

# Latest update

2020-04-16 1600UTC

by BII/GIS, A\*STAR Singapore



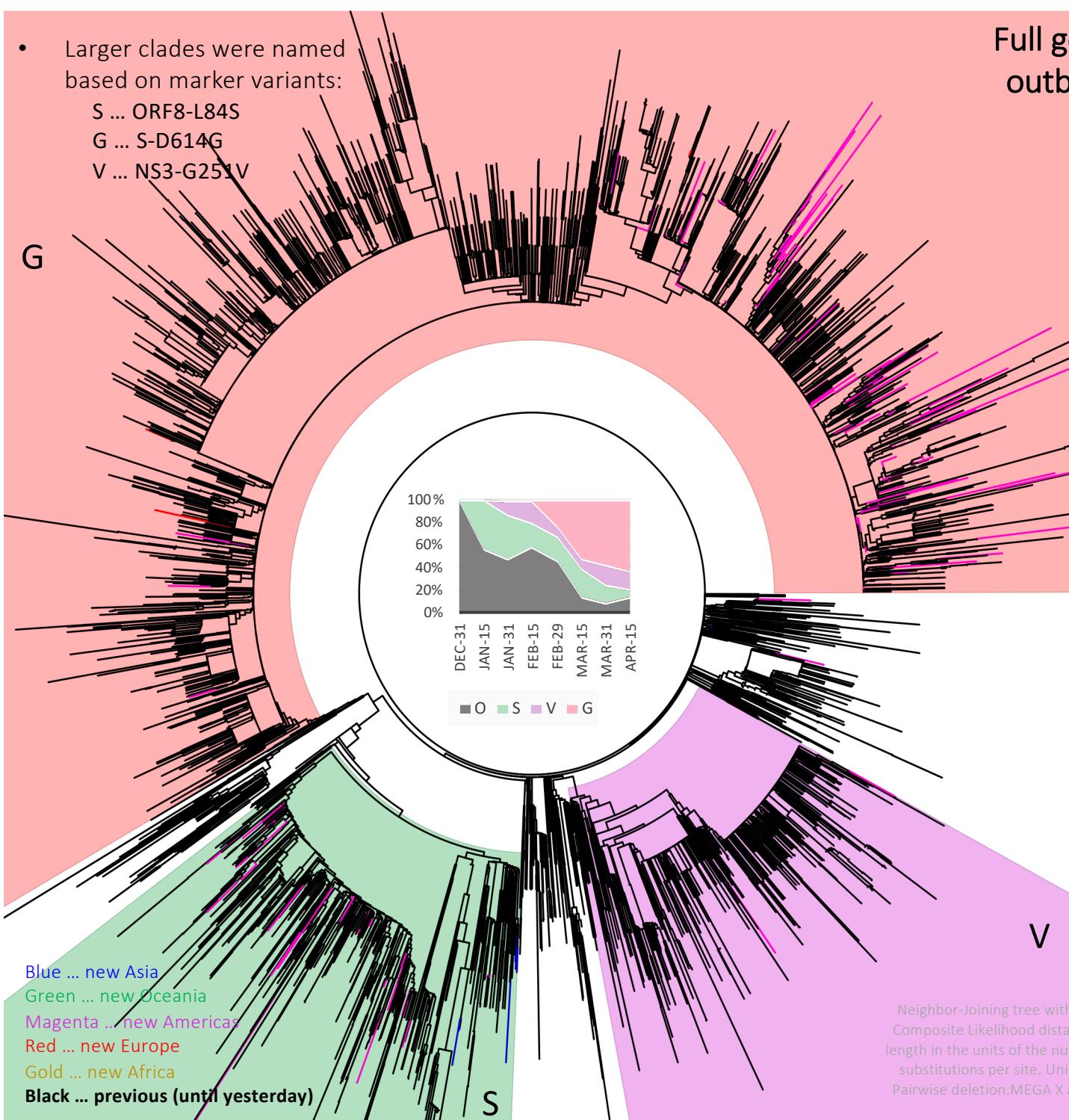
- Larger clades were named based on marker variants:

S ... ORF8-L84S

G ... S-D614G

V ... NS3-G251V

## Full genome tree derived from all outbreak sequences 2020-04-16



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

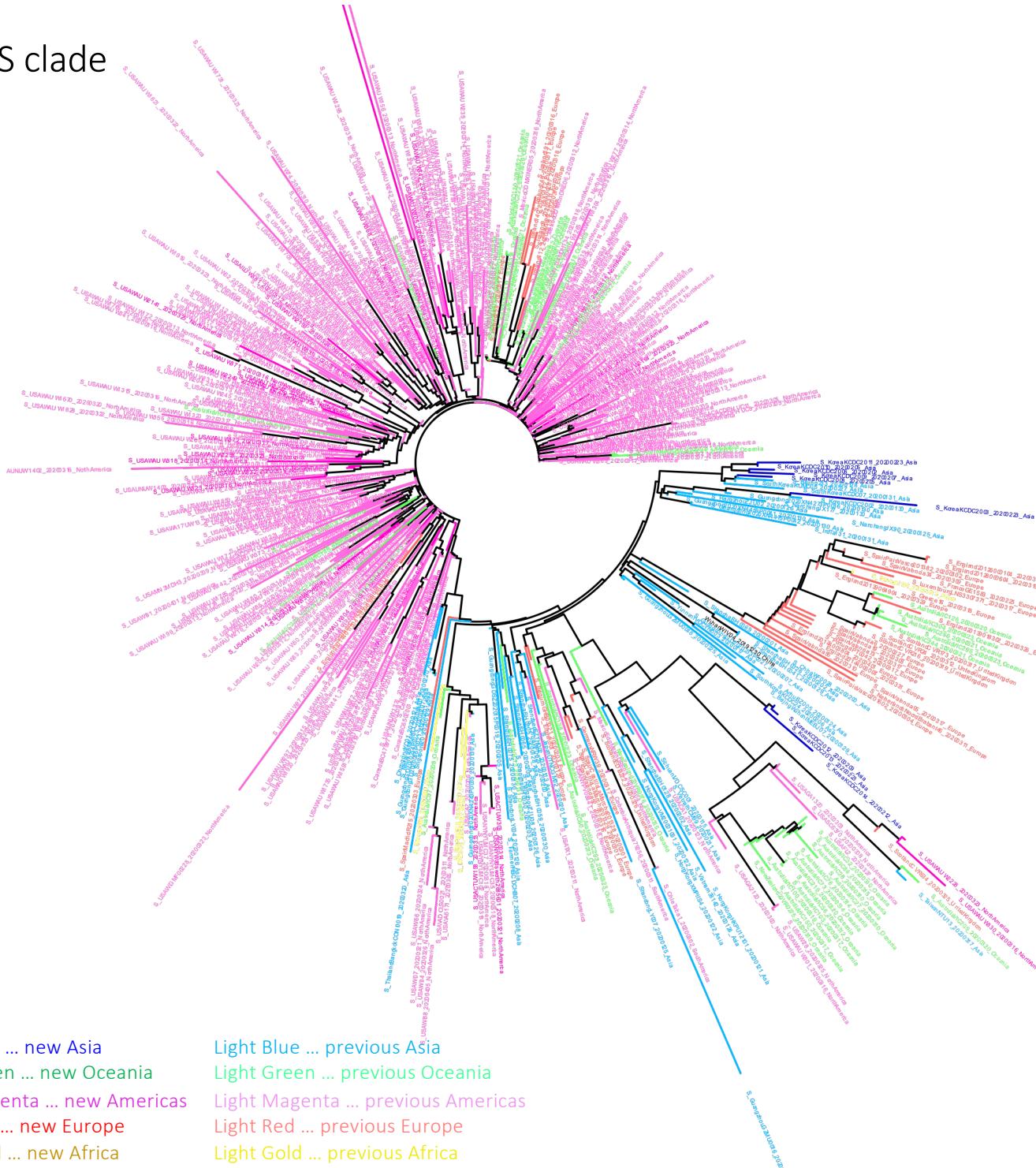
We gratefully acknowledge the Authors from Originating and Submitting laboratories of sequence data on which the analysis is based.



by BII/GIS, A\*STAR Singapore

# Full genome trees of major subclades 2020-04-16

S clade



Notable changes:

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

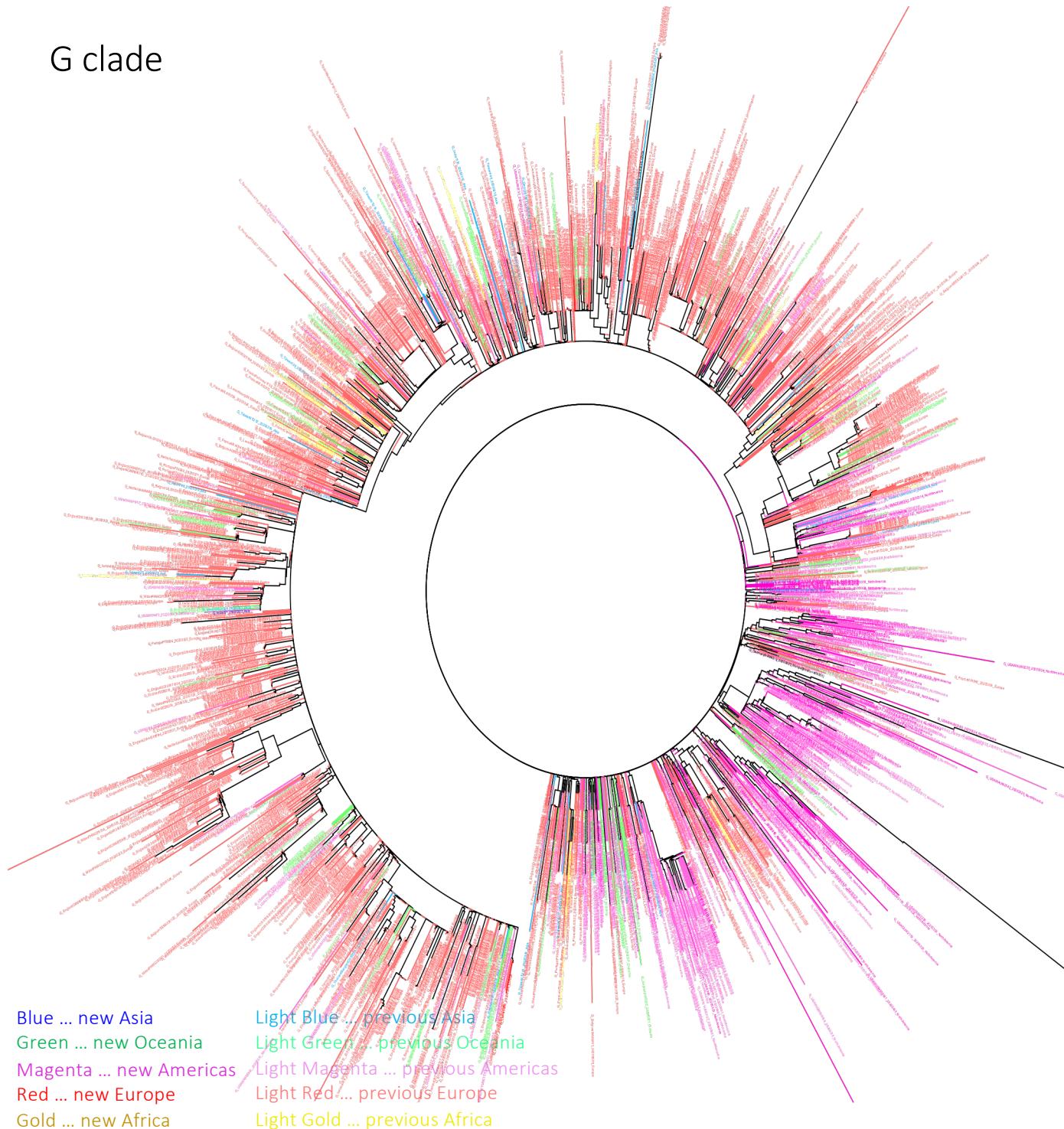
We gratefully acknowledge the Authors from Originating and Submitting laboratories of sequence data on which the analysis is based.

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

G clade

# Full genome trees of major subclades 2020-04-16



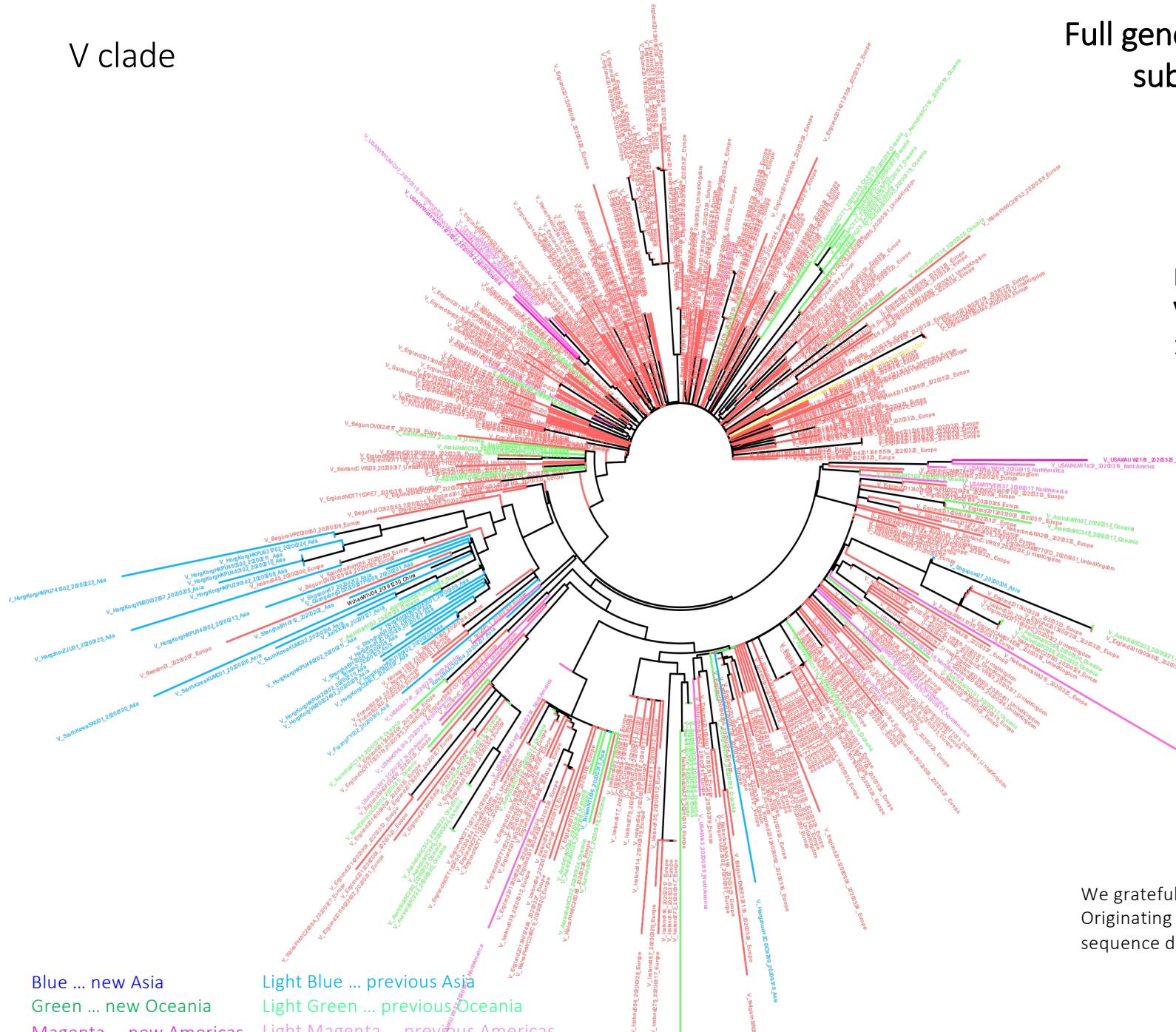
Notable changes:

**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

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

V clade

# Full genome trees of major subclades 2020-04-16



Notable changes:  
**V clade 1,204 (+2):**  
1 USA/WA, 1 USA/NY

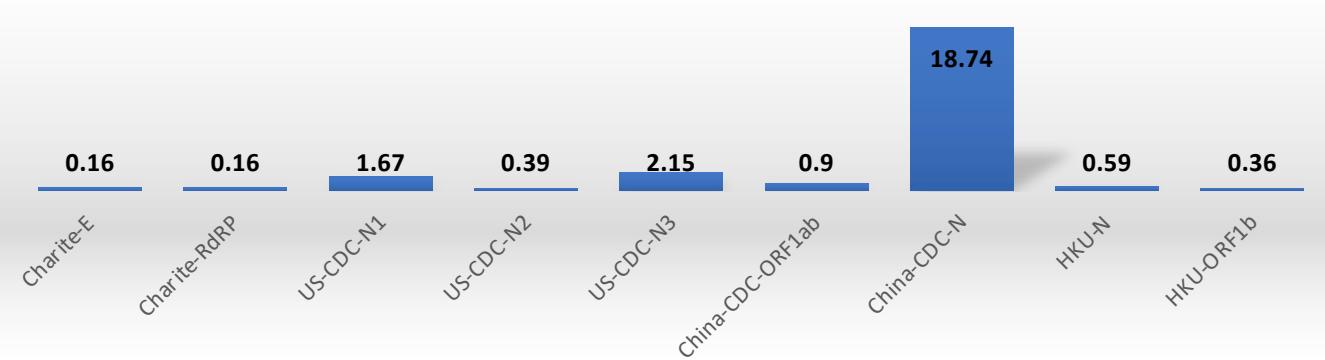
We gratefully acknowledge the Authors from  
Originating and Submitting laboratories of  
sequence data on which the analysis is based.

# Common primer check for high quality genomes

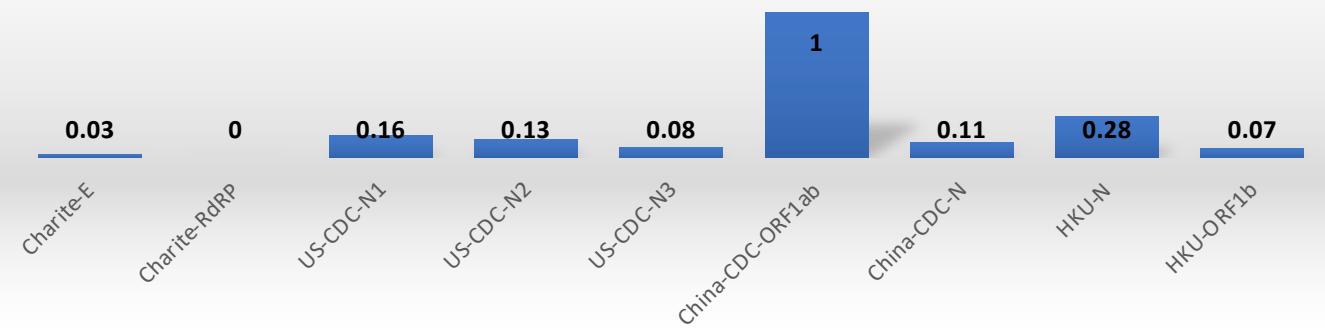
2020-04-15 (updated every 3 days)

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/coronavirus/protocol-v2-1.pdf>  
<https://www.who.int/docs/default-source/coronavirus/peiris-protocol-16-1-20.pdf>  
[http://ivdc.chinacdc.cn/kyjz/202001/t20200121\\_211337.html](http://ivdc.chinacdc.cn/kyjz/202001/t20200121_211337.html)  
<https://www.who.int/docs/default-source/coronavirus/uscdrcrt-pcr-panel-primer-probes.pdf>

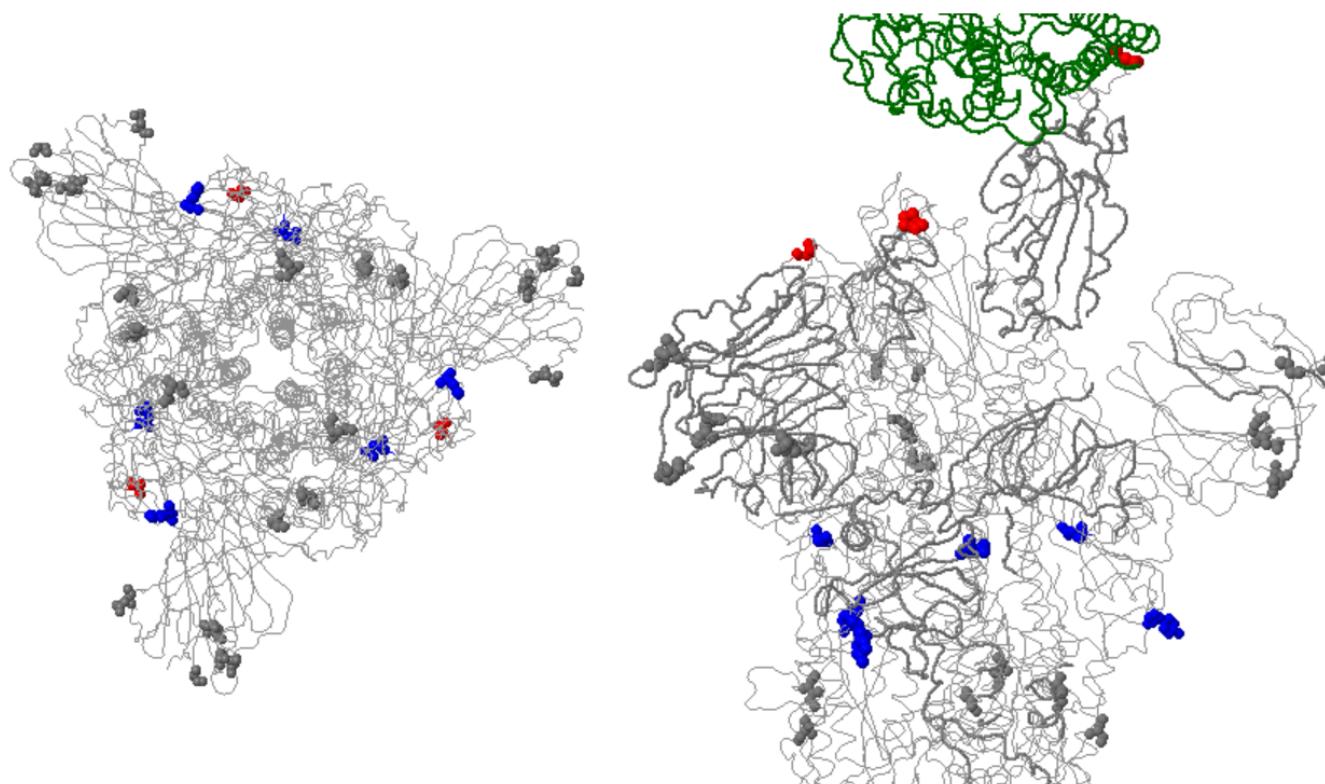
by BII/GIS, A\*STAR Singapore



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

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.

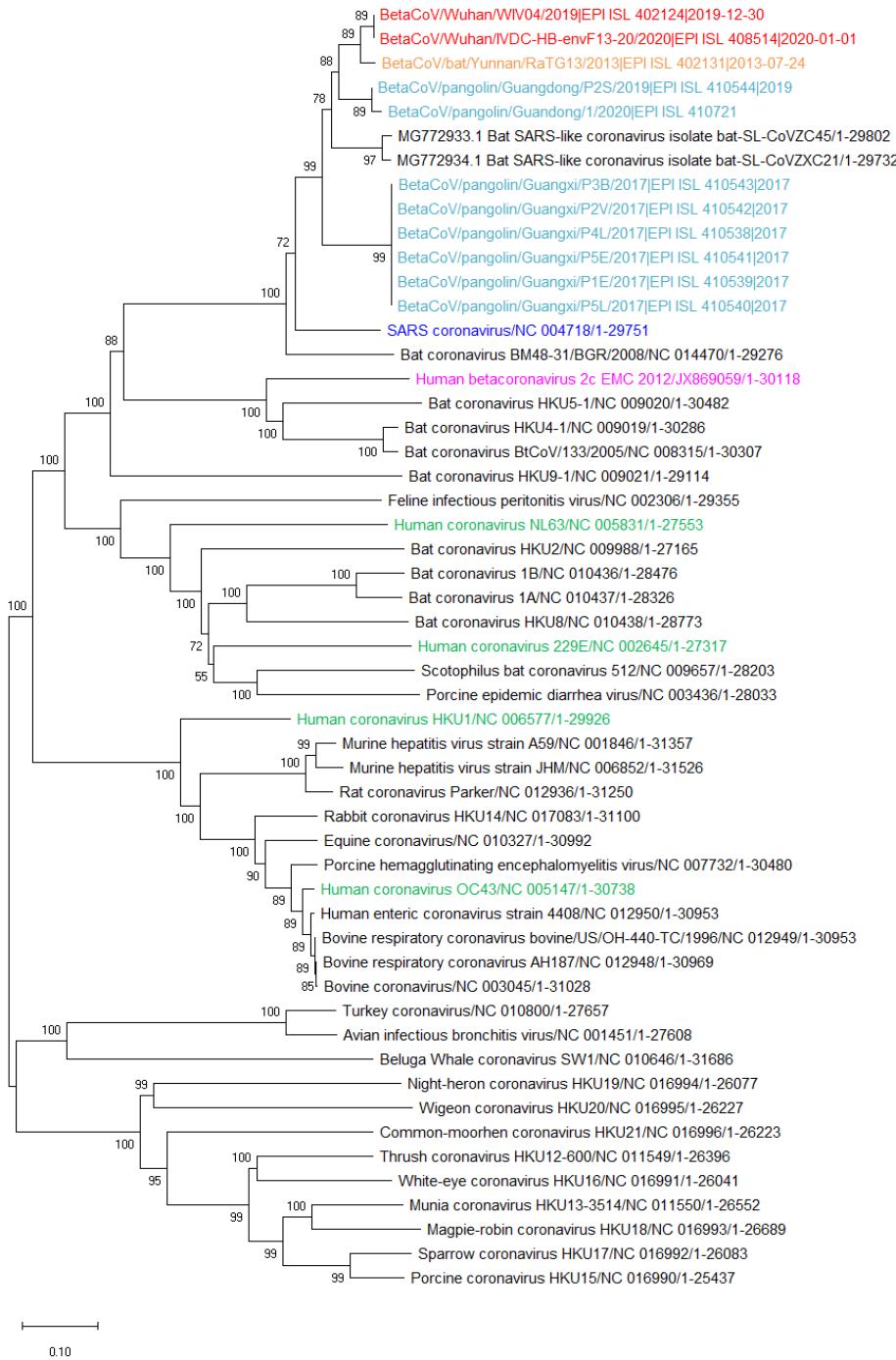
Numbering relative to start codon 21563 in hCoV-19/Wuhan/WIV04/2019

# Summary

First Characterization

by BII/GIS, A\*STAR Singapore





## 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.

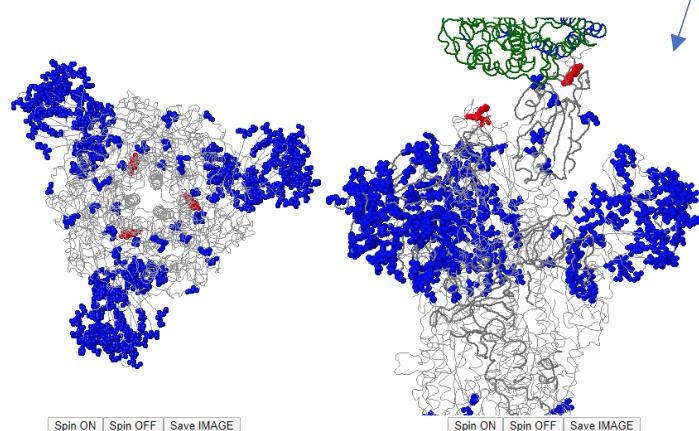
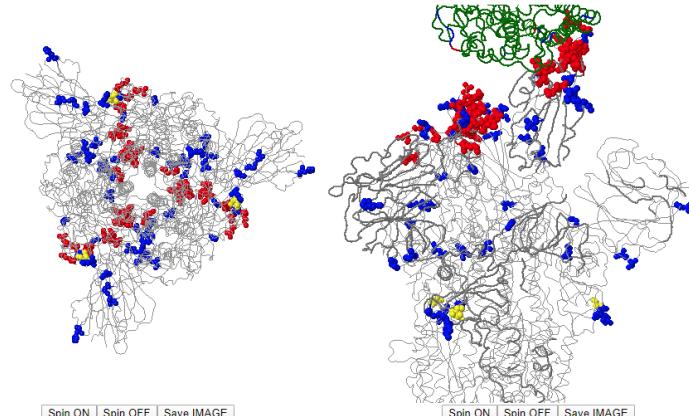
Phylogenetic tree of Wuhan CoV full genome sequences in context of representatives of all CoV families (whole genome Neighbor Joining, Maximum Composite Likelihood, uniform rates, 500 bootstrap, MegaX)

by BII/GIS, A\*STAR Singapore



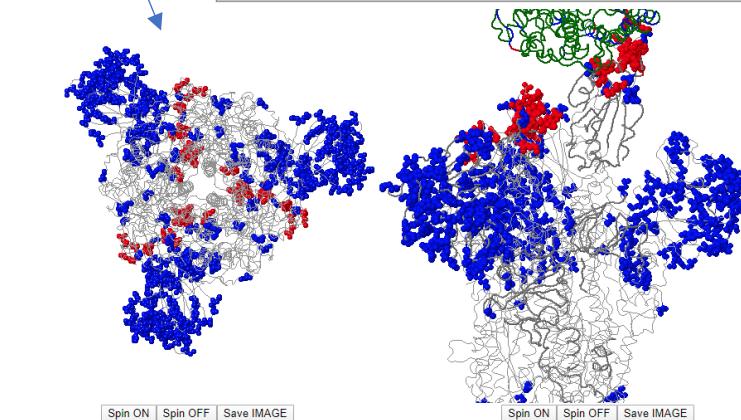
# Spike host receptor changes for nearest bat and nearest pangolin sequences

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



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

Query Sequence: <input type="text" value="BetaCoV/pangolin/Guangdong/1/2019/EPI_ISL_410721/2019"/>	Reference Sequence: <input type="text" value="BetaCoV-2019n-CoV/Wuhan/WIV04/2019"/>	<input type="button" value="Submit"/>
% AA identity: 91.280%		# mutations: 111
List of variations displayed in structure (nearest residue if in loop/termini) <small>Region</small>		
List of mutations not displayed in structure		
<b>\$12N T200 T22A G32A L249 P250 A271 F87 F235 T30Q K41T V42I S46N V47T H49Y S50L T55S D53G L54V F58Y T61S Y66Y H68Y T72G(T69) E76(T77) F79V P85D N78K V90I S39A H10V S112N K13T Y14S V121I E132N M137Y F140Y G142S V143C S151T M153S E154T S155R R156A M163Y M164A Q173K P174S R175P S176T D177N S178P D179N S180P D181N S182P D183N S184P H207Y T210V V212V V213N S214S D215H Q218D D228E L229 I231A R237K Q239A Z243T L244I Y248W S256N S268N A269W G261F E262S Q271A L276M K276N Y294A F306L R346T A372T H402V K417R Q498H M19N K529Q N568S R564S A688 S R689A S708A T747I A1670S A1670E D1684E</b>		



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

Query Sequence <input type="text" value="BetaCoV/pangolin/Guangdong/1/2019[EP1_ISL_A410721]2019"/>	Reference Sequence: <input type="text" value="BetaCoV-2019nCoV-like/bat/Yunnan/RaTG13/2013"/>	<input type="button" value="Submit"/>
% AA identity: 90.307%		# mutations: 123
List of variations displayed in structure (nearest residue if in loop/termini region)		
S17N T20G T22A Q23A L46P P50Q A73T Y28F T30 K41T V61 S69N V97T H49V T51 C55D G49R D59Y T65P R76Q T80I V127I V131M V170I F179V V201I N67Y V80I S94A M10Y S12N K113M V114S E131M N137Y F140Y G142S V143G S151T M153S E154T S156R R158A P183Y N164A Q173K P174S L176M M177L T191W E193S K195R K196R I197Y H207Y T210V L220P V223I V224T V225I V226I V227I V228I V229I V230I G244I Y248D(Z49I) Z250M S255N S256N A260V G261F A262S Q271A L278M K278N T303L D324E M402V K417R I439N H440N I441L A443S E445V F449Y A495S K478T Q484E L487Y A490P Y493Q G494S Y498H D501N H505Y K529Q N565S A604T R634S A684S S687A S704A T743I A1068S A1074T D1089E		
List of mutations not displayed in structure		

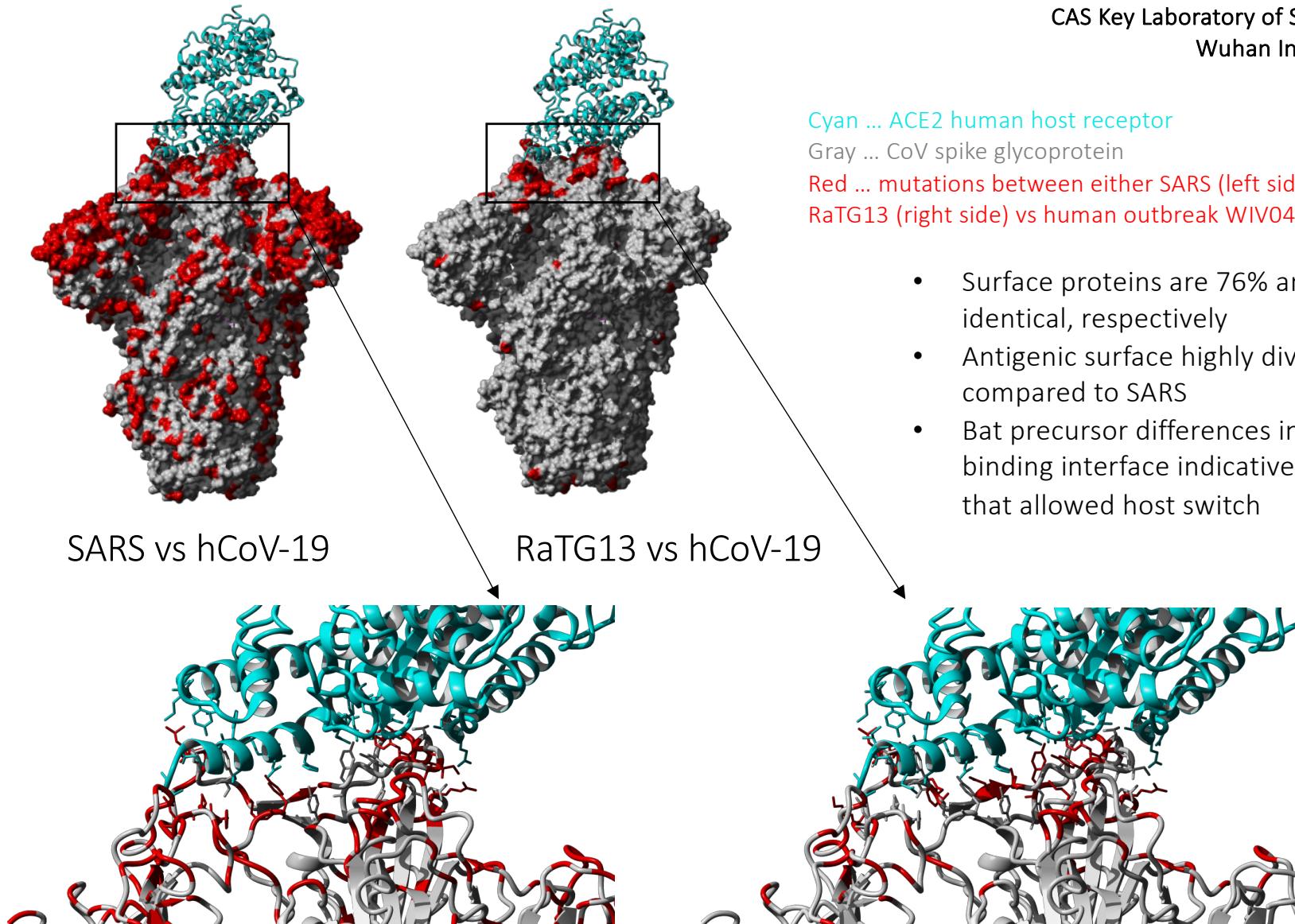
by BII, A\*STAR Singapore

The GISAID logo consists of the word "GISAID" in white, bold, sans-serif letters, set against a green and blue gradient oval background.

# 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



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

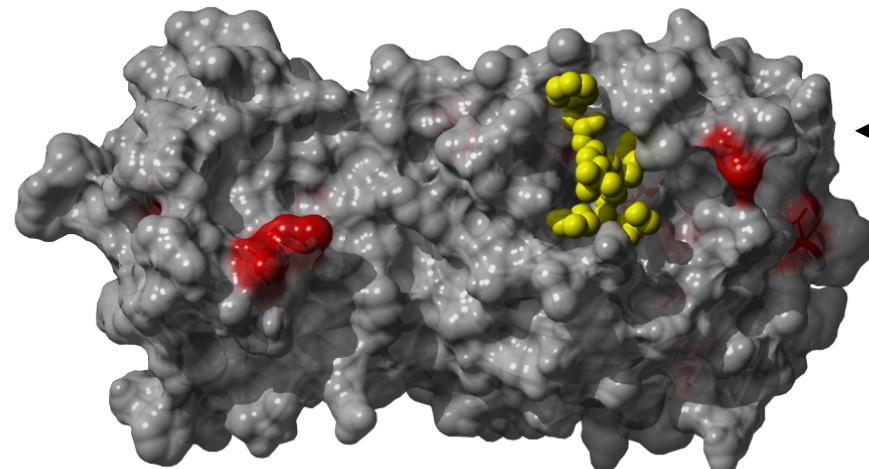
We gratefully acknowledge the Authors from Originating and Submitting laboratories of sequence data on which the analysis is based.

by BII, A\*STAR Singapore



# Potential drug targets highly conserved between hCoV-19 and SARS

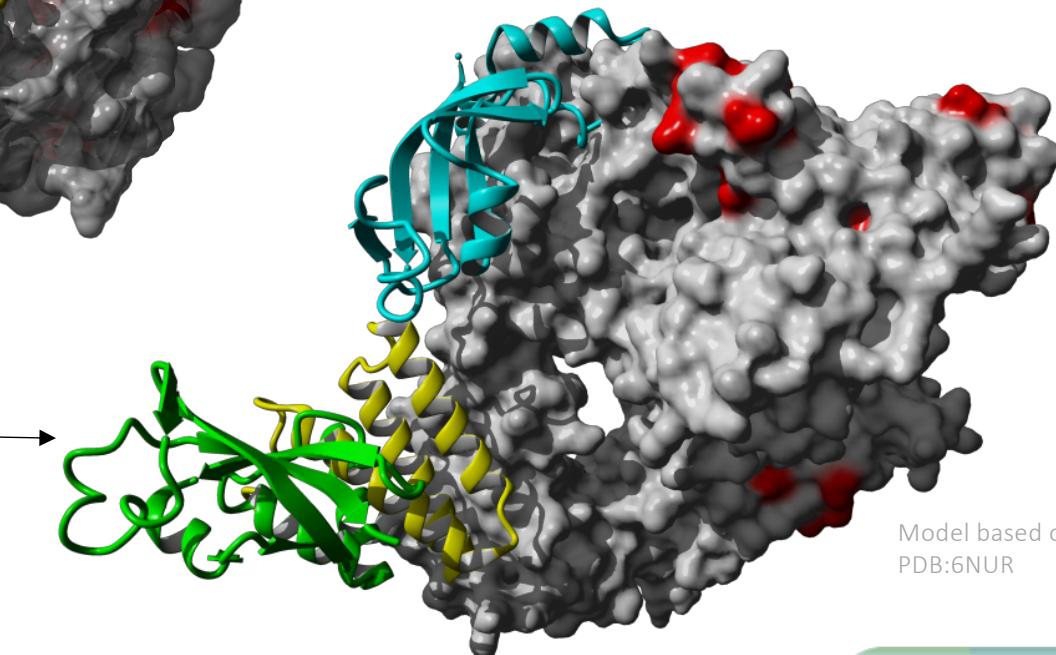
- Both, the main protease and polymerase which are potential drug targets are highly conserved between hCoV-19 and SARS with 96% and 97% overall identity, respectively
- Inhibitors developed against the SARS-CoV main protease or polymerase have good potential to bind similarly to hCoV-19



Model based on PDB:3TNT

## Main protease hCoV-19 vs SARS

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



Model based on  
PDB:6NUR

## Polymerase hCoV-19 vs SARS

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

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

by BII, A\*STAR Singapore

