

SARS-CoV-2, the evolution of structural proteins

Jan Pačes

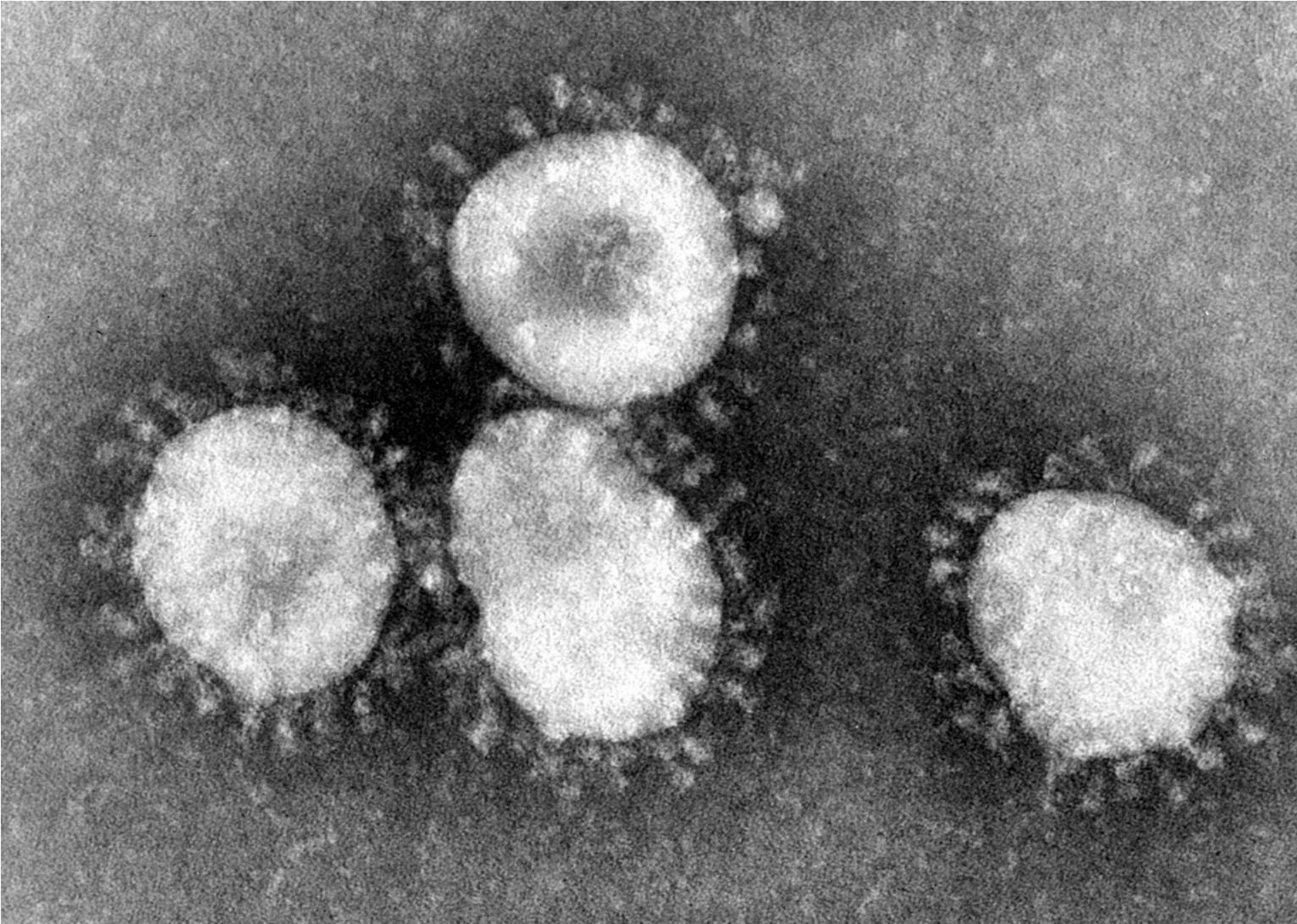
Institute of Molecular Genetics AS CR
University of Chemistry and Technology

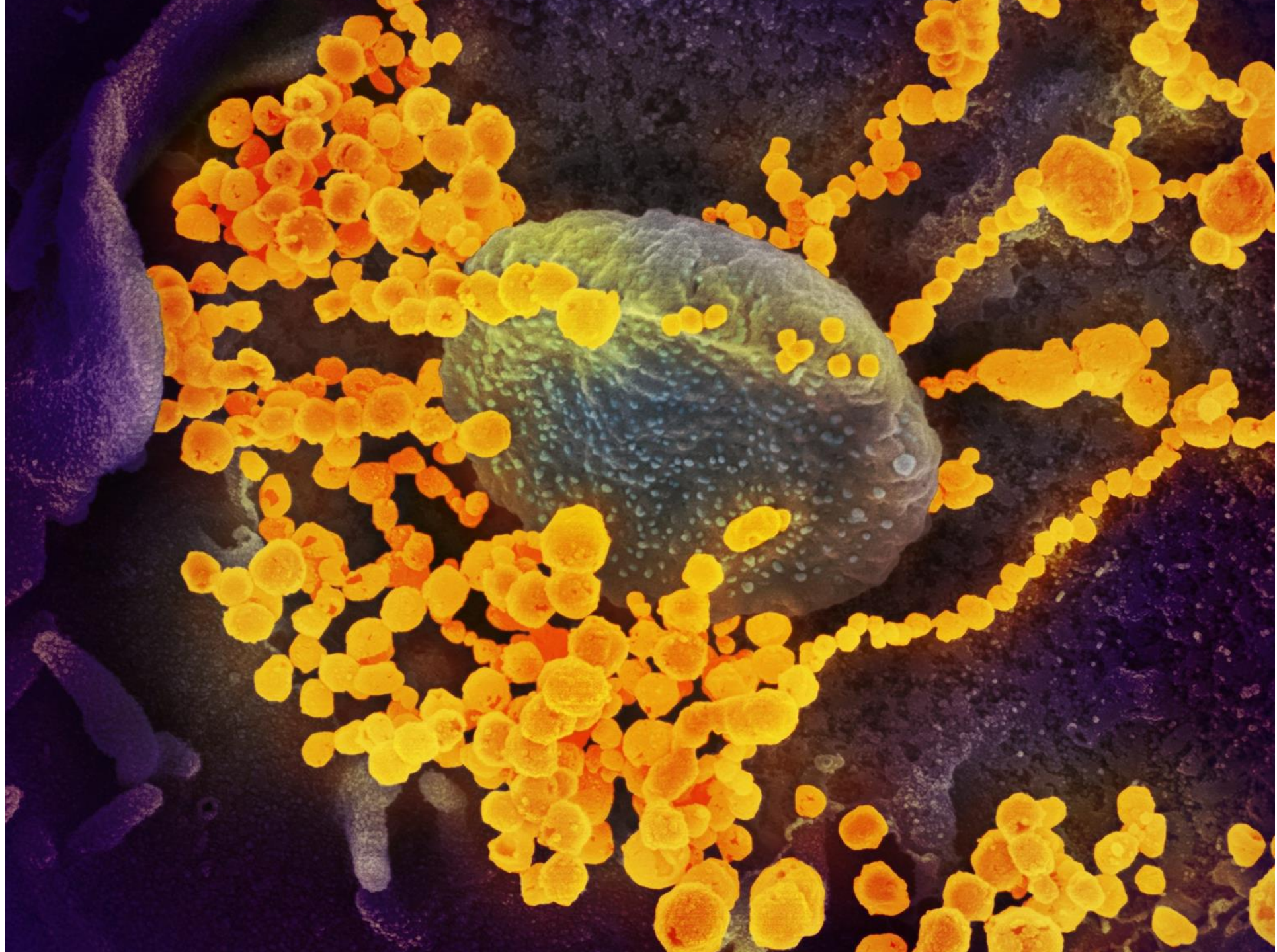
ELIXIR

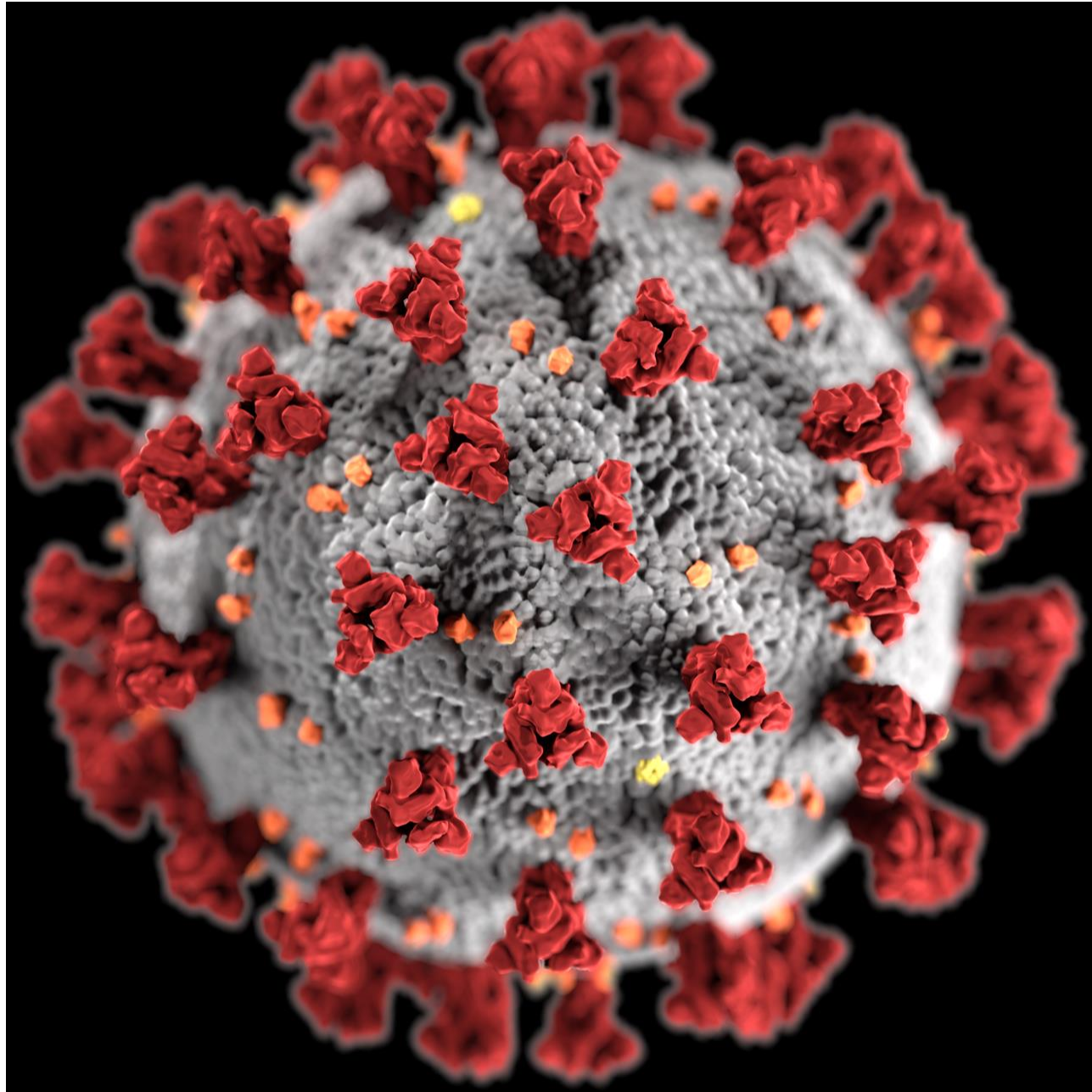


Coronaviruses

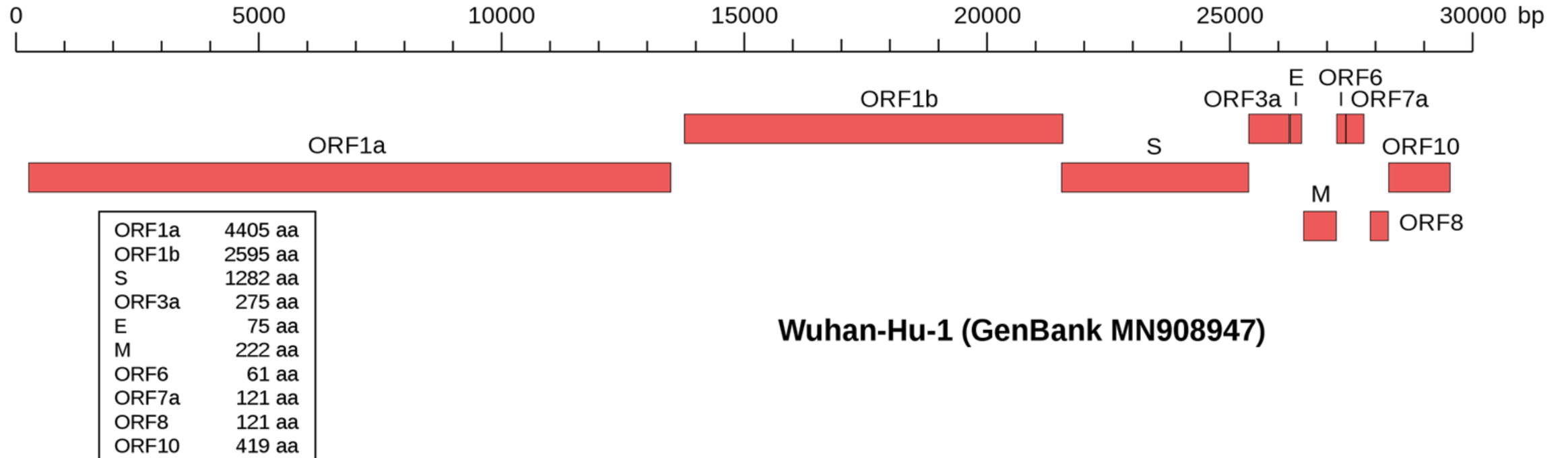
- Single strand RNA+ viruses
- genome around 30 kbp (one of the biggest viral genomes)
- so far 6+1 human coronaviruses:
 - genus: *Alphacoronavirus*, species **HCoV-229E** a **HCoV-NL63**
 - genus: *Betacoronavirus*, species **HCoV-OC43** a **HCoV-HKU1**
 - East respiratory syndrome-related coronavirus (**MERS-CoV**)
 - subgenus: Sarbecovirus severe acute respiratory syndrome-related coronavirus
 - **SARS-CoV**
 - **SARS-CoV-2**



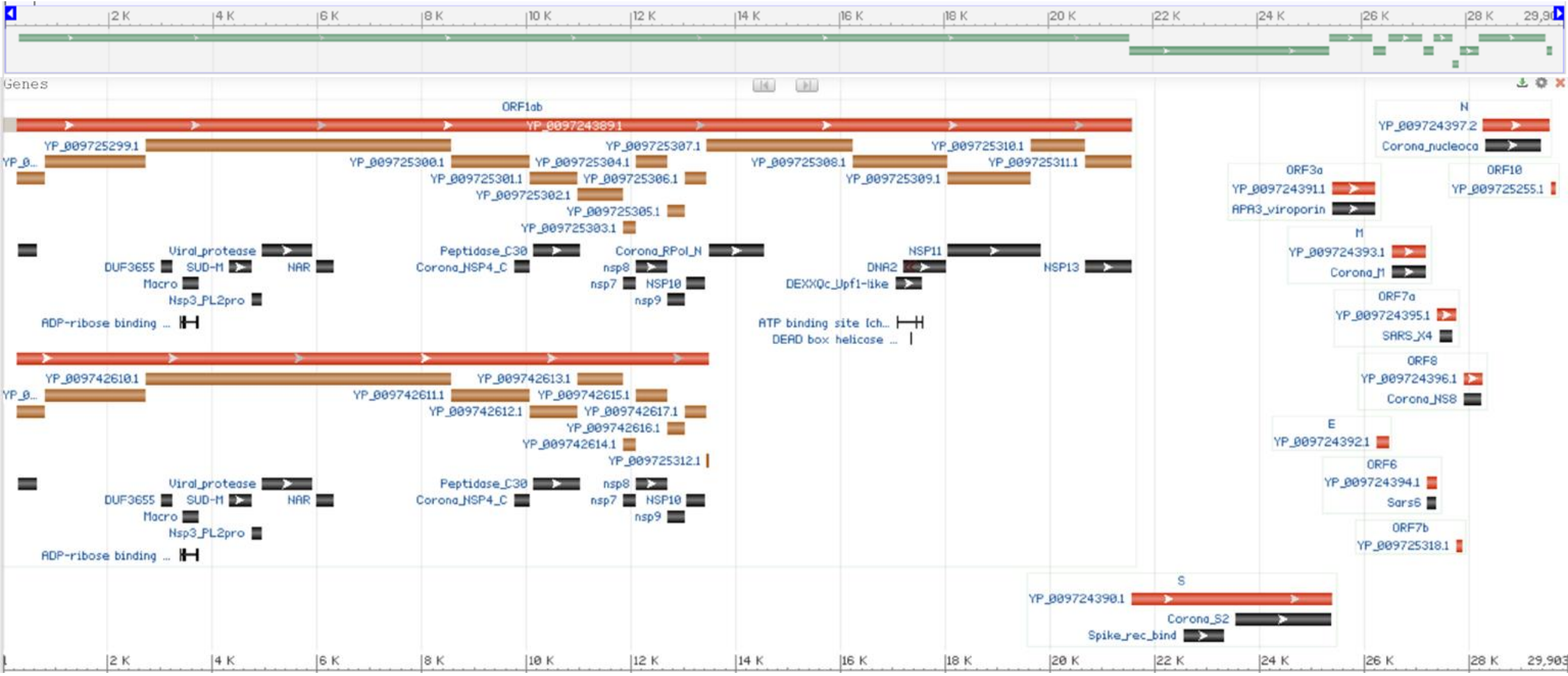


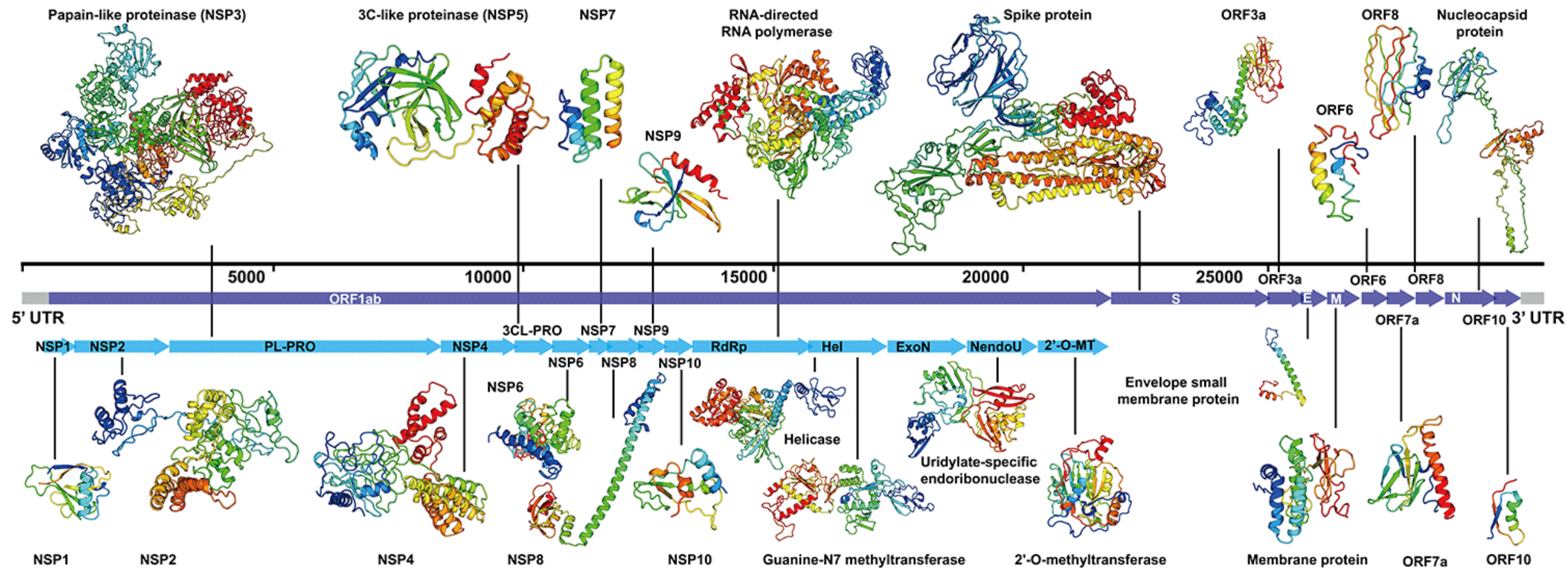


Genomic structure of SARS-CoV-2

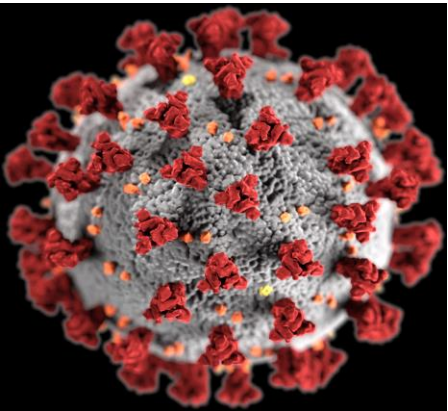
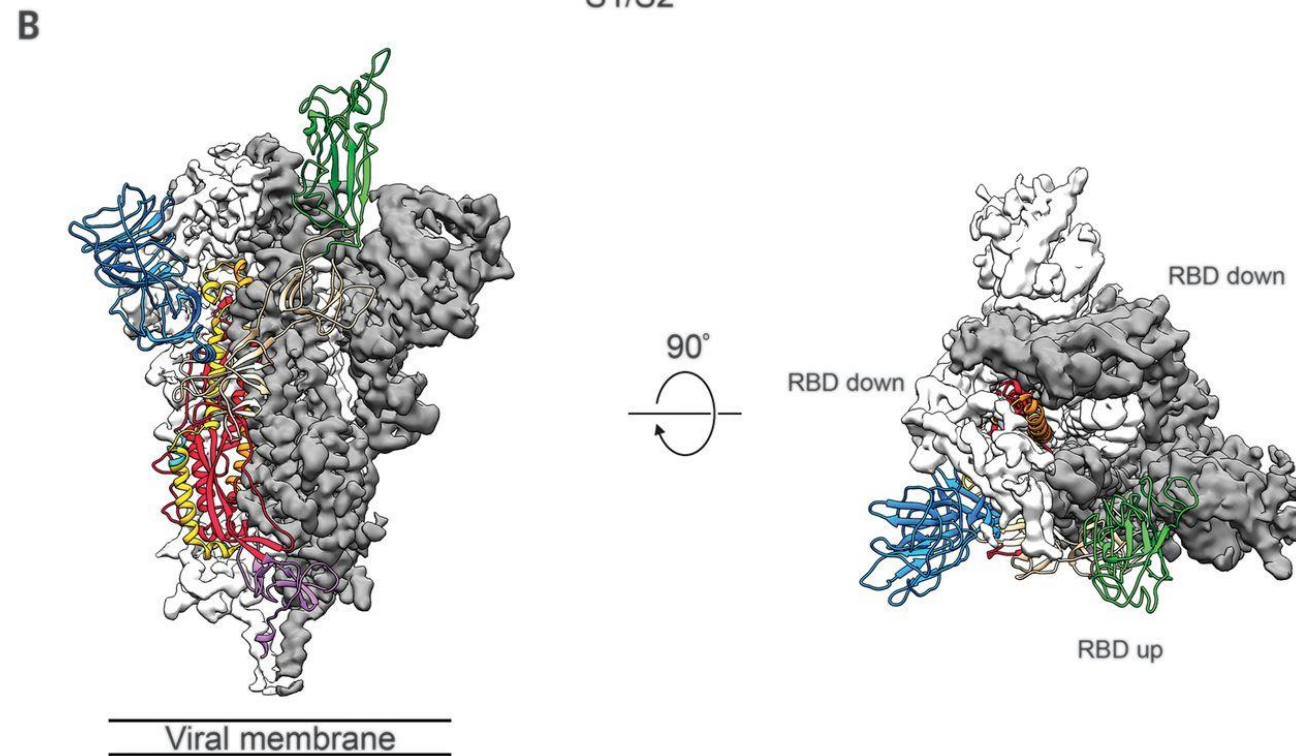
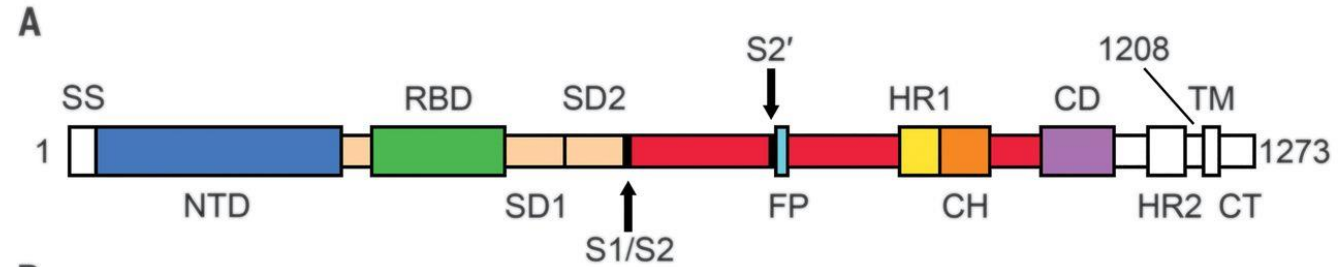


Genomic structure of SARS-CoV-2

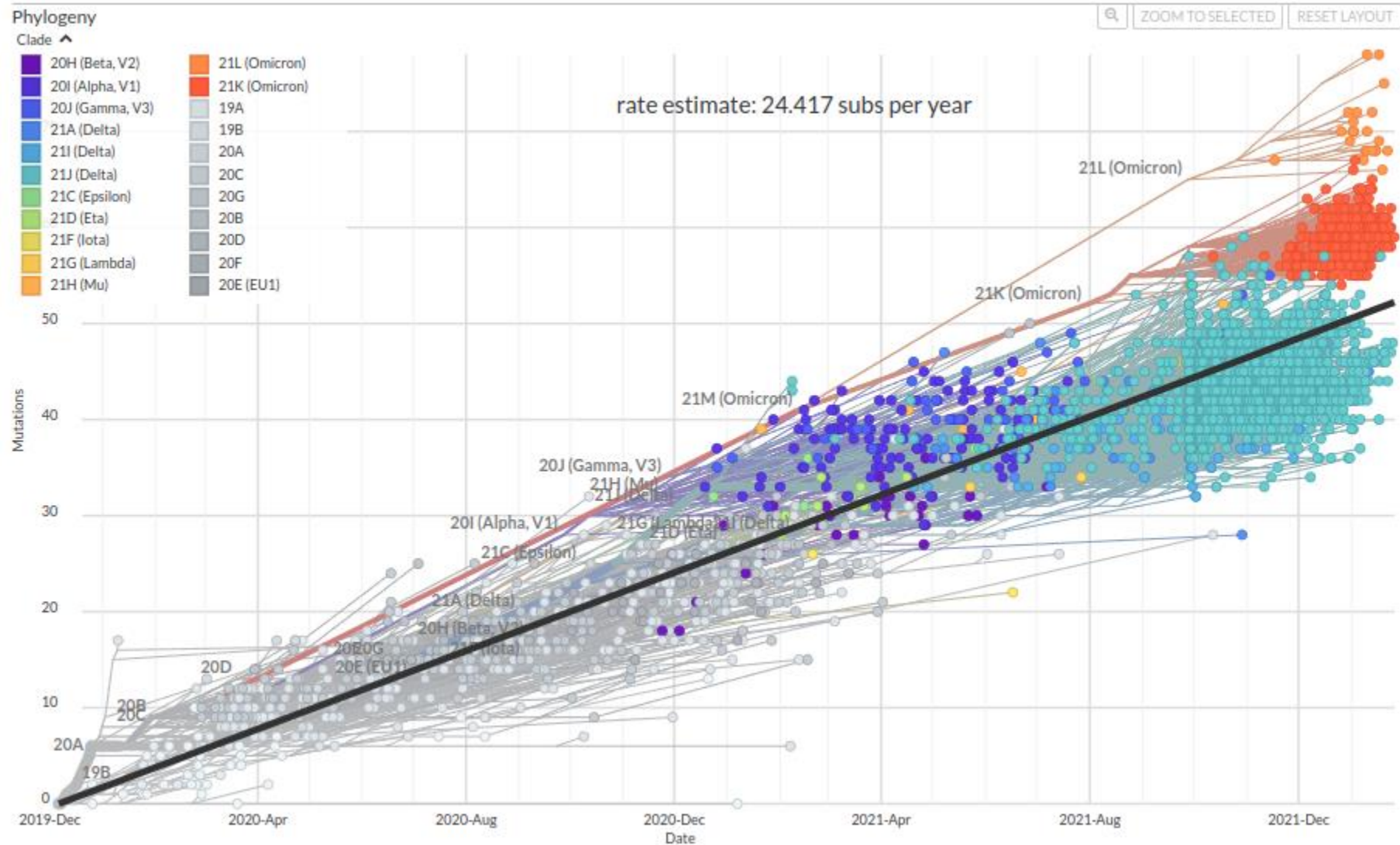




Spike gene and protein

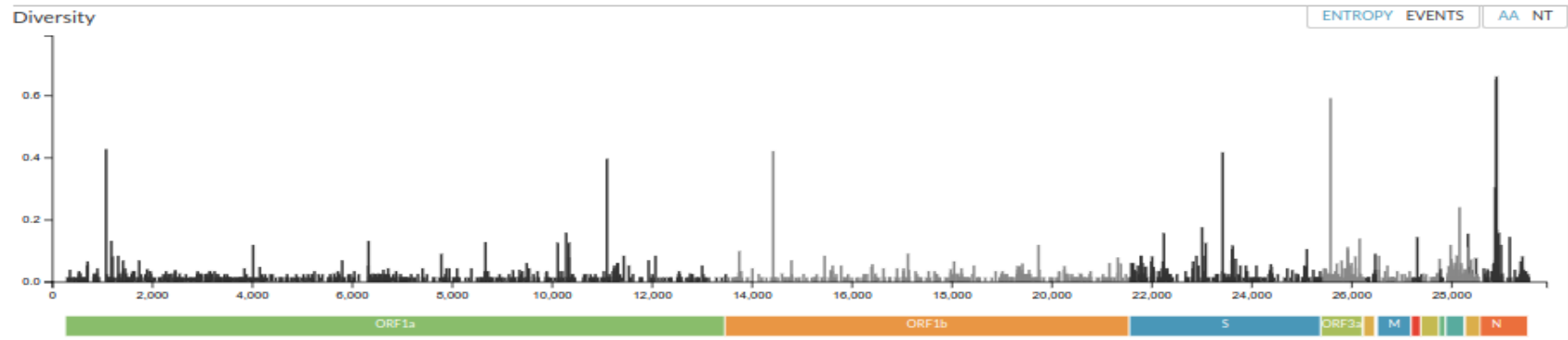


Mutational speed ($\sim 25/\text{year}$)

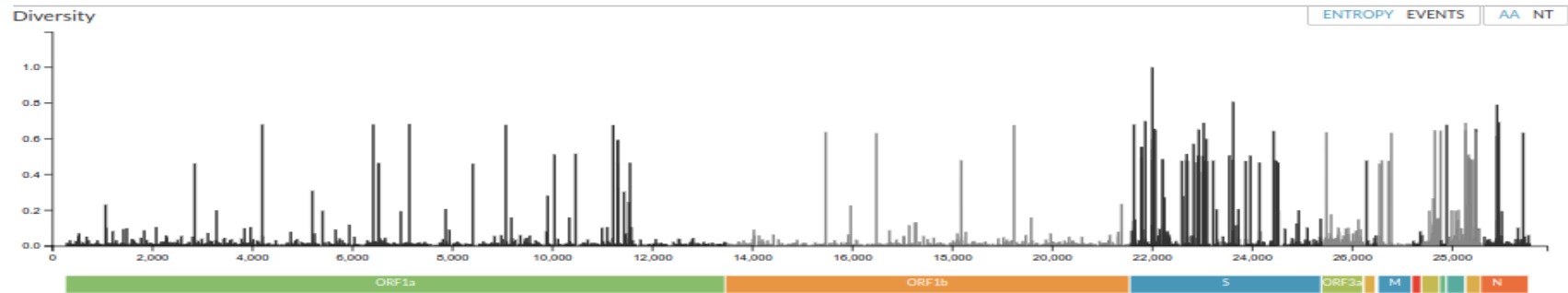


Mutations in SARS-CoV-2

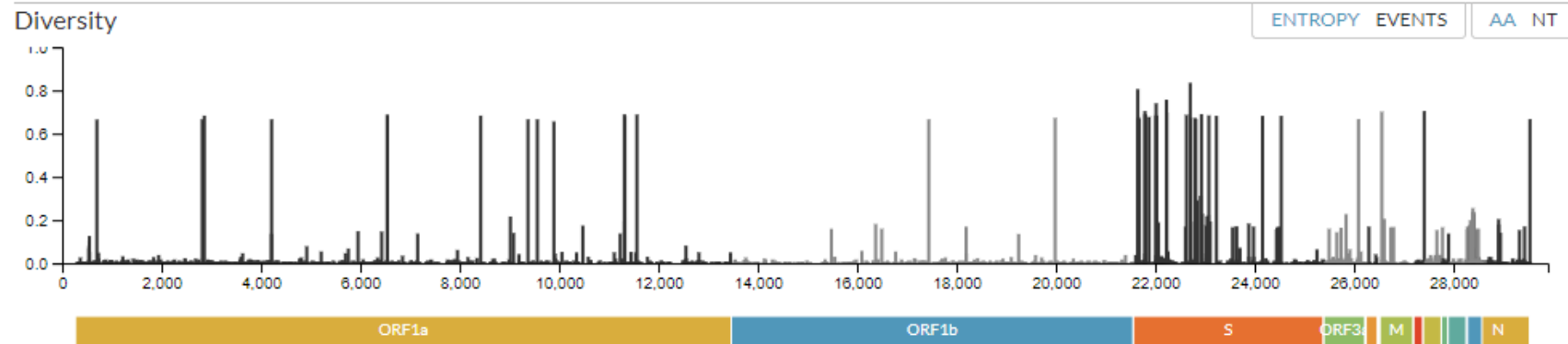
2020



2021



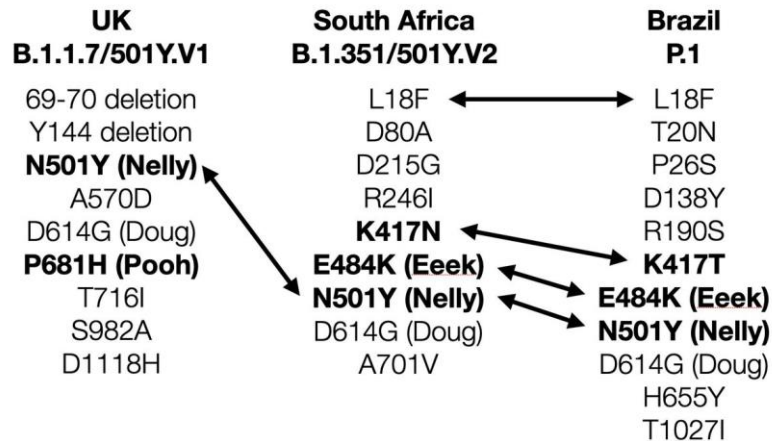
2021



Naming convention of AA changes

S:N501Y

- protein S
- aminoacid asparagin
- position 501 in protein
- change to tyrosin



20I (Alpha, V1) (B.1.1.7)	20H (Beta, V2) (B.1.351)	20J (Gamma, V3) (P.1)	21A (Delta) (B.1.617.2)	21B (Kappa) (B.1.617.1)	21C (Epsilon) (B.1.427/9)	21D (Eta) (B.1.525)
Shared mutations						
Sort by: Commonness <input type="checkbox"/> Position <input checked="" type="checkbox"/>						
	S: L 18 F	S: L 18 F				
		S: P 26 S				
S: H 69 -						S: H 69 -
S: V 70 -						S: V 70 -
S: Y 144 -						S: Y 144 -
	S: L 241 -					
	S: L 242 -					
	S: A 243 -					
	S: K 417 N	S: K 417 T				
			S: L 452 R	S: L 452 R	S: L 452 R	
	S: E 484 K	S: E 484 K		S: E 484 Q		S: E 484 K
S: N 501 Y	S: N 501 Y	S: N 501 Y				
S: D 614 G	S: D 614 G	S: D 614 G	S: D 614 G	S: D 614 G	S: D 614 G	S: D 614 G
S: P 681 H			S: P 681 R	S: P 681 R		
	S: A 701 V					
			S: D 950 N			
		S: T 1027 I				
S: D 1118 H						

and nucleotide

<https://covariants.org/variants/>

Mutation Information

- **S:N501** has appeared multiple times independently: each can be associated with different accompanying mutations
- Amino-acid changes are **S:N501Y** (nucleotide mutation **A23063T**), **S:N501T** (nucleotide mutation **A23064C**), and **S:N501S** (nucleotide mutation **A23064G**)

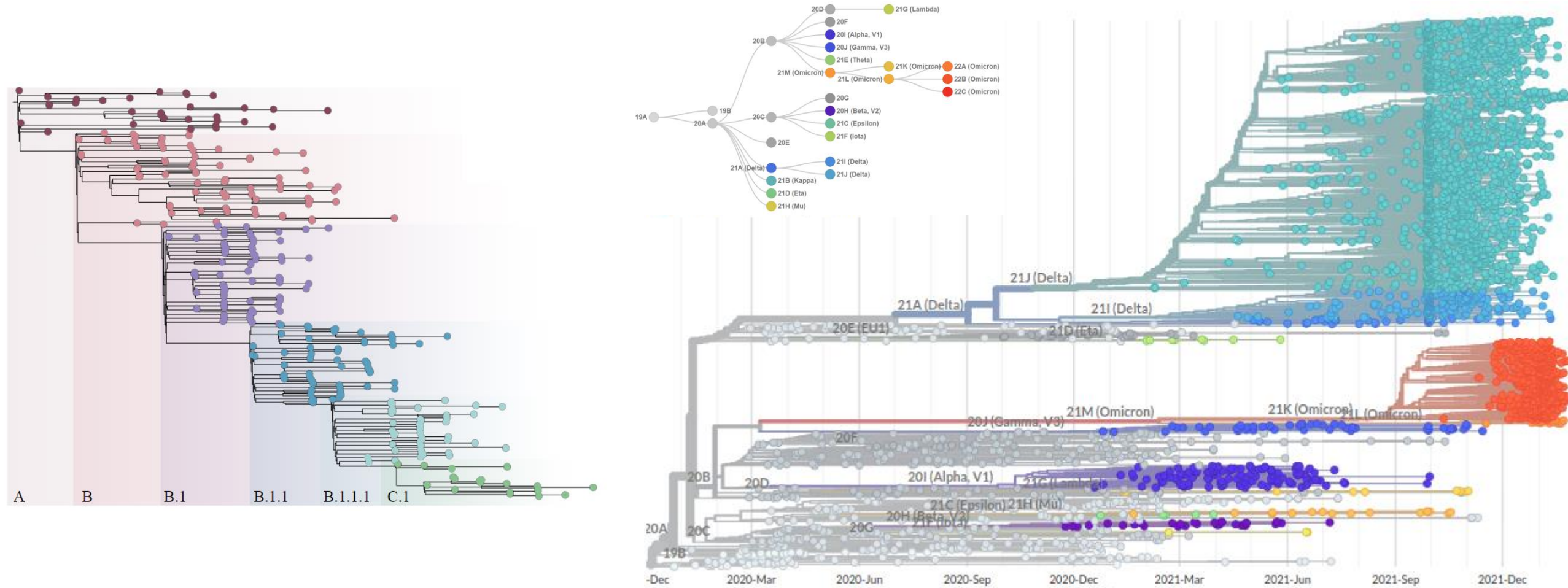
<https://codon2nucleotide.theo.io/>

Codon2Nucleotide

Convert from codon position to genomic coordinates, or vice versa. Currently for SARS-CoV-2. [GitHub](#)

Codon	Nucleotide
Gene: <input type="text" value="S"/>	Nucleotide: <input type="text" value="23063"/> to 23065
Codon: <input type="text" value="501"/>	

How we name variants (Pango, Nextstrain)



Other naming schemes

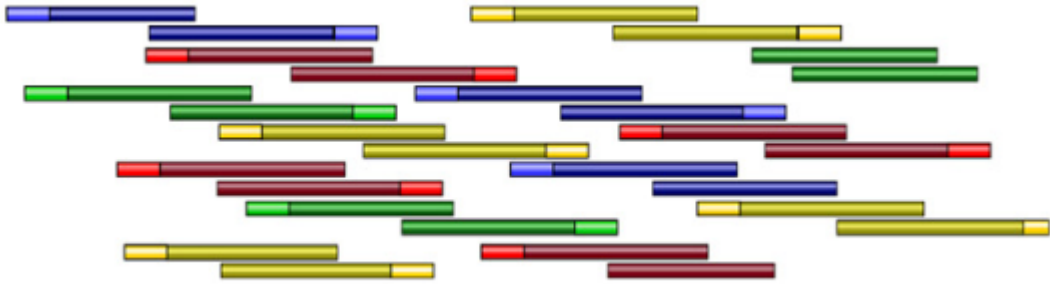
Nextstrain Clade	Pango Lineage	WHO Label ↗	Other
20I (Alpha, V1)	B.1.1.7 ↗	α Alpha	VOC 202012/01
20H (Beta, V2)	B.1.351 ↗	β Beta	501Y.V2
20J (Gamma, V3)	P.1 ↗	γ Gamma	
21A (Delta)	B.1.617.2 ↗	δ Delta	
21B (Kappa)	B.1.617.1 ↗	κ Kappa	
21C (Epsilon)	B.1.427, B.1.429	ε Epsilon	CAL.20C
21D (Eta)	B.1.525 ↗	η Eta	
21F (Iota)	B.1.526	ι Iota	(Part of Pango lineage)
21G (Lambda)	C.37	λ Lambda	
21H (Mu)	B.1.621	μ Mu	
20E (EU1)	B.1.177		EU1
20B/S: 732 A	B.1.1.519		
20A/S: 126 A	B.1.620		
20A.EU2	B.1.160		
20A/S: 439 K	B.1.258		
20A/S: 98 F	B.1.221		
20C/S: 80 Y	B.1.367		
20B/S: 626 S	B.1.1.277		
20B/S: 1122 L	B.1.1.302		

Amplicon sequencing

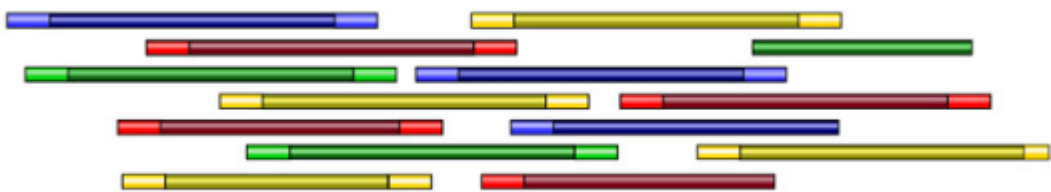
Primer design



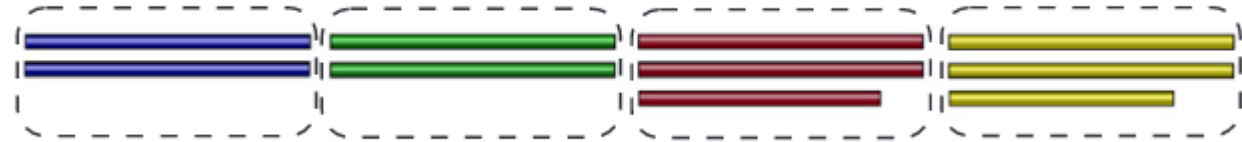
Amplicon sequencing



Merge



Clustering



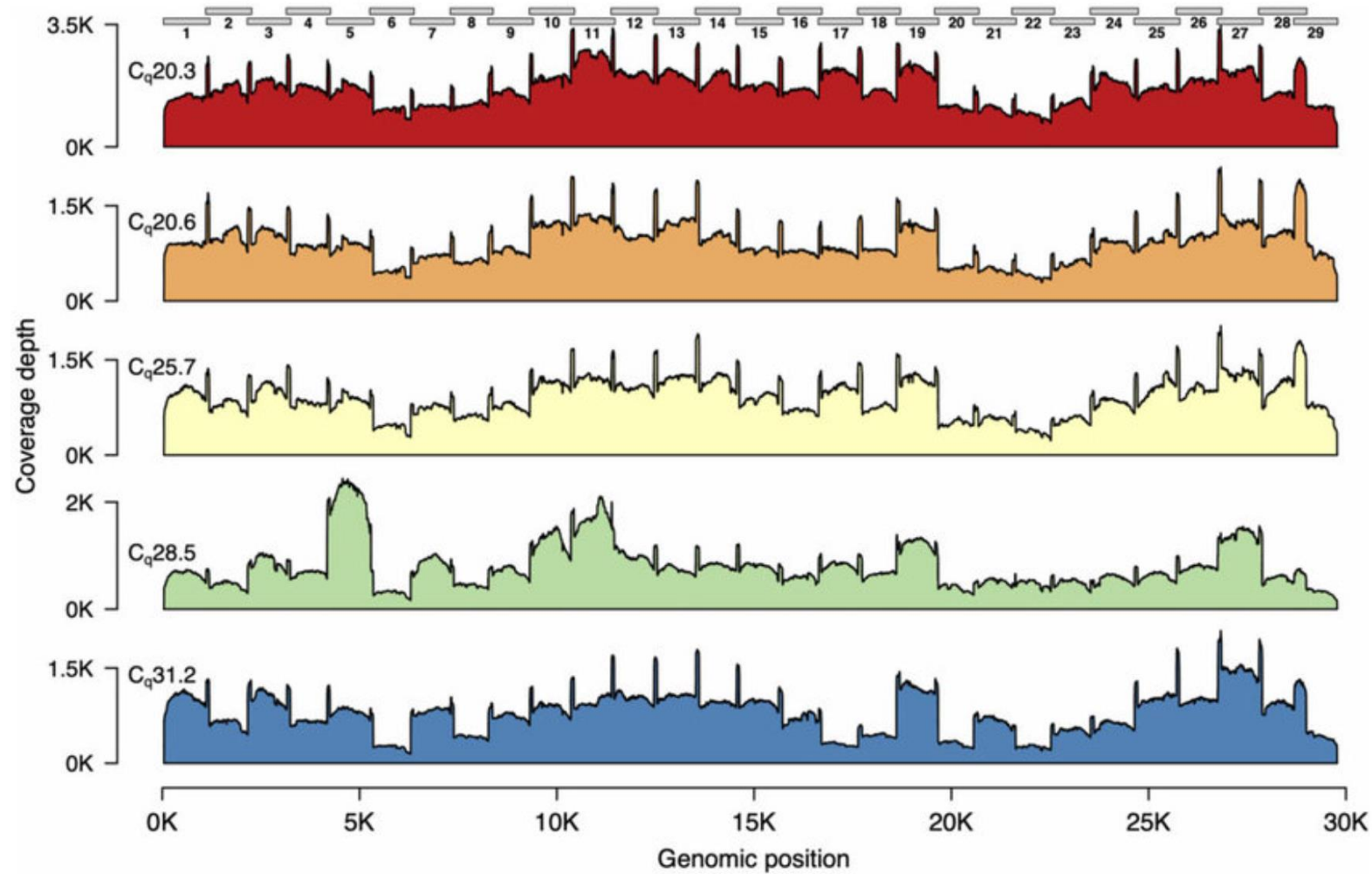
Consensus



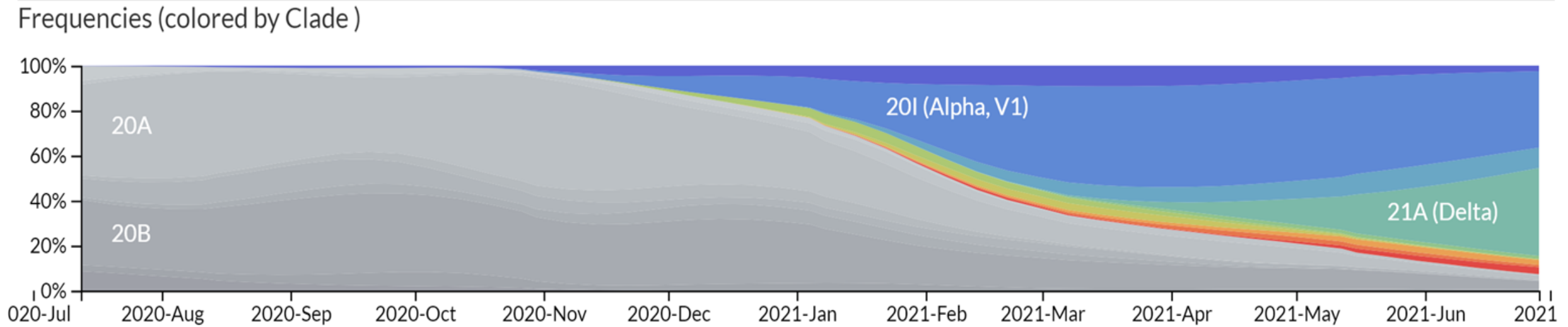
Assembly

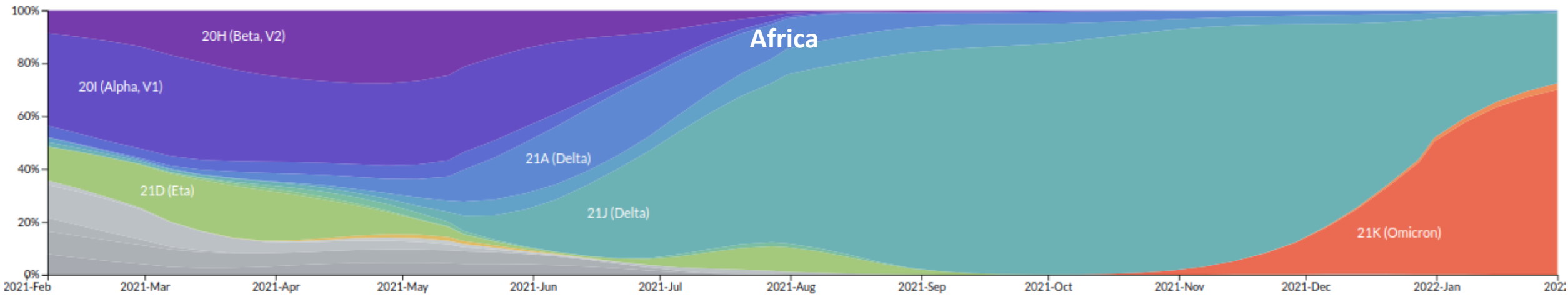
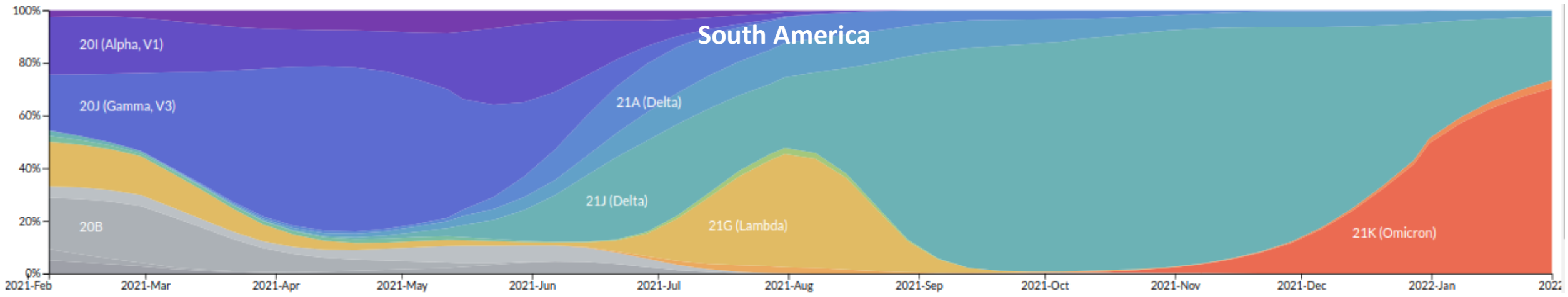
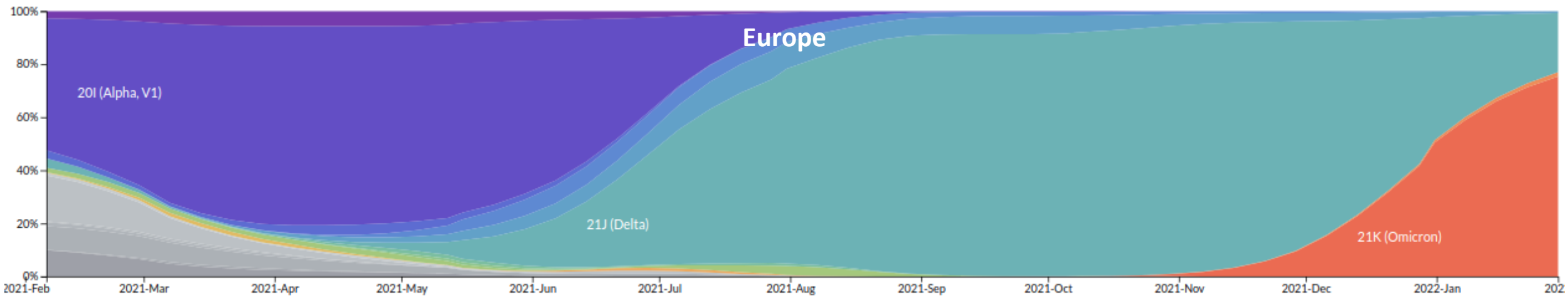


Amplicon sequencing



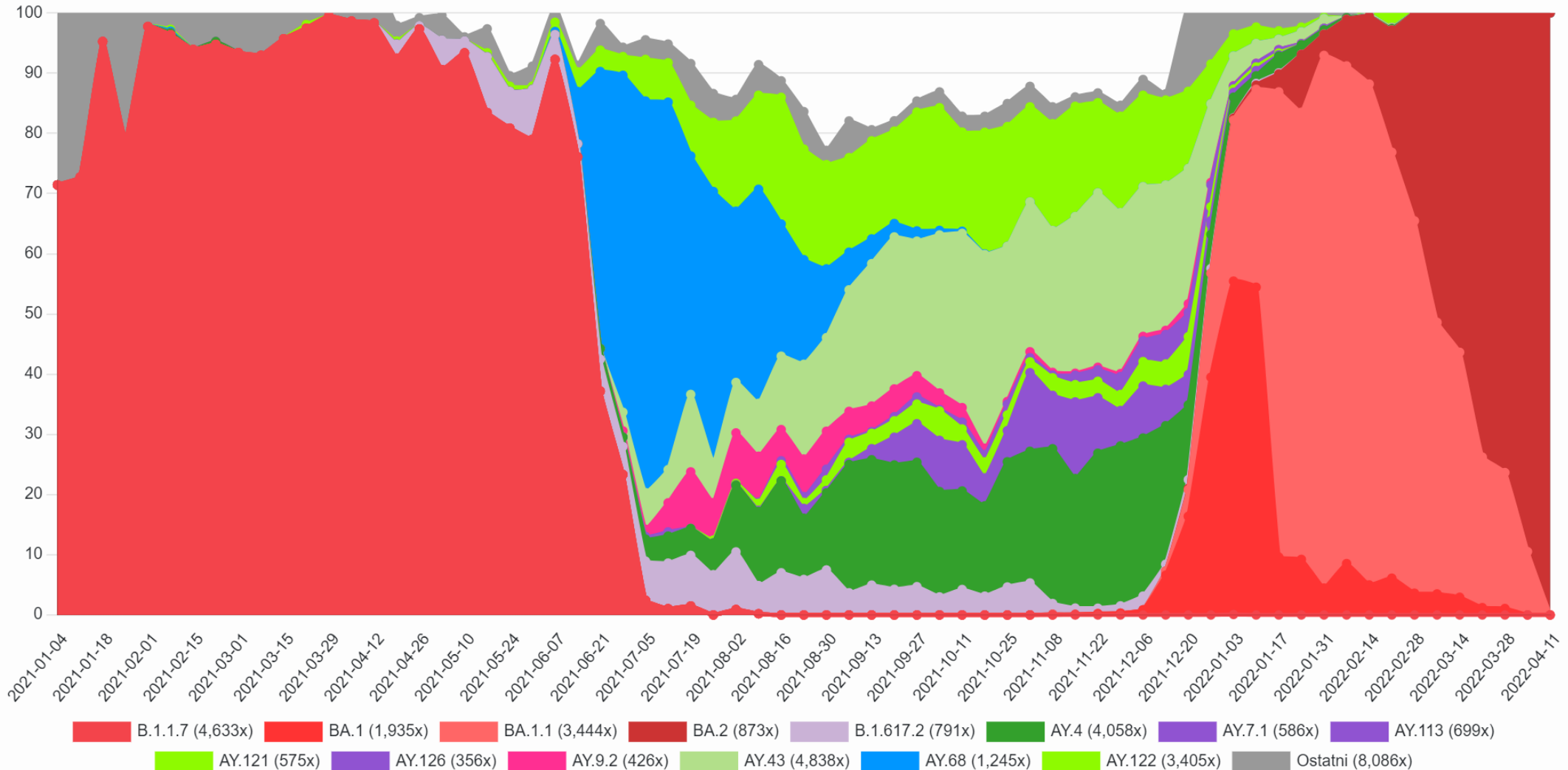
Spread of variants in space and time

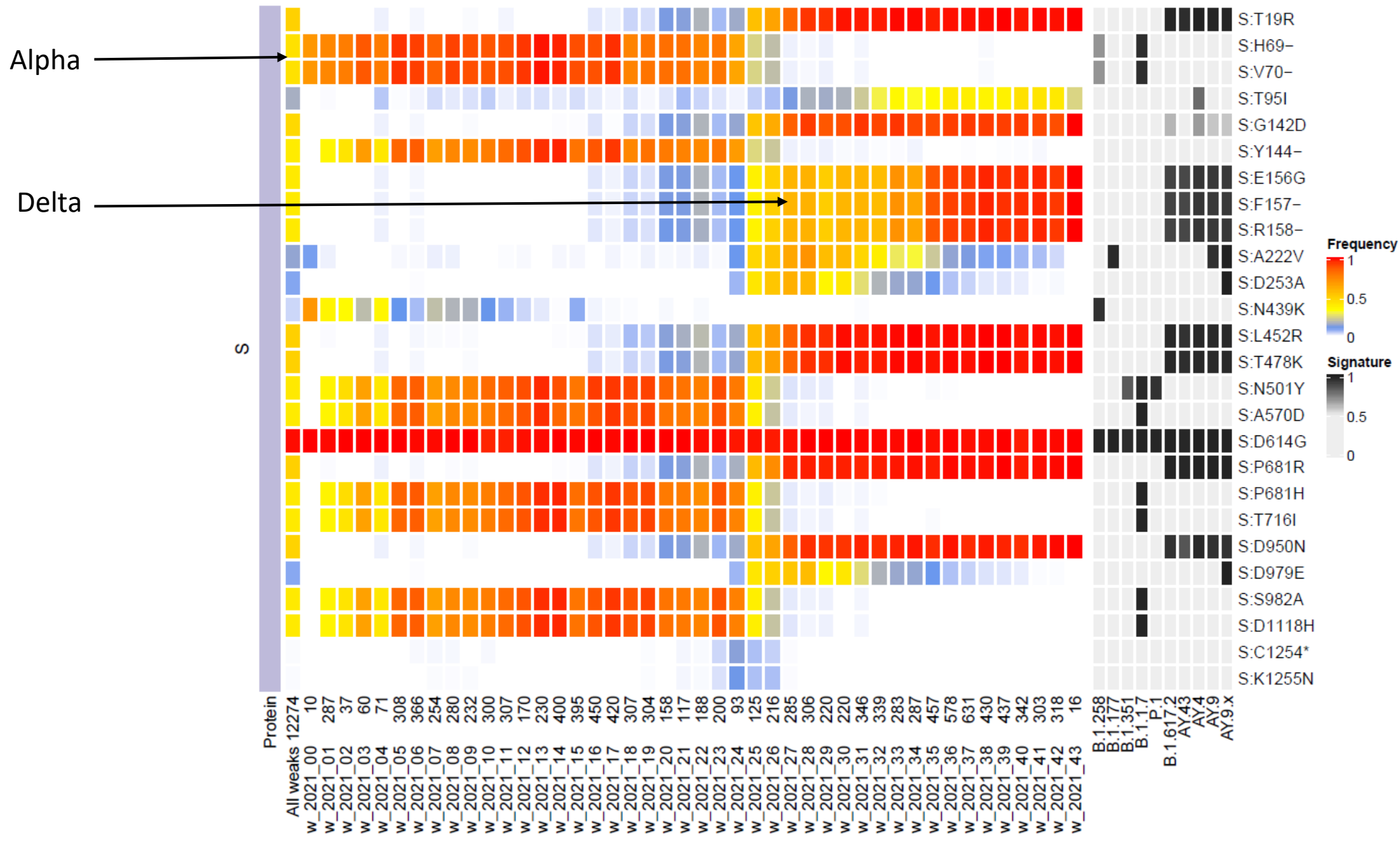




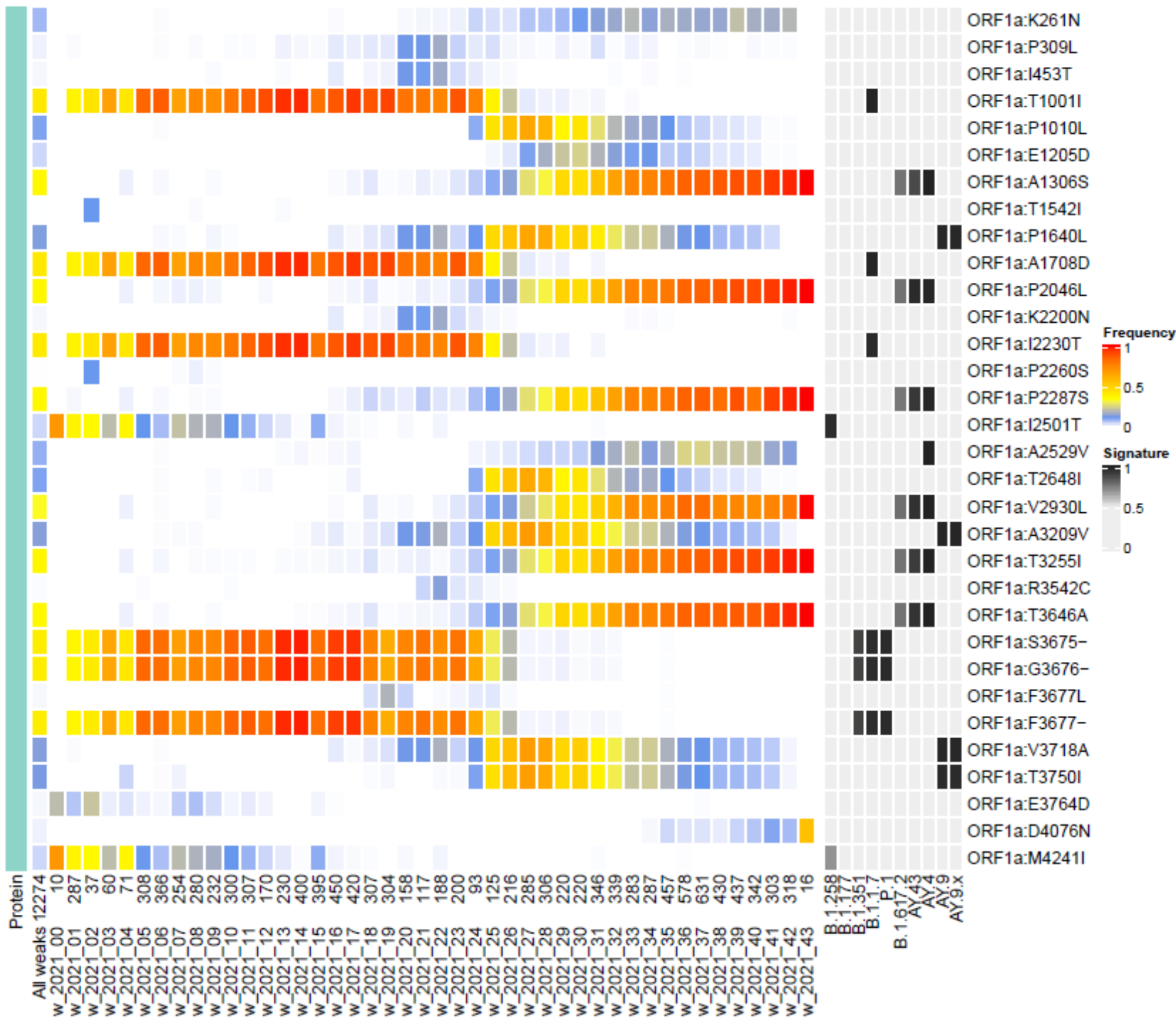
What variants do we have in Czechia?

Linie SARS-CoV-2 detekované v Česku (1. 1. 2021 - 6. 5. 2022)





ORF1a



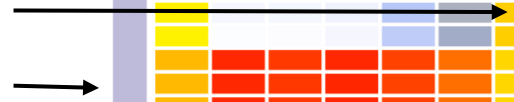
Protein

All weeks 12274
 w_2021_00 10
 w_2021_01 287
 w_2021_02 37
 w_2021_03 60
 w_2021_04 71
 w_2021_05 308
 w_2021_06 366
 w_2021_07 254
 w_2021_08 280
 w_2021_09 232
 w_2021_10 300
 w_2021_11 307
 w_2021_12 170
 w_2021_13 230
 w_2021_14 400
 w_2021_15 395
 w_2021_16 450
 w_2021_17 420
 w_2021_18 307
 w_2021_19 304
 w_2021_20 158
 w_2021_21 117
 w_2021_22 188
 w_2021_23 200
 w_2021_24 93
 w_2021_25 125
 w_2021_26 216
 w_2021_27 285
 w_2021_28 306
 w_2021_29 220
 w_2021_30 220
 w_2021_31 346
 w_2021_32 339
 w_2021_33 283
 w_2021_34 287
 w_2021_35 457
 w_2021_36 578
 w_2021_37 631
 w_2021_38 430
 w_2021_39 437
 w_2021_40 342
 w_2021_41 303
 w_2021_42 318
 w_2021_43 16

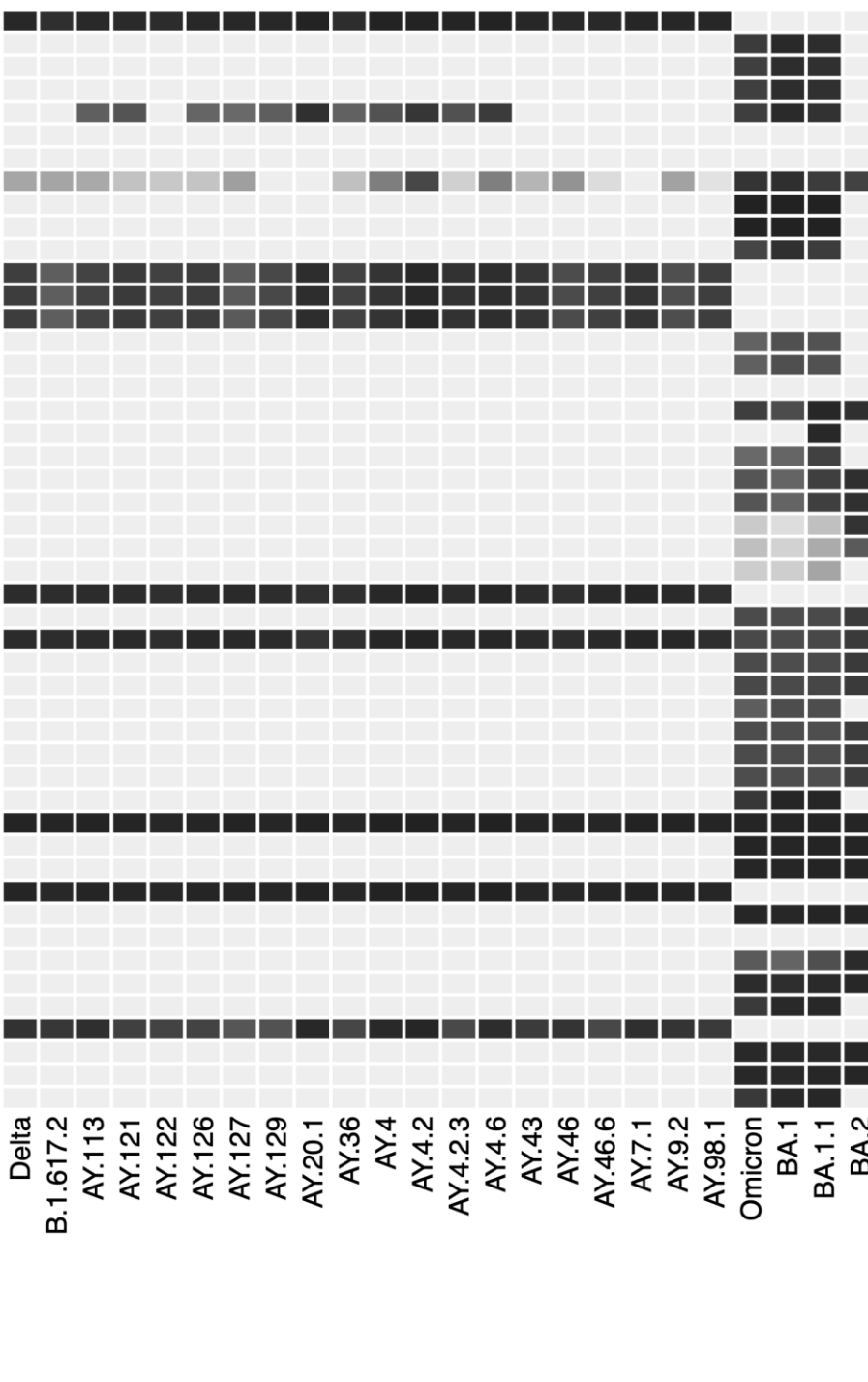
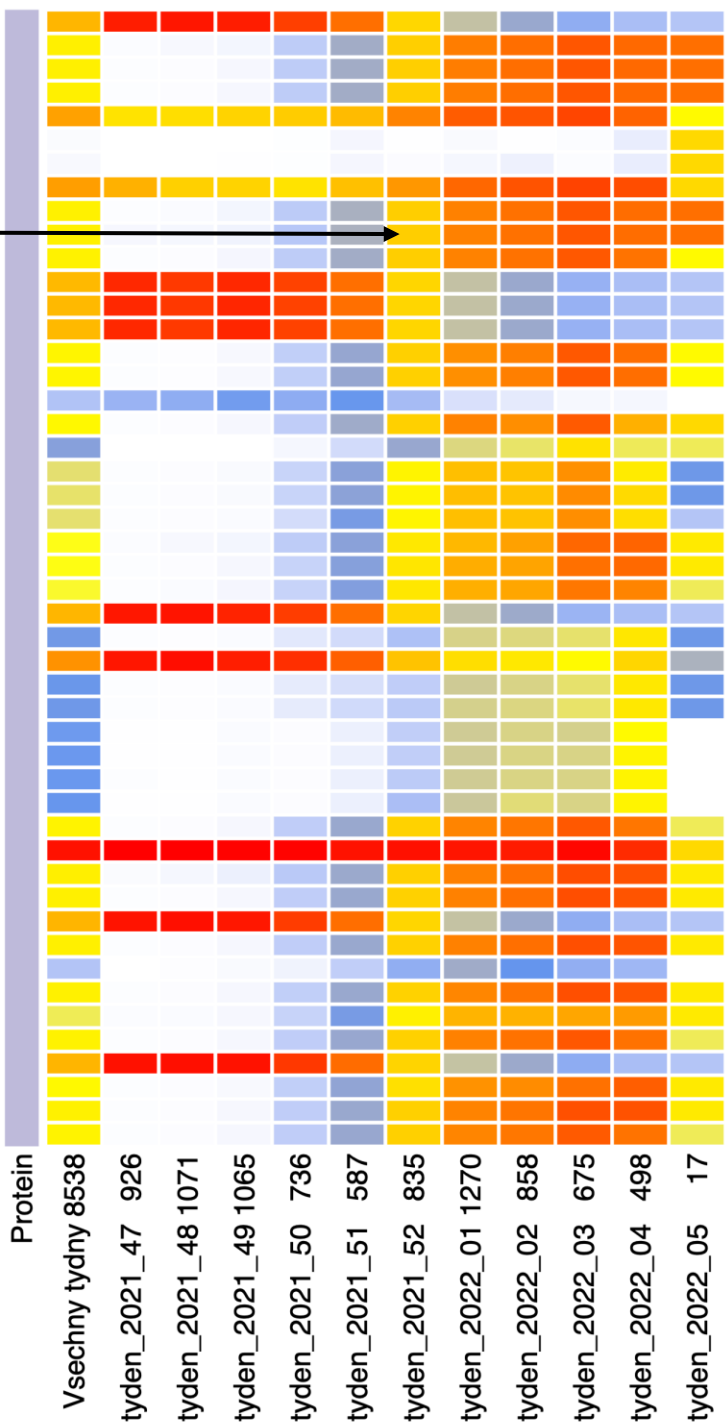
B.1.258
 B.1.177
 B.1.351
 B.1.1.7
 P.1
 B.1.617.2
 AY.43
 AY.44
 AY.9
 AY.9.X

ORF1a:K261N
 ORF1a:P309L
 ORF1a:I453T
 ORF1a:T1001I
 ORF1a:P1010L
 ORF1a:E1205D
 ORF1a:A1306S
 ORF1a:T1542I
 ORF1a:P1640L
 ORF1a:A1708D
 ORF1a:P2046L
 ORF1a:K2200N
 ORF1a:I2230T
 ORF1a:P2260S
 ORF1a:P2287S
 ORF1a:I2501T
 ORF1a:A2529V
 ORF1a:T2648I
 ORF1a:V2930L
 ORF1a:A3209V
 ORF1a:T3255I
 ORF1a:R3542C
 ORF1a:T3646A
 ORF1a:S3675-
 ORF1a:G3676-
 ORF1a:F3677L
 ORF1a:F3677-
 ORF1a:V3718A
 ORF1a:T3750I
 ORF1a:E3764D
 ORF1a:D4076N
 ORF1a:M424I

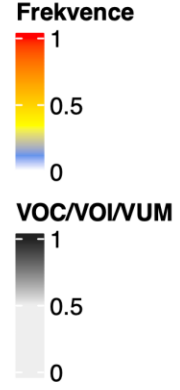
Omicron
Delta

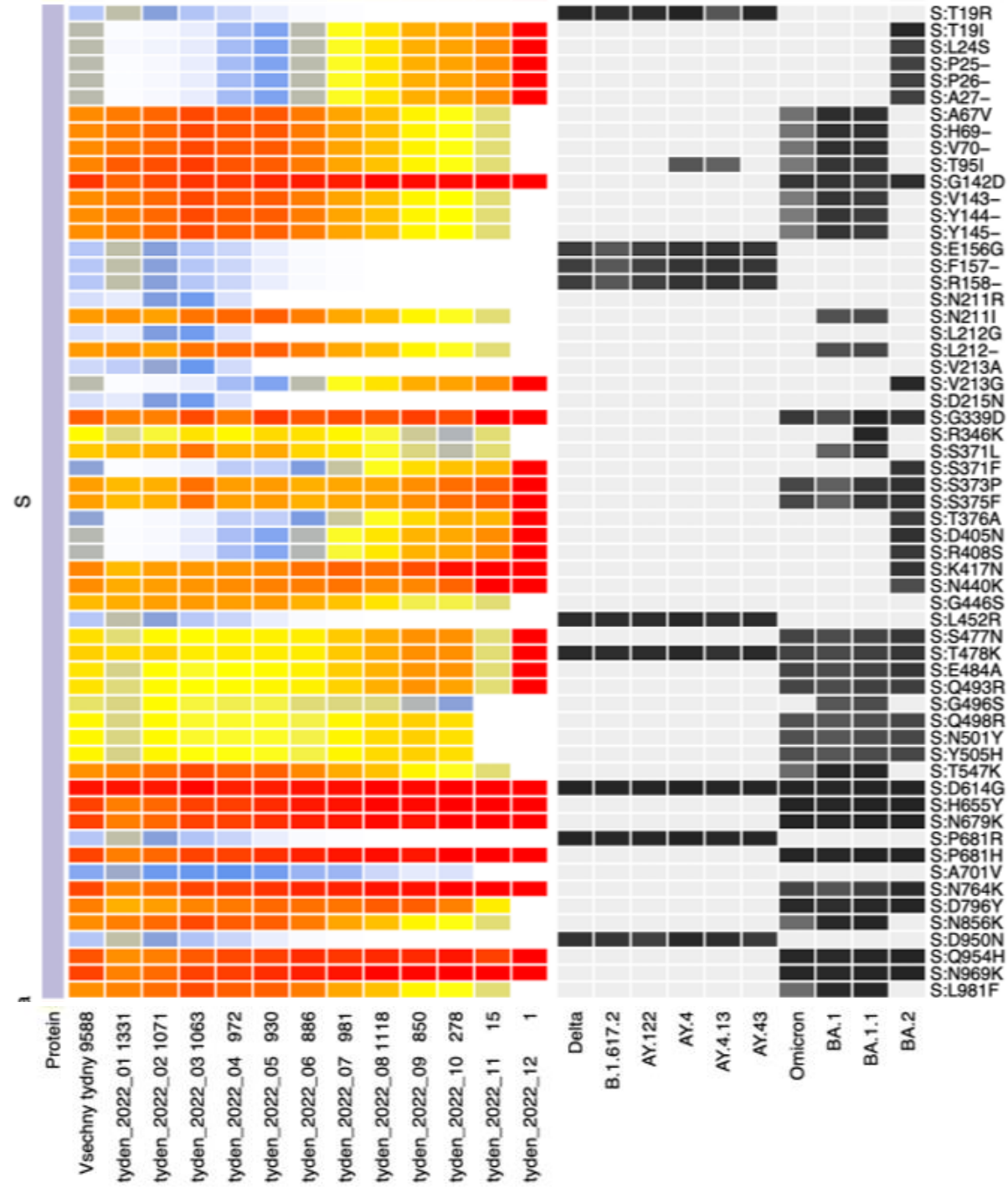


S

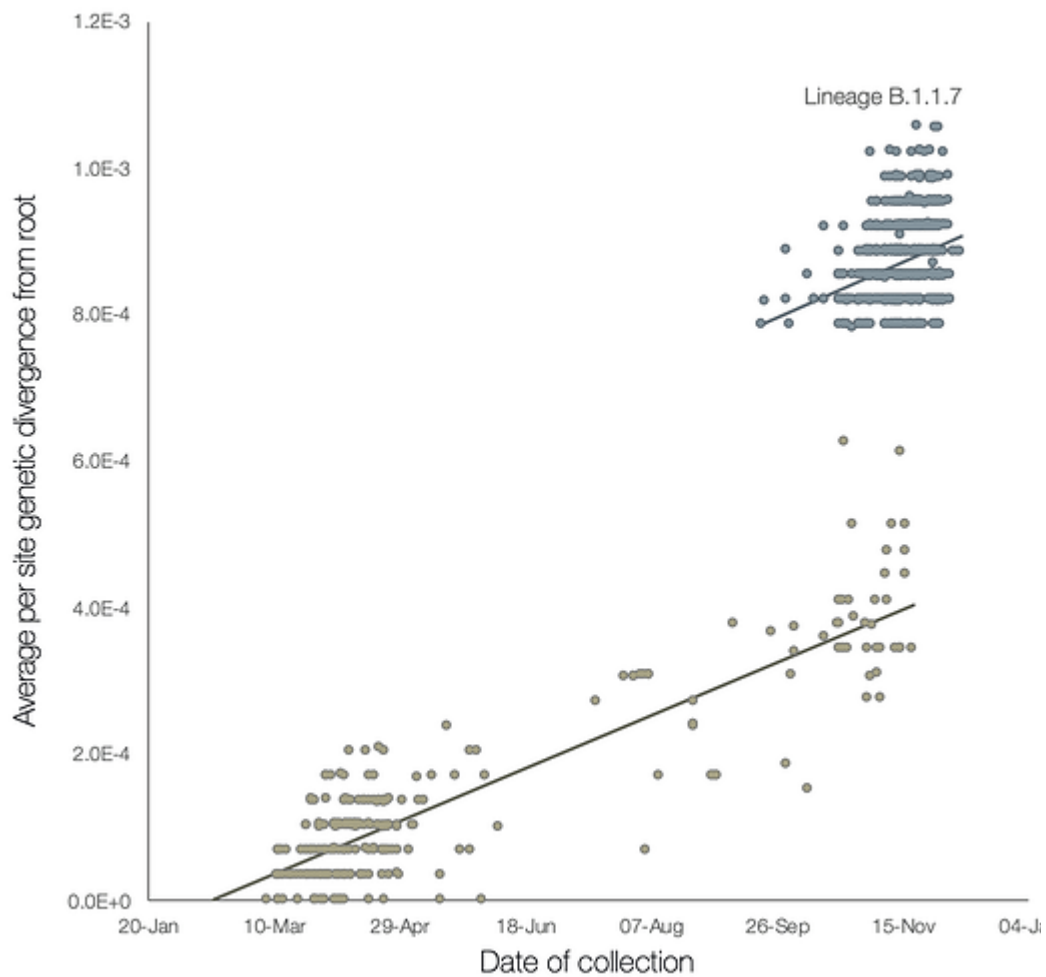


S:T19R
S:A67V
S:H69-
S:V70-
S:T95I
S:L141W
S:G142-
S:G142D
S:V143-
S:Y144-
S:Y145-
S:E156G
S:F157-
S:R158-
S:N211I
S:L212-
S:V213A
S:G339D
S:P346K
S:S371L
S:S373P
S:S375F
S:K417N
S:N440K
S:G446S
S:L452R
S:S477N
S:T478K
S:E484A
S:Q493R
S:G496S
S:Q498R
S:N501Y
S:Y505H
S:T547K
S:D614G
S:H655Y
S:N679K
S:P681R
S:P681H
S:A701V
S:N764K
S:D796Y
S:N856K
S:D950N
S:Q954H
S:N969K
S:L981F



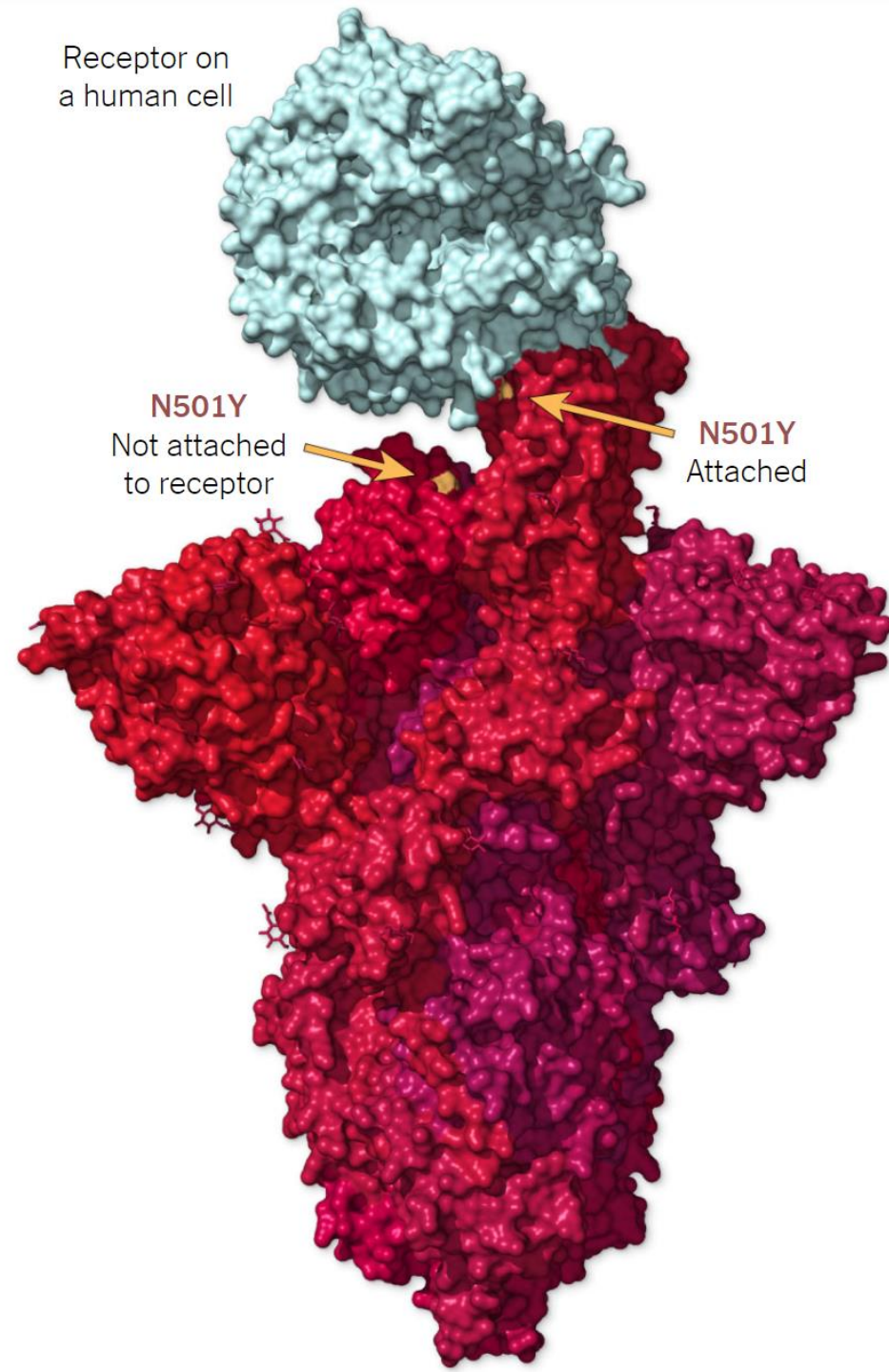
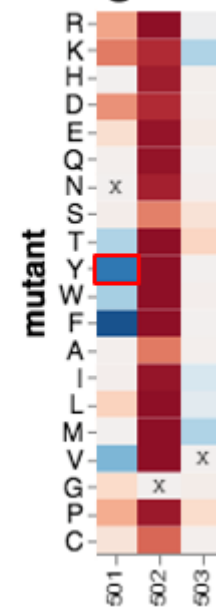
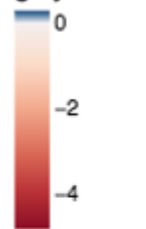


Mutation N501Y

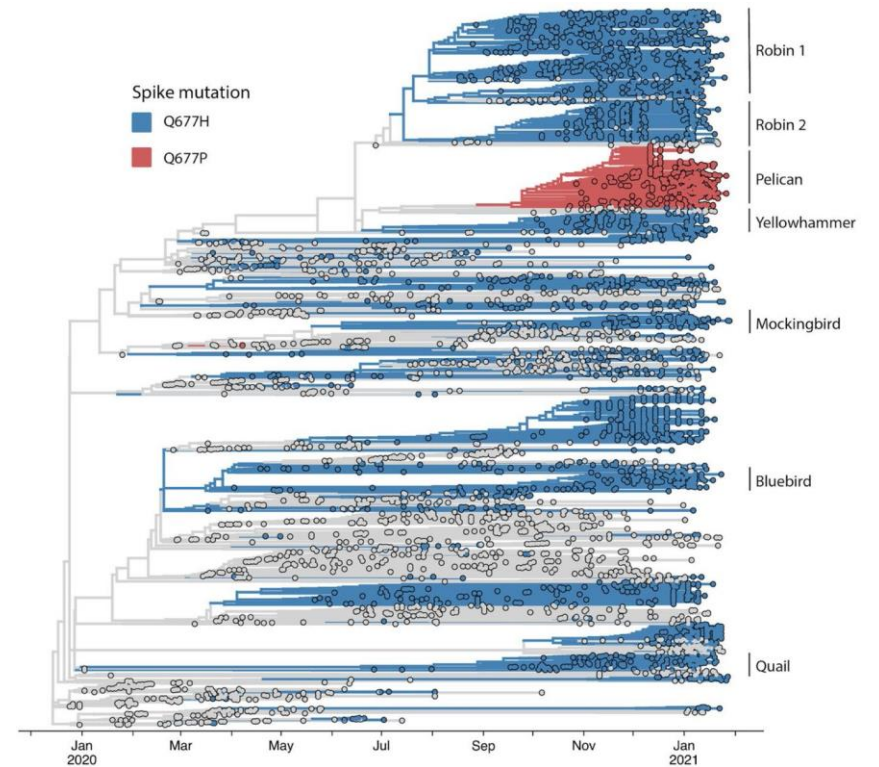
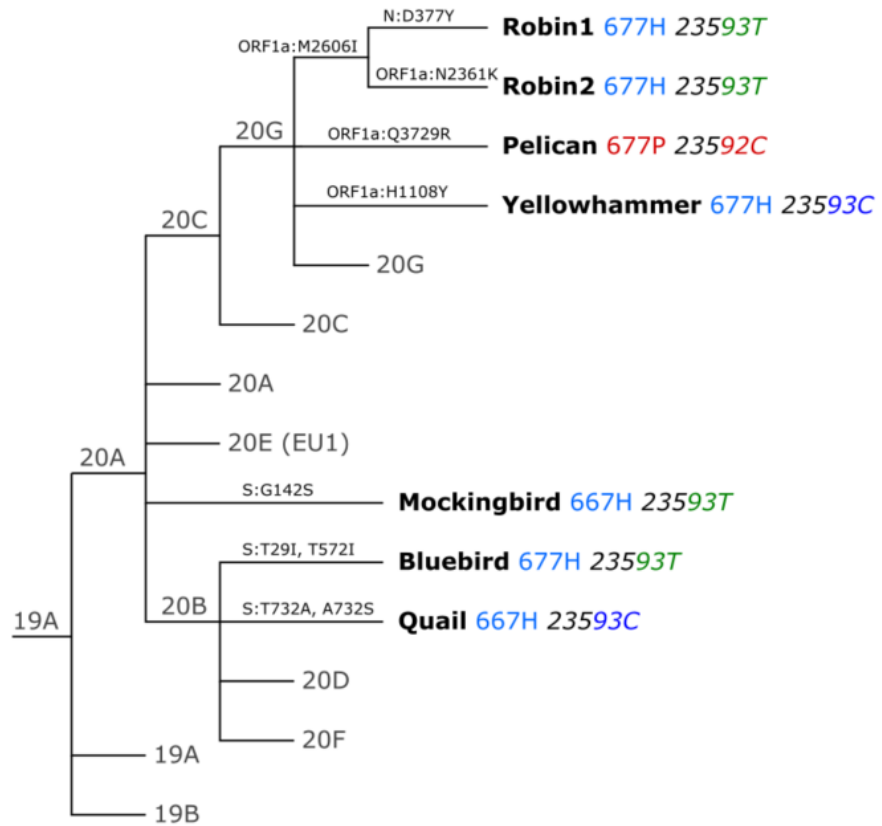
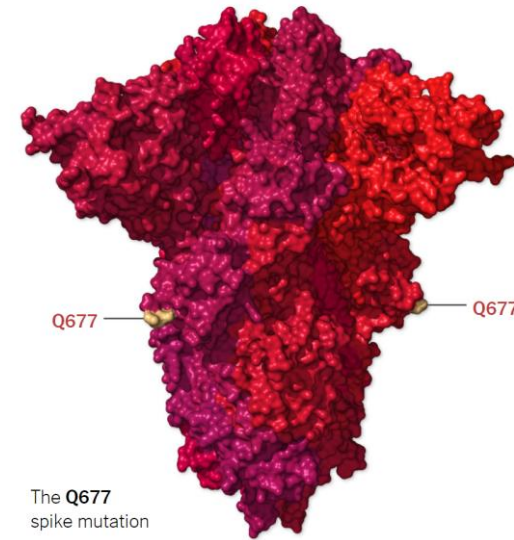


ACE2 binding

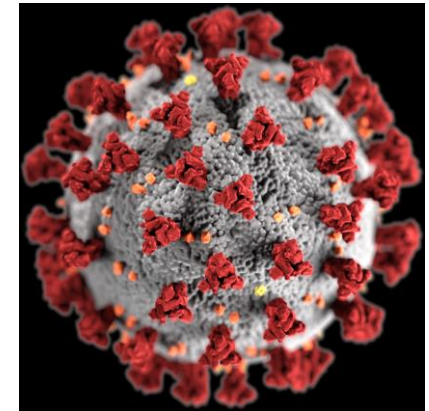
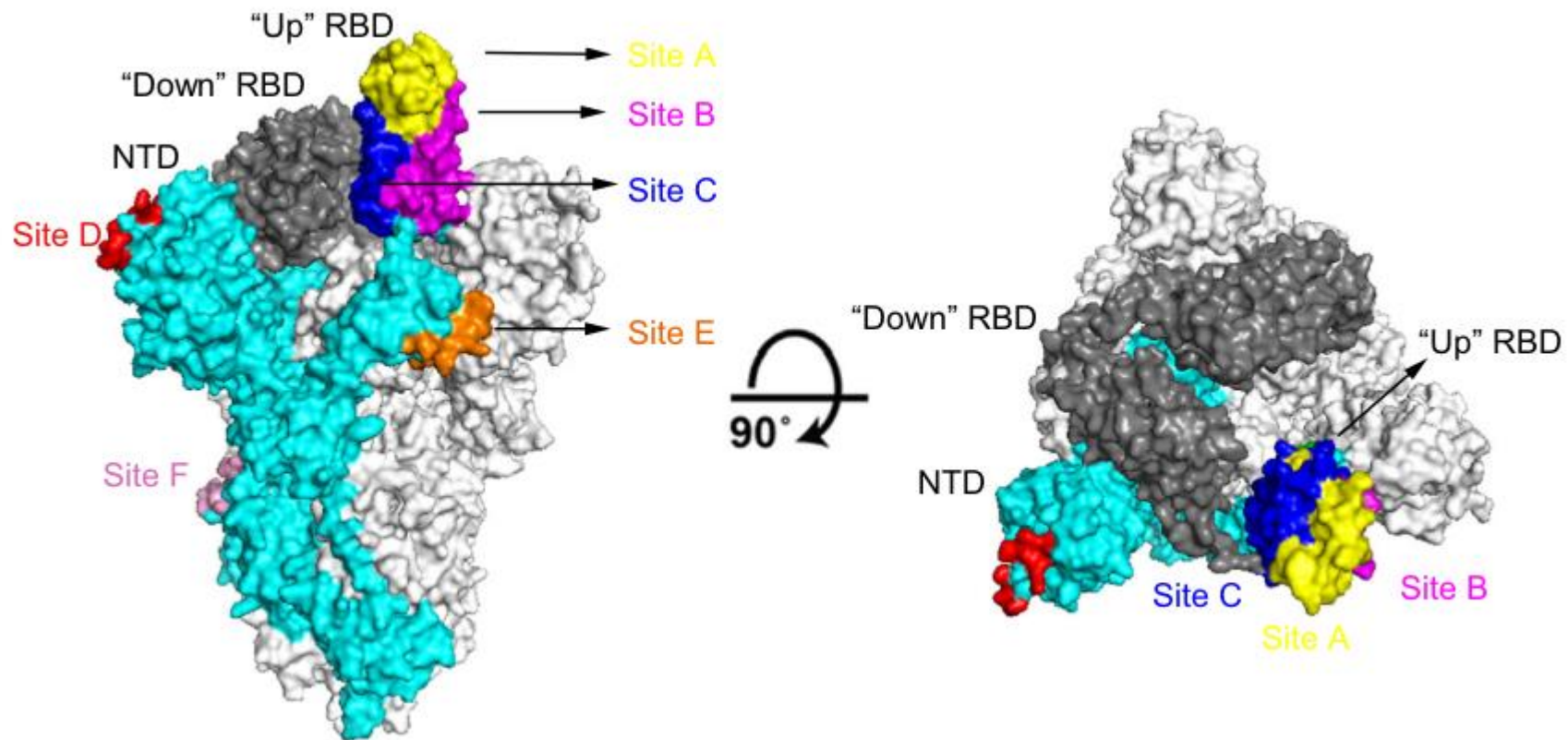
grey is n.d.



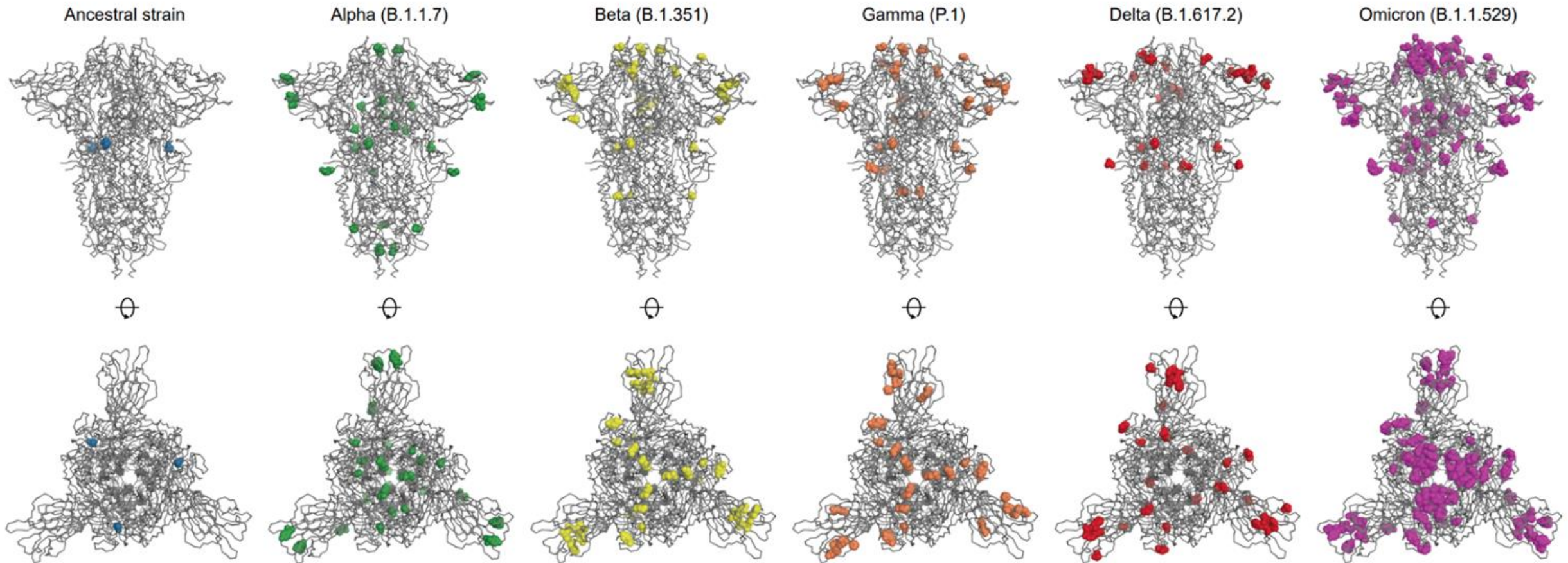
Q677 mutation



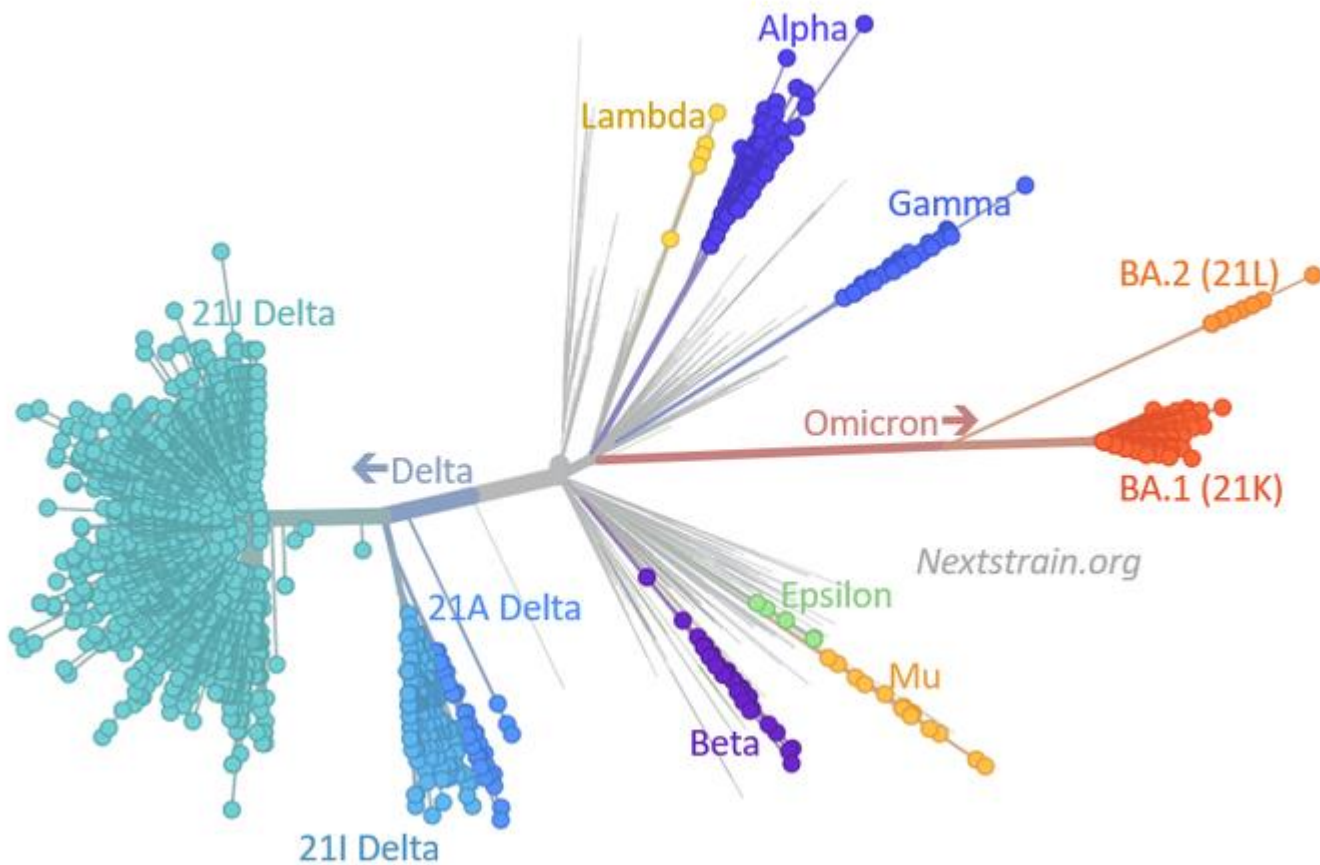
Antibody binding sites on Spike protein



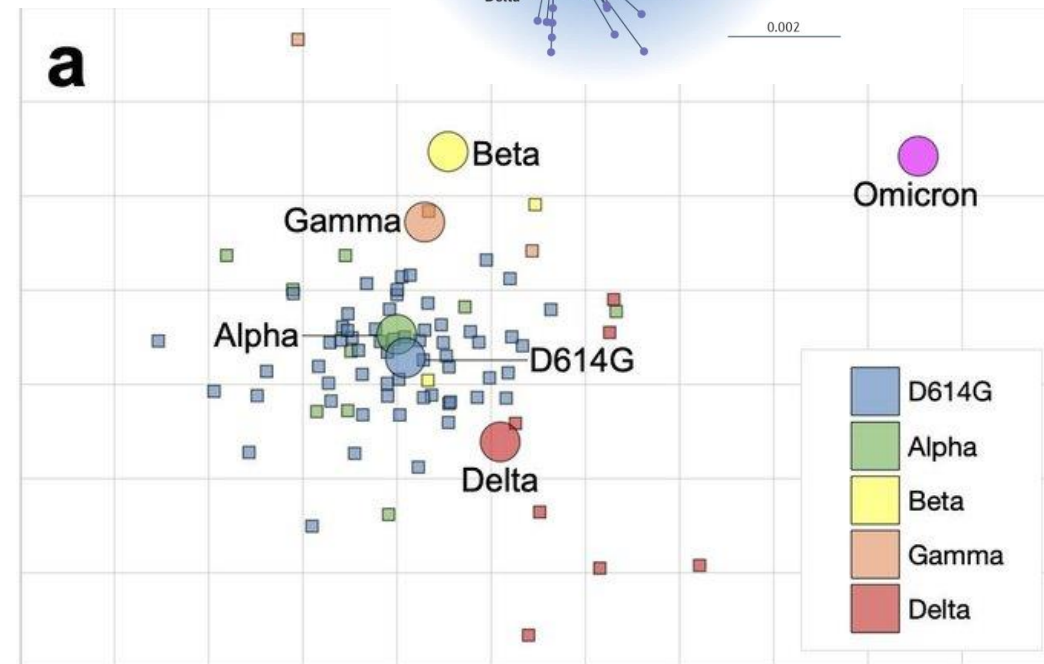
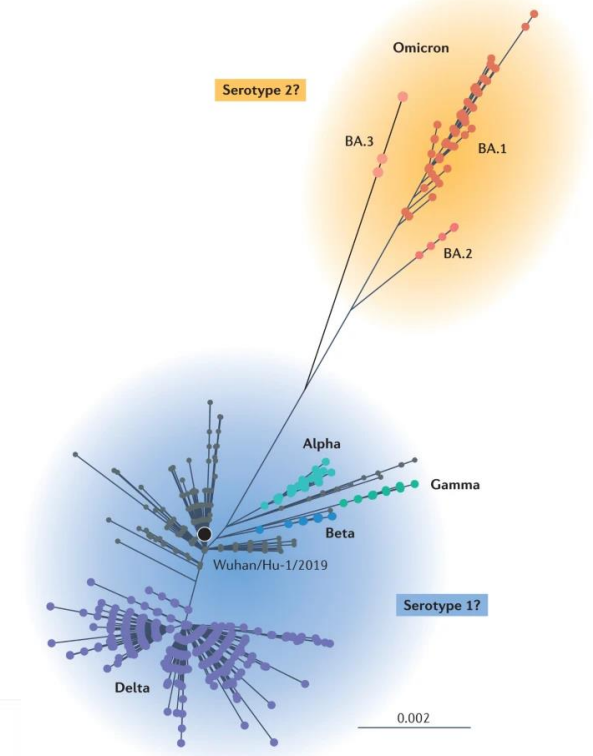
Variant comparison



genetic and immuno relationship



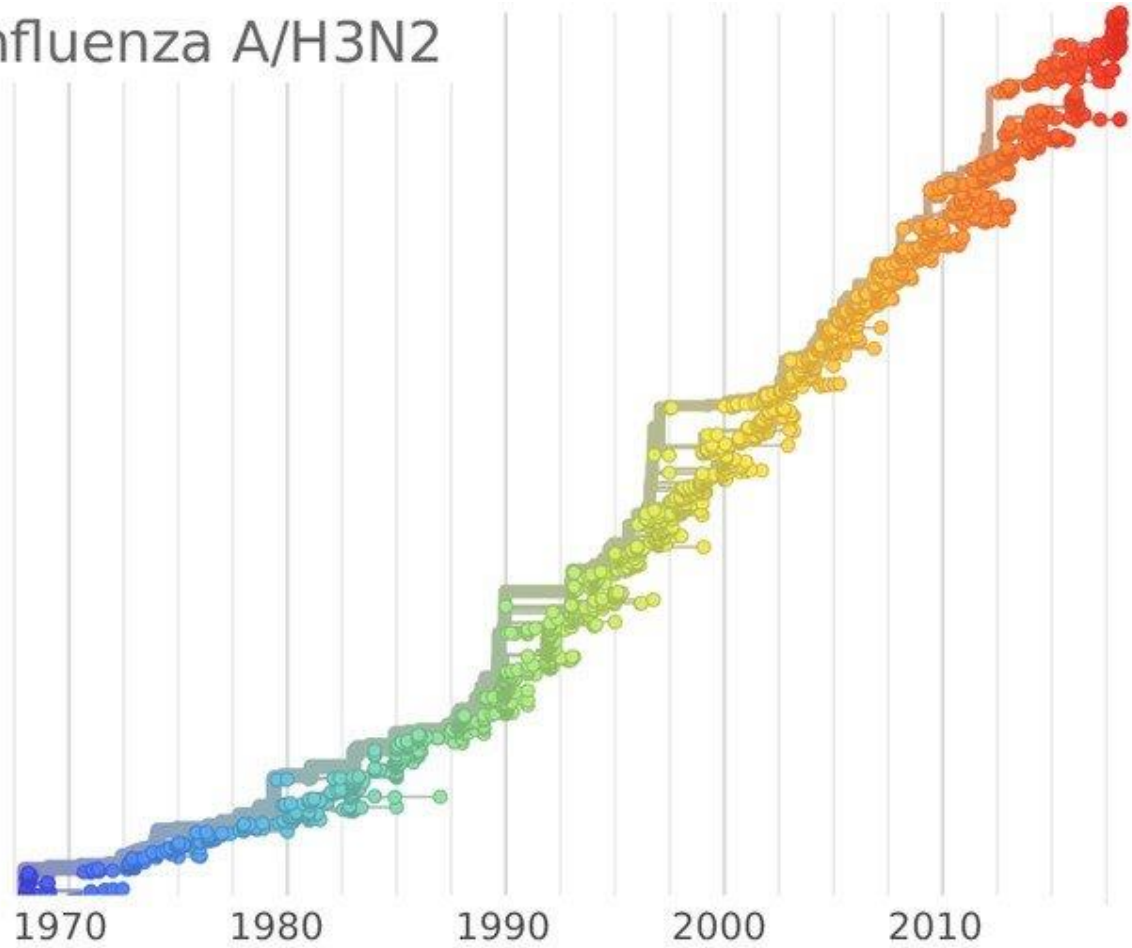
<https://nextstrain.org>



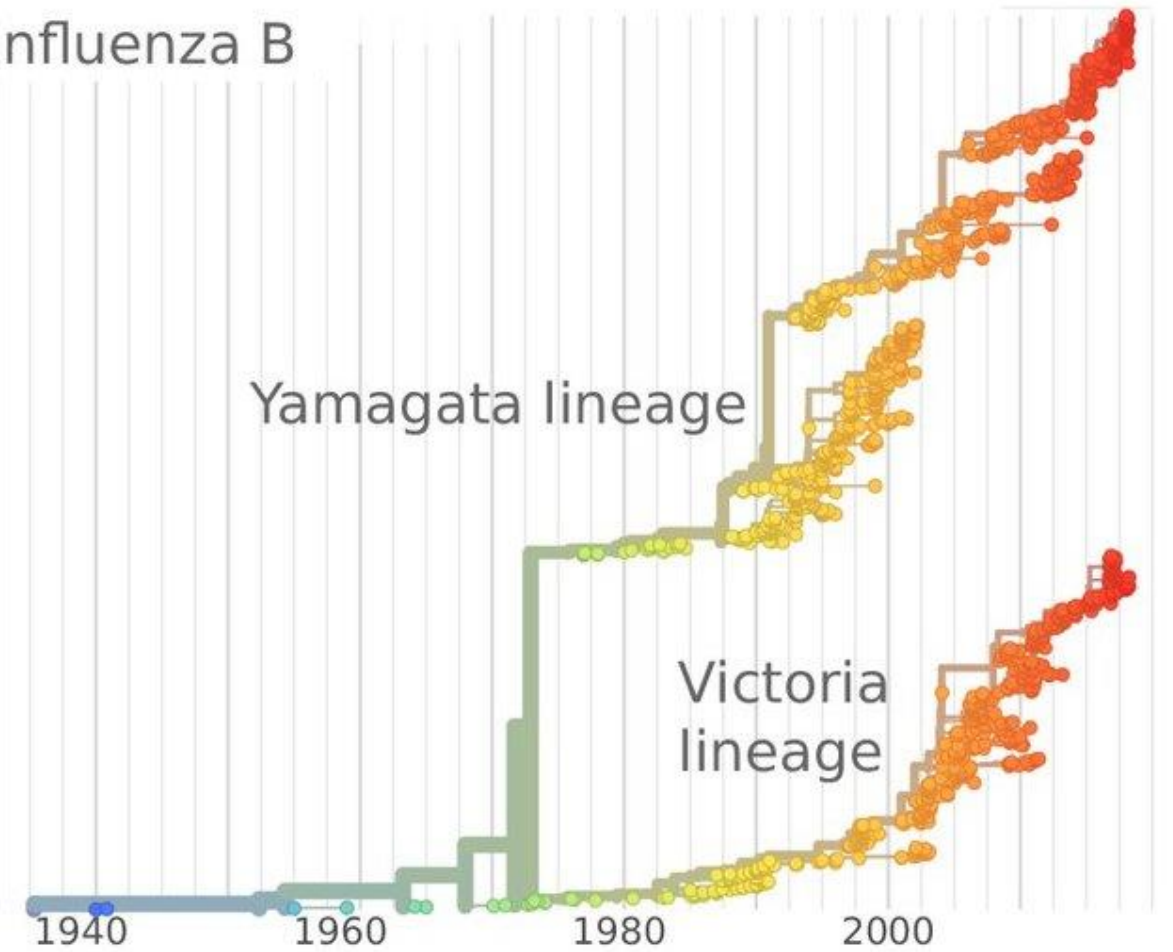
<https://doi.org/10.1101/2022.01.03.21268582>

Adaptive evolution in influenza

Influenza A/H3N2



Influenza B



Influenza subtypes in time

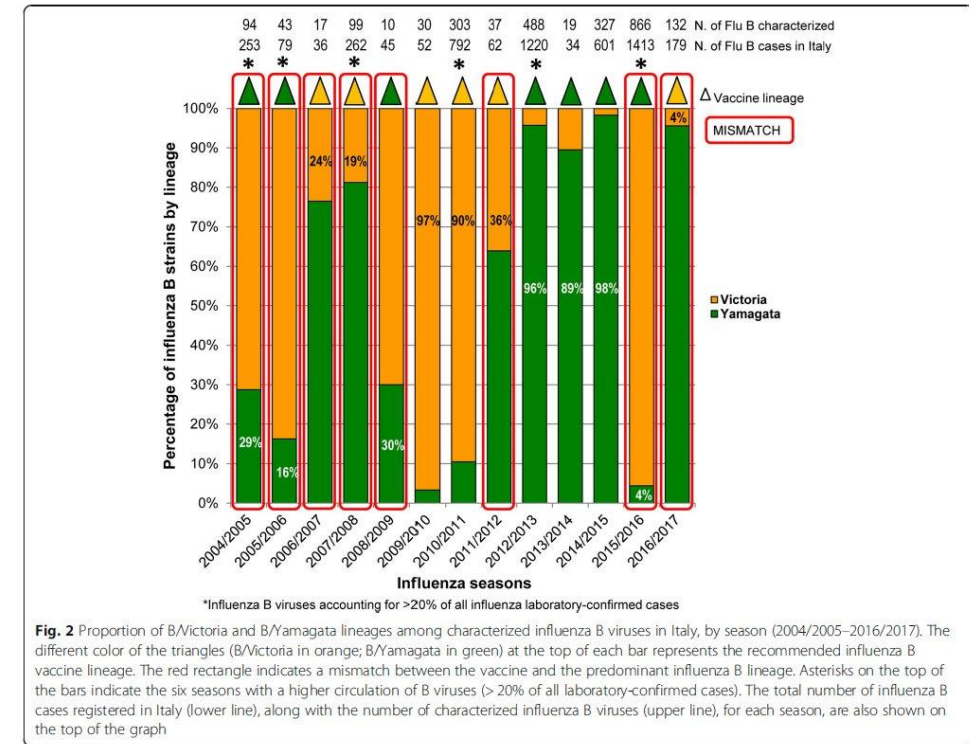
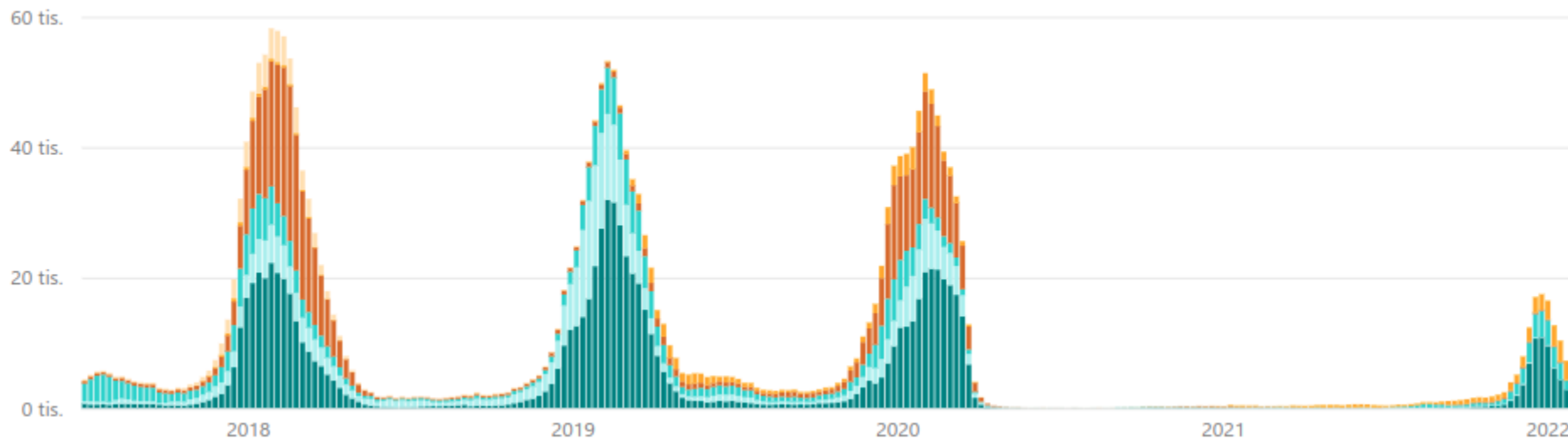
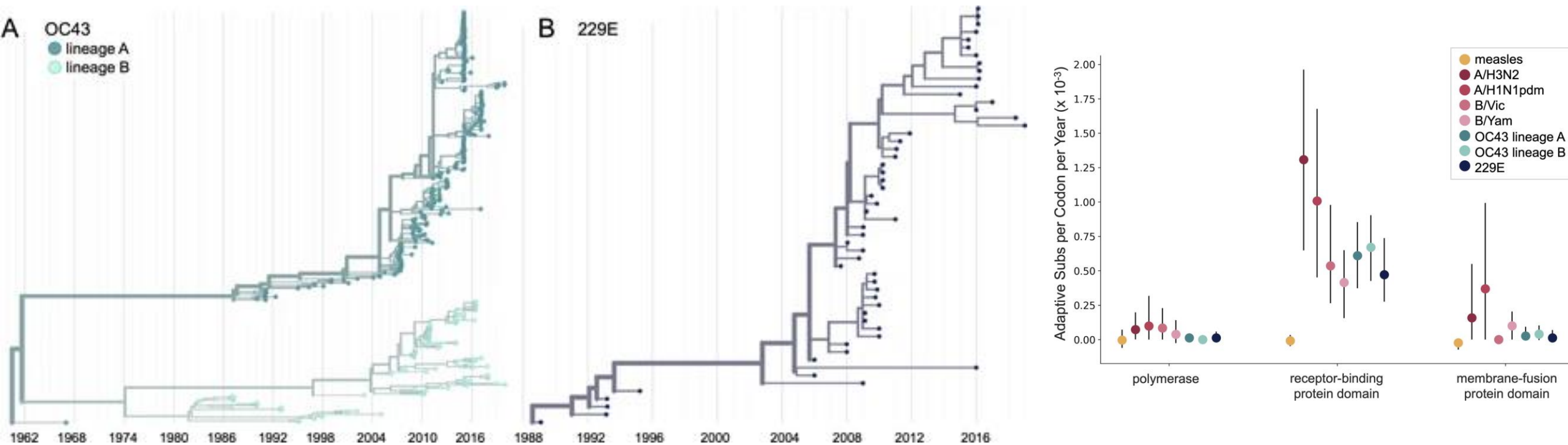


Fig. 2 Proportion of B/Victoria and B/Yamagata lineages among characterized influenza B viruses in Italy, by season (2004/2005–2016/2017). The different color of the triangles (B/Victoria in orange; B/Yamagata in green) at the top of each bar represents the recommended influenza B vaccine lineage. The red rectangle indicates a mismatch between the vaccine and the predominant influenza B lineage. Asterisks on the top of the bars indicate the six seasons with a higher circulation of B viruses (> 20% of all laboratory-confirmed cases). The total number of influenza B cases registered in Italy (lower line), along with the number of characterized influenza B viruses (upper line), for each season, are also shown on the top of the graph

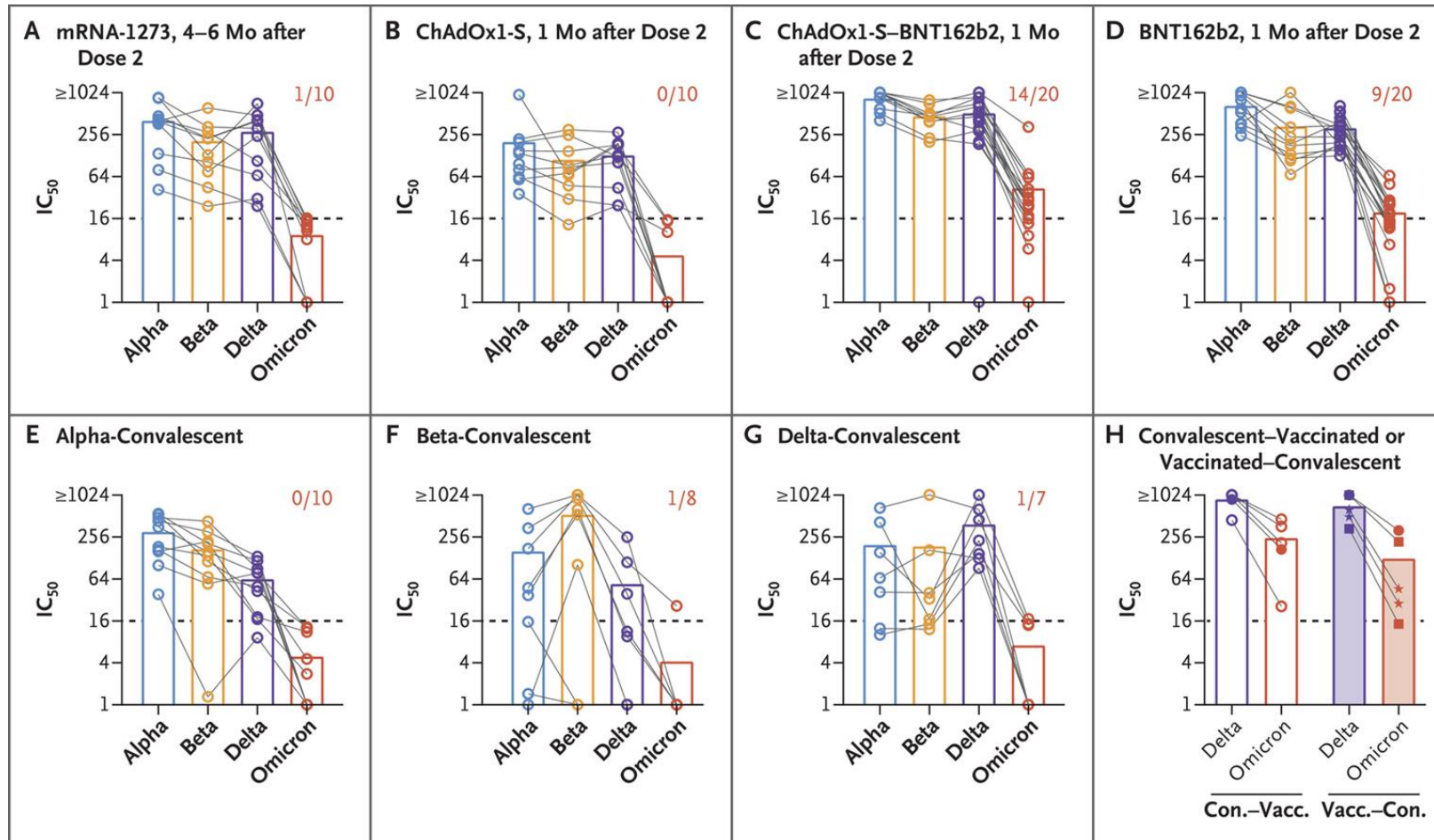
● Influenza A not subtyped ● Influenza A(H1) ● Influenza A(H1N1)pdm09 ● Influenza A(H3) ● Influenza B (lineage not d... ● Influenza B (Victoria) ● Influenza B (Yamagata)



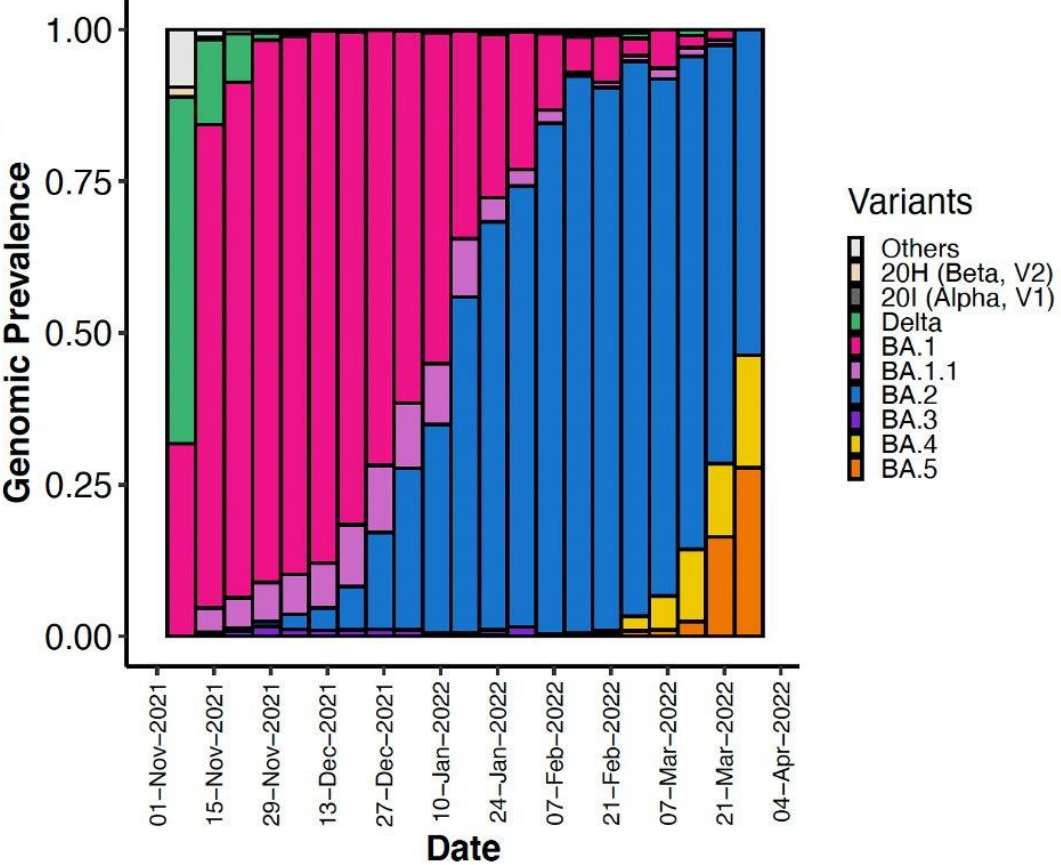
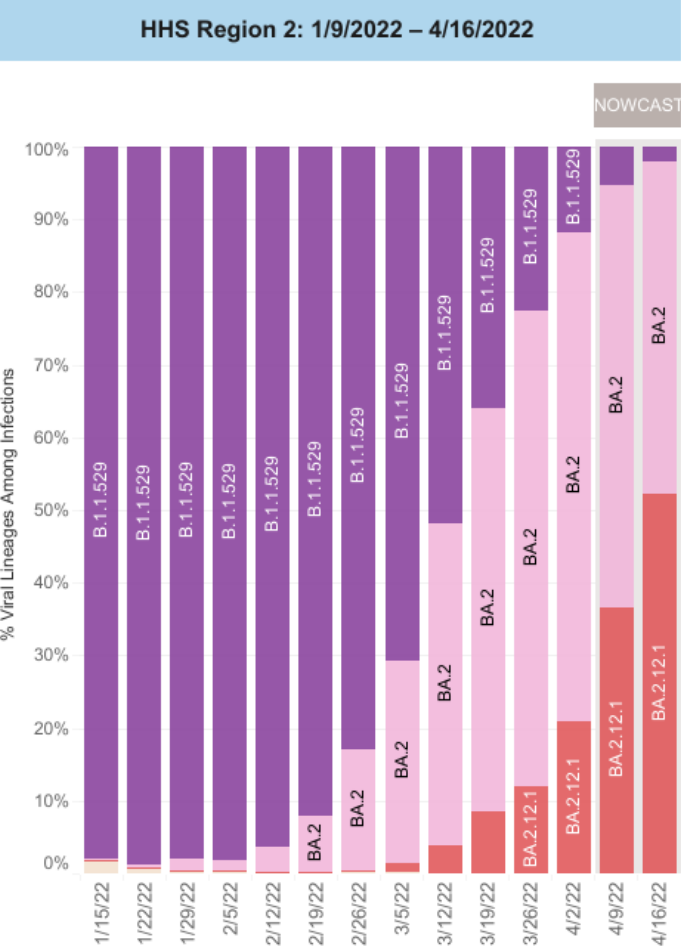
Adaptive evolution is coronaviruses



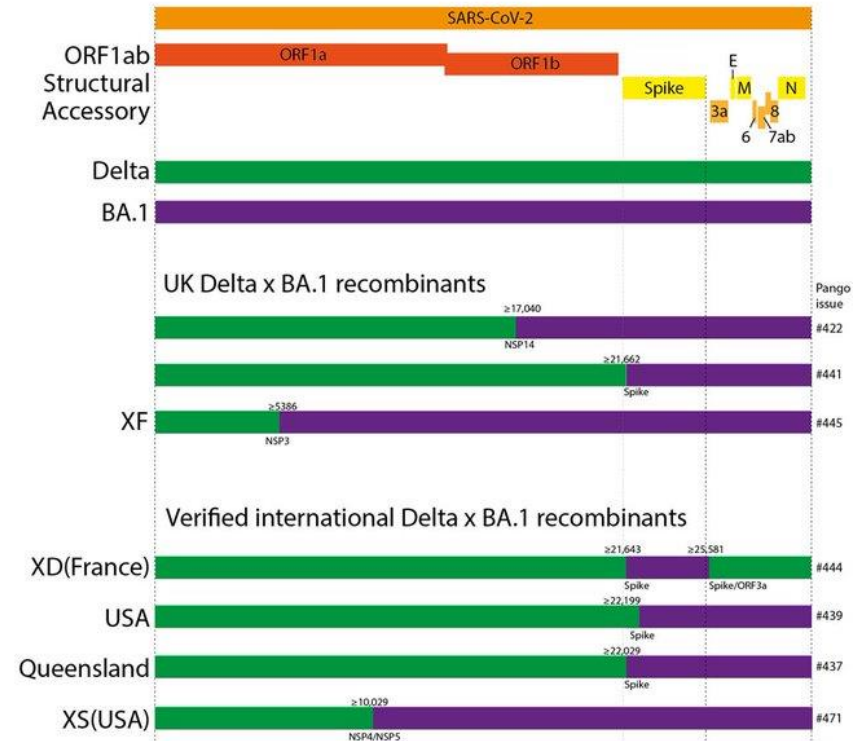
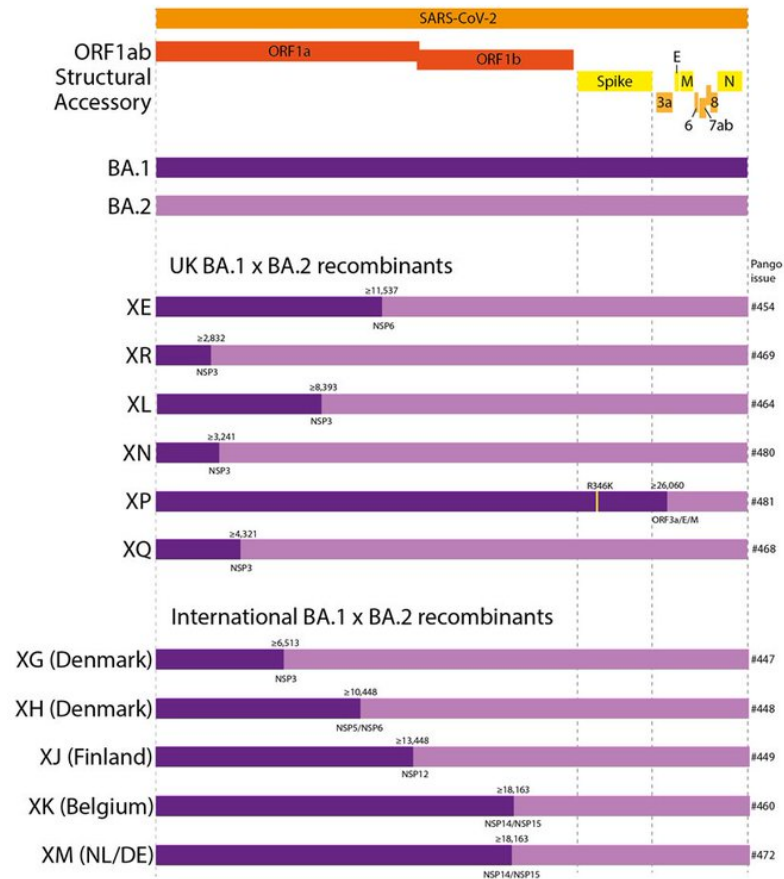
Antibody neutralization of omicron



New omicron subvariants



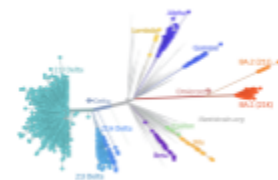
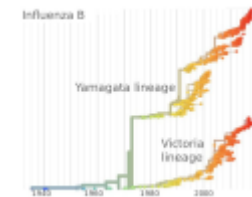
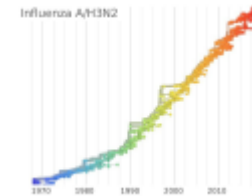
Recombinants



Future

Three possible scenarios

- Current dominant strain (omicron) slowly evolving (Influenza A model)
- Two strains (omicron, delta) periodically dominant (Influenza B model)
- New variant of different serotype (SARS model)



Mutations associated with fitness

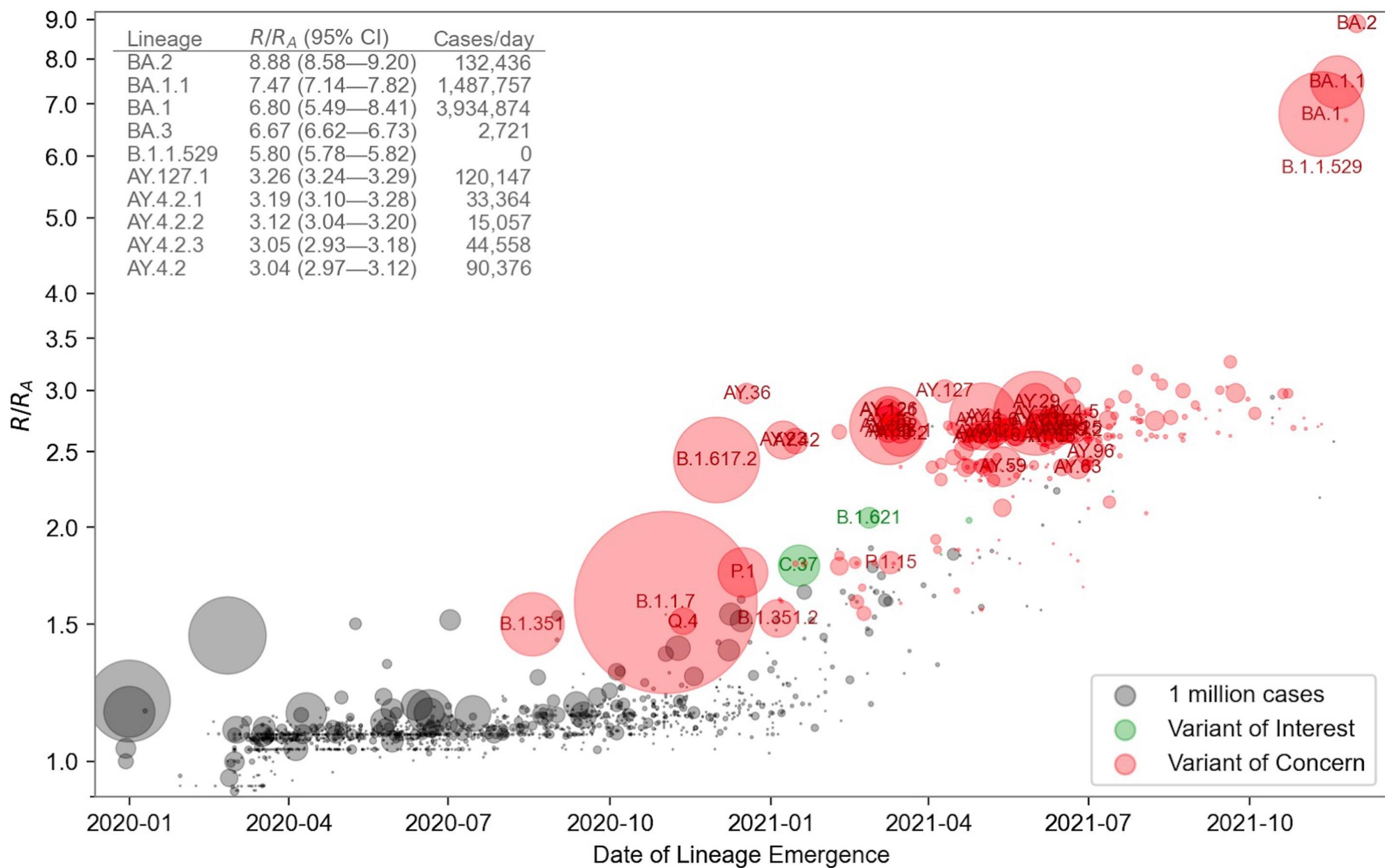
Science

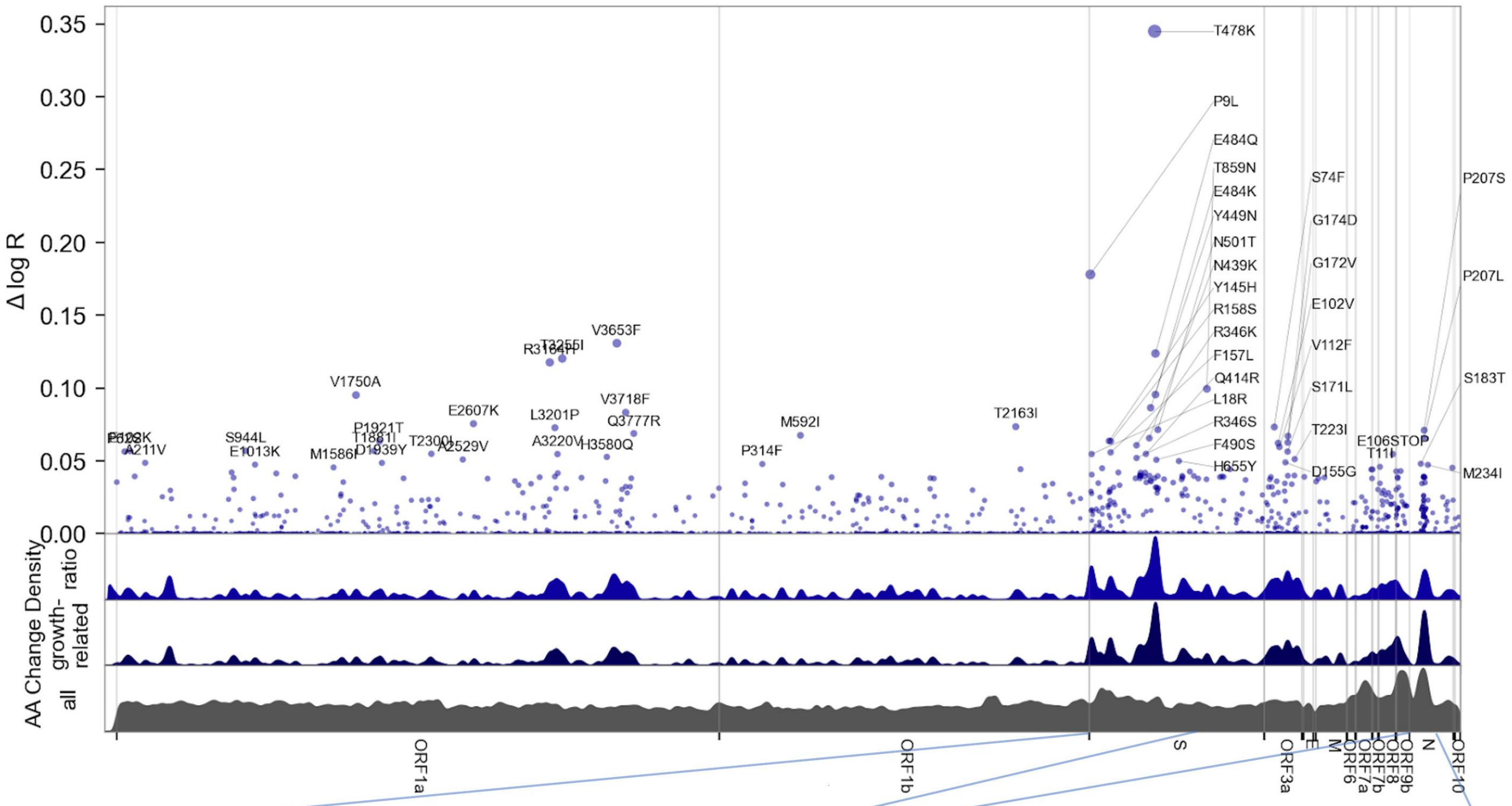
Research Articles

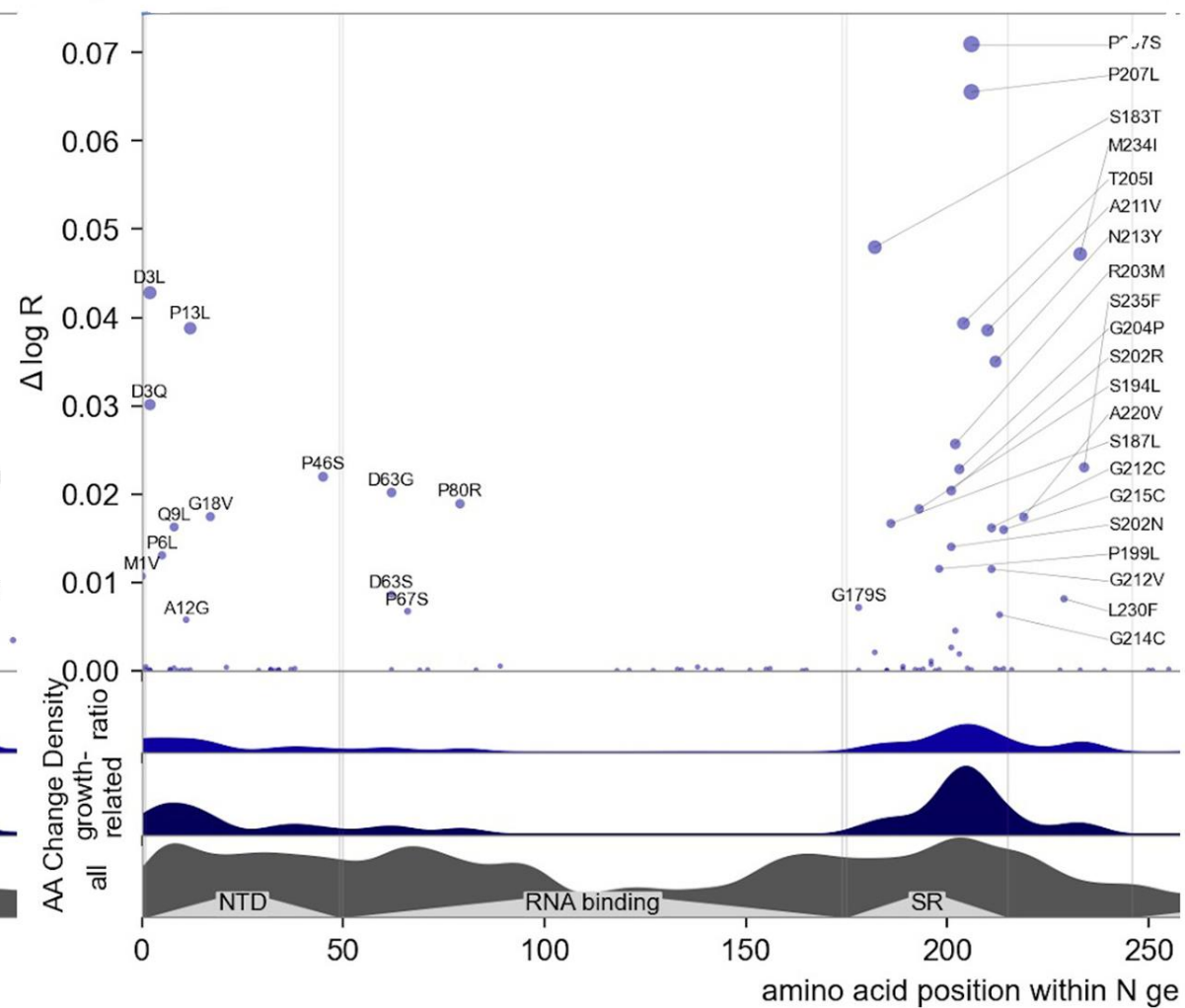
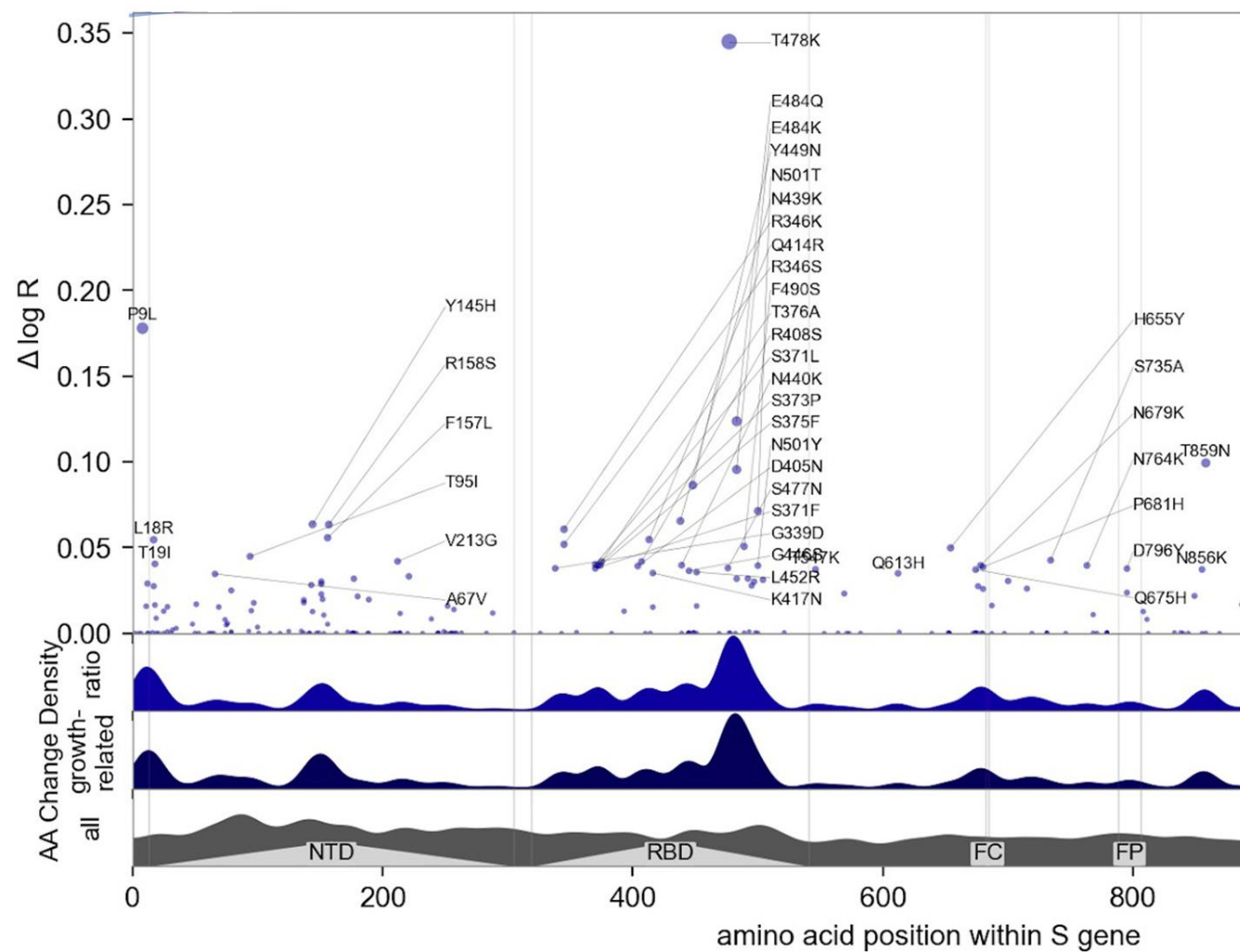
Cite as: F. Obermeyer *et al.*,
Science 10.1126/science.abm1208 (2022).

Analysis of 6.4 million SARS-CoV-2 genomes identifies mutations associated with fitness

Fritz Obermeyer^{1,2*†}, Martin Jankowiak^{1,2}, Nikolaos Barkas¹, Stephen F. Schaffner^{1,3,4}, Jesse D. Pyle^{1,5}, Leonid Yurkovetskiy⁶, Matteo Bosso⁶, Daniel J. Park¹, Mehrtash Babadi¹, Bronwyn L. MacInnis^{1,4,7}, Jeremy Luban^{1,6,7,8}, Pardis C. Sabeti^{1,3,4,7,9‡}, Jacob E. Lemieux^{1,10*‡}

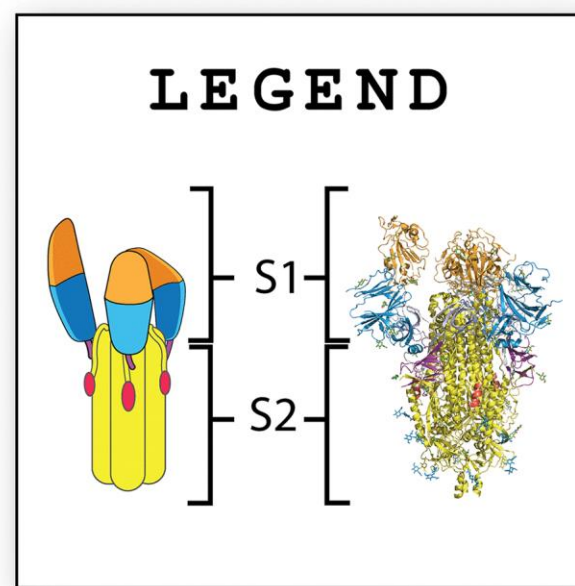
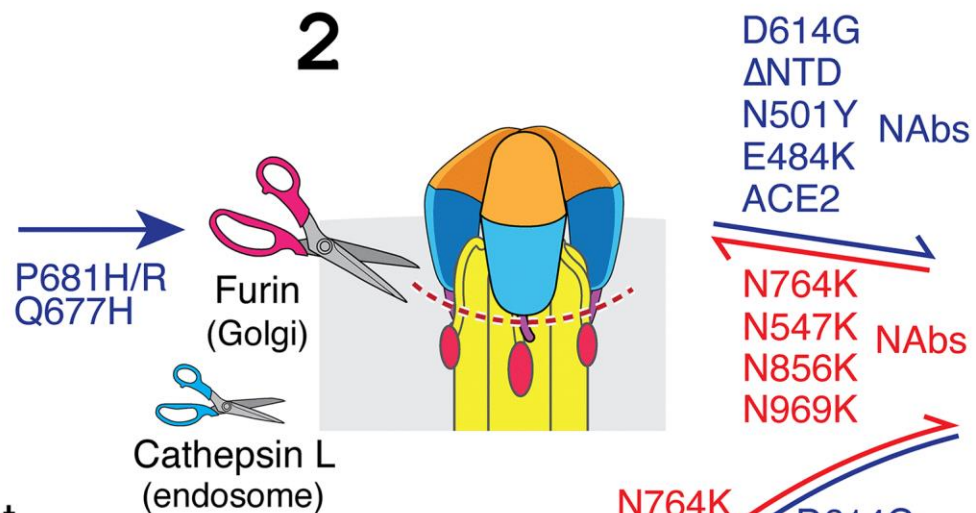
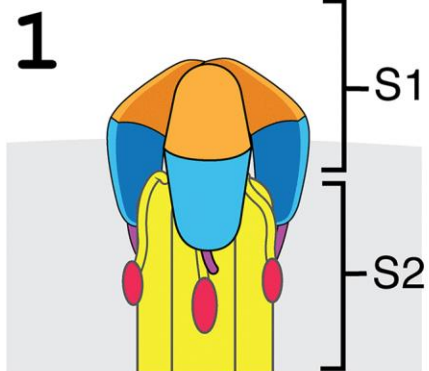




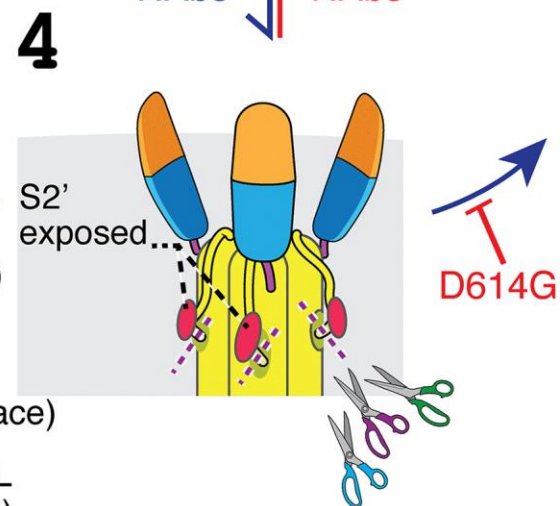
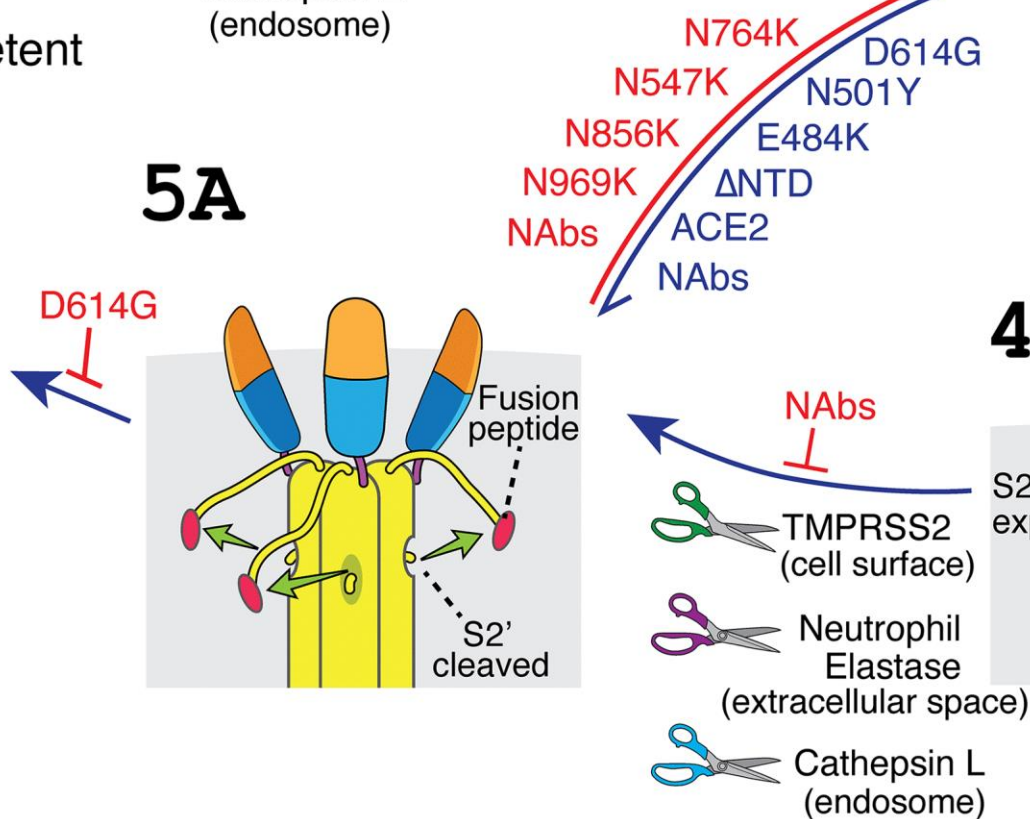
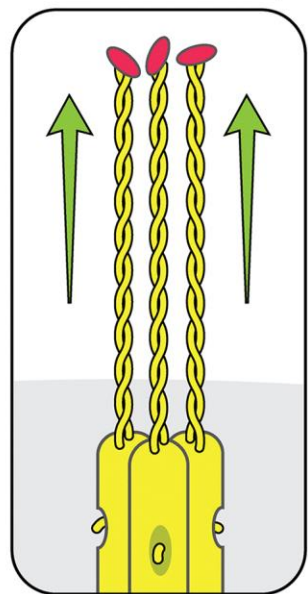


Most significant substitutions

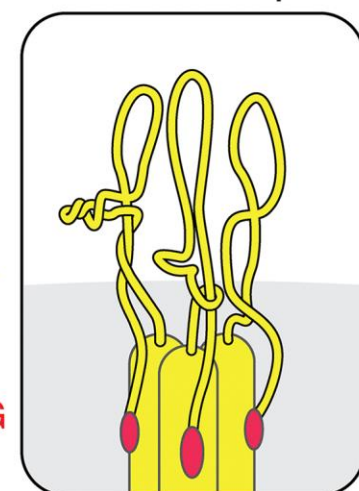
Rank	Gene	Substitution	Fold increase in fitness	Lineages
1	S	H655Y	1.051	33
2	S	T95I	1.046	30
3	ORF1a	P3395H	1.039	5
4	S	N764K	1.04	6
5	ORF1a	K856R	1.039	2
6	S	S371L	1.041	3
7	E	T9I	1.04	5
8	S	Q954H	1.04	5
9	ORF9b	P10S	1.039	25
10	S	L981F	1.04	2
11	N	P13L	1.04	25
12	S	G339D	1.039	4
13	S	S375F	1.04	5
14	S	S477N	1.039	47
15	S	N679K	1.04	11
16	S	S373P	1.04	5
17	M	Q19E	1.039	5
18	S	D796Y	1.038	11
19	S	N969K	1.04	5
20	S	T547K	1.038	3



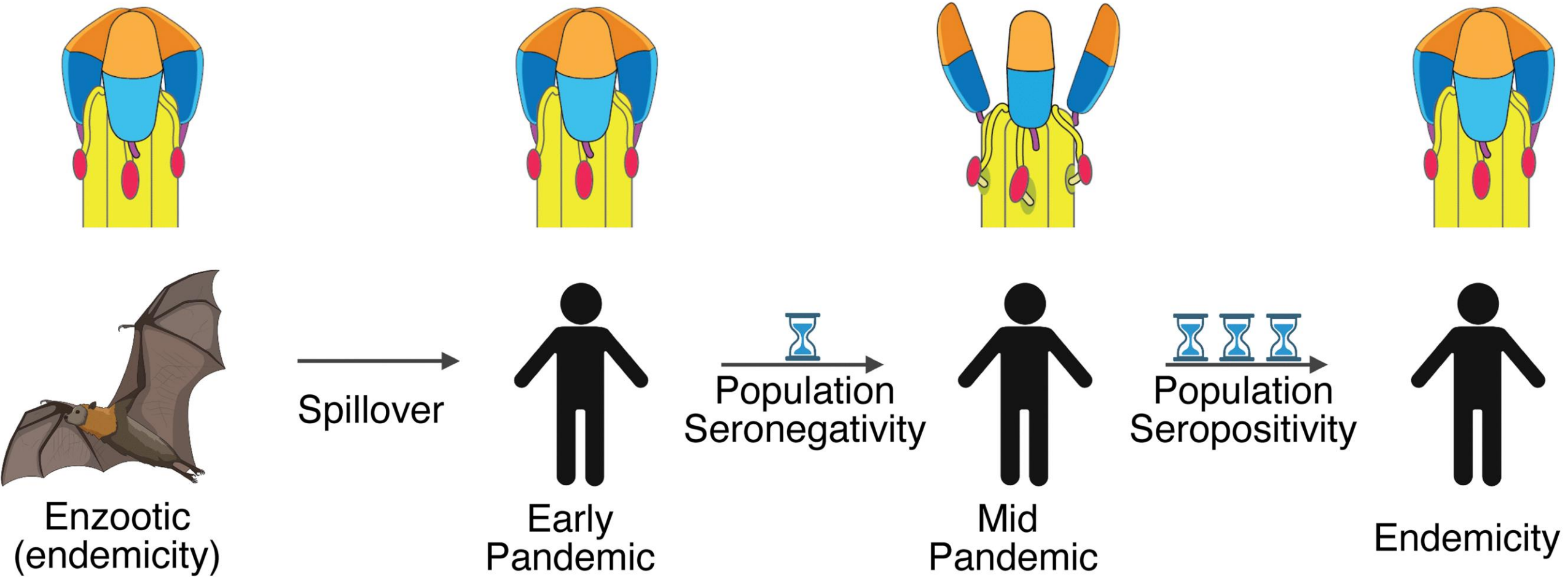
6 Fusion Competent



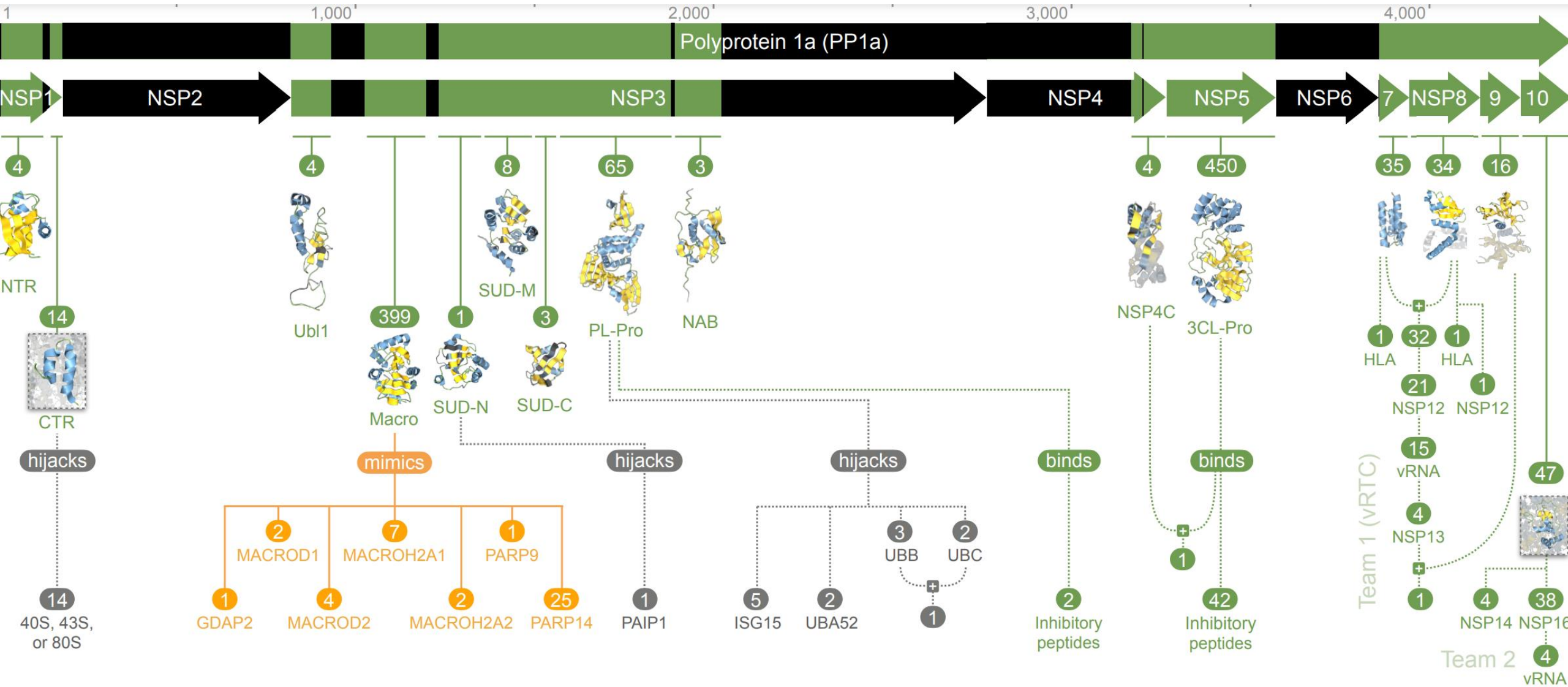
5B Fusion Incompetent



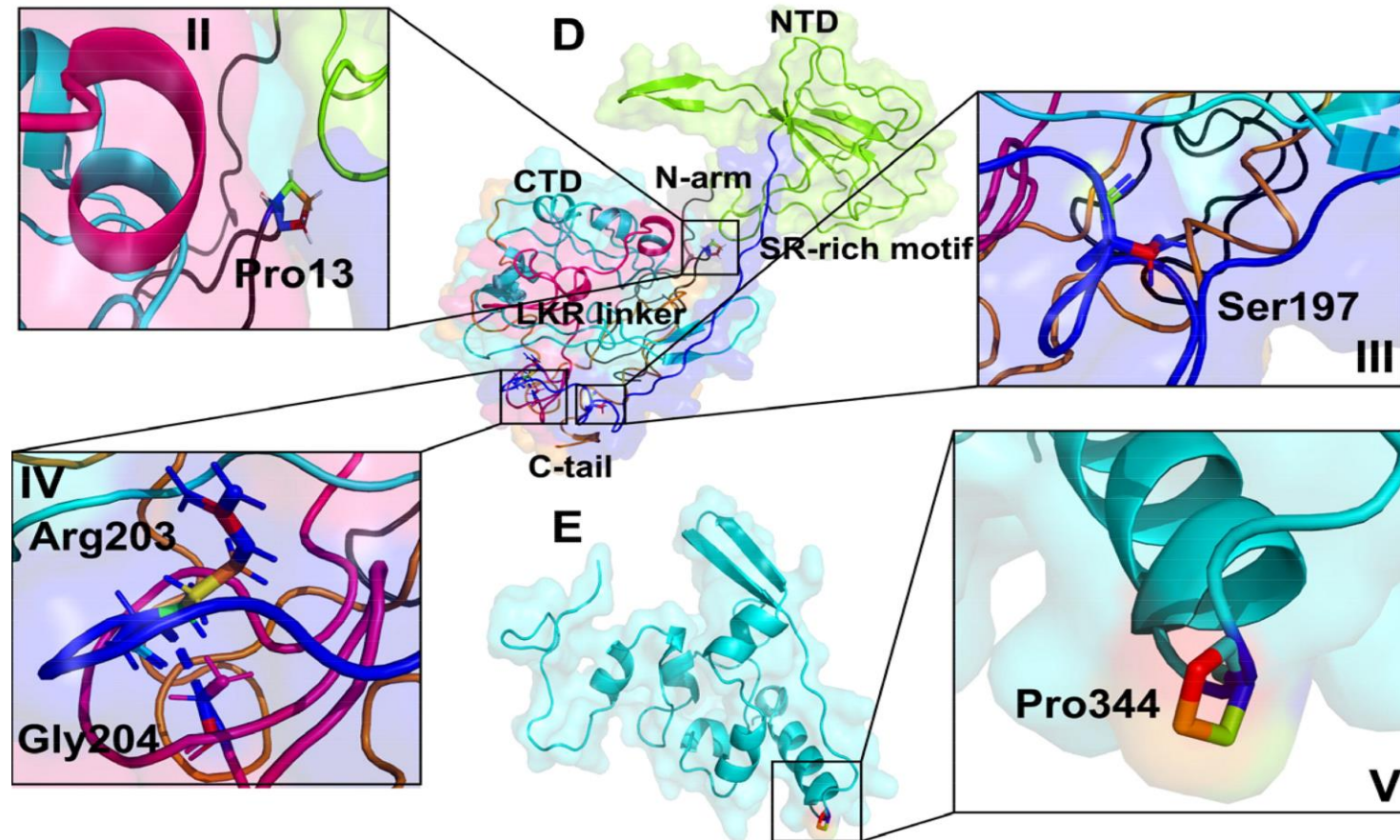
Spike changes



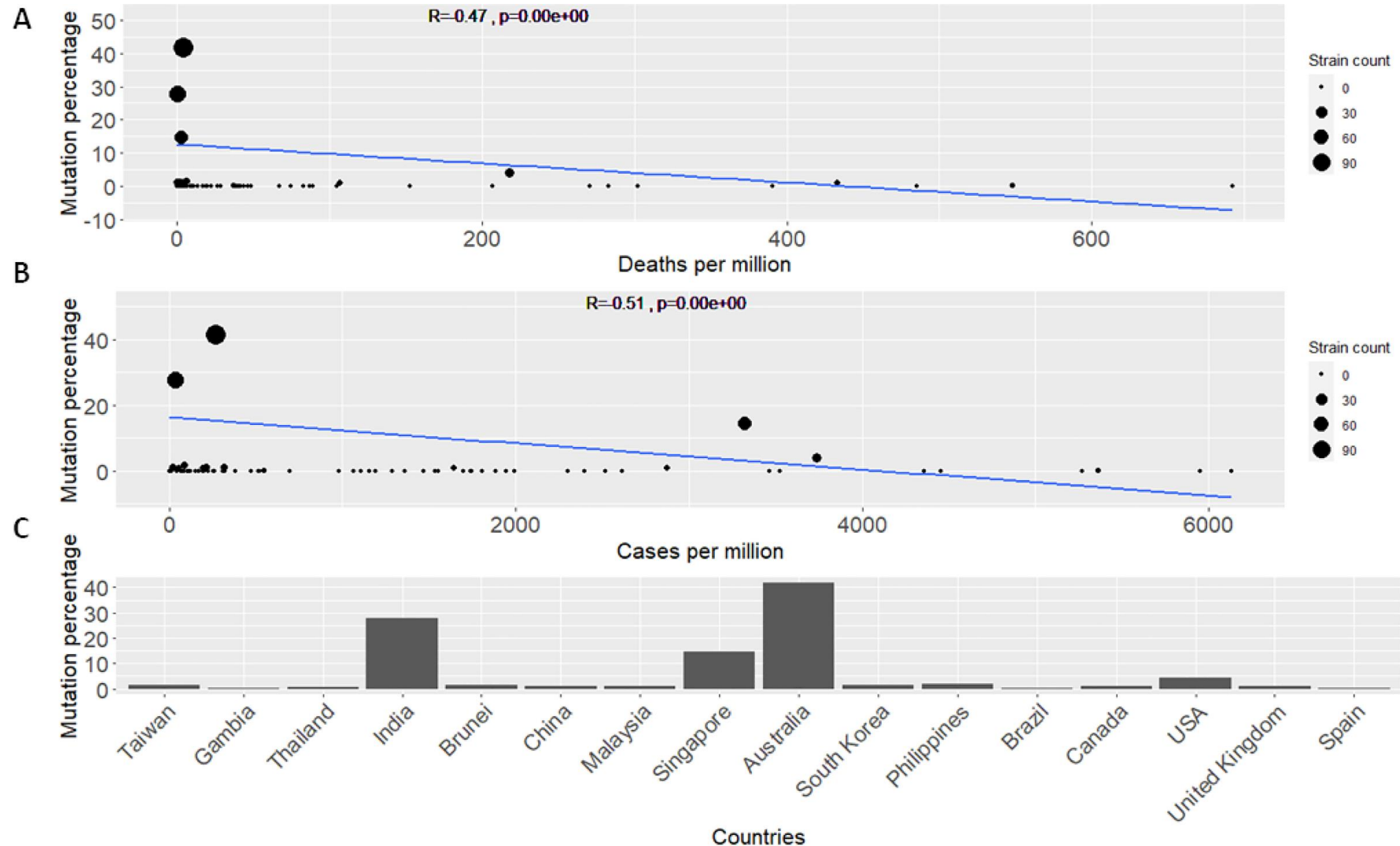
Aquaria covid resource



Important mutations in N protein

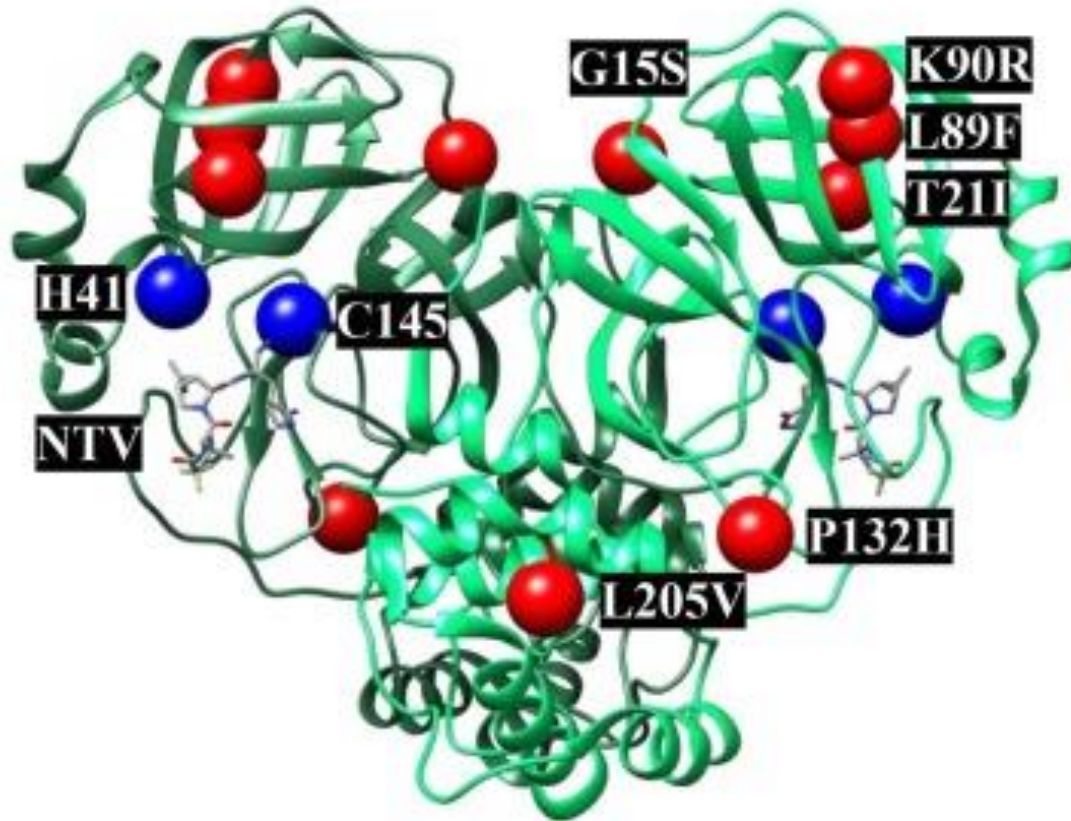


N:P13L



Comparison of Mpro mutants

a

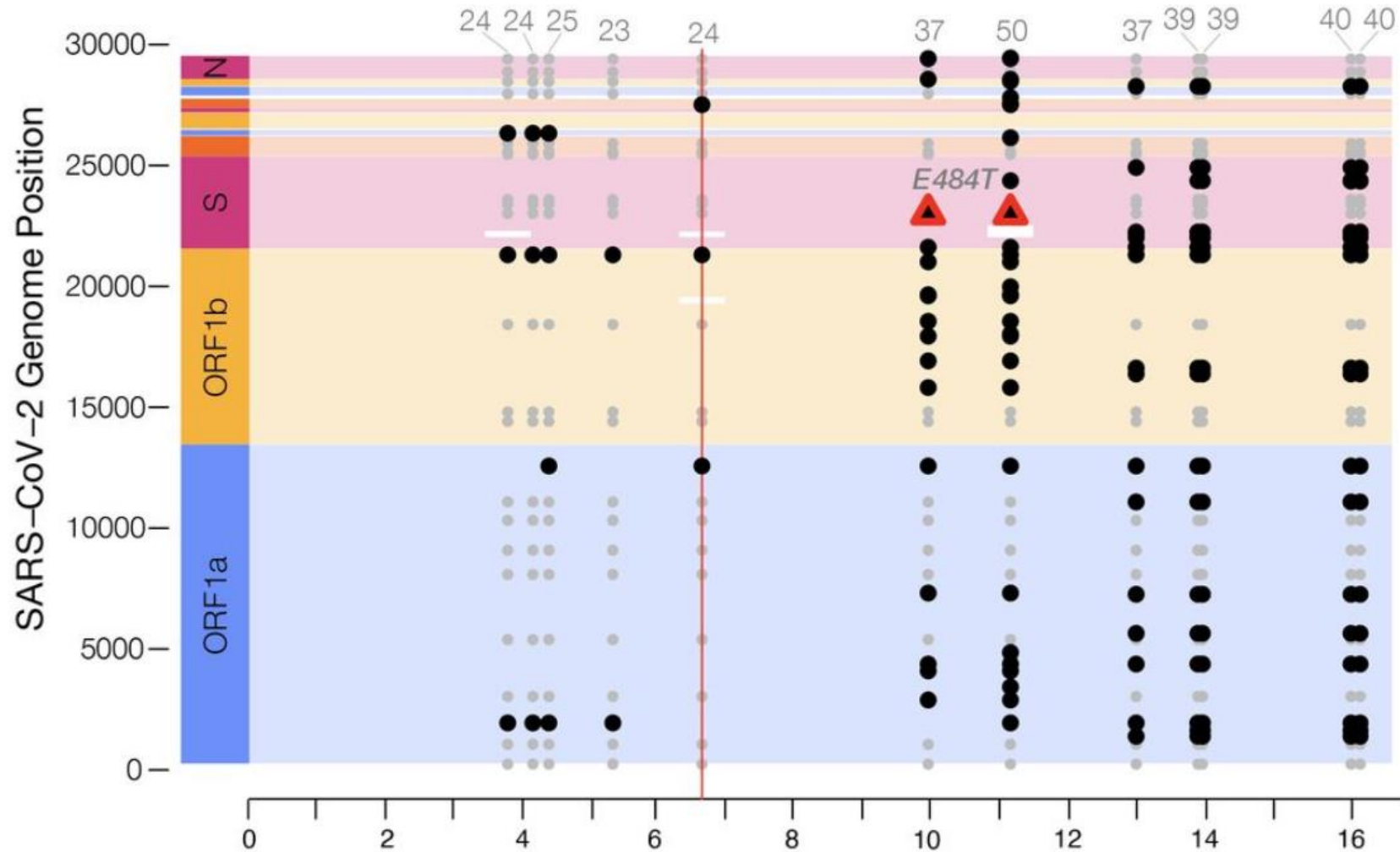


b

Mutation	Lineage
WT	i.a. B.1.617.2 (Delta)
G15S	C.37 (Lambda)
T21I	B.1.1.318
L89F	B.1.2
K90R	B.1.351 (Beta)
P132H	B.1.1.529 (Omicron)
L205V	P.2 (Zeta)

- (a) X-ray co-crystal structure of SARS-CoV-2 Mpro in complex with nirmatrelvir (NTV) (PDB: 7RFW). The sites of mutations (red) and the catalytic dyad (blue) in the two protomers (green) are indicated.
- (b) List of prevalent Mpro mutations and their corresponding SARS-CoV-2 lineage.

Mutations in long term infection



That's it, thanks

COG-CZ

Jan Pačes

Institute of Molecular Genetics AS CR

University of Chemistry and Technology



Project „Enhancing Whole Genome Sequencing (WGS) and/or Reverse Transcription Polymerase Chain Reaction (RT-PCR) national infrastructures and capacities to respond to the Covid-19 pandemic in the European Union and European Economic Area“ had received funding from the European Centre for Disease Prevention and Control under the Grant Agreement number ECDC/HERA/2021/004 ECD.12218.



More information about the project: <http://www.szu.cz/ecdc-1>

NOTE:

Information included reflects only the authors' view and the Centre (ECDC) is not responsible for any use that may be made of the information it contains.