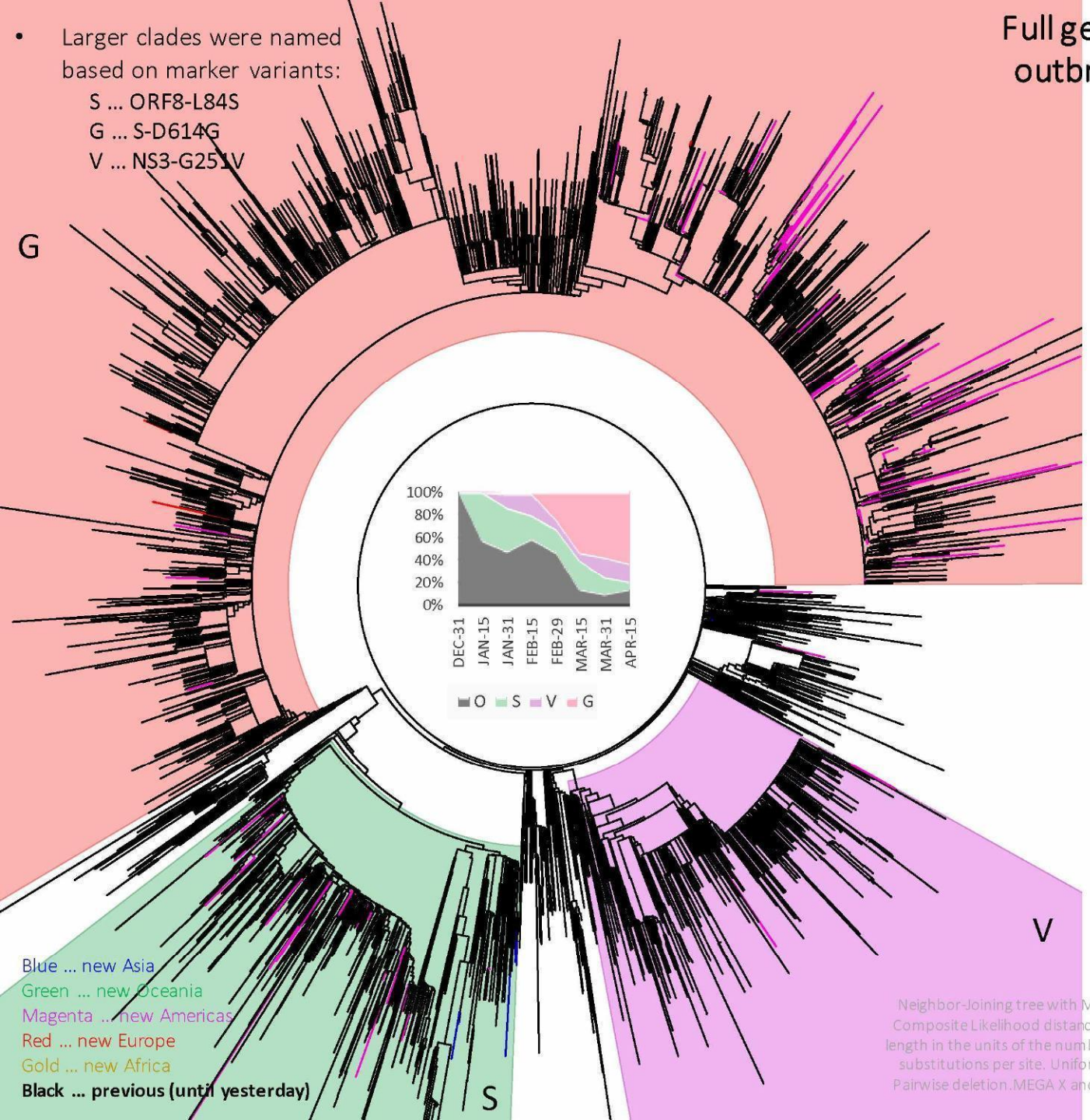
Latest update

2020-04-16 1600UTC

Full genome tree derived from all outbreak sequences 2[illegible]16

- Larger clades were named based on marker variants:
  - S ... ORF8-L84S
  - G ... S-D614G
  - V ... NS3-G251V



Notable changes:
8,646 full genomes (+188)
(excluding low coverage, out of 9,265 entries)

S clade 1,070 (+44):
31 USA/WA, 9 Korea, 2 USA/CT, 1 USA/IL, 1 USA/NY
G clade 4,987 (+137):
53 USA/WA, 36 USA/NY, 19 USA/ID, 11 USA/CT, 5 Latvia, 3 USA/MN, 3 USA/WI, 2 USA/OR, 2 USA/UN, 2 USA/IL, 1 India
V clade 1,204 (+2):
1 USA/WA, 1 USA/NY
Other clades 1,385 (+5):
3 Korea, 1 USA/IL, 1 USA/NY

Blue ... new Asia
Green ... new Oceania
Magenta ... new Americas
Red ... new Europe
Gold ... new Africa
Black ... previous (until yesterday)

Neighbor-Joining tree with Maximum Composite Likelihood distance. Branch length in the units of the number of base substitutions per site. Uniform rates. Pairwise deletion. MEGA X and FigTree.

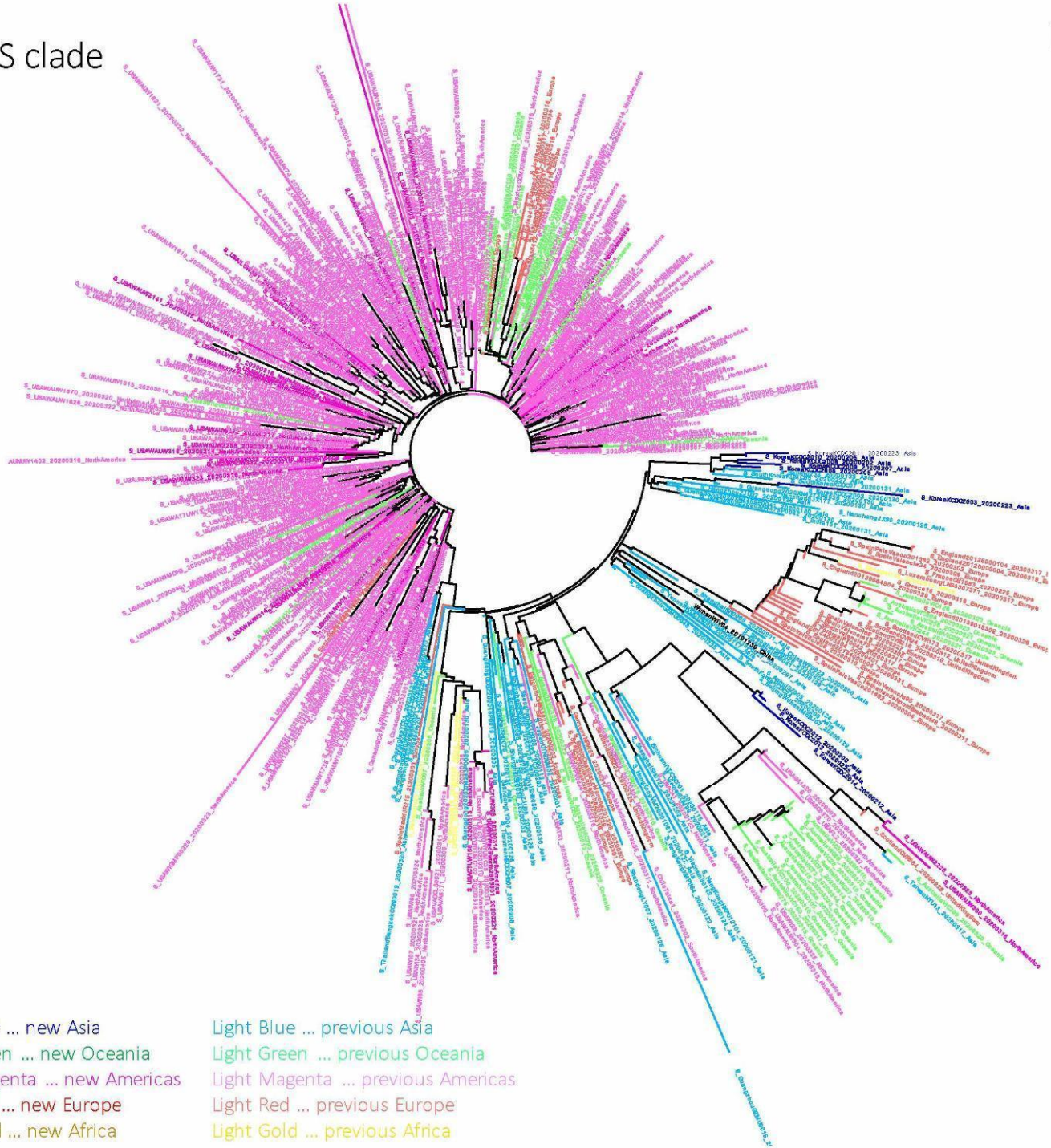
We gratefully acknowledge the Authors from Originating and Submitting laboratories of sequence data on which the analysis is based.

GISAID

by BII/GIS, A*STAR Singapore

Full genome trees of major subclades 201697943.6

S clade



Notable changes:

S clade 1,070 (+44):
31 USA/WA, 9 Korea, 2 USA/CT, 1 USA/IL, 1 USA/NY

Blue ... new Asia
Green ... new Oceania
Magenta ... new Americas
Red ... new Europe
Gold ... new Africa

Light Blue ... previous Asia
Light Green ... previous Oceania
Light Magenta ... previous Americas
Light Red ... previous Europe
Light Gold ... previous Africa

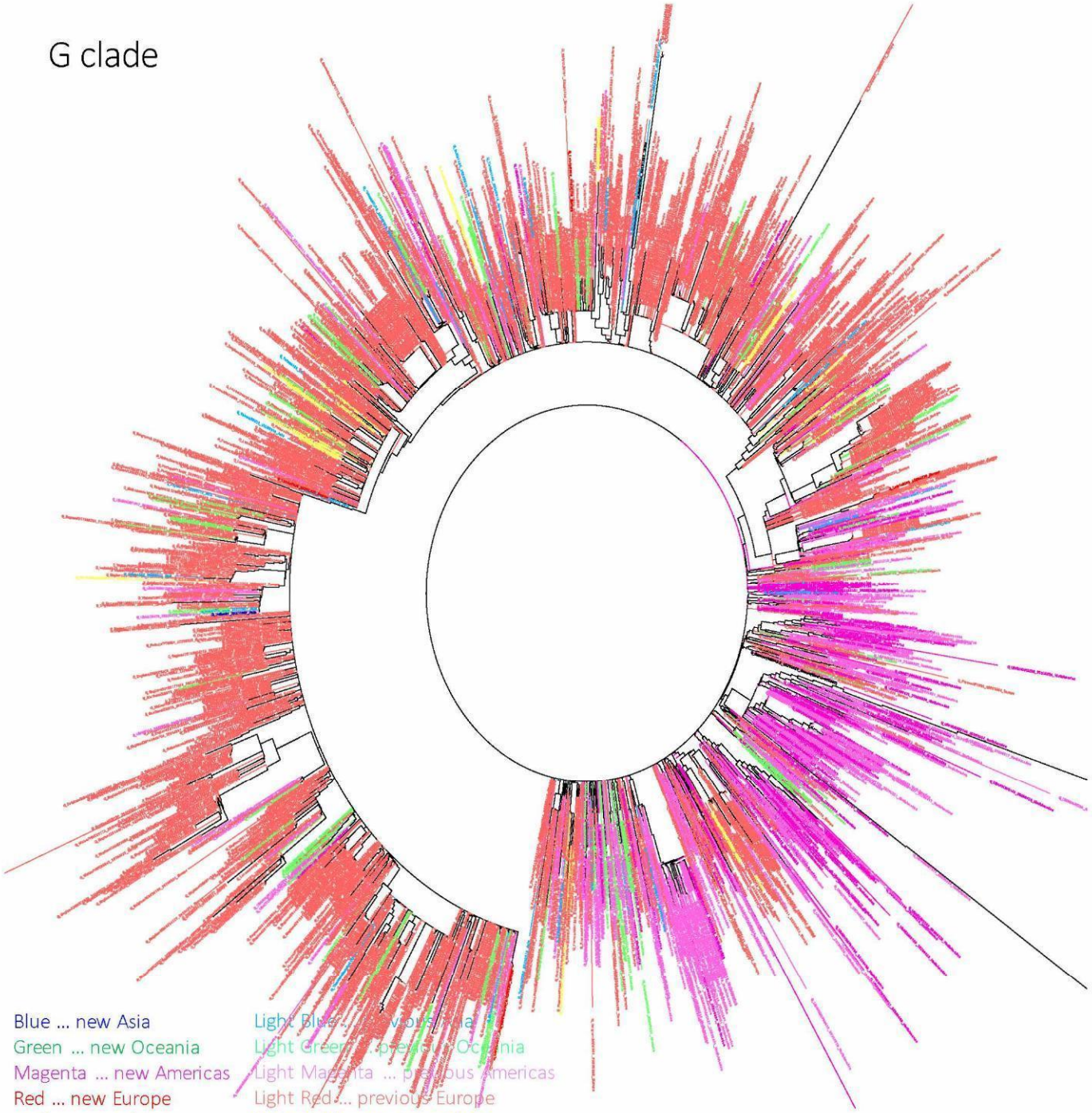
We gratefully acknowledge the Authors from Originating and Submitting laboratories of sequence data on which the analysis is based.

GISAID

by BII/GIS, A*STAR Singapore

G clade

Full genome trees of major subclades 2016979436



Notable changes:

G clade 4,987 (+137):
53 USA/WA, 36 USA/NY, 19 USA/ID, 11 USA/CT, 5 Latvia, 3 USA/MN, 3 USA/WI, 2 USA/OR, 2 USA/UN, 2 USA/IL, 1 India

Blue ... new Asia
Green ... new Oceania
Magenta ... new Americas
Red ... new Europe
Gold ... new Africa

Light Blue ... previous Asia
Light Green ... previous Oceania
Light Magenta ... previous Americas
Light Red ... previous Europe
Light Gold ... previous Africa

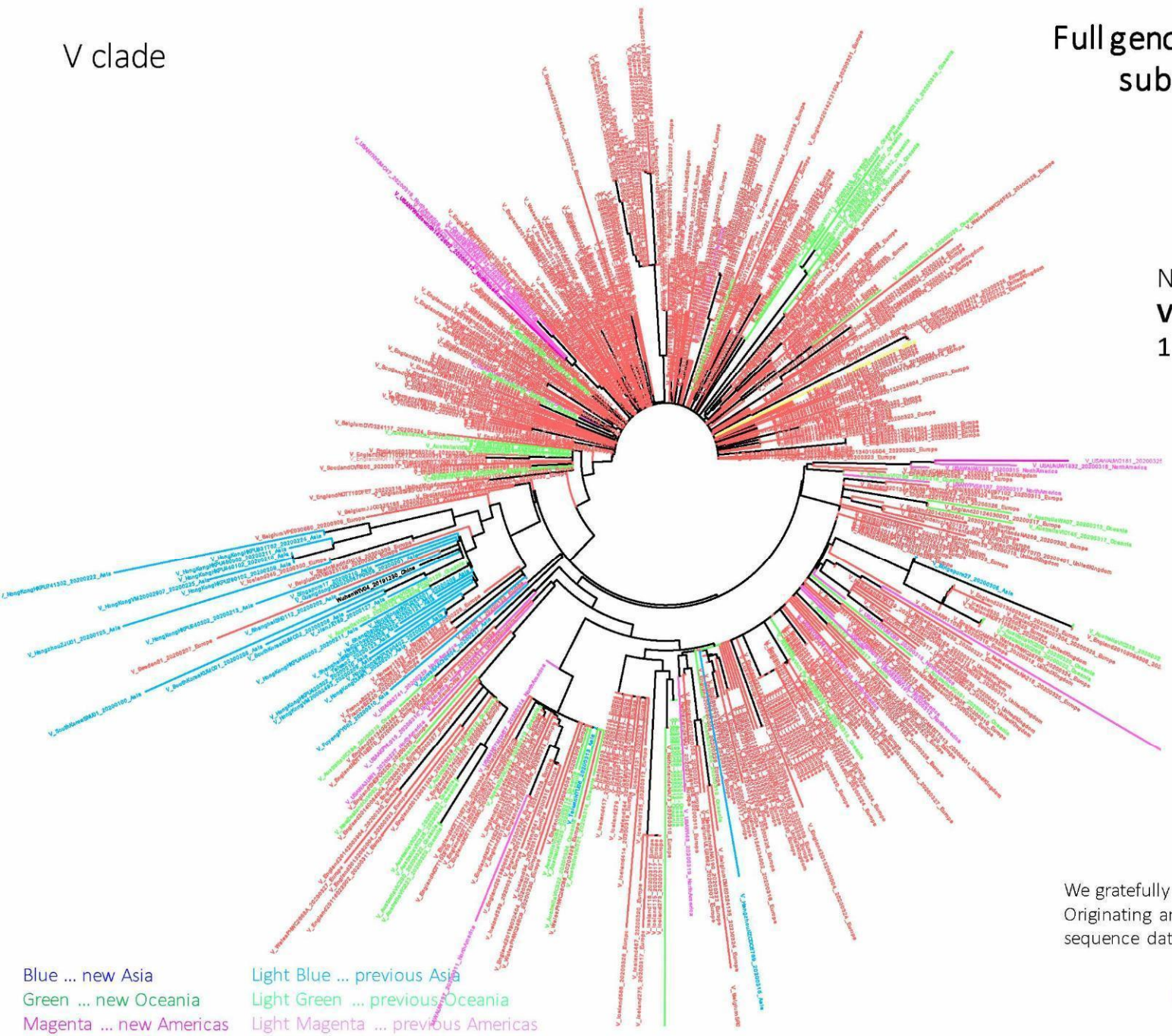
We gratefully acknowledge the Authors from Originating and Submitting laboratories of sequence data on which the analysis is based.

GISAID

by BII/GIS, A*STAR Singapore

V clade

Full genome trees of major subclades 2016979431 16



Notable changes:

V clade 1,204 (+2):

1 USA/WA, 1 USA/NY

Blue ... new Asia
Green ... new Oceania
Magenta ... new Americas
Red ... new Europe
Gold ... new Africa

Light Blue ... previous Asia
Light Green ... previous Oceania
Light Magenta ... previous Americas
Light Red ... previous Europe
Light Gold ... previous Africa

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

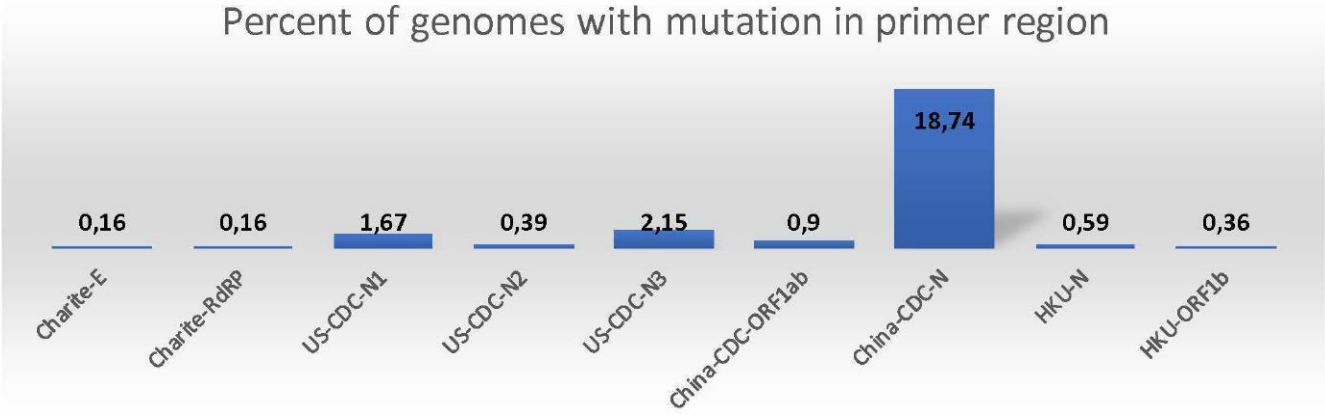
GISAID

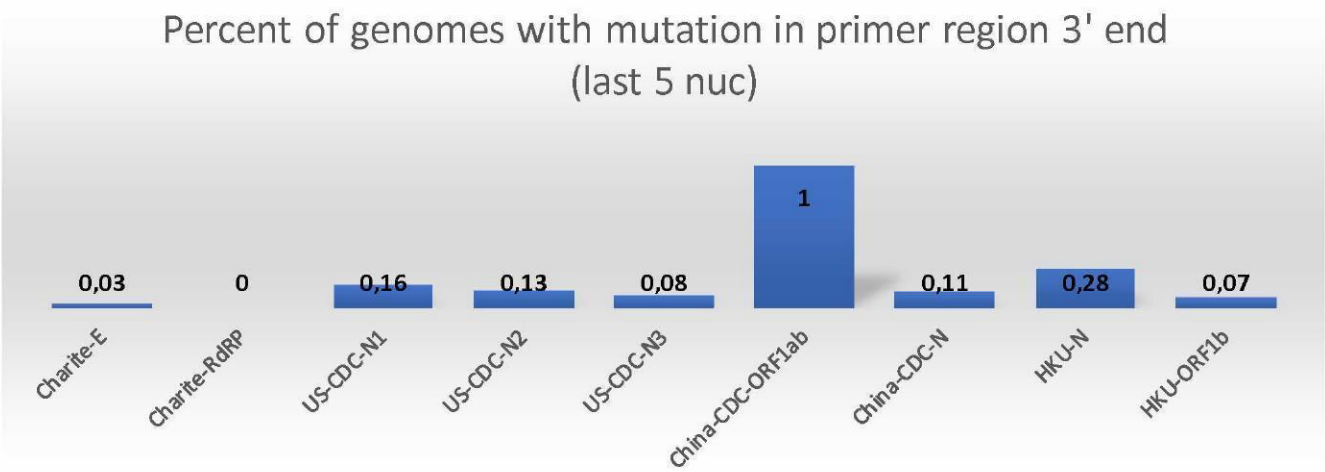
by BII/GIS, A*STAR Singapore

Common primer check for high quality genomes

2020-04-15 (updated ev1697943ys)

To reduce noise of random mutations ~6,100 available high quality genomes (out of 8,458) are considered here





This is a new simplified summary view of the percent of 6,100 high quality genomes (defined as <1% Ns and <0.05% unique mutations) with one or more mutations in either forward, probe or reverse primer region. This does not necessarily indicate a primer would not function but serves as a guide to variability of the targeted region. The second Figure shows the same but with mutations in 3' ends for the primer regions (defined as last 5 nucleotides of the primer sequence) which can affect sensitivity partially.

The results are obtained with a custom Perl script applied to results of BLASTN searches.

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

Sources of primer sequences:
https://www.who.int/docs/default-source/coronaviruse/protocol-v2-1.pdf
https://www.who.int/docs/default-source/coronaviruse/peiris-protocol-16-1-20.pdf
http://ivdc.chinacdc.cn/kyjz/202001/t20200121_211337.html
https://www.who.int/docs/default-source/coronaviruse/uscdcrt-pcr-panel-primer-probes.pdf

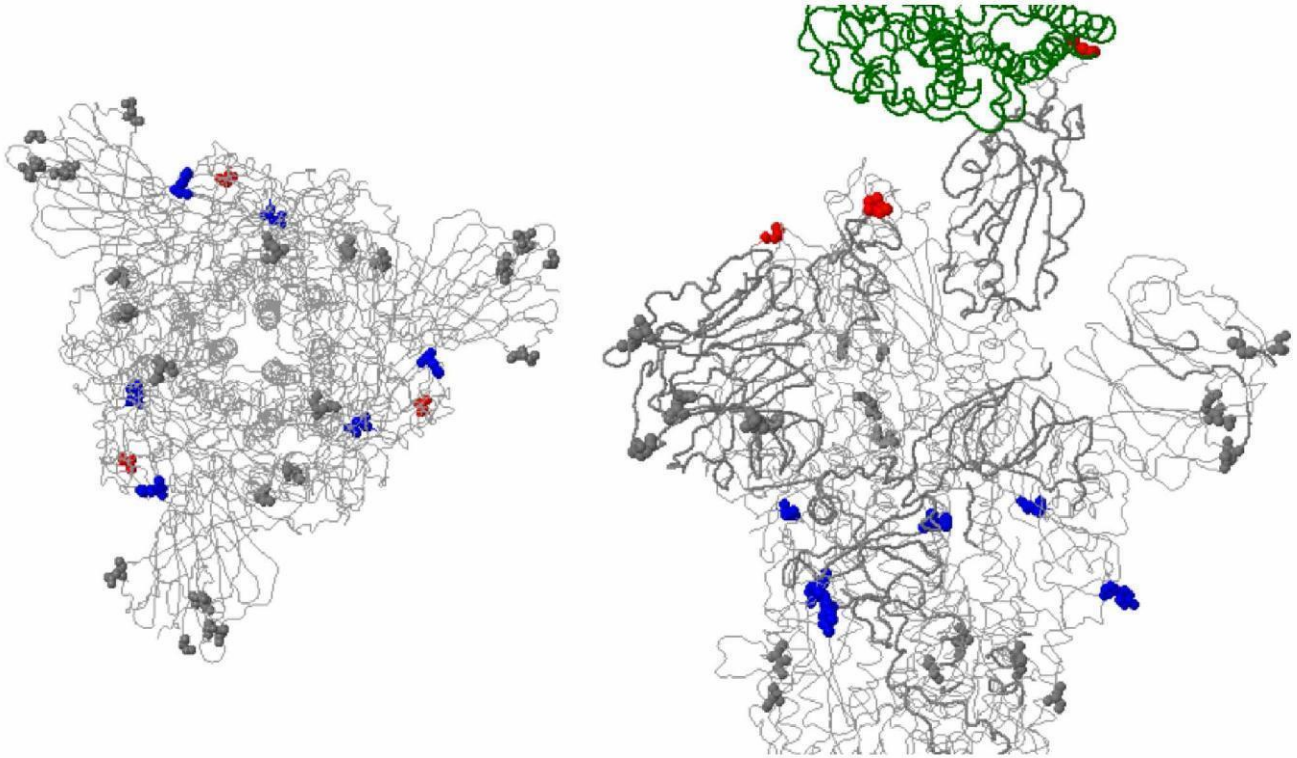
by BII/GIS, A*STAR Singapore

GISAID

Receptor binding surveillance for complete genomes 2020-04-16 1697943

New occurrence of previous receptor binding mutation V483A (3 new in USA/WA)

Total: 7 different rare variants near the binding interface not known to be linked to severity. **V483A** in 26 samples (23 USA/WA, 2 USA/UN, 1 USA/CT), **V483I** in 1 English sample, L455I together with F456V in one Brazilian sample, **G476S** in 18 samples (13 USA/WA, 2 USA/OR, 1 USA/ID, 1 USA/CT, 1 Belgium), **S494P** in 1 English sample and **N439K** in 1 Scottish sample.



Mutations in the spike glycoprotein for the 964 new complete genomes are shown here.

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

Green ... ACE2 human host receptor
Gray ... CoV spike glycoprotein trimer
Gray balls ... Spike glycoprotein variation occurring once (in EpiCoV)
Blue balls ... Spike glycoprotein variation occurring more than once (in EpiCoV)
Red balls ... Spike glycoprotein variation near host receptor
Yellow ... Insertion/deletion

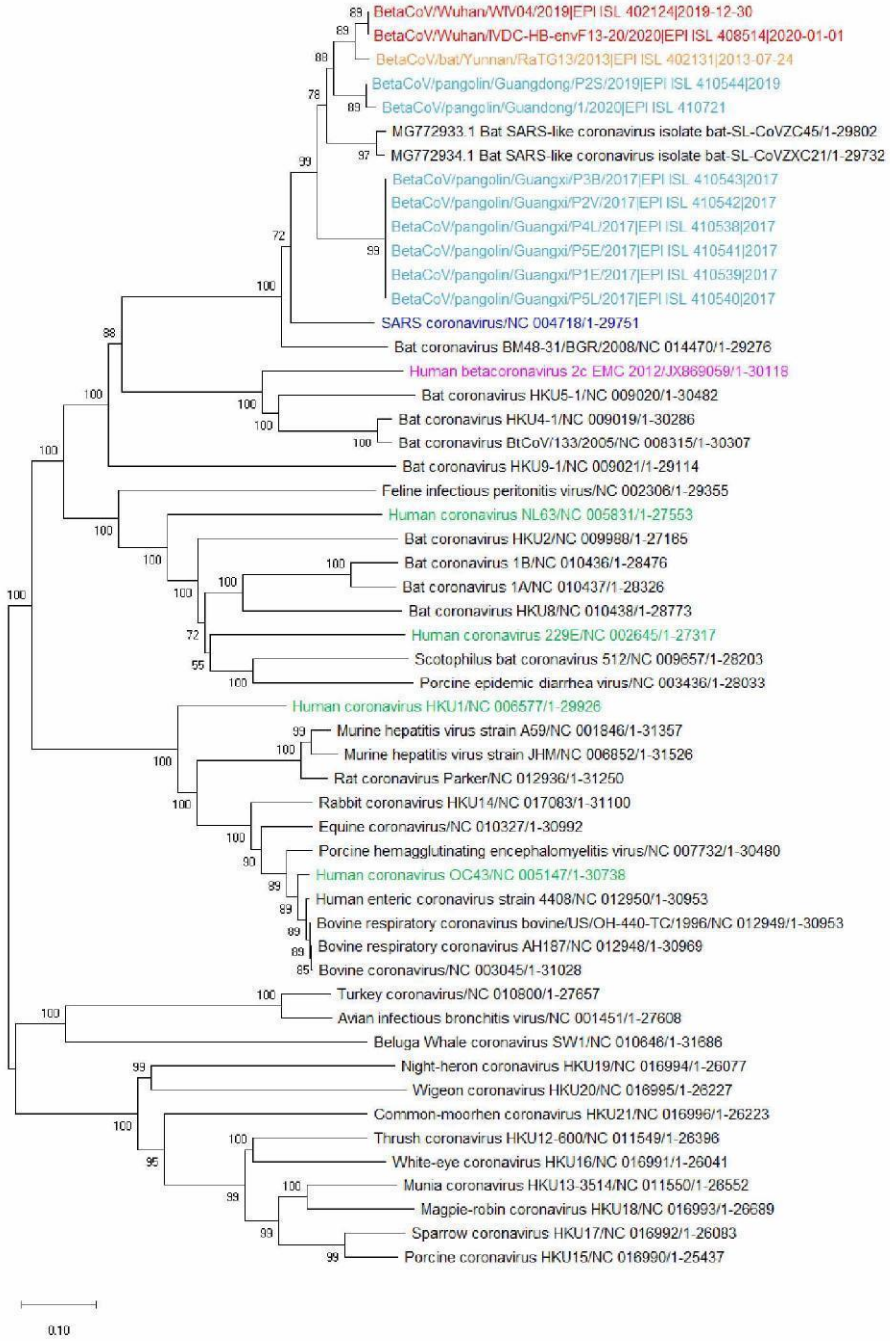
Equivalent positions have been studied for V483A and V483I in MERS (DOI: 10.1128/JVI.01381-18) and G476S, L455I, F456V, S494P and N439K in SARS (DOI: 10.1074/jbc.M111.325803 DOI: 10.1086/651022 DOI:10.1186/1743-422X-2-73) where they most often weakly reduced host receptor binding and altered antigenicity.

Numbering relative to start codon 21563 in hCoV-19/Wuhan/WIV04/2019

GISAID
by BII/GIS, A*STAR Singapore

Summary

First Characterization



Full genome tree of all CoV fam 1697943

- Nearest bat precursor RaTG13
- Nearest pangolin precursors from Guangdong
- Several pangolin-derived sequences part of recent family of related viruses

Genome identity to hCoV-19:

- 96% RaTG13 (nearest bat precursor)
- 90% Guangdong1/P2S (nearest pangolin precursor)
- 88% ZC45/ZXC21 bat precursor
- 80% SARS

Orange ... bat RaTG13
Red ... hCoV-19 2019-2020
Cyan ... pangolin CoV
Blue ... SARS CoV
Purple ... MERS CoV
Green ... common cold CoV

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

Phylogenetic tree of Wuhan CoV full genome sequences in context of representatives of all CoV families (whole genome Neighbor Joining, Maximum Composite Likelihood, uniform rates, 500 bootstrap, MegaX)

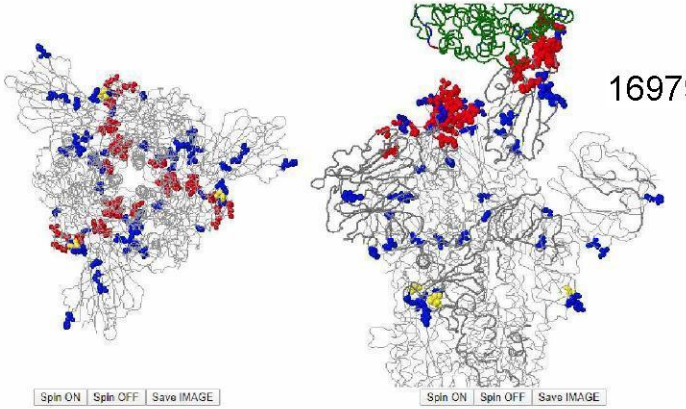
by BII/GIS, A*STAR Singapore

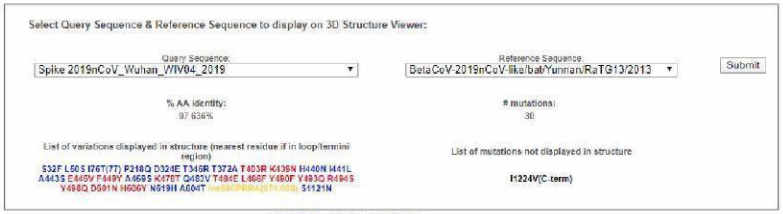
GISAID

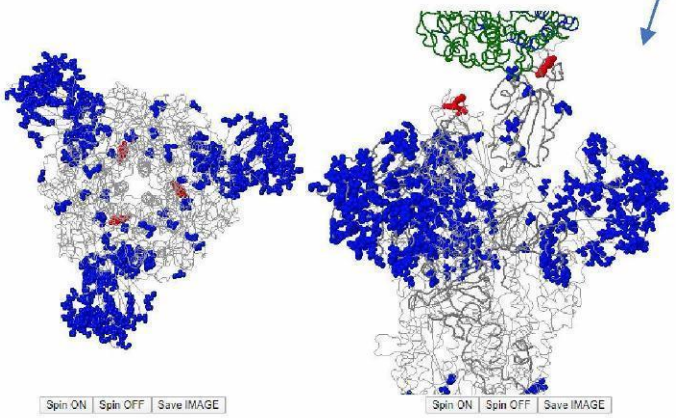
Spike host receptor changes for nearest bat and nearest pangolin sequences

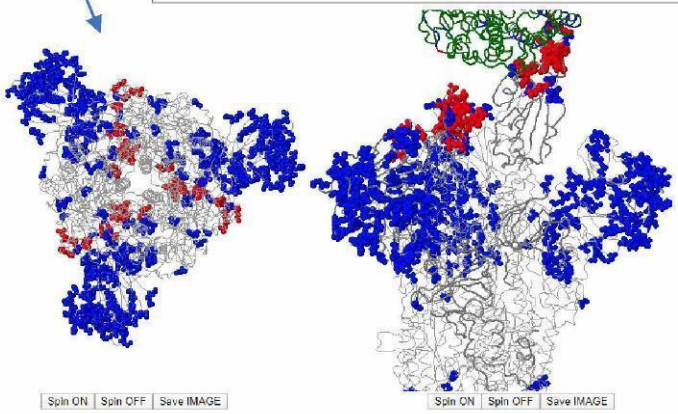
Strain 1	Strain 2	Spike overall identity	Interface mutations
Human Wuhan	Bat Yunnan	98%	13
Pangolin Guangdong	Bat Yunnan	90%	13
Pangolin Guangdong	Human Wuhan	91%	1

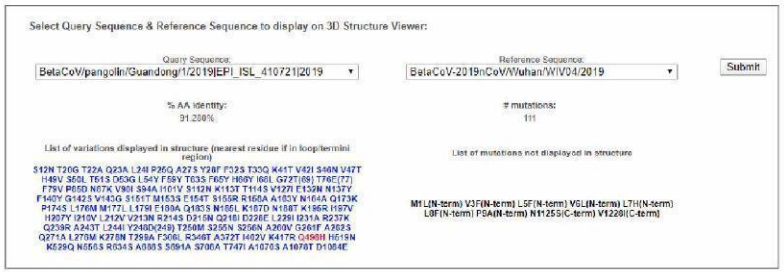
1697943

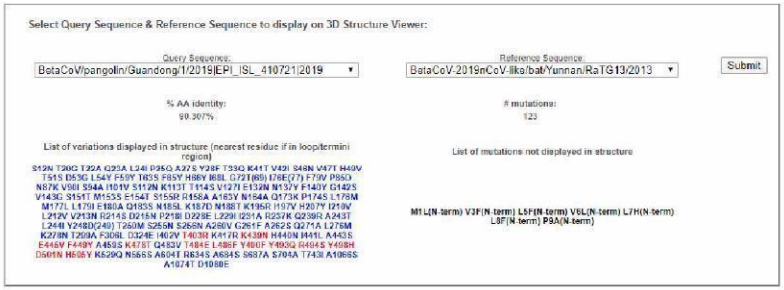










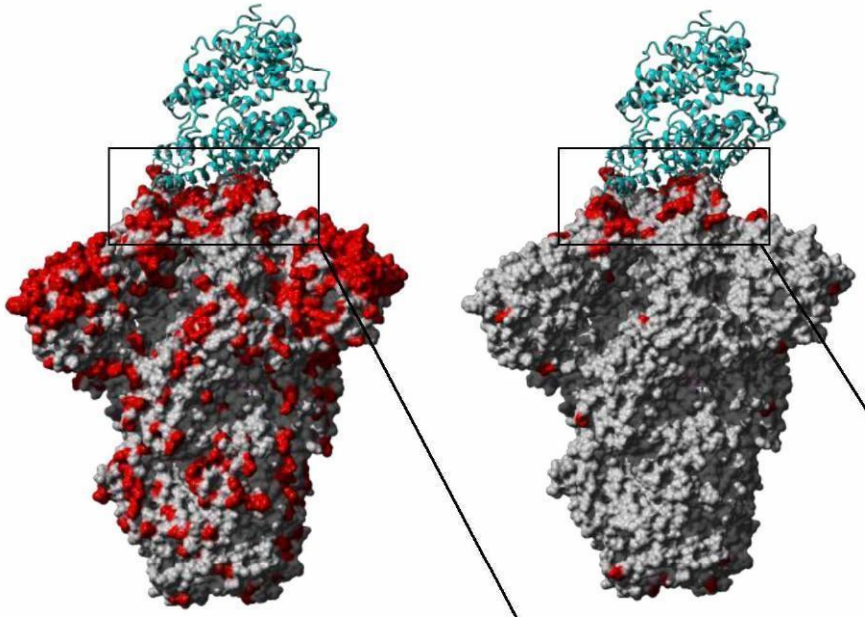


by BII, A*STAR Singapore

Host receptor binding site differences between SARS, bat precursor (RaTG13) and human outbreak hCoV-19

Additional Analysis for EPI_ISL_1697943 sequence from Zhengli Shi's lab

CAS Key Laboratory of Special Pathogens,
Wuhan Institute of Virology

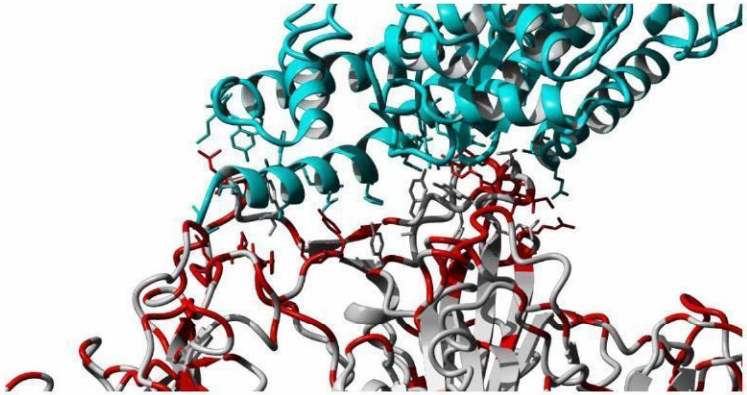


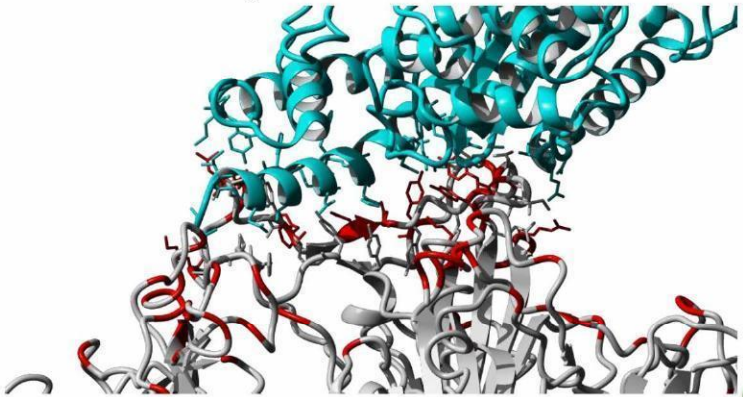
SARS vs hCoV-19

RaTG13 vs hCoV-19

Cyan … ACE2 human host receptor
Gray … CoV spike glycoprotein
Red … mutations between either SARS (left side) or bat precursor RaTG13 (right side) vs human outbreak WIV04 CoV

- Surface proteins are 76% and 98% identical, respectively
- Antigenic surface highly divergent compared to SARS
- Bat precursor differences in receptor binding interface indicative of changes that allowed host switch



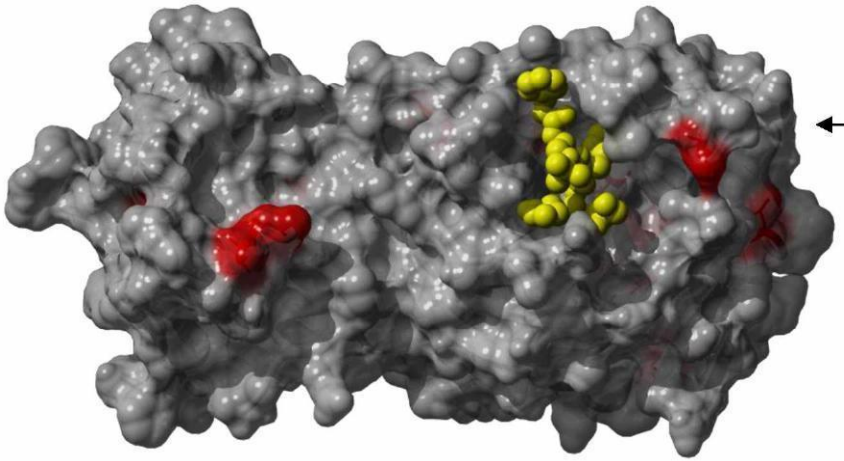


We gratefully acknowledge the Authors from Originating and Submitting laboratories of sequence data on which the analysis is based.

by BII, A*STAR Singapore

Potential drug targets highly conserved between hCoV-19 and S...[1697943]

- Both, the main protease and polymerase which are potential drug targets are highly conserved between hCoV-19 and SARS with 96% and 97% overall identity, respectively
- Inhibitors developed against the SARS-CoV main protease or polymerase have good potential to bind similarly to hCoV-19



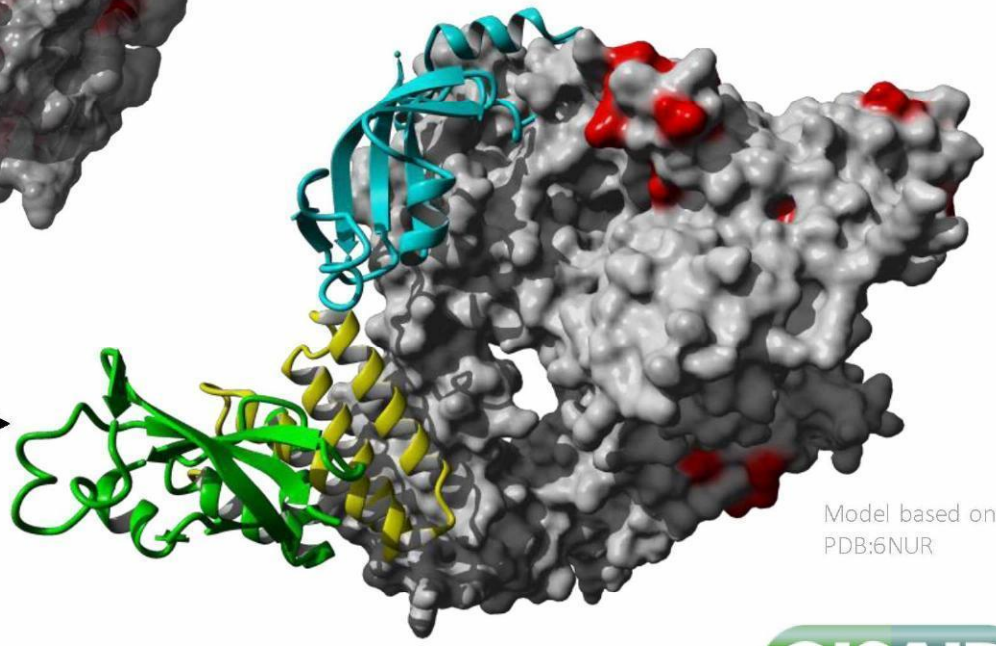
Model based on PDB:3TNT

Main protease hCoV-19 vs SARS

Red ... consensus differences (surface mutations)
Yellow ... substrate analogue/inhibitor

Polymerase hCoV-19 vs SARS

nsp12 (gray=identical, red=mutated) complex with nsp7 (yellow) and nsp8 (cyan, green)



Model based on PDB:6NUR

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

by BII, A*STAR Singapore

GISAID