



**HAL**  
open science

## Bat coronaviruses related to SARS-CoV-2 and infectious for human cells

Sarah Temmam, Khamsing Vongphayloth, Eduard Baquero, Sandie Munier, Massimiliano Bonomi, Béatrice Regnault, Bounsavane Douangboubpha, Yasaman Karami, Delphine Chrétien, Daosavanh Sanamxay, et al.

### ► To cite this version:

Sarah Temmam, Khamsing Vongphayloth, Eduard Baquero, Sandie Munier, Massimiliano Bonomi, et al.. Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature*, 2022, 604 (7905), pp.330-336. 10.1038/s41586-022-04532-4 . pasteur-03661247

**HAL Id: pasteur-03661247**

**<https://pasteur.hal.science/pasteur-03661247>**

Submitted on 19 May 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **Bat coronaviruses related to SARS-CoV-2 and infectious for human cells**

2 Sarah Temmam\* (1,2), Khamsing Vongphayloth\* (3), Eduard Baquero Salazar\* (4), Sandie Munier\* (5),  
3 Massimiliano Bonomi\* (6), Béatrice Regnault (1,2), Bounsavane Douangboubpha (7), Yasaman Karami  
4 (6), Delphine Chrétien (1,2), Daosavanh Sanamxay (7), Vilakhan Xayaphet (7), Phetphoumin  
5 Paphaphanh (7), Vincent Lacoste (3), Somphavanh Somlor (3), Khaithong Lakeomany (3), Nothasin  
6 Phommavanh (3), Philippe Pérot (1,2), Océane Dehan (5,8), Faustine Amara (5), Flora Donati (5,8),  
7 Thomas Bigot (1,9), Michael Nilges (6), Félix A. Rey (4), Sylvie van der Werf (5,8), Paul T. Brey (3), Marc  
8 Eloit + (1,2,10).

9 \* equally contributing authors

10 + corresponding author: Institut Pasteur, Université de Paris, Pathogen Discovery Laboratory, 25-28  
11 rue du Dr. Roux, 75015, Paris, France [marc.eloit@pasteur.fr](mailto:marc.eloit@pasteur.fr)

12

13

14 1) Institut Pasteur, Université de Paris, Pathogen Discovery Laboratory, 25-28 rue du Dr. Roux, 75015,  
15 Paris, France.

16 2) Institut Pasteur, Université de Paris, The OIE Collaborating Center for the detection and  
17 identification in humans of emerging animal pathogens, 25-28 rue du Dr. Roux, 75015, Paris, France

18 3) Institut Pasteur du Laos, Samsenthai Road, Ban Kao-Gnot, Sisattanak District, 3560 Vientiane, Lao  
19 PDR

20 4) Institut Pasteur, Université de Paris, CNRS UMR 3569, Structural Virology Unit, 25-28 rue du Dr.  
21 Roux, 75015, Paris, France.

22 5) Institut Pasteur, Université de Paris, CNRS UMR 3569, Molecular Genetics of RNA Viruses Unit, Paris,  
23 France

24 6) Institut Pasteur, Université de Paris, CNRS UMR 3528, Structural Bioinformatics Unit, Paris, France

25 7) Faculty of Environmental Sciences, National University of Laos, Dong Dok Campus, P.O. Box: 7322,  
26 Xaythany District, Vientiane Capital, Lao PDR

27 8) Institut Pasteur, Université de Paris, National Reference Center for Respiratory Viruses, 25-28 rue  
28 du Dr. Roux, 75015, Paris, France.

29 9) Institut Pasteur, Université de Paris, Bioinformatic and Biostatistic Hub - Computational Biology  
30 Department, 25-28 rue du Dr. Roux, 75015, Paris, France.

31 10) Ecole Nationale Vétérinaire d'Alfort, University of Paris-Est, 94704 Maisons-Alfort, France

32



## 33 Summary

34 The animal reservoir of SARS-CoV-2 is unknown despite reports of various SARS-CoV-2-related viruses  
35 in Asian *Rhinolophus* bats<sup>1,2,3,4</sup>, including the closest virus from *R. affinis*, RaTG13<sup>5,6</sup> and in pangolins<sup>7-</sup>  
36 <sup>9</sup>. SARS-CoV-2 presents a mosaic genome, to which different progenitors contribute. The spike  
37 sequence determines the binding affinity and accessibility of its receptor-binding domain (RBD) to the  
38 cellular angiotensin-converting enzyme 2 (ACE2) receptor and is responsible for host range<sup>10-12</sup>. SARS-  
39 CoV-2 progenitor bat viruses genetically close to SARS-CoV-2 and able to enter human cells through a  
40 human ACE2 pathway have not yet been identified, though they would be key in understanding the  
41 origin of the epidemics. Here we show that such viruses indeed circulate in cave bats living in the  
42 limestone karstic terrain in North Laos, within the Indochinese peninsula. We found that the RBDs of  
43 these viruses differ from that of SARS-CoV-2 by only one or two residues at the interface with ACE2,  
44 bind more efficiently to the hACE2 protein than the SARS-CoV-2 Wuhan strain isolated in early human  
45 cases, and mediate hACE2-dependent entry and replication in human cells, which is inhibited by  
46 antibodies neutralizing SARS-CoV-2. None of these bat viruses harbors a furin cleavage site in the spike.  
47 Our findings therefore indicate that bat-borne SARS-CoV-2-like viruses potentially infectious for  
48 humans circulate in *Rhinolophus* spp. in the Indochinese peninsula.

49

## 50 Introduction

51 The origin of SARS-CoV-2, as its mode of introduction into the human population, is currently unknown.  
52 Since its emergence, numerous animal species have been studied to identify possible reservoirs and/or  
53 intermediate hosts of the virus, including a large diversity of insectivorous bats of the genus  
54 *Rhinolophus*. Despite the recent report of various SARS-CoV-2-related viruses in *Rhinolophus shameli*  
55 (isolated in Cambodia in 2010<sup>13</sup>), *R. pusillus* and *R. malayanus* (China, 2020 and 2019 respectively<sup>2</sup>), in  
56 *R. acuminatus* (Thailand, 2020<sup>3</sup>) and *R. cornutus* (Japan, 2013<sup>4</sup>), the closest SARS-CoV-2 bat-borne  
57 genome still remains the one from *R. affinis*, RaTG13 (China, 2013)<sup>5,6</sup> with 96.1% similarity at the whole  
58 genome level. Several studies also suggested the involvement of pangolin coronaviruses in the  
59 emergence of SARS-CoV-2<sup>7-9</sup>. Since its appearance in humans, SARS-CoV-2 has evolved through  
60 sporadic mutations and recombination events<sup>14</sup>, some of which correspond to gains in fitness allowing  
61 the virus to spread more widely, or to escape neutralizing antibodies<sup>15</sup>.

62 To decipher the origin of SARS-CoV-2, it is therefore essential to ascertain the diversity of animal  
63 coronaviruses and more specifically of bat coronaviruses. Although the identification of SARS-CoV-2 in  
64 bats is a major goal, a more realistic objective is to identify the sequences that contribute to its  
65 mosaicism. The spike sequence appears essential, as it determines the binding affinity and accessibility  
66 of the receptor-binding domain (RBD) to the cellular angiotensin-converting enzyme 2 (ACE2) receptor  
67 and is therefore responsible for host range<sup>10-12</sup>. The closest related bat strain identified to date  
68 (RaTG13) has a low RBD sequence similarity to SARS-CoV-2 and, with only 11/17 human ACE2 (hACE2)  
69 contact amino-acid residues conserved with SARS-CoV-2, its affinity to hACE2 is very limited<sup>16</sup>.  
70 Moreover, SARS-CoV-2 poorly infects bats and bat cells tested so far<sup>17</sup>. In addition, no bat SARS-CoV-  
71 2-like virus has been shown to use hACE2 to efficiently enter human cells, and none presents the furin  
72 cleavage site that is associated with an increased pathogenicity in humans<sup>18</sup>. SARS-CoV-2 RBD binds to  
73 *Rhinolophus macrotis* ACE2 with a lower affinity than to hACE2<sup>19</sup>. An essential piece of information, i.e.  
74 finding bat viruses with an RBD motif close to that of SARS-CoV-2 and capable of binding to hACE2 with  
75 high affinity, is therefore missing.

76 We hypothesized that this type of virus could be identified in bats living in the limestone karstic terrain  
77 common to China, Laos, and Vietnam in the Indochinese peninsula. We report here the presence of  
78 sarbecoviruses close to SARS-CoV-2 whose RBDs differ from that of SARS-CoV-2 by only one or two  
79 contact residues, that strongly bind to the hACE2 protein, and mediate a hACE2-dependent entry and

80 replication into human cells. Despite the absence of the furin cleavage site, these viruses may have  
81 contributed to SARS-CoV-2's origin and may intrinsically pose a future risk of direct transmission to  
82 humans.

### 83 **Main**

#### 84 *Diversity of bat and coronavirus species*

85 A total of 645 bats belonging to six families and 46 species were captured (Table S1). Two hundred and  
86 forty-seven blood samples, 608 saliva, 539 anal/feces, and 157 urine swabs were collected from the  
87 northern part of Laos (Table S2). We first screened all 539 feces samples through a pan-coronavirus RT  
88 nested PCR approach<sup>20</sup>. Overall, 24 individuals of 10 species were positive, and one individual (BANAL-  
89 27) was concomitantly infected by an alphacoronavirus and a betacoronavirus (Table S3). BLAST  
90 analysis of amplicons identified alphacoronavirus sequences of the *Decacovirus*, *Pedacovirus*, and  
91 *Rhinacovirus* subgenera and betacoronavirus sequences of the *Nobecovirus* and *Sarbecovirus*  
92 subgenera. Sequences of the *Sarbecovirus* subgenus were all identified from *Rhinolophus* individuals  
93 belonging to three different species *i.e.*, *R. malayanus*, *R. marshalli*, and *R. pusillus*. Positive individuals  
94 were trapped in three different districts, and those infected with a sarbecovirus were all from the  
95 Fueng district in Vientiane Province (Fig. 1A, site 1).

96 NGS together with Sanger sequencing was used to obtain a complete genomic sequence of five of the  
97 seven sarbecoviruses (Fig. 1, Table S4). The coverage of the genome of the remaining two  
98 sarbecoviruses, *i.e.* BANAL-27 and BANAL-242 sampled from *R. pusillus* and *R. malayanus* bats  
99 respectively, was 90% and therefore they were not included in the final analyses. Phylogenetic analyses  
100 performed on the receptor-binding protein domain of lineages A and B human SARS-CoV-2<sup>21</sup>, and on  
101 representative bat and pangolin sarbecoviruses, placed the Laotian *R. malayanus* BANAL-52, *R. pusillus*  
102 BANAL-103, and *R. marshalli* BANAL-236 coronaviruses close to human SARS-CoV-2 and pangolin  
103 coronaviruses collected in 2019, while *R. malayanus* BANAL-116 and BANAL-247 coronaviruses  
104 belonged to a sister clade with other bat coronaviruses (RmYN02, RacCS203, RpYN06, and PrC31) from  
105 different *Rhinolophus* species. Pangolin coronaviruses sampled in 2017 displayed a basal position  
106 relative to these strains (Fig. 1B). Interestingly, one should note that very similar SARS-CoV-2-like  
107 viruses are shared by different bat species, suggesting a possible circulation of viruses between  
108 different species living sympatrically in the same caves. These results are consistent with the similarity  
109 plot analysis showing that RaTG13 and BANAL-52 bat coronaviruses exhibit high nucleotide identity  
110 with SARS-CoV-2 throughout the length of the genome (96.8% for BANAL-52 and 96.1% for RaTG13).  
111 Interestingly, BANAL-52 presents a higher level of nucleotide conservation than RaTG13 in the S1  
112 domain of the spike, and especially in the spike's N-terminal domain (NTD) and RBD (Fig. 1C). These  
113 observations are congruent with amino-acid identities between human SARS-CoV-2 and  
114 representative bat and pangolin coronaviruses, which present a high level of conservation, except for  
115 ORF8 of bat BANAL-116, BANAL-247, Rc-o319 and RmYN02 (Supp. Figure 1). Interestingly, the S1  
116 domain of the spike (and especially the N-terminal domain) presents a lower degree of conservation  
117 in several bat coronaviruses, suggesting that this domain may reflect a relative degree of adaptation  
118 of the virus to its mammalian host (Fig. 1D, Extended data 1).

#### 119 *Bat Sarbecovirus evolutionary history*

120 Following GARD analysis, we identified 14 recombinant breakpoints during the evolutionary history of  
121 sarbecoviruses, which were further confirmed by phylogenetic analyses performed on the 15  
122 fragments of sequences defined by the breakpoints (Fig. 2, Supp. Figure 2). No specific signature was  
123 identified in the breakpoints (Table S5). SARS-CoV-2 presents a mosaic genome, to which contributed

124 more than five sequences close to sequences published or determined during this study: *R. malayanus*  
125 RmYN02 and *R. pusillus* RpYN06 viruses found in China in 2019, *R. affinis* RaTG13 coronavirus found in  
126 China in 2013, and *R. malayanus* BANAL-52 and *R. pusillus* BANAL-103 found in North Laos in 2020 (this  
127 study). No pangolin coronavirus sequence was immediately associated with a recombination event at  
128 the origin of SARS-CoV-2. Laotian *Rhinolophus* bat coronaviruses presented a lower degree of  
129 recombination compared to SARS-CoV-2. Such recombination events occurred between other BANAL  
130 viruses isolated from bats living sympatrically in caves in the same area.

131 Interestingly, the origin of several fragments of SARS-CoV-2 genomes could be assigned to several  
132 donor strains and not a unique donor sequence. For example, a breakpoint was identified 7 nucleotides  
133 upstream of the RBD region of S1: the downstream fragment of SARS-CoV-2, which comprises the RBD  
134 and the beginning of S2, could involve BANAL-52 *R. malayanus*, BANAL-103 *R. pusillus*, and BANAL-236  
135 *R. marshalli* viruses, which formed a highly supported sister clade of SARS-CoV-2 (fragment 11, Supp.  
136 Figure 2). In a more basal position are *R. shameli* bat coronaviruses and pangolin-2019 coronaviruses,  
137 consistent with the conservation of RBD amino-acid sequences among SARS-CoV-2 and representative  
138 bat and pangolin coronaviruses (Extended data 2). Among the 17 residues that interact with human  
139 ACE2, 16 are conserved between SARS-CoV-2 and BANAL-52 or -103 (one mismatch, H498Q), and  
140 15/17 are conserved for BANAL-236 (two mismatches, K493Q and H498Q) while only 13/17 residues  
141 are conserved for the Cambodian bat *R. shameli* virus and 11/17 for the Chinese bat *R. affinis* RaTG13  
142 virus. At the full spike protein level, bat *R. affinis* RaTG13 and pangolin-2017 P4L viruses looked closer  
143 to SARS-CoV-2 than bat *R. malayanus* BANAL-52, but this effect is due to a higher degree of  
144 conservation in S2. All these viruses shared the absence of a furin cleavage site and the conservation  
145 of the internal fusion peptide (Supp. Figures 1 & 3 and Extended data 3).

#### 146 *Interaction of BANAL RBDs with ACE2*

147 Biolayer interferometry experiments to measure the interaction between hACE2 and the RBDs of  
148 BANAL-52/103 (which have identical residues in the receptor-binding motif) (Extended data 2), BANAL-  
149 236 and SARS-CoV-2 (residues 233-524), resulted in a dissociation constant  $K_d$  three times lower for  
150 the BANAL RBDs compared to SARS-CoV-2 (Fig. 3A). This higher affinity can be attributed to the Q498H  
151 mismatch that has been reported to increase the affinity of SARS-CoV-2 RBD for hACE2, and also to be  
152 involved in the host range expansion of SARS-CoV-2 and SARS-CoV-2 like viruses<sup>22-29</sup>.

153 To study the effect of the mutations at the interface between these RBDs and hACE2, we performed  
154 Molecular Dynamics (MD) simulations of the SARS-CoV-2, BANAL-236 and BANAL-52/103 RBD/hACE2  
155 complexes initiated from the crystal structure and homology models of these systems, respectively  
156 (Table S6). Cluster analysis of the MD trajectories revealed that, at the RBD-hACE2 interface, both  
157 BANAL complexes were identical to the SARS-CoV-2 RBD/hACE2 complex within 2 Å backbone RMSD  
158 (Extended data 4), except for one of the BANAL-52/103 simulations that displayed larger  
159 conformational variability of the RBD residues S443-Y449 (Supp. Methods and Extended data 5).  
160 Empirical scoring functions predicted a similar RBD/hACE2 binding energy in all three complexes  
161 (Extended data 6).

162 The analysis of the persistence of hydrogen bonds and salt bridges provided further insights into the  
163 effect of the substitutions at the RBD/hACE2 interface (Fig. 3B and Extended data 7). The H498Q  
164 mismatch present in both BANAL-52/103 and BANAL-236 RBDs disrupted the hydrogen bonds between  
165 RBD Q498 and both hACE2 K353 and Q42. However, these hydrogen bonds were only transiently  
166 formed in the SARS-CoV-2 complex: more persistent hydrogen bonds in this region (RBD T500 – hACE2  
167 D355, RBD G502 – hACE2 K353, RBD Y505 – hACE2 E37) were not affected. The K493Q mismatch  
168 enabled the formation of two salt bridges between RBD and hACE2 that were not present in the SARS-  
169 CoV-2 complex (RBD K493 – hACE2 E35 and RBD K493 – hACE2 D38).

170 For further insight into the molecular details of these interactions, we determined the crystal structure  
171 of the complex BANAL-236 RBD with the hACE2 peptidase domain to 2.9 Å resolution (Table S7). The  
172 overall structure of this RBD is identical to that of the SARS-CoV-2 (RMSD 0.360 Å, 150 C $\alpha$ ). The only  
173 significant difference is in the region between amino acids D363 and S375 (Fig. 3C, arrow). In this  
174 region, BANAL spikes have the A372T mismatch, which converts the sequence <sup>370</sup>NSA<sup>372</sup> to <sup>370</sup>NST<sup>372</sup>,  
175 an N-linked glycosylation sequon. The crystals indeed showed clear electron density for the first N-  
176 acetylglucosamine residue of glycan attached to N<sup>370</sup>. The main chain at residue T<sup>372</sup> makes a hydrogen  
177 bond with the glycan residue, altering the conformation of the main chain downstream which results  
178 in partial unwinding of helix H4 in RBD 236. The calculation of a simulated annealing composite omit  
179 map for the segment D363-S375 confirms the correct assignment of the structure for this polypeptide  
180 segment and the N-glycosylation at residue N370 (Extended data 8).

181 As expected, most of the interactions observed in the SARS-CoV-2 RBD/hACE2 complex<sup>30</sup> are also  
182 present in the structure of the BANAL-236 RBD/hACE2 complex. In these interfaces, there are three  
183 main clusters of interactions as indicated in Fig. 3C (insets). The sequence mismatches are in clusters  
184 2 (making the salt bridge RBD K493 – hACE2 E35) and 3 (hydrogen bond between RBD H498 and D38).  
185 Although the interaction K493-E35 contributes to stabilizing the complex, it does not seem to affect  
186 drastically the binding to hACE2 because both BANAL RBDs have similar K<sub>d</sub> values.

#### 187 *Virus replication in human cells*

188 To assess whether the BANAL-236 spike protein could mediate entry into cells expressing human ACE2,  
189 we generated lentiviral particles pseudotyped with the Wuhan or the BANAL-236 spikes (Supp. Figure  
190 4). We detected spike-mediated entry of the BANAL-236 spike-pseudotyped lentivirus in 293T-ACE2,  
191 contrarily to control cells not expressing hACE2 (Fig. 4A). Entry was blocked by human sera neutralizing  
192 SARS-CoV-2, but not by control non-neutralizing sera, demonstrating that neutralization of BANAL-236  
193 was specific for epitopes shared with the spike of SARS-CoV-2 (Fig. 4B).

194 In order to isolate infectious viruses, rectal swabs were inoculated on VeroE6 cells. No CPE was  
195 observed 3 and 4 days after infection, but viral RNAs were detected for one of the 2 wells inoculated  
196 with the BANAL-236 sample (C<sub>T</sub> = 25.1 at D3, C<sub>T</sub> = 21.7 at D4). The culture supernatant (C1) formed  
197 plaques on VeroE6 and the titer was 3800 pfu/mL. A C2 viral stock was prepared by amplification on  
198 VeroE6 at a MOI of 10<sup>-4</sup>. The culture supernatant was harvested on day 4 when CPE was observed and  
199 titrated on VeroE6 (Extended data 9). The plaques' phenotype was small, but the titer reached 2.6.10<sup>6</sup>  
200 pfu/mL. The random NGS performed on the RNA extracted from this stock confirmed that the culture  
201 was pure and corresponded to the BANAL-236 virus, without any non-synonymous mutations between  
202 the original BANAL-236 genome and the C2 viral stock. Replication of BANAL-236 in VeroE6 was  
203 efficiently inhibited by soluble human ACE2, thus showing that entry and propagation was largely  
204 ACE2-dependent (Fig. 4C, Extended data 9). Furthermore, BANAL-236 replicated in human cell lines  
205 expressing endogenous levels of ACE2, Calu-3 and Caco-2 (Fig. 4C). The kinetics of RNA synthesis were  
206 slower compared to Wuhan virus. Infectious viral particles were produced at D4 in the supernatant of  
207 Caco-2 and Calu-3 cells (respectively 10<sup>4.7</sup> and 10<sup>2.9</sup> pfu/mL).

208

#### 209 **Discussion**

210 Many sarbecoviruses circulate in *Rhinolophus* colonies living in caves in China and probably also in  
211 neighboring countries further south<sup>31-33</sup>. During the course of a prospective study in Northern Laos,  
212 we have identified, among other coronaviruses, five sarbecoviruses for which we obtained full-length  
213 sequences. Among these, three (BANAL-52, -103, and -236) were considered to be close to SARS-CoV-  
214 2 because of the similarity of one of the S1 domains (NTD, RBD) or S2 with SARS-CoV-2.

215 Because genomic regions subject to recombination are likely contributing to host-virus interactions,  
216 we compared SARS-CoV-2 strains from the two lineages identified at the very onset of the COVID-19  
217 outbreak<sup>21</sup> to these novel bat sarbecoviruses and to pangolin strains within the SARS-CoV-2 clade.  
218 Strains close to *R. pusillus* RpYN06, *R. malayanus* RmYN02, and *Rhinolophus* sp. PrC31 isolated in China  
219 in 2018-2019, along with *R. malayanus* BANAL-52, *R. pusillus* BANAL-103, and *R. marshalli* BANAL-236  
220 isolated in Laos in 2020, contributed to the appearance of SARS-CoV-2 in different regions of the  
221 genome. No closer viral genome has yet been identified as a possible contributor, and pangolin  
222 coronaviruses appear as more distantly related than bat coronaviruses. We identified potential  
223 recombination sites, allowing for the reconstruction of the phylogenetic history of early isolated SARS-  
224 CoV-2 strains between homologous regions defined by recombination points. We identified a  
225 breakpoint at the beginning of the SARS-CoV-2 RBD, resulting in a downstream fragment key for the  
226 virus tropism and host spectrum composed of the RBD, the furin cleavage site, and ending in the N-  
227 terminal region of S2. Despite the absence of the furin site, phylogenetic reconstruction of this  
228 fragment revealed that Laotian *R. malayanus* BANAL-52, *R. pusillus* BANAL-103, and *R. marshalli*  
229 BANAL-236 coronaviruses are the closest ancestors of SARS-CoV-2 known to date. ORF8 was highly  
230 divergent between SARS-CoV-2 related genomes. ORF8 from strains BANAL-52, -103, -236, like that of  
231 RaTG13, were closer to SARS-CoV-2 than to pangolin strains. ORF8 encodes a protein that has been  
232 proposed to participate in immune evasion<sup>34</sup> and is deleted in many human SARS-CoV-2 strains that  
233 appeared after March 2020<sup>35</sup>, which is reminiscent of the deletions identified during the 2003 SARS  
234 epidemic<sup>36</sup>. Therefore, the presence of ORF8 is consistent with bats acting as a natural reservoir of  
235 early strains of SARS-CoV-2.

236 Structural and functional biology studies have identified the RBD domain that mediates the interaction  
237 with hACE2 and host range, as well as the major amino acids that are involved<sup>30,37,38</sup>. The RBDs (BANAL-  
238 52, -103, and -236) are closer to SARS-CoV-2 than that of any other bat strain described so far, in  
239 particular that of RaTG13, the virus identified in *R. affinis* from the Mojiang mineshaft where  
240 pneumonia cases with clinical characteristics *a posteriori* interpreted as similar to COVID-19<sup>6</sup> were  
241 recorded in 2012<sup>39,40</sup>. Overall, one (H498Q (BANAL-103 and -52)) or two (K493Q and H498Q (BANAL-  
242 236)) amino acids interacting with hACE2 are substituted in these strains in comparison to SARS-CoV-  
243 2. These mutations did not destabilize the BANAL-236 / hACE2 interface, as shown by the BLI  
244 experiments (Fig. 3A) and analyzed by MD simulations.

245 Our results therefore support the hypothesis that SARS-CoV-2 could originally result from a  
246 recombination of sequences pre-existing in *Rhinolophus* bats living in the extensive limestone cave  
247 systems of South-East Asia and South China<sup>41,42</sup>. Many species forage in the same cave areas, including  
248 *R. malayanus* and *R. pusillus*<sup>43</sup>. In addition, the distribution of *R. marshalli*, *R. malayanus*, and *R. pusillus*  
249 overlaps in the Indochinese sub-region (Supp. Figure 5), which means they may share caves as roost  
250 sites and foraging habitats<sup>44</sup>. With the novel viruses here described, understanding the emergence of  
251 SARS-CoV-2 does not need to hypothesize recombination or natural selection for increased RBD affinity  
252 for hACE2 in an intermediate host like the pangolin before spillover<sup>45</sup>, nor natural selection in humans  
253 following spillover<sup>46</sup>. However, we found no furin cleavage site in any of these viruses on sequences  
254 determined from original fecal swab samples, devoid of any risk of counterselection of the furin site  
255 by amplification in Vero cells<sup>18</sup>. Lack of furin cleavage site may be explained by insufficient sampling in  
256 bats. Based on comparison of the sequences around the S1/S2 cleavage site (Extended data 3), it has  
257 been suggested that the furin cleavage site present in SARS-CoV-2 could originate from recombination  
258 events between SARS-CoV-2-related coronaviruses co-circulating in bats<sup>2,47</sup>, meaning that BANAL-116,  
259 BANAL-247, bat RmYN02<sup>2</sup> and bat RacCS203<sup>3</sup> coronaviruses may share a common history with SARS-  
260 CoV-2. Alternatively, the furin cleavage site could have been acquired through passages of the virus in

261 an alternate host or during an early poorly symptomatic unreported circulation in humans. Finally, the  
262 epidemiological link between these bat viruses and the first human cases remains to be established.

263 As expected from the high affinity for ACE2 of the S ectodomain of BANAL-236, pseudoviruses  
264 expressing it were able to enter efficiently human cells expressing endogenous hACE2 using an ACE2-  
265 dependent pathway. However, alternative routes of entry may still exist, especially in cells that do not  
266 express ACE2<sup>48</sup>. Entry was blocked by a serum neutralizing SARS-CoV-2. The RaTG13 strain, the closest  
267 to SARS-CoV-2 known before, had never been isolated. In contrast, preliminary studies show that  
268 BANAL-236 replicated in primate VeroE6 cells with a small plaque phenotype compared to SARS-CoV-  
269 2. Further analysis may indicate more clearly which steps shape infectivity.

270 To conclude, our results pinpoint the presence of new bat sarbecoviruses that seem to have the same  
271 potential for infecting humans as early strains of SARS-CoV-2. Guano collectors, or certain ascetic  
272 religious communities who spend time in or very close to caves, as well as tourists visiting caves, are  
273 particularly at risk of being exposed. Further investigations are needed to assess if such exposed  
274 populations have been infected, symptomatically or not, by one of these viruses, and whether  
275 infection could confer protection against subsequent SARS-CoV-2 infections. In this context, it is  
276 noteworthy that SARS-CoV-2 with the furin site deleted replicates in hamsters and in transgenic mice  
277 expressing hACE2, but leads to less severe disease while protecting from rechallenge with wild-type  
278 SARS-CoV-2<sup>18</sup>.

279

#### 280 **Data availability statement:**

281 Sequences data that support the findings of this study have been deposited in the GenBank and GISAID  
282 databases with the following accession numbers: MZ937000 / EPI\_ISL\_4302644 (BANAL-52),  
283 MZ937001 / EPI\_ISL\_4302645 (BANAL-103), MZ937002 / EPI\_ISL\_4302646 (BANAL-116), MZ937003 /  
284 EPI\_ISL\_4302647 (BANAL-236), and MZ937004 / EPI\_ISL\_4302648 (BANAL-247). Raw sarbecovirus NGS  
285 reads and Sanger sequencing .ab1 files were deposited onto the NCBI SRA repository under the  
286 BioProject PRJNA796968 (SAMN24959173-77). The crystal structure presented in this manuscript was  
287 deposited in the Protein Data Bank (PDB) with accession code 7PKI. The GROMACS topology and input  
288 files as well as the analysis scripts used are freely available on PLUMED-NEST ([www.plumed-nest.org](http://www.plumed-nest.org))  
289 under accession ID plumID:21.037.

#### 290 **Code availability statement:**

291 MODELLER v. 10.1 (<https://salilab.org/modeller/>) and GROMACS v. 2020.4 (<http://www.gromacs.org>)  
292 were used for MD simulations. XDS (v. feb 5 2021), CCP4 (v 7.0), COOT (v. 0.8.6) and PHENIX (v. 1.19.2-  
293 4158) softwares were used for X-ray diffraction data processing, model building and refinement of  
294 BANAL 236 RBD-hACE2 complex . PyMOL (v 2.4.2) was used for structural image rendering.

295 PLUMED v. 2.7 (<https://www.plumed.org>); ROSETTA v. 3.11 (<https://www.rosettacommons.org/>);  
296 FoldX v. 4 (<http://foldxsuite.crg.eu/>); MDAnalysis v. 1.0.0 (<https://www.mdanalysis.org/>); MDTraj v.  
297 1.9.5 (<https://www.mdtraj.org/1.9.5/index.html>); and in-house scripts available at  
298 <https://github.com/maxbonomi/bat-MD> were used for MD simulations.

299 hyphy 2.5.31; PhyML & MAFFT implemented through NGPphylogeny (<https://ngpphylogeny.fr/>);  
300 Simplot 3.5.1 were used for phylogenetic and recombination analyses.

301 MSSPE 1 was used for *Betacoronavirus* primer enrichment design and is available at  
302 <https://github.com/chiulab/MSSPE-design>.

303 MICROSEEK is an in-house pipeline that uses ALIEN TRIMMER v. 2.0  
304 (<https://gitlab.pasteur.fr/GIPhy/AlienTrimmer>) for read trimming / clipping, BBNORM from BBMAP v.  
305 38.86 package (<https://sourceforge.net/projects/bbmap/>) for coverage normalization, MEGAHIT v.  
306 1.2.9 (<https://github.com/voutcn/megahit>) for assembly, a in-house ORF-finder  
307 ([https://figshare.com/articles/code/translateReads\\_py/7588592](https://figshare.com/articles/code/translateReads_py/7588592)), and then DIAMOND v. 2.0.4  
308 (<https://github.com/bbuchfink/diamond/>) and NCBI BLAST v. 2.12.0+  
309 (<ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/2.12.0/>) both for sequence searching.

310

311

312

313 **References**

- 314 1. Hul, V. et al. A novel SARS-CoV-2 related coronavirus in bats from Cambodia. bioRxiv  
315 2021.01.26.428212 (2021) doi:10.1101/2021.01.26.428212.
- 316 2. Zhou, H. et al. Identification of novel bat coronaviruses sheds light on the evolutionary origins  
317 of SARS-CoV-2 and related viruses. *Cell* (2021) doi:10.1016/j.cell.2021.06.008.
- 318 3. Wacharapluesadee, S. et al. Evidence for SARS-CoV-2 related coronaviruses circulating in bats  
319 and pangolins in Southeast Asia. *Nat. Commun.* 12, 972 (2021).
- 320 4. Murakami, S. et al. Detection and Characterization of Bat Sarbecovirus Phylogenetically  
321 Related to SARS-CoV-2, Japan. *Emerg. Infect. Dis.* 26, 3025–3029 (2020).
- 322 5. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat  
323 origin. *Nature* 579, 270–273 (2020).
- 324 6. Rahalkar, M. C. & Bahulikar, R. A. Lethal Pneumonia Cases in Mojiang Miners (2012) and the  
325 Mineshaft Could Provide Important Clues to the Origin of SARS-CoV-2. *Front. Public Health* 8, (2020).
- 326 7. Liu, P. et al. Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)?  
327 *PLoS Pathog.* 16, e1008421 (2020).
- 328 8. Xiao, K. et al. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature*  
329 583, 286–289 (2020).
- 330 9. Wahba, L. et al. An Extensive Meta-Metagenomic Search Identifies SARS-CoV-2-Homologous  
331 Sequences in Pangolin Lung Viromes. *mSphere* 5, e00160-20 (2020).
- 332 10. Letko, M., Marzi, A. & Munster, V. Functional assessment of cell entry and receptor usage for  
333 SARS-CoV-2 and other lineage B betacoronaviruses. *Nat. Microbiol.* 5, 562–569 (2020).
- 334 11. Shang, J. et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature* 581, 221–224  
335 (2020).
- 336 12. Wang, Q. et al. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2.  
337 *Cell* 181, 894-904.e9 (2020).
- 338 13. Delaune, D. et al. A novel SARS-CoV-2 related coronavirus in bats from Cambodia. *Nat.*  
339 *Commun.* 12, 6563 (2021).
- 340 14. Jackson, B. et al. Generation and transmission of interlineage recombinants in the SARS-CoV-  
341 2 pandemic. *Cell* 184, 5179-5188.e8 (2021).
- 342 15. Rochman, N. D. et al. Ongoing global and regional adaptive evolution of SARS-CoV-2. *Proc.*  
343 *Natl. Acad. Sci.* 118, (2021).
- 344 16. Liu, K. et al. Binding and molecular basis of the bat coronavirus RaTG13 virus to ACE2 in  
345 humans and other species. *Cell* 184, 3438-3451.e10 (2021).
- 346 17. Aicher, S.-M. et al. Species-specific molecular barriers to SARS-CoV-2 replication in bat cells.  
347 bioRxiv 2021.05.31.446374 (2021) doi:10.1101/2021.05.31.446374.
- 348 18. Johnson, B. A. et al. Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature*  
349 591, 293–299 (2021).



- 350 19. Liu, K. et al. Cross-species recognition of SARS-CoV-2 to bat ACE2. *Proc. Natl. Acad. Sci.* 118,  
351 (2021).
- 352 20. Chu, D. K. W. et al. Avian Coronavirus in Wild Aquatic Birds. *J. Virol.* 85, 12815–12820 (2011).
- 353 21. Rambaut, A. et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist  
354 genomic epidemiology. *Nat. Microbiol.* 5, 1403–1407 (2020).
- 355 22. Wrapp, D. et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation.  
356 *Science* 367, 1260–1263 (2020).
- 357 23. Laffebert, C., de Koning, K., Kanaar, R. & Lebbink, J. H. G. Experimental Evidence for Enhanced  
358 Receptor Binding by Rapidly Spreading SARS-CoV-2 Variants. *J. Mol. Biol.* 433, 167058 (2021).
- 359 24. Lei, C. et al. Neutralization of SARS-CoV-2 spike pseudotyped virus by recombinant ACE2-Ig.  
360 *Nat. Commun.* 11, 2070 (2020).
- 361 25. Walls, A. C. et al. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein.  
362 *Cell* 181, 281–292.e6 (2020).
- 363 26. Niu, S. et al. Molecular basis of cross-species ACE2 interactions with SARS-CoV-2-like viruses  
364 of pangolin origin. *EMBO J.* 40, e107786 (2021).
- 365 27. Zhang, Y. et al. SARS-CoV-2 rapidly adapts in aged BALB/c mice and induces typical  
366 pneumonia. *J. Virol.* JVI.02477-20 (2021) doi:10.1128/JVI.02477-20.
- 367 28. Huang, K. et al. Q493K and Q498H substitutions in Spike promote adaptation of SARS-CoV-2  
368 in mice. *EBioMedicine* 67, 103381 (2021).
- 369 29. Zhang, S. et al. Bat and pangolin coronavirus spike glycoprotein structures provide insights  
370 into SARS-CoV-2 evolution. *Nat. Commun.* 12, 1607 (2021).
- 371 30. Lan, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2  
372 receptor. *Nature* 581, 215–220 (2020).
- 373 31. Hu, B. et al. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new  
374 insights into the origin of SARS coronavirus. *PLOS Pathog.* 13, e1006698 (2017).
- 375 32. Ge, X.-Y. et al. Isolation and characterization of a bat SARS-like coronavirus that uses the  
376 ACE2 receptor. *Nature* 503, 535–538 (2013).
- 377 33. Latinne, A. et al. Origin and cross-species transmission of bat coronaviruses in China. *Nat.*  
378 *Commun.* 11, 4235 (2020).
- 379 34. Novel Immunoglobulin Domain Proteins Provide Insights into Evolution and Pathogenesis of  
380 SARS-CoV-2-Related Viruses. <https://journals.asm.org/doi/epub/10.1128/mBio.00760-20>  
381 doi:10.1128/mBio.00760-20.
- 382 35. Su, Y. C. F. et al. Discovery and Genomic Characterization of a 382-Nucleotide Deletion in  
383 ORF7b and ORF8 during the Early Evolution of SARS-CoV-2. *mBio* 11, e01610-20.
- 384 36. Chinese SARS Molecular Epidemiology Consortium. Molecular evolution of the SARS  
385 coronavirus during the course of the SARS epidemic in China. *Science* 303, 1666–1669 (2004).
- 386 37. Conceicao, C. et al. The SARS-CoV-2 Spike protein has a broad tropism for mammalian ACE2  
387 proteins. *PLoS Biol.* 18, e3001016 (2020).

- 388 38. Damas, J. et al. Broad host range of SARS-CoV-2 predicted by comparative and structural  
389 analysis of ACE2 in vertebrates. *Proc. Natl. Acad. Sci. U. S. A.* 117, 22311–22322 (2020).
- 390 39. Cohen, J. Wuhan coronavirus hunter Shi Zhengli speaks out. *Science* 369, 487–488 (2020).
- 391 40. Ge, X.-Y. et al. Coexistence of multiple coronaviruses in several bat colonies in an abandoned  
392 mineshaft. *Viol. Sin.* 31, 31–40 (2016).
- 393 41. Clements, R., Sodhi, N. S., Schilthuizen, M. & Ng, P. K. L. Limestone Karsts of Southeast Asia:  
394 Imperiled Arks of Biodiversity. *BioScience* 56, 733–742 (2006).
- 395 42. Hassanin, A., Tu, V. T., Curaudeau, M. & Csorba, G. Inferring the ecological niche of bat  
396 viruses closely related to SARS-CoV-2 using phylogeographic analyses of *Rhinolophus* species. *Sci.*  
397 *Rep.* 11, 14276 (2021).
- 398 43. Soisook, P. et al. A taxonomic review of *Rhinolophus steno* and *R. malayanus* (Chiroptera:  
399 Rhinolophidae) from continental Southeast Asia: an evaluation of echolocation call frequency in  
400 discriminating between cryptic species. *Acta Chiropterologica* 10, 221–242 (2008).
- 401 44. Francis, c. *Field Guide to the Mammals of South-east Asia* (2nd Edition). (2019).
- 402 45. Makarenkov, V., Mazouze, B., Rabusseau, G. & Legendre, P. Horizontal gene transfer and  
403 recombination analysis of SARS-CoV-2 genes helps discover its close relatives and shed light on its  
404 origin. *BMC Ecol. Evol.* 21, 5 (2021).
- 405 46. Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of  
406 SARS-CoV-2. *Nat. Med.* 26, 450–452 (2020).
- 407 47. The Sarbecovirus origin of SARS-CoV-2's furin cleavage site - SARS-CoV-2 coronavirus / nCoV-  
408 2019 Evolutionary History. *Virological* [https://virological.org/t/the-sarbecovirus-origin-of-sars-cov-2-  
409 s-furin-cleavage-site/536/6](https://virological.org/t/the-sarbecovirus-origin-of-sars-cov-2-s-furin-cleavage-site/536/6) (2021).
- 410 48. Puray-Chavez, M. et al. Systematic analysis of SARS-CoV-2 infection of an ACE2-negative  
411 human airway cell. *Cell Rep.* 36, 109364 (2021).
- 412
- 413

**Figure 1. Genomic description of bat-borne sarbecoviruses identified in Laos.** (A) Map of sampling sites. All BANAL isolates were collected from the same site (site 1). Map was downloaded from <https://www.diva-gis.org/gdata>. (B) Phylogenetic analysis of the protein sequence of the receptor-binding domain of Laotian and representative human, bat, and pangolin sarbecoviruses. Sequences were aligned with MAFFT<sup>49</sup> in “auto” mode, and maximum likelihood phylogenetic reconstruction was performed with PhyML implemented through the NGPhylogeny portal<sup>50</sup> with the LG+G substitution model. Branch support was evaluated with the aBayes parameter. Accession numbers and bat species are specified in the name of the sequences. Sequences are colored according to Fig. 1C. (C) Similarity plot analysis of Laotian and representative bat and pangolin sarbecoviruses based on the full-length genome sequence of SARS-CoV-2 human prototype strain (NC\_045512, Wuhan-Hu-1) used as reference. The analysis was performed with the Kimura-2 parameter model, a window size of 1,000 base pairs, and a step size of 100 base pairs with SimPlot program, version 3.5.1<sup>51</sup>. (D) Heatmap of identities at the protein level of representative human, bat, and pangolin sarbecoviruses compared to human SARS-CoV-2 lineage B (NC\_045512). Spike protein has been divided into functional domains, and the sequences are ordered according to percentage of identity of the RBD domain. “\*”: absence of a functional ORF10 in Thai bat RacCS203 (accession number MW251308). Heatmap was created using the gplots package in R (version 3.6.3).

**Figure 2. Recombination events in the evolutionary history of sarbecoviruses.** Schematic representation of the 15 recombinant fragments of relevant *Sarbecovirus* genomes compared to SARS-CoV-2 human prototype strain (NC\_045512). Breakpoint positions are referred to the position in the alignment. Where possible, the closest viral sequence is indicated for each fragment. In other cases, “MULT” (group of multiple sequences) is mentioned. “\$”: unresolved fragment phylogeny (fragment 13, from positions 27344 to 27800 in the alignment). Sequences are colored as in Fig. 1. The complete phylogenetic analyses are presented in Supp. Figure 2.

**Figure 3. Dynamics of the binding of hACE2 to bat-sarbecovirus borne RBDs and structural insight of the complex.** (A) Biolayer interferometry binding analysis of the hACE2 peptidase domain to immobilized BANAL52/103, BANAL-236 RBDs or SARS-CoV-2. Black lines correspond to global fit of the data using a 1:1 binding model. (B) Frequency of formation of salt bridges close to the RBD/ACE2 interface (from left to right: D30/K417, E35/K493, D38/K493, K31/E35, and D38/K353) during the course of the MD simulations. The analysis is performed for 9 different MD simulations (3 replicates for each complex) of hACE2 in complex with SARS-CoV-2 (shades of green), BANAL-236 (shades of red) and BANAL-52/103 RBDs (shades of blue). (C) Ribbon representations of the crystal structures of hACE2 peptidase domain (cyan) in complex with SARS-CoV-2 (PDB 6M0J) or BANAL-236 (this study, PDB 7PKI) RBDs (pink). Black arrows in the overall structures indicate the structural difference between the two complexes at the level of helix H4. The insets show the main interactions in the ACE2-RBD interfaces. Residues in the RBM mutated between SARS-CoV-2 and BANAL-236 are indicated with colored boxes.

**Figure 4. BANAL-236 entry and propagation in human cells.** (A) Results of spike-pseudotyped BANAL-236 (square) and Wuhan-Hu-1 (diamond) pseudoviruses entry assay in HEK-293T cells expressing (purple) or not (grey) the hACE2 receptor, expressed in Relative Luminescence unit (RLU) produced by the firefly luciferase present in the lentiviral backbone and the Bright-Glo luciferase substrate. A single experiment performed in triplicate representative of two experiments is shown. Center values represent the average of the 3 replicates and error bars indicate SD. (B) Results of spike-pseudotyped BANAL-236 (black) and Wuhan-Hu-1 (grey) neutralization assay expressed in % of neutralization of luciferase activity in the absence of serum. SARS-CoV-2 neutralizing sera were from patients with confirmed infections while non-neutralizing sera samples were collected before SARS-CoV-2 spread in Laos. Dashed line: neutralization threshold. A single experiment representative of three independent experiments is shown. (C) Human cell lines expressing endogenous ACE2, Calu-3 (blue line) and Caco-2 (green line), were infected at a MOI of 0.01 with BANAL-236 (square dots) and Wuhan (diamond dots) viruses. VeroE6 cells (red line) were infected at a MOI of 0.0001 with BANAL-236 and Wuhan viruses pre-incubated or not (grey line) with soluble human ACE2 (sACE2) at 25 µg/mL for 30 min. Genome copy number was quantified by RT-qPCR in the supernatants recovered 3- and 4-days post-infection. A single experiment performed in triplicate is shown. Center values represent the average of the 3 replicates and error bars indicate SD.

**Extended data 1. Spike identity matrices at the genus level of representative sarbecoviruses.** Amino-acid (lower) and nucleotide (upper) identity matrices of Laotian and representative human, bat, and pangolin sarbecoviruses. Spike N-terminal (NTD), Receptor-binding (RBD) and S2 nucleotide and amino-acid sequences were aligned with MAFFT, and identity matrices were constructed using CLC Main Workbench 21.0.4 (Qiagen). Matrices were colored according to the identity scale, from 25% (red) to 100% (green) of nucleotide or amino-acid identity.

**Extended data 2. Alignment of spike RBD domain.** Protein alignment of the Receptor Binding Domain (RBD) of Laotian and representative human, bat and pangolin sarbecoviruses. Sequences were aligned with MAFFT in G-INS-I mode. Residues interacting with human ACE2 receptor are highlighted in grey. The domain used for interactions modeling, based on the X-ray structure 6MOJ (residues T333 to G526), is highlighted by a black line.

**Extended data 3. Nucleotide and amino-acid alignments of the furin cleavage site region.** Complete nucleotide and amino-acid spike sequences of representative bat SARS-CoV-2-like coronaviruses were downloaded from GenBank and GISAID and aligned with MAFFT (G-INS-I parameter) (**A & C**). Alignments were manually edited as proposed by Zhou<sup>2</sup> and Lytras<sup>47</sup> with CLC Main Workbench (Qiagen) (**B & D**). Alignments of the furin cleavage region are presented at the nucleotide (**A & B**) and the amino-acid (**C & D**) level, respectively.

**Extended data 4. Analysis of the stability and conformational heterogeneity of RBD/hACE2 complexes.** Time series (left column) and violin plots (right column) of backbone Root Mean Square Deviation (RMSD) from the initial, energy-minimized model calculated on the residues in RBD (**A**), hACE2 (**B**), at the interface of RBD and hACE2 (**C**), and on the entire complex (**D**). In the violin plots, the white circle corresponds to the median value, the black rectangle extends from the first to the third quantiles, and the thin black line represents the 95% confidence intervals. Population of the 3 most significant clusters visited during the course of the MD simulations (**E**). The analysis is performed for 9 different MD simulations: 3 replicates of the SARS-CoV-2 (shades of green), BANAL-236 (shades of red), and BANAL-52/103 (shades of blue) RBD/hACE2 complexes.

**Extended data 5. Additional analysis of the MD simulations of the BANAL-52/103 RBD/hACE2 complex.** Comparison of the time series of interface RMSD during the course of two MD simulations of the BANAL-52/103 RBD/hACE2 complex with short (BANAL-52/103-CoV.1) and long (BANAL-52/103-CoV.1\*) equilibration phase. The large fluctuations of the interface RMSD are due to the flexibility of the RBD loop between residues S443 and Y449 (insets, in yellow). When these residues were not included in the calculation of the interface RMSD, the time series displayed a more stable behavior (BANAL-52/103-CoV.1-L and BANAL-52/103-CoV.1\*-L).

**Extended data 6. Estimation of RBD-hACE2 binding energy.** Time series (left column) and violin plots (right column) of the RBD-hACE2 binding energy estimated using ROSETTA (**A**) and FoldX (**B**). In the violin plots, the white circle corresponds to the median value, the black rectangle extends from the first to the third quantiles, and the thin black line represents the 95% confidence intervals. The analysis is performed for 9 different MD simulations: 3 replicates of the SARS-CoV-2 (shades of green), BANAL-236 (shades of red), and BANAL-52/103 (shades of blue) RBD/hACE2 complexes.

**Extended data 7. Analysis of the inter-subunits hydrogen bonds at the interface of RBD and hACE2.** Frequency of formation of hydrogen bonds at the interface of RBD and hACE2 in the knob (**A**), base (**B**), and tip regions (**C**). The analysis is performed for 9 different MD simulations: 3 replicates of the SARS-CoV-2 (shades of green), BANAL-236 (shades of red), and BANAL-52/103 (shades of blue) RBD/hACE2 complexes.

**Extended data 8. Sticks representation** of segments D364-S375 of BANAL 236 (left panel) and SARS-CoV-2 (right panel) RBDs. A 2Fo-Fc composite omit map (contoured at  $3\sigma$ ) is shown for this region in BANAL 236 RBD.

**Extended data 9. Isolation of BANAL-236 on VeroE6 cells.** (**A**) CPE observed on VeroE6 4 days after inoculation at an MOI of  $10^{-4}$  from the C1. (**B**) Uninfected VeroE6 cells layer. (**C**) Plaque assay performed from the C2 stock on VeroE6 cells. (**D**) Comparative CPE observed on VeroE6 cells infected with Wuhan (top panel) or BANAL-236 (bottom panel) in absence (left) or in presence (right) of soluble ACE2. Photos were taken with a EVOS XL Core microscope at  $\times 10$ . A single experiment performed in triplicate is shown for each cell line.

## Methods

### **Ethical and legal statements**

The bat study was approved by the wildlife authorities of the Department of Forest Resource Management (DFRM), and the Ministry of Agriculture and Forestry, Lao PDR, No. 2493/DFRM, issued on May 21, 2020; No. 0755/MAF issued on June 2, 2020. All animals were captured, handled, and sampled following previously published protocols and ASM guidelines<sup>52,53</sup>. Exportation from Lao and importation in France were conducted according to national regulations. Human serum samples used for neutralization assays were already available<sup>54</sup> and selected based on their status regarding seroneutralization of SARS-CoV-2. They were collected following a protocol approved by the Lao National Ethics Committee for Health Research (NECHR) (Ref #052/2020).

### **Biosafety**

Both the Institut Pasteur du Laos (IPL) and the Faculty of Environmental Science have extensive experience in safely collecting bats (appropriate biosafety training and Personal Protective Equipment (PPE) for collectors). In this study, training prior to field work was organized once for field work participants. The aim was to teach participants on the transmission risks of infectious agents from bats and how to identify, assess, and mitigate these risks, as well to practice the use of PPE.

During field collection, sampling stations were selected to minimize potential exposure to infectious agents and stress on the animals during the handling time by selecting (1) an area easy to disinfect, (2) out of view of the general public, (3) a location that would not expose the general population (like a picnic area), and (4) procedures that reduce time and stress on bats caused by handling. For handling bats and bat samples the following minimum PPE was required: eye protection, an N95 respirator, long clothing/cover all, and latex gloves (2pairs). All wastes were disposed of in biohazard bags and were transported to a disposal site of IPL in Vientiane capital. Each sample box containing 81 cryovial tubes was put inside individual plastic ziplock and transported from field to IPL using dry ice by packing them in a polystyrene foam box. At IPL, samples were stored in a specific -80°C freezer until analysis.

For initial sample analysis at IPL, samples were transferred to BSL-3 room for nucleic acid extraction. Initial Coronavirus screening by RT-PCR was then performed under internal regulation on biosafety and security of IPL. Sample extracted products (50 µl) were stored in NucleoSpin 8 sample boxes (8x12) and sent to IP-Paris by packing them in the individual plastic ziplock and using dry ice. Aliquots of anal swab samples were sent in a separated box by individually triple packaging in dry ice.

All experiments on potentially infectious samples done at Institut Pasteur (Paris) were conducted in BSL-3 laboratories according to procedures adapted to respiratory viruses.

### **Bat sampling areas and sample collection**

Trapping sessions were conducted on four sites, in Fueng and Meth Districts, Vientiane Province, and in Namor and Xay Districts, Oudomxay Province, between July 2020 and January 2021 (Fig. 1A, Tables S1 and S2). Bats were captured using four-bank harp traps<sup>55</sup> and mist nets set in forest patches between rice fields/orange/banana plantations and karst limestone formations, for 5-8 nights depending on accessibility. Harp traps were set across natural trails in patches of forest understory. Mist nets were set across natural trails, at the edges of forests, at entrances of caves, and in areas near cave entrances, as well as in open areas or those with high forest canopy. Bats were morphologically identified following morphological criteria<sup>55-57</sup>. Other data such as forearm length (FA), sex, developmental stage (adult or juvenile) and reproductive condition (pregnant or lactating) were also recorded. Bats were sampled for saliva, feces and/or urine, and blood before release at the capture

site. Species identification of PCR-positive individuals was confirmed by sequencing the mitochondrial Cytochrome oxidase 1 <sup>32</sup>.

### Initial Coronavirus screening

Total RNA was extracted from feces samples using the NucleoSpin 8 virus kit (Macherey Nagel). cDNA was synthesized using the Maxima H minus first strand cDNA synthesis kit (Thermo Scientific) and random hexamers following the manufacturer's instructions. The presence of coronaviruses was tested by a nested PCR approach using PCR master mix (Promega) and by targeting the RNA-dependent RNA Polymerase gene using combinations of degenerate and non-degenerate consensus primers as previously described<sup>20</sup>. PCR products of the expected size were directly sequenced on both strands by Sanger sequencing using the nested PCR primers. The sequences obtained were confirmed by similarity analysis using the NCBI BLASTn search (<http://www.ncbi.nlm.nih.gov/BLAST>).

### Primer design for *Betacoronavirus* enrichment before Next-Generation Sequencing

*Betacoronavirus* enrichment was performed at the genus level by adapting a previously described protocol<sup>58</sup> based on k-mers for targeted-sequence enrichment prior to NGS. Briefly, 2,000 complete *Betacoronavirus* genomes were downloaded from the GenBank and GISAID databases and then clustered to a 95% sequence identity using CD-HIT-EST<sup>59</sup>. Overall, 185 representative sequences of all betacoronaviruses were used for further analysis. Due to the high diversity within the genus *Betacoronavirus*, the full genomes belonging to the subgenera *Sarbecovirus*, *Nobecovirus*, and *Merbecovirus* plus *Embecovirus* were separately aligned using MAFFT multiple sequence alignment software and used to design 13-mer spiked primers per cluster. The genomic position of the full set of primers was extracted from some representative subgenera sequences and close primers were removed. Finally, 416 spiked primers were synthesized by Eurofins Genomics Germany GmbH.

### Sample preparation for sequencing

Reverse transcription was performed using the mix of spiked primers and random hexamers at a 10:1 ratio using the SuperScript™ IV First-Strand Synthesis System (Invitrogen). After a denaturation step at 95°C for 3 min in the presence of dNTPs (500 μM), DTT (5 mM), and RNaseOUT inhibitor, the RT reaction was incubated for 10 thermal cycles consisting of 6 steps at 8°C for 12 s, 15°C for 45 s, 20°C for 45 s, 30°C for 30 s, 35°C for 2 min and 42°C for 3 min, followed by a final incubation step at 42°C for 20 min, as previously described<sup>60</sup>. Double-stranded cDNA was generated using the Sequenase2.0 DNA Polymerase (Applied Biosystems™) in presence of dNTPs and then purified using the Beckman Coulter™ Agencourt AMPure XP.

For samples with a low nucleic acid content, a random amplification step was performed using the MALBAC Single Cell WGA kit (Yikon Genomics, Promega). The amplified product was then purified using AMPure XP Beads, eluted in a final volume of 20 μL of low TE (10 mM Tris-HCl (pH 8,0), 0,1 mM EDTA), and quantified with the Qubit™ DNA HS Assay (Life Technologies, Thermo Fisher Scientific Inc.).

Libraries were generated using the NEBNext Ultra II DNA Library Prep kit (New England Biolabs) after a fragmentation step using the Covaris M220 Focused-ultrasonicator using microTUBE-15 (Peak Incident power (W)=18, Duty Factor=20%, Cycles per Burst=50, Treatment time (sec)=60). The PCR amplified libraries were cleaned up using 0.9X AMPure XP Beads and checked on the 2100 Bioanalyzer system with the High-Sensitivity DNA kit (Agilent Technologies, France) and quantified with the Qubit™ DNA HS Assay. Finally, the dual-multiplexed libraries were pooled (six samples/pool) and run on the Illumina NextSeq500 platform with High Output Kit v2.5 (150 Cycles).

## Amplicon sequencing

In addition to enrichment-based sequencing, cDNA was amplified using the AmpliSeq for Illumina SARS-CoV-2 Research Panel (cat# 20020496), applying twenty-six amplification cycles in PCR1 and nine cycles in PCR2. Primers at the end of the amplicons were partially digested during the library preparation, following manufacturer's instructions. Libraries were barcoded individually using the Illumina UD dual-indexes and normalized with the AmpliSeq Library Equalizer for Illumina (cat# 20019171), then pooled and sequenced on the Illumina NextSeq500 instrument using a Mid Output v2.5 kit (SR 150 cycles). To eliminate residual PCR primer sequences, raw reads were trimmed by 15 bases at each end and special attention was paid to checking that internal sequences corresponding to primer regions in overlapping amplicons did not derive from primer sequences of the multiplex PCR. As an internal control, we verified that sequences of the complete genome of sample BANAL-236 obtained from the enrichment-based sequencing approach and from the AmpliSeq approach were identical.

## Genome assembly and finishing

Raw reads from the enrichment-based sequencing were processed with an in-house bioinformatics pipeline (Microseek; Bigot et al., submitted) comprising quality check and trimming, reads normalization, *de novo* assembly, and ORF prediction of contigs and singletons, followed by 3 levels of taxonomic assignment<sup>61</sup>. Sequences identified as *Sarbecovirus* were then mapped onto appropriate reference sequences using CLC Genomics Workbench 20.0 (Qiagen). Trimmed reads from the amplicon sequencing were mapped to the SARS-CoV-2 genome first, then mapped again (refined mapping) to the closest genome relative. When needed, complete genomes were obtained by conventional PCR and Sanger sequencing. Briefly, viral RNA was reverse transcribed using the SuperScript IV reverse transcriptase (Invitrogen, USA) and cDNA was subsequently used to fill the gaps in the genomes using the Phusion High Fidelity DNA polymerase (New England Biolabs, France) and specific primers flanking the missing regions. Positive PCR products were further purified and sequenced by Sanger sequencing at Eurofins Genomics.

## Recombination and phylogenetic analyses

Identification of recombination events occurring during the evolutionary history of bat sarbecoviruses was performed using the IDPlot package<sup>62</sup>, a web-based workflow that includes multiple sequence alignment and phylogeny-based breakpoints prediction using the GARD algorithm from the HyPhy genetic analysis suite<sup>63</sup>. First, a comprehensive analysis comprising 106 sequences and covering all non-human *Sarbecovirus* and *Sarbecovirus*-related complete genomes available in GenBank and GISAID databases was performed, including prototype strains of SARS-CoV-2 isolated in 2019. Special attention was paid to including bat-borne and pangolin-borne viral sequences in order to maximize the ability to capture a large diversity of the sarbecoviruses. Then, a reduced set of 36 sequences was chosen because of their phylogenetic proximity with SARS-CoV-2, and the GARD algorithm was run to identify recombination breakpoints in Laotian and representative human, bat, and pangolin sarbecoviruses. Breakpoint coordinates were confirmed by performing phylogenetic analyses on the corresponding fragments using PhyML implemented through the NGPhylogeny portal<sup>50</sup>. Branch support was evaluated with the aBayes parameter.

## Generation of lentiviral pseudoviruses

The BANAL-236 and Wuhan synthetic spike genes were cloned into the pVAX1 vector with a cytoplasmic tail truncation of 19 amino acids. Pseudotyped lentiviral particles were prepared using HEK-293T cells (ATCC CRL-3216) seeded in 10-cm dishes. 293T cells were co-transfected with a 5 µg of spike-encoding plasmid, 10 µg of lentiviral backbone plasmid expressing the firefly luciferase (pHAGE-CMV-Luc2-IRES-ZsGreen-W), and 3.3 µg of each lentiviral helper plasmid expressing HIV Gag-Pol (HDM-

Hgpm2), Tat (HDM-tat1b), and Rev (pRC-CMV-Rev1b) using calcium-phosphate precipitation<sup>64</sup>. The medium was replaced 5 h post-transfection by 6 mL of DMEM without fetal calf serum (FCS) and phenol red. Pseudotyped particles were harvested 48 h post-transfection, clarified by centrifugation at 2500 g for 5 min and frozen at  $-80^{\circ}\text{C}$ . Mock pseudotyped lentivirus was generated as above but in the absence of an S-expressing plasmid.

### Spike-pseudotyped lentivirus entry assays

HEK-293T cells stably expressing human ACE2 (293T-ACE2) were transduced in suspension by mixing 50  $\mu\text{L}$  of 3-fold serial dilutions of S-pseudotyped lentiviruses with 50  $\mu\text{L}$  of cells at  $4.10^5$  cells/mL in 96-well white culture plates<sup>65</sup>. At 60-72 h post-transduction, 100  $\mu\text{L}$  of Bright-Glo luciferase substrate (Promega) were added to the wells and luminescence was measured using a Berthold Centro XS luminometer.

### Neutralization assays

Sera neutralizing or not SARS-CoV-2 were described in<sup>54</sup>. They were decomplexed at  $56^{\circ}\text{C}$  for 30 min and 2.5  $\mu\text{L}$  were incubated with 0.5  $\mu\text{L}$  of S-pseudotyped lentiviruses in a final volume of 50  $\mu\text{L}$  of DMEM-10% FCS without phenol red in 96-well white culture plates. After 30 min at room temperature, 50  $\mu\text{L}$  of 293T-ACE2 cells in suspension at  $4.10^5$  cells/mL were mixed to the wells. Luminescence was measured at 60-72 h post-transduction as described above. Neutralization was calculated using the following formula:  $1 - (\text{RLU in presence of serum} / (\text{mean of RLU in absence of serum determined in 12 wells} - 3 \times \text{STD}))$ .

### Virus isolation and multiplication

Rectal swabs were inoculated in duplicate in 24-well plates containing VeroE6 cells (ATCC CRL-1586) (1/5 dilution in 100  $\mu\text{L}$  of DMEM without FCS supplemented with 1% Penicillin-Streptomycin, 1% Fungizone, and 1  $\mu\text{g}/\text{mL}$  TPCK-treated trypsin). After 1 h of adsorption at  $37^{\circ}\text{C}$ , inoculum was removed and 1 mL of the medium described above was added. Three and 4 days after infection, cytopathic effect (CPE) was monitored and 100  $\mu\text{L}$  of supernatant was collected for RNA extraction. RT-qPCR targeting a conserved sequence in the E gene was performed as described<sup>66</sup>. Culture supernatant (C1) was harvested at day 4 and titrated by plaque assay on VeroE6 overlaid with 0.5% carboxymethylcellulose containing 1  $\mu\text{g}/\text{mL}$  TPCK-treated trypsin. A viral stock was prepared by amplification on VeroE6 cells at a MOI of  $10^{-4}$ . Culture supernatant (C2) was harvested at day 4 when massive CPE was observed and titrated by plaque assay on VeroE6 as described above. A C3 viral stock was produced by a subsequent viral amplification at a MOI of  $10^{-4}$  for 3 days on VeroE6. RNA was extracted from the viral stock and submitted to random NGS analysis using the SMARTer® Stranded Total RNA-Seq Kit v3 - Pico Input Mammalian (Takara Bio USA, Inc.). Raw reads were processed with Microseek pipeline, as described above.

For infection experiments, VeroE6 or human cell lines Calu-3 (lung cells, ATCC HTB-55) and Caco-2 (intestinal cells, ATCC HTB-37) were infected in triplicate in 24-well plates with BANAL-236 (C3) or Wuhan viral stocks at a MOI of 0.0001 and 0.01 respectively. Human soluble ACE2 was pre-incubated at 25  $\mu\text{g}/\text{mL}$ <sup>67</sup> for 30 min with the viral inoculum before VeroE6 infection. Infections were carried out without TPCK in the medium described above for 3 and 4 days. Supernatants were recovered at 0-, 3- and 4-days post-infection and genome copy number and viral titers were quantified as described above.

### Protein expression and purification

BANAL-52/103, BANAL-236 and SARS-CoV-2 RBDs (residues 233-524, with C-terminal 8xHis-Strep and AVI tags) and hACE2 peptidase domain (residues 19-615, with C-terminal 8xHis tag) were expressed in Expi293F cells at  $37^{\circ}\text{C}$  and 8%  $\text{CO}_2$  (GnTI- Expi293™, ThermoFisher Scientific). Cell culture supernatants



were collected five days post transfection and purified by affinity chromatography followed by size exclusion chromatography (SEC) using a 200 10/300 GL column pre-equilibrated in 20 mM Tris-HCl pH 8.0, 100 mM NaCl.

For crystallization experiments, the same constructs were expressed in Expi293F GnTI cells. The protein tags were cleaved overnight with thrombin and deglycosylated with EndoH. The RBD was mixed with a 1.3 molar excess of hACE2 and the complex was purified by SEC.

### Biolayer Interferometry

Purified Avi-tagged RBD was biotinylated using a BirA biotin-protein ligase kit according to manufacturer's instructions (Avidity). The biotinylated RBDs at 100 nM were immobilized to SA sensors. A 1:2 dilution series of hACE2 starting at 100 nM in PBS-BSA buffer was used in cycles of 200 s association followed by 200 s dissociation steps to determine protein-protein affinity. The data were baseline-subtracted and the plots fitted using the Pall FortéBio/Sartorius analysis software (version 12.0). Data were plotted in Prism 9.1.0.

### Crystallization and data collection

Crystals of complex BANAL-236 RBD/hACE2 were obtained at 4°C in sitting drops by mixing 200 nl of the protein complex at 8 mg/mL with 200 nl of reservoir solution containing 0.2 M lithium sulfate, 0.1 M Tris 8.5, 30 % w/v PEG 4000. The crystals were soaked in reservoir solution containing 20% glycerol as cryoprotectant before being flash-frozen in liquid nitrogen. X-ray diffraction data was collected on the beamline PROXIMA 1 at the SOLEIL synchrotron (St Aubin, France) and reduced using the XDS package<sup>68</sup>. The structure of the complex was determined by molecular replacement with Phaser software<sup>69</sup> using the coordinates of SARS-CoV-2 RBD in complex with hACE2 as search template (Protein Data Bank (PDB) 6M0J). The model was manually corrected in COOT<sup>70</sup> and refined with phenix.refine<sup>71</sup>. The final coordinates were deposited in the PDB with the entry code 7PKI.

### Molecular Dynamics simulations of RBD/hACE2 complexes

**Generation of homology models.** Homology models of the BANAL-236 and BANAL-52/103 RBD/hACE2 complexes were constructed using the X-ray structure of the SARS-CoV-2 RBD/hACE2 complex (PDB code 6M0J; resolution 2.45 Å) as template using MODELLER v. 10.1<sup>72</sup>. These models included the fragments S19-D615 and T333-G526 of hACE2 and RBD, respectively, which were the regions resolved in the template. In these regions, BANAL-236 and BANAL-52/103 RBDs have a sequence identity to SARS-CoV-2 RBD equal to 96.9% and 97.4%, respectively. The alignment reported in Extended data 3 was used. The zinc and chloride atoms present in the template were retained during homology modeling; *N*-Acetylglucosamine (NAG) and water residues were removed. 7 disulfide bonds were detected by MODELLER in the template (3 in hACE2 and 4 in the RBD) and enforced in the generation of the homology models using CHARMM-like distance and dihedral angles restraints. For each construct, 100 homology models were built and ranked based on the normalized DOPE score<sup>73</sup>. The top 3 scoring models of each complex were used as starting points of 3 independent MD simulations, as described in the following section.

**Details of the MD simulations: setup, equilibration, and production.** The 6 homology models described in the previous section along with the X-ray structure of the SARS-CoV-2 RBD/hACE2 complex (PDB code 6M0J) were used as input to the CHARMM-GUI server<sup>74</sup>. In the case of the X-ray structure, for consistency with the homology models, the zinc and chloride atoms present in 6M0J were retained while the NAG and water residues were removed. The 7 systems were solvated in a triclinic box of initial x-y-z dimensions of ~13.5 nm \* 9.2 nm \* 8.3 nm. Potassium and chloride ions were added to ensure charge neutrality at a salt concentration of 0.15 M. The total number of atoms was

~104000. Additional details of the systems are reported in Table S4. The CHARMM36m force field<sup>75</sup> was used for the protein and ions and the TIP3P model<sup>76</sup> was used for the water molecules. CHARMM36m force field parameters for the 7 pairs of cysteines linked by disulfide bonds were used. The CHARMM-GUI models were first energy-minimized using the steepest descent algorithm. After minimization, the systems were equilibrated using a 1-ns long simulation in the NPT ensemble followed by a 1-ns long simulation in the NVT ensemble. The temperature T was set at 300 K and the pressure P at 1 atm using the Bussi-Donadio-Parrinello thermostat<sup>77</sup> and the Berendsen barostat<sup>78</sup>, respectively. During equilibration, harmonic restraints on the positions of the protein backbone and sidechains heavy atoms were applied. For each system studied, production simulations were performed in the NVT ensemble for 1  $\mu$ s. A time step of 2 fs was used together with LINCS constraints on h-bonds<sup>79</sup>. The van der Waals interactions were gradually switched off at 1.0 nm and cut off at 1.2 nm; the particle-mesh Ewald method was used to calculate electrostatic interactions with cutoff at 1.2 nm<sup>80</sup>. Production simulations were performed at room temperature for consistency with the BLI experiments.

**Details of the analysis.** To evaluate the stability of the starting model during the production simulations, we calculated the backbone Root Mean Square Deviation (RMSD) with respect to the energy-minimized structure for each frame of the trajectories. The RMSD was calculated separately for the residues in: RBD (Extended data 5A), hACE2 (Extended data 5B), RBD-hACE2 interface (Extended data 5C), and in the entire complex (Extended data 5D). Interfacial residues were defined as the residues in one subunit closer than 0.8 nm to the residues in the other subunit in the X-ray structure of the SARS-Cov-2 RBD/hACE2 complex. RMSD calculations were performed using the *driver* utility of PLUMED v. 2.7<sup>81</sup>. To characterize conformational heterogeneity at the RBD-hACE2 interface, we performed a cluster analysis of the MD trajectories using as similarity metrics the backbone RMSD of the interfacial residues and the GROMOS clustering approach<sup>82</sup> with a cutoff equal to 0.2 nm. The 9 trajectories were first concatenated, clustering was then performed, and finally the population of each cluster was calculated separately for each trajectory (Extended data 5E). To estimate the binding energy between RBD and hACE2, we used the *InterfaceAnalyzer* tool in ROSETTA v. 3.11<sup>83</sup> (Extended data 6A) and the *AnalyseComplex* tool in FoldX v. 4<sup>84</sup> (Extended data 6B). To identify relevant interactions at the RBD and hACE2 interface that could contribute to the binding affinity of the complex, we quantified the frequency of formation of inter-subunits salt bridges (Fig. 3B) and hydrogen bonds (Extended data 7) during the course of the MD simulations. For each frame of the trajectories, we used PLUMED to calculate the distances between the sidechain charged groups of aspartic acids (OD1/OD2), glutamic acids (OE1/OE2), lysines (NZ), and arginines (NH1/NH2). An inter-subunits salt bridge was then defined as formed if the distance between groups with opposite charge was lower than 0.32 nm. We confirmed this calculation using the *Salt Bridges* tool available in VMD<sup>85</sup>. To monitor the formation of inter-subunits hydrogen bonds, we used the *Hydrogen Bond Analysis* module of the MDAnalysis library v. 1.0.0<sup>86</sup>. A donor-acceptor distance and angular cutoffs of 0.3 nm and 150° were used to define the formation of a hydrogen bond. We confirmed this calculation using the *Hydrogen Bonds* tool available in VMD.

## References

49. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 20, 1160–1166 (2019).

50. Lemoine, F. et al. NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Res.* 47, W260–W265 (2019).
51. Lole, K. S. et al. Full-Length Human Immunodeficiency Virus Type 1 Genomes from Subtype C-Infected Seroconverters in India, with Evidence of Intersubtype Recombination. *J. Virol.* 73, 152–160 (1999).
52. Predict. PREDICT One Health Consortium 2013. Protocol for Bat and Rodent Sampling Methods.
53. Sikes, R. S., Gannon, W. L., & the Animal Care and Use Committee of the American Society of Mammalogists. Guidelines of the American Society of Mammalogists for the use of wild mammals in research. *J. Mammal.* 92, 235–253 (2011).
54. Virachith, S. et al. Low seroprevalence of COVID-19 in Lao PDR, late 2020. *Lancet Reg. Health - West. Pac.* 13, 100197 (2021).
55. Francis, C. A Comparison of Mist Nets and Two Designs of Harp Traps for Capturing Bats. *Journal of Mammalogy* 865–970.
56. Hutson, A. M. *Mammals of the Indomalayan Region: A Systematic Review* by G. B. Corbet and J. E. Hill (Oxford University Press, Oxford, and Natural History Museum, London, 1992, ISBN 019 854693 9, 488 pp. HB £60.00). *Oryx* 27, 124–125 (1993).
57. Csorba, G., Ujhelyi, P. & Thomas, N. *Horseshoe Bats of the World (Chiroptera : rhinolophidae)*. Alana Books Bishop's Castle Shropsh. U. K. 160 (2003).
58. Deng, X. et al. Metagenomic sequencing with spiked primer enrichment for viral diagnostics and genomic surveillance. *Nat. Microbiol.* 5, 443–454 (2020).
59. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinforma. Oxf. Engl.* 28, 3150–3152 (2012).
60. Regnault, B. et al. Deep Impact of Random Amplification and Library Construction Methods on Viral Metagenomics Results. *Viruses* 13, 253 (2021).
61. Bratuleanu, B. E. et al. The virome of Rhipicephalus, Dermacentor and Haemaphysalis ticks from Eastern Romania includes novel viruses with potential relevance for public health. *Transbound. Emerg. Dis.* n/a,.
62. Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H. & Frost, S. D. W. Automated Phylogenetic Detection of Recombination Using a Genetic Algorithm. *Mol. Biol. Evol.* 23, 1891–1901 (2006).
63. Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H. & Frost, S. D. W. GARD: a genetic algorithm for recombination detection. *Bioinformatics* 22, 3096–3098 (2006).
64. Crawford, K. H. D. et al. Protocol and Reagents for Pseudotyping Lentiviral Particles with SARS-CoV-2 Spike Protein for Neutralization Assays. *Viruses* 12, E513 (2020).
65. Anna, F. et al. High seroprevalence but short-lived immune response to SARS-CoV-2 infection in Paris. *Eur. J. Immunol.* 51, 180–190 (2021).
66. Corman, V. M. et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill. Bull. Eur. Sur Mal. Transm. Eur. Commun. Dis. Bull.* 25, (2020).

67. Monteil, V. et al. Inhibition of SARS-CoV-2 Infections in Engineered Human Tissues Using Clinical-Grade Soluble Human ACE2. *Cell* 181, 905-913.e7 (2020).
68. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* 66, 125–132 (2010).
69. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674 (2007).
70. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132 (2004).
71. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. Sect. Struct. Biol.* 75, 861–877 (2019).
72. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815 (1993).
73. Shen, M.-Y. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci. Publ. Protein Soc.* 15, 2507–2524 (2006).
74. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* 29, 1859–1865 (2008).
75. Huang, J. et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* 14, 71–73 (2017).
76. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935 (1983).
77. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126, 014101 (2007).
78. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690 (1984).
79. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* 18, 1463–1472 (1997).
80. Essmann, U. et al. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103, 8577–8593 (1995).
81. PLUMED consortium. Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* 16, 670–673 (2019).
82. Daura, X. et al. Peptide folding: When simulation meets experiment. *Angew. Chem.-Int. Ed. Engl.* 38, 236–240 (1999).
83. Stranges, P. B. & Kuhlman, B. A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Protein Sci. Publ. Protein Soc.* 22, 74–82 (2013).
84. Schymkowitz, J. et al. The FoldX web server: an online force field. *Nucleic Acids Res.* 33, W382-388 (2005).
85. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38, 27–28 (1996).

86. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* 32, 2319–2327 (2011).

### **Acknowledgments:**

The authors want to thank Souand Mohamed Ali, Nicolas Da Rocha, Angela Brisebarre, Thonglakhone Xaybounsou, and Somsanith Chonephetsarath for their help at the bench. Next-generation sequencing was performed with the help of Biomix Platform, C2RT, Institut Pasteur, Paris, France, supported by France Génomique (ANR-10-INBS-09-09), IBISA, and the Illumina COVID-19 Projects' offer. The work was granted access to the HPC resources of IDRIS under the allocation 2020-101592 made by GENCI. We thank the Ministry of Health and the Ministry of Natural Resources and Environments, Lao PDR, for their authorization of the field work and the Faculty of Environmental Science for its authorization of the field research collaboration. The work was funded by an Institut Pasteur "Covid Taskforce" and in part by the H2020 project 101003589 (RECOVER) and Labex IBEID (ANR-10-LABX62-IBEID) grants. Field and laboratory work at IP-Laos was also funded by a UK embassy grant (Grant No. INT 2021/LOV C19 02) and Luxembourg Development special grant (Grant No. LAO/030•202324).

### **Author contributions**

ST, KV, PTB, ME conceived the study design.

KV, BD, KL, NPDS, VX, PPa carried out sample procurement and bat species identification.

SM, FD performed virus isolation, entry, and neutralization assays.

BR, DC, PP prepared the NGS libraries and carried out the next-generation sequencing.

ST, BR, TB carried out the genome assembly, recombination, and phylogenetic analyses.

VL, SS, KL, NP implemented the pan-coronavirus PCR testing.

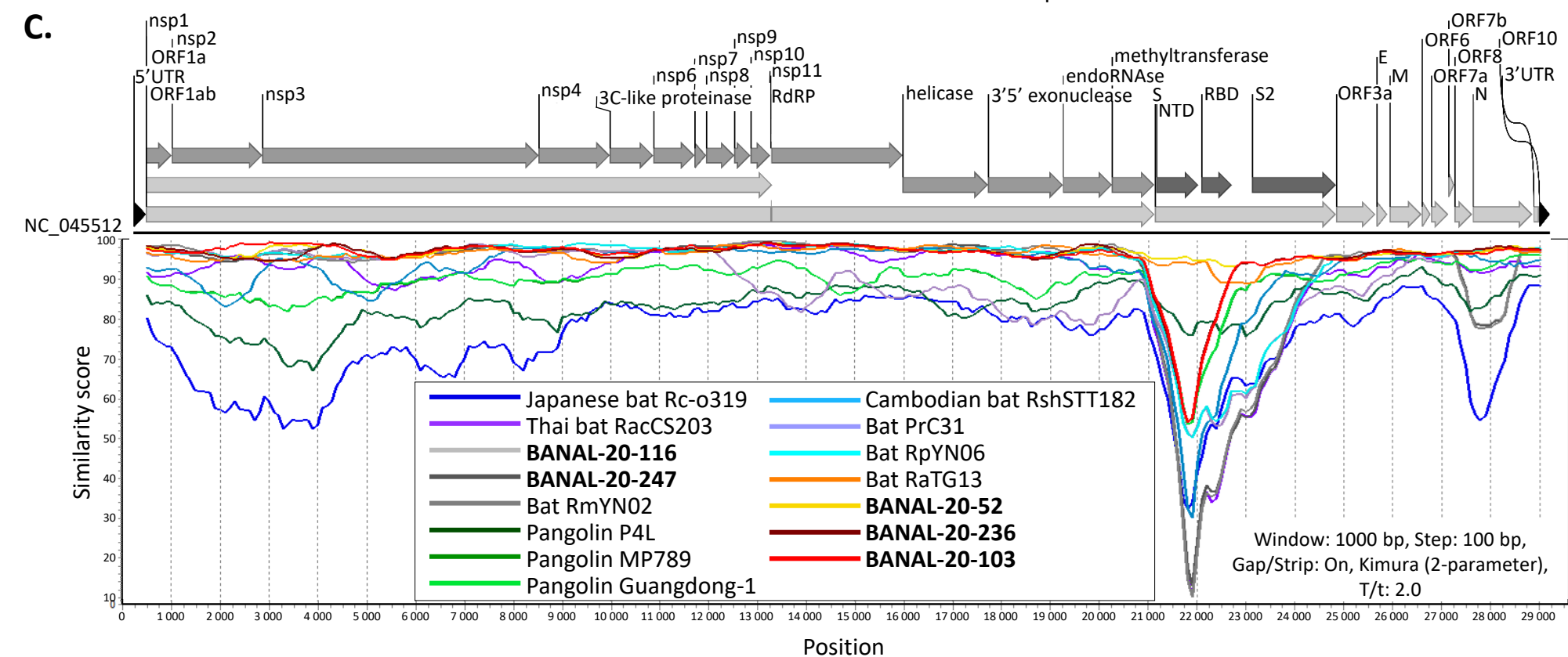
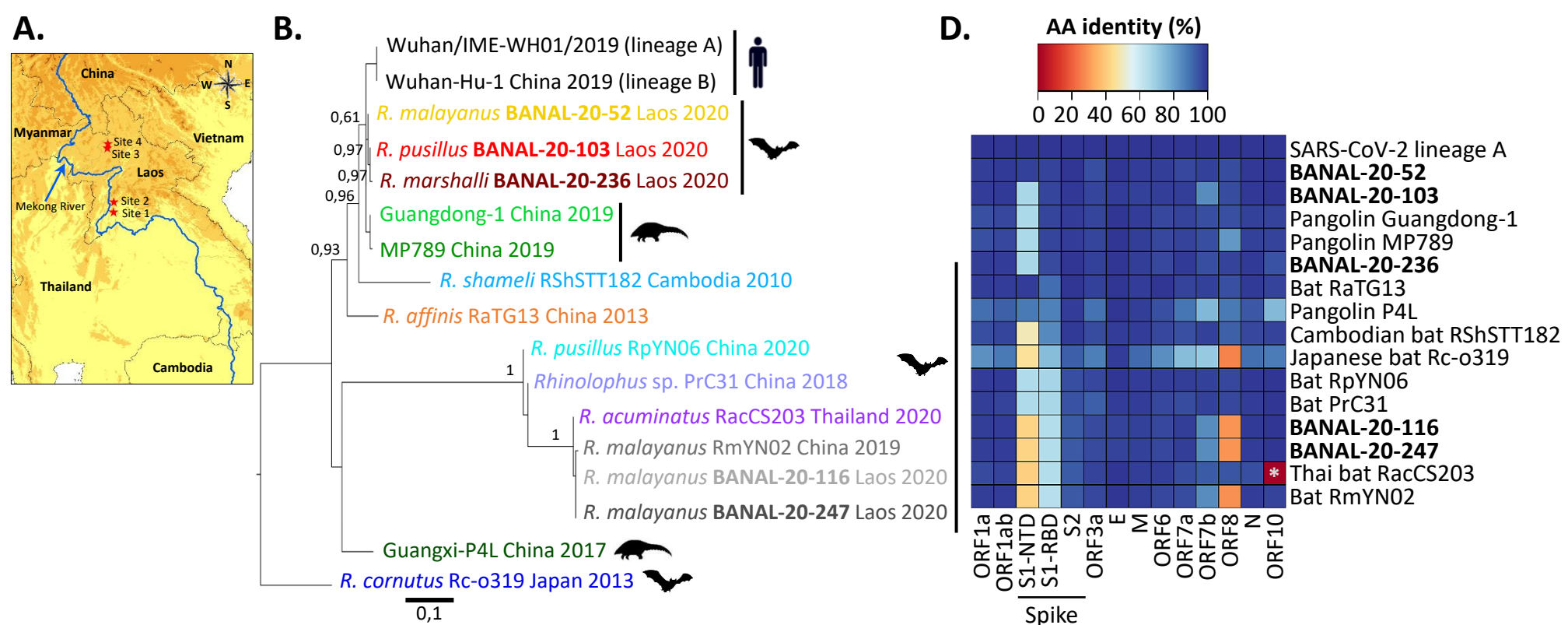
EBS and FR performed and analyzed the structure and binding studies.

MB, YK, MN performed and analyzed the molecular dynamic simulations.

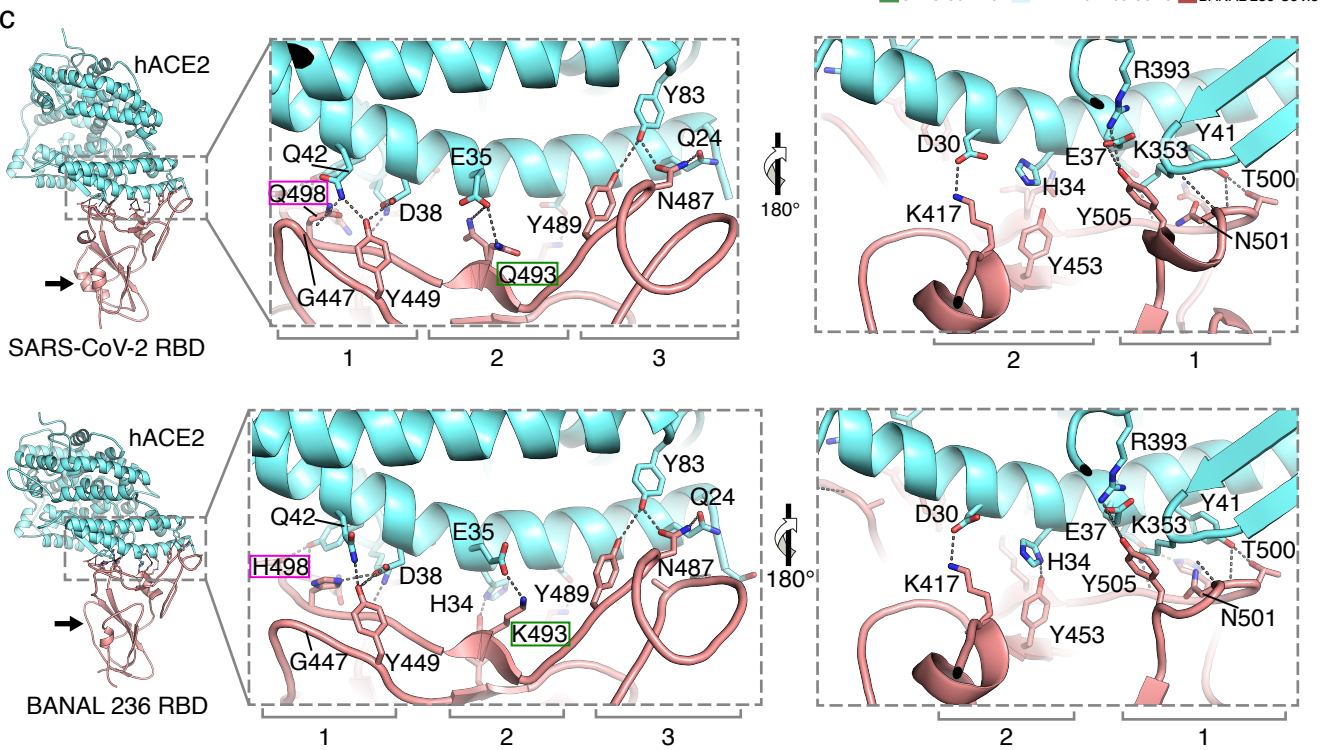
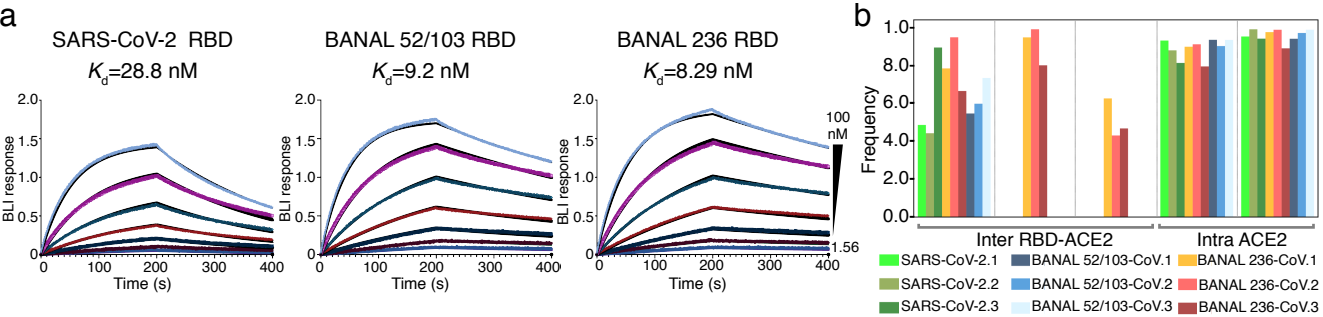
ME and ST wrote the manuscript with inputs from all other authors.

### **Competing interests**

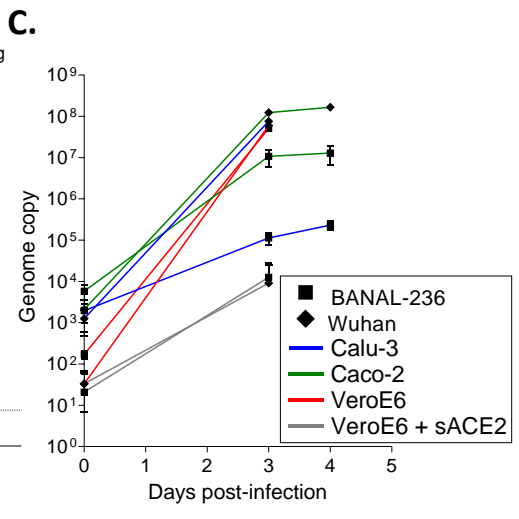
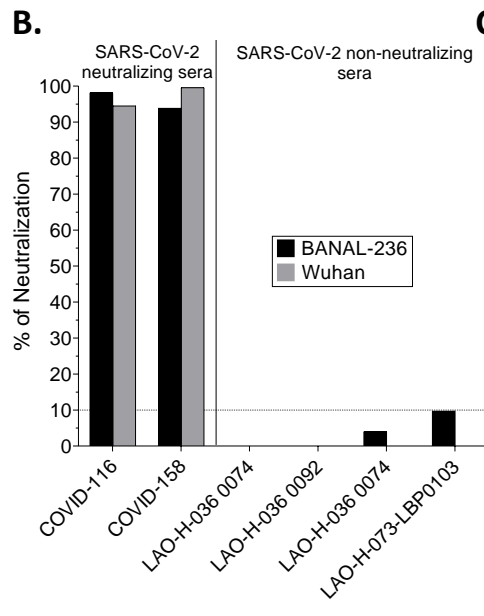
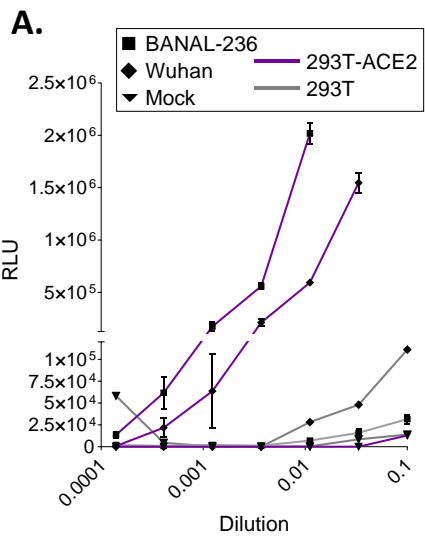
All authors of the manuscript are cited as inventors of a U.S. Provisional patent application number 63/240.535 filed on September 03, 2021 entitled: NEW ISOLATED BAT SARS-COV ABLE TO INFECT HUMAN CELLS AND HARBORING RECEPTOR-BINDING DOMAINS CLOSE TO SARS-COV-2. The patent covers medical applications (diagnostic, vaccination) of the sequences described in the manuscript, more specifically in the part " Diversity of bat and coronavirus species"





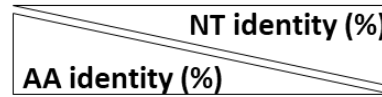






# Extended data 1.

## Similarity scale



S1-NTD		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	SARS-CoV-2 lineage A (EPI_ISL_529213)	-	100,00	94,29	94,52	64,50	64,50	47,13	47,13	46,46	46,79	63,36	63,47	54,36	64,16	64,16	77,63	51,01
2	SARS-CoV-2 lineage B (NC_045512)	100,00	-	94,29	94,52	64,50	64,50	47,13	47,13	46,46	46,79	63,36	63,47	54,36	64,16	64,16	77,63	51,01
3	Bat RaTG13 (MN996532)	98,64	98,64	-	93,04	65,64	65,64	47,92	47,92	47,24	47,13	63,93	63,70	53,79	64,27	64,27	76,83	50,90
4	BANAL20-52	97,96	97,96	99,32	-	66,44	65,98	48,71	48,71	48,03	48,37	64,27	64,16	54,93	65,41	65,41	77,85	50,45
5	BANAL-20-236	67,12	67,12	67,80	68,14	-	97,33	51,31	51,31	50,97	50,06	78,40	77,93	54,95	85,83	85,83	65,29	49,66
6	BANAL-20-103	67,12	67,12	67,80	68,14	99,65	-	50,86	50,86	50,74	49,60	78,86	78,63	54,84	86,18	86,18	65,64	49,54
7	BANAL-20-116	41,95	41,95	42,28	42,62	45,58	45,24	-	99,88	97,79	93,48	50,63	50,51	50,63	50,63	50,63	46,73	53,60
8	BANAL-20-247	41,95	41,95	42,28	42,62	45,58	45,24	99,63	-	97,66	93,36	50,63	50,51	50,63	50,63	50,63	46,73	53,60
9	Bat RmYN02 (EPI_ISL_412977)	41,61	41,61	41,95	42,28	44,56	44,22	98,16	97,79	-	94,22	50,51	50,40	50,52	50,17	50,17	46,50	54,08
10	Thai bat RacCS203 (MW251308)	40,94	40,94	41,28	41,61	44,56	44,22	95,22	95,22	94,85	-	50,63	50,29	50,17	50,06	50,06	47,07	53,72
11	Bat RpYN06 (EPI_ISL_1699446)	65,42	65,42	65,76	65,42	84,08	84,43	43,20	43,20	42,86	42,18	-	97,10	52,76	80,14	80,14	64,72	50,80
12	Bat PrC31 (EPI_ISL_1098866)	65,76	65,76	66,10	65,76	83,74	84,08	42,86	42,86	42,52	41,84	98,96	-	53,11	79,79	79,79	64,95	50,80
13	Cambodian bat RShSTT182 (EPI_ISL_852604)	48,48	48,48	49,16	49,49	47,26	46,92	46,05	46,05	45,70	44,67	47,26	47,60	-	55,18	55,18	53,41	47,11
14	Pangolin Guangdong-1 (EPI_ISL_410721)	66,78	66,78	67,46	67,12	94,46	94,46	44,22	44,22	43,54	43,20	85,81	85,47	47,26	-	100,00	66,09	50,69
15	Pangolin MP789 (MT121216)	66,78	66,78	67,46	67,12	94,46	94,46	44,22	44,22	43,54	43,20	85,81	85,47	47,26	100,00	-	66,09	50,69
16	Pangolin P4L (EPI_ISL_410538)	87,76	87,76	88,44	88,78	66,55	66,55	42,91	42,91	42,57	41,89	63,82	64,16	48,47	66,55	66,55	-	50,17
17	Japanese bat Rc-o319 (LC556375)	45,64	45,64	45,97	45,64	43,69	43,34	52,88	52,88	53,24	52,88	44,37	44,03	39,12	44,71	44,71	45,61	-

NT identity (%)

NT identity (%)

NT identity (%)

### AA identity (%)

S1-RBD		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	SARS-CoV-2 lineage A (EPI_ISL_529213)	-	100,00	86,25	93,87	93,87	94,02	60,99	60,84	61,43	60,99	62,03	61,29	77,43	86,85	86,70	79,82	68,46
2	SARS-CoV-2 lineage B (NC_045512)	100,00	-	86,25	93,87	93,87	94,02	60,99	60,84	61,43	60,99	62,03	61,29	77,43	86,85	86,70	79,82	68,46
3	Bat RaTG13 (MN996532)	90,18	90,18	-	86,40	86,85	86,85	60,69	60,54	60,24	60,84	62,63	62,03	76,23	82,06	81,91	79,22	68,46
4	BANAL20-52	97,32	97,32	90,62	-	99,40	99,55	61,58	61,58	61,73	61,29	63,38	63,23	77,58	87,74	87,59	79,07	68,76
5	BANAL-20-236	96,43	96,43	91,07	99,11	-	99,85	61,88	61,88	62,03	61,58	63,68	63,53	77,58	87,74	87,59	79,37	69,36
6	BANAL-20-103	96,88	96,88	91,07	99,55	99,55	-	61,73	61,73	61,88	61,43	63,53	63,38	77,73	87,89	87,74	79,37	69,21
7	BANAL-20-116	64,73	64,73	63,84	64,29	64,29	64,29	-	99,84	99,18	95,59	78,76	79,25	62,39	59,94	59,79	62,03	60,44
8	BANAL-20-247	64,73	64,73	63,84	64,29	64,29	64,29	100,00	-	99,02	95,42	78,92	79,08	62,23	59,79	59,64	61,88	60,28
9	Bat RmYN02 (EPI_ISL_412977)	64,29	64,29	63,39	63,84	63,84	63,84	99,51	99,51	-	95,42	78,76	79,25	62,23	59,79	59,64	61,88	60,90
10	Thai bat RacCS203 (MW251308)	64,73	64,73	63,84	64,29	64,29	64,29	99,51	99,51	99,02	-	80,23	80,39	61,16	59,79	59,64	63,23	61,06
11	Bat RpYN06 (EPI_ISL_1699446)	67,41	67,41	66,07	66,96	66,96	66,96	89,81	89,81	89,32	90,29	-	94,93	61,31	63,98	63,83	63,68	63,86
12	Bat PrC31 (EPI_ISL_1098866)	66,96	66,96	65,62	66,52	66,52	66,52	90,78	90,78	90,29	91,26	99,02	-	61,47	63,23	63,08	62,78	63,40
13	Cambodian bat RShSTT182 (EPI_ISL_852604)	84,38	84,38	82,59	85,27	84,82	85,27	64,09	64,09	64,09	64,09	63,80	64,25	-	76,38	76,23	72,05	65,02
14	Pangolin Guangdong-1 (EPI_ISL_410721)	96,88	96,88	90,18	98,66	97,77	98,21	64,73	64,73	64,29	64,73	67,41	66,96	84,82	-	99,85	78,92	69,96
15	Pangolin MP789 (MT121216)	96,88	96,88	89,73	98,21	97,32	97,77	64,29	64,29	63,84	64,29	66,96	66,52	84,38	99,55	-	78,77	69,96
16	Pangolin P4L (EPI_ISL_410538)	86,61	86,61	87,95	87,95	87,95	87,95	64,29	64,29	63,84	64,73	66,52	66,52	79,91	88,39	87,95	-	69,51
17	Japanese bat Rc-o319 (LC556375)	73,66	73,66	76,34	75,89	76,79	76,34	63,76	63,76	63,30	63,76	64,68	64,22	73,54	75,89	75,45	74,11	-

### AA identity (%)

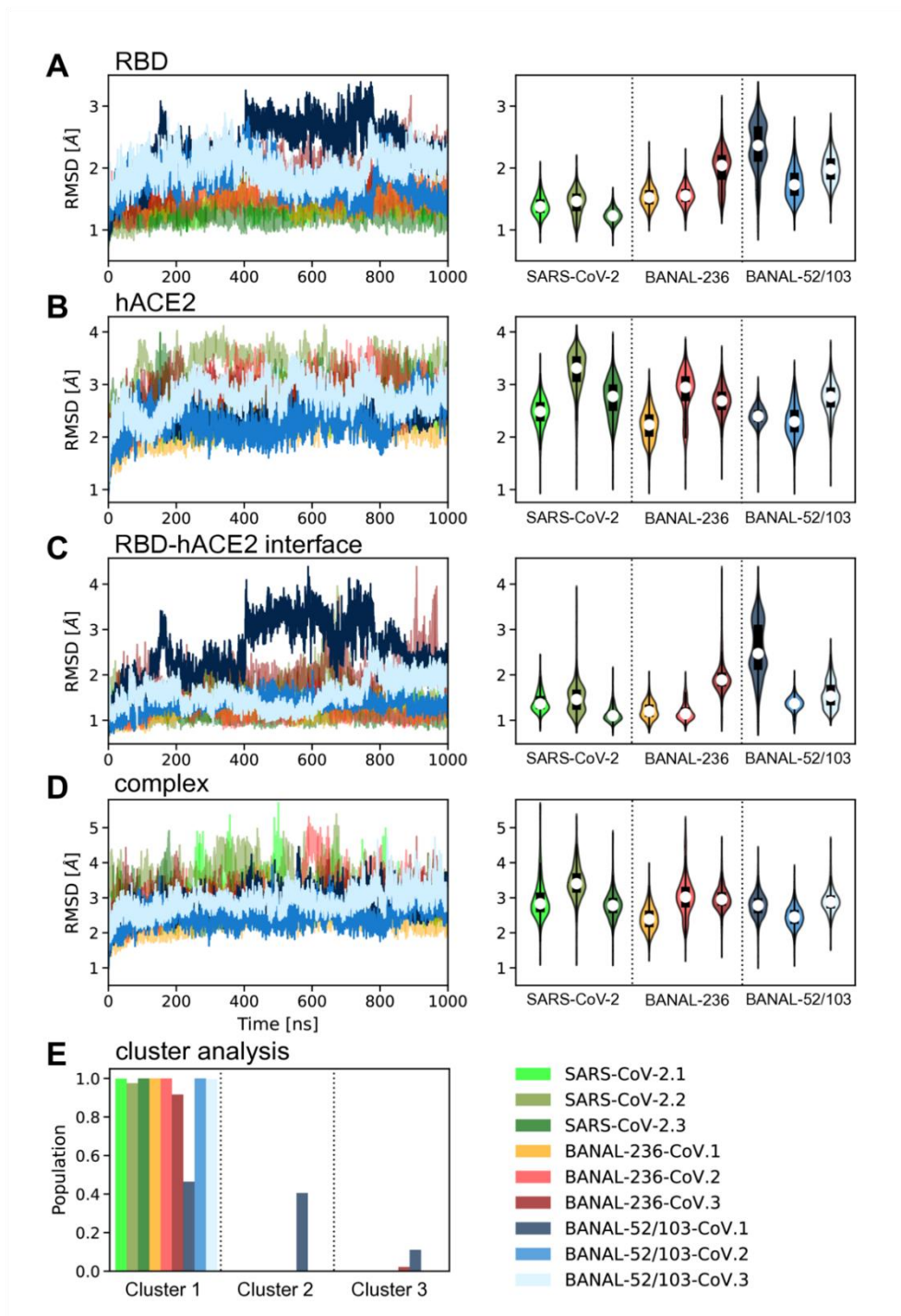
S2		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	SARS-CoV-2 lineage A (EPI_ISL_529213)	-	100,00	94,91	95,59	95,42	95,19	89,81	89,87	90,21	87,32	88,91	85,51	92,93	91,11	91,06	88,00	81,64
2	SARS-CoV-2 lineage B (NC_045512)	100,00	-	94,91	95,59	95,42	95,19	89,81	89,87	90,21	87,32	88,91	85,51	92,93	91,11	91,06	88,00	81,64
3	Bat RaTG13 (MN996532)	99,66	99,66	-	95,93	95,76	95,47	88,68	88,68	88,91	87,66	90,10	86,30	92,30	91,74	91,68	87,83	82,03
4	BANAL20-52	99,66	99,66	99,32	-	99,49	98,81	89,19	89,25	89,25	86,98	89,81	86,36	92,87	91,57	91,51	87,72	82,37
5	BANAL-20-236	99,66	99,66	99,32	99,66	-	99,09	89,02	89,08	89,08	86,93	89,76	86,47	92,81	91,51	91,45	87,89	82,26
6	BANAL-20-103	99,49	99,49	99,49	99,49	99,49	-	88,91	88,96	89,19	86,64	89,19	86,36	92,47	91,91	91,85	87,61	82,32
7	BANAL-20-116	93,20	93,20	93,20	93,03	93,03	92,86	-	99,77	98,41	93,77	88,51	84,72	86,64	85,00	84,95	83,81	79,94
8	BANAL-20-247	93,37	93,37	93,20	93,20	93,20	93,03	99,66	-	98,41	93,77	88,57	84,72	86,64	85,00	84,95	83,76	80,00
9	Bat RmYN02 (EPI_ISL_412977)	93,37	93,37	93,71	93,20	93,20	93,37	99,15	99,15	-	94,05	88,62	84,83	86,93	85,46	85,40	84,15	80,45
10	Thai bat RacCS203 (MW251308)	93,20	93,20	93,54	93,03	93,03	93,20	98,81	98,81	99,32	-	87,38	84,78	85,06	85,23	85,17	83,76	80,17
11	Bat RpYN06 (EPI_ISL_1699446)	95,41	95,41	95,07	95,07	95,07	94,90	93,03	93,20	93,03	93,37	-	91,00	87,38	85,96	85,91	85,00	82,03
12	Bat PrC31 (EPI_ISL_1098866)	94,56	94,56	94,22	94,22	94,22	94,05	91,84	91,84	91,67	91,84	97,96	-	85,40	84,72	84,66	84,89	82,26
13	Cambodian bat RShSTT182 (EPI_ISL_852604)	99,15	99,15	98,98	98,81	98,81	98,64	92,86	93,03	93,03	92,86	94,73	93,88	-	90,38	90,32	87,61	81,58
14	Pangolin Guangdong-1 (EPI_ISL_410721)	98,47	98,47	98,81	98,13	98,13	98,30	92,35	92,35	92,86	93,37	95,41	94,73	97,96	-	99,94	87,10	81,58
15	Pangolin MP789 (MT121216)	98,47	98,47	98,81	98,13	98,13	98,30	92,35	92,35	92,86	93,37	95,41	94,73	97,96	100,00	-	87,15	81,64
16	Pangolin P4L (EPI_ISL_410538)	98,47	98,47	98,81	98,13	98,13	98,30	92,86	92,69	93,20	93,37	95,07	94,39	97,96	98,47	98,47	-	80,85
17	Japanese bat Rc-o319 (LC556375)	92,87	92,87	93,04	92,70	92,70	92,87	89,64	89,81	89,81	89,98	93,04	94,23	92,36	93,04	93,04	93,38	-

### AA identity (%)

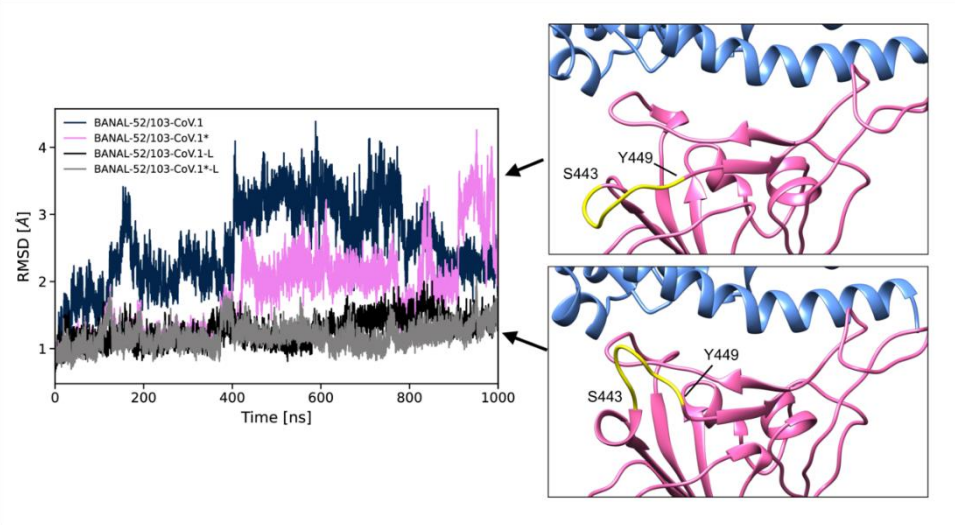




Extended data 4.

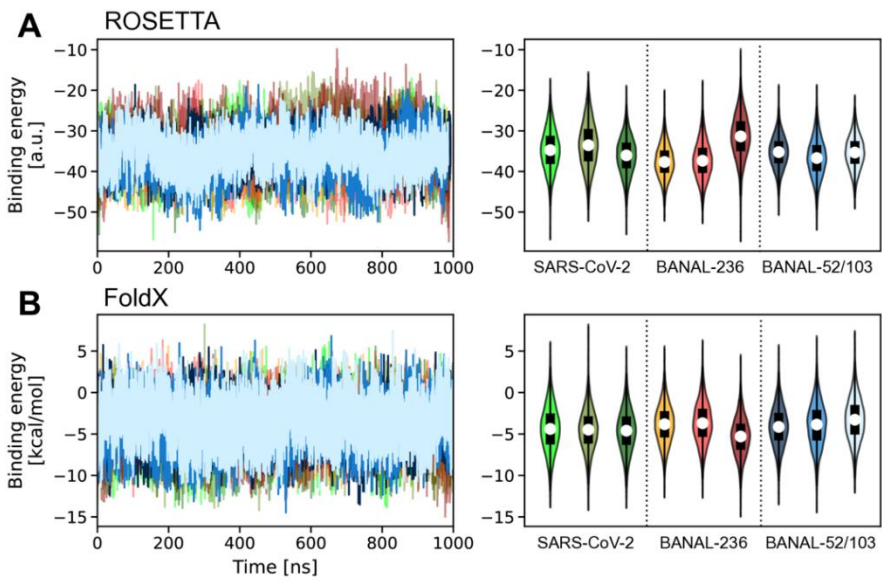


Extended data 5.

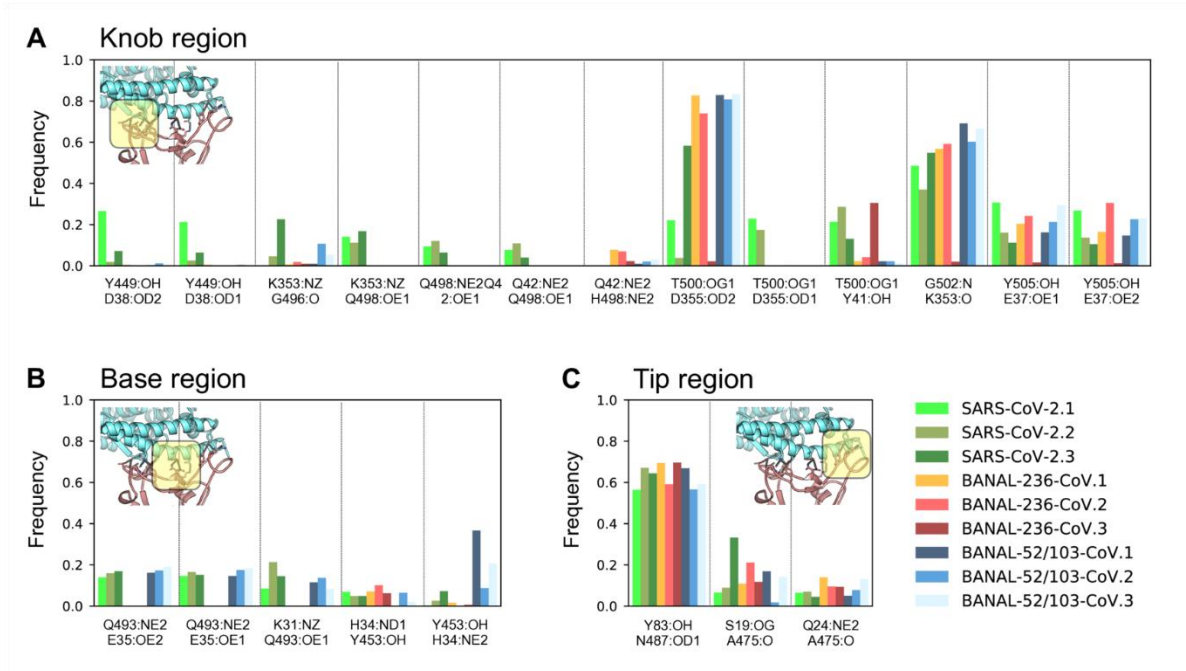




Extended data 6.

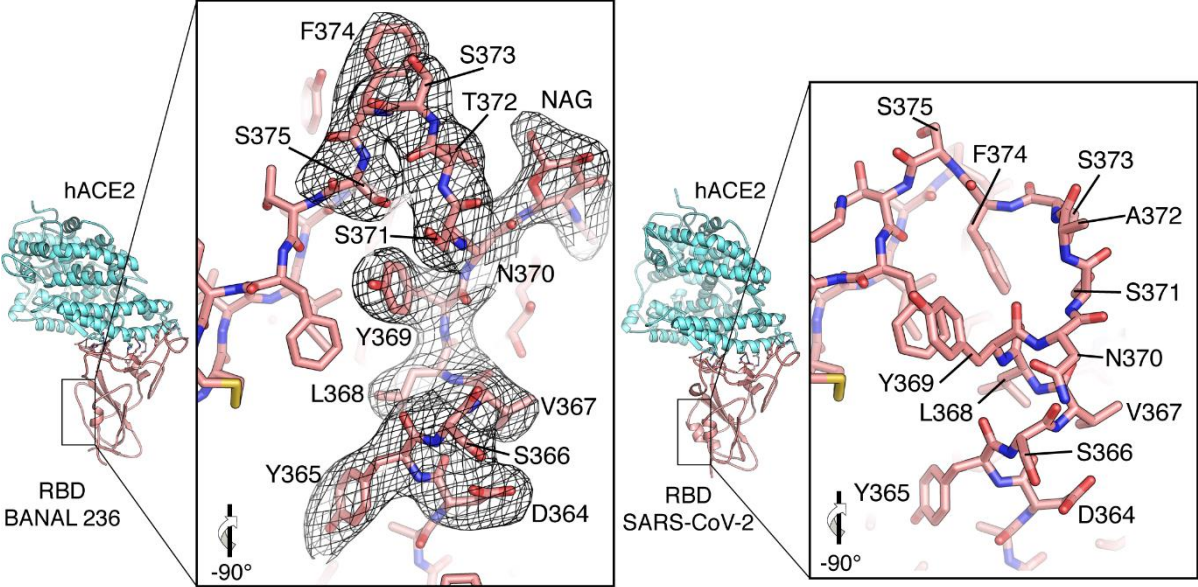


## Extended data 7.



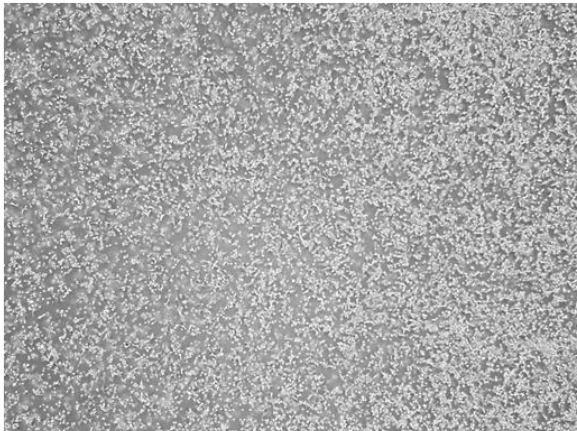


Extended data 8.

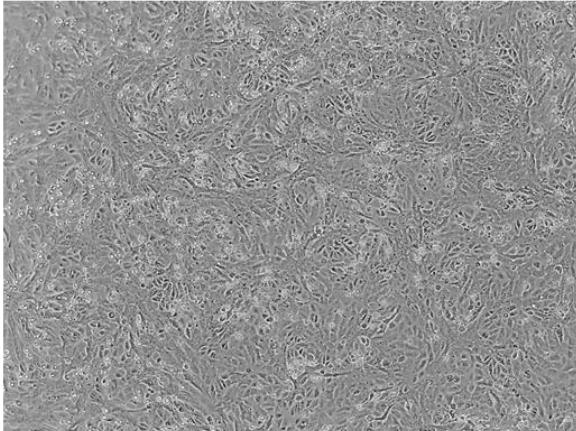


Extended data 9.

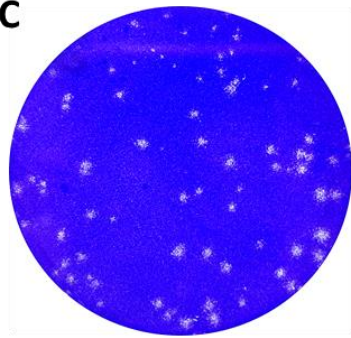
**A**



**B**



**C**

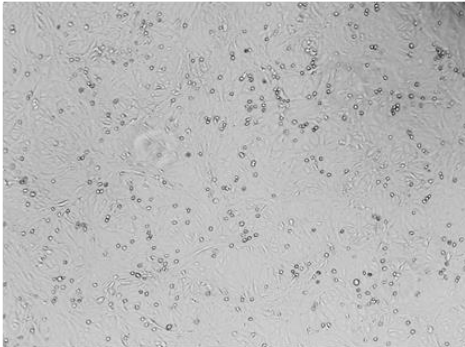
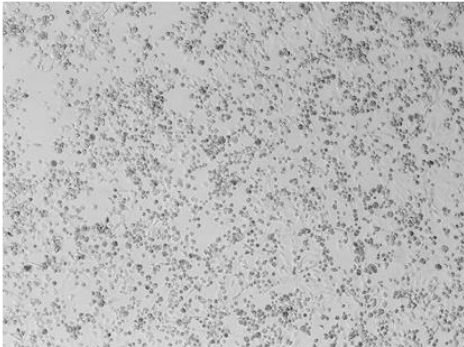


**D**

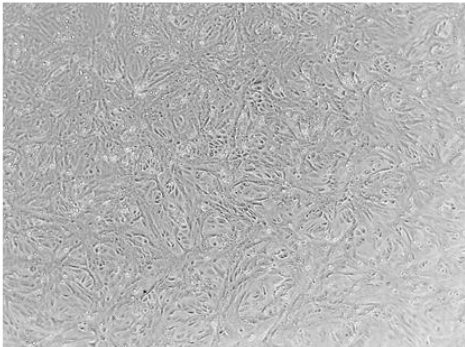
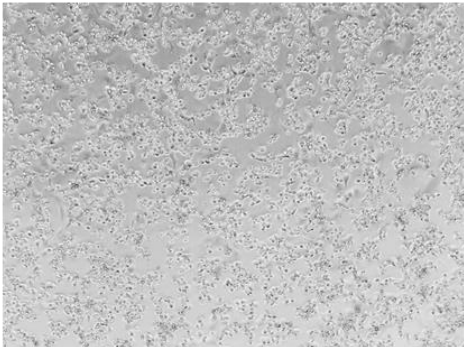
Control

+ soluble ACE2 (25 µg/mL)

Wuhan



BANAL-236



## Bat coronaviruses related to SARS-CoV-2 and infectious for human cells

Sarah Temmam\* (1,2), Khamsing Vongphayloth\* (3), Eduard Baquero Salazar\* (4), Sandie Munier\* (5), Massimiliano Bonomi\* (6), Béatrice Regnault (1,2), Bounsavane Douangboubpha (7), Yasaman Karami (6), Delphine Chrétien (1,2), Daosavanh Sanamxay (7), Vilakhan Xayaphet (7), Phetphoumin Paphaphanh (7), Vincent Lacoste (3), Somphavanh Somlor (3), Khaithong Lakeomany (3), Nothasin Phommavanh (3), Philippe Pérot (1,2), Océane Dehan (5,8), Faustine Amara (5), Flora Donati (5,8), Thomas Bigot (1,9), Michael Nilges (6), Félix A. Rey (4), Sylvie van der Werf (5,8), Paul T. Brey (3), Marc Eloit + (1,2,10).

\* co-first authors

+ corresponding author

## SI guide

<b>Supplementary Methods</b> .....	2
<b>Supplementary Tables</b> .....	3
<b>Table S1: Total number of bats collected and morphological identification.</b> ....	3
<b>Table S2: List of study sites in Laos with coordinates</b> .....	5
<b>Table S3: Results from pan-coronavirus RT nested PCR and Sanger sequencing</b> .....	6
<b>Table S4. Vertical coverage of BANAL genomes presented in the study</b> .....	8
<b>Table S5. Characteristics of the recombination breakpoints</b> .....	9
<b>Table S6. List of the MD simulations performed.</b> .....	10
<b>Table S8: Acknowledgment of sharing of SARS-CoV-2-related genome sequences from the GISAID database related to Figures 1-2 and Supp. Figures 1-4.</b> .....	12
<b>Supplementary Figures</b> .....	18
<b>Figure S1: Identity matrices at the genus level of representative sarbecoviruses</b> .....	18
<b>Figure S2: Complete phylogenetic reconstructions of recombinant fragments, as described in Figure 2.</b> .....	22
<b>Figure S3: Alignment of the whole spike amino-acid sequences.</b> .....	39
<b>Figure S4: Western-Blot analysis of concentrated lentiviral particles pseudotyped with Wuhan or BANAL-236 spike proteins.</b> .....	41
<b>Figure S5: Distribution of bats species from which SARS-CoV-2-like viruses were detected in this study.</b> .....	43
<b>References</b> .....	44

## Supplementary Methods

**Additional analysis of the MD simulations of the BANAL-52/103 RBD/hACE2 complex.** One of our 3 simulations of the BANAL-52/103 RBD/hACE2 complex (BANAL-52/103-CoV.1 in Extended data 5) showed large fluctuations of the RBD and RBD-hACE2 interface RMSDs. As a consequence of this larger conformational variability, the cluster analysis of BANAL-52/103-CoV.1 identified 2 major populated clusters (Extended data 5E). To exclude the possibility of having initialized the simulation from a poorly equilibrated structure, we performed a 10 times longer equilibration phase of the initial model followed by an additional 1  $\mu$ s-long production simulation (BANAL-52/103-CoV.1\*). The time series of the RBD-hACE2 interface RMSD in this additional simulation (Extended data 8, pink line) still displayed large RMSD fluctuations, especially towards the end of the simulation. After careful inspection of the trajectory, we concluded that the cause of the observed RMSD behavior in both BANAL-52/103-CoV.1 and BANAL-52/103-CoV.1\* was the structural variability of the loop formed by RBD residues S443 to Y449 (Extended data 8, insets, yellow residues). According to our definition, these residues were considered at the RBD/hACE2 interface; however, only a transiently formed hydrogen bond maintains this loop close to hACE2 (Extended data 7A). The removal of this stretch of 7 residues (out of 110 interfacial residues) from the calculation of the interface RMSD (Extended data 8, black and gray lines) clearly shows that the motion of the S443-Y449 loop was the cause of the large RMSD fluctuations observed in the BANAL-52/103-CoV.1 simulation.

## Supplementary Tables

Table S1: Total number of bats collected and morphological identification.

Family	Genus	Species	Vientiane Province		Oudomxay Province		Total No. of bats
			Fueng (Site 1)	Meth (Site 2)	Xay (Site 3)	Na Mor (Site 4)	
<b>Hipposideridae</b>	<b>Aselliscus</b>	<i>stoliczkanus</i>	5	5	30		40
	<b>Hipposideros</b>	<i>cineraceus</i>	14	3			17
		<i>khaokhouayensis</i>	12	4			16
		<i>larvatus</i>	47				47
		<i>gentilis</i>	55	7	1		63
		<i>rotalis</i>	1				1
<b>Megadermatidae</b>	<b>Megaderma</b>	<i>spasma</i>	3	6		2	11
<b>Molossidae</b>	<b>Chaerephon</b>	<i>plicatus</i>	45				45
<b>Pteropodidae</b>	<b>Cynopterus</b>	<i>horsfieldii</i>			2		2
		<i>sphinx</i>		10	1	1	12
	<b>Eonycteris</b>	<i>spelaea</i>	1	9		1	11
	<b>Macroglossus</b>	<i>sobrinus</i>	5	45	7	3	60
	<b>Megaerops</b>	<i>niphanae</i>		1		1	2
	<b>Sphaerias</b>	<i>blanfordi</i>		1			1
<b>Rhinolophidae</b>	<b>Rhinolophus</b>	<i>affinis</i>	2				2
		<i>coelophyllus</i>	10	2			12
		<i>luctus</i>	1				1
		<i>malayanus</i>	46	1	9	1	57
		<i>marshalli</i>	1				1
		<i>microglobosus</i>	5	1	1		7
		<i>pearsonii</i>	46	19	1		66
		<i>pusillus</i>	7	11			18
		<i>rex</i>			2		2
		<i>shameli</i>			1		1
		<i>siamensis</i>			6	4	10
		sp.	2				2
		<i>thomasi</i>	1	5	2		8
<b>Vespertilionidae</b>	<b>Arielulus</b>	<i>aureocollaris</i>			1		1
	<b>Glischropus</b>	<i>bucephalus</i>		2	2		4
	<b>la</b>	<i>io</i>			1		1
	<b>Kerivoula</b>	<i>cf.hardwickii</i>		9			9
		<i>cf.titania</i>		1			1
		<i>depressa</i>			8		8
		<i>hardwickii</i>	5				5
	<b>Murina</b>	<i>feae</i>	1	1	1		3
		sp1		4	2		6
		sp2		1	2		3
	<b>Myotis</b>	sp.	7				7

		sp1			45		45
		sp2			3		3
		sp3			1		1
	<b><i>Pipistrellus</i></b>	sp.	8				8
	<b><i>Scotomanes</i></b>	<i>ornatus</i>			1		1
	<b><i>Tylonycteris</i></b>	<i>fulvida</i>		3			3
		<i>malayana</i>		14			14
		sp.	5		2		7
<b>Grand Total</b>			<b>335</b>	<b>165</b>	<b>132</b>	<b>13</b>	<b>645</b>

Table S2: List of study sites in Laos with coordinates

Province	District	Village	Locality	Latitude	Longitude
Vientiane Province	Site 1 : Fueng	Na Ang	Pha Tong	18° 32.879'N	101° 58.938'E
			Tham Nam cave	18° 32.914'N	101° 58.459'E
			Tham Pha	18° 34.018'N	101° 58.338'E
		Noon Hin Hae Pha Louang	Pha Nouk Kok	18° 31.743'N	101° 58.648'E
			Tham Pha Louang	18° 30.495'N	101° 57.421'E
	Site 2 : Meth	Had Yao	Pha Keo Phong Mon	18° 54.426'N	101° 58.789'E
			Pha Khee Chear (Keo Tao)	18° 54.898'N	101° 59.048'E
		Houay Pa Mak	Pha Mum	18° 53.717'N	101° 56.957'E
			Pha Mum, Keo Tao	18° 54.787'N	101° 58.292'E
			Pha Yao, Keo Tao	18° 55.386'N	101° 57.723'E
Oudomxay Province	Site 3 : Xay	Chom Ong	Chom Ong Cave	20° 43.071'N	101° 45.811'E
			Houay Om Loung	20° 43.111'N	101° 46.577'E
			Houay Pak Veuk	20° 41.647'N	101° 47.179'E
			Pa Noo Khom	20° 44.143'N	101° 46.770'E
			Site 4 : Na Mor	Na Thong	Nam Ook Hoo Cave
		20° 51.907'N			101° 46.926'E

Table S3: Results from pan-coronavirus RT nested PCR and Sanger sequencing

Anal Swab ID	Bat Genus	Bat Species	Genus	Subgenus	Closest sequence	% blastN similarity	District	Village	Locality
BANAL-20-27	<i>Rhinolophus</i>	<i>pusillus</i>	co-infection <i>Alpha/Betacoronavirus</i>	unclassified <i>Decacovirus/ Sarbecovirus</i>	Rhinolophus bat coronavirus HKU32 / SARS-CoV-2	80%	Feuang (Site 1)	Nonhinhae	Pha Nok Kok
BANAL-20-51	<i>Hipposideros</i>	<i>pomona</i>	<i>Alphacoronavirus</i>	<i>Decacovirus</i>	Bat coronavirus HKU10	98%	Feuang (Site 1)	Nonhinhae	Pha Nok Kok
BANAL-20-52	<i>Rhinolophus</i>	<i>malayanus</i>	<i>Betacoronavirus</i>	<i>Sarbecovirus</i>	SARS-CoV-2	99%	Feuang (Site 1)	Nonhinhae	Pha Nok Kok
BANAL-20-103	<i>Rhinolophus</i>	<i>pusillus</i>	<i>Betacoronavirus</i>	<i>Sarbecovirus</i>	SARS-CoV-2	99%	Feuang (Site 1)	Na Ang	Tham Pha
BANAL-20-115	<i>Hipposideros</i>	<i>pomona</i>	<i>Alphacoronavirus</i>	<i>Decacovirus</i>	Bat coronavirus HKU10	98%	Feuang (Site 1)	Pha Luang	Tham Pha Luang
BANAL-20-116	<i>Rhinolophus</i>	<i>malayanus</i>	<i>Betacoronavirus</i>	<i>Sarbecovirus</i>	SARS-CoV-2	98%	Feuang (Site 1)	Pha Luang	Tham Pha Luang
BANAL-20-178	<i>Hipposideros</i>	<i>pomona</i>	<i>Alphacoronavirus</i>	<i>Decacovirus</i>	Bat coronavirus HKU10	98%	Feuang (Site 1)	Pha Luang	Tham Pha Luang
BANAL-20-191	<i>Chaerephon</i>	<i>plicatus</i>	<i>Alphacoronavirus</i>	<i>Pedacovirus</i>	Porcine epidemic diarrhea virus (PEDV)	93%	Feuang (Site 1)	Pha Luang	Tham Pha Luang
BANAL-20-197	<i>Chaerephon</i>	<i>plicatus</i>	<i>Alphacoronavirus</i>	<i>Pedacovirus</i>	Porcine epidemic diarrhea virus (PEDV)	93%	Feuang (Site 1)	Pha Luang	Tham Pha Luang
BANAL-20-212	<i>Chaerephon</i>	<i>plicatus</i>	<i>Alphacoronavirus</i>	<i>Pedacovirus</i>	Porcine epidemic diarrhea virus (PEDV)	93%	Feuang (Site 1)	Pha Luang	Tham Pha Luang
BANAL-20-213	<i>Chaerephon</i>	<i>plicatus</i>	<i>Alphacoronavirus</i>	<i>Pedacovirus</i>	Porcine epidemic diarrhea virus (PEDV)	93%	Feuang (Site 1)	Pha Luang	Tham Pha Luang
BANAL-20-236	<i>Rhinolophus</i>	<i>marshalli</i>	<i>Betacoronavirus</i>	<i>Sarbecovirus</i>	SARS-CoV-2	99%	Feuang (Site 1)	Na Ang	Pha Tong
BANAL-20-242	<i>Rhinolophus</i>	<i>malayanus</i>	<i>Betacoronavirus</i>	<i>Sarbecovirus</i>	SARS-CoV-2	99%	Feuang (Site 1)	Na Ang	Pha Tong
BANAL-20-247	<i>Rhinolophus</i>	<i>malayanus</i>	<i>Betacoronavirus</i>	<i>Sarbecovirus</i>	SARS-CoV-2	99%	Feuang (Site 1)	Na Ang	Pha Tong



BANAL-20-251	<i>Hipposideros</i>	<i>khaokhouayensis</i>	<i>Alphacoronavirus</i>	<i>Decacovirus</i>	Bat coronavirus HKU10	97%	Feuang (Site 1)	Na Ang	Pha Tong
BANAL-20-273	<i>Rhinolophus</i>	<i>pusillus</i>	<i>Alphacoronavirus</i>	unclassified <i>Decacovirus</i>	Rhinolophus bat coronavirus HKU32	92%	Feuang (Site 1)	Na Ang	Pha Tong
BANAL-20-289	<i>Rhinolophus</i>	<i>affinis</i>	<i>Alphacoronavirus</i>	<i>Rhinacovirus</i>	Rhinolophus affinis bat coronavirus HKU2-related / Swine acute diarrhea syndrome related coronavirus (SADS-CoV)	98%	Feuang (Site 1)	Na Ang	Pha Tong
BANAL-20-290	<i>Eonycteris</i>	<i>spelaea</i>	<i>Betacoronavirus</i>	<i>Nobecovirus</i>	Rousettus bat coronavirus GCCDC1	99%	Maed (Site 2)	Had Yao	Keo tao + Had Yao
BANAL-20-390	<i>Eonycteris</i>	<i>spelaea</i>	<i>Betacoronavirus</i>	<i>Nobecovirus</i>	Rousettus bat coronavirus GCCDC1	97%	Maed (Site 2)	Huay Pa Mak	Keo Tao
BANAL-20-395	<i>Eonycteris</i>	<i>spelaea</i>	<i>Betacoronavirus</i>	<i>Nobecovirus</i>	Rousettus bat coronavirus GCCDC1	99%	Maed (Site 2)	Huay Pa Mak	Keo Tao
BANAL-20-398	<i>Eonycteris</i>	<i>spelaea</i>	<i>Betacoronavirus</i>	<i>Nobecovirus</i>	Rousettus bat coronavirus GCCDC1	99%	Maed (Site 2)	Huay Pa Mak	Keo Tao
BANAL-20-403	<i>Eonycteris</i>	<i>spelaea</i>	<i>Betacoronavirus</i>	<i>Nobecovirus</i>	Rousettus bat coronavirus GCCDC1	99%	Maed (Site 2)	Huay Pa Mak	Keo Tao
BANAL-20-416	<i>Cynopterus</i>	<i>sphinx</i>	<i>Betacoronavirus</i>	<i>Nobecovirus</i>	Rousettus bat coronavirus GCCDC1	99%	Maed (Site 2)	Huay Pa Mak	Pha Moum
BANAL-20-497	<i>Myotis</i>	sp.	<i>Alphacoronavirus</i>	<i>Pedacovirus</i>	Bat alphacoronavirus*	99%	Xay (Site 3)	Chom-Ong	Near Chom-Ong Cave

\* Due to the low genome coverage of BANAL-20-497 (< 1%), the taxonomic assignment is not robust.

Table S4. Vertical coverage of BANAL genomes presented in the study

<b>BANAL sample</b>	<b>Average coverage for whole genome (SD)</b>	<b>Average coverage in Spike (SD)</b>	<b>Average coverage in RBD (SD)</b>
BANAL-52	83720 X (89558) *	75247 X (97210) *	36623 X (64721) *
BANAL-103	43 X (47) *	19 X (16) *	0.42 X (0.74) *
BANAL-236	2872 X (2477)	4100 X (2862)	2847 X (1530)
BANAL-116	662 X (512)	1038 X (511)	679 X (309)
BANAL-247	15 X (13) *	19 X (14) *	7 X (3) *

The table above presents the average vertical coverage obtained after NGS for all BANAL samples. Of note, several key regions (noted by a \* in the table) were obtained by Sanger sequencing after multiple overlapping PCRs, resulting in a lower vertical coverage of the region but with higher confidence of the consensus sequence. In addition, BANAL-52 was sequenced in three independent NGS runs, including one AmpliSeq, to achieve a good genome coverage. Therefore, the coverage presented here for BANAL-52 results from the combination of the three runs.

Table S5. Characteristics of the recombination breakpoints

Breakpoint no	Downstream fragment no	Coordinates of the fragment	Breakpoint located in coding region?	Breakpoint position in the codon	Sequences before (aa)	Breakpoint sequence	Sequences after (aa)
1	2	1 – 1,820	yes (ORF1a)	3rd	GGTGGTGTGTTCA (GGVVQ)	G	TTGACTTCGCAGTGG (LTSQW)
2	3	1,821 – 2,689	yes (ORF1a)	2nd	CCACTGGGCATTG (PLGID)	A	TTAGATGAGTGGAGT (LDEWS)
3	4	3,971 – 7,176	yes (ORF1a)	2nd	GTGGTTATACCTA (VVIPT)	C	TAAAAAGGCTGGTGGC (KKAGG)
4	5	7,177 – 11,461	yes (ORF1a)	1st	ATGTGGTTAATAATT (MWLIII)	A	ATCTTGACAAATG (NLVQM)
5	6	11,462 – 12562	yes (ORF1a)	2nd	TTTGGCCTCTTTT (FGLFC)	G	TTTACTCAACCGCTAC (LLNRY)
6	7	12,563 – 17694	yes (ORF1a)	2nd	ACTGCTTGCACTG (TACTD)	A	TGACAATGCGTTAGCT (DNALA)
7	8	17,695 – 19,791	yes (ORF1ab)	2nd	ACCACTGAAACAG (TTETA)	C	AGCTCACTCTTGTAAT (AHSCN)
8	9	19,792 – 21,204	yes (ORF1ab)	2nd	CTCACTGTCTTTT (LTVFF)	T	TGATGGTAGAGTTGAT (DGRVD)
9	10	21,205 – 22,388	yes (ORF1ab)	2nd	TTATTTGACATGA (LFDMS)	G	TAAATTTCCCCTTAAA (KFPLK)
10	11	22,389 – 24,229	yes (S)	3rd	ACAGAATCTATTGT (TESIV)	T	AGATTTCCCTAATATT (RFPNI)
11	12	24,230 – 27,343	yes (S)	2nd	CAAGACTCACTTT (QDSL S)	C	TTCCACAGCAAGTGCA (STASA)
12	13	27,344 – 27,800	yes (ORF7a)	2nd	GAGTGTGTTAGAG (ECVRG)	G	TACAACAGTACTTTTA (TTVLL)
13	14	27,801 – 28,463	yes (ORF8)	1st	CTTGTTTTCTTAGGA (LVFLG)	A	ATCATCACAAGTGTGTA (IITTV)
14	15	28,464 – 29,597	yes (N)	3rd	AGAGCTACCAGACG (RATRR)	A	ATTCGTGGTGGTGAC (IRGGD)

**Table S6. List of the MD simulations performed.**

The 3 production simulations of the SARS-CoV-2 RBD/hACE2 complex were initiated from the same equilibrated model with different random velocities. Production simulations of the BANAL-236 and BANAL-52/103 RBD/hACE2 complexes were initiated from different equilibrated models obtained, for each construct, from the top 3 scoring homology models.

<b>ID</b>	<b>Construct</b>	<b>Initial model</b>	<b># K/Cl ions</b>	<b># waters</b>	<b>Total # atoms</b>	<b>Production time [<math>\mu</math>s]</b>
SARS-CoV-2.1	SARS-CoV-2 RBD/hACE2	X-ray	121/97	30621	104592	1
SARS-CoV-2.2	SARS-CoV-2 RBD/hACE2	X-ray	121/97	30621	104592	1
SARS-CoV-2.3	SARS-CoV-2 RBD/hACE2	X-ray	121/97	30621	104592	1
BANAL-236-CoV.1	BANAL-236 RBD/hACE2	Homology model (top1)	121/97	30502	104228	1
BANAL-236-CoV.2	BANAL-236 RBD/hACE2	Homology model (top2)	120/96	30140	103140	1
BANAL-236-CoV.3	BANAL-236 RBD/hACE2	Homology model (top3)	120/96	30090	102990	1
BANAL-52/103-CoV.1	BANAL-52/103 RBD/hACE2	Homology model (top1)	123/98	30808	105144	1
BANAL-52/103-CoV.2	BANAL-52/103 RBD/hACE2	Homology model (top2)	121/96	30135	103121	1
BANAL-52/103-CoV.3	BANAL-52/103 RBD/hACE2	Homology model (top3)	124/99	31307	106643	1

**Table S7. Data collection and refinement statistics of the crystal structure of RBD-236 in complex with hACE2.**

	Crystal 1
<b>Data collection</b>	
Space group	P 65 2 2
Cell dimensions <sup>27</sup>	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	341.53 341.53 68.15
$\alpha$ , $\beta$ , $\gamma$ (°)	90 90 120
Resolution (Å)	48.63 - 2.94 (3.047 -2.94) *
<i>R</i> <sub>sym</sub> Or <i>R</i> <sub>merge</sub>	0.05642 (0.4364)
<i>I</i> / $\sigma$ <i>I</i>	11.42 (1.50)
Completeness (%)	86.98 (12.53)
Redundancy	2.0 (2.0)
<b>Refinement</b>	
Resolution (Å)	48.63 - 2.94 (3.047 -2.94) *
No. reflections	
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.214/0.227
No. atoms	6610
Protein	6418
Ligand/ion	184
Water	8
<i>B</i> -factors	64.68
Protein	63.72
Ligand/ion	99.31
Water	42.82
R.m.s. deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.65

\*Values in parentheses are for highest-resolution shell.

Data processing was carried out using XDS<sup>67</sup>. Molecular replacement and model refinement were performed using PHENIX<sup>68,70</sup>. Model building was carried out using COOT<sup>69</sup>. Structural figures were made using PYMOL (The PyMOL Molecular Graphics System, v.2.0 Schrödinger). We used 2 crystals for X-ray data collection. Each crystal resulted in one set of X-ray data. The best dataset (as judged by data statistics) was used for structure determination and refinement.

\*Values in parentheses are for the highest-resolution shell.

Table S8: Acknowledgment of sharing of SARS-CoV-2-related genome sequences from the GISAID database related to Figures 1-2 and Supp. Figures 1-4. We thank the authors listed in the table below for sharing their genomic sequences of coronaviruses analyzed in this study.

Accession ID	Virus name	Location	Host	Collection date	Originating lab	Submitting lab	Authors
EPI_ISL_529214	hCoV-19/Wuhan/IME-WH02/2019	Asia / China / Hubei / Wuhan	Human	2019-12-30	Beijing Institute of Microbiology and Epidemiology	Beijing Institute of Microbiology and Epidemiology	Fan, Hang; Qin, E.; Wu, Y.; Guo, Y.; Zhang, X.; Yong, Y.; Hou, J.; Xu, Z.; Mu, J.; Teng, Yue; Mi, Z.; Yang, R.; Song, Yajun.; Li, B.; Cui, Y.
EPI_ISL_529213	hCoV-19/Wuhan/IME-WH01/2019	Asia / China / Hubei / Wuhan	Human	2019-12-30	Beijing Institute of Microbiology and Epidemiology	Beijing Institute of Microbiology and Epidemiology	Fan, Hang; Qin, E.; Wu, Y.; Guo, Y.; Zhang, X.; Yong, Y.; Hou, J.; Xu, Z.; Mu, J.; Teng, Yue; Mi, Z.; Yang, R.; Song, Yajun.; Li, B.; Cui, Y.
EPI_ISL_402132	hCoV-19/Wuhan/HBCDC-HB-01/2019	Asia / China / Hubei / Wuhan	Human	2019-12-30	Wuhan Jinyintan Hospital	Hubei Provincial Center for Disease Control and Prevention	Bin Fang, Xiang Li, Xiao Yu, Linlin Liu, Bo Yang, Faxian Zhan, Guojun Ye, Xixiang Huo, Junqiang Xu, Bo Yu, Kun Cai, Jing Li, Yongzhong Jiang.
EPI_ISL_402123	hCoV-19/Wuhan/IPBCAMS-WH-01/2019	Asia / China / Hubei / Wuhan	Human	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
EPI_ISL_402119	hCoV-19/Wuhan/IVDC-HB-01/2019	Asia / China / Hubei / Wuhan	Human	2019-12-30	National Institute for Viral Disease Control and	National Institute for Viral Disease Control and Prevention, China CDC	Wenjie Tan, Xiang Zhao, Wenling Wang, Xuejun Ma, Yongzhong Jiang, Roujian Lu, Ji

					Prevention, China CDC		Wang, Weimin Zhou, Peihua Niu, Peipei Li, 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_406798	hCoV-19/Wuhan/WH01/2019	Asia / China / Hubei / Wuhan	Human	2019-12-26	General Hospital of Central Theater Command of People's Liberation Army of China	BGI & Institute of Microbiology, Chinese Academy of Sciences & Shandong First Medical University & Shandong Academy of Medical Sciences & General Hospital of Central Theater Command of People's Liberation Army of China	Weijun Chen, Yuhai Bi, Weifeng Shi and Zhenhong Hu
EPI_ISL_408515	hCoV-19/env/Wuhan/IVDC-HBF13-21/2020	Asia / China / Hubei / Wuhan	Environment (Huanan Seafood Market)	2020-01-01	Institute of Viral Disease Control and Prevention, China CDC	Institute of Viral Disease Control and Prevention, China CDC	William J. Liu, Peipei Liu, Xiang Zhao, Peihua Niu, Yingze Zhao, Wenwen Lei, Ziqian Xu, Shumei Zou, Wei Zhen, Beiwei Ye, Mengjie Yang, Weifeng Shi, Roujian Lu, Wenjie Tan, Zhixiao Chen, Yuchao Wu, Juan Song, Weimin Zhou, Dayan Wang, Jun Han,

							Wenbo Xu, George F. Gao, Guizhen Wu
EPI_ISL_408514	hCoV-19/env/Wuhan/IVDC-HBF13-20/2020	Asia / China / Hubei / Wuhan	Environment (Huanan Seafood Market)	2020-01-01	Institute of Viral Disease Control and Prevention, China CDC	Institute of Viral Disease Control and Prevention, China CDC	William J. Liu, Peipei Liu, Xiang Zhao, Peihua Niu, Yingze Zhao, Wenwen Lei, Ziqian Xu, Shumei Zou, Wei Zhen, Beiwei Ye, Mengjie Yang, Weifeng Shi, Roujian Lu, Wenjie Tan, Zhixiao Chen, Yuchao Wu, Juan Song, Weimin Zhou, Dayan Wang, Jun Han, Wenbo Xu, George F. Gao, Guizhen Wu
EPI_ISL_412977	hCoV-19 /bat/Yunnan/RmYN02/2019	Asia / China / Yunnan / Xishuangbanna	<i>Rhinolophus malayanus</i>	2019-06-25	Shandong First Medical University & Shandong Academy of Medical Sciences	Institute of Microbiology, Chinese Academy of Sciences	Weifeng Shi, Tao Hu, Hong Zhou, Juan Li, Xing Chen, Alice Catherine Hughes, Yuhai Bi
EPI_ISL_1699446	hCoV-19 /bat/Yunnan/RpYN06/2020	Asia / China / Yunnan / Xishuangbanna	<i>Rhinolophus pusillus</i>	2020-05-25	Shandong First Medical University & Shandong Academy of Medical Sciences	Shandong First Medical University & Shandong Academy of Medical Sciences	Weifeng Shi, Edward C. Holmes, Alice C. Hughes, Hong Zhou, Jingkai Ji, Xing Chen, Yuhai Bi, Juan Li, Tao Hu, Yanhua Chen
EPI_ISL_1098866	hCoV-19 /bat/Yunnan/PrC31/2018	Asia / China / Yunnan / Pu'er City	<i>Rhinolophus</i> sp.	2018-08	Yunnan Tropical and Subtropical Animal Viral Disease Laboratory, Yunnan Animal Science and	Diarrhea department, National Institute for Viral Disease Control and Prevention, China CDC	Lili Li



					Veterinary Institute		
EPI_ISL_402131	hCoV-19 /bat/Yunnan/RaTG13/2013	Asia / China / Yunnan / Pu'er	<i>Rhinolophus affinis</i>	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_852604	hCoV-19 /bat/Cambodia/RShSTT182 /2010	Asia / Cambodia / Steung Treng	<i>Rhinolophus shameli</i>	2010-12-06	Virology Unit, Institut Pasteur du Cambodge	G5 Evolutionary Genomics of RNA viruses, Virology Department, Institut Pasteur	Vibol Hul, Deborah Delaune, Erik A Karlsson, Ou Tey Putita, Alexandre Hassanin, Artem Baidaliuk, Fabiana Gámbaro, Vuong Tan Tu, Lucy Keatts, Jonna Mazet, Christine Johnson, Philippe Buchy, Philippe Dussart, Tracey Goldstein, Etienne Simon-Lorière, Veasna Duong
EPI_ISL_852605	hCoV-19 /bat/Cambodia/RShSTT200 /2010	Asia / Cambodia / Steung Treng	<i>Rhinolophus shameli</i>	2010-12-06	Virology Unit, Institut Pasteur du Cambodge	G5 Evolutionary Genomics of RNA viruses, Virology Department, Institut Pasteur	Vibol Hul, Deborah Delaune, Erik A Karlsson, Ou Tey Putita, Alexandre Hassanin, Artem Baidaliuk, Fabiana Gámbaro, Vuong Tan Tu, Lucy Keatts, Jonna Mazet, Christine Johnson, Philippe Buchy, Philippe Dussart, Tracey Goldstein, Etienne Simon-Lorière, Veasna Duong

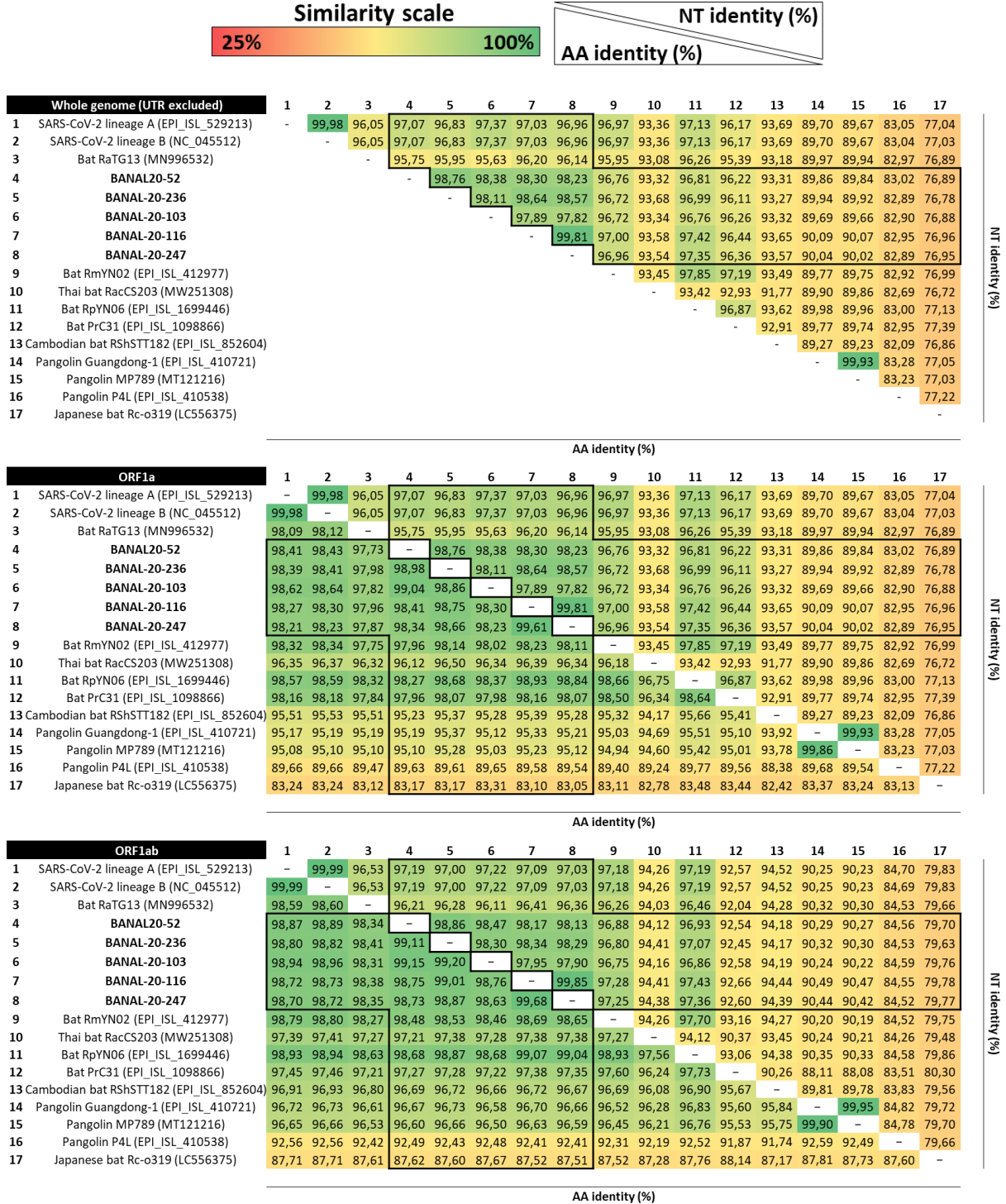
EPI_ISL_410721	hCoV-19 /pangolin/Guangdong/1/2019	Asia / China / Guangdong	<i>Manis javanica</i>	2019	South China Agricultural University	South China Agricultural University	Yongyi Shen, Lihua Xiao, Wu Chen
EPI_ISL_471467	hCoV-19 /pangolin/Guangdong/A22-2 /2019	Asia / China / Guangdong	<i>Manis javanica</i>	2019	South China Agricultural University	South China Agricultural University	Yongyi Shen, Wu Chen
EPI_ISL_410541	hCoV-19 /pangolin/Guangxi/P5E/2017	Asia / China / Guangxi	<i>Manis javanica</i>	2017	Beijing Institute of Microbiology and Epidemiology	Beijing Institute of Microbiology and Epidemiology	Wu-Chun Cao; Tommy Tsan-Yuk Lam; Na Jia; Ya-Wei Zhang; Jia-Fu Jiang; Bao-Gui Jiang
EPI_ISL_410540	hCoV-19 /pangolin/Guangxi/P5L/2017	Asia / China / Guangxi	<i>Manis javanica</i>	2017	Beijing Institute of Microbiology and Epidemiology	Beijing Institute of Microbiology and Epidemiology	Wu-Chun Cao; Tommy Tsan-Yuk Lam; Na Jia; Ya-Wei Zhang; Jia-Fu Jiang; Bao-Gui Jiang
EPI_ISL_410539	hCoV-19 /pangolin/Guangxi/P1E/2017	Asia / China / Guangxi	<i>Manis javanica</i>	2017	Beijing Institute of Microbiology and Epidemiology	Beijing Institute of Microbiology and Epidemiology	Wu-Chun Cao; Tommy Tsan-Yuk Lam; Na Jia; Ya-Wei Zhang; Jia-Fu Jiang; Bao-Gui Jiang
EPI_ISL_410538	hCoV-19 /pangolin/Guangxi/P4L/2017	Asia / China / Guangxi	<i>Manis javanica</i>	2017	Beijing Institute of Microbiology and Epidemiology	Beijing Institute of Microbiology and Epidemiology	Wu-Chun Cao; Tommy Tsan-Yuk Lam; Na Jia; Ya-Wei Zhang; Jia-Fu Jiang; Bao-Gui Jiang
EPI_ISL_410542	hCoV-19 /pangolin/Guangxi/P2V/2017	Asia / China / Guangxi	<i>Manis javanica</i>	2017	Beijing Institute of Microbiology and Epidemiology	Beijing Institute of Microbiology and Epidemiology	Wu-Chun Cao; Tommy Tsan-Yuk Lam; Na Jia; Ya-Wei Zhang; Jia-Fu Jiang; Bao-Gui Jiang
EPI_ISL_1699444	hCoV-19 /bat/Yunnan/RsYN04/2020	Asia / China / Yunnan / Xishuangbanna	<i>Rhinolophus stheno</i>	2020-04-18	Shandong First Medical University & Shandong Academy of Medical Sciences	Shandong First Medical University & Shandong Academy of Medical Sciences	Weifeng Shi, Edward C. Holmes, Alice C. Hughes, Hong Zhou, Jingkai Ji, Xing Chen, Yuhai Bi, Juan Li, Tao Hu, Yanhua Chen

EPI_ISL_1699448	hCoV-19 /bat/Yunnan/RmYN08/2020	Asia / China / Yunnan / Xishuangbanna	<i>Rhinolophus malayanus</i>	2020-07-14	Shandong First Medical University & Shandong Academy of Medical Sciences	Shandong First Medical University & Shandong Academy of Medical Sciences	Weifeng Shi, Edward C. Holmes, Alice C. Hughes, Hong Zhou, Jingkai Ji, Xing Chen, Yuhai Bi, Juan Li, Tao Hu, Yanhua Chen
EPI_ISL_1699445	hCoV-19 /bat/Yunnan/RmYN05/2020	Asia / China / Yunnan / Xishuangbanna	<i>Rhinolophus malayanus</i>	2020-05-25	Shandong First Medical University & Shandong Academy of Medical Sciences	Shandong First Medical University & Shandong Academy of Medical Sciences	Weifeng Shi, Edward C. Holmes, Alice C. Hughes, Hong Zhou, Jingkai Ji, Xing Chen, Yuhai Bi, Juan Li, Tao Hu, Yanhua Chen
EPI_ISL_1699447	hCoV-19 /bat/Yunnan/RmYN07/2020	Asia / China / Yunnan / Xishuangbanna	<i>Rhinolophus malayanus</i>	2020-06-03	Shandong First Medical University & Shandong Academy of Medical Sciences	Shandong First Medical University & Shandong Academy of Medical Sciences	Weifeng Shi, Edward C. Holmes, Alice C. Hughes, Hong Zhou, Jingkai Ji, Xing Chen, Yuhai Bi, Juan Li, Tao Hu, Yanhua Chen
EPI_ISL_1699449	hCoV-19 /bat/Yunnan/RsYN09/2020	Asia / China / Yunnan / Xishuangbanna	<i>Rhinolophus stheno</i>	2020-07-28	Shandong First Medical University & Shandong Academy of Medical Sciences	Shandong First Medical University & Shandong Academy of Medical Sciences	Weifeng Shi, Edward C. Holmes, Alice C. Hughes, Hong Zhou, Jingkai Ji, Xing Chen, Yuhai Bi, Juan Li, Tao Hu, Yanhua Chen
EPI_ISL_1699443	hCoV-19 /bat/Yunnan/RsYN03/2019	Asia / China / Yunnan / Xishuangbanna	<i>Rhinolophus sinicus</i>	2019-10-22	Shandong First Medical University & Shandong Academy of Medical Sciences	Shandong First Medical University & Shandong Academy of Medical Sciences	Weifeng Shi, Edward C. Holmes, Alice C. Hughes, Hong Zhou, Jingkai Ji, Xing Chen, Yuhai Bi, Juan Li, Tao Hu, Yanhua Chen

## Supplementary Figures

Figure S1: Identity matrices at the genus level of representative sarbecoviruses.

Amino-acid (lower) and nucleotide (upper) identity matrices of Laotian and representative human, bat, and pangolin sarbecoviruses. Individual genes and proteins were aligned with MAFFT, and identity matrices were constructed using CLC Main Workbench 21.0.4 (Qiagen). Matrices were colored according to the identity scale, from 25% (red) to 100% (green) of nucleotide or amino-acid identity.



Whole Spike		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	SARS-CoV-2 lineage A (EPI_ISL_529213)	-	100,00	92,86	94,61	87,16	87,14	71,53	71,55	71,73	70,43	76,34	74,72	79,66	83,71	83,66	83,07	70,81
2	SARS-CoV-2 lineage B (NC_045512)	100,00	-	92,86	94,61	87,16	87,14	71,53	71,55	71,73	70,43	76,34	74,72	79,66	83,71	83,66	83,07	70,81
3	Bat RaTG13 (MN996532)	97,41	97,41	-	93,54	86,86	86,78	71,20	71,20	71,17	70,81	77,29	75,40	79,17	83,19	83,14	82,78	71,16
4	BANAL20-52	98,43	98,43	97,79	-	91,32	91,00	71,85	71,91	71,85	70,86	77,34	75,71	79,88	84,40	84,34	83,04	71,19
5	BANAL-20-236	90,51	90,51	89,93	91,50	-	98,82	72,47	72,52	72,55	71,29	80,94	79,25	79,84	89,36	89,31	80,16	71,19
6	BANAL-20-103	90,51	90,51	90,01	91,50	99,60	-	72,26	72,31	72,50	71,03	80,73	79,30	79,66	89,65	89,60	80,08	71,19
7	BANAL-20-116	72,69	72,69	72,76	72,76	73,54	73,39	-	99,81	98,34	94,00	75,33	73,60	71,79	70,03	69,98	69,21	69,36
8	BANAL-20-247	72,77	72,77	72,76	72,84	73,62	73,46	99,67	-	98,32	93,97	75,41	73,60	71,79	70,03	69,98	69,19	69,39
9	Bat RmYN02 (EPI_ISL_412977)	72,61	72,61	72,84	72,68	73,31	73,31	99,02	99,02	-	94,38	75,41	73,68	71,85	70,19	70,14	69,37	69,71
10	Thai bat RacCS203 (MW251308)	72,38	72,38	72,61	72,45	73,23	73,23	97,88	97,96	98,04	-	75,17	73,81	70,61	70,11	70,06	69,42	69,47
11	Bat RpyN06 (EPI_ISL_1699446)	80,78	80,78	80,65	80,65	85,45	85,45	77,96	78,12	77,88	77,96	-	93,83	72,77	79,81	79,75	75,35	70,61
12	Bat PrC31 (EPI_ISL_1098866)	80,55	80,55	80,41	80,41	85,06	85,06	77,32	77,40	77,16	77,16	98,23	-	72,03	78,99	78,94	75,07	70,69
13	Cambodian bat RShSTT182 (EPI_ISL_852604)	82,77	82,77	82,72	83,11	82,73	82,65	73,87	73,95	73,87	73,48	75,65	75,57	-	78,37	78,32	75,35	69,55
14	Pangolin Guangdong-1 (EPI_ISL_410721)	89,96	89,96	89,30	90,32	97,08	97,23	72,99	72,99	72,99	73,15	86,25	85,93	82,41	-	99,89	79,90	71,24
15	Pangolin MP789 (MT121216)	89,96	89,96	89,22	90,24	97,00	97,15	72,91	72,91	72,91	73,07	86,17	85,85	82,33	99,76	-	79,90	71,26
16	Pangolin P4L (EPI_ISL_410538)	92,30	92,30	93,07	92,83	87,16	87,23	72,88	72,88	72,96	72,96	80,54	80,14	80,72	87,47	87,39	-	70,46
17	Japanese bat Rc-o319 (LC556375)	76,15	76,15	77,10	76,78	76,54	76,46	74,22	74,30	74,30	74,30	74,76	75,24	74,86	76,61	76,54	76,67	-

NT Identity (%)

AA Identity (%)

ORF3a		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	SARS-CoV-2 lineage A (EPI_ISL_529213)	-	100,00	96,26	95,53	96,62	96,26	95,65	95,77	96,38	91,91	96,26	89,13	94,32	93,24	93,24	86,84	83,15
2	SARS-CoV-2 lineage B (NC_045512)	100,00	-	96,26	95,53	96,62	96,26	95,65	95,77	96,38	91,91	96,26	89,13	94,32	93,24	93,24	86,84	83,15
3	Bat RaTG13 (MN996532)	97,82	97,82	-	95,65	96,74	96,62	96,26	96,38	96,98	92,03	96,14	88,65	93,72	93,60	93,60	86,23	83,87
4	BANAL20-52	95,64	95,64	96,73	-	96,98	96,74	96,50	96,62	96,86	91,55	96,62	88,04	92,63	92,39	92,39	86,84	82,55
5	BANAL-20-236	97,82	97,82	98,55	96,00	-	99,03	98,07	98,19	96,74	91,55	96,50	89,37	93,72	93,48	93,48	87,20	82,91
6	BANAL-20-103	97,09	97,09	97,82	96,00	98,55	-	98,67	98,67	96,38	91,55	96,38	89,25	93,48	93,60	93,60	87,08	83,15
7	BANAL-20-116	96,36	96,36	97,09	95,27	97,82	98,55	-	99,88	96,01	91,67	96,14	89,01	93,24	92,87	92,87	86,84	82,19
8	BANAL-20-247	96,36	96,36	97,09	95,27	97,82	98,55	100,00	-	96,14	91,79	96,26	89,01	93,36	93,00	93,00	86,84	82,19
9	Bat RmYN02 (EPI_ISL_412977)	96,73	96,73	97,82	96,73	97,09	96,36	95,64	95,64	-	92,03	97,10	88,41	93,84	93,12	93,12	87,20	82,31
10	Thai bat RacCS203 (MW251308)	97,45	97,45	97,45	96,00	97,45	97,09	96,36	96,36	96,73	-	92,63	87,80	91,67	90,94	90,94	86,84	82,31
11	Bat RpyN06 (EPI_ISL_1699446)	96,73	96,73	97,09	96,00	97,09	96,36	95,64	95,64	97,09	96,73	-	88,65	93,60	92,87	92,87	87,44	82,91
12	Bat PrC31 (EPI_ISL_1098866)	92,73	92,73	93,09	91,27	93,45	92,73	92,00	92,00	91,64	93,09	91,64	-	87,56	90,58	90,58	85,14	79,42
13	Cambodian bat RShSTT182 (EPI_ISL_852604)	96,00	96,00	96,00	93,82	96,00	95,27	94,55	94,55	94,91	95,64	94,91	90,91	-	91,79	91,79	84,90	82,79
14	Pangolin Guangdong-1 (EPI_ISL_410721)	97,09	97,09	97,09	94,91	97,09	96,36	95,64	95,64	96,00	97,45	96,00	93,45	95,64	-	100,00	86,96	82,91
15	Pangolin MP789 (MT121216)	97,09	97,09	97,09	94,91	97,09	96,36	95,64	95,64	96,00	97,45	96,00	93,45	95,64	100,00	-	86,96	82,91
16	Pangolin P4L (EPI_ISL_410538)	89,45	89,45	89,82	88,73	90,55	90,55	89,45	89,45	90,18	90,18	89,45	86,18	87,64	89,45	89,45	-	83,27
17	Japanese bat Rc-o319 (LC556375)	86,96	86,96	86,59	86,59	86,59	87,32	86,23	86,23	85,87	86,59	85,51	83,33	84,78	86,96	86,96	86,23	-

NT Identity (%)

AA Identity (%)

E		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	SARS-CoV-2 lineage A (EPI_ISL_529213)	-	100,00	99,56	99,56	98,25	97,81	97,37	97,37	98,68	99,12	99,12	99,12	99,12	99,12	99,12	97,37	97,37
2	SARS-CoV-2 lineage B (NC_045512)	100,00	-	99,56	99,56	98,25	97,81	97,37	97,37	98,68	99,12	99,12	99,12	99,12	99,12	99,12	97,37	97,37
3	Bat RaTG13 (MN996532)	100,00	100,00	-	100,00	98,68	98,25	97,81	97,81	99,12	98,68	99,56	98,68	99,56	99,56	99,56	97,81	96,93
4	BANAL20-52	100,00	100,00	100,00	-	98,68	98,25	97,81	97,81	99,12	98,68	99,56	98,68	99,56	99,56	99,56	97,81	96,93
5	BANAL-20-236	100,00	100,00	100,00	100,00	-	99,56	99,12	99,12	98,68	97,37	98,25	97,37	98,25	99,12	99,12	96,93	96,49
6	BANAL-20-103	98,67	98,67	98,67	98,67	98,67	-	99,56	99,56	98,25	96,93	97,81	96,93	98,68	98,68	98,68	96,49	96,05
7	BANAL-20-116	97,33	97,33	97,33	97,33	97,33	98,67	-	99,12	97,81	96,49	97,37	96,49	98,25	98,25	98,25	96,05	95,61
8	BANAL-20-247	97,33	97,33	97,33	97,33	97,33	98,67	97,33	-	97,81	96,49	97,37	96,49	98,25	98,25	98,25	96,05	95,61
9	Bat RmYN02 (EPI_ISL_412977)	100,00	100,00	100,00	100,00	100,00	98,67	97,33	97,33	-	97,81	99,56	97,81	98,68	98,68	98,68	97,37	96,49
10	Thai bat RacCS203 (MW251308)	100,00	100,00	100,00	100,00	100,00	98,67	97,33	97,33	100,00	-	98,25	98,25	98,25	98,25	98,25	96,49	96,49
11	Bat RpyN06 (EPI_ISL_1699446)	100,00	100,00	100,00	100,00	100,00	98,67	97,33	97,33	100,00	100,00	-	98,25	99,12	99,12	99,12	97,81	96,93
12	Bat PrC31 (EPI_ISL_1098866)	98,67	98,67	98,67	98,67	98,67	97,33	96,00	96,00	98,67	98,67	98,67	-	98,25	98,25	98,25	96,49	96,05
13	Cambodian bat RShSTT182 (EPI_ISL_852604)	98,67	98,67	98,67	98,67	98,67	100,00	98,67	98,67	98,67	98,67	97,33	-	99,12	99,12	97,37	96,49	96,49
14	Pangolin Guangdong-1 (EPI_ISL_410721)	100,00	100,00	100,00	100,00	100,00	98,67	97,33	97,33	100,00	100,00	100,00	98,67	98,67	-	100,00	97,81	97,37
15	Pangolin MP789 (MT121216)	100,00	100,00	100,00	100,00	100,00	98,67	97,33	97,33	100,00	100,00	100,00	98,67	98,67	100,00	-	97,81	97,37
16	Pangolin P4L (EPI_ISL_410538)	100,00	100,00	100,00	100,00	100,00	98,67	97,33	97,33	100,00	100,00	100,00	98,67	98,67	100,00	100,00	-	97,81
17	Japanese bat Rc-o319 (LC556375)	98,67	98,67	98,67	98,67	98,67	97,33	96,00	97,33	98,67	98,67	98,67	97,33	97,33	98,67	98,67	98,67	-

NT Identity (%)

AA Identity (%)

M		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	SARS-CoV-2 lineage A (EPI_ISL_529213)	-	100,00	95,07	95,22	95,22	94,92	94,92	94,92	94,47	94,32	94,17	93,57	95,07	93,27	93,27	91,33	86,55
2	SARS-CoV-2 lineage B (NC_045512)	100,00	-	95,07	95,22	95,22	94,92	94,92	94,92	94,47	94,32	94,17	93,57	95,07	93,27	93,27	91,33	86,55
3	Bat RaTG13 (MN996532)	98,65	98,65	-	94,17	95,05	94,44	94,44	94,44	94,74	94,59	94,44	93,42	94,44	93,12	93,12	90,58	85,35
4	BANAL20-52	98,65	98,65	99,10	-	97,16	98,06	98,36	98,36	95,52	92,83	95,22	92,97	93,87	92,83	92,83	90,73	86,40
5	BANAL-20-236	98,20	98,20	99,55	98,65	-	98,20	98,20	98,20	96,25	93,54	96,25	92,83	94,89	92,38	92,38	90,13	86,10
6	BANAL-20-103	97,75	97,75	99,10	98,20	99,55	-	99,70	99,70	95,95	93,39	95,65	91,93	94,59	91,78	91,78	89,84	85,80
7	BANAL-20-116																	



ORF6		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	SARS-CoV-2 lineage A (EPI_ISL_529213)	-	100,00	98,39	96,77	96,77	96,77	96,24	96,24	96,77	96,24	96,77	96,77	95,70	95,70	95,70	90,86	86,56
2	SARS-CoV-2 lineage B (NC_045512)	100,00	-	98,39	96,77	96,77	96,77	96,24	96,24	96,77	96,24	96,77	96,77	95,70	95,70	95,70	90,86	86,56
3	Bat RaTG13 (MN996532)	100,00	100,00	-	97,31	98,39	97,31	96,77	96,77	98,39	94,62	98,39	98,39	96,24	94,09	94,09	91,40	87,10
4	BANAL20-52	98,39	98,39	98,39	-	96,77	100,00	99,46	99,46	96,77	95,16	96,77	96,77	93,55	93,55	93,55	89,78	84,41
5	BANAL20-236	96,77	96,77	96,77	95,16	-	96,77	96,24	96,24	97,85	94,09	97,85	96,77	94,62	93,01	93,01	90,32	86,56
6	BANAL-20-103	98,39	98,39	98,39	100,00	95,16	-	99,46	99,46	96,77	95,16	96,77	96,77	93,55	93,55	93,55	89,78	84,41
7	BANAL-20-116	96,77	96,77	96,77	98,39	93,55	98,39	-	98,92	96,24	94,62	96,24	96,24	93,01	93,01	93,01	89,25	83,87
8	BANAL-20-247	96,77	96,77	96,77	98,39	93,55	98,39	96,77	-	96,24	94,62	96,24	96,24	94,09	93,01	93,01	89,25	83,87
9	Bat RmYN02 (EPI_ISL_412977)	96,77	96,77	96,77	98,39	93,55	98,39	96,77	96,77	-	95,16	98,92	97,85	94,62	93,55	93,55	90,86	86,56
10	Thai bat RacCS203 (MW251308)	98,39	98,39	98,39	100,00	95,16	100,00	98,39	98,39	98,39	-	95,16	94,09	91,94	95,16	95,16	89,25	85,48
11	Bat RpyN06 (EPI_ISL_1699446)	98,39	98,39	98,39	100,00	95,16	100,00	98,39	98,39	98,39	100,00	-	97,85	94,62	93,55	93,55	90,86	86,56
12	Bat PrC31 (EPI_ISL_1098866)	96,77	96,77	96,77	98,39	93,55	98,39	96,77	96,77	96,77	98,39	98,39	-	94,62	93,55	93,55	90,86	87,63
13	Cambodian bat RShSTT182 (EPI_ISL_852604)	96,77	96,77	96,77	95,16	93,55	95,16	93,55	96,77	93,55	95,16	95,16	93,55	-	92,47	92,47	89,25	84,95
14	Pangolin Guangdong-1 (EPI_ISL_410721)	96,77	96,77	96,77	98,39	93,55	98,39	96,77	96,77	96,77	98,39	98,39	96,77	93,55	-	100,00	92,47	86,02
15	Pangolin MP789 (MT121216)	96,77	96,77	96,77	98,39	93,55	98,39	96,77	96,77	96,77	98,39	98,39	96,77	93,55	100,00	-	92,47	86,02
16	Pangolin P4L (EPI_ISL_410538)	95,16	95,16	95,16	96,77	91,94	96,77	95,16	95,16	95,16	96,77	96,77	95,16	91,94	95,16	95,16	-	85,48
17	Japanese bat Rc-o319 (LC556375)	83,87	83,87	83,87	82,26	80,65	82,26	80,65	80,65	82,26	82,26	82,26	83,87	80,65	82,26	82,26	80,65	-

NT Identity (%)

AA Identity (%)

ORF7a		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	SARS-CoV-2 lineage A (EPI_ISL_529213)	-	100,00	95,63	96,17	97,54	97,54	97,54	97,54	96,17	92,35	96,72	95,93	96,17	93,44	93,44	86,61	76,96
2	SARS-CoV-2 lineage B (NC_045512)	100,00	-	95,63	96,17	97,54	97,54	97,54	97,54	96,17	92,35	96,72	95,93	96,17	93,44	93,44	86,61	76,96
3	Bat RaTG13 (MN996532)	97,52	97,52	-	93,44	96,17	95,90	94,81	94,81	95,08	92,90	95,63	93,77	95,90	91,80	91,80	86,61	75,88
4	BANAL20-52	96,69	96,69	95,87	-	95,90	96,99	98,63	98,63	94,54	90,98	95,08	94,85	94,54	91,53	91,53	85,52	75,88
5	BANAL20-236	99,17	99,17	98,35	97,52	-	98,36	97,27	97,27	96,72	92,62	96,99	96,21	96,99	92,90	92,90	87,43	77,78
6	BANAL-20-103	98,35	98,35	97,52	98,35	99,17	-	98,36	98,36	96,45	92,08	97,54	96,21	96,99	92,35	92,35	86,89	77,51
7	BANAL-20-116	98,35	98,35	97,52	98,35	99,17	100,00	-	100,00	95,90	91,80	96,45	96,21	95,90	92,35	92,35	86,07	77,24
8	BANAL-20-247	98,35	98,35	97,52	98,35	99,17	100,00	100,00	-	95,90	91,80	96,45	96,21	95,90	92,35	92,35	86,07	77,24
9	Bat RmYN02 (EPI_ISL_412977)	95,87	95,87	95,04	94,21	96,69	95,87	95,87	95,87	-	91,53	95,08	95,39	95,08	92,90	92,90	86,07	77,24
10	Thai bat RacCS203 (MW251308)	95,87	95,87	95,04	94,21	96,69	95,87	95,87	95,87	95,04	-	91,26	90,79	92,35	90,44	90,44	83,33	74,80
11	Bat RpyN06 (EPI_ISL_1699446)	98,35	98,35	97,52	96,69	99,17	98,35	98,35	98,35	95,87	95,87	-	94,85	95,63	92,08	92,08	86,34	77,24
12	Bat PrC31 (EPI_ISL_1098866)	95,90	95,90	95,08	94,26	96,72	95,90	95,90	95,90	93,44	95,08	95,90	-	94,31	91,60	91,60	85,37	76,34
13	Cambodian bat RShSTT182 (EPI_ISL_852604)	95,04	95,04	94,21	95,04	95,87	96,69	96,69	96,69	92,56	93,39	95,04	92,62	-	92,08	92,08	86,34	76,96
14	Pangolin Guangdong-1 (EPI_ISL_410721)	97,52	97,52	95,04	94,21	96,69	95,87	95,87	95,87	95,04	95,04	95,87	93,44	92,56	100,00	-	85,79	76,96
15	Pangolin MP789 (MT121216)	97,52	97,52	95,04	94,21	96,69	95,87	95,87	95,87	95,04	95,04	95,87	93,44	92,56	100,00	-	85,79	76,96
16	Pangolin P4L (EPI_ISL_410538)	88,43	88,43	89,26	87,60	89,26	88,43	88,43	88,43	86,78	85,95	88,43	86,07	86,78	86,78	86,78	-	72,09
17	Japanese bat Rc-o319 (LC556375)	73,77	73,77	72,95	72,95	74,59	73,77	73,77	73,77	75,41	74,59	73,77	72,36	73,77	73,77	73,77	68,85	-

NT Identity (%)

AA Identity (%)

ORF7b		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
1	SARS-CoV-2 lineage A (EPI_ISL_529213)	-	100,00	99,24	97,73	96,97	92,42	92,42	92,42	92,42	93,94	97,73	97,73	98,48	91,67	91,67	82,58	77,27	
2	SARS-CoV-2 lineage B (NC_045512)	100,00	-	99,24	97,73	96,97	92,42	92,42	92,42	92,42	93,94	97,73	97,73	98,48	91,67	91,67	82,58	77,27	
3	Bat RaTG13 (MN996532)	97,73	97,73	-	98,48	97,73	93,18	93,18	93,18	93,18	94,70	98,48	98,48	99,24	92,42	92,42	83,33	78,03	
4	BANAL20-52	97,73	97,73	100,00	-	97,73	94,70	94,70	94,70	94,70	93,18	96,97	98,48	99,24	92,42	92,42	83,33	79,55	
5	BANAL-20-236	95,45	95,45	97,73	97,73	-	93,18	93,18	93,18	93,18	93,94	97,73	97,73	98,48	93,18	93,18	83,33	78,03	
6	BANAL-20-103	84,09	84,09	86,36	86,36	84,09	-	100,00	100,00	100,00	88,64	92,42	93,18	93,94	87,88	87,88	83,33	81,06	
7	BANAL-20-116	84,09	84,09	86,36	86,36	84,09	100,00	-	100,00	100,00	88,64	92,42	93,18	93,94	87,88	87,88	83,33	81,06	
8	BANAL-20-247	84,09	84,09	86,36	86,36	84,09	100,00	100,00	100,00	-	88,64	92,42	93,18	93,94	87,88	87,88	83,33	81,06	
9	Bat RmYN02 (EPI_ISL_412977)	93,18	93,18	95,45	95,45	93,18	81,82	81,82	81,82	81,82	-	95,45	93,18	93,94	90,15	90,15	83,33	75,00	
10	Thai bat RacCS203 (MW251308)	93,18	93,18	95,45	95,45	93,18	81,82	81,82	81,82	81,82	81,82	-	95,45	93,18	93,94	90,15	90,15	83,33	75,00
11	Bat RpyN06 (EPI_ISL_1699446)	97,73	97,73	100,00	100,00	97,73	86,36	86,36	86,36	86,36	95,45	-	96,97	97,73	92,42	92,42	83,33	78,03	
12	Bat PrC31 (EPI_ISL_1098866)	97,73	97,73	100,00	100,00	97,73	86,36	86,36	86,36	86,36	95,45	100,00	-	99,24	93,94	92,42	84,85	79,79	
13	Cambodian bat RShSTT182 (EPI_ISL_852604)	97,73	97,73	100,00	100,00	97,73	86,36	86,36	86,36	86,36	95,45	100,00	100,00	-	93,18	93,18	84,09	78,79	
14	Pangolin Guangdong-1 (EPI_ISL_410721)	95,45	95,45	97,73	97,73	95,45	84,09	84,09	84,09	84,09	97,73	97,73	97,73	97,73	-	100,00	81,82	74,24	
15	Pangolin MP789 (MT121216)	95,45	95,45	97,73	97,73	95,45	84,09	84,09	84,09	84,09	97,73	97,73	97,73	97,73	100,00	-	81,82	74,24	
16	Pangolin P4L (EPI_ISL_410538)	75,00	75,00	77,27	77,27	75,00	79,55	79,55	79,55	79,55	75,00	77,27	77,27	77,27	75,00	75,00	-	75,00	
17	Japanese bat Rc-o319 (LC556375)	70,45	70,45	72,73	72,73	70,45	75,00	75,00	75,00	75,00	68,18	72,73	72,73	72,73	70,45	70,45	65,91	-	

NT Identity (%)

AA Identity (%)

ORF8		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	SARS-CoV-2 lineage A (EPI_ISL_529213)	-	99,73	97,27	96,72	98,36	98,09	47,96	47,96	48,72	92,35	97,81	98,36	93,99	92,62	80,05	80,87	43,83
2	SARS-CoV-2 lineage B (NC_045512)	99,73	-	96,99	96,45	98,09	97,81	47,70	47,70	48,47	92,08	97,54	98,09	93,72	92,35	79,78	80,60	43,58
3	Bat RaTG13 (MN996532)	95,87	95,04	-	96,17	97,81	98,09	48,21	48,21	48,98	92,35	97,27	97,81	93,44	92,08	79,78	80,87	43,32
4	BANAL20-52	96,69	95,87	97,52	-	97,27	96,99	47,96	47,96	48,72	92,90	96,99	97,27	93,72	93,72	81,69	81,97	43,32
5	BANAL-20-236	98,35	97,52	97,52	98,35	-	99,18	48,21	48,21	48,98	92,90	98,91	99,45	95,08	93,17	80,60	80,87	43,32
6	BANAL-20-103	96,69	95,87	97,52	96,69	98,35	-	48,72	48,72	49,49	92,62	98,63	99,18	94,81	92,90	80,33	80,60	43,83
7	BANAL-20-116	29,77																

N		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	SARS-CoV-2 lineage A (EPI_ISL_529213)	-	100,00	96,90	97,94	97,70	97,62	97,86	97,62	97,30	93,17	97,70	96,98	93,89	96,19	96,19	90,95	88,33
2	SARS-CoV-2 lineage B (NC_045512)	100,00	-	96,90	97,94	97,70	97,62	97,86	97,62	97,30	93,17	97,70	96,98	93,89	96,19	96,19	90,95	88,33
3	Bat RaTG13 (MN996532)	99,05	99,05	-	97,54	97,14	97,38	97,46	97,30	97,38	93,89	97,78	96,59	94,52	96,51	96,51	91,19	87,78
4	BANAL20-52	99,28	99,28	99,28	-	98,97	99,05	99,29	99,05	98,10	93,89	98,33	97,62	94,68	96,43	96,43	90,95	88,10
5	BANAL-20-236	99,05	99,05	99,05	99,28	-	99,13	99,05	98,81	98,02	93,41	98,41	97,70	94,52	96,19	96,19	90,87	87,94
6	BANAL-20-103	99,28	99,28	99,28	99,52	99,28	-	99,29	99,05	98,25	93,81	98,49	97,78	94,68	96,27	96,27	90,79	87,78
7	BANAL-20-116	99,05	99,05	99,05	99,28	99,05	99,28	-	99,76	98,25	93,65	98,41	97,70	94,76	96,35	96,35	91,03	88,02
8	BANAL-20-247	98,81	98,81	98,81	99,05	98,81	99,05	99,76	-	98,10	93,73	98,25	97,54	94,84	96,43	96,43	91,03	87,86
9	Bat RmYN02 (EPI_ISL_412977)	98,57	98,57	98,57	98,81	98,57	98,81	98,81	98,57	-	93,73	98,81	98,10	94,84	96,35	96,35	91,27	87,86
10	Thai bat RacCS203 (MW251308)	95,70	95,70	95,70	95,94	95,70	95,94	95,70	95,47	95,70	-	94,13	93,65	95,79	93,41	93,41	90,00	85,95
11	Bat RpYN06 (EPI_ISL_1699446)	99,28	99,28	99,28	99,52	99,28	99,52	99,28	99,05	99,28	95,94	-	98,65	95,16	96,90	96,90	91,59	88,25
12	Bat PrC31 (EPI_ISL_1098866)	98,57	98,57	98,57	98,81	98,57	98,81	98,57	98,33	98,57	95,47	99,28	-	94,29	96,19	96,19	91,11	87,62
13	Cambodian bat RShSTT182 (EPI_ISL_852604)	96,66	96,66	96,66	96,90	96,66	96,90	96,66	96,42	96,66	98,57	96,90	96,42	-	94,52	94,52	91,59	87,54
14	Pangolin Guangdong-1 (EPI_ISL_410721)	97,85	97,85	97,85	98,09	97,85	98,09	97,85	97,61	97,37	94,99	98,09	97,85	96,42	-	100,00	91,27	87,14
15	Pangolin MP789 (MT121216)	97,85	97,85	97,85	98,09	97,85	98,09	97,85	97,61	97,37	94,99	98,09	97,85	96,42	100,00	-	91,27	87,14
16	Pangolin P4L (EPI_ISL_410538)	93,79	93,79	94,27	94,03	93,79	94,03	93,79	93,56	93,56	93,56	94,03	94,03	94,27	93,56	93,56	-	86,48
17	Japanese bat Rc-o319 (LC556375)	89,50	89,50	89,74	89,50	89,74	89,50	89,26	89,02	89,26	88,31	89,74	89,26	89,26	89,50	89,50	88,54	-

NT identity (%)

AA identity (%)

ORF10		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	SARS-CoV-2 lineage A (EPI_ISL_529213)	-	100,00	99,15	100,00	98,29	100,00	100,00	100,00	99,15	99,15	100,00	99,15	99,15	99,15	99,15	88,89	94,87
2	SARS-CoV-2 lineage B (NC_045512)	100,00	-	99,15	100,00	98,29	100,00	100,00	100,00	99,15	99,15	100,00	99,15	99,15	99,15	99,15	88,89	94,87
3	Bat RaTG13 (MN996532)	97,44	97,44	-	99,15	97,44	99,15	99,15	99,15	98,29	98,29	99,15	98,29	98,29	98,29	98,29	88,03	94,02
4	BANAL20-52	100,00	100,00	97,44	-	98,29	100,00	100,00	100,00	99,15	99,15	100,00	99,15	99,15	99,15	99,15	88,89	94,87
5	BANAL-20-236	94,87	94,87	92,31	94,87	-	98,29	98,29	98,29	97,44	97,44	98,29	97,44	97,44	99,15	99,15	88,89	94,87
6	BANAL-20-103	100,00	100,00	97,44	100,00	94,87	-	100,00	100,00	99,15	99,15	100,00	99,15	99,15	99,15	99,15	88,89	94,87
7	BANAL-20-116	100,00	100,00	97,44	100,00	94,87	100,00	-	100,00	99,15	99,15	100,00	99,15	99,15	99,15	99,15	88,89	94,87
8	BANAL-20-247	100,00	100,00	97,44	100,00	94,87	100,00	100,00	-	99,15	99,15	100,00	99,15	99,15	99,15	99,15	88,89	94,87
9	Bat RmYN02 (EPI_ISL_412977)	97,44	97,44	94,87	97,44	92,31	97,44	97,44	97,44	-	98,29	99,15	98,29	98,29	98,29	98,29	88,03	94,02
10	Thai bat RacCS203 (MW251308)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	-	99,15	98,29	98,29	98,29	98,29	88,03	94,02
11	Bat RpYN06 (EPI_ISL_1699446)	100,00	100,00	97,44	100,00	94,87	100,00	100,00	100,00	97,44	N/A	-	99,15	99,15	99,15	99,15	88,89	94,87
12	Bat PrC31 (EPI_ISL_1098866)	100,00	100,00	97,44	100,00	94,87	100,00	100,00	100,00	97,44	N/A	100,00	-	98,29	98,29	98,29	89,74	95,73
13	Cambodian bat RShSTT182 (EPI_ISL_852604)	97,44	97,44	94,87	97,44	94,87	97,44	97,44	97,44	94,87	N/A	97,44	97,44	-	98,29	98,29	88,03	94,02
14	Pangolin Guangdong-1 (EPI_ISL_410721)	97,44	97,44	94,87	97,44	97,44	97,44	97,44	97,44	94,87	N/A	97,44	97,44	97,44	-	100,00	89,74	95,73
15	Pangolin MP789 (MT121216)	97,44	97,44	94,87	97,44	97,44	97,44	97,44	97,44	94,87	N/A	97,44	97,44	97,44	100,00	-	89,74	95,73
16	Pangolin P4L (EPI_ISL_410538)	74,36	74,36	71,79	74,36	74,36	74,36	74,36	74,36	71,79	N/A	74,36	74,36	74,36	76,92	76,92	-	91,45
17	Japanese bat Rc-o319 (LC556375)	87,18	87,18	84,62	87,18	89,74	87,18	87,18	87,18	84,62	N/A	87,18	87,18	87,18	89,74	89,74	79,49	-

NT identity (%)

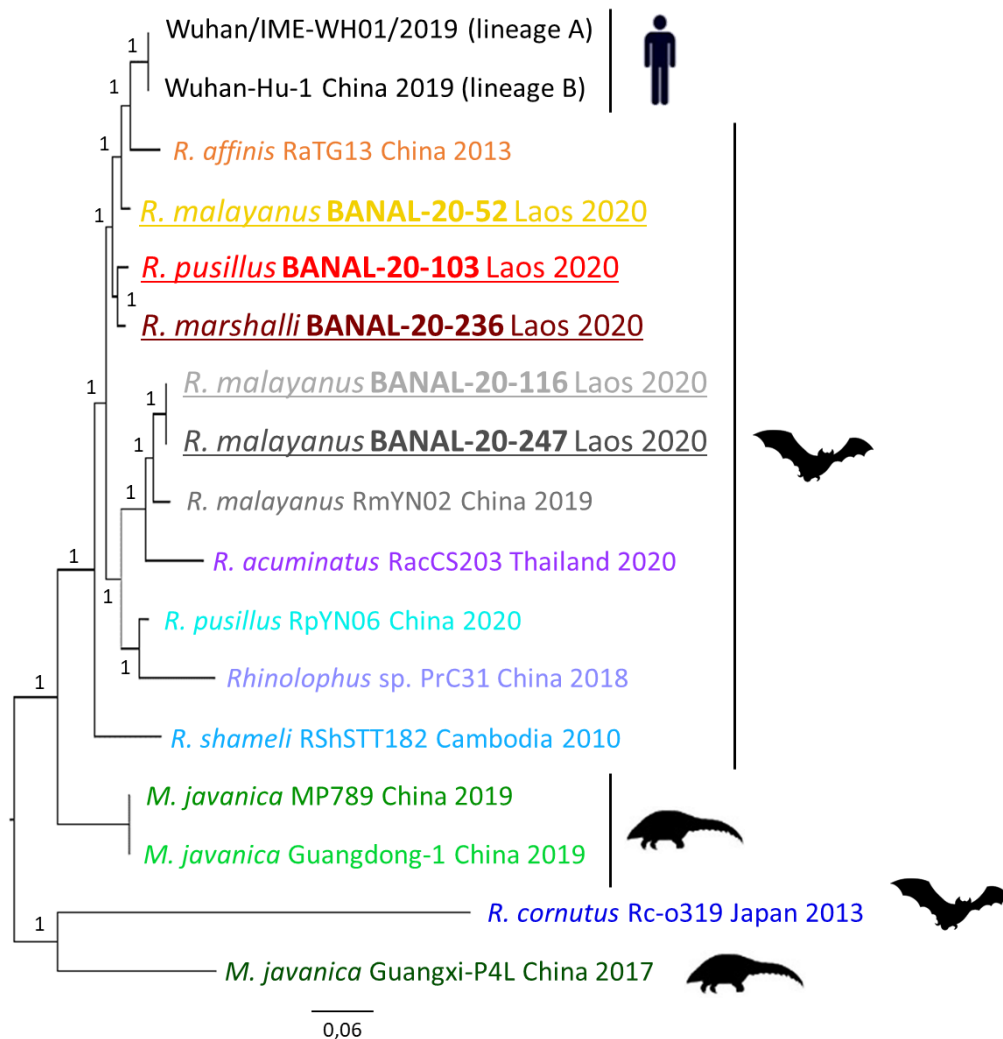
AA identity (%)

Figure S2: Complete phylogenetic reconstructions of recombinant fragments, as described in Figure 2.

The identification of the breakpoints is highly dependent on the programs used, the sequences datasets and the quality of the alignments, leading to difficulties in comparing independent studies performed with different tools. Phylogenetic analyses performed on different fragments are therefore necessary to confirm the identification of breakpoints, especially by confirming that the position of a given sequence differs between fragments bracketing the breakpoint. This is the strategy presented below: sequences are colored according to Figure 2.

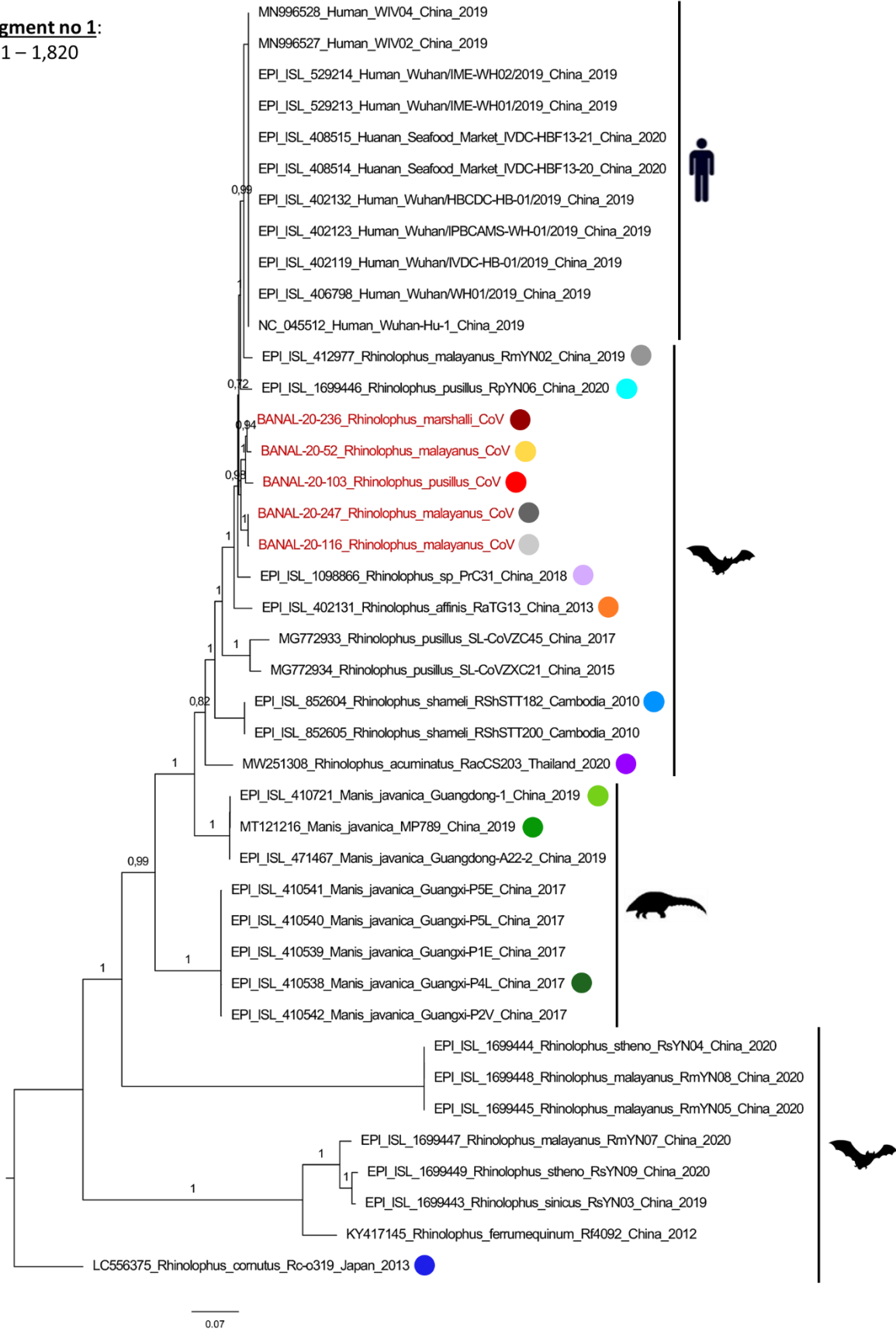
The substitution model considered for all phylogenetic reconstructions presented below was GTR+G. PhyML was used to infer phylogenies both at the whole genome and at the breakpoints nucleotide level. The tree topology search was performed by selecting the best of NNI (Nearest Neighbor Interchange) and SPR (Subtree Pruning and Regraphing). The statistical test selected for branch support was the approximate Bayes parameter, as proposed in the NGPhylogeny portal. Only posterior probabilities >0.5 were considered.

### Whole genome



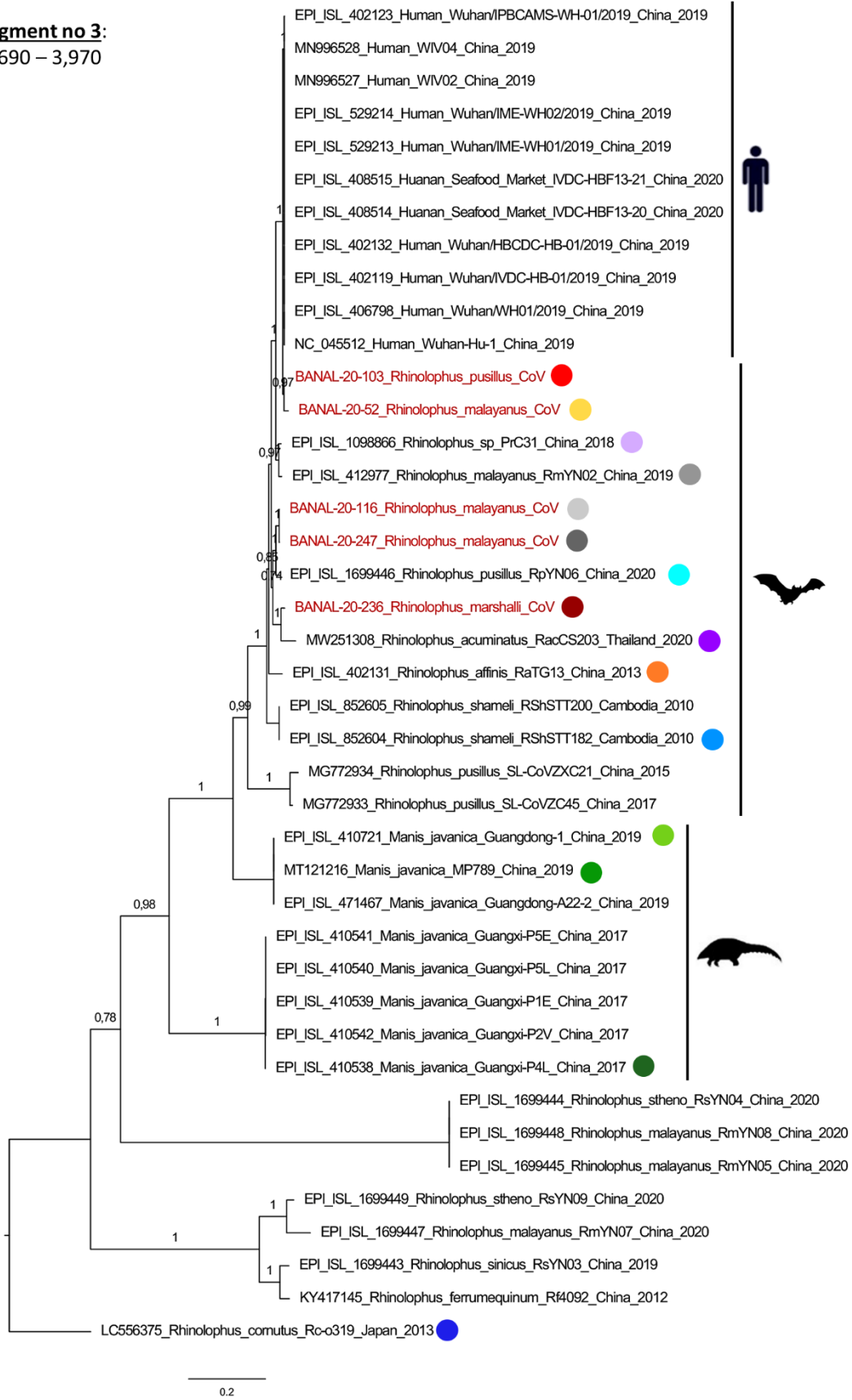


**Fragment no 1:**  
1 – 1,820

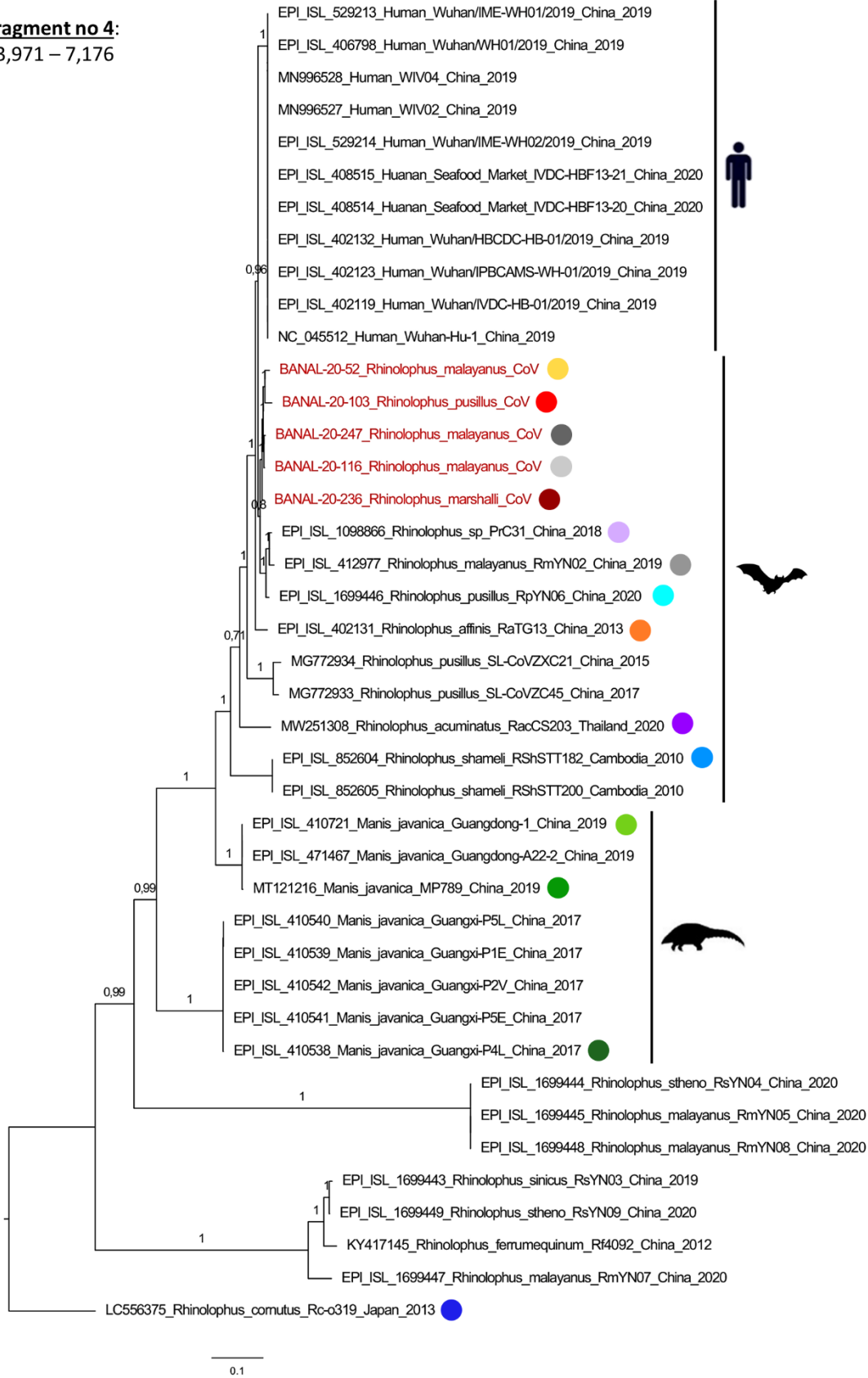




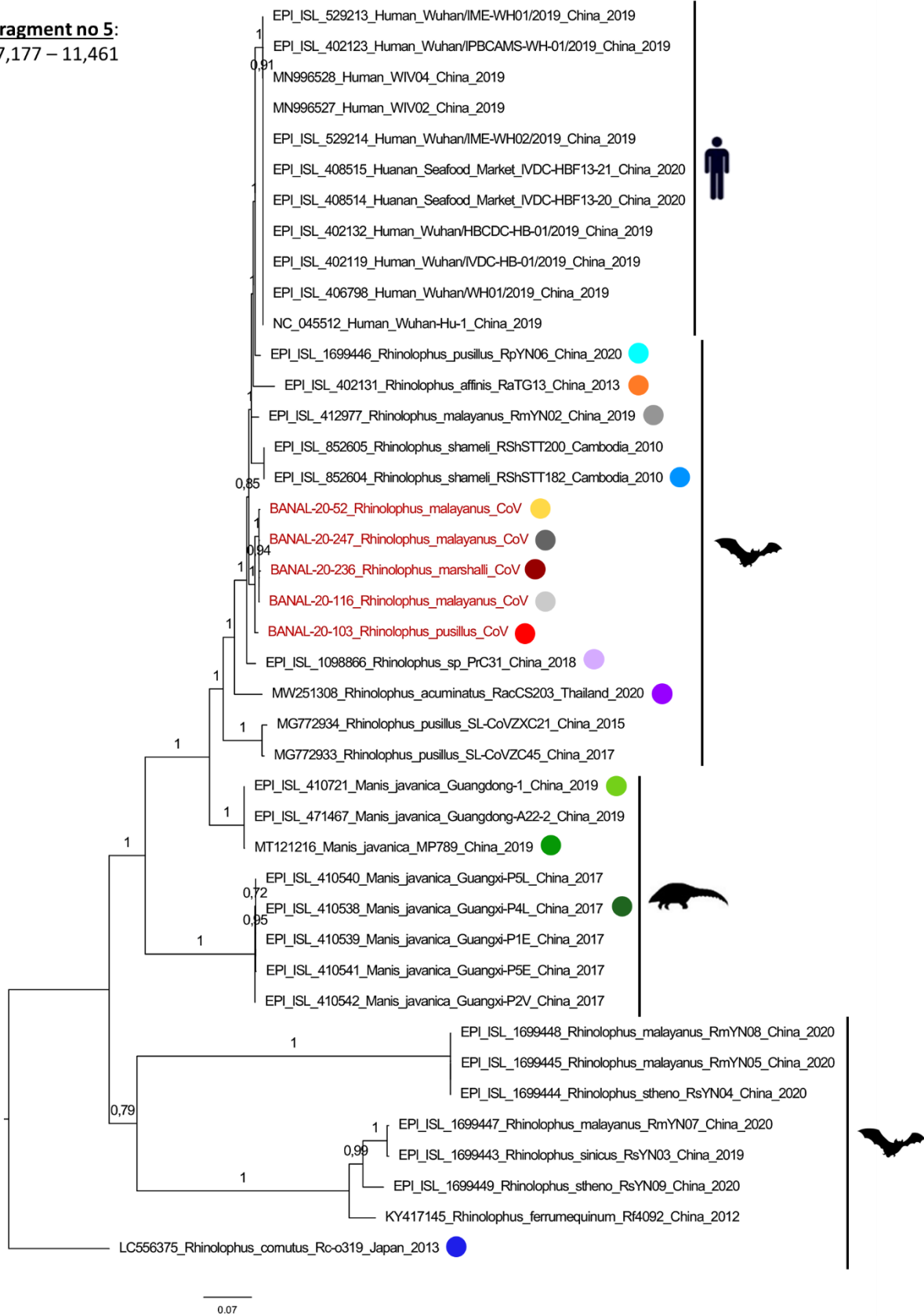
**Fragment no 3:**  
2,690 – 3,970



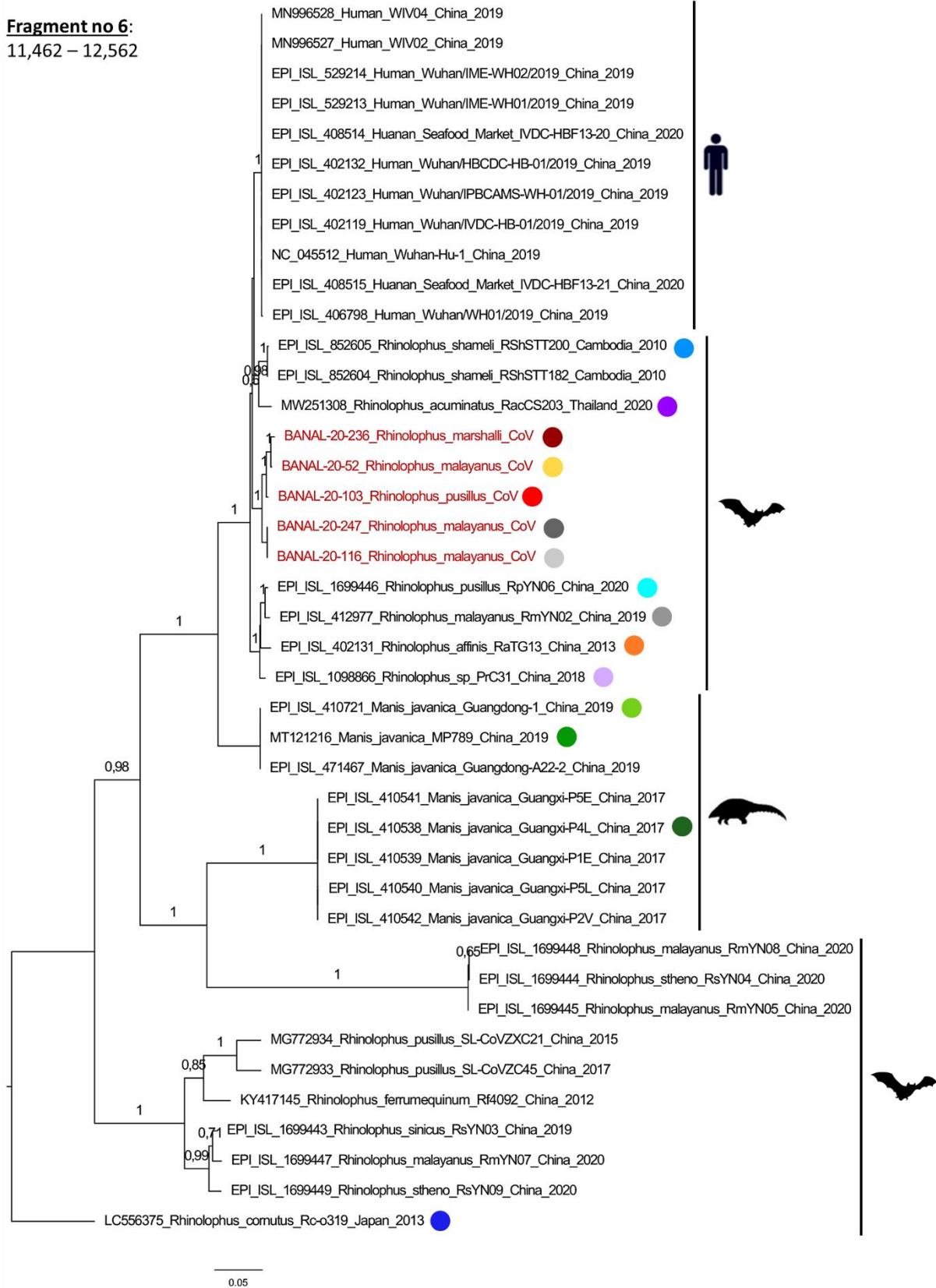
**Fragment no 4:**  
3,971 – 7,176



**Fragment no 5:**  
7,177 – 11,461

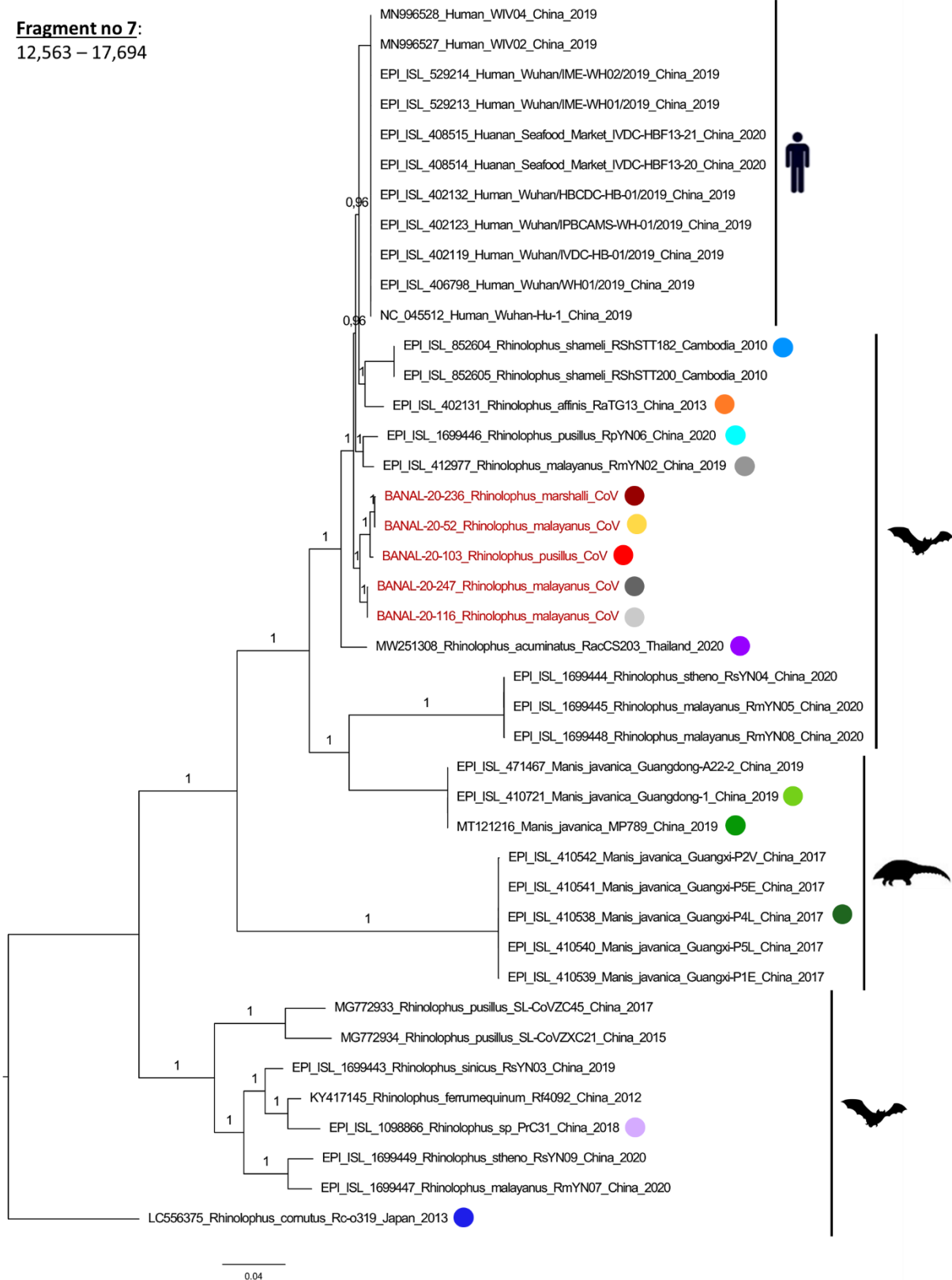


**Fragment no 6:**  
11,462 – 12,562

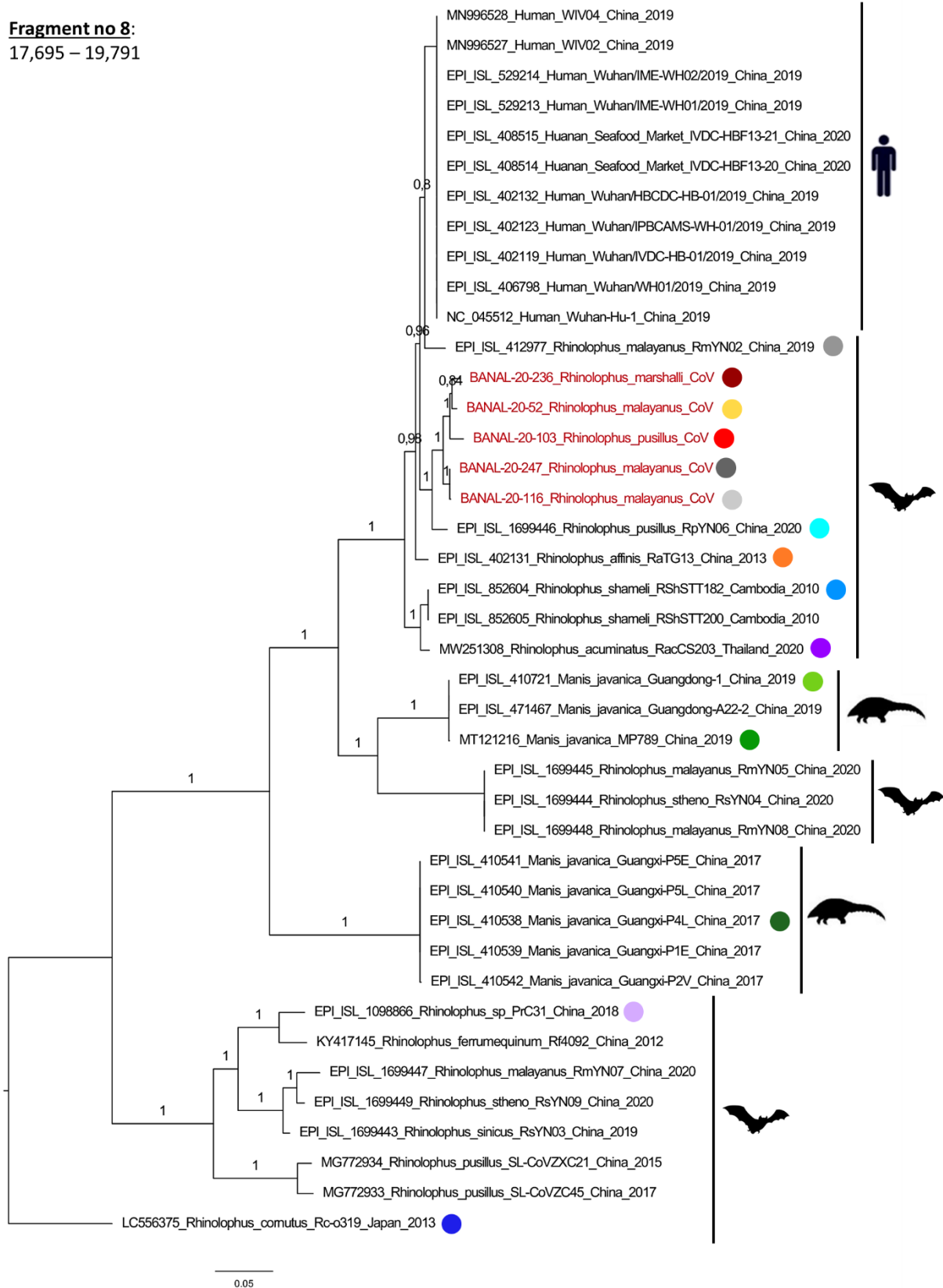




**Fragment no 7:**  
12,563 – 17,694

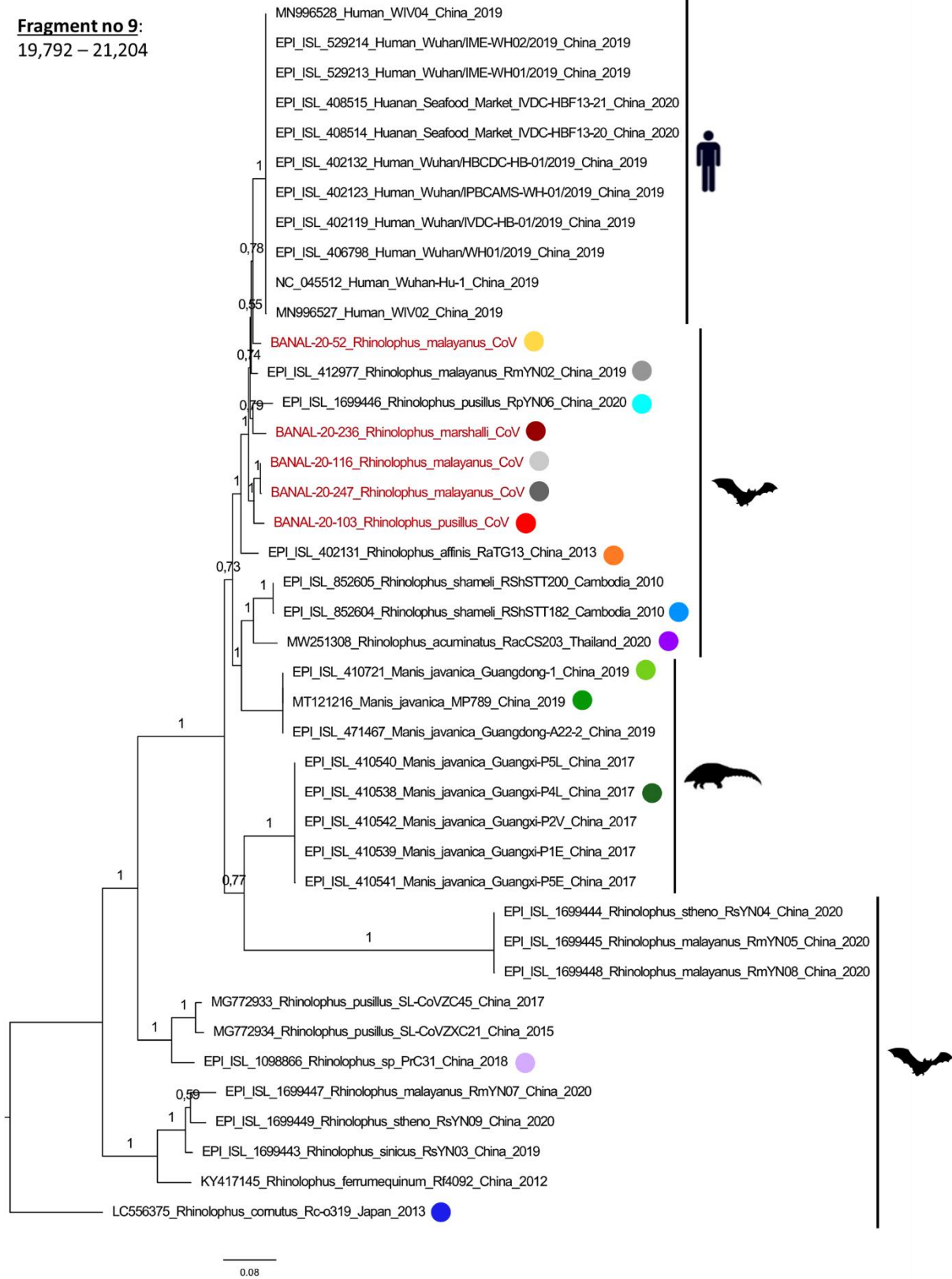


**Fragment no 8:**  
17,695 – 19,791

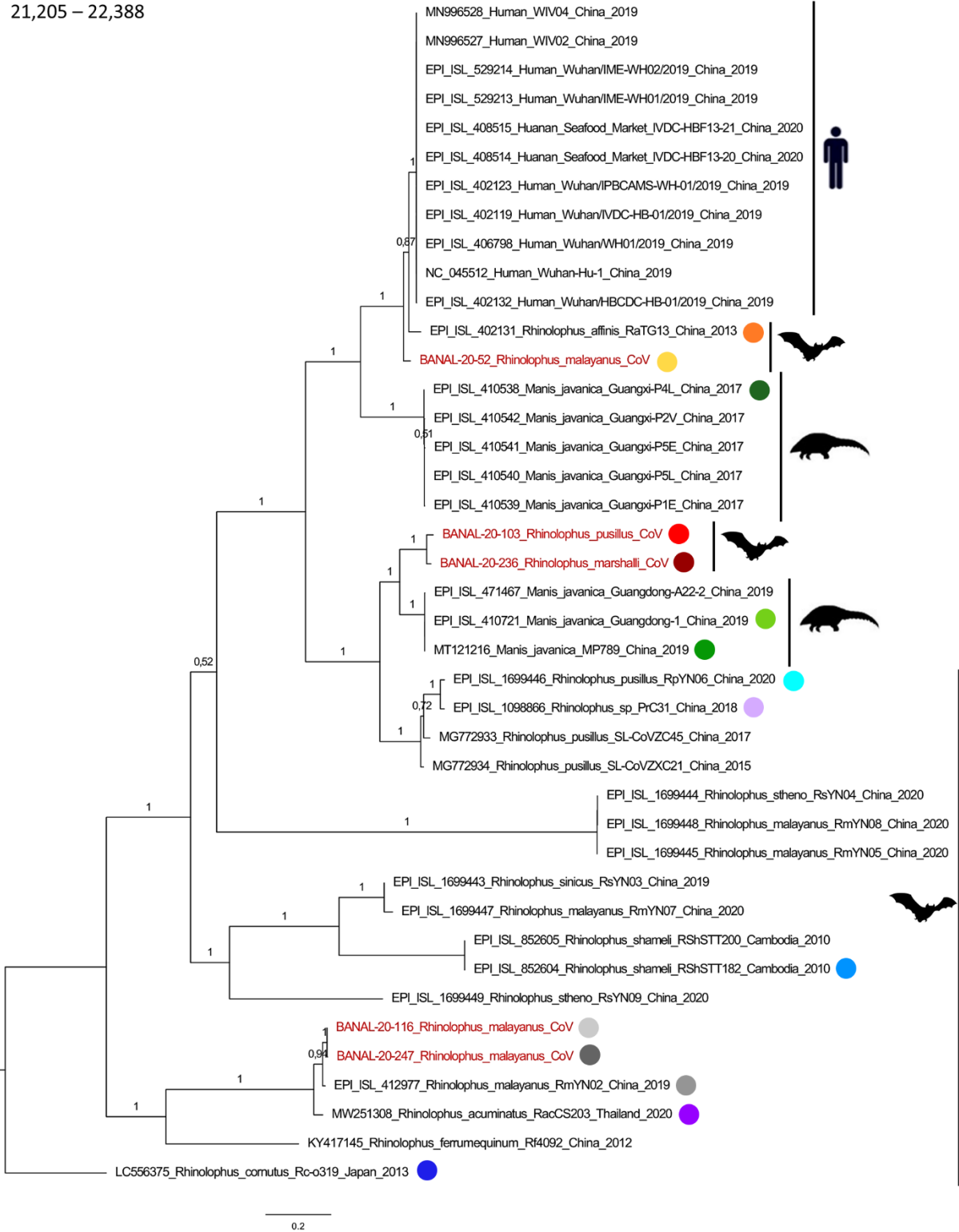




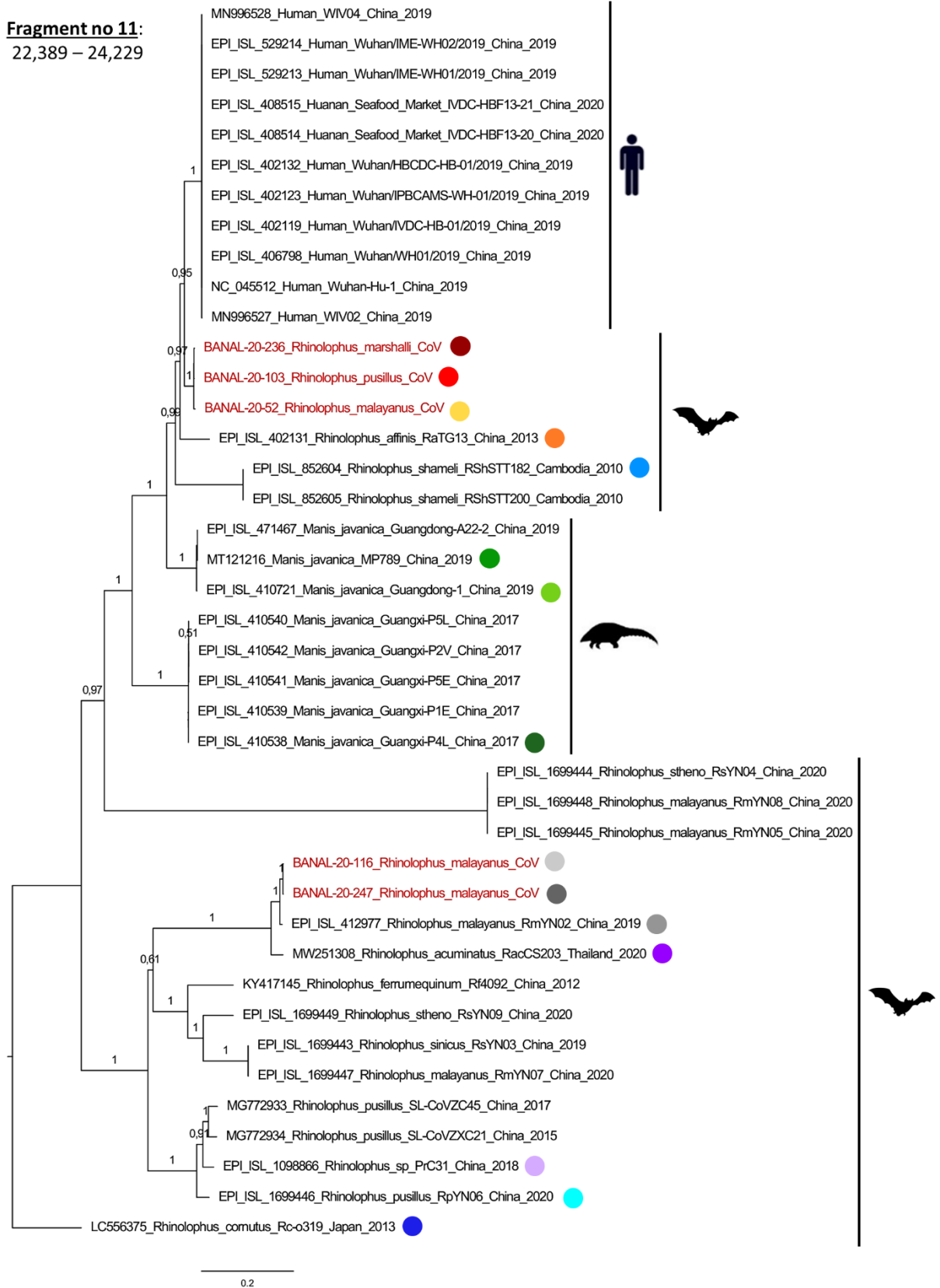
**Fragment no 9:**  
19,792 – 21,204



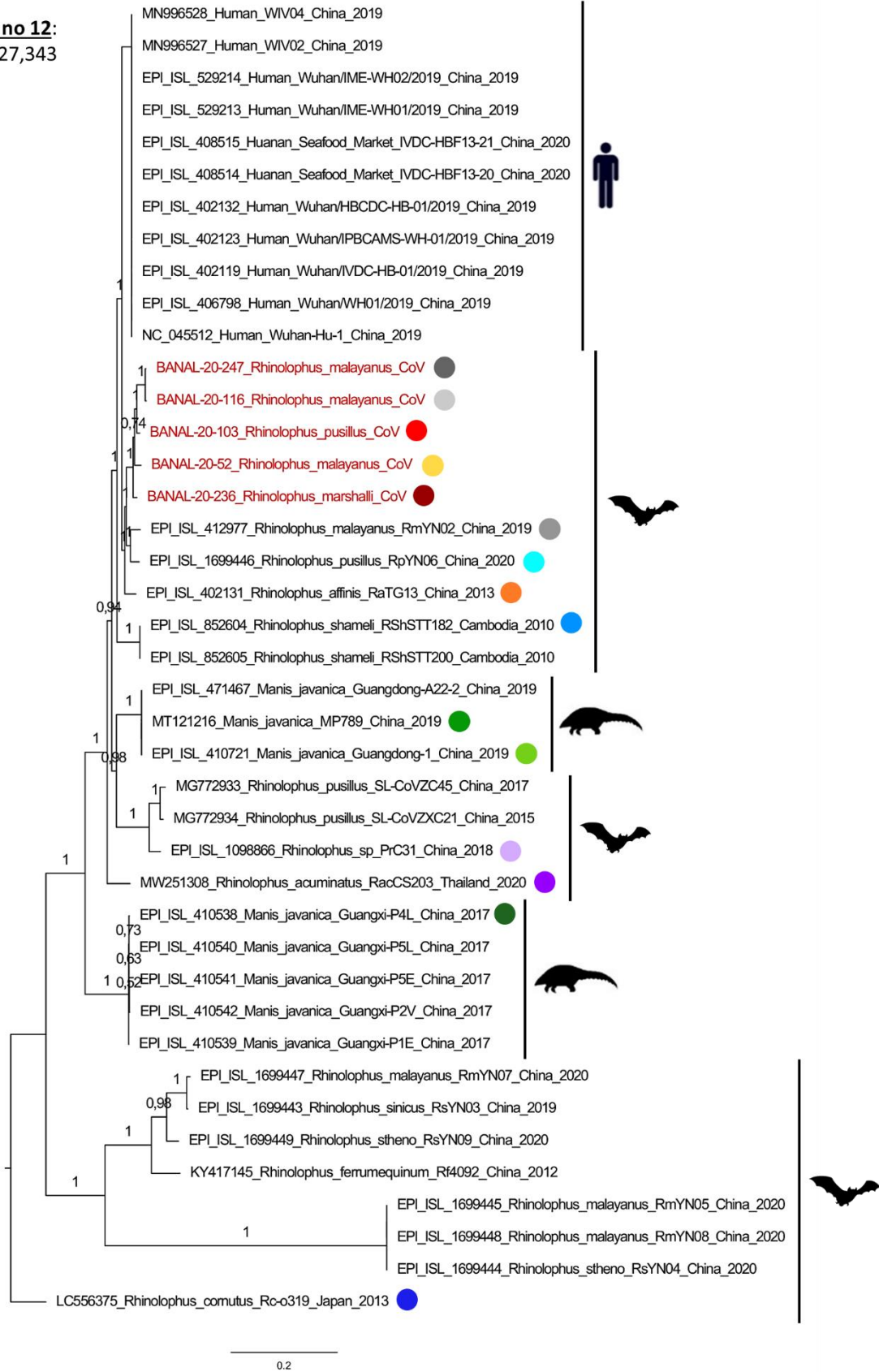
**Fragment no 10:**  
21,205 – 22,388



**Fragment no 11:**  
22,389 – 24,229

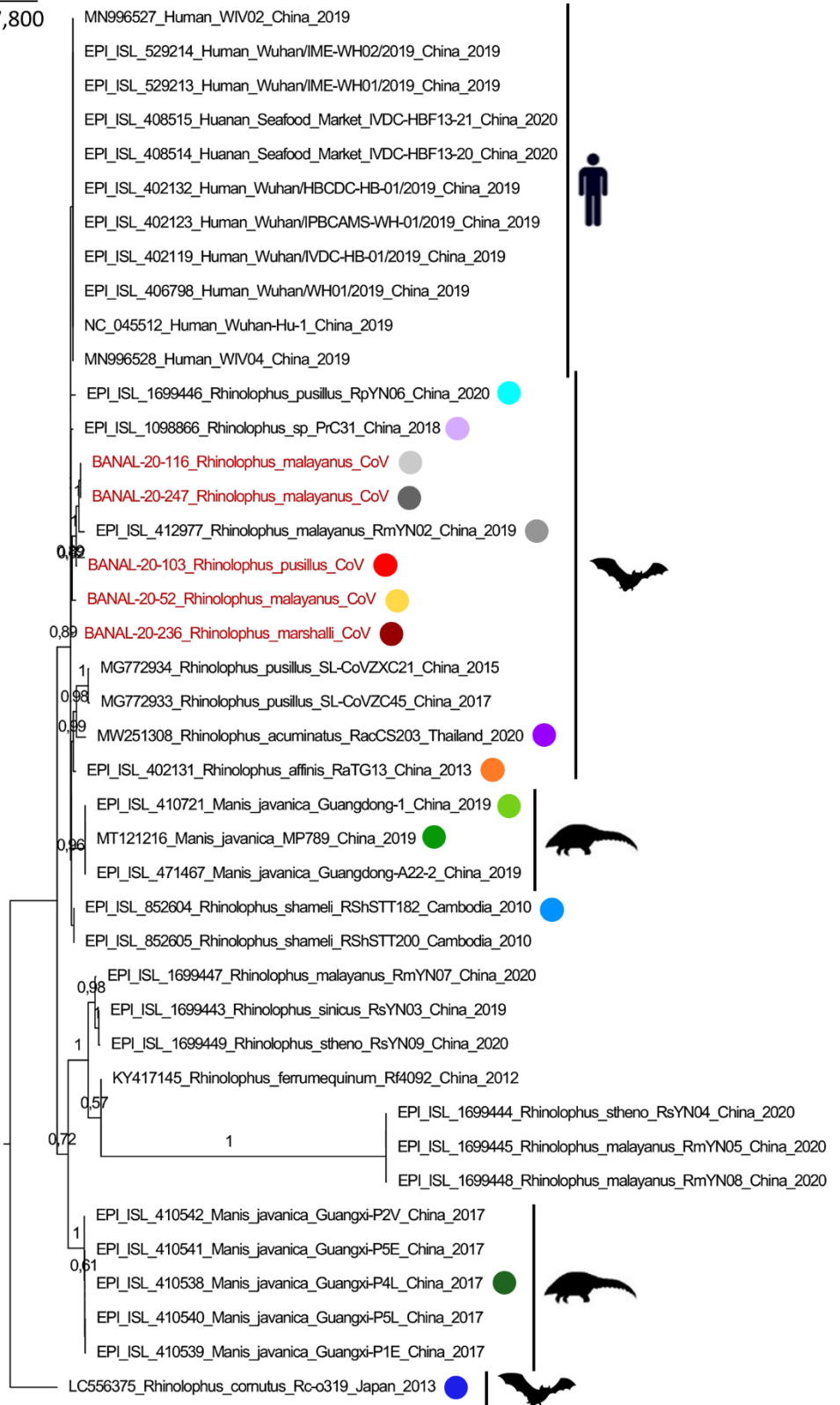


**Fragment no 12:**  
24,230 – 27,343



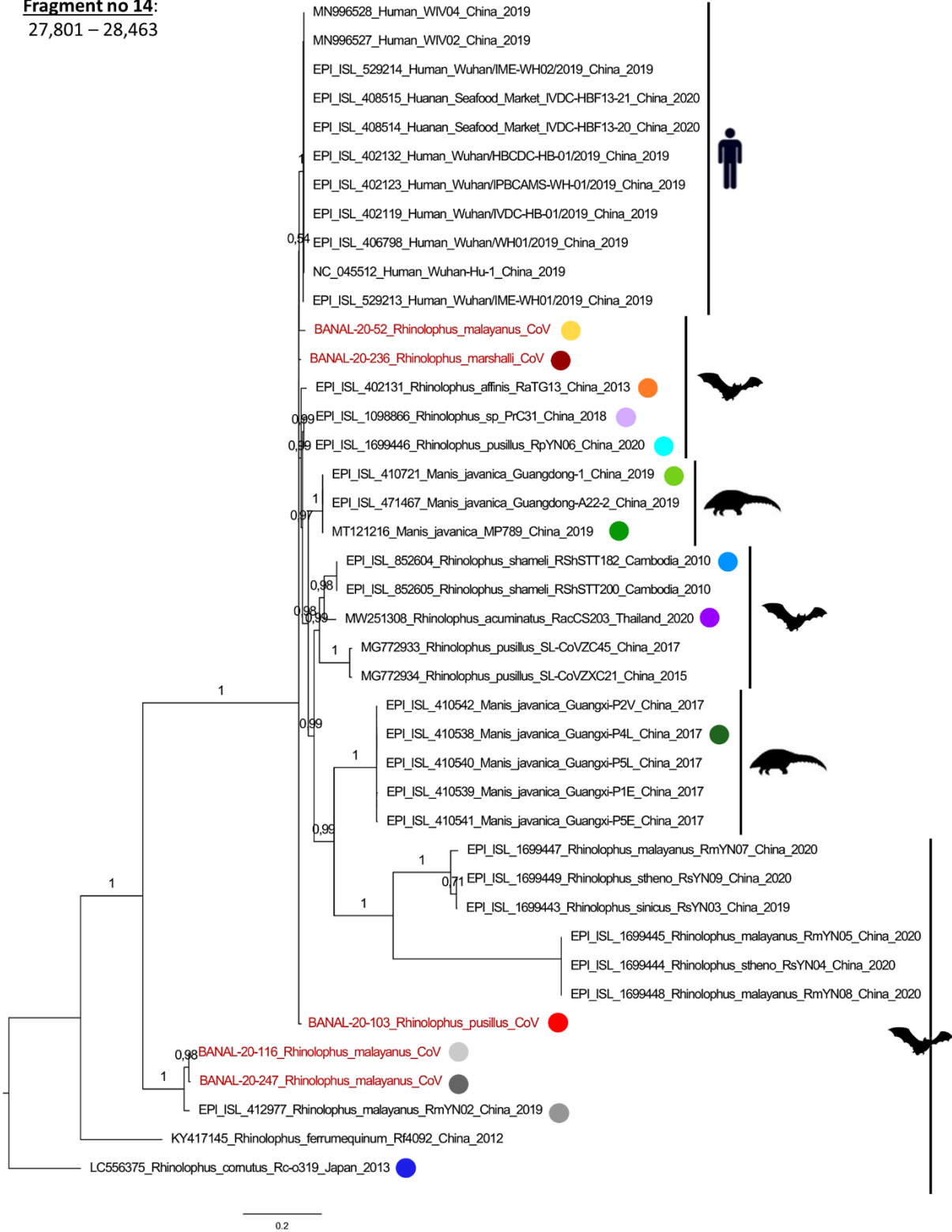
**Fragment no 13:**

27,344 – 27,800



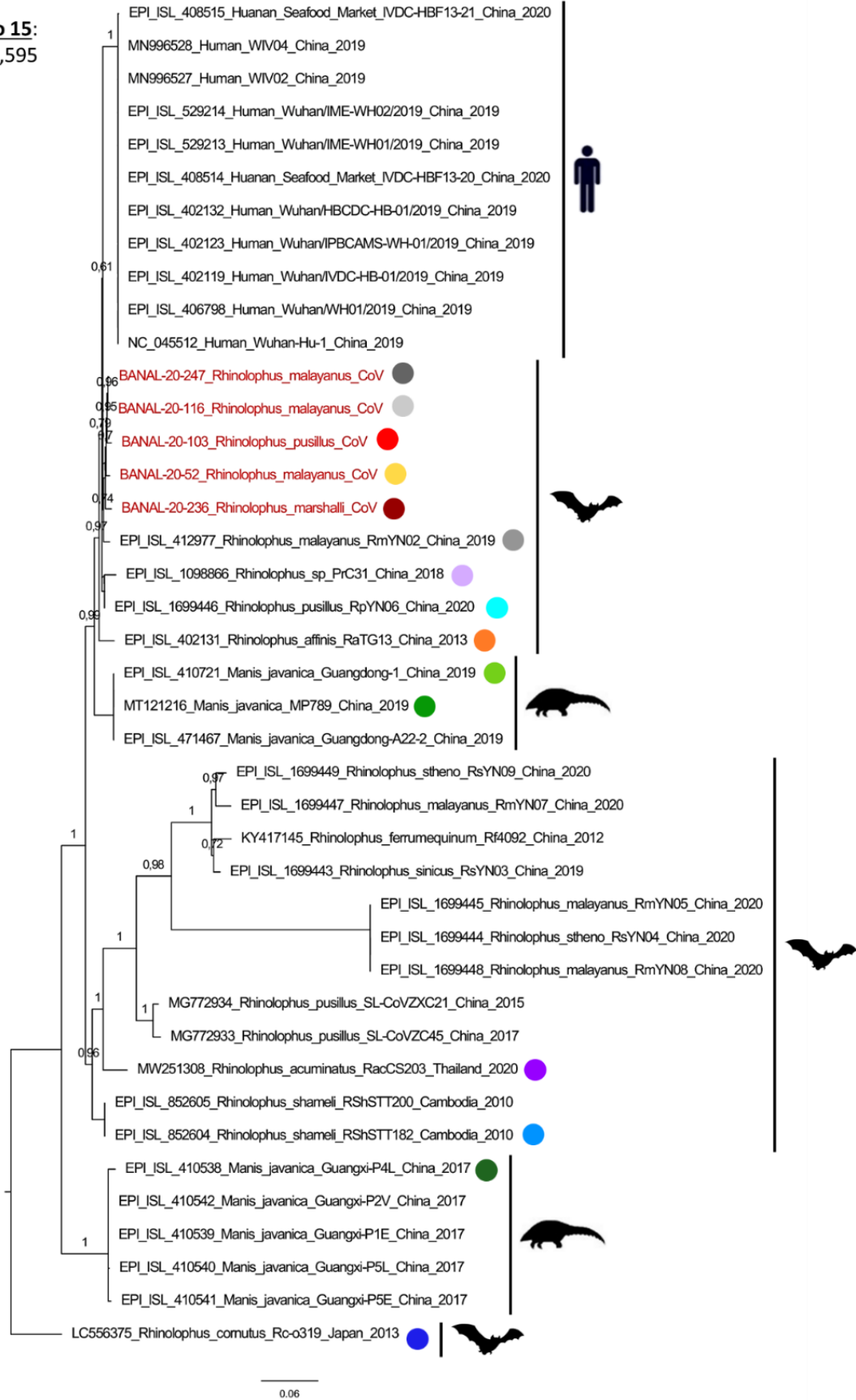
0.4

**Fragment no 14:**  
27,801 – 28,463





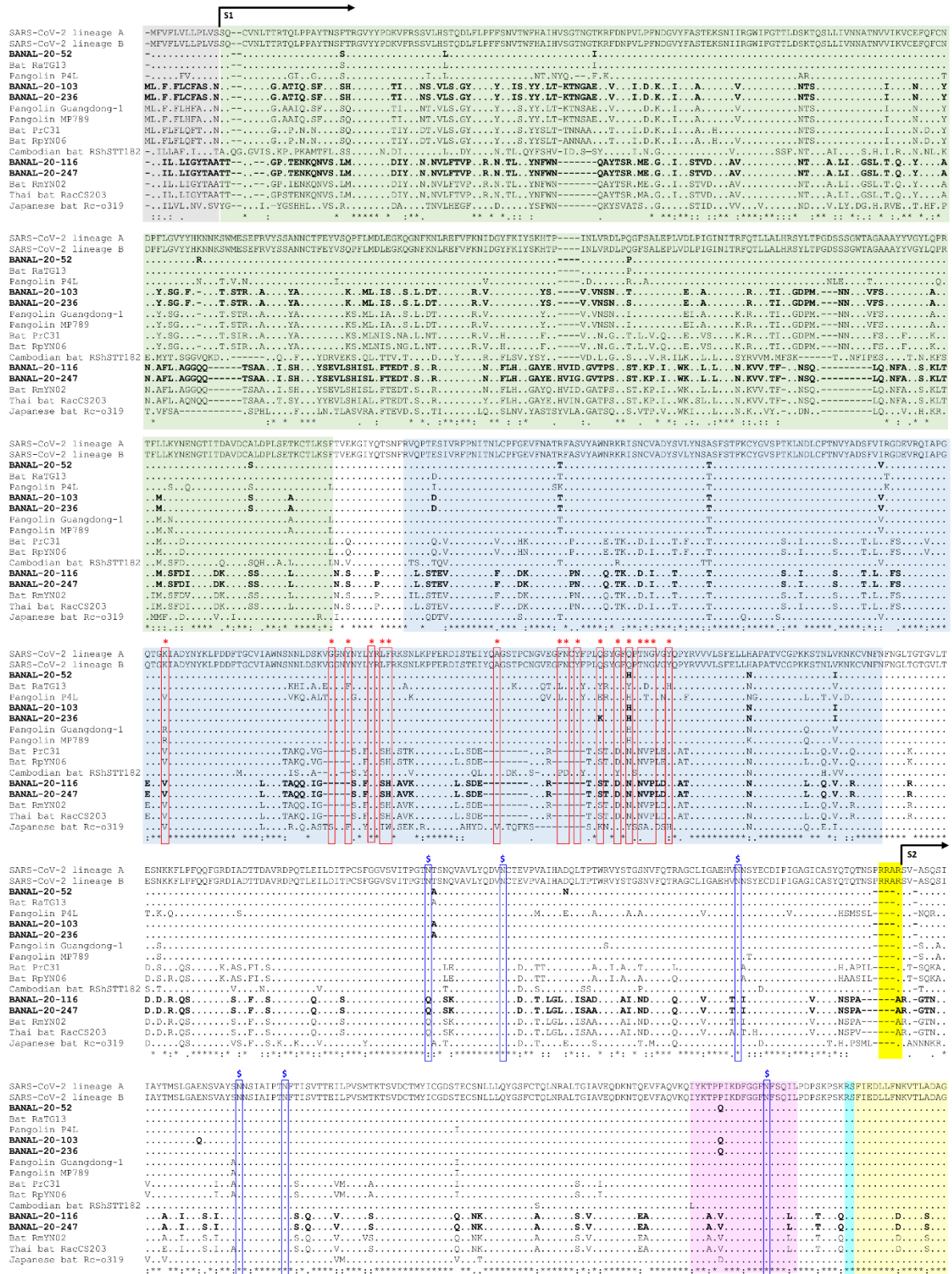
**Fragment no 15:**  
28,464 – 29,595



One should note that other studies have characterized the mosaicism of sarbecoviruses through recombination analysis. For example the program 3SEQ was used by to study the interlineage recombinations of SARS-CoV-2 variants<sup>1</sup>. Boni et al. demonstrated using 3SEQ that Sarbecovirus evolution was shaped by multiple recombination events, including several at the beginning and within the spike, as we observed in our study<sup>2</sup>. Forni et al. used 3SEQ and identified hotspots of recombination in sarbecoviruses, most of them located at the beginning of the spike<sup>3</sup>. Lytras et al. used (as in our study) GARD as a more conservative method than RDP5 (which includes a large set of programs) to identify breakpoints<sup>4</sup>. They identified 21 breakpoints through the genome of SARS-CoV-2, among which some are also found in our analysis (for example the one at the beginning of the spike, which seems to correspond to our breakpoint no 9).



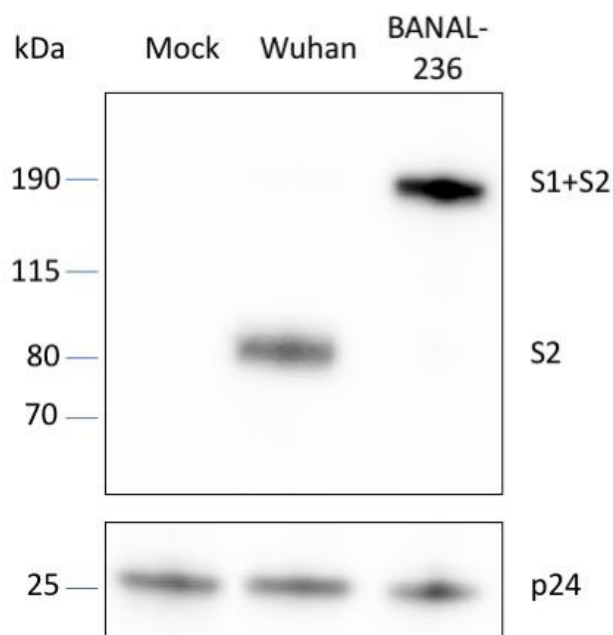
Figure S3: Alignment of the whole spike amino-acid sequences.





**Figure S4: Western-Blot analysis of concentrated lentiviral particles** pseudotyped with Wuhan or BANAL-236 spike proteins.

Pseudotyped lentiviral particles were concentrated by ultracentrifugation at 40,000g for 2h through a 20% sucrose cushion (MLS-50 rotor in an Optima MAX-XP ultracentrifuge). Pellets were resuspended in 250  $\mu$ L of DMEM for 30 min at 4°C. An equal amount of 2X NuPAGE LDS sample buffer was added and different volumes were loaded onto a 4-12% NuPAGE Bis-Tris gel (ThermoFisher). PVDF membranes (Hybond®, Amersham) were incubated with primary antibodies directed against Spike (ABIN199984, 1/2,000) or p24 (NIH, 1/2000) and revealed with peroxidase-conjugated secondary antibodies (GE Healthcare) and the ECL2 substrate (ThermoFisher). The chemiluminescence signals were acquired using a ChemiDoc Imager (Bio-Rad) (uncropped blot below). A single experiment was performed, and cropped and un-cropped blots are presented below.



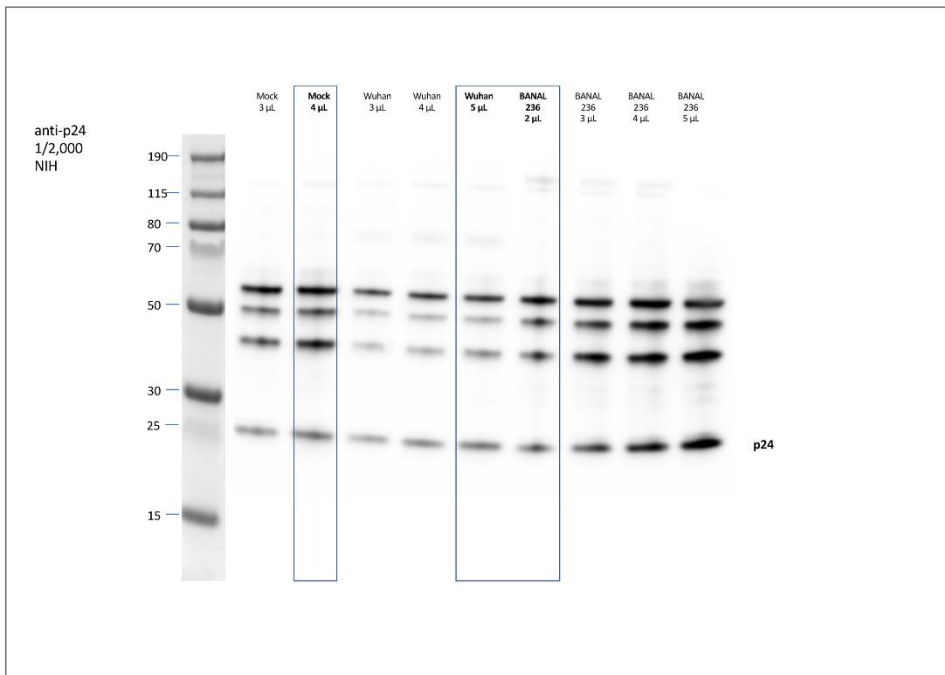
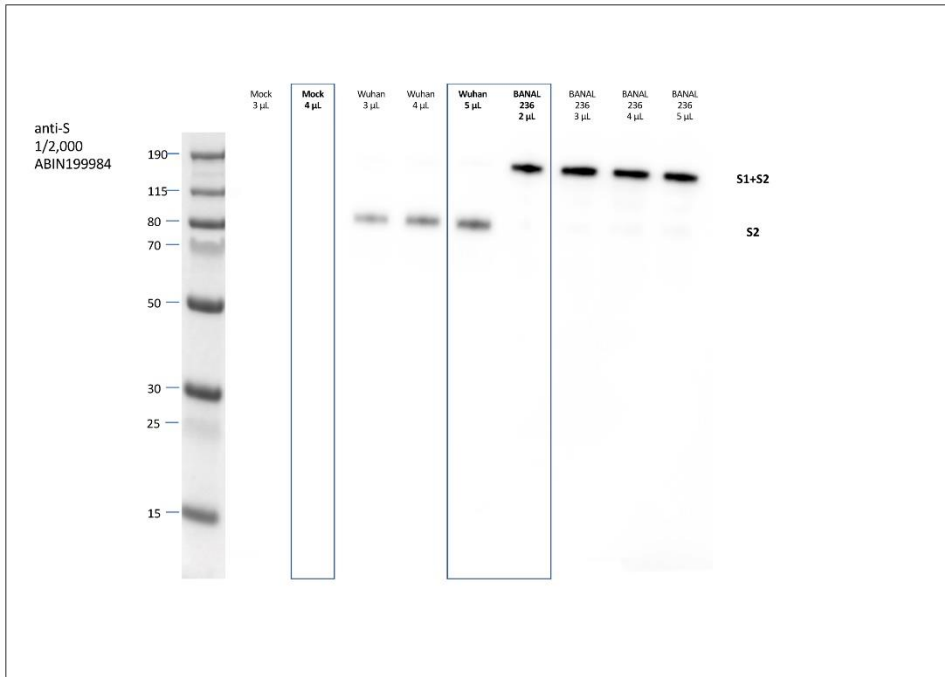
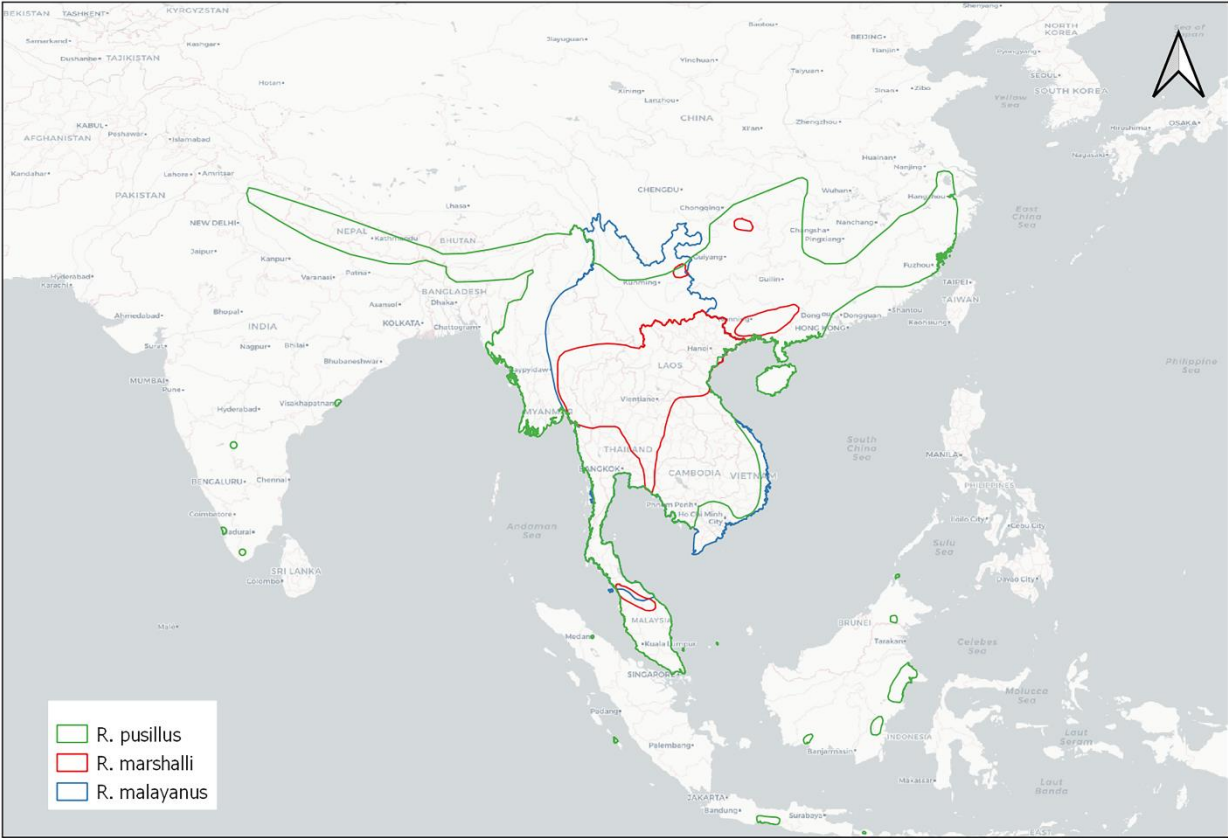


Figure S5: Distribution of bats species from which SARS-CoV-2-like viruses were detected in this study. Distribution of *R. pusillus* (green), *R. marshalli* (red), and *R. malayanus* (blue). Data were retrieved from the IUCN Red List of Threatened Species (<https://www.iucnredlist.org/>) and map was downloaded from <https://www.diva-gis.org/gdata>.



## References

1. Jackson, B. *et al.* Generation and transmission of interlineage recombinants in the SARS-CoV-2 pandemic. *Cell* **184**, 5179-5188.e8 (2021).
2. Boni, M. F. *et al.* Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat Microbiol* **5**, 1408–1417 (2020).
3. Forni, D., Cagliani, R. & Sironi, M. Recombination and Positive Selection Differentially Shaped the Diversity of Betacoronavirus Subgenera. *Viruses* **12**, E1313 (2020).
4. Lytras, S. *et al.* *Exploring the natural origins of SARS-CoV-2 in the light of recombination.* 2021.01.22.427830 <https://www.biorxiv.org/content/10.1101/2021.01.22.427830v3> (2021) doi:10.1101/2021.01.22.427830.