

 Open access • Posted Content • DOI:10.1101/2020.03.15.991844

Emergence of SARS-CoV-2 spike RBD mutants that enhance viral infectivity through increased human ACE2 receptor binding affinity — [Source link](#)

Junxian Ou, Zhonghua Zhou, Ruixue Dai, Shan Zhao ...+10 more authors

Institutions: Southern Medical University, Jinan University, Fudan University, George Mason University ...+1 more institutions

Published on: 12 Sep 2020 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Infectivity and Mutant

Related papers:

- [V367F mutation in SARS-CoV-2 spike RBD emerging during the early transmission phase enhances viral infectivity through increased human ACE2 receptor binding affinity](#)
- [Emergence of RBD mutations in circulating SARS-CoV-2 strains enhancing the structural stability and human ACE2 receptor affinity of the spike protein](#)
- [Emergence of RBD mutations from circulating SARS-CoV-2 strains with enhanced structural stability and higher human ACE2 receptor affinity of the spike protein](#)
- [Altered interaction between RBD and ACE2 receptor contributes towards the increased transmissibility of SARS CoV-2 delta, kappa, beta, and gamma strains with RBD double mutations.](#)
- [The high infectivity of SARS-CoV-2 B.1.1.7 is associated with increased interaction force between Spike-ACE2 caused by the viral N501Y mutation](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/emergence-of-sars-cov-2-spike-rbd-mutants-that-enhance-viral-2og4l75kwj>

1 **V367F mutation in SARS-CoV-2 spike RBD emerging during the early transmission phase**
2 **enhances viral infectivity through increased human ACE2 receptor binding affinity**

3
4 Junxian Ou,^{1†} Zhonghua Zhou,^{2†} Ruixue Dai,³ Shan Zhao,¹ Xiaowei Wu,¹ Jing Zhang,⁵ Wendong
5 Lan,¹ Lilian Cui,⁴ Jianguo Wu,⁵ Donald Seto,⁶ James Chodosh,⁷ Gong Zhang,^{2#} Qiwei Zhang^{1,5#}

6
7 ¹ Guangdong Provincial Key Laboratory of Tropical Disease Research, School of Public Health,
8 Southern Medical University, Guangzhou, Guangdong 510515, China

9 ² Key Laboratory of Functional Protein Research of Guangdong Higher Education Institutes,
10 Institute of Life and Health Engineering, College of Life Science and Technology, Jinan
11 University, Guangzhou, Guangdong 510632, China.

12 ³ Department of Environmental Science and Engineering, Fudan University, Shanghai 200433,
13 China

14 ⁴ Novoprotein Scientific Inc. Shanghai 201203, China

15 ⁵ Guangdong Provincial Key Laboratory of Virology, Institute of Medical Microbiology, Jinan
16 University, Guangzhou, Guangdong 510632, China

17 ⁶ Bioinformatics and Computational Biology Program, School of Systems Biology, George Mason
18 University, Manassas, VA 20110, USA

19 ⁷ Department of Ophthalmology, Howe Laboratory Massachusetts Eye and Ear Infirmary, Harvard
20 Medical School, Boston, MA 02114, USA

21
22 **Running Head:** SARS-CoV-2 V367F mutation enhances viral infectivity

23
24 **#Address correspondence** to Qiwei Zhang, zhangqw@jnu.edu.cn, ORCID: 0000-0002-2770-111X;
25 Gong Zhang, zhanggong@jnu.edu.cn, ORCID: 0000-0003-0418-3433.

26 J.O. and Z.Z. contributed equally to this work.

27

28

29 **Abstract**

30 The current global pandemic of COVID-19 is caused by a novel coronavirus SARS-CoV-2. The
31 SARS-CoV-2 spike protein receptor-binding domain (RBD) is the critical determinant of viral
32 tropism and infectivity. To investigate whether naturally occurring mutations in the RBD during the
33 early transmission phase have altered the receptor binding affinity and infectivity, firstly we
34 analyzed *in silico* the binding dynamics between mutated SARS-CoV-2 RBDs and the human ACE2
35 receptor. Among 32,123 genomes of SARS-CoV-2 isolates (January through March, 2020), 302
36 non-synonymous RBD mutants were identified and clustered into 96 mutant types. The six
37 dominant mutations were analyzed applying molecular dynamics simulations. The mutant type
38 V367F continuously circulating worldwide displayed higher binding affinity to human ACE2 due to
39 the enhanced structural stabilization of the RBD beta-sheet scaffold. The increased infectivity of
40 V367 mutants was further validated by performing receptor-ligand binding ELISA, surface plasmon
41 resonance, and pseudotyped virus assays. Genome phylogenetic analysis of V367F mutants showed
42 that during the early transmission phase, most V367F mutants clustered more closely with the
43 SARS-CoV-2 prototype strain than the dual-mutation variants (V367F + D614G) which emerged
44 later and formed a distinct sub-cluster. The analysis of critical RBD mutations provides further
45 insights into the evolutionary trajectory of SARS-CoV-2 under negative selection pressure and
46 supports the continuing surveillance of spike mutations to aid in the development of new
47 COVID-19 drugs and vaccines.

48

49 **Importance**

50 A novel coronavirus SARS-CoV-2 has caused the pandemic of COVID-19. The origin of
51 SARS-CoV-2 was associated with zoonotic infections. The spike protein receptor-binding domain
52 (RBD) is identified as the critical determinant of viral tropism and infectivity. Thus, whether the
53 mutations in the RBD of the circulating SARS-CoV-2 isolates have altered the receptor binding
54 affinity and caused them more infectious, should be paid more attentions to. Given that
55 SARS-CoV-2 is a novel coronavirus, the significance of our research is in identifying and validating
56 the RBD mutant types emerging during the early transmission phase that have increased human

57 ACE2 receptor binding affinity and infectivity. The RBD mutation analysis provides insights into
58 SARS-CoV-2 evolution. The continuous surveillance of RBD mutations with increased human
59 ACE2 affinity in human or other animals is important and necessary, particularly when the direct
60 correlation between the virus variations and vaccine effectiveness is underdetermined during the
61 sustained COVID-19 pandemic.

62

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

65

66 **1. Introduction**

67 A novel coronavirus SARS-CoV-2 has caused outbreaks of Coronavirus Disease 2019 (COVID-19)
68 globally beginning in mid-December 2019, with an epicenter in Wuhan, China(1–3) As of
69 November 22, 2020, SARS-CoV-2 had infected 57.8 million people world-wide and caused 1.3
70 million deaths with an estimated fatality rate of 2.25%(4). This on-going pandemic of COVID-19
71 has become the most serious threat to public health in this century.

72

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

77

78 Angiotensin-converting enzyme 2 (ACE2) is the host cellular receptor for the SARS-CoV-2 spike
79 glycoprotein, which is similar to its counterpart in SARS-CoV. The receptor-binding domain (RBD)
80 of the spike protein subunit S1 interacts directly with ACE2, providing for tight binding to the
81 peptidase domain of ACE2(7–10). Therefore, RBD is the critical determinant of virus-receptor
82 interaction and reflects viral host range, tropism, and infectivity. Although the RBD sequences of
83 different SARS-CoV-2 strains circulating globally are largely conserved, mutations have appeared;
84 these may account for differences in viral infectivity and also contribute to its spread(10–14).

85

86 Meanwhile, S glycoprotein participates in antigenic recognition and is expressed on its protein
87 surface, likely to be immunogenic as for carrying both T-cell and B-cell epitopes. The potential
88 antibody binding sites that have been identified indicates RBD has important B-cell epitopes. The
89 main antibody binding sites substantially overlap with RBD, and the antibody binding to these sites
90 is able to block viral entry into cells(14–16).

91

92 To investigate whether mutations in RBD emergent during the early transmission phase have altered
93 the receptor binding affinities and whether these isolates may have been selected for higher
94 infectivity, the binding dynamics and the infectivity between the SARS-CoV-2 RBD mutants and

95 human ACE2 receptor were modeled and assessed computationally. The evolutionary trajectory of
96 SARS-CoV-2 during the early transmission phase under negative selection pressure was also
97 analyzed. In addition, experimental validation of the enhanced affinity and infectivity of the V367F
98 mutant was performed.

99

100

101 **2. Results and discussion**

102 **2.1 SARS-CoV-2 Spike and RBD mutation mapping and scanning during the early** 103 **transmission phase**

104

105 32,123 SARS-CoV-2 isolates with whole-genome sequences available in public databases and
106 sampled before March 31, 2020 were analyzed. Among them, 302 isolates with RBD mutations
107 were identified when compared with reference strain Wuhun-Hu-1 that was the firstly isolated in
108 Wuhan in December 2019. All the mutants were clustered into 96 mutant types, six of which were
109 dominant mutant types that were found in more than ten isolates (**Table 1**): V483A (35x), V367F
110 (34x), V341I (23x), N439K (16x), A344S (15x), G476S (12x) (**Figure 2C**). V483A and V367F
111 accounted for 11.59% and 11.26% of 302 mutants, respectively. V367F mutants occurred on Jan 22,
112 2020, which was the earliest dominant mutant type. During the early transmission phase, there was
113 no D614G mutation in 35 V483A mutants; however, there were 11 of 34 V367F mutants that
114 harbored D614G mutation. The detailed alignment of amino acid sequences in the RBD of the
115 mutants were shown in **Supplementary Figure 1**. These mutants were emergent in multiple
116 continents, including Asia, Europe, North America and Oceania. Most RBD mutants were
117 circulating in Europe and North America (**Figure. 1**). V367F mutants were found in all the four
118 continents.

119

120

121 **Table 1. Dominant mutations in the spike RBD (>10 isolates) and their binding affinity change**
122 **(as of March 31, 2020).**

123 The numbers of dominant mutations in the spike RBD (as of March 31, 2020 collected) are shown.
124 The date of mutant occurrence, ID of the first mutant, and concurrence of D614 or G614 mutants
125 are identified. V483A mutants (35x) were only found in D614 isolates. V367F mutants were first
126 identified in D614 isolates (23x), then identified in G614 mutants (11x). Receptor binding
127 efficiencies of dominant mutants were evaluated by both amino acid property and binding free
128 energy change using the MM-PBSA method (18, 19). The number of total mutants are 302.

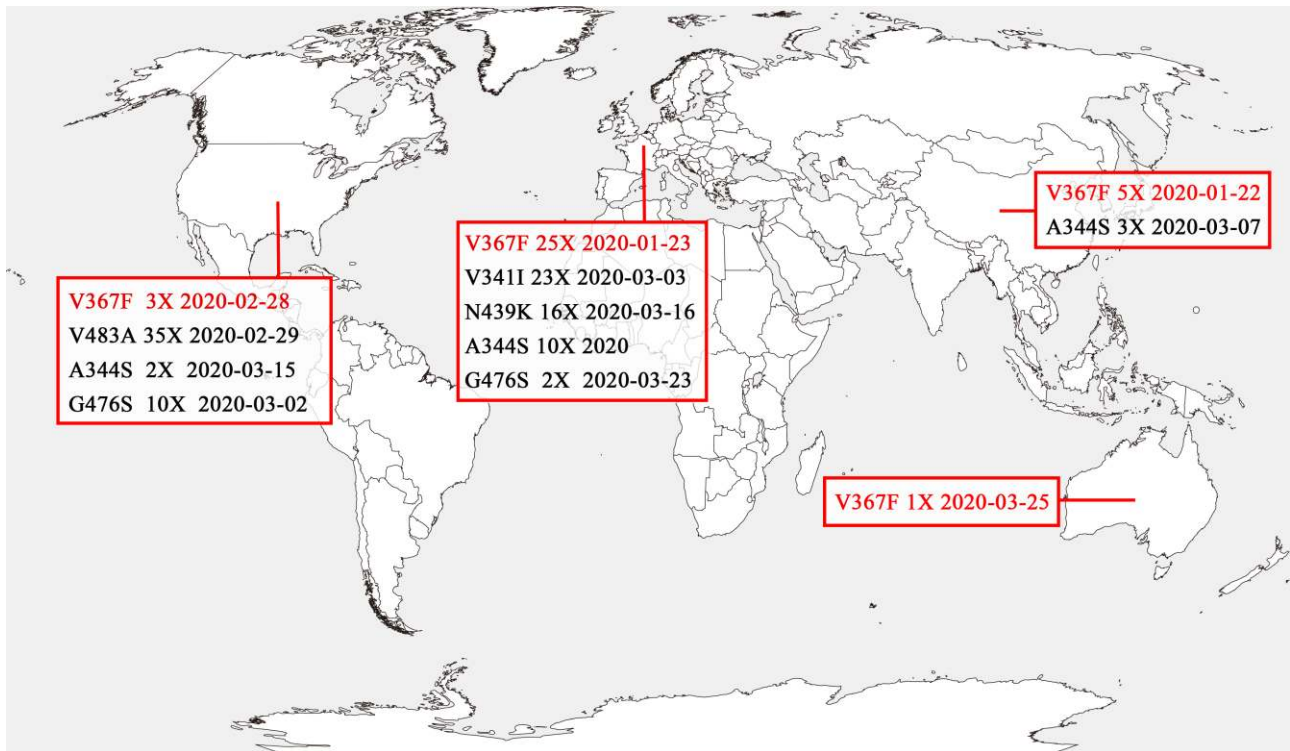
129

Amino acid position in spike gene	Amino acid change	Percentage of mutants	Date of occurrence	ID of the first mutant	D614	G614	Total	Binding free energy increase
483	V to A	11.59%	2020/2/29	EPI_ISL_417159	35	0	35	No
367	V to F	11.26%	2020/1/22	EPI_ISL_408975	23	11	34	Yes
341	V to I	7.95%	2020/3/3	EPI_ISL_454450	0	23	23	No
439	N to K	5.30%	2020/3/16	EPI_ISL_425684	0	16	16	No
344	A to S	4.97%	2020/3/7	EPI_ISL_506954	2	13	15	No
476	G to S	3.97%	2020/3/2	EPI_ISL_417081	2	10	12	No

130

131

132 **Fig. 1: Geographical Distribution of the SARS-CoV-2 RBD mutants.** The geographic
133 distribution of the RBD mutants in four continents (>10 isolates) is displayed. The mutants marked
134 in black are mutants with similar binding affinities with strain Wuhan-Hu-1. V367F mutants with
135 the enhanced binding affinity were found in all the four continents and marked in red. The mutants
136 analyzed were isolated as of March 31, 2020.



137

138

139

140 **2.2 Nucleotide diversity and selective pressure on SARS-CoV-2 spike gene and RBD**

141

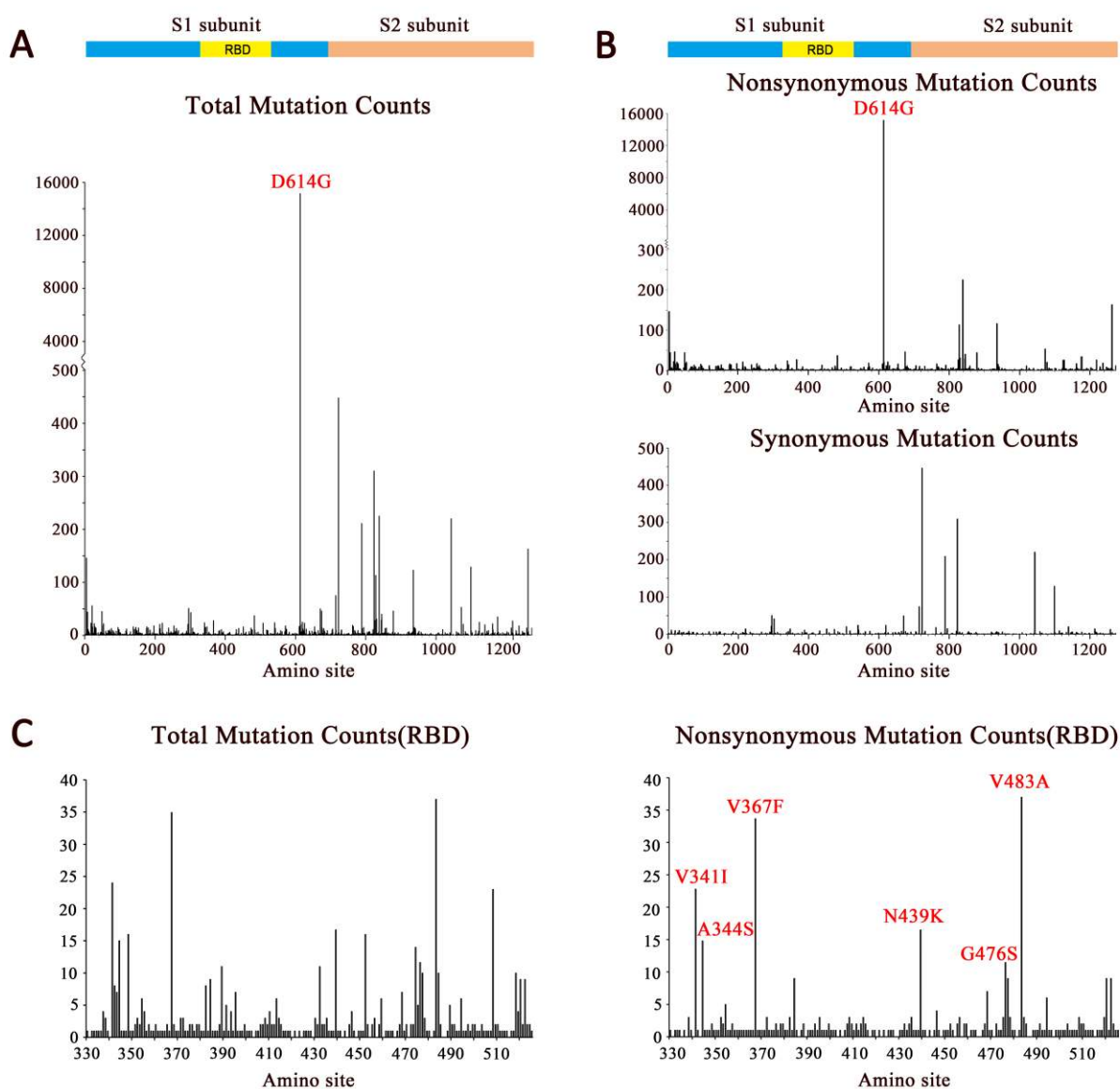
142 As SARS-CoV-2 S protein mediates attachment of the virus to cell-surface receptors and fusion
143 between virus and cell membranes, the polymorphism and divergence of S gene were analyzed by
144 DnaSP6 (version 6.12.03) (15). Overall, S2 subunit was more diverse than S1 subunit (**Figure 2A**).
145 D614G mutation in S1 subunit accounted for most of the polymorphism and divergence. Most of
146 the other synonymous and nonsynonymous mutations occurred in S2 subunit (**Figure 2B**), of which
147 synonymous mutations were more than nonsynonymous mutations. During the early transmission
148 phase, RBD was conserved and not as diverse as S2 subunit (**Figure 2A and 2B**). The six major
149 mutant types were shown in **Figure 2C**.

150

151 Since RBD is the only domain to bind human ACE2 receptor and, in turn, initiates cell entry, it is
152 believed that the RBD should be highly conserved. To test this hypothesis, we investigated the
153 selective pressures of the S gene during the early transmission phrase by calculating
154 nonsynonymous/synonymous substitution rate ratios (dN/dS ratios) for various segments of the S
155 gene from the 32,123 SARS-CoV-2 isolates (20, 21). The entire S gene exhibited a dN/dS of 0.86,
156 close to 0.83 reported previously (22), showing that the S gene is indeed under negative selection
157 (**Table 2**). However, different regions of the S gene may be subject to different selection pressures.
158 The S1 gene exhibited unusually high dN/dS (2.05) due to the widely spreading non-synonymous
159 mutation D614G in the S1 gene (23). The S2 gene exhibited a lower dN/dS (0.2388551), indicating
160 that the S2 gene was more conserved than S1 under negative selection. The S1-RBD showed lower
161 dN/dS (0.7258567) compared to the entire S gene (0.8569338). Therefore, the functional relevance
162 of the RBD mutations may be inferred.

163

164 **Fig. 2. Mutation scanning graph of SARS-CoV-2 S gene.** Mutation sites comparing to the strain
165 Wuhun-Hu-1 were analyzed by BioAider (V1.314) (17). **(A)** The total mutation (synonymous and
166 nonsynonymous) in spike gene; **(B)** The nonsynonymous and synonymous mutations in spike gene;
167 **(C)** The total mutations and nonsynonymous mutations in RBD. All positions containing gaps and
168 missing data were eliminated. Structural domains are annotated. The peak signal of D614G and the
169 other dominant mutations are marked in red. The ordinate showed the number of the mutants. The
170 isolates analyzed were isolated as of March 31, 2020.



171

172

173

174 **Table 2: Nucleotide substitution rates and selection pressures for S gene.**

175 The numbers of nonsynonymous and synonymous differences per sequence from averaging over all
176 sequence pairs are shown. Analyses were conducted using the Nei-Gojobori method(Jukes-cantor
177 model) in Mega X (10.0.2) (20, 21). The analyses involved 32,123 SARS-CoV-2 S gene sequences.
178 All positions containing gaps and missing data were discarded.

179

Gene	Length(bp)	Mean Non-synonymous Substations/site	Mean Synonymous Substations/site	dN/dS
S	3822	0.0002342	0.0002733	0.8569338
S1	2043	0.0003519	0.0001715	2.0518950
S1-RBD	585	0.0000699	0.0000963	0.7258567
S2	1779	0.0000943	0.0003948	0.2388551

180

181

182

183

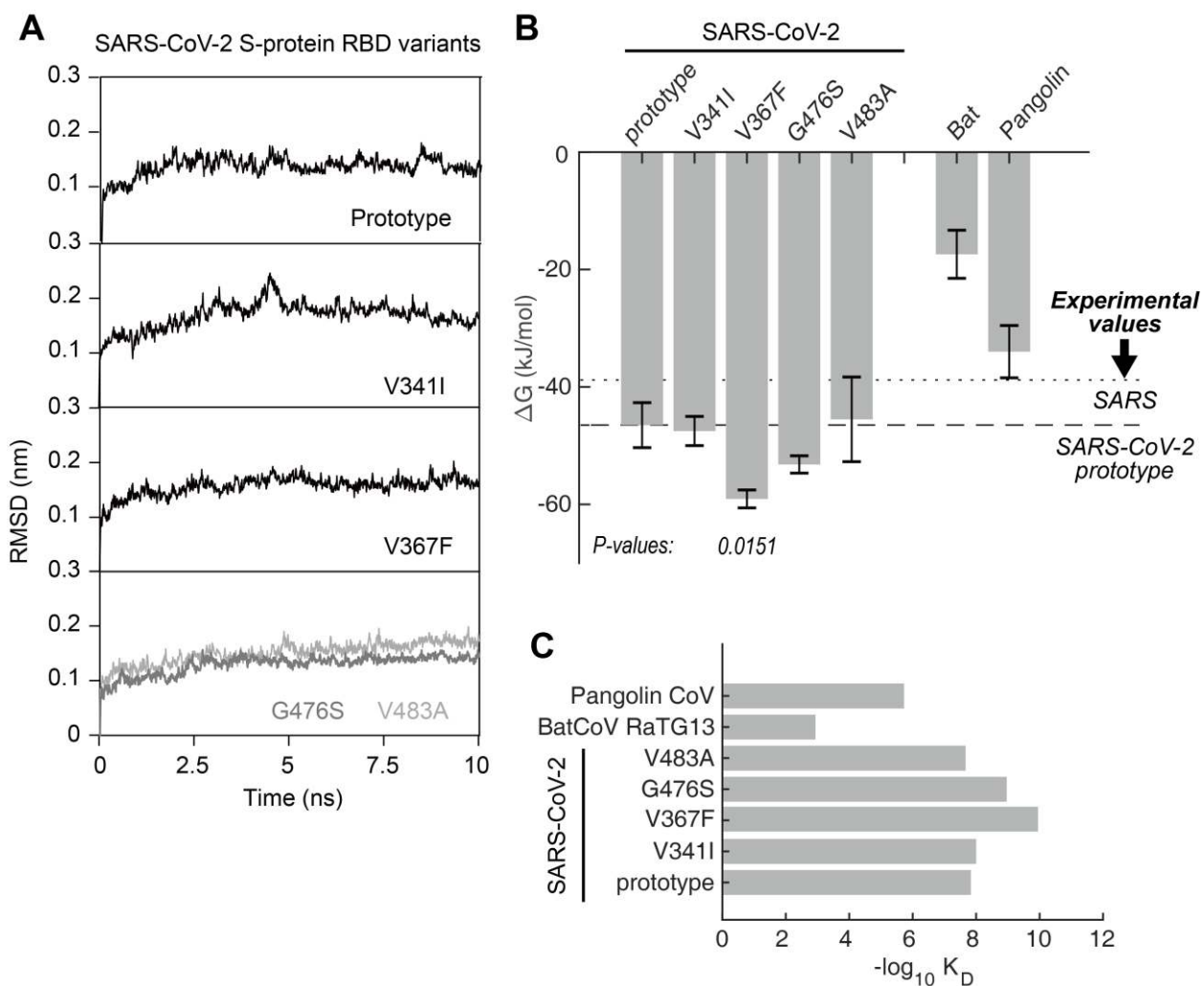
184 **2.3 V367F mutant emergent during the early transmission phase binds human ACE2 receptor** 185 **with higher affinity**

186

187 To estimate the functional changes suggested by the RBD mutations, we performed molecular
188 dynamics (MD) simulations for the prototype SARS-CoV-2 (Wuhan-Hu-1 strain) and the RBD
189 mutants in order to assess their binding energy to human ACE2 receptor. Each simulation was
190 performed at 10ns and each model was simulated in triplicate. All trajectories reached a plateau of
191 RMSD after 2~5ns (**Figure 3A**), indicating that their structures reached equilibrium. All of the
192 subsequent computations on thermodynamics were based on the 5~10ns trajectories. The ΔG of
193 V367F mutant was significantly low (~ 60 kJ/mol) ($P=0.0151$), approximately 25% lower than the
194 prototype strain (-46.5 kJ/mol, calculated from the experimentally measured KD) (**Figure 3B**),
195 suggesting a significantly increased affinity to human ACE2; the other mutants showed a similar
196 ΔG compared to the prototype (**Figure 3B**). Compared to the KD (14.7 nM) of the prototype RBD,
197 the equilibrium dissociation constant (KD) of V367F mutant was calculated as 0.11 nM (**Figure 3C**),
198 which was two orders of magnitude lower than the prototype strain, indicating a remarkably
199 increased affinity to the human ACE2 receptor. In comparison, the bat CoV RBD (strain RaTG13,
200 with the highest genome similarity) showed a much lower binding affinity (KD=1.17mM;
201 $\Delta G=-17.4$ kJ/mol) to human ACE2 than the pangolin CoV (KD=1.89 μ M; $\Delta G=-33.9$ kJ/mol). The
202 affinity of the pangolin CoV to human ACE2 was lower than the SARS-CoV-2 prototype strain
203 (KD=14.7nM; $\Delta G=-46.5$ kJ/mol) (**Figure 3B, 3C**).

204

205 **Fig. 3: Binding free energy calculated for the SARS-CoV-2 S-RBD to human ACE2 receptor.**
 206 (A) RMSD of typical MD trajectories of the SARS-CoV-2 prototype and the mutants. (B)
 207 Comparison of the binding free energy (ΔG) of the RBDs of human SARS-CoV-2, Bat CoV, and
 208 pangolin CoV, with the human ACE2. Note, the ΔG is inversely proportional to the binding affinity.
 209 Data are presented as mean \pm SD. *P*-values were calculated using single-tailed student t-test. The
 210 *P*-values are shown for those with *P* < 0.05. The ΔG calculated from experimental K_D values of
 211 SARS and SARS-CoV-2 prototype are marked in dotted and dashed lines, respectively. (C)
 212 Comparison of the equilibrium dissociation constants (K_D) as calculated with the ΔG .



213

214

215

216

217

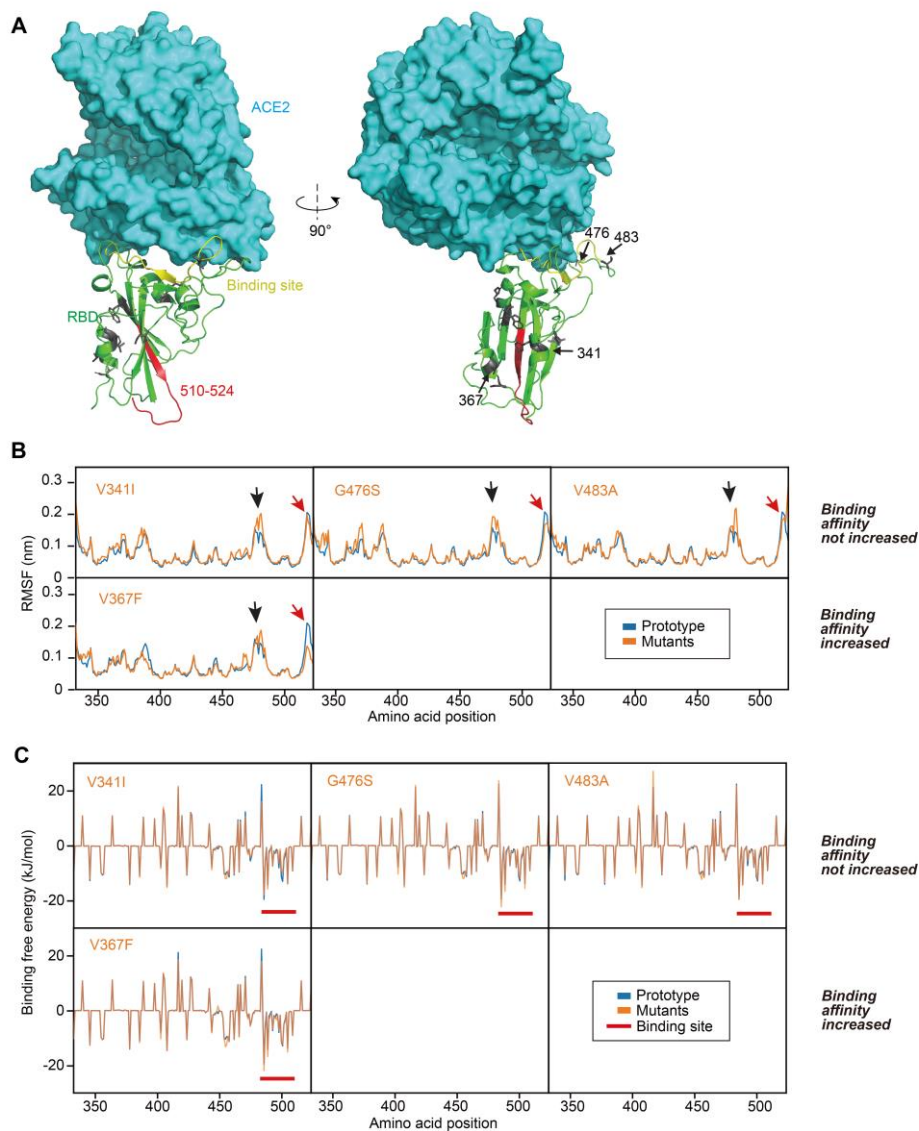
218 **2.4 Structural basis for the increased affinity of the SARS-CoV RBD mutants**

219

220 To elucidate the structural basis of the increased affinity of V367F mutant, we investigated the
221 dynamics of the residues in these structures in greater detail. The binding surface of the RBD to
222 ACE2 is largely arrayed in random coil conformation, which lacks structural rigidity. Logically, a
223 firm scaffold should be necessary to maintain this conformation of the interaction surface, and thus
224 may facilitate the binding affinity. The beta-sheet structure scaffold, centered by residues 510-524
225 (Figure 4A, marked as red), apparently provides this rigidity. “Higher affinity” mutant V367F
226 showed a considerable decrease of the RMSF (Root Mean Square of Fluctuation) at this region,
227 demonstrating a more rigid structure; this was not observed for other mutants (Figure 4B).
228 Coincidentally, the substitutions that accounts for the affinity increase of V367F are all located near
229 this fragment. Indeed, residues 475-485, which is a random coil near the binding site, showed a
230 remarkably higher RMSF for the “similar affinity” mutants, in contrast to the “higher affinity”
231 V367F mutant (Figure 4B). Moreover, the “higher affinity” mutant V367F exhibited a general
232 decreased ΔG in the binding site region in contrast to the “similar affinity” mutants (Figure 4C).

233

234 **Fig. 4: Structural analysis of RBD mutants and the effects on their binding affinity.** (A) The
235 binding surface and interaction of the RBD to ACE2, with the locations of the mutant amino acids
236 noted. Beta-sheet structure scaffold was centered by residues 510-524 (in red). (B) Root Mean
237 Square of Fluctuation (RMSF) of the mutants were compared to that of the prototype. Red arrows
238 denote the fragment of residues 510-524. Black arrows denote the fragment of residues 475-485. (C)
239 Contribution of each amino acid to the binding free energy. Red bars denote the binding site.



240

241

242

243

244

245 **2.5 Experimental validation of the enhanced affinity and infectivity of the V367F mutant**

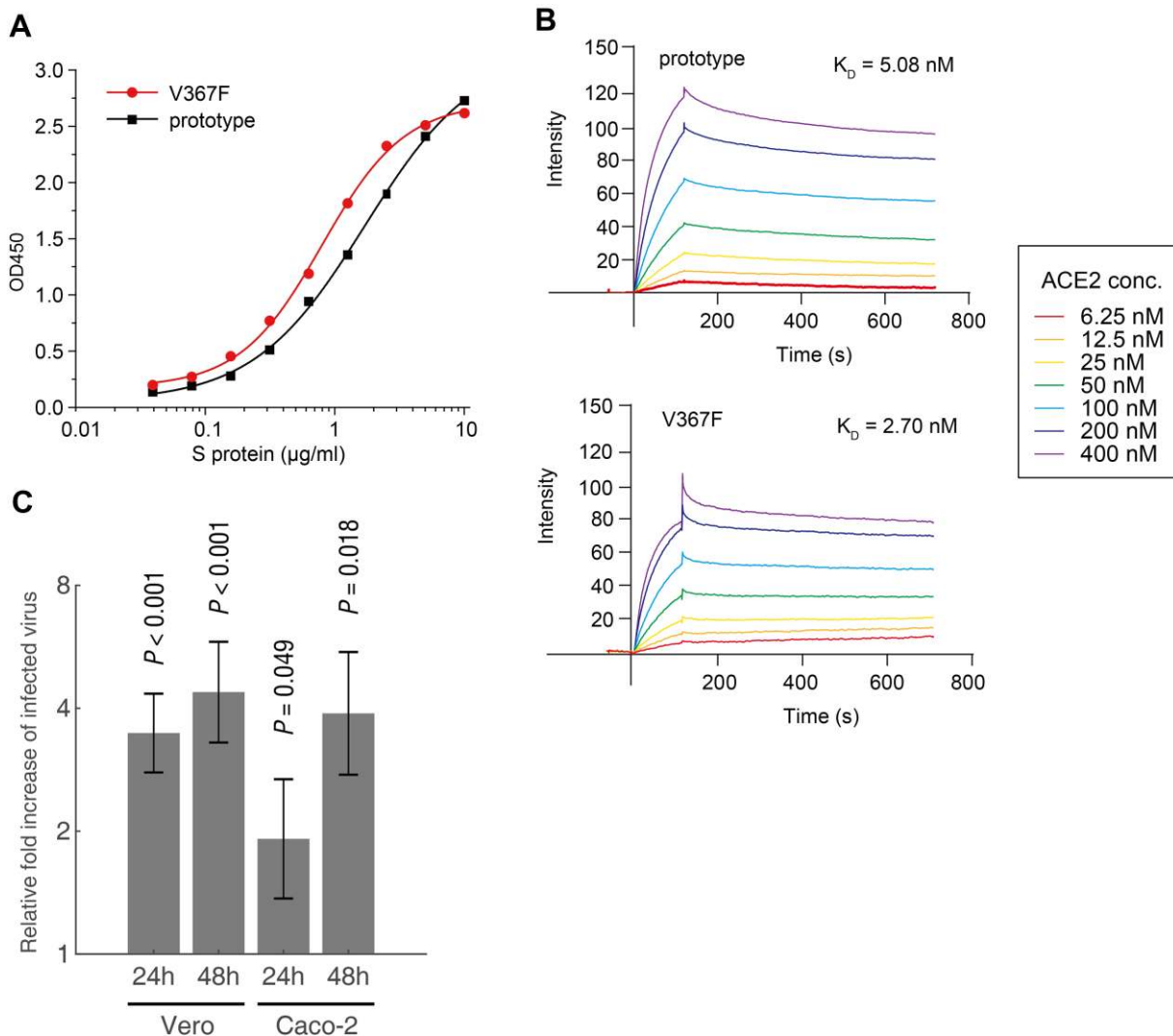
246

247 As of March 31, 2020, among all of the mutants with dominant mutations in spike RBD (>10
248 isolates), most mutations resulted in a substitution of amino acids with similar properties. V367F is
249 the only mutant with higher binding affinity, as calculated by MD simulation (**Figure 3**).
250 Therefore, the binding affinity and the infectivity of V367F mutant were further validated
251 experimentally. First, we performed experiments to assess the binding affinity in vitro with a
252 receptor-ligand binding ELISA assay using purified S proteins and human ACE2 protein. The result
253 showed that the V367F mutation significantly lowered the ED₅₀ concentration (ED₅₀ = 0.8±0.04
254 µg/ml), as compared to the prototype (ED₅₀ = 1.7±0.14 µg/ml) (**Figure 5A**). This demonstrates that
255 the V367F mutant has higher affinity with human ACE2 than the prototype. Second, we performed
256 surface plasmon resonance (SPR) experiments, which yielded the same conclusion: the prototype
257 had a KD of 5.08 nM, compared to the V367F mutant with a KD of 2.70 nM (**Figure 5B**).
258 Additionally, we performed a virus-cell interaction experiment to investigate the invasion efficiency
259 of S proteins using a HIV-backbone pseudovirus assay. Same amounts of S protein-containing
260 pseudovirus were subjected to infection of ACE2-overexpressed Vero and Caco-2 cells. A higher
261 infection efficiency is represented by the increased copy number of the integrated lentivirus genome.
262 At 24 hours post-infection (h.p.i.), the V367F mutant pseudovirus showed 6.08x higher copy
263 numbers than the prototype in Caco-2 cells (P<0.01). At 48 h.p.i, the V367F mutant pseudovirus
264 showed 4.38x and 3.88x higher copy number than the prototype in Vero and Caco-2 cells,
265 respectively (P<0.01) (**Figure 5C**). Therefore, the computation, protein and cell validations were
266 consistent with each other: the V367F mutant had enhanced affinity and infectivity.

267

268 **Figure 5. Experimental validation of the enhanced affinity and infectivity of the V367F**
269 **mutant.**

270 (A) Comparison of the binding affinity of prototype S protein and V367F mutant to human ACE2
271 receptor by ligand-receptor binding ELISA assay. (B) Comparison of the binding affinity of
272 prototype S protein and V367F mutant to human ACE2 protein by SPR. (C) Quantification of the
273 genome copy number of the V367F mutant vs. the prototype using pseudovirus infection assay. The
274 relative fold increases of viruses infecting the cells are shown by the pseudoviral DNA copy number
275 of V367F mutant in both Vero and Caco-2 cells at 24h and 48h post-infection. Experiments were
276 performed in triplicates, and the P-values were calculated using two-tailed t-test for two samples
277 with different variances.



278

279 **2.6 The convergence of SARS-CoV-2 RBD and D614G mutations in dominant mutation** 280 **isolates**

281

282 The D614G mutation is located in the S1 region and is outside the RBD of the SARS-CoV-2 spike
283 protein. It has been confirmed that the D614G mutant increases virus infectivity by elevating its
284 sensitivity to protease, which mutant has spread widely (23). Among all of the dominant RBD
285 mutations (>10 isolates), V367F, A344S and G476S mutations were detected along with the D614
286 prototype during the early transmission phase. Later, the D614G mutant emerged along with these
287 RBD mutants (**Table 1**).

288

289 The V367F mutants were initially discovered in January in Hong Kong. Afterwards, the V367F
290 mutants emerged mainly in Europe, including United Kingdom, the Netherlands, Austria, and
291 Iceland, as well as in the USA, Australia, and China. The D614G+V367F dual mutant was initially
292 emergent in March in the Netherlands (**Supplementary Table 1**).

293

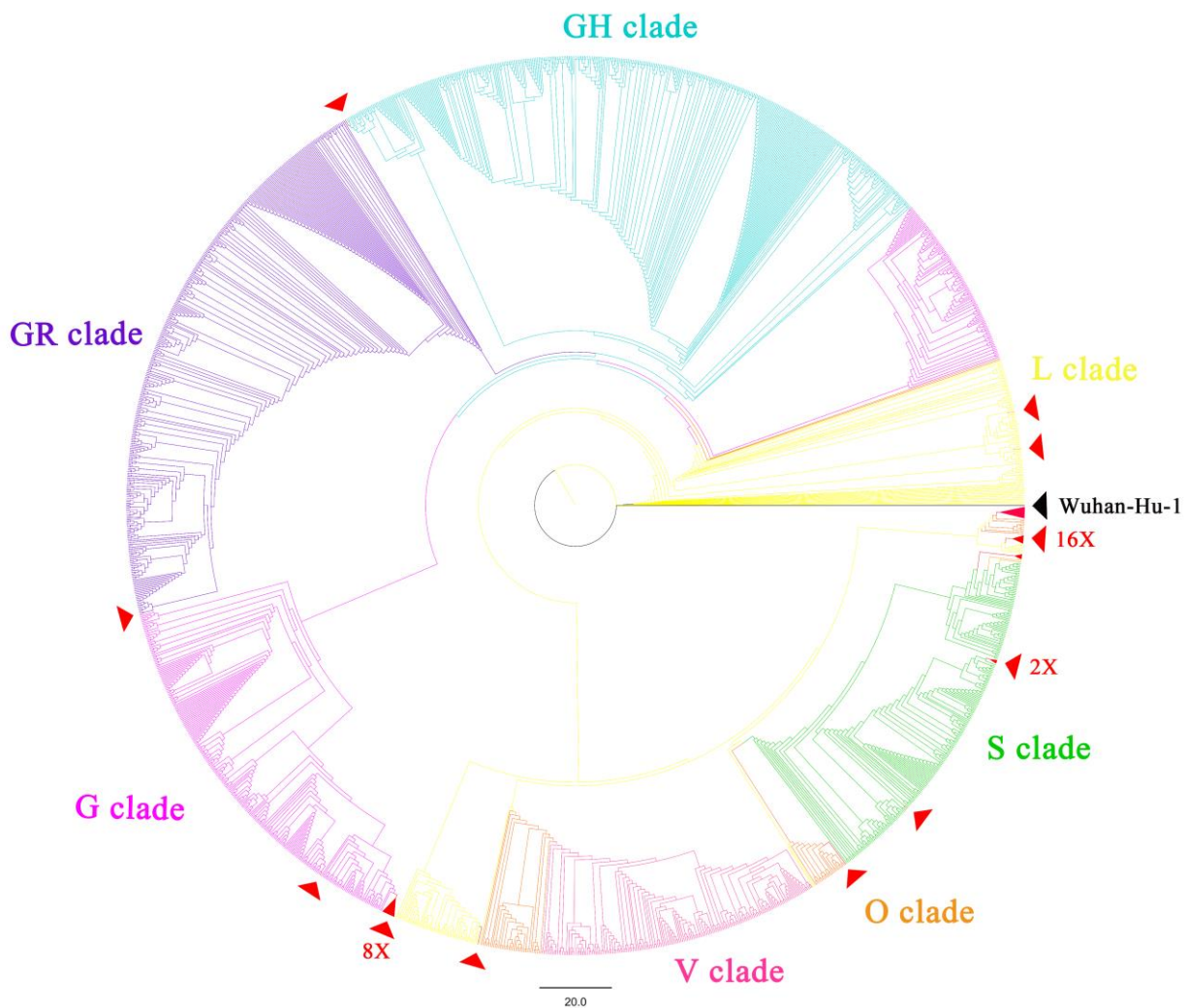
294 The phylogenetic analysis of the V367F mutant genomes during the early transmission phase
295 showed that V367F mutants clustered more closely with the SARS-CoV-2 prototype strain in L
296 clade. Intriguingly, all of the dual-mutation variants (V367F + D614G) emerging later formed a
297 distinct sub-cluster in GH, GR, G clades, separate from the prototype sub-cluster (**Figure 6**)(27). This
298 indicates that the V367F mutation may have evolved along with D614G mutation, suggesting a
299 synergistic effect of increased infectivity. Compared with all the single mutation (V367F) variants,
300 dual mutation (V367F + D614G) variants were emergent in both GH and GR clades. Multiple dual
301 mutation (V367F + D614G) variants located in different clades had mutations which were not
302 detected in early V367F mutants. This indicates that the dual mutations may have evolved through
303 several individual evolution events (**Figure 6**) (23, 28).

304

305

306 **Figure 6. The whole genome phylogenetic analysis of the SARS-CoV-2 variants and mutants**
307 (December 2019 through March 2020). The whole genome phylogenetic tree was constructed by IQ
308 tree 2.02 using the maximum likelihood method with GTR+F+R3 model and 1000 bootstrap
309 replicates, and applying default parameters. All 34 V367F mutants were included. For reference, the
310 branch of Wuhan-Hu-01 is marked in black. V367F mutants are marked in red; sampled referenced
311 sequences are annotated in different colors by clades using Figtree 1.4.4
312 (<https://github.com/rambaut/figtree/releases>).

313



314

315

316

317

318 **3. Discussion**

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

329

330 First, our analysis of molecular dynamics simulation indicated a remarkable enhancement of the
331 affinity efficiency of multiple mutant S proteins. Compared to the prototype strain Wuhan-Hu-1, the
332 ΔG of mutants decreased $\sim 25\%$. These mutants bind to ACE2 more stably due to the enhancement of
333 the base rigidity of the S protein structure. Potential and recent animal-to-human transmission events
334 of SARS-CoV-2 may explain the strong positive selection and enhancement of the affinity during the
335 pandemic. Ongoing adaptation to transmission and replication in humans, including mutation events
336 in the RBD may boost the binding affinity and lead to an increase in the basic reproduction number
337 (R_0), and in theory, further enable human to human transmission.

338

339 The origin of this virus has been of considerable interest and speculation since the outbreak. Due to
340 the high sequence similarity of the bat SARS-like CoV genome and the pangolin CoV RBD sequence
341 to the SARS-CoV-2 sequences, these hosts were thought to have initiated the infection in humans(5, 6,
342 30) Our results provide more clues to support this hypothesis. Our results suggest that the binding
343 energy of the bat SARS-like CoV RBD is too high to bind human ACE2 effectively (K_D in
344 millimolar range). In contrast, the pangolin CoV showed a binding K_D for human ACE2 at the
345 micromolar range, just $\sim 6x$ higher than that of human SARS virus ($K_D = 0.326\mu M$)(Figure. 3), which

346 indicates that the pangolin CoV has the potential to infect humans in unprotected close contact.
347 Alignments of the genomic sequences of SARS-CoV-2 and pangolin CoV viruses suggest
348 recombination events, particularly in the RBD domain, between pangolin and bat viruses (28).

349

350 The V367F mutants were found during the early transmission phase. As RBD is conserved in
351 SARS-CoV-2, the coincidence of V367F mutants across large geographic distances indicates that this
352 mutation is more robust and that these variants originated as a novel sub-lineage, given the close
353 isolation dates (January 22 and 23, respectively). An alternate view is that asymptomatic individuals
354 with the same mutation were “super-infecting” travelers. Along with the epidemiological data,
355 mutation surveillance is of critical importance as it can reveal more exact transmission routes of the
356 epidemic and provide early warnings of additional outbreaks. Emergence of SARS-CoV-2 mutants in
357 Hong Kong, France, and other countries with RBD mutations allowing higher binding affinity to
358 human ACE2 receptor suggests an increased risk of emergence of other mutants with “high-affinity”
359 and increased infectivity during a sustained pandemic of COVID-19, particularly if effective
360 precautions are not completely implemented. By performing assays comparing the prototype spike
361 protein to the V367F mutation containing counterpart, we confirmed the significantly enhanced
362 binding affinity and likely higher infectivity of the V367F mutant, and showed that the mutation
363 stabilizes the RBD structure.

364

365 By tracking mutation types in the SARS-CoV-2 spike RBD, most of the dominant RBD mutants were
366 first detected within the prototype D614 strain, and then together with the G614 variants. This is
367 possibly due to multiple individual recombination events between the D614 mutants and G614
368 mutants, although it is difficult to determine the exact position and time point due to the high
369 sequence identities among SARS-CoV-2 mutants. D614G is distinct from the RBD mutations – it is
370 not located in the RBD but enhances viral infectivity by elevating its sensitivity to protease and
371 increasing stability (12, 13, 23). Furthermore, D614G and V367F may function independently and
372 have synergistic effects on viral infectivity. Recombination is known to play an important role in
373 natural coronavirus evolution, which may contribute to the convergence of dual enhancing mutants

374 (D614G + V367F). The phylogenetic and substitution analyses indicated that the divergent clade of
375 dual mutation mutants were possibly from multiple individual recombination events which
376 introduced new mutations. Whether these new mutations may produce more aggressive viruses
377 remains further research. More attention should be paid to the risk of the advantage accumulation
378 evolution through recombination among the variants. In this case the recombinants may be more
379 infectious and also better at immune escape.

380

381 The S protein is also important for antigen recognition. By tracking dominant RBD mutants up to
382 March 31, 2020, multiple mutations with more than 10 isolates putatively related to host receptor
383 binding and affinity have been detected in this study. The equivalent positions in more than 500
384 SARS and MERS genomes were compared. Among them, V483A in the MERS-CoV and N439K in
385 SARS-CoV resulted in weakly reduced host receptor binding and altered antigenicity, revealing
386 possible immune escape driving virus evolution (32). Given the alanine shares the similar chemical
387 and structural properties with serine, A344S variant is expected to have similar affinity to human
388 ACE2 with the prototype strain. However, since the RBD contains important antigenic epitopes, RBD
389 mutations, especially those that change the amino acid properties, may weaken the binding affinity of
390 any antibody raised against the prototype strain. This may lead to decreased vaccine efficacy and
391 should be further studied.

392

393 In summary, we have found 302 RBD mutants clustering into 96 dominant mutant types during the
394 early transmission phase. Among the six dominant mutation types, V367F mutants that emerged in
395 Asia and Europe displayed enhanced structural stability of the spike protein along with higher
396 binding affinities to the human ACE2 receptor. V367F mutants have been continuously circulating
397 along with the D614G mutants. This indicates that the V367F mutants are stable and may have
398 acquired increased infectivity for humans during the COVID-19 pandemic. The emergence of dual
399 mutants (V367F+D614G) possibly due to the recombination may be a hint of emergence of other
400 variants with increased infectivity or with enhanced escape from the host immune response. These
401 findings support the continuing surveillance of spike mutations to aid in the development of new

402 COVID-19 drugs and vaccines.

403

404 **4. Methods and materials**

405 **4.1 Genome sequence dataset in this study**

406 Full-length gene sequences of SARS-CoV-2 were downloaded from the NCBI GenBank Database,
407 China 2019 Novel Coronavirus Resource (<https://bigd.big.ac.cn/ncov>), and GISAID EpiFluTM
408 Database (<http://www.GISAID.org>). 34,702 SARS-CoV-2 full-genome sequences isolated during
409 early transmission phase (sample collected before March 31, 2020) were downloaded and the
410 sequences with amino acid mutations in S protein and RBD region were parsed. The genome
411 sequences with either the V367F mutation or the V367F/D614G dual mutations in the S protein
412 RBD were screened and analyzed in this study (**Supplementary Table1**). For evolution analysis,
413 1,753 full genome sequences from each Nextstrain (27) clade were filtered and cluster random
414 sampled from GISAID EpiFluTM Database.

415

416 **4.2 Mutation analyses and phylogenetic analyses**

417 Alignment of S protein sequences from different sources and comparison of ACE2 proteins among
418 different species were performed using MAFFT version 7, with default parameters
419 (<https://mafft.cbrc.jp/alignmloadnt/server/>) and BioEdit (33, 34). Selection pressure analyses
420 were conducted using the Nei-Gojobori method (jukes-cantor model) with Mega X (version
421 10.0.2)(20, 21) . Substitution mutation analyses of mutants are performed compared with
422 Wuhan-Hu-01 sequence using BioAider (version 1.314)(17). Phylogenetic analyses were conducted
423 with Iqtree2, applying the maximum-likelihood method with 1000 bootstrap replicates using
424 GTR+F+R3 model by MFP ModelFinder (35). The trees were annotated and modified using Figtree
425 (version 1.4.4) (<https://github.com/rambaut/figtree/releases>).

426

427 **4.3 Molecular dynamics (MD) simulation**

428 The complex structure of the SARS-CoV-2 S-protein RBD domain and human ACE2 was obtained
429 from the Nation Microbiology Data Center (ID: NMDCS0000001) (PDB ID: 6LZG)

430 (<https://www.rcsb.org/structure/6LZG>)(25). Mutated amino acids of the SARS-CoV-2 RBD mutants
431 were directly replaced in the model, and the bat/pangolin CoV RBD domain was modelled using
432 SWISS-MODEL(26). Molecular dynamics simulation was performed using GROMACS 2019 with
433 the following options and parameters: explicit solvent model, system temperature 37°C, OPLS/AA
434 all-atoms force field, and LINCS restraints. With 2fs steps, each simulation was performed at 10ns,
435 and each model was simulated three times to generate three independent trajectory replications.
436 Binding free energy (ΔG) was calculated using MM-PBSA method (GitHub;
437 <https://github.com/Jerkwin/gmxtool>), with the trajectories after structural equilibrium assessed
438 using RMSD (Root Mean Square Deviation)(24,36). To calculate the equilibrium dissociation
439 constant (KD) and ΔG , the formula $\Delta G=RT\ln K_D$ was used. Estimated ΔG s of the RBD mutants
440 were normalized using the ΔG of the prototype strain which was derived from experimental data.

441

442 **4.4 Expression of recombinant S protein mutant**

443 The SARS-CoV-2 prototype S gene was cloned into pNPM5 vector (Novoprotein, NJ, USA), and
444 fused with a C-terminal His6-tag. The V367F mutation was introduced using site-directed
445 mutagenesis, consistent with the nucleotide sequence of the actual isolate. These two constructs
446 were transfected into HEK293 cells using polyethyleneimine. Since the S protein included a signal
447 peptide in its N-terminal 14 amino acids, it was secreted into the medium. The expressed proteins
448 were purified from filtered cell supernatants with a Ni-NTA column. Eluted protein solution was
449 then dialyzed against PBS (pH7.4) for subsequent assays.

450

451 **4.5 Ligand-receptor binding ELISA assay**

452 Human ACE2 protein was immobilized onto a microtiter plate at 5 $\mu\text{g/ml}$ (100 $\mu\text{l/well}$). Each S
453 protein (prototype and V367F) was added as a ligand at different concentrations, ranging from 0.03
454 $\mu\text{g/ml}$ to 10 $\mu\text{g/ml}$, and then incubated for 2 hours at 37°C to allow receptor-ligand interaction. The
455 ligand / receptor mixture was then washed three times. 100 μl of HRP anti-His Tag Antibody
456 (BioLegend, USA) (diluted 1:20000) was added to each well, and allowed to react for 1 hour. After
457 three washes, the signal was visualized using TMB solution (Sigma-Aldrich, USA) and a microtiter

458 plate reader recording OD450.

459

460 **4.6 Surface Plasmon Resonance (SPR) experiments**

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

471

472 **4.7 Production and titration of SARS-CoV-2 pseudoviruses bearing V367F S protein**

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

478

479 Generation of SARS-CoV-2 S HIV-backbone pseudovirus was done as previously described with
480 some modifications(36, 37). Briefly, 293T cells, at about 70-90% confluence, were co-transfected
481 with 9 ug of the transfer vector (pLv-CMV), 6 ug packaging plasmid (psPAX-lentiviral), and 6 ug
482 envelope plasmid (pCB-spike or pCB-spikeV367F). Pseudoviruses were harvested at 48 h
483 post-transfection, using a 2.5 ml sterile syringe and subsequently filtered into either Falcon or
484 microcentrifuge tubes via a syringe driven 0.45 µm filter. Virus titration was measured by RT-qPCR
485 targeting the WPRE gene of pseudoviruses and using the Hifair® III One Step RT-qPCR SYBR

486 Green Kit (Yeasen), as per the manufacturer's protocol. After reverse transcription (10 min at 42°C)
487 and initial denaturation (5 min at 95°C), PCR amplification was performed for 40 cycles (15 s at
488 95°C, 60 s at 60°C). Primers were as follows: WPRE-F, 5'- CACCACCTGTCAGCTCCTTT-3';
489 WPRE -R, 5'- ACGGAATTGTCAGTGCCCAA-3'.

490

491 **4.8 SARS-CoV-2 spike-mediated pseudovirus entry assay.**

492 To detect S variants mediated viral entry, VERO E6 and Caco2 cells (5×10^4) grown in 24-well
493 plates were respectively infected with either 20 MOI of S-V367 or S-F367-bearing pseudovirus
494 (1×10^7 pseudovirus copies per ml). The cell medium was replaced by 500 μ l fresh DMEM medium
495 6h post-infection. Relative fold increase of infected virus titers was calculated according to the
496 WPRE DNA copy number of the lentiviral proviruses measured by TB Green® Premix Ex Taq™ II
497 (Takara) at 24 and 48 h post-infection. Data from all of the samples were obtained from three
498 independent experiments, and each sample was tested in triplicate.

499

500 **Acknowledgments**

501 We gratefully acknowledge the authors, originating and submitting laboratories of the sequences
502 from GISAID's EpiFlu™ Database on which this research is based. All submitters of data may be
503 contacted directly via www.gisaid.org.

504

505 Data and acknowledgement of GISAID sequences are available in [Supplementary Table 2](#).

506

507 **Funding statement:**

508 This work was supported by grants from the National Key Research and Development Program of
509 China (2017YFA0505001/2018YFC0910200/2018YFE0204503), the National Natural Science
510 Foundation of China (81730061), the Guangdong Key Research and Development Program
511 (2019B020226001), the Natural Science Foundation of Guangdong Province (2018B030312010),
512 and the Guangzhou Healthcare Collaborative Innovation Major Project (201803040004 and
513 201803040007).

514

515 **Conflict of interest**

516 The authors declare that they have no conflicts of interest.

517

518

519

520

521

522

523

524

525

526

527

528 **Reference**

- 529 1. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KSM, Lau EHY, Wong JY,
530 Xing X, Xiang N, Wu Y, Li C, Chen Q, Li D, Liu T, Zhao J, Liu M, Tu W, Chen C, Jin L,
531 Yang R, Wang Q, Zhou S, Wang R, Liu H, Luo Y, Liu Y, Shao G, Li H, Tao Z, Yang Y,
532 Deng Z, Liu B, Ma Z, Zhang Y, Shi G, Lam TTY, Wu JT, Gao GF, Cowling BJ, Yang B,
533 Leung GM, Feng Z. 2020. Early Transmission Dynamics in Wuhan, China, of Novel
534 Coronavirus–Infected Pneumonia. *N Engl J Med* 1–9.
- 535 2. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan
536 ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ. 2020. A
537 new coronavirus associated with human respiratory disease in China. *Nature* 579:265–269.
- 538 3. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P,
539 Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W. 2020. A Novel Coronavirus from
540 Patients with Pneumonia in China, 2019. *N Engl J Med* 727–733.
- 541 4. WHO. 2020. COVID-19 Weekly Epidemiological Update
542 1;4.([https://www.who.int/publications/m/item/weekly-epidemiological-update---24-novemb](https://www.who.int/publications/m/item/weekly-epidemiological-update---24-november-2020)
543 [er-2020](https://www.who.int/publications/m/item/weekly-epidemiological-update---24-november-2020))
- 544 5. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, Si H-R, Zhu Y, Li B, Huang C-L,
545 Chen H-D, Chen J, Luo Y, Guo H, Jiang R-D, Liu M-Q, Chen Y, Shen X-R, Wang X, Zheng
546 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.
547 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin.
548 *Nature*.
- 549 6. Xiao K, Zhai J, Feng Y, Zhou N, Zhang X, Zou JJ, Li N, Guo Y, Li X, Shen X, Zhang Z, Shu
550 F, Huang W, Li Y, Zhang Z, Chen RA, Wu YJ, Peng SM, Huang M, Xie WJ, Cai QH, Hou
551 FH, Chen W, Xiao L, Shen Y. 2020. Isolation of SARS-CoV-2-related coronavirus from
552 Malayan pangolins. *Nature* 583:286–289.
- 553 7. Mathewson AC, Bishop A, Yao Y, Kemp F, Ren J, Chen H, Xu X, Berkhout B, van der
554 Hoek L, Jones IM. 2008. Interaction of severe acute respiratory syndrome-coronavirus and
555 NL63 coronavirus spike proteins with angiotensin converting enzyme-2. *J Gen Virol*

- 556 89:2741–2745.
- 557 8. Letko M, Marzi A, Munster V. 2020. Functional assessment of cell entry and receptor usage
558 for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol* 5.
- 559 9. Chen Y, Guo Y, Pan Y, Zhao ZJ. 2020. Structure analysis of the receptor binding of
560 2019-nCoV. *Biochem Biophys Res Commun* 2:0–5.
- 561 10. D.Wrapp, N.Wang, K.S.Corbett, J.A.Goldsmith, C.L.Hsieh, O.Abiona, B.S.Graham JSM.
562 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*
563 (80-) 2507:1–9.
- 564 11. Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, Zhao C, Zhang Q, Liu H, Nie L, Qin H, Wang M,
565 Lu Q, Li X, Sun Q, Liu J, Zhang L, Li X, Huang W, Wang Y. 2020. The Impact of Mutations
566 in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell* 1–11.
- 567 12. Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, Zhang X, Muruato AE, Zou J,
568 Fontes-Garfias CR, Mirchandani D, Scharton D, Bilello JP, Ku Z, An Z, Kalveram B,
569 Freiberg AN, Menachery VD, Xie X, Plante KS, Weaver SC, Shi PY. 2020. Spike mutation
570 D614G alters SARS-CoV-2 fitness. *Nature*.
- 571 13. Yurkovetskiy L, Wang X, Pascal KE, Tomkins-Tinch C, Nyalile TP, Wang Y, Baum A,
572 Diehl WE, Dauphin A, Carbone C, Veinotte K, Egri SB, Schaffner SF, Lemieux JE, Munro
573 JB, Rafique A, Barve A, Sabeti PC, Kyratsous CA, Dudkina N V., Shen K, Luban J. 2020.
574 Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell*
575 183:739-751.e8.
- 576 14. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, Zhang Q, Shi X, Wang Q, Zhang L, Wang X.
577 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2
578 receptor. *Nature* 581:215–220.
- 579 15. Glycoprotein C-S, Walls AC, Park Y, Tortorici MA, Wall A, McGuire AT, Veessler D, Walls
580 AC, Park Y, Tortorici MA, Wall A, McGuire AT. 2020. Structure , Function , and
581 Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 181:281-292.e6.
- 582 16. Ahmed SF, Quadeer AA, McKay MR. 2020. Preliminary identification of potential vaccine
583 targets for 2019-nCoV based on SARS-CoV immunological studies. *Viruses*

- 584 2020.02.03.933226.
- 585 17. Zhou ZJ, Qiu Y, Pu Y, Huang X, Ge XY. 2020. BioAider: An efficient tool for viral genome
586 analysis and its application in tracing SARS-CoV-2 transmission. *Sustain Cities Soc*
587 63:102466.
- 588 18. Sanders JM, Wampole ME, Thakur ML, Wickstrom E. 2013. Molecular Determinants of
589 Epidermal Growth Factor Binding : A Molecular Dynamics Study 8:8–10.
- 590 19. Rifai EA, Van Dijk M, Vermeulen NPE, Yanuar A, Geerke DP. 2019. A Comparative Linear
591 Interaction Energy and MM/PBSA Study on SIRT1-Ligand Binding Free Energy Calculation.
592 *J Chem Inf Model* 59:4018–4033.
- 593 20. Nei M, Gojoborit T. 1986. Simple methods for estimating the numbers of synonymous and
594 nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418–426.
- 595 21. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular evolutionary
596 genetics analysis across computing platforms. *Mol Biol Evol* 35:1547–1549.
- 597 22. Lai A, Bergna A, Caucci S, Clementi N, Vicenti I, Dragoni F, Cattelan AM, Menzo S, Pan A,
598 Callegaro A, Tagliabracci A, Caruso A, Caccuri F, Ronchiadin S, Balotta C, Zazzi M,
599 Vaccher E, Clementi M, Galli M, Zehender G. 2020. Molecular tracing of SARS-CoV-2 in
600 Italy in the first three months of the epidemic. *Viruses* 12:1–13.
- 601 23. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N,
602 Giorgi EE, Bhattacharya T, Foley B, Hastie KM, Parker MD, Partridge DG, Evans CM,
603 Freeman TM, de Silva TI, Angyal A, Brown RL, Carrilero L, Green LR, Groves DC,
604 Johnson KJ, Keeley AJ, Lindsey BB, Parsons PJ, Raza M, Rowland-Jones S, Smith N,
605 Tucker RM, Wang D, Wyles MD, McDanal C, Perez LG, Tang H, Moon-Walker A, Whelan
606 SP, LaBranche CC, Sapphire EO, Montefiori DC. 2020. Tracking Changes in SARS-CoV-2
607 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*
608 182:812-827.e19.
- 609 24. Homeyer N, Gohlke H. 2012. Free energy calculations by the Molecular Mechanics
610 Poisson-Boltzmann Surface Area method. *Mol Inform* 31:114–122.
- 611 25. Wang Q, Zhang Y, Wu L, Niu S, Song C, Zhang Z, Lu G, Qiao C, Hu Y, Yuen KY, Wang Q,

- 612 Zhou H, Yan J, Qi J. 2020. Structural and Functional Basis of SARS-CoV-2 Entry by Using
613 Human ACE2. *Cell* 181:894-904.e9.
- 614 26. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, De Beer
615 TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. 2018. SWISS-MODEL: Homology
616 modelling of protein structures and complexes. *Nucleic Acids Res* 46:W296–W303.
- 617 27. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T,
618 Neher RA. 2018. NextStrain: Real-time tracking of pathogen evolution. *Bioinformatics*
619 34:4121–4123.
- 620 28. Li X, Giorgi EE, Marichannelgowda MH, Foley B, Xiao C, Kong XP, Chen Y, Gnanakaran S,
621 Korber B, Gao F. 2020. Emergence of SARS-CoV-2 through recombination and strong
622 purifying selection. *Sci Adv* 6:1–12.
- 623 29. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A,
624 Lanfear R, Teeling E. 2020. IQ-TREE 2: New Models and Efficient Methods for
625 Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* 37:1530–1534.
- 626 30. Li X, Giorgi EE, Marichannelgowda MH, Foley B, Xiao C, Kong X, Chen Y, Gnanakaran S,
627 Korber B, Gao F. 2020. Emergence of SARS-CoV-2 through recombination and strong
628 purifying selection 1–11.
- 629 31. Zost SJ, Gilchuk P, Case JB, Binshtein E, Chen RE, Nkolola JP, Schäfer A, Reidy JX,
630 Trivette A, Nargi RS, Sutton RE, Suryadevara N, Martinez DR, Williamson LE, Chen EC,
631 Jones T, Day S, Myers L, Hassan AO, Kafai NM, Winkler ES, Fox JM, Shrihari S, Mueller
632 BK, Meiler J, Chandrashekar A, Mercado NB, Steinhardt JJ, Ren K, Loo YM, Kallewaard
633 NL, McCune BT, Keeler SP, Holtzman MJ, Barouch DH, Gralinski LE, Baric RS, Thackray
634 LB, Diamond MS, Carnahan RH, Crowe JE. 2020. Potently neutralizing and protective
635 human antibodies against SARS-CoV-2. *Nature* 584:443–449.
- 636 32. Kleine-Weber H, Elzayat T, Wang L, Elzayat MT, Wang L, Graham BS, Muller MA,
637 Drosten C, Pohlmann S, Hoffmann M. 2019. Mutations in the Spike Protein of Middle East
638 Respiratory Syndrome Coronavirus Transmitted in Korea Increase. *J Virol* 93:1–14.
- 639 33. Katoh K, Rozewicki J, Yamada KD. 2018. MAFFT online service: Multiple sequence

- 640 alignment, interactive sequence choice and visualization. *Brief Bioinform* 20:1160–1166.
- 641 34. Kuraku S, Zmasek CM, Nishimura O, Katoh K. 2013. aLeaves facilitates on-demand
642 exploration of metazoan gene family trees on MAFFT sequence alignment server with
643 enhanced interactivity. *Nucleic Acids Res* 41:22–28.
- 644 35. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermin LS. 2017. ModelFinder:
645 fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589.
- 646 36. Zhao G, Du L, Ma C, Li Y, Li L, Poon VK, Wang L, Yu F, Zheng BJ, Jiang S, Zhou Y. 2013.
647 A safe and convenient pseudovirus-based inhibition assay to detect neutralizing antibodies
648 and screen for viral entry inhibitors against the novel human coronavirus MERS-CoV. *Virol*
649 *J* 10:1–8.
- 650 37. Nie J, Li Q, Wu J, Zhao C, Hao H, Liu H, Zhang L, Nie L, Qin H, Wang M, Lu Q, Li X, Sun
651 Q, Liu J, Fan C, Huang W, Xu M, Wang Y. 2020. Establishment and validation of a
652 pseudovirus neutralization assay for SARS-CoV-2. *Emerg Microbes Infect* 9:680–686.
- 653
- 654
- 655
- 656
- 657
- 658
- 659
- 660
- 661
- 662
- 663
- 664
- 665
- 666

1

2 **Supplementary data**

3 **Supplementary Table 1:** Meta data of the isolates with V367F mutations in the RBD of spike glycoprotein (January through March, 2020)

4 **Supplementary Table 2:** Acknowledgement table for GISAID sequences

5

6

7

8

9

10

11

12

13

14

15

16

17

18

1 **Supplementary Figure 1:** Multiple alignments of the RBD amino acid sequences. SARS-CoV-2 Wuhan-Hu-1, the first reported genome, is used as
2 reference. Bat and pangolin SARS-like coronaviruses are also included. Amino acid substitutions are marked. Dots indicate identical amino acids.
3
4