

Latest update

2020-04-16 1600UTC



Full genome tree derived from all outbreak sequences 2016-2019

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):

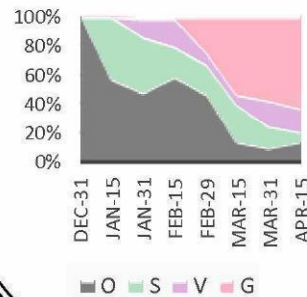
53 USA/WA, 36 USA/NY, 19 USA/ID, 11 USA/CT, 5 Latvia, 3 USA/MN, 3 USA/WI, 2 USA/OR, 2 USA/UN, 2 USA/IL, 1 India

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.

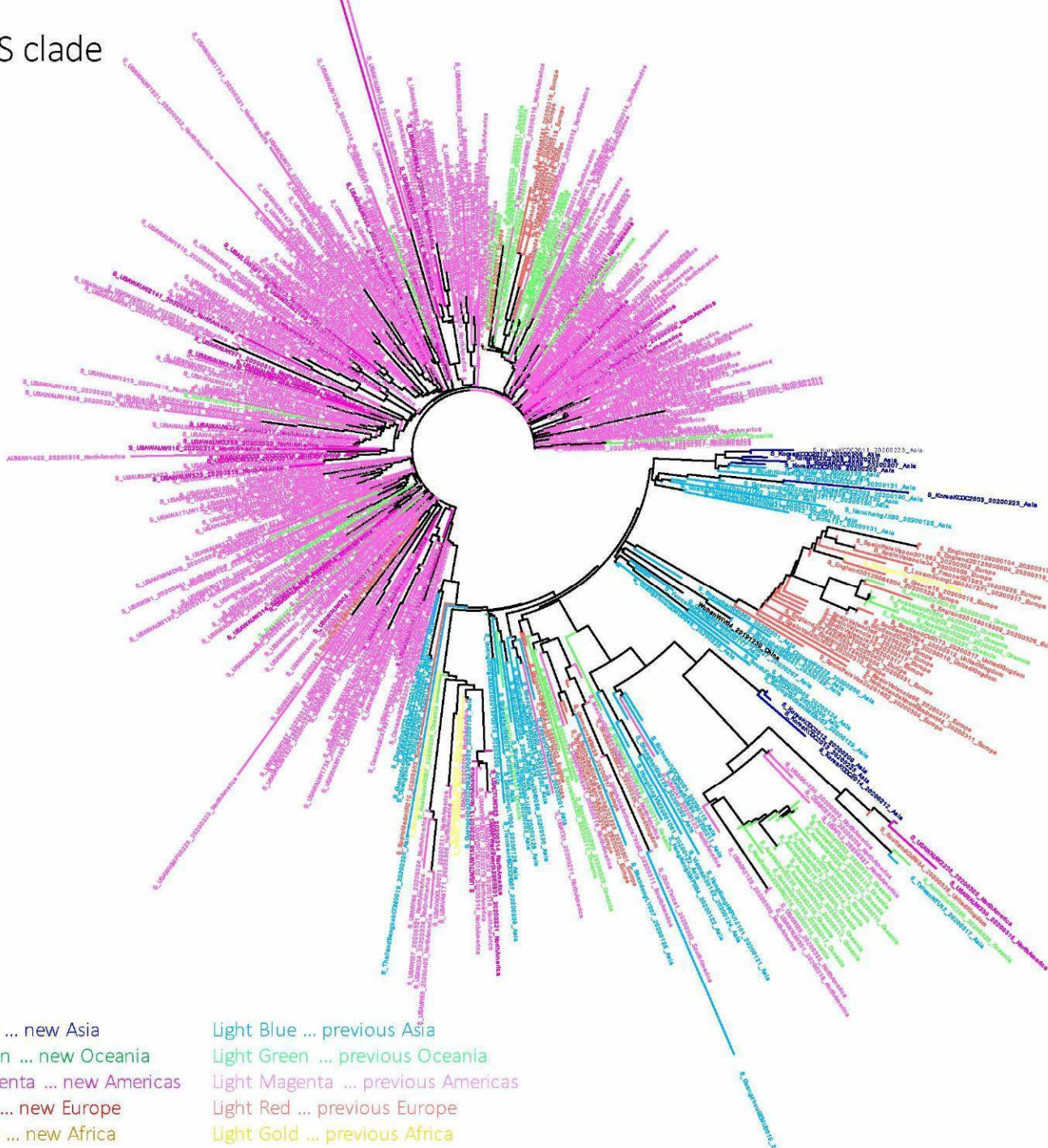
We gratefully acknowledge the Authors from Originating and Submitting laboratories of sequence data on which the analysis is based.



by BII/GIS, A*STAR Singapore

S clade

Full genome trees of major subclades 201697943_6



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.



by BII/GIS, A*STAR Singapore

G clade

Full genome trees of major subclades 201697943_6

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

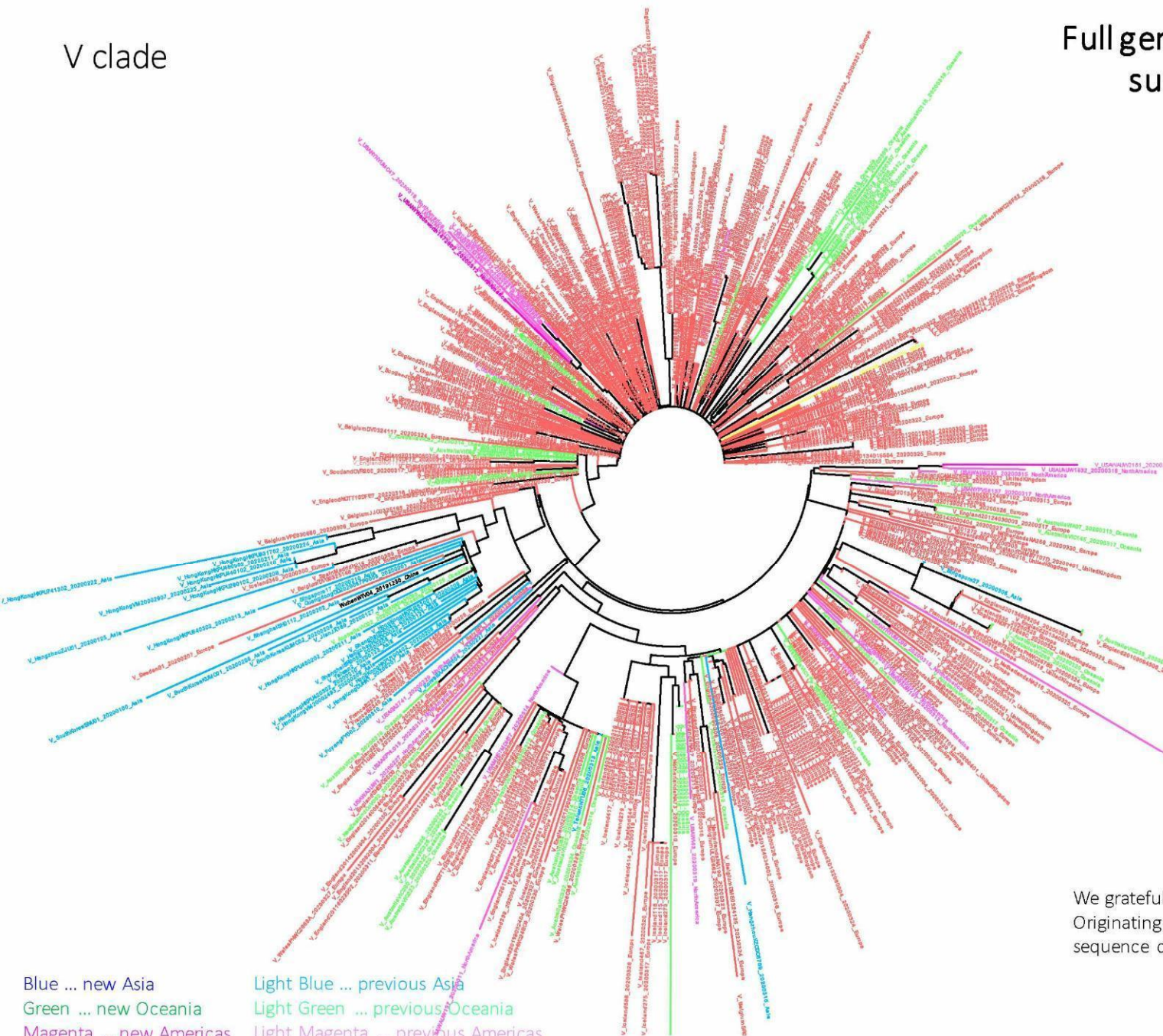
We gratefully acknowledge the Authors from Originating and Submitting laboratories of sequence data on which the analysis is based.



by BII/GIS, A*STAR Singapore

V clade

Full genome trees of major subclades 20190916



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.



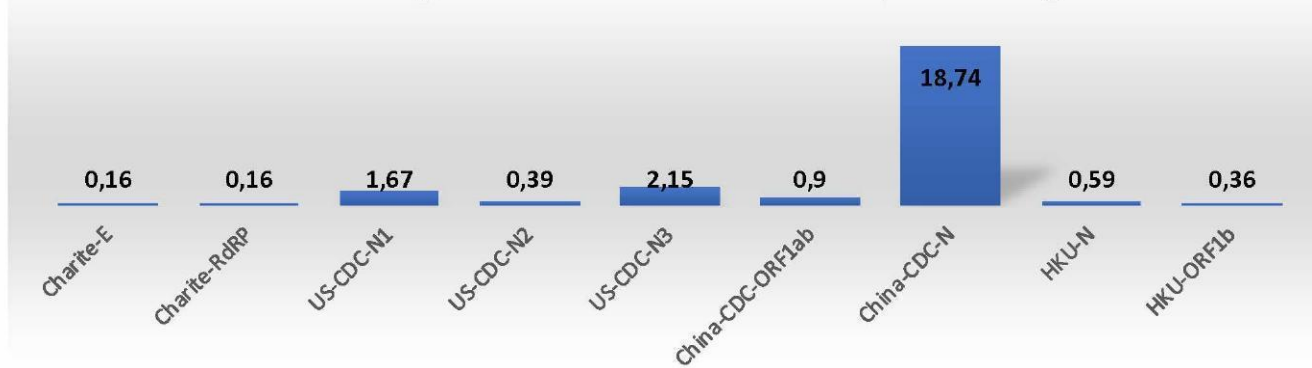
by BII/GIS, A*STAR Singapore

Common primer check for high quality genomes

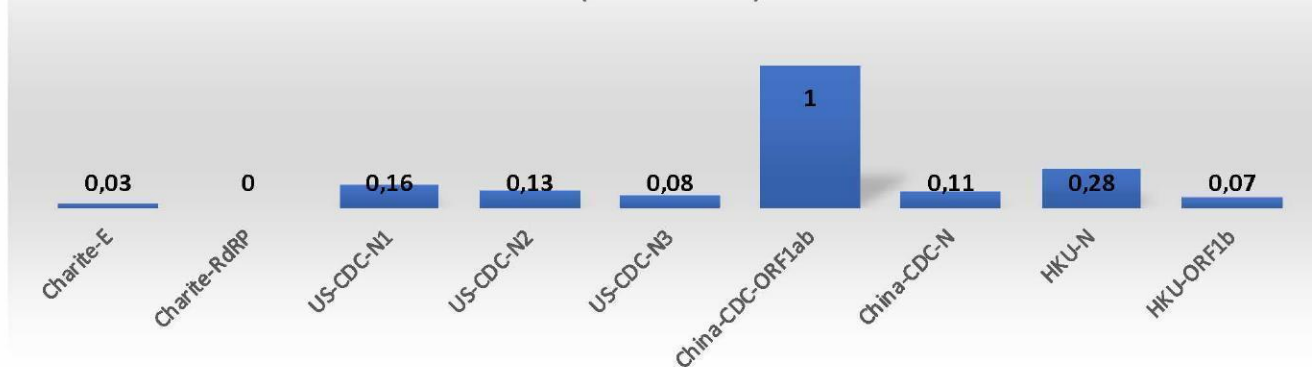
2020-04-15 (updated 1697943 days)

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

Percent of genomes with mutation in primer region



Percent of genomes with mutation in primer region 3' end (last 5 nuc)



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.

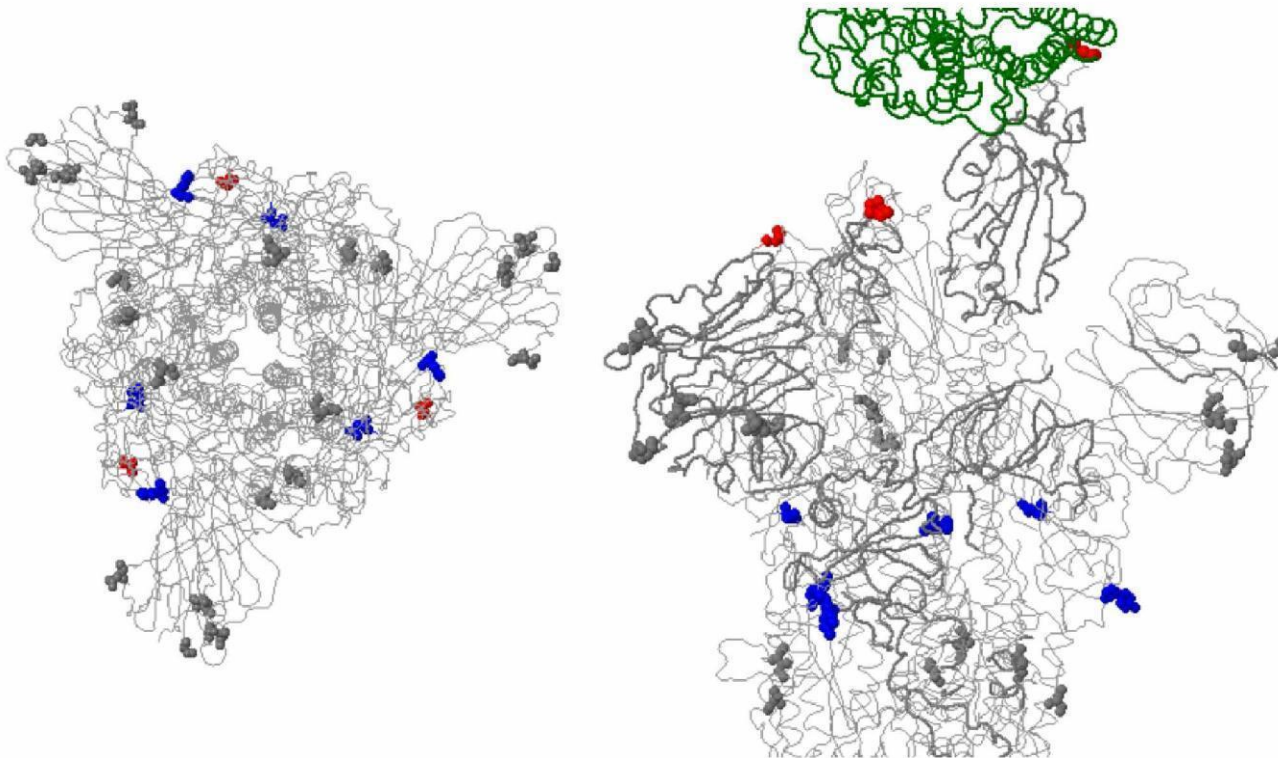
Sources of primer sequences:

- <https://www.who.int/docs/default-source/coronaviruse/protocolv2-1.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/uscdrt-pcr-panel-primer-probes.pdf>



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](https://doi.org/10.1128/JVI.01381-18)) and G476S, L455I, F456V, S494P and N439K in SARS (DOI: [10.1074/jbc.M111.325803](https://doi.org/10.1074/jbc.M111.325803) DOI: [10.1086/651022](https://doi.org/10.1086/651022) DOI: [10.1186/1743-422X-2-73](https://doi.org/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

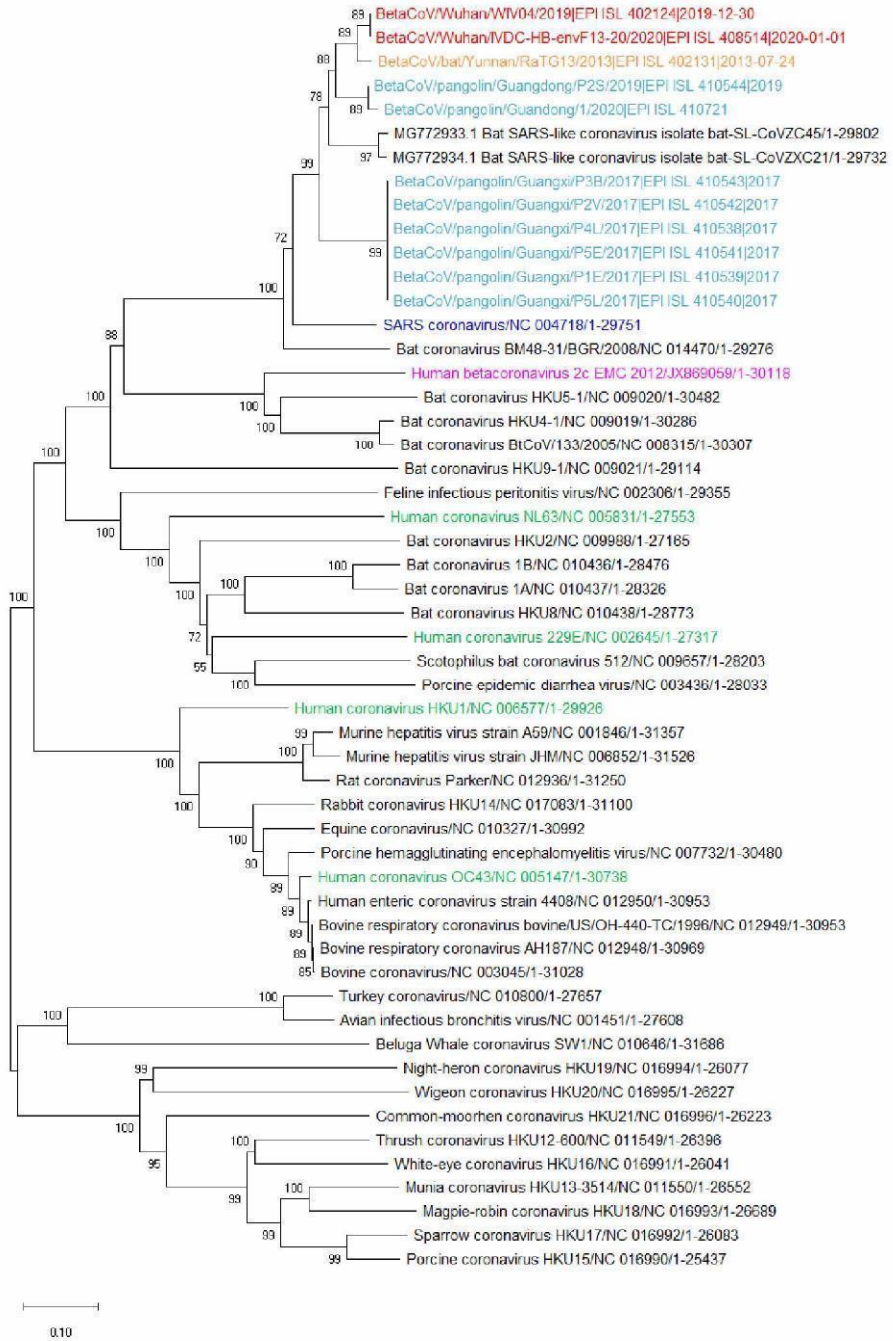


by BII/GIS, A*STAR Singapore

Summary

First Characterization

Full genome tree of all CoV far1697943



- 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.

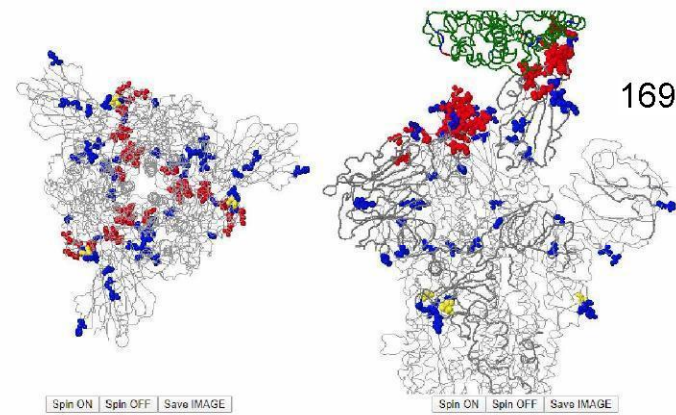
Phylogenetic tree of Wuhan CoV full genome sequences in context of representatives of all CoV families (whole genome NeighborJoining, Maximum Composite Likelihood, uniform rates, 500 bootstrap, MegaX)



Spike host receptor changes for nearest bat and nearest pangolin sequences

1697943

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



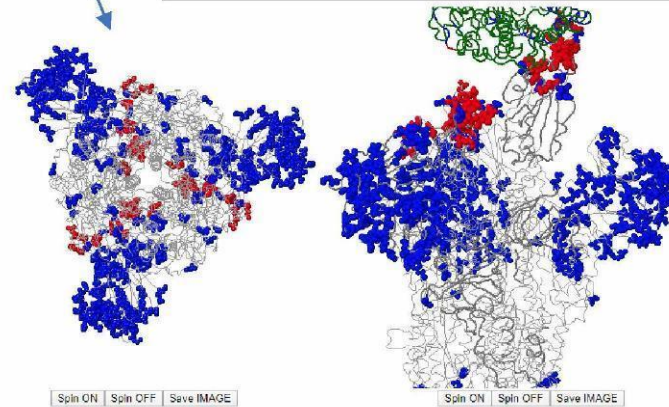
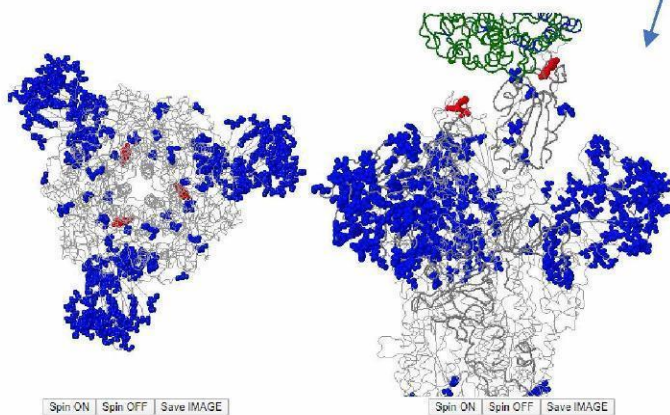
Select Query Sequence & Reference Sequence to display on 3D Structure Viewer:

Query Sequence: Spike 2019nCoV_Wuhan_WIV04_2019
 Reference Sequence: BetaCoV-2019nCoV-llka/bat/Yunnan/RaTG13/2013

% AA identity: 97.636%
 # mutations: 36

List of variations displayed in structure (nearest residue if in loop/terminal region):
 S32F L56 S 176T 777 P218Q D324E T345R T372A T403R I638N H640N H641L A643S E645Y F649Y A653S A675Y G681V T682C L683P V689F Y693Q R694S V698Q D691N H690Y N619H A604T

List of mutations not displayed in structure:
 H1224W(C-term)



Select Query Sequence & Reference Sequence to display on 3D Structure Viewer:

Query Sequence: BetaCoV/pangolin/Guangdong/1/2019/EPI_ISL_410721/2019
 Reference Sequence: BetaCoV-2019nCoV/Wuhan/WIV04/2019

% AA identity: 91.208%
 # mutations: 111

List of variations displayed in structure (nearest residue if in loop/terminal region):
 S12N T20G T22A Q23A L24I P26Q A27 S Y28F F32S T33Q K41T V43I S48N V47T H48V S50L T61S Q62G L64Y F69Y T635 F69Y H69Y I60L G72T(I68) T76E(77) F78V R650 N676 V691 S664 H610N S110N K110T T144S V122I E153W N157Y F140Y Q142S V143Q S116T M153S C154T S159R R165A A165V N166A Q173K P174S L176M M177L L179I E180A Q183S N186L K187D N188T K196R I197V H197Y I210V L219Y V238N P244S Q218I D222E L236I Q231A Q237K Q239R A243T L244I Y249D(Q249) T266M S265N S266N A269V Q281F A282S Q271A L270M K278N T268A F266L R248T A322T H629V K417R C456H H614M K629Q N656S R634S A693S S691A S708A T747I A1076S A1078T D1064E

List of mutations not displayed in structure:
 M1(LN-term) V3(FIN-term) L5(FIN-term) V5(LIN-term) L7(HN-term) L3(FIN-term) P3(AIN-term) N1125(SIC-term) V1228(I-C-term)

Select Query Sequence & Reference Sequence to display on 3D Structure Viewer:

Query Sequence: BetaCoV/pangolin/Guangdong/1/2019/EPI_ISL_410721/2019
 Reference Sequence: BetaCoV-2019nCoV-llka/bat/Yunnan/RaTG13/2013

% AA identity: 89.307%
 # mutations: 123

List of variations displayed in structure (nearest residue if in loop/terminal region):
 S12N T20G T22A Q23A L24I P26Q A27 S Y28F T33Q K41T V43I S48N V47T H48V I515 D53G L54Y P55Y T635 F69Y H69Y I60L G72T(I68) I76E(77) F78V P850 N676 V691 S664 H610N S110N K110T T144S V122I E153W N157Y F140Y Q142S V143Q S116T M153S E154I S159R R165A A165V N166A Q173K P174S L176M M177L L179I E180A Q183S N186L K187D N188T K196R I197V H197Y I210V L219Y V238N P244S Q218I D222E L236I Q231A Q237K Q239R A243T L244I Y249D(Q249) T266M S265N S266N A269V G281F A282S Q271A L270M K278N T268A F266L R248T A322T H629V K417R C456H H614M K629Q N656Y N629Q N656S A691E I654S A654S S691A S708A T743A A1065S A1074T D1068E

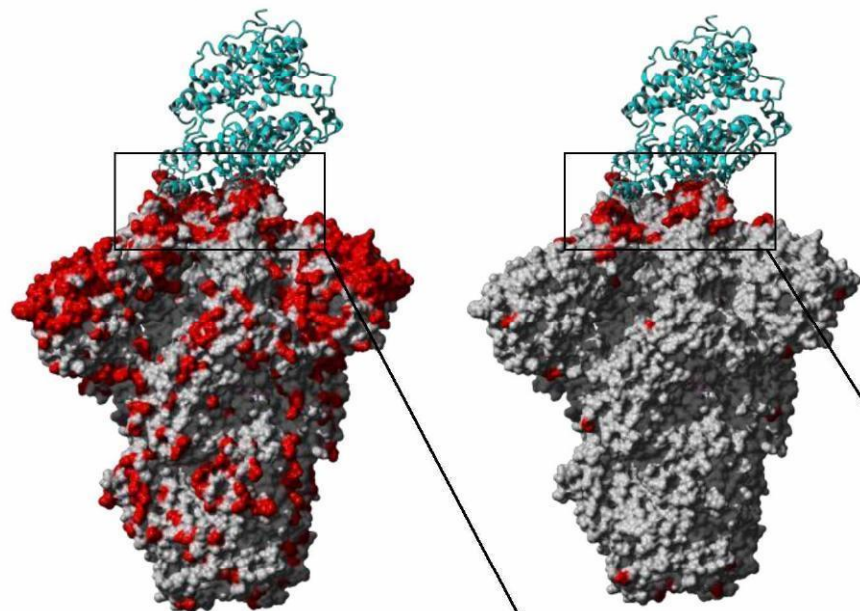
List of mutations not displayed in structure:
 M1(LN-term) V3(FIN-term) L5(FIN-term) V5(LIN-term) L7(HN-term) L3(FIN-term) P3(AIN-term)



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

Additional Analysis for F₁₆₉₇₉₄₃ sequence from Zhengli Shi's lab

CAS Key Laboratory of Special Pathogens,
Wuhan Institute of Virology



Cyan ... ACE2 human host receptor

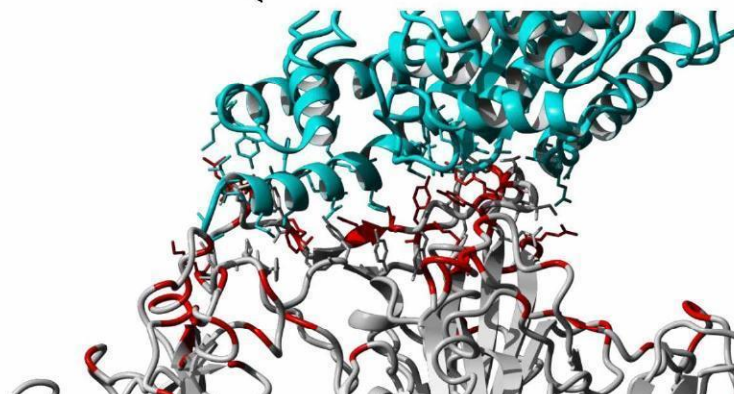
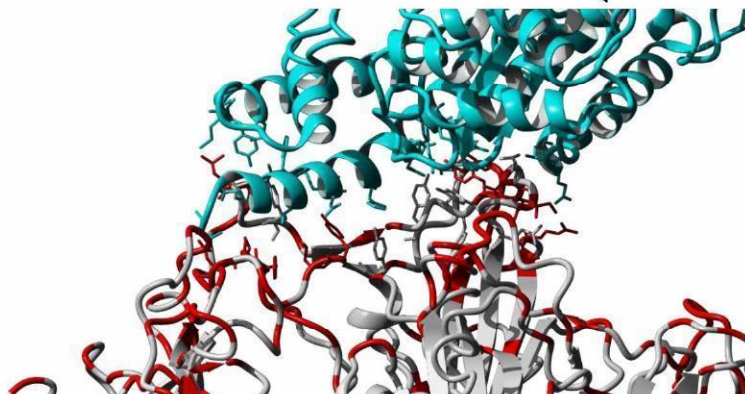
Gray ... CoV spike glycoprotein

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

SARS vs hCoV-19

RaTG13 vs hCoV-19



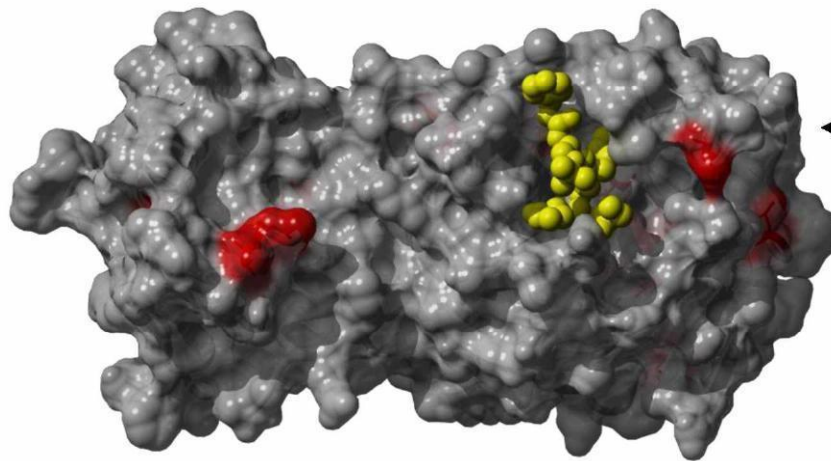
We gratefully acknowledge the Authors from Originating and Submitting laboratories of sequence data on which the analysis is based.

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



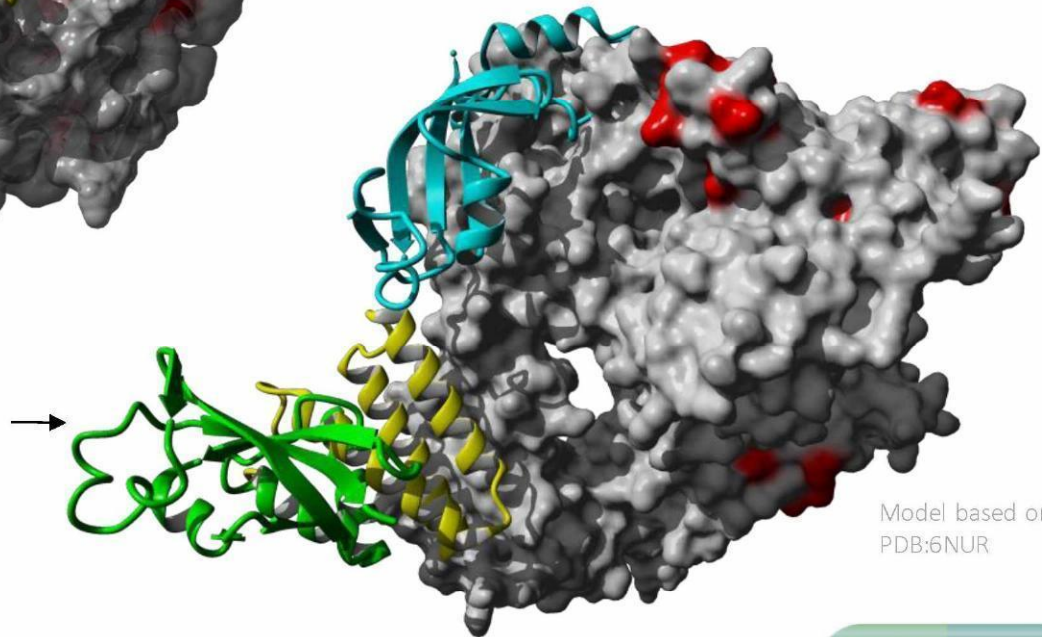
Model based on PDB:3TNT

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

