Emergence of RBD mutations in circulating SARS-CoV-2 strains enhancing the structural stability and human ACE2 receptor affinity of the spike protein

Junxian Ou,[a] Zhonghua Zhou,[b] Ruixue Dai,[c,f] Jing Zhang,[d] Wendong Lan,[a] Shan Zhao,[a] Jianguo Wu,[d] Donald Seto,[e] Lilian Cui,[f] Gong Zhang,[b#] Qiwei Zhang[a,d#]

[a] Guangdong Provincial Key Laboratory of Tropical Disease Research, School of Public Health, Southern Medical University, Guangzhou, Guangdong 510515, China

[b] 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.

[c] Department of Environmental Science and Engineering，Fudan University, Shanghai 200433, China

[d] Guangdong Provincial Key Laboratory of Virology, Institute of Medical Microbiology, Jinan University, Guangzhou, Guangdong 510632, China

[e] Bioinformatics and Computational Biology Program, School of Systems Biology, George Mason University, Manassas, VA 20110, USA

[f] Novoprotein Scientific Inc. Shanghai 201203, China

Running Head: RBD mutations enhance human ACE2 receptor affinity

#Address correspondence to Qiwei Zhang, zhangqw@smu.edu.cn; Gong Zhang, zhanggong@jnu.edu.cn.

J.O., Z.Z., and R.D. contributed equally to this work.

Abstract: 250 words

Importance: 150 words

Text: 2694 words

Abstract

A novel coronavirus SARS-CoV-2 is associated with the current global pandemic of Coronavirus Disease 2019 (COVID-19). Spike protein receptor-binding domain (RBD) of SARS-CoV-2 is the critical determinant of viral tropism and infectivity. To investigate whether the mutations in the RBD have altered the receptor binding affinity and caused these strains more infectious, we performed molecular dynamics simulations of the binding affinity between the mutant SARS-CoV-2 RBDs to date and the human ACE2 receptor. Among 1609 genomes of global SARS-CoV-2 strains, 32 non-synonymous RBD mutants were identified and clustered into 9 mutant types under high positive selection pressure. Three mutant types (V367F, W436R, and D364Y) emerging in Wuhan, Shenzhen, Hong Kong, and France, displayed higher human ACE2 affinity, and probably higher infectivity. This is due to the enhanced structural stabilization of the RBD beta-sheet scaffold. High frequencies of RBD mutations were identified: V367F from five France and one Hong Kong mutants, 13 V483A and 7 G476S mutants from the U.S.A. This suggested they originated as novel sub-lineages. The enhancement of the binding affinity of the mutant type (V367F) was further validated by the receptor-ligand binding ELISA assay. The molecular dynamics simulations also indicated that it would be difficult for bat SARS-like CoV to infect humans. However, the pangolin CoV is potentially infectious to humans. The analysis of critical RBD mutations provides further insights into the evolutionary history of SARS-CoV-2 under high selection pressure. An enhancement of the SARS-CoV-2 binding affinity to human ACE2 receptor reveals higher infectivity of the mutant strains.

Importance

A novel coronavirus SARS-CoV-2 has caused the pandemic of COVID-19. The origin of SARS-CoV-2 was associated with zoonotic infections. The spike protein receptor-binding domain (RBD) is identified as the critical determinant of viral tropism and infectivity. Thus, whether the mutations in the RBD of the circulating SARS-CoV-2 strains have altered the receptor binding affinity and caused these strains more infectious, should be paid more attentions to. Here, 32 non-synonymous RBD mutants were identified and clustered into 9 mutant types under high positive

selection pressure, suggesting they originated as novel sub-lineages. Three mutant types displayed higher human ACE2 affinity, and probably higher infectivity, one of which (V367F) was validated by wet bench. The RBD mutation analysis provides insights into SARS-CoV-2 evolution. The emergence of RBD mutations with increased human ACE2 affinity reveals higher risk of severe morbidity and mortality during a sustained COVID-19 pandemic, particularly if no effective precautions are implemented.

Keywords: COVID-19; SARS-CoV-2; ACE2; RBD; mutations; affinity; infectivity; spike protein

1. Introduction

A novel coronavirus SARS-CoV-2 has caused the outbreaks of Coronavirus Disease 2019 (COVID-19) globally since the first report in mid-December 2019 in Wuhan, China (1–4). As of April 14, 2020, SARS-CoV-2 has infected 1,844,863 people world-wide and caused 117,021 deaths (5) with the estimated fatality rate of 6.34%. This on-going pandemic of COVID-19 has become the most serious threat to public health in recent times.

The origin of SARS-CoV-2 remains elusive. However, the initial cases were largely associated with the seafood market, which indicated this were potential zoonotic infections(2). Although bats and pangolins are most likely the reservoir hosts and the intermediate hosts in the wild, more evidences are in need to support the zoonotic infections and track the origin of this new coronavirus(6–8).

Angiotensin-converting enzyme 2 (ACE2) is the cellular receptor of SARS-CoV-2 (9), which is the same as for SARS-CoV. The receptor-binding domain (RBD) of the subunit S1 directly interacts with ACE2, which provides for tight binding to the peptidase domain of ACE2. Therefore, RBD is the critical determinant of virus-receptor interaction and reflects viral host range, tropism and infectivity(6, 10–12). Although the RBD sequences of different SARS-CoV-2 strains circulating globally are conserved, mutations have appeared, which might account for differences in viral infectivity and contribute to its spread.

Meanwhile, S protein participates in antigenic recognition expressed on its protein surface, likely to be immunogenic as for carrying both T-cell and B-cell epitopes. The potential antibody binding sites that have been identified indicates RBD has important B-cell epitopes. The main antibody binding sites substantially overlap with RBD, and the antibody binding to these sites is likely to block viral entry into cells(13, 14).

To investigate whether these mutations in RBD have altered the receptor binding affinities and whether these strains may have been selected for higher infectivity, the binding dynamics between

the SARS-CoV-2 RBDs of the mutant strains to date and human ACE2 receptor were modelled and assessed.

2. Results and discussion

2.1 SARS-CoV-2 RBD mutation mapping

Among the 1609 SARS-CoV-2 strains with whole genome sequences available in the public databases, 32 strains contained non-synonymous 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 (Fig. 1). Most mutants deviate from the original reported genome (SARS-CoV-2 Wuhan-Hu-1) by only one amino acid (Supplementary Figure 1). These 32 mutations parse into 9 mutant types. Mutation V367F was found in six individual strains isolated from four patients: Three in France and one in Hong Kong. Similarly, high frequencies of RBD mutations were also identified the U.S.A: 13 V483A mutants and 7 G476S mutants (Fig. 1). This suggested that these mutant strains may have originated as novel sub-lineages.

2.2 Nucleotide diversity indicates strong positive selective pressure on RBD

Since RBD is the only domain to bind human ACE2 and initiate cell entry, it is believed that the RBD should be highly conserved. However, polymorphism and divergence analysis by DnaSP6 (version 6.12.03) (15) showed that the RBD sequences were as diverse as the other regions of the S protein (Fig. 2). The peak signals for diversity distribute across the entire S protein, and the multiple peaks in the RBD also reached the Pi value of ~0.0008, similar to Pi values in the other regions. Therefore, we hypothesize that the RBD would be selected to maintain or even improve its binding affinity to human ACE2.

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 SARS-CoV-2 strains. With respect to our hypothesis, 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 bind human ACE2 receptor with higher affinity

To estimate the functional changes suggested by the RBD mutations, we performed molecular dynamics 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, which was performed using GROMACS 2019. The complex structure of the SARS-CoV-2 S-protein RBD domain and human ACE2 was obtained from National Microbiology Data Center (ID: NMDCS0000001) (PDB ID: 6LZG) (https://www.rcsb.org/structure/6LZG). Mutant 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(16). Each simulation was performed at 10ns and each model was simulated in triplicates. All trajectories reached a plateau of RMSD after 2~5ns (Fig. 3A), indicating that their structures reached an 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 (Fig. 3B). 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 K_D) (Fig. 3B). Compared to the K_D = 14.7 nM of the prototype RBD(17), the equilibrium dissociation constant (K_D) of the three mutants are calculated as 0.12 nM, 0.11 nM, and 0.13 nM, respectively (Fig. 3C), which were two orders of magnitude lower than for the prototype strain, indicating a remarkably increased affinity to the human ACE2 receptor. In the only double amino acid mutant (N354D, D364Y), the N354D substitution decreased the affinity, while the D364Y provided an even higher affinity than the overall double mutations (Fig. 3B). This indicated that the D364Y is the major contributor of the enhanced affinity.

To validate the change of the binding affinity of the mutant S protein (V367F) experimentally, a receptor-ligand binding ELISA assay of the S proteins and the ACE2 was performed. Fig. 3D showed that the V367F mutant significantly lowered the ED50 concentration (ED50 = 0.8±0.04 μg/ml), as compared to the prototype (ED50 = 1.7±0.14 μg/ml), demonstrating that the V367F mutant has higher affinity than the prototype. This result qualitatively validated our computational simulation.

In comparison, the bat CoV RBD (strain RaTG13, with the highest genome similarity) showed a much lower binding affinity (K_D=1.17mM; ΔG=-17.4kJ/mol) to human ACE2 than the pangolin CoV (K_D=1.89μM; ΔG=-33.9kJ/mol). For comparison, the affinity of the pangolin CoV was slightly lower than the SARS-CoV-2 prototype strain (K_D=14.7nM; ΔG=-46.5kJ/mol) (Fig. 3B, 3C).

2.4 Structural basis for the increased affinity

The 9 mutant types were divided into two groups: the "similar affinity" group (V341I, F342L, R408I, A435S, G476S, V483A), whose affinity is not significantly increased, and the "higher affinity" group (N354D D364Y, V367F, W436R), whose affinity is significantly increased. To explain the structural basis of the increased affinity, we investigated deeper into the dynamics of the residues of these structures. The binding surface of the RBD to ACE2 is largely in random coil conformation, which lacks structural rigidity. Therefore, 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 (Fig. 4A, marked as red), 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 (Fig. 4B). 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 (Fig. 4B). Moreover, the "higher affinity" group exhibited a general

decreased ΔG in the binding site region, but not the "similar affinity" group mutants (Fig. 4C). In addition, the D364Y and W436R themselves directly contributed to the ΔG decrease. In contrast, the N354D mutation directly elevated the ΔG, which coincides its consequence (Fig. 4B). The mutation W436R provides a positively charged Arg in the proximity of the complementing highly negative charged ACE2 surface. This potential electrostatic attraction may contribute to the higher affinity (Fig. 4D).

3. Discussion

Due to the lengthening pandemic and evolving nature of the SARS-CoV-2 virus globally, identifying changes in viral infectivity is crucial to containing the COVID-19 spread. Quarantine policies need to be adapted with respect to the changes in virus infectivity. This report provides computational insight into the functional outcome of mutations in RBD: RBD mutants under positive selection pressure, and several mutants acquired increased binding affinity to human ACE2 receptor, implying higher infectivity to humans (noted for one mutant with experimental validation).

It should be noted that the mutation V367F enhancing the 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 the geographic distance indicates that this mutant may have evolved to be more robust and that these strains originated as a novel sub-lineage, given the close isolation dates (January 22 and 23, respectively). Combined with the epidemiological data, mutation surveillance is of critical importance as it can reveal more exact transmission routes of the epidemic and provide early warning for additional outbreaks. Emergence of SARS-CoV-2 strains in Hong Kong, France, and other countries with RBD mutations providing higher binding affinity to human ACE2 receptor suggests a higher risk of more severe morbidity and mortality during a sustained pandemic of COVID-19, particularly if no effective precautions are implemented.

Our analysis of molecular dynamics simulation indicates the remarkable enhancement of the affinity efficiency of mutant S protein. Compared to the prototype strain Wuhan-Hu-1, the ΔG of mutants decreased ~25%. Mutants bind ACE2 more stably due to the enhancement of the base rigidity. 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. The viruses have been adapting to transmission and replication in humans; mutation or recombination events in RBD may boost the binding affinity and cause the basic reproduction number (R0) to climb in theory, i.e., the human to human transmission more easily.

The origination of the virus is a constant hot topic since the virus outbreak. Due to the high homology of the bat SARS-like CoV genome and pangolin CoV RBD to the SARS-CoV-2, these wild animals were thought to initiate the infection in human. Our results provided more clues on this postulation. In our study, the binding energy of the bat SARS-like CoV RBD suggests it is too high to bind human ACE2 effectively (K_D in millimolar range). In contrast, the pangolin CoV showed a K_D of binding to human ACE2 at the micromolar range, just ~6x higher than that of human SARS virus (K_D = 0.326μM)(17) (Fig. 3), indicating that the pangolin CoV has the potential to infect human in unprotected close contact. Alignment of the genomic sequences of SARS-CoV-2 and pangolin CoV viruses indicated the evidence for recombination events in RBD domain between pangolin and bat viruses. The pangolin CoV has been detected among the smuggled Malayan pangolins in multiple provinces in China(7, 8), suggesting a risk of zoonotic infection from wild animals to human constantly and widely.

The S protein is also important for antigen recognition. In this survey of 1609 strains, 32 had amino acid mutations in the RBD. High frequencies of RBD mutations were identified: V367F from five France and one Hong Kong mutants, 13 V483A and 7 G476S mutants from the U.S.A. Since the RBD contains important antigenic epitopes, frequent mutations in RBD, especially those which change the amino acid properties, may weaken the binding affinity of the antibody raised against the prototype strain. This may lead to decreased vaccine efficacy and should be further validated.

In summary, we have identified 32 RBD mutant strains clustering into 9 mutant types under high positive selection pressure. This suggested that they originated as novel sub-lineages. Three mutant types emerging in Asia and Europe displays enhanced structural stability of the spike protein along with higher binding affinities to human ACE2 receptor, which indicates that these mutants may have acquired increased infectivity to humans.

4. Methods and materials

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 were downloaded and the sequences with amino acid mutations in S protein and RBD region were screened. The genome sequences with amino acid mutations in S protein and the RBD were analyzed in this study (Supplementary Table 1).

4.2 Sequences alignment and polymorphism analyses

Alignment of S protein sequences from different sources and comparison of ACE2 proteins among different species were accomplished by MAFFT version 7 online serve with default parameter （https://mafft.cbrc.jp/alignmeloadnt/server/）and Bioedit(18, 19). Polymorphism and divergence were analyzed by DnaSP6 (version 6.12.03) (15). Analyses were conducted using the Nei-Gojobori model(20). All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in Mega X（version 10.0.2）(21).

4.3 Molecular dynamics (MD) simulation

The complex structure of the SARS-CoV-2 S-protein RBD domain and human ACE2 was obtained from Nation Microbiology Data Center (ID: NMDCS0000001) (PDB ID: 6LZG). Mutant 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(16). 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, LINCS restraints. With 2fs steps, each simulation was performed at 10ns, and each model was simulated 3 times to generate 3 independent trajectory replications. Binding free energy (ΔG) was calculated using MM-PBSA method (software downloaded from GitHub: https://github.com/Jerkwin/gmxtool) with the trajectories after structural equilibrium assessed using RMSD (Root Mean Square Deviation)(22).

The formula $\Delta G = RTlnK_D$ was used to calculate between equilibrium dissociation constant (K_D) and ΔG. The estimated ΔG of the RBD mutants were normalized using the ΔG of the prototype strain which was derived from experimental data(23).

4.4 Recombinant S protein mutant expression

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

4.5 Ligand-receptor binding ELISA assay

The human ACE2 was immobilized in the microtiter plate at 5 μg/ml (100μl/well). The S proteins (prototype and V367F, respectively) was added as ligand at different concentrations, 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 was then washed 3 times. 100μl of HRP anti-His Tag Antibody (BioLegend, USA) (diluted 1:20000) was added to each well for 1 hour. After 3 times washing, the signal was visualized using TMB solution (Sigma-Aldrich, USA). OD450 was recorded using microtiter plate reader.

Acknowledgments

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.

Data available in Supplementary material.

Funding statement:

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).

Conflict of interest

The authors declare that they have no conflicts of interest.

Reference

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W. 2020. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 727–733.
2. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KSM, Lau EHY, Wong JY, Xing X, Xiang N, Wu Y, Li C, Chen Q, Li D, Liu T, Zhao J, Liu M, Tu W, Chen C, Jin L, Yang R, Wang Q, Zhou S, Wang R, Liu H, Luo Y, Liu Y, Shao G, Li H, Tao Z, Yang Y, Deng Z, Liu B, Ma Z, Zhang Y, Shi G, Lam TTY, Wu JT, Gao GF, Cowling BJ, Yang B, Leung GM, Feng Z. 2020. Early Transmission Dynamics in Wuhan, China, of Novel

Coronavirus–Infected Pneumonia. N Engl J Med 1–9.

3. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z, Xiong Y, Zhao Y, Li Y, Wang X, Peng Z. 2020. Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA - J Am Med Assoc 1–9.

4. Chan JFW, Yuan S, Kok KH, To KKW, Chu H, Yang J, Xing F, Liu J, Yip CCY, Poon RWS, Tsoi HW, Lo SKF, Chan KH, Poon VKM, Chan WM, Ip JD, Cai JP, Cheng VCC, Chen H, Hui CKM, Yuen KY. 2020. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 395:514–523.

5. World Health Organization. 2020. Coronavirus disease (COVID-2019) situation reports, 2020-04-14. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/.

6. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, Si H-R, Zhu Y, Li B, Huang C-L, Chen H-D, Chen J, Luo Y, Guo H, Jiang R-D, Liu M-Q, Chen Y, Shen X-R, Wang X, Zheng X-S, Zhao K, Chen Q-J, Deng F, Liu L-L, Yan B, Zhan F-X, Wang Y-Y, Xiao G-F, Shi Z-L. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature.

7. Lam TT-Y, Shum MH-H, Zhu H-C, Tong Y-G, Ni X-B, Liao Y-S, Wei W, Cheung WY-M, Li W-J, Li L-F, Leung GM, Holmes EC, Hu Y-L, Guan Y. 2020. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. Nature.

8. Xiao K, Zhai J, Feng Y, Zhou N, Zhang X, Zou J-J, Li N, Guo Y, Li X, Shen X, Zhang Z, Shu F, Huang W, Li Y, Zhang Z, Chen R-A, Wu Y-J, Peng S-M, Huang M, Xie W-J, Cai Q-H, Hou F-H, Liu Y, Chen W, Xiao L, Shen Y. 2020. Isolation and Characterization of 2019-nCoV-like Coronavirus from Malayan Pangolins. bioRxiv 2020.02.17.951335.

9. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu N-H, Nitsche A, Müller MA, Drosten C, Pöhlmann S. 2020. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell S0092-8674(20)30229–4.

10. Letko M, Marzi A, Munster V. 2020. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nat Microbiol 1–8.
11. Chen Y, Guo Y, Pan Y, Zhao ZJ. 2020. Structure analysis of the receptor binding of 2019-nCoV. Biochem Biophys Res Commun 2:0–5.
12. Wan Y, Shang J, Graham R, Baric RS, Li F. 2020. Receptor recognition by novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS. J Virol.
13. Fast E, Altman RB, Chen B. 2020. Potential T-cell and B-cell Epitopes of 2019-nCoV. bioRxiv 2020.02.19.955484.
14. Ahmed SF, Quadeer AA, McKay MR. 2020. Preliminary identification of potential vaccine targets for 2019-nCoV based on SARS-CoV immunological studies. Viruses 2020.02.03.933226.
15. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sanchez-Gracia A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol 34:3299–3302.
16. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, De Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. 2018. SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Res 46:W296–W303.
17. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, Graham BS, McLellan JS. 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science (80-) 367:1260 LP – 1263.
18. Katoh K, Rozewicki J, Yamada KD. 2018. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform 20:1160–1166.
19. Kuraku S, Zmasek CM, Nishimura O, Katoh K. 2013. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. Nucleic Acids Res 41:22–28.
20. Nei M, Gojoborit T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426.

375 21. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular evolutionary
376 genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549.
377 22. Homeyer N, Gohlke H. 2012. Free energy calculations by the Molecular Mechanics
378 Poisson-Boltzmann Surface Area method. Mol Inform 31:114–122.
379 23. Rifai EA, Van Dijk M, Vermeulen NPE, Yanuar A, Geerke DP. 2019. A Comparative Linear
380 Interaction Energy and MM/PBSA Study on SIRT1-Ligand Binding Free Energy Calculation.
381 J Chem Inf Model 59:4018–4033.
382

Figure

Fig. 1: Distribution of the SARS-CoV-2 strains mutant in the RBD of the S protein. The geographic distribution of the 32 RBD mutant strains clustering into 9 mutant types is displayed. The 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.

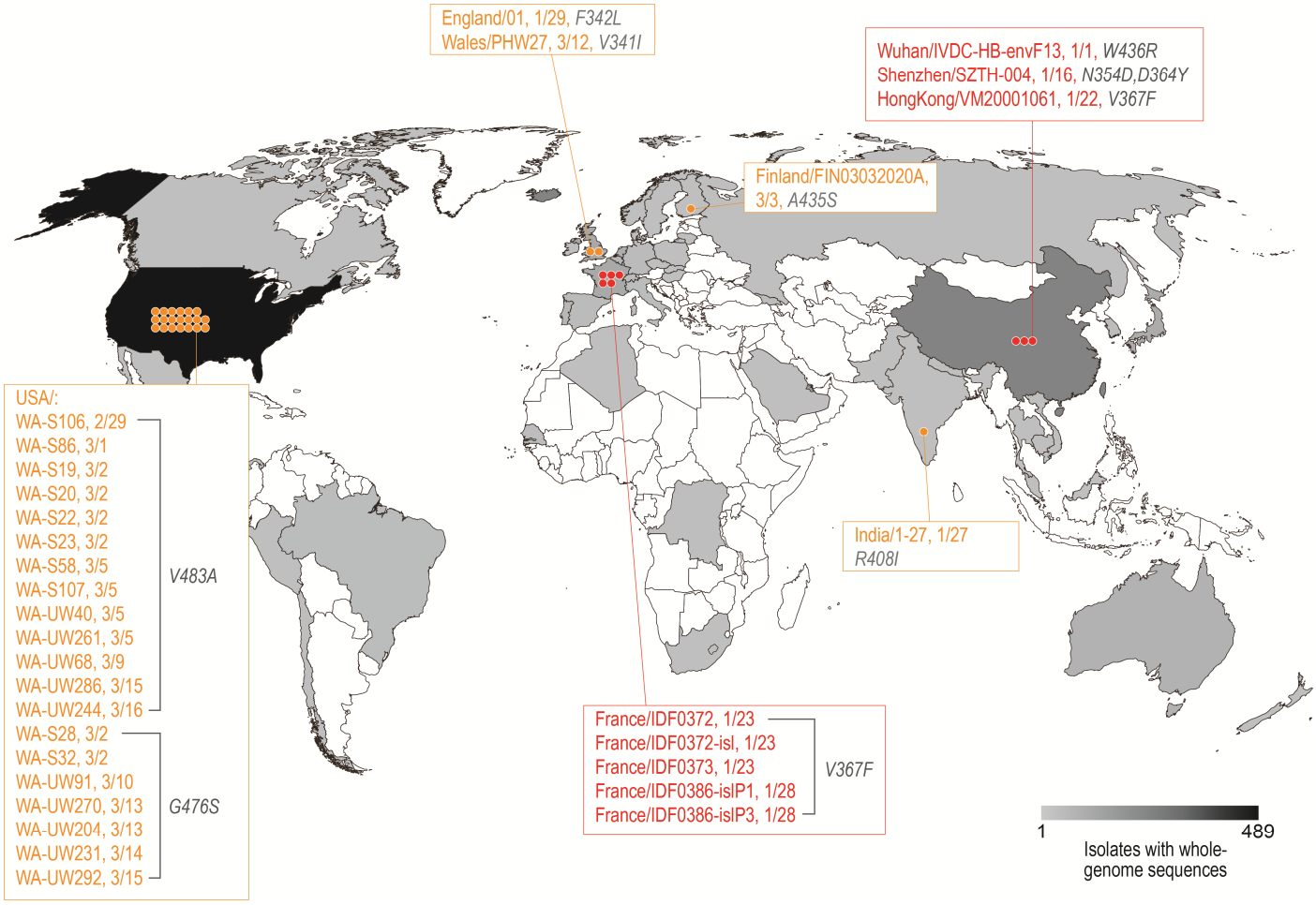


Fig. 2: Polymorphism and divergence graph of SARS-CoV-2 S gene. Polymorphism and divergence were analyzed by DnaSP6 (version 6.12.03). Analyses were conducted using the Nei-Gojobori model. All positions containing gaps and missing data were eliminated. Structural domains are annotated. The Pi values are calculated with window size: 20 nt, step size: 5.

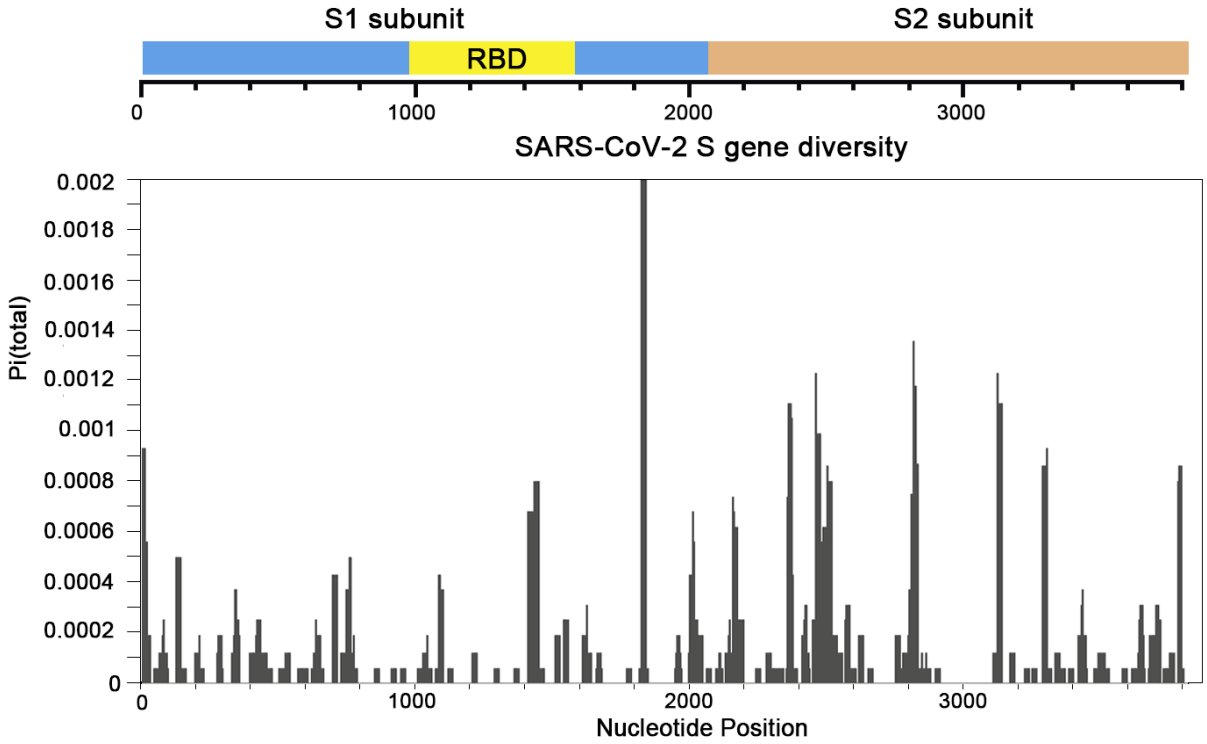
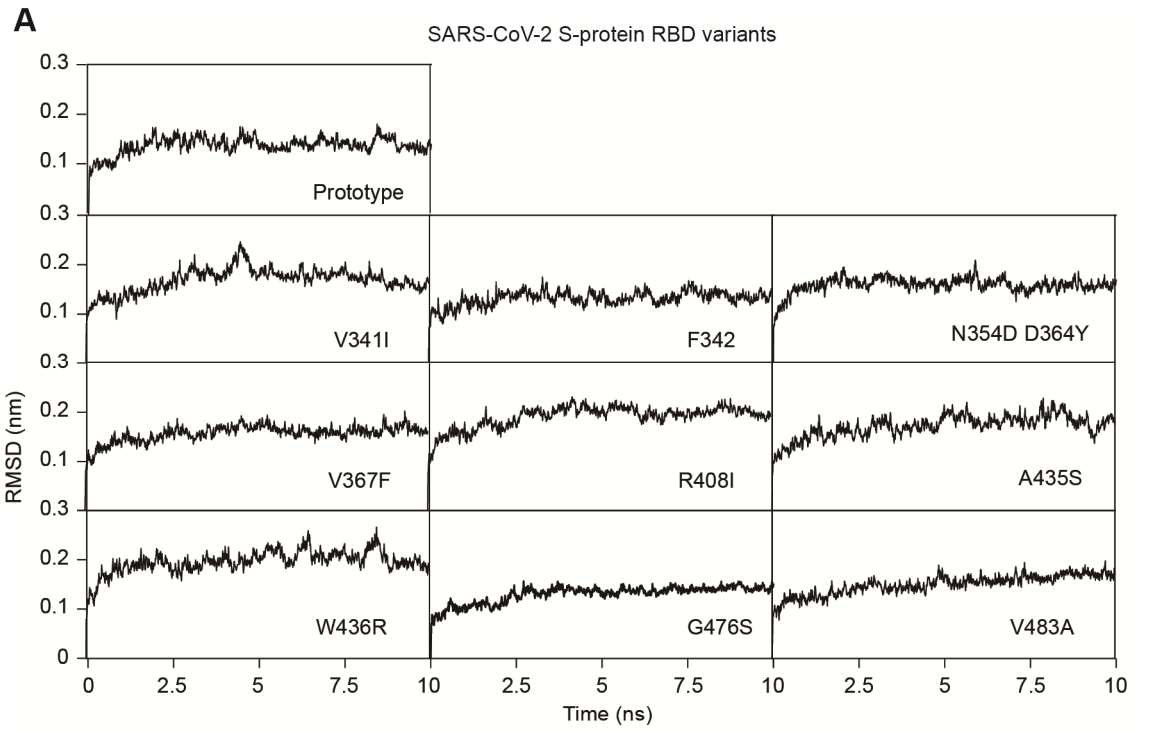
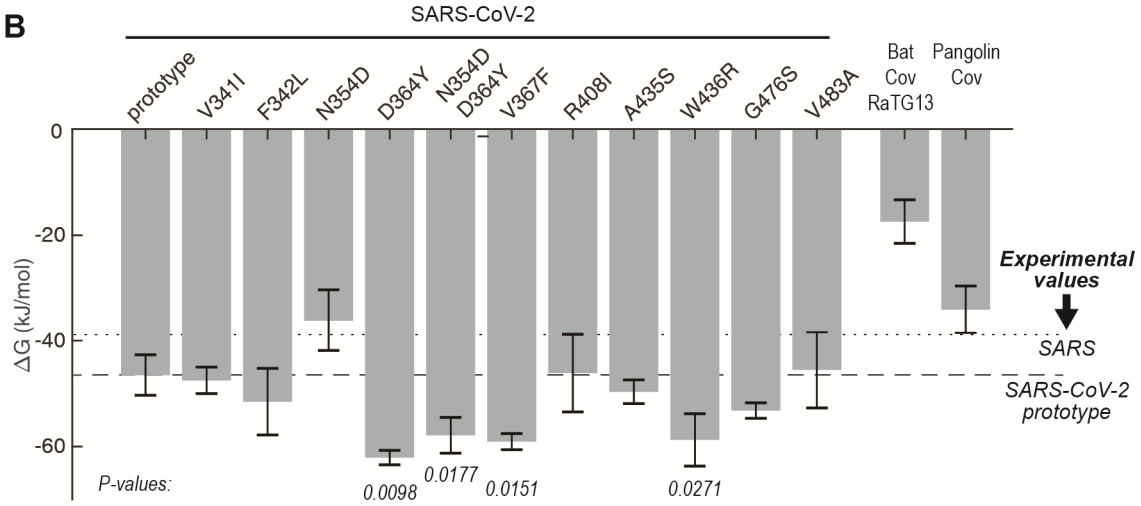
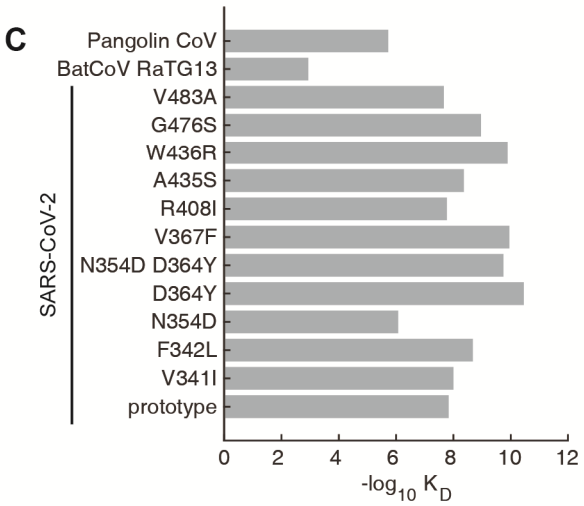
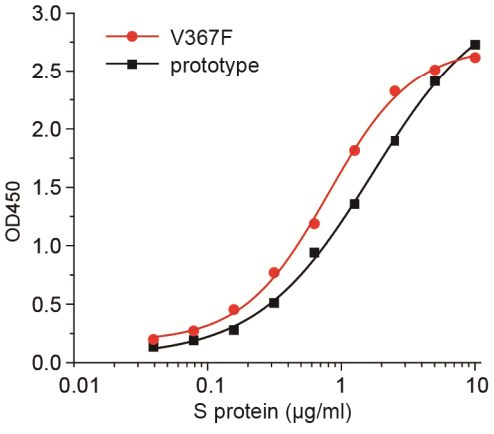


Fig. 3: Binding free energy calculated for the SARS-CoV-2 S-RBD to human ACE2. (A) RMSD of typical MD trajectories of the SARS-CoV-2 prototype and the mutant strains. **(B)** Comparison of the binding free energy (ΔG) of the RBDs and the human ACE2. Note, the ΔG is inversely proportional to the binding 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. **(D)** Comparison of the binding affinity of prototype S protein and V367F mutant to human ACE2 by ligand-receptor binding ELISA assay.









407

Fig. 4: Structural analysis of RBD mutants and the effects on their binding affinity. (A) Spatial location of the mutant amino acids and the fragment 510-524. **(B)** RMSF of the nine mutants were compared to that of the prototype. 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. **(D)** View of the interaction surface of ACE2, with charge density noted. The arginine of the W436R mutant is in the proximity of the negatively charged amino acids. The electrostatic surface charges of the ACE2 are calculated using Pymol, with the charge unit K_bT/e_c, as noted in the Pymol manual.

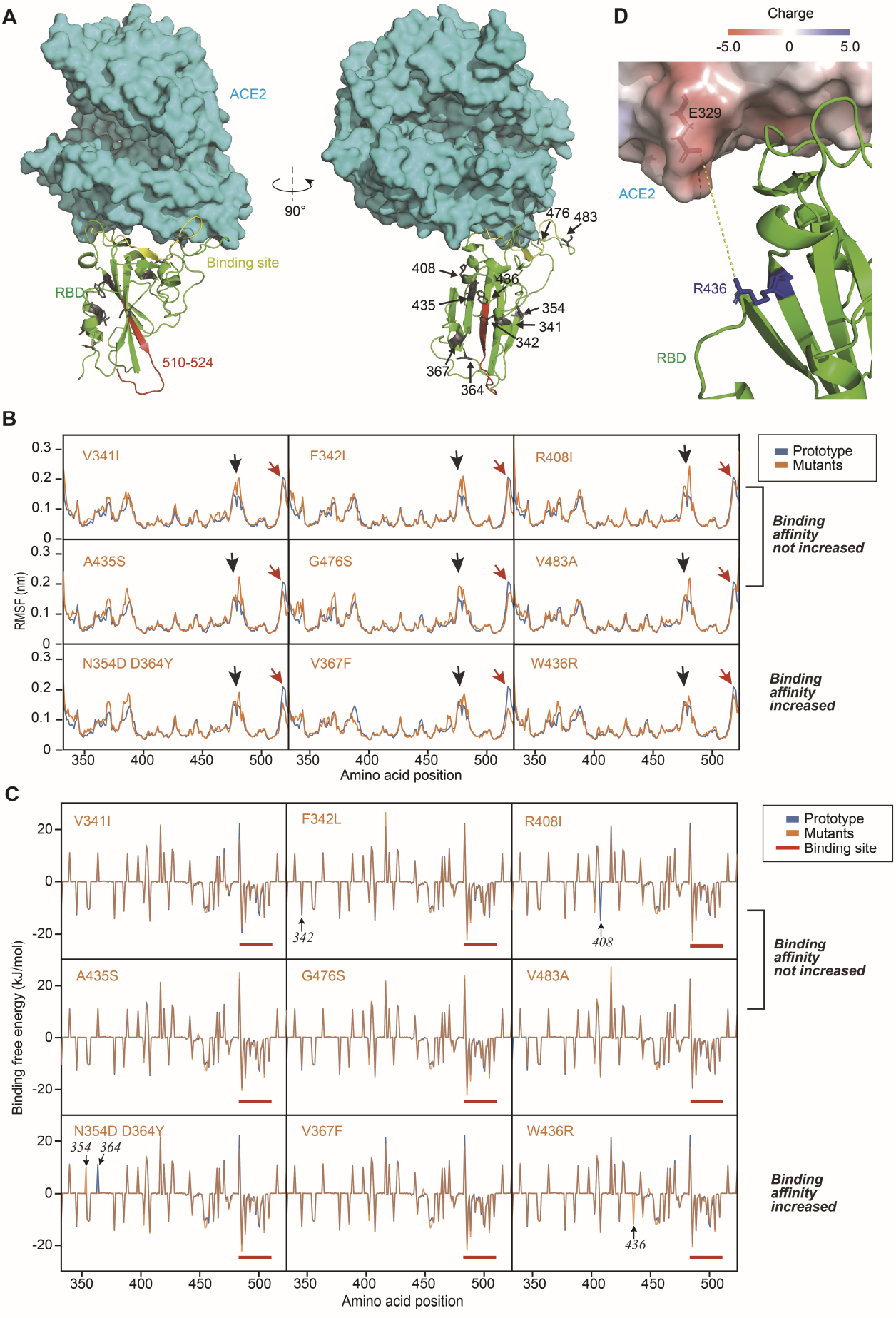


Table 1: Nucleotide substitution rates and selection pressures for S gene.

The numbers of nonsynonymous and synonymous differences per sequence from averaging over all sequence pairs are shown. Analyses were conducted using the Nei-Gojobori model. The analysis involved 1609 SARS-CoV-2 S gene sequences. All positions containing gaps and missing data were discarded.

Gene	Length(bp)	Mean Non-syonymous Substations/site	Mean Syonymous 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

Supplementary data

Supplementary Table 1: Meta data of the strains with non-synonymous mutations in the RBD of spike glycoprotein.

GISAID Virus name	RBD mutation	Collection date	Location	Gender	Age	Specimen source	Additional information	Accession ID
hCoV-19/Wuhan/IVDC-HB-envF13/2020	W436R	2020/1/1	Asia/China/Hubei/Wuhan	Unknown	Unknown	Environment	HuananSeafoodMarket	EPI_ISL_408511
hCoV-19/Shenzhen/SZTH-004/2020	N354D, D364Y	2020/1/16	Asia/China/Guandong/Shenzhen	Male	63	Alveolarlavage fluid		EPI_ISL_406595
hCoV-19/HongKong/VM20001061/2020	V367F	2020/1/22	Asia/HongKong	Male	39	Nasopharyngealaspirate&Throatswab		EPI_ISL_412028
hCoV-19/France/IDF0372/2020	V367F	2020/1/23	Europe/France/Ile-de-France/Paris	Female	31	Oro-Pharyngealswab		EPI_ISL_406596
hCoV-19/France/IDF0372-isl/2020	V367F	2020/1/23	Europe/France/Ile-de-France/Paris	Female	31	Oro-Pharyngealswab		EPI_ISL_410720
hCoV-19/France/IDF0373/2020	V367F	2020/1/23	Europe/France/Ile-de-France/Paris	Male	32	Orao-pharungealswab		EPI_ISL_406597
hCoV-19/India/1-27/2020	R408I	2020/1/27	Asia/India/Kerala	Female	20	Throatswab	TravelhistorytoChina	EPI_ISL_413522
hCoV-19/France/IDF0386-islP1/2020	V367F	2020/1/28	Europe/France/Ile-de-France/Paris	Female	30	Naso-pharyngealswab	relatedtoEPI_ISL_406596	EPI_ISL_411219
hCoV-19/France/IDF0386-islP3/2020	V367F	2020/1/28	Europe/France/Ile-de-France/Paris	Female	30	Naso-pharyngealswab	relatedtoEPI_ISL_406596	EPI_ISL_411220
hCoV-19/England/01/2020	F342L	2020/1/29	Europe/England	Female	50	swab	Englandclusterpatient1	EPI_ISL_407071
hCoV-19/USA/WA-S106/2020	V483A	2020/2/29	NorthAmerica/USA/Washington	Unknown	Unknown			EPI_ISL_417159
hCoV-19/USA/WA-S86/2020	V483A	2020/3/1	NorthAmerica/USA/Washington	Unknown	Unknown			EPI_ISL_417139
hCoV-19/USA/WA-S19/2020	V483A	2020/3/2	NorthAmerica/USA/Washington	Unknown	Unknown			EPI_ISL_417072

hCoV-19/USA/WA-S20/2020	V483A	2020/3/2	NorthAmerica/USA/Washington	Unknown	Unknown	EPI_ISL_417073
hCoV-19/USA/WA-S22/2020	V483A	2020/3/2	NorthAmerica/USA/Washington	Unknown	Unknown	EPI_ISL_417075
hCoV-19/USA/WA-S23/2020	V483A	2020/3/2	NorthAmerica/USA/Washington	Unknown	Unknown	EPI_ISL_417076
hCoV-19/USA/WA-S28/2020	G476S	2020/3/2	NorthAmerica/USA/Washington	Unknown	Unknown	EPI_ISL_417081
hCoV-19/USA/WA-S32/2020	G476S	2020/3/2	NorthAmerica/USA/Washington	Unknown	Unknown	EPI_ISL_417085
hCoV-19/Finland/FIN03032020A/ 2020	A435S	2020/3/3	Europe/Finland/Helsinki	Male	40	EPI_ISL_413602
hCoV-19/USA/WA-S58/2020	V483A	2020/3/5	NorthAmerica/USA/Washington	Unknown	Unknown	EPI_ISL_417111
hCoV-19/USA/WA-S107/2020	V483A	2020/3/5	NorthAmerica/USA/Washington	Unknown	Unknown	EPI_ISL_417160
hCoV-19/USA/WA-UW40/2020	V483A	2020/3/5	NorthAmerica/USA/Washington	Unknown	Unknown	EPI_ISL_415605
hCoV-19/USA/WA-UW261/2020	V483A	2020/3/5	NorthAmerica/USA/Washington	Unknown	Unknown	EPI_ISL_418046
hCoV-19/USA/WA-UW68/2020	V483A	2020/3/9	NorthAmerica/USA/Washington	Unknown	Unknown	EPI_ISL_415596
hCoV-19/USA/WA-UW91/2020	G476S	2020/3/10	NorthAmerica/USA/Washington	Unknown	Unknown	EPI_ISL_416447
hCoV-19/Wales/PHW27/2020	V341I	2020/3/12	Europe/UnitedKingdom/Wales	Male	49	EPI_ISL_415655
hCoV-19/USA/WA-UW270/2020	G476S	2020/3/13	NorthAmerica/USA/Washington	Unknown	Unknown	EPI_ISL_418055
hCoV-19/USA/WA-UW204/2020	G476S	2020/3/13	NorthAmerica/USA/Washington	Unknown	Unknown	EPI_ISL_417353
hCoV-19/USA/WA-UW231/2020	G476S	2020/3/14	NorthAmerica/USA/Washington	Unknown	Unknown	EPI_ISL_417380
hCoV-19/USA/WA-UW286/2020	V483A	2020/3/15	NorthAmerica/USA/Washington	Unknown	Unknown	EPI_ISL_418071
hCoV-19/USA/WA-UW292/2020	G476S	2020/3/15	NorthAmerica/USA/Washington	Unknown	Unknown	EPI_ISL_418077
hCoV-19/USA/WA-UW244/2020	V483A	2020/3/16	NorthAmerica/USA/Washington	Unknown	Unknown	EPI_ISL_418029

426

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

```
330       340       350       360       370
|....|....|....|....|....|....|....|....|....|....
PNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKC SARS-CoV-2 Wuhan-Hu-1
.....................................F..... SARS-CoV-2 France/IDF0372
.....................................F..... SARS-CoV-2 France/IDF0373
.....................................F..... SARS-CoV-2 France/IDF0372-isl
.....................................F..... SARS-CoV-2 France/IDF0386-islP1
.....................................F..... SARS-CoV-2 France/IDF0386-islP3
.....................................F..... SARS-CoV-2 Hong Kong/VM20001061
.......................................... SARS-CoV-2 Wuhan/IVDC-HB-envF13
.....................D.........Y.......... SARS-CoV-2 Shenzhen/SZTH-004
.......................................... SARS-CoV-2 Finland/FIN03032020A
..........I............................... SARS-CoV-2 Wales/PHW27
...........L............................... SARS-CoV-2 England/01
.......................................... SARS-CoV-2 India/1-27
.......................................... SARS-CoV-2 USA/WA-UW31
.......................................... SARS-CoV-2 USA/WA-S19
.......................................... SARS-CoV-2 USA/WA-S20
.......................................... SARS-CoV-2 USA/WA-S22
.......................................... SARS-CoV-2 USA/WA-S23
.......................................... SARS-CoV-2 USA/WA-S58
.......................................... SARS-CoV-2 USA/WA-S86
.......................................... SARS-CoV-2 USA/WA-S106
.......................................... SARS-CoV-2 USA/WA-S107
.......................................... SARS-CoV-2 USA/WA-UW68
.......................................... SARS-CoV-2 USA/WA-UW244
.......................................... SARS-CoV-2 USA/WA-UW286
.......................................... SARS-CoV-2 USA/WA-UW40
.......................................... SARS-CoV-2 USA/WA-UW292
.......................................... SARS-CoV-2 USA/WA-S28
.......................................... SARS-CoV-2 USA/WA-S32
.......................................... SARS-CoV-2 USA/WA-UW91
.......................................... SARS-CoV-2 USA/WA-UW270
.......................................... SARS-CoV-2 USA/WA-UW204
.......................................... SARS-CoV-2 USA/WA-UW231
..............T.......................T..... Bat SARS-like Yunnan/RaTG13
..............T.......................T..... Pangolin SARS-like Guandong/1
```

```
380       390       400       410       420
|....|....|....|....|....|....|....|....|....|....
YGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDF SARS-CoV-2 Wuhan-Hu-1
.......................................................... SARS-CoV-2 France/IDF0372
.......................................................... SARS-CoV-2 France/IDF0373
.......................................................... SARS-CoV-2 France/IDF0372-isl
.......................................................... SARS-CoV-2 France/IDF0386-islP1
.......................................................... SARS-CoV-2 France/IDF0386-islP3
.......................................................... SARS-CoV-2 Hong Kong/VM20001061
.......................................................... SARS-CoV-2 Wuhan/IVDC-HB-envF13
.......................................................... SARS-CoV-2 Shenzhen/SZTH-004
.......................................................... SARS-CoV-2 Finland/FIN03032020A
.......................................................... SARS-CoV-2 Wales/PHW27
.......................................................... SARS-CoV-2 England/01
.......................I.................................. SARS-CoV-2 India/1-27
.......................................................... SARS-CoV-2 USA/WA-UW31
.......................................................... SARS-CoV-2 USA/WA-S19
.......................................................... SARS-CoV-2 USA/WA-S20
.......................................................... SARS-CoV-2 USA/WA-S22
.......................................................... SARS-CoV-2 USA/WA-S23
.......................................................... SARS-CoV-2 USA/WA-S58
.......................................................... SARS-CoV-2 USA/WA-S86
.......................................................... SARS-CoV-2 USA/WA-S106
.......................................................... SARS-CoV-2 USA/WA-S107
.......................................................... SARS-CoV-2 USA/WA-UW68
.......................................................... SARS-CoV-2 USA/WA-UW244
.......................................................... SARS-CoV-2 USA/WA-UW286
.......................................................... SARS-CoV-2 USA/WA-UW40
.......................................................... SARS-CoV-2 USA/WA-UW292
.......................................................... SARS-CoV-2 USA/WA-S28
.......................................................... SARS-CoV-2 USA/WA-S32
.......................................................... SARS-CoV-2 USA/WA-UW91
.......................................................... SARS-CoV-2 USA/WA-UW270
.......................................................... SARS-CoV-2 USA/WA-UW204
.......................................................... SARS-CoV-2 USA/WA-UW231
.....................T.................................. Bat SARS-like Yunnan/RaTG13
.....................V.............R.......... Pangolin SARS-like Guandong/1
```

```
510  430        440        450        460        470
511  |....|....|....|....|....|....|....|....|....|....
512  TGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTP SARS-CoV-2 Wuhan-Hu-1
513  ............................................................ SARS-CoV-2 France/IDF0372
514  ............................................................ SARS-CoV-2 France/IDF0373
515  ............................................................ SARS-CoV-2 France/IDF0372-isl
516  ............................................................ SARS-CoV-2 France/IDF0386-islP1
517  ............................................................ SARS-CoV-2 France/IDF0386-islP3
518  ............................................................ SARS-CoV-2 Hong Kong/VM20001061
519  ......R..................................................... SARS-CoV-2 Wuhan/IVDC-HB-envF13
520  ............................................................ SARS-CoV-2 Shenzhen/SZTH-004
521  .....S...................................................... SARS-CoV-2 Finland/FIN03032020A
522  ............................................................ SARS-CoV-2 Wales/PHW27
523  ............................................................ SARS-CoV-2 England/01
524  ............................................................ SARS-CoV-2 India/1-27
525  ............................................................ SARS-CoV-2 USA/WA-UW31
526  ............................................................ SARS-CoV-2 USA/WA-S19
527  ............................................................ SARS-CoV-2 USA/WA-S20
528  ............................................................ SARS-CoV-2 USA/WA-S22
529  ............................................................ SARS-CoV-2 USA/WA-S23
530  ............................................................ SARS-CoV-2 USA/WA-S58
531  ............................................................ SARS-CoV-2 USA/WA-S86
532  ............................................................ SARS-CoV-2 USA/WA-S106
533  ............................................................ SARS-CoV-2 USA/WA-S107
534  ............................................................ SARS-CoV-2 USA/WA-UW68
535  ............................................................ SARS-CoV-2 USA/WA-UW244
536  ............................................................ SARS-CoV-2 USA/WA-UW286
537  ............................................................ SARS-CoV-2 USA/WA-UW40
538  ..........................................S... SARS-CoV-2 USA/WA-UW292
539  ..........................................S... SARS-CoV-2 USA/WA-S28
540  ..........................................S... SARS-CoV-2 USA/WA-S32
541  ..........................................S... SARS-CoV-2 USA/WA-UW91
542  ..........................................S... SARS-CoV-2 USA/WA-UW270
543  ..........................................S... SARS-CoV-2 USA/WA-UW204
544  ..........................................S... SARS-CoV-2 USA/WA-UW231
545  .........KHI.A.E...F.........A..................K. Bat SARS-like Yunnan/RaTG13
546  ............................................................ Pangolin SARS-like Guandong/1
547
548
```

```
549  480       490       500       510       520
550  |....|....|....|....|....|....|....|....|....
551  CNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATV      SARS-CoV-2 Wuhan-Hu-1
552  .....................................................      SARS-CoV-2 France/IDF0372
553  .....................................................      SARS-CoV-2 France/IDF0373
554  .....................................................      SARS-CoV-2 France/IDF0372-isl
555  .....................................................      SARS-CoV-2 France/IDF0386-islP
556  .....................................................      SARS-CoV-2 France/IDF0386-islP
557  .....................................................      SARS-CoV-2 Hong Kong/VM2000106
558  .....................................................      SARS-CoV-2 Wuhan/IVDC-HB-envF1
559  .....................................................      SARS-CoV-2 Shenzhen/SZTH-004
560  .....................................................      SARS-CoV-2Finland/FIN03032020A
561  .....................................................      SARS-CoV-2 Wales/PHW27
562  .....................................................      SARS-CoV-2 England/01
563  .....................................................      SARS-CoV-2 India/1-27
564  ...A.................................................      SARS-CoV-2 USA/WA-UW31
565  ...A.................................................      SARS-CoV-2 USA/WA-S19
566  ...A.................................................      SARS-CoV-2 USA/WA-S20
567  ...A.................................................      SARS-CoV-2 USA/WA-S22
568  ...A.................................................      SARS-CoV-2 USA/WA-S23
569  ...A.................................................      SARS-CoV-2 USA/WA-S58
570  ...A.................................................      SARS-CoV-2 USA/WA-S86
571  ...A.................................................      SARS-CoV-2 USA/WA-S106
572  ...A.................................................      SARS-CoV-2 USA/WA-S107
573  ...A.................................................      SARS-CoV-2 USA/WA-UW68
574  ...A.................................................      SARS-CoV-2 USA/WA-UW244
575  ...A.................................................      SARS-CoV-2 USA/WA-UW286
576  ...A.................................................      SARS-CoV-2 USA/WA-UW40
577  .....................................................      SARS-CoV-2 USA/WA-UW292
578  .....................................................      SARS-CoV-2 USA/WA-S28
579  .....................................................      SARS-CoV-2 USA/WA-S32
580  .....................................................      SARS-CoV-2 USA/WA-UW91
581  .....................................................      SARS-CoV-2 USA/WA-UW270
582  .....................................................      SARS-CoV-2 USA/WA-UW204
583  .....................................................      SARS-CoV-2 USA/WA-UW231
584  ...QT.L...Y..YR...Y..D...H..........N.....      Bat SARS-like Yunnan/RaTG13
585  ................H..................N.....      Pangolin SARS-like Guandong/1
586
587
```