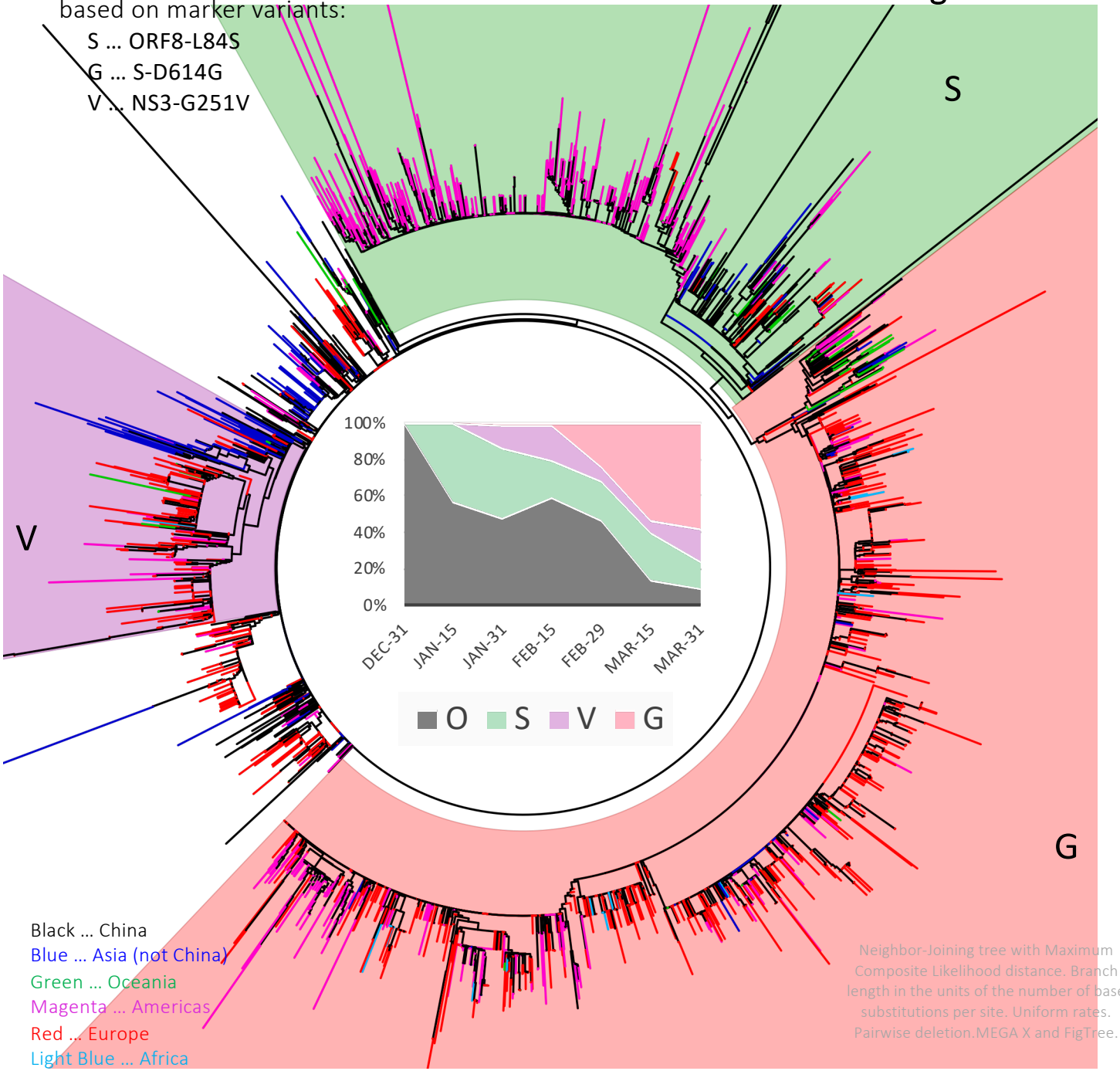
Latest update

2020-04-04 1500UTC

by BII/GIS, A*STAR Singapore

Full genome tree of all outbreak sequences
2020-04-04

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



Black … China
Blue … Asia (not China)
Green … Oceania
Magenta … Americas
Red … Europe
Light Blue … Africa

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.

Notable changes:

3,459 full genomes (+357)
(excluding low coverage, out of 3,795 entries)

S clade 763 (+80):
67 Australia/VIC, 9 Spain, 2 USA/VA, 2 USA/NY

G clade 1,576 (+186):
151 Australia/VIC, 16 Austria, 6 USA/NY, 4 USA/VA, 4 Spain, 3 USA/WI, 1 Belarus, 1 Latvia

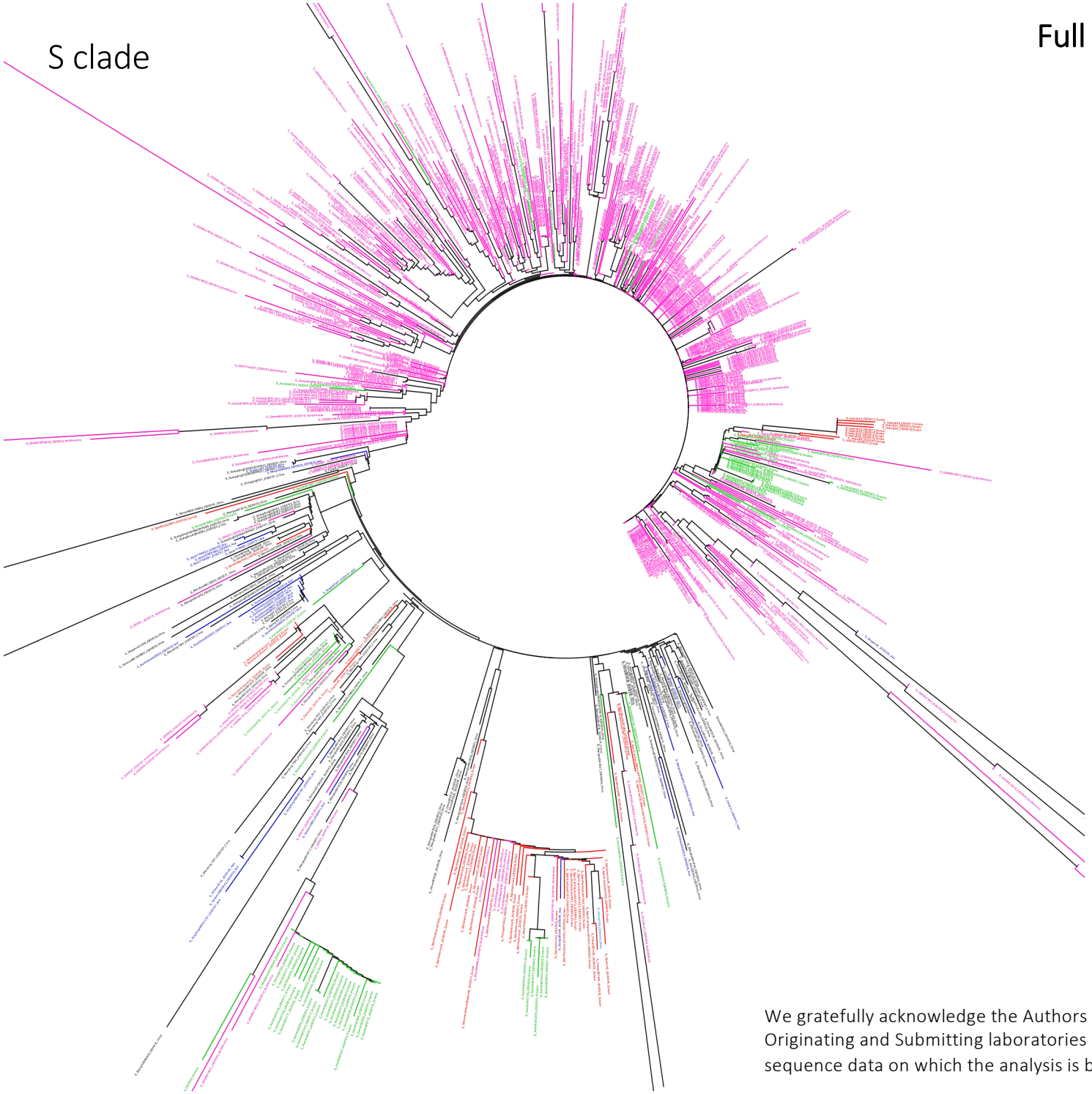
V clade 389 (+51):
48 Australia/VIC, 1 Austria, 1 Spain, 1 USA/NY

Other clades 731 (+40):
31 Australia/VIC, 4 Australia/NT0, 4 Austria, 1 Belarus

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 2020-04-04

Black … China
Blue … Asia (not China)
Green … Oceania
Magenta … Americas
Red … Europe
Light Blue … Africa

Notable changes:
S clade 763 (+80):
67 Australia/VIC, 9 Spain, 2 USA/VA, 2 USA/NY

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

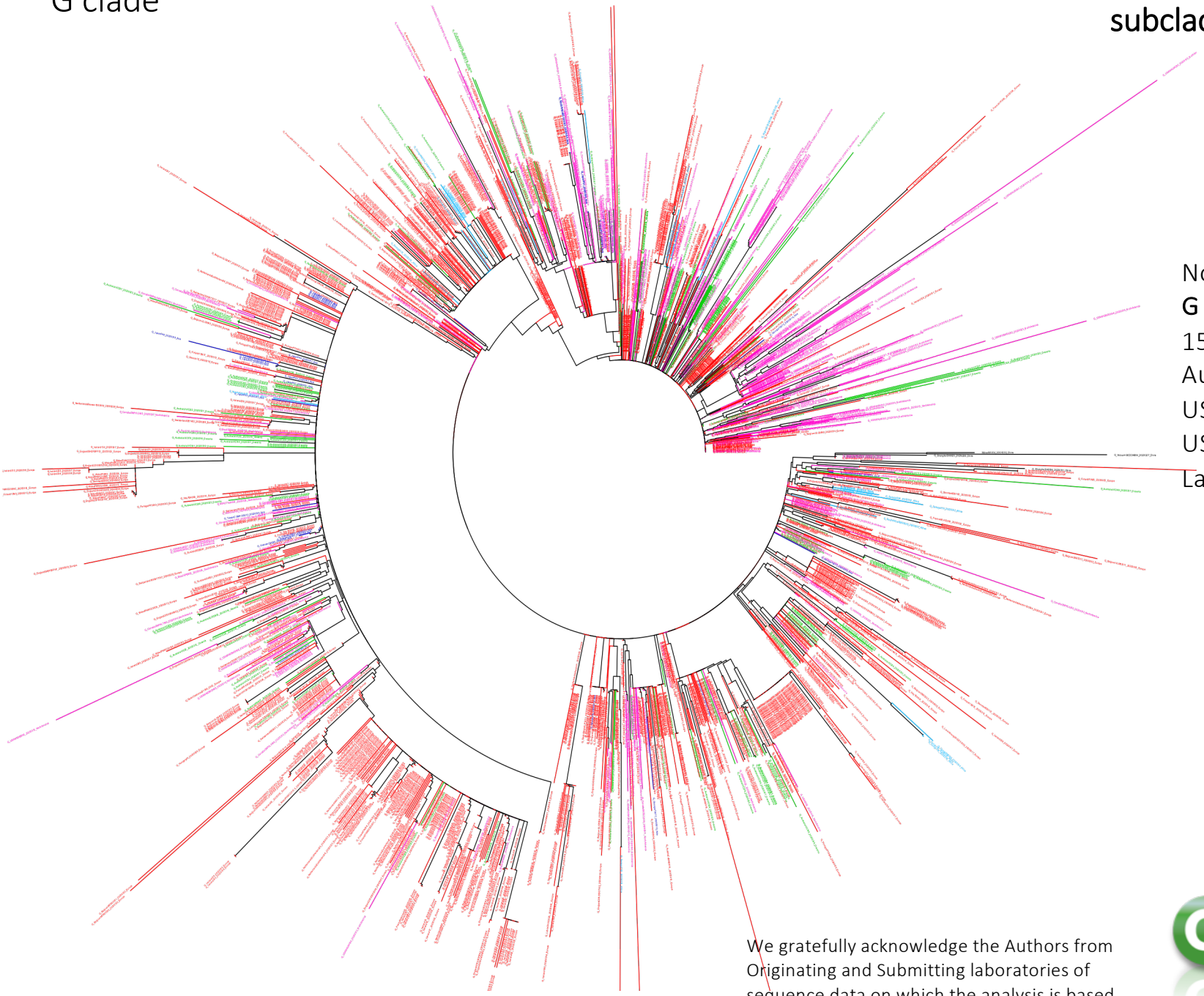
GISAID

by BII/GIS, A*STAR Singapore

G clade

Full genome trees of major subclades 2020-04-04

Black … China
Blue … Asia (not China)
Green … Oceania
Magenta … Americas
Red … Europe
Light Blue … Africa

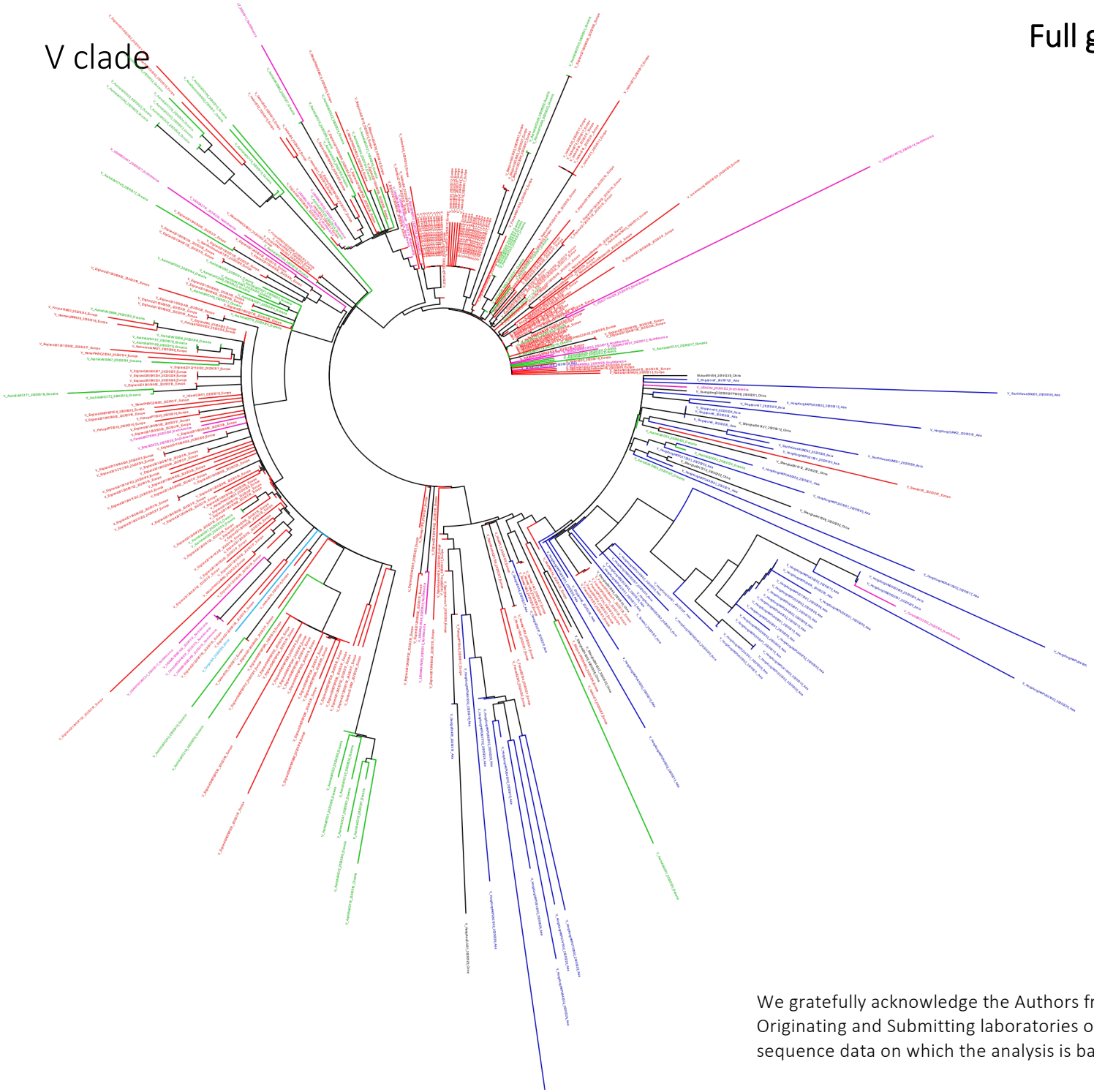


Notable changes:
G clade 1,576 (+186):
151 Australia/VIC, 16 Austria, 6 USA/NY, 4 USA/VA, 4 Spain, 3 USA/WI, 1 Belarus, 1 Latvia

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 2020-04-04



Black … China
Blue … Asia (not China)
Green … Oceania
Magenta … Americas
Red … Europe
Light Blue … Africa

Notable changes:
V clade 389 (+51):
48 Australia/VIC, 1 Austria, 1 Spain, 1 USA/NY

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

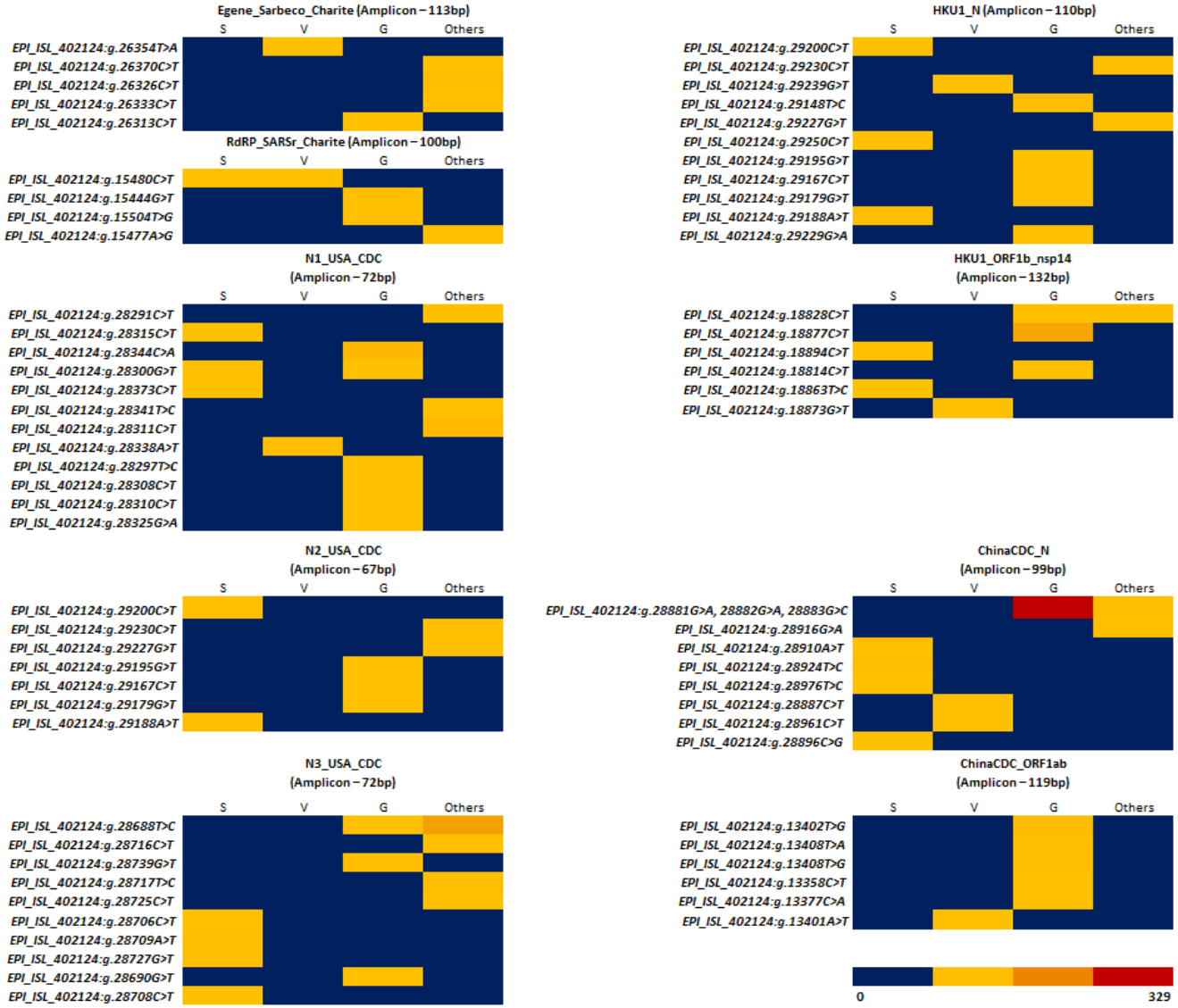
GISAID

by BII/GIS, A*STAR Singapore

Full genome nucleotide alignments for high quality genomes

2020-04-03 (updated every 3 days)

To reduce noise of random mutations all 2,402 available high quality genomes (out of 2,434) are considered here



The color codes in the heatmaps represent the number of genomes carrying a positional mismatch to different PCR primers and probes – Presented here is data analysed using high quality genomes on EpiCoV with PCR primers and probes (amplicons) from Charite, HKU, ChinaCDC and USACDC.

The colors Blue-Yellow-Red are used to represent a range of no genomes carrying a mismatch on amplicon positions to the highest number of genomes carrying a mismatch on amplicon positions.

All the genomes are aligned to the reference genome EPI_ISL_402124 and the amplicon co-ordinates are as per the reference genome.

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

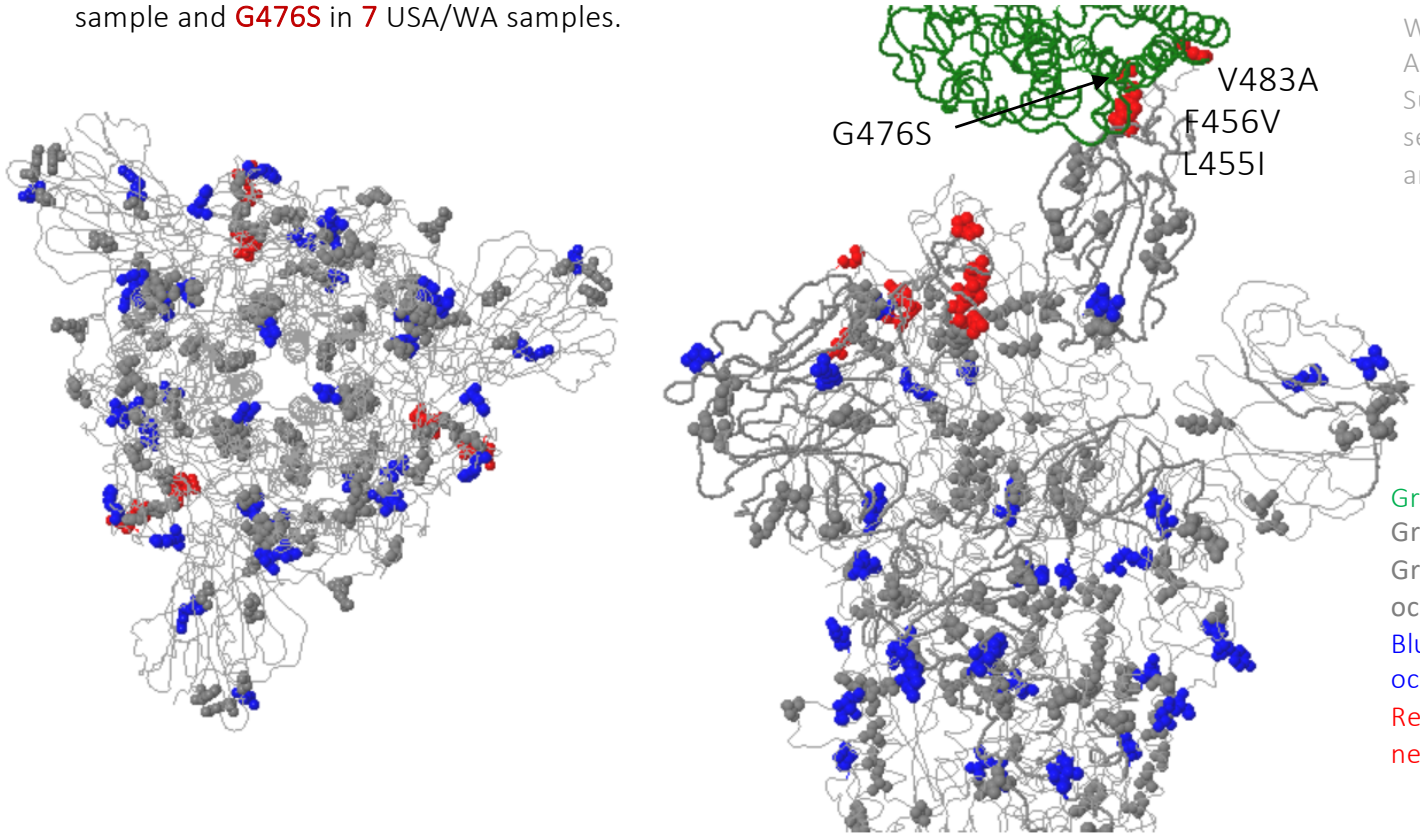
https://www.who.int/docs/default-source/coronaviruse/protocol-v2-1.pdf
https://www.who.int/docs/default-source/coronaviruse/peiris-protocol-16-1-20.pdf
http://ivdc.chinacdc.cn/kyjz/202001/t20200121_211337.html
https://www.who.int/docs/default-source/coronaviruse/uscdcrt-pcr-panel-primer-probes.pdf

by BII/GIS, A*STAR Singapore

GISAID

Receptor binding surveillance for high quality genomes 2020-04-04

4 different rare variants near the binding interface not known to be linked to severity. V483A in 16 USA/WA samples, L455I together with F456V in one Brazilian sample and G476S in 7 USA/WA samples.



To reduce noise of random mutations all 2,264 available high quality genomes (out of 2,919) are considered 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
Blue balls … Spike glycoprotein variation occurring more than once
Red balls … Spike glycoprotein variation near host receptor

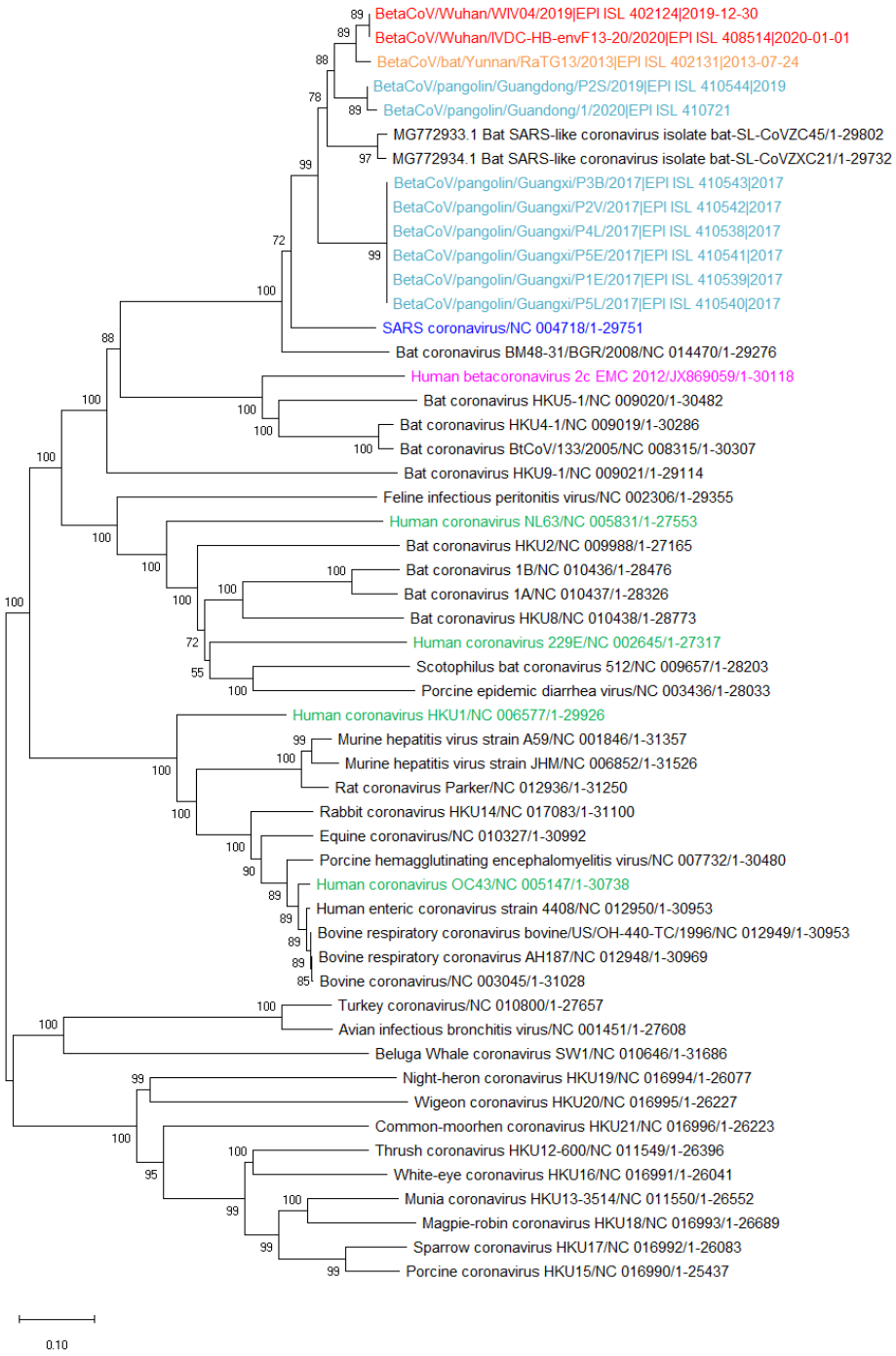
Equivalent positions have been studied for V483A in MERS (I529T, DOI: 10.1128/JVI.01381-18) and L455I, F456V and G476S in SARS (Y442F, DOI: 10.1074/jbc.M111.325803 and L443R, D463G DOI: 10.1086/651022) where they weakly reduced host receptor binding and altered antigenicity.

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

by BII/GIS, A*STAR Singapore

Summary

First Characterization



Full genome tree of all CoV families

- 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.

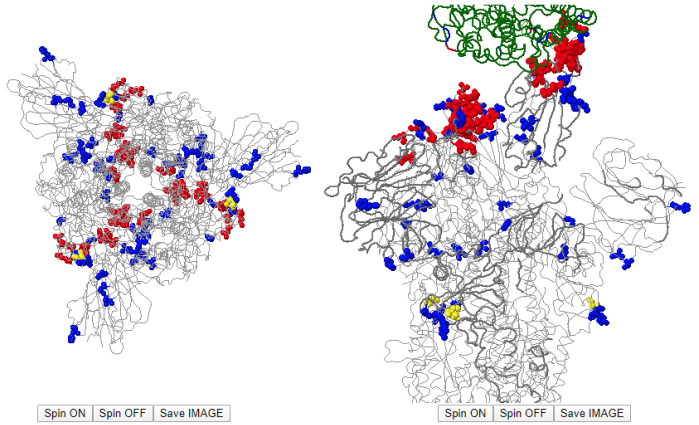
Phylogenetic tree of Wuhan CoV full genome sequences in context of representatives of all CoV families (whole genome Neighbor Joining, Maximum Composite Likelihood, uniform rates, 500 bootstrap, MegaX)

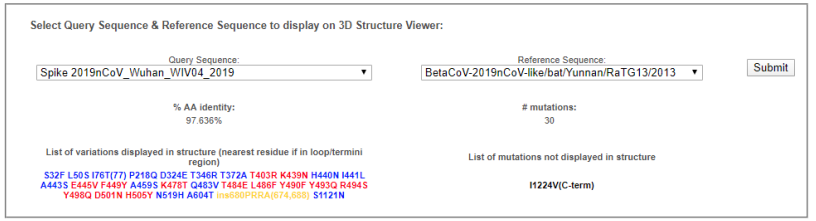
by BII/GIS, A*STAR Singapore

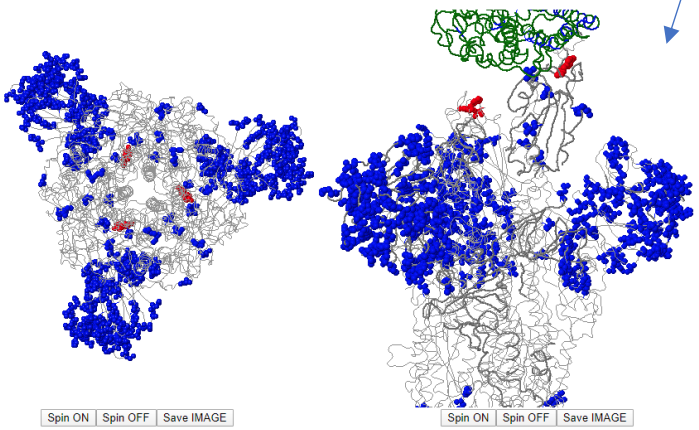
GISAID

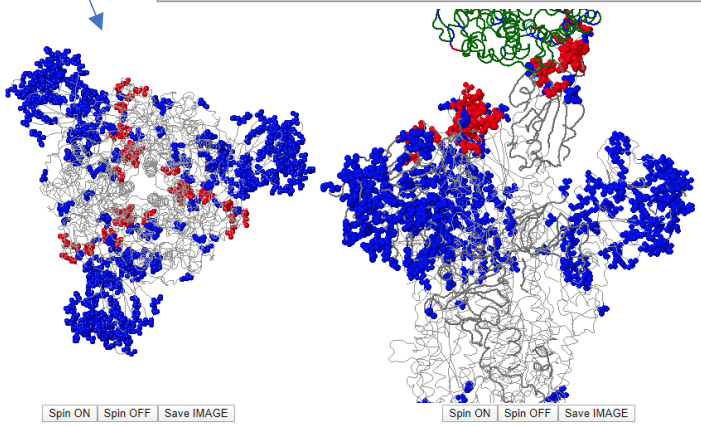
Spike host receptor changes for nearest bat and nearest pangolin sequences

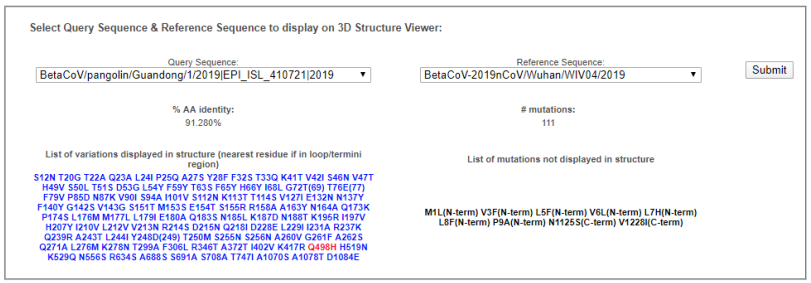
Strain 1	Strain 2	Spike overall identity	Interface mutations
Human Wuhan	Bat Yunnan	98%	13
Pangolin Guangdong	Bat Yunnan	90%	13
Pangolin Guangdong	Human Wuhan	91%	1

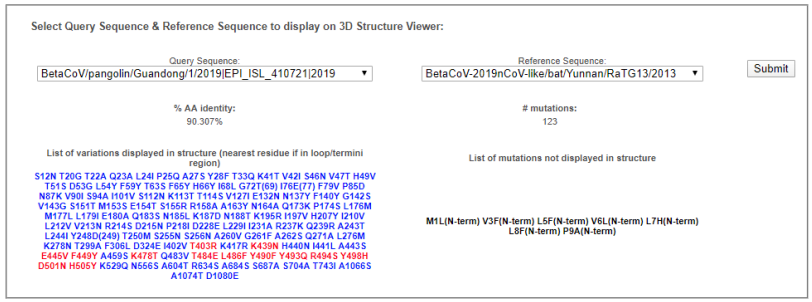








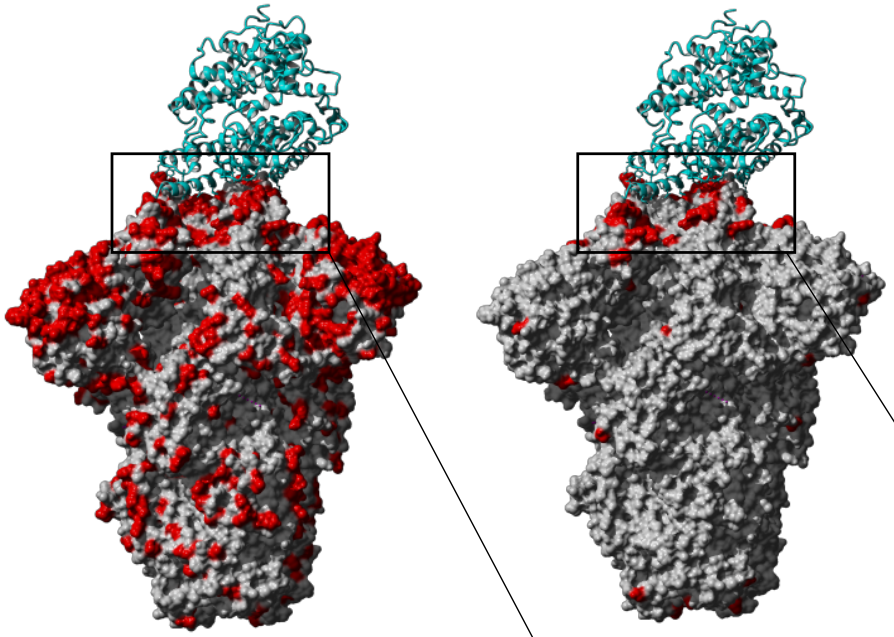




Host receptor binding site differences between SARS, bat precursor (RaTG13) and human outbreak hCoV-19

Additional Analysis for RaTG13 sequence from Zhengli Shi's lab

CAS Key Laboratory of Special Pathogens,
Wuhan Institute of Virology



SARS vs hCoV-19

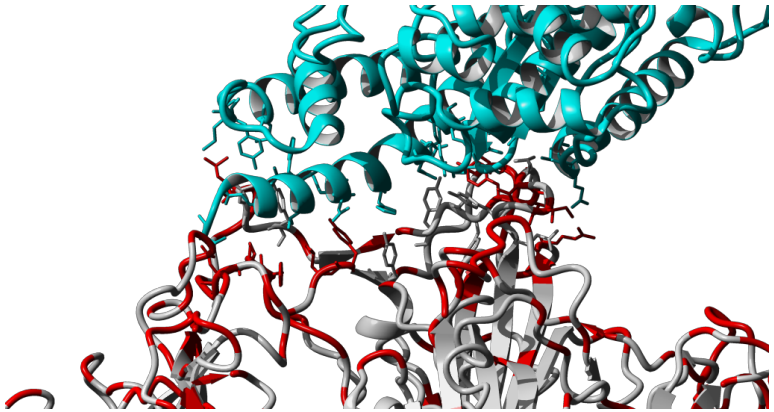
RaTG13 vs hCoV-19

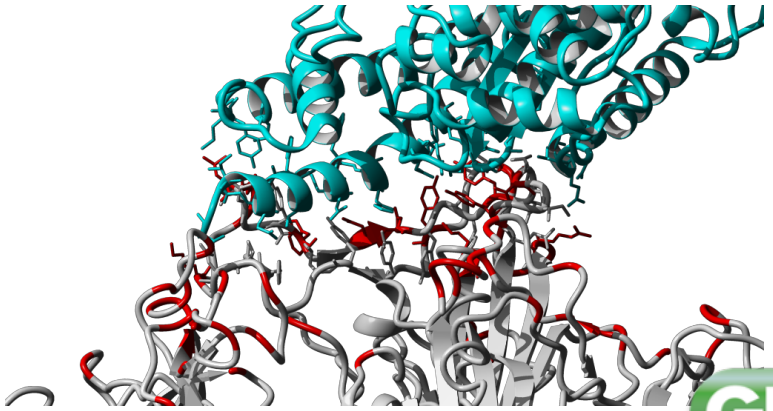
Cyan … ACE2 human host receptor

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





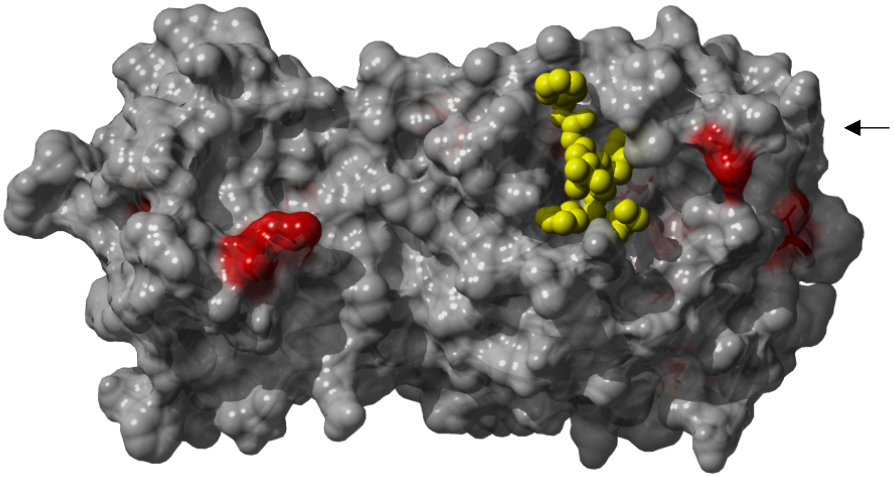
We gratefully acknowledge the Authors from Originating and Submitting laboratories of sequence data on which the analysis is based.

by BII, A*STAR Singapore

GISAID

Potential drug targets highly conserved between hCoV-19 and SARS

- 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



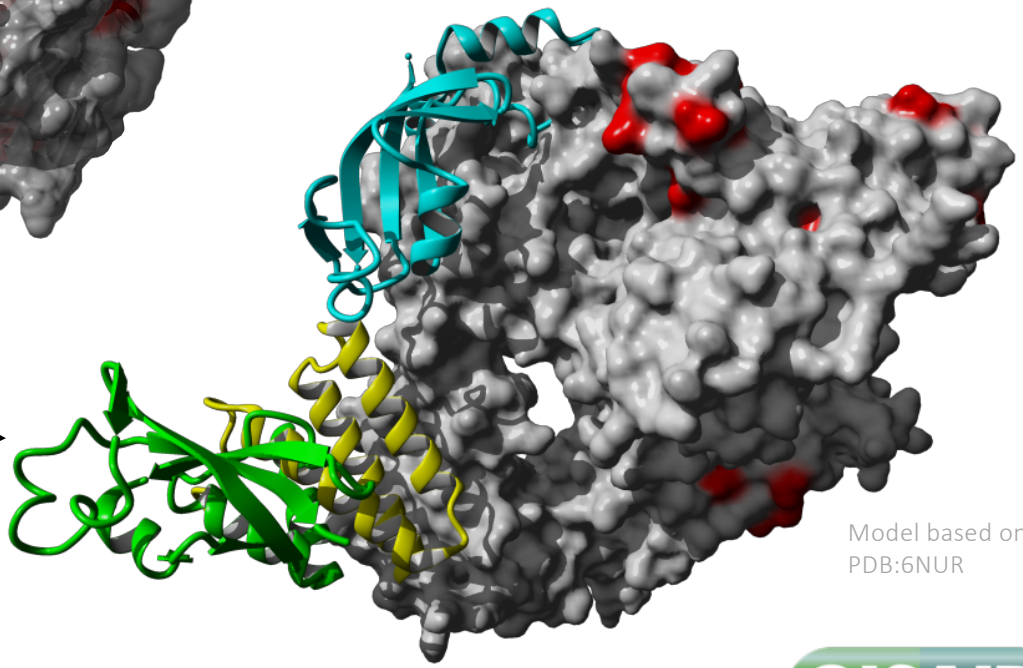
Main protease hCoV-19 vs SARS

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

Model based on PDB:3TNT

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