ONE KIT ONE PRODUCT
WORKSHOP

YAKIN DOĞU ÜNİVERSİTESİ
DESAM ENSTİTÜSÜ

DATABASE MINING AND TARGET SELECTION

Assoc. Prof. Buket Baddal

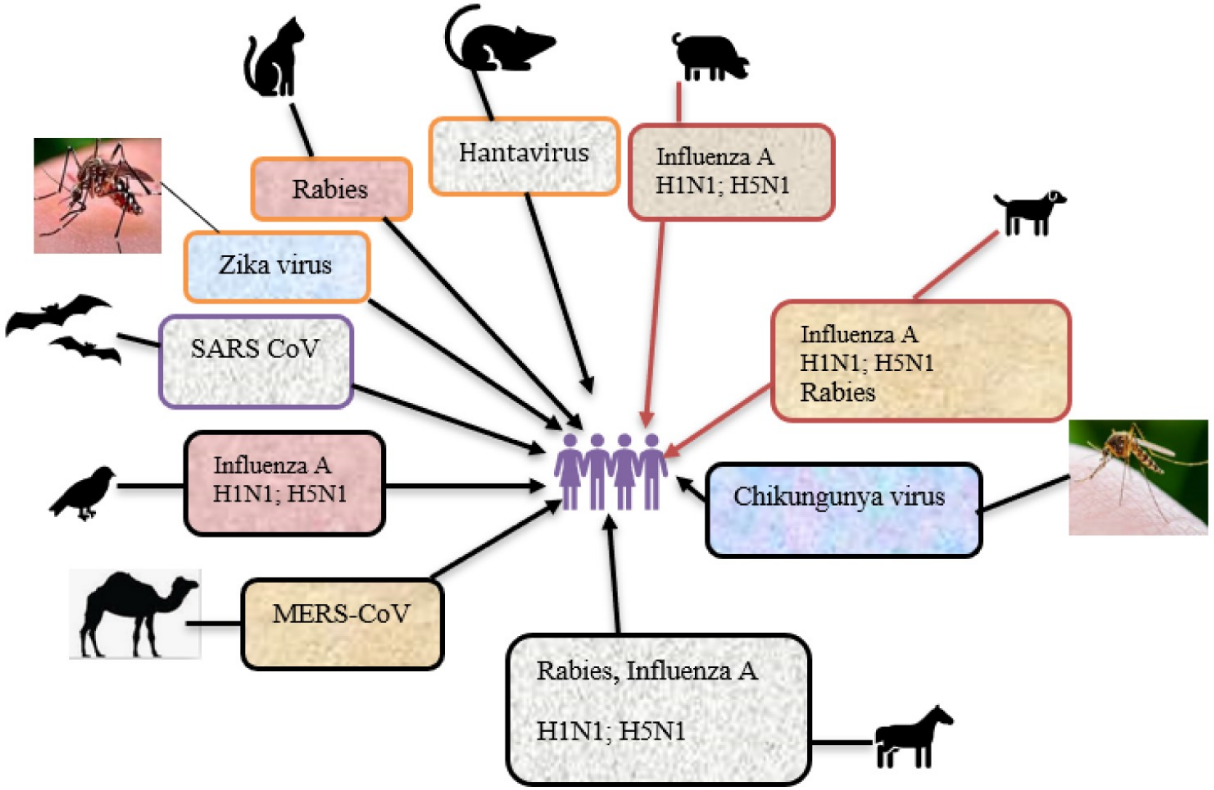
Near East University, Faculty of Medicine, Department of Medical Microbiology and Clinical Microbiology
Near East University Hospital, Molecular Microbiology Laboratory

29 July 2022, Nicosia

Emerging, re-emerging infectious agents and variants

- Emerging and Re-Emerging Infectious Diseases (EIDs) are infections that have newly appeared in a population or have existed previously but are rapidly increasing in incidence or geographic range

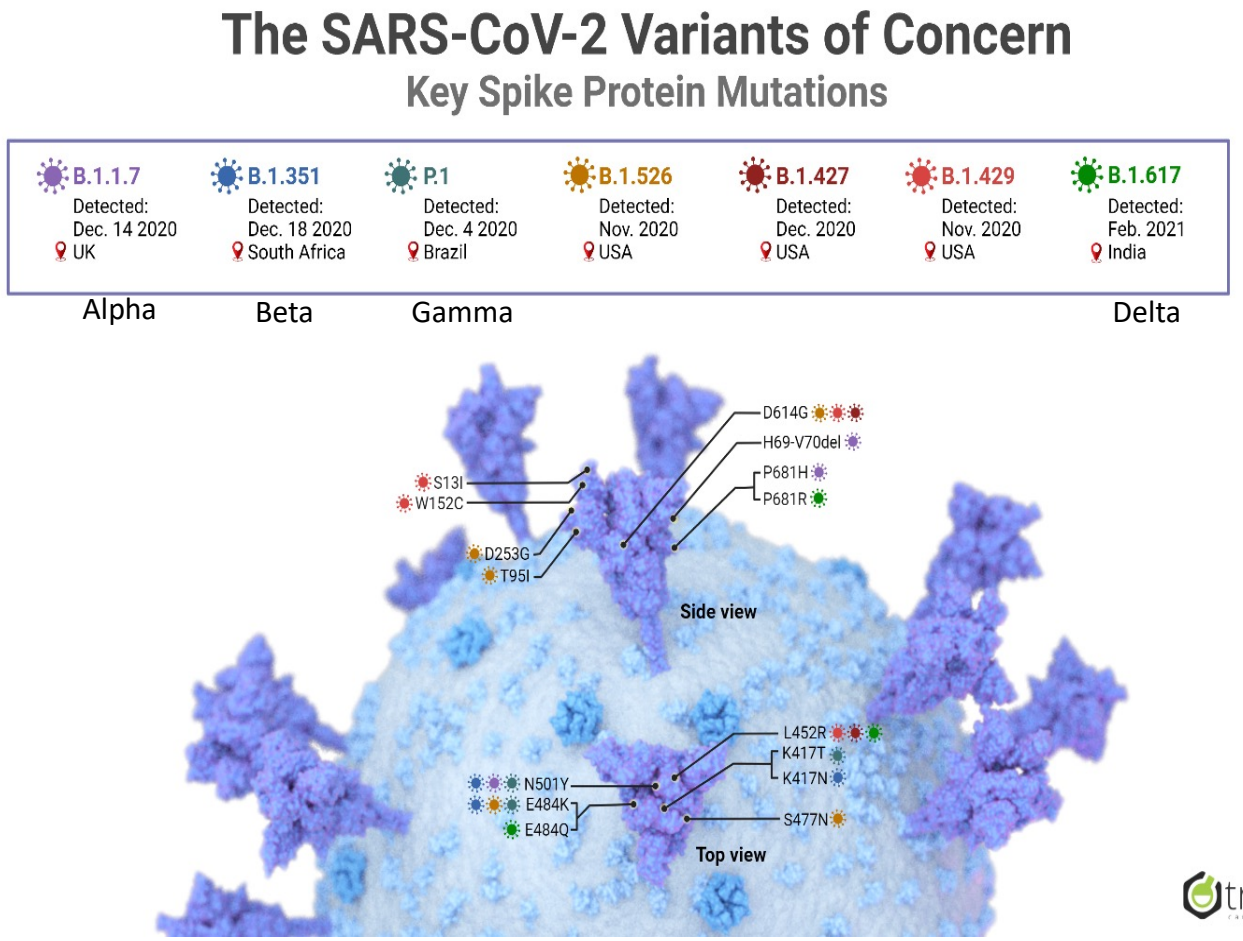
- Examples include HIV, Zika virus, West Nile virus and SARS as well as re-emerging diseases such as chikungunya, influenza and monkeypox



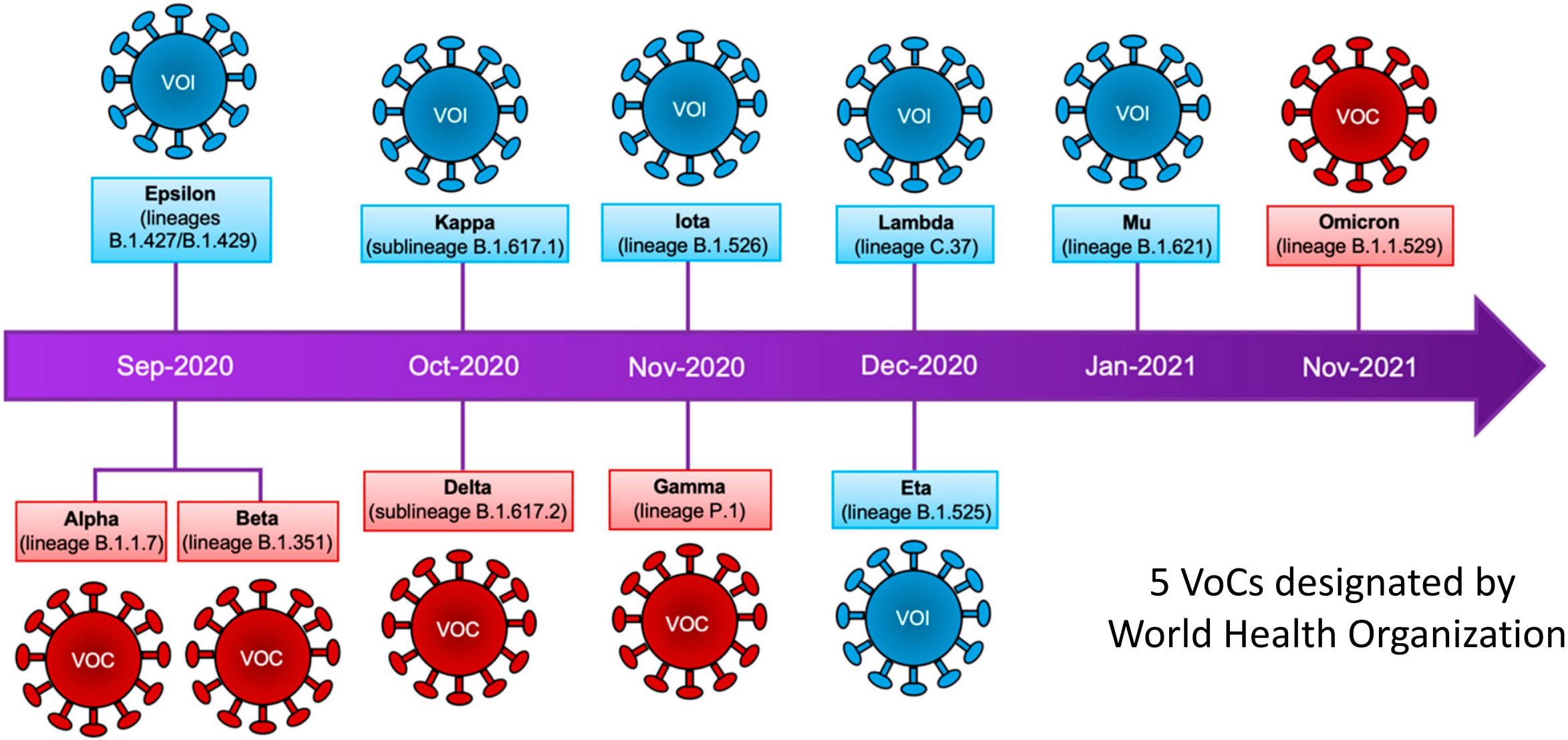
Emerging, re-emerging infectious agents and variants

- Viruses such as SARS-CoV-2 continuously evolve as changes in the genetic code (via genetic mutations or viral recombination) occur during replication of the genome.

- A variant has one or more mutations that differentiate it from other variants of the SARS-CoV-2 viruses. As expected, multiple variants of SARS-CoV-2 have been documented globally throughout the COVID-19 pandemic.

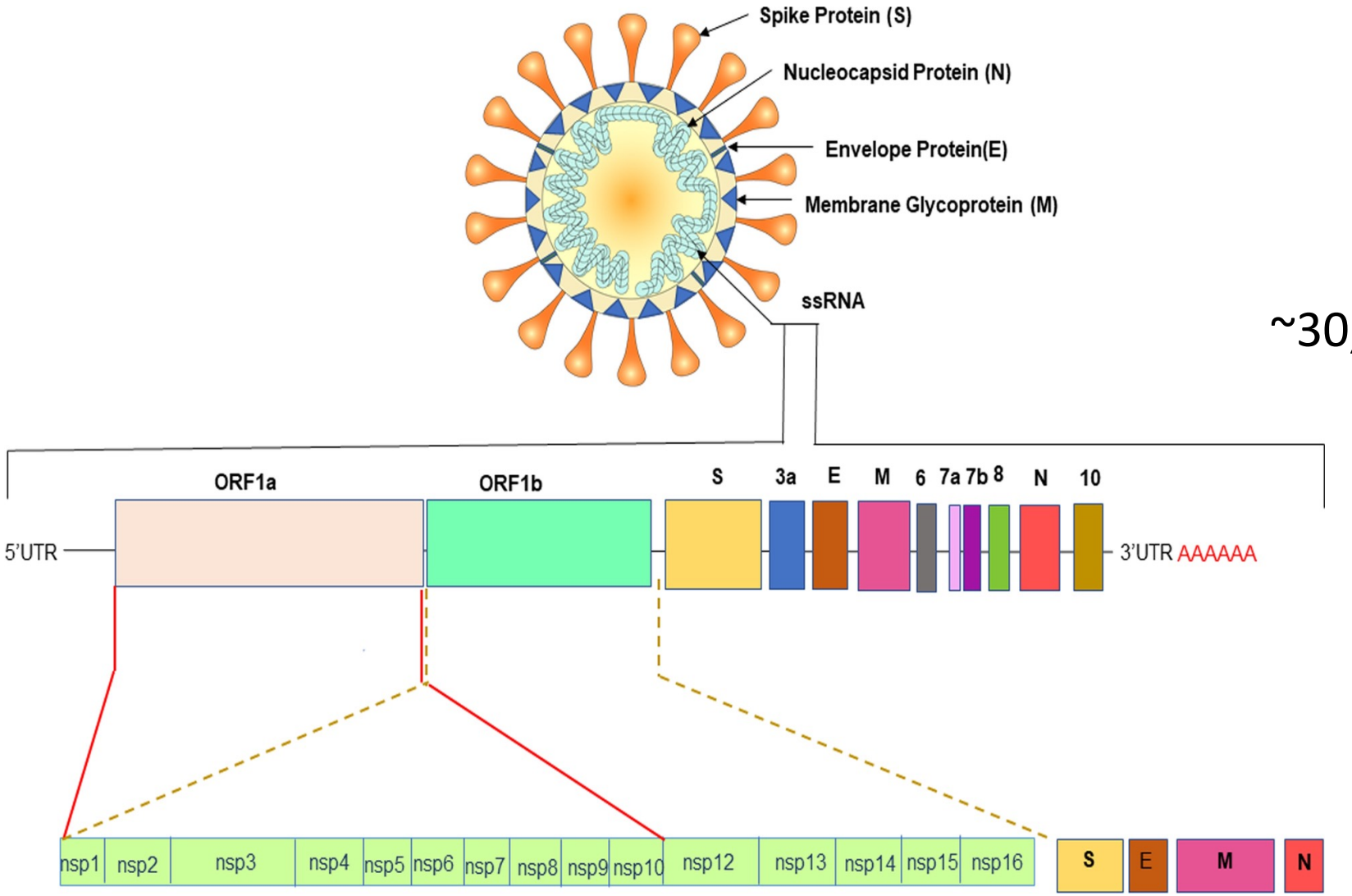


SARS-CoV-2: emerging variants of concern (VoC)



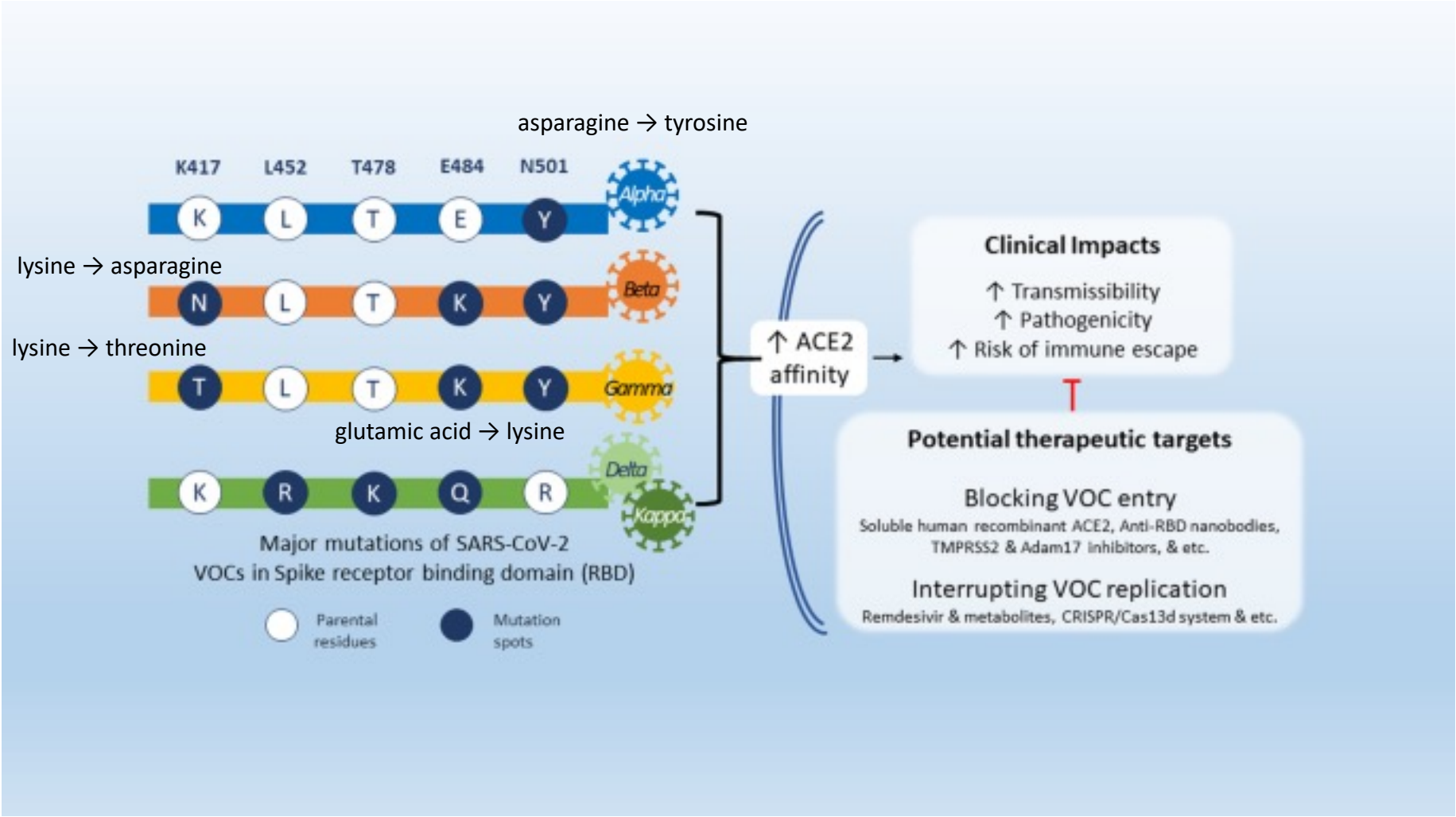
5 VoCs designated by World Health Organization

SARS-CoV-2 VoCs: What do we know?

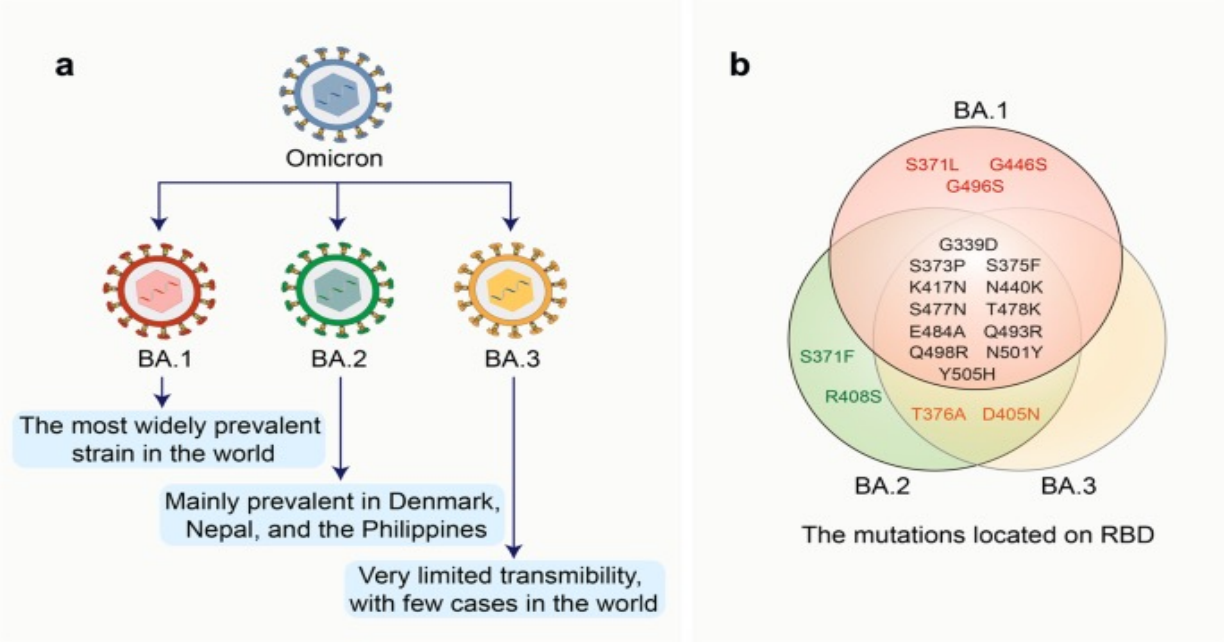


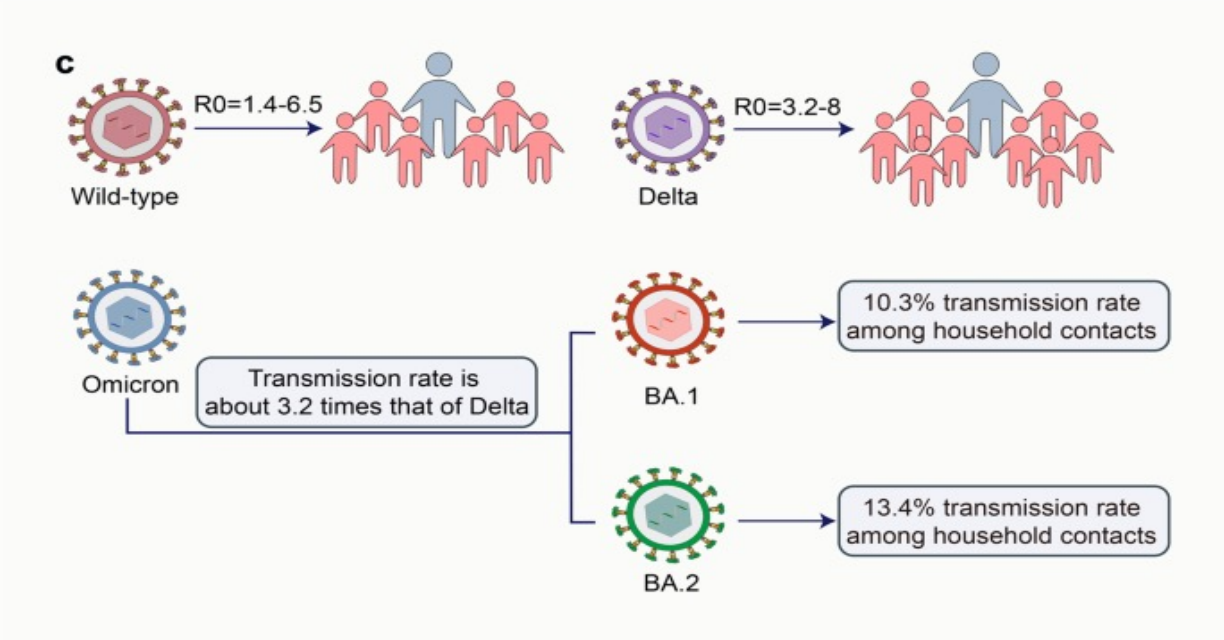
~30,000 nucleotides

Reported mutations in SARS-CoV-2 VoCs



Reported mutations in SARS-CoV-2 VoCs:





Change in Proportions of Omicron Sublineages

2022-07-05

Prevalence of Omicron sublineages collected 07 Jun 2022-05 Jul 2022
compared with sublineages collected 10 May 2022-07 Jun 2022

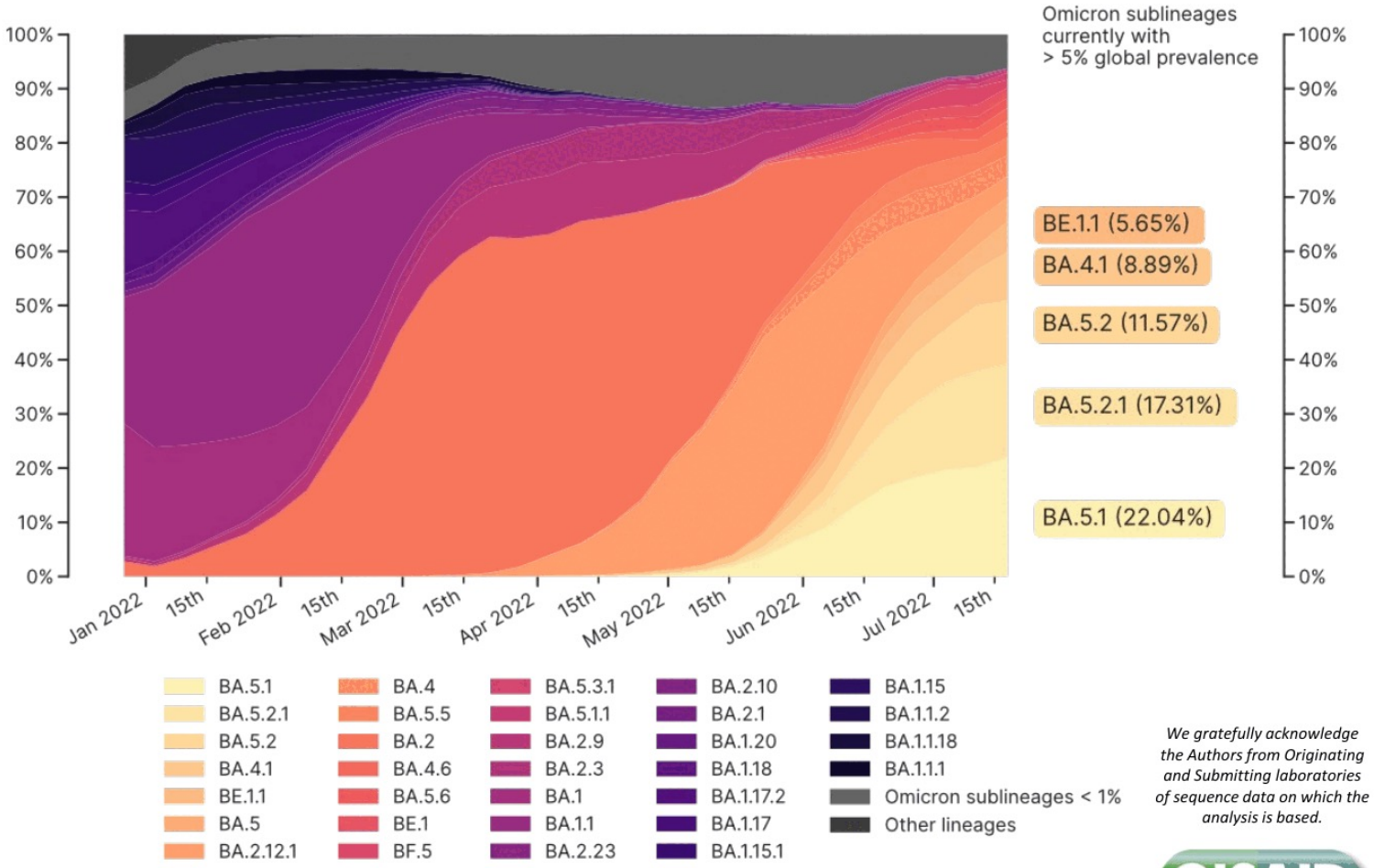
	Africa	Asia	Europe	Oceania	North America	South America
BA.2	-6.6%	-14.0%	-33.7%	-18.7%	-12.9%	-21.5%
BA.2.10	-0.7%	-1.9%	-0.2%	-0.9%	-0.2%	-0.0%
BA.2.10.1	+0.0%	-0.5%	-0.0%	-0.4%	-0.0%	-0.0%
BA.2.12	-0.1%	-0.2%	-0.1%	-1.3%	-0.6%	-0.1%
BA.2.12.1	-0.6%	-1.0%	+2.8%	+1.7%	-7.3%	+6.9%
BA.2.18	+0.0%	-0.0%	-0.3%	-0.1%	-0.1%	-0.0%
BA.2.23	-0.7%	-0.1%	-0.7%	+0.7%	-0.2%	-1.0%
BA.2.3	-0.6%	-13.9%	-1.5%	-4.1%	-2.9%	-2.6%
BA.2.31	+0.3%	-0.3%	+0.0%	+0.0%	-0.1%	-0.0%
BA.2.38	-0.1%	-0.4%	+0.0%	+0.1%	+0.0%	+0.0%
BA.2.9	-0.5%	-2.0%	-8.9%	-0.4%	-1.9%	-2.2%
BA.4	+4.9%	+8.6%	+8.2%	+7.6%	+7.0%	+12.0%
BA.5	+5.7%	+30.9%	+22.4%	+15.2%	+16.3%	+8.7%
BA.5.1	-0.8%	+7.7%	+15.6%	+1.6%	+4.0%	+7.8%

GISAID
by BII/GIS, A*STAR Singapore

YAKIN DOĞU ÜNİVERSİTESİ
DESAM ENSTİTÜSÜ

Timecourse of Omicron variant sublineage distribution

2022-07-26



We gratefully acknowledge the Authors from Originating and Submitting laboratories of sequence data on which the analysis is based.

GISAID

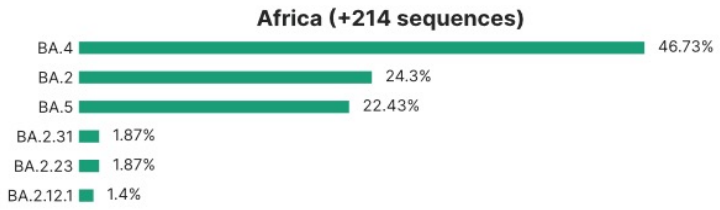
by BII/GIS, A*STAR Singapore

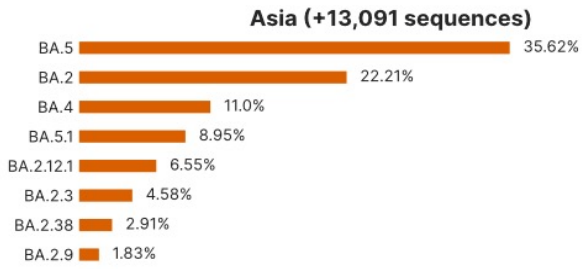
See https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/ for variant information and definitions.

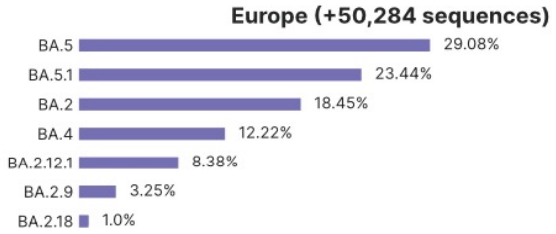
YAKIN DOĞU ÜNİVERSİTESİ
DESAM ENSTİTÜSÜ

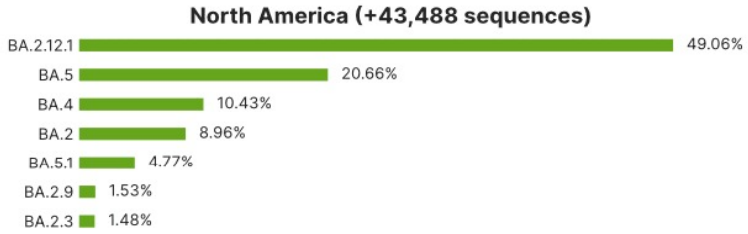
Regional trends of Omicron variant sublineages

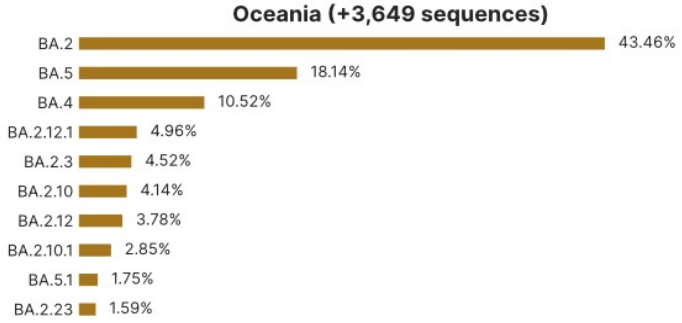
in sequences collected from 2022-06-07 to 2022-07-05

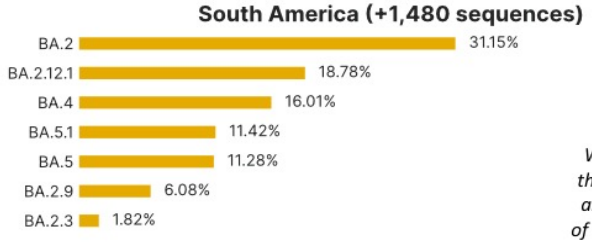












We gratefully acknowledge the Authors from Originating and Submitting laboratories of sequence data on which the analysis is based.

GISAID

SARS-CoV-2 variants BA.4 and BA.5 show substantial immune escape compared with BA.1 and BA.2

A

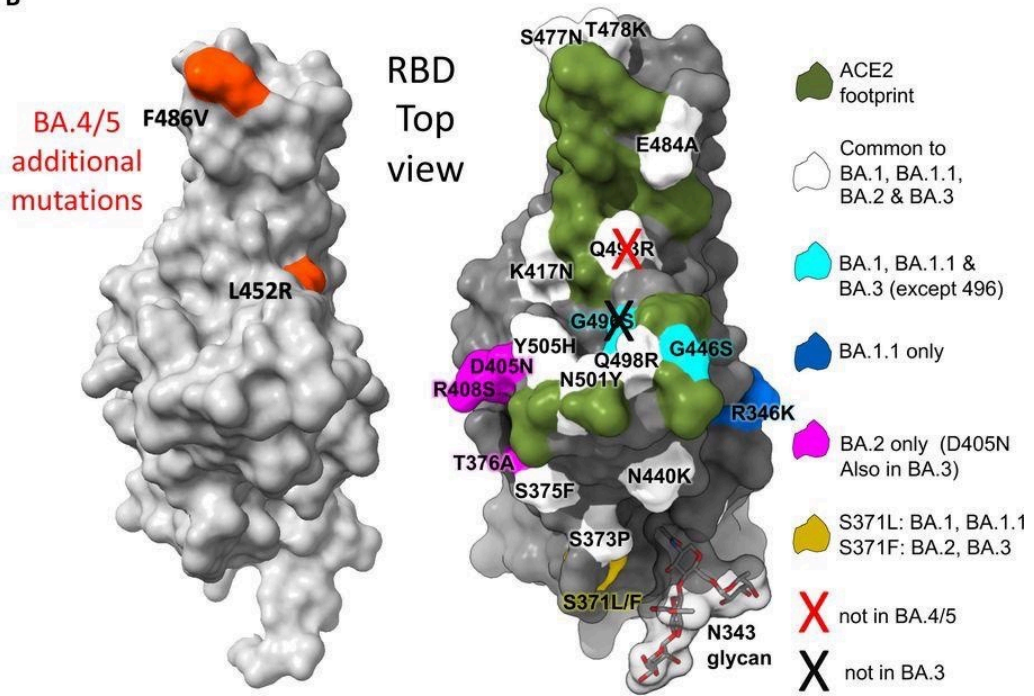
BA.1			A67V,Δ69-70,T95I,G142D,Δ143-145,N211I,Δ212,								ins214EPE	NTD
BA.1.1			A67V,Δ69-70,T95I,G142D,Δ143-145,N211I,Δ212,								ins214EPE	
BA.2	T19I,Δ24-26,A27S,					G142D,				V213G		
BA.3			A67V,Δ69-70,T95I,G142D, Δ143-145,N211I,Δ212									
BA.4/5	T19I,Δ24-26,A27S,			Δ69-70,		G142D,				V213G		

BA.1	G339D,	S371L,	S373P,S375F,					K417N,N440K,G446S		RBD
BA.1.1	G339D,R346K,	S371L,	S373P,S375F,					K417N,N440K,G446S		
BA.2	G339D,	S371F,	S373P,S375F,	T376A,D405N,R408S,				K417N,N440K		
BA.3	G339D,	S371F,	S373P,S375F,		D405N,			K417N,N440K,G446S		
BA.4/5	G339D,	S371F,	S373P,S375F,	T376A,D405N,R408S,				K417N,N440K,		

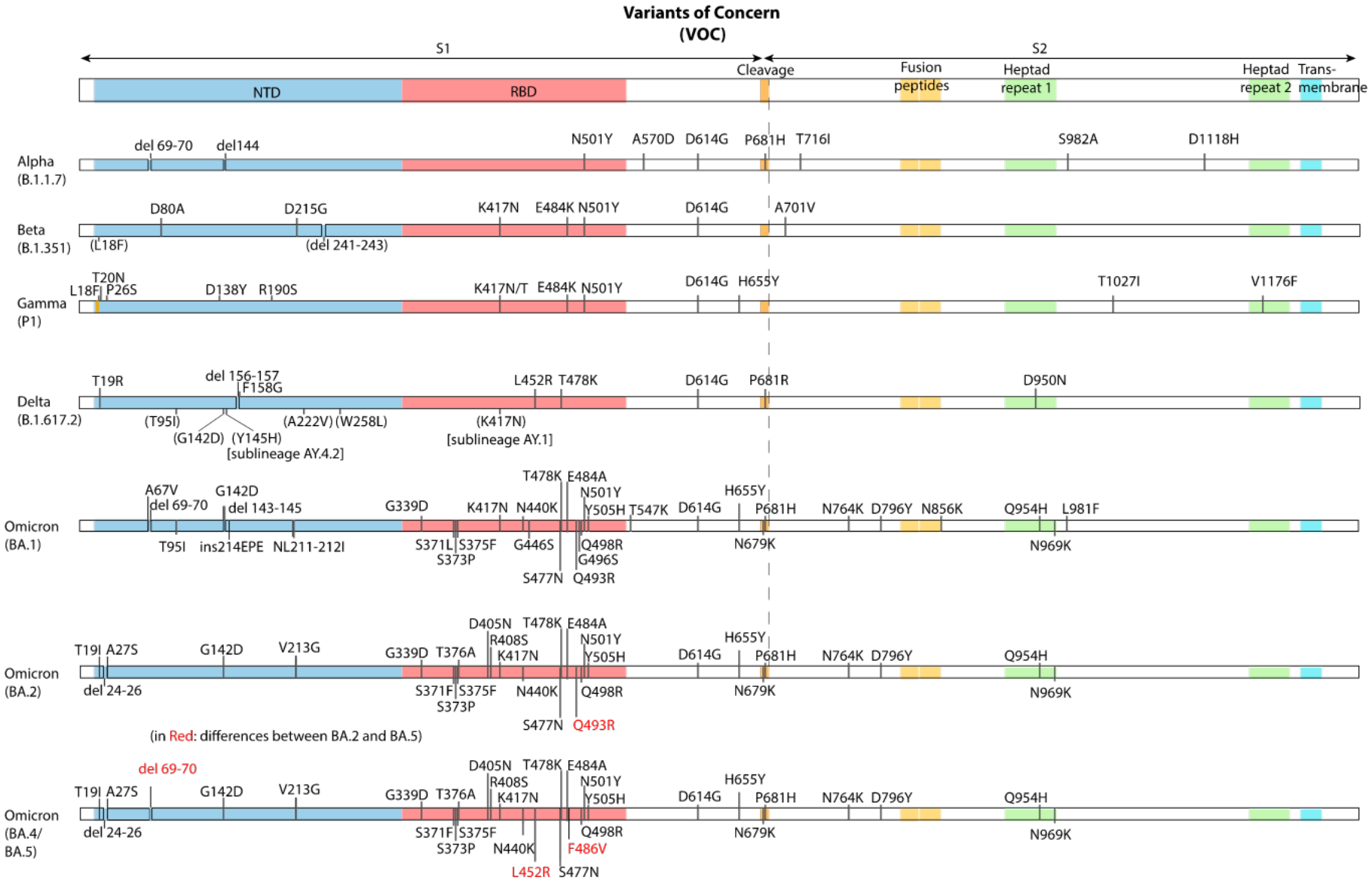
BA.1	S477N,T478K,E484A,	Q493R,G496S,	Q498R,N501Y,Y505H			
BA.1.1	S477N,T478K,E484A,	Q493R,G496S,	Q498R,N501Y,Y505H			
BA.2	S477N,T478K,E484A,	Q493R,	Q498R,N501Y,Y505H			
BA.3	S477N,T478K,E484A,	Q493R,	Q498R,N501Y,Y505H			
BA.4/5	L452R,S477N,T478K,E484A,F486V,		Q498R,N501Y,Y505H			

BA.1	T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F					
BA.1.1	T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F					
BA.2	D614G,H655Y,N679K,P681H,N764K,D796Y,				Q954H,N969K	
BA.3	D614G,H655Y,N679K,P681H,N746K,D796Y,				Q954H,N969K	
BA.4/5	D614G,H655Y,N679K,P681H,N764K, D796Y,				Q954H,N969K	

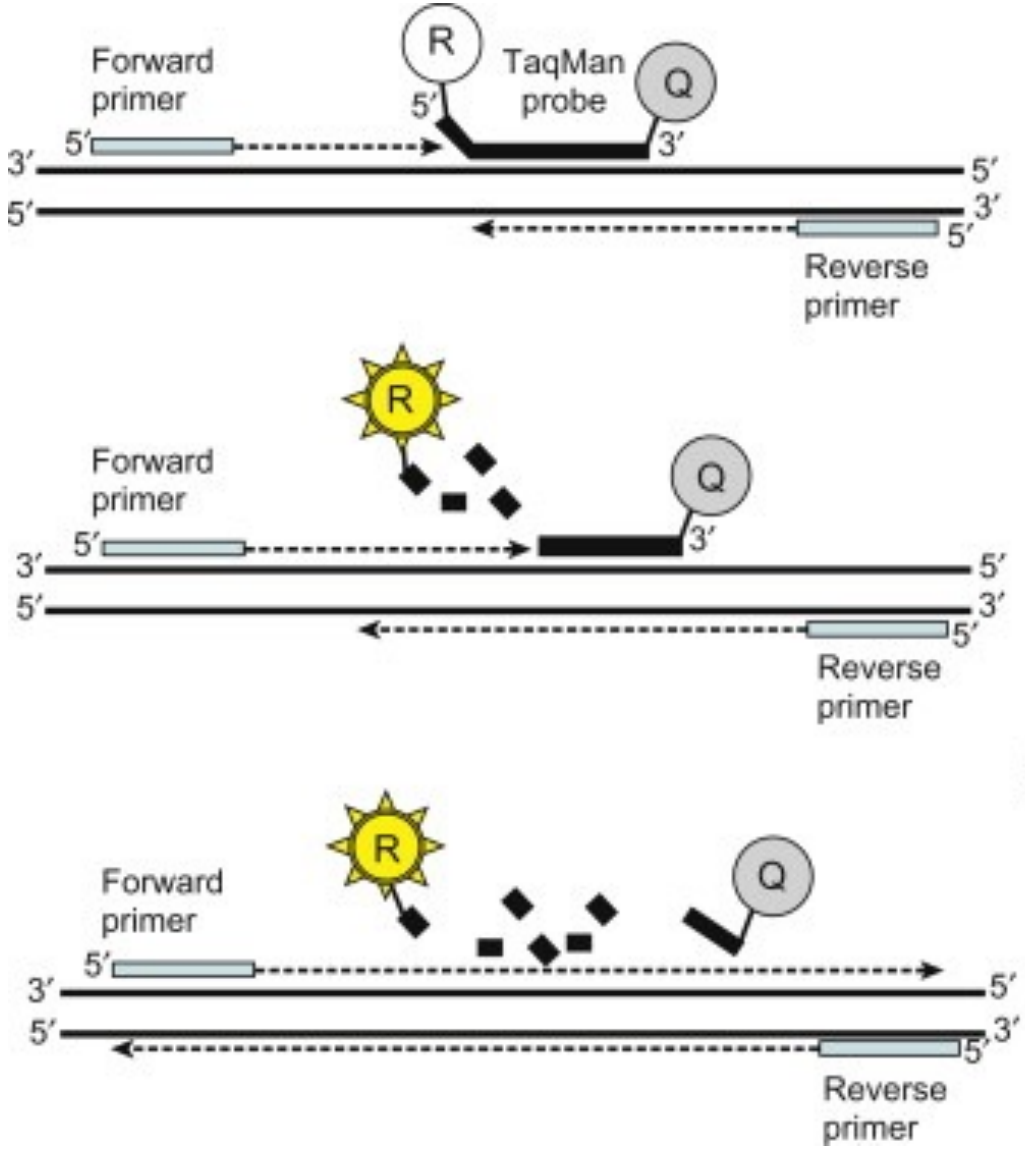
B



SARS-CoV-2 VoC Mutation Profile



RT-PCR Detection of Pathogen DNA/RNA



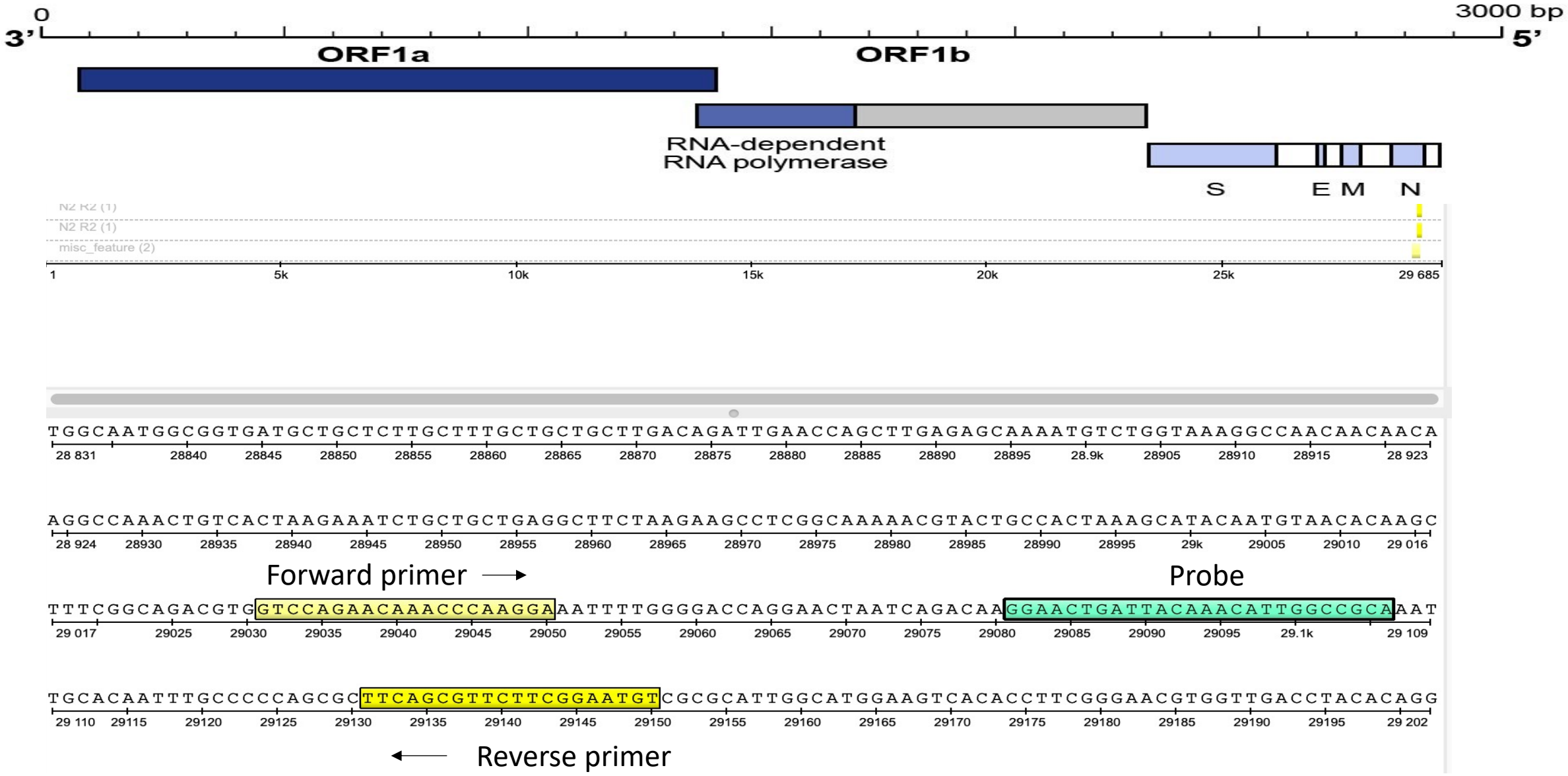
Polymerization and Strand Displacement

Probe Cleavage
(release of reporter dye)

Fluorescence occurs when reporter dye and quencher dye are no longer in close proximity

Completion of Polymerization

RT-PCR Detection of Pathogen DNA/RNA

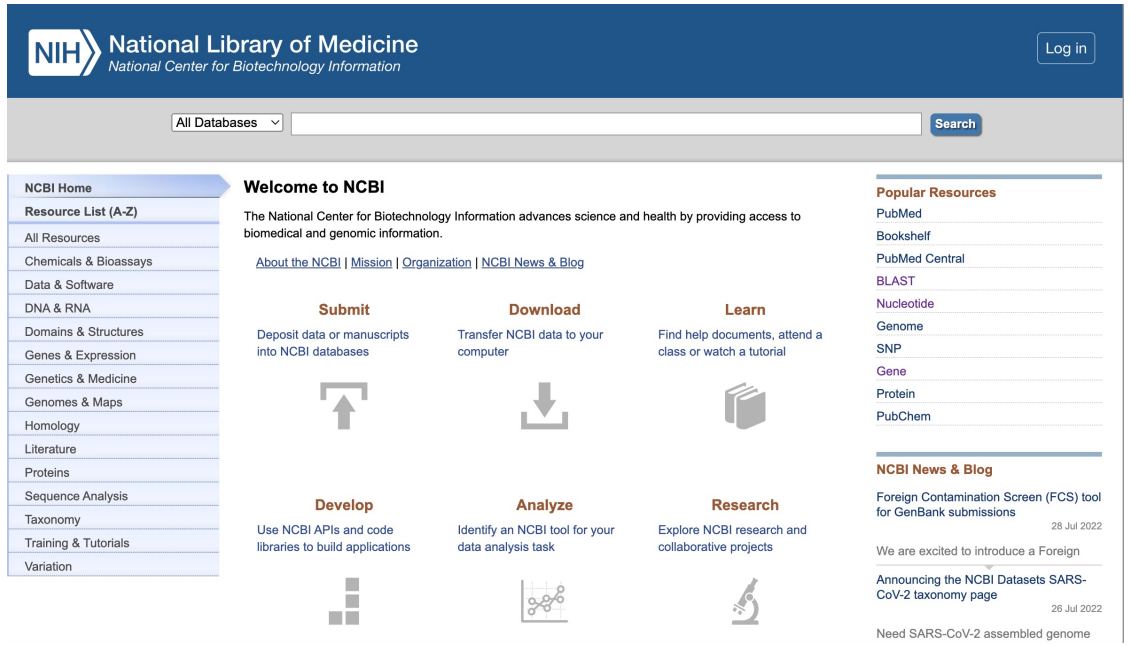


Databases for Obtaining Pathogen Genome Data

- www.gisaid.org GISAID

Global initiative on sharing all influenza data

- www.ncbi.nlm.nih.gov



- www.ncbi.nlm.nih.gov/labs/virus/vssi NCBI Virus *Sequences for discovery*

- https://covariants.org/ CoVariants Enabled by data from GISAID

- www.fludb.org/ IRD Influenza Research Database

- https://www.hiv.lanl.gov/
  HIV Sequence Database

GISAID Database

- The GISAID Initiative promotes the rapid sharing of data from all influenza viruses, SARS-CoV-2 and recently monkeypox virus
- This includes genetic sequence and related clinical and epidemiological data associated with human viruses, and geographical data associated with avian and other animal viruses
- Aim is to help researchers understand how viruses evolve and spread during epidemics and pandemics

GISAID

YAKIN DOĞU ÜNİVERSİTESİ
DESAM ENSTİTÜSÜ

NCBI Database

ncbi.nlm.nih.gov/gene/43740575

National Library of Medicine
National Center for Biotechnology Information

Log in

Gene | Gene | Search
Advanced | Help

Full Report | Send to: | Hide sidebar >>

N nucleocapsid phosphoprotein [*Severe acute respiratory syndrome coronavirus 2*]

Download Datasets

Gene ID: 43740575, updated on 3-Jul-2022

Summary

Gene symbol	N
Gene description	nucleocapsid phosphoprotein
Locus tag	GU280_gp10
Gene type	protein coding
RefSeq status	PROVISIONAL
Organism	Severe acute respiratory syndrome coronavirus 2 (isolate: Wuhan-Hu-1, nat-host: Homo sapiens)
Lineage	Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes; Nidovirales; Cornidovirineae; Coronaviridae; Orthocoronavirinae; Betacoronavirus; Sarbecovirus
Summary	Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is an enveloped, positive-sense, single-stranded RNA virus that causes coronavirus disease 2019 (COVID-19). Virus particles include the RNA genetic material and structural proteins needed for invasion of host cells. Once inside the cell the infecting RNA is used to encode structural proteins that make up virus particles, nonstructural proteins that direct virus assembly, transcription, replication and host control and accessory proteins whose function has not been determined.~ The structural proteins of SARS-CoV-2 include the envelope protein (E), spike or surface glycoprotein (S), membrane protein (M) and the nucleocapsid protein (N). The nucleocapsid phosphoprotein is a structural protein that binds to, protects the viral RNA genome and is involved in packaging the RNA into virus particles. The N protein has been suggested as an antiviral drug target.

Table of contents

Related information

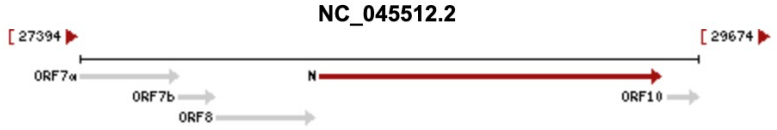
3D structures
BioProjects

NCBI Database

YAKIN DOĞU ÜNİVERSİTESİ
DESAM ENSTİTÜSÜ

Genomic context

Sequence: NC_045512.2 (28274..29533)

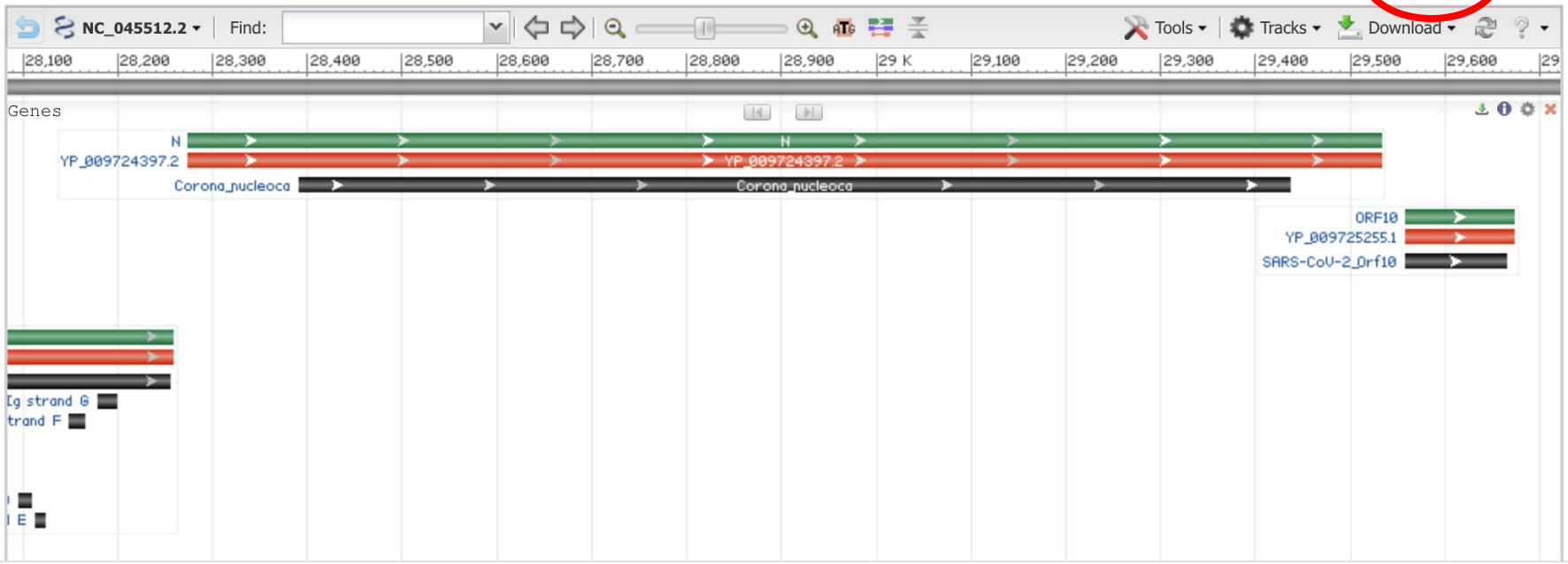


Genomic regions, transcripts, and products

Genomic Sequence: NC_045512.2

Go to reference sequence details

Go to nucleotide: Graphics FASTA GenBank



Full text in PMC_nucleotide
Gene neighbors
Genome
Nucleotide
Protein
PubMed
PubMed (GeneRIF)
PubMed(nucleotide/PMC)
RefSeq Proteins
Taxonomy

General information

About Gene
FAQ
FTP site
Help
My NCBI help
NCBI Handbook
Statistics

Related sites

BLAST
Genome

NCBI Database

ncbi.nlm.nih.gov/nuccore/NC_045512.2?report=fasta&from=28274&to=29533

National Library of Medicine
National Center for Biotechnology Information

Log in

Nucleotide | Nucleotide | Search
Advanced | Help

FASTA | Send to:

Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome

NCBI Reference Sequence: NC_045512.2

GenBank Graphics

```
>NC_045512.2:28274-29533 Severe acute respiratory syndrome coronavirus 2 isolate
Wuhan-Hu-1, complete genome
ATGTCTGATAATGGACCCCAAAATCAGCGAAATGCACCCCGCATTACGTTTGGTGGACCCTCAGATTCAA
CTGGCAGTAACCAGAATGGAGAACGCAGTGGGGCGCGATCAAAACAACGTCGGCCCCAAGGTTTACCCAA
TAATACTGCGTCTTGGTTCACCGCTCTCACTCAACATGGCAAGGAAGACCTTAAATTCCCTCGAGGACAA
GGCGTTCCAATTAACACCAATAGCAGTCCAGATGACCAAATTGGCTACTACCGAAGAGCTACCAGACGAA
TTCGTGGTGGTGACGGTAAAATGAAAGATCTCAGTCCAAGATGGTATTTCTACTACCTAGGAACTGGGCC
AGAAGCTGGACTTCCCTATGGTGCTAACAAAGACGGCATCATATGGGTTGCAACTGAGGGAGCCTTGAAT
ACACCAAAAGATCACATTGGCACCCGCAATCCTGCTAACAATGCTGCAATCGTGCTACAACTTCCTCAAG
GAACAACATTGCCAAAAGGCTTCTACGCAGAAGGGAGCAGAGGCGGCAGTCAAGCCTCTTCTCGTTCCTC
ATCACGTAGTCGCAACAGTTCAAGAAATTCAACTCCAGGCAGCAGTAGGGGAACTTCTCCTGCTAGAATG
GCTGGCAATGGCGGTGATGCTGCTCTTGCTTTGCTGCTGCTTGACAGATTGAACCAGCTTGAGAGCAAAA
TGTCTGGTAAAGGCCAACAACAACAAGGCCAAACTGTCACTAAGAAATCTGCTGCTGAGGCTTCTAAGAA
GCCTCGGCAAAAACGTACTGCCACTAAAGCATACAATGTAACACAAGCTTTCGGCAGACGTGGTCCAGAA
CAAACCCAAGGAAATTTTGGGGACCAGGAACTAATCAGACAAGGAACTGATTACAAACATTGGCCGCAAA
TTGCACAATTTGCCCCCAGCGCTTCAGCGTTCTTCGGAATGTCGCGCATTGGCATGGAAGTCACACCTTC
GGGAACGTGGTTGACCTACACAGGTGCCATCAAATTGGATGACAAAGATCCAAATTTCAAAGATCAAGTC
ATTTTGCTGAATAAGCATATTGACGCATACAAAACATTCCCACCAACAGAGCCTAAAAAGGACAAAAAGA
AGAAGGCTGATGAAACTCAAGCCTTACCGCAGAGACAGAAGAAACAGCAAACTGTGACTCTTCTTCCTGC
TGCAGATTTGGATGATTTCTCCAAACAATTGCAACAATCCATGAGCAGTGCTGACTCAACTCAGGCCTAA
```

Change region shown

- ○ Whole sequence
- ◉ Selected region

from: 28274 to: 29533

Update View

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

NCBI Virus

Retrieve, view, and download SARS-CoV-2 coronavirus genomic and protein sequences.

Related information

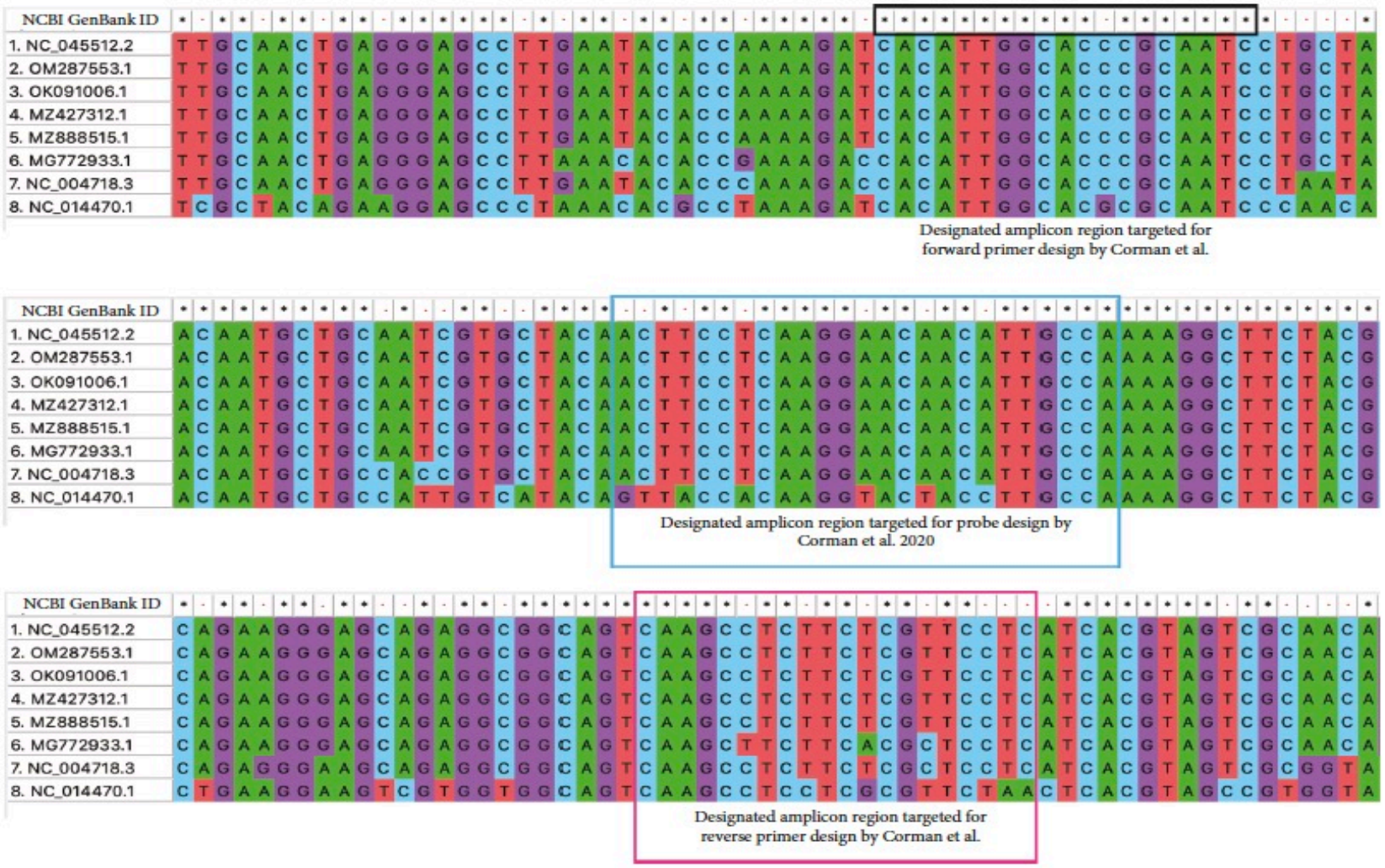
After obtaining genome data…

- We can analyze microorganism genome data using multiple software including Snapgene, Ugene, CLC sequence viewer

- We can run Multiple Sequence Alignment analysis to compare overall sequence similarity of multiple genome sequences using tools such as MUSCLE, Clustal Omega, EMBOSS Cons, Mview

Multiple Sequence Alignment

Nucleocapsid (N) gene



* denotes perfect nucleotide position match

- red dot denotes mismatch

FIGURE 2: N gene designated fragment—MSA (MUSCLE) by MEGA11 version 0.1. The region of primers and probe is designed by [1]. Star (*) signs denote perfect nucleotide position match, and red dots (.) denote mismatch. The black rectangle shape denotes the amplicon region for forward primer design, the blue rectangle shape denotes the amplicon region for probe design, and the pink rectangle shape denotes amplicon position for reverse primer position. Analyzed NCBI GenBank accession IDs for MSA for the entire E gene represent as follows:

Multiple Sequence Alignment

Envelope (E) gene

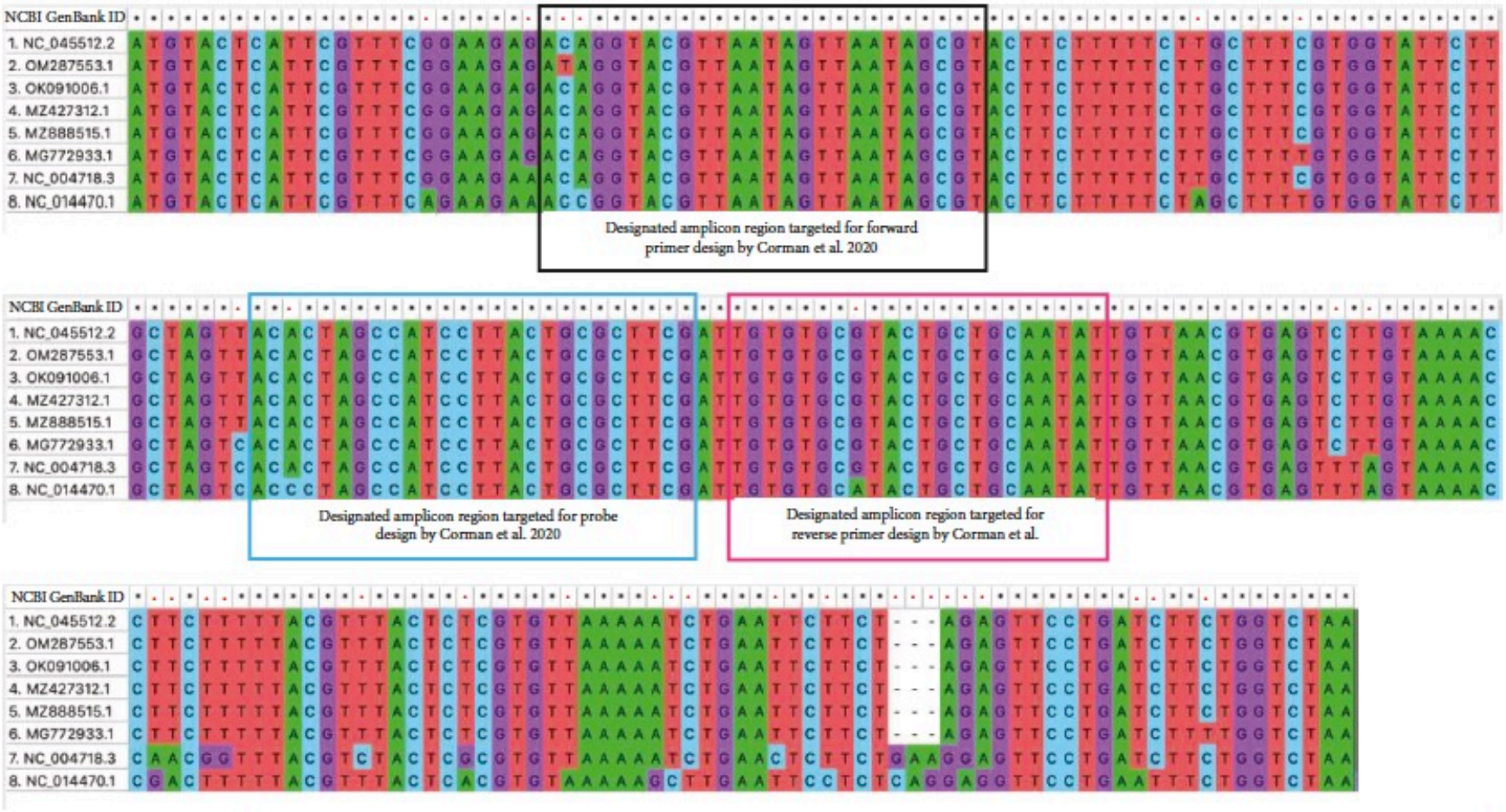


FIGURE 1: E gene multiple sequence alignment (MUSCLE) by MEGA11 version 0.1. The region of primers and probe was designed by [1]. Star (*) signs denote perfect nucleotide position match, and red dots (.) denote mismatch. The black rectangle shape denotes the amplicon region for forward primer design, the blue rectangle shape denotes the amplicon region for probe design, and the pink rectangle shape denotes amplicon position for reverse primer position. Analyzed NCBI GenBank accession IDs for MSA for the entire E gene represent as

* denotes perfect nucleotide position match

- red dot denotes mismatch

Target Selection for RT-PCR Detection of Pathogens

- The gene target for PCR amplification assays should be:

- **conserved**
- **expressed during infection cycle**
- **involved in pathogenesis/replication**
- **less prone to mutations**

YAKIN DOĞU ÜNİVERSİTESİ
DESAM ENSTİTÜSÜ

Target Selection for RT-PCR Detection of Pathogens

Example: Omicron variant BA.4

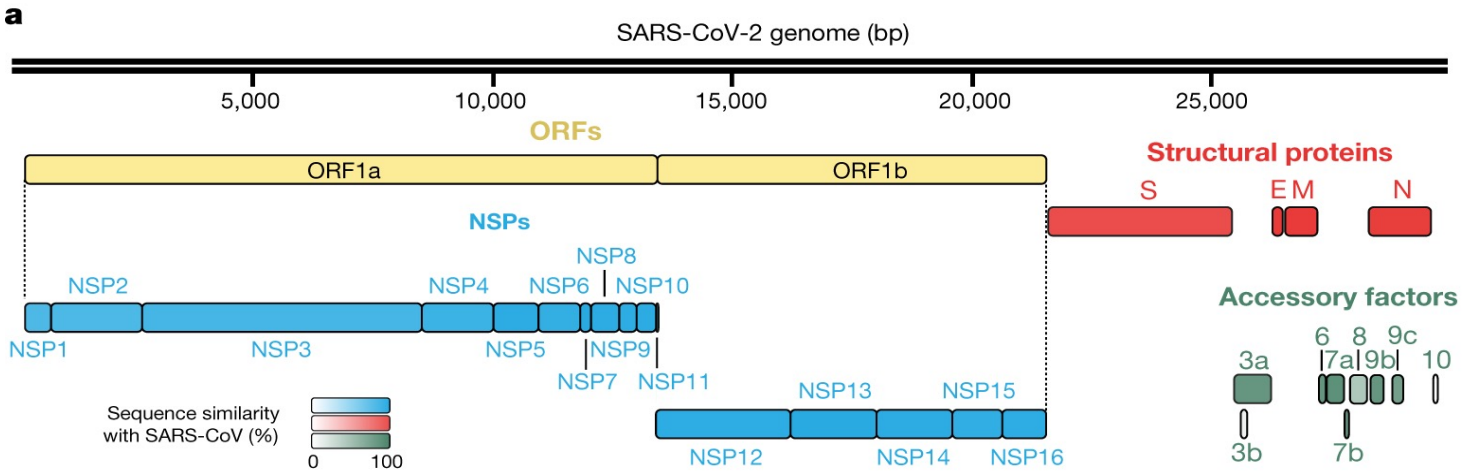
Defining mutations

Nonsynonymous

S: T 19 I
S: L 24 -
S: P 25 -
S: P 26 -
S: A 27 S
S: H 69 -
S: V 70 -
S: G 142 D
S: V 213 G
S: G 339 D
S: S 371 F
S: S 373 P
S: S 375 F
S: T 376 A
S: D 405 N
S: R 408 S
S: K 417 N
S: N 440 K
S: L 452 R
S: S 477 N
S: T 478 K
S: E 484 A
S: F 486 V
S: Q 498 R

S: N 501 Y
S: Y 505 H
S: D 614 G
S: H 655 Y
S: N 679 K
S: P 681 H
S: N 764 K
S: D 796 Y
S: Q 954 H
S: N 969 K
N: P 13 L
N: E 31 -
N: R 32 -
N: S 33 -
N: P 151 S
N: R 203 K
N: G 204 R
N: S 413 R
ORF1a: S 135 R
ORF1a: K 141 -
ORF1a: S 142 -
ORF1a: F 143 -
ORF1a: T 842 I
ORF1a: G 1307 S
ORF1a: L 3027 F
ORF1a: T 3090 I
ORF1a: T 3255 I
ORF1a: P 3395 H

ORF1a: S 135 R
ORF1a: K 141 -
ORF1a: S 142 -
ORF1a: F 143 -
ORF1a: T 842 I
ORF1a: G 1307 S
ORF1a: L 3027 F
ORF1a: T 3090 I
ORF1a: T 3255 I
ORF1a: P 3395 H
ORF1a: S 3675 -
ORF1a: G 3676 -
ORF1a: F 3677 -
ORF1b: P 314 L
ORF1b: R 1315 C
ORF1b: I 1566 V
ORF1b: T 2163 I
ORF3a: T 223 I
ORF6: D 61 L
ORF7b: L 11 F
ORF9b: P 10 S
ORF9b: E 27 -
ORF9b: N 28 -
ORF9b: A 29 -
E: T 9 I
M: Q 19 E
M: A 63 T

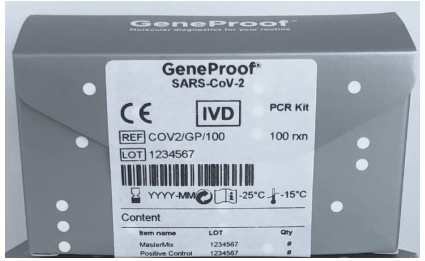


Gene Targets Used for Commercial RT-PCR Assay Kits for SARS-CoV-2

- As SARS-CoV-2 variants continue to emerge worldwide, diagnostic developers face increasing challenges to demonstrate that SARS-CoV-2 assays will continue to detect the virus variant that may be circulating in the population being tested.

- At the beginning of the COVID-19 pandemic, many RT-PCR kits were designed to detect the viral spike (S) gene

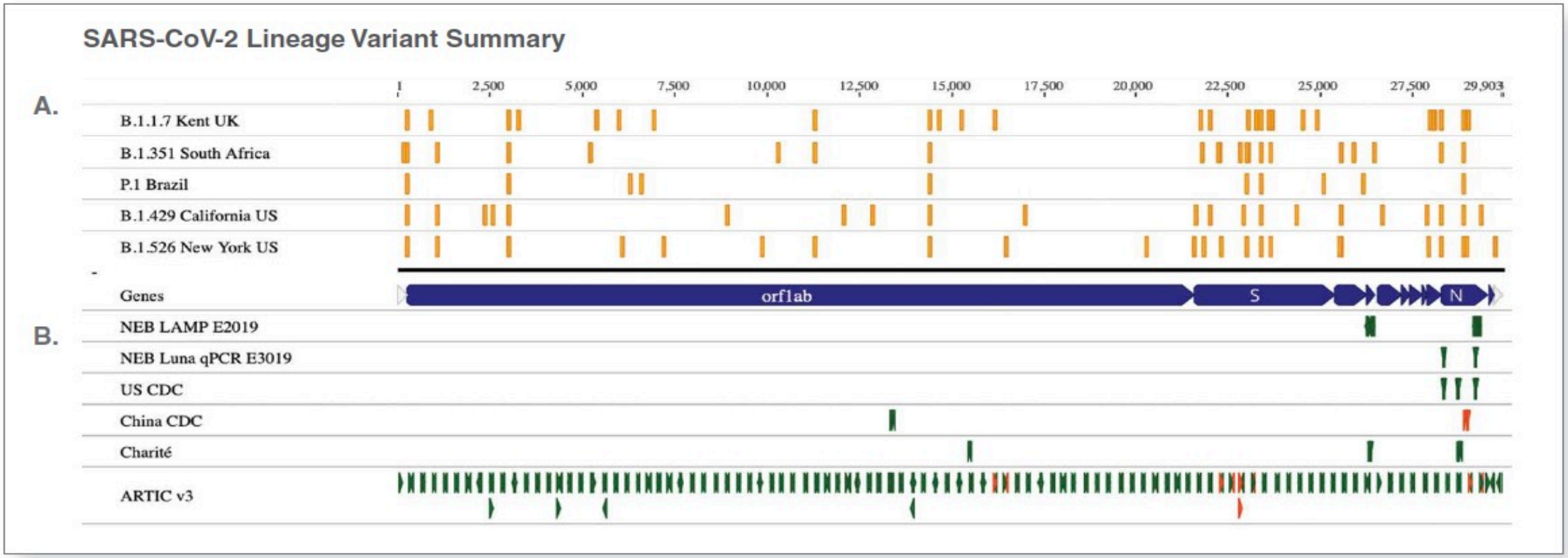
- With the increasing S gene mutations in emerging variants (S gene dropout), scientists moved to other gene targets such as ORF1ab, N gene, E gene, RdRp gene



Primer Monitor: an online tool to track SARS-CoV-2 variants that may impact primers used in diagnostic assays

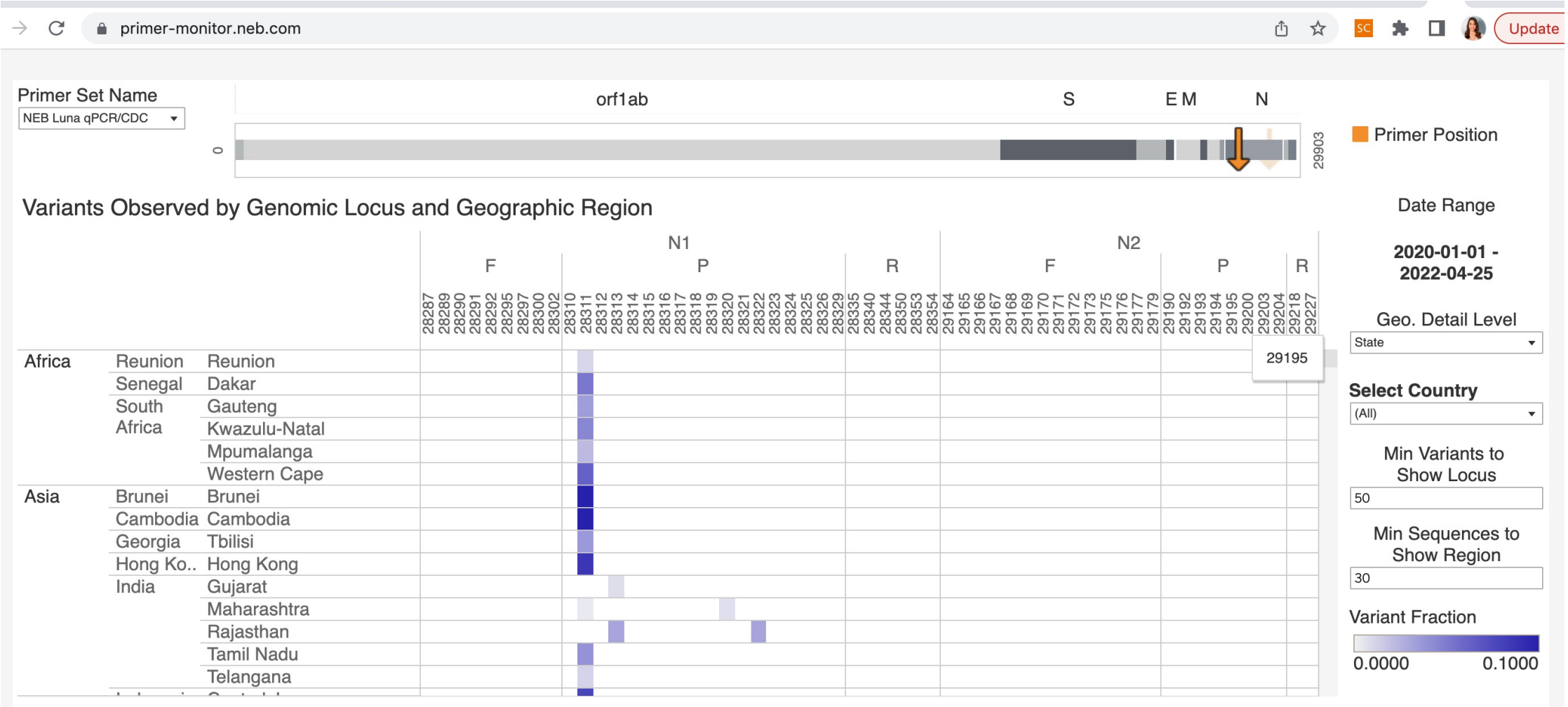
https://primer-monitor.neb.com/

Commonly discussed variants of interest or concern are depicted along with specific mutational loci (A). Below the reference SARS-CoV-2 genome (blue), commonly used primer sets that overlap variants of interest/concern are highlighted in orange.



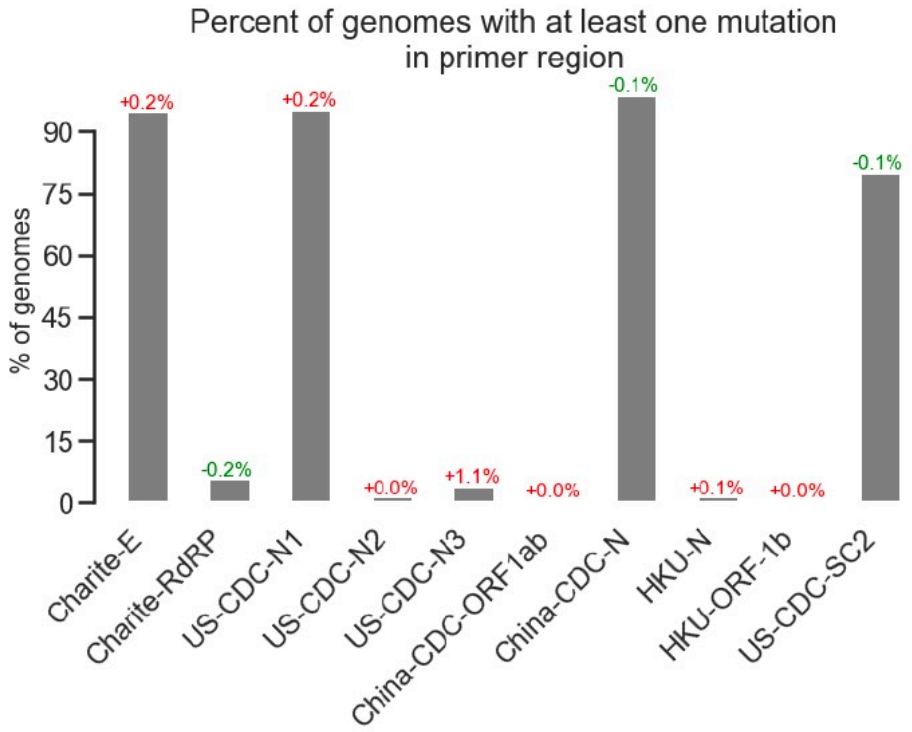
Primer Monitor

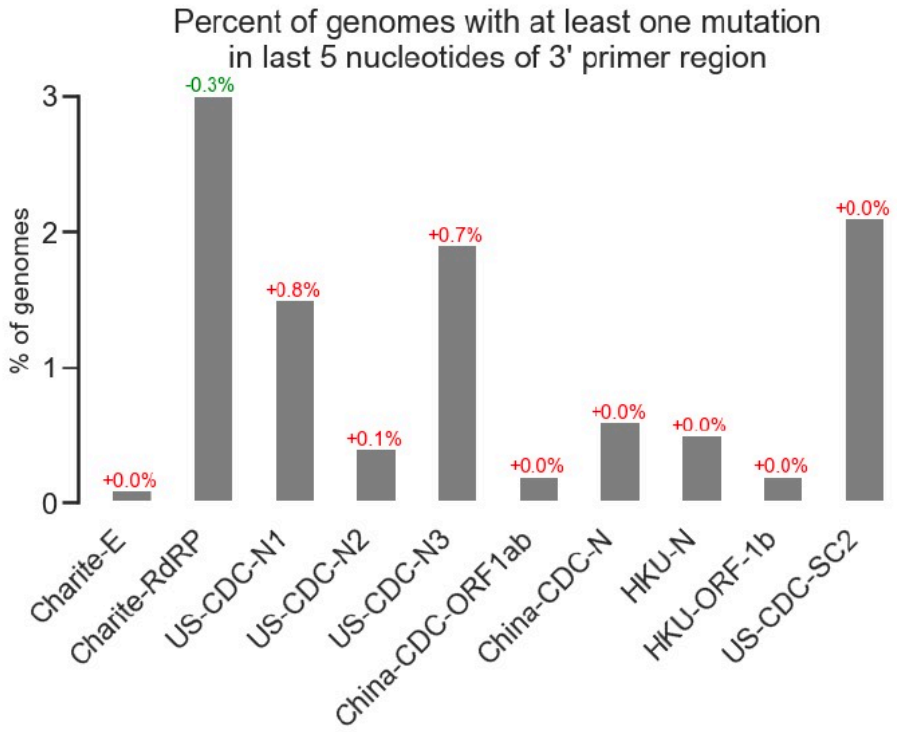
https://primer-monitor.neb.com/



Impact of Mutations on Diagnostic Assays

Common Primer Check for High Quality Genomes 2022-07-05





GISAID

Developing your own diagnostic assay…

- For an accurate and efficient diagnostic assay for emerging and re-emerging pathogens, periodic screening for mutations in the genome should be performed using shared databases.

- With the right tools, theoretical and practical knowledge, you can design molecular diagnostic assays for any pathogen of interest..


thank you