

1 **Identification of 2019-nCoV related coronaviruses in Malayan pangolins in**
2 **southern China**

3

4 Tommy Tsan-Yuk Lam^{1,2}, Marcus Ho-Hin Shum², Hua-Chen Zhu^{1,2}, Yi-Gang Tong³, Xue-
5 Bing Ni², Yun-Shi Liao², Wei Wei⁴, William Yiu-Man Cheung², Wen-Juan Li³, Lian-Feng
6 Li⁴, Gabriel M Leung², Edward C. Holmes⁵, Yan-Ling Hu^{4*}, Yi Guan^{1,2*}

7

8 ¹ Joint Institute of Virology (Shantou University / The University of Hong Kong) &
9 Guangdong-Hongkong Joint Laboratory of Emerging Infectious Diseases, Shantou
10 University, Shantou, Guangdong, 515063, P. R. China.

11 ² State Key Laboratory of Emerging Infectious Diseases, School of Public Health, The
12 University of Hong Kong, Hong Kong SAR, P. R. China.

13 ³ Beijing Advanced Innovation Center for Soft Matter Science and Engineering (BAIC-SM),
14 College of Life Science and Technology, Beijing University of Chemical Technology,
15 Beijing, 100029, P. R. China.

16 ⁴ Life Sciences Institute, Guangxi Medical University, Nanning, Guangxi, 530021, P. R.
17 China.

18 ⁵ Marie Bashir Institute for Infectious Diseases and Biosecurity, School of Life and
19 Environmental Sciences and School of Medical Sciences, The University of Sydney, Sydney,
20 Australia.

21

22 * Corresponding authors:

23 Prof. Yi Guan, State Key Laboratory of Emerging Infectious Diseases, School of Public
24 Health, The University of Hong Kong, Hong Kong SAR, P. R. China; Email: yguan@hku.hk.

25 Prof. Yan-Ling Hu, Life Sciences Institute, Guangxi Medical University, Nanning, Guangxi
26 530021, China; Email: huyanling@gxmu.edu.cn.

27 **Abstract**

28 The ongoing outbreak of viral pneumonia in China and beyond is associated with a novel
29 coronavirus, provisionally termed 2019-nCoV. This outbreak has been tentatively associated
30 with a seafood market in Wuhan, China, where the sale of wild animals may be the source of
31 zoonotic infection. Although bats are likely reservoir hosts for 2019-nCoV, the identity of
32 any intermediate host facilitating transfer to humans is unknown. Here, we report the
33 identification of 2019-nCoV related coronaviruses in pangolins (*Manis javanica*) seized in
34 anti-smuggling operations in southern China. Metagenomic sequencing identified pangolin
35 associated CoVs that belong to two sub-lineages of 2019-nCoV related coronaviruses,
36 including one very closely related to 2019-nCoV in the receptor-binding domain. The
37 discovery of multiple lineages of pangolin coronavirus and their similarity to 2019-nCoV
38 suggests that pangolins should be considered as possible intermediate hosts for this novel
39 human virus and should be removed from wet markets to prevent zoonotic transmission.

40 Text

41 An outbreak of serious pneumonia disease was reported in Wuhan, China on 30 December
42 2019. The causative agent was soon identified as a novel coronavirus - 2019-nCoV¹. Case
43 numbers grew rapidly from 27 in December 2019 to 7,818 globally as of 30 January 2020²,
44 leading to the WHO to declare a public health emergency. Many of the early cases were
45 linked to Huanan seafood market in Wuhan city, Hubei province, from where the probable
46 zoonotic source is speculated to originate³. Currently, only environmental samples taken from
47 the market have been reported as positive for 2019-nCoV by the China CDC⁴. However, as
48 similar wet markets were already implicated in the SARS outbreak of 2002-2003⁵, it seems
49 likely that wild animals are also involved in the emergence of 2019-nCoV. Indeed, a number
50 of non-aquatic mammals were available for purchase in the Huanan seafood market prior to
51 the outbreak⁴. Unfortunately, because the market was cleared soon after the outbreak began,
52 determining the source virus in the wild animals from the market is challenging. Although a
53 coronavirus closely related to 2019-nCoV sampled from a *Rhinolophus affinis* bat in Yunnan
54 in 2013 has now been identified⁶, closely related viruses have not yet been detected in other
55 wildlife species. Here, we present the identification of 2019-nCoV related viruses in
56 pangolins smuggled into southern China.

57

58 We investigated the virome composition of pangolins (mammalian order Pholidota). These
59 animals are of growing importance and interest because they are the most illegally trafficked
60 of any group of mammal: they are used as both a food source and their scales are utilized in
61 traditional Chinese medicine. A number of pangolin species now are regarded as critically
62 endangered on the International Union for Conservation of Nature Red List of Threatened
63 Species. We received frozen tissue (lungs, intestine, blood) samples that were collected from
64 18 Malayan pangolins (*Manis javanica*) during August 2017-January 2018. These pangolins
65 were obtained during the anti-smuggling operations by Guangxi Customs. Strikingly, high-
66 throughput sequencing of their RNA revealed the presence of coronaviruses in six (two lung,
67 two intestine, one lung-intestine mix, one blood) of 43 samples. With the sequence read data,
68 and by filling gaps with amplicon sequencing, we were able to obtain six full or nearly full
69 genome sequences - denoted GX/P1E, GX/P2V, GX/P3B, GX/P4L, GX/P5E and GX/P5L -
70 that fall into the 2019-nCoV lineage (within the genus *Betacoronavirus*) in a phylogenetic
71 analysis (Figure 1a). These viruses also have similar genomic organization as 2019-nCoV,
72 with nine predicted open reading frames (Figure 1b; Extended Data Table S5). We were also
73 able to successfully isolate the virus using the Vero E6 cell line (Extended Data Figure S1).

74 Based on these new genome sequences, we designed primers for qPCR detection to confirm
75 that the raw samples were positive for the coronavirus. We conducted further qPCR testing
76 on another batch of archived pangolin samples collected between May-July 2018. Among the
77 19 samples (nine intestine tissues, ten lung tissues) tested from 12 animals, three lung tissue
78 samples were coronavirus positive.

79

80 In addition to the animals from Guangxi, after the start of the 2019-nCoV outbreak the
81 Guangzhou Customs Technology Center re-examined their five archived pangolin samples
82 (two skin swabs, one unknown tissue, one scale) obtained in previous anti-smuggling
83 operations in March 2019. Following high-throughput sequencing the scale sample was
84 found to contain coronavirus reads, and from these data we were able to assemble a partial
85 genome sequence of 21,505bp (denoted as GD/P2S), representing approximately 72% of the
86 2019-nCoV genome. This virus sequence on the scale may in fact come from contaminants of
87 other infected pangolin tissues. Notably, another study of diseased pangolins in Guangdong
88 performed in 2019 also identified viral contigs from lung samples that were similarly related
89 to 2019-nCoV⁷. Different assembly methods and manual curation were performed to generate
90 a partial genome sequence that comprised about 86.3% of the full-length virus genome
91 (denoted as GD/P1L in the Figure 1a tree).

92

93 These novel pangolin coronavirus genomes have approximately 85.5% to 92.4% similarity to
94 2019-nCoV, and represent two sub-lineages of 2019-nCoVs in the phylogenetic tree, one of
95 which (GD/P1L and GDP2S) is extremely closely related to 2019-nCoV (Figure 1; red
96 circles). It has previously been noted that members of the subgenus *Sarbecovirus* have
97 experienced widespread recombination⁸. In support of this, a recombination analysis
98 performed here revealed that bat coronaviruses ZC45 and ZCS21 are likely recombinants,
99 containing genome fragments from multiple SARS-CoV related lineages (genome regions 2,
100 5, 7) and 2019-nCoV related lineages including that from pangolins (regions 1, 3, 4, 6, 8).

101

102 More notable, however, was the observation of putative recombination signals between the
103 pangolins coronaviruses, bat coronaviruses RaTG13, and human 2019-nCoV (Figure 1c, d).
104 In particular, 2019-nCoV exhibits very high sequence similarity to the Guangdong pangolin
105 coronaviruses in the receptor-binding domain (RBD; 97.4% amino acid similarity; indicated
106 by red arrow in Figure 1c and Figure 2a), even though it is most closely related to bat
107 coronavirus RaTG13 in the remainder of the viral genome. Bat CoV RaTG and the human

108 2019-nCoV have only 89.2% amino acid similarity in RBD. Indeed, the Guangdong pangolin
109 coronaviruses and 2019-nCoV possess identical amino acids at the five critical residues of the
110 RBD, whereas RaTG13 only shares one amino acid with 2019-nCoV (residue 442, human
111 SARS-CoV numbering⁹). Interestingly, a phylogenetic analysis of synonymous sites alone in
112 the RBD revealed that the phylogenetic position of the Guangdong pangolin is consistent
113 with that in the remainder of the viral genome, rather than being the closest relative of 2019-
114 nCoV (Figure 2b). Hence, it is possible that the amino acid similarity between the RBD of
115 the Guangdong pangolin coronaviruses and 2019-nCoV is due to selectively-mediated
116 convergent evolution rather than recombination, although it is difficult to choose between
117 these scenarios on current data. Although the drivers of any convergent evolution are
118 unknown, its possible occurrence, as well as that of recombination, would further highlight
119 the role played by intermediate animal hosts in human virus emergence.

120

121 To date, pangolins are the only mammals other than bats documented to be infected by a
122 2019-nCoV related coronavirus. It is striking that two related lineages of CoVs are found in
123 pangolins and that both are also related to 2019-nCoV. This suggests that these animals may
124 be long-term reservoir hosts for these viruses, which is surprising as pangolins are solitary
125 animals with relatively small population sizes, reflecting their endangered status¹⁰. However,
126 it cannot be excluded that pangolins acquired their 2019-nCoV related viruses independently
127 from bats or another animal host. It is also notable that both lineages of pangolin
128 coronaviruses were obtained from trafficked Malayan pangolins, likely originating from
129 Southeast Asia, and there is a marked lack of knowledge of the viral diversity maintained by
130 this animal in regions where it is indigenous. Undoubtedly, the extent of virus transmission in
131 pangolin populations requires additional investigation, but the repeated occurrence of
132 infections with 2019-nCoV related coronaviruses in Guangxi and Guangdong provinces
133 suggests that this animal may be a potentially important host in coronavirus emergence.

134

135 Coronaviruses, including those related to 2019-nCoV, are clearly present in many wild
136 mammals in Asia^{5,6,7,11}. Although the epidemiology, pathogenicity, interspecies infectivity
137 and transmissibility of coronaviruses in pangolins remains to be studied, the data presented
138 here strongly suggests that handling these animals requires considerable caution, and that
139 their sale in wet markets should be strictly prohibited. Further surveillance on pangolins in
140 the natural environment in China and Southeast Asia are clearly needed to understand their
141 role in the emergence of 2019-nCoV and the risk of future zoonotic transmission.

142 **Methods**

143 **Ethics Statement**

144 The animals studied here were rescued and treated by the Guangxi Zhuang Autonomous
145 Region Terrestrial Wildlife Medical-aid and Monitoring Epidemic Diseases Research Center
146 under the ethics approval (wild animal treatment regulation No. [2011] 85). The samples
147 were collected following the procedure guideline (Pangolins Rescue Procedure, November,
148 2016).

149

150 **Sample collection, viral detection and sequencing of pangolins in Guangxi**

151 We received frozen tissue samples of 18 pangolins (*Manis javanica*) from Guangxi Medical
152 University, China, which were collected between August 2017 – January 2018. These
153 pangolins were seized by the Guangxi Customs during their routine anti-smuggling
154 operations. All animal individuals comprised samples from multiple organs including lungs,
155 intestine and blood, with the exception of six individuals for which only lung tissues were
156 available, five with mixed intestine and lung tissues only, one with intestine tissues only, and
157 one comprising two blood samples. Using the intestine-lung mixed sample we were able to
158 isolate a novel *Betacoronavirus* using the Vero-E6 cell line (Extended Data Figure S1). A
159 High Pure Viral RNA Kit (Roche, Switzerland) was used for RNA extraction on all 43
160 samples. For RNA sequencing, a sequencing library was constructed using an Ion Total
161 RNA-Seq Kit v2 (Thermo Fisher Scientific, MA, USA), and the library was subsequently
162 sequenced using an Ion Torrent S5 sequencer (Thermo Fisher Scientific). Reverse
163 Transcription was performed using an SuperScript III First-Strand Synthesis System for RT-
164 PCR (Thermo Fisher Scientific, MA, USA). DNA libraries were constructed using the
165 NEBNext Ultra II DNA Library Prep Kit and sequenced on a MiSeq sequencer. The NGS
166 QC Toolkit V2.3.3 was used to remove low-quality and short reads. Both BLASTn and
167 BLASTx were used to search against a local virus database, utilizing the data available at
168 NCBI/GenBank. Genome sequences were assembled using the CLC Genomic Workbench
169 v.9.0. To fill gaps in high throughput sequencing and obtain the whole viral genome
170 sequence, amplicon primers based on the bat SARS-like coronavirus ZC45 (GenBank
171 accession number MG772933) sequence were designed for amplicon-based sequencing.

172

173 A total of six samples (including the virus isolate) contained reads that matched a
174 *Betacoronavirus* (Extended Data Table S1). We obtained near complete genomes from these
175 samples (98%, compared to 2019-nCoV), with the virus genomes denoted as GX/P1E,

176 GX/P2V, GX/P3B, GX/P4L, GX/P5E and GX/P5L. Based on these genome sequences, we
177 designed primers for qPCR to confirm the positivity of the original tissue samples (Extended
178 Data Table S4). This revealed an original lung tissue sample that was also qPCR positive, in
179 addition to the six original samples with coronavirus reads. We further tested an addition 19
180 samples (nine intestine tissues, ten lung tissues), from 12 smuggled pangolins sampled
181 between May-July 2018 by the group from Guangxi Medical University. The genome
182 sequences of GX/P1E, GX/P2V, GX/P3B, GX/P4L, GX/P5E and GX/P5L were submitted to
183 GenBank and GISAID databases, their accession numbers will be available as soon as it is
184 generated.

185

186 **Sample collection, viral detection and sequencing of pangolins in Guangdong**

187 After the start of the 2019-nCoV outbreak, the Guangzhou Customs Technology Center re-
188 examined their five archived pangolin samples (two skin swabs, one unknown tissue, one
189 scale) obtained in anti-smuggling operations undertaken in March 2019. RNA was extracted
190 from all five samples (Qiagen, USA), and was subjected to high-throughput RNA sequencing
191 on the Illumina HiSeq platform by Vision medicals, Guangdong, China. The scale sample
192 was found to contain coronavirus reads using BLAST methods. These reads were quality
193 assessed, cleaned and assembled into contigs by both *de novo* (MEGAHIT v1.1.3¹²) and
194 using reference (BWA v0.7.13¹³) assembly methods, with BetaCoV/Wuhan/WIV04/2019 as
195 reference. The contigs were combined, and approximately 72% of the coronavirus genome
196 (21,505bp) was obtained. This sequence was denoted as pangolin CoV GD/P2S.

197

198 Liu *et al.* recently published a meta-transcriptomic study of pangolins⁷ and deposited 21
199 RNA-seq raw files on the SRA database (<https://www.ncbi.nlm.nih.gov/sra>). We screened
200 these raw read files using BLAST methods and found that five (SRR10168374,
201 SRR10168376, SRR10168377, SRR10168378 and SRR10168392) contained reads that
202 mapped to 2019-nCoV. These reads were subjected to quality assessment, cleaning and then
203 *de novo* assembly using MEGAHIT¹² and reference assembly using BWA¹³. These reads
204 were then merged and curated in a pileup alignment file to obtain the consensus sequences.
205 This combined consensus sequence is 25,753bp in length (about 86.3% of
206 BetaCoV/Wuhan/WIV04/2019) and denoted pangolin CoV GD/P1L. Notably, it has 66.8%
207 overlap and only 0.21% divergence (i.e. a sequence identity of 99.79%) with the GD/P2S
208 sequence. Since their genetic difference is so low, for the recombination analysis we merged

209 the GD/P1L and GD/P2S sequences into a single consensus sequence to minimize gap
210 regions within any sequences.

211

212 The viral genome organizations of Guangxi and Guangdong pangolin coronaviruses were
213 similar to 2019-nCoV. They had a total number of 9 open reading frames (ORFs) and shared
214 the same gene order of ORF1ab replicase, envelope glycoprotein spike (S), envelope (E),
215 membrane (M), nucleocapsid (N), plus other predicted ORFs. Detailed comparison of the
216 ORF length and similarity with 2019-nCoV and bat coronavirus RaTG13 is provided in
217 Extended Table S5.

218

219 **Sequence, phylogenetic and recombination analyses**

220 The human 2019-nCoV and bat RaTG13 coronavirus genome sequences were downloaded
221 from Virological.org (<http://virological.org>) and the GISAID (<https://www.gisaid.org>)
222 databases on 17 January 2020, with the data kindly shared by the submitters (Extended Data
223 Table S2). Other coronaviruses (subgenus *Sarbecovirus*) were downloaded from GenBank
224 (Extended Data Table S3) and compared to those obtained here. We constructed a multiple
225 sequence alignment of their complete genomes and individual genes using MAFFT v7.273¹⁴.
226 Maximum likelihood phylogenies were estimated using PhyML v3.1¹⁵, utilizing the
227 GTR+I+ Γ model of nucleotide substitution with 1,000 bootstrap replicates. To investigate
228 potential recombination events, we implemented a window sliding approach to determine the
229 changing patterns of sequence similarity and phylogenetic clustering between the query and
230 the reference sequences, as well as a scanning of phylogenetic clusters performed directly
231 from the multiple sequence alignment. Maximum likelihood trees were estimated from each
232 window extraction (i.e. genome regions 1 to 8) using PhyML as described above.

233 **References:**

- 234 1. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China.
235 *Nature* (2020). <https://doi.org/10.1038/s41586-020-2008-3>.
- 236 2. World Health Organization. WHO Novel Coronavirus (2019-nCoV) situation reports.
237 <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>
- 238 3. Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus:
239 implications for virus origins and receptor binding. *Lancet*, (2020). doi:10.1016/S0140-
240 6736(20)30251-8.
- 241 4. Cohen, J. Mining coronavirus genomes for clues to the outbreak's origins. *Science*,
242 (2020) [https://www.sciencemag.org/news/2020/01/mining-coronavirus-genomes-clues-](https://www.sciencemag.org/news/2020/01/mining-coronavirus-genomes-clues-outbreak-s-origins)
243 [outbreak-s-origins](https://www.sciencemag.org/news/2020/01/mining-coronavirus-genomes-clues-outbreak-s-origins).
- 244 5. Wang, M. *et al.* SARS-CoV infection in a restaurant from palm civet. *Emerg Infect Dis*
245 11:1860-5, (2005).
- 246 6. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat
247 origin. *Nature* (2020). <https://doi.org/10.1038/s41586-020-2012-7>.
- 248 7. Liu, P. *et al.* Viral Metagenomics Revealed Sendai Virus and Coronavirus Infection of
249 Malayan Pangolins (*Manis javanica*). *Viruses* 11, (2019).
- 250 8. Hon, C.C. *et al.* Evidence of the recombinant origin of a bat severe acute respiratory
251 syndrome (SARS)-like coronavirus and its implications on the direct ancestor of SARS
252 coronavirus. *J Virol* 82, 1819-1826, (2008).
- 253 9. Wan, Y. *et al.* Receptor recognition by novel coronavirus from Wuhan: An analysis
254 based on decade-long structural studies of SARS. *J Virol* , (2020).
- 255 10. Heinrich, S. *et al.* The global trafficking of pangolins: A comprehensive summary of
256 seizures and trafficking routes from 2010–2015. *Trafficking*, (2017).
- 257 11. Wang, W. *et al.* 2017. Discovery of a Highly Divergent Coronavirus in the Asian House
258 Shrew from China Illuminates the Origin of the Alphacoronaviruses. *J Virol* 91, (2017).
- 259 *References associated in Methods only:*
- 260 12. Li, D. *et al.* MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics
261 assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674-1676, (2015).
- 262 13. Li, H. and Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform.
263 *Bioinformatics* 26, 589-595, (2010).
- 264 14. Rozewicki, J. *et al.* MAFFT-DASH: integrated protein sequence and structural alignment.
265 *Nucleic Acids Res* 47, W5-W10, (2019).

- 266 15. Guindon, S. *et al.* PHYML Online--a web server for fast maximum likelihood-based
267 phylogenetic inference. *Nucleic Acids Res* 33, W557-559, (2010).

268

269 **Acknowledgements:**

270 We thank Prof. Wu-Chun Cao, Dr. Na Jia, Dr. Ya-Wei Zhang, Dr. Jia-Fu Jiang, Dr. Bao-Gui
271 Jiang, and their team in State Key Laboratory of Pathogen and Biosecurity, Beijing Institute
272 of Microbiology and Epidemiology, Beijing for their substantial contributions to this study,
273 including coordinating among research parties, conducting virus isolation, qPCR and
274 sequencing. We also thank Guangzhou Customs Technology Center, Guangzhou for re-
275 examining their archived pangolin samples and providing the data. All the above persons
276 agree on the content of the study, authorship arrangement and publication. We thank the staff
277 of the Guangxi and Guangdong Custom Bureau for their laborious anti-smuggling operations.
278 We thank all the scientists who kindly shared their genomic sequences of the coronaviruses
279 used in this study.

280

281 This work was supported by research grants from National Key Plan for Scientific Research
282 and Development of China (2016YFD0500302; 2017YFE0190800), funding for Guangdong-
283 Hongkong-Macau Joint Laboratory (2019B121205009), The National Natural Science
284 Foundation of China (NSFC) Excellent Young Scientists Fund (Hong Kong and Macau)
285 (31922087), Li Ka Shing Foundation, National Institutes of Health (HHSN272201400006C),
286 and the Australian Research Council (FL170100022).

287

288 **Author contributions:** Y.G. and Y.L.H. designed and supervised research. W.W., W.J.L.
289 and L.F.L. collected samples and conducted genome sequencing. M.H.H.S, X.B.N. and
290 T.T.Y.L. performed genome assembly and annotation. Y.G.T., T.T.Y.L., M.H.H.S, X.B.N,
291 W.Y.M.C. and Y.S.L. performed genome analysis and interpretation. T.T.Y.L. and
292 E.C.H. wrote the paper. H.C.Z., Y.L.H., G.M.L and Y.G. joined the data interpretation and
293 edited the paper.

294

295 **Competing interests:** No conflict of interest declared.

296

297 **Materials & correspondence:**

298 Correspondence and requests for materials should be addressed to Prof. Yi Guan
299 (yguan@hku.hk)

300

301 **Data availability:**

302 Data that support the findings of this study have been deposited in GenBank database with
303 the xxxx-xxxx accession codes and SRA database with xxxx-xxxx accession codes.

304 **Figure Legends**

305

306 **Figure 1.** Phylogenetic analyses depicting the evolutionary relationship between human
307 2019-nCoV, the pangolin coronavirus sequences obtained in this study, and the other
308 reference coronaviruses. The phylogenies were estimated using a maximum likelihood
309 approach employing the GTR+I+ Γ nucleotide substitution model and 1,000 bootstrap
310 replicates. (A) Phylogeny of the subgenus *Sarbecovirus* (genus *Betacoronavirus*) estimated
311 from the concatenated ORF1ab-S-E-M-N genes. Red circles indicate the pangolin
312 coronavirus sequences generated in this study. Note that GD/P1L is the consensus sequence
313 re-assembled from the raw data previously published⁷. (B) Genome organization of
314 coronaviruses including the pangolin coronaviruses, with the predicted ORFs shown in
315 different colors. (C) Sliding window analysis of changing patterns of sequence similarity
316 between 2019-nCoV, pangolin coronaviruses and bat coronavirus RaTG13. The name of the
317 query sequences are shown vertically on the right of the analysis boxes. The similarities to
318 different reference sequences are indicated by different colors shown in the legend box at the
319 top. Guangdong pangolin coronaviruses GD/P1L and GD/P2S were merged for this analysis.
320 The blue arrows at the top indicate the position of the ORFs in the alignment analyzed. The
321 potential recombination breakpoints are shown in pink dash lines, which together slice the
322 genomes into eight regions (regions with <200bp were omitted; indicated by the red line at
323 bottom) for phylogenetic analysis. (D) Phylogenetic trees of different genomic regions.
324 SARS-CoV and 2019-nCoV related lineages are shown in blue and red tree branches. Branch
325 scale bars are 0.1 substitutions/site.

326

327 **Figure 2.** (A) Sequence alignment showing the receptor-binding domain (RBD) in human,
328 pangolin and bat coronaviruses. The five critical residues for binding between SARS-CoV
329 RBD and human ACE2 protein are indicated in red boxes, and ACE2-contacting residues are
330 indicated with yellow boxes, following Wan et al.⁹. Note that in Guangdong pangolin
331 sequence, the codon positions coding for amino acid 337 proline, 420 aspartic acid, 499
332 proline and 519 asparagine have ambiguous nucleotide compositions, resulting to possibly
333 alternative amino acids threonine, glycine, threonine and lysine respectively. GD:
334 Guangdong, GX: Guangxi. (B) Phylogenetic trees of 2019-nCoV related lineage estimated
335 from the whole RBD region (upper) and synonymous sites only (lower). Branch supports
336 obtained from 1,000 bootstrap replicates are shown.

Figure 1

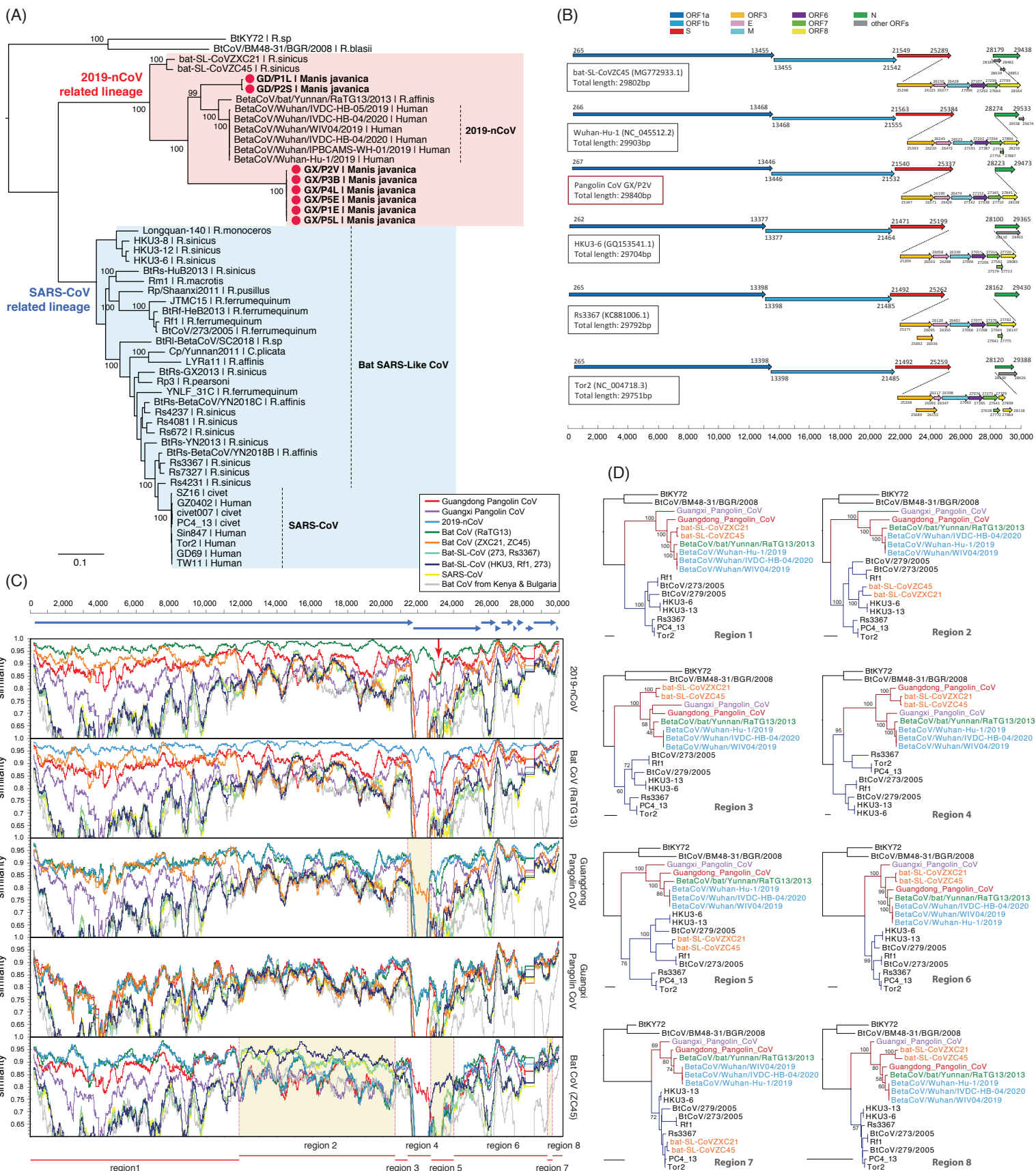
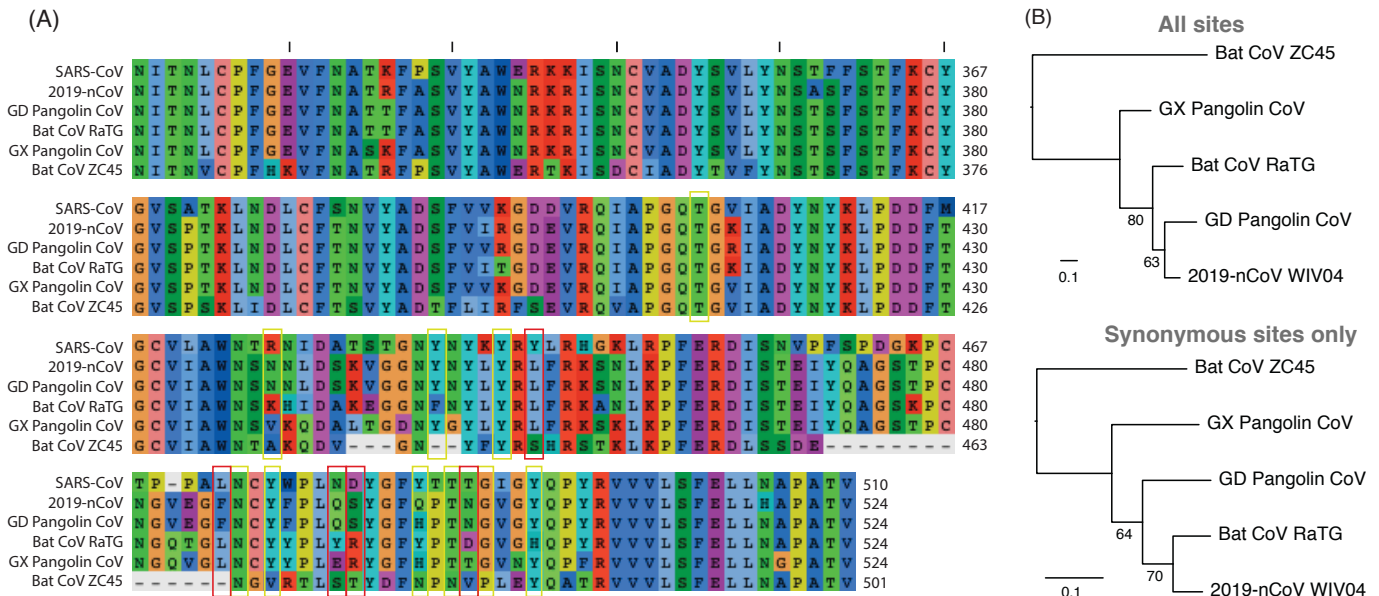


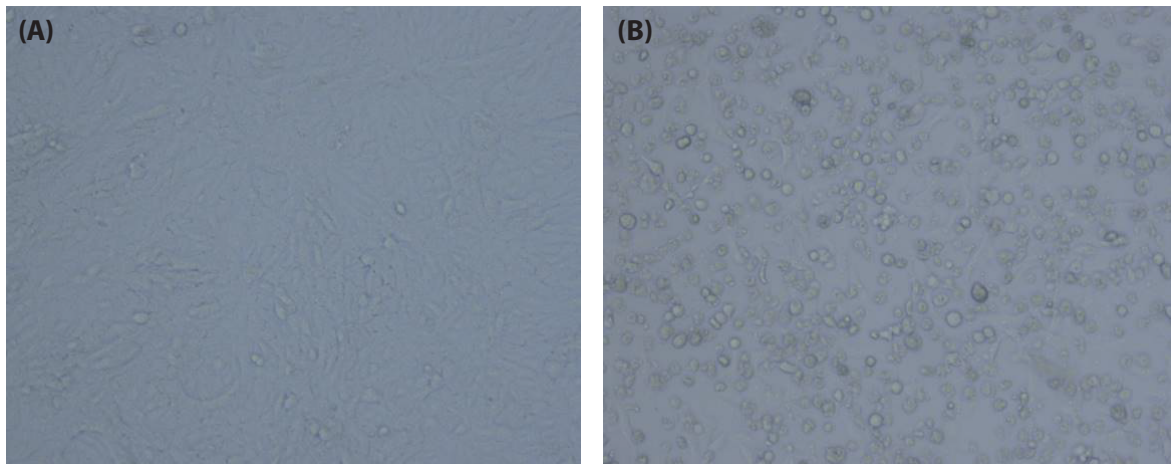
Figure 2



337 **Extended Data**

338

339 **Figure S1.** Microscopic image of the cytopathic effect in virus isolation using Vero E6. (A)
340 Negative control of Vero E6 cell line. (B) Cytopathic effect seen in viral culture (5 days post
341 inoculation).



342

343 **Table S1.** High-throughput sequencing results of the samples with coronavirus reads

Source location	Animal	Sample type	Sample number	Sequencing raw data ID
Guangxi	Pangolin	Intestine	GX/P1E	Data submission in process; Identifier will be available as soon as it is generated.
Guangxi	Pangolin	Virus isolate from intestine-lung mixed samples	GX/P2V	Data submission in process; Identifier will be available as soon as it is generated.
Guangxi	Pangolin	Blood	GX/P3B	Data submission in process; Identifier will be available as soon as it is generated.
Guangxi	Pangolin	Lung	GX/P4L	Data submission in process; Identifier will be available as soon as it is generated.
Guangxi	Pangolin	Intestine	GX/P5E	Data submission in process; Identifier will be available as soon as it is generated.
Guangxi	Pangolin	Lung	GX/P5L	Data submission in process; Identifier will be available as soon as it is generated.
Guangdong	Pangolin	Scale	GD/P2S	Data submission in process; Identifier will be available as soon as it is generated.

344

345 **Table S2.** Acknowledgement of sharing of 2019-nCoV genome sequences from the
 346 Virological.org and the GISAID databases. We gratefully thank the authors listed below for
 347 sharing their genomic sequences of coronaviruses analyzed in this study.

Accession ID	Virus name	Location	Collection date	Originating lab	Submitting lab	Authors
Virological.org sequence (NC_045512.2)	BetaCoV/Wuhan- Hu-1/2019	China / Wuhan	2019-12	National Institute for Communicable Disease Control and Prevention (ICDC) Chinese Center for Disease Control and Prevention (China CDC)	National Institute for Communicable Disease Control and Prevention (ICDC) Chinese Center for Disease Control and Prevention (China CDC)	Zhang,Y.-Z., Wu,F., Chen,Y.-M., Pei,Y.-Y., Xu,L., Wang,W., Zhao,S., Yu,B., Hu,Y., Tao,Z.-W., Song,Z.-G., Tian,J.-H., Zhang,Y.-L., Liu,Y., Zheng,J.-J., Dai,F.-H., Wang,Q.-M., She,J.-L. and Zhu,T.-Y.
EPI_ISL_ 402131	BetaCoV/bat/ Yunnan/ RaTG13/2013	China / Yunnan Province / Pu'er City	2013-07- 24	Wuhan Institute of Virology, Chinese Academy of Sciences	Wuhan Institute of Virology, Chinese Academy of Sciences	Yan Zhu, Ping Yu, Bei Li, Ben Hu, Hao-Rui Si, Xing-Lou Yang, Peng Zhou, Zheng-Li Shi
EPI_ISL_ 402121	BetaCoV/Wuhan/ IVDC-HB- 05/2019	China / Hubei Province / Wuhan City	2019-12- 30	National Institute for Viral Disease Control and Prevention, China CDC	National Institute for Viral Disease Control and Prevention, China CDC	Wenjie Tan, Xuejun Ma, Xiang Zhao, Wenling Wang, Yongzhong Jiang, Roujian Lu, Ji Wang, Peihua Niu, Weimin Zhou, Faxian Zhan, Weifeng Shi, Baoying Huang, Jun Liu, Li Zhao, Yao Meng, Fei Ye, Na Zhu, Xiaozhou He, Peipei Liu, Yang Li, Jing Chen, Wenbo Xu, George F. Gao, Guizhen Wu
EPI_ISL_ 402120	BetaCoV/Wuhan/ IVDC-HB- 04/2020	China / Hubei Province / Wuhan City	2020-01- 01	National Institute for Viral Disease Control and Prevention, China CDC	National Institute for Viral Disease Control and Prevention, China CDC	Wenjie Tan, Xiang Zhao, Wenling Wang, Xuejun Ma, Yongzhong Jiang, Roujian Lu, Ji Wang, Weimin Zhou, Peihua Niu, Peipei Liu, Faxian Zhan, Weifeng Shi, Baoying Huang, Jun Liu, Li Zhao, Yao

						Meng, Xiaozhou He, Fei Ye, Na Zhu, Yang Li, Jing Chen, Wenbo Xu, George F. Gao, Guizhen Wu
EPI_ISL_ 402124	BetaCoV/Wuhan/ WIV04/2019	China / Hubei Province / Wuhan City	2019-12- 30	Wuhan Jinyintan Hospital	Wuhan Institute of Virology, Chinese Academy of Sciences	Peng Zhou, Xing-Lou Yang, Ding-Yu Zhang, Lei Zhang, Yan Zhu, Hao-Rui Si, Zhengli Shi
EPI_ISL_ 402123	BetaCoV/Wuhan /IPBCAMS-WH- 01/2019	China / Hubei Province / Wuhan City	2019-12- 24	Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College	Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College	Lili Ren, Jianwei Wang, Qi Jin, Zichun Xiang, Zhiqiang Wu, Chao Wu, Yiwei Liu

348

349 **Table S3.** GenBank accession numbers of coronavirus sequences used in this study.

Accession ID	Strain name	Host
NC_004718.3	Tor2	Human
AY313906.1	GD69	Human
MK211377.1	BtRs-BetaCoV/YN2018C	<i>R. affinis</i>
MK211376.1	BtRs-BetaCoV/YN2018B	<i>R. affinis</i>
MK211374.1	BtRI-BetaCoV/SC2018	<i>R. sp</i>
KY352407.1	BtKY72	<i>R. sp</i>
MG772934.1	bat-SL-CoVZXC21	<i>R. sinicus</i>
MG772933.1	bat-SL-CoVZC45	<i>R. sinicus</i>
KY417151.1	Rs7327	<i>R. sinicus</i>
KY417147.1	Rs4237	<i>R. sinicus</i>
KY417146.1	Rs4231	<i>R. sinicus</i>
KY417143.1	Rs4081	<i>R. sinicus</i>
KJ473816.1	BtRs-YN2013	<i>R. sinicus</i>
KJ473815.1	BtRs-GX2013	<i>R. sinicus</i>
KJ473814.1	BtRs-HuB2013	<i>R. sinicus</i>
KJ473812.1	BtRf-HeB2013	<i>R. ferrumequinum</i>
JX993988.1	Cp/Yunnan2011	<i>C. plicata</i>
JX993987.1	Rp/Shaanxi2011	<i>R. pusillus</i>
KU182964.1	JTMC15	<i>R. ferrumequinum</i>
KP886808.1	YNLF_31C	<i>R. ferrumequinum</i>
KF569996.1	LYRa11	<i>R. affinis</i>
KC881006.1	Rs3367	<i>R. sinicus</i>
DQ412043.1	Rm1	<i>R. macrotis</i>
DQ412042.1	Rf1	<i>R. ferrumequinum</i>
GU190215.1	BtCoV/BM48-31/BGR/2008	<i>R. blasii</i>
GQ153547.1	HKU3-12	<i>R. sinicus</i>
GQ153543.1	HKU3-8	<i>R. sinicus</i>
GQ153541.1	HKU3-6	<i>R. sinicus</i>
FJ588686.1	Rs672	<i>R. sinicus</i>
DQ071615.1	Rp3	<i>R. pearsoni</i>
AY304488.1	SZ16	civet
DQ648856.1	BtCoV/273/2005	<i>R. ferrumequinum</i>
AY572034.1	civet007	civet
AY502924.1	TW11	Human
AY613948.1	PC4_13	civet
AY613947.1	GZ0402	Human
AY559095.1	Sin847	Human
KF294457.1	Longquan-140	<i>R. monoceros</i>
DQ648857.1	BtCoV/279/2005	<i>R. macrotis</i>

350

351 **Table S4.** Primers used for qPCR detection of pangolin associated coronavirus

pCov-Forward	AGGTGACGAGGTTAGACAAATAG
pCov-Reverse	CCAAGCAATAACACAACCAGTAA
pCov-Probe	ACCCGGACAAACTGGTGTTATTGCT

352

353 **Table S5.** Genomic comparison of 2019-nCoV with Bat-Cov RaTG13, Guangdong pangolin
 354 CoV and Guangxi pangolin CoV.

	Bat-Cov RaTG13 [#]			Guangdong pangolin CoV [#]			Guangxi pangolin CoV [#]		
	Length bat/2019- nCoV(bp)	nt Identity %	aa Identity %	Length GD/2019- nCoV(bp)	nt Identity %	aa Identity %	Length GX/2019- nCoV (bp)	nt Identity %	aa Identity %
ORF1ab	21287/21290	96.5	98.6	20076*/21290	90.8	97.1	21266/21290	84.8	92.5
S	3852/3864	92.9	97.7	3648/3864	89.3	90.7	3804/3864	83.5	92.5
ORF3a	828/828	96.3	97.8	828/828	93.4	97.4	828/828	86.9	89.3
E	228/228	99.6	100	228/228	98.3	100	228/228	97.4	100
M	693/693	95.7	99.5	693/693	93.1	98.6	693/693	91.6	98.2
ORF6	186/186	98.4	100	186/186	94.6	96.6	186/186	90.9	95
ORF7a	366/366	95.6	97.5	366/366	93.4	97.5	366/366	86.6	87.7
ORF8	366/366	96.9	94.9	366/366	92.1	94.9	366/366	81.5	86.8
N	1260/1260	96.8	98.8	1260/1260	96.1	97.6	1254/1260	90.9	94.1

355 [#]: Wuhan-Hu-1 2019-nCoV (NC_045512.2) was used for comparison with Bat-CoV RaTG13

356 (EPI_ISL_402131), Guangdong pangolin CoV (merged of GD/P1L and GD/P2S), and Guangxi

357 pangolin CoV (GX/P5L)

358 * partial sequence