THIRD M.K. PAL MEMORIAL LECTURE
Saha Institute of Nucleur Physics Alumni Association

Sprints & Hops
The Global Journey of a Mutant SARS-CoV-2 *Coronavirus*

Partha P. Majumder
NATIONAL SCIENCE CHAIR

March 20, 2021

This study could not be accomplished without the ideas provided by and the work carried out by

Nidhan K. Biswas

Arnab Ghosh

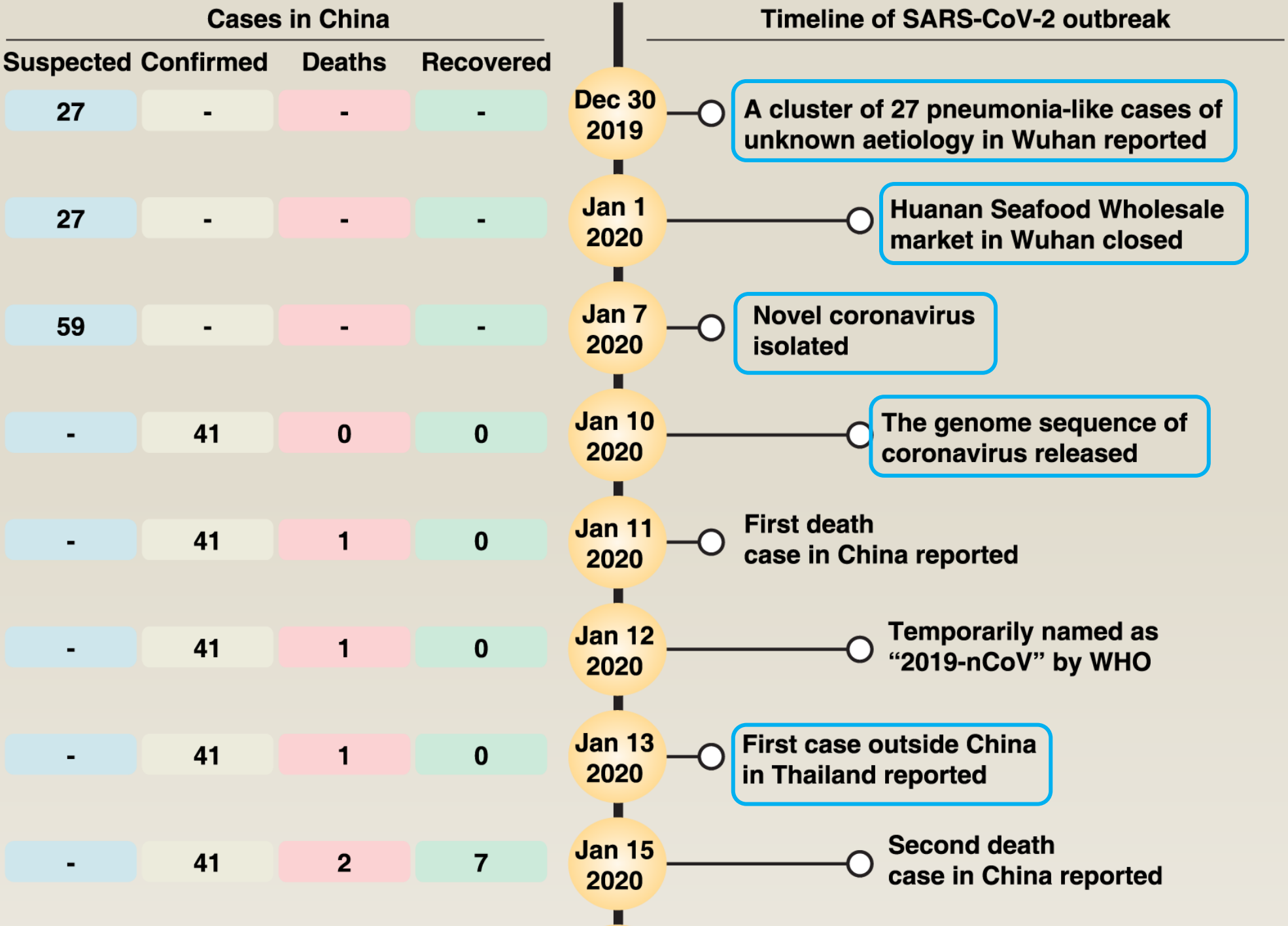
Chandrika Bhattacharyya

Chitrarpita Das

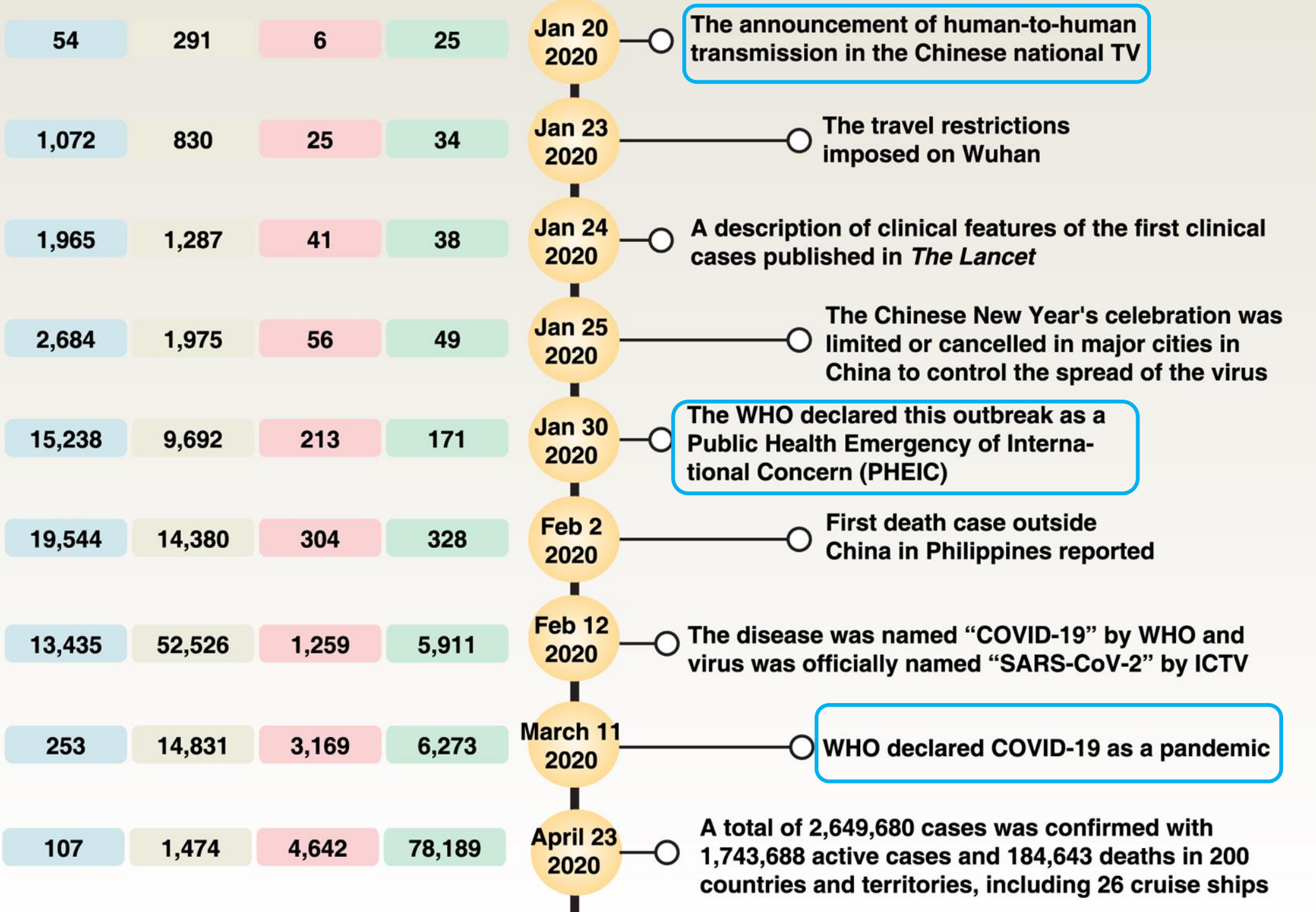
Analabha Basu

Souvik Mukherjee

Animesh K. Singh



Tang D, Comish P, Kang R (2020) The hallmarks of COVID-19 disease. PLOS Pathogens 16(5): e1008536. https://doi.org/10.1371/journal.ppat.1008536 https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1008536



Tang D, Comish P, Kang R (2020) The hallmarks of COVID-19 disease. PLOS Pathogens 16(5): e1008536. https://doi.org/10.1371/journal.ppat.1008536 https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1008536

Global Cases: 122,416,036

29,730,475 US
11,871,390 Brazil
11,555,284 India
4,397,816 Russia
4,299,200 United Kingdom
4,242,156 France
3,332,418 Italy
3,212,332 Spain
2,971,633 Turkey
2,654,734 Germany
2,324,426 Colombia
2,234,913 Argentina
2,187,910 Mexico
2,036,700 Poland
…..

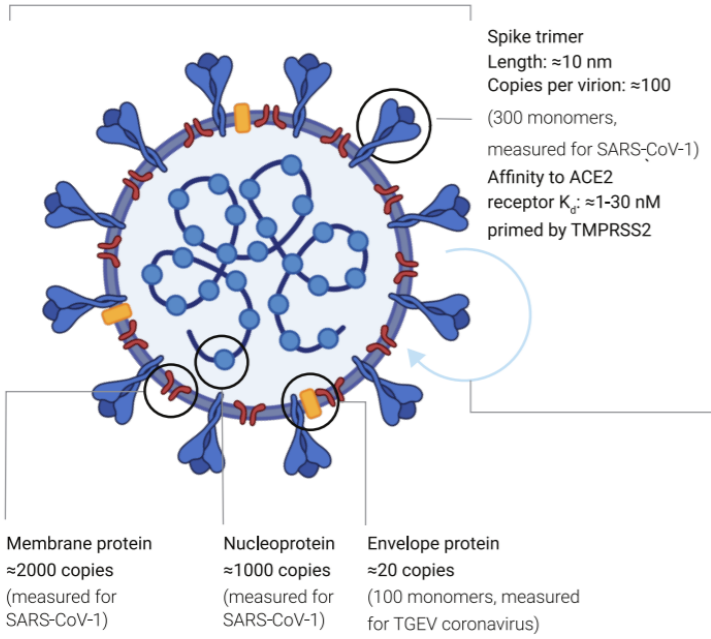
20th March 2021

https://coronavirus.jhu.edu/map.html
Johns Hopkins University and Medicine
Coronavirus Resource Center

Global Deaths: 2,703,126
(2.21%)

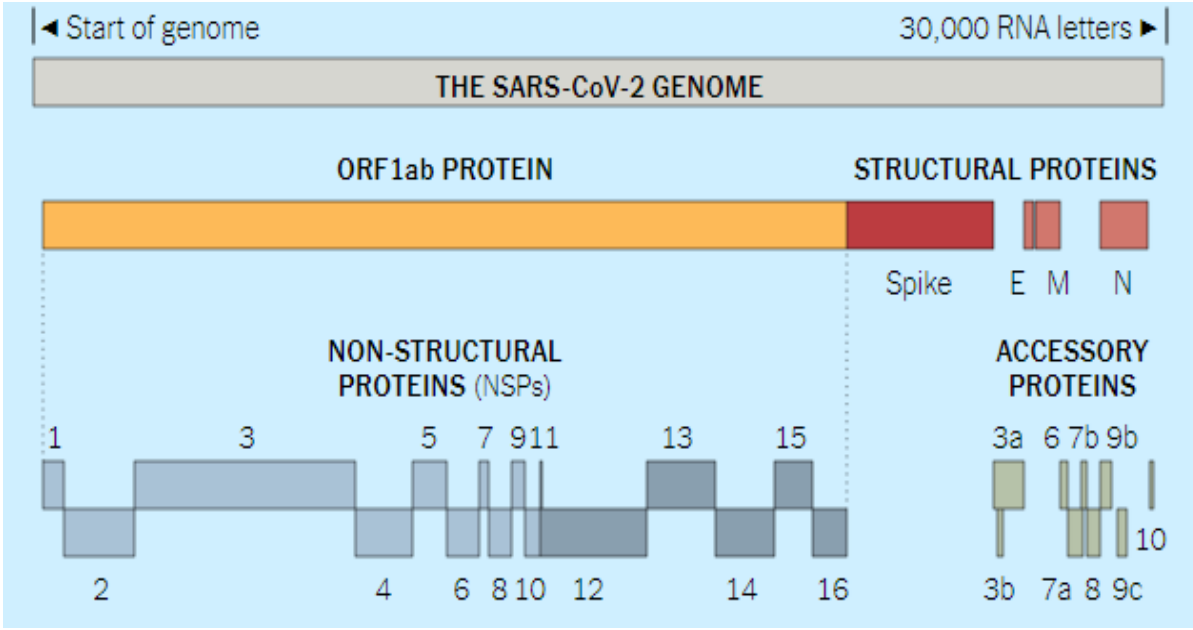
SARS-CoV-2 is a single stranded (+) sense RNA virus with a genome length of about 30000 bases

Diameter: ≈100 nm
Volume: ~10^6 nm^3 = 10^{-3} fL
Mass: ~10^3 MDa ≈ 1 fg



Replication Timescales

in tissue-culture
Virion entry into cell: ~10 min (measured for SARS-CoV-1)
Eclipse period: ~10 hrs (time to make intracellular virions)
Burst size: ~10^3 virions (measured for MHV coronavirus)



It encodes **29 structural and non-structural proteins**, including ORF1a/b polyprotein, which includes a RNA dependent RNA polymerase (RdRp) and other non-structural proteins, **spike (S) glycoprotein**, envelope (E), membrane (M) and the nucleocapsid (N) proteins

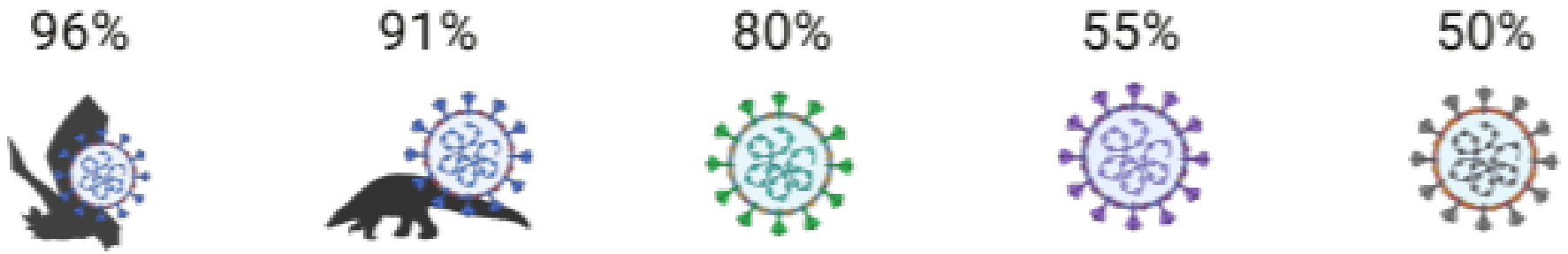
Nucleotide Identity: Inferring evolutionary descent

Genome

Source: 10.7554/eLife.57309

```
                                                    614
SARS_CoV2-QHD43416      VITPGTNTSNQVAVLYQDVNCTEVPVA
QIA48614-Pangolin       VITPGTNTSNQVAVLYQDVNCTEVPMA
SPIKE_BCHK3-Q3LZX1-Bat  VITPGTNASSEVAVLYQDVNCTDVPTA
SPIKE_CVHSA-P59594      VITPGTNASSEVAVLYQDVNCTDVSTA
AAU04646-Civet          VITPGTNASSEVAVLYQDVNCTDVSTL
                        *******:* .:***********:*
```

Nucleotide identity to SARS-CoV-2



- **Viruses cannot survive without their hosts.**
- **To replicate, they have to enter host cells and use host-cell machinery.**
- **Viruses don’t want to kill their hosts.**
- **Host mortality: SARS-CoV-1=11%, MERS=34%, SARS-CoV-1 <3%**

SARS-CoV-2's Preference for Host Cell Types

Host Cells

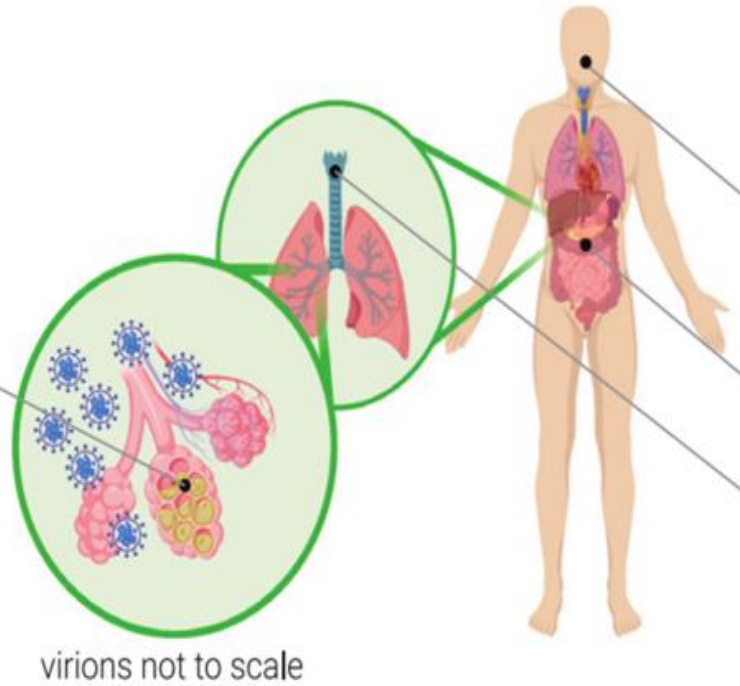
(tentative list; number of cells per person)

Type I & II pneumocytes (~10^{11} cells)

Alveolar macrophage (~10^{10} cells)

Mucous cell in nasal cavity (~10^{9} cells)

Host cell volume: ~10^{3} μm^3 = 10^{3} fL



virions not to scale

Concentration

maximal observed values following diagnosis

(Woelfel et al. 2020; Kim et al. 2020; Pan et al. 2020)

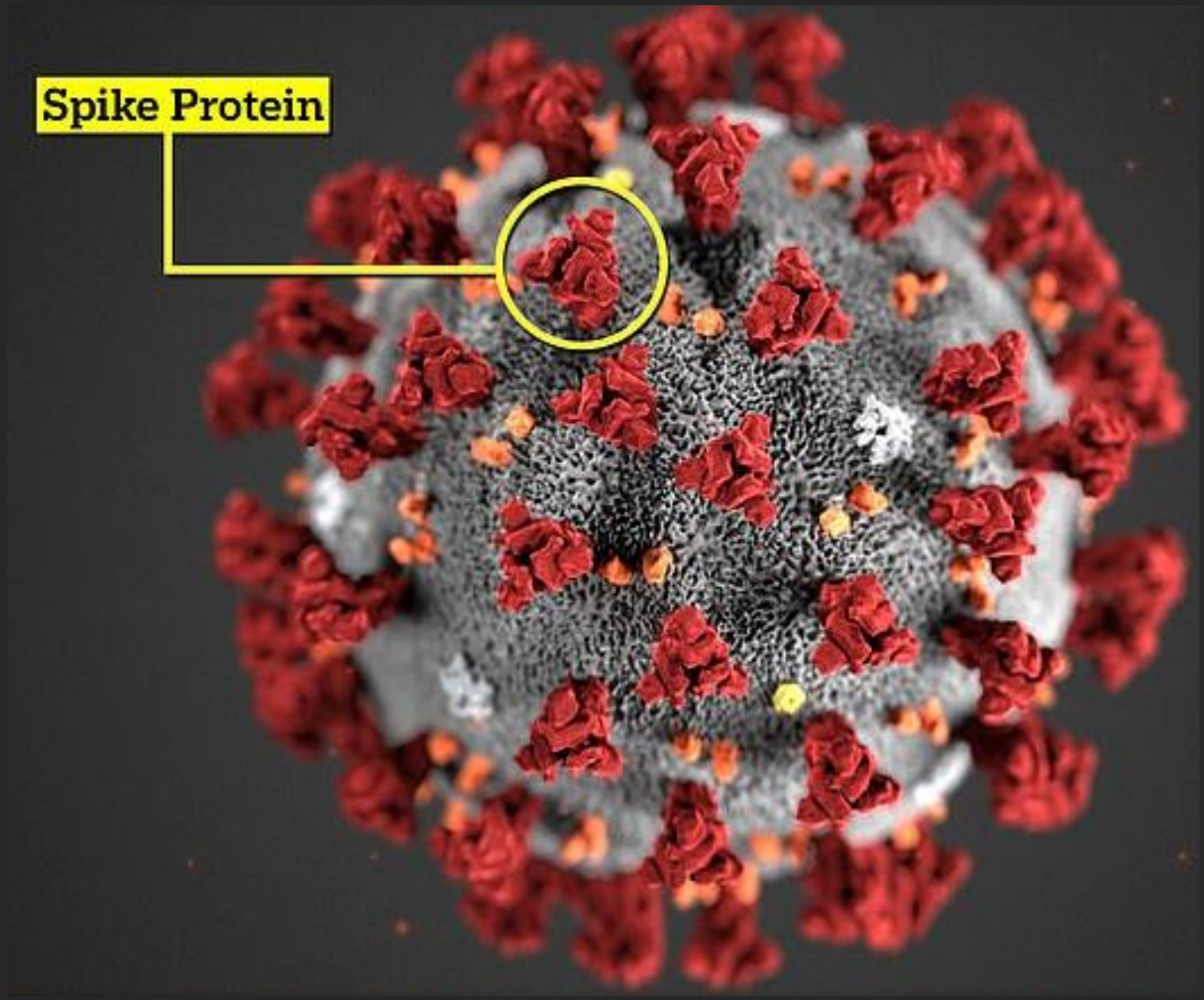
Nasopharynx: 10^{6}-10^{9} RNAs/swab

Throat: 10^{4}-10^{8} RNAs/swab

Stool: 10^{4}-10^{8} RNAs/g

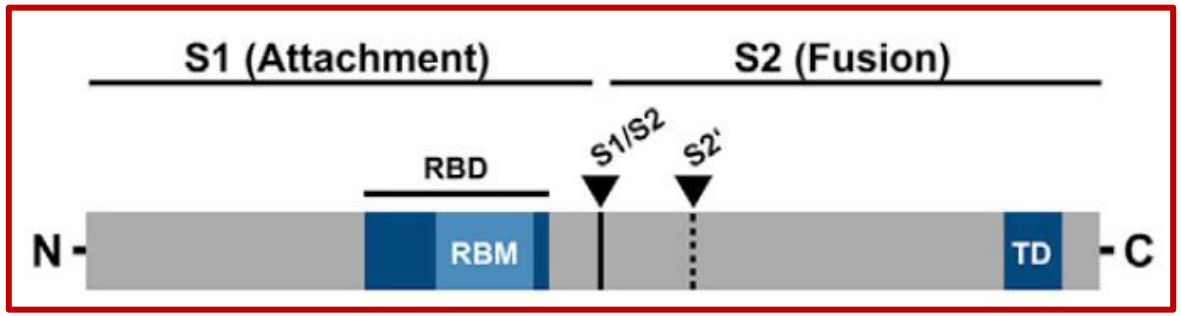
Sputum: 10^{6}-10^{11} RNAs/mL

RNA counts can markedly overestimate infectious virions

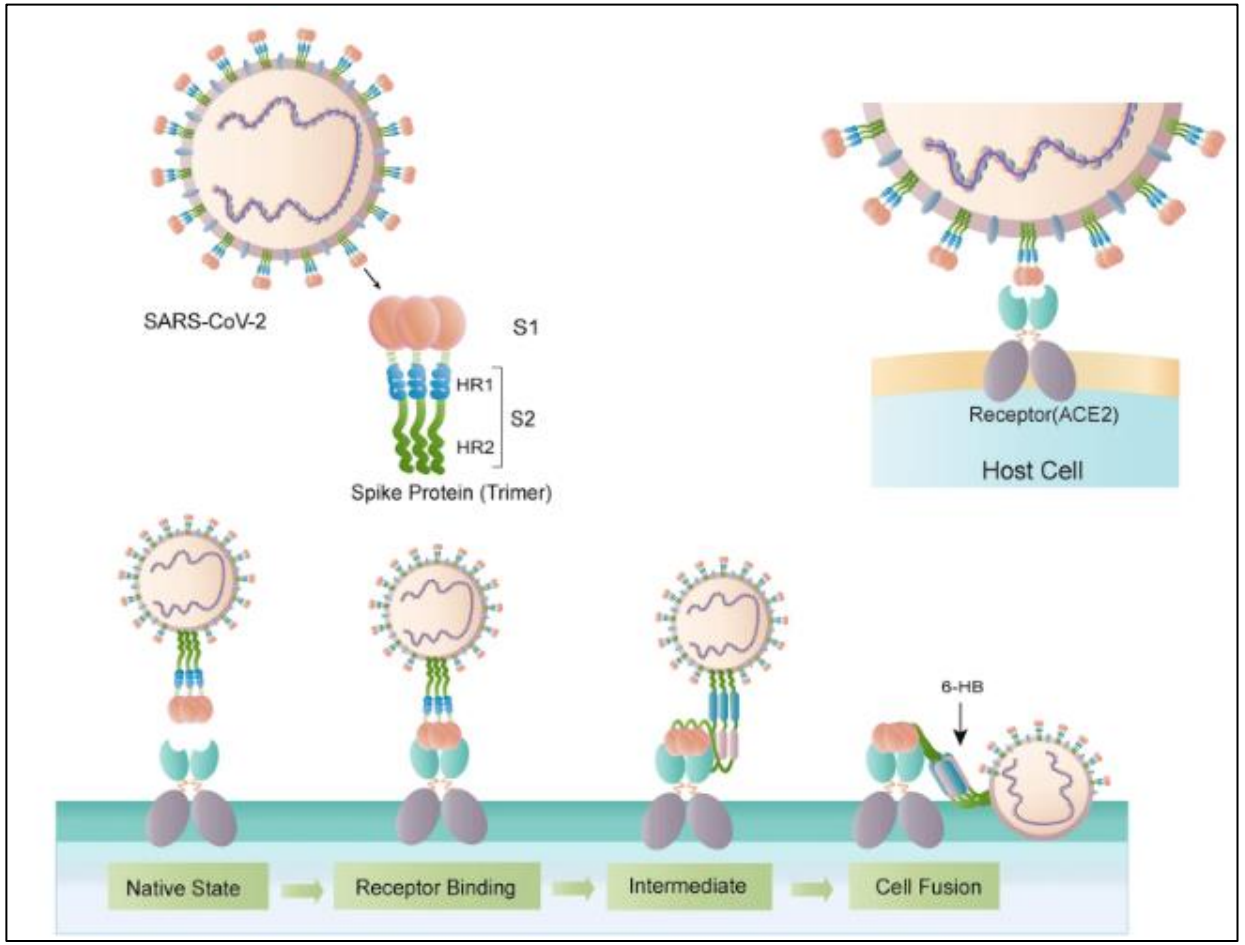

Spike Protein

The spike (S) protein plays a key role in the receptor recognition and cell membrane fusion process.

- S is composed of two subunits, S1 and S2.

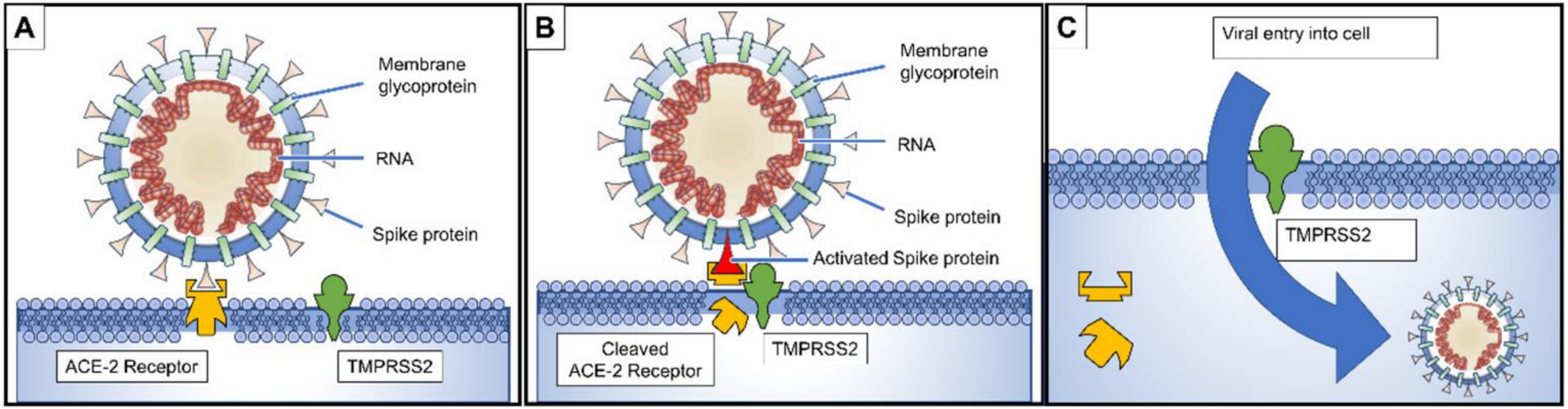


- The S1 subunit contains a receptor-binding domain that recognizes and binds to the host receptor angiotensin-converting enzyme 2.
- The S2 subunit facilitates fusion of the viral membrane with a cellular membrane



DOI https://doi.org/10.1038/s41401-020-0485-4

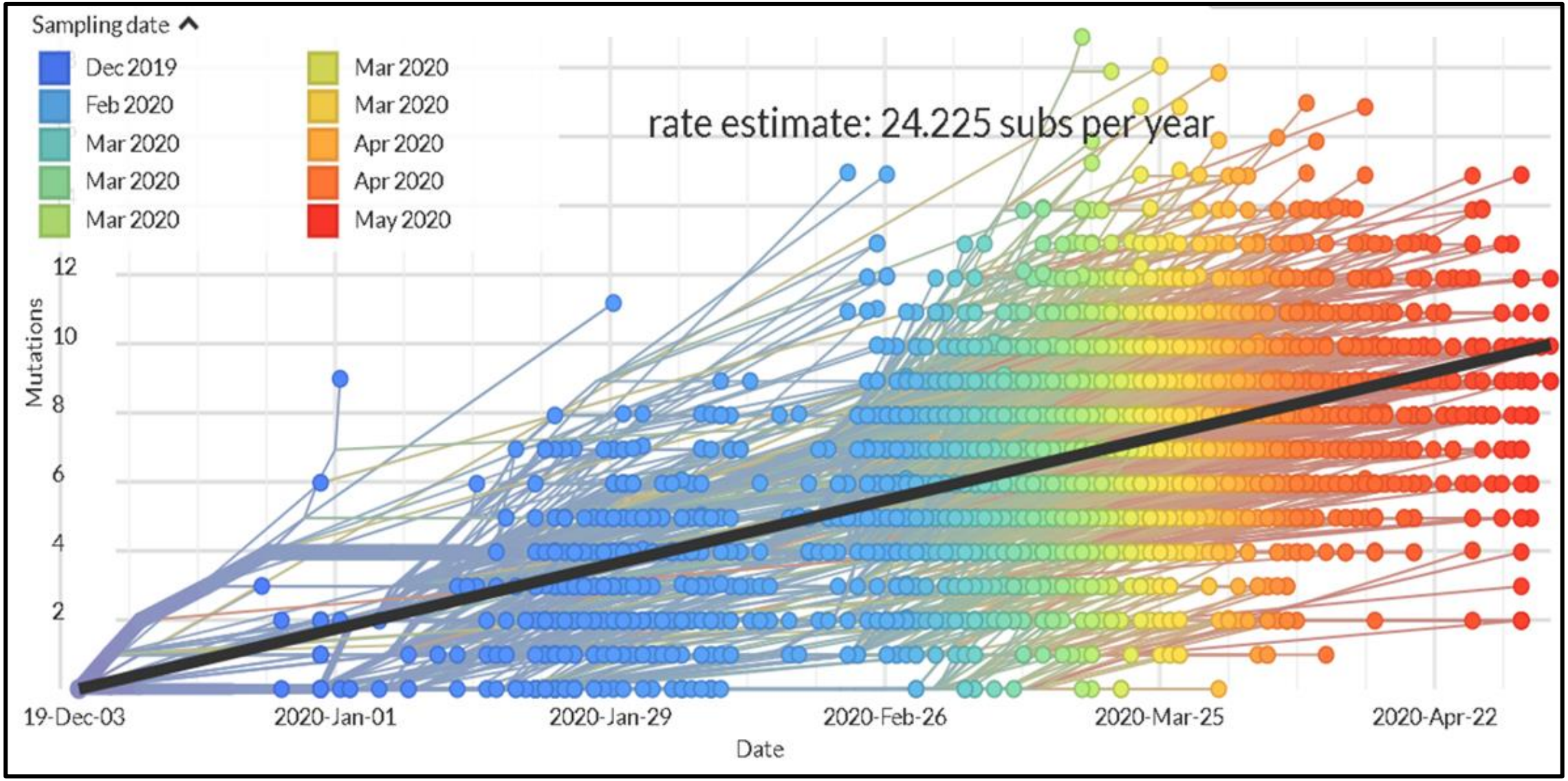
Host Cell Partners that Facilitate SARS-CoV-2 to Enter



TMPRSS2 protein is a serine protease which contains a type II transmembrane domain, a receptor class A domain, a scavenger receptor cysteine-rich domain and a protease domain.

https://doi.org/10.3390/pathogens9030231

GISAID & Nextstrain: Two major sequence databases that permit evolutionary and some epidemiological analyses



www.nextstrain.org

Indian J Med Res 151, May 2020, pp 450-458
DOI: 10.4103/ijmr.IJMR_1125_20

Quick Response Code:

Analysis of RNA sequences of 3636 SARS-CoV-2 collected from 55 countries reveals selective sweep of one virus type

Nidhan K. Biswas & Partha P. Majumder

National Institute of Biomedical Genomics, Kalyani, West Bengal, India

Gudbjartsson et al.: Spread of SARS-CoV-2 in the Icelandic Population *New England Journal Medicine* 382;24 June 11, 2020

Korber et al.: Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. July 2020. *Cell* 182(4)

Classification of SARS-CoV-2 into Subtypes Based on Frequencies of Genomic Changes and Similarities of Genomes

Table 1: Numbers of SARS-CoV-2 sequences belonging to a specific phylogenetic clade.

Phylogenetic clade	Clade Order	Defining mutation(s) for clade	Numbers of viral sequence belonging to clade (Total no. of sequences used = 3636)
O	1	Ancestral Clade	582
B	2	ORF8 - L84S	191
B1	3	ORF8 - L84S , nt - C18060T	505
B2	4	ORF8 - L84S , nt - C29095T	20
B4	5	ORF8 - L84S , N - S202N	24
A3	6	ORF1a - V378I , ORF1a - L3606F	87
A6	7	nt - T514C	53
A7	8	ORF1a - A3220V	4
A1a	9	ORF3a - G251V , ORF1a - L3606F	321
A2	10	S - D614G	1
A2a	11	S - D614G, ORF1b - P314L	1848

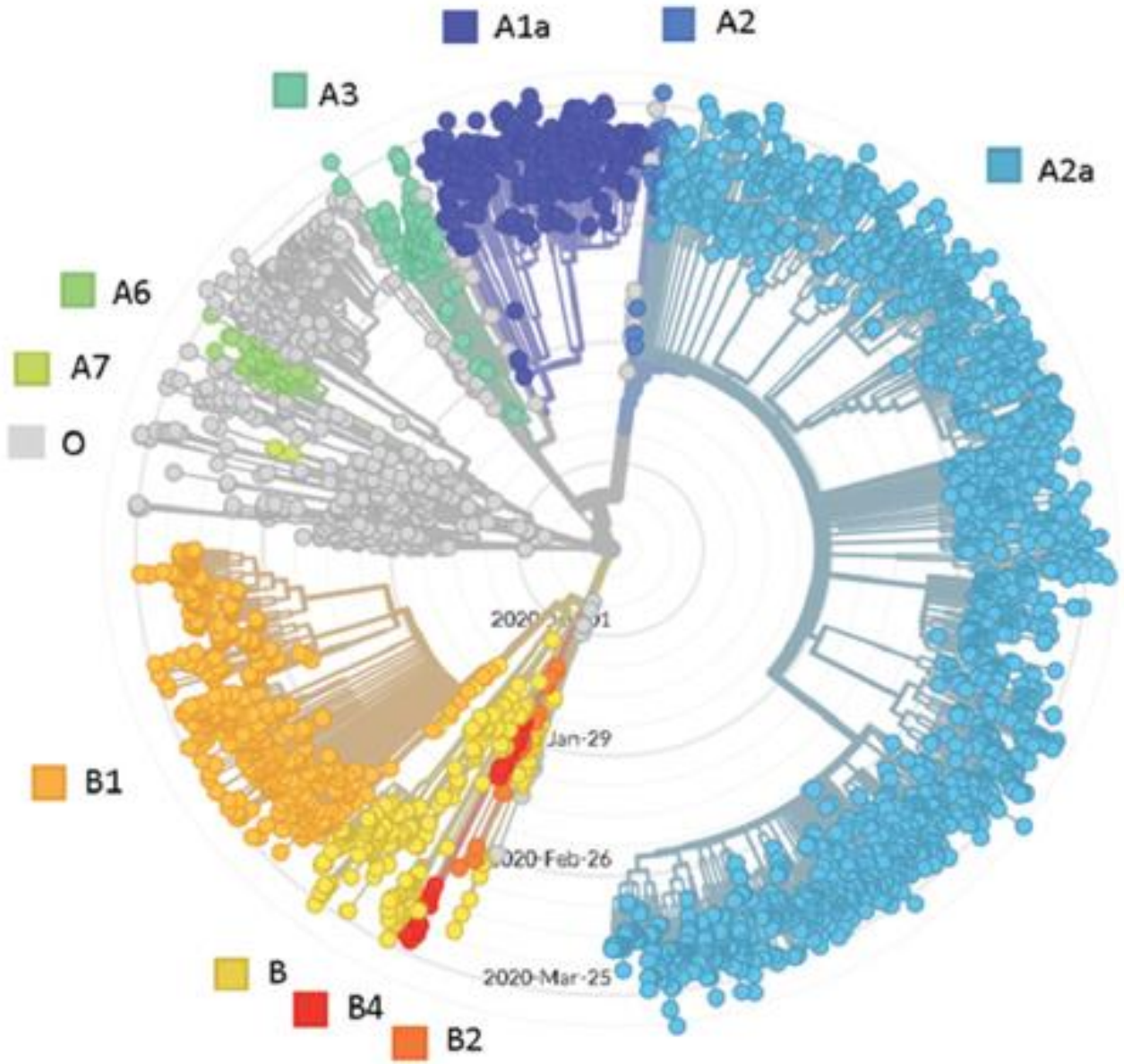
614 Aspartate (D)

614 Glycine (G)

nt: Nucleotide

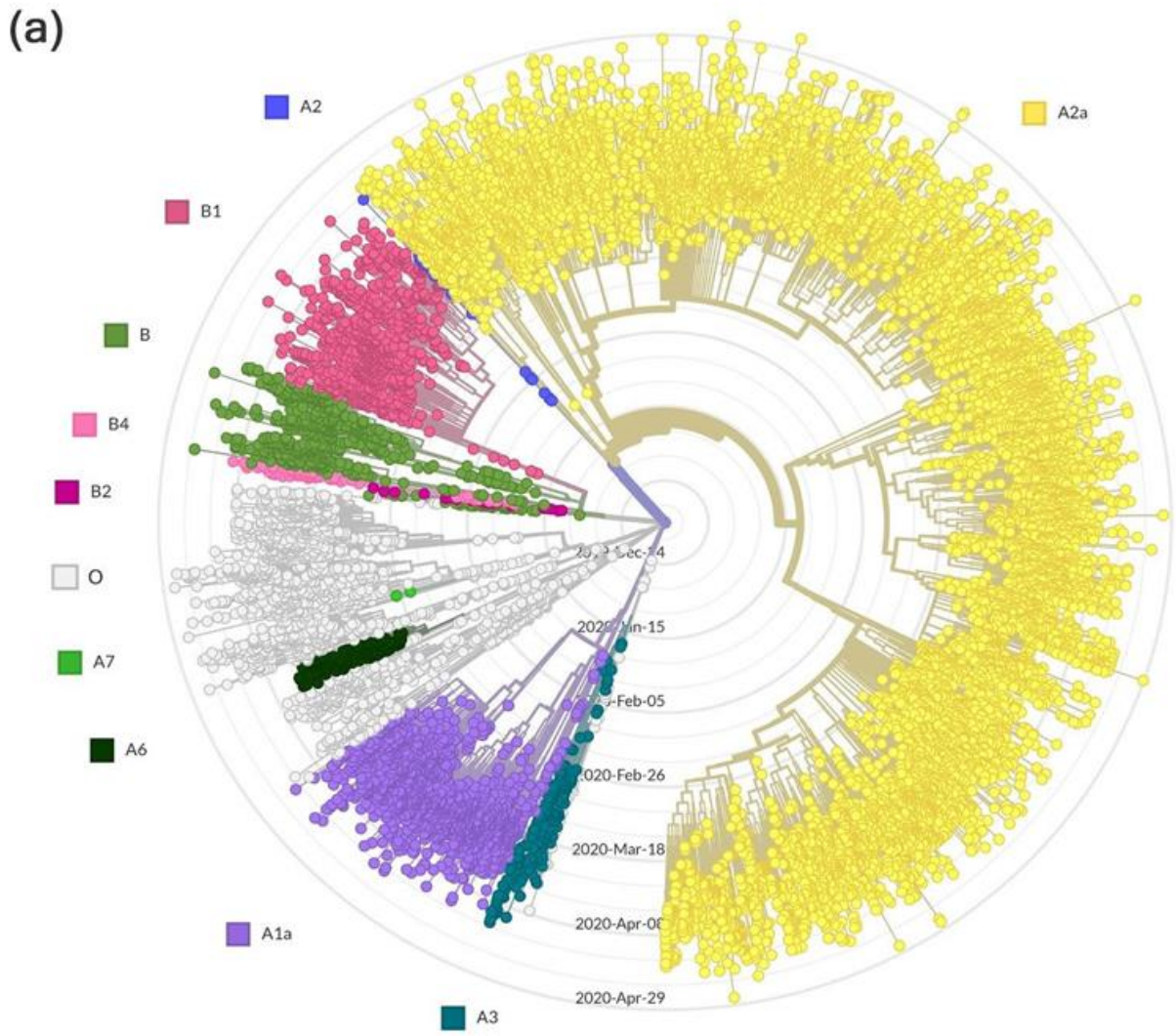
A total of 11 distinct mutations define the 10 derived clades.

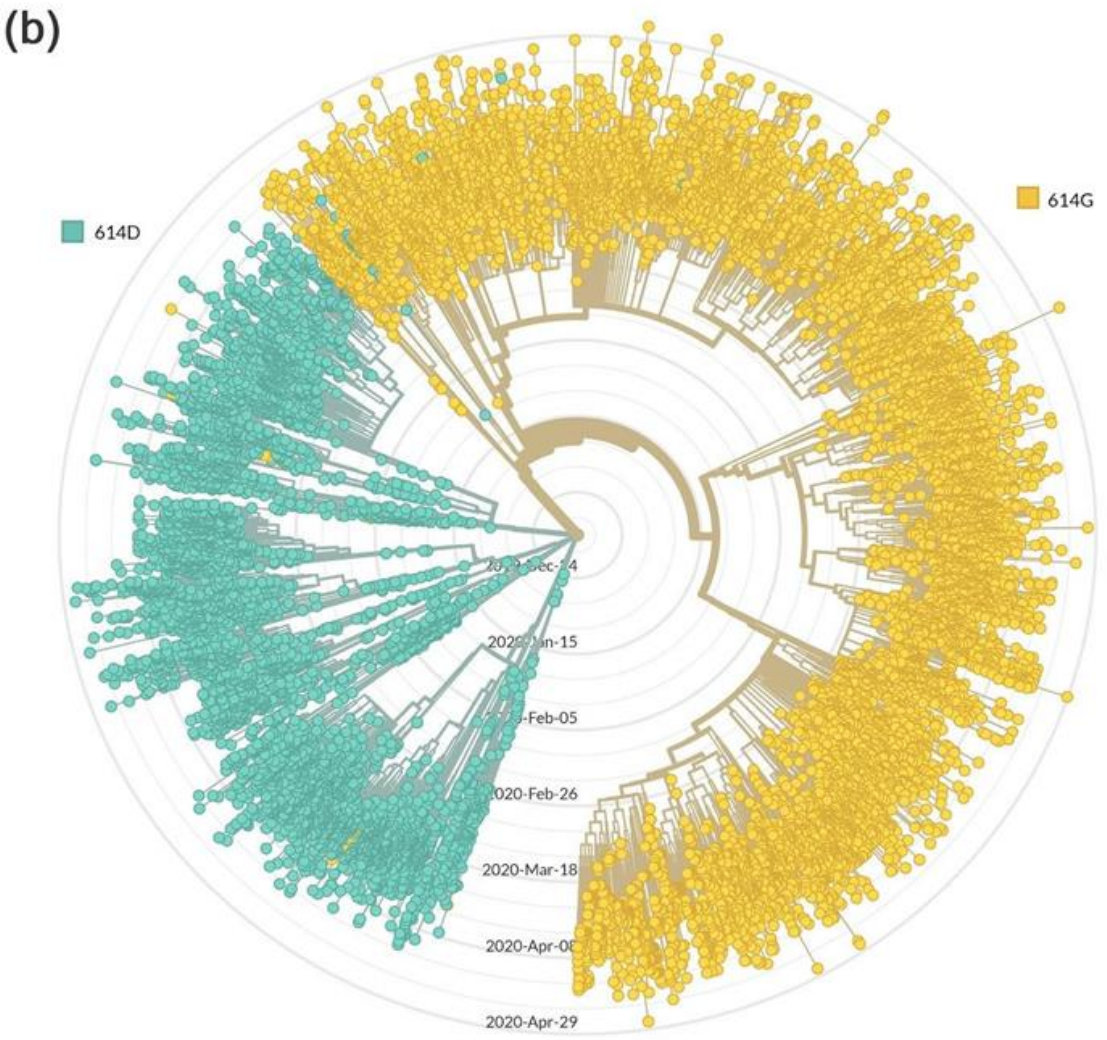
BISWAS & MAJUMDER: SELECTIVE SWEEP OF ONE SARS-CoV-2 TYPE



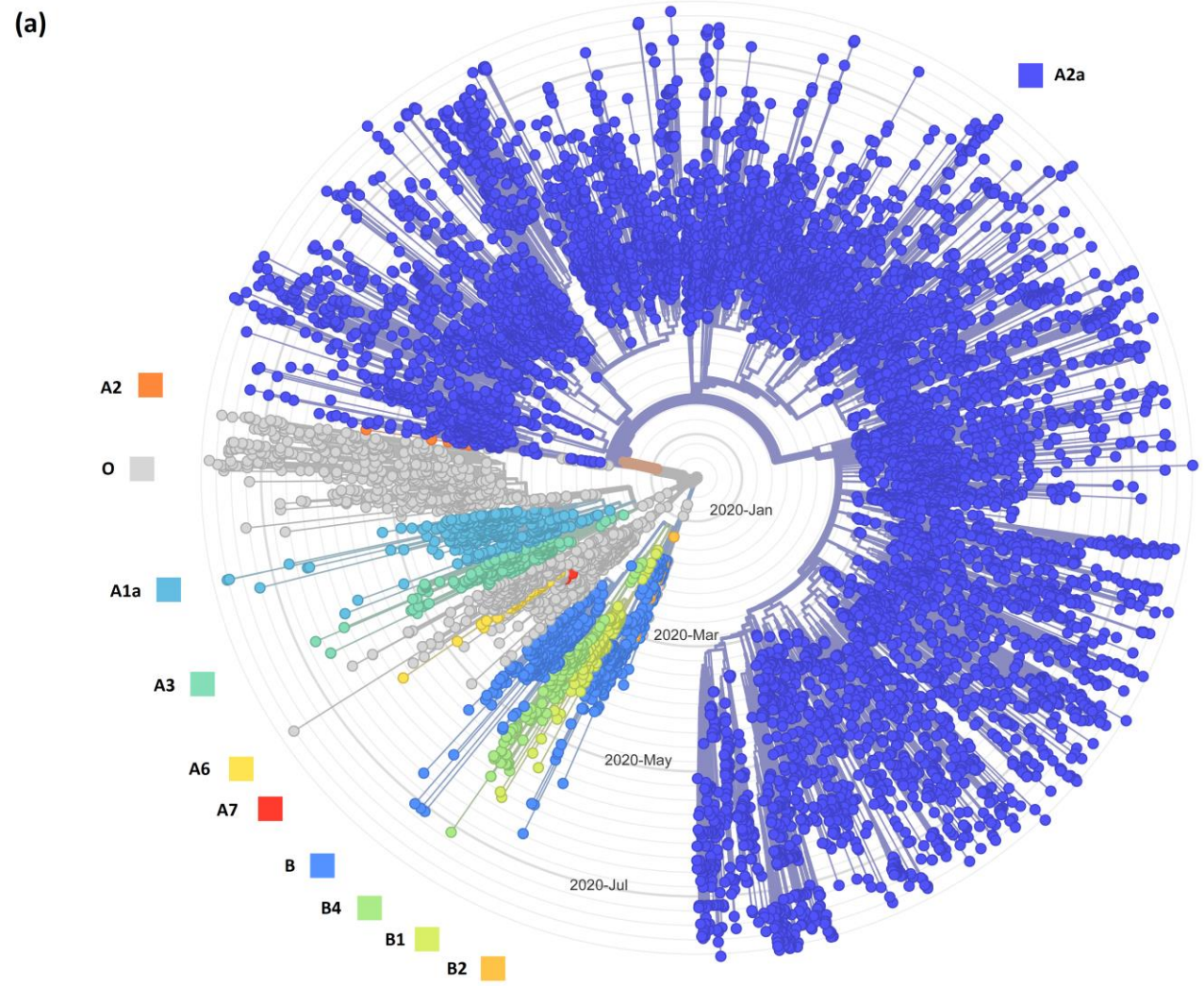
31st March 2020

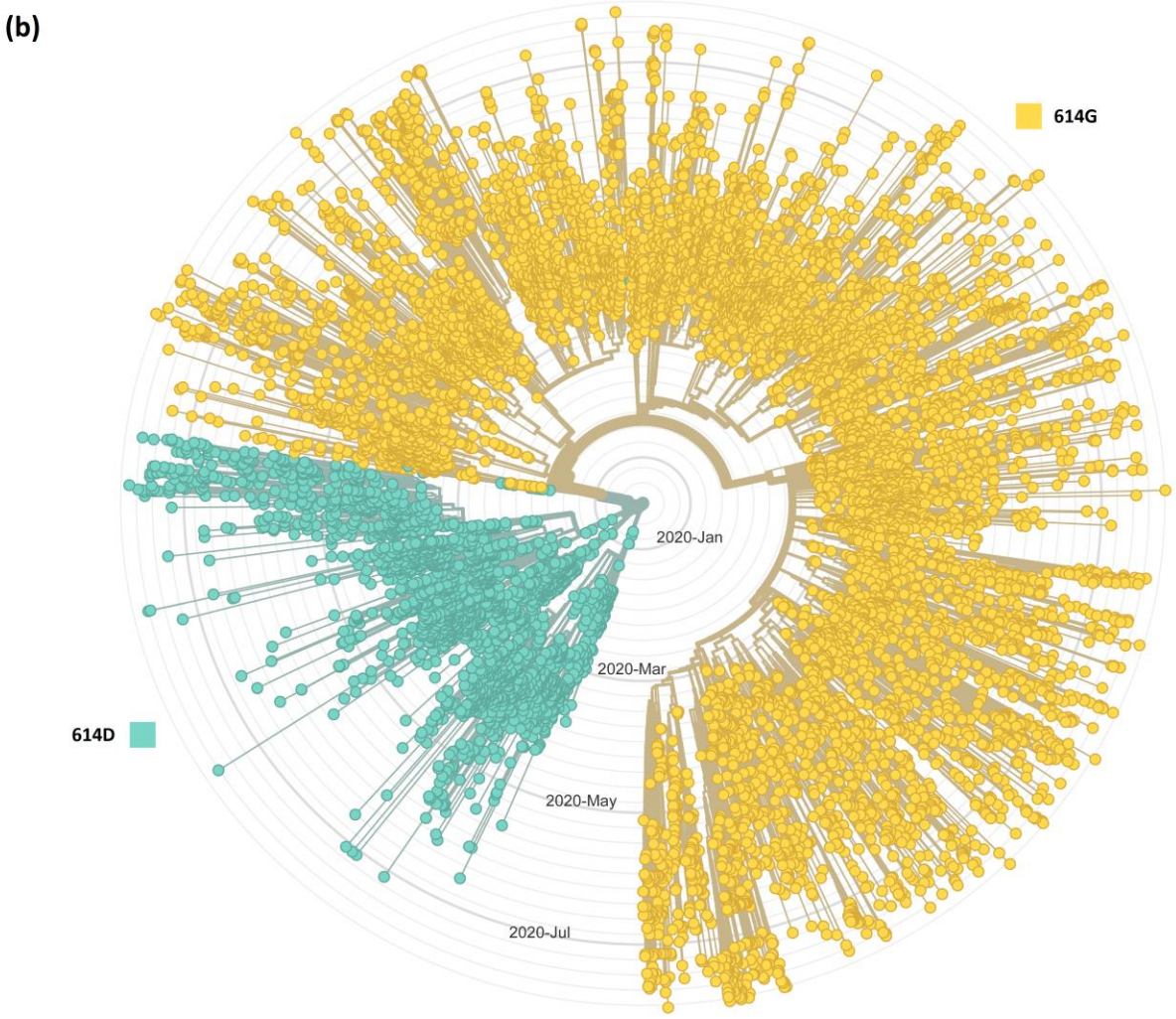
Early May 2020



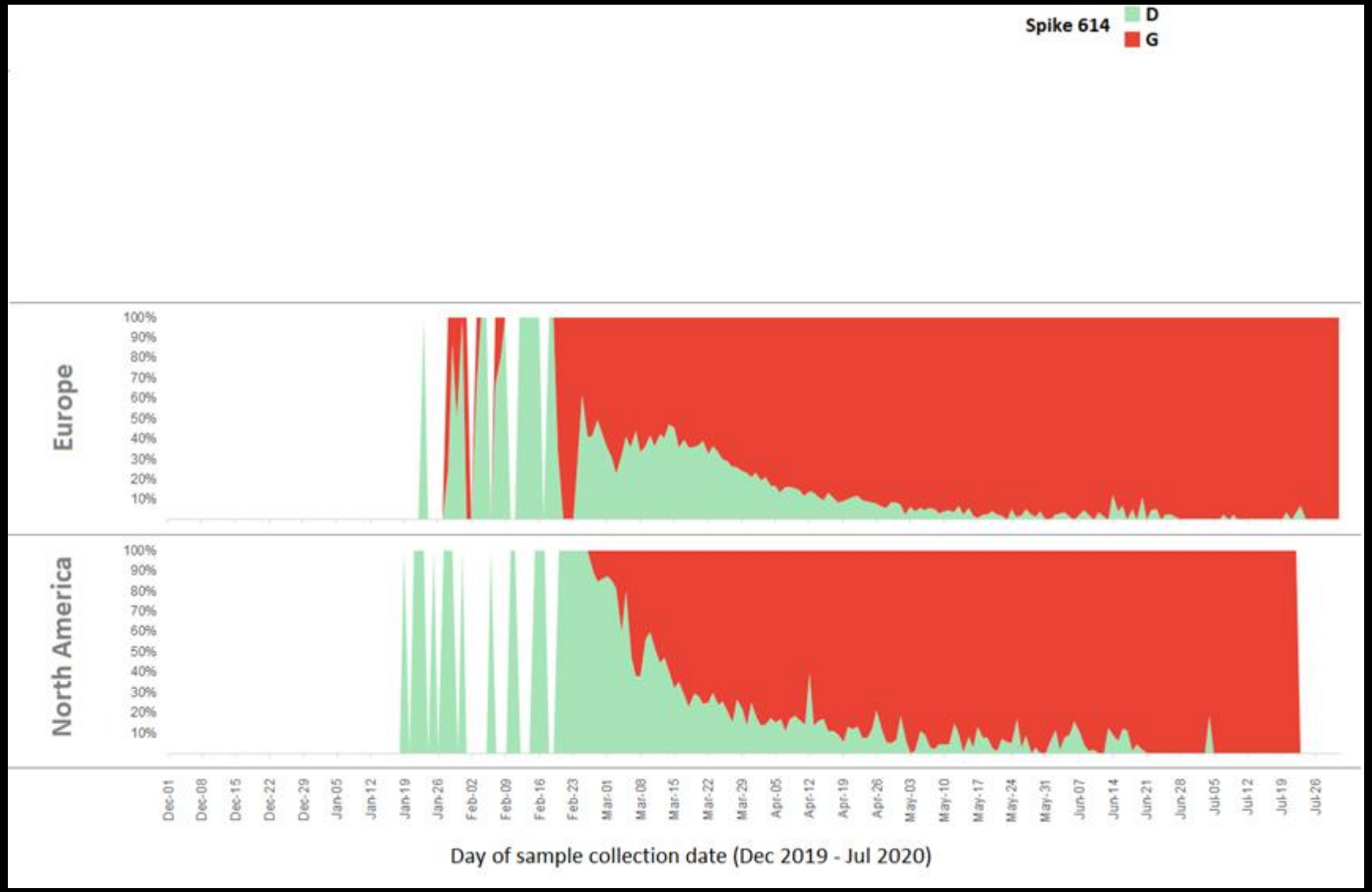


July 2020

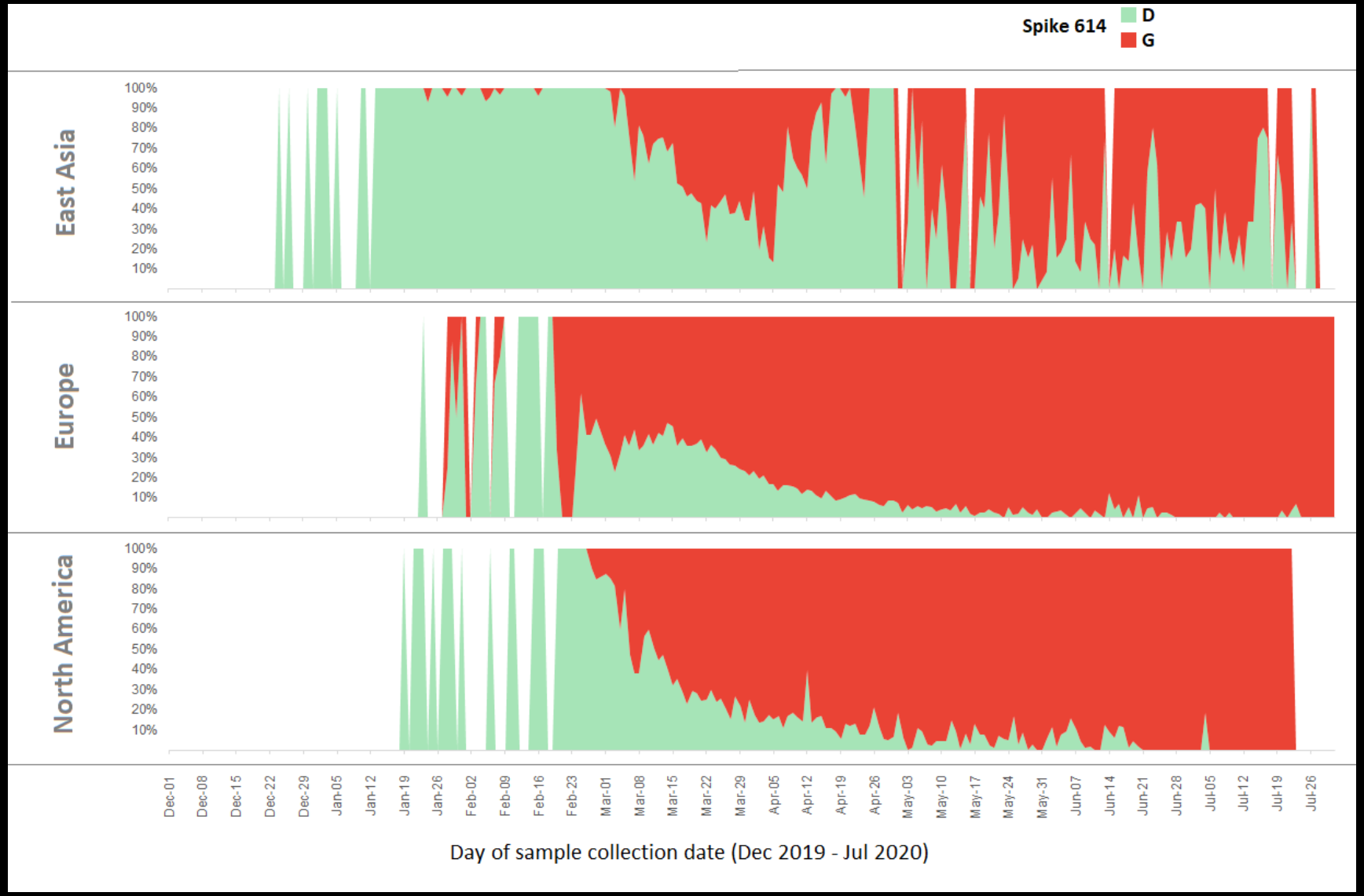

(a)
A2a
A2
O
A1a
A3
A6
A7
B
B4
B1
B2
2020-Jan
2020-Mar
2020-May
2020-Jul


(b)
614G
614D
2020-Jan
2020-Mar
2020-May
2020-Jul

The mutant 614G has swept through Caucasian populations of North-America and Europe

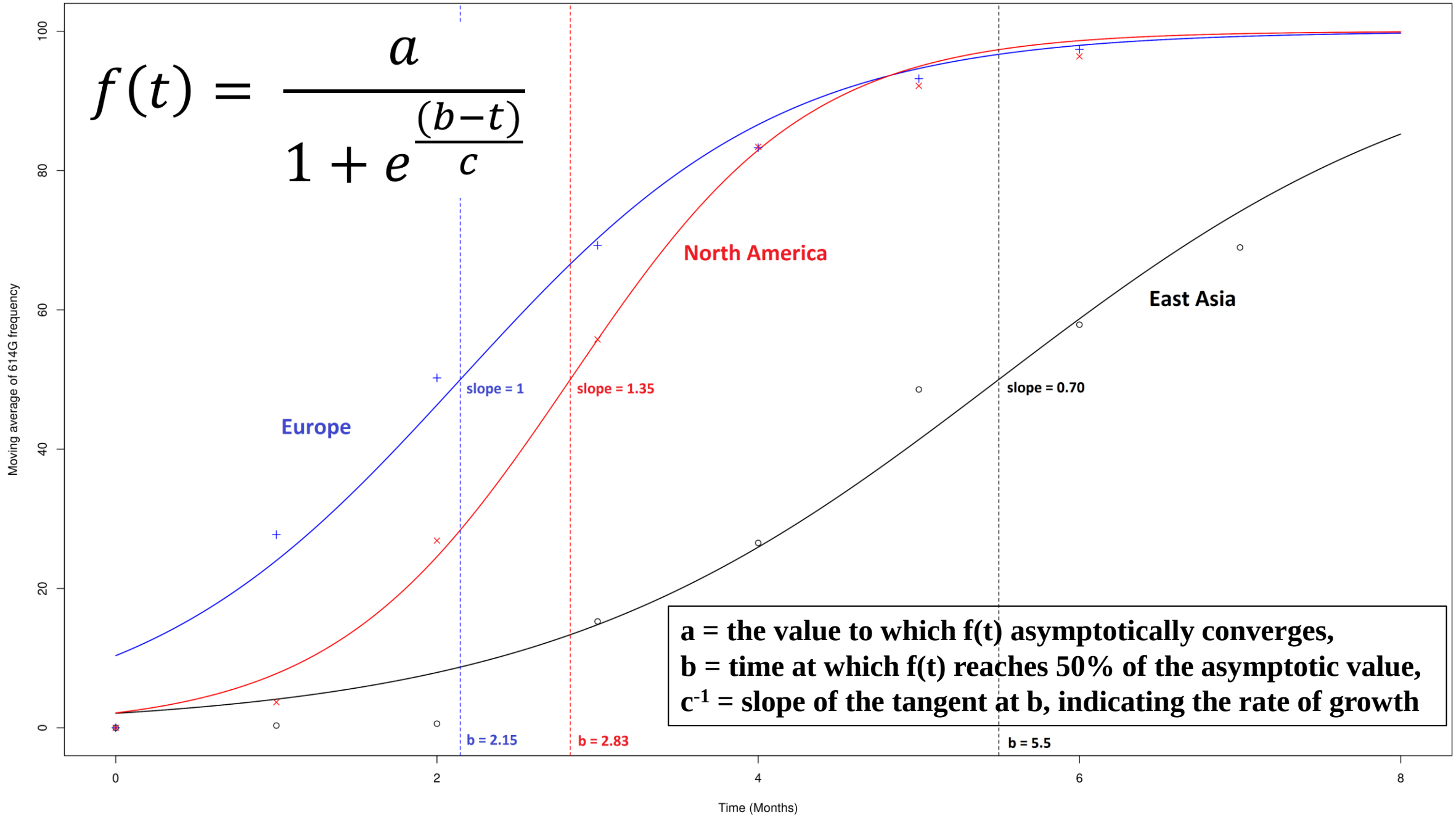


The mutant 614G has sprinted through Caucasian populations of North-America and Europe … but is only hopping in East Asia



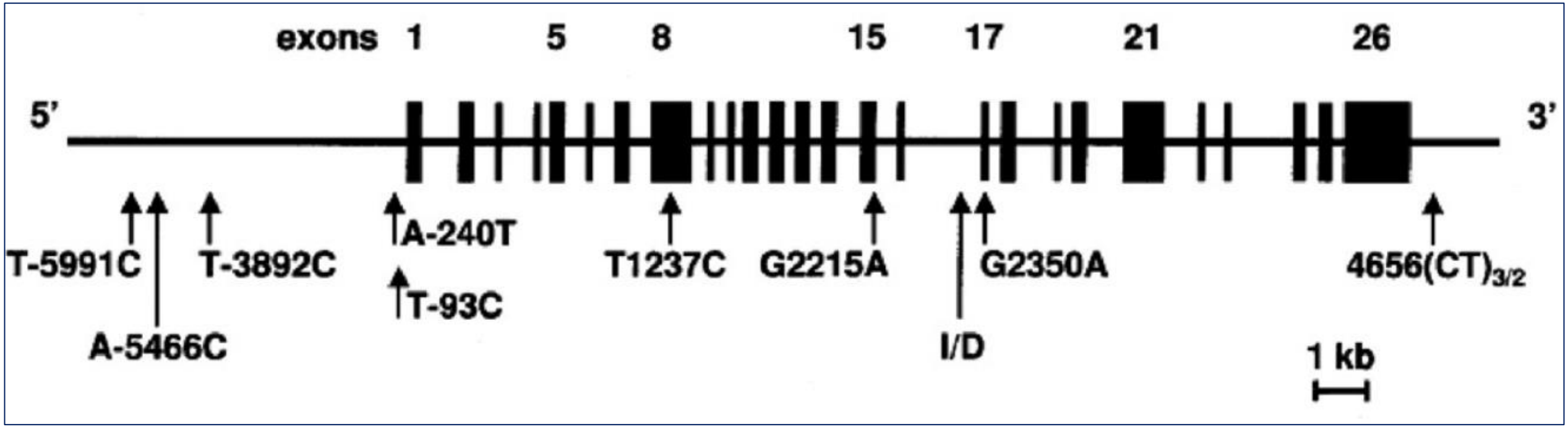
On May 5th of last year, **Reuters announced** that "*Asia coronavirus cases hit 250,000 but pace much slower than U.S., Europe.*"

The **Reuters report** stated "*The region where the COVID-19 pandemic started has fared better overall than North America and Europe since the first case was reported in Wuhan, China on January 10 (2020) … It has taken Asia almost four months to reach the 250,000 infection milestone, a level that Spain alone is approaching just a little over two months since reporting its first case.*"


$f(t) = \frac{a}{1 + e^{\frac{(b-t)}{c}}}$
Moving average of 614G frequency
Time (Months)
0
20
40
60
80
100
0
2
4
6
8
Europe
North America
East Asia
slope = 1
slope = 1.35
slope = 0.70
b = 2.15
b = 2.83
b = 5.5
a = the value to which f(t) asymptotically converges,
b = time at which f(t) reaches 50% of the asymptotic value,
c^{-1} = slope of the tangent at b, indicating the rate of growth

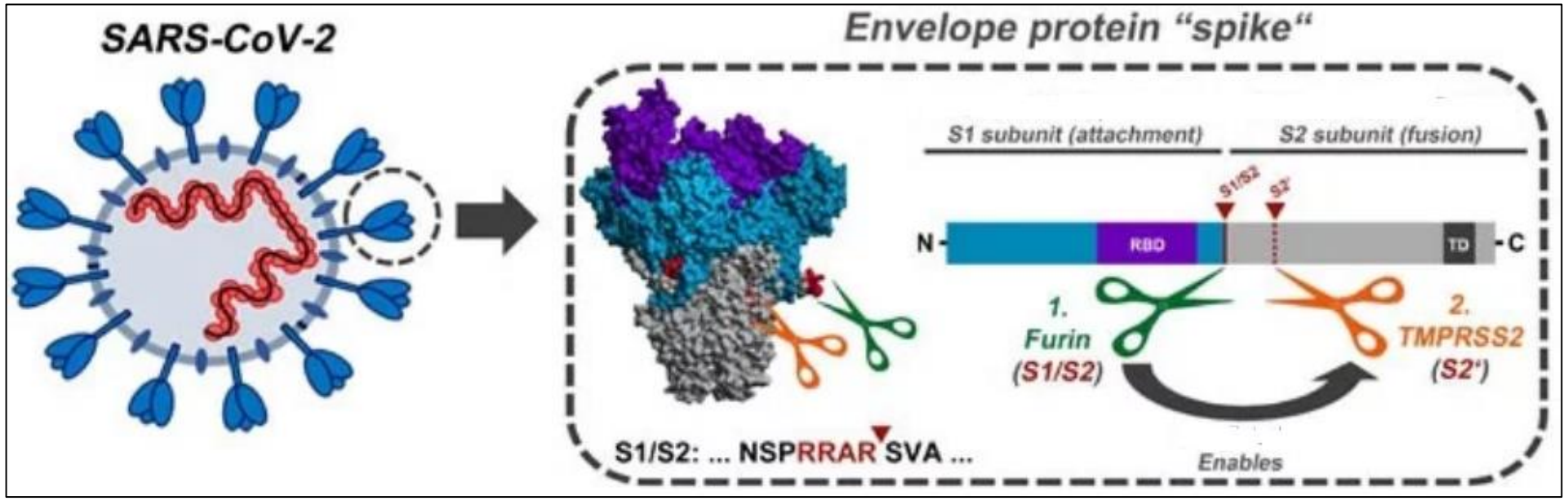
Why is the coronavirus finding it difficult to sweep through non-Caucasian populations of East Asia?

- Spike protein latches on to ACE protein on cell surface to gain entry
- Some variants impact on ACE expression levels
- There is no significant difference between Caucasian and non-Caucasian populations in the frequencies of such variants



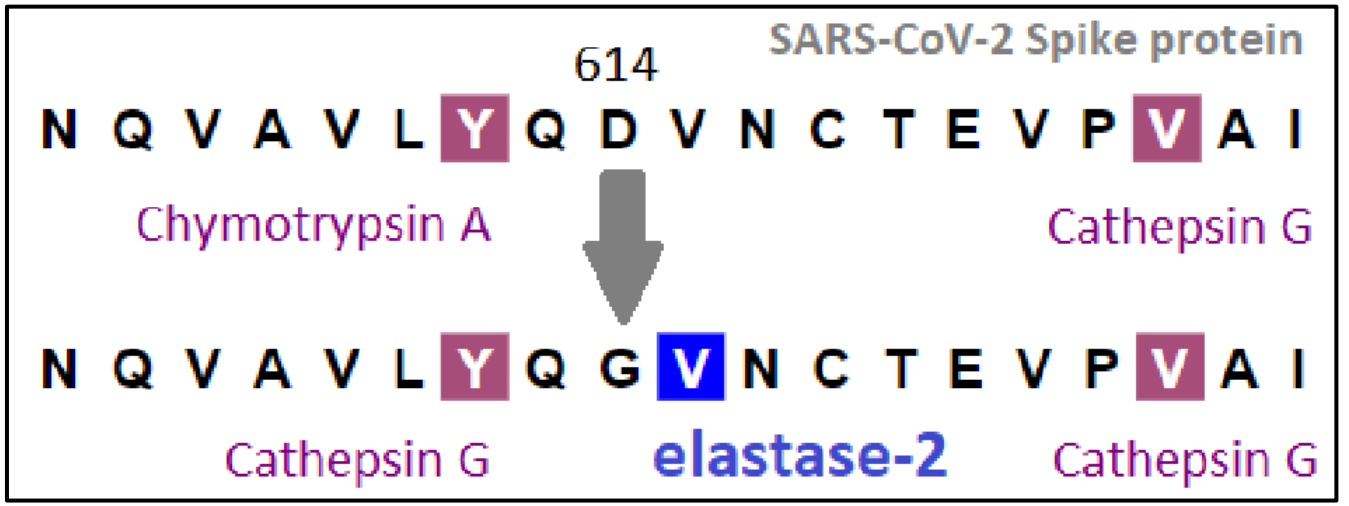
Membrane fusion depends on S protein cleavage by host cell proteases at the S1/S2 and the S2' site which results in S protein activation

- The Spike protein is inactive, until cleaved

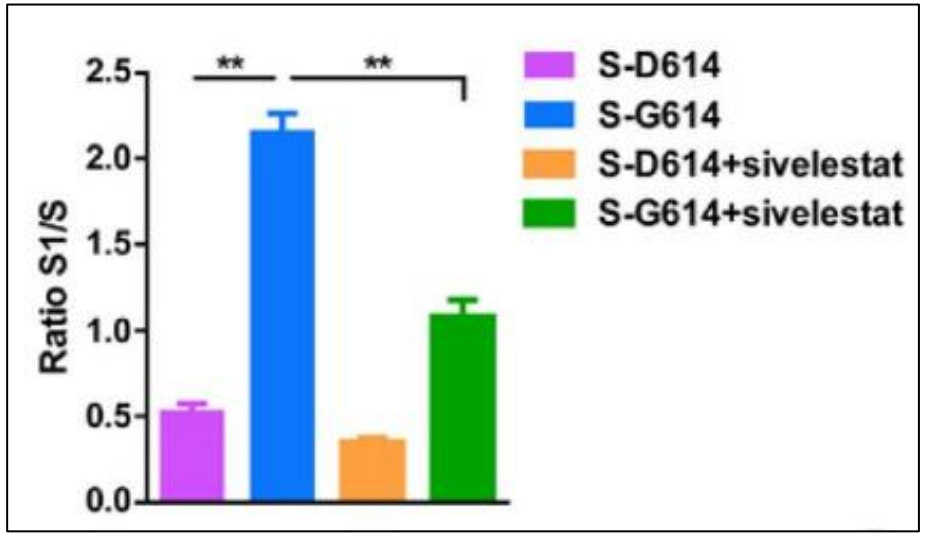


Cleavage sites are present in both ancestral D614 and the mutant 614G.

The D to G mutation creates an additional cleavage site at 615-616 by Neutrophil Elastase



The 614G mutant is cleaved by neutrophil elastase 4-fold more efficiently than D614 that is cleaved only by TMPRSS2



D614G mutation of SARS-CoV-2 spike protein enhances viral infectivity
doi: https://doi.org/10.1101/2020.06.20.161323

Higher level of cleavage implies greater level of activation of the Spike protein, facilitating entry of SARS-CoV-2 into the host cell.

Higher level of neutrophil elastase can accrue if there is a higher number of neutrophils.

Is there evidence that Caucasians have a higher count of neutrophils than non-Caucasians?

American Journal of Human Genetics **99**: 22–39 (2016)
Large-Scale Exome-wide Association Analysis Identifies Loci for White Blood Cell Traits and Pleiotropy with Immune-Mediated Diseases.

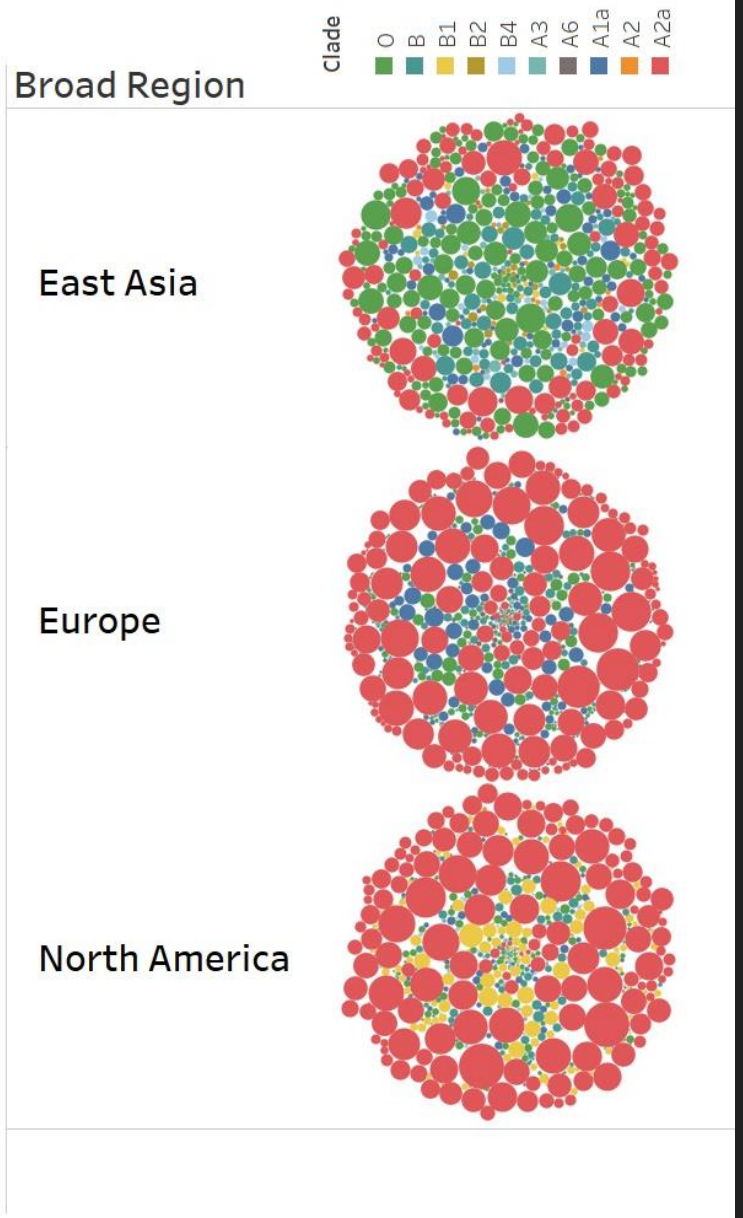
- Individuals of five different ethnicities investigated: European, African, Hispanic American, East Asian and South Asian.
- **W.r.t. Neutrophil count, no significant differences were attributable to differences in ancestries.**

α1-antitrypsin (AAT) inhibits neutrophil elastase

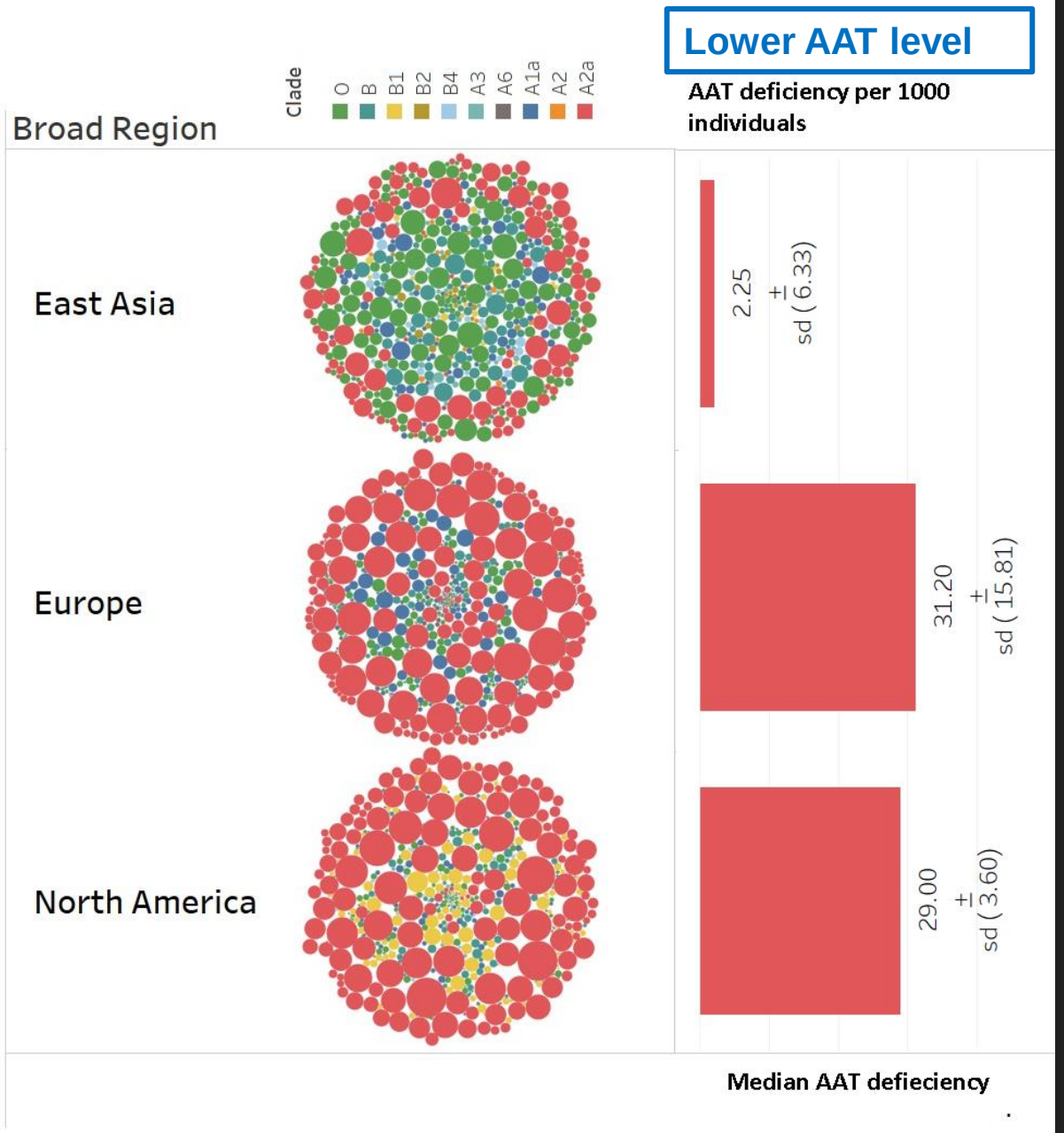
- AAT inhibits elastase around normal tissue.
- AAT deficiency is caused by mutations in the *SERPINA1* gene, located on chromosome 14.
- *SERPINA1* has many alleles that produce different amounts of AAT.
  - M allele produces normal levels of the AAT protein
  - S allele produces moderately low levels
  - Z allele produces very low levels.

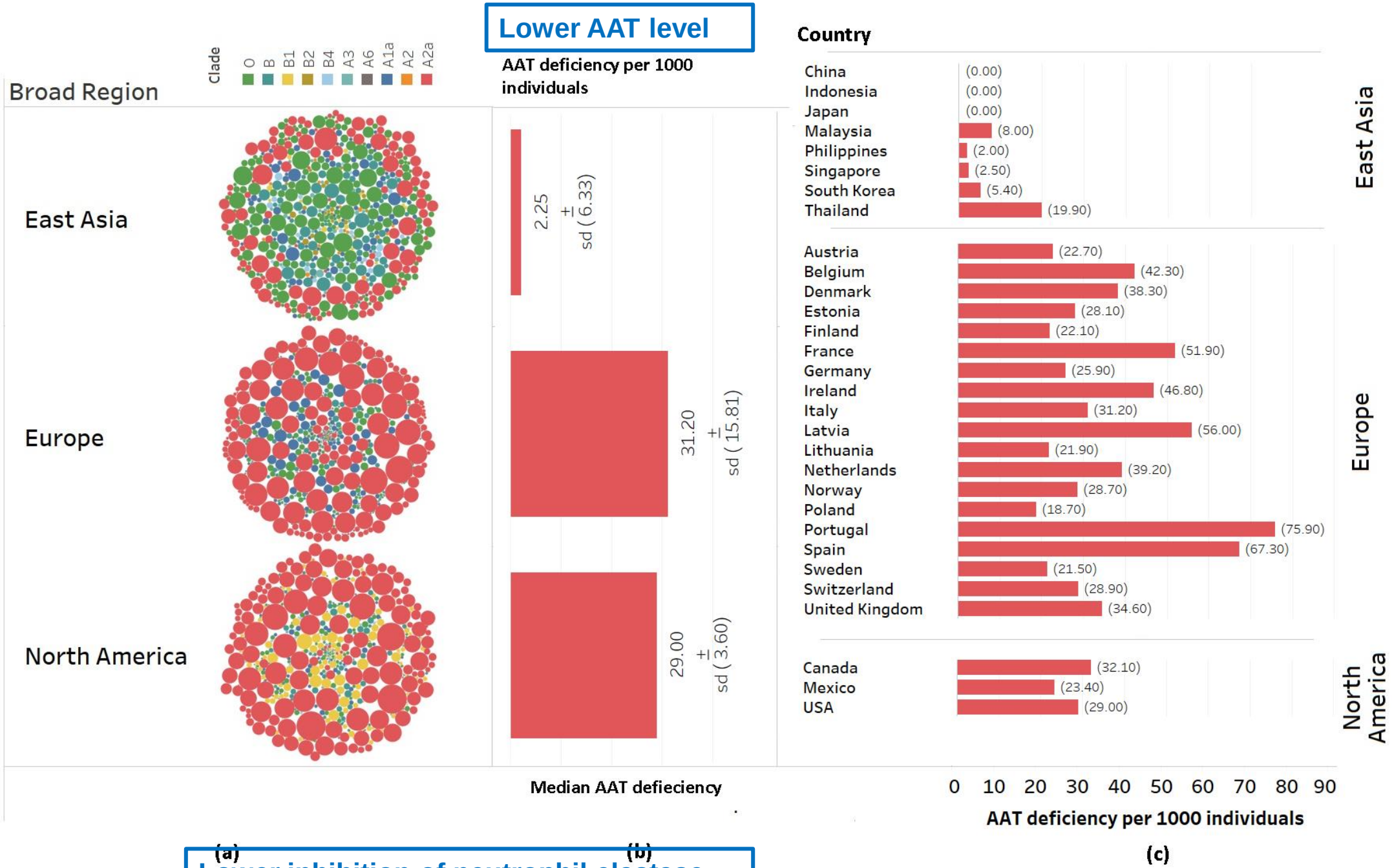
MM	**Normal**
MS, MZ	**Slightly Deficient**
SZ, ZZ	**Deficient**

❑ **Lower level of AAT** ➡ **Lower inhibition of neutrophil elastase**
➡ **Higher level of neutrophil elastase**
➡ **Greater level of Spike protein activation of SARS-CoV-2**
➡ **Higher level of infectivity**
➡ **Better spread in population**


Clade
O
B
B1
B2
B4
A3
A6
A1a
A2
A2a
Broad Region
East Asia
Europe
North America

(a)


Lower AAT level
Clade
O
B
B1
B2
B4
A3
A6
A1a
A2
A2a
Broad Region
AAT deficiency per 1000 individuals
East Asia
Europe
North America
2.25
± sd (6.33)
31.20
± sd (15.81)
29.00
± sd (3.60)
Median AAT defieciency
(a)
(b)
Lower inhibition of neutrophil elastase
Better spread in population

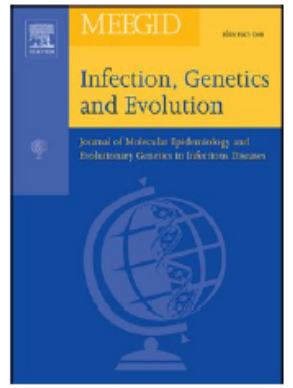

Lower AAT level
Clade
O
B
B1
B2
B4
A3
A6
A1a
A2
A2a
Broad Region
East Asia
Europe
North America
AAT deficiency per 1000 individuals
2.25
± sd (6.33)
31.20
± sd (15.81)
29.00
± sd (3.60)
Median AAT defieciency
(a)
(b)
Country
China (0.00)
Indonesia (0.00)
Japan (0.00)
Malaysia (8.00)
Philippines (2.00)
Singapore (2.50)
South Korea (5.40)
Thailand (19.90)
East Asia
Austria (22.70)
Belgium (42.30)
Denmark (38.30)
Estonia (28.10)
Finland (22.10)
France (51.90)
Germany (25.90)
Ireland (46.80)
Italy (31.20)
Latvia (56.00)
Lithuania (21.90)
Netherlands (39.20)
Norway (28.70)
Poland (18.70)
Portugal (75.90)
Spain (67.30)
Sweden (21.50)
Switzerland (28.90)
United Kingdom (34.60)
Europe
Canada (32.10)
Mexico (23.40)
USA (29.00)
North America
0 10 20 30 40 50 60 70 80 90
AAT deficiency per 1000 individuals
(c)
Lower inhibition of neutrophil elastase
Better spread in population

Infection, Genetics and Evolution 90 (2021) 104760

Contents lists available at ScienceDirect

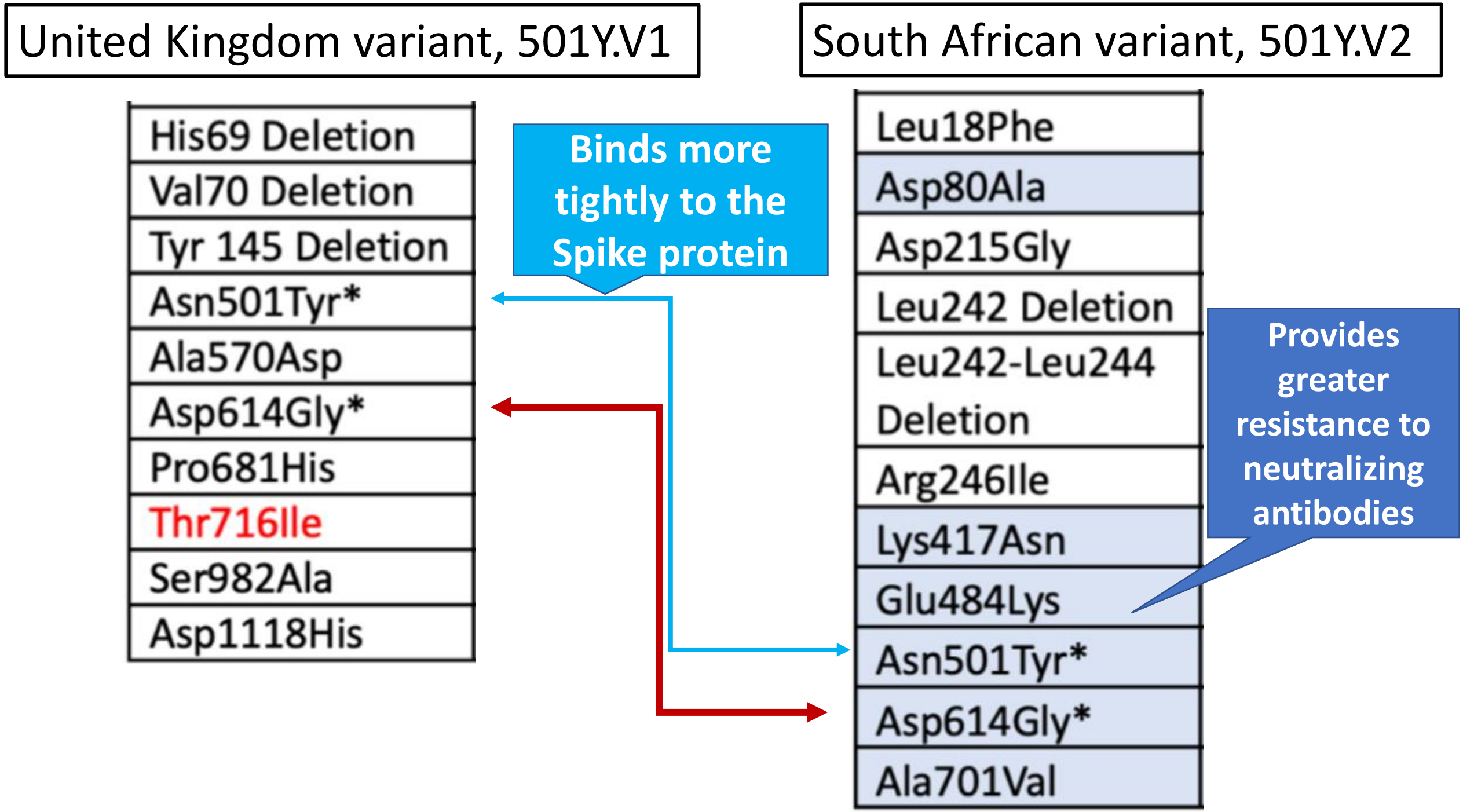
Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid



Research paper

SARS-CoV-2 mutation 614G creates an elastase cleavage site enhancing its spread in high AAT-deficient regions


United Kingdom variant, 501Y.V1
South African variant, 501Y.V2
His69 Deletion
Val70 Deletion
Tyr 145 Deletion
Asn501Tyr*
Ala570Asp
Asp614Gly*
Pro681His
Thr716Ile
Ser982Ala
Asp1118His
Binds more tightly to the Spike protein
Leu18Phe
Asp80Ala
Asp215Gly
Leu242 Deletion
Leu242-Leu244 Deletion
Arg246Ile
Lys417Asn
Glu484Lys
Asn501Tyr*
Asp614Gly*
Ala701Val
Provides greater resistance to neutralizing antibodies

FRONT MATTER

Title

- Increased elastase sensitivity and decreased intramolecular interactions in the more transmissible SARS-CoV-2 variants' spike protein
- Analysis of the new UK and SA SARS-CoV-2 variants

Authors

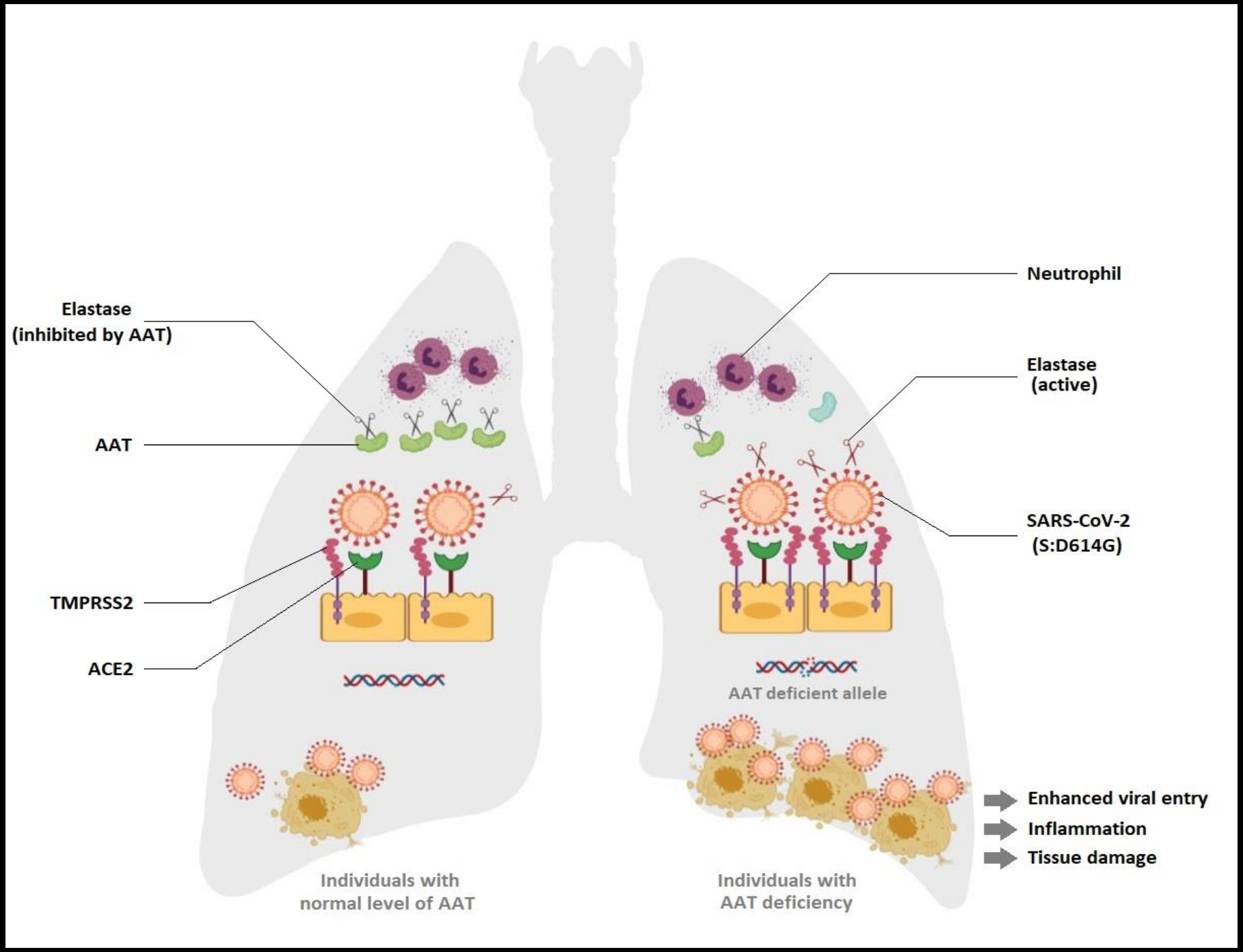
S. Pokhrel[1]†, L. Lee[1]†, B.R. Kraemer[1]†, K. Samardzic[1], D. Mochly-Rosen[1*]

Affiliations

1. Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA

† These authors contributed equally to this work

* mochly@stanford.edu


Elastase
(inhibited by AAT)
AAT
TMPRSS2
ACE2
Neutrophil
Elastase
(active)
SARS-CoV-2
(S:D614G)
AAT deficient allele
Enhanced viral entry
Inflammation
Tissue damage
Individuals with
normal level of AAT
Individuals with
AAT deficiency