SARS-CoV-2 Spike protein variant D614G increases infectivity and retains sensitivity to antibodies that target the receptor binding domain

Leonid Yurkovetskiy[1], Kristen E. Pascal[2], Christopher Tomkins-Tinch[3,4], Thomas Nyalile[1], Yetao Wang[1], Alina Baum[2], William E. Diehl[1], Ann Dauphin[1], Claudia Carbone[1], Kristen Veinotte[1], Shawn B. Egri[1], Stephen F. Schaffner[3,4], Jacob E. Lemieux[3,7], James Munro[5,6], Pardis C. Sabeti[3,4,8,9,10,11], Christos A. Kyratsous[2,11], Kuang Shen[1,11], Jeremy Luban[1,3,6,10,11]

[1]Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA 01605, USA
[2]Regeneron Pharmaceutical, Inc., 777 Old Saw Mill River Rd, Tarrytown, NY 10591
[3]Broad Institute of Harvard and MIT, 75 Ames Street, Cambridge, MA 02142, USA
[4]Harvard University, 52 Oxford Street, Cambridge, MA 02138, USA
[5]Department of Microbiology and Physiological Systems, University of Massachusetts Medical School, Worcester, MA 01605, USA
[6]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA
[7]Massachusetts General Hospital, 55 Fruit Stree, Boston, MA, 02114
[8]Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, 02115 Boston, MA
[9]Howard Hughes Medical Institute, 4000 Jones Bridge Rd, Chevy Chase, MD 20815
[10]Massachusetts Consortium on Pathogen Readiness, Boston, MA, 02115

[11]Correspondence:
Pardis.Sabeti@broadinstitute.org (P.C.S.), christos.kyratsous@regeneron.com (C.A.K.), Kuang.Shen@umassmed.edu (K.S.), jeremy.luban@umassmed.edu (J.L.)

Abstract

Virus genome sequence variants that appear over the course of an outbreak can be exploited to map the trajectory of the virus from one susceptible host to another. While such variants are usually of no functional significance, in some cases they may allow the virus to transmit faster, change disease severity, or confer resistance to antiviral therapies. Since the discovery of SARS-CoV-2 as the cause of COVID-19, the virus has spread around the globe, and thousands of SARS-CoV-2 genomes have been sequenced. The rate of sequence variation among SARS-CoV-2 isolates is modest for an RNA virus but the enormous number of human-to-human transmission events has provided abundant opportunity for selection of sequence variants. Among these, the SARS-CoV-2 Spike protein variant, D614G, was not present in the presumptive common ancestor of this zoonotic virus but was first detected in late January in Germany and China. The D614G variant steadily increased in frequency and now constitutes >97% of isolates world-wide, raising the question whether D614G confers a replication advantage to SARS-CoV-2. Structural models predict that D614G would disrupt contacts between the S1 and S2 domains of the Spike protein and cause significant shifts in conformation. Using single-cycle vectors we showed that D614G is three to nine-fold more infectious than the ancestral form on human lung and colon cell lines, as well as on other human cell lines rendered permissive by ectopic expression of human ACE2 and TMPRSS2, or by ACE2 orthologues from pangolin, pig, dog, or cat. Nonetheless, monoclonal antibodies targeting the receptor binding domain of the SARS-CoV-2 Spike protein retain full neutralization potency. These results suggest that D614G was selected for increased human-to-human transmission, that it contributed to the rapidity of SARS-CoV-2 spread around the world, and that it does not confer resistance to antiviral therapies targeting the receptor binding domain.

Introduction

Next-generation sequencing permits real-time detection of genetic variants that appear in pathogens during disease outbreaks. Tracking viral variants now constitutes a requisite component of the epidemiologist's toolkit that may pinpoint the origins of a zoonotic virus and the trajectory it takes from one susceptible host to another (Hadfield et al., 2018; Shu and McCauley, 2017). Lagging behind sequence-based modeling of virus phylogeny and transmission chains is the ability to understand the effect of viral variants on the efficiency of transmission to new hosts or on the clinical severity of infection. Most sequence variants that arise during virus replication are either detrimental to the fitness of the virus or without consequence. Even so, such variants can increase in frequency over the course of an outbreak by chance (Grubaugh et al., 2020). More rarely, though, increasing frequency of a variant could reflect competitive advantage due to higher intrinsic replication capacity with increased viral load and transmissibility.

In December 2019, an outbreak of unexplained fatal pneumonia became apparent in Wuhan City, Hubei Province, China. By early January 2020, SARS-CoV-2 was identified as the virus causing the disease (Huang et al., 2020; Lu et al., 2020; Wu et al., 2020a, 2020b; Zhou et al., 2020b; Zhu et al., 2020). After SARS-CoV (Drosten et al., 2003; Ksiazek et al., 2003) and MERS-CoV (Zaki et al., 2012), SARS-CoV-2 is the third human coronavirus this century known to cause pneumonia with a significant case-fatality rate (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020). Hundreds of coronaviruses have been identified in bats, including at least 50 SARS-like *Sarbecoviruses* (Lu et al., 2020; Zhou et al., 2020a). The virus closest in sequence to SARS-CoV-2 observed to date was isolated from a bat (Zhou et al., 2020b) though the most proximal animal reservoir for SARS-CoV-2 remains unknown (Andersen et al., 2020; Lam et al., 2020).

Over the course of the SARS-CoV-2 pandemic, there have been reports of super-spreaders and transmission chains that have been more difficult to interrupt in some locations than in others (Hu et al., 2020). As few as 10% of SARS-CoV-2 infected people account for the majority of virus transmission events (Endo et al., 2020). Reported case fatality rates have varied by more than 10-fold (Center for Systems Science and Engineering at Johns Hopkins University; COVID-19 National Emergency Response Center, Epidemiology and Case Management Team, Korea Centers for Disease Control and Prevention, 2020; Grasselli et al., 2020; Onder et al., 2020). Such differences may reflect regional differences in the age of susceptible populations or the prevalence of comorbidities such as hypertension. Regional variation in diagnostic testing frequency, or in rate of reporting, is another likely explanation for these differences, since a wider net of diagnostic testing will capture milder cases (Zhao et al., 2020). Understanding the cause of these observed differences is critical for rational utilization of limited medical resources to counteract the pandemic.

Coronaviruses have the largest genomes of any known RNA viruses (Saberi et al., 2018) and they encode a 3'-to-5'-exoribonuclease required for high-fidelity replication by the viral RNA-dependent RNA polymerase (Denison et al., 2011; Smith et al., 2014). By preventing otherwise lethal mutagenesis (Smith et al., 2013) the repair function of the coronavirus exonuclease is thought necessary for the coronavirus genome size to extend beyond the theoretical limit imposed by error rates of viral RNA polymerases (Holmes, 2003). Though the rate of sequence variation among SARS-CoV-2 isolates is modest, over the course of the pandemic the virus has had opportunity to generate sequence variants, many of which have been identified among the thousands of SARS-CoV-2 genomes posted for public access (https://www.gisaid.org/) (Hadfield et al., 2018). 6,754 unique single nucleotide polymorphisms have been identified from high-quality sequence data, 99.4% of which are within viral gene sequences and 64% of these encode amino acid changes (Phelan et al., 2020). Here we investigate the potential consequences of one of these variants, the Spike protein variant D614G.

Results and Discussion

Over the course of the COVID-19 pandemic, the SARS-CoV-2 D614G S protein variant supplanted the ancestral virus

SARS-CoV-2 genomes sequenced from the earliest COVID-19 cases had minimal genetic variation. This suggests that they were recent descendants of a common ancestor that gave rise to all SARS-CoV-2 genomes seen during the outbreak (Rambaut et al., 2020). At the end of January 2020, a nonsynonymous nucleotide change from the ancestral virus, A23403G, first appeared in virus genomes reported in China (hCoV-19/Zhejiang/HZ103/2020; 24 January 2020) and in Germany (hCoV-19/Germany/BavPat1-ChVir929/2020; 28 January 2020). This nucleotide change encodes the spike (S) protein variant D614G. Although these sequences suggest that the D614G variant first emerged in China or Germany, given how few virus genomes have been sequenced from early in the outbreak, its geographic origin cannot be determined.

Since its first appearance, the frequency of the D614G variant has increased steadily over time in the global SARS-CoV-2 genomic dataset released by GISAID, to the point that it is nearly fixed in the viruses sampled globally (Figure 1A) or specifically in Europe, North America, Oceania, and Asia (Figure 1B). The D614G variant emerged at higher frequency in Africa and South America, suggesting a founder effect where a plurality of introductions carried D614G to these regions. Although recent data for South America are sparse, available sequences suggest D614G is approaching fixation in this continent as well.

Over the course of the whole pandemic, the D614G variant is present in approximately 74% of published sequences (Figure 1A). It should be noted, however, that the roughly 53,000 high-quality SARS-CoV-2 genomic sequences available from GISAID as of 25 June 2020, are a narrow snapshot of the global pandemic, approximately 0.5% of the 9.611M cases confirmed as of the same date (Dong et al., 2020). Analysis of D614G variant frequency is expected to gain resolution as sequences are released for samples banked across the duration of the pandemic. While understanding will improve around the historical trajectory of its ascendence to fixation, the D614G variant is now universally common on every continent except Antarctica.

The SARS-CoV-2 Spike protein variant D614G, then, was not present in the common ancestor of the sampled isolates but has risen to high frequency across many regions. This raises the question whether D614G confers a selective replication advantage to SARS-CoV-2, or whether it may explain regional differences in rates of death or other clinical parameters. Answering this question will require correlation with rates of transmission through populations, or with disease outcome, in combination with detailed assessment of its effects on viral biochemistry and replication, some of which are examined in experiments described below. To date, more severe disease outcome with D614G has not been reported, but several studies suggest that it is associated with a modest increase in viral load (Korber et al., 2020; McNamara et al., 2020; Wagner et al.).

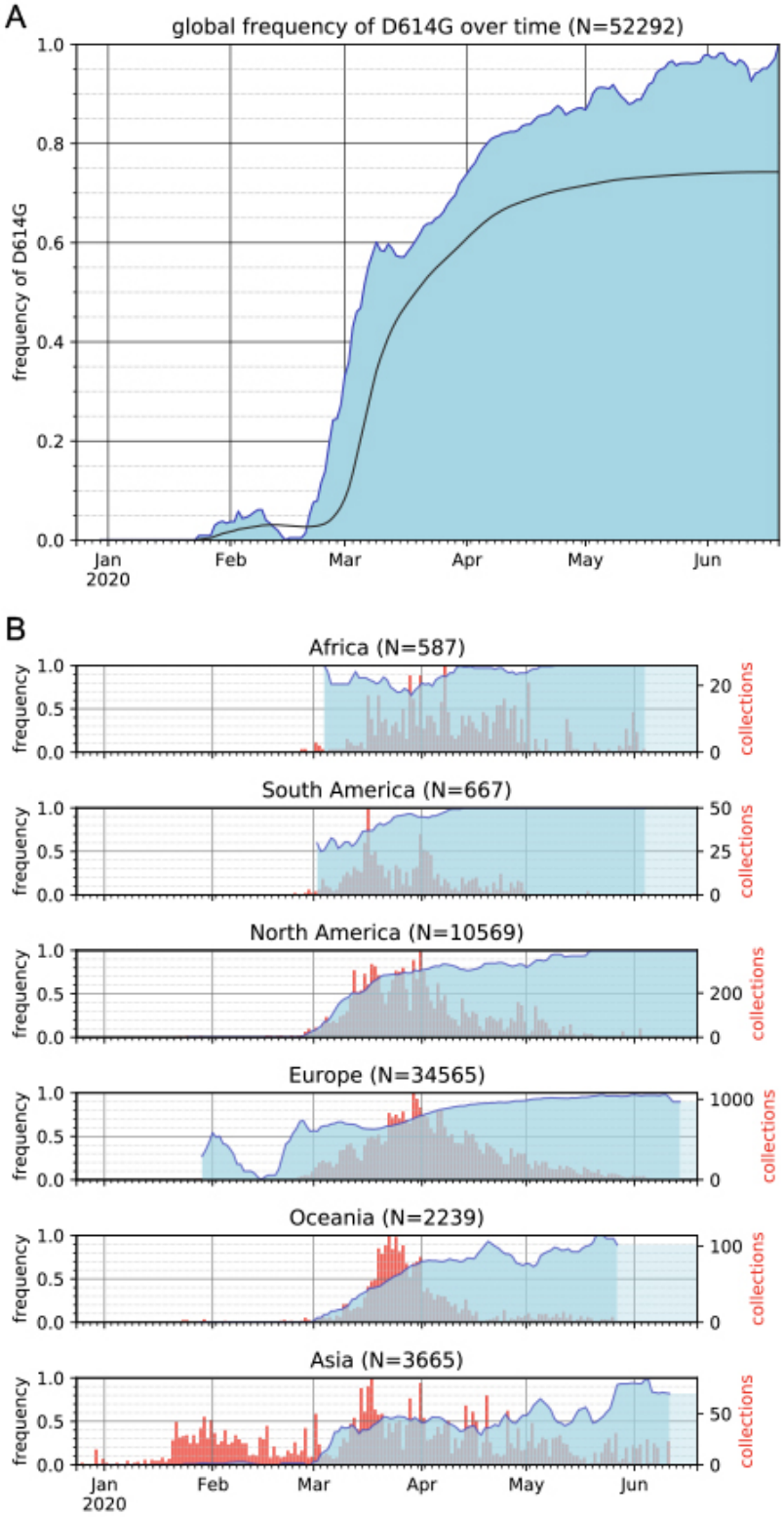


Figure 1. Frequency of the SARS-CoV-2 S protein D614G variant over the course of the COVID-19 pandemic.
The frequency of the S protein D614G variant over time in sequences published via the GISAID SARS-CoV-2 database, as of 25 June 2020. (A) Global dataset: The filled plot (blue) represents a 7-day rolling average of the fraction of sequences bearing the D614G variant for each collection date. Dates without published sequences are linearly interpolated. The cumulative frequency of the D614G variant (black line) is overlaid, showing the frequency of D614G in sequences from samples collected up to and including each date. (B) The frequency of the D614G variant over time (blue) in sequences collected from six continental regions, plotted as a 7-day rolling average. Frequencies are interpolated for dates without published sequence. The frequency of the last date with data is carried forward where recent dates lack data to indicate the most recent calculated frequency (light blue). The number of samples collected, sequenced, and published is overlaid (red bars) to provide the denominator used in calculating the frequency for each date, and to illustrate the non-uniformity of sampling.

The SARS-CoV-2 D614G S protein variant increases infectivity

The ability of the SARS-CoV-2 D614G S protein variant to target virions for infection of ACE2-positive cells was assessed using single-cycle lentiviral vector pseudotypes in tissue culture. Mammalian expression plasmids were engineered to encode the ancestral S protein (D614) and the D614G variant, as described previously (Moore et al., 2004). Each S protein expression plasmid was separately transfected into HEK-293 cells with plasmids encoding HIV-1 structural proteins and enzymes. Separate plasmids were transfected that encode RNAs with HIV-1 *cis*-acting signals for packaging and replication, and either GFP or luciferase (Luc) reporter cassettes. For each condition tested, multiple virus stocks were produced and tested individually after vector particle normalization using reverse transcriptase activity. 48 hours after challenge with the vectors, the transduction efficiency of each virion preparation was assessed by measuring the percent GFP-positive cells by flow cytometry, or by target cell-associated luciferase activity.

When Calu-3 human lung epithelial cells were used as targets, challenge with lentivirus bearing D614G resulted in 6-fold more GFP-positive cells, or 5-fold more bulk luciferase activity, than did particles bearing D614 S protein (Figure 2). When Caco-2 human colon epithelial cells were used as the target cells 4-fold higher infectivity was observed with D614G (Figure 2). Additionally, when HEK-293 cells or SupT1 cells had been rendered infectable by stable expression of exogenous ACE2 and TMPRSS2, D614G was 9-fold more infectious than D614 (Figure 2).

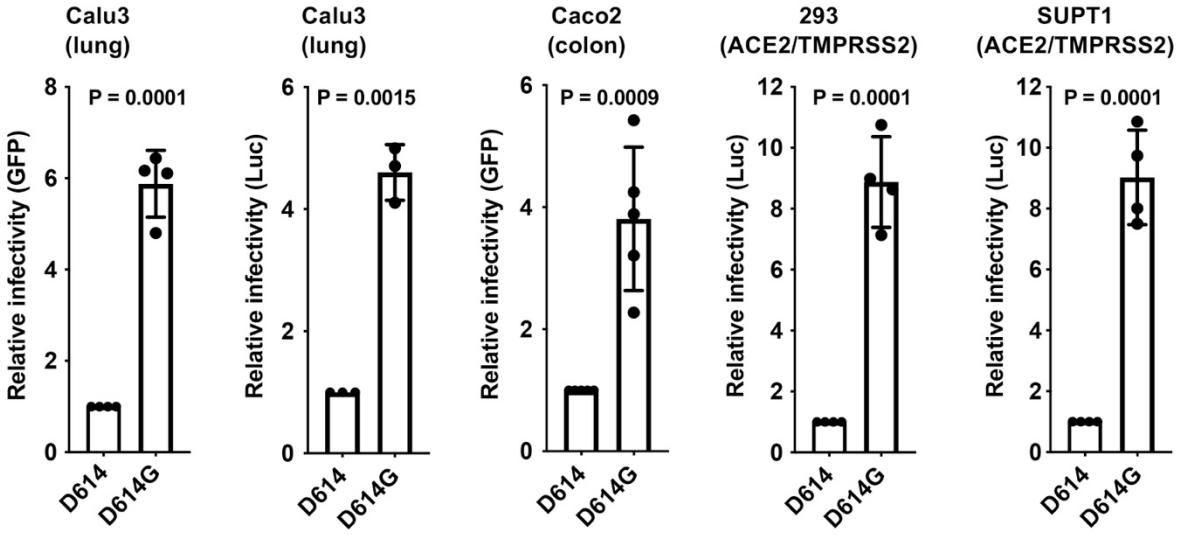


Figure 2. SARS-CoV-2 D614G S protein variant enhances infectivity

Lentiviral virions bearing either GFP or Luciferase transgenes, and pseudotyped with either SARS-CoV-2 D614 or D614G S proteins, were produced by transfection of HEK293 cells, and used to transduce human Calu3 lung cells, Caco2 colon cells, and either HEK293 or SupT1 cells stably expressing ACE2 and TMPRSS2. Relative infectivity of D614G vs D614, with D614 set at 1, was determined based on flow cytometry for percent GFP-positivity or bulk luciferase activity. Each point represents transduction with a lentiviral stock that was derived from an independent transfection. Shown are the mean +/- SD.

Effect of D614G on infectivity of different ACE2 orthologues

If SARS-CoV-2 Spike D614G is an adaptive mutant that was selected for increased human-to-human transmission following spillover from an animal reservoir, increased infectivity might only be evident on cells bearing ACE2 orthologues similar to that of the human. A SARS-like coronavirus isolated from a horseshoe bat (*Rhinopholus sp.*), with 96% identity along the whole virus genome and 93% in the S protein coding sequence, has the greatest sequence similarity to SARS-CoV-2 (Zhou et al., 2020b). Nonetheless, the most proximal source of the SARS-CoV-2 zoonosis remains unknown. Structural and biochemical studies (Letko et al., 2020; Walls et al., 2020; Wrapp et al., 2020; Zhou et al., 2020b) show that SARS-CoV-2 has a receptor binding domain with high affinity for ACE2 from humans, ferrets, cats, and other species that encode ACE2 with high homology (Wan et al., 2020). In a heterologous expression system on HeLa cells, human, civet, horse-shoe bat (Rhinolophus sinicus), and pig orthologs conferred susceptibility to infection, while the mouse ortholog did not (Zhou et al., 2020b).

As a first step to examine whether the increased infectivity of D614G is species-specific, plasmids were generated to express ACE2 orthologues from human (*Homo sapiens*), Malayan pangolin (*Manis javanica*), pig (*Sus scrofa*), cat (*Felis catus*), dog (*Canis lupis*), rat (*Rattus norvegicus*), and mouse (*Mus musculus*). HEK-293 were transfected separately with each of the ACE2 expression plasmids and challenged with luciferase-reporter-lentiviral pseudotypes bearing SARS-CoV-2 Spike protein, either D614 or D614G. Relative increase in infectivity due to D614G was comparable in cells expressing human, pangolin, pig, cat, or dog ACE2 orthologues (Figure 3). The relative effect of D614G on infectivity using cells bearing rat or mouse ACE2 orthologues could not be determined since in both cases the luciferase activity with D614 was not clearly above background. Examination of D614G infectivity using ACE2s encoded by other mammalian species may provide further insight into the structure and function of this Spike variant.

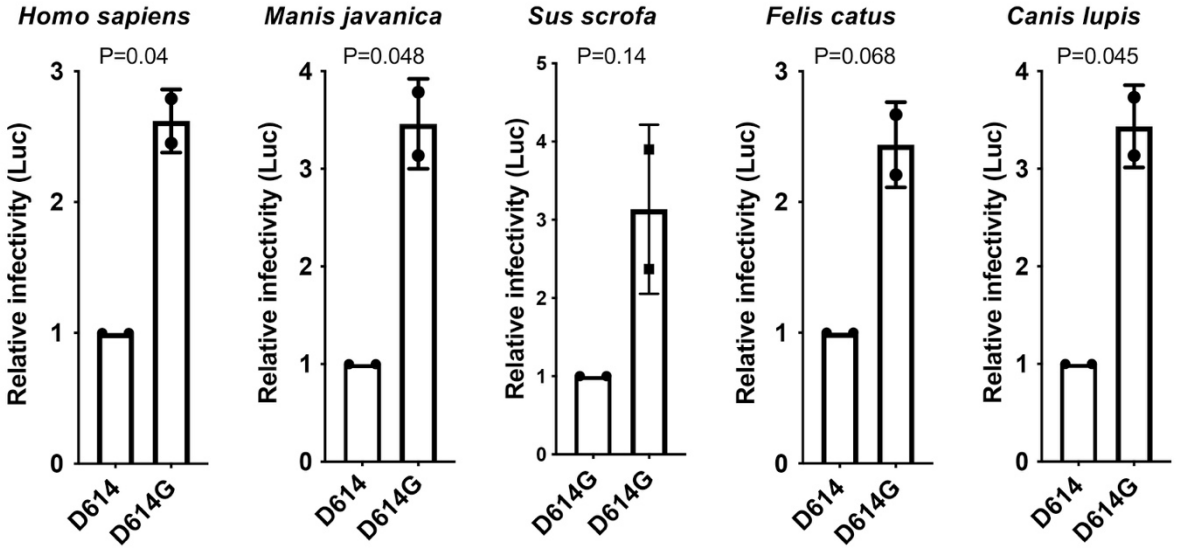
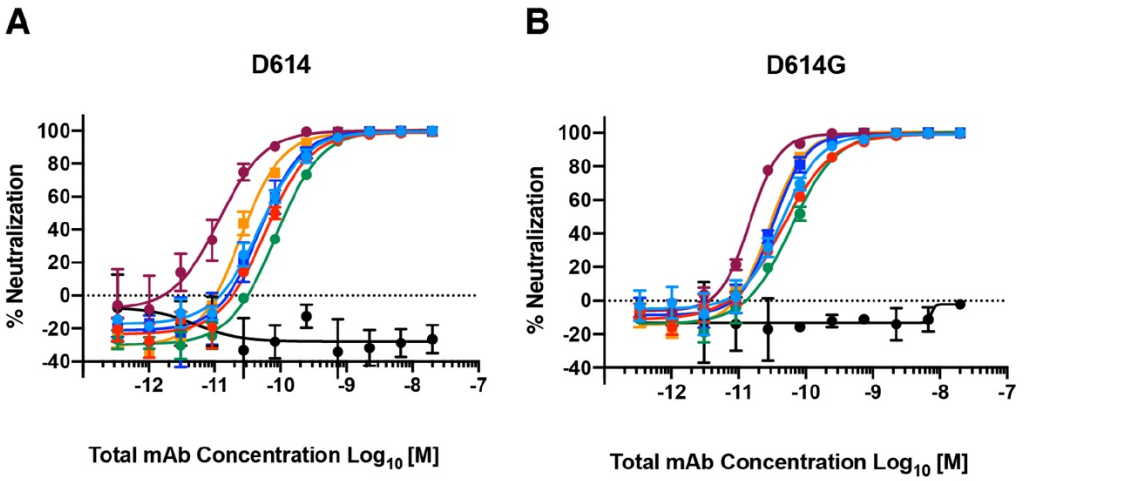


Figure 3. SARS-CoV-2 D614G S protein variant enhances infectivity on ACE2 orthologues

Lentiviral virions bearing Luciferase transgenes, and pseudotyped with either SARS-CoV-2 D614 or D614G S proteins, were produced by transfection of HEK293 cells, and used to transduce human HEK293 cells transiently transfected with plasmids encoding the indicated ACE2 orthologues. Relative infectivity of D614G vs D614, with D614 set at 1, was determined based on bulk luciferase activity. Each point represents transduction using lentiviral stock derived from an independent transfection. Shown are the mean +/- SD.

D614G is sensitive to neutralization by monoclonal antibodies targeting the receptor binding domain

The global spread and enhanced infectivity of the SARS-CoV-2 D614G variant raises the question of whether this change in Spike protein structure may compromise the effectiveness of antiviral therapies targeting that protein. To determine if this is the case, the neutralization potency was assessed of four monoclonal antibodies that target the SARS-CoV-2 Spike protein receptor binding domain. These fully human monoclonals are currently under evaluation in clinical trials (NCT04425629, NCT04426695) as therapeutics for COVID-19 (Hansen et al., 2020). As seen in Figure 4, all monoclonal antibodies tested, individually or in combinations, demonstrated similar neutralization potency against both SARS-CoV-2 Spike protein variants. As all four of these monoclonal antibodies bind epitopes within the Spike protein receptor binding domain, it remains important to determine whether the D614G variants changes neutralization sensitivity to other classes of anti-Spike monoclonal antibodies.



C

Antibody	D614 IC_{50} (M)	D614G IC_{50} (M)
REGN10989	1.17E-11	1.47E-11
REGN10987	5.76E-11	4.57E-11
REGN10933	4.81E-11	4.34E-11
REGN10934	8.35E-11	6.04E-11
REGN10933+REGN10987	4.82E-11	3.28E-11
REGN10989+REGN10987	2.33E-11	2.46E-11

Figure 4. Neutralization potency of monoclonal antibodies targeting SARS-CoV-2 Spike protein variants. Vero cells were challenged with pVSV-SARS-CoV-2-S-mNeon pseudoparticles encoding either (A) 614D or (B) 614G spike variants, in the presence of serial dilutions of the indicated anti-Spike monoclonal antibodies or IgG1 isotype control. mNeon protein fluorescence was measured 24 hours post-infection as a read-out for virus infectivity. Data is graphed as percent neutralization relative to virus only infection control. (C) Neutralization potency (IC_{50}) of individual monoclonals and of monoclonal combinations against the 614G and 614D variants, is indicated.

Modeling the effect of D614G on structure

Structures of the soluble trimeric S protein ectodomain in pre-fusion conformations demonstrate that the receptor-binding domains (RBD) of each protomer can independently adopt either a "down" position or an "up" position, giving rise to asymmetric trimer conformations (Walls et al., 2020; Wrapp et al., 2020). Previous structural studies of SARS-CoV S and MERS-CoV S identified similar conformations (Gui et al., 2017; Pallesen et al., 2017; Walls et al., 2019; Yuan et al., 2017). These studies suggested that only upon transition of an RBD to the up position is the protomer capable of binding the ACE2 receptor. Models of coronavirus S-mediated membrane fusion describe ACE2 binding to all three RBD domains in the up position as destabilizing the pre-fusion S trimer, leading to dissociation of S1 from S2 and promoting transition to the post-fusion conformation (Pallesen et al., 2017; Walls et al., 2019). Thus, the up position likely reflects an intermediate conformation that is on-pathway to S-mediated membrane fusion.

The structures of SARS-CoV-2 S (Walls et al., 2020; Wrapp et al., 2020) predict that D614 of a protomer with RBD in the up position forms a stabilizing contact with T859 of the neighboring protomer (Figure 2). Specifically, the delta-O of D614 in the up protomer is within hydrogen-bonding distance (2.7A) of the gamma-O of T859 of the neighboring protomer. The other two inter-protomer distances for D614-T859 are 3.4A and 4.4A, suggesting that the electrostatic interaction may be labile and only form during transition to the up conformation. It is predicted that the D614G variant would abrogate this interaction and destabilize the inter-protomer interface.

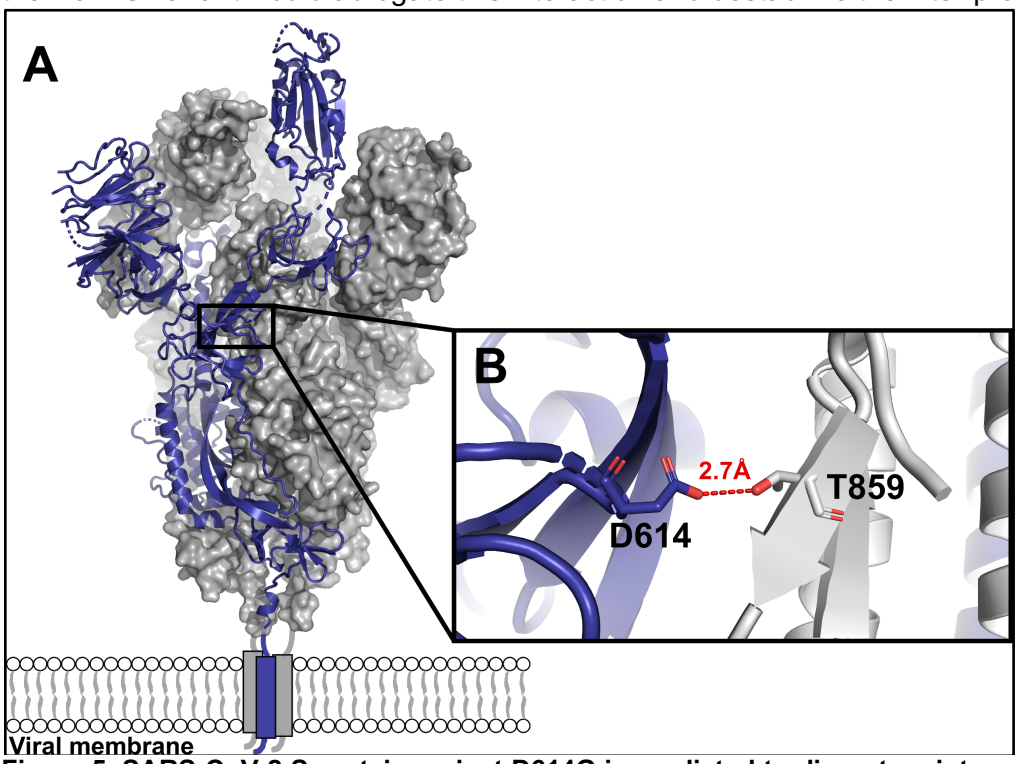


Figure 5. SARS-CoV-2 S protein variant D614G is predicted to disrupt an inter-protomer hydrogen bond with T859.

(A) The cryo-EM structure of SARS-CoV-2 S in an asymmetric conformation. The single protomer in the up conformation is shown in blue cartoon representation. The two protomers in the down position are shown in grey surface representations (PDB: 6VSB).

(B) Blow-up showing the predicted stabilizing electrostatic interaction between the delta-O of D614 of the up protomer (blue) and the gamme-O of T859 of the neighboring down protomer (grey). O atoms are shown in red. The inter-atomic distance of 2.7A, which is within the H-bonding distance, is shown in red. The D614G mutant would remove this electrostatic interaction.

Methods

Analysis of D614G frequency in published data

The frequency of the SARS-CoV-2 D614G S protein variant in published genomic data was examined using the full Nextstrain-curated set of sequences available from GISAID as of 25 June 2020 (Hadfield et al., 2018; Shu and McCauley, 2017). Sequences were aligned to the ancestral reference sequence (NCBI GenBank accession NC_045512.2) using mafft v7.464 (Katoh and Standley, 2013) with the "--keeplength" and "--addfragments" parameters, which preserve the coordinate space of the reference sequence. To remove lower-quality sequences from the dataset, all sequences in the alignment were masked with ambiguous bases ('N') in the regions spanning the first 100bp and the last 50bp, as well as at error-prone sites located at the (1-indexed, NC_045512.2 coordinate space) positions 13402, 24389, 24390. Sequences shorter than 28kb or with >2% ambiguous bases were removed from the alignment. The frequency of D614G was calculated in the resulting data by extracting the sequence region corresponding to the gene for the S protein, spanning 21563-25384bp. These sequences were processed using a script importing biopython (Cock et al., 2009) to remove any gaps introduced by the alignment process and translate the sequence to protein space. The identity of the variant at amino acid position 614 was tabulated for the full dataset and reported as frequency by date using the collection dates reported in the Nextstrain-curated metadata file available from GISAID (Hadfield et al., 2018; Shu and McCauley, 2017). The frequency was calculated as (# sequences with D614G)/(# sequences). Frequency within the six continental regions was calculated based on the "region" geographic classification associated with each sample in the metadata. Frequency values were linearly interpolated for dates surrounded by valid data. The frequency of the last date with data was carried forward where recent dates lack data. The resulting values were rendered as plots using matplotlib (Hunter, 2007). The script for analyzing and plotting D614G variant frequency is available via GitHub: https://gist.github.com/tomkinsc/c3fa656b0c833db6d3a2c579b74dcc4f

Plasmids

The plasmids used here were either previously described or generated using standard cloning methods. The full list of plasmids used here, along with their purpose and characteristics, is provided in Table 1. All newly engineered plasmids are available, along with full sequences, at https://www.addgene.org/Jeremy_Luban/.

Cell culture

All cells were cultured in humidified incubators with 5% CO_2 at 37º C, and monitored for mycoplasma contamination using the Mycoplasma Detection kit (Lonza LT07-318). HEK293 cells (ATCC CRL-1573), and HEK-293T cells (CRL-3216 or CRL-3216) were cultured in DMEM supplemented with 10% heat-inactivated FBS, 1 mM sodium pyruvate, 20 mM GlutaMAX, 1× MEM non-essential amino acids, and 25 mM HEPES, pH 7.2. Calu3 cells (ATCC HTB-55) were maintained in EMEM supplemented with 10% FBS. Caco2 cells (ATCC HTB-37) were maintained in EMEM supplemented with 20% FBS. SUP-T1 [VB] cells (ATCC CRL-1942) were cultured in RPMI supplemented with 10% heat-inactivated FBS, 1mM sodium pyruvate, 20mM GlutaMAX, 1xMEM non-essential amino acids, and 25 mM HEPES, pH7.2. Vero cells (ATCC CCL-81) were cultured in DMEM high glucose media containing 10% heat-inactivated fetal bovine serum, and 1X Penicillin/Streptomycin/L-Glutamine.

Virus production

24 hrs prior to transfection, 6×10^5 HEK-293 cells were plated per well in 6 well plates. All transfections used 2.49 µg plasmid DNA with 6.25 µL TransIT LT1 transfection reagent (Mirus, Madison, WI) in 250 µL Opti-MEM (Gibco).

Single-cycle HIV-1 vectors pseudotyped with SARS-CoV-2 Spike protein, either D614 or D614G, were produced by transfection of either HIV-1 pNL4-3 Δenv Δvpr luciferase reporter plasmid (pNL4-3.Luc.R-E-), or pUC57mini NL4-3 Δenv eGFP reporter plasmid, in combination with the indicated Spike expression plasmid, at a ratio of 4:1.

ACE2 expression vectors were produced by transfecting cells with one of the pscALPSpuro-ACE2 plasmids, along with the HIV-1 *gag-pol* expression plasmid psPAX2, and the VSV glycoprotein expression plasmid pMD2.G (4:3:1 ratio of plasmids).. 16 hrs post-transfection, culture media was changed. Viral supernatant was harvested 48 hours after media change, passed through a 0.45 µm filter, and stored at 4ºC. TMPRSS2 expression transfer vector was produced similarly but with pscALPSblasti-TMPRSS2.

Exogenous reverse transcriptase assay

5 µL transfection supernatant was mixed with 5 µL 0.25% Triton X-100, 50 mM KCl, 100 mM Tris-HCl pH 7.4, and 0.4 U/µL RiboLock RNase inhibitor, and then diluted 1:100 in 5 mM $(NH_4)_2SO_4$, 20 mM KCl, and 20 mM Tris-HCl pH 8.3. 10 µL of this was then added to a single-step, RT-PCR assay with 35 nM MS2 RNA (IDT) as template, 500 nM of each primer (5'-TCCTGCTCAACTTCCTGTCGAG-3' and 5'-CACAGGTCAAACCTCCTAGGAATG-3'), and 0.1 µL hot-start Taq DNA polymerase (Promega, Madison, WI) in 20 mM Tris-Cl pH 8.3, 5 mM $(NH_4)_2SO_4$, 20 mM KCl, 5 mM $MgCl_2$, 0.1 mg/ml BSA, 1/20,000 SYBR Green I (Invitrogen), and 200 µM dNTPs in total 20 µL reaction. The RT-PCR reaction was carried out in a Biorad CFX96 real-time PCR detection system with the following parameters: 42ºC for 20 min, 95ºC for 2 min, and 40 cycles [95ºC for 5 sec, 60ºC for 5 sec, 72ºC for 15 sec, and acquisition at 80ºC for 5 sec].

Transduction with lentiviral vectors

<u>HEK-293 cells expressing ACE2 and TMPRSS2</u>: 2.5 x 10^5 cells were plated per well in a 12 well plate. The next day cells were transduced with 250 uL of supernatant containing TMPRSS2-encoding lentivirus for 16 hr at 37°C, after which fresh media was added to cells. 48 hrs after transduction cells were replated and selected with blasticidin (InvivoGen, catalogue #ant-bl-1) at 10 ug/ml. After selection, cells were transduced similarly with supernatant containing ACE2-encoding lentivirus and selected with 1 ug/mL of puromycin (InvivoGen, San Diego, CA, catalogue #ant-pr-1).

<u>SupT1 cells expressing ACE2 and TMPRSS2</u>: 1 x 10^6 SupT1 cells were transduced with 400 uL of supernatant containing TMPRSS2-encoding virus followed by selection with 10 ug/mL blasticidin 48 hrs later. TMPRSS2 expressing SupT1 cells were then transduced with a second vector expressing ACE2, followed by puromycin selection at 1 ug/mL.

Virus Infectivity Assays

16 hours prior to transduction, adherent cells were seeded in 96 well plates. HEK-293 cells were plated at 5 x 10^4 cells per well. Calu3 and Caco2 were plated at 3 x 10^4 per well. Cells were incubated in virus-containing media for 16 hrs at 37°C when fresh medium was added to cells. 48 to 72 hours after transduction cells were assessed for luciferase activity or for GFP by flow cytometry. For transfection of SupT1 cells, 1 x 10^5 cells were plated in a U-bottom 96 well plate and spinfected at 30ºC with 100 uL of virus-containing media for 2 hours at 1,200 x g. Fresh media was added after spinfection and cells were incubated 48-72 hours prior to analysis. Cells transduced with GFP virus were fixed with BD Cytofix (BD Biosciences, San Jose, CA, Cat number 554655) and analyzed using the Accuri C6system. Data was analyzed using FlowJo 10.5 (FlowJo, LLC, Ashland, OR). Cells transduced with luciferase expressing virus were assessed using Promega Steady-Glo system (Promega Madison, WI)

Generation of VSV pseudoparticles and neutralizations assays

VSV-SARS-CoV-2-S pseudoparticle generation and neutralization assays were performed as previously described (Baum et al., 2020; Hansen et al., 2020). HEK293T cells were seeded overnight in DMEM high glucose media (Life Technologies) containing 10% heat-inactivated fetal bovine serum (Life Technologies), and Penicillin/- Streptomycin-L-Glutamine (Life Technologies). The following day, Spike expression plasmids were transfected with Lipofectamine LTX (Life Technologies) following the manufacturer's protocol. At 24 hours post transfection, the cells were washed with phosphate buffered saline (PBS) and infected at an MOI of 1 with the VSVΔG:mNeon/VSV-G virus diluted in 10 mL Opti-MEM (Life Technologies). The cells were incubated 1 hour at 37°C with 5% CO2. Cells were washed three times with PBS to remove residual input virus and overlaid with DMEM high glucose media (Life Technologies) with 0.7% Low IgG BSA (Sigma), sodium pyruvate (Life Technologies), and Gentamicin (Life Technologies). After 24 hours at 37° C with 5% CO2, the supernatant containing pseudoparticles was collected, centrifuged at 3,000 x g for 5 minutes to clarify, aliquoted, and frozen at -80° C. For neutralization assays, Vero cells were seeded in 96-well plates 24 hours prior to assay and grown to 85% confluence before challenge. Antibodies were diluted in DMEM high 3 glucose media containing 0.7% Low IgG BSA (Sigma), 1X Sodium Pyruvate, and 0.5% Gentamicin (this will be referred to as "Infection Media") to 2X assay concentration and diluted 3-fold down in Infection Media, for an 11-point dilution curve in the assay beginning at 10 ug/mL (66.67 nM). pVSV-SARS-CoV-2-S pseudoparticles were diluted 1:1 in Infection Media for a fluorescent focus (ffu) count in the assay of ~1000 ffu. Antibody dilutions were mixed 1:1 with pseudoparticles for 30 minutes at room temperature prior to addition onto Vero cells. Cells were incubated at 37° C, 5% CO2 for 24 hours. Supernatant was removed from cells and replaced with 100 uL PBS, and fluorescent foci were quantitated using the SpectraMax i3 plate reader with MiniMax imaging cytometer.

Acknowledgements

This work was supported by NIH grants R37AI147868 and R01AI148784 to J.L., NIH/NIAID U19AI110818 to P.C.S., 7DP2AI124384-02 to J.B.M., a grant from the Evergrande COVID-19 Response Fund Award from the Massachusetts Consortium on Pathogen Readiness to J.L., a grant from The Worcester Foundation to K.S., a Sara Elizabeth O'Brien Fellowship Award/King Trust to L.Y., and a National Science Foundation Graduate Research Fellowship (Grant No. 1745303) to C.H.T.-T. A portion of this project has been funded in whole or in part with Federal funds from the Department of Health and Human Services; Office of the Assistant Secretary for Preparedness and Response; Biomedical Advanced Research and Development Authority, under OT number: HHSO100201700020C.

References

Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., and Garry, R.F. (2020). The proximal origin of SARS-CoV-2. Nat. Med. *26*, 450–452.

Baum, A., Fulton, B.O., Wloga, E., Copin, R., Pascal, K.E., Russo, V., Giordano, S., Lanza, K., Negron, N., Ni, M., et al. (2020). Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. Science.

Center for Systems Science and Engineering at Johns Hopkins University Coronavirus COVID-19 Global Cases.

Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics *25*, 1422–1423.

Coronaviridae Study Group of the International Committee on Taxonomy of Viruses (2020). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol *5*, 536–544.

COVID-19 National Emergency Response Center, Epidemiology and Case Management Team, Korea Centers for Disease Control and Prevention (2020). Coronavirus Disease-19: The First 7,755 Cases in the Republic of Korea. Osong Public Health Res Perspect *11*, 85–90.

Denison, M.R., Graham, R.L., Donaldson, E.F., Eckerle, L.D., and Baric, R.S. (2011). Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. RNA Biol. *8*, 270–279.

Dong, E., Du, H., and Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. *20*, 533–534.

Drosten, C., Günther, S., Preiser, W., van der Werf, S., Brodt, H.-R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R.A.M., et al. (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N. Engl. J. Med. *348*, 1967–1976.

Endo, A., Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Abbott, S., Kucharski, A.J., and Funk, S. (2020). Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. Wellcome Open Res *5*, 67.

Grasselli, G., Pesenti, A., and Cecconi, M. (2020). Critical Care Utilization for the COVID-19 Outbreak in Lombardy, Italy: Early Experience and Forecast During an Emergency Response. JAMA.

Grubaugh, N.D., Petrone, M.E., and Holmes, E.C. (2020). We shouldn't worry when a virus mutates during disease outbreaks. Nat Microbiol *5*, 529–530.

Gui, M., Song, W., Zhou, H., Xu, J., Chen, S., Xiang, Y., and Wang, X. (2017). Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. Cell Res. *27*, 119–129.

Hadfield, J., Megill, C., Bell, S.M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., and Neher, R.A. (2018). Nextstrain: real-time tracking of pathogen evolution.

Bioinformatics *34*, 4121–4123.

Hansen, J., Baum, A., Pascal, K.E., Russo, V., Giordano, S., Wloga, E., Fulton, B.O., Yan, Y., Koon, K., Patel, K., et al. (2020). Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. Science.

Holmes, E.C. (2003). Error thresholds and the constraints to RNA virus evolution. Trends Microbiol. *11*, 543–546.

Hu, K., Zhao, Y., Wang, M., Zeng, Q., Wang, X., Wang, M., Zheng, Z., Li, X., Zhang, Y., Wang, T., et al. (2020). Identification of a super-spreading chain of transmission associated with COVID-19 (medRxiv).

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet *395*, 497–506.

Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment. Comput. Sci. Eng. *9*, 90–95.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., et al. (2020). Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell.

Ksiazek, T.G., Erdman, D., Goldsmith, C.S., Zaki, S.R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J.A., Lim, W., et al. (2003). A novel coronavirus associated with severe acute respiratory syndrome. N. Engl. J. Med. *348*, 1953–1966.

Lam, T.T.-Y., Jia, N., Zhang, Y.-W., Shum, M.H.-H., Jiang, J.-F., Zhu, H.-C., Tong, Y.-G., Shi, Y.-X., Ni, X.-B., Liao, Y.-S., et al. (2020). Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. Nature.

Letko, M., Marzi, A., and Munster, V. (2020). Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nat Microbiol *5*, 562–569.

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet *395*, 565–574.

McNamara, R.P., Caro-Vegas, C., Landis, J.T., Moorad, R., Pluta, L.J., Eason, A.B., Thompson, C., Bailey, A., Villamor, F.C.S., Lange, P.T., et al. (2020). High-density amplicon sequencing identifies community spread and ongoing evolution of SARS-CoV-2 in the Southern United States.

Moore, M.J., Dorfman, T., Li, W., Wong, S.K., Li, Y., Kuhn, J.H., Coderre, J., Vasilieva, N., Han, Z., Greenough, T.C., et al. (2004). Retroviruses pseudotyped with the severe acute respiratory syndrome coronavirus spike protein efficiently infect cells expressing angiotensin-converting enzyme 2. J. Virol. *78*, 10628–10635.

Onder, G., Rezza, G., and Brusaferro, S. (2020). Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. JAMA.

Pallesen, J., Wang, N., Corbett, K.S., Wrapp, D., Kirchdoerfer, R.N., Turner, H.L., Cottrell, C.A., Becker, M.M., Wang, L., Shi, W., et al. (2017). Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. Proc. Natl. Acad. Sci. U. S. A. *114*, E7348–E7357.

Phelan, J., Deelder, W., Ward, D., Campino, S., Hibberd, M.L., and Clark, T.G. (2020). Controlling the SARS-CoV-2 outbreak, insights from large scale whole genome sequences generated across the world.

Rambaut, A., Holmes, E.C., Hill, V., O'Toole, Á., McCrone, J.T., Ruis, C., du Plessis, L., and Pybus, O.G. (2020). A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology.

Saberi, A., Gulyaeva, A.A., Brubacher, J.L., Newmark, P.A., and Gorbalenya, A.E. (2018). A planarian nidovirus expands the limits of RNA genome size. PLoS Pathog. *14*, e1007314.

Shu, Y., and McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data – from vision to reality. Eurosurveillance *22*, ii=30494.

Smith, E.C., Blanc, H., Surdel, M.C., Vignuzzi, M., and Denison, M.R. (2013). Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: evidence for proofreading and potential therapeutics. PLoS Pathog. *9*, e1003565.

Smith, E.C., Sexton, N.R., and Denison, M.R. (2014). Thinking Outside the Triangle: Replication Fidelity of the Largest RNA Viruses. Annu Rev Virol *1*, 111–132.

Wagner, C., Roychoudhury, P., Hadfield, J., Hodcroft, E.B., Lee, J., Moncla, L.H., Müller, N.F., Behrens, C., Huang, M.L., Mathias, P., et al. Comparing viral load and clinical outcomes in Washington State across D614G mutation in spike protein of SARS-CoV-2. Github.

Walls, A.C., Xiong, X., Park, Y.-J., Tortorici, M.A., Snijder, J., Quispe, J., Cameroni, E., Gopal, R., Dai, M., Lanzavecchia, A., et al. (2019). Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion. Cell *176*, 1026–1039.e15.

Walls, A.C., Park, Y.-J., Tortorici, M.A., Wall, A., McGuire, A.T., and Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell *181*, 281–292.e6.

Wan, Y., Shang, J., Graham, R., Baric, R.S., and Li, F. (2020). Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. J. Virol. *94*.

Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.-L., Abiona, O., Graham, B.S., and McLellan, J.S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science *367*, 1260–1263.

Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., Meng, J., Zhu, Z., Zhang, Z., Wang, J., et al. (2020a). Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. Cell Host Microbe *27*, 325–328.

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., et al. (2020b). A new coronavirus associated with human respiratory disease in China. Nature *579*, 265–269.

Yuan, Y., Cao, D., Zhang, Y., Ma, J., Qi, J., Wang, Q., Lu, G., Wu, Y., Yan, J., Shi, Y., et al.

(2017). Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. Nat. Commun. *8*, 15092.

Zaki, A.M., van Boheemen, S., Bestebroer, T.M., Albert D M, and Fouchier, R.A.M. (2012). Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. New England Journal of Medicine *367*, 1814–1820.

Zhao, S., Lin, Q., Ran, J., Musa, S.S., Yang, G., Wang, W., Lou, Y., Gao, D., Yang, L., He, D., et al. (2020). Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. Int. J. Infect. Dis. *92*, 214–217.

Zhou, H., Chen, X., Hu, T., Li, J., Song, H., Liu, Y., Wang, P., Liu, D., Yang, J., Holmes, E.C., et al. (2020a). A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein. Current Biology *30*, 2196–2203.e3.

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., et al. (2020b). A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature *579*, 270–273.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., et al. (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. N. Engl. J. Med. *382*, 727–733.