Emergence of SARS-CoV-2 spike RBD mutants that enhance viral infectivity through increased human ACE2 receptor binding affinity

Junxian Ou, Zhonghua Zhou, Ruixue Dai, Shan Zhao, Xiaowei Wu, Jing Zhang, Wendong Lan, Lilian Cui, Jianguo Wu, Donald Seto, James Chodosh, Gong Zhang,[*] and Qiwei Zhang[*]

J. Ou, S. Zhao, X. Wu, W. Lan, Prof. Q. Zhang

Guangdong Provincial Key Laboratory of Tropical Disease Research, School of Public Health, Southern Medical University,

Guangzhou, Guangdong 510515, China

E-mail: zhangqw@jnu.edu.cn

Z. Zhou, Prof. G. Zhang

Key Laboratory of Functional Protein Research of Guangdong Higher Education Institutes,

Institute of Life and Health Engineering, College of Life Science and Technology,

Jinan University,

Guangzhou, Guangdong 510632, China

E-mail: zhanggong@jnu.edu.cn

R. Dai, L. Cui

Department of Environmental Science and Engineering,

Fudan University,

Shanghai 200433, China

And Novoprotein Scientific Inc.

Shanghai 201203, China

Dr. J. Zhang, Prof. J. Wu, Prof. Q. Zhang

Guangdong Provincial Key Laboratory of Virology,

Institute of Medical Microbiology,

Jinan University,

Guangzhou, Guangdong 510632, China

Prof. D. Seto

Bioinformatics and Computational Biology Program,

School of Systems Biology,

George Mason University,

Manassas, VA 20110, USA

Prof. J. Chodosh

Department of Ophthalmology, Howe Laboratory

Massachusetts Eye and Ear Infirmary

Harvard Medical School

Boston, MA 02114, USA

Abstract

The current global pandemic of COVID-19 is caused by a novel coronavirus SARS-CoV-2. The SARS-CoV-2 spike protein receptor-binding domain (RBD) is the critical determinant of viral tropism and infectivity. To investigate whether naturally occurring mutations in the RBD have altered the receptor binding affinity and infectivity, firstly we analyzed *in silico* the binding dynamics between mutated SARS-CoV-2 RBDs and the human ACE2 receptor. Among 1609 genomes of SARS-CoV-2 strains isolated during the early transmission phase, 32 non-synonymous RBD mutants were identified and found clustered into nine mutant types under high positive selection pressure. Applying molecular dynamics simulations, three mutant types (V367F, W436R, N354D/D364Y) displayed higher binding affinity to human ACE2, likely due to the enhanced structural stabilization of the RBD beta-sheet scaffold. The increased infectivity of one mutant (V367F) circulating worldwide was further validated by performing receptor-ligand binding ELISA, surface plasmon resonance, and pseudotyped virus assays. Genome phylogenetic analysis of V367F mutants showed that during the early transmission phase, most V367F mutants clustered more closely with the SARS-CoV-2 prototype strain than the dual-mutation variants (V367F + D614G), which emerged later and formed a distinct sub-cluster. The analysis of critical RBD mutations provides further insights into the evolutionary trajectory of SARS-CoV-2 under high selection pressure and supports the continuing surveillance of spike mutations to aid in the development of COVID-19 drugs and vaccines.

Keywords: COVID-19; SARS-CoV-2; receptor-binding domain (RBD); viral infectivity; ACE2 receptor

1. Introduction

A novel coronavirus SARS-CoV-2 has caused outbreaks of Coronavirus Disease 2019 (COVID-19) globally beginning in mid-December 2019, with an epicenter in Wuhan, China[1–3]. As of September 8, 2020, SARS-CoV-2 had infected 27,236,916 people world-wide and caused 891,031 deaths with an estimated fatality rate of 3.27%[4]. This on-going pandemic of COVID-19 has become the most serious threat to public health in this century.

The origin of SARS-CoV-2 remains elusive. However, the initial cases were largely associated with a seafood market, which indicated potential zoonotic transmissions[1,5]. Although bats and pangolins are most likely the reservoir and the intermediate hosts in the wild, more evidence is needed to support zoonotic transmission and to track the origin of this new coronavirus[6–8].

Angiotensin-converting enzyme 2 (ACE2) is the host cellular receptor for the SARS-CoV-2 spike protein, which is similar to its counterpart in SARS-CoV[9–12]. The receptor-binding domain (RBD) of the spike protein subunit S1 interacts directly with ACE2, providing for tight binding to the peptidase domain of ACE2[13,14]. Therefore, RBD is the critical determinant of virus-receptor interaction and reflects viral host range, tropism, and infectivity. Although the RBD sequences of different SARS-CoV-2 strains circulating globally are largely conserved, mutations have appeared; these may account for differences in viral infectivity and also contribute to its spread[15–18].

The S protein expresses antigenic sites on its protein surface, likely with both T-cell and B-cell epitopes. The main antibody binding sites substantially overlap with RBD, and an antibody binding to these sites is likely to block viral entry into cells[19–21].

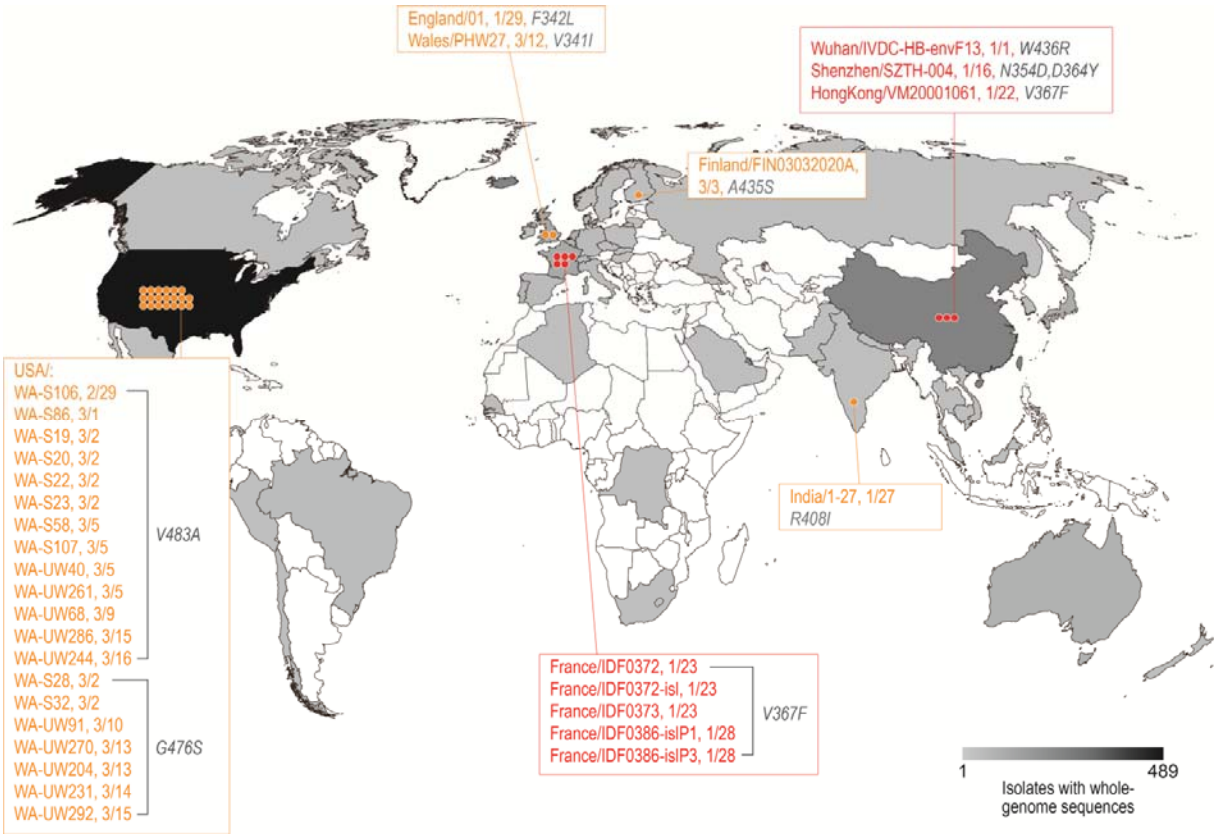
To investigate whether mutations in RBD have altered the receptor binding affinities and whether these strains may have been selected for higher infectivity, the binding dynamics and the infectivity between the SARS-CoV-2 RBDs of the mutant strains to date and human ACE2 receptor were modelled and assessed computationally. In addition, experimental validation of the enhanced affility and infectivity of the V367F mutant was assessed. The experimental results supported and validated our computational simulations.

2. Results

2.1. SARS-CoV-2 RBD mutation mapping during the early transmission phase

Among the 1609 SARS-CoV-2 strains with whole genome sequences available in public databases collected before April 2020, 32 strains contained amino acid mutations in the RBD (**Supplementary Table 1**). These strains were reported from multiple locations, including China, U.K., Finland, France, Belgium, U.S.A., and India (**Figure. 1)**. Most mutants in RBD deviated from the original reported genome (SARS-CoV-2 Wuhan-Hu-1) by only one amino acid (**Supplementary Figure 1**). The 32 mutations parse into nine types. Mutation V367F was found in six individual strains isolated from four patients: three in France and one in Hong Kong. This suggested that these variants may have originated as a sub-lineage. Similarly, novel sub-lineages were also identified including the 13 V483A mutants and seven G476S mutants isolated in the U.S.A. (**Figure. 1**).

Figure. 1: Distribution of the SARS-CoV-2 RBD mutants during the early transmission phase (January through March, 2020).



The geographic distribution of the 32 RBD mutants clustering into nine mutant types is displayed. Strains with names highlighted in red are mutants with the enhanced binding affinity. The strains with names noted in yellow are mutants with similar binding affinities.

2.2. Strong positive selective pressure on SARS-CoV-2 RBD

Since RBD is the only domain to bind human ACE2 and, in turn, initiates cell entry, it is believed that the RBD should be highly conserved. To test this hypothesis, we investigated the selective pressures of the S gene by calculating nonsynonymous/synonymous substitution rate ratios (dN/dS ratios) for various segments of the S gene in the 1609 early SARS-CoV-2 strains[22]. Interestingly, the entire S gene exhibited a dN/dS of 4.1197, remarkably greater than 1, showing that the S gene is indeed under positive selective pressure (**Table 1**). The RBD showed a similar dN/dS (3.3545) as the entire S protein, indicating that high selective pressure was also applied to this essential domain. Therefore, the functional relevance of these RBD mutations may be inferred.

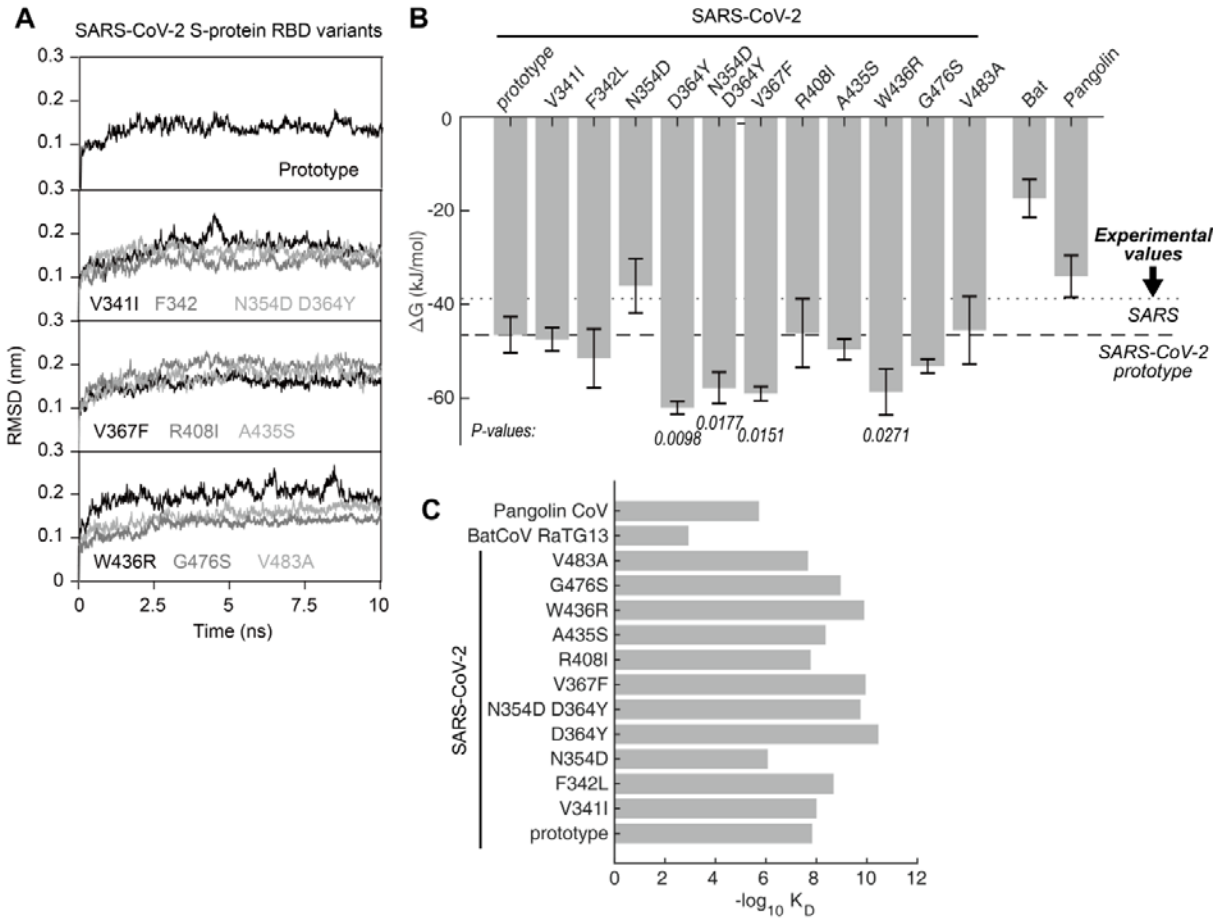
2.3. Three mutant types during the early transmission phase bind human ACE2 receptor with higher affinity

To estimate the functional changes suggested by the RBD mutations, we performed molecular dynamics (MD) simulations for the prototype SARS-CoV-2 (Wuhan-Hu-1 strain) and the RBD mutants in order to assess their binding energy to human ACE2. These were performed using GROMACS 2019[23,24]. The complex structure of the SARS-CoV-2 RBD domain and human ACE2 was obtained from the National Microbiology Data Center (ID: NMDCS0000001) (PDB ID: 6LZG) (https://www.rcsb.org/structure/6LZG). Mutated amino acids of the SARS-CoV-2 RBD mutants were directly replaced in the model, and the bat/pangolin CoV RBD domain was modeled using SWISS-MODEL[25]. Each simulation was performed at 10ns and each model was simulated in triplicate. All trajectories reached a plateau of RMSD after 2~5ns

(**Figure. 2A**), indicating that their structures reached equilibrium. All of the subsequent computations on thermodynamics were based on the 5~10ns trajectories. Three RBD mutant types (N354D and D364Y, V367F, W436R) exhibited significantly lowered ΔG, suggesting a significantly increased affinity to human ACE2; the other mutants showed a similar ΔG compared to the prototype (**Figure. 2B**). The ΔG of these three mutant types were around -58 kJ/mol, approximately 25% lower than the prototype strain (-46.5 kJ/mol, calculated from the experimentally measured KD) (**Figure. 2B**). Compared to the KD = 14.7 nM of the prototype RBD[9], the equilibrium dissociation constant (KD) of the three mutants are calculated as 0.12 nM, 0.11 nM, and 0.13 nM, respectively (**Figure. 2C**), which were two orders of magnitude lower than for the prototype strain, indicating a remarkably increased affinity to the human ACE2 receptor. For the only double amino acid mutant (N354D, D364Y), the N354D substitution decreased the affinity, while the D364Y provided a higher affinity than the double mutation (**Figure. 2B**). This indicated that the D364Y is the major contributor to the enhanced affinity.

In comparison, the bat CoV RBD (strain RaTG13, with the highest genome similarity) showed a much lower binding affinity (KD=1.17mM; ΔG=-17.4kJ/mol) to human ACE2 than the pangolin CoV (KD=1.89μM; ΔG=-33.9kJ/mol). For comparison, the affinity of the pangolin CoV was slightly lower than the SARS-CoV-2 prototype strain (KD=14.7nM; ΔG=-46.5kJ/mol) (**Figure. 2B, 2C**).

Figure. 2: Binding free energy calculated for the SARS-CoV-2 RBD variants to human ACE2 from January through March, 2020.

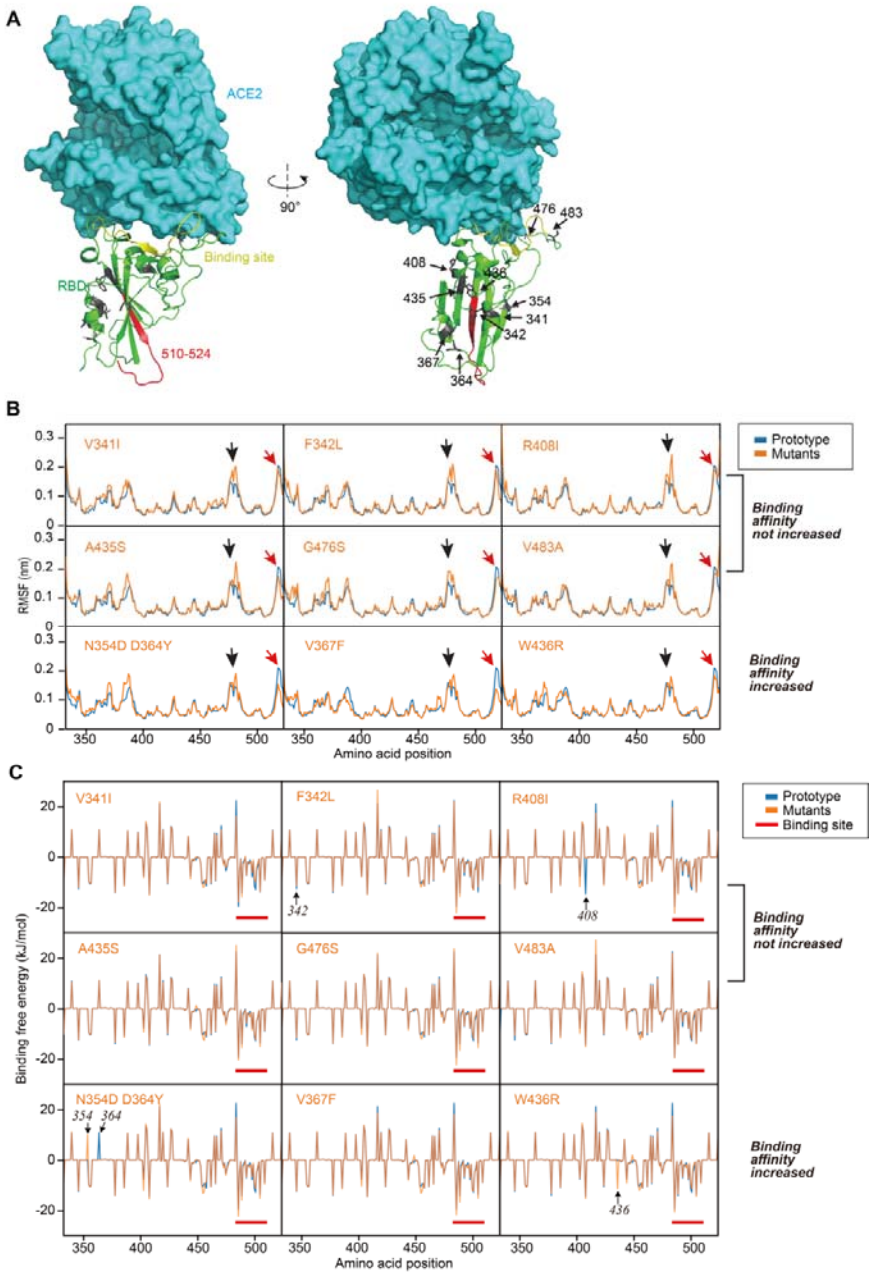


(A) Root Mean Square Deviation (RMSD) of typical MD trajectories of the SARS-CoV-2 prototype and the mutant strains (V341I, F342L, R408I, A435S, G476S, V483A, N354D D364Y, V367F, W436R). **(B)** Comparison of the binding free energy (ΔG) of the RBDs and the human ACE2. Note, the ΔG is inversely proportional to the binding affinity. Data are presented as mean±SD. *P*-values were calculated using single-tailed student t-test. The *P*-values are shown for those with $P < 0.05$. The ΔG calculated from experimental K_D values of SARS and SARS-CoV-2 prototype are marked in dotted and dashed lines, respectively. **(C)** Comparison of the equilibrium dissociation constants (K_D) as calculated with the ΔG.

2.4. Structural basis for the increased affinity of the SARS-CoV RBD mutants

The nine RBD mutant types that emerged during the early transmission phase were divided into two groups: a “similar affinity” group (V341I, F342L, R408I, A435S, G476S, V483A), whose affinity is not significantly increased, and a “higher affinity” group (N354D D364Y, V367F, W436R), whose affinity is significantly increased. To explain the structural basis of the increased affinity, we investigated the dynamics of the residues in these structures in greater detail. The binding surface of the RBD to ACE2 is largely arrayed in random coil conformation, which lacks structural rigidity. Logically, a firm scaffold should be necessary to maintain this conformation of the interaction surface, and thus may facilitate the binding affinity. The beta-sheet structure scaffold, centered by residues 510-524 (**Figure. 3A**, marked as red), apparently provides this rigidity. “Higher affinity” mutants (N354D D364Y, V367F, and W436R) showed a considerable decrease of the RMSF (Root Mean Square of Fluctuation) at this region, demonstrating a more rigid structure; this was not observed for the “similar affinity” mutants (**Figure. 3B**). Coincidentally, the substitutions that account for the affinity increase (D364Y, V367F, and W436R) are all located near this fragment. Indeed, residues 475-485, which is a random coil near the binding site, showed a remarkably higher RMSF for the “similar affinity” group mutants, in contrast to the “higher affinity” group mutants (**Figure. 3B**). Moreover, the “higher affinity” group exhibited a general decreased ΔG in the binding site region, in contrast to the “similar affinity” group mutants (**Figure. 3C**). In addition, the D364Y and W436R mutations directly contributed to the ΔG decrease. In contrast, the N354D mutation directly elevated the ΔG (**Figure. 3B**).

Figure. 3: Structural analysis of RBD mutants and the effects on their binding affinity.



(A) The binding surface and interaction of the RBD to ACE2, with the locations of the mutant amino acids noted. Beta-sheet structure scaffold was centered by residues 510-524 (in red). **(B)** Root Mean Square of Fluctuation (RMSF) of the nine mutants were compared to that of the prototype (V341I, F342L, R408I, A435S, G476S, V483A, N354D D364Y, V367F, and W436R). Red arrows denote the fragment of residues

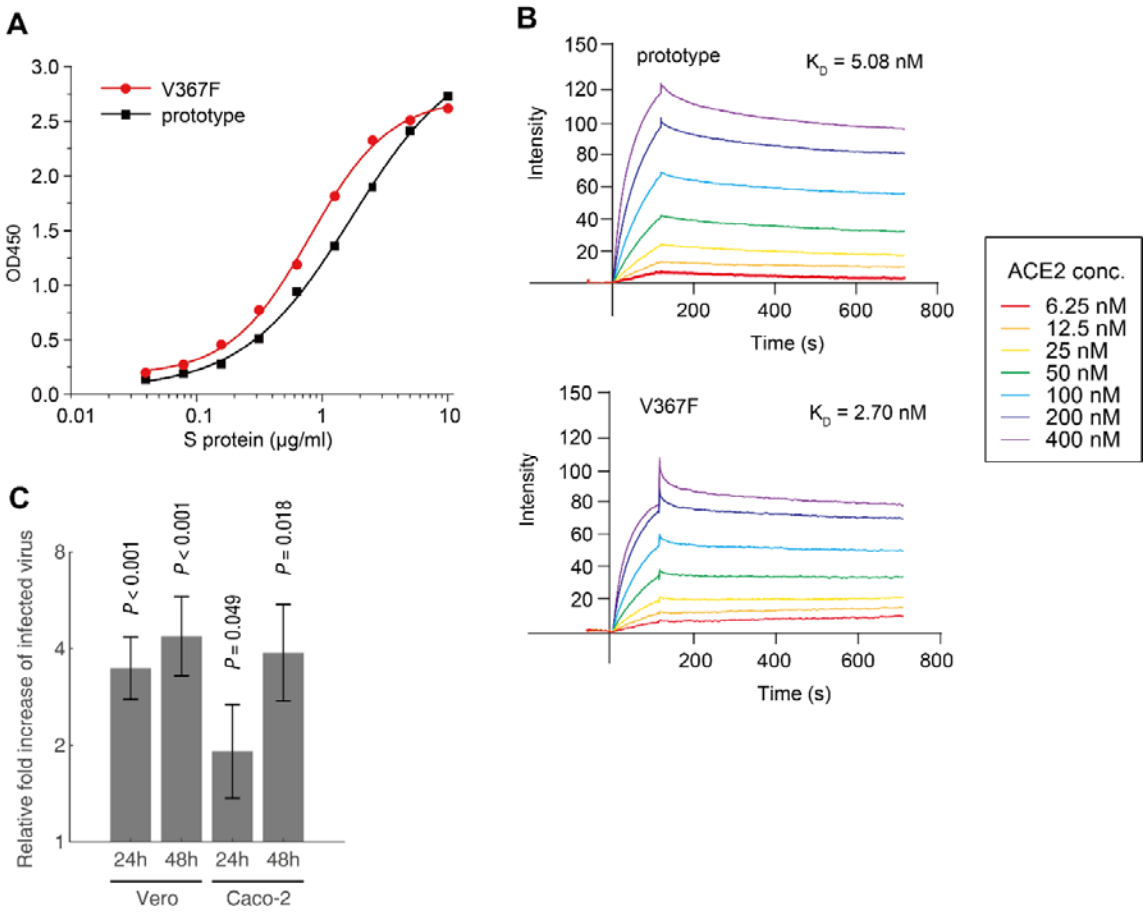
510-524. Black arrows denote the fragment of residues 475-485. **(C)** Contribution of each amino acid to the binding free energy. Red bars denote the binding site.

2.5. Experimental validation of the enhanced affility and infectivity of the V367F mutant

As of August 24, 2020, among all of the mutants with dominant mutations in spike RBD (>20 strains), most mutations resulted in a substitution of amino acids with similar properties. V367F is the only mutant with higher binding affinity, as calculated by MD simulation (**Table 2**). The other "higher affinity" mutants (N354D D364Y, and W436R) that were observed during the early transmission phase could not be detected again. Therefore, the binding affinity and the infectivity of V367F mutant were further validated experimentally. First, we performed experiments to assess the binding affinity *in vitro* with a receptor-ligand binding ELISA assay using purified S proteins and human ACE2 protein. The result showed that the V367F mutation significantly lowered the ED_{50} concentration (ED_{50} = 0.8±0.04 μg/ml), as compared to the prototype (ED_{50} = 1.7±0.14 μg/ml) (**Figure. 4A**). This demonstrates that the V367F mutant has higher affinity than the prototype. Second, we performed surface plasmon resonance (SPR) experiments, which yielded the same conclusion: the prototype had a K_D of 5.08 nM, compared to the V367F mutant with a K_D of 2.70 nM (**Figure. 4B**). These results qualitatively validated our computational simulations. Next, we performed an *in vivo* experiment to investigate the invasion efficiency of S proteins using a pseudovirus assay. Same amounts of S protein-containing pseudovirus were subjected to infection of ACE2-overexpressed Vero and Caco-2 cells. A higher infection efficiency is represented by the increased copy number of the integrated lentivirus genome. At 24 hours post-infection, the V367F mutant pseudovirus showed 6.08x higher copy numbers than the prototype in Caco-2 cells. After 48 hours, the V367F mutant

pseudovirus showed 4.38x and 3.88x higher copy number than the prototype in Vero and Caco-2 cells, respectively (**Figure. 4C)**. These enhancements were statistically significant ($P<0.01$, t-test). Therefore, the *in vitro* and *in vivo* results were consistent with enhanced affinity and infectivity for the V367F mutant.

Figure.4: Experimental validation of the enhanced affinity and infectivity of the V367F mutant.

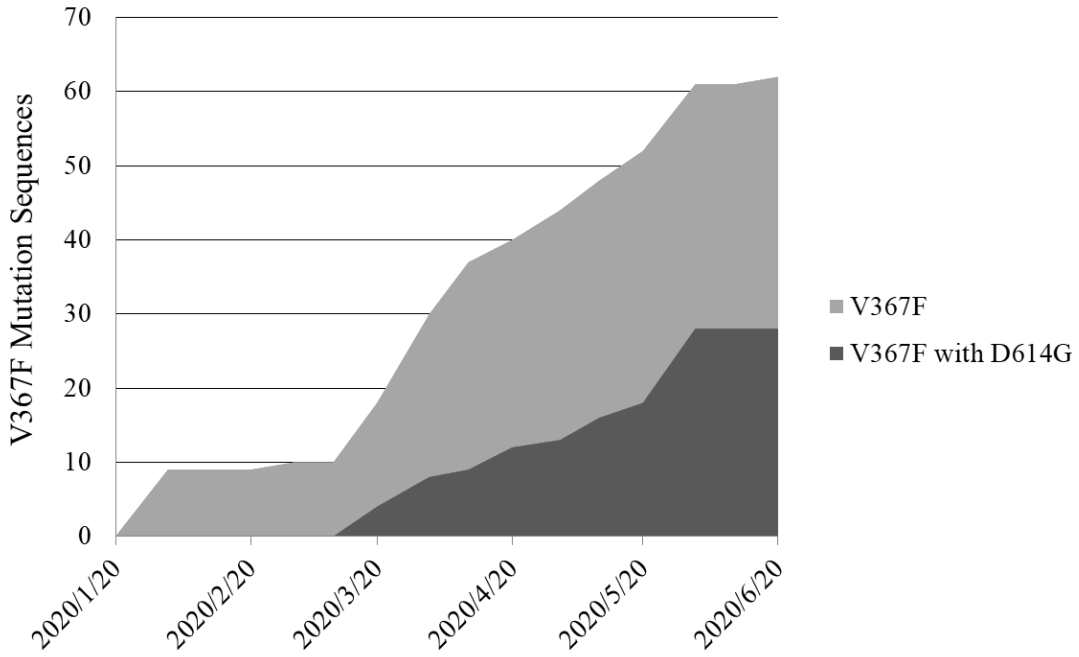


(A) Comparison of the binding affinity of prototype S protein and V367F mutant to human ACE2 by ligand-receptor binding ELISA assay. **(B)** Comparison of the binding affinity of prototype S protein and V367F mutant to human ACE2 protein by SPR. **(C)** Quantification of the genome copy number of the V367F mutant vs. the prototype using pseudovirus infection assay. The relative fold increases of infected virus are shown. Relative fold increase of pseudoviral DNA copy number of V367F mutant can be detected in both Vero and Caco-2 cells at 24h and 48h post-infection. Experiments were performed in triplicates, and the *P*-values were calculated using two-tailed t-test for two samples with different variances.

2.6. The convergence of SARS-CoV-2 RBD and D614G mutation in dominant mutation strains

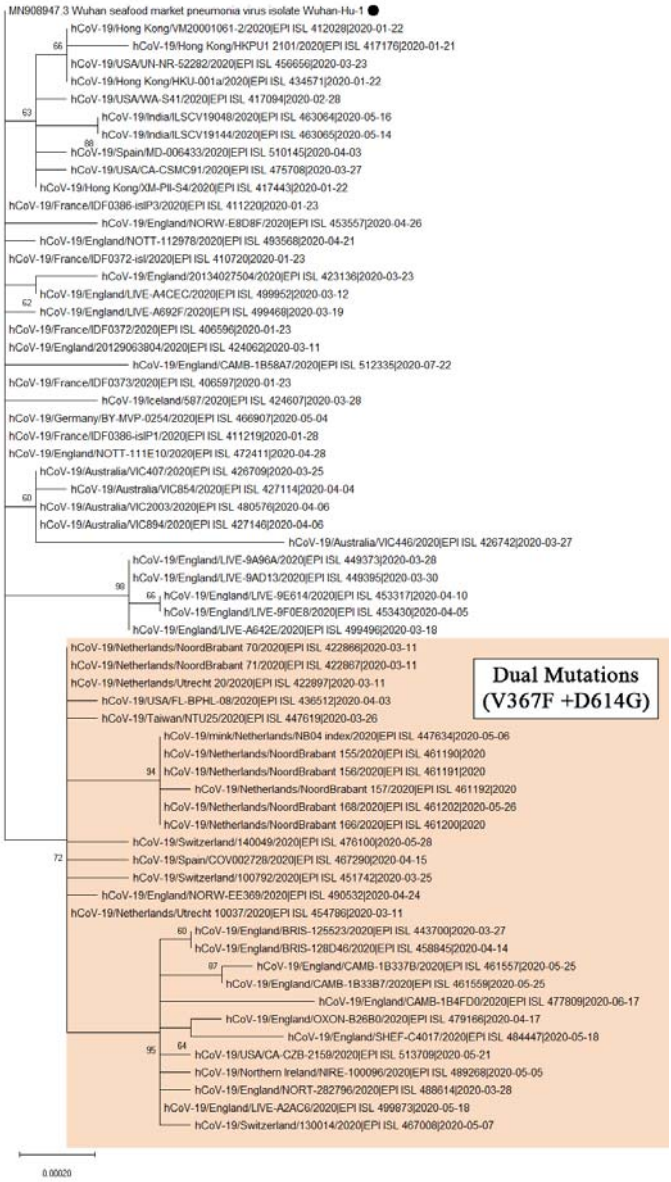
The D614G mutation is located in the S1 region and is outside of the RBD of the SARS-CoV-2 spike protein. It has been confirmed that the D614G change increases virus infectivity by elevating its sensitivity to protease and that this mutation has spread widely[26]. Among all of the dominant mutations (>20 strains) in the spike RBD, V367F, S477N, V483A, G485R, A520S, P384L, A522V, and P330S were detected along with the D614 prototype during the early transmission phase. Later, these mutations were detected with the D614G mutant (**Table 2**). The V367F mutants were initially discovered in January in Hong Kong, and the D614G+V367F dual mutant was initially discovered in March in the Netherlands. Afterwards, detection of both the V367F single mutant and the dual mutant increased rapidly (**Figure. 5**). Additionally, the distribution of V367F mutants and D614G/V367F dual mutants emerged mainly in Europe, including United Kingdom, the Netherlands, Spain, Northern Ireland, Switzerland, and Iceland, as well as in the USA, Australia, and Taiwan (**Supplementary Table 2**). The phylogenetic analysis of the V367F mutant genomes showed that most V367F mutants during the early transmission phase clustered more closely with the SARS-CoV-2 prototype strain. Intriguingly, all of the dual-mutation variants (V367F + D614G) formed a distinct sub-cluster separate from the V367 sub-cluster (**Figure. 6**). This indicates that the V367F mutation may have evolved along with D614G mutation, suggesting a synergistic effect of increased infectivity.

Figure. 5: Emergence of V367F and V367F + D614G mutants (January through June, 2020).



The number of V367F mutation and dual-mutation with D614G sequences are shown with respect to the early phase of the pandemic. The number of the V367F and V367F+D614G mutants and the sample collection dates are shown. The V367F+D614G mutants were first detected in March and increased rapidly.

Figure. 6: The whole genome phylogenetic analysis of the V367F mutants (January through August, 2020).



The whole genome phylogenetic analysis, including all V367F mutants. For reference, the prototype strain Wuhan-HU-1 is marked with a filled circle. A phylogenetic tree was constructed using the maximum likelihood method with 1000 bootstrap replicates (MEGA X), and applying default parameters[27]. The distinct sub-cluster formed by all of the dual-mutation variants (V367F + D614G) is indicated in the orange box.

3. Discussion

Due to the challenging and on-going pandemic and given the evolving nature of the SARS-CoV-2 virus globally, identifying changes in viral infectivity may prove crucial to containing the COVID-19 pandemic. Quarantine policies need to be adapted with respect to the viral infectivity change. It is always a dilemma of quarantine and economy. Any government would balance the lost due to the quarantine lockdown versus the lost due to the disease. Numerous models have been rasied to estimate the two costs. For example, if the viral strain is more infectious, more stringent lockdown measure would be expected. This report provides computational and experimental insights into the functional outcome of mutations in RBD. As RBD mutations are under positive selection pressure, we identified the mutants that acquired increased binding affinity to human ACE2 receptor and presumably higher infectivity for human cells.

First, our analysis of molecular dynamics simulation indicated a remarkable enhancement of the affinity efficiency of multiple mutant S proteins. Compared to the prototype strain Wuhan-Hu-1, the ΔG of mutants decreased ~25%. These mutants bind to ACE2 more stably due to the enhancement of the base rigidity of the S protein structure. Potential and recent animal-to-human transmission events of SARS-CoV-2 may explain the strong positive selection and enhancement of the affinity during the pandemic. Ongoing adaptation to transmission and replication in humans, including mutation events in the RBD may boost the binding affinity and lead to an increase in the basic reproduction number (R0), and in theory, further enable human to human transmission.

The origin of this virus has been of considerable interest and speculation since the outbreak. Due to the high sequence similarity of the bat SARS-like CoV genome and the pangolin CoV RBD sequence to the SARS-CoV-2 sequences, these hosts were thought to have initiated the infection in humans[6–8]. Our results provide more clues to support this hypothesis. Our results suggest that the binding energy of the bat SARS-like CoV RBD is too high to bind human ACE2 effectively (K_D in millimolar range). In

contrast, the pangolin CoV showed a binding K_D for human ACE2 at the micromolar range, just ~6x higher than that of human SARS virus (K_D = 0.326μM) [28](**Figure. 3**), which indicates that the pangolin CoV has the potential to infect humans in unprotected close contact. Alignments of the genomic sequences of SARS-CoV-2 and pangolin CoV viruses suggest recombination events, particularly in the RBD domain, between pangolin and bat viruses[6]. The pangolin CoV has been detected among intercepted smuggled Malayan pangolins in multiple provinces in China, suggesting and supporting a high risk of zoonotic infections to humans, constantly and widely[7,8].

It should be noted that during the early transmission phase, the mutation V367F that enhances the binding affinity was found in six strains: one in Hong Kong and five in France. As RBD is conserved in SARS-CoV-2, the coincidence of six strains with the same mutation across large geographic distances indicates that this mutant is more robust and that these strains originated as a novel sub-lineage, given the close isolation dates (January 22 and 23, respectively). An alternate view is that asymptomatic individuals with the same mutation were "super-infecting" travellers. Along with the epidemiological data, mutation surveillance is of critical importance as it can reveal more exact transmission routes of the epidemic and provide early warnings of additional outbreaks. Emergence of SARS-CoV-2 strains in Hong Kong, France, and other countries with RBD mutations allowing higher binding affinity to human ACE2 receptor suggests an increased risk and more severe morbidity and mortality during a sustained pandemic of COVID-19, particularly if effective precautions are not implemented. By performing assays comparing the prototype spike protein to the V367F mutation containing counterpart, we confirmed the significantly enhanced binding affinity and likely higher infectivity of the V367F mutant, and showed that the mutation stabilizes the RBD structure.

By tracking mutation types in the SARS-CoV-2 spike RBD, only the V367F was continuously observed. The N354D, D364Y, and W436R mutations seemed to have disappeared during the course of the

pandemic, perhaps in part due to the extremely strict isolation policy set in China. Most of the dominant RBD mutants were first detected with the prototype D614, and then together with the G614 mutation. This is possibly due to multiple individual recombination events between the D614+V367F mutants and G614 mutants, although it is difficult to determine the exact position and time point due to the high sequence identities among SARS-CoV-2 mutants. D614G is distinct from the RBD mutationsV367F – it is not located in the RBD but enhances viral infectivity by elevating its sensitivity to protease[26]. Therefore, D614G and V367F may function independently and have synergistic effects on viral infectivity. Recombination is known to play an important role in natural coronavirus evolution, which may contribute to the convergence of dual enhancing mutants (D614G + V367F)[16,18,29,30]. More attention should be paid to the risk of the advantage accumulation evolution through recombination among the variants, in which the recombinants may be more infectious and also better at immune escape.

The S protein is also important for antigen recognition. By tracking dominant RBD mutation sequences up to August, 2020, multiple mutants with more than 20 sequences putatively related to weakly reduced host receptor binding and altered antigenicity have been detected[17]. These were found by comparing equivalent positions in more than 500 sequences of SARS and MERS genomes, in particular the N439K mutation[17]. Equivalent positions have been studied for V483A in the MERS-CoV genomes and for N439K in the SARS-CoV genomes. These mutations are hypothesized to result in weakly reduced host receptor binding and altered antigenicity, revealing possible immune escape driving virus evolution[31],[32]. Since the RBD contains important antigenic epitopes, frequent mutations in RBD, especially those that change the amino acid properties, may weaken the binding affinity of any antibody raised against the prototype strain[20,21]. This may lead to decreased vaccine efficacy and should be further studied.

In summary, we have identified 32 RBD mutant strains clustering into nine mutant types during the early transmission phase. Three mutant types that emerged in Asia and Europe display apparent enhanced

structural stability of the spike protein along with higher binding affinities to the human ACE2 receptor. Multiple mutants apparently disappeared under high positive selection pressure; only the V367F mutation, which showed higher infectivity, was persistently observed along with the D614G mutation, the latter not in the RBD. This indicates that the V367F mutant is stable and may have acquired increased infectivity for humans during the COVID-19 pandemic. The emergence of dual mutants (V367F+D614G) suggests recombination among SARS-CoV-2 genomes may increase the infectivity of the virus along with enhanced escape from the host immune response. These findings support the continuing surveillance of spike mutations to aid in the development of COVID-19 drugs and vaccines.

4. Methods

4.1. Genome sequence dataset in this study

Full-length protein sequences of S protein RBD were downloaded from the NCBI GenBank Database, China 2019 Novel Coronavirus Resource (https://bigd.big.ac.cn/ncov), and GISAID EpiFluTM Database (http://www.GISAID.org). 1609 SARS-CoV-2 full-genome sequences isolated during early transmission phase (before April 5, 2020) were downloaded and the sequences with amino acid mutations in S protein and RBD region were parsed . The genome sequences with amino acid mutations in S protein and the RBD were analyzed in this study (**Supplementary Table 1**).

For tracking the V367F mutation sequences and the dominant spike mutations “to date”, 83,758 SARS-CoV-2 spike amino sequences isolated before August 24, 2020 were retrieved from the GISAID EpiFluTM Database (http://www.GISAID.org) and aligned. The genome sequences with either theV367F mutation or the V367F/D614G dual mutations in the S protein RBD were screened and analyzed in this study (**Supplementary Table 2)**.

4.2. Sequences alignment and phylogenetic analyses

Alignment of S protein sequences from different sources and comparison of ACE2 proteins among different species were performed using MAFFT version 7, with default parameters（https://mafft.cbrc.jp/alignmeloadnt/server/ ）and Bioedit[33,34]. Selection pressure analyses were conducted

using the Nei-Gojobori model[22]. Phylogentic analyses were conducted with Mega X（version 10.0.2, applying the maximum-likelihood method with 500 bootstrap replicates and default parameters[27].

4.3. Molecular dynamics (MD) simulation

The complex structure of the SARS-CoV-2 S-protein RBD domain and human ACE2 was obtained from the Nation Microbiology Data Center (ID: NMDCS0000001) (PDB ID: 6LZG). Mutated amino acids of the SARS-CoV-2 RBD mutants were directly replaced in the model, and the bat/pangolin CoV RBD domain was modelled using SWISS-MODEL[25]. Molecular dynamics simulation was performed using GROMACS 2019 with the following options and parameters: explicit solvent model, system temperature 37°C, OPLS/AA all-atoms force field, and LINCS restraints. With 2fs steps, each simulation was performed at 10ns, and each model was simulated three times to generate three independent trajectory replications. Binding free energy (ΔG) was calculated using MM-PBSA method (GitHub; https://github.com/Jerkwin/gmxtool), with the trajectories after structural equilibrium assessed using RMSD (Root Mean Square Deviation)[24,35]. To calculate the equilibrium dissociation constant (K_D) and ΔG, the formula ? $G = RTln K_D$ was used. Estimated ΔGs of the RBD mutants were normalized using the ΔG of the prototype strain which was derived from experimental data[28].

4.4. Recombinant S protein mutant expression

The SARS-CoV-2 prototype S gene was cloned into pNPM5 vector (Novoprotein, NJ, USA), and fused with a C-terminal His_6-tag. The V367F mutation was introduced using site-directed mutagenesis, consistent with the nucleotide sequence of the actual isolate. These two constructs were transfected into HEK293 cells using polyethyleneimine. Since the S protein includeed a signal peptide in its N-terminal 14 amino acids, it was secreted into the medium. The expressed proteins were purified from filtered cell supernatants with a Ni-NTA column. Eluted protein solution was then dialyzed against PBS (pH7.4) for subsequent assays.

4.5. Ligand-receptor binding ELISA assay

Human ACE2 protein was immobilized onto a microtiter plate at 5 μg/ml (100μl/well). Each S protein (prototype and V367F) was added as a ligand at different concentrations, ranging from 0.03 μg/ml to 10

μg/ml, and then incubated for 2 hours at 37°C to allow receptor-ligand interaction. The ligand / receptor mixture was then washed three times. 100μl of HRP anti-His Tag Antibody (BioLegend, USA) (diluted 1:20000) was added to each well, and allowed to react for 1 hour. After three washes, the signal was visualized using TMB solution (Sigma-Aldrich, USA) and a microtiter plate reader recording OD450.

4.6. Surface Plasmon Resonance (SPR) experiments

The SPR experiments were performed in a BIAcore T200 instrument (GE, USA). For this, the SARS-CoV-2 S-proteins, either prototype or V367F, were immobilized on the Sensor Chip NTA (GE, USA), according to the manufacturer's protocol. Human ACE2 protein was injected in each experiment in seven concentrations (6.25, 12.5, 25, 50, 100, 200, and 400 nM). For each cycle, the absorption phase lasted for 120 seconds and the dissociation phase lasted for 600 seconds. After each cycle, the chip was regenerated using 350mM EDTA and 50mM NaOH wash for 120 seconds,. Blank controls with 0 nM ACE2 were performed, with the blank signals subtracted from the cycle signals. All experiments were performed at 37°C. K_D values were calculated by fitting the curves using the software provided with the instrument.

4.7. Production and titration of SARS-CoV-2 pseudoviruses bearing V367F S protein

The full-length S gene of SARS-CoV-2 strain Wuhan-Hu-1(NC_045512.2) was cloned into a SARS-CoV-2 Spike vector (PackGene, Guangzhou, China), and confirmed by DNA sequencing. Plasmid SARS-CoV-2 Spike (G1099T), incorporating the V367F mutation in the S gene, was constructed by site-directed mutagenesis using the ClonExpress MultiS One Step Cloning Kit (Vazyme), as per the manufacturer's protocol.

Generation of SARS-CoV-2 S pseudovirus was done as previously described with some modifications[36]. Briefly, 293T cells, at about 70-90% confluence, were co-transfected with 9 ug of the transfer vector (pLv-CMV), 6 ug packaging plasmid (psPAX-lentiviral), and 6 ug envelope plasmid (pCB-spike or pCB-spikeV367F). Pseudoviruses were harvested at 48 h post-transfection, using a 2.5 ml sterile syringe and subsequently filtered into either Falcon or microcentrifuge tubes via a syringe driven 0.45 μm filter. Virus titration was measured by RT-qPCR targeting the WPRE gene of pseudoviruses and using the Hifair® □

One Step RT-qPCR SYBR Green Kit (Yeasen), as per the manufacturer's protocol. After reverse transcription (10 min at 42°C) and initial denaturation (5 min at 95°C), PCR amplification was performed for 40 cycles (15 s at 95°C, 60 s at 60°C). Primers were as follows: WPRE-F, 5'-CACCACCTGTCAGCTCCTTT-3'; WPRE -R, 5'- ACGGAATTGTCAGTGCCCAA-3'.

4.8. SARS-CoV-2 spike-mediated pseudovirus entry assay.

To detect S variants mediated viral entry, VERO E6 and Caco2 cells (5×10^4) grown in 24-well plates were respectively infected with either 20 MOI of S-V367 or S-F367-bearing pseudovirus ($1x10^7$ pseudovirus copies per ml). The cell medium was replaced by 500 ul fresh DMEM medium 6h post-infection. Relative fold increase of infected virus titers was calculated according to the WPRE DNA copy number of the lentiviral proviruses measured by TB Green® Premix Ex Taq ™ II (Takara) at 24 and 48 h post-infection. Data from all of the samples were obtained from three independent experiments, and each sample was tested in triplicate.

Acknowledgements

We gratefully acknowledge the authors, originating and submitting laboratories of the sequences from GISAID's EpiFlu™ Database on which this research is based. All submitters of data may be contacted directly via www.gisaid.org. This work was supported by grants from the National Key Research and Development Program of China (2017YFA0505001/2018YFC0910200/2018YFE0204503), the National Natural Science Foundation of China (81730061), the Guangdong Key Research and Development Program (2019B020226001), the Natural Science Foundation of Guangdong Province (2018B030312010), and the Guangzhou Healthcare Collaborative Innovation Major Project (201803040004 and 201803040007). J. Ou, Z. Zhou & R. Dai contributed equally to this work.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

[1] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G. F. Gao, W. Tan, *N. Engl. J. Med.* **2020**, 727.

[2] F. Wu, S. Zhao, B. Yu, Y. M. Chen, W. Wang, Z. G. Song, Y. Hu, Z. W. Tao, J. H. Tian, Y. Y. Pei, M. L. Yuan, Y. L. Zhang, F. H. Dai, Y. Liu, Q. M. Wang, J. J. Zheng, L. Xu, E. C. Holmes, Y. Z. Zhang, *Nature* **2020**, *579*, 265.

[3] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, Y. Zhao, Y. Li, X. Wang, Z. Peng, *JAMA - J. Am. Med. Assoc.* **2020**, 1.

[4] WHO, Coronavirus disease (COVID-19) Weekly Epidemiological Update. 2020, 24 August. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200824-weekly-epi-update.pdf?sfvrsn=806986d1_4

[5] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. M. Leung, E. H. Y. Lau, J. Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T. T. Y. Lam, J. T. Wu, G. F. Gao, B. J. Cowling, B. Yang, G. M. Leung, Z. Feng, *N. Engl. J. Med.* **2020**, 1.

[6] T. T. Y. Lam, N. Jia, Y. W. Zhang, M. H. H. Shum, J. F. Jiang, H. C. Zhu, Y. G. Tong, Y. X. Shi, X. B. Ni, Y. S. Liao, W. J. Li, B. G. Jiang, W. Wei, T. T. Yuan, K. Zheng, X. M. Cui, J. Li, G. Q. Pei, X. Qiang, W. Y. M. Cheung, L. F. Li, F. F. Sun, S. Qin, J. C. Huang, G. M. Leung, E. C. Holmes, Y. L. Hu, Y. Guan, W. C. Cao, *Nature* **2020**, *583*, 282.

[7] K. Xiao, J. Zhai, Y. Feng, N. Zhou, X. Zhang, J. J. Zou, N. Li, Y. Guo, X. Li, X. Shen, Z. Zhang, F. Shu, W. Huang, Y. Li, Z. Zhang, R. A. Chen, Y. J. Wu, S. M. Peng, M. Huang, W. J. Xie, Q. H. Cai, F. H. Hou, W. Chen, L. Xiao, Y. Shen, *Nature* **2020**, *583*, 286.

[8] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G.-F. Xiao, Z.-L. Shi, *Nature* **2020**, DOI: 10.1038/s41586-020-2012-7.

[9] D.Wrapp, N.Wang, K.S.Corbett, J.A.Goldsmith, C.L.Hsieh, O.Abiona, B.S.Graham, *Science (80-.).* **2020**, *2507*, 1.

[10] Y. Wan, J. Shang, R. Graham, R. S. Baric, F. Li, *J. Virol.* **2020**, DOI: 10.1128/jvi.00127-20.

[11] Y. Chen, Y. Guo, Y. Pan, Z. J. Zhao, *Biochem. Biophys. Res. Commun.* **2020**, *2*, 0.

[12] J. B. Reese, S. L. Bober, M. B. Daly, C. Program, F. Chase, B. Sciences, P. Care, F. Chase, S. Behavior, B. Neoplasms, *Nature* **2016**, *531*, 118.

[13] A. C. Walls, M. A. Tortorici, B. Bosch, B. Frenz, P. J. M. Rottier, F. Dimaio, F. A. Rey, D. Veesler, **2016**, *531*, 114.

[14] M. Letko, A. Marzi, V. Munster, *Nat. Microbiol.* **2020**, *5*, DOI: 10.1038/s41564-020-0688-y.

[15] X. Xu, P. Chen, J. Wang, J. Feng, H. Zhou, X. Li, W. Zhong, P. Hao, *Sci. China Life Sci.* **2020**, *63*, 457.

[16] Q. Rehman, S. U., Shafique, L., Ihsan, A. & Liu, *Pathogens* **2020**, *9*.

[17] Q. Li, J. Wu, J. Nie, L. Zhang, H. Hao, S. Liu, C. Zhao, Q. Zhang, H. Liu, L. Nie, H. Qin, M. Wang, Q. Lu, X. Li, Q. Sun, J. Liu, L. Zhang, X. Li, W. Huang, Y. Wang, *Cell* **2020**, 1.

[18] X. Li, E. E. Giorg, M. H. Marichannegowda, B. Foley, C. Xiao, X. P. Kong, Y. Chen, S. Gnanakaran, B. Korber, F. Gao, *Sci. Adv.* **2020**, *6*, 1.

[19] A. Grifoni, J. Sidney, Y. Zhang, R. H. Scheuermann, B. Peters, A. Sette, *Cell Host Microbe* **2020**, *27*, 671.

[20] D. Wang, J. Mai, W. Zhou, W. Yu, Y. Zhan, N. Wang, N. D. Epstein, Y. Yang, *Vaccines* **2020**, *8*, 1.

[21] S. F. Ahmed, A. A. Quadeer, M. R. McKay, *Viruses* **2020**, 2020.02.03.933226.

[22] M. Nei, T. Gojoborit, *Mol. Biol. Evol.* **1986**, *3*, 418.

[23] N. Homeyer, H. Gohlke, *Mol. Inform.* **2012**, *31*, 114.

[24] J. M. Sanders, M. E. Wampole, M. L. Thakur, E. Wickstrom, **2013**, *8*, 8.

[25] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. De Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, *Nucleic Acids Res.* **2018**, *46*, W296.

[26] B. Korber, W. M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, N. Hengartner, E. E. Giorgi, T. Bhattacharya, B. Foley, K. M. Hastie, M. D. Parker, D. G. Partridge, C. M. Evans, T. M. Freeman, T. I. de Silva, A. Angyal, R. L. Brown, L. Carrilero, L. R. Green, D. C. Groves, K. J. Johnson, A. J. Keeley, B. B. Lindsey, P. J. Parsons, M. Raza, S. Rowland-Jones, N. Smith, R. M. Tucker, D. Wang, M. D. Wyles, C. McDanal, L. G. Perez, H. Tang, A. Moon-Walker, S. P. Whelan, C. C. LaBranche, E. O. Saphire, D. C. Montefiori, *Cell* **2020**, *182*, 812.

[27] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, *Mol. Biol. Evol.* **2018**, *35*, 1547.

[28] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, J. Shaman, *Science (80-.).* **2020**, *368*, 489.

[29] R. T. Y. So, D. K. W. Chu, E. Miguel, R. A. P. M. Perera, J. O. Oladipo, O. Fassi-Fihri, G. Aylet, R. L. W. Ko, Z. Zhou, M.-S. Cheng, S. A. Kuranga, F. L. Roger, V. Chevalier, R. J. Webby, P. C. Y. Woo, L. L. M. Poon, M. Peiris, *J. Virol.* **2019**, *93*, 1.

[30] S. K. P. Lau, Y. Feng, H. Chen, H. K. H. Luk, W. Yang, K. S. M. Li, Y. Zhang, **2015**, *89*, 10532.

[31] H. Kleine-Weber, T. Elzayat, L. Wang, M. T. Elzayat, L. Wang, B. S. Graham, M. A. Muller, C. Drosten, S. Pohlmann, M. Hoffmann, *J. Virol.* **2019**, *93*, 1.

[32] B. Rockx, E. Donaldson, M. Frieman, T. Sheahan, D. Corti, A. Lanzavecchia, R. S. Baric, *J. Infect. Dis.* **2010**, *201*, 946.

[33] K. Katoh, J. Rozewicki, K. D. Yamada, *Brief. Bioinform.* **2018**, *20*, 1160.

[34] S. Kuraku, C. M. Zmasek, O. Nishimura, K. Katoh, *Nucleic Acids Res.* **2013**, *41*, 22.

[35] E. A. Rifai, M. Van Dijk, N. P. E. Vermeulen, A. Yanuar, D. P. Geerke, *J. Chem. Inf. Model.* **2019**, *59*, 4018.

[36] G. Zhao, L. Du, C. Ma, Y. Li, L. Li, V. K. Poon, L. Wang, F. Yu, B. J. Zheng, S. Jiang, Y. Zhou, *Virol. J.* **2013**, *10*, 1.

Tables

Table 1: Nucleotide substitution rates and selection pressures for S gene.

Gene	Length (bp)	Mean Non-synonymous Substations/site	Mean Synonymous Substations/site	dN/dS
S	3822	0.7726	0.1875	4.1197
S1	2043	0.6207	0.0571	10.8636
S1-RBD	585	0.0458	0.0137	3.3545
S2	1779	0.1519	0.1304	1.1646

The number of nonsynonymous and synonymous differences per sequence from averaging over all sequence pairs are shown. Analyses were conducted using the Nei-Gojobori model[22]. The analysis involved 1609 SARS-CoV-2 S gene sequences.

Table 2: Dominant mutations in the spike RBD (>20 strains) and their binding affinity change (As of August 24, 2020).

aa Position (in Spike)	aa Prototype	aa Mutation	Sequences	First Occurrence Date	First Occurrence ID	D614 count	G614 count	Total	Receptor Binding increase
439	N	K	557	2020.3.16	EPI_ISL_425684	0	557	557	No
477	S	N	291	2020.3.28	EPI_ISL_460393	1	287	288	No
478	T	I	106	2020.3.18	EPI_ISL_440368	0	104	104	No
479	P	S	76	2020.4.13	EPI_ISL_470698	0	66	66	No
367	V	F	63	2020.1.21	EPI_ISL_417176	35	28	63	Yes
483	V	A	38	2020.2.29	EPI_ISL_417159	36	1	37	No
485	G	R	32	2020.2.6	EPI_ISL_450418	1	31	32	No
520	A	S	28	2019.12.30	EPI_ISL_402124	6	14	20	No
341	V	I	27	2020.3.10	EPI_ISL_440476	0	27	27	No
384	P	L	25	2020.3.26	EPI_ISL_425960	3	18	21	No
522	A	V	24	2020.3.25	EPI_ISL_421654	5	19	24	No
330	P	S	22	2020.3.13	EPI_ISL_454225	2	20	22	No

The number of dominant mutations in the spike RBD (as of August 24, 2020) are shown. The First Sequence Occurrence date and ID, and D614 sequences number and G614 sequences number are identified. S477N, V367F, G485R, A520S, P384L, A522V, and P330S mutations were first detected in D614 sequences, and then with the G614 mutation. Receptor binding efficiencies of dominant mutation are evaluated by amino acid property change and by binding free energy using the MM-PBSA method[35]. aa: amino acid.

Supplementary Figure 1

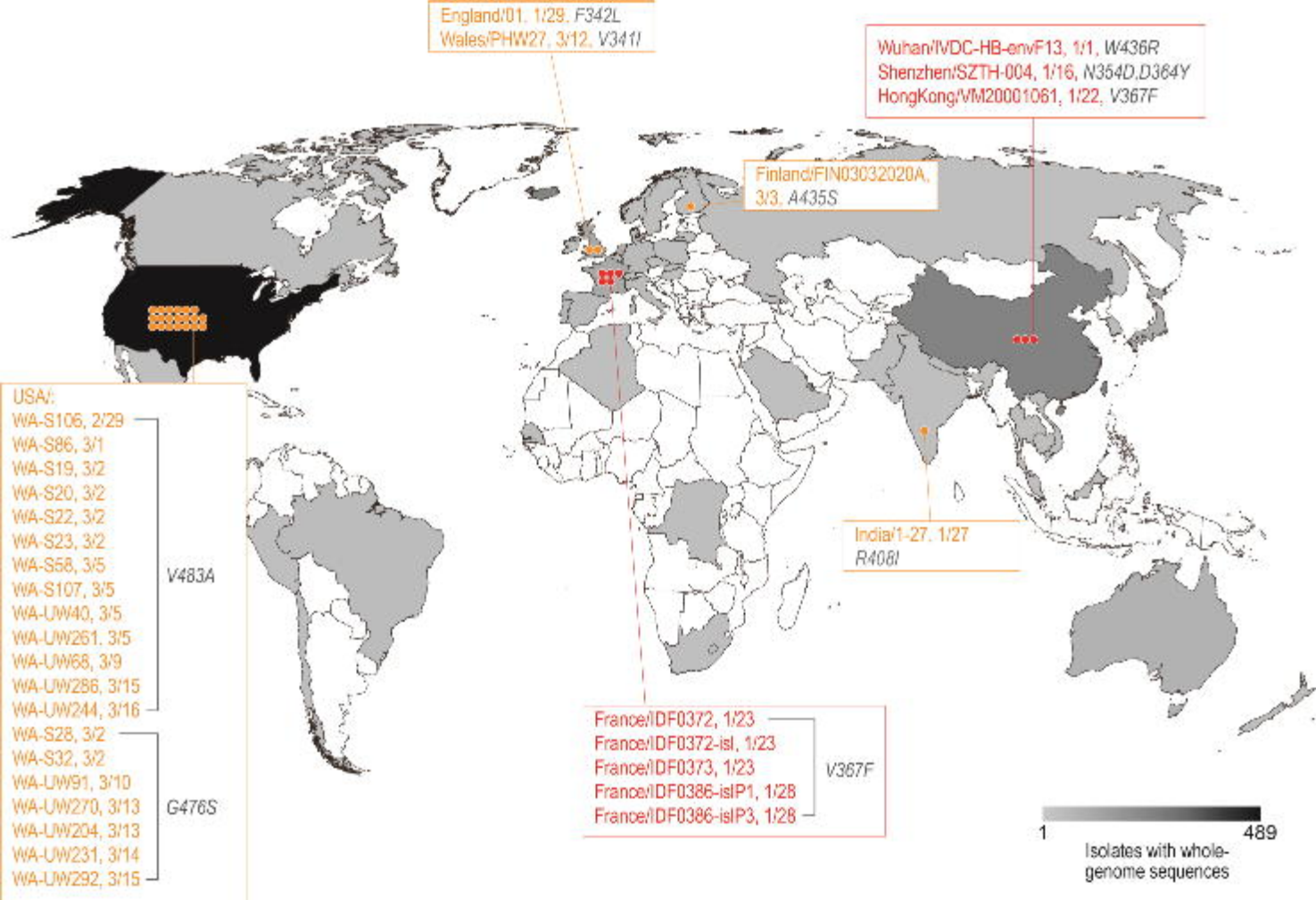
Multiple alignments of the RBD amino acid sequences from January through March, 2020. SARS-CoV-2 Wuhan-Hu-1, the first reported genome, is used as reference. A bat and a pangolin SARS-like coronavirus sequences are also included. Amino acid substitutions are marked. Dots indicate identical amino acids.

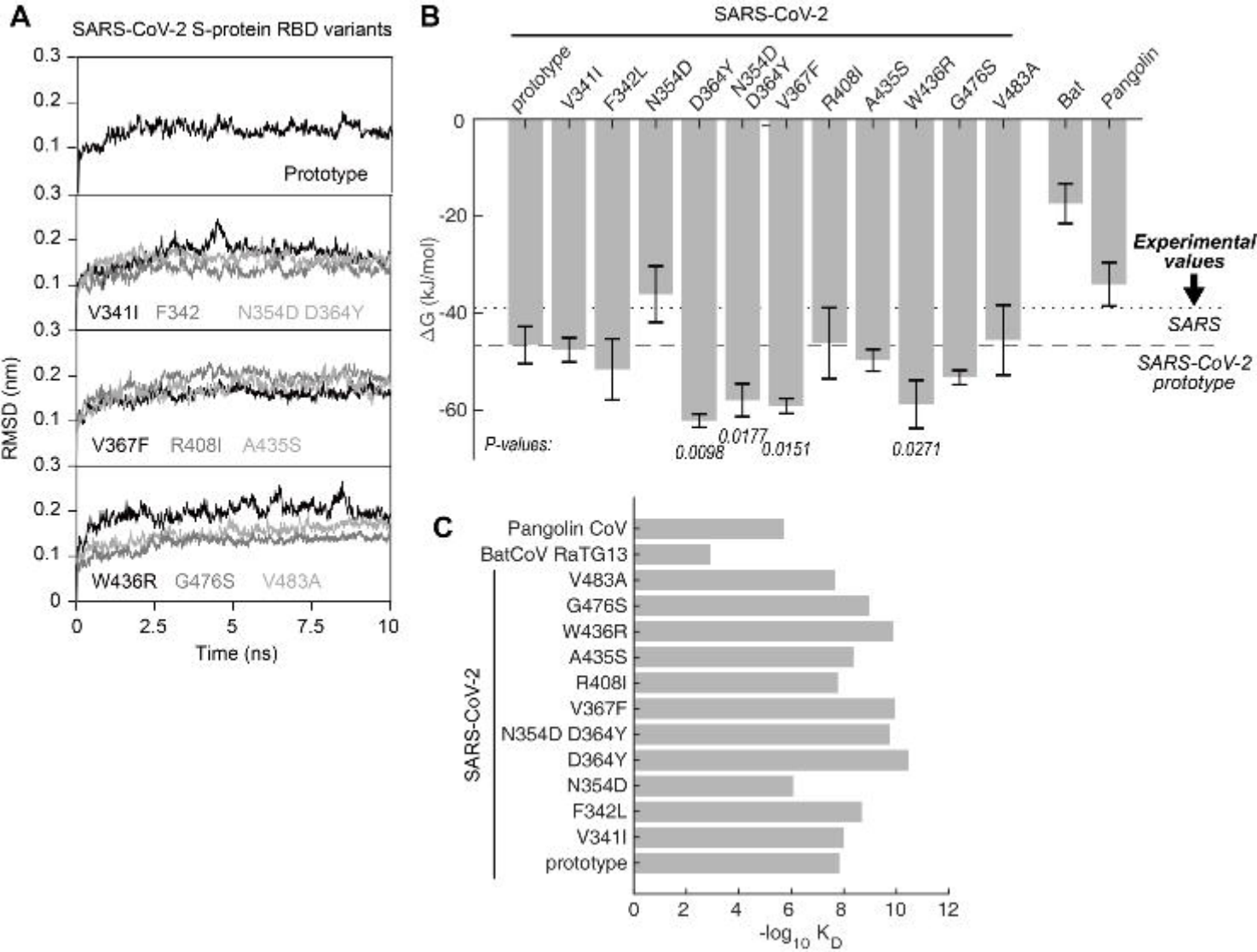
Supplementary Table 1

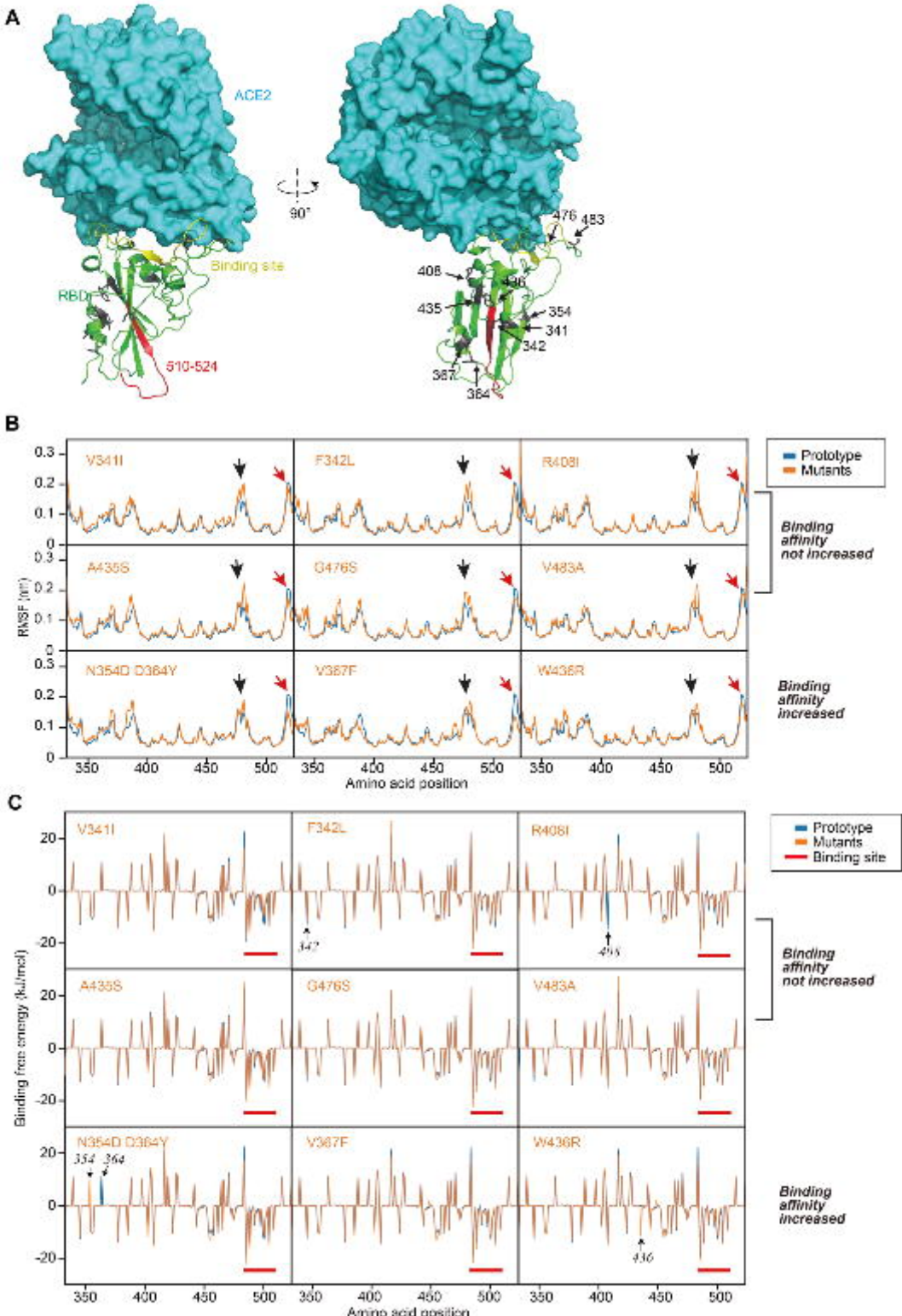
Metadata of the strains with mutations in the RBD of spike glycoprotein (January through March, 2020).

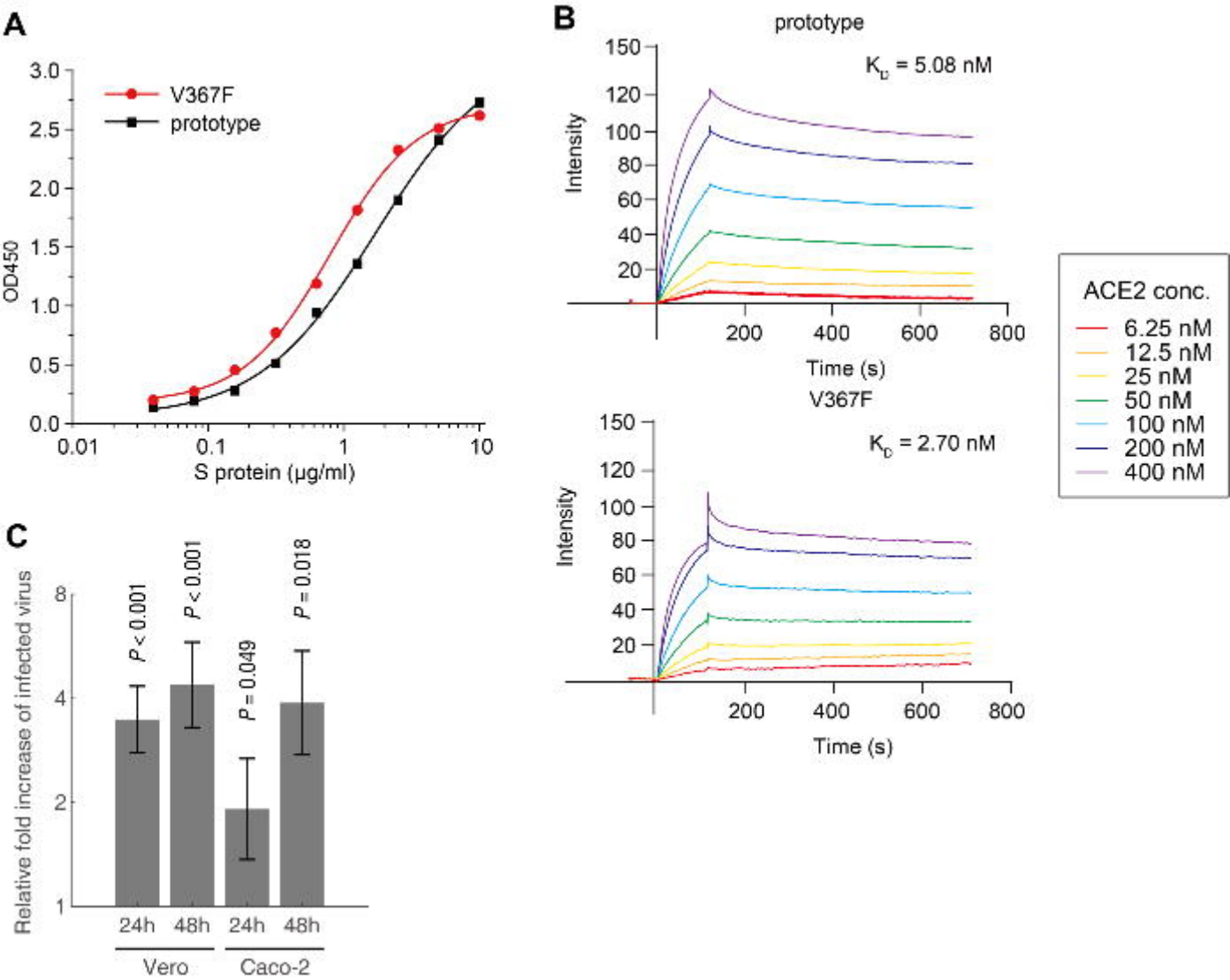
Supplementary Table 2

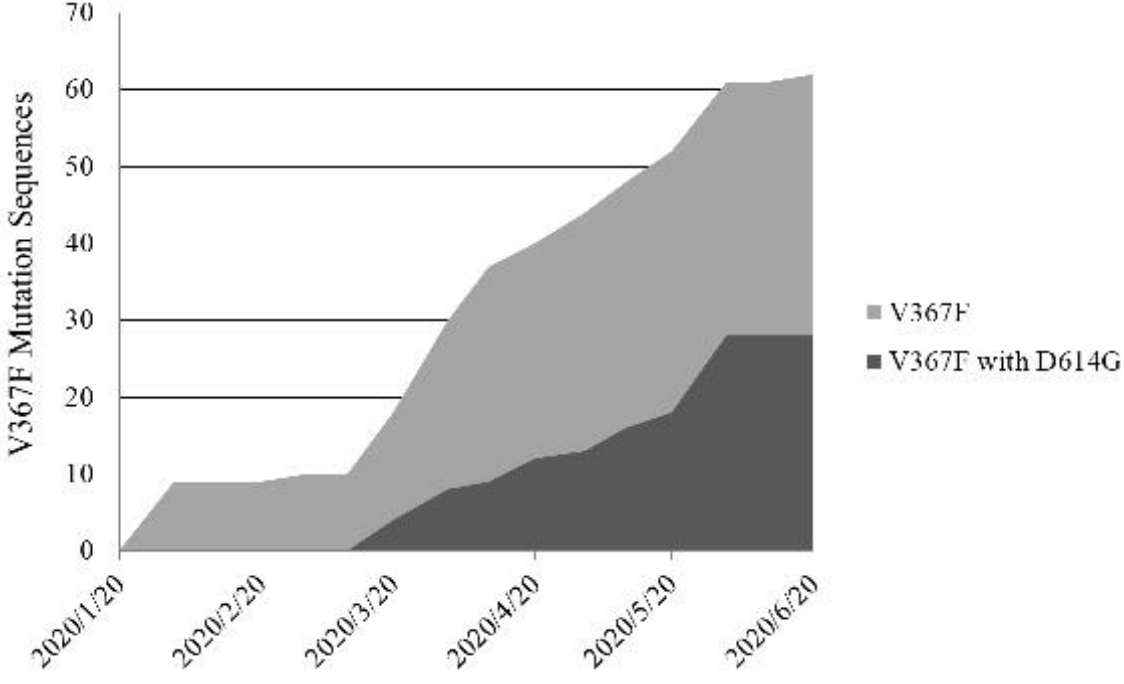
Metadata of the strains with V367F mutations in the RBD of spike glycoprotein (January through August, 2020).


England/01, 1/29, F342L
Wales/PHW27, 3/12, V341I
Wuhan/IVDC-HB-envF13, 1/1, W436R
Shenzhen/SZTH-004, 1/16, N354D,D364Y
HongKong/VM20001061, 1/22, V367F
Finland/FIN03032020A, 3/3, A435S
India/1-27, 1/27
R408I
USA/:
WA-S106, 2/29
WA-S86, 3/1
WA-S19, 3/2
WA-S20, 3/2
WA-S22, 3/2
WA-S23, 3/2
WA-S58, 3/5
WA-S107, 3/5
WA-UW40, 3/5
WA-UW261, 3/5
WA-UW68, 3/9
WA-UW286, 3/15
WA-UW244, 3/16
V483A
WA-S28, 3/2
WA-S32, 3/2
WA-UW91, 3/10
WA-UW270, 3/13
WA-UW204, 3/13
WA-UW231, 3/14
WA-UW292, 3/15
G476S
France/IDF0372, 1/23
France/IDF0372-isl, 1/23
France/IDF0373, 1/23
France/IDF0386-islP1, 1/28
France/IDF0386-islP3, 1/28
V367F
1
489
Isolates with whole-genome sequences


A
SARS-CoV-2 S-protein RBD variants
RMSD (nm)
Prototype
V341I F342 N354D D364Y
V367F R408I A435S
W436R G476S V483A
Time (ns)
B
SARS-CoV-2
prototype
V341I
F342L
N354D
D364Y
N354D D364Y
V367F
R408I
A435S
W436R
G476S
V483A
Bat
Pangolin
ΔG (kJ/mol)
Experimental values
SARS
SARS-CoV-2 prototype
P-values:
0.0098
0.0177
0.0151
0.0271
C
Pangolin CoV
BatCoV RaTG13
SARS-CoV-2
V483A
G476S
W436R
A435S
R408I
V367F
N354D D364Y
D364Y
N354D
F342L
V341I
prototype
-log10 KD


A
ACE2
90°
Binding site
RBD
510-524
476
483
408
436
435
354
341
342
367
364
B
V341I
F342L
R408I
A435S
G476S
V483A
N354D D364Y
V367F
W436R
Prototype
Mutants
Binding affinity not increased
Binding affinity increased
RMSF (nm)
Amino acid position
C
Prototype
Mutants
Binding site
Binding affinity not increased
Binding affinity increased
Binding free energy (kJ/mol)
Amino acid position


A
V367F
prototype
OD450
S protein (μg/ml)
B
prototype
K_D = 5.08 nM
Intensity
Time (s)
V367F
K_D = 2.70 nM
ACE2 conc.
6.25 nM
12.5 nM
25 nM
50 nM
100 nM
200 nM
400 nM
C
Relative fold increase of infected virus
P < 0.001
P < 0.001
P = 0.049
P = 0.018
24h
48h
24h
48h
Vero
Caco-2


V367F Mutation Sequences
70
60
50
40
30
20
10
0
2020/1/20
2020/2/20
2020/3/20
2020/4/20
2020/5/20
2020/6/20
V367F
V367F with D614G

MN908947.3 Wuhan seafood market pneumonia virus isolate Wuhan Hu 1 ●

hCoV 19/Hong Kong/VM20001061 2/2020|EPI_ISL 412029|2020 01 22

hCoV 19/Hong Kong/HKPU1 2101/2020|EPI_ISL 417176|2020 01 21

hCoV 19/USA/WUN NR 52282/2020|EPI_ISL 466656|2020 03 23

hCoV 19/Hong Kong/HKU 001a/2020|EPI_ISL 404571|2020 01 22

hCoV 19/USA/WA 5411/2020|EPI_ISL 417064|2020 02 29

hCoV 19/India/LSCV19046/2020|EPI_ISL 483081|2020 05 16

hCoV 19/India/LSCV19144/2020|EPI_ISL 483085|2020 05 14

hCoV 19/Spain/MD 008100/2020|EPI_ISL 510145|2020 04 03

hCoV 19/USA/CA CSMC91/2020|EPI_ISL 475709|2020 03 27

hCoV 19/Hong Kong/XM PH 54/2020|EPI_ISL 417443|2020 01 22

hCoV 19/France/DF0386 IP3/2020|EPI_ISL 411220|2020 01 23

hCoV 19/England/NORW E9D6F/2020|EPI_ISL 463557|2020 04 28

hCoV 19/England/NOTT 112979/2020|EPI_ISL 463566|2020 04 21

hCoV 19/France/DF0372 IP1/2020|EPI_ISL 410720|2020 01 23

hCoV 19/England/20134027504/2020|EPI_ISL 423138|2020 03 23

hCoV 19/England/LIVE A9CEC/2020|EPI_ISL 499952|2020 03 12

hCoV 19/England/LIVE A692F/2020|EPI_ISL 499160|2020 03 19

hCoV 19/France/DF0372/2020|EPI_ISL 408598|2020 01 23

hCoV 19/England/20129063004/2020|EPI_ISL 424032|2020 03 11

hCoV 19/England/CAMB 1D56A7/2020|EPI_ISL 512335|2020 07 22

hCoV 19/France/DF0373/2020|EPI_ISL 408597|2020 01 23

hCoV 19/Iceland/597/2020|EPI_ISL 424607|2020 03 29

hCoV 19/Germany/BY MVP C264/2020|EPI_ISL 468907|2020 05 04

hCoV 19/France/DF0386 IP1/2020|EPI_ISL 411219|2020 01 29

hCoV 19/England/NOTT 111E10/2020|EPI_ISL 472411|2020 04 29

hCoV 19/Australia/VIC407/2020|EPI_ISL 428709|2020 03 25

hCoV 19/Australia/VIC054/2020|EPI_ISL 427111|2020 01 04

hCoV 19/Australia/VIC003/2020|EPI_ISL 490578|2020 01 06

hCoV 19/Australia/VIC084/2020|EPI_ISL 427148|2020 01 06

hCoV 19/Australia/VIC118/2020|EPI_ISL 426742|2020 03 27

hCoV 19/England/LIVE 9A86A/2020|EPI_ISL 449373|2020 03 29

hCoV 19/England/LIVE 9AD13/2020|EPI_ISL 449385|2020 03 30

hCoV 19/England/LIVE 9E811/2020|EPI_ISL 453317|2020 04 10

hCoV 19/England/LIVE 9F0C6/2020|EPI_ISL 453100|2020 04 05

hCoV 19/England/LIVE A642D/2020|EPI_ISL 499198|2020 03 16

hCoV 19/Netherlands/NoordBrabant 70/2020|EPI_ISL 422666|2020 03 11

hCoV 19/Netherlands/NoordBrabant 71/2020|EPI_ISL 422667|2020 03 11

hCoV 19/Netherlands/Utrecht 20/2020|EPI_ISL 422997|2020 03 11

hCoV 19/USA/FL BPHL 09/2020|EPI_ISL 436512|2020 04 03

hCoV 19/Taiwan/NTU25/2020|EPI_ISL 447619|2020 03 28

hCoV 19/mink/Netherlands/ND04 index/2020|EPI_ISL 447801|2020 05 06

hCoV 19/Netherlands/NoordBrabant 155/2020|EPI_ISL 461190|2020

hCoV 19/Netherlands/NoordBrabant 158/2020|EPI_ISL 461191|2020

hCoV 19/Netherlands/NoordBrabant 157/2020|EPI_ISL 461192|2020

hCoV 19/Netherlands/NoordBrabant 189/2020|EPI_ISL 461202|2020 05 28

hCoV 19/Netherlands/NoordBrabant 188/2020|EPI_ISL 461200|2020

hCoV 19/Switzerland/100019/2020|EPI_ISL 476100|2020 05 29

hCoV 19/Spain/CCV002726/2020|EPI_ISL 487290|2020 04 15

hCoV 19/Switzerland/100792/2020|EPI_ISL 491742|2020 03 25

hCoV 19/England/NORW EC099/2020|EPI_ISL 490632|2020 04 21

hCoV 19/Netherlands/Utrecht 10037/2020|EPI_ISL 454798|2020 03 11

hCoV 19/England/DRIS 125523/2020|EPI_ISL 443700|2020 03 27

hCoV 19/England/DRIS 129D46/2020|EPI_ISL 459915|2020 04 14

hCoV 19/England/CAMB 1D337D/2020|EPI_ISL 481557|2020 05 26

hCoV 19/England/CAMB 1D33D7/2020|EPI_ISL 481559|2020 05 26

hCoV 19/England/CAMB 1D4FD0/2020|EPI_ISL 477009|2020 03 17

hCoV 19/England/OXON D26D0/2020|EPI_ISL 479183|2020 04 17

hCoV 19/England/SHEF C4017/2020|EPI_ISL 464447|2020 05 19

hCoV 19/USA/CA CZD 2159/2020|EPI_ISL 513706|2020 05 21

hCoV 19/Northern Ireland/NIRC 100098/2020|EPI_ISL 469260|2020 05 05

hCoV 19/England/NORT 292798/2020|EPI_ISL 466614|2020 03 29

hCoV 19/England/LIVE A2AC6/2020|EPI_ISL 499973|2020 05 16

hCoV 19/Switzerland/10001 1/2020|EPI_ISL 467000|2020 05 07

Dual Mutations (V367F +D614G)

Table 2: Dominant mutation in spike RBD （>20 strains） and their binding ability

Amino Position in Spike	Protype Amino	Mutation Amino	SequencesNumber
478	T	I	106
479	P	S	76
367	V	F	63
483	V	A	38
485	G	R	32
520	A	S	28
341	V	I	27
384	P	L	25
522	A	V	24
330	P	S	22

(As of August 24th ,2020)

Sequnce First Occurance Date	First Sequence Occurance ID	D614 Number	G614 Number
2020.3.18	EPI_ISL_440368	0	104
2020.4.13	EPI_ISL_470698	0	66
2020.1.21	EPI_ISL_417176	35	28
2020.2.29	EPI_ISL_417159	36	1
2020.2.6	EPI_ISL_450418	1	31
2019.12.30	EPI_ISL_402124	6	14
2020.3.10	EPI_ISL_440476	0	27
2020.3.26	EPI_ISL_425960	3	18
2020.3.25	EPI_ISL_421654	5	19
2020.3.13	EPI_ISL_454225	2	20

Total Number	Increasing Receptor Binding
104	No
66	No
63	Yes
37	No
32	No
20	No
27	No
21	No
24	No
22	No

Table 1: Nucleotide substitution rates and selection pressures for S gene.

Gene	Length(bp)	Mean Non-synonymous Substations/site	Mean Synonymous Substations/site
S	3822	0.7726	0.1875
S1	2043	0.6207	0.0571
S1-RBD	585	0.0458	0.0137
S2	1779	0.1519	0.1304

dN/dS
4.1197
10.8636
3.3545
1.1646