

1 **SARS-CoV-2 originated from SARS-CoV-1-related Bat-CoVs through Pan-CoVs**
2 **rather than from SARS-CoV-2-related Bat-CoVs**

3 Perumal Arumugam Desingu^{1#}, K. Nagarajan²

4 ¹Department of Microbiology and Cell Biology, Indian Institute of Science, Bengaluru, India

5 ²Department of Veterinary Pathology, Madras Veterinary College, Vepery, Chennai, 600007,
6 Tamil Nadu Veterinary and Animal Sciences University (TANUVAS),

7

8

9

10

11 **Running title:** SARS-CoV-2 originated from the Pan-CoVs

12

13

14

15

16

17

18

19

20

21

22

23

24 # Corresponding authors,

25 Perumal Arumugam Desingu, Department of Microbiology and Cell Biology, Division of
26 Biological Sciences, Indian Institute of Science, Bangalore – 560012, India. Email:

27 perumald@iisc.ac.in; padesingu@gmail.com Phone: +91 80 2293 2068, Fax: +91 80 2360
28 2697

29

30 **Abstract**

31 The emergence of the novel SARS-CoV-2 in 2019 sparked a dispute concerning its origin.
32 Here, we report that the SARS-CoV-2 originated through pangolin-coronavirus (Pan-CoVs)
33 from the SARS-CoV-related-bat-coronaviruses (SARS-CoV-1-rB-CoVs) rather than from
34 SARS-CoV-2-related-bat-coronaviruses (SARS-CoV-2-rB-CoVs), in contrast to the previous
35 thought. Further, our analyses strongly suggest that the Pan-CoVs evolved from the SARS-
36 CoV-1-rB-CoVs without recombination. Further, our results suggest that the SARS-CoV-1-
37 rB-CoVs' perhaps jumped into the pangolin, which forced the viruses to mutate and adapt to
38 the new host, and resulted in the origin of Pan-CoVs. Surprisingly, the Pan-CoVs formed an
39 evolutionary intermediate between SARS-CoV-2 and SARS-CoV-2-rB-CoVs at the spike
40 gene. Our findings also suggest that the Pan-CoV/GX and Pan-CoV/Guangdong lineages
41 recombined to form the SARS-CoV-2 spike gene. We also found evidence that the SARS-
42 CoV-2-rB-CoVs spike gene evolved via recombination between Pan-CoV/Guangdong and
43 SARS-CoV-1-rB-CoVs. Overall, our findings suggest that the SARS-CoV-2 emerged from
44 SARS-CoV-1-rB-CoVs through host jumping.

45 **Keywords:** SARS-CoVs, the origin of SARS-CoV-2, evolutionary missing-link spike
46 protein, pangolin coronavirus, bat SARS coronavirus

47 **Introduction**

48 SARS-CoV-2, a readily human adaptable, transmissible, and infective virus, first surfaced in
49 China in December 2019 and quickly spread over the world (1-3), posing significant
50 evolutionary questions about its genesis. The majority of the investigations, on the other
51 hand, focused on the SARS-like CoV-2's genesis from SARS-CoV-2-rB-CoVs (4-7) through
52 Pan-CoVs (7-9) via genetic recombination (8,10-13). Furthermore, SARS-CoV-2 is also the
53 seventh coronavirus to infect humans, and it is closely linked to SARS-CoVs (SARS-CoV-1).
54 Although SARS-CoV-2 and related bat coronaviruses (SARS-CoV-2-rB-CoVs) have been
55 identified (2,4-7), none of them have developed an evolutionary intermediate connecting link
56 between SARS-CoV-1/SARS-CoV-1-rB-CoVs and SARS-CoV-2/SARS-CoV-2-rB-CoVs
57 across their genome (8,10-13). Similarly, when compared to SARS-CoV-2, all the SARS-
58 CoV-2-rB-CoVs, except RaTG13, displayed a lot of genetic heterogeneity in the spike gene
59 (5-7,14), which is critical for defining host range, transmissibility, virus entry, human-human
60 infection, and host-mediated immune responses (15-19). In light of this, the evolution of
61 SARS-CoV-2-rB-CoVs in bats from SARS-CoV-1-rB-CoVs, as well as the origin of SARS-
62 CoV-2 from SARS-CoV-1 is unknown.

63 The SARS-CoV-2 is designated as a new/novel coronavirus due to its genetic diversity in the
64 pp1a, spike protein, ORF3, and protein 8 genes (1,3,4). The CoVs in general, the spike
65 protein influences the host range, virus transmission, attachment, and entry (15-19); pp1a
66 (nsp3, largest gene of pp1a) which determines the viral replication/transcription complex
67 (RTC) and also regulates various other functions such as ssRNA binding, nucleocapsid
68 binding, de-MARylation, dePARylation, ADPr binding, G-quadruples binding, protease, and
69 deISGylation (20-27); ORF3 is a gene that controls virus release via viral ion channels

70 (viroporins) (28,29), and regulate autophagy to favour viral replication (30-32); and ORF8
71 (protein 8) controls antigen presentation through MHC-I and regulates host immune
72 surveillance (33,34). Moreover, in the pp1a, ORF3, and protein 8 areas, the SARS-CoV-2-rB-
73 CoVs shared a close genetic link with the SARS-CoV-2. When compared to SARS-CoV-2,
74 the SARS-CoV-2-rB-CoVs (except RaTG13) revealed a lot of genetic variation in the spike
75 gene. Overall, the origins of the novel pp1a, ORF3, protein 8, and spike genes in SARS-CoV-
76 2/SARS-CoV-2-rB-CoVs are unknown.

77 In this study, we found evidence that the SARS-CoV-2 evolved from SARS-CoV-1/SARS-
78 CoV-1-rB-CoVs through Pan-CoVs as an evolutionary intermediate between SARS-CoV-1/
79 SARS-CoV-1-rB-CoVs and SARS-CoV-2/SARS-CoV-2-rB-CoVs across their genome
80 without recombination events. Our data further suggest that SARS-CoV-2 arose through the
81 recombination of two different Pan-CoVs lineages. We also present evidence that the SARS-
82 CoV-2-rB-CoVs are the result of recombination between Pan-CoVs and SARS-CoV-1-rB-
83 CoVs.

84 **Results**

85 **Pan-CoVs genetic diversity at the complete genome levels**

86 In our phylogenetic, net between-group mean distance (NBGM), and similarity plot-based
87 analysis, the SARS-CoV-2-rB-CoVs of RaTG13, RShSTT200, RShSTT182, RacCS203,
88 RmYN02, and RpYN06 formed a close genetic link with the SARS-CoV-2 (**Figure 1A, 1B**
89 **and 1C**). On the other hand, in our phylogenetic tree analysis, we also observed that the
90 SARS-CoV-2-rB-CoVs of bat/PrC31, bat-SL-CoVZC45, and bat-SL-CoVZXC21 viruses
91 constituted out-group from SARS-CoV-2 (**Figure 1A**). Consistent with these results, our
92 NBGM analysis revealed that these viruses (bat/PrC31, bat-SL-CoVZC45, and bat-SL-
93 CoVZXC21) had displayed around 10% genetic diversity when compared to SARS-CoV-2
94 (**Figure 1B**). Furthermore, in our similarity plot-based analysis, the bat/PrC31, bat-SL-
95 CoVZC45, and bat-SL-CoVZXC21 revealed a great deal of variability in the pp1b gene when
96 compared to SARS-CoV-2 (**Figure 1C**).

97 Next, in our phylogenetic analysis, the Pan-CoVs are divided into two lineages: Pan-
98 CoV/Guangdong and Pan-CoV/GX (**Figure 1A**), and it is also supported by NBGM
99 (**Figure 1B**) and similarity plot-based analysis (**Figure 1D; Supplementary Figure 1A-1B**).
100 Surprisingly, these Pan-CoVs lineages formed an evolutionary bridge between SARS-CoV-
101 1/SARS-CoV-1-rB-CoVs and SARS-CoV-2/SARS-CoV-2-rB-CoVs throughout their
102 genome (**Figure 1D; Supplementary Figure 1A-1B**). The Pan-CoV/Guangdong lineage
103 developed a closer relationship with the SARS-CoV-2 and SARS-CoV-2-rB-CoVs than the
104 Pan-CoV/GX lineage in the phylogenetic tree (**Figure 1A**) and also in NBGM analyses
105 (**Figure 1B**). In our similarity plot-based analysis, the Pan-CoV/Guangdong lineage shared a
106 stronger genetic resemblance with the SARS-CoV-2/SARS-CoV-2-rB-CoVs except for the
107 spike gene S1-N-terminal domain (S1-NTD) (**Figure 1D; Supplementary Figure 1A-1B**).
108 Remarkably in our further analysis, the Pan-CoV/GX showed an evolutionary intermediate
109 between the SARS-CoV-1/SARS-CoV-1-rB-CoVs and SARS-CoV-2/SARS-CoV-2-rB-

110 CoVs throughout their genome (**Figure 1D; Supplementary Figure 1A-1B**). Collectively,
111 these findings suggest that the SARS-CoV-1/SARS-CoV-1-rB-CoVs jumped into pangolins
112 and evolved into Pan-CoV/GX in pangolins as a result of host selection and adaptation
113 pressure. The Pan-CoV/GX in the pangolin further evolved into the Pan-CoV/Guangdong
114 lineage, which is strikingly similar to the SARS-CoV-2/SARS-CoV-2-rB-CoVs due to
115 probable better host adaptation and immune evasion selection pressure.

116 **Evolution of pp1a in SARS-CoV-2 from SARS-CoV-1 through Pan-CoVs**

117 According to earlier studies, the pp1a gene was reported as a novel gene in SARS-CoV-2
118 (1,3,4). To explore the evolutionary genesis of the pp1a gene, we used the SARS-CoV-2
119 nucleotide sequences from the first nucleotide to 11725nt (nucleotide numbering according to
120 NC 045512.2/Wuhan-Hu-1). Because the SARS-CoV-1 viruses displayed a lot of diversity
121 from 1-11725nt and displayed a closer identity from 11726-21561nt with the SARS-CoV-
122 2/SARS-CoV-2-rB-CoVs, therefore we divided the pp1ab into two different regions for the
123 analysis as follows: 1- 11725nt and 11726-21561nt for analysis (**Figure 1D; Supplementary**
124 **Figure 1A-1B**). The SARS-CoV-2-rB-CoVs of bat-SL-CoVZC45 and bat-SL-CoVZXC21
125 have nearly identical sequences to the SARS-CoV-1 from the 11725nt to the end of the pp1ab
126 gene (**Figure 1D; Supplementary Figure 1A-1B**). The SARS-CoV-2-rB-CoVs of the
127 RpYN06, RmYN02, and Bat/PrC31 viruses formed a closer relationship with SARS-CoV-2
128 in this region (1-11725nt) than the RaTG13 virus (**Figure 2A-2C; Supplementary Figure**
129 **2A-2B**). Surprisingly in this location, the Pan-CoVs formed an evolutionary bridge between
130 SARS-CoV-1/SARS-CoV-1-rB-CoVs and SARS-CoV-2/SARS-CoV-2-rB-CoVs (**Figure**
131 **2A-2C; Supplementary Figure 2A-2B**).

132 Next, in comparison to the previous region, the genetic diversity between SARS-CoV-1 and
133 SARS-CoV-2 is lowest in the pp1ab regions of 11726nt to 21561 nucleotides (**Figure 2D-2F;**
134 **Supplementary Figure 3A-3B**). Surprisingly, the bat/PrC31, bat-SL-CoVZC45, and bat-SL-
135 CoVZXC21 SARS-CoV-2-rB-CoVs had nearly identical sequence similarity to the SARS-
136 CoV-1-rB-CoVs of bat-SARS-coronavirus-HKU3 viruses (**Figure 2D-2F; Supplementary**
137 **Figure 3A-3B**). Remarkably, Pan-CoVs also created an evolutionary intermediate between
138 SARS-CoV-1/SARS-CoV-1-rB-CoVs and SARS-CoV-2/SARS-CoV-2-rB-CoVs in this
139 region (**Figure 2D-2F; Supplementary Figure 3A-3B**). Overall, our findings imply that
140 throughout the pp1ab gene, Pan-CoVs formed an evolutionary intermediate between SARS-
141 CoV-1 and SARS-CoV-2. Altogether, our results suggest that the Pan-CoVs gained this pp1a
142 gene from SARS-CoV-1/SARS-CoV-1-rB-CoVs and it might have evolved into a novel pp1a
143 gene in pangolin through new host adaptation and immune evasion mediated evolution.
144 Further, the jumping of Pan-CoVs into humans/other related hosts possibly evolved into the
145 SARS-CoV-2-specific pp1a which is seen in the ongoing outbreak. On the other hand, the
146 SARS-CoV-2-rB-CoVs (bat/PrC31, bat-SL-CoVZC45, and bat-SL-CoVZXC21) potentially
147 gained this pp1ab gene by recombination of Pan-CoVs and the SARS-CoV-1-rB-CoVs of
148 bat-SARS-coronavirus-HKU3 viruses.

149

150 **Origin of novel ORF3 in SARS-CoV-2 from SARS-CoV-1 through Pan-CoVs**

151 We investigated the evolutionary history of the novel ORF3 in SARS-CoV-2 and SARS-
152 CoV-2-rB-CoVs in light of recent studies that suggested they acquired it (1,3,4). In this
153 location, the Pan-CoV/GX lineage formed an evolutionary link between SARS-CoV-1 and
154 SARS-CoV-2, while the Pan-CoV/Guangdong lineage shared a high degree of sequence
155 similarity with SARS-CoV-2 (**Figure 3A-3C; Supplementary Figure 4A-4B**). It was
156 surprising to find that SARS-CoV-2-rB-CoVs of bat/PrC31, bat-SL-CoVZC45, and bat-SL-
157 CoVZXC21 viruses differed considerably from SARS-CoV-2 and other SARS-CoV-2-rB-
158 CoVs (**Figure 3A-3C; Supplementary Figure 4A-4B**). These results indicated that as
159 similar to the pp1a gene, the novel ORF3 also evolved in the SARS-CoV-2 from the SARS-
160 CoV-1 through the Pan-CoVs by the host jump mediated virus adaptation and evolution.

161 **Genetic diversity of the envelope (E), membrane (M), protein 6, protein 7a, and protein** 162 **7b**

163 Next, the envelope (E), membrane (M), protein 6, protein 7a, and protein 7b portions of the
164 Pan-CoV/Guangdong, SARS-CoV-2, and SARS-CoV-2-rB-CoVs had a significant degree of
165 sequence similarity, whereas the Pan-CoV/GX lineage lay between the SARS-CoV-1 and
166 SARS-CoV-2 (**Figure 3D-3F; Supplementary Figure 5A-5B**). These results suggest that
167 the SARS-CoV-2 and other SARS-CoV-2-rB-CoVs gained E, M, protein 6, protein 7a, and
168 protein 7b from the SARS-CoV-1 via the Pan-CoV lineages.

169 **Gain of novel Protein 8 in SARS-CoV-2 from SARS-CoV-1 through Pan-CoVs**

170 Although it is widely known that SARS-CoV-2 and SARS-CoV-2-rB-CoVs have the unique
171 protein 8 gene in their genomes (1,3,4), the evolutionary origin of this gene remains
172 unknown. We observed that similar to ORF3, the SARS-CoV-2-rB-CoVs of bat-SL-
173 CoVZC45 and bat-SL-CoVZXC21 viruses showed significant genetic diversity from SARS-
174 CoV-2 and other SARS-CoV-2-rB-CoVs in the protein-8 region (**Figure 4A-4B**).
175 Remarkably, SARS-CoV-2-rB-CoVs of bat/PrC31 on the other hand demonstrated a stronger
176 genetic identity with SARS-CoV-2 and other SARS-CoV-2-rB-CoVs. More importantly, the
177 Pan-CoV/Guangdong lineage shared a high level of sequence similarity with SARS-CoV-2,
178 whereas the Pan-CoV/GX lineage was an evolutionary bridge between SARS-CoV-1 and
179 SARS-CoV-2 (**Figure 4A-4B**). These results collectively suggest that similar to pp1a and
180 ORF3, the SARS-CoV-2 gained the novel protein 8 from the SARS-CoV-1 through the Pan-
181 CoVs by the host jump mediated virus adaptation and evolution.

182 **Genetic diversity of the nucleocapsid (N), ORF10 and 3'UTR**

183 Next, the nucleocapsid (N), ORF10 and 3'UTR region, the Pan-CoV/Guangdong, SARS-
184 CoV-2, and SARS-CoV-2-rB-CoVs had a closer degree of sequence similarity; however, the
185 Pan-CoV/GX lineage formed evolutionary intermediate between SARS-CoV-1 and SARS-
186 CoV-2 in phylogenetic, NBGMD, and similarity plot-based analysis (**Figure 4C-4E;**
187 **Supplementary Figure 6A-6B**). These results indicate that the SARS-CoV-2 and other

188 SARS-CoV-2-rB-CoVs acquired the nucleocapsid (N), ORF10, and 3'UTR region from the
189 SARS-CoV-1 via the Pan-CoV lineages through the host jump based evolution.

190 **Recombination-based origin of SARS-CoV-2 spike gene from Pan-CoVs**

191 The spike gene which is known to determine the host range, transmission, virus entry, and
192 host immunity (15-19) were reported to display huge genetic diversity between the SARS-
193 CoV-2 and SARS-CoV-2-rB-CoVs (5-7,14). Consistent with this, we are also found huge
194 genetic diversity between the SARS-CoV-2 and SARS-CoV-2-rB-CoVs (**Figure 5A-5B**) in
195 the spike gene. Interestingly in our phylogenetic and similarity plot analysis, the Pan-CoVs
196 displayed as an evolutionary intermediate between SARS-CoV-2 and SARS-CoV-2-rB-CoVs
197 at the entire spike gene level (**Figure 5A&5C**). Fascinatingly, we also observed that among
198 the Pan-CoV lineages, the Pan-CoV/Guangdong lineage had a closer evolutionary link with
199 SARS-CoV-2 (**Figure 5C-5D; Supplementary Figure 7A-7B**). In confirmation with this, at
200 the spike protein S1-Receptor binding domain (S1-RBD) levels, Pan-CoV/Guangdong
201 lineage showed essentially identical sequence similarity to SARS-CoV-2 (**Figure 5C-5D;**
202 **Supplementary Figure 7A-7B**). In contrast, when compared to SARS-CoV-2, the Pan-
203 CoV/Guangdong lineage revealed a lot of genetic variation in the spike protein S1-N-
204 Terminal domains (S1-NTD) (**Figure 5C-5D; Supplementary Figure 7A-7B**). Further, our
205 similarity plot analysis revealed that, in comparison to SARS-CoV-2, the Pan-CoV/GX
206 lineage showed limited variation at the S1-NTD and S1-RBD (**Figure 5C-5D;**
207 **Supplementary Figure 7A-7B**). These findings point to recombination between the Pan-
208 CoV/Guangdong and Pan-CoV/GX lineages as a probable evolutionary source for the SARS-
209 CoV-2 virus. More interestingly, our similarity plot and recombination detection program
210 (RDP) analyses also confirmed this recombination of Pan-CoV/Guangdong and Pan-CoV/GX
211 lineage in the S1-NTD (**Figure 5C-5F; Supplementary Figure 7A-7B**). Collectively, our
212 results strongly suggest that the SARS-CoV-2 spike gene was evolved through recombination
213 of the Pan-CoV/Guangdong and Pan-CoV/GX lineages. Overall, our results strongly suggest
214 that the SARS-CoV-2 gained all the novel genes such as pp1a, spike, ORF3, and protein 8
215 from the SARS-CoV-1 through the Pan-CoVs by the host jump mediated virus adaptation
216 and evolution.

217 **Recombination-based evolution of SARS-CoV-2-rB-CoVs spike gene from Pan-CoVs**

218 Except for RaTG13, all SARS-CoV-2-rB-CoVs had a lot of genetic variabilities when
219 compared to SARS-CoV-2 at the entire spike (S) gene levels (5-7,14). In our phylogenetic
220 analysis, the Pan-CoVs formed an evolutionary intermediate between the SARS-CoV-2 and
221 SARS-CoV-2-rB-CoVs in the spike gene (**Figure 5A**). On the other hand, in NBGM
222 analysis, the SARS-CoV-2-rB-CoVs displayed almost equal genetic diversity with SARS-
223 CoV-2 and SARS-CoV-1/SARS-CoV-1-rB-CoVs (**Figure 5B**). However, the SARS-CoV-2-
224 rB-CoVs of Bat/China/RpYN06, bat/PrC31, bat-SL-CoVZC45, and bat-SL-CoVZXC21
225 displayed a somewhat closer identity with SARS-CoV-2 and Pan-CoVs than SARS-CoV-
226 1/SARS-CoV-1-rB-CoVs (**Figure 5A-5B**). Further, similarity plot analysis displayed the
227 sequence similarities between the Pan-CoV/Guangdong lineages and SARS-CoV-2-rB-CoVs
228 (Bat/China/RpYN06, bat/PrC31, bat-SL-CoVZC45, and bat-SL-CoVZXC21) at the S1-NTD

229 levels (**Figure 5D; Supplementary Figure 7A-7B**). Similarly, the SARS-CoV-2-rB-CoVs
230 displayed high sequence similarity with SARS-CoV-1-rB-CoVs of bat-SARS-coronavirus-
231 HKU3 viruses in the S1-RBD (**Figure 5D; Supplementary Figure 7A-7B**). These results
232 indicate that the SARS-CoV-2-rB-CoVs originated through recombination of the Pan-
233 CoV/Guangdong lineages and the SARS-CoV-1-rB-CoVs of bat-SARS-coronavirus-HKU3
234 viruses.

235 To explore the possible recombination, we performed RDP analysis. The recombination
236 events between the Pan-CoV/Guangdong lineages and the SARS-CoV-1-rB-CoVs of bat-
237 SARS-coronavirus-HKU3 viruses were confirmed by our analysis as follows. We
238 characterized that the bat-SL-CoVZXC21 virus is a recombinant of Pan-CoV/Guangdong
239 lineages with part of pp1ab and S1-RBD of SARS-CoV-1-rB-CoVs of bat-SARS-
240 coronavirus-HKU3-1 virus (**Figure 6A-6B**). Similarly, the bat-SL-CoVZC45 virus was
241 identified as a recombinant Pan-CoV/Guangdong lineage with the pp1ab and S1-RBD of
242 SARS-CoV-1-rB-CoVs of bat-SARS-coronavirus-HKU3-8 virus (**Figure 6C-6D**).
243 Furthermore, our results indicate that the SARS-CoV-2-rB-CoVs of bat/PrC31 as a
244 recombinant Pan-CoV/Guangdong lineage with the part of pp1ab and S1-RBD of SARS-
245 CoV-1-rB-CoVs of Bat SARS coronavirus Rp3 (**Figure 6E-6F**). Finally, our results revealed
246 that the Bat/China/RpYN06 virus is the most closely related to SARS-CoV-2 at the complete
247 genome level next to the bat/RaTG13 virus, which evolved through the recombination of
248 Pan-CoV/Guangdong lineages' and bat-SL-CoVZC45 virus (**Figure 6G-6H**). Overall, our
249 findings strongly suggest that SARS-CoV-2-rB-CoVs arose from a series of recombination
250 between Pan-CoVs and SARS-CoV-1-rB-CoVs.

251 **Discussion**

252 In this study, we show that Pan-CoVs represent an intermediate stage of evolution between
253 SARS-CoV-1/SARS-CoV-1-rB-CoVs and SARS-CoV-2/SARS-CoV-2-rB-CoVs without
254 any recombination across their genome. Interestingly, our results indicated that the host
255 receptor binding spike gene, the viral replication and translation controlling gene of pp1a, the
256 virus release regulating gene of ORF3, and the host immune regulating gene of ORF8
257 showed considerable evolution in Pan-CoVs when compared to SARS-CoV-1/SARS-CoV-1-
258 rB-CoVs. A virus that jumps from natural host to another must overcome the following host
259 genetic barriers such as host surface receptors, host antiviral proteins for viral replication,
260 virus protein translation, virus assembly, release, and the host's innate and adaptive immune
261 protection (35-39). Viruses used to mutate and bypass all these host barriers to accomplish in
262 the new species through host jump (35-37,40). In majority of the time, accidental virus
263 introduction to an unnatural host does not result in effective infection (35-37,40,41).
264 However, in rare cases/continuous/forced exposure to an unnatural host, the viruses begin to
265 adapt to the new host through virus mutations to tackle the host selection pressure and
266 prolong their lifecycle (35-37,41). Viruses must then continue to evolve to increase their
267 adaptability. Finally, it becomes highly adapted to new species to produce a large viral load
268 that can be transmitted within the new hosts (35,37,38,40). Furthermore coronaviruses in
269 general, the quasispecies evolutionary processes have been well described (39,42-49). To
270 support this, the ongoing SARS-CoV-2 outbreak in humans displayed widespread

271 mutation/deletions and evolution in the critical genes of ORF1a (46,50-54); spike gene
272 (46,50,51,55,56); ORF3 (50,57-59); and ORF8 (33,50,60-65) those viral proteins regulates
273 the virus transmission, better adaptability to host and evasion of the host immunity.
274 Furthermore, the presence of selection pressure for the evolution of the SARS-CoV-2 has
275 been highly characterized (10,52,66,67). More importantly, among the Pan-CoV/GX and
276 Pan-CoV/Guangdong lineages, Pan-CoV/Guangdong lineage displayed high sequence
277 identity with SARS-CoV-2 and was isolated from the clinically ill pangolins (8,9). Therefore,
278 collectively it can be inferred that the SARS-CoV-1/SARS-CoV-1-rB-CoVs were jumped
279 into the pangolin (new host), and then these viruses mutated/evolved in the host receptor
280 binding spike gene, viral replication, and translation controlling gene of pp1a, virus release
281 regulating gene of ORF3 and host immune regulating gene of ORF8 and which then evolved
282 as Pan-CoV/GX lineage for better host adaptation. In the second stage of enhanced pangolin
283 host adaptability, the Pan-CoV/GX lineage was then transformed to Pan-CoV/Guangdong
284 lineage (**Figure 7**). This notion is further supported by the fact that pangolins were unwell
285 and deceased as a result of the Pan-CoV/Guangdong lineage (8,9).

286 In addition, huge genetic diversity was documented between the SARS-CoV-2 and SARS-
287 CoV-2-rB-CoVs in the spike gene (5-7,14) which is responsible for critical host range,
288 transmissibility, and virus entry. Furthermore, the occurrence of widespread recombination
289 events in the spike gene of SARS-CoV-2-rB-CoVs has been extensively studied for this
290 gene's evolution (8,10-13). As per our findings in this study, revealed that the progenitor
291 SARS-CoV-2/RaTG13-like virus is recombinant of the Pan-CoV/Guangdong and Pan-
292 CoV/GX lineages (**Figure 7**). In this perspective, our study implies that a recombinant of the
293 Pan-CoV/Guangdong and Pan-CoV/GX lineages may have entered humans and begun
294 evolving in these genes for improved adaptation (**Figure 7**). Due to the mild illness or short-
295 chains of human-human infection, the SARS-CoV-2 viral infection may have gone unnoticed
296 in the early stages of the virus jump (38). The virus then evolved in humans as a result of
297 better adaptation, resulting in the SARS-CoV-2 pandemics (**Figure 7**).

298 Furthermore, RaTG13 is the only bat virus that shared a spike gene with SARS-CoV-2. Our
299 findings further demonstrated that the SARS-CoV-2-rB-CoVs of bat/PrC31, bat-SL-
300 CoVZC45, and bat-SL-CoVZXC21 were formed by recombination of Pan-CoVs and SARS-
301 CoV-1-rB-CoVs in the part of pp1a and S1-RBD (**Figure 7**). Surprisingly, Our findings
302 found evidence that the Bat/China/RpYN06 virus evolved by recombination of Pan-
303 CoV/Guangdong lineages and bat-SL-CoVZC45 virus, which is the next one to bat/RaTG13
304 virus in terms of overall genome similarity to SARS-CoV-2 (**Figure 7**).

305 In conclusion, the SARS-CoV-1/SARS-CoV-1-rB-CoVs were perhaps jumped into the
306 pangolin, further in the presence of host adaptation pressure and immune evasion, these
307 viruses might have evolved in the critical virus entry (Spike gene), replication and translation
308 (pp1a-nsp3), virus release (ORF3) and host immune evasion (protein 8) genes which then
309 lead to the origin of novel Pan-CoVs. The presence of two different lineages in the Pan-CoVs
310 indicates the different stages of host adaptation mediated evolution. The Pan-CoV/GX
311 lineage may be originated from the first stage of host adaptation and Pan-CoV/Guangdong
312 lineage might be the later stage of better-adapted virus in pangolins to cause disease. Further,

313 the progenitor RaTG13/SARS-CoV-2-like virus evolved through the recombination of the
314 Pan-CoV/GX and Pan-CoV/Guangdong lineages in the S1-NTD. Similarly, progenitor
315 SARS-CoV-2-rB-CoVs-like viruses were evolved through recombination of Pan-CoVs and
316 SARS-CoV-1-rB-CoVs in the part of pp1ab and S1-RBD. Further, the progenitor SARS-
317 CoV-2-like virus and SARS-CoV-2-rB-CoVs-like virus's gave raise to SARS-CoV-2 and
318 SARS-CoV-2-rB-CoVs respectively through host jump mediated evolution. This study
319 demonstrates that the SARS-CoV-2 outbreak emerged through host jump mediated virus
320 evolution (**Figure 7**). The close monitoring and large-scale surveillance of SARS-CoV-
321 2/SARS-CoV-2-rB-CoVs in different domestic and wild species are warranted to predict their
322 potential to evolve as a new virus that may cause the next pandemic outbreak, and also get
323 prepared with the vaccines and antivirals.

324 **Materials and Methods**

325 **Genetic diversity analysis**

326 **I. Phylogenetic analysis**

327 SARS-CoV-2 and other bat coronavirus sequences were obtained using the NCBI and
328 GISAID databases. MEGA7 was used for the phylogenetic analysis, with the Neighbor-
329 Joining method used to infer evolutionary history, bootstrap tests (1000 replicates), the
330 Maximum Composite Likelihood method used to compute evolutionary distances, and
331 gamma distribution (shape parameter = 5) used to compute the evolutionary distances being
332 employed to model the rate variation among sites. For evolutionary comparisons, differences
333 in composition bias among sequences were analyzed, and all ambiguous sites were discarded
334 in the analysis for each sequence pair.

335 **II. Net Between Group Mean Distance (NBGM) analysis**

336 The NBGM were measured using MEGA7 tool and the Kimura 2-parameter model, with a
337 gamma distribution (shape parameter=5) used as a model to measure rate variation among
338 sites, a bootstrap test (1000 replicates) was used to estimate the Standard error and all
339 ambiguous sites for each sequence pair were removed from the analysis. Above the diagonal,
340 standard error estimates were displayed.

341 **Recombination analysis**

342 **I. SimPlot analysis**

343 SimPlot 3.5.1 was used to calculate the percent identity between the query and reference
344 sequences. MEGA7 was used to align the nucleotide sequences before they were exported to
345 SimPlot 3.5.1 for further analysis. We used the Kimura two-parameter method to measure the
346 identity between the query and reference sequences, using the 500 base pair of the window at
347 a 50 base pair step.

348

349

350 **II. Recombination Detection Program (RDP) analysis**

351 In SARS-CoV-2 and similar viruses, RDP4 was utilized to detect possible recombination
352 events. The nucleotides were aligned in MEGA7 before being transferred to RDP4 for further
353 processing. For the BOOTSCAN, GENECONV, Chimaera, RDP, MaxChi, SISCAN, and
354 3seq methods, default parameter values were used, and a minimum of four or more
355 approaches were examined for probable recombination using a Bonferroni adjusted p-value
356 cut-off (0.05). The images created using the BOOTSCAN approach are shown in the figures.

357 **Funding**

358 P.A.D is a DST-INSPIRE faculty is supported by research funding from the Government of
359 India (DST/INSPIRE/04/2016/001067). P.A.D is supported by research funding from the
360 Science and Engineering Research Board, Department of Science and Technology,
361 Government of India (CRG/2018/002192).

362 **Acknowledgments**

363 We thank Director, IISc, Bangalore, India, The Chair, Department of Microbiology and Cell
364 Biology, Indian Institute of Science, Bengaluru, India, and Prof. Nagalingam R. Sundaresan,
365 Department of Microbiology and Cell Biology, Indian Institute of Science, Bengaluru, India
366 for the providing working place at IISc, Bangalore.

367 **Author contributions**

368 PAD performed most of the bioinformatics experiments. KN assisted PAD for most of the
369 bioinformatics works. PAD and KN wrote the first draft of the manuscript. PAD conceived
370 the study, designed experiments, and wrote the final version of the manuscript.

371 **Conflict of interest:** None.

372 **Data Availability**

373 We have retrieved the nucleotide sequences from publically available NCBI and GISAID
374 databases. Further, all the nucleotide sequences accession numbers and names are indicated
375 in the respective figures.

376 **References**

- 377 1. Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N. *et al.*
378 (2020) Genomic characterisation and epidemiology of 2019 novel coronavirus: implications
379 for virus origins and receptor binding. *Lancet*, **395**, 565-574.
- 380 2. Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei,
381 Y.Y. *et al.* (2020) A new coronavirus associated with human respiratory disease in China.
382 *Nature*, **579**, 265-269.
- 383 3. Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R. *et al.*
384 (2020) A Novel Coronavirus from Patients with Pneumonia in China, 2019. *The New England*
385 *journal of medicine*, **382**, 727-733.

- 386 4. Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang,
387 C.L. *et al.* (2020) A pneumonia outbreak associated with a new coronavirus of probable bat
388 origin. *Nature*, **579**, 270-273.
- 389 5. Zhou, H., Ji, J., Chen, X., Bi, Y., Li, J., Wang, Q., Hu, T., Song, H., Zhao, R., Chen, Y. *et al.* (2021)
390 Identification of novel bat coronaviruses sheds light on the evolutionary origins of SARS-CoV-
391 2 and related viruses. *Cell*.
- 392 6. Zhou, H., Chen, X., Hu, T., Li, J., Song, H., Liu, Y., Wang, P., Liu, D., Yang, J., Holmes, E.C. *et al.*
393 (2020) A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at
394 the S1/S2 Cleavage Site of the Spike Protein. *Current biology : CB*, **30**, 2196-2203 e2193.
- 395 7. Wacharapluesadee, S., Tan, C.W., Maneeorn, P., Duengkae, P., Zhu, F., Joyjinda, Y.,
396 Kaewpom, T., Chia, W.N., Ampoot, W., Lim, B.L. *et al.* (2021) Evidence for SARS-CoV-2
397 related coronaviruses circulating in bats and pangolins in Southeast Asia. *Nature*
398 *communications*, **12**, 972.
- 399 8. Lam, T.T., Jia, N., Zhang, Y.W., Shum, M.H., Jiang, J.F., Zhu, H.C., Tong, Y.G., Shi, Y.X., Ni, X.B.,
400 Liao, Y.S. *et al.* (2020) Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins.
401 *Nature*, **583**, 282-285.
- 402 9. Xiao, K., Zhai, J., Feng, Y., Zhou, N., Zhang, X., Zou, J.J., Li, N., Guo, Y., Li, X., Shen, X. *et al.*
403 (2020) Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature*, **583**,
404 286-289.
- 405 10. Li, X., Giorgi, E.E., Marichannegowda, M.H., Foley, B., Xiao, C., Kong, X.P., Chen, Y.,
406 Gnanakaran, S., Korber, B. and Gao, F. (2020) Emergence of SARS-CoV-2 through
407 recombination and strong purifying selection. *Science advances*, **6**.
- 408 11. Zhu, Z., Meng, K. and Meng, G. (2020) Genomic recombination events may reveal the
409 evolution of coronavirus and the origin of SARS-CoV-2. *Scientific reports*, **10**, 21617.
- 410 12. Makarenkov, V., Mazouze, B., Rabusseau, G. and Legendre, P. (2021) Horizontal gene
411 transfer and recombination analysis of SARS-CoV-2 genes helps discover its close relatives
412 and shed light on its origin. *BMC ecology and evolution*, **21**, 5.
- 413 13. Shahhosseini, N., Wong, G., Kobinger, G.P. and Chinikar, S. (2021) SARS-CoV-2 spillover
414 transmission due to recombination event. *Gene reports*, **23**, 101045.
- 415 14. York, A. (2021) Searching for relatives of SARS-CoV-2 in bats. *Nature reviews. Microbiology*.
- 416 15. Walls, A.C., Park, Y.J., Tortorici, M.A., Wall, A., McGuire, A.T. and Velesler, D. (2020)
417 Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, **181**, 281-
418 292 e286.
- 419 16. Chi, X., Yan, R., Zhang, J., Zhang, G., Zhang, Y., Hao, M., Zhang, Z., Fan, P., Dong, Y., Yang, Y.
420 *et al.* (2020) A neutralizing human antibody binds to the N-terminal domain of the Spike
421 protein of SARS-CoV-2. *Science*, **369**, 650-655.
- 422 17. Li, F. (2016) Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annual review*
423 *of virology*, **3**, 237-261.
- 424 18. Li, F. (2015) Receptor recognition mechanisms of coronaviruses: a decade of structural
425 studies. *Journal of virology*, **89**, 1954-1964.
- 426 19. Hulswit, R.J., de Haan, C.A. and Bosch, B.J. (2016) Coronavirus Spike Protein and Tropism
427 Changes. *Advances in virus research*, **96**, 29-57.
- 428 20. Li, M., Ye, G., Si, Y., Shen, Z., Liu, Z., Shi, Y., Xiao, S., Fu, Z.F. and Peng, G. (2021) Structure of
429 the multiple functional domains from coronavirus nonstructural protein 3. *Emerging*
430 *microbes & infections*, **10**, 66-80.
- 431 21. Lei, J., Kusov, Y. and Hilgenfeld, R. (2018) Nsp3 of coronaviruses: Structures and functions of
432 a large multi-domain protein. *Antiviral research*, **149**, 58-74.
- 433 22. Alhammad, Y.M.O., Kashipathy, M.M., Roy, A., Gagne, J.P., McDonald, P., Gao, P., Nonfoux,
434 L., Battaile, K.P., Johnson, D.K., Holmstrom, E.D. *et al.* (2021) The SARS-CoV-2 Conserved
435 Macrodomain Is a Mono-ADP-Ribosylhydrolase. *Journal of virology*, **95**.

- 436 23. Frick, D.N., Viridi, R.S., Vuksanovic, N., Dahal, N. and Silvaggi, N.R. (2020) Molecular Basis for
437 ADP-Ribose Binding to the Mac1 Domain of SARS-CoV-2 nsp3. *Biochemistry*, **59**, 2608-2615.
- 438 24. Mariano, G., Farthing, R.J., Lale-Farjat, S.L.M. and Bergeron, J.R.C. (2020) Structural
439 Characterization of SARS-CoV-2: Where We Are, and Where We Need to Be. *Frontiers in*
440 *molecular biosciences*, **7**, 605236.
- 441 25. Neuman, B.W. (2016) Bioinformatics and functional analyses of coronavirus nonstructural
442 proteins involved in the formation of replicative organelles. *Antiviral research*, **135**, 97-107.
- 443 26. V'Kovski, P., Kratzel, A., Steiner, S., Stalder, H. and Thiel, V. (2021) Coronavirus biology and
444 replication: implications for SARS-CoV-2. *Nature reviews. Microbiology*, **19**, 155-170.
- 445 27. Khan, M.T., Zeb, M.T., Ahsan, H., Ahmed, A., Ali, A., Akhtar, K., Malik, S.I., Cui, Z., Ali, S.,
446 Khan, A.S. *et al.* (2021) SARS-CoV-2 nucleocapsid and Nsp3 binding: an in silico study.
447 *Archives of microbiology*, **203**, 59-66.
- 448 28. Tan, Y., Schneider, T., Shukla, P.K., Chandrasekharan, M.B., Aravind, L. and Zhang, D. (2021)
449 Unification and extensive diversification of M/Orf3-related ion channel proteins in
450 coronaviruses and other nidoviruses. *Virus evolution*, **7**, veab014.
- 451 29. Lu, W., Zheng, B.J., Xu, K., Schwarz, W., Du, L., Wong, C.K., Chen, J., Duan, S., Deubel, V. and
452 Sun, B. (2006) Severe acute respiratory syndrome-associated coronavirus 3a protein forms
453 an ion channel and modulates virus release. *Proceedings of the National Academy of*
454 *Sciences of the United States of America*, **103**, 12540-12545.
- 455 30. Zhang, Y., Sun, H., Pei, R., Mao, B., Zhao, Z., Li, H., Lin, Y. and Lu, K. (2021) The SARS-CoV-2
456 protein ORF3a inhibits fusion of autophagosomes with lysosomes. *Cell discovery*, **7**, 31.
- 457 31. Miao, G., Zhao, H., Li, Y., Ji, M., Chen, Y., Shi, Y., Bi, Y., Wang, P. and Zhang, H. (2021) ORF3a
458 of the COVID-19 virus SARS-CoV-2 blocks HOPS complex-mediated assembly of the SNARE
459 complex required for autolysosome formation. *Developmental cell*, **56**, 427-442 e425.
- 460 32. Koepke, L., Hirschenberger, M., Hayn, M., Kirchhoff, F. and Sparrer, K.M. (2021)
461 Manipulation of autophagy by SARS-CoV-2 proteins. *Autophagy*, 1-3.
- 462 33. Flower, T.G., Buffalo, C.Z., Hooy, R.M., Allaire, M., Ren, X. and Hurley, J.H. (2021) Structure of
463 SARS-CoV-2 ORF8, a rapidly evolving immune evasion protein. *Proceedings of the National*
464 *Academy of Sciences of the United States of America*, **118**.
- 465 34. Zhang, Y., Chen, Y., Li, Y., Huang, F., Luo, B., Yuan, Y., Xia, B., Ma, X., Yang, T., Yu, F. *et al.*
466 (2021) The ORF8 protein of SARS-CoV-2 mediates immune evasion through down-regulating
467 MHC-Iota. *Proceedings of the National Academy of Sciences of the United States of America*,
468 **118**.
- 469 35. Parrish, C.R., Holmes, E.C., Morens, D.M., Park, E.C., Burke, D.S., Calisher, C.H., Laughlin,
470 C.A., Saif, L.J. and Daszak, P. (2008) Cross-species virus transmission and the emergence of
471 new epidemic diseases. *Microbiology and molecular biology reviews : MMBR*, **72**, 457-470.
- 472 36. Geoghegan, J.L. and Holmes, E.C. (2017) Predicting virus emergence amid evolutionary
473 noise. *Open biology*, **7**.
- 474 37. Warren, C.J. and Sawyer, S.L. (2019) How host genetics dictates successful viral zoonosis.
475 *PLoS biology*, **17**, e3000217.
- 476 38. Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C. and Garry, R.F. (2020) The proximal
477 origin of SARS-CoV-2. *Nature medicine*, **26**, 450-452.
- 478 39. Song, H.D., Tu, C.C., Zhang, G.W., Wang, S.Y., Zheng, K., Lei, L.C., Chen, Q.X., Gao, Y.W., Zhou,
479 H.Q., Xiang, H. *et al.* (2005) Cross-host evolution of severe acute respiratory syndrome
480 coronavirus in palm civet and human. *Proceedings of the National Academy of Sciences of*
481 *the United States of America*, **102**, 2430-2435.
- 482 40. Antia, R., Regoes, R.R., Koella, J.C. and Bergstrom, C.T. (2003) The role of evolution in the
483 emergence of infectious diseases. *Nature*, **426**, 658-661.
- 484 41. Wolfe, N.D., Dunavan, C.P. and Diamond, J. (2007) Origins of major human infectious
485 diseases. *Nature*, **447**, 279-283.

- 486 42. Xu, D., Zhang, Z. and Wang, F.S. (2004) SARS-associated coronavirus quasispecies in
487 individual patients. *The New England journal of medicine*, **350**, 1366-1367.
- 488 43. Tang, J.W., Cheung, J.L., Chu, I.M., Sung, J.J., Peiris, M. and Chan, P.K. (2006) The large 386-nt
489 deletion in SARS-associated coronavirus: evidence for quasispecies? *The Journal of infectious*
490 *diseases*, **194**, 808-813.
- 491 44. Zhang, X., Hasoksuz, M., Spiro, D., Halpin, R., Wang, S., Vlasova, A., Janies, D., Jones, L.R.,
492 Ghedin, E. and Saif, L.J. (2007) Quasispecies of bovine enteric and respiratory coronaviruses
493 based on complete genome sequences and genetic changes after tissue culture adaptation.
494 *Virology*, **363**, 1-10.
- 495 45. Jary, A., Leducq, V., Malet, I., Marot, S., Klement-Frutos, E., Teyssou, E., Soulie, C., Abdi, B.,
496 Wirden, M., Pourcher, V. *et al.* (2020) Evolution of viral quasispecies during SARS-CoV-2
497 infection. *Clinical microbiology and infection : the official publication of the European Society*
498 *of Clinical Microbiology and Infectious Diseases*, **26**, 1560 e1561-1560 e1564.
- 499 46. Armero, A., Berthet, N. and Avarre, J.C. (2021) Intra-Host Diversity of SARS-Cov-2 Should Not
500 Be Neglected: Case of the State of Victoria, Australia. *Viruses*, **13**.
- 501 47. Nyayanit, D.A., Yadav, P.D., Kharde, R. and Shete-Aich, A. (2020) Quasispecies analysis of the
502 SARS-CoV-2 from representative clinical samples: A preliminary analysis. *The Indian journal*
503 *of medical research*, **152**, 105-107.
- 504 48. Al Khatib, H.A., Benslimane, F.M., Elbashir, I.E., Coyle, P.V., Al Maslamani, M.A., Al-Khal, A.,
505 Al Thani, A.A. and Yassine, H.M. (2020) Within-Host Diversity of SARS-CoV-2 in COVID-19
506 Patients With Variable Disease Severities. *Frontiers in cellular and infection microbiology*, **10**,
507 575613.
- 508 49. Karamitros, T., Papadopoulou, G., Bousali, M., Mexias, A., Tsiodras, S. and Mentis, A. (2020)
509 SARS-CoV-2 exhibits intra-host genomic plasticity and low-frequency polymorphic
510 quasispecies. *Journal of clinical virology : the official publication of the Pan American Society*
511 *for Clinical Virology*, **131**, 104585.
- 512 50. Omotoso, O.E., Babalola, A.D. and Mataareek, A. (2021) Mutational hotspots and conserved
513 domains of SARS-CoV-2 genome in African population. *Beni-Suef University journal of basic*
514 *and applied sciences*, **10**, 11.
- 515 51. Kim, J.S., Jang, J.H., Kim, J.M., Chung, Y.S., Yoo, C.K. and Han, M.G. (2020) Genome-Wide
516 Identification and Characterization of Point Mutations in the SARS-CoV-2 Genome. *Osong*
517 *public health and research perspectives*, **11**, 101-111.
- 518 52. Velazquez-Salinas, L., Zarate, S., Eberl, S., Gladue, D.P., Novella, I. and Borca, M.V. (2020)
519 Positive Selection of ORF1ab, ORF3a, and ORF8 Genes Drives the Early Evolutionary Trends
520 of SARS-CoV-2 During the 2020 COVID-19 Pandemic. *Frontiers in microbiology*, **11**, 550674.
- 521 53. Banerjee, S., Seal, S., Dey, R., Mondal, K.K. and Bhattacharjee, P. (2021) Mutational spectra
522 of SARS-CoV-2 orf1ab polyprotein and signature mutations in the United States of America.
523 *Journal of medical virology*, **93**, 1428-1435.
- 524 54. Srivastava, S., Banu, S., Singh, P., Sowpati, D.T. and Mishra, R.K. (2021) SARS-CoV-2
525 genomics: An Indian perspective on sequencing viral variants. *Journal of biosciences*, **46**.
- 526 55. Harvey, W.T., Carabelli, A.M., Jackson, B., Gupta, R.K., Thomson, E.C., Harrison, E.M.,
527 Ludden, C., Reeve, R., Rambaut, A., Consortium, C.-G.U. *et al.* (2021) SARS-CoV-2 variants,
528 spike mutations and immune escape. *Nature reviews. Microbiology*, **19**, 409-424.
- 529 56. Guruprasad, L. (2021) Human SARS CoV-2 spike protein mutations. *Proteins*, **89**, 569-576.
- 530 57. Lam, J.Y., Yuen, C.K., Ip, J.D., Wong, W.M., To, K.K., Yuen, K.Y. and Kok, K.H. (2020) Loss of
531 orf3b in the circulating SARS-CoV-2 strains. *Emerging microbes & infections*, **9**, 2685-2696.
- 532 58. Hassan, S.S., Choudhury, P.P., Basu, P. and Jana, S.S. (2020) Molecular conservation and
533 differential mutation on ORF3a gene in Indian SARS-CoV2 genomes. *Genomics*, **112**, 3226-
534 3237.
- 535 59. Bianchi, M., Borsetti, A., Ciccozzi, M. and Pascarella, S. (2021) SARS-Cov-2 ORF3a: Mutability
536 and function. *International journal of biological macromolecules*, **170**, 820-826.

- 537 60. Tan, Y., Schneider, T., Leong, M., Aravind, L. and Zhang, D. (2020) Novel Immunoglobulin
538 Domain Proteins Provide Insights into Evolution and Pathogenesis of SARS-CoV-2-Related
539 Viruses. *mBio*, **11**.
- 540 61. Zinzula, L. (2021) Lost in deletion: The enigmatic ORF8 protein of SARS-CoV-2. *Biochemical
541 and biophysical research communications*, **538**, 116-124.
- 542 62. Mohammad, S., Bouchama, A., Mohammad Alharbi, B., Rashid, M., Saleem Khatlani, T.,
543 Gaber, N.S. and Malik, S.S. (2020) SARS-CoV-2 ORF8 and SARS-CoV ORF8ab: Genomic
544 Divergence and Functional Convergence. *Pathogens*, **9**.
- 545 63. Su, Y.C.F., Anderson, D.E., Young, B.E., Linster, M., Zhu, F., Jayakumar, J., Zhuang, Y.,
546 Kalimuddin, S., Low, J.G.H., Tan, C.W. *et al.* (2020) Discovery and Genomic Characterization
547 of a 382-Nucleotide Deletion in ORF7b and ORF8 during the Early Evolution of SARS-CoV-2.
548 *mBio*, **11**.
- 549 64. Hassan, S.S., Aljabali, A.A.A., Panda, P.K., Ghosh, S., Attrish, D., Choudhury, P.P., Seyran, M.,
550 Pizzol, D., Adadi, P., Abd El-Aziz, T.M. *et al.* (2021) A unique view of SARS-CoV-2 through the
551 lens of ORF8 protein. *Computers in biology and medicine*, **133**, 104380.
- 552 65. Pereira, F. (2020) Evolutionary dynamics of the SARS-CoV-2 ORF8 accessory gene. *Infection,
553 genetics and evolution : journal of molecular epidemiology and evolutionary genetics in
554 infectious diseases*, **85**, 104525.
- 555 66. Frank, H.K., Enard, D. and Boyd, S.D. (2020) Exceptional diversity and selection pressure on
556 SARS-CoV and SARS-CoV-2 host receptor in bats compared to other mammals. *bioRxiv*.
- 557 67. MacLean, O.A., Lytras, S., Weaver, S., Singer, J.B., Boni, M.F., Lemey, P., Kosakovsky Pond,
558 S.L. and Robertson, D.L. (2021) Natural selection in the evolution of SARS-CoV-2 in bats
559 created a generalist virus and highly capable human pathogen. *PLoS biology*, **19**, e3001115.

560

561 **Figures and Legends**

562 **Figure 1. SARS-CoV-2 and Pan-CoVs genetic diversity at the complete genome. (A &B)**
563 The Phylogenetic tree (A) and net between-group mean distance (B) respectively for the Pan-
564 CoVs, SARS-CoV-2, and related viruses using the complete genome sequence. (C) Similarity
565 plot analysis using the complete genomes of the SARS-CoV-2 and SARS-CoV-2-rB-CoVs.
566 SARS-CoV-2 is set as a query sequence (D) Similarity plot analysis using the complete
567 genomes of the SARS-CoV-2, SARS-CoV-2-rB-CoVs, Pan-CoVs, SARS-CoV-1, and
568 SARS-CoV-1-rB-CoVs. SARS-CoV-2 is set as a query sequence.

569 **Figure 2. The genetic diversity analysis of the pp1ab gene. (A-C)** The Phylogenetic tree
570 (A), net between-group mean distance (B), and similarity plot analysis (C) respectively for
571 the Pan-CoVs, SARS-CoV-2, and related viruses using part of pp1ab (1-11725nt). SARS-
572 CoV-2 is set as a query sequence. (D-F) The Phylogenetic tree (D), net between-group mean
573 distance (E), and similarity plot analysis (F) respectively for the Pan-CoVs, SARS-CoV-2,
574 and related viruses using part of pp1ab from 11726nt to 21561 nucleotides. SARS-CoV-2 is
575 set as a query sequence.

576 **Figure 3. The genetic diversity analysis of ORF3, E, M, protein 6, protein 7a, and**
577 **protein 7b. (A-C)** The Phylogenetic tree (A), net between-group mean distance (B), and
578 similarity plot analysis (C) respectively for the Pan-CoVs, SARS-CoV-2, and related viruses
579 using nucleotide sequences of complete ORF3. (D-F) The Phylogenetic tree (D), net
580 between-group mean distance (E), and similarity plot analysis (F) respectively for the Pan-

581 CoVs, SARS-CoV-2, and related viruses using the nucleotide sequences of E, M, protein 6,
582 protein 7a, and protein 7b. SARS-CoV-2 is set as a query sequence.

583 **Figure 4. The genetic diversity analysis of ORF8 (protein 8), nucleocapsid (N), ORF10,**
584 **and 3'UTR. (A-B)** The Phylogenetic tree (A), and net between-group mean distance (B)
585 respectively, for the Pan-CoVs, SARS-CoV-2, and related viruses using nucleotide sequences
586 of complete ORF8 (protein 8). **(C-E)** The Phylogenetic tree (C), net between-group mean
587 distance (D), and similarity plot analysis (E) respectively for the Pan-CoVs, SARS-CoV-2,
588 and related viruses using the nucleotide sequences of the nucleocapsid (N), ORF10, and
589 3'UTR. SARS-CoV-2 is set as a query sequence.

590 **Figure 5. Origin SARS-CoV-2 spike gene through recombination of Pan-CoVs. (A-B)**
591 The Phylogenetic tree (A), and net between-group mean distance (B) respectively, for the
592 Pan-CoVs, SARS-CoV-2, and related viruses using nucleotide sequences of complete spike
593 gene. **(C)** Similarity plot analysis for the SARS-CoV-2, bat/RaTG13, Pan-CoV/GX, and Pan-
594 CoV/Guangdong lineages using the complete genome nucleotide sequences. SARS-CoV-2 is
595 used as a query sequence. **(D)** Similarity plot analysis for the Pan-CoVs, SARS-CoV-2, and
596 related viruses using nucleotide sequences of complete spike gene. SARS-CoV-2 is used as a
597 query sequence. **(E)** The recombination analysis with BOOTSCAN method using complete
598 genomes of the SARS-CoV-2, PCoV-GX-P2V, and Pangolin/Guangdong/MP789. **(F)** The
599 recombination analysis with BOOTSCAN method using complete genomes of the
600 Bat/RaTG13, PCoV-GX-P2V, and Pangolin/Guangdong/MP789.

601 **Figure 6. Origin SARS-CoV-2-rB-CoVs spike gene through recombination. (A&B)** The
602 recombination analysis with Similarity plot (A) and BOOTSCAN (B) using the complete
603 genomes the Bat-SL-CoVZXC21, Pangolin/Guangdong/A22-2, and Bat/SARS/HKU3-1;
604 Bat-SL-CoVZXC21 used as a query sequence. **(C&D)** The recombination analysis with
605 Similarity plot (C) and BOOTSCAN (D) using the complete genomes the Bat-SL-CoVZC45,
606 Pangolin/Guangdong/A22-2, and Bat/SARS/HKU3-8; Bat-SL-CoVZC45 used as a query
607 sequence. **(E&F)** The recombination analysis: Similarity plot (E) and BOOTSCAN (F) using
608 complete genomes of the Bat/PrC31, Pangolin/Guangdong/A22-2, and Bat/SARS/Rp3;
609 Bat/PrC31 used as a query sequence. **(G&H)** The recombination analysis with Similarity plot
610 (G) and BOOTSCAN (H) using complete genomes of the Bat/China/RpYN06,
611 Pangolin/Guangdong/A22-2, and Bat-SL-CoVZC45; Bat/China/RpYN06 was used as a query
612 sequence.

613 **Figure 7. Schematic representation of origin of the SARS-CoV-2 from SARS-CoV-1**
614 **through Pan-CoVs.** The SARS-CoV-1/SARS-CoV-1-rB-CoVs were perhaps jumped into
615 the pangolin, and further evolved into a novel Pan-CoVs (two different lineages namely Pan-
616 CoV/GX and Pan-CoV/Guangdong) by new host pressure. The SARS-CoV-2 evolved from
617 the recombination of Pan-CoV/GX and Pan-CoV/Guangdong. Further, SARS-CoV-2-rB-
618 CoVs were evolved through recombination of Pan-CoVs and SARS-CoV-1-rB-CoVs.

619

620

621 **Supplementary Figures and Legends**

622 **Supplementary Figure 1. Similarity plot analysis using the complete genomes of SARS-**
623 **CoV-2 and its related viruses. (A)** Similarity plot analysis using the complete genomes of
624 the SARS-CoV-2, SARS-CoV-2-rB-CoVs, Pan-CoVs, SARS-CoV-1, and SARS-CoV-1-rB-
625 CoVs. The Pan-CoV/GX lineages are set as a query sequence. **(B)** Similarity plot analysis
626 using the complete genomes of the SARS-CoV-2, SARS-CoV-2-rB-CoVs, Pan-CoVs,
627 SARS-CoV-1, and SARS-CoV-1-rB-CoVs. The SARS-CoV-1 is set as a query sequence.

628 **Supplementary Figure 2. Similarity plot analysis for the part of pp1ab (1-11725nt). (A)**
629 Similarity plot analysis for the Pan-CoVs, SARS-CoV-2, and related viruses using part of
630 pp1ab (1-11725nt). The SARS-CoV-1 is set as a query sequence. **(B)** Similarity plot analysis
631 for the Pan-CoVs, SARS-CoV-2, and related viruses using part of pp1ab (1-11725nt). The
632 Pan-CoV/GX lineages are set as a query sequence.

633 **Supplementary Figure 3. Similarity plot analysis for the part of pp1ab (11726nt to**
634 **21561nt). (A)** Similarity plot analysis for the Pan-CoVs, SARS-CoV-2, and related viruses
635 using part of pp1ab (11726nt to 21561nt). The Pan-CoV/GX lineages are set as a query
636 sequence. **(B)** Similarity plot analysis for the Pan-CoVs, SARS-CoV-2, and related viruses
637 using part of pp1ab (11726nt to 21561nt). The SARS-CoV-1 is set as a query sequence.

638 **Supplementary Figure 4. Similarity plot analysis for the ORF3. (A)** Similarity plot
639 analysis for the Pan-CoVs, SARS-CoV-2, and related viruses using nucleotide sequence of
640 complete ORF3. The SARS-CoV-1 is set as a query sequence. **(B)** Similarity plot analysis for
641 the Pan-CoVs, SARS-CoV-2, and related viruses using nucleotide sequence of complete
642 ORF3. The Pan-CoV/GX lineages are set as a query sequence.

643 **Supplementary Figure 5. Similarity plot analysis for the E, M, protein 6, protein 7a, and**
644 **protein 7b genes. (A)** Similarity plot analysis for the Pan-CoVs, SARS-CoV-2, and related
645 viruses using the nucleotide sequences of E, M, protein 6, protein 7a, and protein 7b. The
646 SARS-CoV-1 is set as a query sequence. **(B)** Similarity plot analysis for the Pan-CoVs,
647 SARS-CoV-2, and related viruses using the nucleotide sequences of E, M, protein 6, protein
648 7a, and protein 7b. The Pan-CoV/GX lineages are set as a query sequence.

649 **Supplementary Figure 6. Similarity plot analysis for the complete nucleocapsid (N),**
650 **ORF10, and 3'UTR. (A)** Similarity plot for the Pan-CoVs, SARS-CoV-2, and related
651 viruses using the nucleotide sequences of the nucleocapsid (N), ORF10, and 3'UTR. The
652 SARS-CoV-1 is set as a query sequence. **(B)** Similarity plot for the Pan-CoVs, SARS-CoV-2,
653 and related viruses using the nucleotide sequences of the nucleocapsid (N), ORF10, and
654 3'UTR. The Pan-CoV/GX lineages are set as a query sequence.

655 **Supplementary Figure 7. Similarity plot analysis for the complete spike gene. (A)**
656 Similarity plot for the Pan-CoVs, SARS-CoV-2, and related viruses using the nucleotide
657 sequences of complete spike gene. The SARS-CoV-1 is set as a query sequence. **(B)**
658 Similarity plot for the Pan-CoVs, SARS-CoV-2, and related viruses using the nucleotide
659 sequences of complete spike gene. The Pan-CoV/GX lineages are set as a query sequence.

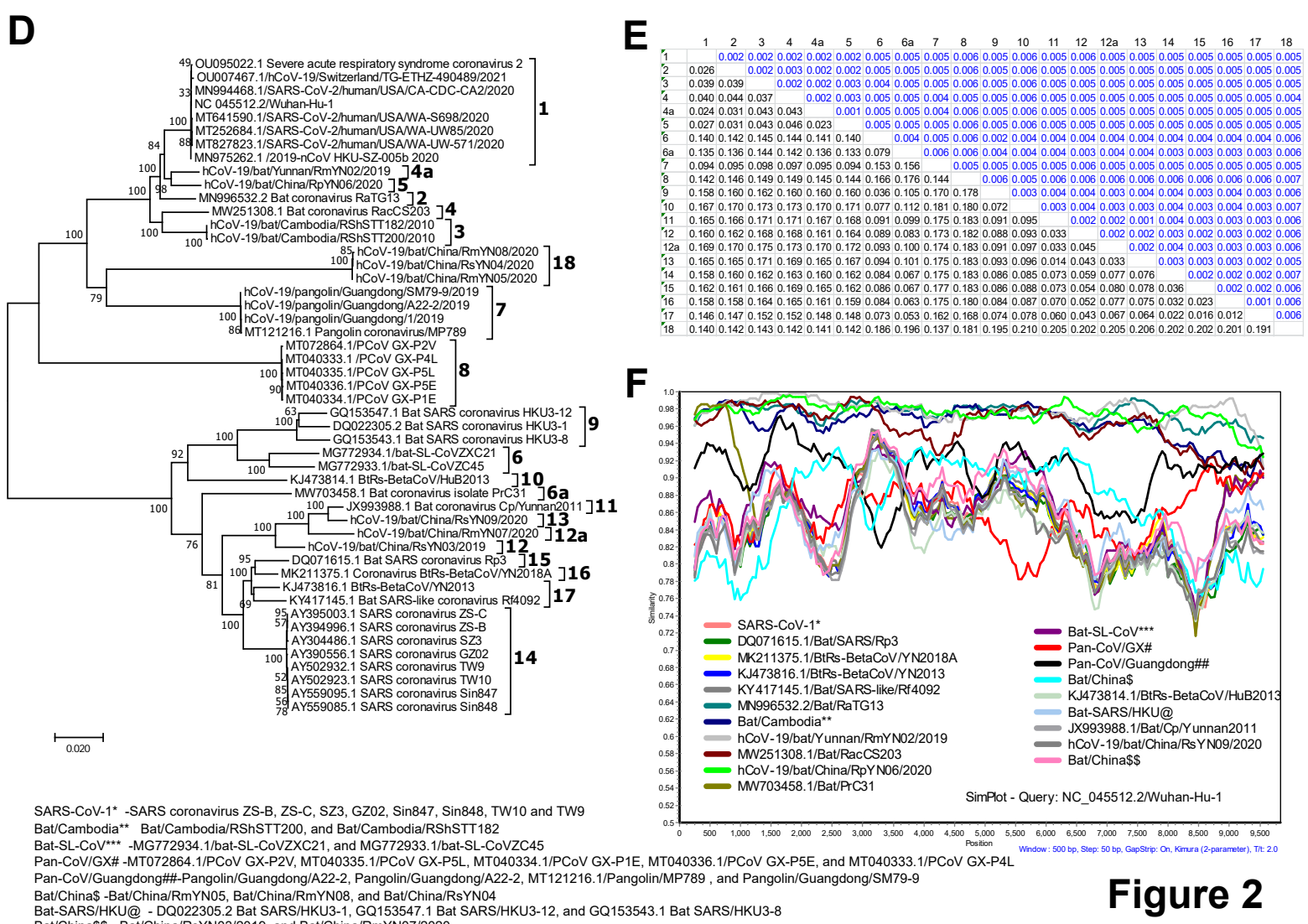
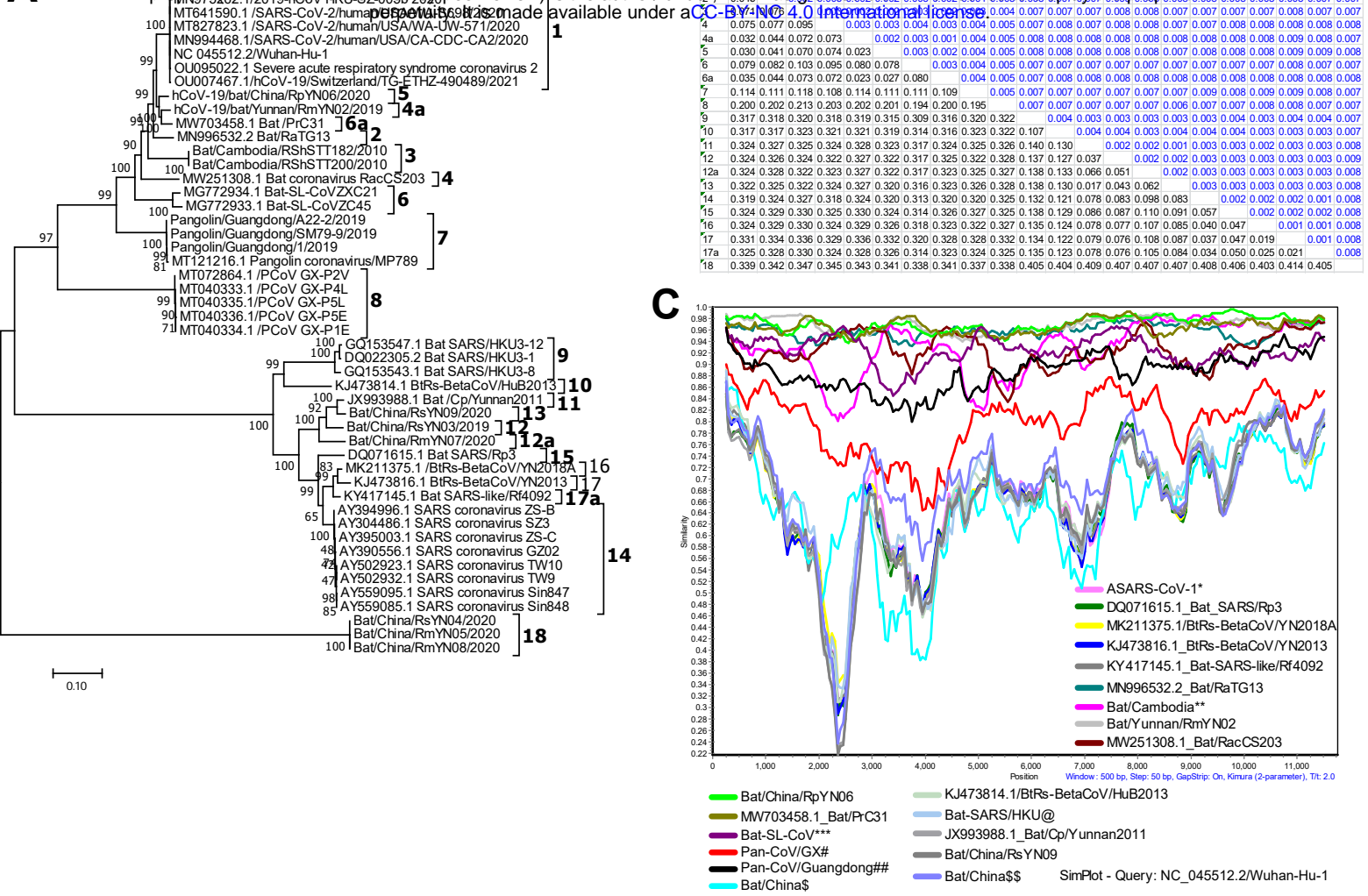
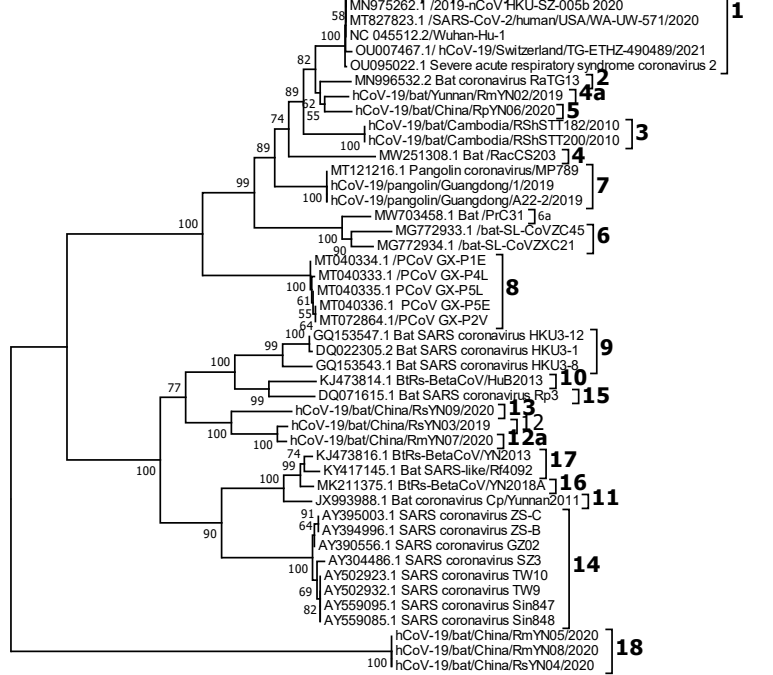
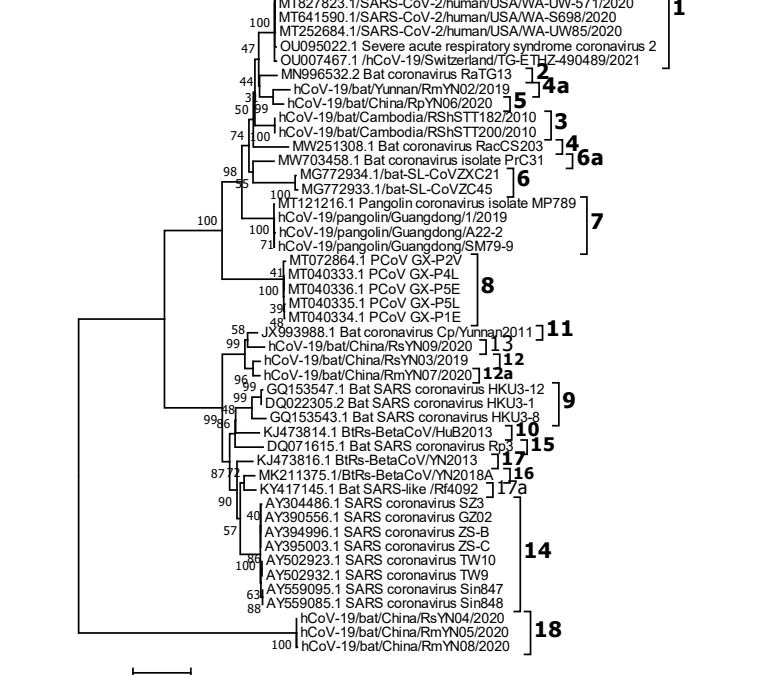


Figure 2

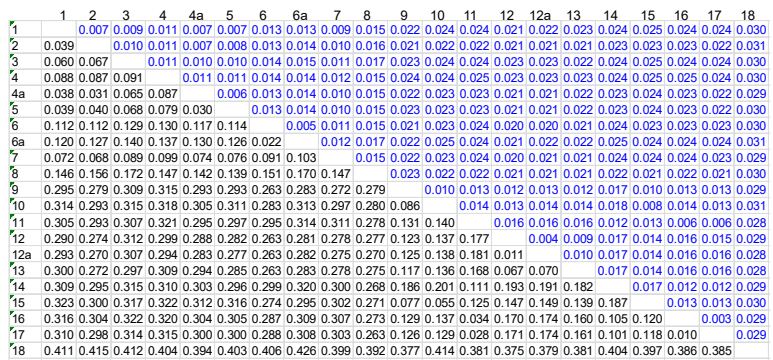
A



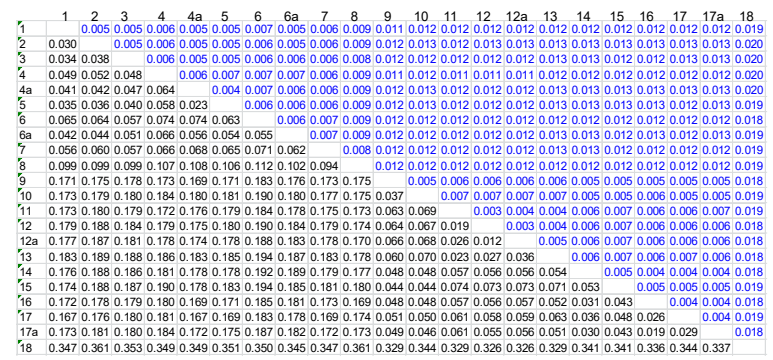
D



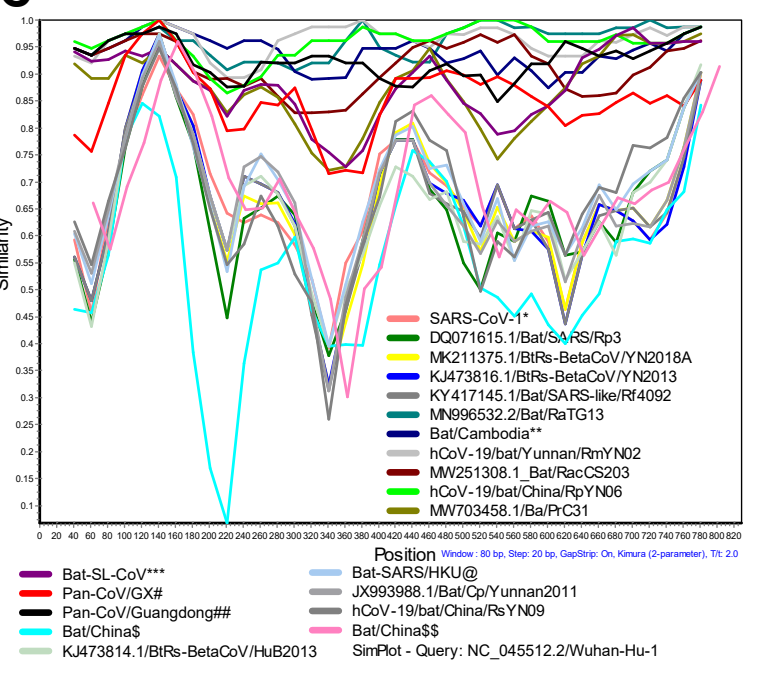
B



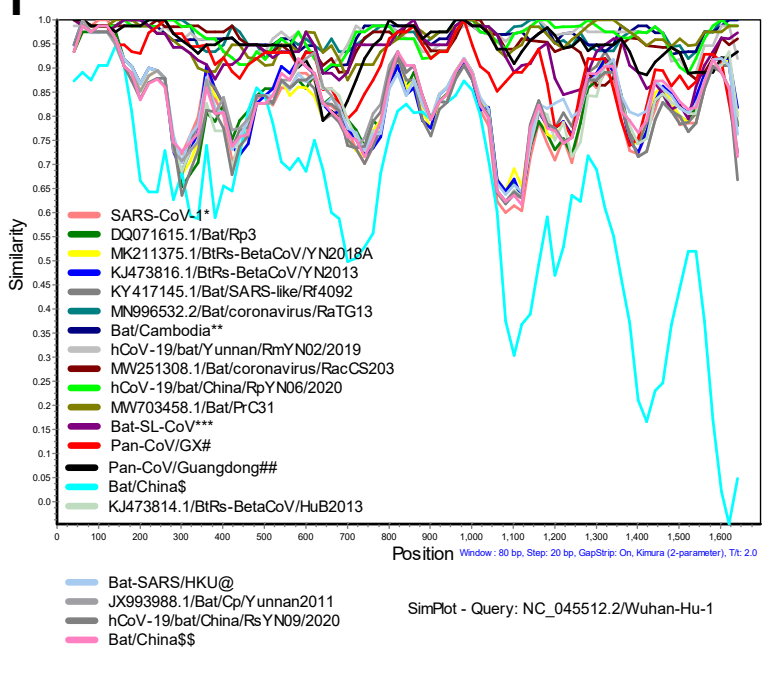
E



C



F



SARS-CoV-1* -SARS coronavirus ZS-B, ZS-C, SZ3, GZ02, Sin847, Sin848, TW10 and TW9
 Bat/Cambodia** Bat/Cambodia/RShSTT200, and Bat/Cambodia/RShSTT182
 Bat-SL-CoV*** -MG772934.1/bat-SL-CoVZXC21, and MG772933.1/bat-SL-CoVZC45
 Pan-CoV/GX# -MT072864.1/PCoV GX-P2V, MT040335.1/PCoV GX-P5L, MT040334.1/PCoV GX-P4L
 Pan-CoV/Guangdong## -Pangolin/Guangdong/A22-2, Pangolin/Guangdong/A22-2, MT121216.1/Pangolin/MP789, and Pangolin/Guangdong/SM79-9
 Bat/China\$ -Bat/China/RmYN05, Bat/China/RmYN08, and Bat/China/RsYN04
 Bat-SARS/HKU@ - DQ071615.2 Bat SARS/HKU3-1, GQ153547.1 Bat SARS/HKU3-12, and GQ153543.1 Bat SARS/HKU3-8
 Bat/China\$\$ -Bat/China/RsYN03/2019, and Bat/China/RmYN07/2020

Figure 3

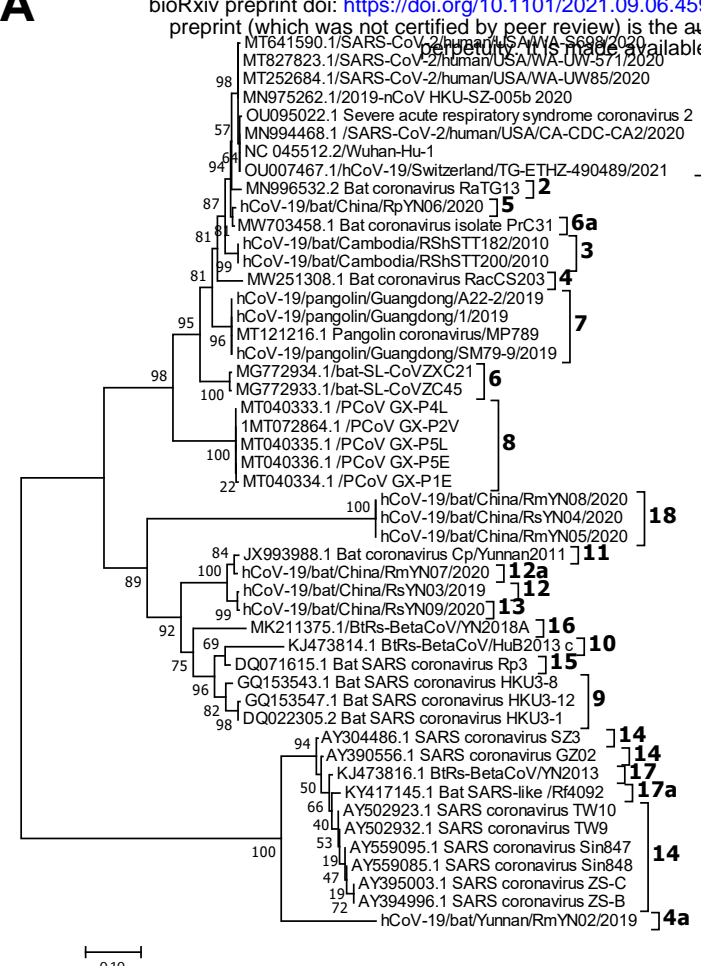


Table (B) showing pairwise similarity values between sequences. The table is a lower triangular matrix with columns labeled 1 through 18. The values represent the similarity between sequences in the corresponding rows and columns.

	1	2	3	4a	5	6	6a	7	8	9	10	11	12	12a	13	14	15	16	17	17a	18	
1	1																					
2	0.047	1																				
3	0.083	0.083	1																			
4a	1.040	1.016	0.933	1																		
5	0.023	0.028	0.040	0.079	1																	
6	0.122	0.131	0.123	0.130	0.975	1																
6a	0.017	0.022	0.034	0.076	1.004	0.005	1															
7	0.081	0.087	0.097	0.086	1.027	0.078	0.098	1														
8	0.229	0.230	0.243	0.238	1.072	0.229	0.201	0.229	1													
9	0.464	0.458	0.455	0.456	0.965	0.466	0.496	0.467	0.473	1												
10	0.567	0.563	0.552	0.556	1.014	0.569	0.604	0.569	0.562	0.591	1											
11	0.473	0.476	0.483	0.487	0.987	0.477	0.505	0.476	0.443	0.515	0.210	1										
12	0.479	0.482	0.501	0.476	1.034	0.483	0.499	0.483	0.426	0.503	0.205	0.268	1									
12a	0.473	0.464	0.483	0.477	1.025	0.476	0.443	0.515	0.198	0.244	0.104	0.043	0.111	1								
13	0.473	0.476	0.495	0.470	1.047	0.477	0.506	0.477	0.432	0.509	0.201	0.264	0.037	0.003	0.046	1						
14	0.990	0.996	0.921	1.010	0.225	0.962	0.945	0.970	0.968	0.932	0.922	0.898	0.927	0.955	0.924	0.969	1					
15	0.472	0.475	0.476	0.486	0.932	0.470	0.505	0.475	0.452	0.448	0.080	0.114	0.214	0.206	0.202	0.210	0.868	1				
16	0.479	0.475	0.467	0.463	1.027	0.486	0.526	0.486	0.475	0.548	0.164	0.228	0.224	0.220	0.212	0.216	0.977	0.186	1			
17	0.962	0.977	0.911	1.036	0.243	0.944	0.944	0.954	0.966	0.908	0.923	0.922	0.944	0.976	0.932	0.988	0.016	0.886	1.011	1		
17a	0.996	0.987	0.944	1.060	0.247	0.977	0.956	0.988	0.988	0.938	0.924	0.956	0.988	0.920	1.001	0.027	0.887	1.027	0.026	0.148	1	
18	0.739	0.713	0.704	0.745	1.343	0.747	0.767	0.748	0.730	0.694	0.525	0.603	0.608	0.612	0.572	0.605	1.182	0.524	0.587	1.157	1.157	1

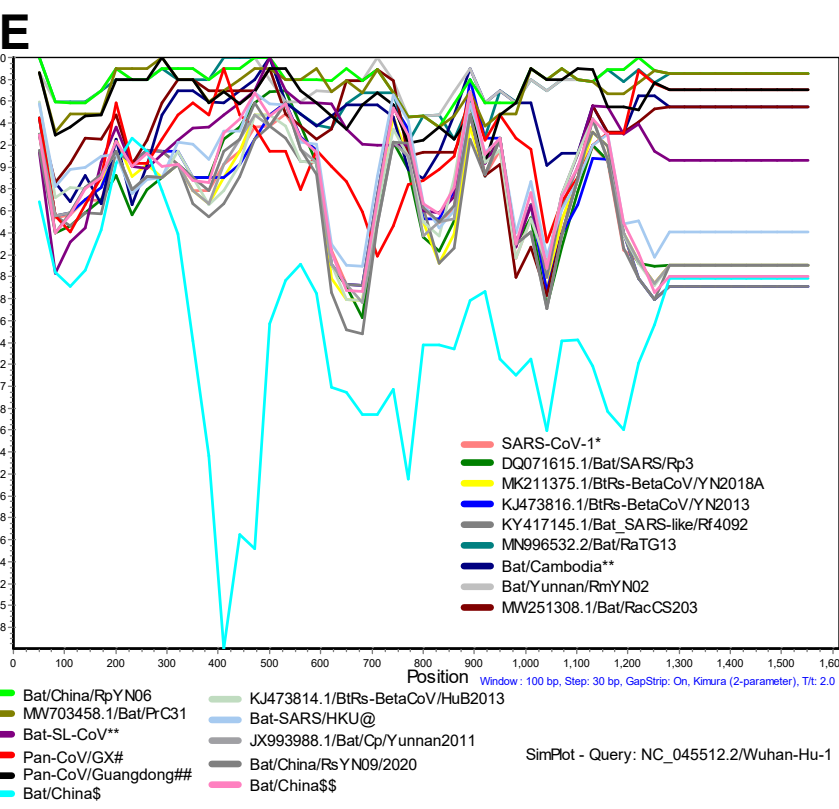
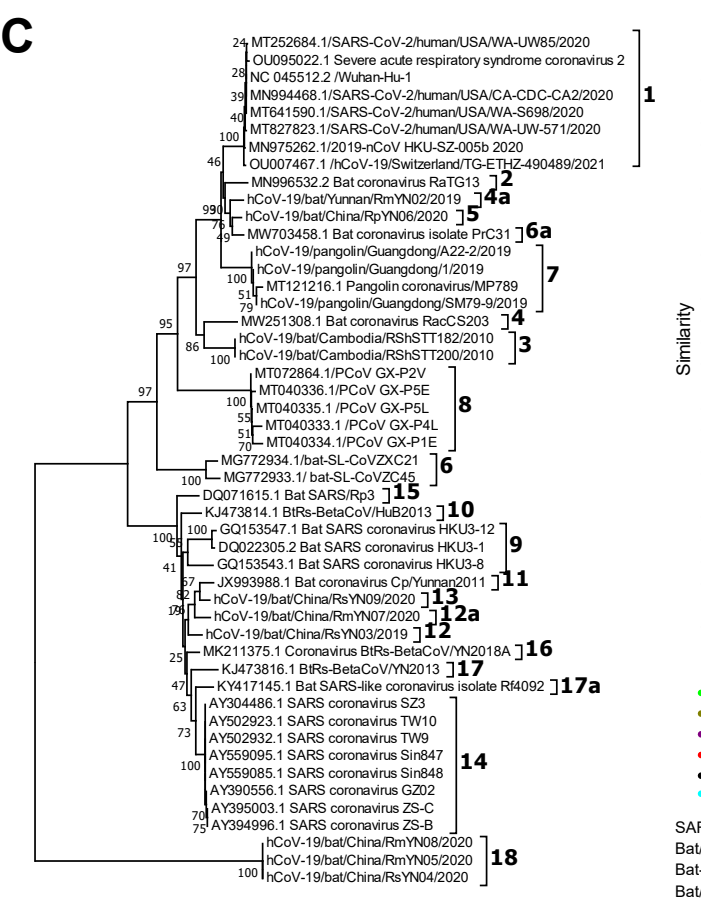
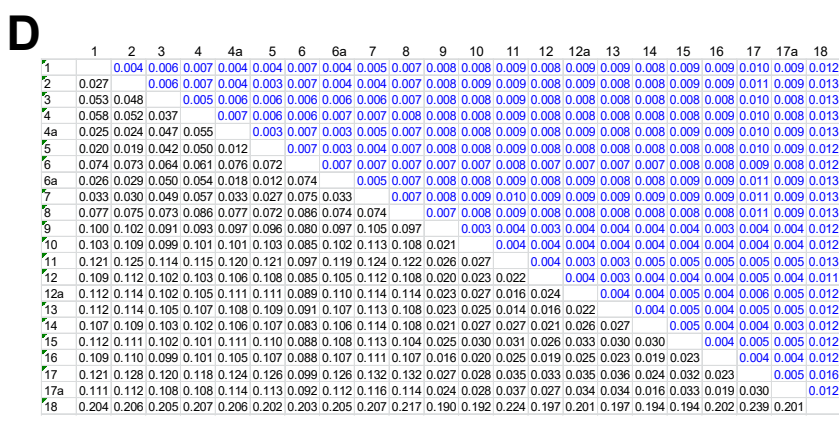
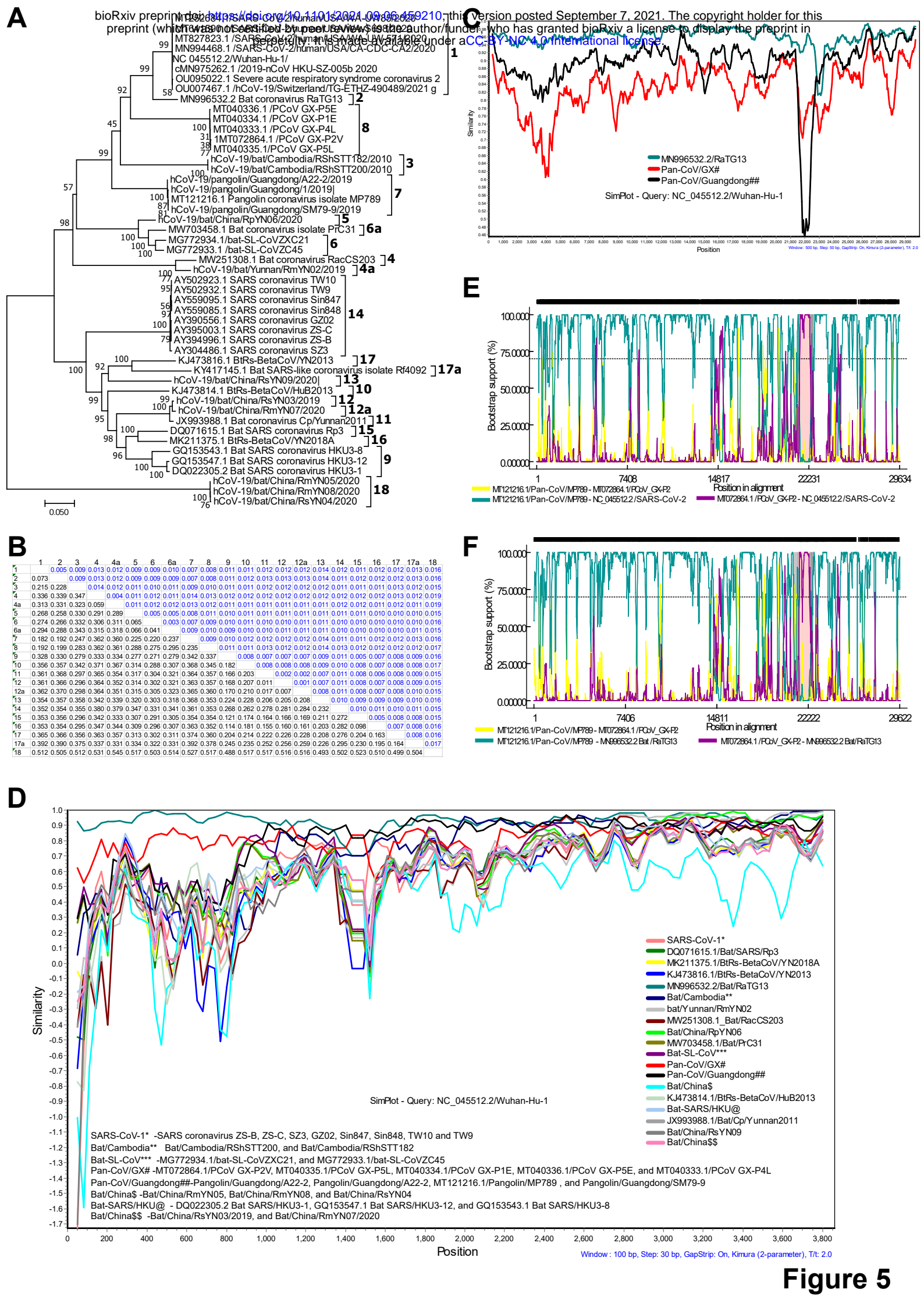


Figure 4
 SARS-CoV-1* -SARS coronavirus ZS-B, ZS-C, SZ3, GZ02, Sin847, Sin848, TW10 and TW9
 Bat/Cambodia** -Bat/Cambodia/RShSTT200, and Bat/Cambodia/RShSTT182
 Bat-SL-CoV*** -MG772934.1/bat-SL-CoVZXC21, and MG772933.1/bat-SL-CoVZC45
 Bat/China\$ -Bat/China/RmYN05, Bat/China/RmYN08, and Bat/China/RsYN04
 Bat-SARS/HKU@ - DQ0722305.2 Bat SARS/HKU3-1, GQ153547.1 Bat SARS/HKU3-12, and GQ153543.1 Bat SARS/HKU3-8
 Bat/China\$\$ - Bat/China/RsYN03/2019, and Bat/China/RmYN07/2020
 Pan-CoV/Guangdong## - Pangolin/Guangdong/A22-2, Pangolin/Guangdong/A22-2, MT121216.1/Pangolin/MP789, and Pangolin/Guangdong/SM79-9
 Pan-CoV/GX# -MT072864.1/PCoV GX-P2V, MT040335.1/PCoV GX-P5L, MT040334.1/PCoV GX-P1E, MT040336.1/PCoV GX-P5E, and MT040333.1/PCoV GX-P4L



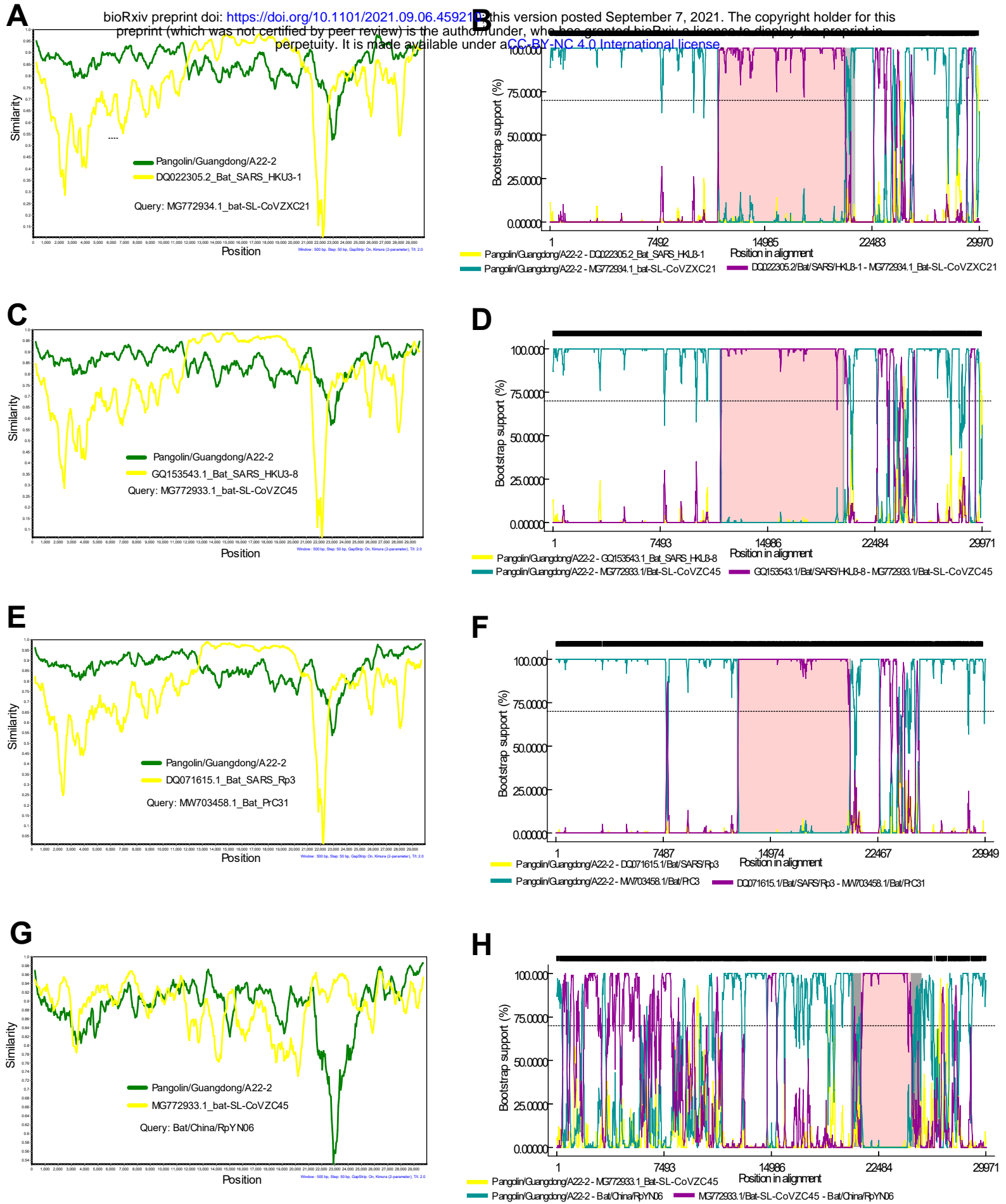


Figure 6

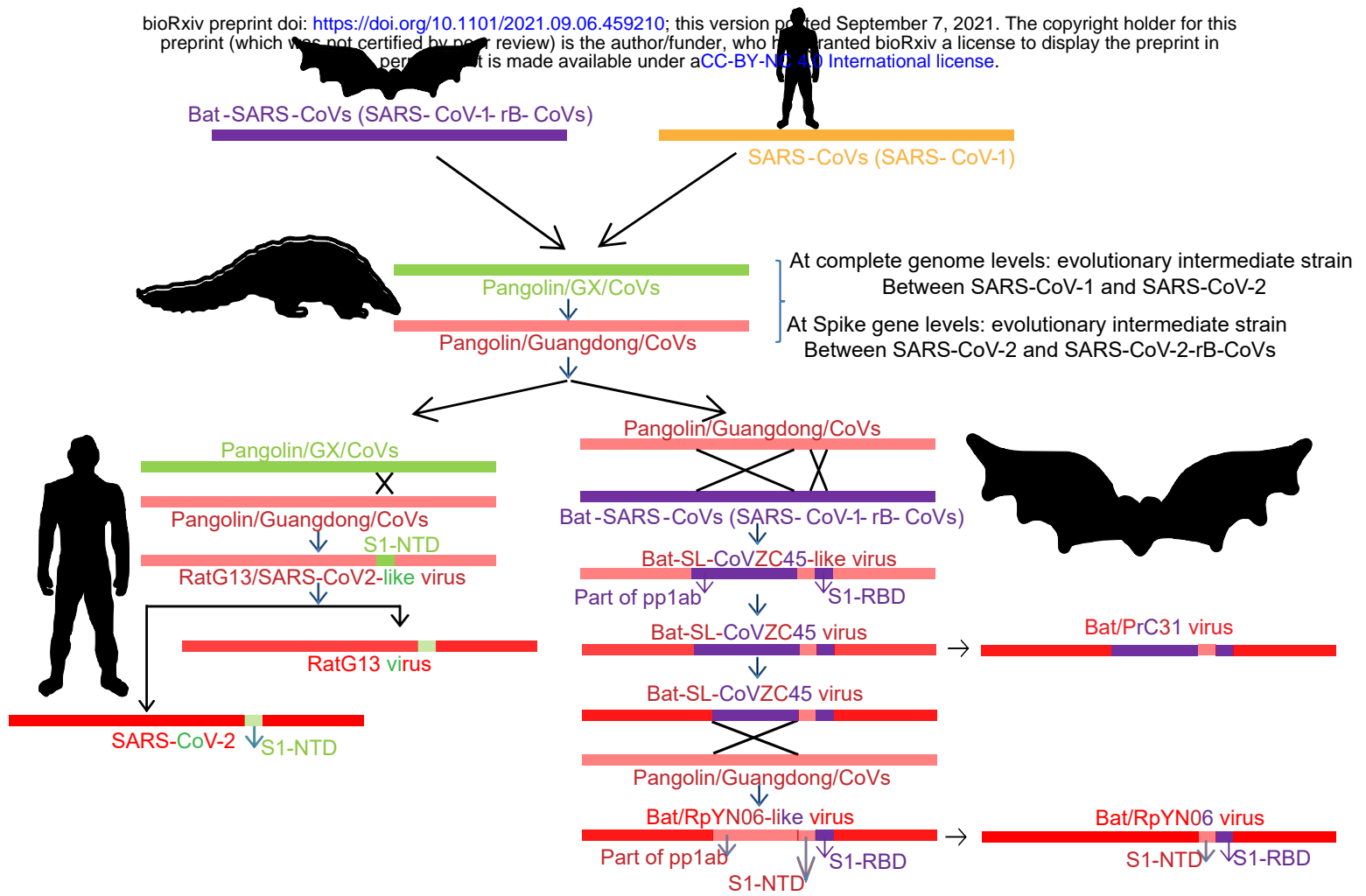
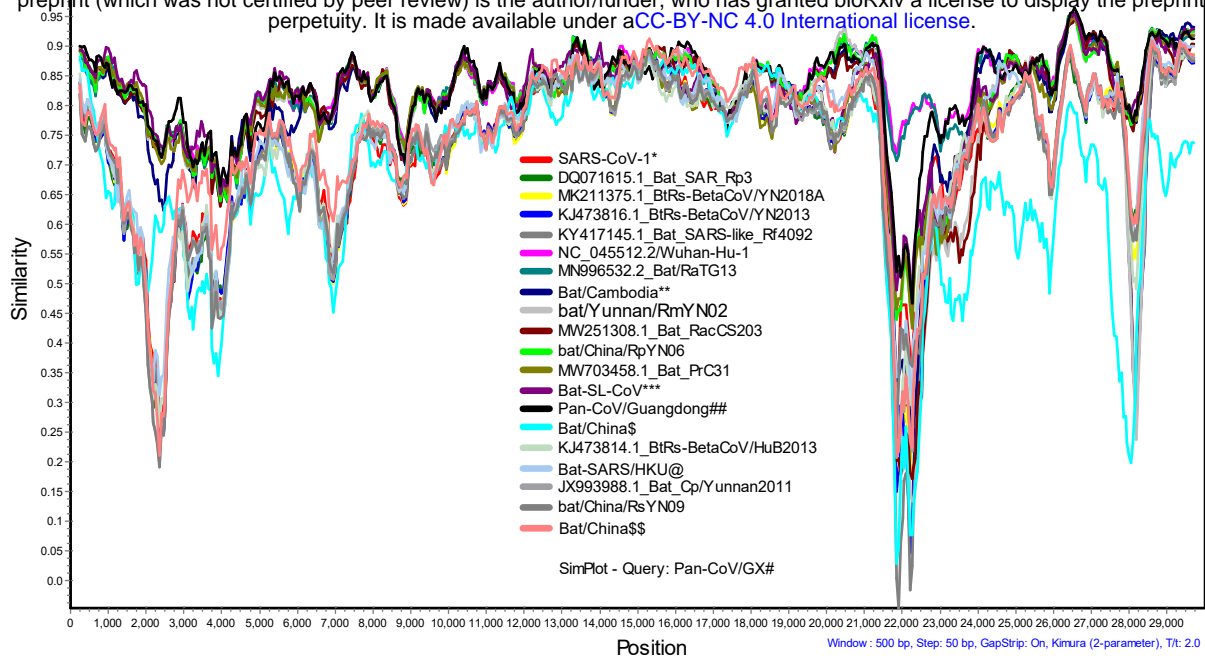
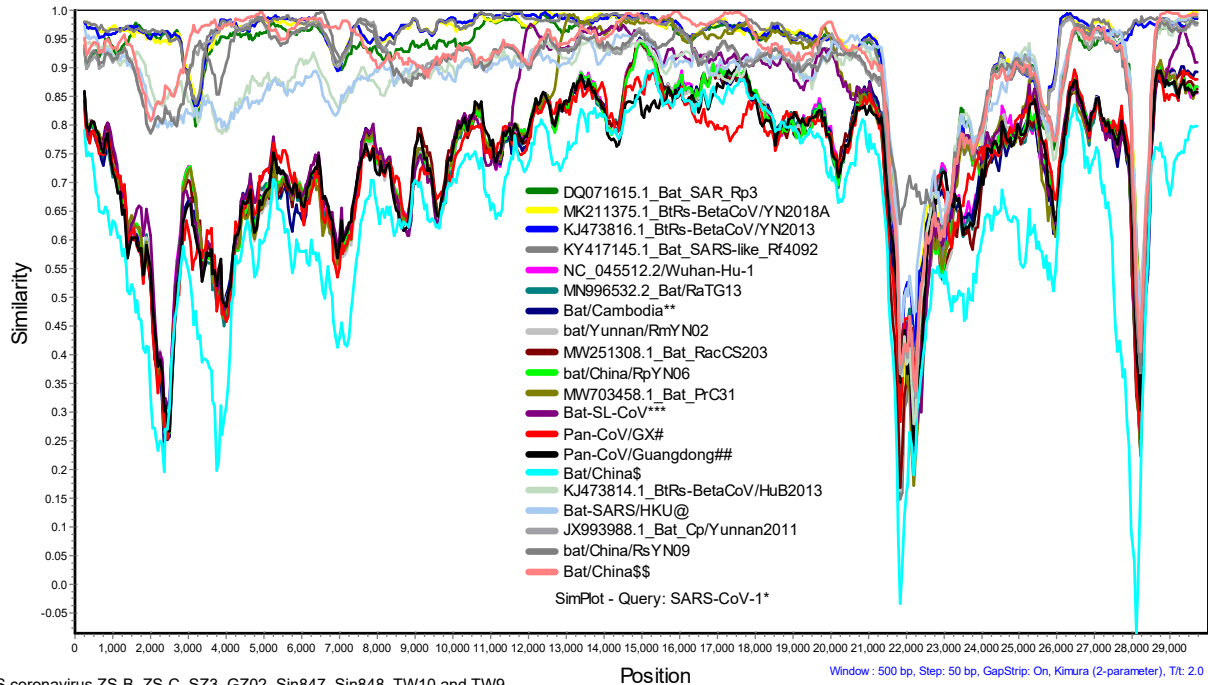


Figure 7

A



B



SARS-CoV-1* -SARS coronavirus ZS-B, ZS-C, SZ3, GZ02, Sin847, Sin848, TW10 and TW9

Bat/Cambodia** Bat/Cambodia/RShSTT200, and Bat/Cambodia/RShSTT182

Bat-SL-CoV*** -MG772934.1/bat-SL-CoVZXC21, and MG772933.1/bat-SL-CoVZC45

Pan-CoV/GX# -MT072864.1/PCoV GX-P2V, MT040335.1/PCoV GX-P5L, MT040334.1/PCoV GX-P1E, MT040336.1/PCoV GX-P5E, and MT040333.1/PCoV GX-P4L

Pan-CoV/Guangdong##-Pangolin/Guangdong/A22-2, Pangolin/Guangdong/A22-2, MT121216.1/Pangolin/MP789, and Pangolin/Guangdong/SM79-9

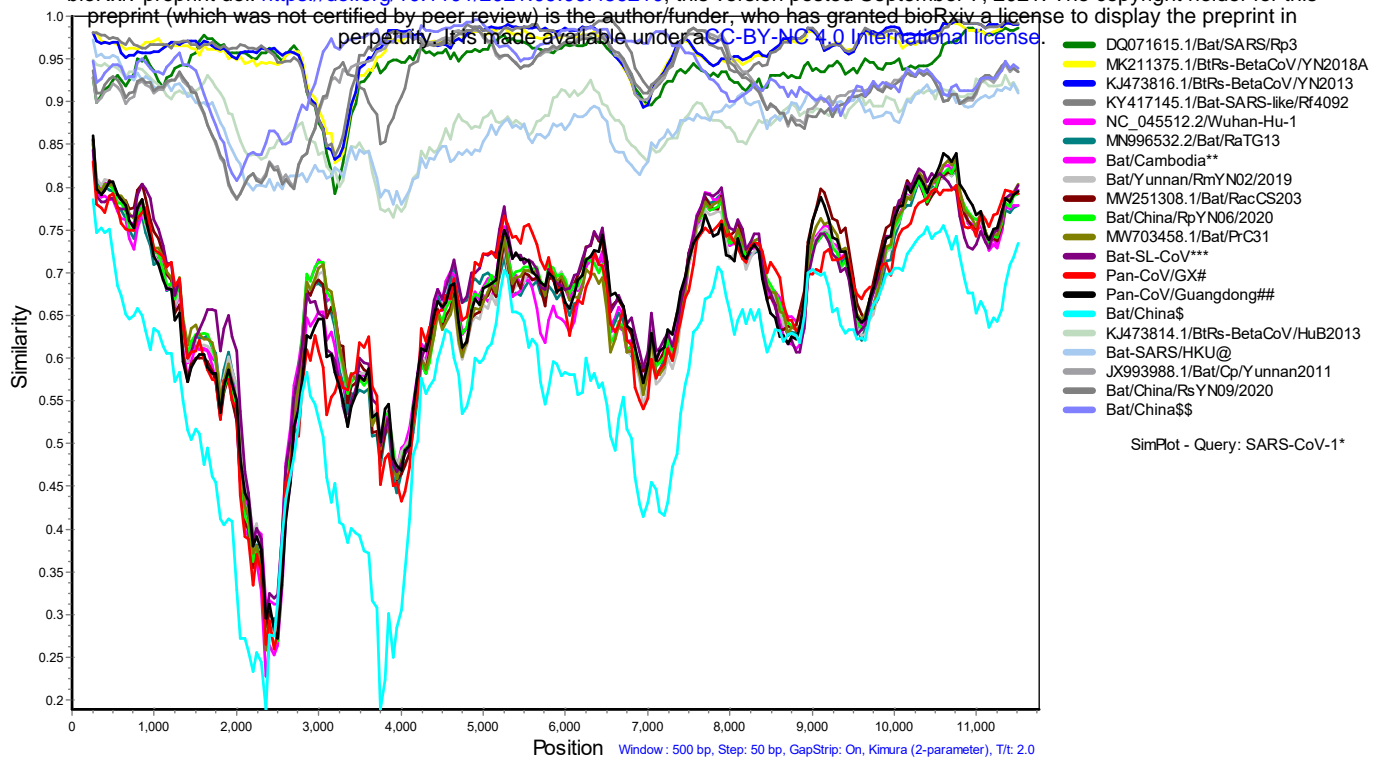
Bat/China\$ -Bat/China/RmYN05, Bat/China/RmYN08, and Bat/China/RsYN04

Bat-SARS/HKU@ - DQ022305.2 Bat SARS/HKU3-1, GQ153547.1 Bat SARS/HKU3-12, and GQ153543.1 Bat SARS/HKU3-8

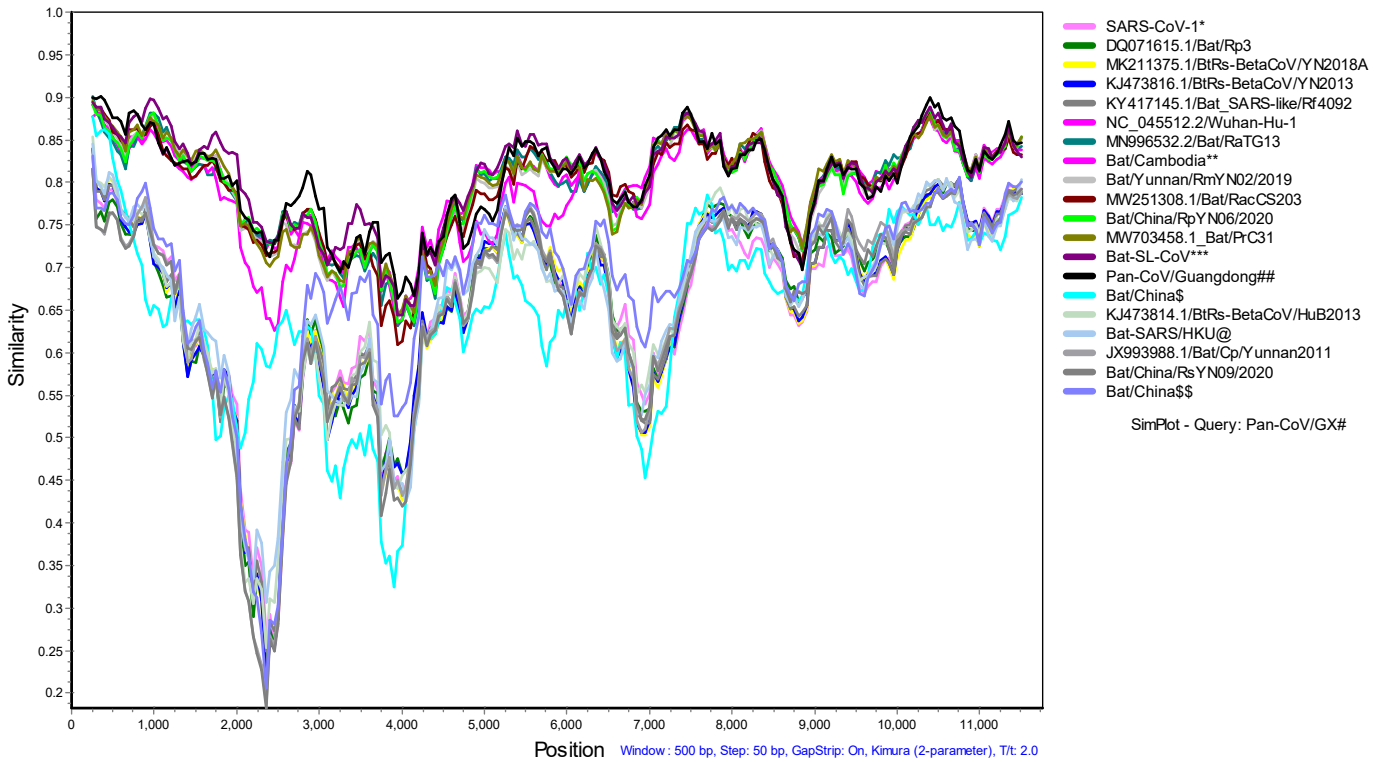
Bat/China\$\$ -Bat/China/RsYN03/2019, and Bat/China/RmYN07/2020

Supplementary Figure 1

A



B



SARS-CoV-1* -SARS coronavirus ZS-B, ZS-C, SZ3, GZ02, Sin847, Sin848, TW10 and TW9

Bat/Cambodia** -Bat/Cambodia/RShSTT200, and Bat/Cambodia/RShSTT182

Bat-SL-CoV*** -MG772934.1/bat-SL-CoVZXC21, and MG772933.1/bat-SL-CoVZC45

Pan-CoV/GX# -MT072864.1/PCoV GX-P2V, MT040335.1/PCoV GX-P5L, MT040334.1/PCoV GX-P1E, MT040336.1/PCoV GX-P5E, and MT040333.1/PCoV GX-P4L

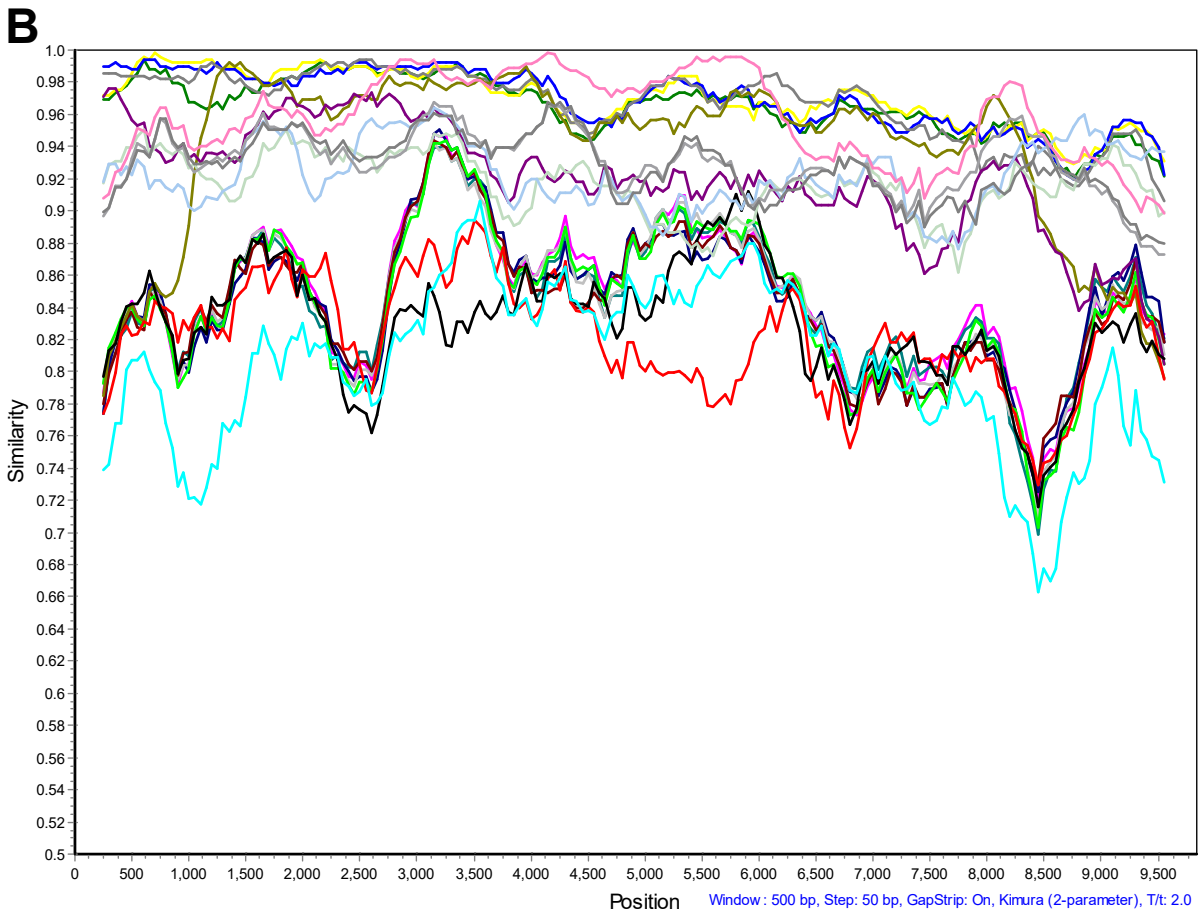
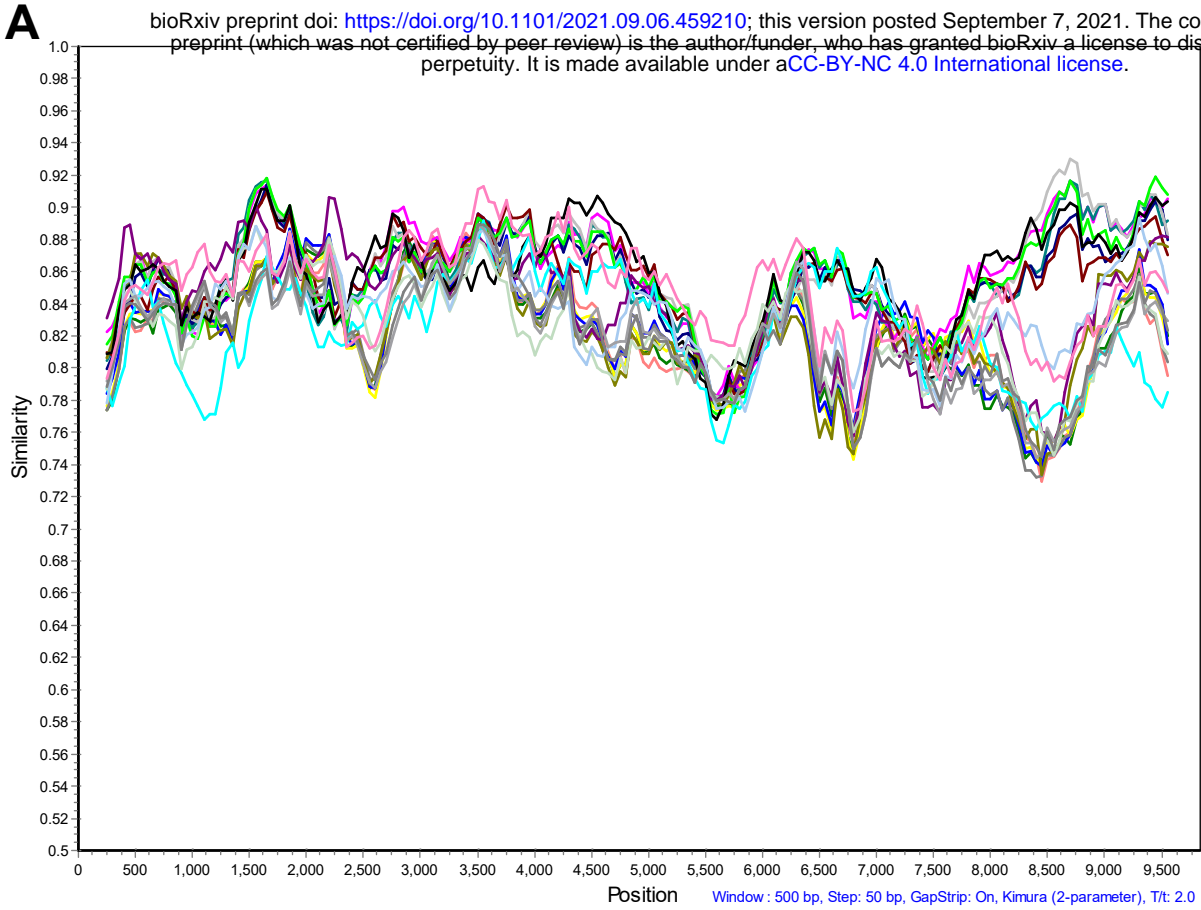
Pan-CoV/Guangdong## -Pangolin/Guangdong/A22-2, Pangolin/Guangdong/A22-2, MT121216.1/Pangolin/MP789, and Pangolin/Guangdong/SM79-9

Bat/China\$ -Bat/China/RmYN05, Bat/China/RmYN08, and Bat/China/RsYN04

Bat-SARS/HKU@ - DQ022305.2 Bat SARS/HKU3-1, GQ153547.1 Bat SARS/HKU3-12, and GQ153543.1 Bat SARS/HKU3-8

Bat/China\$\$ -Bat/China/RsYN03/2019, and Bat/China/RmYN07/2020

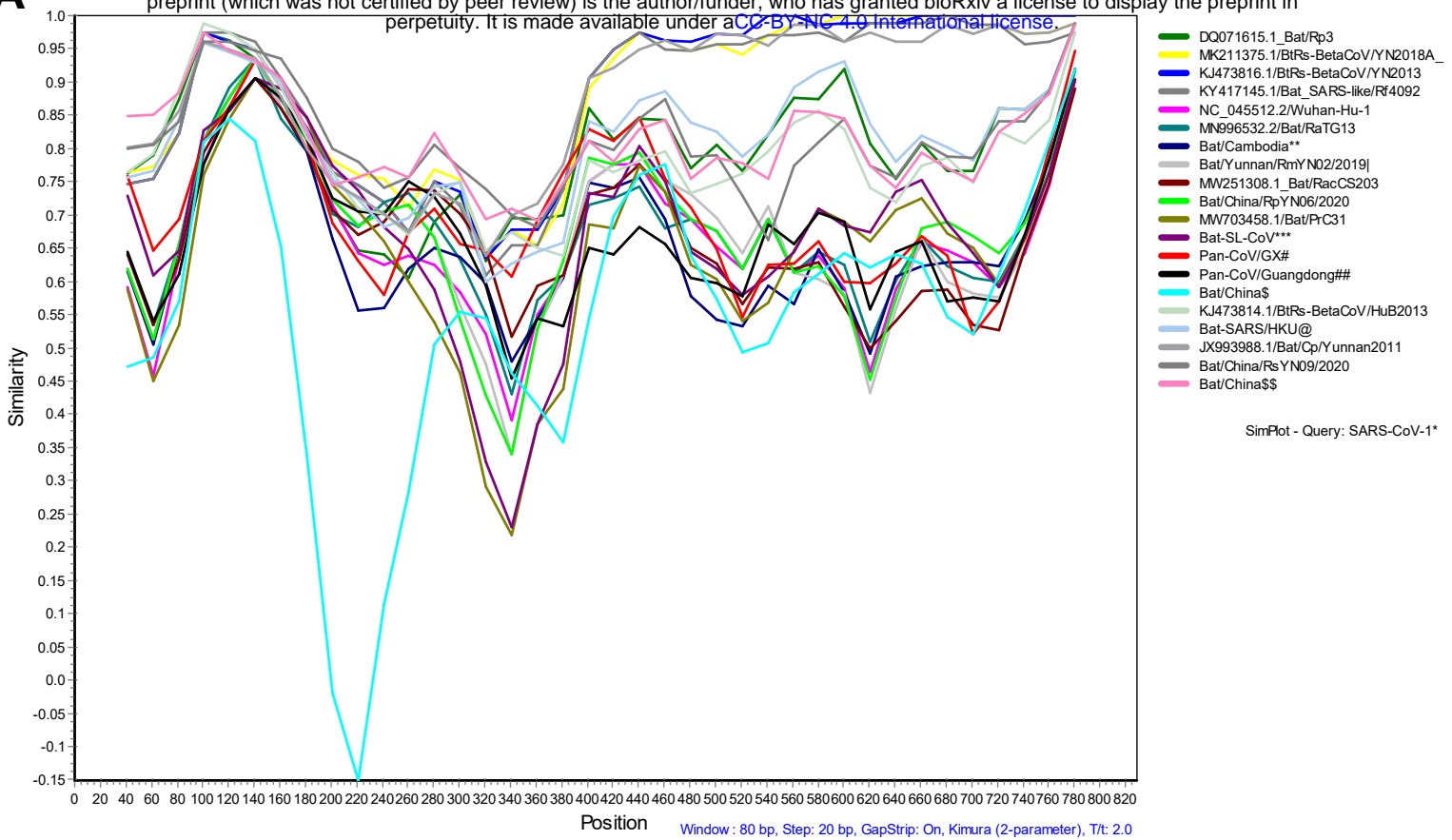
Supplementary Figure 2



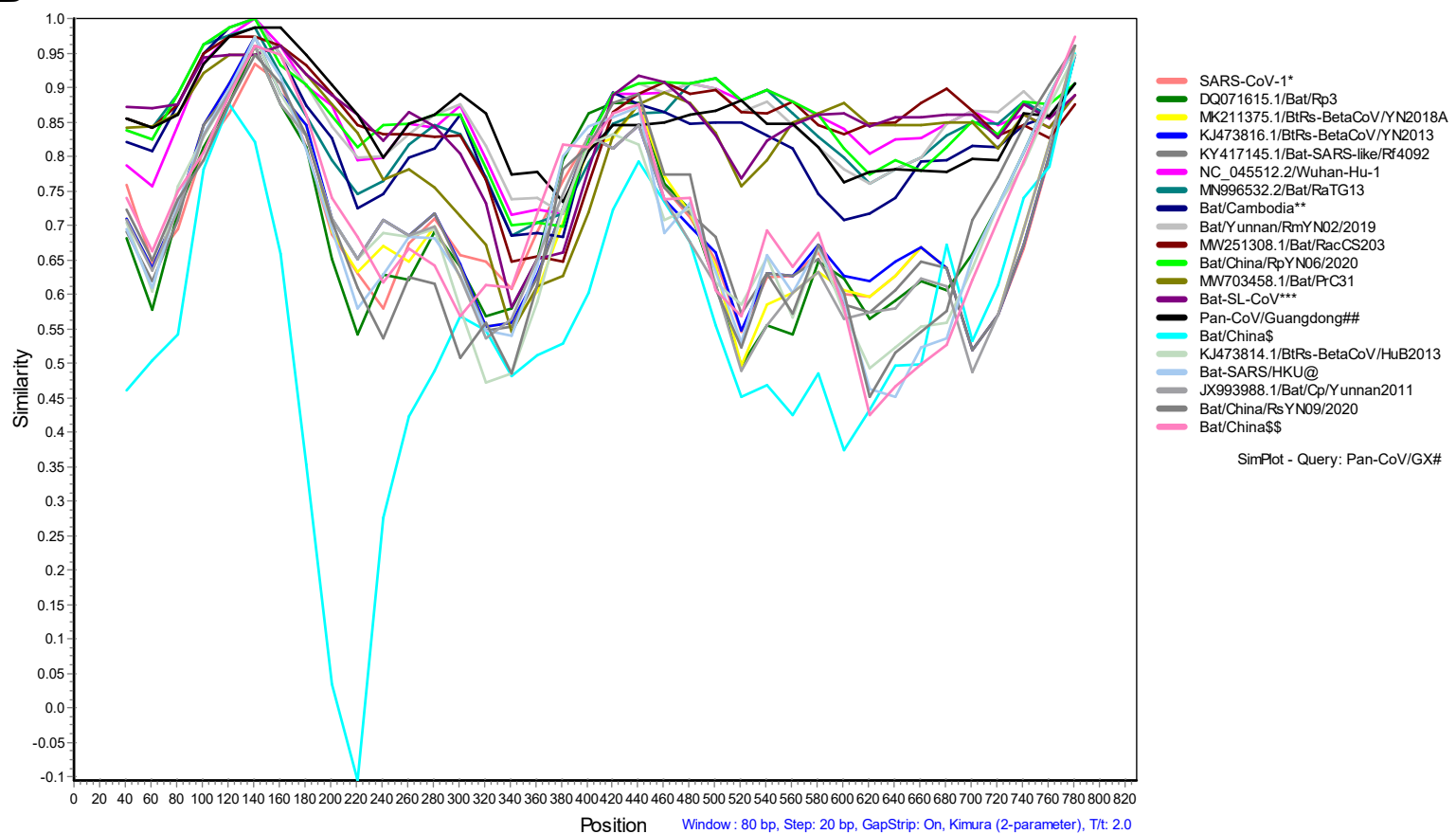
SARS-CoV-1* -SARS coronavirus ZS-B, ZS-C, SZ3, GZ02, Sin847, Sin848, TW10 and TW9
 Bat/Cambodia** -Bat/Cambodia/RShSTT200, and Bat/Cambodia/RShSTT182
 Bat-SL-CoV*** -MG772934.1/bat-SL-CoVZXC21, and MG772933.1/bat-SL-CoVZC45
 Pan-CoV/GX# -MT072864.1/PCoV GX-P2V, MT040335.1/PCoV GX-P5L, MT040334.1/PCoV GX-P1E, MT040336.1/PCoV GX-P5E, and MT040333.1/PCoV GX-P4L
 Pan-CoV/Guangdong## -Pangolin/Guangdong/A22-2, Pangolin/Guangdong/A22-2, MT121216.1/Pangolin/MP789 , and Pangolin/Guangdong/SM79-9
 Bat/China\$ -Bat/China/RmYN05, Bat/China/RmYN08, and Bat/China/RsYN04
 Bat-SARS/HKU@ - DQ022305.2 Bat SARS/HKU3-1, GQ153547.1 Bat SARS/HKU3-12, and GQ153543.1 Bat SARS/HKU3-8
 Bat/China\$\$ -Bat/China/RsYN03/2019, and Bat/China/RmYN07/2020

Supplementary Figure 3

A

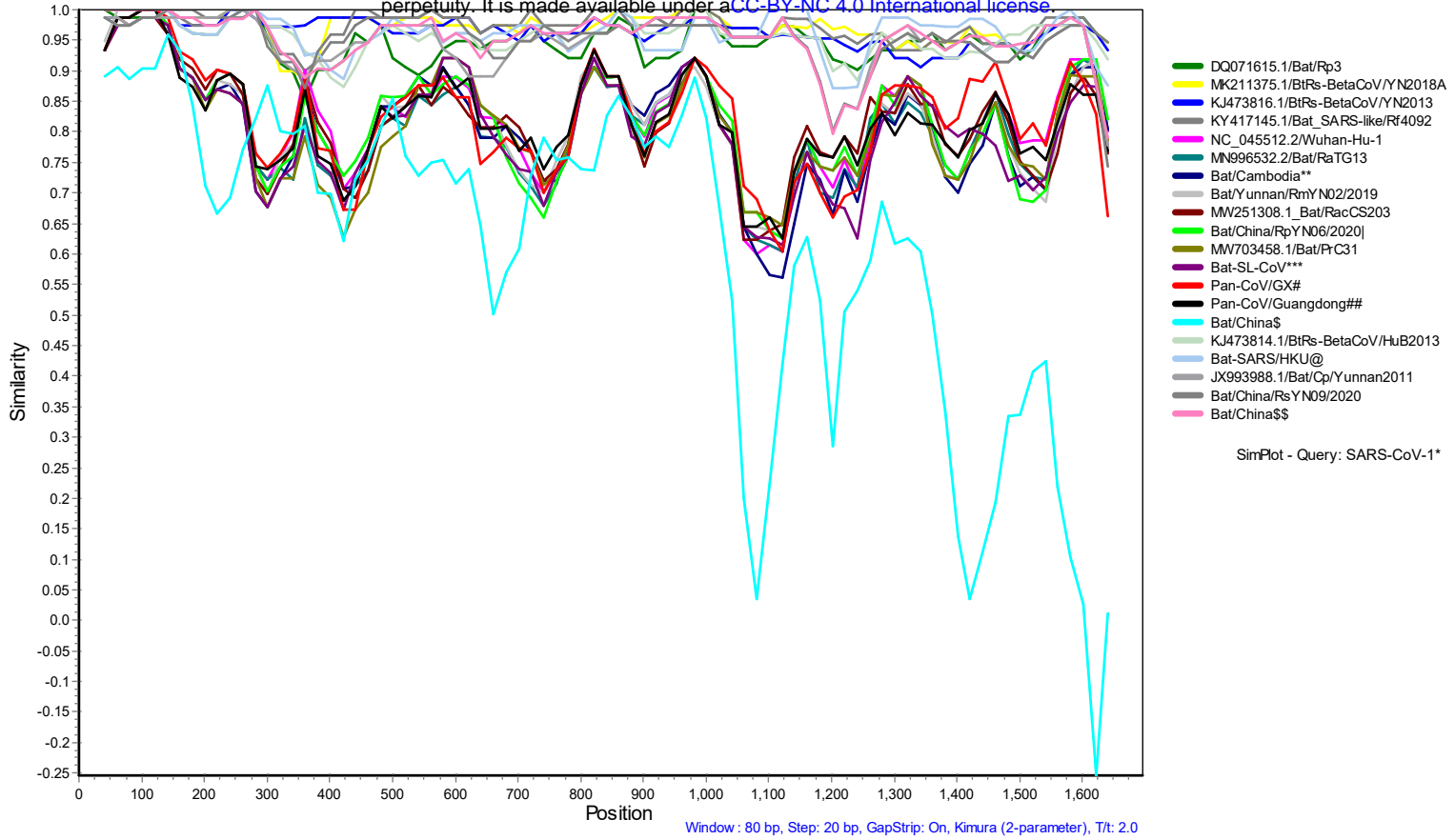


B

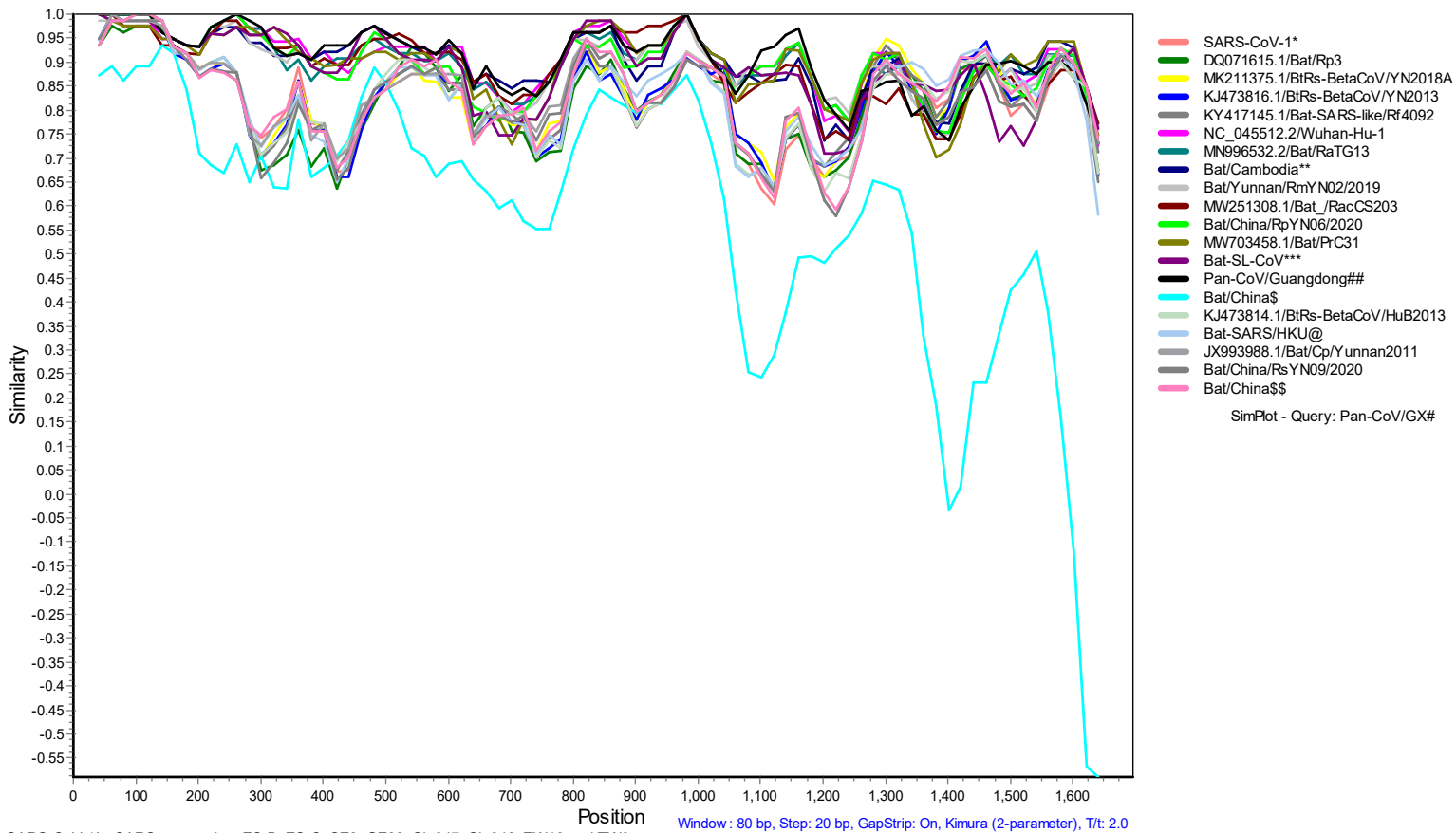


SARS-CoV-1* -SARS coronavirus ZS-B, ZS-C, SZ3, GZ02, Sin847, Sin848, TW10 and TW9
 Bat/Cambodia** -Bat/Cambodia/RShSTT200, and Bat/Cambodia/RShSTT182
 Bat-SL-CoV*** -MG772934.1/bat-SL-CoVZXC21, and MG772933.1/bat-SL-CoVZC45
 Pan-CoV/GX# -MT072864.1/PCoV GX-P2V, MT040335.1/PCoV GX-P5L, MT040334.1/PCoV GX-P1E, MT040336.1/PCoV GX-P5E, and MT040333.1/PCoV GX-P4L
 Pan-CoV/Guangdong## -Pangolin/Guangdong/A22-2, Pangolin/Guangdong/A22-2, MT121216.1/Pangolin/MP789, and Pangolin/Guangdong/SM79-9
 Bat/China\$ -Bat/China/RmYN05, Bat/China/RmYN08, and Bat/China/RsYN04
 Bat-SARS/HKU@ -DQ022305.2 Bat SARS/HKU3-1, GQ153547.1 Bat SARS/HKU3-12, and GQ153543.1 Bat SARS/HKU3-8
 Bat/China\$\$ -Bat/China/RsYN03/2019, and Bat/China/RmYN07/2020

A



B



SARS-CoV-1* -SARS coronavirus ZS-B, ZS-C, SZ3, GZ02, Sin847, Sin848, TW10 and TW9

Bat/Cambodia** -Bat/Cambodia/RShSTT200, and Bat/Cambodia/RShSTT182

Bat-SL-CoV*** -MG772934.1/bat-SL-CoVZXC21, and MG772933.1/bat-SL-CoVZC45

Pan-CoV/GX# -MT072864.1/PCoV GX-P2V, MT040335.1/PCoV GX-P5L, MT040334.1/PCoV GX-P1E, MT040336.1/PCoV GX-P5E, and MT040333.1/PCoV GX-P4L

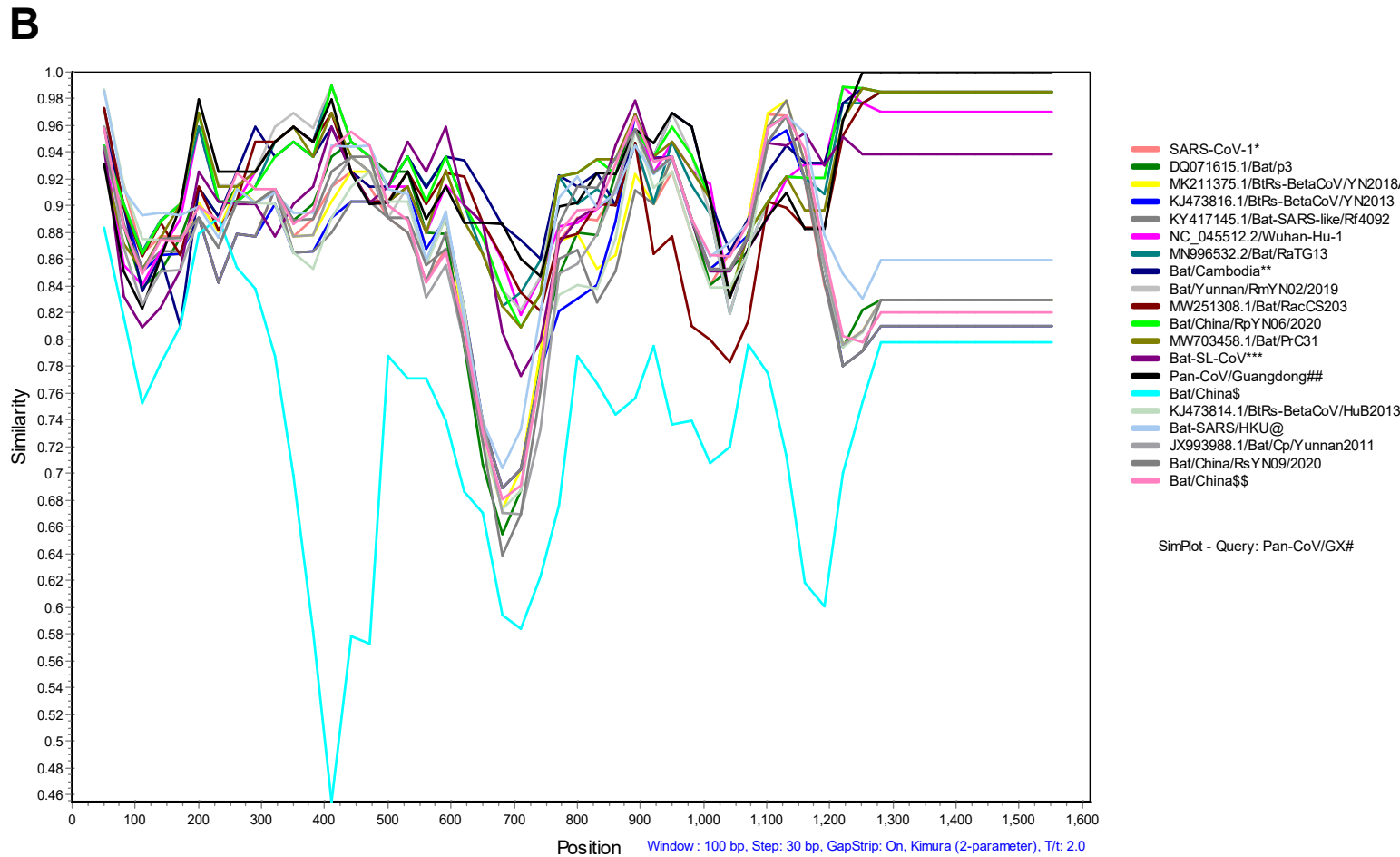
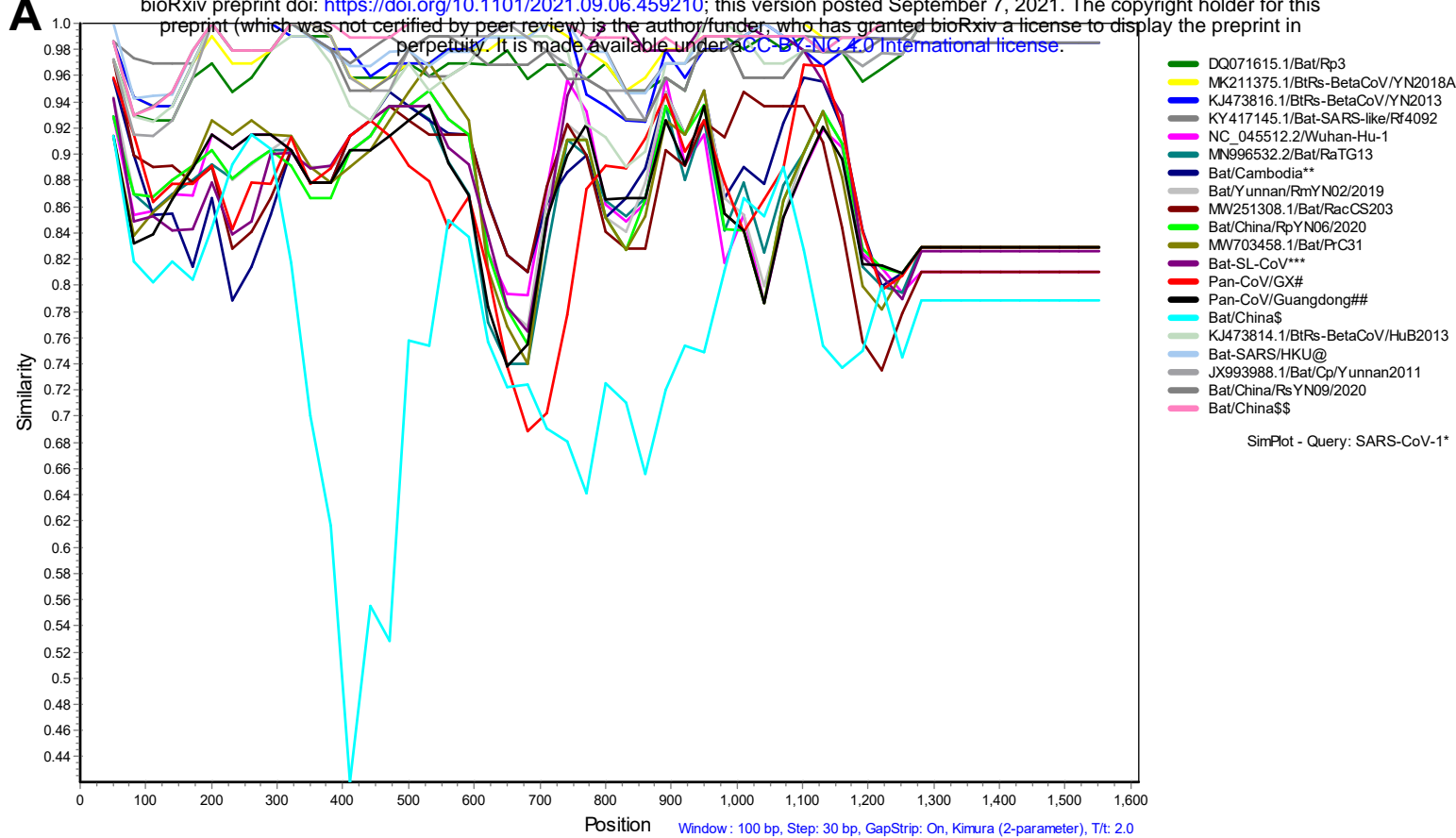
Pan-CoV/Guangdong## -Pangolin/Guangdong/A22-2, Pangolin/Guangdong/A22-2, MT121216.1/Pangolin/MP789, and Pangolin/Guangdong/SM79-9

Bat/China\$ -Bat/China/RmYN05, Bat/China/RmYN08, and Bat/China/RsYN04

Bat-SARS/HKU@ - DQ022305.2 Bat SARS/HKU3-1, GQ153547.1 Bat SARS/HKU3-12, and GQ153543.1 Bat SARS/HKU3-8

Bat/China\$\$ -Bat/China/RsYN03/2019, and Bat/China/RmYN07/2020

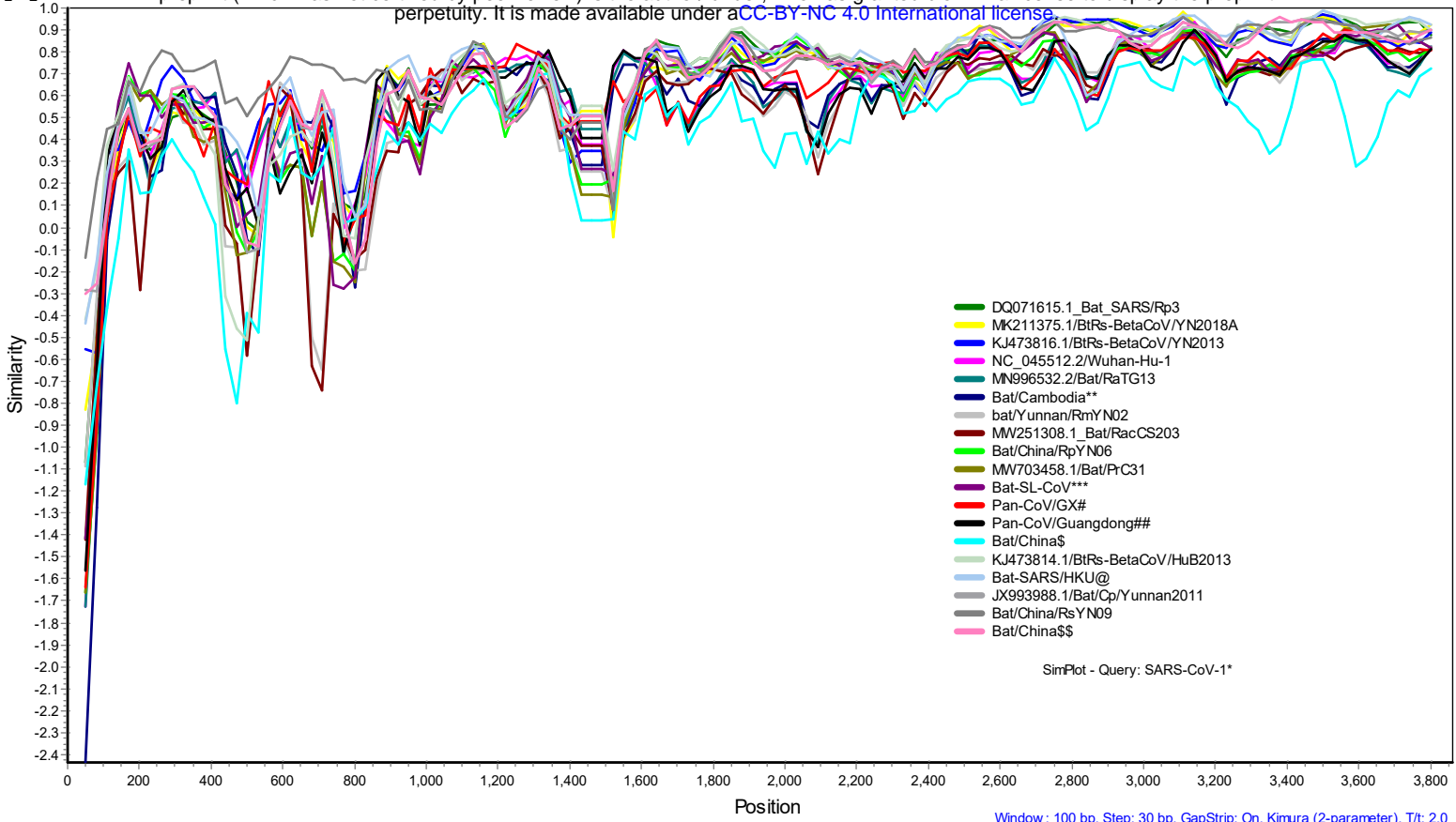
Supplementary Figure 5



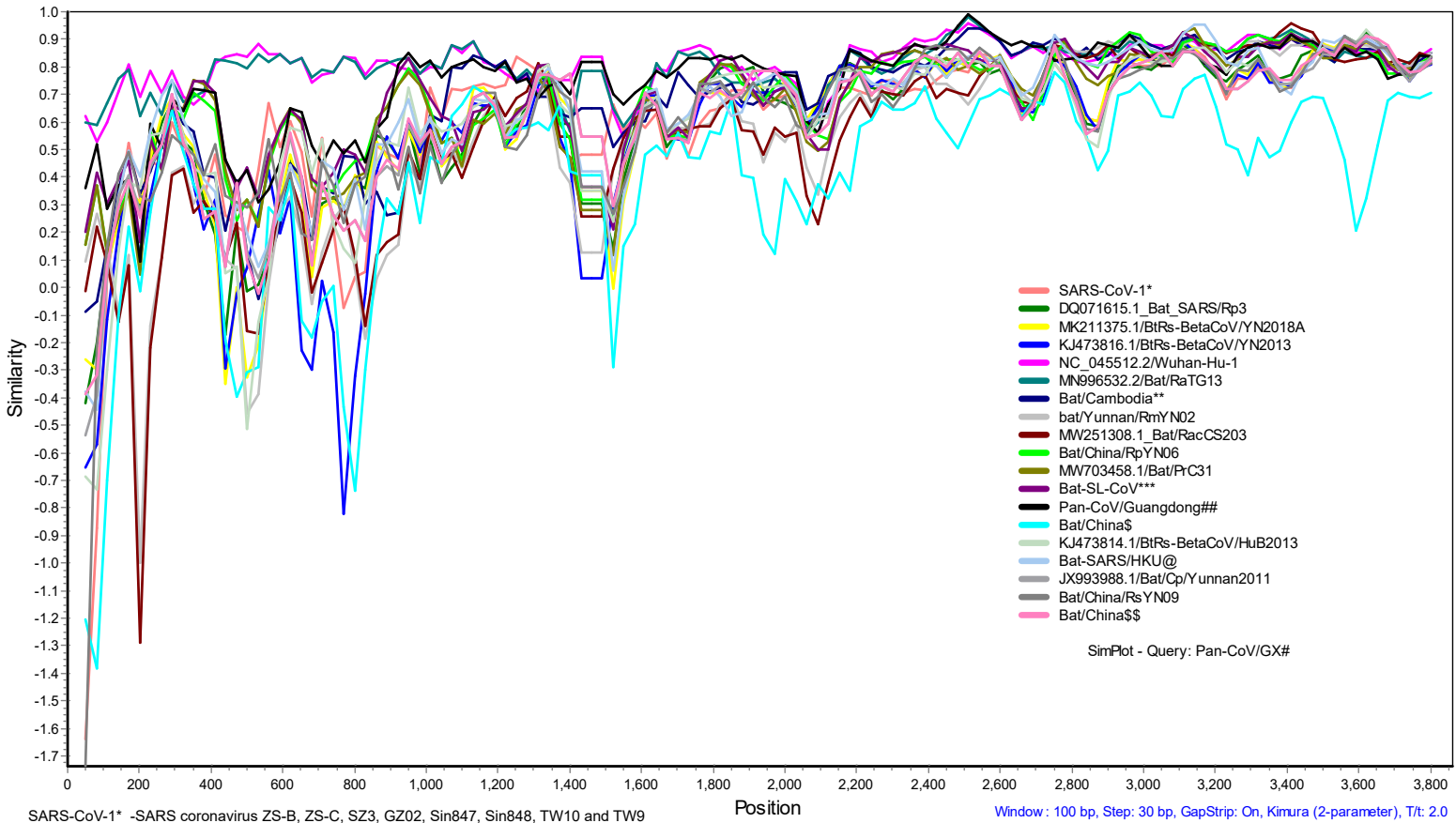
SARS-CoV-1* -SARS coronavirus ZS-B, ZS-C, SZ3, GZ02, Sin847, Sin848, TW10 and TW9
 Bat/Cambodia** -Bat/Cambodia/RShSTT200, and Bat/Cambodia/RShSTT182
 Bat-SL-CoV*** -MG772934.1/bat-SL-CoVZXC21, and MG772933.1/bat-SL-CoVZC45
 Pan-CoV/GX# -MT072864.1/PCoV GX-P2V, MT040335.1/PCoV GX-P5L, MT040334.1/PCoV GX-P1E, MT040336.1/PCoV GX-P5E, and MT040333.1/PCoV GX-P4L
 Pan-CoV/Guangdong##-Pangolin/Guangdong/A22-2, Pangolin/Guangdong/A22-2, MT121216.1/Pangolin/MP789, and Pangolin/Guangdong/SM79-9
 Bat/China\$ -Bat/China/RmYN05, Bat/China/RmYN08, and Bat/China/RsYN04
 Bat-SARS/HKU@ - DQ022305.2 Bat SARS/HKU3-1, GQ153547.1 Bat SARS/HKU3-12, and GQ153543.1 Bat SARS/HKU3-8
 Bat/China\$\$ -Bat/China/RsYN03/2019, and Bat/China/RmYN07/2020

Supplementary Figure 6

A



B



SARS-CoV-1* -SARS coronavirus ZS-B, ZS-C, SZ3, GZ02, Sin847, Sin848, TW10 and TW9

Bat/Cambodia** Bat/Cambodia/RShSTT200, and Bat/Cambodia/RShSTT182

Bat-SL-CoV*** -MG772934.1/bat-SL-CoVZXC21, and MG772933.1/bat-SL-CoVZC45

Pan-CoV/GX# -MT072864.1/PCoV GX-P2V, MT040335.1/PCoV GX-P5L, MT040334.1/PCoV GX-P1E, MT040336.1/PCoV GX-P5E, and MT040333.1/PCoV GX-P4L

Pan-CoV/Guangdong## -Pangolin/Guangdong/A22-2, Pangolin/Guangdong/A22-2, MT121216.1/Pangolin/MP789 , and Pangolin/Guangdong/SM79-9

Bat/China\$ -Bat/China/RmYN05, Bat/China/RmYN08, and Bat/China/RsYN04

Bat-SARS/HKU@ - DQ022305.2 Bat SARS/HKU3-1, GQ153547.1 Bat SARS/HKU3-12, and GQ153543.1 Bat SARS/HKU3-8

Bat/China\$\$ -Bat/China/RsYN03/2019, and Bat/China/RmYN07/2020

Supplementary Figure 7