

Coronaviruses with a SARS-CoV-2-like receptor-binding domain allowing ACE2-mediated entry into human cells isolated from bats of Indochinese peninsula

Sarah Temmam

Institut Pasteur

Khamsing Vongphayloth

Institut Pasteur

Eduard Baquero Salazar

Institut Pasteur

Sandie Munier

Institut Pasteur

Max Bonomi

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

Béatrice Régnault

Institut Pasteur

Bounsavane Douangboubpha

University Lao

Yasaman Karami

Institut Pasteur <https://orcid.org/0000-0001-8413-2665>

Delphine Chretien

Institut Pasteur

Daosavanh Sanamxay

University Lao

Vilakhan Xayaphet

University Lao

Phetphoumin Paphaphanh

University Lao

Vincent Lacoste

Institut Pasteur Lao <https://orcid.org/0000-0002-3173-4053>

Somphavanh Somlor

Institut Pasteur Lao

Khaithong Lakeomany

Institut Pasteur Lao

Nothasin Phommavanh

Institut Pasteur Lao

Philippe Pérot

institut pasteur <https://orcid.org/0000-0002-5194-8200>

Flora Donati

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

Thomas Bigot

institut pasteur

Michael Nilges

Institut Pasteur, Paris <https://orcid.org/0000-0002-1451-8092>

Félix Rey

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

Sylvie van der Werf

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

Paul Brey

Institut Pasteur Lao

Marc Eloit (✉ marc.eloit@pasteur.fr)

Institut Pasteur

Biological Sciences - Article

Keywords: Coronaviruses, SARS-CoV-2, spike sequence, angiotensin-converting enzyme 2 pathway, receptor-binding domain, cave bats, Indochinese peninsula

Posted Date: September 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-871965/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The animal reservoir of SARS-CoV-2 is unknown despite reports of various SARS-CoV-2-related viruses in Asian *Rhinolophus* bats, including the closest virus from *R. affinis*, RaTG13. Several studies have suggested the involvement of pangolin coronaviruses in SARS-CoV-2 emergence. SARS-CoV-2 presents a mosaic genome, to which different progenitors contribute. The spike sequence determines the binding affinity and accessibility of its receptor-binding domain (RBD) to the cellular angiotensin-converting enzyme 2 (ACE2) receptor and is responsible for host range. SARS-CoV-2 progenitor bat viruses genetically close to SARS-CoV-2 and able to enter human cells through a human ACE2 pathway have not yet been identified, though they would be key in understanding the origin of the epidemics. Here we show that such viruses indeed circulate in cave bats living in the limestone karstic terrain in North Laos, within the Indochinese peninsula. We found that the RBDs of these viruses differ from that of SARS-CoV-2 by only one or two residues, bind as efficiently to the hACE2 protein as the SARS-CoV-2 Wuhan strain isolated in early human cases, and mediate hACE2-dependent entry into human cells, which is inhibited by antibodies neutralizing SARS-CoV-2. None of these bat viruses harbors a furin cleavage site in the spike. Our findings therefore indicate that bat-borne SARS-CoV-2-like viruses potentially infectious for humans circulate in *Rhinolophus* spp. in the Indochinese peninsula.

Introduction

The origin of SARS-CoV-2, as its mode of introduction into the human population, is currently unknown. Since its emergence, numerous animal species have been studied to identify possible reservoirs and/or intermediate hosts of the virus, including a large diversity of insectivorous bats of the genus *Rhinolophus*. Despite the recent report of various SARS-CoV-2-related viruses in *Rhinolophus shameli* (isolated in Cambodia in 2010¹), *R. pusillus* and *R. malayanus* (China, 2020 and 2019 respectively²), in *R. acuminatus* (Thailand, 2020³) and *R. cornutus* (Japan, 2013⁴), the closest SARS-CoV-2 bat-borne genome still remains the one from *R. affinis*, RaTG13 (China, 2013)^{5,6}. Several studies also suggested the role of pangolin-related coronaviruses in the emergence of SARS-CoV-2⁷⁻⁹. Since its appearance in humans, SARS-CoV-2 has evolved only through sporadic mutations, some of which correspond to gains in fitness allowing the virus to spread more widely, or to escape neutralizing antibodies¹³. To decipher the origin of SARS-CoV-2, it is therefore essential to ascertain the diversity of animal coronaviruses and more specifically of bat coronaviruses. Although the identification of SARS-CoV-2 in bats is a major goal, it may be unattainable.

A more realistic objective is to identify the sequences that contribute to its mosaicism. Among these, the spike sequence appears essential, as it determines the binding affinity and accessibility of the receptor-binding domain (RBD) to the cellular angiotensin-converting enzyme 2 (ACE2) receptor, and is therefore responsible for host range¹⁰⁻¹². The closest related bat strain identified to date (RaTG13) has a low RBD sequence similarity to SARS-CoV-2 and, with only 11/17 human ACE2 (hACE2) contact amino-acid residues conserved with SARS-CoV-2, its affinity to hACE2 is very limited¹⁴. Moreover, SARS-CoV-2 poorly

infects bats and bat cells tested so far¹⁵. In addition, no SARS-CoV-2-like virus has been shown to use hACE2 to enter human cells, and none presents the furin cleavage site that is associated with an increased pathogenicity in humans¹⁶. SARS-CoV-2 RBD binds to *Rhinolophus macrotis* ACE2 with a lower affinity than to hACE2¹⁷. An essential piece of information, i.e. finding bat viruses with an RBD motif close to that of SARS-CoV-2 and capable of binding to hACE2 with high affinity, is therefore missing.

We hypothesized that this type of virus could be identified in bats living in the limestone karstic terrain common to China, Laos, and Vietnam in the Indochinese peninsula. We report here the presence of sarbecoviruses close to SARS-CoV-2 whose RBDs differ from that of SARS-CoV-2 by only one or two residues, that strongly bind to the hACE2 protein, and mediate a hACE2-dependent entry in human cells. Despite the absence of the furin cleavage site, these viruses may have contributed to SARS-CoV-2's origin and may intrinsically pose a future risk of direct transmission to humans.

Main

Diversity of bats and coronaviruses in caves

A total of 645 bats belonging to six families and 46 species were captured (Table S1). Two hundred and forty-seven blood samples, 608 saliva, 539 anal/feces, and 157 urine swabs were collected from the northern part of Laos (Table S2). We first attempted to amplify part of the RNA-dependent RNA polymerase gene from all feces samples (n = 539) through a pan-coronavirus RT nested PCR approach¹⁸. Amplification products were obtained from 24 individuals of 10 species, and one individual (BANAL-27) was concomitantly infected by an alphacoronavirus and a betacoronavirus (Table S3). BLAST analysis of obtained sequences identified alphacoronavirus sequences of the *Decacovirus*, *Pedacovirus*, and *Rhinacovirus* subgenera and betacoronavirus sequences of the *Nobecovirus* and *Sarbecovirus* subgenera. Sequences of the *Sarbecovirus* subgenus were all identified from *Rhinolophus* individuals belonging to three different species i.e., *R. malayanus*, *R. marshalli*, and *R. pusillus*. Positive individuals were trapped in three different districts, and those infected with a sarbecovirus were all from the Fueng district in Vientiane Province (Fig. 1A, site 1).

The complete genome sequence of five of the seven sarbecoviruses identified was then obtained by next-generation sequencing (Fig. 1). The coverage of the genome of the remaining two sarbecoviruses, i.e. BANAL-27 and BANAL-242 sampled from *R. pusillus* and *R. malayanus* bats respectively, was 90% and therefore they were not included in the final analyses. Phylogenetic analyses performed on the complete genome sequence of lineages A and B human SARS-CoV-2¹⁹, and on representative bat and pangolin sarbecoviruses, placed the Laotian *R. malayanus* BANAL-52, *R. pusillus* BANAL-103, and *R. marshalli* BANAL-236 coronaviruses close to human SARS-CoV-2 and *R. affinis* RaTG13 coronaviruses, while *R. malayanus* BANAL-116 and BANAL-247 coronaviruses belonged to a sister clade with other bat coronaviruses (RmYN02, RacCS203, RpYN06, and PrC31) from different *Rhinolophus* species. Pangolin coronaviruses possessed a basal position relative to these strains (Fig. 1B). Interestingly, one should note that very similar SARS-CoV-2-like viruses are shared by different bat species, suggesting a possible

circulation of viruses between different species living sympatrically in the same caves. These results are consistent with the similarity plot analysis showing that RaTG13 and BANAL-52 bat coronaviruses exhibit a high nucleotide identity with human SARS-CoV-2 throughout the length of the genome. Interestingly, BANAL-52 presents a higher level of nucleotide conservation than RaTG13 in the S1 domain of the spike, and especially in the spike's N-terminal domain (NTD) and RBD (Fig. 1C). These observations are congruent with amino-acid identities between human SARS-CoV-2 and representative bat and pangolin coronaviruses which present a high level of conservation, except for ORF8 of bat BANAL-116, BANAL-247, Rc-o319 and RmYN02. Interestingly, the S1 domain of the spike (and especially the N-terminal domain) presents a lower degree of conservation in several bat coronaviruses, suggesting that this domain may reflect a relative degree of adaptation of the virus to its mammalian host (Fig. 1D, Supp. Figure 1).

Recombination events and phylogenetic analyses of sarbecoviruses

Following GARD analysis, we identified 14 recombinant breakpoints during the evolutionary history of sarbecoviruses, which were further confirmed by phylogenetic analyses performed on the 15 consecutive fragments of sequences defined by the breakpoints (Fig. 2, Supp. Figure 2). SARS-CoV-2 presents a mosaic genome, to which more than five sequences close to sequences published or determined during this study contributed: *R. malayanus* RmYN02 and *R. pusillus* RpYN06 viruses found in China in 2019, *R. affinis* RaTG13 coronavirus found in China in 2013, and *R. malayanus* BANAL-52 and *R. pusillus* BANAL-103 found in North Laos in 2020 (this study). No pangolin sequence was immediately associated with a recombination event at the origin of SARS-CoV-2, despite previous hypotheses proposed before the availability of the sequences described in this paper. Laotian *Rhinolophus* bat coronaviruses presented a lower degree of recombination compared to SARS-CoV-2 and when present, recombination events occurred between other BANAL viruses, in line with the fact that all viruses have been isolated from bats living sympatrically in caves in the same area.

Interestingly, the origin of several fragments of SARS-CoV-2 genomes could be assigned to several donor strains and not a unique donor sequence. For example, a breakpoint was identified at the beginning of the RBD region of S1 which revealed that the downstream fragment of SARS-CoV-2, which comprises the RBD and the beginning of S2 domains, could involve BANAL-52 *R. malayanus*, BANAL-103 *R. pusillus*, and BANAL-236 *R. marshalli* viruses, which formed a highly supported sister clade of SARS-CoV-2 (fragment 11, Fig. 2). In a more basal position, *R. shameli* bat coronaviruses and pangolin-2019 coronaviruses are found. These results are consistent with the conservation of RBD amino-acid sequences among SARS-CoV-2 and representative bat and pangolin coronaviruses (Supp. Figure3). Among the 17 residues that interact with human ACE2, 16 are conserved between SARS-CoV-2 and BANAL-52 or -103 (one mismatch, H498Q), and 15/17 are conserved for BANAL-236 (two mismatches, K493Q and H498Q) while only 13/17 residues are conserved for the Cambodian bat *R. shameli* virus and 11/17 for the Chinese bat *R. affinis* RaTG13 virus. At the full spike protein level, bat *R. affinis* RaTG13 and pangolin-2017 P4L viruses looked closer to SARS-CoV-2 than bat *R. malayanus* BANAL-52 but this was shaped by the higher degree of conservation of the S2 domain of the spike. All these viruses shared the

same features, such as the absence of a furin cleavage site or the conservation of the internal fusion peptide (Supp. Figure 4).

Interaction of RBDs with ACE2 and functional dynamics

To characterize the interaction of the RBDs of the BANAL-52/103, which are identical for the receptor-binding motif residues (Supp. Figure 3), and BANAL-236 spikes (residues 233-524) with hACE2, we performed biolayer interferometry assays and found that both RBDs display a binding affinity for hACE2 in the low nanomolar range (Fig. 3A), with values comparable to those previously reported for SARS-CoV-2^{20–23}.

To study the effect of the mutations at the interface between these RBDs and hACE2 on their binding affinity, we constructed homology models of the BANAL-236 and BANAL-52/103 RBD/hACE2 complexes using the crystal structure of the SARS-CoV-2 RBD/hACE2 complex as template (PDB id 6M0J). Of note, the amino-acid sequence in the RBD region covered by these models is identical in BANAL-103 and BANAL-52 (Supp. Figure 3). We performed all-atom explicit solvent Molecular Dynamics (MD) simulations of these complexes for a total aggregated time of 9 μ s (Table S4). A cluster analysis of the MD trajectories revealed that, at the RBD-hACE2 interface, both BANAL-236 and BANAL-52/103 complexes were identical to the SARS-CoV-2 RBD/hACE2 complex within 2 Å backbone RMSD (Supp. Figure 5). ROSETTA and FoldX empirical scoring functions predicted a similar RBD-hACE2 binding energy in all three complexes (Supp. Figure 6).

The analysis of the persistence of hydrogen bonds and salt bridges during the course of the MD simulations provided further insights into the effect of the substitutions at the RBD-hACE2 interface (Fig. 3B and Supp. Figure 7). The H498Q mismatch present in both BANAL-52/103 and BANAL-236 RBDs negatively affected the formation of hydrogen bonds between RBD Q498 and both hACE2 K353 and Q42. However, in the SARS-CoV-2 complex, these hydrogen bonds were only transiently formed on a microsecond timescale. More persistent hydrogen bonds in this region (RBD T500 – hACE2 D355, RBD G502 – hACE2 K353, and RBD Y505 – hACE2 E37) were not affected. The K493Q mismatch present only in BANAL-236 RBD enabled the formation of two salt bridges between the RBD and hACE2 that were not present in the SARS-CoV-2 complex (RBD K493 – hACE2 E35 and RBD K493 – hACE2 D38, the former being the most persistent).

For further insight into the molecular details of these interactions, we determined the crystal structure of the complex BANAL-236 RBD with the hACE2 peptidase domain to 2.9 Å resolution. The overall structure of the BANAL-236 RBD is identical to that of the SARS-CoV-2 RBD within the accuracy expected at this resolution (RMSD 0.360 Å, 150 Ca). The only significant difference is the unwinding of helix H4 that participates in lateral contacts between RBDs in the SARS-CoV-2 spike (Fig. 3C, arrow) which might influence the dynamics of the opening and closing of the RBDs in the spike.

As expected, most of the interactions observed in the SARS-CoV-2 RBD/hACE2 complex²⁴ are also present in the structure of the BANAL-236 RBD/hACE2 complex. In these interfaces, there are three main

clusters of interactions in which the salt bridge RBD K493 – hACE2 E35 and the hydrogen bond between the side chains of residues RBD H498 and D38 are located in clusters 1 and 2, respectively, as predicted by the MD simulations (Fig. 3B and C, insets). Although these interactions contribute to stabilizing the complex, they do not seem to affect drastically the binding to hACE2 because both RBDs have affinities similar to hACE2 and are in the same range of affinity of SARS-CoV-2 RBD.

Virus isolation and entry into human cells expressing ACE2

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

In order to isolate infectious virus, rectal swabs were inoculated on VeroE6 cells. Cytopathic effect (CPE) and presence of viral RNA were monitored daily. No CPE was observed 3 and 4 days after infection, but viral RNAs were detected for one of the 2 wells inoculated with the BANAL-236 sample ($C_T = 25.1$ at D3, $C_T = 21.7$ at D4). The culture supernatant (C1) formed plaques on VeroE6 and the titer was 3800 pfu/mL. A C2 viral stock was prepared by amplification on VeroE6 at a MOI of 10^{-4} . The culture supernatant was harvested at day 4 when CPE was observed and titrated on VeroE6 (Supp. Figure 8). The plaques' phenotype was small, but the titer reached $2.6 \cdot 10^6$ pfu/mL. The random NGS performed on the RNA extracted from this stock confirmed that the culture was pure and corresponded to the BANAL-236 virus. No non-synonymous mutations were observed between the original BANAL-236 genome and the C2 viral stock.

Discussion

Many sarbecoviruses circulate in *Rhinolophus* colonies living in caves in China and probably also in neighboring countries further south (Laos, Myanmar, Thailand, and Vietnam)^{25–27}. During the course of a prospective study in Northern Laos, we have identified in 10 bat species 25 different coronaviruses belonging to the *Alphacoronavirus* and *Betacoronavirus* genera. We then focused our study on the five sarbecoviruses for which we obtained full-length sequences. Among these, three (BANAL-52, -103, and -236) were considered to be close to SARS-CoV-2 because of the similarity in amino acids of one of the S domains (S1-NTD, S1-RBD, or S2) with the homologous domain of SARS-CoV-2. The similarity plot analysis revealed that the evolution history of SARS-CoV-2 is more complex than expected and that *R. affinis* RaTG13, isolated in Yunnan in 2013, is no longer considered the proximal ancestor of SARS-CoV-2: strains close to *R. pusillus* RpYN06, *R. malayanus* RmYN02, and *Rhinolophus* sp. PrC31 isolated in China in 2018-2019, along with *R. malayanus* BANAL-52, *R. pusillus* BANAL-103, and *R. marshalli* BANAL-236 isolated in Laos in 2020, contributed to the appearance of SARS-CoV-2 in different regions of the

genome. No closer viral genome has yet been identified as a possible contributor, and pangolin coronaviruses appear as more distantly related than bat coronaviruses.

Because genomic regions subject to recombination are likely contributing to host-virus interactions and adaptation following spillover events, we compared SARS-CoV-2 strains from the two lineages identified at the very onset of the COVID-19 outbreak¹⁹ to these novel bat sarbecoviruses and to pangolin strains within the SARS-CoV-2 clade. We thus identified potential recombination sites, allowing for the reconstruction of the phylogenetic history of early isolated SARS-CoV-2 strains between homologous regions defined by recombination points. The interaction of the SARS-CoV-2 spike with hACE2 is a key event in cell infection. The spike is divided into two subunits, S1 and S2: S1 contains an RBD that specifically binds to hACE2, whereas the S2 subunit contains the fusion peptide. Regarding the spike, we identified a breakpoint at the beginning of the SARS-CoV-2 RBD, resulting in a downstream fragment composed of the RBD, the furin cleavage site, and ending in the N-terminal region of S2. Despite the absence of the furin site in these novel bat sarbecoviruses, phylogenetic reconstruction of this fragment, key for the virus tropism and host spectrum, revealed that Laotian *R. malayanus* BANAL-52, *R. pusillus* BANAL-103, and *R. marshalli* BANAL-236 coronaviruses are the closest ancestors of SARS-CoV-2 known to date. Identification of strains of animal origin with a furin cleavage site may require additional sampling. As seen by others, ORF8 was highly divergent between SARS-CoV-2 related genomes. ORF8 from strains BANAL-52, -103, -236, like that of RaTG13, were closer to SARS-CoV-2 than to pangolin strains. ORF8 encodes a protein that has been proposed to participate in immune evasion²⁸. It is noteworthy that ORF8 was deleted in many human SARS-CoV-2 strains that appeared after March 2020²⁹, which is reminiscent of the deletions identified during the 2003 SARS epidemic³⁰. Therefore, ORF8 can be a marker of SARS-CoV-2 adaptation to humans and its presence in bat strains is consistent with bats acting as a natural reservoir of early strains of SARS-CoV-2.

Structural and functional biology studies have identified the RBD domain that mediates the interaction with hACE2, as well as the major amino acids that are involved²⁴. The host range is dependent on this interaction^{31,32}. We show that the spike of SARS-CoV-2 is a mosaic of sequences close to the following bat viruses: S1-NTD (BANAL-52 and RaTG13), S1-RBD (BANAL-52, -103, and -236), and S2 (BANAL-52, -236, -103, and RaTG13). Notably, the RBDs (BANAL-52, -103, and -236) are closer to SARS-CoV-2 than that of any other bat strain described so far. Overall, one (H498Q (strains BANAL-103 and -52)) or two (K493Q and H498Q (strain BANAL-236)) amino acids interacting with hACE2 are substituted in these strains in comparison to SARS-CoV-2. These mutations did not destabilize the BANAL-236 / hACE2 interface, as shown by the BLI experiments (Fig. 3A).

Our results contribute to understanding the origin of SARS-CoV-2: they show that sequences very close to those of the early strains of SARS-CoV-2 responsible for the pandemic exist in nature and are found in several *Rhinolophus* bat species. The RBDs of the viruses found in our study are closer to that of SARS-CoV-2 than to the RaTG13 RBD, the virus identified in *R. affinis* from the Mojiang mineshaft where pneumonia cases with clinical characteristics strikingly similar to COVID-19⁶ were recorded in 2012^{33,34}.

We found here sarbecoviruses with RBDs closest to that of SARS-CoV-2 in three different bat species, *R. marshalli*, *R. malayanus*, and *R. pusillus*. Our results therefore support the hypothesis that SARS-CoV-2 could originally result from a recombination of sequences pre-existing in *Rhinolophus* bats living in the extensive limestone cave systems of South-East Asia and South China³⁵, which provides ideal conditions for interspecies interactions among *Rhinolophus* bats. They are restricted to limestone caves for their roosting sites and forage in the vicinity of these caves, and many species have been found foraging in the same cave areas, including *R. malayanus* and *R. pusillus*³⁶. In addition, the distribution of *R. marshalli*, *R. malayanus*, and *R. pusillus* overlaps in the Indo-Chinese subregion, which means they may share caves as roost sites and foraging habitats³⁷.

The pangolin has been suspected as being an intermediate host of SARS-CoV-2. The pangolin CoV-2017 and -2019 genomes have a high overall protein identity with SARS-CoV-2 and RaTG13 (up to 100% for certain proteins). In particular, RBD was highly conserved between pangolin-CoV-2019 and SARS-CoV-2. On the other hand, the amino-acid sequence identity of S1-NTD is only 63.1% identical between pangolin-CoV-2017 and SARS-CoV-2⁷. With the novel viruses here described, the pool of sequences found in *Rhinolophus* spp. allows the reconstitution of a genome sufficiently close to that of SARS-CoV-2 without the need to hypothesize recombination or natural selection for increased RBD affinity for hACE2 in an intermediate host before spillover³⁸, nor natural selection in humans following spillover³⁹. However, we found no furin site in any of these viruses on sequences determined from original fecal swab samples, devoid of any bias associated with counterselection of the furin site by amplification in Vero cells¹⁶. Lack of furin cleavage may be explained by insufficient sampling in bats, or by acquisition of the furin cleavage site through passages of the virus in an alternate host or during an early poorly symptomatic unreported circulation in humans. Finally, where these intergenomic recombinations arose and the epidemiological link with the first human cases remains to be established.

As expected from the high affinity for ACE2 of the S ectodomain of the BANAL-236, pseudoviruses expressing it were able to enter efficiently human cells expressing hACE2 using an ACE2-dependent pathway. Entry was blocked by a serum neutralizing SARS-CoV-2. The RaTG13 strain, the closest to SARS-CoV-2 known before, had never been isolated. In contrast, preliminary studies show that BANAL-236 replicated in primate VeroE6 cells with a small plaque phenotype compared to SARS-CoV-2. Further analysis may indicate more clearly whether post-entry steps also shape infectivity.

To conclude, our results pinpoint the presence of new bat sarbecoviruses that seem to have the same potential for infecting humans as early strains of SARS-CoV-2. People working in caves, such as guano collectors, or certain ascetic religious communities who spend time in or very close to caves, as well as tourists who visit the caves, are particularly at risk of being exposed. Further investigations are needed to assess if such exposed populations have been infected by one of these viruses, if these infections are associated with symptoms, and whether they could confer protection against subsequent SARS-CoV-2 infections. In this context, it is noteworthy that SARS-CoV-2 with the furin site deleted replicates in

hamsters and in transgenic mice expressing hACE2, but leads to ablated disease while protecting from rechallenge with wild-type SARS-CoV-2¹⁶.

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^{40,41}.

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 traps⁴² 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^{42–44}. 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 1²⁶.

Initial Coronavirus screening

Total RNA was extracted from feces samples and the presence of coronaviruses was tested by PCR targeting the RNA-dependent RNA polymerase gene using combinations of degenerate and non-degenerate consensus primers followed by amplicon sequencing (see Supplementary Methods).

Next-Generation Sequencing

Betacoronavirus enrichment was performed before sequencing at the reverse transcription step by adapting a previously described protocol (Supplementary Methods)⁴⁵. For samples with a low nucleic acid content, a random amplification step was then performed using the MALBAC Single Cell WGA kit (Yikon Genomics, Promega). Libraries were generated using the NEBNext Ultra II DNA Library Prep kit (New England Biolabs) and run on the Illumina NextSeq500 platform with High Output Kit v2.5 (150 Cycles). 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. Raw reads from the enrichment-based sequencing were processed with an in-

house bioinformatics pipeline (Microseek; Bigot et al., submitted). 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 (see Supplementary Methods).

Recombination and phylogenetic analyses

Identification of recombination events occurring during the evolutionary history of bat sarbecoviruses was performed by the IDPlot package⁴⁶, a web-based workflow that includes multiple sequence alignment and phylogeny-based breakpoint prediction using the GARD algorithm from the HyPhy genetic analysis suite⁴⁷. Coordinates for the breakpoints were confirmed by performing phylogenetic analyses on the corresponding fragments using PhyML implemented through the NGPhylogeny portal⁴⁸. Branch support was evaluated with the aBayes parameter (see Supplementary Methods).

Molecular Dynamics simulations of RBD/ACE2 complexes

We performed a total of 9 all-atom explicit-solvent molecular dynamics (MD) simulations of the SARS-CoV-2, BANAL-236, and BANAL-52/103 RBD/hACE2 complexes (Table S4). All simulations were performed with GROMACS 2020.4⁴⁹ for a total aggregated time of 9 μ s. The GROMACS topology and input files as well as the analysis scripts used are available on PLUMED-NEST (www.plumed-nest.org)⁵⁰ under accession ID plumID:21.037. Additional details of the MD simulations and their analysis are provided in Supplementary Methods.

Generation of lentiviral pseudoviruses and entry assays

The BANAL-236 and Wuhan spikes with a cytoplasmic tail truncation of 19 amino acids were expressed at the surface of lentiviral particles. HEK-293T cells stably expressing human ACE2 (293T-ACE2) were transduced in suspension by mixing 50 μ L of 3-fold serial dilutions of S-pseudotyped lentiviruses with cells in 96-well white culture plates. At 60-72 h post-transduction, Bright-Glo luciferase substrate (Promega) was added to the wells and luminescence was measured using a Berthold Centro XS luminometer (see Supplementary Methods).

Virus isolation

Rectal swabs were inoculated in duplicate in 24-well plates containing VeroE6 cells. Three and four days after infection, a cytopathic effect (CPE) was monitored and 100 μ L of supernatant was collected for RNA extraction. RT-qPCR targeting the E gene was performed as described⁵¹. Culture supernatant (C1) was harvested at day 4 and titrated by plaque assay on VeroE6. A viral stock was prepared by amplification on VeroE6 cells. Culture supernatant (C2) was harvested at day 4 when massive CPE was observed and titrated. RNA was extracted from the viral stock and submitted to random NGS analysis. Raw reads were processed with Microseek pipeline, as described above.

Protein expression and purification

BANAL-52/103 and BANAL-236 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°C and 8% CO₂. 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 package⁵². The structure of the complex was determined by molecular replacement with Phaser software⁵³ 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 COOT⁵⁴ and refined with phenix.refine⁵⁵. The final coordinates were deposited in the PDB with the entry code 7PKI.

Data availability statement:

Sequences data that support the findings of this study have been deposited in the GenBank database with the following accession numbers: MZ937000 (BANAL-52), MZ937001 (BANAL-103), MZ937002 (BANAL-116), MZ937003 (BANAL-236), and MZ937004 (BANAL-247).

Declarations

Acknowledgments:

The authors want to thank Souand Mohamed Ali, Nicolas Da Rocha, Thonglakhone Xaybounsou, and Somsanith Chonephetsarath for their help at the bench. Next-generation sequencing was performed with the help of Biomics 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) 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.

References

1. Hul, V. *et al.* A novel SARS-CoV-2 related coronavirus in bats from Cambodia. *bioRxiv* 2021.01.26.428212 (2021) doi:10.1101/2021.01.26.428212.
2. Zhou, H. *et al.* 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. Wacharapluesadee, S. *et al.* Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. *Nat. Commun.* **12**, 972 (2021).
4. Murakami, S. *et al.* Detection and Characterization of Bat Sarbecovirus Phylogenetically Related to SARS-CoV-2, Japan. *Emerg. Infect. Dis.* **26**, 3025–3029 (2020).

5. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
6. Rahalkar, M. C. & Bahulikar, R. A. 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).
7. Liu, P. *et al.* Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? *PLoS Pathog.* **16**, e1008421 (2020).
8. Xiao, K. *et al.* Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature* **583**, 286–289 (2020).
9. Wahba, L. *et al.* An Extensive Meta-Metagenomic Search Identifies SARS-CoV-2-Homologous Sequences in Pangolin Lung Viromes. *mSphere* **5**, e00160-20 (2020).
10. Letko, M., Marzi, A. & Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat. Microbiol.* **5**, 562–569 (2020).
11. Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221–224 (2020).
12. Wang, Q. *et al.* Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* **181**, 894-904.e9 (2020).
13. Rochman, N. D. *et al.* Ongoing global and regional adaptive evolution of SARS-CoV-2. *Proc. Natl. Acad. Sci.* **118**, (2021).
14. Liu, K. *et al.* Binding and molecular basis of the bat coronavirus RaTG13 virus to ACE2 in humans and other species. *Cell* **184**, 3438-3451.e10 (2021).
15. Aicher, S.-M. *et al.* Species-specific molecular barriers to SARS-CoV-2 replication in bat cells. *bioRxiv* 2021.05.31.446374 (2021) doi:10.1101/2021.05.31.446374.
16. Johnson, B. A. *et al.* Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature* **591**, 293–299 (2021).
17. Liu, K. *et al.* Cross-species recognition of SARS-CoV-2 to bat ACE2. *Proc. Natl. Acad. Sci.* **118**, (2021).
18. Chu, D. K. W. *et al.* Avian Coronavirus in Wild Aquatic Birds. *J. Virol.* **85**, 12815–12820 (2011).
19. Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **5**, 1403–1407 (2020).

20. Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020).
21. Laffebber, C., de Koning, K., Kanaar, R. & Lebbink, J. H. G. Experimental Evidence for Enhanced Receptor Binding by Rapidly Spreading SARS-CoV-2 Variants. *J. Mol. Biol.* **433**, 167058 (2021).
22. Lei, C. *et al.* Neutralization of SARS-CoV-2 spike pseudotyped virus by recombinant ACE2-Ig. *Nat. Commun.* **11**, 2070 (2020).
23. Walls, A. C. *et al.* Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281-292.e6 (2020).
24. Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).
25. Hu, B. *et al.* Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLOS Pathog.* **13**, e1006698 (2017).
26. Ge, X.-Y. *et al.* Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **503**, 535–538 (2013).
27. Latinne, A. *et al.* Origin and cross-species transmission of bat coronaviruses in China. *Nat. Commun.* **11**, 4235 (2020).
28. Novel Immunoglobulin Domain Proteins Provide Insights into Evolution and Pathogenesis of SARS-CoV-2-Related Viruses. <https://journals.asm.org/doi/epub/10.1128/mBio.00760-20>
doi:10.1128/mBio.00760-20.
29. Su, Y. C. F. *et al.* Discovery and Genomic Characterization of a 382-Nucleotide Deletion in ORF7b and ORF8 during the Early Evolution of SARS-CoV-2. *mBio* **11**, e01610-20.
30. Chinese SARS Molecular Epidemiology Consortium. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. *Science* **303**, 1666–1669 (2004).
31. Conceicao, C. *et al.* The SARS-CoV-2 Spike protein has a broad tropism for mammalian ACE2 proteins. *PLoS Biol.* **18**, e3001016 (2020).
32. Damas, J. *et al.* Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 22311–22322 (2020).
33. Cohen, J. Wuhan coronavirus hunter Shi Zhengli speaks out. *Science* **369**, 487–488 (2020).
34. Ge, X.-Y. *et al.* Coexistence of multiple coronaviruses in several bat colonies in an abandoned mineshaft. *Virol. Sin.* **31**, 31–40 (2016).

35. Clements, R., Sodhi, N. S., Schilthuizen, M. & Ng, P. K. L. Limestone Karsts of Southeast Asia: Imperiled Arks of Biodiversity. *BioScience* **56**, 733–742 (2006).
36. Soisook, P. *et al.* A taxonomic review of *Rhinolophus steno* and *R. malayanus* (Chiroptera: Rhinolophidae) from continental Southeast Asia: an evaluation of echolocation call frequency in discriminating between cryptic species. *Acta Chiropterologica* **10**, 221–242 (2008).
37. Francis, c. *Field Guide to the Mammals of South-east Asia (2nd Edition)*. (2019).
38. Makarenkov, V., Mazouze, B., Rabusseau, G. & Legendre, P. Horizontal gene transfer and recombination analysis of SARS-CoV-2 genes helps discover its close relatives and shed light on its origin. *BMC Ecol. Evol.* **21**, 5 (2021).
39. Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nat. Med.* **26**, 450–452 (2020).
40. Predict. PREDICT One Health Consortium 2013. Protocol for Bat and Rodent Sampling Methods.
41. 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).
42. Francis, C. A Comparison of Mist Nets and Two Designs of Harp Traps for Capturing Bats. *Journal of Mammalogy* 865–970.
43. 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).
44. Csorba, G., Ujhelyi, P. & Thomas, N. Horseshoe Bats of the World (Chiroptera : rhinolophidae). *Alana Books Bishop's Castle Shropsh. U. K.* 160 (2003).
45. Deng, X. *et al.* Metagenomic sequencing with spiked primer enrichment for viral diagnostics and genomic surveillance. *Nat. Microbiol.* **5**, 443–454 (2020).
46. 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).
47. 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).
48. Lemoine, F. *et al.* NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Res.* **47**, W260–W265 (2019).

49. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
50. PLUMED consortium. Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* **16**, 670–673 (2019).
51. 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).
52. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
53. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
54. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
55. 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).
56. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **20**, 1160–1166 (2019).
57. 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).
58. Virachith, S. *et al.* Low seroprevalence of COVID-19 in Lao PDR, late 2020. *Lancet Reg. Health - West. Pac.* **13**, 100197 (2021).

Figures

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 3.5.157. (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, Wuhan-Hu-1). 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).

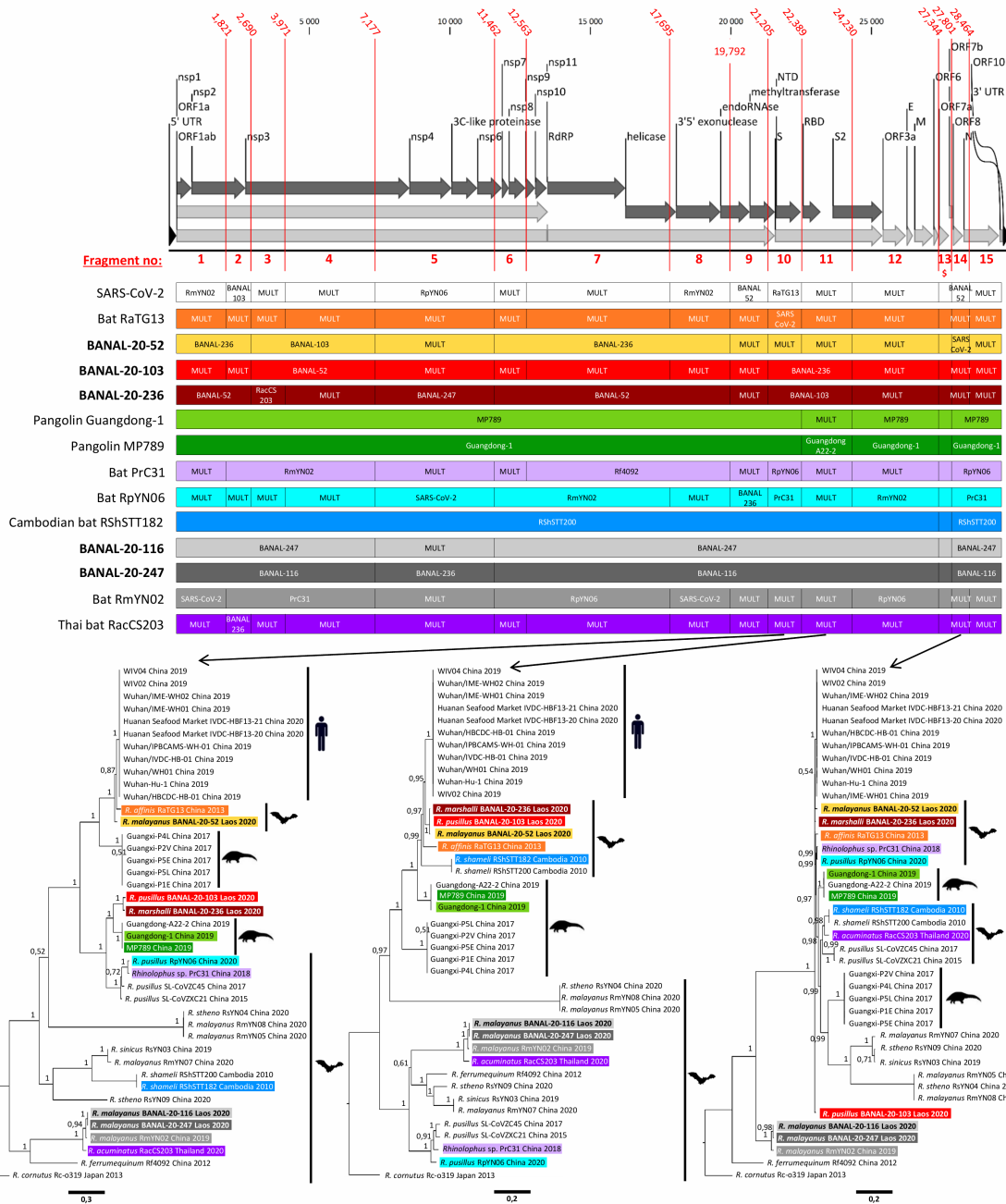


Figure 2

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) and associated phylogenetic reconstructions performed with PhyML implemented through the NGPhylogeny portal48. Branch support was evaluated with the aBayes parameter. Posterior probabilities >0.5 are mentioned. 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.

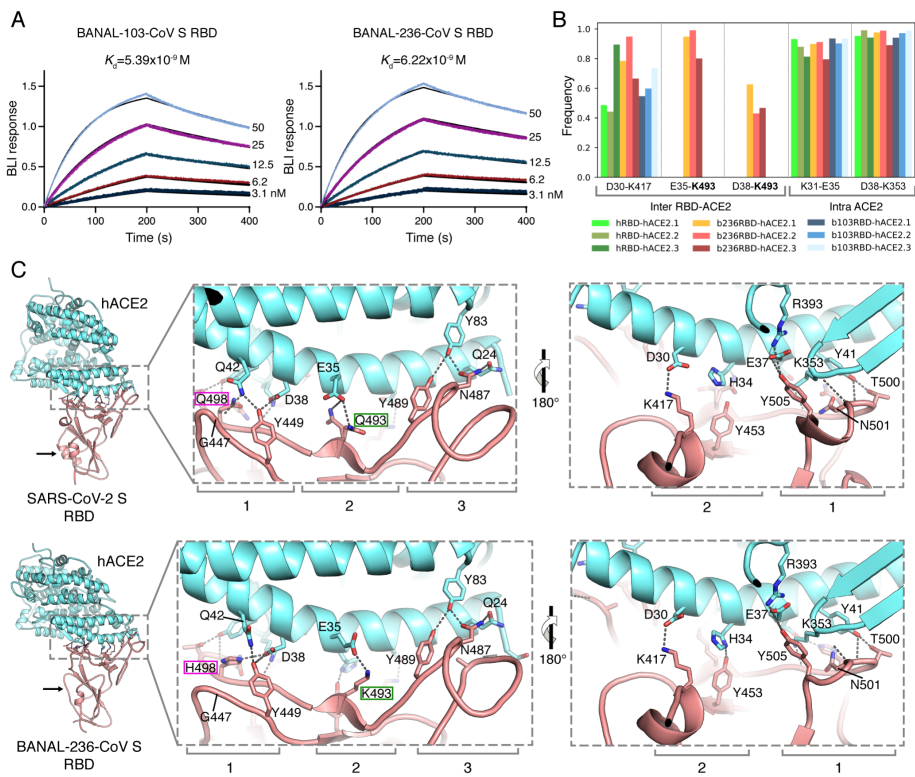


Figure 3

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 or BANAL-236 RBDs. Black lines correspond to global fit of the data using a 1:1 binding model. (B) Frequency of formation of salt bridges at the interface of RBD and hACE2 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.

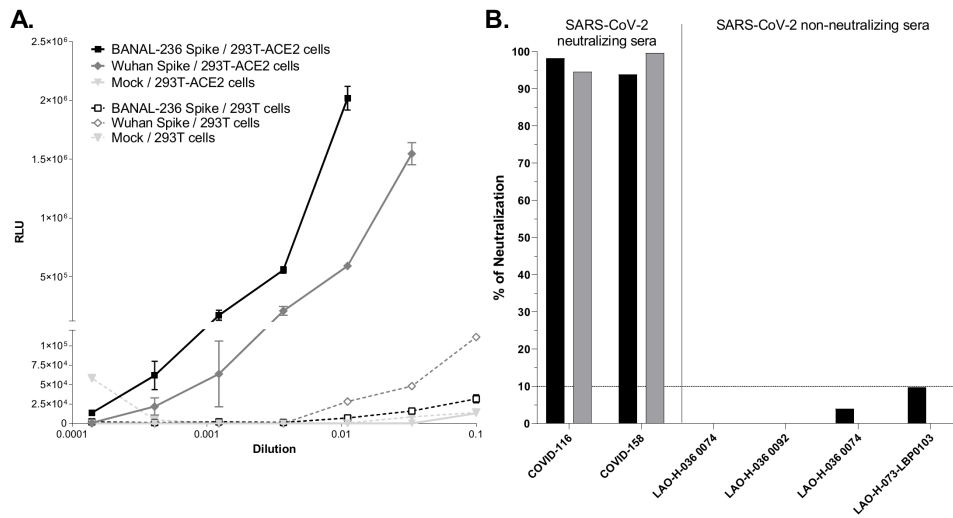


Figure 4

Pseudoviruses expressing BANAL-236 spike entry in human cells. (A) Results of spike-pseudotyped BANAL-236 (black) and Wuhan-Hu-1 (dark grey) pseudoviruses entry assay in HEK-293T cells expressing or not 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. (B) Results of spike-pseudotyped BANAL-236 (black) and Wuhan-Hu-1 (dark grey) neutralization assay expressed in % of

neutralization of the input pseudovirus 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 Laos⁵⁸. Dashed line: neutralization threshold.

Supplementary Files

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

- [supplementarymethodsanddatav9forsubmission.docx](#)