SARS-CoV-2-Related Bat Virus in Human Relevant Models Sheds Light on the Proximal Origin of COVID-19

Sarah Temmam

   Institut Pasteur   https://orcid.org/0000-0003-3655-9220

Xavier Montagutelli

   Institut Pasteur   https://orcid.org/0000-0002-9372-5398

Cécile Hérate

   Université Paris-Saclay, Inserm, CEA, Center for Immunology of Viral, Auto-immune, Hematological and Bacterial diseases

Flora Donati

   Institut Pasteur   https://orcid.org/0000-0003-2862-9751

Beatrice Regnault

   Institut Pasteur

Mikaël Attia

   Institut Pasteur

Eduard Baquero Salazar

   Institut Pasteur

Delphine Chrétien

   Institut Pasteur

Laurine Conquet

   Institut Pasteur

Grégory Jouvion

   Ecole nationale vétérinaire d'Alfort

Juliana Pipoli da Fonseca

   Institut Pasteur

Thomas Cokelaer

   Institut Pasteur   https://orcid.org/0000-0001-6286-1138

Faustine Amara

   Institut Pasteur

Francis Relouzat

   CEA

Thibaut Naninck

  Université Paris-Saclay, Inserm, CEA, Center for Immunology of Viral, Auto-immune, Hematological and Bacterial diseases   https://orcid.org/0000-0002-6261-6412

Julien Lemaitre

  CEA – Université Paris Saclay – INSERM U1184

Nathalie Dereuddre-Bosquet

  CEA   https://orcid.org/0000-0001-6682-6313

Quentin Pascal

  Université Paris-Saclay, INSERM, CEA

Max Bonomi

  Institut Pasteur   https://orcid.org/0000-0002-7321-0004

Thomas Bigot

  Institut Pasteur

Sandie Munier

  Institut Pasteur

Félix Rey

  Institut Pasteur   https://orcid.org/0000-0002-9953-7988

Roger Le Grand

  CEA-Université Paris Sud-Inserm, U1184, IDMIT Department, IBFJ, Fontenay-aux-Roses

Sylvie van der Werf

  Institut Pasteur   https://orcid.org/0000-0002-1148-4456

Marc Eloit (✉ marc.eloit@pasteur.fr)

  Institut Pasteur   https://orcid.org/0000-0002-1853-7207

Article

Keywords:

SARS-CoV-2-RELATED BAT VIRUS IN HUMAN RELEVANT MODELS SHEDS LIGHT ON THE PROXIMAL ORIGIN OF COVID-19

Sarah Temmam*[1,2], Xavier Montagutelli*[3], Cecile Herate*[4], Flora Donati*[5,6], Beatrice Regnault[1,2], Mikael Attia[5], Eduard Baquero Salazar[7], Delphine Chretien[1,2], Laurine Conquet[3], Grégory Jouvion[8,9], Juliana Pipoli Da Fonseca[10], Thomas Cokelaer[10], Faustine Amara[5], Francis Relouzat[4], Thibaut Naninck[4], Julien Lemaitre[4], Nathalie Derreudre-Bosquet[4], Quentin Pascal[4], Massimiliano Bonomi[11], Thomas Bigot[1,12], Sandie Munier[5], Felix Rey[7], Roger Le Grand[4], Sylvie van der Werf[5,6], Marc Eloit[1,2,13,+]

*These authors contributed equally to this work.

+ Corresponding author: Institut Pasteur, Paris, France. marc.eloit@pasteur.fr

1 Institut Pasteur, Université Paris Cité, Pathogen Discovery Laboratory, Paris, France.

2 Institut Pasteur, Université Paris Cité, The OIE Collaborating Center for the detection and identification in humans of emerging animal pathogens, Paris, France.

3 Institut Pasteur, Université Paris Cité, Mouse Genetics Laboratory, Paris, France.

4 Center for Immunology of Viral, Auto-immune, Hematological and Bacterial Diseases (IMVA-HB/IDMIT), Université Paris-Saclay, Inserm, CEA, Fontenay-aux-Roses, France.

5 Institut Pasteur, Université Paris Cité, CNRS UMR 3569, Molecular Genetics of RNA Viruses Unit, Paris, France.

6 Institut Pasteur, Université Paris Cité, National Reference Center for Respiratory Viruses, Paris, France.

7 Institut Pasteur, Université Paris Cité, CNRS UMR 3569, Structural Virology Unit, Paris, France.

8 Ecole nationale vétérinaire d'Alfort, Unité d'Histologie et d'Anatomie Pathologique, Maisons-Alfort, France.

9 Université Paris Est Créteil, EnvA, ANSES, Unité DYNAMIC, Créteil, France.

10 Biomics Platform, C2RT, Institut Pasteur, Université Paris Cité, Paris, France.

11 Institut Pasteur, Université Paris Cité, CNRS UMR 3528, Structural Bioinformatics Unit, Paris, France.

12 Institut Pasteur, Université Paris Cité, Bioinformatic and Biostatistic Hub - Computational Biology Department, Paris, France.

13 Ecole nationale vétérinaire d'Alfort, University of Paris-Est, Maisons-Alfort, France.

34   **Abstract:** Bat sarbecovirus BANAL-236 is highly related to SARS-CoV-2 and infects human

35   cells, albeit lacking the furin cleavage site in its spike protein. To inform on the origin of SARS-

36   CoV-2, we evaluated the clinical, epidemiological and evolutionary consequences of a potential

37   BANAL-236 spillover into humans using animal models. The virus replicates efficiently and

38   pauci-symptomatically in humanized mice and in macaques, where its tropism is enteric, strongly

39   differing from that of SARS-CoV-2. BANAL-236 infection leads to protection against

40   superinfection by a more virulent strain like Wuhan SARS-CoV-2. Yet we found no evidence of

41   antibodies recognizing bat sarbecoviruses in populations highly exposed to bats, indicating that

42   such infections, if they occur, are rare. Six passages in mice or in human intestinal cells, mimicking

43   putative early spillover events, selected adaptive mutations without appearance of a furin cleavage

44   site and not change in virulence. We thus conclude that the hypothesis of the SARS-CoV-2

45   pandemic being preceded by silent circulation in humans of BANAL-236-like strains leading to

46   the acquisition of a furin cleavage site is unlikely. Our studies suggest that a specific search for a

47   furin cleavage site in sarbecoviruses in the wild should be pursued to understand the origin of the

48   SARS-CoV-2 pandemics.

49

50   **One-Sentence Summary:**

51   BAT SARS-CoV-2-related coronavirus is mainly enterotropic in human relevant models and
52   protects against SARS-CoV-2 infection.

Main Text

The origin of SARS-CoV-2, as well as its mode of introduction into the human population, is currently unknown. Various SARS-CoV-2-related viruses have been described in *Rhinolophus shameli* (isolated in Cambodia in 2010[1]), *R. pusillus* and *R. malayanus* (China, 2020 and 2019 respectively[2]), in *R. acuminatus* (Thailand, 2020[3]), *R. cornutus* (Japan, 2013[4]) and *R. affinis,* RaTG13 (China, 2013)[5, 6]. Very recently we reported the discovery in fecal swabs of bats from Laos coronaviruses highly similar to SARS-CoV-2: BANAL-52 in *R. malayanus,* BANAL-103 in *R. pusillus* and BANAL-236 in *R. marshalli*. These were shown to bind human angiotensin-converting enzyme 2 (hACE2) more efficiently than early strains of SARS-CoV-2, to enter primate cells by an ACE2-dependent mechanism and, for the virus that could be isolated (BANAL-236), to replicate in human Caco-2 and Calu-3 cells expressing physiological levels of hACE2[7]. They lack a furin cleavage site that has been shown to reduce the pathogenicity and transmissibility of SARS-CoV-2 in animal models[8], suggesting that these bat viruses could be less pathogenic in humans. These results do not rule out these viruses from having initiated the pandemic, as SARS-CoV-2 showed very high adaptability after the pandemic start. The emergence of variants of concern (VOC) paralleled their increase in transmissibility and/or pathogenicity in humans without any decrease of the host range, as evaluated in wild-type mice[9–11].

Our objective was to investigate the possibility that infection of human populations with strains like BANAL-236 with potential low pathogenicity and/or transmissibility occurred prior to the detection of the first COVID-19 clinical cases (end of 2019), preceding adaptation to humans of more transmissible and pathogenic strains. We asked whether such human infections could have been detected clinically, could have led to a cycle of human-to-human transmission, and whether adaptive mutations could have occurred in a first phase of silent circulation in humans before the emergence and detection of symptomatic cases.

RESULTS

Bat BANAL-236 replicates in transgenic mice expressing hACE2

In this study, we aimed to analyze whether mice that express hACE2 (K18-hACE2) were susceptible to BANAL-236. To this end, K18-hACE2 mice were inoculated with BANAL-236 or SARS-CoV-2 strains (strain BetaCoV/France/IDF0372/2020, GISAID accession number EPI_ISL_406596, referred below as "Wuhan-372"), while BALB/cJRj mice were used as controls (Fig. 1A). Mice inoculated with BANAL-236 showed moderate weight loss at 10^4 PFU (Fig. 1B), but not at 10^3 PFU. Weight loss was more pronounced in mice infected with the Wuhan-372 strain than the BANAL-236 strain at 10^4 PFU. Compared with Wuhan-372, BANAL-236-infected K18-hACE2 mice showed a lower genome copy number and virus titer in the lungs (-1.66 and -1.65 $\log_{10}$, respectively, Fig. 1C). Infectious BANAL-236 virus, however, could not be detected in the liver and spleen of K18-hACE2 mice, and the genome copy numbers were 2.53 $\log_{10}$ (spleen) and 5.53 $\log_{10}$ (liver) lower than in the lungs. Significant differences between the two viruses were not observed (-0.53 $\log_{10}$ in the spleen and -0.44 $\log_{10}$ in the liver between BANAL-236 and Wuhan-372). BALB/c mice could be infected by BANAL-236, but at lower levels compared with K18-hACE2 mice and without any clinical sign of infection (Fig. 1B, 1C). Mice infected with BANAL-236 developed an antibody response, which neutralized the homologous strain more efficiently than the Wuhan-372 strain (Fig. 1D, day D30; Fig. S1).

Mice infected by BANAL-236 at 10^4 PFU were fully protected against lethality and weight loss after challenge with 10^4 PFU of a virulent SARS-CoV-2 strain (hereafter called "Wuhan-D614G", strain hCoV-19/France/GES-1973/2020, GISAID accession number EPI_ISL_414631), while naive K18-hACE2 mice infected by Wuhan-D614G presented a 100% mortality rate at day 8 post-infection (Fig. 1E, logrank test, p=0.00095). At 10^3 PFU, 1 of 6 mice primo-infected by BANAL-236 and challenged by Wuhan-D614G was euthanized due to severe symptoms while 5/6 experienced moderate and transient body weight loss (Fig. S2). This cross-protection (comparison

4

102   with Wuhan-D614G, logrank test, p=0.0019) was less efficient than that conferred by the Wuhan-

103   372 strain (logrank test, p=0.00095), for which, at the 10^3 PFU dose, no body weight loss was

104   recorded after challenge (Fig. 1E). The neutralizing antibody responses against BANAL-236 and

105   Wuhan-372 were boosted after the challenge with the Wuhan-D614G strain (Fig. 1D, D45).

106   Histopathological analysis of the lung at day 3 post-inoculation of K18-hACE2 mice infected with

107   10^4 PFU revealed less severe lesions BANAL-236- than in Wuhan-372-infected mice, with

108   interstitial inflammation often centered on blood vessels or bronchioles and endothelial cell injury

109   (Fig. 1F, Fig. S3).

110   Therefore, BANAL-236 replicates in hACE2 transgenic mice, but its pneumotropism and its

111   pathogenicity are lower than those of the SARS-CoV-2 Wuhan strain.

112   **Pneumotropism of BANAL-236 is not increased by serial passages in hACE2 transgenic**
113   **mice**

114   We conducted six passages of the initial BANAL-236 strain by inoculating three K18-hACE2

115   mice intranasally with pooled lung homogenates from the previous passage (Fig. 1G). The pooled

116   lung homogenates from the last passage (P6, 10^4 PFU) were then inoculated into K18-hACE2 mice

117   to compare the pathogenicity of P6 with that of P1. We found no significant differences in body

118   weight variation between the two strains (Fig. 1G). The number of genome copies per g of lung

119   increased upon passages up to 9.7 log10 gc/g at P4 and remained constant during the next passages,

120   never reaching the levels observed for the Wuhan-372 strain at the first passage (11.3 log10 gc/g).

121   We did not detect any viral RNA in the liver of mice inoculated with P6 (data not shown), whereas

122   the viral load was 4.1 log10 gc/g for P1. Genome copy numbers in the spleen decreased by 1.4

123   $\log_{10}$ between mice infected by P1 and P6. One mouse of three infected by BANAL-236 P6

124   presented a detectable but non quantifiable viral load in the feces or in the rectum (data not shown).

125   These results show that the six serial passages in humanized mice did not modify the

126   pneumotropism of BANAL-236 nor steer the virus towards a more virulent phenotype.

127   BANAL-236 behaves as an enteric virus in macaques

128   Two cynomolgus macaques (#MF1 and # MF2) were exposed to 0,6 10^6 PFU of BANAL-236 by

129   the intranasal and intratracheal routes simultaneously and followed for 43 days (Fig. 2A). Body

130   temperature and clinical score remained within normal values but #MF2 showed progressive

131   weight loss (Fig. 2B). None of the animals had detectable genomic viral loads in broncho-alveolar

132   lavage (BAL) and in nasopharyngeal and tracheal samples (swabs), as assessed by RT-qPCR for

133   viral genomic and sub-genomic RNAs, except a low transient blip for the two macaques at D2-3

134   post-infection in nasopharyngeal fluids (2.9-3.2 $\log_{10}$ copies/mL) and in BALs (4.1-3.4 $\log_{10}$

135   copies/mL) (Fig. 2C). Viral sub-genomic RNAs (sgRNAs), which are indirect evidence for viral

136   replication, were not detectable in nasopharyngeal swabs nor BALs. In contrast, the two macaques

137   shed the virus in the feces (rectal swabs) at D11 (3.52 $\log_{10}$ copies/mL) for animal #MF1 and from

138   D7 to D14 for animal #MF2 that peaked at D11 at high level (7.84 $\log_{10}$ copies/mL). Both animals

139   had undetectable viral loads at D21. Viral sub-genomic RNA peaked at D11 for animal #MF2 (Fig.

140   2C). Of note, this individual presented progressive weight loss from D3 to D28, suggesting a

141   possible link with viral excretion of BANAL-236 virus in the feces. Results were compared to

142   historical data obtained for SARS-CoV-2 (Wuhan-372 strain) when administered by the same

143   standardized protocol to 33 or 11 macaques at 1.10^6 or 1.10^5 PFU, respectively. The replication

144   profile of SARS-CoV-2 and BANAL-236 was strikingly different, with lung tropism (BAL) and

145   respiratory shedding (tracheal and nasopharyngeal swab samples) being dominant for SARS-CoV-

146   2, and enteric tropism for BANAL-236 (Fig. 2C).

147   The two macaques infected by BANAL-236 showed very mild pulmonary lesions (Fig. S4)

148   characterized by non-extended ground-glass opacities detected by chest computed tomography

149   (CT score of 1), while macaques infected by SARS-CoV-2 showed CT scores of 5 or more (Fig.

150   2B)(*12*). BANAL-236-infected animals experienced a transient lymphopenia at D2 or D3 while

151   other hematological parameters remained within normal ranges (Fig. S5). Sera of the immunized

152   macaques exhibited BANAL-236 and cross-reactive SARS-CoV-2 neutralizing responses (Fig.

153   2B).

154   Therefore BANAL-236 largely behaves like an enteric virus, and displays low pathogenicity in

155   macaques.

156   **Mutations selected upon passages in human cellular or animal models**
157   We attempted to model the human transmission cycles of BANAL-236 by passaging it 6 times in

158   Caco-2 cells and in transgenic mice expressing hACE2, beginning with the strain initially isolated

159   and amplified by 2 passages in Vero E6 cells (Fig. 3A). We did not passage the virus further as

160   the objective was not to adapt the virus to human cells but to mimic a worst-case scenario

161   considering the low transmissibility of the virus, as informed by previous animal studies for SARS-

162   CoV-2 and as developed below for BANAL-236. We did not find evidence for any variation within

163   or around the furin cleavage site despite high vertical sequencing depth (Table S1). We confirmed

164   with standard PCR that at the same locus no insertion was detectable (Fig. S6).

165   However, we identified a set of mutations that were not detectable in the original BANAL-236

166   fecal swab despite a high sequencing depth (Table S2). Notably, a missense mutation V391I (V395

167   in SARS-CoV-2 Wuhan strains), already detected after the first passages in Vero E6 cells, became

168   dominant after the first passage in intestinal Caco-2 cells. The same mutation was positively

169   selected in the course of macaque infection (from day 7 to day 14) while being counter-selected in

170   K18-hACE2 mice (Fig. 3B, Fig. S7). This mutation was associated with other mutations appearing

171   during passages in Caco-2 cells (NSP10 N4350K [N4358 in SARS-CoV-2], endoRNAse NSP15

172   N6448T [N6456 in SARS-CoV-2], ORF3a silent mutation T248 [T248 in SARS-CoV-2]) or

173   during macaque infection (spike NTD S52R [T51 in SARS-CoV-2], NSP14 3'-5' exonuclease

174   H5943Q [H5951 in SARS-CoV-2]), suggesting a possible complementarity of these mutations

175   affecting different proteins or genes.

176   On the other hand, passaging the virus in transgenic mice selected a cluster of mutations located

177   in the spike (P627L [P631 in SARS-CoV-2], silent T108 [T108 in SARS-CoV-2]); in the RdRp

178   (silent N4998 [N4992 in SARS-CoV-2]; in ORF6 (A12T [A12 in SARS-CoV-2]), and in ORF3a

179   (L101R [L101 in SARS-CoV-2]) that were specific to the mouse model (Fig. 3B). These mutations

180   were identified after the first passage and became predominant thereafter. Remarkably, these

181   mutations were different from those observed in Caco-2 cells (Fig. 3C), suggesting that they

182   probably do not impact infection pathways shared by transgenic mice and Caco-2 cells. Silent

183   mutations were also identified but the codon used in SARS-CoV-2 for the corresponding residue

184   was most generally the original codon of BANAL-236 sequenced from the bat feces and not the

185   mutated one (Table S3).

186   To gain insight into the phenotype of the missense mutations of the spike RBD-V391I [V395 in

187   SARS-CoV-2 Wuhan strains] and in particular its effect on binding to hACE2, we performed all-

188   atom explicit solvent Molecular Dynamics (MD) simulations of the wild type (WT) BANAL-236

189   RBD/hACE2 complex and the mutant (Table S4). The MD simulations were initiated from the X-

190   ray crystal structure of the WT complex (PDB code 7PKI)(*13*) and an homology model of the

191   mutant and were extended for a total aggregated simulation time of 6 μs. The MD simulations did

192   not reveal significant differences upon mutation in the stability of the RBD or in the interaction

193   strength with hACE2 in the microsecond timescale (Fig. S8). The FoldX scoring function predicted

194   similar binding affinities for all the RBD-hACE2 complexes studied (Fig. S9). Moreover,

195   persistent hydrogen bonds (Fig. S10) and salt bridges (Fig. S11) observed at the WT RBD-hACE2

196   interface were conserved in the mutated complex.

197 Regarding the S1-P627L mutation, this residue belongs to a loop in S1 that shares high amino acid

198 identity with SARS-CoV-2. The loop is at the interface with the NTD and is not resolved in most

199 cryo-EM structures of the spike except for the Omicron variant (*14*), which indicates a high

200 flexibility in this region of the spike. Therefore, we do not expect that the P627L mutation will

201 drastically affect the overall structure and function of the spike besides inducing minor local

202 conformational changes.

203 Finally, the S52R mutation (T51 in SARS-CoV-2) located in the NTD, which was positively

204 selected during the course of macaque infection, is located at the interface of the trimer but is not

205 fully surface exposed. This residue does not seem to interact with any other residue in the vicinity

206 but may possibly affect the spike overall dynamic.

207 **Sarbecoviruses serosurvey in humans exposed to bats**

208 In order to determine whether humans in close contact with bats could have been infected by bat

209 sarbecoviruses, we compared reactivity of sera from a population in close contact with bats in this

210 sampling area (guano collectors, bat hunters or sellers) and with that of sera from the general

211 population in Laos collected at the same period (2020)(*14,15*) or earlier, before the emergence of

212 COVID-19 in China (2019). Pre-pandemic sera collected in France and matched for sex and age

213 with the 2019 Lao pre-pandemic sera served as a negative control for subjects exposed to these

214 viruses, and sera from Lao people infected with SARS-CoV-2 were used as a positive control. Sera

215 were tested for their ability to neutralize pseudotypes of BANAL-236, -52 and -103 viruses or to

216 bind their S proteins by two independent techniques (LIPS and LuLISA). None of the sera from

217 the three Lao non-COVID populations showed neutralizing activity against any of the three viruses

218 above the background signals identified in the control French population (Fig. 4 A-C, PNT). Sera

219 from the general Lao population in the pre-2019 period showed higher binding to S-52 than the

220 control sera collected in France (Fig. 4B, LIPS). However, this binding was not confirmed by

221   LuLISA for the same population, nor by any of the techniques for the population most exposed to

222   bats and we interpreted this signal as non-specific. Control sera from patients infected with SARS-

223   CoV-2 showed cross-neutralization and cross-binding to bat viruses BANAL-236, -52 and -103 as

224   expected, given the proximity of their sequences (Fig. 4 A-C). Therefore, no human infection by

225   bat sarbecoviruses was detected, with an upper limit of infection rate detection of 4.1% (bat-

226   exposed population, 2020, n=74, p=0.05) and 1.5% (general population 2019 and 2020, n=200,

227   p=0.05).

228   **DISCUSSION**

229   The origin of the COVID-19 pandemic could have proceeded from direct human infection by bat

230   viruses like BANAL-236 or BANAL-52, which is supported by their very close proximity with

231   SARS-CoV-2 and the failure to identify intermediate hosts contrary to their rapid detection for

232   SARS-CoV[16] and MERS-CoV[17, 18]. According to this hypothesis, people may have been

233   infected in the South of China or neighboring countries by bats while visiting caves such as for

234   trapping purposes, or by their products like meat or guano, and travelled to Wuhan. Another

235   circumstance could be a leakage from a laboratory that propagated bat sarbecoviruses in Wuhan.

236   To aid in evaluating the likelihood of these scenarios, we have taken advantage of our successful

237   isolation of the BANAL-236 virus to anticipate the clinical consequence of such infection and the

238   subsequent evolution of the virus. Our results add three important elements. First, data in

239   transgenic mice and macaques suggest that human infections, if any, could have been sub-clinical,

240   or led to mild symptoms that could have been easily missed. Second, we show that BANAL-236

241   was mainly enterotropic and shed at high titer in the feces of macaques infected by the nasal and

242   tracheal routes. This is reminiscent of SARS-CoV, which was enterotropic in addition to being

243   pneumotropic[19, 20]. Notably, viral load in mouse lungs was around 2 logs lower than upon

244   SARS-CoV-2 infection, suggesting that transmission through the respiratory route was less

245   efficient for BANAL-236 than for SARS-CoV-2. Therefore, if a similar strain had circulated

246   initially in humans, it was likely an enteric virus that remained undetected until acquiring a more

247   efficient spreading ability, becoming also more pathogenic because of its adaptation to the

248   respiratory tract. This spreading ability became further optimized in the variants that emerged later

249   (delta, omicron). Spreading of enteric viruses is generally much less efficient than for respiratory

250   viruses and does not lead to pandemics. In addition, people infected by such strains might be

251   protected against more virulent strains (Fig. 1E). Third, we did not evidence human infection in a

252   population in contact with bats, in contrast with previous observations that used less specific

253   nucleocapsid-derived antigen(*21*), which supports the hypothesis that BANAL viruses are highly

254   transmissible neither from bats to humans nor between humans.

255   We have mimicked the likely evolution of these bat viruses in humans by performing 6 passages

256   in transgenic mice through the respiratory route and in Caco-2 cells, a human intestinal epithelial

257   cell line. We evidenced mutations along the genome, but none was found among hACE2

258   contacting residues of the RBD, nor were predicted by our MD simulations to increase affinity for

259   hACE2, which suggests that the very high affinity of BANAL-236 RBD for hACE2 did not benefit

260   from additional selection pressure. Within the spike, V391I, which has no impact on RBD binding

261   to hACE2(*22*), was positively selected in intestinal Caco-2 cells and in monkey feces but not in

262   the lungs of transgenic mice. This mutation could therefore reflect either an adaptation to the host

263   species or the intestinal tropism of BANAL-236 in primates. Other mutations selected in

264   transgenic mice (NSP12 RdRP N4984N, S1 P627L, NS6 A12T) and in monkey feces (NSP14

265   ExoN H5943Q and S1-NTD S52R) warrant further investigation regarding their phenotype and

266   complementarity as it is noteworthy that they were selected synchronously during passages in mice

267   or in the course of infection in macaques. Nevertheless, all residues mutated during passages, like

268   most of the silent mutations, are identical in the Wuhan strain and in BANAL-236 sequenced from

269   bat fecal swabs. Moreover, these mutations were rare in the variants described in the GISAID

270   database (Table S5), suggesting that they could correspond to an early adaptive process but were

271   not stabilized in the consensus of early and late SARS-CoV-2 strains. The increase in fitness that

272   they impart in human cells therefore could only occur when the spike lacks a furin cleavage site.

273   The acquisition of a furin cleavage site could therefore have been the key event leading to the

274   SARS-CoV-2 epidemic and pathogenic Wuhan strain. Furin cleavage sites have a well-known role

275   in avian influenza subtypes H5 and H7 pathogenesis. Indeed, acquisition of a furin cleavage site

276   composed of arginine (R) or lysine (K) in avian influenza virus hemagglutinin after replication in

277   chickens is a documented event leading to the emergence of highly pathogenic avian influenza

278   viruses (HPAIVs) from low pathogenic avian influenza viruses (LPAIVs) after spillover from wild

279   aquatic birds. Such a mechanism can be reproduced by passages *in ovo*(*23–25*). The furin site is

280   thought to be created by duplication of lysine and arginine residues by polymerase slippage, as

281   these amino acids are encoded by purine-rich codons(*26*). In our experiments, 6 passages in mice

282   by the respiratory route or in human intestinal cells did not lead to the selection of a furin site. It

283   can be argued that more passages could have allowed the selection of mutants with a polybasic

284   site, as for the influenza virus which requires up to 11 *in vivo* passages from LPAIV strains to see

285   HPAIVs strains emerge(*24*). This high number of passages is consistent with the efficient

286   circulation of LPAIVs in high-density poultry flocks. Regarding bat sarbecoviruses, in the context

287   of the demonstration of, at worst, a low circulation in humans exposed to bats, our results make

288   the spontaneous appearance of a furin site during silent circulation in humans unlikely.

289   The hypothesis of intense circulation in a different animal species remains open, but there is no

290   reason to believe that such circulation would have been more efficient than in humans given the

291   high affinity of the spike of BANAL viruses for human ACE2(*13*). Based on our work, it appears

292   that the most probable hypotheses regarding the natural origin of the virus are the existence of bat

293 viruses harboring a furin site that could have infected humans directly or via other animal species.

294 Interestingly, infection with BANAL-236 conferred cross-protection in mice against a SARS-

295 CoV-2 challenge, which suggests that herd immunity against putative more virulent bat

296 sarbecovirus strains could exist in local human populations frequently infected by bat viruses,

297 albeit not detected in our study. Alternatively, human infection might have occurred in the course

298 of experimental virus isolation in the laboratory(*27*). Concerning the first hypothesis, we are

299 pursuing investigations in caves located within this vast karstic biotope shared by China and

300 neighboring countries.

301 **Materiel and Methods**
302 Ethics

303 All work done in mice was approved by the Institut Pasteur Ethics Committee (project dap 210050)

304 and authorized by the French Ministry of Research under #31816 in compliance with the European

305 and French regulations on the protection of live vertebrates and the Standards for Human Care and

306 Use of Laboratory Animals, of the Office for Laboratory Animal Welfare (OLAW, assurance

307 number #F16-00110 A5476-01).

308 Cynomolgus macaques (*Macaca fascicularis*) originating from Mauritian AAALAC certified

309 breeding centers were used in this study. All animals were housed within IDMIT animal facilities

310 at CEA, Fontenay-aux-Roses under BSL-3 containment when necessary (Animal facility

311 authorization #D92-032-02, Préfecture des Hauts de Seine, France) and in compliance with

312 European Directive 2010/63/EU, the French regulations and the Standards for Human Care and

313 Use of Laboratory Animals, of the Office for Laboratory Animal Welfare (OLAW, assurance

314 number #A5826-01, US). Animals tested negative for *Campylobacter*, *Yersinia*, *Shigella* and

315 *Salmonella* before being used in the study.

316   The protocols were approved by the institutional ethical committee "Comité d'Ethique en

317   Expérimentation Animale du Commissariat à l'Energie Atomique et aux Energies Alternatives"

318   (CEtEA #44) under statement number A20-037. The study was authorized by the "Research,

319   Innovation and Education Ministry" under registration number APAFIS#24434-2020030216532863

320   v3. All information on the ethics committee is available at https://cache.media.enseignementsup-

321   recherche.gouv.fr/file/utilisation_des_animaux_fins_scientifiques/22/1/comiteethiqueea17_juin2

322   013_257221.pdf.

323   <u>Human samples</u>

324   Pre-COVID-19 human serum samples (n=100) were collected in the context of a hospital-based

325   serostudy in Saravan province, southern Laos, 2017 (Lao National Ethics Committee for Health

326   Research reference 018/NECHR/2017(*27*). Serum samples from late 2020 (n=100) were collected

327   from the general Lao population, healthcare workers and bat contacts from six provinces (Ref

328   052/NECHR/2020(*15*)). The bat contacts group (n=74) consisted of guano collectors and bat

329   hunters or sellers from four villages located in areas alongside isolated limestone karsts where

330   BANAL-20-236 and other SARS-like coronaviruses were detected from bats. SARS-CoV-2

331   positive serum samples (n=15) were obtained from patients during routine COVID-19

332   surveillance.

333   <u>Cells</u>

334   Vero E6 and Caco-2 cells (ATCC) were grown in Dulbecco's modified Eagle's medium (DMEM)

335   with 5% fetal bovine serum (Thermo Fisher Scientific) and 1% penicillin-streptomycin (Thermo

336   Fisher Scientific). All cell lines were tested negative for mycoplasma.

337   <u>Mice infection</u>

338 B6.Cg-Tg(K18-ACE2)2Prlmn/J transgenic mice were imported from The Jackson Laboratory (SN

339 #34860) and bred at the Institut Pasteur. BALB/cJRj were purchased from Janvier Labs (Le Genest

340 St Isle, France). Mice were maintained under specific-pathogen-free conditions with a 14-h light

341 and 10-h dark cycle and ad libitum food and water in the Institut Pasteur animal facility.

342 Infection studies were performed in animal biosafety level 3 (BSL-3) facilities at the Institut

343 Pasteur, on 11 to 19-week-old female mice. Forty to 60 microliters of either viral suspension or

344 lung homogenate were inoculated intranasally to ketamine-xylazine-anesthetized mice. Group size

345 and inoculation dose are indicated in the figure legends. Clinical signs of disease and weight loss

346 were monitored daily. Mice reaching predefined humane endpoints, the end of the observation

347 period or the time point for sample collection were euthanized by ketamine/xylazine overdose.

348 Blood was collected by puncture of the retroorbital sinus. For viral quantification, tissue samples

349 were placed in lysing matrix D tubes (MP Biomedical) and stored at -80°C.

350 For serial passaging, three mice (P1) were inoculated with 10^3 PFU of BANAL-236 and euthanized

351 at 3 days post-infection. The right lung lobe was placed on a 70µ cell strainer (Falcon), minced

352 with fine scissors and ground with a syringe plunger using 400 µL of PBS. The three lung

353 homogenates were mixed in equal volumes and 40 µL of the mix was inoculated to each of three

354 P2 mice. Viral load and viral titers were measured on the mix. The procedure was repeated until

355 P6. Lung homogenates from P6 mice were frozen until viral titer was obtained. P7 mice were

356 inoculated with 10^4 PFU of P6 mix.

357 <u>Histology</u>

358 Mouse lungs were fixed for one week in 10% neutral-buffered formalin for biosafety reasons and

359 transferred into 70% ethanol. Samples were embedded in paraffin. Four µm-thick sections were

360 cut and stained with hematoxylin and eosin (HE).

361 Macaques infection

362 Two female cynomolgus macaques aged 4-5 years were exposed to a total dose of $0,6.10^6$ PFU of

363 BANAL-236 virus isolated and passaged twice in VeroE6 cells via the combination of intranasal

364 and intra-tracheal routes (0.25 mL in each nostril and 4.5 mL in the trachea, i.e. a total of 5 mL;

365 day 0), using atropine (0.04 mg/kg) for pre-medication and ketamine (5 mg/kg) with medetomidine

366 (0.05 mg/kg) for anesthesia, with sample collection as previously described(*29*). Blood cell counts,

367 haemoglobin and haematocrit were determined from EDTA blood using a DXH800 analyzer

368 (Beckman Coulter).

369 BANAL-236 genomic and sub-genomic RNA RT-qPCR for viral load assessment in NHPs

370 Upper respiratory (nasopharyngeal and tracheal) and rectal specimens were collected with swabs

371 (Viral Transport Medium, CDC, DSR-052-01). Tracheal swabs were performed by insertion of the

372 swab above the tip of the epiglottis into the upper trachea at approximately 1.5 cm of the epiglottis.

373 All specimens were stored between 2°C and 8°C until analysis by RT-qPCR with a plasmid

374 standard concentration range containing an RdRp gene fragment including the RdRp-IP4 RT-PCR

375 target sequence. The limit of detection was estimated to be 2.67 $\log_{10}$ copies of SARS-CoV-2

376 gRNA per mL and the limit of quantification was estimated to be 3.67 $\log_{10}$ copies per mL. SARS-

377 CoV-2 sub-genomic mRNA (sgRNA) levels were assessed by RT-qPCR targeting the E gene using

378 primers and probes previously described(*30*): leader-specific primer sgLeadSARSCoV2-F

379 CGATCTCTTGTAGATCTGTTCTC, E-Sarbeco-R primer ATATTGCAGCAGTACGCACACA

380 and E-Sarbeco probe HEX579 ACACTAGCCATCCTTACTGCGCTTCG-BHQ1. The protocol

381 describing the procedure for the detection of SARS-CoV-2 is available on the WHO website

382 (https://www.who.int/docs/default-source/coronaviruse/real-time-rt-pcr-assays-for-the-detection-of-sars-

383 cov-2-institut-pasteur-paris.pdf?sfvrsn=3662fcb6_2). The limit of detection was estimated to be 2.87

384    $\log_{10}$ copies of SARS-CoV-2 sgRNA per mL and the limit of quantification was estimated to be

385   3.87 $\log_{10}$ copies per mL.

386   Chest CT Protocol

387   All imaging acquisitions were performed on the Digital Photon Counting (DPC) PET-CT system

388   (Vereos-Ingenuity, Philips) implemented in BSL3 laboratory. For imaging sessions, animals were

389   first anesthetized with Ketamine (10mg/kg) + Metedomidine (0.05mg/kg) and then maintained

390   under isofluorane 2% in a supine position on a patient warming blanket (Bear Hugger, 3M) on the

391   machine bed with cardiac rate, oxygen saturation and temperature monitoring.

392   CT was performed under breath-hold with a detector collimation of 64×0.6 mm, the tube voltage

393   was 120 kV and intensity of about 150mAs. Automatic dose optimization tools (Dose Right, Z-

394   DOM, 3D-DOM by Philips Healthcare) regulated the intensity. CT images were reconstructed

395   with a slice thickness of 1.25 mm and an interval of 0.25 mm.

396   Images were analyzed using INTELLISPACE PORTAL 8 (Philips healthcare) and 3DSlicer (Open

397   source tool). Pulmonary lesions were defined as Ground Glass Opacity, Crazy603 paving pattern

398   or consolidation as previously described. Lesion features detected by CT imaging were assessed

399   by two analyzers independently and final CT score results were obtained by consensus.

400   Generation of lentiviral pseudoviruses and neutralization assay

401   The lentiviruses expressing BANAL-236 and Wuhan-372 synthetic spike genes were described

402   previously(*13*). Those expressing the spikes of BANAL-52 and BANAL-103 were constructed

403   similarly. Neutralization assays were conducted as described(*13*).

404   Luciferase Immunoprecipitation Assays

405   Recombinant antigens comprising the RBD or the ectodomain of the full spike of BANAL-236, -

406   52, and -103 viruses were designed in fusion with the nanoluciferase as follows: the foldon domain

407   (YIPEAPRDGQAYVRKDGEWVLLSTFL) was added to the C-terminus of each ectodomain to

408   allow the S protein to trimerize(*30*), resembling the native spike state of the virion. The

409   nanoluciferase was added to the carboxy-terminal end of each construct spaced by a 3-residues

410   GSG linker. An exogenous signal peptide was added to RBD constructs to ensure efficient protein

411   secretion. Synthetic genes were ordered from GenScript with codon usage optimized for protein

412   expression in mammalian cells and cloned in the pcDNA3.1(+) vector. Plasmids were transfected

413   in Expi293F Gn-TI cells using the ExpiFectamine293 Transfection kit (Fisher Scientific)

414   according to the manufacturer's recommendations, recombinant proteins were harvested at day 4

415   in the supernatant without any purification step and quantified by adding the substrate of the

416   nanoluciferase to serial dilutions of the supernatant. Luminescence was measured onto a Centro

417   XS^3 LB 960 luminometer (Berthold Technologies, France). LIPS assay was conducted as

418   previously described(*32*) except that 10 µL of sera diluted to ¼ in fetal calf serum (FCS, Fisher

419   scientific) were engaged per reaction. The signal-to-noise ratio of each serum was calculated by

420   dividing the RLU by the mean of 10 negative controls (with 10 µL of FCS) + five standard

421   deviations.

422   <u>LuLISA (Luciferase-Linked Immunoscorbent Assay)</u>

423   The purified full-length N and S protein stabilized in prefusion conformation of BANAL-236,

424   BANAL-52 and BANAL-103 as described previously for SARS-CoV-2 (*33*) was diluted in PBS

425   1X at 50 ng/µL. White plates (Thermo Scientific 437796 /470372) were coated with 50 µL/well

426   during 4 hours at room temperature. After 3 washes in PBS 1X 0,1%Tween, 50 µL of heat-

427   inactivated serums diluted 1/200 in PBS 1X, 0,1%Tween 1% milk were added/well, and incubated

428   for 1 hour at room temperature. After 3 washes in PBS 1X 0,1%Tween, anti-IgG nanobody fused

429   to the nanoluciferase was added diluted in PBS 1X, 0,1%Tween, 1% milk (dilution depending

430   upon the preparation) and incubated for 20 minutes at room temperature. After 3 washes in PBS

431   1X 0,1%Tween, bioluminescent detection of antibody levels was performed by using 50 µL of 12

432   µM of the nanoluciferase substrate Furimazine, and measuring RLU for 0,5sec/well in a

433   luminometer (Bertold centro XS).

434   <u>Mutation analysis</u>

435   The quantification of viral loads in genome copy per g of lung of each infected transgenic mouse

436   expressing hACE-2 was carried out to constitute seven pools of replicates, each of 3-4 samples at

437   equivalent concentration while individual macaque rectal samples were sequenced. A custom

438   panel (251 probes) based on the MZ9370003.2 BANAL-236 genome was designed by TWIST

439   Bioscience for NGS-based Target Capture. The double stranded cDNA (dscDNA) of each pool

440   were constructed using the Protoscript II First strand cDNA synthesis kit followed by NEBNext

441   Ultra II Non-Directional RNA Second Strand Synthesis Module. Purified dscDNA was then used

442   as input for the Illumina DNA Prep with Enrichment library preparation protocol. Briefly, each

443   purified dscDNA sample was tagmented and fragments were barcoded. The barcoded libraries

444   were pooled and hybridized with the TWIST custom panel for BANAL-236 following the

445   specifications for third-party panels of the Illumina DNA Prep with Enrichment protocol. The

446   captured libraries were sequenced on an Illumina MiSeq Micro format in a paired-end 300-cycle

447   run. For the infected VeroE6 and Caco-2 cells, the libraries were performed using the SMARTer

448   Stranded Total RNA-Seq Kit v3 - Pico Input Mammalian kit (TaKaRa Bio, Ozyme) was used

449   according to the manufacturer instructions and sequenced on an Illumina NextSeq 500 in a paired-

450   end 300-cycle run, except for the Caco-2_C4 sample (single-read sequencing).

451   Reads were trimmed with AlienTrimmer, then cropped using a custom python script in order to

452   clean the sequence ends. Cleaned reads were mapped on the BANAL-20-236 genome (GenBank:

453   MZ9370003.2) using CLC Genomics Workbench 20.0.4 with read Length fraction=1 and

454   Similarity fraction=0.985. After adding InDel qualities, variants were called using LoFreq v.2.1.4

455   and variant annotation was performed with SnpEff v4_5covid19_core. Variants were checked for

456   depth and strand bias. SNVs detected in at least three serial passages were first selected. Mutations

457   with low frequency and weak variation were discarded.

458   <u>RT-qPCR</u> <u>on</u> <u>mice</u> <u>tissues</u>

459   Organs were disrupted in 1 mL of DMEM with the TissueLyser LT homogenizer (Qiagen) and 2

460   mL-reinforced tubes containing beads (Bertin, France). Total RNAs were extracted from 100 µL

461   of clarified supernatant with the Maxwell® RSC simplyRNA Tissue kit (Promega) with DNAse

462   treatment, according to the manufacturer's instructions. Samples were tested with a one-step RT-

463   qPCR Taqman system using a set of primers and probe targeting the E gene, as previously

464   described(*34*) and [https://www.who.int/docs/default-source/coronaviruse/real-time-rt-pcr-assays-](https://www.who.int/docs/default-source/coronaviruse/real-time-rt-pcr-assays-for-the-detection-of-sars-cov-2-institut-pasteur-paris.pdf?sfvrsn=3662fcb6_2)

465   [for-the-detection-of-sars-cov-2-institut-pasteur-paris.pdf?sfvrsn=3662fcb6_2](https://www.who.int/docs/default-source/coronaviruse/real-time-rt-pcr-assays-for-the-detection-of-sars-cov-2-institut-pasteur-paris.pdf?sfvrsn=3662fcb6_2).

466   <u>Titration</u>

467   Approximately 10^6 Vero E6 cells were seeded to each well of a 6-well plate and cultured overnight

468   at 37°C, 5% CO_2. The next day, the cells were infected with 400 µL of 10-fold serial dilutions of

469   viral supernatants in DMEM without fetal bovine serum (FBS), supplemented with 1 µg/mL

470   TPCK-Trypsin and 1% penicillin-streptomycin. The viruses were incubated with the cells at 37 °C

471   with 5% CO_2 for 1 hour. An overlay of MEM-avicel supplemented with 1 µg/mL TPCK-Trypsin

472   was added to the cells. Three days later, the plates were stained with 30 % crystal violet, 20%

473   ethanol and 10% formaldehyde. Plaques were counted and viral titers in PFU/mL were calculated.

Molecular Dynamics simulations of BANAL-236 RBD/human ACE2 complexes

We performed 6 all-atom explicit-solvent molecular dynamics (MD) simulations of the human ACE2 (hACE2) in complex with the wild type BANAL-236 RBD (WT-hACE2) and V391I mutant (V391I-hACE2). All simulations were performed with GROMACS 2021.4(*35*) for a total aggregated time of 6 µs. The GROMACS topology and input files as well as the analysis scripts used are available on PLUMED-NEST (www.plumed-nest.org)(*36*) under accession ID plumID:22.027. This work was granted access to the HPC resources of IDRIS under the allocation 2022-102408 made by GENCI.

Details of the MD simulations: setup, equilibration, and production.

The X-ray structure of the wild type BANAL-236 RBD/hACE2 complex (PDB code 7PKI) was used as input to the CHARMM-GUI server(*37*). The zinc and chloride atoms present in 7PKI as well as the NAG and water residues were retained. The system was solvated in a triclinic box of initial x-y-z dimensions equal to 12.8 nm * 8.9 nm * 8.7 nm. Potassium and chloride ions were added to ensure charge neutrality at a salt concentration of 0.15 M. The total number of atoms were 96937. CHARMM-GUI was also used to construct a homology model of the V391I mutant starting from the X-ray structure of the wild type. Additional details of the systems are reported in Table S4. The CHARMM36m force field(*38*) was used for the protein and ions and the TIP3P model(*39*) was used for the water molecules. The CHARMM-GUI models were first energy-minimized using the steepest descent algorithm. After minimization, the wild type and mutant complexes were equilibrated using a 10-ns long simulation in the NPT ensemble followed by a 10-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(*40*) and the Berendsen barostat(*41*) respectively. During equilibration, harmonic restraints on the positions of the protein backbone

497   and sidechains heavy atoms were applied. From the NVT equilibration runs, 3 conformations of

498   the wild type and V391I mutant were extracted and used as starting point for production

499   simulations. The production simulations were performed in the NVT ensemble for 1 μs. A time

500   step of 2 fs was used together with LINCS constraints on h-bonds(*42*). The van der Waals

501   interactions were gradually switched off at 1.0 nm and cut off at 1.2 nm; the particle-mesh Ewald

502   method was used to calculate electrostatic interactions with cutoff at 1.2 nm(*43*).

503   *Details of the analysis.*

504   To evaluate the stability of the starting model during the production runs, we calculated the

505   backbone Root Mean Square Deviation (RMSD) with respect to the energy-minimized structure

506   for each frame of the trajectories. The RMSD was calculated separately for: RBD, hACE2, the

507   residues at the RBD-hACE2 interface, which were defined as the residues in one subunit closer

508   than 0.8 nm to the residues in the other subunit in the X-ray structure 7PKI, and the entire complex

509   (Fig. S8). RMSD calculations were performed using the *driver* utility of PLUMED v. 2.7(*36*). To

510   estimate the binding energy between RBD and hACE2, we used the *AnalyseComplex* tool in FoldX

511   v. 4(*44*) (Fig. S9). To identify relevant interactions at the RBD-hACE2 interface, we quantified

512   the frequency of formation of inter-subunits hydrogen bonds (Fig. S10) and salt bridges (Fig. S11)

513   during the course of the MD simulations. To monitor the formation of hydrogen bonds, we used

514   the Hydrogen Bond Analysis module of the MDAnalysis library v. 2.1.0(*45*). A donor–acceptor

515   distance and angular cutoff of 0.3 nm and 150°, respectively, were used to define the formation of

516   a hydrogen bond. For salt bridges, for each frame of the trajectories we used MDAnalysis to

517   compute the distances between the sidechain charged groups of aspartic acids (OD1/OD2),

518   glutamic acids (OE1/OE2), lysines (NZ), and arginines (NH1/NH2). A inter-subunits salt bridge

519   was defined as formed if the distance between groups with opposite charge was lower than 0.32

520   nm.

521   Statistical analyses

522   Statistical analysis was performed with Prism (GraphPad Software, LLC) and R software. Survival

523   curves were compared with a logrank test. Body weight loss curves were compared day-by-day

524   with a Student's t-test.

525   Illustrations

526   Illustrations for mice and primate infections were created with BioRender.com.

References and Notes

1. D. Delaune, V. Hul, E. A. Karlsson, A. Hassanin, T. P. Ou, A. Baidaliuk, F. Gámbaro, M. Prot, V. T. Tu, S. Chea, L. Keatts, J. Mazet, C. K. Johnson, P. Buchy, P. Dussart, T. Goldstein, E. Simon-Lorière, V. Duong, A novel SARS-CoV-2 related coronavirus in bats from Cambodia. *Nat Commun*. **12**, 6563 (2021).

2. H. Zhou, J. Ji, X. Chen, Y. Bi, J. Li, Q. Wang, T. Hu, H. Song, R. Zhao, Y. Chen, M. Cui, Y. Zhang, A. C. Hughes, E. C. Holmes, W. Shi, Identification of novel bat coronaviruses sheds light on the evolutionary origins of SARS-CoV-2 and related viruses. *Cell* (2021), doi:10.1016/j.cell.2021.06.008.

3. S. Wacharapluesadee, C. W. Tan, P. Maneeorn, P. Duengkae, F. Zhu, Y. Joyjinda, T. Kaewpom, W. N. Chia, W. Ampoot, B. L. Lim, K. Worachotsueptrakun, V. C.-W. Chen, N. Sirichan, C. Ruchisrisarod, A. Rodpan, K. Noradechanon, T. Phaichana, N. Jantarat, B. Thongnumchaima, C. Tu, G. Crameri, M. M. Stokes, T. Hemachudha, L.-F. Wang, Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. *Nat Commun*. **12**, 972 (2021).

4. S. Murakami, T. Kitamura, J. Suzuki, R. Sato, T. Aoi, M. Fujii, H. Matsugo, H. Kamiki, H. Ishida, A. Takenaka-Uema, M. Shimojima, T. Horimoto, Detection and Characterization of Bat Sarbecovirus Phylogenetically Related to SARS-CoV-2, Japan. *Emerg Infect Dis*. **26**, 3025–3029 (2020).

5. P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G.-F. Xiao, Z.-L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. **579**, 270–273 (2020).

6. M. C. Rahalkar, R. A. Bahulikar, Lethal Pneumonia Cases in Mojiang Miners (2012) and the Mineshaft Could Provide Important Clues to the Origin of SARS-CoV-2. *Front. Public Health*. **8** (2020), doi:10.3389/fpubh.2020.581569.

7. C. Pommerenke, U. Rand, C. C. Uphoff, S. Nagel, M. Zaborski, V. Hauer, M. Kaufmann, C. Meyer, S. A. Denkmann, P. Riese, K. Eschke, Y. Kim, Z. M. Safranko, I.-C. Kurolt, A. Markotic, L. Cicin-Sain, L. Steenpass, Identification of cell lines CL-14, CL-40 and CAL-51 as suitable models for SARS-CoV-2 infection studies. *PLoS One*. **16**, e0255622 (2021).

8. B. A. Johnson, et al, A. L. Bailey, B. Kalveram, K. G. Lokugamage, A. Muruato, J. Zou, X. Zhang, T. Juelich, J. K. Smith, L. Zhang, N. Bopp, C. Schindewolf, M. Vu, A. Vanderheiden, E. S. Winkler, D. Swetnam, J. A. Plante, P. Aguilar, K. S. Plante, V. Popov, B. Lee, S. C. Weaver, M. S. Suthar, A. L. Routh, P. Ren, Z. Ku, Z. An, K. Debbink, M. S. Diamond, P.-Y. Shi, A. N. Freiberg, V. D. Menachery, Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature*. **591**, 293–299 (2021).

9. M. Hoffmann, N. Krüger, S. Schulz, A. Cossmann, C. Rocha, A. Kempf, I. Nehlmeier, L. Graichen, A.-S. Moldenhauer, M. S. Winkler, M. Lier, A. Dopfer-Jablonka, H.-M. Jäck, G. M. N. Behrens, S. Pöhlmann, The Omicron variant is highly resistant against antibody-

567 mediated neutralization – implications for control of the COVID-19 pandemic. *Cell* (2021),
568 doi:10.1016/j.cell.2021.12.032.

569 10. E. Cameroni, J. E. Bowen, L. E. Rosen, C. Saliba, S. K. Zepeda, K. Culap, D. Pinto, L. A.
570   VanBlargan, A. De Marco, J. di Iulio, F. Zatta, H. Kaiser, J. Noack, N. Farhat, N.
571   Czudnochowski, C. Havenar-Daughton, K. R. Sprouse, J. R. Dillen, A. E. Powell, A. Chen,
572   C. Maher, L. Yin, D. Sun, L. Soriaga, J. Bassi, C. Silacci-Fregni, C. Gustafsson, N. M.
573   Franko, J. Logue, N. T. Iqbal, I. Mazzitelli, J. Geffner, R. Grifantini, H. Chu, A. Gori, A.
574   Riva, O. Giannini, A. Ceschi, P. Ferrari, P. E. Cippà, A. Franzetti-Pellanda, C. Garzoni, P. J.
575   Halfmann, Y. Kawaoka, C. Hebner, L. A. Purcell, L. Piccoli, M. S. Pizzuto, A. C. Walls, M.
576   S. Diamond, A. Telenti, H. W. Virgin, A. Lanzavecchia, G. Snell, D. Veesler, D. Corti,
577   Broadly neutralizing antibodies overcome SARS-CoV-2 Omicron antigenic shift. *Nature*, 1–
578   9 (2021).

579 11. P. J. Halfmann, S. Iida, K. Iwatsuki-Horimoto, T. Maemura, M. Kiso, S. M. Scheaffer, T. L.
580   Darling, A. Joshi, S. Loeber, G. Singh, S. L. Foster, B. Ying, J. B. Case, Z. Chong, B.
581   Whitener, J. Moliva, K. Floyd, M. Ujie, N. Nakajima, M. Ito, R. Wright, R. Uraki, P.
582   Warang, M. Gagne, R. Li, Y. Sakai-Tagawa, Y. Liu, D. Larson, J. E. Osorio, J. P.
583   Hernandez-Ortiz, A. R. Henry, K. Ciouderis, K. R. Florek, M. Patel, A. Odle, L.-Y. R.
584   Wong, A. C. Bateman, Z. Wang, V.-V. Edara, Z. Chong, J. Franks, T. Jeevan, T. Fabrizio, J.
585   DeBeauchamp, L. Kercher, P. Seiler, A. S. Gonzalez-Reiche, E. M. Sordillo, L. A. Chang,
586   H. van Bakel, V. Simon, D. C. Douek, N. J. Sullivan, L. B. Thackray, H. Ueki, S.
587   Yamayoshi, M. Imai, S. Perlman, R. J. Webby, R. A. Seder, M. S. Suthar, A. García-Sastre,
588   M. Schotsaert, T. Suzuki, A. C. M. Boon, M. S. Diamond, Y. Kawaoka, SARS-CoV-2
589   Omicron virus causes attenuated disease in mice and hamsters. *Nature*, 1–1 (2022).

590 12. P. Maisonnasse, J. Guedj, V. Contreras, S. Behillil, C. Solas, R. Marlin, T. Naninck, A.
591   Pizzorno, J. Lemaitre, A. Gonçalves, N. Kahlaoui, O. Terrier, R. H. T. Fang, V. Enouf, N.
592   Dereuddre-Bosquet, A. Brisebarre, F. Touret, C. Chapon, B. Hoen, B. Lina, M. R. Calatrava,
593   S. van der Werf, X. de Lamballerie, R. Le Grand, Hydroxychloroquine use against SARS-
594   CoV-2 infection in non-human primates. *Nature*. **585**, 584–587 (2020).

595 13. S. Temmam, K. Vongphayloth, E. B. Salazar, S. Munier, M. Bonomi, B. Regnault, B.
596   Douangboubpha, Y. Karami, D. Chrétien, D. Sanamxay, V. Xayaphet, P. Paphaphanh, V.
597   Lacoste, S. Somlor, K. Lakeomany, N. Phommavanh, P. Pérot, O. Dehan, F. Amara, F.
598   Donati, T. Bigot, M. Nilges, F. A. Rey, S. van der Werf, P. T. Brey, M. Eloit, Bat
599   coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature*, 1–10 (2022).

600 14. W. Yin, Y. Xu, P. Xu, X. Cao, C. Wu, C. Gu, X. He, X. Wang, S. Huang, Q. Yuan, K. Wu,
601   W. Hu, Z. Huang, J. Liu, Z. Wang, F. Jia, K. Xia, P. Liu, X. Wang, B. Song, J. Zheng, H.
602   Jiang, X. Cheng, Y. Jiang, S.-J. Deng, H. E. Xu, Structures of the Omicron spike trimer with
603   ACE2 and an anti-Omicron antibody. *Science*. **375**, 1048–1053 (2022).

604 15. S. Virachith, V. Pommelet, E. Calvez, V. Khounvisith, S. Sayasone, S. Kounnavong, M.
605   Mayxay, P. Xangsayarath, S. Temmam, M. Eloit, N. Escriou, T. Rose, K. Vongphayloth, J.
606   M. Hübschen, V. Lacoste, S. Somlor, D. Phonekeo, P. T. Brey, A. P. Black, Low
607   seroprevalence of COVID-19 in Lao PDR, late 2020. *The Lancet Regional Health - Western
608   Pacific*. **13**, 100197 (2021).

609   16. Y. Guan, B. J. Zheng, Y. Q. He, X. L. Liu, Z. X. Zhuang, C. L. Cheung, S. W. Luo, P. H. Li,
610   L. J. Zhang, Y. J. Guan, K. M. Butt, K. L. Wong, K. W. Chan, W. Lim, K. F. Shortridge, K.
611   Y. Yuen, J. S. M. Peiris, L. L. M. Poon, Isolation and characterization of viruses related to
612   the SARS coronavirus from animals in southern China. *Science*. **302**, 276–278 (2003).

613   17. M. A. Müller, B. Meyer, V. M. Corman, M. Al-Masri, A. Turkestani, D. Ritz, A. Sieberg, S.
614   Aldabbagh, B.-J. Bosch, E. Lattwein, R. F. Alhakeem, A. M. Assiri, A. M. Albarrak, A. M.
615   Al-Shangiti, J. A. Al-Tawfiq, P. Wikramaratna, A. A. Alrabeeah, C. Drosten, Z. A. Memish,
616   Presence of Middle East respiratory syndrome coronavirus antibodies in Saudi Arabia: a
617   nationwide, cross-sectional, serological study. *Lancet Infect Dis*. **15**, 559–564 (2015).

618   18. C. B. Reusken, B. L. Haagmans, M. A. Müller, C. Gutierrez, G.-J. Godeke, B. Meyer, D.
619   Muth, V. S. Raj, L. S.-D. Vries, V. M. Corman, J.-F. Drexler, S. L. Smits, Y. E. El Tahir, R.
620   De Sousa, J. van Beek, N. Nowotny, K. van Maanen, E. Hidalgo-Hermoso, B.-J. Bosch, P.
621   Rottier, A. Osterhaus, C. Gortázar-Schmidt, C. Drosten, M. P. Koopmans, Middle East
622   respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a
623   comparative serological study. *The Lancet Infectious Diseases*. **13**, 859–866 (2013).

624   19. W. K. Leung, K.-F. To, P. K. S. Chan, H. L. Y. Chan, A. K. L. Wu, N. Lee, K. Y. Yuen, J. J.
625   Y. Sung, Enteric involvement of severe acute respiratory syndrome-associated coronavirus
626   infection. *Gastroenterology*. **125**, 1011–1017 (2003).

627   20. J. S. M. Peiris, C. M. Chu, V. C. C. Cheng, K. S. Chan, I. F. N. Hung, L. L. M. Poon, K. I.
628   Law, B. S. F. Tang, T. Y. W. Hon, C. S. Chan, K. H. Chan, J. S. C. Ng, B. J. Zheng, W. L.
629   Ng, R. W. M. Lai, Y. Guan, K. Y. Yuen, HKU/UCH SARS Study Group, Clinical
630   progression and viral load in a community outbreak of coronavirus-associated SARS
631   pneumonia: a prospective study. *Lancet*. **361**, 1767–1772 (2003).

632   21. H. Li, E. Mendelsohn, C. Zong, W. Zhang, E. Hagan, N. Wang, S. Li, H. Yan, H. Huang, G.
633   Zhu, N. Ross, A. Chmura, P. Terry, M. Fielder, M. Miller, Z. Shi, P. Daszak, Human-animal
634   interactions and bat coronavirus spillover potential among rural residents in Southern China.
635   *Biosaf Health*. **1**, 84–90 (2019).

636   22. T. N. Starr, A. J. Greaney, S. K. Hilton, D. Ellis, K. H. D. Crawford, A. S. Dingens, M. J.
637   Navarro, J. E. Bowen, M. A. Tortorici, A. C. Walls, N. P. King, D. Veesler, J. D. Bloom,
638   Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints
639   on Folding and ACE2 Binding. *Cell*. **182**, 1295-1310.e20 (2020).

640   23. T. Horimoto, Y. Kawaoka, Molecular changes in virulent mutants arising from avirulent
641   avian influenza viruses during replication in 14-day-old embryonated eggs. *Virology*. **206**,
642   755–759 (1995).

643   24. A. T. Laleye, C. Abolnik, Emergence of highly pathogenic H5N2 and H7N1 influenza A
644   viruses from low pathogenic precursors by serial passage in ovo. *PLoS One*. **15**, e0240290
645   (2020).

646   25. A. H. Seekings, W. A. Howard, A. Nuñéz, M. J. Slomka, A. C. Banyard, D. Hicks, R. J.
647   Ellis, J. Nuñéz-García, L. C. Hartgroves, W. S. Barclay, J. Banks, I. H. Brown, The

648   Emergence of H7N7 Highly Pathogenic Avian Influenza Virus from Low Pathogenicity
649   Avian Influenza Virus Using an in ovo Embryo Culture Model. *Viruses*. **12**, E920 (2020).

650 26. M. L. Perdue, M. García, D. Senne, M. Fraire, Virulence-associated sequence duplication at
651   the hemagglutinin cleavage site of avian influenza viruses. *Virus Res*. **49**, 173–186 (1997).

652 27. N. L. Harrison, J. D. Sachs, A call for an independent inquiry into the origin of the SARS-
653   CoV-2 virus. *Proceedings of the National Academy of Sciences*. **119**, e2202769119 (2022).

654 28. K. Xaydalasouk, K. Sayasinh, J. M. Hübschen, V. Khounvisith, S. Keomany, C. P. Muller,
655   A. P. Black, Age-stratified seroprevalence of vaccine-preventable infectious disease in
656   Saravan, Southern Lao People's Democratic Republic. *Int J Infect Dis*. **107**, 25–30 (2021).

657 29. P. Maisonnasse, Y. Aldon, A. Marc, R. Marlin, N. Dereuddre-Bosquet, N. A. Kuzmina, A.
658   W. Freyn, J. L. Snitselaar, A. Gonçalves, T. G. Caniels, J. A. Burger, M. Poniman, I.
659   Bontjer, V. Chesnais, S. Diry, A. Iershov, A. J. Ronk, S. Jangra, R. Rathnasinghe, P. J. M.
660   Brouwer, T. P. L. Bijl, J. van Schooten, M. Brinkkemper, H. Liu, M. Yuan, C. E. Mire, M. J.
661   van Breemen, V. Contreras, T. Naninck, J. Lemaître, N. Kahlaoui, F. Relouzat, C. Chapon,
662   R. Ho Tsong Fang, C. McDanal, M. Osei-Twum, N. St-Amant, L. Gagnon, D. C. Montefiori,
663   I. A. Wilson, E. Ginoux, G. J. de Bree, A. García-Sastre, M. Schotsaert, L. Coughlan, A.
664   Bukreyev, S. van der Werf, J. Guedj, R. W. Sanders, M. J. van Gils, R. Le Grand, COVA1-
665   18 neutralizing antibody protects against SARS-CoV-2 in three preclinical models. *Nat
666   Commun*. **12**, 6097 (2021).

667 30. T. Naninck, N. Kahlaoui, J. Lemaitre, P. Maisonnasse, A. De Mori, Q. Pascal, V. Contreras,
668   R. Marlin, F. Relouzat, B. Delache, C. Hérate, Y. Aldon, M. van Gils, N. Zabaleta, R. Ho
669   Tsong Fang, N. Bosquet, R. W. Sanders, L. H. Vandenberghe, C. Chapon, R. Le Grand,
670   Computed tomography and [18F]-FDG PET imaging provide additional readouts for
671   COVID-19 pathogenesis and therapies evaluation in non-human primates. *iScience*. **25**,
672   104101 (2022).

673 31. J. Stevens, A. L. Corper, C. F. Basler, J. K. Taubenberger, P. Palese, I. A. Wilson, Structure
674   of the uncleaved human H1 hemagglutinin from the extinct 1918 influenza virus. *Science*.
675   **303**, 1866–1870 (2004).

676 32. L. Grzelak, S. Temmam, C. Planchais, C. Demeret, L. Tondeur, C. Huon, F. Guivel-
677   Benhassine, I. Staropoli, M. Chazal, J. Dufloo, D. Planas, J. Buchrieser, M. M. Rajah, R.
678   Robinot, F. Porrot, M. Albert, K.-Y. Chen, B. Crescenzo-Chaigne, F. Donati, F. Anna, P.
679   Souque, M. Gransagne, J. Bellalou, M. Nowakowski, M. Backovic, L. Bouadma, L. Le
680   Fevre, Q. Le Hingrat, D. Descamps, A. Pourbaix, C. Laouénan, J. Ghosn, Y. Yazdanpanah,
681   C. Besombes, N. Jolly, S. Pellerin-Fernandes, O. Cheny, M.-N. Ungeheuer, G. Mellon, P.
682   Morel, S. Rolland, F. A. Rey, S. Behillil, V. Enouf, A. Lemaitre, M.-A. Créach, S. Petres, N.
683   Escriou, P. Charneau, A. Fontanet, B. Hoen, T. Bruel, M. Eloit, H. Mouquet, O. Schwartz, S.
684   van der Werf, A comparison of four serological assays for detecting anti-SARS-CoV-2
685   antibodies in human serum samples from different populations. *Science Translational
686   Medicine*. **12** (2020), doi:10.1126/scitranslmed.abc3103.

687 33. T. Woudenberg, S. Pelleau, F. Anna, M. Attia, F. Donnadieu, A. Gravet, C. Lohmann, H.
688   Seraphin, R. Guiheneuf, C. Delamare, K. Stefic, J. Marlet, E. Brochot, S. Castelain, O.

Augereau, J. Sibilia, F. Dubos, D. Meddour, C. G.-L. Guen, M. Coste-Burel, B.-M. Imbert-Marcille, A. Chauvire-Drouard, C. Schweitzer, A. Gatin, S. Lomazzi, A. Joulié, H. Haas, A. Cantais, F. Bertholon, M.-F. Chinazzo-Vigouroux, M. S. Abdallah, L. Arowas, P. Charneau, B. Hoen, C. Demeret, S. V. D. Werf, A. Fontanet, M. White, Humoral immunity to SARS-CoV-2 and seasonal coronaviruses in children and adults in north-eastern France. *EBioMedicine*. **70**, 103495 (2021).

34. V. M. Corman, O. Landt, M. Kaiser, R. Molenkamp, A. Meijer, D. K. Chu, T. Bleicker, S. Brünink, J. Schneider, M. L. Schmidt, D. G. Mulders, B. L. Haagmans, B. van der Veer, S. van den Brink, L. Wijsman, G. Goderski, J.-L. Romette, J. Ellis, M. Zambon, M. Peiris, H. Goossens, C. Reusken, M. P. Koopmans, C. Drosten, Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill*. **25** (2020), doi:10.2807/1560-7917.ES.2020.25.3.2000045.

35. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. **1–2**, 19–25 (2015).

36. M. Bonomi, G. Bussi, C. Camilloni, G. A. Tribello, P. Banáš, A. Barducci, M. Bernetti, P. G. Bolhuis, S. Bottaro, D. Branduardi, R. Capelli, P. Carloni, M. Ceriotti, A. Cesari, H. Chen, W. Chen, F. Colizzi, S. De, M. De La Pierre, D. Donadio, V. Drobot, B. Ensing, A. L. Ferguson, M. Filizola, J. S. Fraser, H. Fu, P. Gasparotto, F. L. Gervasio, F. Giberti, A. Gil-Ley, T. Giorgino, G. T. Heller, G. M. Hocky, M. Iannuzzi, M. Invernizzi, K. E. Jelfs, A. Jussupow, E. Kirilin, A. Laio, V. Limongelli, K. Lindorff-Larsen, T. Löhr, F. Marinelli, L. Martin-Samos, M. Masetti, R. Meyer, A. Michaelides, C. Molteni, T. Morishita, M. Nava, C. Paissoni, E. Papaleo, M. Parrinello, J. Pfaendtner, P. Piaggi, G. Piccini, A. Pietropaolo, F. Pietrucci, S. Pipolo, D. Provasi, D. Quigley, P. Raiteri, S. Raniolo, J. Rydzewski, M. Salvalaglio, G. C. Sosso, V. Spiwok, J. Šponer, D. W. H. Swenson, P. Tiwary, O. Valsson, M. Vendruscolo, G. A. Voth, A. White, The PLUMED consortium, Promoting transparency and reproducibility in enhanced molecular simulations. *Nat Methods*. **16**, 670–673 (2019).

37. S. Jo, T. Kim, V. G. Iyer, W. Im, CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem*. **29**, 1859–1865 (2008).

38. J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, A. D. MacKerell, CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods*. **14**, 71–73 (2017).

39. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).

40. G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling. *J Chem Phys*. **126**, 014101 (2007).

41. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).

727   42. B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, LINCS: A linear constraint
728   solver for molecular simulations. *Journal of Computational Chemistry*. **18**, 1463–1472
729   (1997).

730   43. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, A smooth
731   particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).

732   44. J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, L. Serrano, The FoldX web
733   server: an online force field. *Nucleic Acids Res*. **33**, W382-388 (2005).

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

Author contributions

- Conceptualization: ST, XM, SW, ME
- Methodology: ST, XM, FR, SW, MB, ME
- Investigation: ST, XM, CH, FD, BR, MA, AB, KV, EBS, DC, LC, JPDF, TC, FA, FR, TN, JL, NDB, QP, MB, TB, SM
- Funding acquisition: PB, SW, ME

768   •   Project administration: ME

769   •   Supervision: PB, FR, RLG, SW, ME

770   •   Writing – original draft: ST, ME

771   •   Writing – review & editing: all

772

773 **Competing interests**

774 Authors declare that they have no competing interests.

775

776 **Data and materials availability**

777 All data are available in the main text or the supplementary materials.

778

779 **Supplementary Materials**

780   •   Materials and Methods

781   •   Figs. S1 to S11

782   •   Tables S1 to S5

783   •   References (28-45)

Fig. 1. BANAL-236 infection in mice expressing hACE2. **A.** Schematic representation of infection of K18 transgenic mice expressing human ACE2 (hACE2, black) or wild-type BALB/c mice (white) with BANAL-236 (red) or SARS-CoV-2 Wuhan-372 (blue) viruses. Sample size shown as 'x'. Lung, liver, spleen and brain were collected at day 3 (D3) post-infection. Blood was collected at D30 post-infection before challenge with SARS-CoV-2 Wuhan-D614G (yellow) virus. A second blood sample was collected at D14 post-challenge. Number of mice per group is shown as "x". **B.** Body weight variation of K18-hACE2 or BALB/c mice infected by BANAL-236 (red) or Wuhan-372 (blue) viruses at 10^4 PFU (solid lines) or 10^3 PFU (dotted lines) expressed as the mean (+/- SEM) body weight variation. Significant differences observed between K18-hACE2 mice infected at 10^4 PFU by BANAL-236 or Wuhan-372 at day 4 is noted by an asterisk. **C.** Quantification of BANAL-236 and Wuhan-372 in lung, spleen and liver of mice infected at 10^4 PFU. Viral loads are expressed as the mean (+/- SD) genome copies of genomic RNA per g organ (solid bars) or as PFUs per g organ (dotted bars). The limit of detection of the RT-qPCR is represented as a dashed line. **D.** Neutralizing antibody titer (expressed in ED_{50}) of mice primo-infected at 10^4 PFU by BANAL-236 (red) or SARS-CoV-2 Wuhan-372 (blue) viruses. The antibody response against the Wuhan-D614G challenge strain of mice primo-infected by BANAL-236 is presented as a dashed line. PNT means pseudo-neutralization. **E.** Survival curves of mice challenged with SARS-CoV-2 Wuhan-D614G at 10^4 PFU, after a primo-infection by BANAL-236 or Wuhan-372 viruses at 10^3 or 10^4 PFU. Survival of K18-hACE2 naive mice infected by SARS-CoV-2 Wuhan-D614G at 10^4 PFU is presented as a reference (yellow). **F.** Histopathological analysis, 3 days post-inoculation, of the lung of K18-hACE2 mice infected with 10^4 PFU of BANAL-236 (a-d) or SARS-CoV-2 Wuhan-372 (e,f) viruses. Interstitial pneumonia characterized by interstitial inflammation, often centered on blood vessels (perivasculitis, black arrows) or bronchi/bronchioles (black arrowheads), and endothelial cell injury and inflammation (endothelitis). Lesions were globally more severe after Wuhan-372 infection (n=4; all of moderate severity) than after BANAL-236 infection (n=4; 3/4 of minimal severity (a-c) and 1/4 of moderate severity (d)). Low magnification images are provided in Figure S3. **G.** Clinical and biological results of the 6 serial passages of BANAL-236 in K18-hACE2 mice as a pool of lung homogenates of the previous passage. The last passage (P6) was then inoculated into K18-hACE2 mice at 10^4 PFU. *Middle panel*: Viral load in the lung is expressed as the mean copies (+/- SEM) of genomic RNA per g of lung (solid line) or as PFUs per g of lung (dashed line). *Right panel*: Body weight variation of K18-hACE2 mice infected by BANAL-236 P1 (red, same results as those of Fig. 1B)

816   or P6 (green) viruses at 10^4 PFU. Significant differences observed between K18-hACE2 mice

817   infected by P1 or P6 are noted by an asterisk.

818   **Fig. 2. BANAL-236 infection in cynomolgus macaques.** **A.** Schematic representation of

819   infection of cynomolgus macaques by BANAL-236 virus. CT scan, clinical scoring and biological

820   samples were realized 10 and 3 days before infection, respectively, to serve as baseline (BL)

821   reference. The body temperature and weight were monitored at each anesthesia. Clinical scoring

822   was realized at BL and then from D2 to D11 post-infection and CT scan at BL and at D2 and D11

823   post-infection. Broncho-alveolar lavage (BAL) was realized at D3 and D11 post-infection and

824   biological sampling of nasopharyngeal, tracheal and rectal swabs at D2 to 28 post-infection to

825   monitor the virus load and antibody profile. Blood samples were collected 10 days before infection

826   to 43 days post-infection and were used for hematology, cytokine quantification and antibody

827   profiling. **B.** Effects of infection by BANAL-236 virus in non-human primates. Animal #MF1 is

828   presented in green and animal #MF2 in purple. Clinical score, body weight variation (in %), CT

829   scores (compared to Wuhan-372) and neutralizing antibodies are measured from 10 days before

830   infection to 43 days post-infection. **C.** Viral load expressed in $\log_{10}$ copies of genomic RNA (solid

831   line) or sub-genomic RNA (dashed line) per mL of tracheal, nasopharyngeal, BAL and rectal

832   fluids. Lower limit of quantification is presented in red for both genomic (solid line) and sub-

833   genomic (dashed line) RNA, and limits of detection are presented in green for both genomic (solid

834   line) and sub-genomic (dashed line) RNA. Results were compared to historical data of SARS-

835   CoV-2 Wuhan-372 virus inoculated at 10^6 or 10^5 PFU (presented as the mean virus load +/- 95%

836   CI expressed in $\log_{10}$ copies of genomic RNA/mL).

837   **Fig. 3. Mutations selected upon passages in K18-hACE2 mice and Caco-2 cells.** **A.** Schematic

838   representation of serial passages of BANAL-236 virus in Vero E6 cells (C1 to C3) followed by 6

839   passages in K18-hACE2 mice or by 6 passages in Caco-2 cells at a constant multiplicity of

840   infection (MOI) (C4 to C9). **B.** Mutation profiles of BANAL-236 after passages in VeroE6 cells,

841   Caco-2 cells, K18-hACE2 mice and in the feces of macaque during the course of infection.

842   Mutations reported in the heatmap were detected in at least three samples at a frequency above

843   10% in at least one sample. Rows are ordered according to the different passages and columns are

844   clustered by SNVs frequencies. **C.** Stacked histogram to compare the SNV frequencies of

845   BANAL-236 from the Caco-2 and mice serial passages along the MZ9370003.2 genome.

846   Mutations reported were detected in at least three serial samples. K18-hACE2 lung passages are

847 color coded in red in the upper panel and Caco-2 passages are color coded in green in the lower

848 panel. Each colored square height summarizes the allelic frequency (AF).

849 **Fig. 4. Antibody testing against bat sarbecoviruses in Lao human populations.** Results of

850 Pseudoneutralisation (PNT), Luciferase immunoprecipitation (LIPS) and LuLISA tests are shown.

851 Pre-pandemic French sera (black) were used as negative controls and Laotian sera samples from

852 confirmed SARS-CoV-2 infection (green) were used as cross-reacting positive controls. Laotian

853 sera samples collected in the general population before 2019 (blue, n=100) or late 2020 (gray,

854 n=100) or in people exposed to bats (purple, n=74) were tested for BANAL-236 (**A**), BANAL-52

855 (**B**) and BANAL-103 (**C**) antibody responses. The ANOVA non-parametric Kruskal-Wallis test

856 was conducted to compare each sub-population to the reference French population.
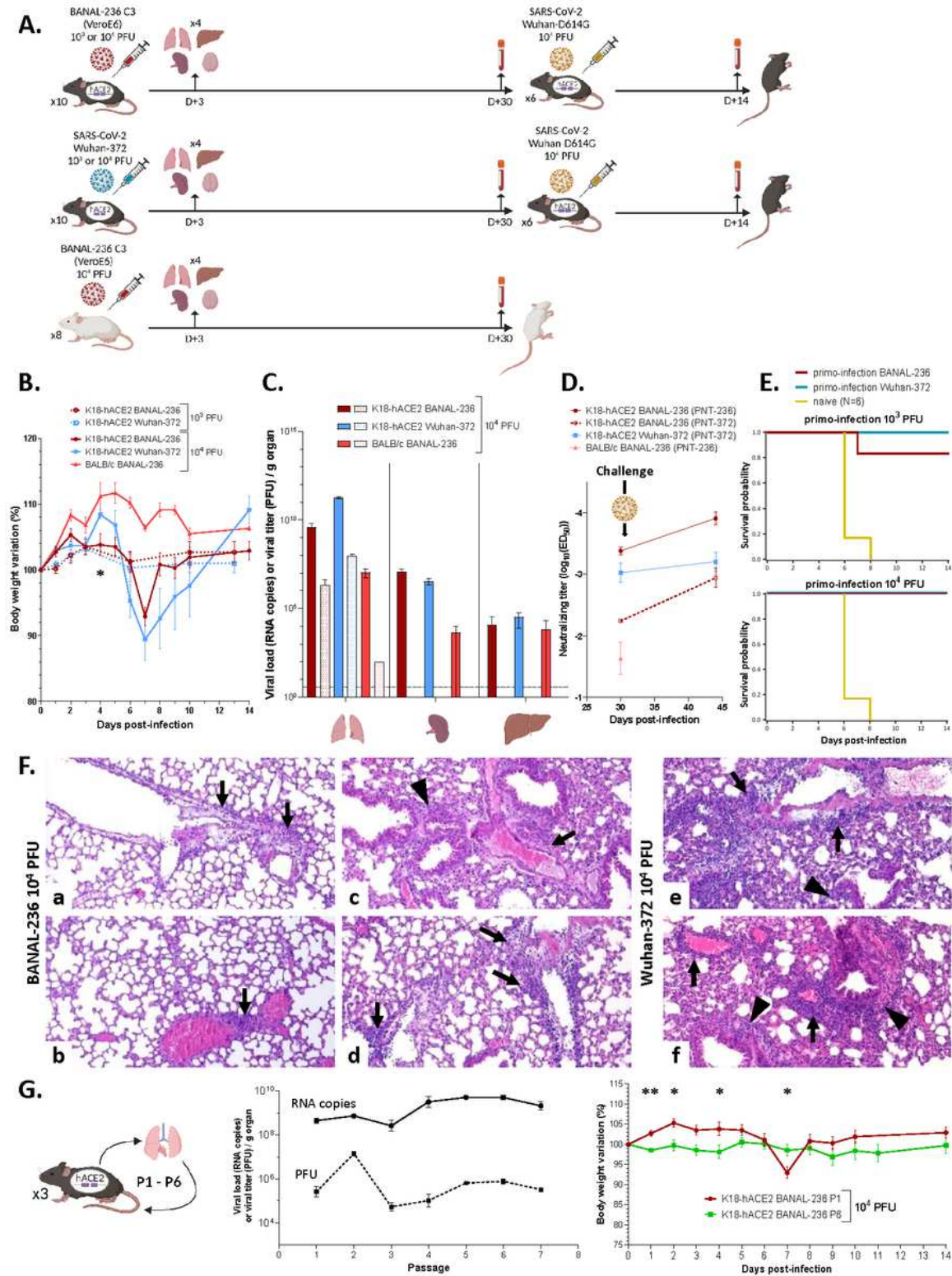
Figures

Figure 1

BANAL-236 infection in mice expressing hACE2. A. Schematic representation of infection of K18 transgenic mice expressing human ACE2 (hACE2, black) or wild-type BALB/c mice (white) with BANAL-236 (red) or SARS-CoV-2 Wuhan-372 (blue) viruses. Sample size shown as 'x'. Lung, liver, spleen and

brain were collected at day 3 (D3) post-infection. Blood was collected at D30 post-infection before challenge with SARS-CoV-2 Wuhan-D614G (yellow) virus. A second blood sample was collected at D14 post-challenge. Number of mice per group is shown as "x". **B.** Body weight variation of K18-hACE2 or BALB/c mice infected by BANAL-236 (red) or Wuhan-372 (blue) viruses at 104 PFU (solid lines) or 103 PFU (dotted lines) expressed as the mean (+/- SEM) body weight variation. Significant differences observed between K18-hACE2 mice infected at 104 PFU by BANAL-236 or Wuhan-372 at day 4 is noted by an asterisk. **C.** Quantification of BANAL-236 and Wuhan-372 in lung, spleen and liver of mice infected at 104 PFU. Viral loads are expressed as the mean (+/- SD) genome copies of genomic RNA per g organ (solid bars) or as PFUs per g organ (dotted bars). The limit of detection of the RT-qPCR is represented as a dashed line. **D.** Neutralizing antibody titer (expressed in ED50) of mice primo-infected at 104 PFU by BANAL-236 (red) or SARS-CoV-2 Wuhan-372 (blue) viruses. The antibody response against the Wuhan-D614G challenge strain of mice primo-infected by BANAL-236 is presented as a dashed line. PNT means pseudo-neutralization. **E.** Survival curves of mice challenged with SARS-CoV-2 Wuhan-D614G at 104 PFU, after a primo-infection by BANAL-236 or Wuhan-372 viruses at 103 or 104 PFU. Survival of K18-hACE2 naive mice infected by SARS-CoV-2 Wuhan-D614G at 104 PFU is presented as a reference (yellow). **F.** Histopathological analysis, 3 days post-inoculation, of the lung of K18-hACE2 mice infected with 104 PFU of BANAL-236 (a-d) or SARS-CoV-2 Wuhan-372 (e,f) viruses. Interstitial pneumonia characterized by interstitial inflammation, often centered on blood vessels (perivasculitis, black arrows) or bronchi/bronchioles (black arrowheads), and endothelial cell injury and inflammation (endothelitis). Lesions were globally more severe after Wuhan-372 infection (n=4; all of moderate severity) than after BANAL-236 infection (n=4; 3/4 of minimal severity (a-c) and 1/4 of moderate severity (d)). Low magnification images are provided in Figure S3. **G.** Clinical and biological results of the 6 serial passages of BANAL-236 in K18-hACE2 mice as a pool of lung homogenates of the previous passage. The last passage (P6) was then inoculated into K18-hACE2 mice at 104 PFU. *Middle panel*: Viral load in the lung is expressed as the mean copies (+/- SEM) of genomic RNA per g of lung (solid line) or as PFUs per g of lung (dashed line). *Right panel*: Body weight variation of K18-hACE2 mice infected by BANAL-236 P1 (red, same results as those of Fig. 1B) or P6 (green) viruses at 104 PFU. Significant differences observed between K18-hACE2 mice infected by P1 or P6 are noted by an asterisk.
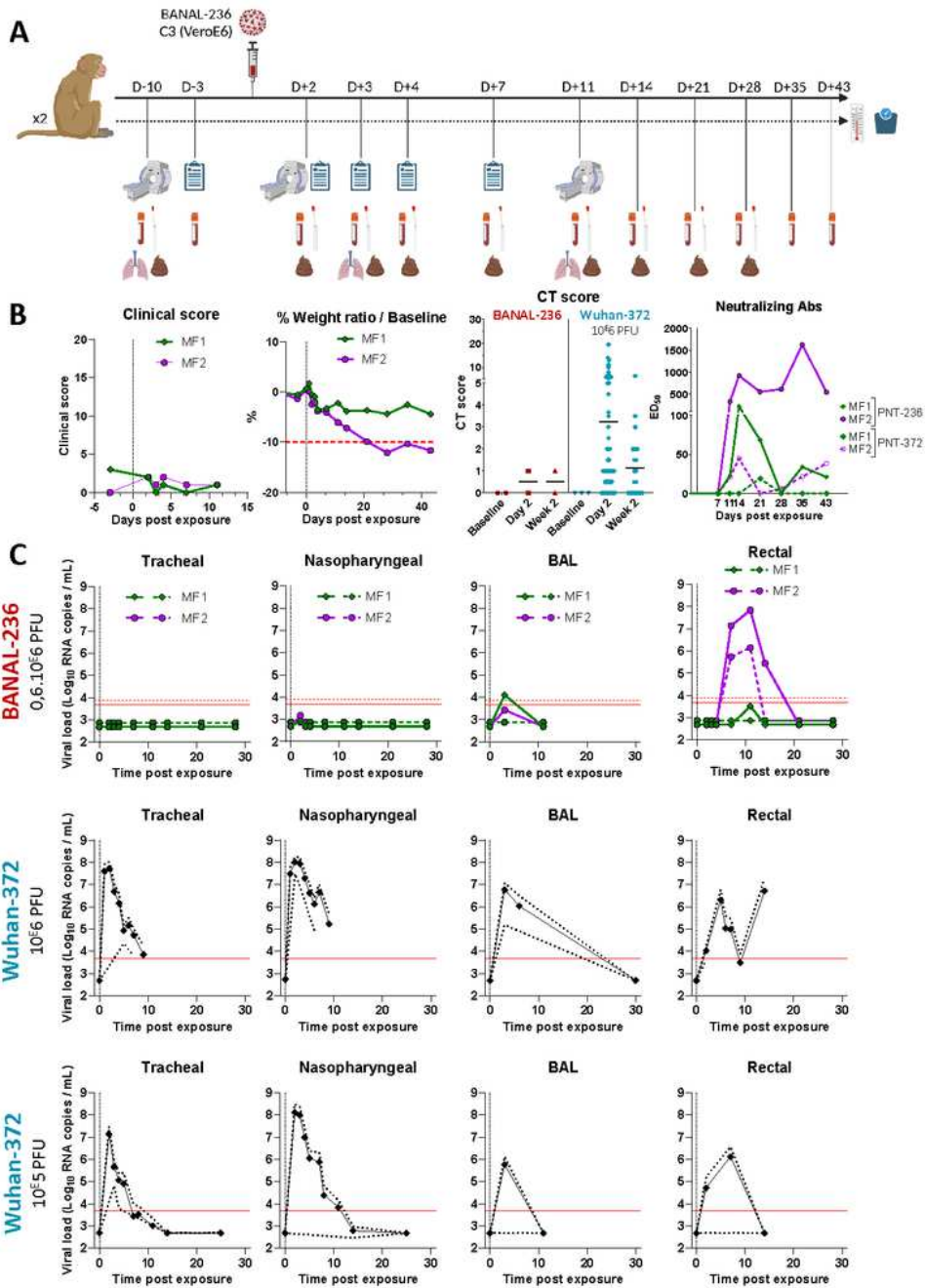
Figure 2

BANAL-236 infection in cynomolgus macaques. A. Schematic representation of infection of cynomolgus macaques by BANAL-236 virus. CT scan, clinical scoring and biological samples were realized 10 and 3 days before infection, respectively, to serve as baseline (BL) reference. The body temperature and weight were monitored at each anesthesia. Clinical scoring was realized at BL and then from D2 to D11 post-infection and CT scan at BL and at D2 and D11 post-infection. Broncho-alveolar lavage (BAL) was

realized at D3 and D11 post-infection and biological sampling of nasopharyngeal, tracheal and rectal swabs at D2 to 28 post-infection to monitor the virus load and antibody profile. Blood samples were collected 10 days before infection to 43 days post-infection and were used for hematology, cytokine quantification and antibody profiling. **B.** Effects of infection by BANAL-236 virus in non-human primates. Animal #MF1 is presented in green and animal #MF2 in purple. Clinical score, body weight variation (in %), CT scores (compared to Wuhan-372) and neutralizing antibodies are measured from 10 days before infection to 43 days post-infection. **C.** Viral load expressed in log10 copies of genomic RNA (solid line) or sub-genomic RNA (dashed line) per mL of tracheal, nasopharyngeal, BAL and rectal fluids. Lower limit of quantification is presented in red for both genomic (solid line) and sub-genomic (dashed line) RNA, and limits of detection are presented in green for both genomic (solid line) and sub-genomic (dashed line) RNA. Results were compared to historical data of SARS-CoV-2 Wuhan-372 virus inoculated at 106 or 105 PFU (presented as the mean virus load +/- 95% CI expressed in log10 copies of genomic RNA/mL).
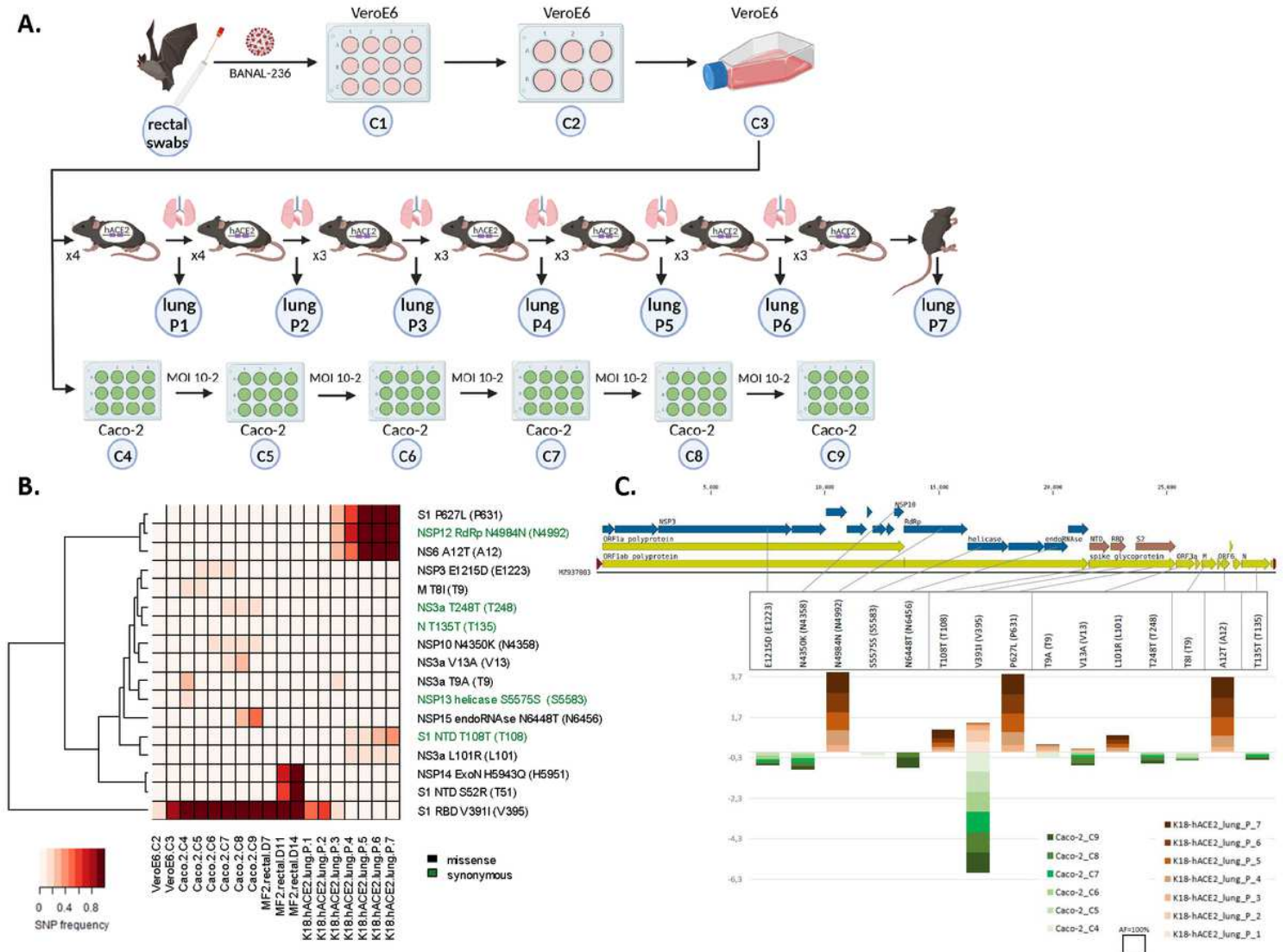

Figure 3

Mutations selected upon passages in K18-hACE2 mice and Caco-2 cells. A. Schematic representation of serial passages of BANAL-236 virus in Vero E6 cells (C1 to C3) followed by 6 passages in K18-hACE2 mice or by 6 passages in Caco-2 cells at a constant multiplicity of infection (MOI) (C4 to C9). **B.** Mutation profiles of BANAL-236 after passages in VeroE6 cells, Caco-2 cells, K18-hACE2 mice and in the feces of macaque during the course of infection. Mutations reported in the heatmap were detected in at least three samples at a frequency above 10% in at least one sample. Rows are ordered according to the different passages and columns are clustered by SNVs frequencies. **C.** Stacked histogram to compare the SNV frequencies of BANAL-236 from the Caco-2 and mice serial passages along the MZ9370003.2 genome. Mutations reported were detected in at least three serial samples. K18-hACE2 lung passages are color coded in red in the upper panel and Caco-2 passages are color coded in green in the lower panel. Each colored square height summarizes the allelic frequency (AF).

Figure 4

Antibody testing against bat sarbecoviruses in Lao human populations. Results of Pseudoneutralisation (PNT), Luciferase immunoprecipitation (LIPS) and LuLISA tests are shown. Pre-pandemic French sera (black) were used as negative controls and Laotian sera samples from confirmed SARS-CoV-2 infection (green) were used as cross-reacting positive controls. Laotian sera samples collected in the general population before 2019 (blue, n=100) or late 2020 (gray, n=100) or in people exposed to bats (purple,

n=74) were tested for BANAL-236 (**A**), BANAL-52 (**B**) and BANAL-103 (**C**) antibody responses. The ANOVA non-parametric Kruskal-Wallis test was conducted to compare each sub-population to the reference French population.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Suppdatav2.pdf