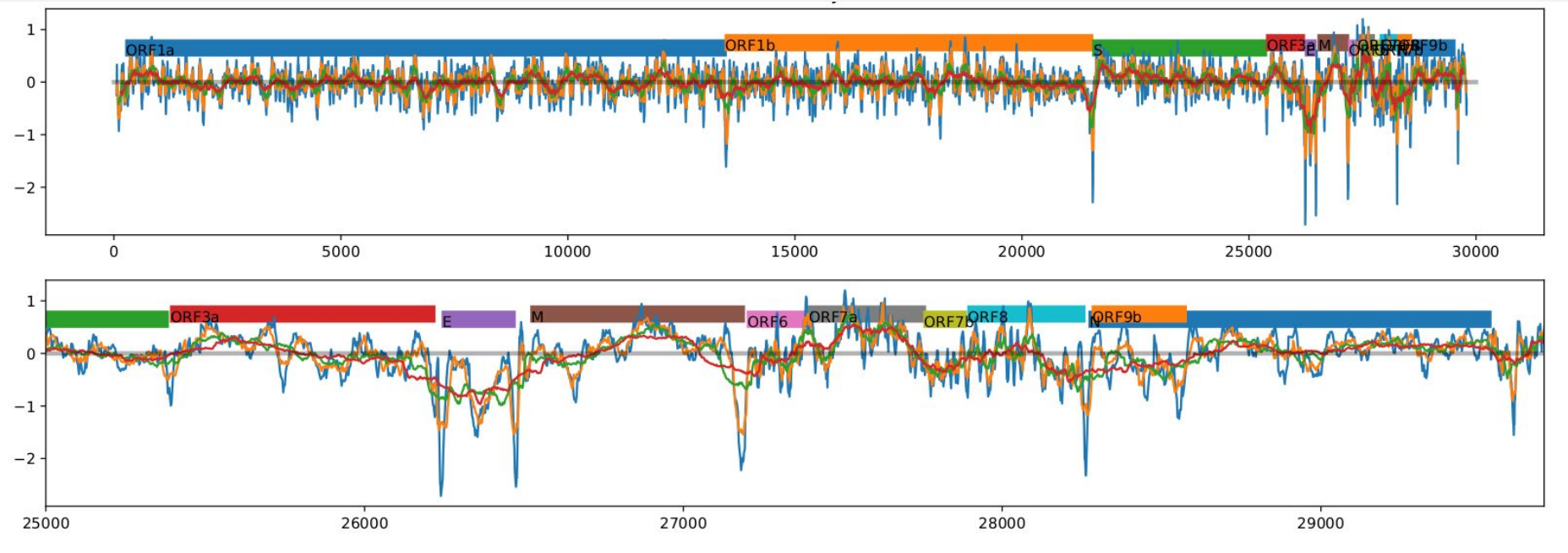




- Different wet lab protocols
- Different bioinformatic pipelines

- Independent phylogenetic structures
- Gradual decorrelation due to epistastis

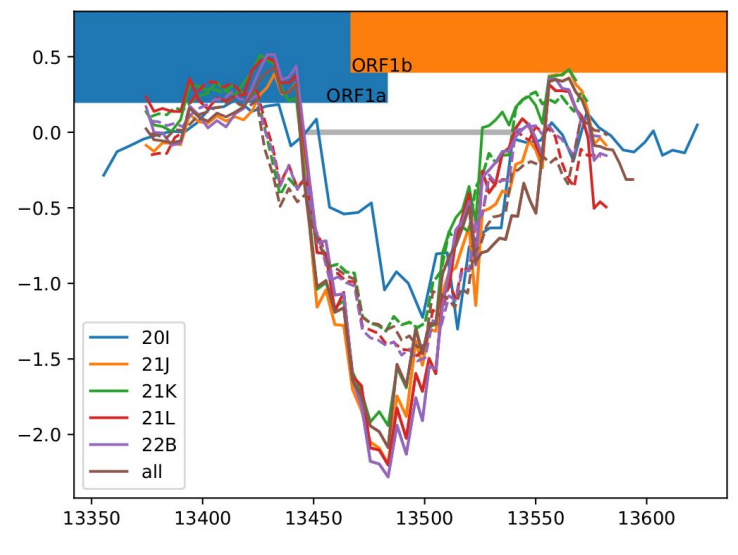
# Selection beyond the coding sequence

- Mutation counts at synonymous sites and non-coding regions
- Constraint is concentrated in a few specific regions
- Most of these regions are well characterized elements

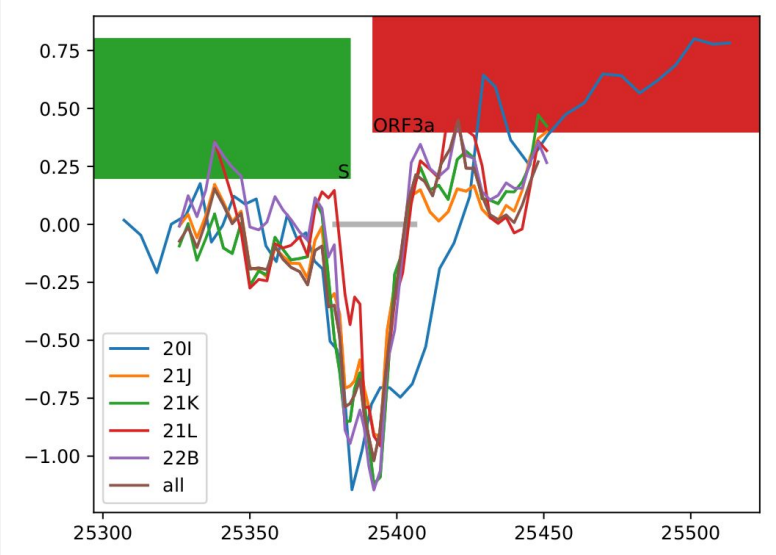


# Well known RNA elements are clearly visible

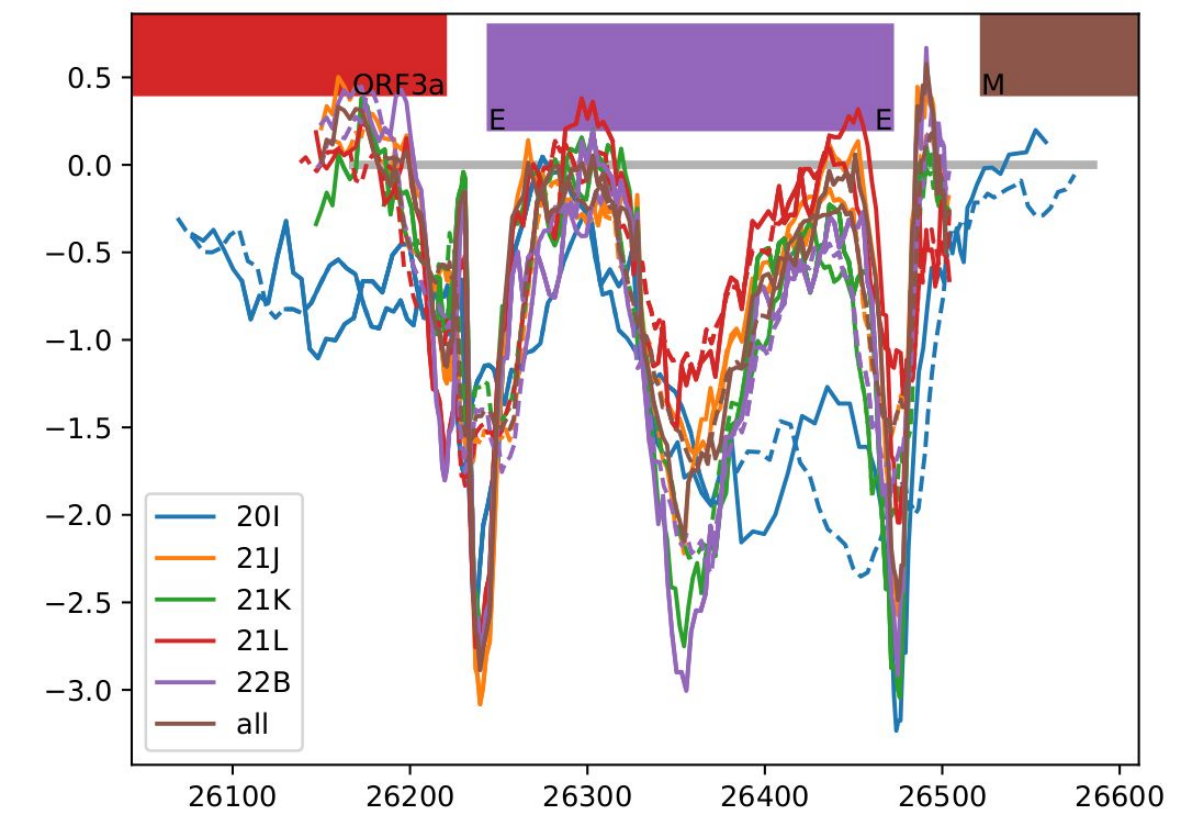
## Ribosomal slippage site




## Transcription regulatory sequences



# Strong signal in E



# Acknowledgements


- Nextstrain team (my lab and Trevor Bedford's lab)
    - Ivan Aksamentov, Cornelius Roemer, Emma Hodcroft, Moira Zuber
    - John Huddleston, Jover Lee, Tom Sibley, James Hadfield, Victor Lin
- Sequence data contributors around the world (shared via GISAID or INSDC)
- Jesse Bloom and his lab


# Comparison with deep mutational scanning data