Emergence of RBD mutations from circulating SARS-CoV-2 strains with enhanced structural stability and higher human ACE2 receptor affinity of the spike protein

Junxian Ou[1]†, Zhonghua Zhou[2]†, Ruixue Dai[3,6]†, Jing Zhang[4], Wendong Lan[1], Shan Zhao[1], Jianguo Wu[4], Donald Seto[5], Lilian Cui[6], Gong Zhang[2]*, Qiwei Zhang[1,4]*

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

[2] 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.

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

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

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

[6] Novoprotein Scientific Inc. Shanghai 201203, China

†These authors contributed equally to this work.

***Correspondence**:

Qiwei Zhang, Guangdong Provincial Key Laboratory of Tropical Disease Research, School of Public Health, Southern Medical University, Guangzhou, Guangdong 510515, China, Tel: 86-20-61648649, Fax: 86-20-61648324, Email: zhangqw@smu.edu.cn;

Gong Zhang, Key Laboratory of Functional Protein Research of Guangdong Higher Education

Institutes and MOE Key Laboratory of Tumor Molecular Biology, Institute of Life and Health Engineering, College of Life Science and Technology, Jinan University, Guangzhou, Guangdong 510632, China, Tel: +86-20-85224031; Email: zhanggong@jnu.edu.cn.

Abstract

Spike protein receptor binding domain (RBD) of SARS-CoV-2 is the critical determinant of viral tropism and infectivity. Genomes of global SARS-CoV-2 strains were analyzed, yielding 32 RBD mutant strains clustering into 10 mutation types under high positive selection pressure. Three mutation types circulating in Wuhan, Shenzhen, Hong Kong, and France, displayed enhanced structural stability along with higher human ACE2 receptor affinity of the spike protein, indicating these mutants may have acquired increased infectivity to humans.

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

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 5, 2020, SARS-CoV-2 has infected 1,051,635 people world-wide and caused 56,985 deaths[5] with the estimated fatality rate of 5.41%. This on-going pandemic of COVID-19 has become the most serious threat to public health in recent times.

Angiotensin-converting enzyme 2 (ACE2) is the cellular receptor of SARS-CoV-2[6], 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[6,7]. 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.

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 mutated strains to date and human ACE2 receptor were modelled and assessed.

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 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 10 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 strains may have originated as a sub-lineage. Similarly, novel sub-lineages may account for the 13 V483A mutants and 7 G476S

mutants isolated in the U.S.A. (Fig. 1).

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) [10] 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 4.6526, remarkably greater than 1, showing that the S gene is indeed under positive selective pressure (Supplementary Table 2). The RBD showed a similar dN/dS (4.3966) 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.

Three mutation types bind human ACE2 receptor with higher affinity

To estimate the functional changes suggested by the RBD mutations, we performed molecular dynamics simulation 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). 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[11]. 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 mutation 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 mutation 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[8], 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 affinity change 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).

Structural basis for the increased affinity

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). One 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).

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 mutates 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 suggests a higher risk of more severe morbidity and mortality during a sustained pandemic of COVID-19 if no effective precautions are implemented.

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 [8]) (Fig. 3), indicating that the pangolin CoV has the potential to infect human in unprotected close contact. The pangolin CoV has been detected among the smuggled Malayan pangolins in multiple provinces in China[12,13], 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. Since the RBD contains important antigenic epitopes, mutations in RBD, especially those which change the amino acid properties, may weaken the binding of the antibody raised against the prototype strain. This may lead to decreased vaccine efficacy.

In summary, we have identified 32 RBD mutant strains clustering into 10 mutation types under high positive selection pressure. Three mutation types emerged in Asia and Europe; these display enhanced structural stability of the spike protein along with higher binding affinities to human ACE2 receptor, and indicate these mutants may have acquired increased infectivity to humans.

Reference

1. Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med*. 2020:727-733. doi:10.1056/nejmoa2001017
2. Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *N Engl J Med*. 2020:1-9. doi:10.1056/nejmoa2001316

3. Wang D, Hu B, Hu C, et al. Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA - J Am Med Assoc*. 2020:1-9. doi:10.1001/jama.2020.1585
4. Chan JFW, Yuan S, Kok KH, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet*. 2020;395(10223):514-523. doi:10.1016/S0140-6736(20)30154-9
5. Mariana N. Coronavirus disease 2019 (COVID-19). 2020;2019(April).
6. Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*. March 2020:S0092-8674(20)30229-4. doi:10.1016/j.cell.2020.02.052
7. Chen Y, Guo Y, Pan Y, Zhao ZJ. Structure analysis of the receptor binding of 2019-nCoV. *Biochem Biophys Res Commun*. 2020;2(xxxx):0-5. doi:10.1016/j.bbrc.2020.02.071
8. Wrapp D, Wang N, Corbett KS, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science (80-)*. 2020;367(6483):1260 LP - 1263. doi:10.1126/science.abb2507
9. Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor recognition by novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS. *J Virol*. 2020;(January). doi:10.1128/jvi.00127-20
10. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol*. 2017;34(12):3299-3302. doi:10.1093/molbev/msx248
11. Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;46(W1):W296-W303. doi:10.1093/nar/gky427
12. Lam TT-Y, Shum MH-H, Zhu H-C, et al. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature*. 2020. doi:10.1038/s41586-020-2169-0
13. Xiao K, Zhai J, Feng Y, et al. Isolation and Characterization of 2019-nCoV-like Coronavirus from Malayan Pangolins. *bioRxiv*. January 2020:2020.02.17.951335.

doi:10.1101/2020.02.17.951335

Figure legends:

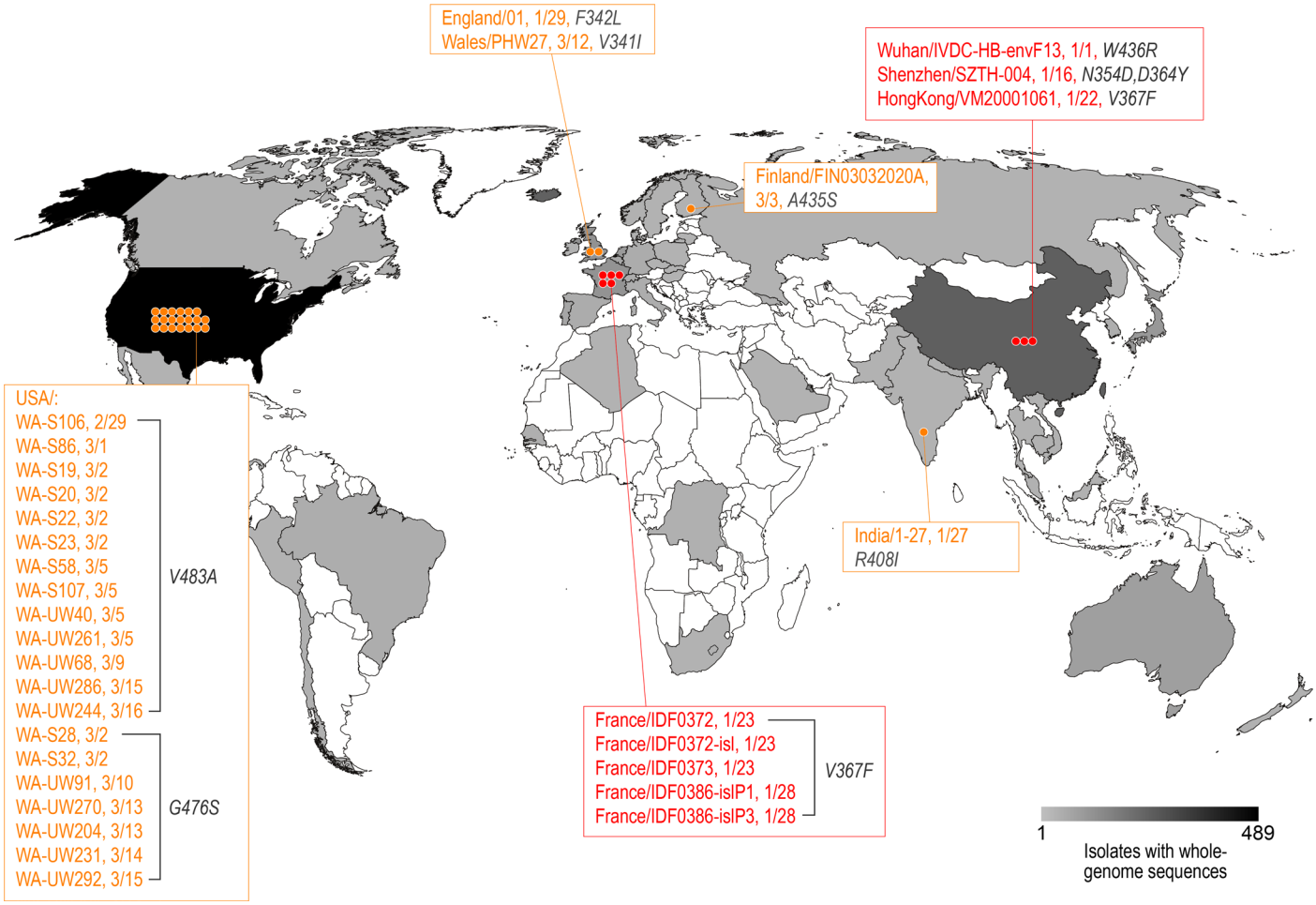

Fig. 1: Distribution of the SARS-CoV-2 strains mutated in the RBD of the S protein. The geographic distribution of the RBD mutant strains is displayed. The strains with names highlighted in red are mutants with the enhanced binding affinity. The strains with names noted in yellow are mutants with similar binding affinities.

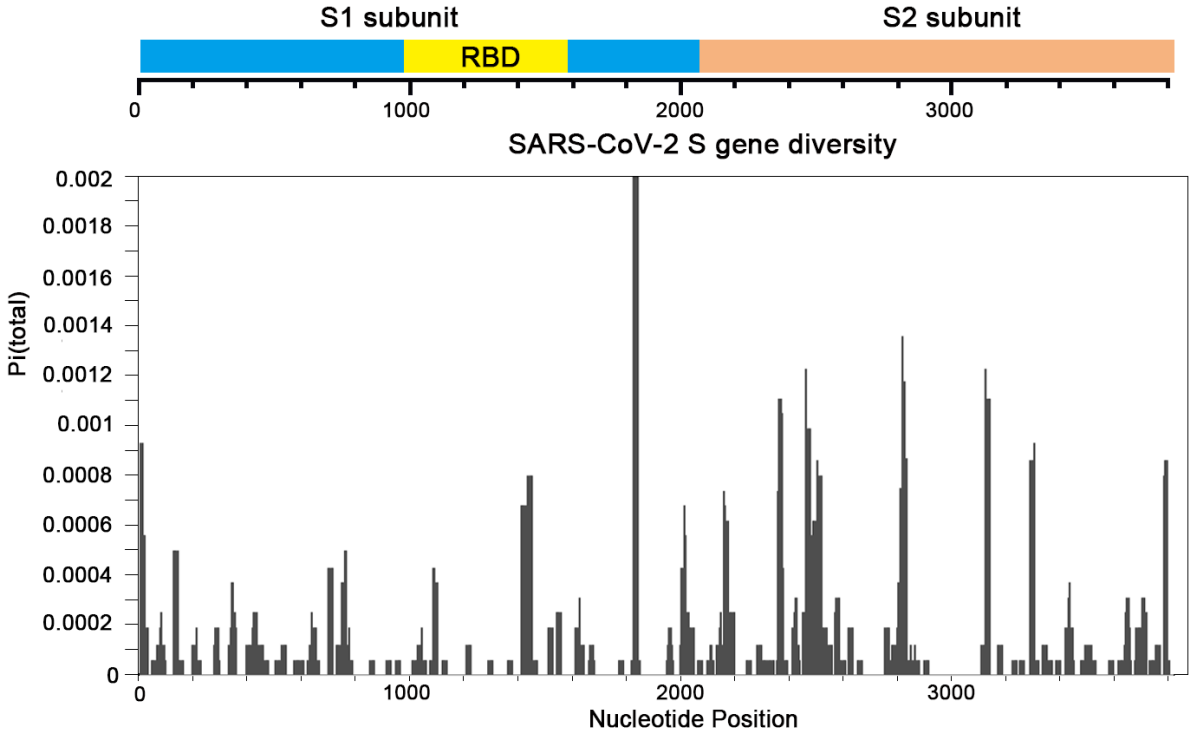

Fig. 2: Polymorphism and divergence graph of SARS-CoV-2 S gene. Polymorphism and divergence were analyzed by DnaSP6 (version 6.12.03). Analyses were conducted using the Nei-Gojobori model. All positions containing gaps and missing data were eliminated. Structural domains are annotated. The Pi values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

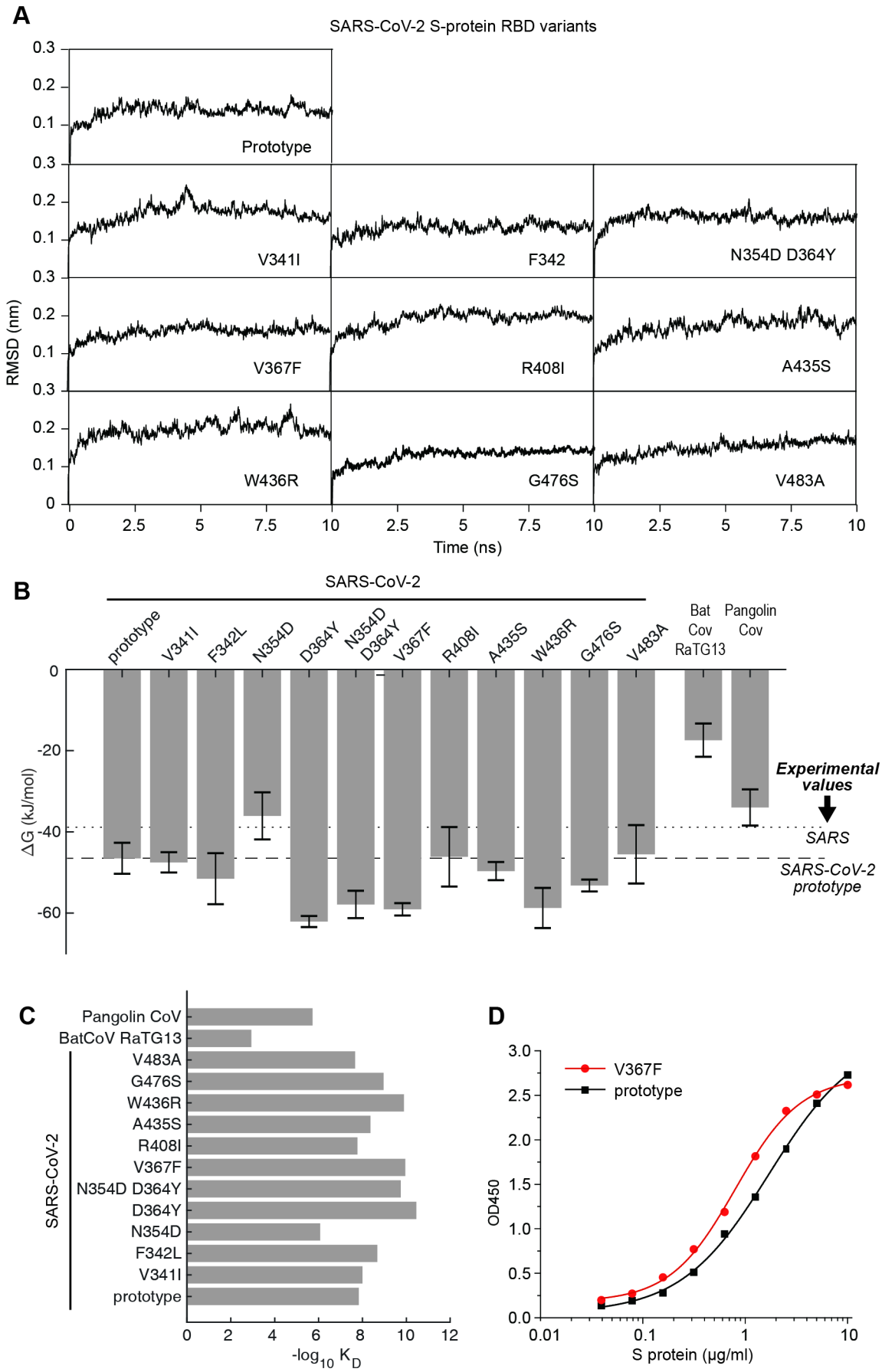
A
SARS-CoV-2 S-protein RBD variants
RMSD (nm)
Prototype
V341I
F342
N354D D364Y
V367F
R408I
A435S
W436R
G476S
V483A
Time (ns)
B
SARS-CoV-2
prototype
V341I
F342L
N354D
D364Y
N354D D364Y
V367F
R408I
A435S
W436R
G476S
V483A
Bat Cov RaTG13
Pangolin Cov
ΔG (kJ/mol)
Experimental values
SARS
SARS-CoV-2 prototype
C
Pangolin CoV
BatCoV RaTG13
SARS-CoV-2
V483A
G476S
W436R
A435S
R408I
V367F
N354D D364Y
D364Y
N354D
F342L
V341I
prototype
-log10 KD
D
V367F
prototype
OD450
S protein (µg/ml)

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.

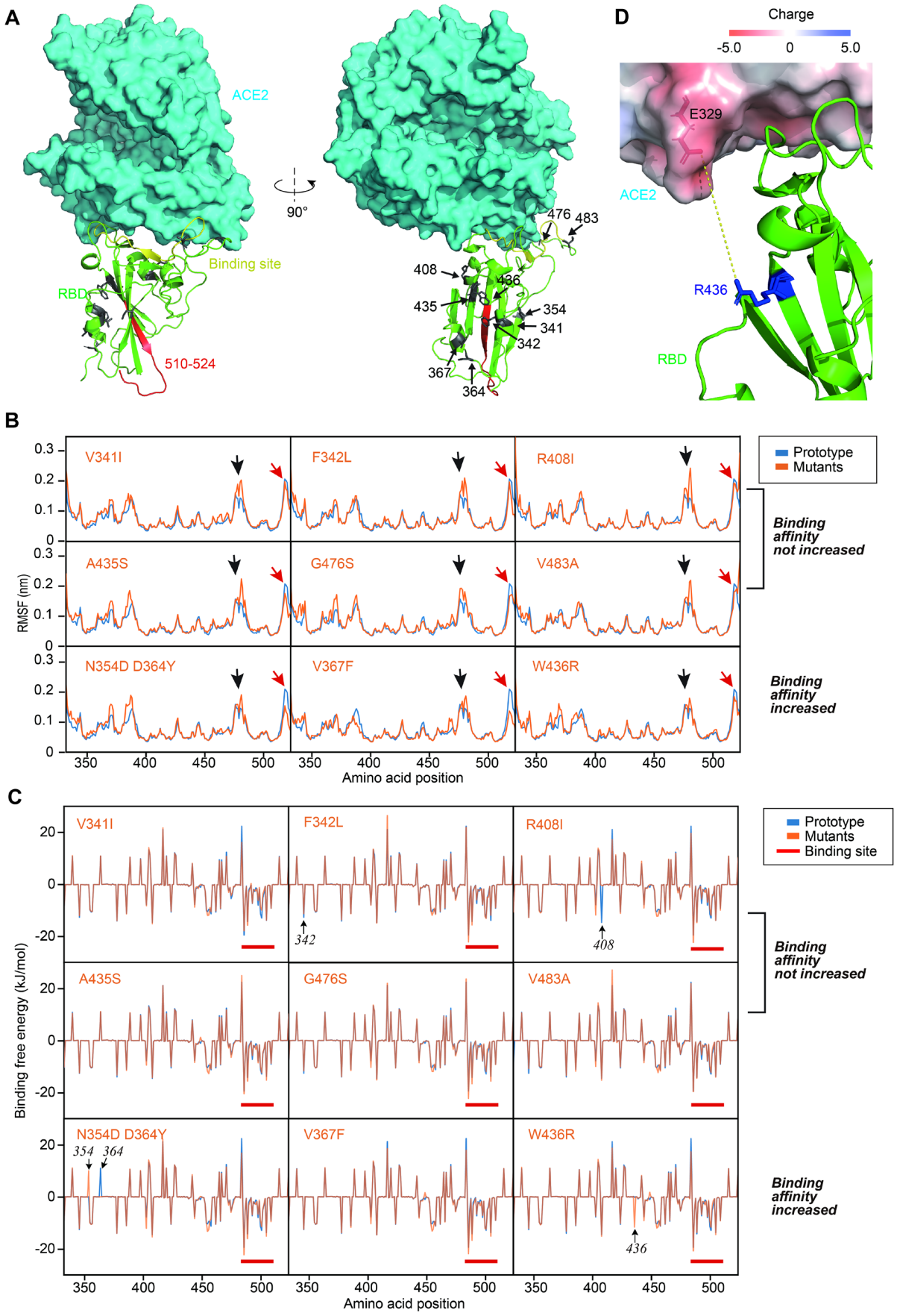
A
ACE2
Binding site
RBD
510-524
90°
476
483
408
436
435
354
341
342
367
364
D
Charge
-5.0
0
5.0
E329
ACE2
R436
RBD
B
V341I
F342L
R408I
A435S
G476S
V483A
N354D D364Y
V367F
W436R
RMSF (nm)
Amino acid position
Prototype
Mutants
Binding affinity not increased
Binding affinity increased
C
Binding free energy (kJ/mol)
342
408
354
364
436
Binding site

Fig. 4: Structural analysis of RBD mutants and the effects on their binding affinity. (A) Spatial location of the mutated 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.

Supplementary Table 1: Meta data of the strains with 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	notapplicable		Environment	HuananSeafoodMarket	EPI_ISL_408511
hCoV-19/Shenzhen/SZTH-004/2020	N354D,D364Y	2020/1/16	Asia/China/Guandong/Shenzhen	Male	63	Alveolarlavagefluid		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

Supplementary Table 2: Estimates of Average Codon-based Evolutionary Divergence over S gene Pairs.

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 660 SARS-CoV-2 nucleotide 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	0.7233	0.1555	4.6526
S1	2043	0.5885	0.0544	10.8094
S1-RBD	585	1.0514	0.2391	4.3966
S2	1779	0.1348	0.1010	1.3342

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
..............S..................................... SARS-CoV-2 USA/WA-UW226
.................................................... 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-UW226
.................................................. 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
```

```
430       440       450       460       470
|....|....|....|....|....|....|....|....|....|....
TGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTP 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
......R........................................... SARS-CoV-2 Wuhan/IVDC-HB-envF13
.................................................. SARS-CoV-2 Shenzhen/SZTH-004
.....S............................................ SARS-CoV-2 Finland/FIN03032020A
.................................................. SARS-CoV-2 Wales/PHW27
.................................................. SARS-CoV-2 England/01
.................................................. SARS-CoV-2 India/1-27
.................................................. SARS-CoV-2 USA/WA-UW226
.................................................. 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
...............................................S... SARS-CoV-2 USA/WA-UW292
...............................................S... SARS-CoV-2 USA/WA-S28
...............................................S... SARS-CoV-2 USA/WA-S32
...............................................S... SARS-CoV-2 USA/WA-UW91
...............................................S... SARS-CoV-2 USA/WA-UW270
...............................................S... SARS-CoV-2 USA/WA-UW204
...............................................S... SARS-CoV-2 USA/WA-UW231
........KHI.A.E...F........A................K. Bat SARS-like Yunnan/RaTG13
.................................................. Pangolin SARS-like Guandong/1
```

```
480       490       500       510       520
|....|....|....|....|....|....|....|....|....
CNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATV      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-islP
.........................................      SARS-CoV-2 France/IDF0386-islP
.........................................      SARS-CoV-2 Hong Kong/VM2000106
.........................................      SARS-CoV-2 Wuhan/IVDC-HB-envF1
.........................................      SARS-CoV-2 Shenzhen/SZTH-004
.........................................      SARS-CoV-2 Finland/FIN03032020A
.........................................      SARS-CoV-2 Wales/PHW27
.........................................      SARS-CoV-2 England/01
.........................................      SARS-CoV-2 India/1-27
.........................................      SARS-CoV-2 USA/WA-UW226
...A.....................................      SARS-CoV-2 USA/WA-UW31
...A.....................................      SARS-CoV-2 USA/WA-S19
...A.....................................      SARS-CoV-2 USA/WA-S20
...A.....................................      SARS-CoV-2 USA/WA-S22
...A.....................................      SARS-CoV-2 USA/WA-S23
...A.....................................      SARS-CoV-2 USA/WA-S58
...A.....................................      SARS-CoV-2 USA/WA-S86
...A.....................................      SARS-CoV-2 USA/WA-S106
...A.....................................      SARS-CoV-2 USA/WA-S107
...A.....................................      SARS-CoV-2 USA/WA-UW68
...A.....................................      SARS-CoV-2 USA/WA-UW244
```

```
...A.......................................      SARS-CoV-2 USA/WA-UW286
...A.......................................      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
...QT.L...Y..YR...Y..D...H............N.....      Bat SARS-like Yunnan/RaTG13
................H....................N.....      Pangolin SARS-like Guandong/1
```