SARS CoV-2 Variation Update

Bette Korber, Will Fischer, Hyejin Yoon, Werner Abfalterer, Kien Nguyen, Gnana Gnanakaran

Los Alamos National Lab

GISAID data sampled from August 24 - August 31. depending on the slide

Updates available through cov.lanl.gov

Slides 2-7) Brief Spike D614G update, and data supporting *enhanced* neutralization sensitivity of the G614 form

Slides 8-15) Current summary of most common Spike mutations

Slide 16-34) Current summary of mutations in the SARS-CoV-2 proteome, major clades and their shifting frequencies, and how this data overlays with the BEI reagents for live virus stocks.

Thanks to GISAID, the Bloom lab for their great interactive github website, and the LANL team.

https://www.gisaid.org

https://jbloomlab.github.io/SARS-CoV-2-RBD_DMS/

cov.lanl.gov

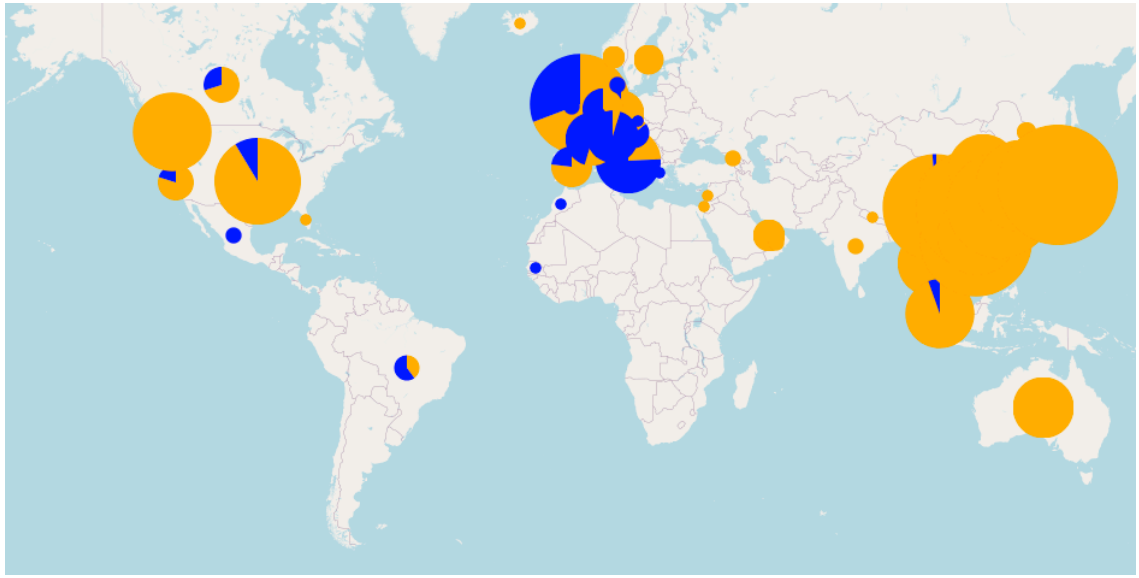
New points

- The original viruses that carried D614 are not very rarely sampled. Among the G614 G clade viruses, the GR clade is now globally the most common, and is tending to increase in frequency relative to the G and GH clade.
  - I don't see GR as an option among the viral stocks available from BEI, perhaps it would be good to get a reference stock in place?
  - The defining amino acid substitutions for the GR clade are outside of Spike, so this is not a worry for Spike reagents
- The Spike S477N mutation virus has become very common in Australia, particularly in sequences from Victoria, and is now >4% of the global GISAID sample; Australia was heavily sampled July, so this biases the global sample some, but even so this merits a deeper look.
  - S S477N is in the Receptor Binding Domain, and arose in the context of the GR clade.
- The Spike D936Y mutation has stayed stable at about 1%
- All other Spike mutations are still <1%, and are summarized in the spreadsheets, but not discussed here.
- The Spike spreadsheet now has more extensive annotation regarding epitopes and functional regions coupled to the sequence variation. We are going to start building a relational database with this data soon…
- After helpful conversations with Mark Lewis, I think it might be still be worth resolving if the RdRp P323L mutation that is carried along with D614G has a functional impact. I think some people are using the BavPat1 as a G614 virus, and it doesn't have the RdRp P323L mutation.

The original Spike D614 form is now rarely sampled in GISAID, G614 is globally dominant

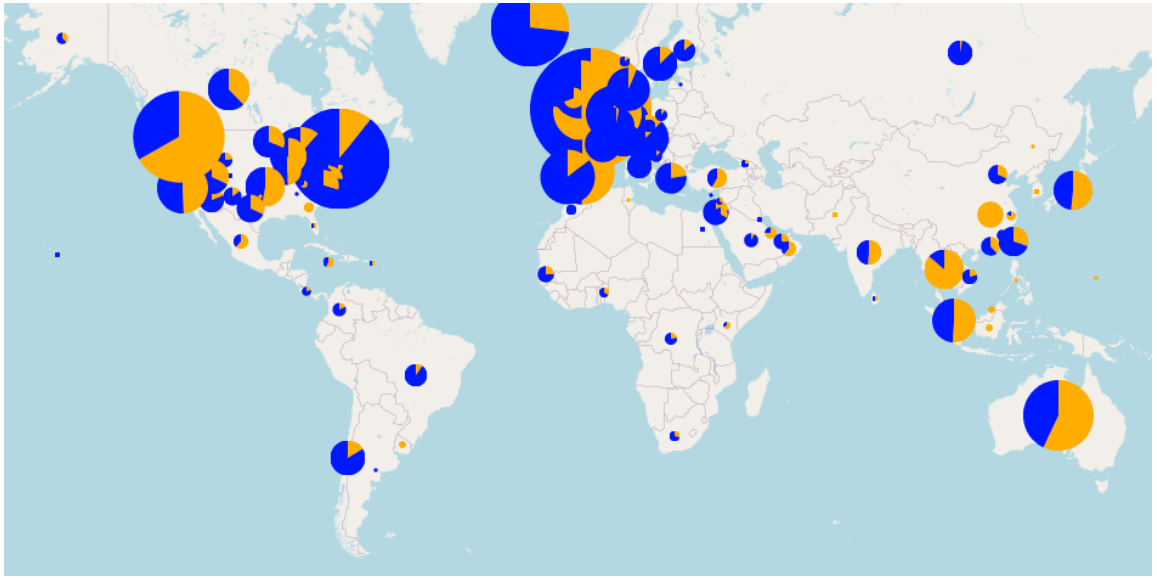
D614
G614

Prior to March 1

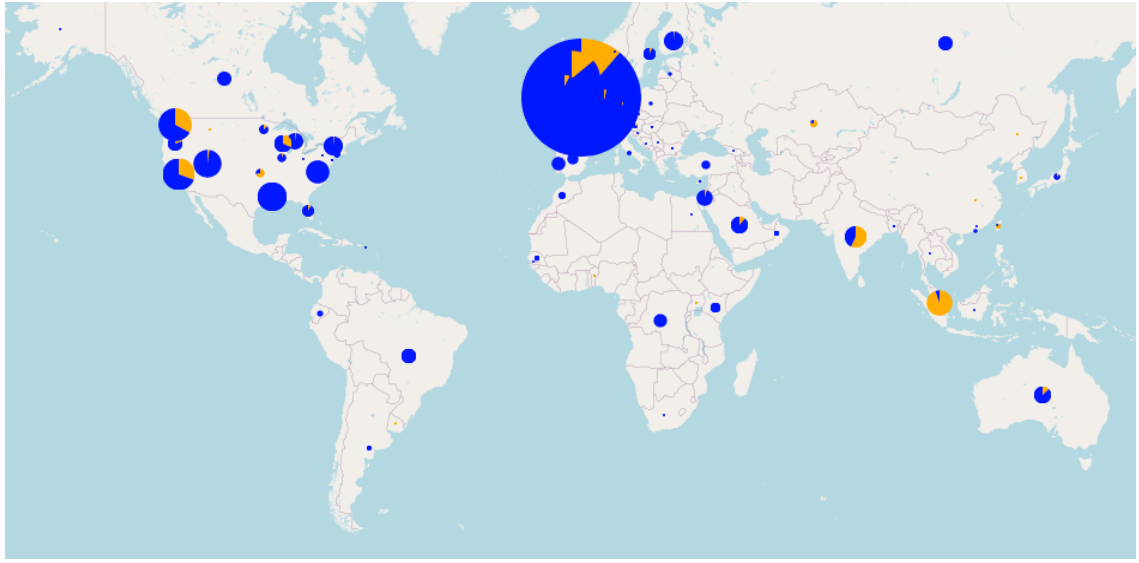


Sampled Aug 25th, 2020

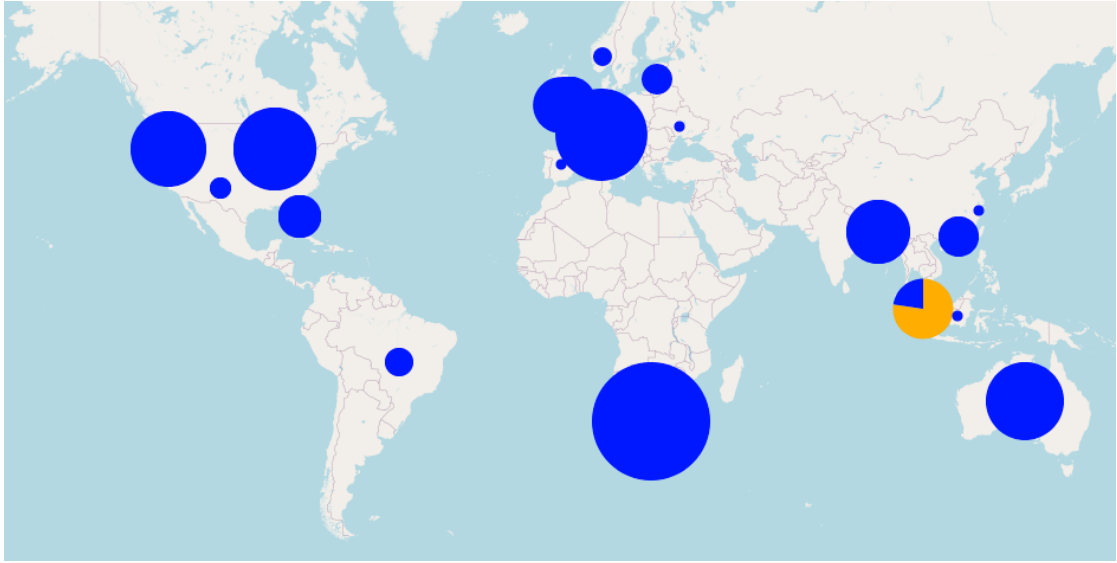
March 11-20



April 11-20



July 11-20



Circle size reflects sampling within a country, pie chart relative frequency of the original Spike D614 and G clade Spike G614

The D614G Spike Mutation *Increases* SARS CoV-2 Susceptibility to Neutralization

MEDRXIV/2020/159905, manuscript submitted for peer review

Drew Weissman, Mohamad-Gabriel Alameh, Thushan de Silva, Paul Collini, Hailey Hornsby, Rebecca Brown, Celia C. LaBranche, Robert J Edwards, Laura Sutherland, Sampa Santra, Katayoun Mansouri, **Sophie Gobeil**, Charlene McDanal, Norbert Pardi, Nick Hengartner, COVID-19 Genomics Consortium UK, Paulo J.C. Lin, Ying Tam, Pamela A. Shaw, Mark G. Lewis, Carsten Boesler, Uğur Şahin, **Priyamvada Acharya**, Barton F. Haynes, Bette Korber, **David C. Montefiori**

1) Effect seen in sera from vaccinated mice, primates and people immunized with the nucleoside-modified mRNA-LNP vaccine platform
   - vaccine data illustrations, next 2 slides

2) An average of 2-fold enhanced neutralization sensitivity to G614 was also found in 70 convalescent sera from recovered subjects

3) Depending on the antibody, sensitivity to neutralizing was sometimes increased (2-150 fold). Some neutralizing antibodies are not impacted at all, it depends on the antibody.

Vaccines: Four different variants of the Spike immunogen:

1 monomeric secreted RBD
2 trimeric secreted RBD
3 diProline stabilized D614 Spike
4 Furin mutant D614 Spike, S1 and S2 subunit associations maintained

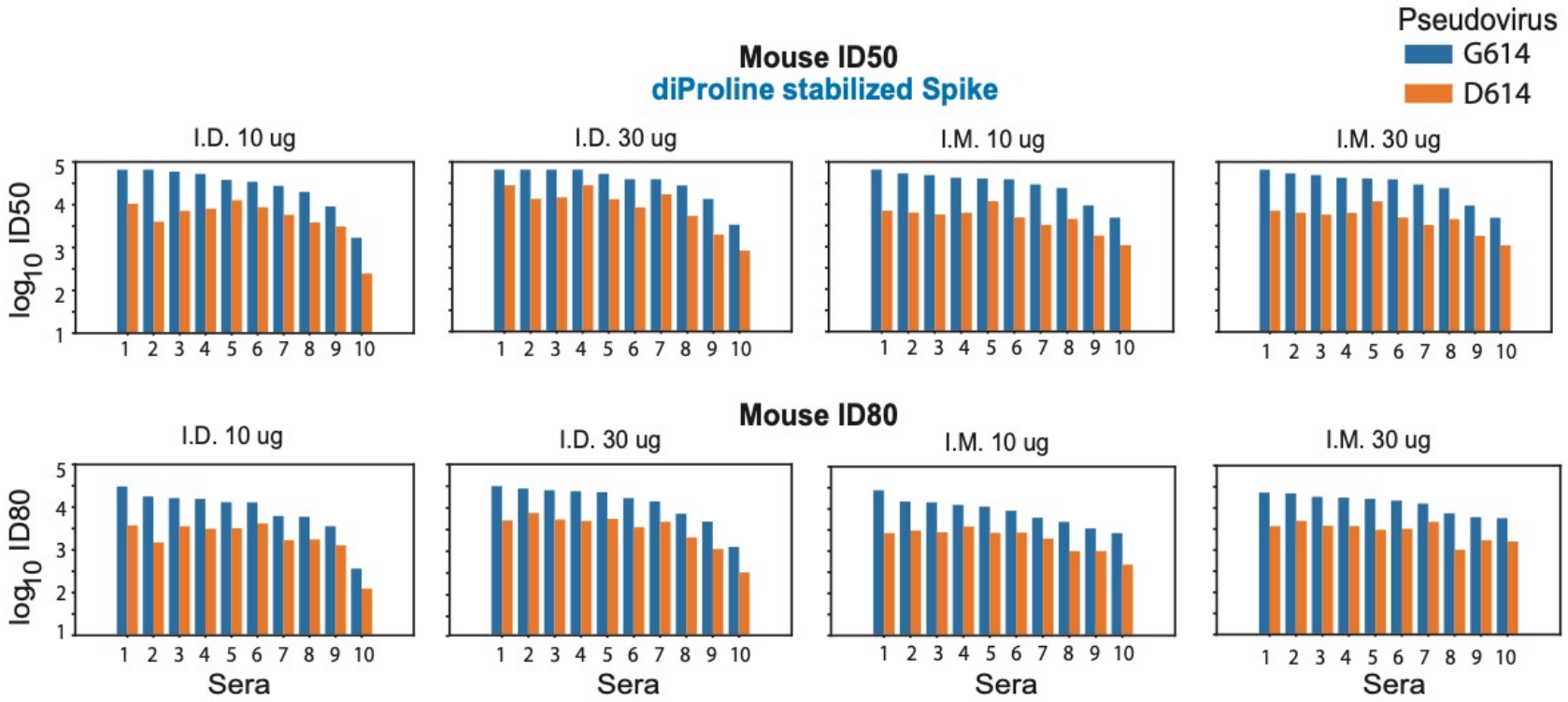
Pseudovirus and neutralization assay:

SARS-CoV-2 neutralization was assessed with Spike-pseudotyped viruses in 293T/ACE2 cells as a function of reductions in luciferase (Luc) reporter activity.

Spike D614 and G614 pseudotype viruses were created in a lentivirus backbone.

4 groups of 10 mice each, comparing dose and delivery

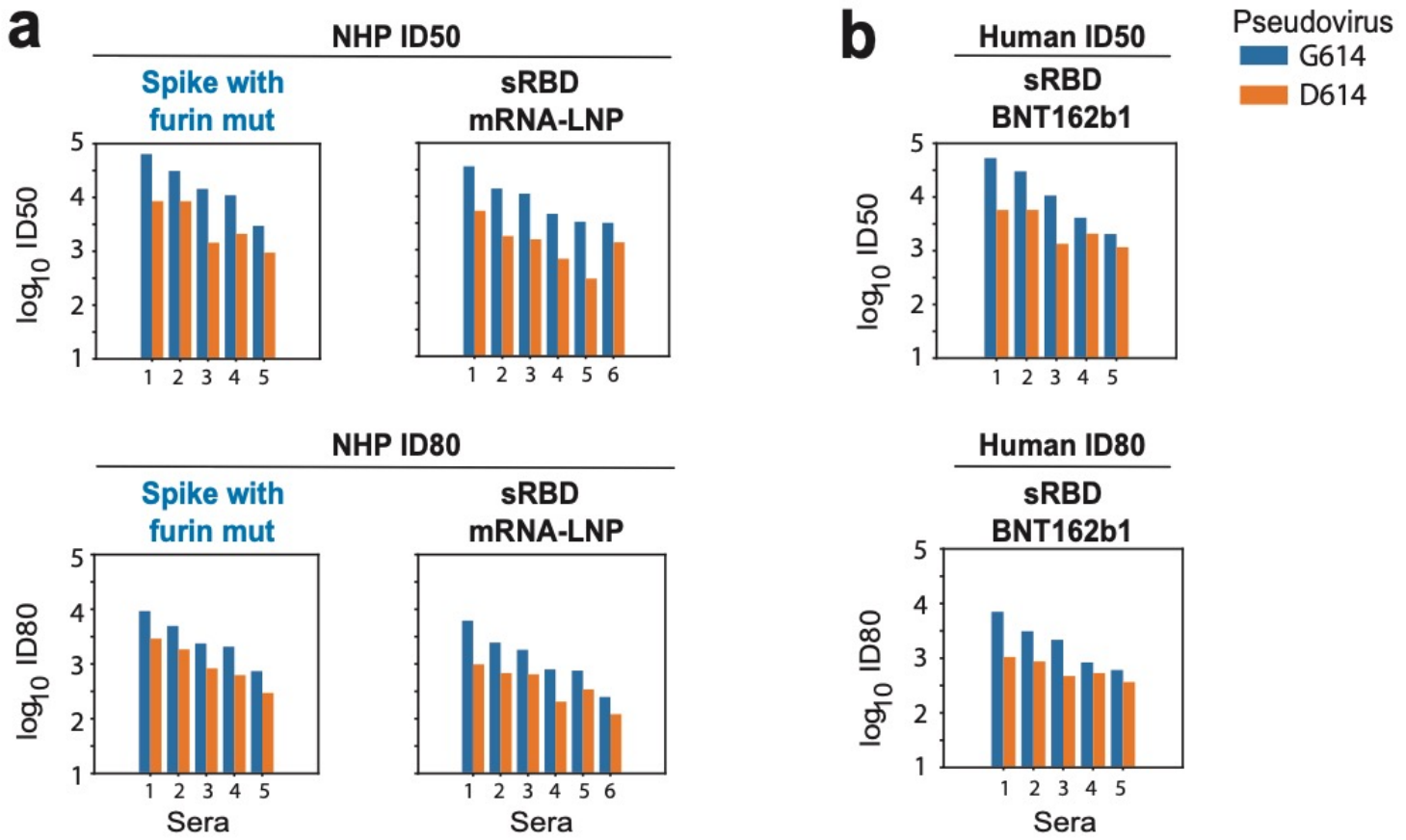
The sera from vaccinated mice were evaluated for neutralization potency, comparing D614 and G614 pseudoviruses.
Each pair of bar graphs represents one serum, G614 is always more sensitive



Route	Dose (µg)	ID50 Geometric mean G:D ratio, 95% CI	p-value	ID80 Geometric mean G:D ratio, (95% CI)	p-value
I.D.	10	5.6 (3.9 — 8.1)	2.0e-06	4.4 (3.2 — 6.2)	3.6e-06
I.D.	30	3.9 (2.9 — 5.1)	1.5e-06	4.3 (3.7 — 5.0)	3.8e-09
I.M.	10	6.5 (5.1 — 8.3)	3.3e-08	4.5 (3.5 — 5.8)	3.5e-07
I.M.	30	5.2 (3.8 — 7.0)	7.8e-07	4.6 (3.7 — 5.6)	4.8e-08
All 40 mice		5.2 (4.5 — 6.0)	2.2e-16	4.4 (4.0 — 4.9)	2.2e-16

David Montefiori, Drew Weissman et al.

G614 enhanced sensitivity also seen in Non-Human Primates (NHPs) and People



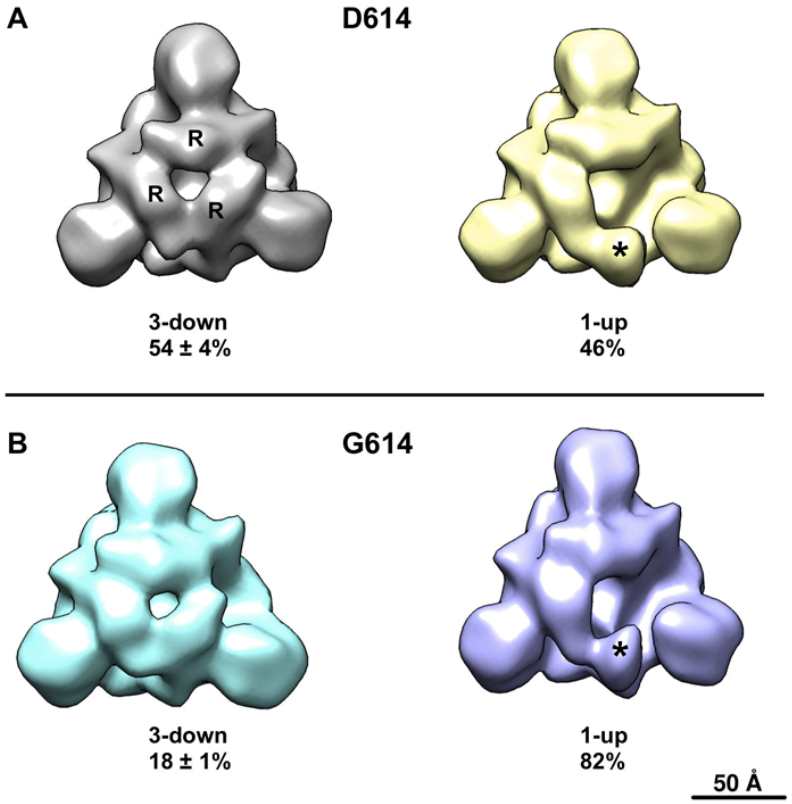
Species	Vaccine	ID50 Geometric mean G:D ratio, 95% CI	p-value	ID80 Geometric mean G:D ratio, (95% CI)	p-value
NHP	Spike furin mut	5.4 (3.0 — 9.8)	0.0015	2.9 (2.5 — 3.3)	<0.001
NHP	sRBD mRNA-LNP	6.5 (3.7 — 11.4)	<0.001	3.2 (2.1 — 5.0)	<0.001
Human	sRBD BNT162b1	4.2 (1.6 — 11.0)	0.014	3.1 (1.4 — 6.8)	0.017

We think the mechanism for both the enhanced infectivity and neturalization sensitivity is that the D614G mutant Spike prefers the “one up” conformation which allows ACE2 interactions and exposes the RBD epitope regions

Negative stain electron microscopy reconstructions

Sophie Gobeil, Priyamvada Acharya

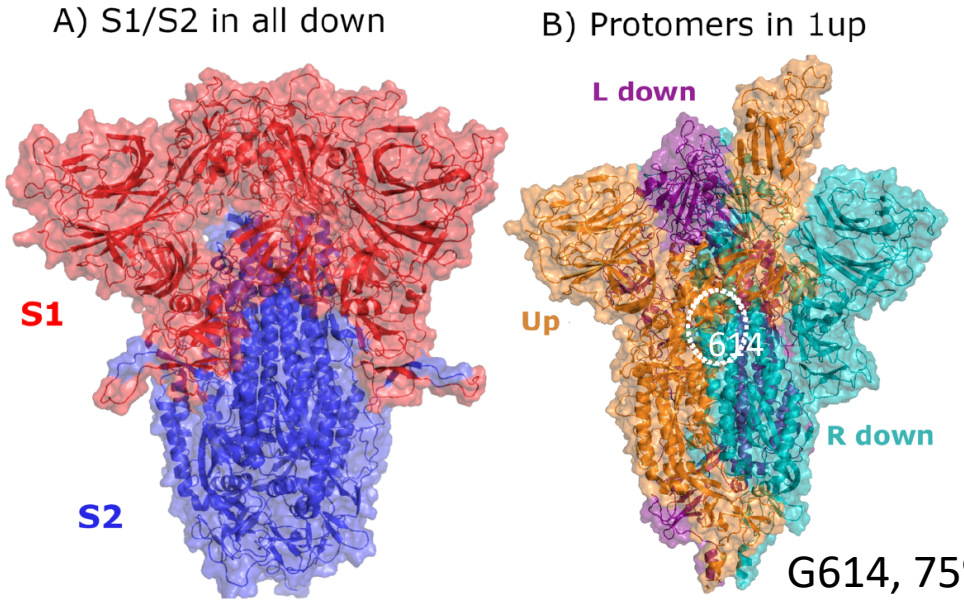
From Weissman et al.



The SARS-CoV-2 Spike Variant D614G Favors an Open Conformational State

Rachael Mansbach, Srirupa Chakraborty, Kien Nguyen, David Montefiori, Bette Korber, **S Gnanakaran**

bioRxiv



G614, 75% 1-up

https://biorxiv.org/cgi/content/short/2020.07.26.219741v1

Summary of Spike Mutations: Spreadsheet

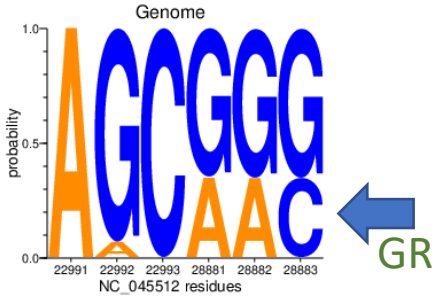
- A unannotated version of both of the spreadsheet tables that summarizes variation are provided with a daily updates from GISAID at *cov.lanl.gov*

- The tab labeled "Spike Variation" contains a row for each position in Spike that includes:
  - The number of each variant, the entropy of each site, the local entropy of each 10 amino acid stretch is shown
  - Sites with > 0.3% variation in GISAID are highlighted in red.
  - Sites and local linear regions that have relatively high entropy are highlighted in yellow
  - We are working on a Genome browser for this information
  - Sites are annotated with Spike regions and mAb features annotated from the following sources:
    - Starr TN, et al. … Bloom JD. Cell. 2020 Aug 11:S0092-8674(20)31003-5. doi: 10.1016/j.cell.2020.08.012. PMID: 32841599
      Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding
      Annotation is based on: https://jbloomlab.github.io/SARS-CoV-2-RBD_DMS/
    - Weisblum Y et al. …Bieniasz PD. bioRxiv. 2020 Jul 22:2020.07.21.214759. doi: 10.1101/2020.07.21.214759. Preprint. PMID: 32743579
      Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants
    - Barnes CO, et al. … Bjorkman PJ. Cell. 2020 Aug 20;182(4):828-842.e16. doi: 10.1016/j.cell.2020.06.025 .PMID: 32645326
      Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies

- The tab labeled "Sites of Interest", summarizes amino acids that on August 31 had >0.3% variants
  - We exclude D614G
  - For other varying sites we provide counts, codons, amino acid variants, and geographic regions
  - A rough version of this table is provided with a daily update at cov.lanl.gov

Summary of Spike Mutations

- We are highlighting 2 sites in Spike:
  - Spike 936: We have been tracking this site as it maintained in GISAID at frequency >1%
  - Spike S477N: We are adding this site as it is very common in Australia, and circulating in England
    - Bases: codon bases 22991-3: AGC -> AAC, base G22992A, encodes S477N
    - Arose as part of the GR clade
    - LOGO shows the global frequencies of bases in the S477N codon, and of the three bases that define the GR subclade of the G clade.
    - The mini alignment shows the global frequencies of there forms.

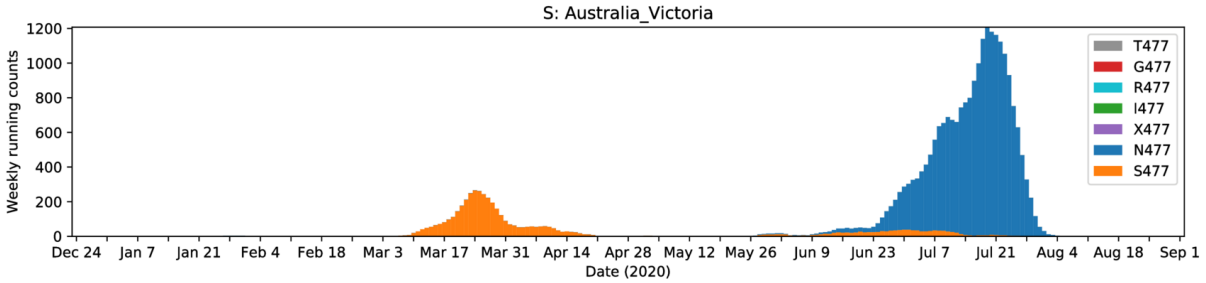


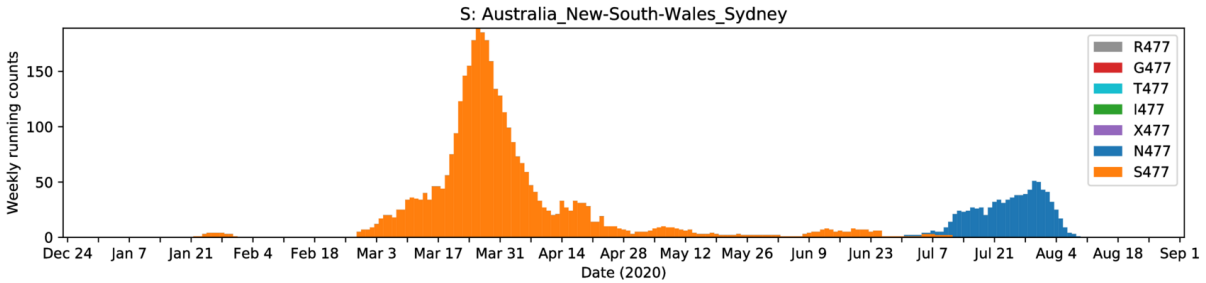
GR clade

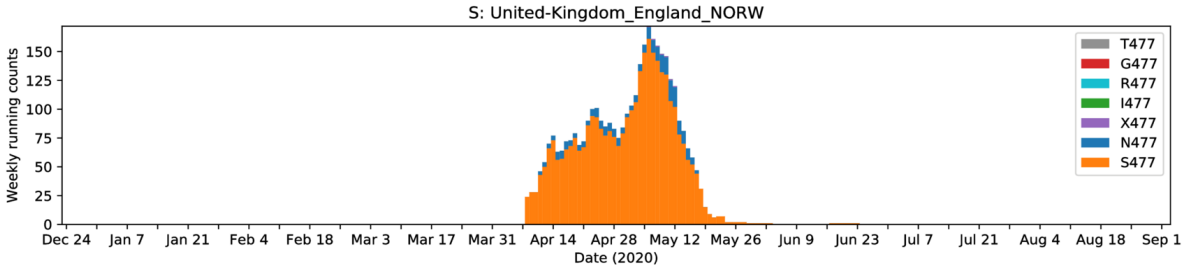
AGCGGG	Count	%	
------	32045	64.86	
---AAC	13778	27.89	GR
-A-AAC	3369	6.82	S477N,GR
-A----	9	0.02	S477N

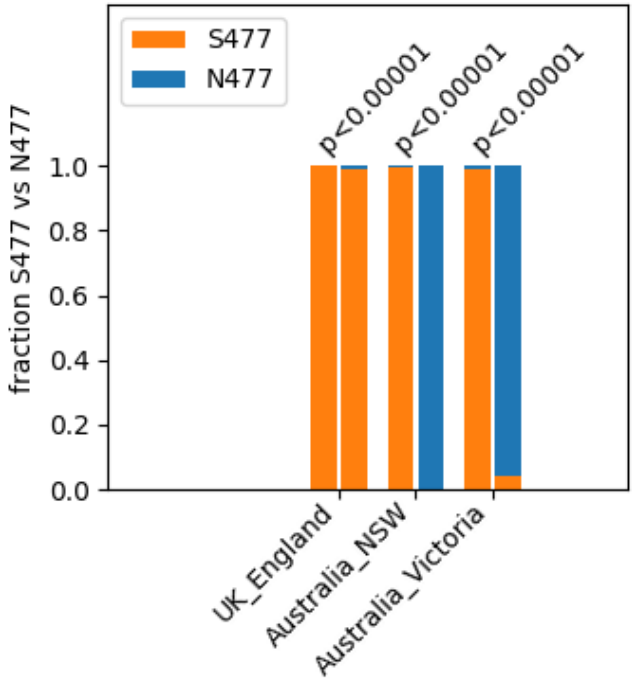
The mutation S477N has recently increased to 4.7% of GISAID.

This can be traced to a large influx of sequences from Australia, where it
Is now the most common form; its rise is in conjunction with the GR clade.
This style of plot shows the weekly average running count of variants found at S 477.









This style of graph shows *all* of the geographic regions in GISAID where amino acids at position 477 have significantly changed frequency over time.

S477N is essentially mostly found in Australia so far, with a small set of cases in England.

This may just be a founder effect in Australia, but worth keeping an eye on, and testing.

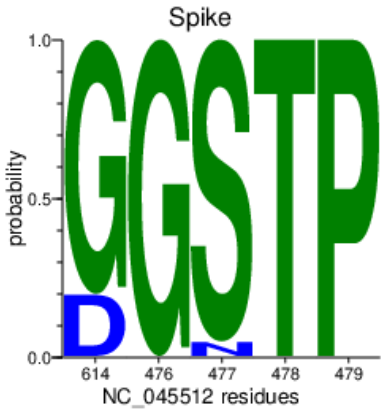
D614 are sequences excluded from these graphs

Recent sampling from Australia shows the S477N (bases 22991-3 mutation is becoming common there.

Spike position

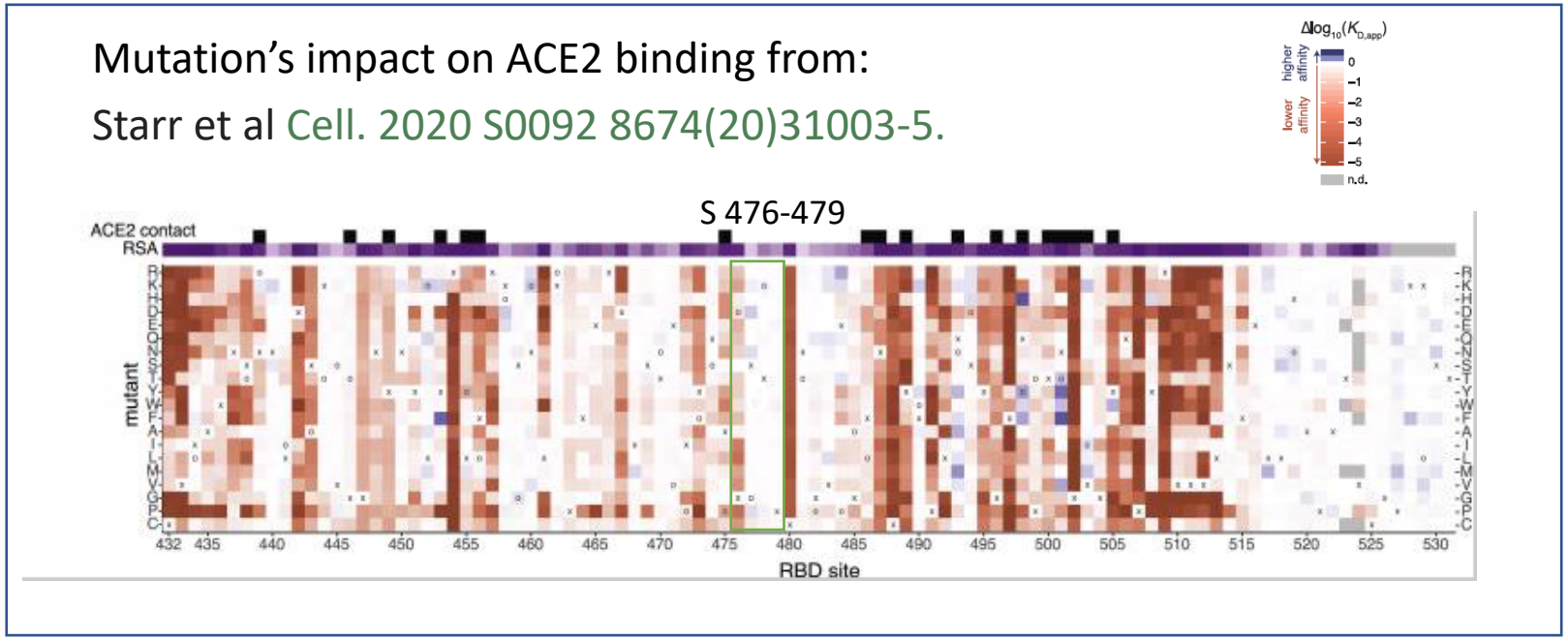
614	476	477	478	479	Count	%
G	GSTP				55394	75.31
D	----				14369	19.54
-	-N--				3410	4.64
-	--I-				97	0.13
-	---S				62	0.08
D	S---				12	0.02
-	S---				7	
-	-I--				7	
-	--K-				3	
-	---L				3	
-	-T--				2	
-	-R--				2	

Australia: 3208 in Victoria, 122 Sydney; UK: 57 in Norwich
UK: 89 in Cambridge
UK: 58 in Wales
US: 9 in Washington state

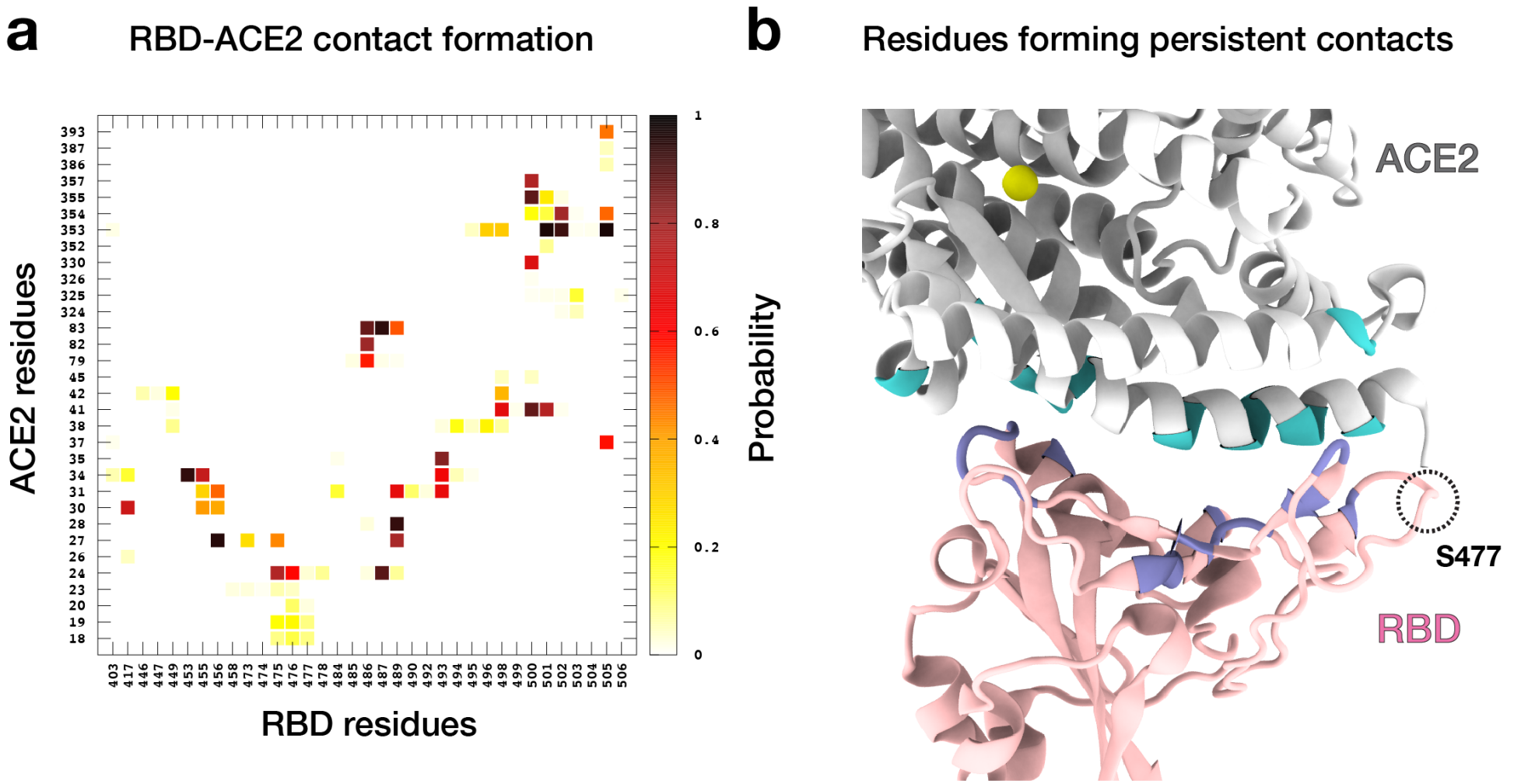


Mutation's impact on ACE2 binding from:

Starr et al Cell. 2020 S0092 8674(20)31003-5.



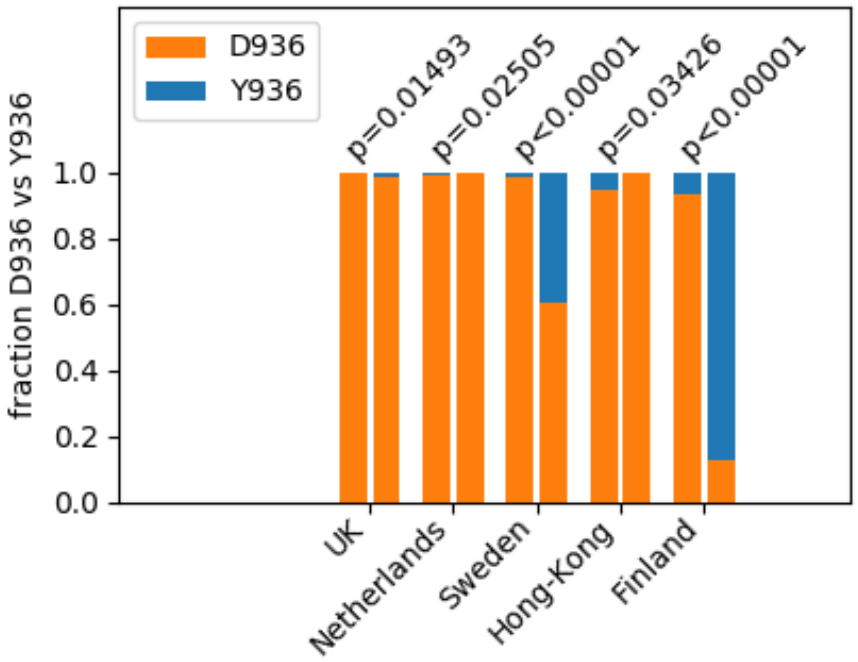
RBD –ACE2 Contacts



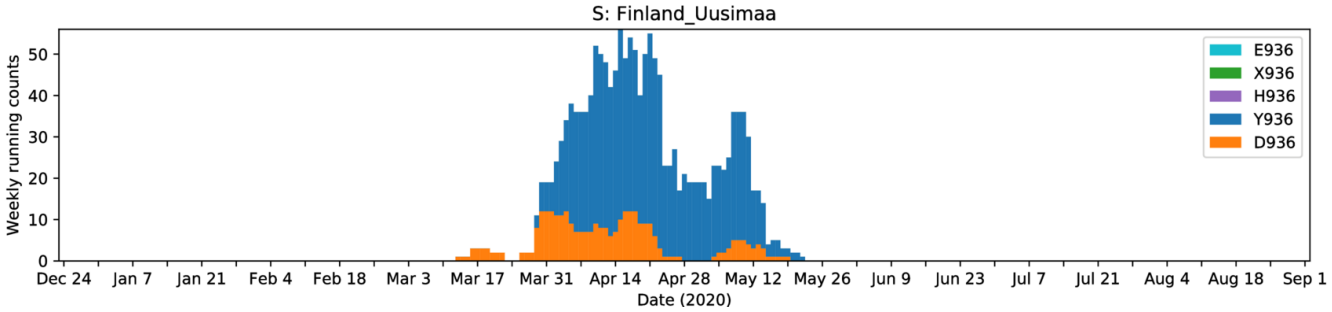
Contact signature between ACE2 and RBD from hundreds of microseconds simulations. (a) Probability of forming a residue-residue contact between ACE2 and RBD. (b) Structural representation of the RBD-ACE2 binding interface highlighting the residues that form persistent contacts (cyan in ACE2 and ice blue in RBD). Dashed black circle indicates the region of S477.

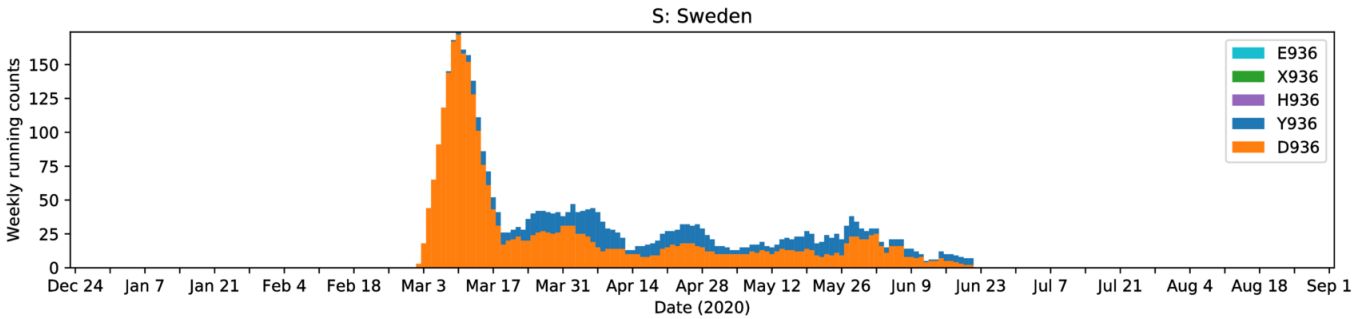
The mutation D936Y continues to maintain at ~1% of GISAID samples, and to persist in Finland, Sweden

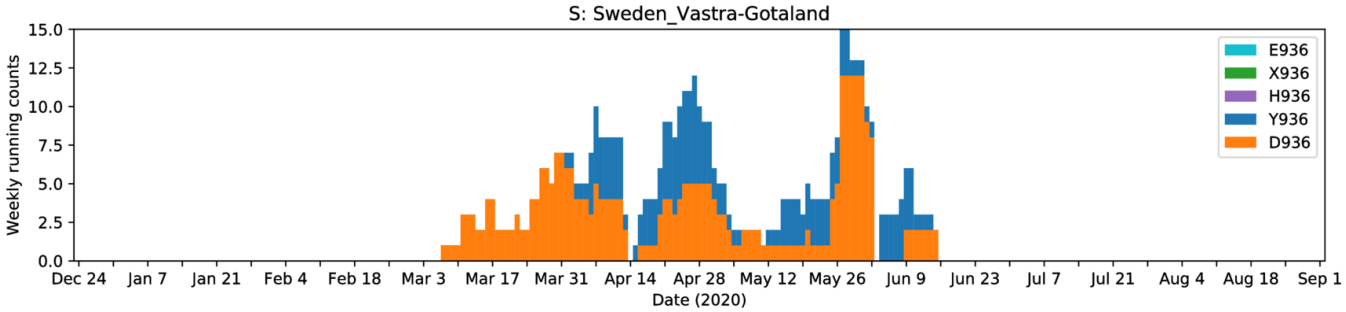
Rare but present in the UK

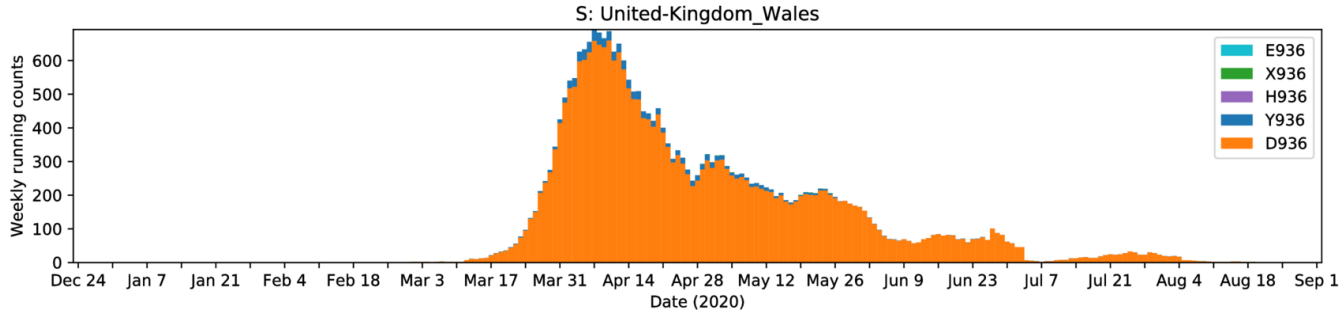


No consistent pattern of increase found In GISAID samples.



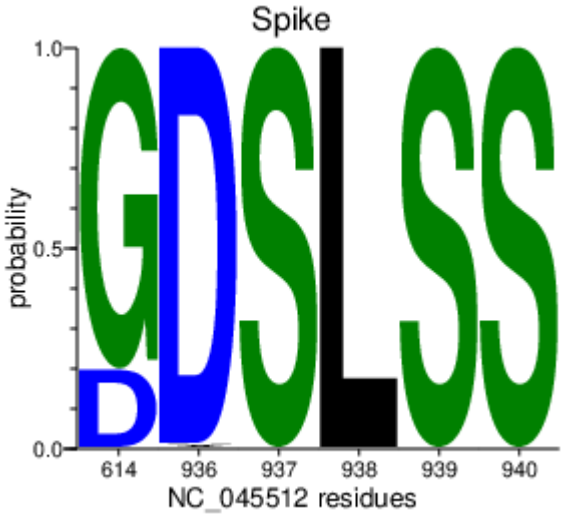






HR1 D936Y

614 + Variable region 936-940



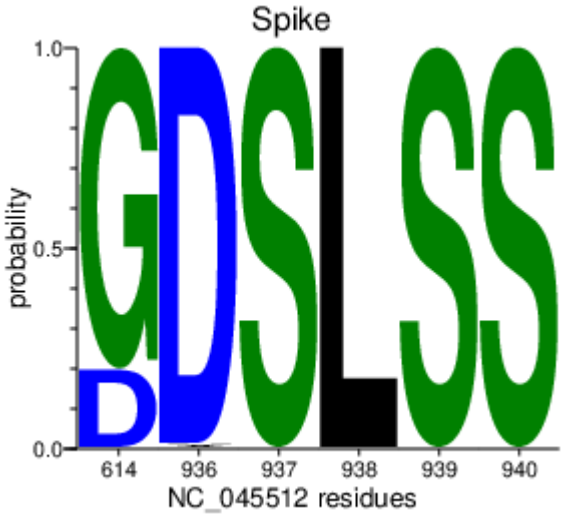
Position number under the LOGO

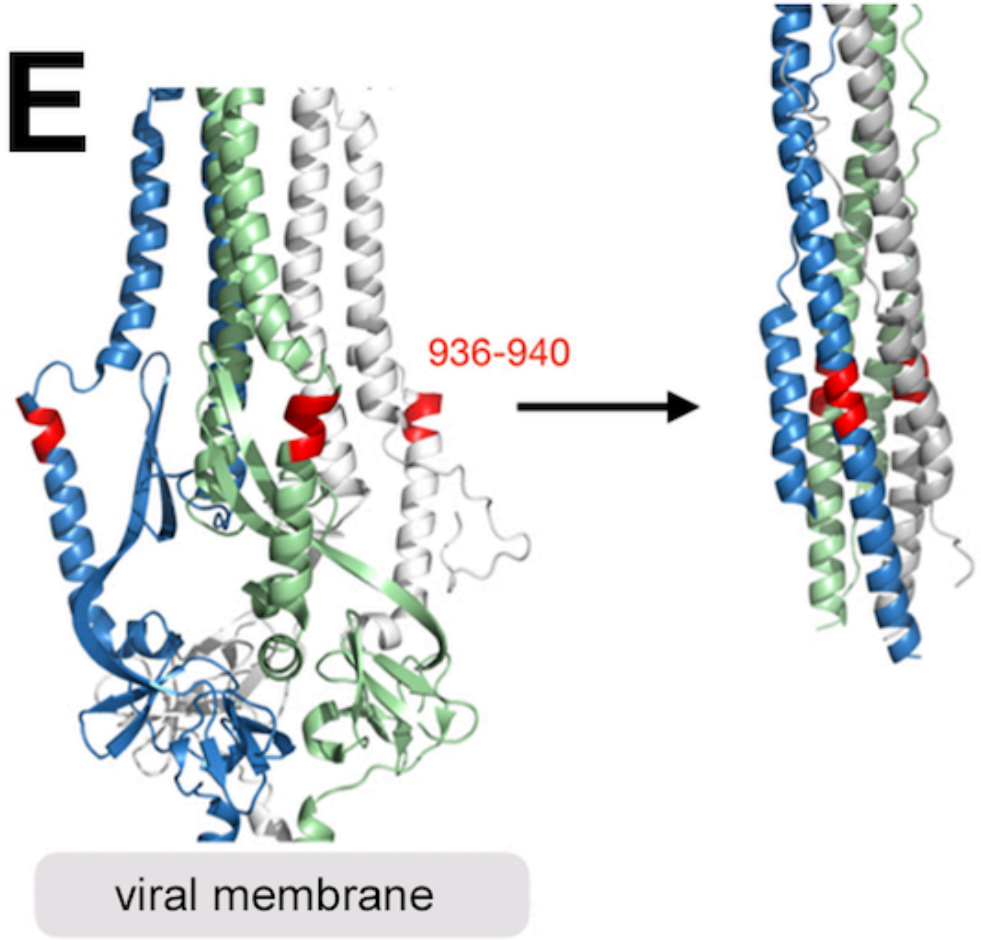
Variant	Count	%	Common Locations
G DSLSS			
– -----	58319	79.1	Global
D -----	14349	19.5	Early D614 form, Global
– Y----	735	1.0	Finland 179 (Uusimaa 179), Sweden 142 (Stockholm 34), UK 354 (Wales 132)
– ---F-	68	0.09	USA 48 (Utah 41)
– ----F	17	0.02	USA 6
– H----	16		
D ----F	14		
– --F--	11		
+ other rare variants			

1. Country lists noted below are not complete, just the most common
2. I'm not including ambiguous base calls that result uncertain amino acids
3. I'm not including very rare variants

HR1 D936Y

614 + Variable region:





From Korber et al:

"Variable cluster 936-940 (red), in the HR1 region of S_2. These residues occur in a region that undergoes conformational transition during fusion: pre-fusion (PDB:6VSB) and post-fusion (PDB: 6LXT) conformations of HR1 are shown, left and right."

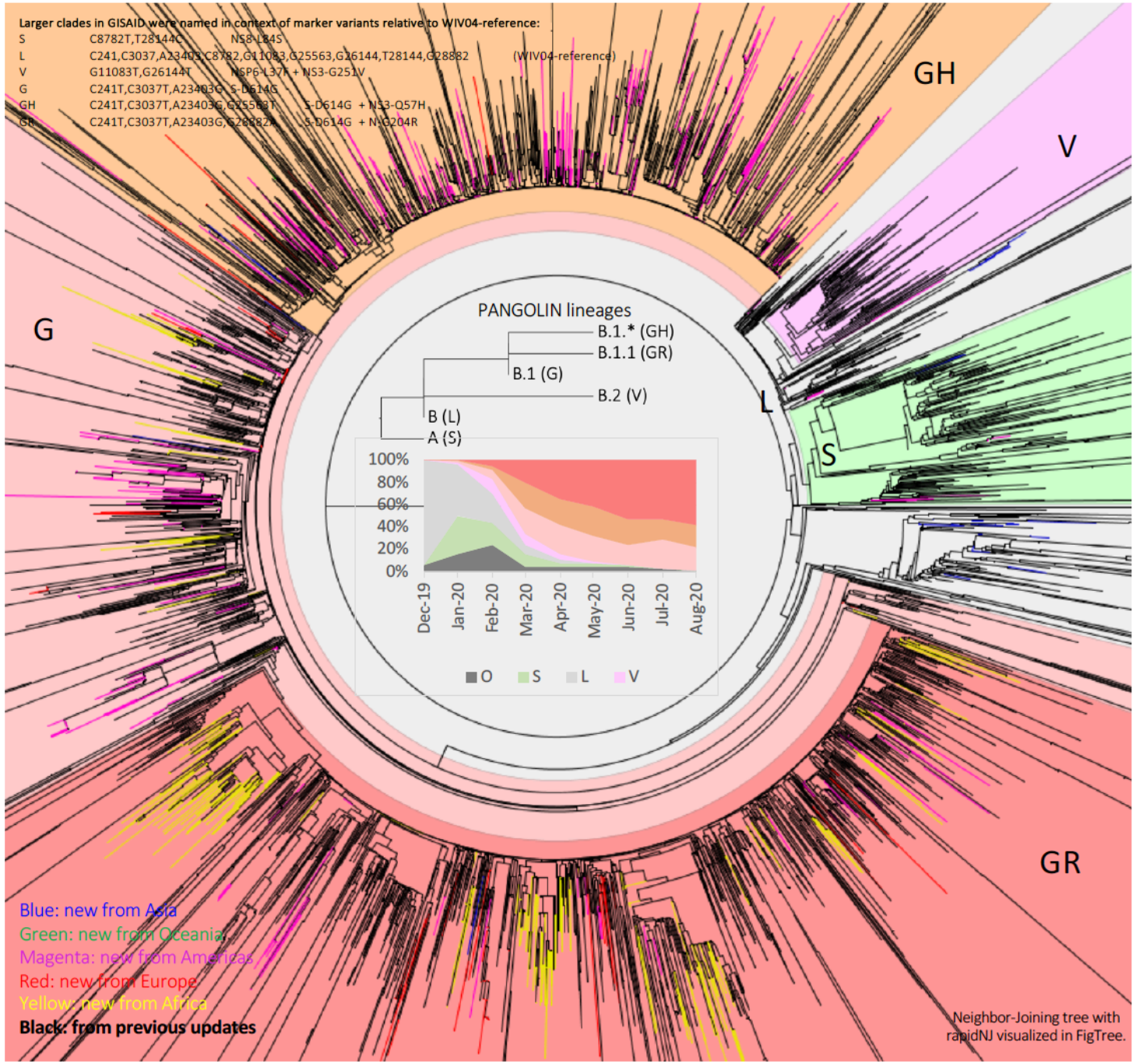
The O, S, L, and V clade are rarely sampled after June 1, during the summer of 2020

G has two sub-lineages, GR and GH.

GR is the most frequently sampled, but is very common in the UK which is highly sampled.

GH is also frequently sampled, and is common in the US.

L is complex, may include recombinants?



Full genome tree derived from all outbreak sequences 2020-08-25

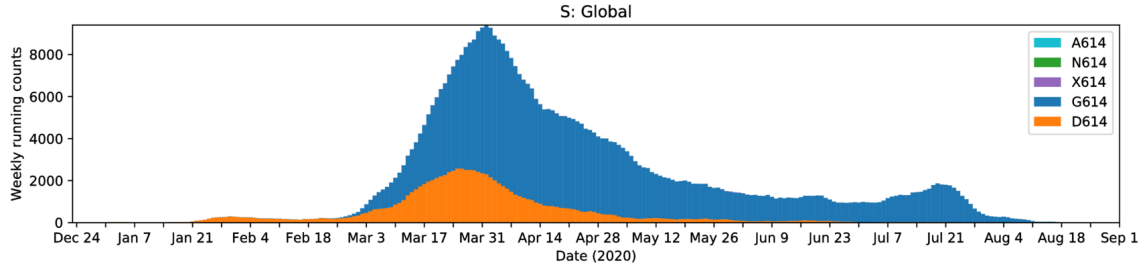
Notable changes:
77,909 full genomes (+971) (excluding low coverage, out of 84,426 entries)

Updated clades:
S clade 5,121 (+14)
L clade 3,869 (+7)
V clade 4,643 (+0)
G clade 18,101 (+227)
GR clade 24,764 (+372)
GH clade 17,850 (+324)
Other clades 3,561 (+27)

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

GISAID

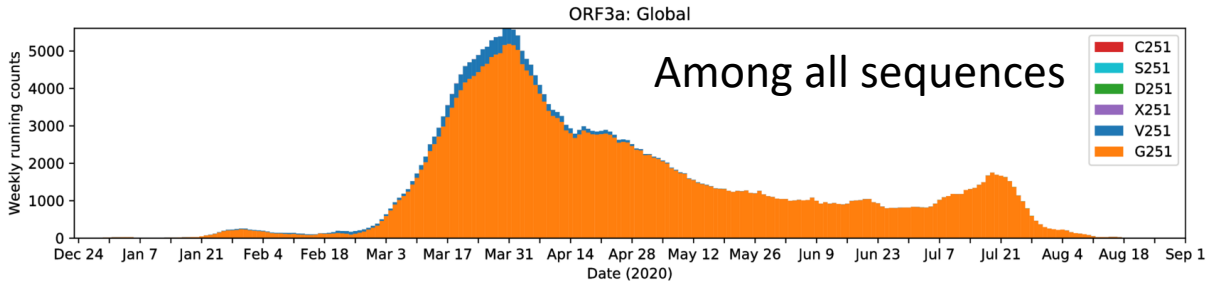
by BII/GIS, A*STAR Singapore

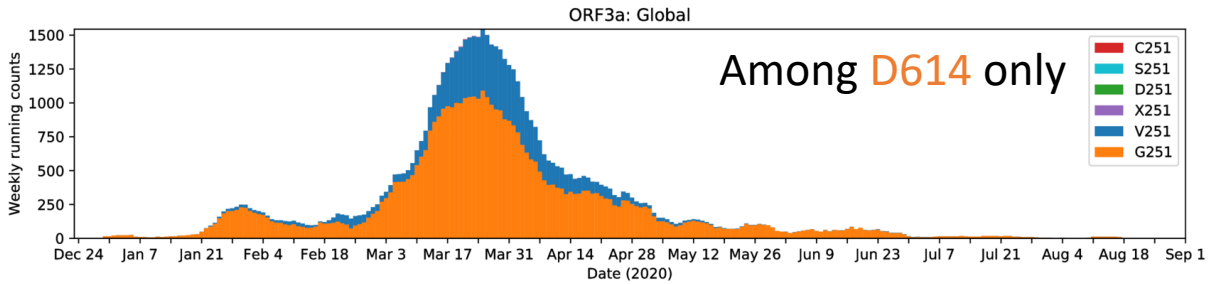


D614 is rarely sampled in GISAID outside of Singapore, summer 2020

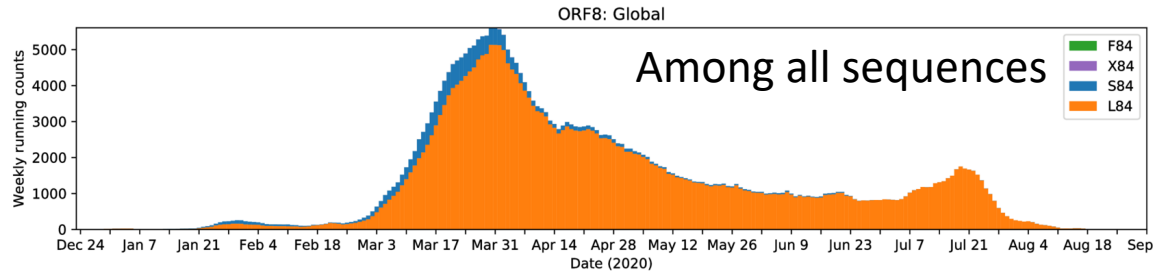
This also goes for the V clade and S clade, which carry both carry D614, are very rarely sampled.

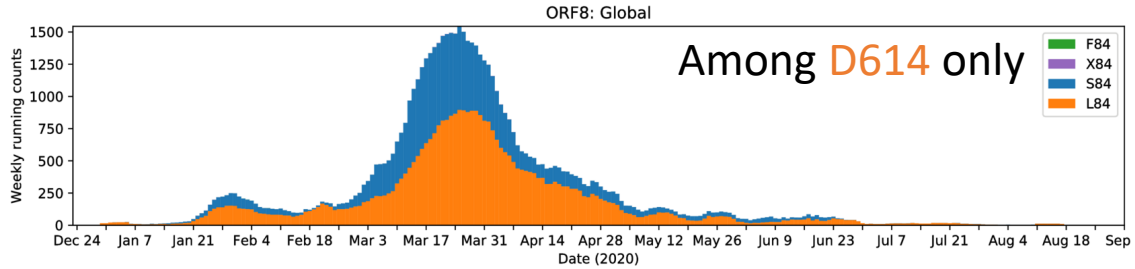
V clade





S clade



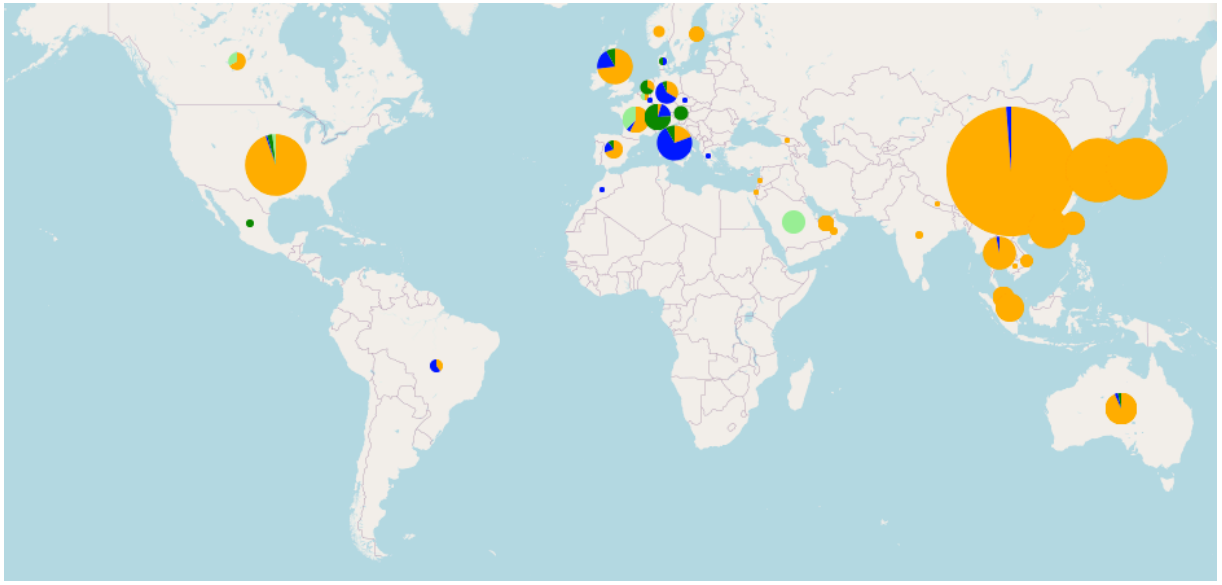


L clade GISAID, appears to be a mixture of other clades.

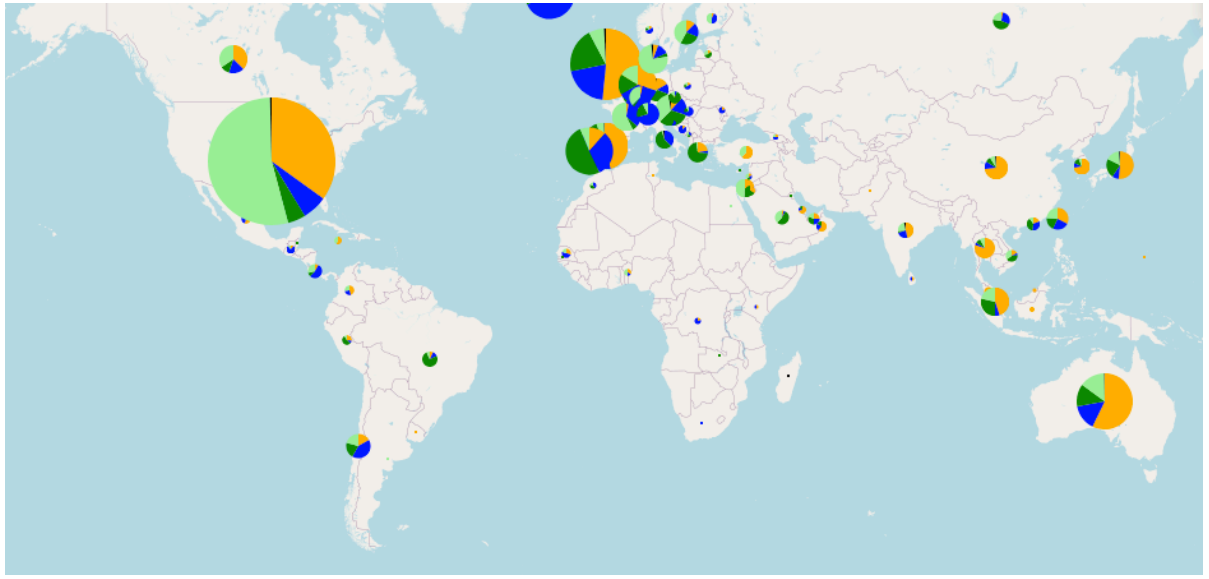
There seems to be a global tendency for the GR clade to be increasing, relative to GH and G.

Sampled Sept 2nd, 2020

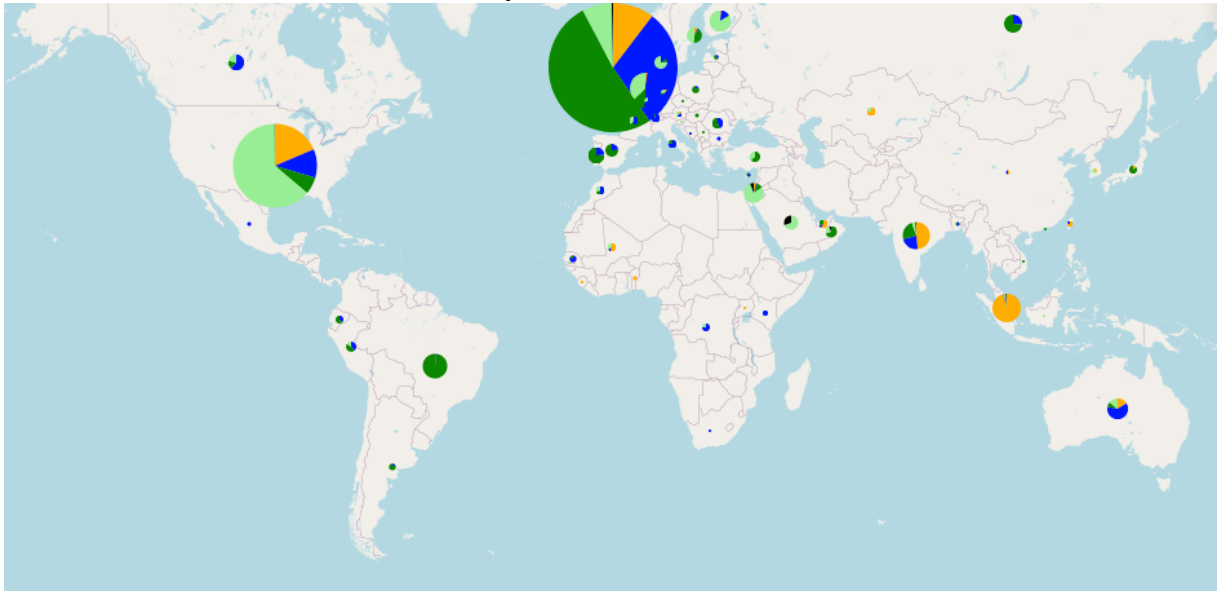
Prior to March 1



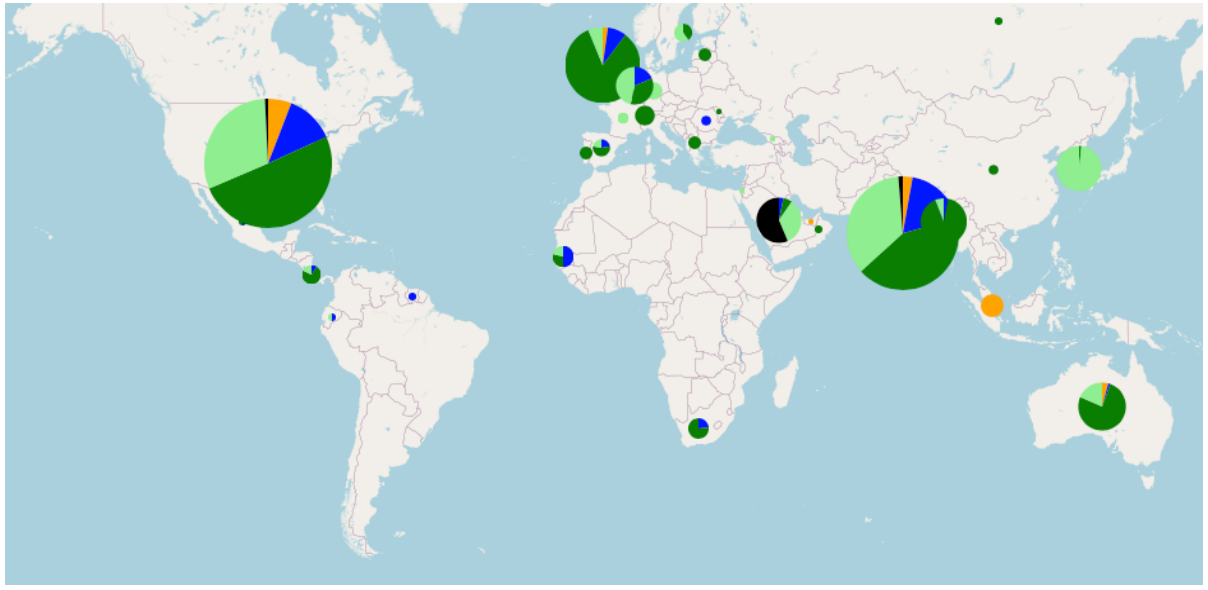
March 11-20



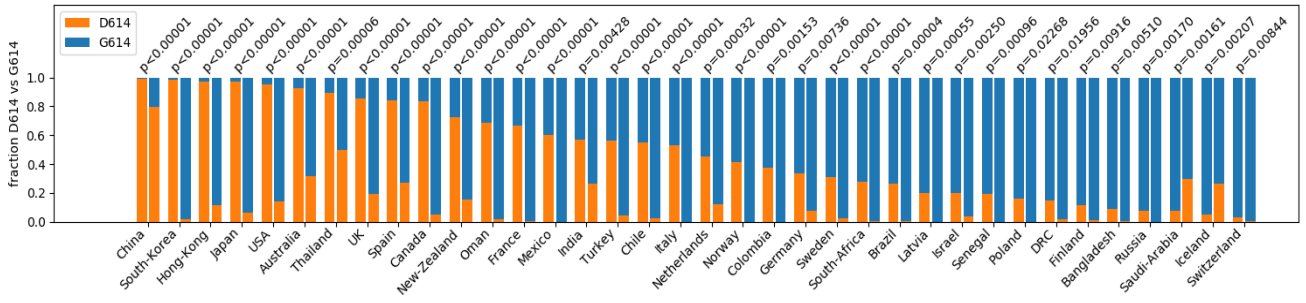
April 11-20



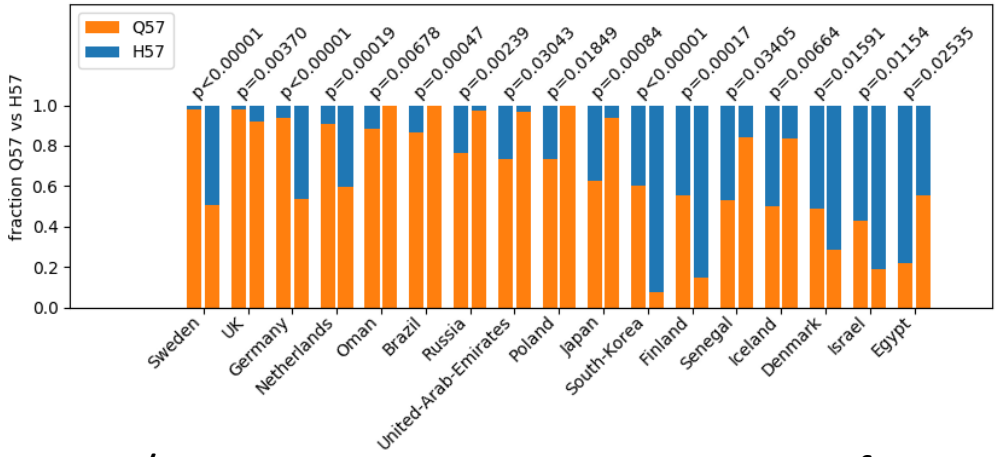
June 11-20



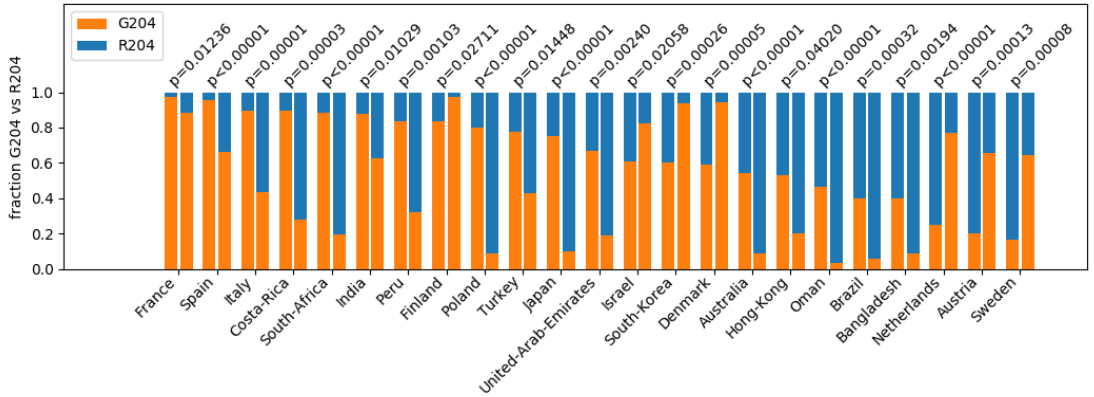
G clade: G614 increasing in frequency, 34/36 countries, p = 1.9e-08



GH subclade, no consistent pattern, is increasing in 8/17, p = 1

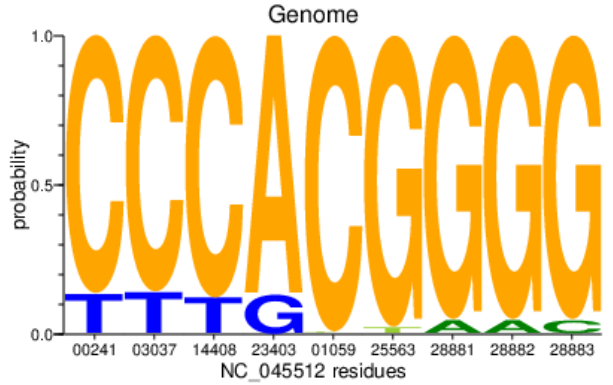


GR, 16/23 are increasing, p = 0.09, not significant, trend?

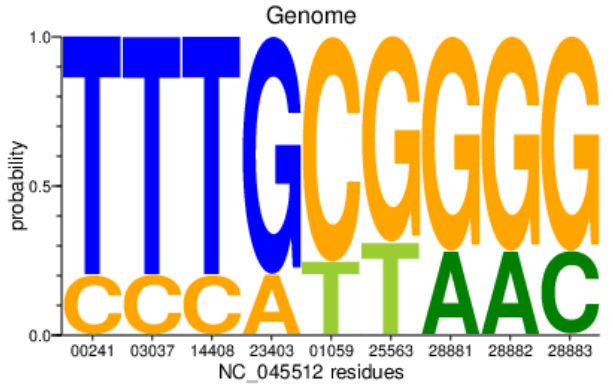


Near complete genomes, 36,388, Aug. 25, 2020

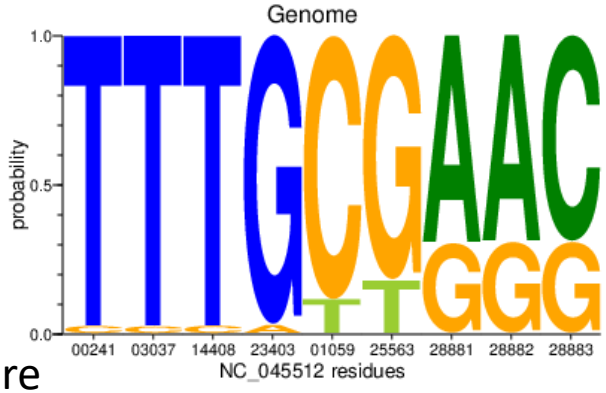
To Mar 1
N = 1,147



Mar 2 – Jun 1
N = 29,989



Jun 2 – Aug 31
N = 8,562



Yakima or Singapore

G GH GR

The mutations defining the G, GR, and GH clades tend to stay together

original CCCA: 19.59
G TTTG: 78.79
other: < 2%

Original CCCA_CG_GGG: 19.54

G	TTTG_CG_GGG	19.70	
GR	TTTG_CG_AAC:	30.59	**UK** USA
GH	TTTG_TT_GGG:	21.81	**USA** UK
GH	TTTG_CT_GGG:	6.12	**USA** UK

~98% of sequences

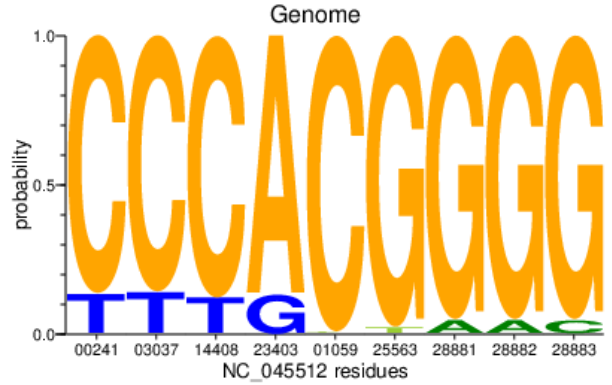
Example of a possible recombination event

GHR TTTG_TT_AAC 0.02

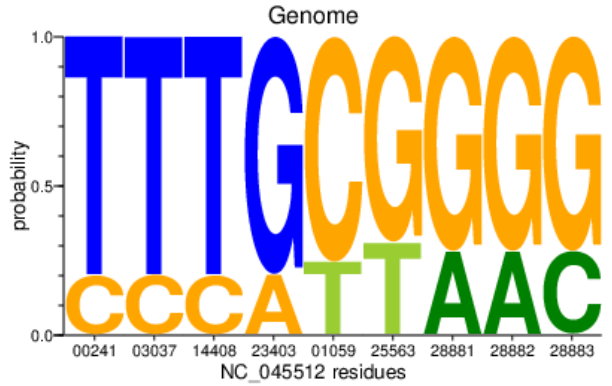
Belgium 1 Adenned
Isreal 1 Bat-Yam
USA S. Carolina 3, Washington Yakima 1

Near complete genomes, 36,388, Aug. 25, 2020

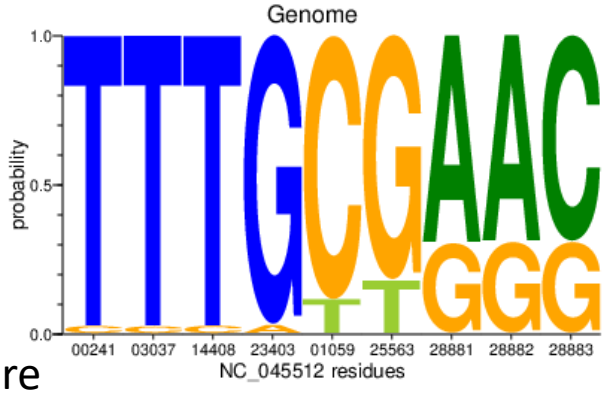
To Mar 1
N = 1,147



Mar 2 – Jun 1
N = 29,989

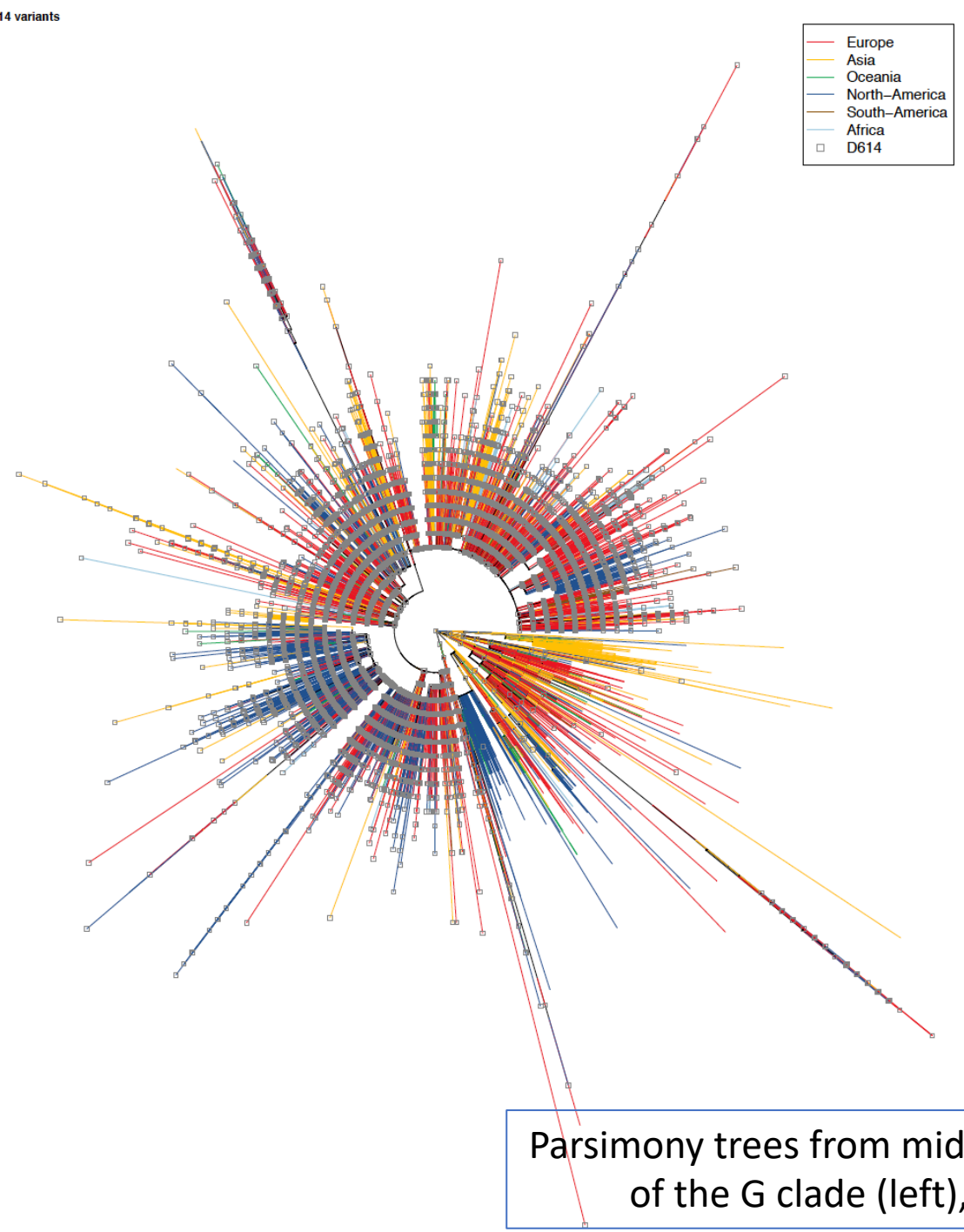


Jun 2 – Aug 31
N = 8,562



Yakima or Singapore

G GH GR



Parsimony trees from mid August, highlighting the relationships of the G clade (left), to the GR and GH clades (right)

Parsimony trees from mid August, highlighting the relationships of the G , GR, and GH clades to recurring mutations in Spike

GR clade origins

Three adjacent nucleotide changes spanning two residues in SARS-CoV-2 nucleoprotein: possible homologous recombination from the transcription-regulating sequence

Shay Leary, Silvana Gaudieri, Abha Chopra, Suman Pakala, Eric Alves, Mina John, Suman Das, Simon Mallal, Elizabeth Phillips bioRxiv 2020.04.10.029454; doi: https://doi.org/10.1101/2020.04.10.029454

From their abstract:

1) This new strain may have arisen by a form of homologous recombination from the core sequence (CS-B) of the transcription-regulating sequences of SAS-CoV-2 itself

2) It rapidly increased to approximately one third of reported sequences from Europe during the month of March 2020.

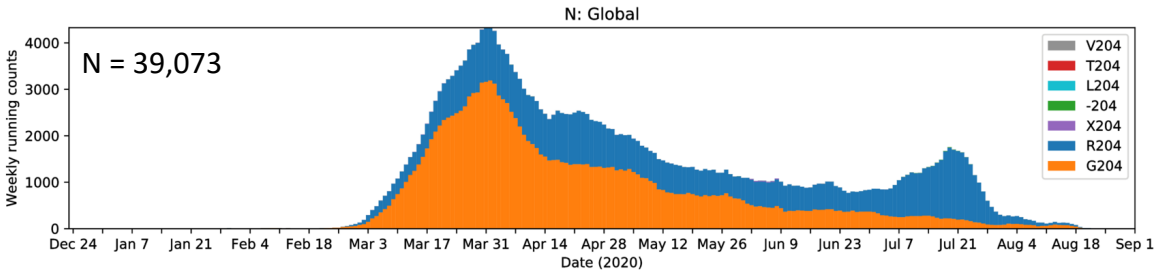
3) These polymorphisms are predicted to reduce the binding of an overlying putative HLA-C*07-restricted epitope and that HLA-C*07 is prevalent in Caucasians being carried by >40% of the population.

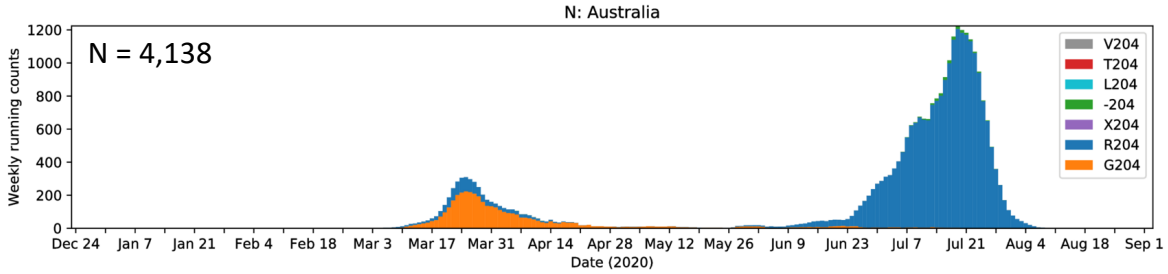
4) Homologous recombination may have occurred since its introduction into humans and be a mechanism for increased viral fitness and adaptation of SARS-CoV-2 to human populations.

The GR clade, represented by Nucleocapsid G204R, and Spike S477N are *both* part of the new dominant form found in Australia

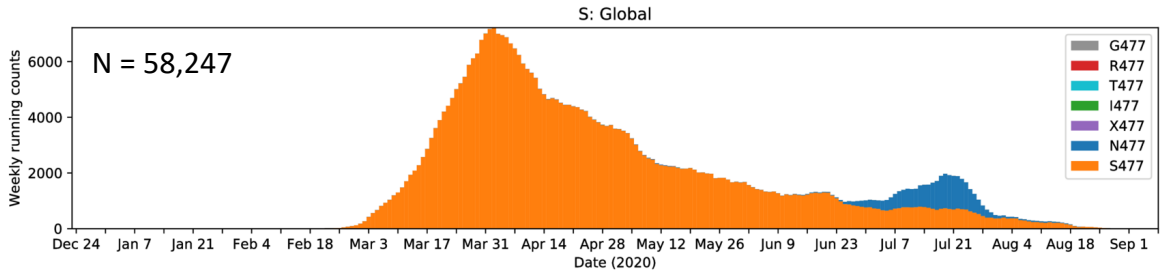
- The global “bump” in the GR clade corresponds to an infusion of sequences from Australia into GISAID over the summer.
- The shift in Australia corresponds to a shift the GR clade, and is due a the form of the virus that also carries Spike S477N
- If positive selection was occurring, it could be at either or both locations

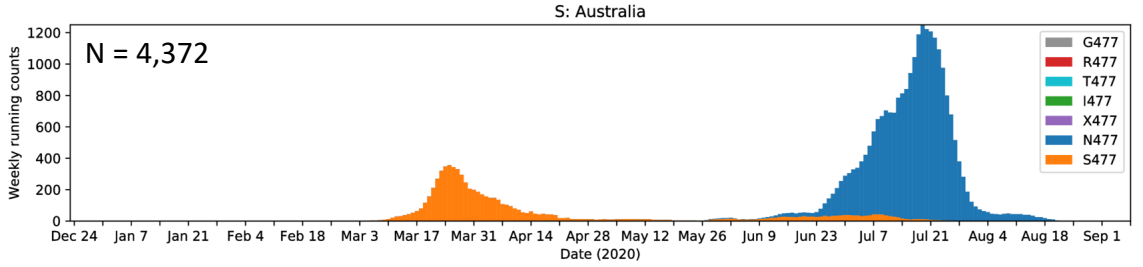
Full alignment, N G204R, representing the GR clade



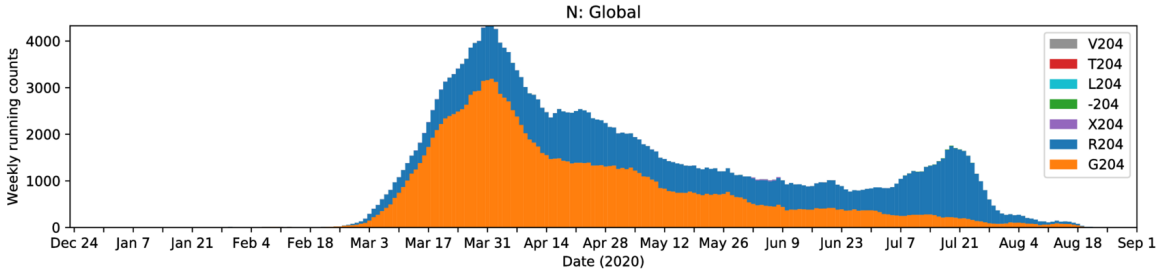


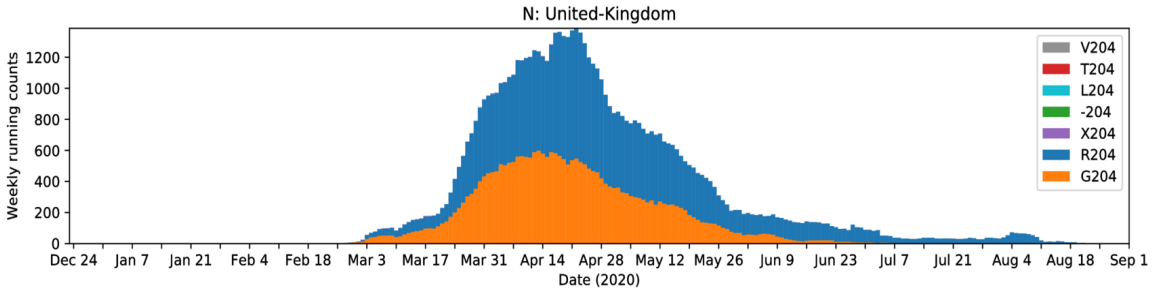
Spike alignment S S477N

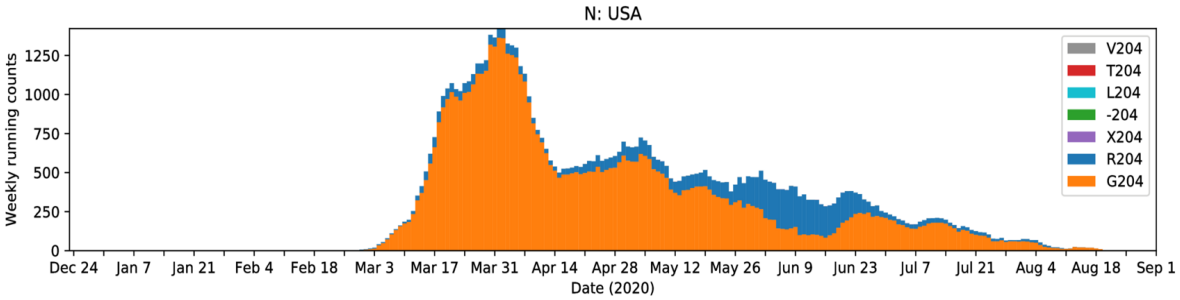


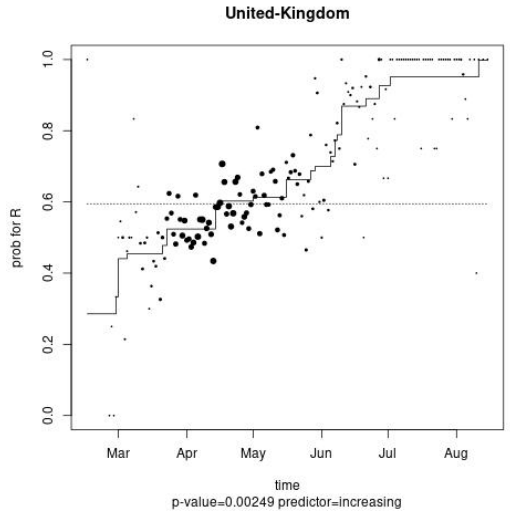


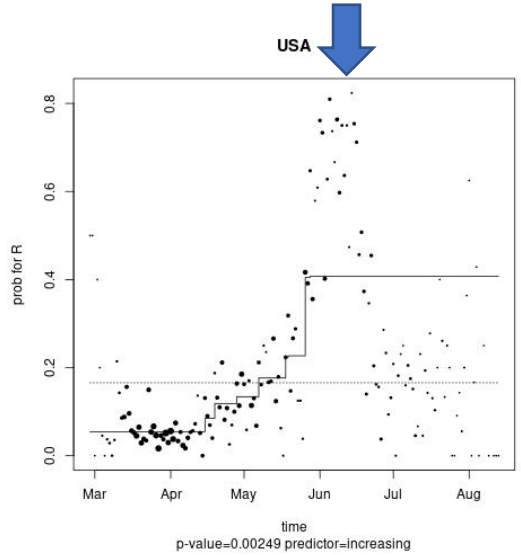
GR, Nucleocapsid G204R, tends to increase within the G614 set, but not always.



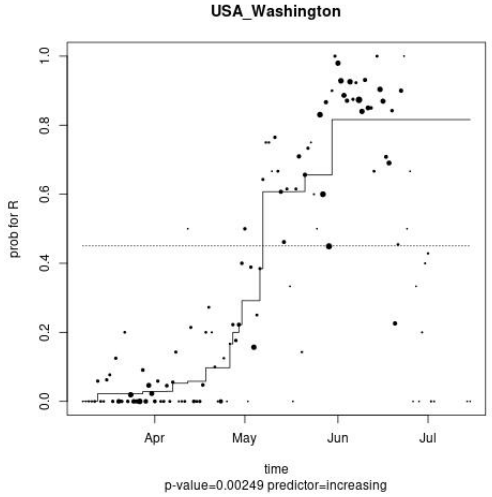






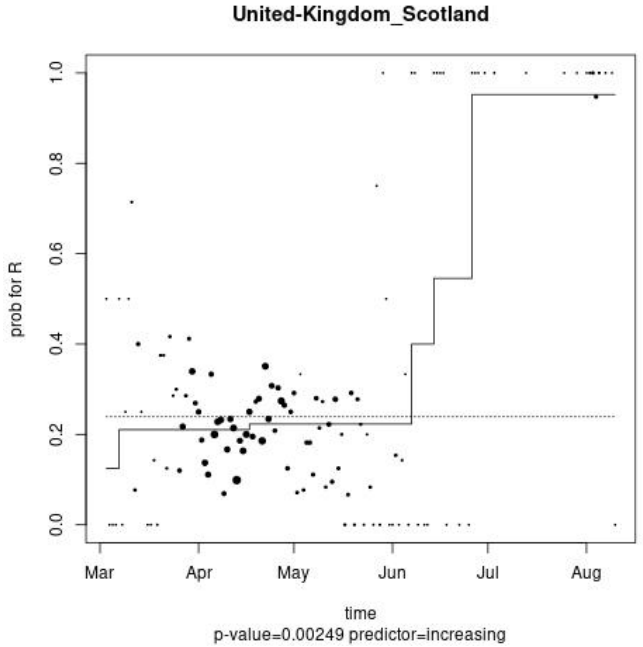


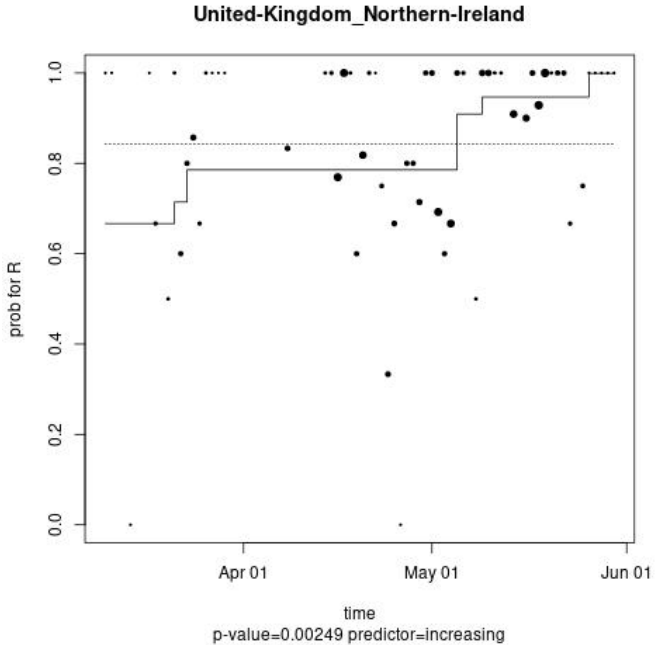
Peak and drop of in the USA may be due to intense sampling from Washington state during the period, where the GR clade is increasing

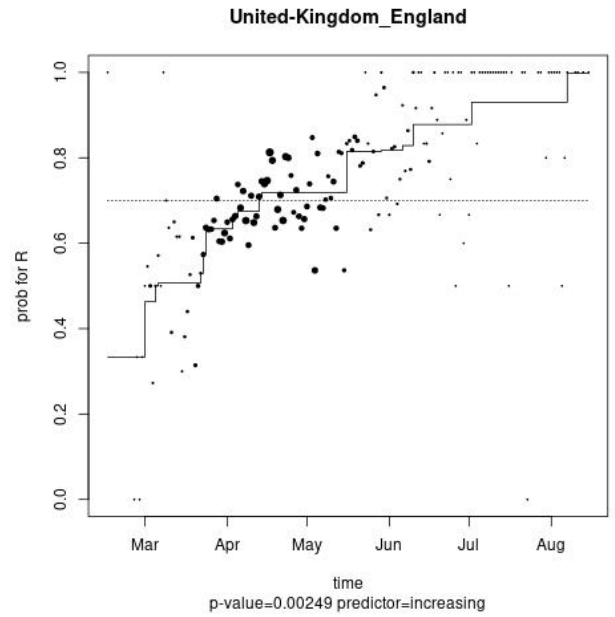


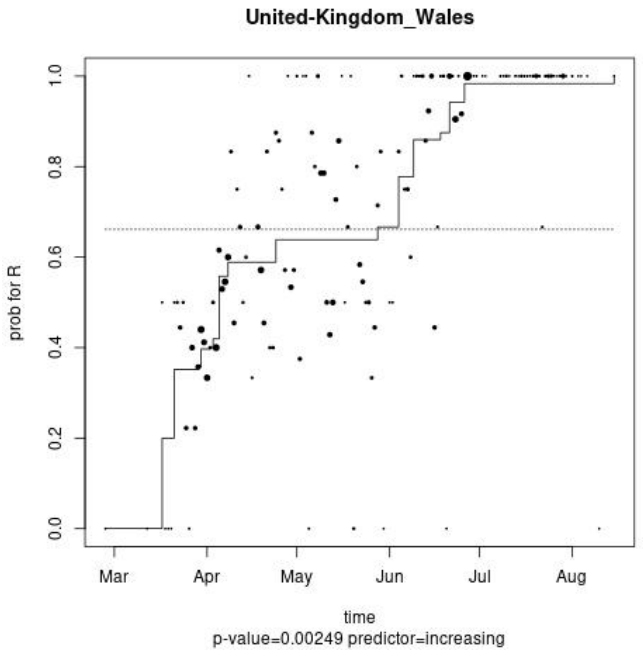
GR is the most common form in the UK, which is the most heavily sampled nation globally, so also most common globally.

The GR clade, represented by Nucleocapsid G204R, is increasing within the G clade G614 set in the UK:





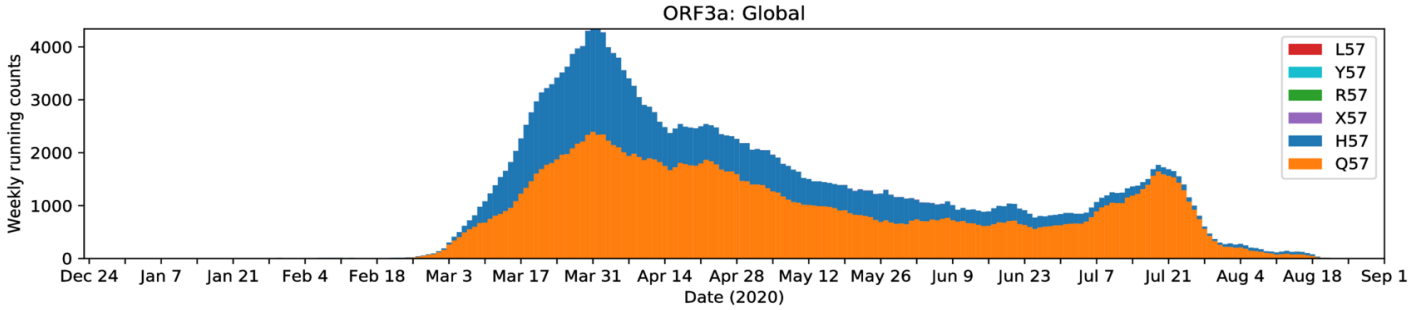


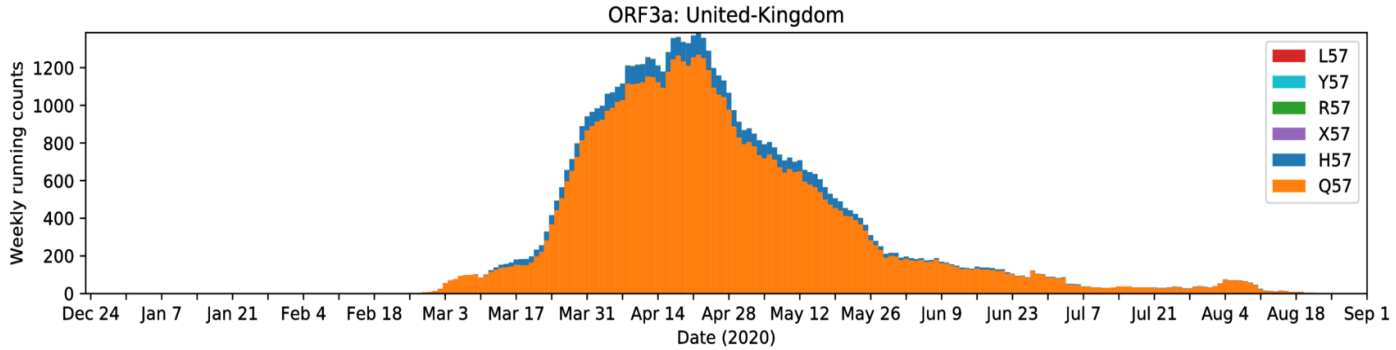


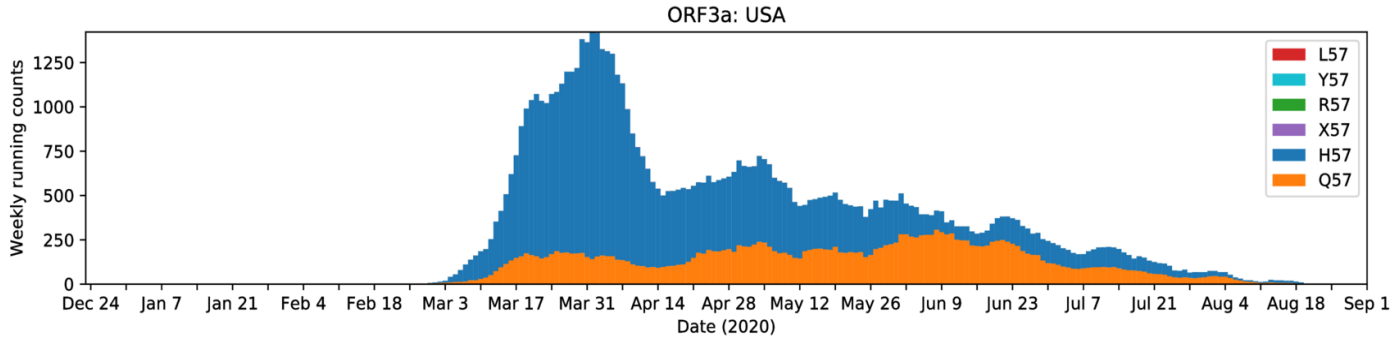
The pattern of increase seen in the UK is not always consistent globally,.

The UK is the most commonly sampled clade in GISAID, so the G clade's high frequency in the UK bias the global sample.

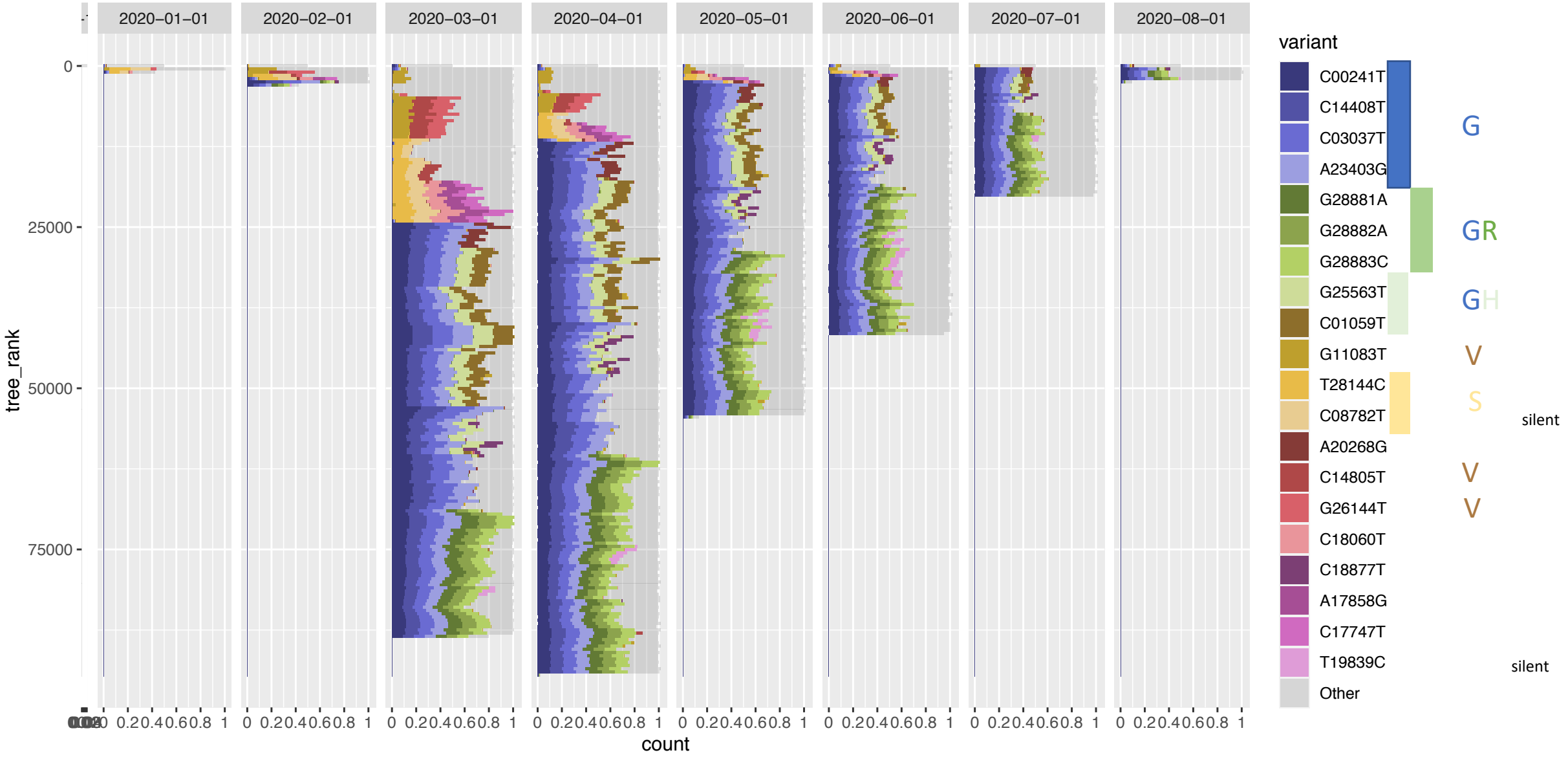
GH, may be diminishing within the G614 set.







The twenty most common base changes in the SARS-CoV-2 genome, organized in a “tree order”, so by clade, indicating frequencies by month. The GISAID common clade is indicated on the right.



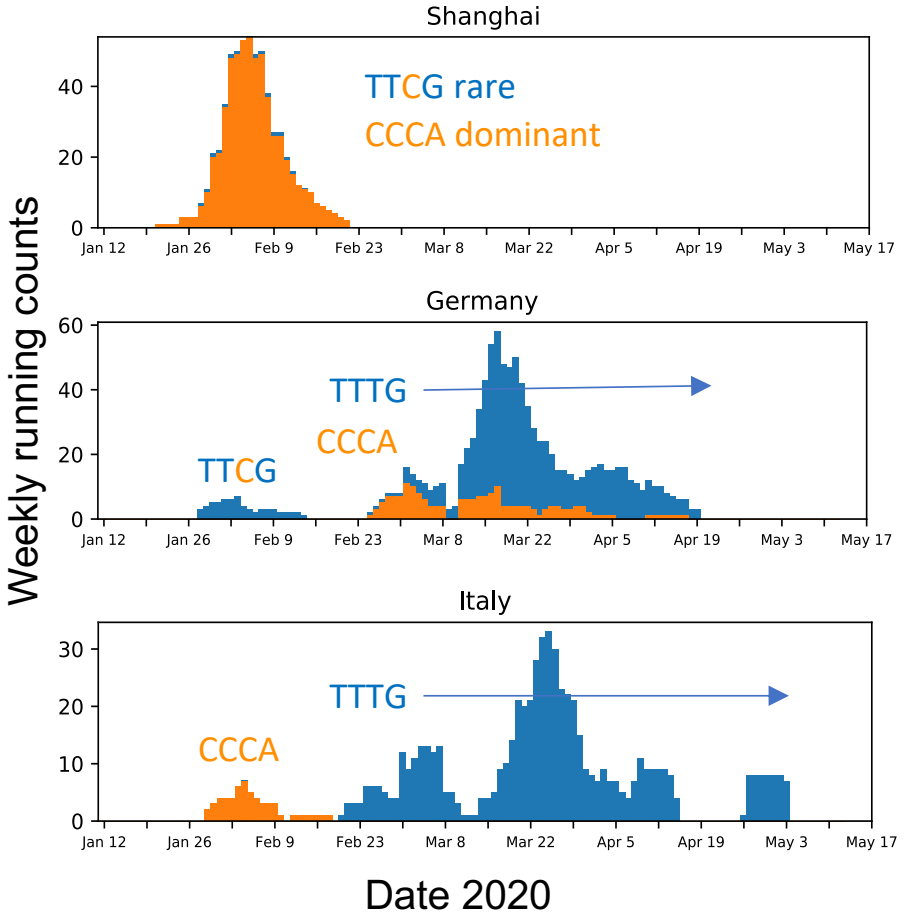
NOTE: this summary was made prior to the Spike 477 mutation infusion of sequences from Australia so does not show base G22992A.

Will Fischer

Origins: S D614G is almost always part of a clade defined by 4 bases

The 3 mutation form was present early in China and Germany, lacked RdRp P323L, and did not take off.
The 4 mutation form, including both the RdRp P323L mutation and the Spike D614G, did rapidly spread globally.
Thus it is possible that the RdRp mutation may have contributed to a selective advantage.

A



B

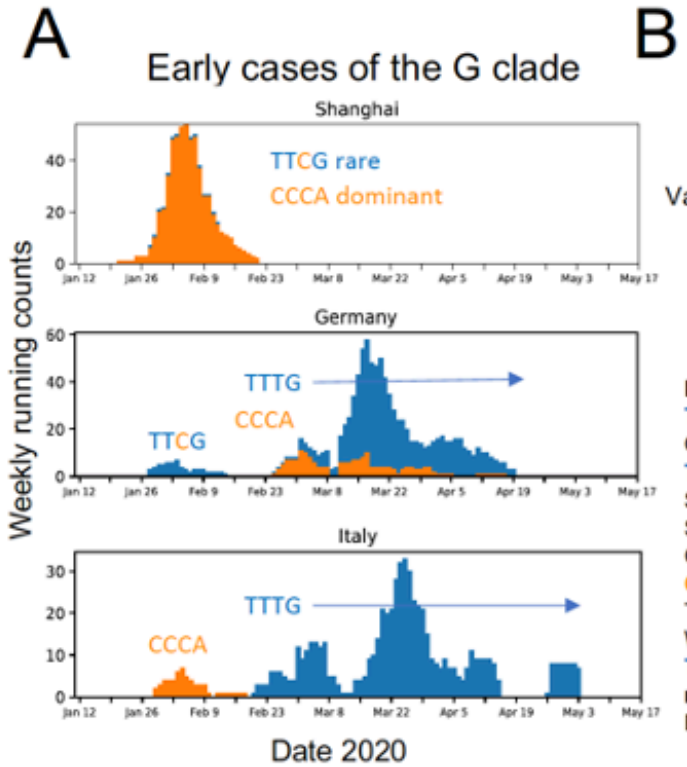
G-clade mutations (C3037T, C14408T, A23403G) CCA -> TTG
Plus the linked mutation in the UTR: C241T CCCA -> TTTG

11805	TTG	(72.03%)	9692	TTTG	(71.65%)
4582	CCA	(27.96%)	3835	CCCA	(28.35%)

Variants:

53	CTG		51	TCTG	5	CTCG
39	TCG		32	TTCG	4	CCTA
16	CCG		13	CTTG	3	TCTA
9	TTA	0.76%	11	TCCA	2	CTTA
8	CTA		9	TCCG	2	CTCA
5	TCA		7	CCCG	1	TTCA
1	ACA		6	TTTA	1	CCTG

Earliest examples in GISAID:
TTCG: Germany, Jan 2020: cluster of cases late Jan.-Feb.
One example: Germany/BavPat1/EPI_ISL_406862|2020-01-28
TTCG: Sampled several times in China, *e.g.*:
Sichuan/SC-PHCC1-022/EPI_ISL_451345|2020-01-24
Shanghai/SH0025/EPI_ISL_416334|2020-02-06
Guangzhou/GZMU0019/EPI_ISL_429080|2020-02-05
CCCG: Sampled twice in early Feb., Wuhan and Thailand
Thailand/Samut_prakarn_840/EPI_ISL_447919|2020-02-04
Wuhan/HBCDC-HB-06/EPI_ISL_412982|2020-02-07
TTTG: First identified in Italy; within 10 days sampled in many in countries in Europe, the USA, Mexico
First sample: Italy/CDG1/2020|EPI_ISL_412973|2020-02-20



B

G-clade mutations (C3037T, C14408T, A23403G) CCA -> TTG
Plus the linked mutation in the UTR: C241T CCCA -> TTTG

11805	TTG	(72.03%)	9692	TTTG	(71.65%)
4582	CCA	(27.96%)	3835	CCCA	(28.35%)

Variants:

53	CTG	51	TCTG	5	CTCG
39	TCG	32	TTCG	4	CCTA
16	CCG	13	CTTG	3	TCTA
9	TTA	11	TCCA	2	CTTA
8	CTA	9	TCCG	2	CTCA
5	TCA	7	CCCG	1	TTCA
1	ACA	6	TTTA	1	CCTG

Earliest examples in GISAID:

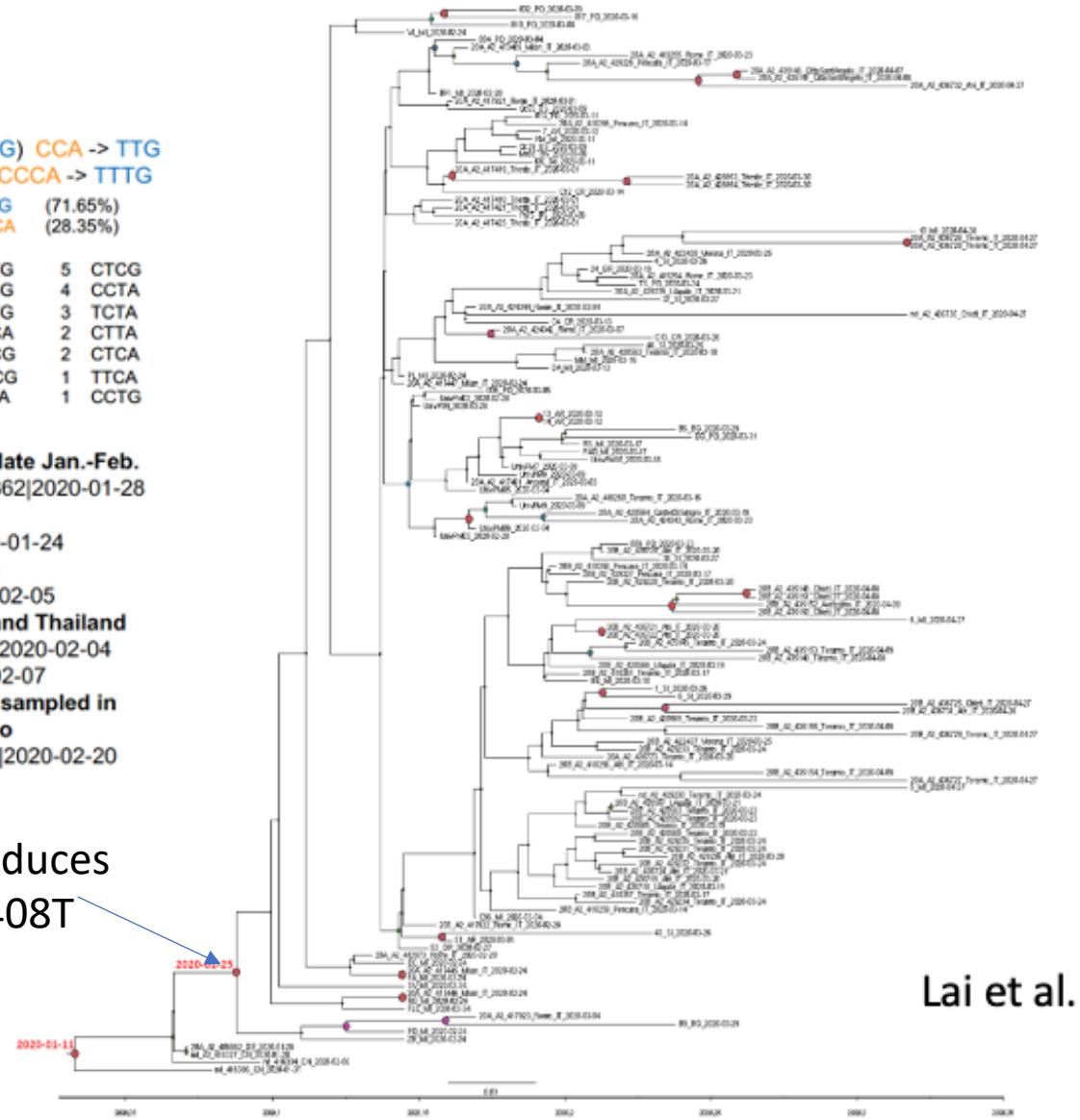
TTCG: Germany, Jan 2020: cluster of cases late Jan.-Feb.
One example: Germany/BavPat1/EPI_ISL_406862|2020-01-28

TTCG: Sampled several times in China, *e.g.*:
Sichuan/SC-PHCC1-022/EPI_ISL_451345|2020-01-24
Shanghai/SH0025/EPI_ISL_416334|2020-02-06
Guangzhou/GZMU0019/EPI_ISL_429080|2020-02-05

CCCG: Sampled twice in early Feb., Wuhan and Thailand
Thailand/Samut_prakarn_840/EPI_ISL_447919|2020-02-04
Wuhan/HBCDC-HB-06/EPI_ISL_412982|2020-02-07

TTTG: First identified in Italy; within 10 days sampled in many in countries in Europe, the USA, Mexico
First sample: Italy/CDG1/2020|EPI_ISL_412973|2020-02-20

Korber et al., Cell 2020, supplementary figure.



Lai et al.

Our findings are consistent with Lai et al., a history of the Italian epidemic

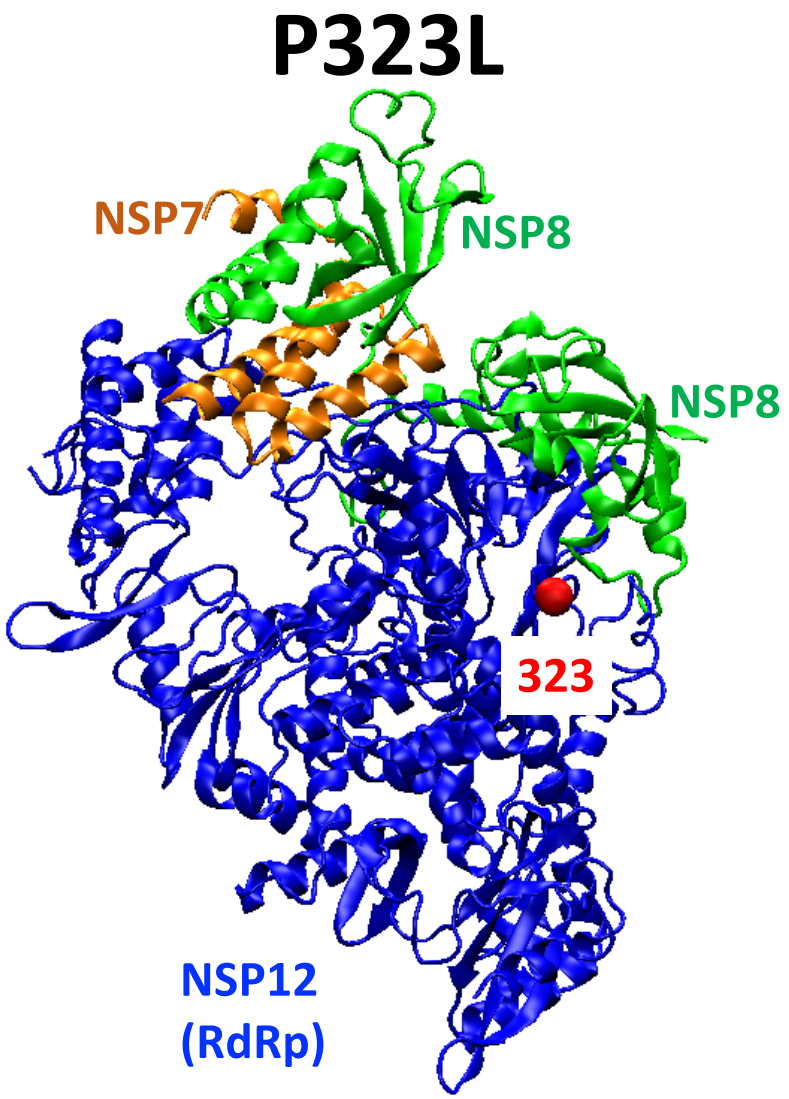
Figure 2. SARS-CoV-2 tree of 136 Italian strains plus one German and three Chinese isolates from Shanghai, showing statistically significant support for clades along the branches (posterior probability > 0.7).Large red and purple circles indicated highest posterior probability. Calendar dates of the tree root and the Italian clade were showed in red.

Lai A, Bergna A, Caucci S, Clementi N, Vicenti I, Dragoni F, et al. Molecular tracing of SARS-CoV-2 in Italy in the first three months of the epidemic. doi:10.1101/2020.07.06.20147140

We don’t know if RdRp P323L is functionally relevant.

The epidemiological evidence raises the possibility.

Here is where is located in the RdRp structure.



Away from the catalytic residues and Motifs A to G in palm domain Which are critical for activity

Proximal to NSP8 cofactor

Located in a hydrophobic cavity

Strains available at BEI Resources

1. O, S are declining, and no longer sampled much in GISAID.

2. Commonly used WA1 is a D614 form, and is an S Clade.

3. G614 is sampled in BEI as:
- BavPat1, which is missing the RdRp mutation,
- within the GH clade, which includes the RdRp mutation.

4. The GR clade is not available (at least at the time of this listing), but is becoming the most prevalent form globally.
- Perhaps an ancestral GR clade isolate would be a helpful reagent to include.

BEI Resources is prioritizing and fast tracking all SARS-CoV-2 registrations and orders. We anticipate a 12-72 hour turn-around time for all SARS-CoV-2 related registrations and a 24-48 hour turn-around time on approved orders. Please indicate SARS-CoV-2 in your scope of use in your registration paperwork. Please contact BEI Resources at contact@beiresources.org for questions.

BEI Resources is working to accession strains of the 2019 novel coronavirus, recently named SARS-CoV-2, identified as the causative agent of an outbreak of viral pneumonia, COVID-19. We understand how important it is to share virus strains and derivatives with researchers, especially during an outbreak.

Currently Available SARS-CoV-2 Materials

BEI Number	Description	Lineage	GISAID Clade	GISAID ID	Clinical Information Available	Registration
Virus						
NR-52281	SARS-CoV-2, Isolate USA-WA1/2020	A	S	EPI_ISL_404895	Male in 30s, returning traveler from Wuhan. Mild disease; recovered.	BEI Level 3
NR-52282	SARS-CoV-2, Isolate Hong Kong/VM20001061/2020	A	S	EPI_ISL_412028	Isolated from a nasopharyngeal aspirate and throat swab from an adult male patient on January 22, 2020 in Hong Kong	BEI Level 3
NR-52284	SARS-CoV-2, Isolate Italy-INMI1	None	O	EPI_ISL_406959 (fragment)	Isolated from sputum of a patient with a respiratory illness who had recently returned from travel to the affected region of China and developed clinical disease (COVID-19) in January 2020 in Rome, Italy.	BEI Level 3
NR-52359	SARS-CoV-2, Isolate England/02/2020	A	S	EPI_ISL_407073	39 yr old Male; Isolated from Nasopharyngeal aspirate & Throat swab	BEI Level 3
NR-52368	SARS-CoV-2, Isolate New York 1-PV08001/2020	B.4	O	EPI_ISL_414476	39 yr old Female; history of travel to Iran	BEI Level 3
NR-52369	SARS-CoV-2, Isolate Singapore/2/2020	B	L	EPI_ISL_407987	Isolated from a throat swab. Patient has respiratory illness, fever and cough	BEI Level 3
NR-52370	SARS-CoV-2, Isolate Germany/BavPat1/2020	B	G	EPI_ISL_406862	Isolated from Nasopharyngea swab. Typical symptoms of mild upper respiratory tract disease (D614G mutation)	BEI Level 3
NR-52381	SARS-CoV-2, Isolate USA-IL1/2020	B	O	EPI_ISL_404253	63 yr old Female; Isolated from sputum	BEI Level 3
NR-52382	SARS-CoV-2, Isolate USA-CA1/2020	A	S	EPI_ISL_406034	38 yr old Male; Isolated from nasopharyngeal swab	BEI Level 3
NR-52383	SARS-CoV-2, Isolate USA-AZ1/2020	A	S	EPI_ISL_406223	26 yr old Male; Isolated from bucal swab	BEI Level 3
NR-52384	SARS-CoV-2, Isolate USA-WI1/2020	B	L	EPI_ISL_408670	52 yr old Female; Isolated from nasopharyngeal swab	BEI Level 3
NR-52385	SARS-CoV-2, Isolate USA-CA3/2020	B	L	EPI_ISL_408008	72 yr old Female; Isolated from oropharyngeal swab	BEI Level 3
NR-52386	SARS-CoV-2, Isolate USA-CA4/2020	B	L	EPI_ISL_408009	57 yr old Male; Isolated from nasopharyngeal swab	BEI Level 3
NR-52387	SARS-CoV-2, Isolate USA-CA2/2020	B.2	O	EPI_ISL_406036	54 yr old Male; Isolated from nasopharyngeal swab	BEI Level 3
NR-52439	SARS-CoV-2, Isolate Chile/Santiago_op4d1/2020	A.2	S	EPI_ISL_415661	Isolated from a Nasal Swab. Patient has respiratory tract infection. History of travel to Europe	BEI Level 3
NR-53514	SARS-CoV-2, Isolate New York-PV08410/2020	B.1	GH	EPI_ISL_421374	63 yr old Male; severe COVID19 with fatal outcome	BEI Level 3
NR-53515	SARS-CoV-2, Isolate New York-PV08449/2020	B.1	GH	EPI_ISL_421400	88 yr old Female; severe COVID19 with fatal outcome	BEI Level 3
NR-53516	SARS-CoV-2, Isolate New York-PV09158/2020	B.1.3	GH	EPI_ISL_422525	62 yr old Male; severe COVID19 with fatal outcome	BEI Level 3
NR-53517	SARS-CoV-2, Isolate New York-PV09197/2020	B.1.3	GH	EPI_ISL_422552	90 yr old Male; severe COVID19 with fatal outcome	BEI Level 3

All mutations relative to the Wuhan reference strain found in the BEI resources listing:

Shaded boxes are associated with common clades and are lined up.

Non-shaded boxes include rare mutations that occur between the common clade mutations, all mutations are listed but these are not lined up

		Clade																										
BEI Reference strains	**Lineage**	**GISAID**	**UTR**		**nsp2**		**silent**		**silent**		**RdRp**			**Spike**		**ORF3A**			**ORF8**			**N**	**N**	**N**		
					T85I						**P323L**			**D614G**		**Q57H**			**L84S**			**R203K**	**R203K**	**G204R**		
hCoV-19/USA/WA1/2020|EPI_ISL_404895|2020-01-19	A	**S**							c08782t			c18060t							t28144c									
hCoV-19/Germany/BavPat1/2020|EPI_ISL_406862|2020-01-28	B	**G**	c00241t				c03037t							a23403g														
hCoV-19/USA/NY-PV08410/2020|EPI_ISL_421374|2020-03-16	B.1	**GH**	c00241t		c01059t		c03037t				c14408t			a23403g		g25563t												
hCoV-19/USA/NY-PV08449/2020|EPI_ISL_421400|2020-03-17*	B.1	**GH**	c00241t		c01059t		c03037t			c10851t	c14408t			a23403g		g25563t												
hCoV-19/USA/NY-PV09158/2020|EPI_ISL_422525|2020-03-22	B.1.3	**GH**	c00241t		c01059t		c03037t			c11916t	c14408t	c18998t		a23403g		g25563t											g29540a	
hCoV-19/USA/NY-PV09197/2020|EPI_ISL_422552|2020-03-20	B.1.3	**GH**	c00241t		c01059t		c03037t			c11916t	c14408t	c18998t	g22225a	a23403g		g25563t											g29540a	
hCoV-19/Chile/Santiago_op4d1/2020|EPI_ISL_415661|2020-03-08									c08782t	t09477a		c14805t					g25979t		t28144c	c28657t	c28863t							
hCoV-19/England/02/2020|EPI_ISL_407073|2020-01-29	A	**S**							c08782t			t18488c			t23605g				t28144c							a29596g		
hCoV-19/Hong_Kong/VM20001061-2/2020|EPI_ISL_412028|2020-01-22	A	**S**				c01663t			c08782t			g22661t					t26729c	g28077c	t28144c									
hCoV-19/USA/AZ1/2020|EPI_ISL_406223|2020-01-22	A	**S**							c08782t	g11083t									t28144c							c29095t		
hCoV-19/USA/CA1/2020|EPI_ISL_406034|2020-01-23	A	**S**				g01548a			c08782t						c24034t		t26729c	g28077c	t28144c	a28792t								
hCoV-19/USA/CA2/2020|EPI_ISL_406036|2020-01-22	B2	**O**								c17000t							g26144t											
hCoV-19/USA/NY1-PV08001/2020|EPI_ISL_414476|2020-02-29	B.4	**O**				g01397a		g03242a												t28688c						g29027t	g29742t	
hCoV-19/USA/CA3/2020|EPI_ISL_408008|2020-01-29	B	**L**		g00614a				a05084g												c28854t								
hCoV-19/USA/CA4/2020|EPI_ISL_408009|2020-01-29	B	**L**		g00614a				a05084g												c28854t								
hCoV-19/Singapore/2/2020|EPI_ISL_407987|2020-01-25	B	**L**															g27147c											
hCoV-19/USA/WI1/2020|EPI_ISL_408670|2020-01-31										c17373t																		
Ancestral form of common clades		**G+**	c00241t				c03037t				c14408t			a23403g														
		GH	c00241t		c01059t		c03037t				c14408t			a23403g		g25563t												
		GR	c00241t				c03037t				c14408t			a23403g								g28881a	g28881a	g28881c				
		S							c08782t										t28144c									

Shaded boxes with a mutation indicate common muations associated with a clade

* hCoV-19/USA/NY-PV08449/2020|EPI_ISL_421400|2020-03-17 has a poor sequence: 1052 N's, 168 gaps

Commonly used forms and two alternatives are the top 4:

USA/WA1: is an S clade, a good representative of the early form of the virus

BavPat1: is an ancestral G clade that carries the D614G Spike, but it missing the RdRp 4th mutation.

- We do not know if this matters, it may not, there is a hint in the global sequence data that it might

2 Alternatives that both include the RdRp mutation that was part of the G expansion

hCoV-19/USA/NY-PV08410/2020|EPI_ISL_421374|2020-03-16

hCoV-19/USA/NY-PV08449/2020|EPI_ISL_421400|2020-03-17*

Isolate reagent choices?

- Given that the RdRp mutation may be relevant for fitness, it might be good to have it included in live viral challenge stocks. Among the current BEI catalogue this would be choices that include the all four mutations that define the G clade expansion, as well as a few additional mutations, as they are GH clade:

  hCoV-19/USA/NY-PV08410/2020|EPI_ISL_421374|2020-03-16
  hCoV-19/USA/NY-PV08449/2020|EPI_ISL_421400|2020-03-17*

  * This sequence, NY-PV08449, was missing many bases: 1052 N's, 168 gaps.
  NY-PV08410 was fully sequenced, and had one amino acid change that is rare, but we have no particular reason to be concerned about. I would use NY-PV08410.

- Given that the GR clade is globally highly prevalent, it might be worth including a GR ancestral stock among the BEI options. We can provide a list of GISAID sequences that have all of the 7 base mutations of the ancestral GR clade, but no additional rare mutation.

- The Spike S477N change emerging in Australia is worth further investigation
  - Potentially beneficial in terms of ACE2 binding and expression
  - Potential escape
  - Dominant form in Australia, and currently sampled other places as well.