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.

Index

Supplementary figures

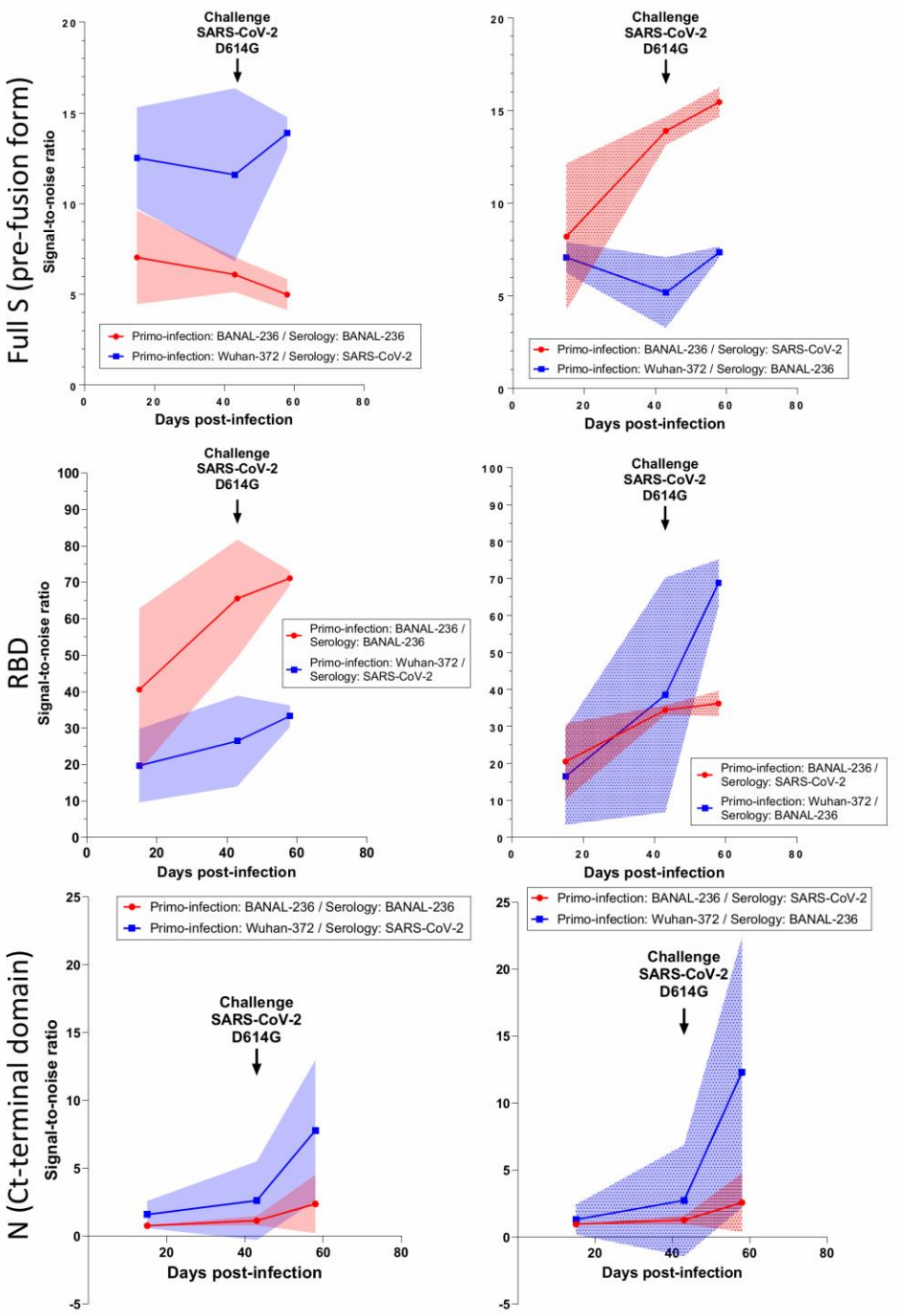

Fig. S1. Immunoprofilling of BANAL-236 and SARS-CoV-2 responses of transgenic mice. K18-hACE2 transgenic mice primo-infected at 10^3 PFU by BANAL-236 or SARS-CoV-2 Wuhan-372 were challenged by an infection at 10^4 PFU of SARS-CoV-2 Wuhan-D614G. Serological responses were assessed by LIPS targeting the full Spike ectodomain (stabilized in pre-fusion form), the RBD or the nucleoprotein (C-terminal domain) at D15 and D43 post-primo-infection, and at D14 post-challenge. The mean signal-to-noise ratio is presented as continuous lines while the standard deviation is presented as filled areas.

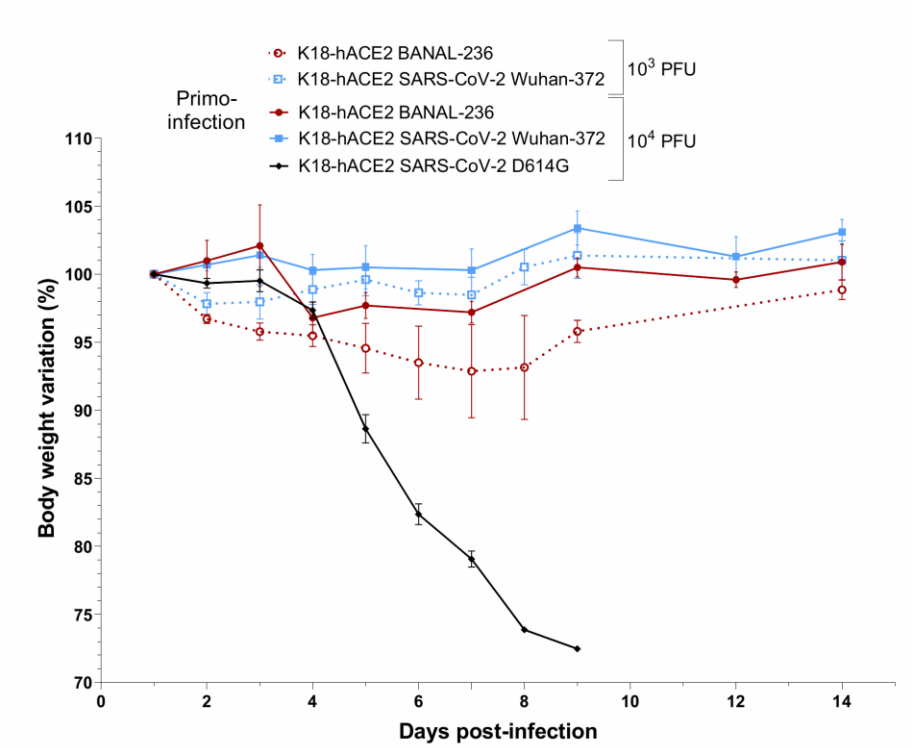

Fig. S2. Body weight variation of transgenic mice challenged by SARS-CoV-2 Wuhan-D614G strain after a primo-infection by BANAL-236 (red) or SARS-CoV-2 Wuhan-372 (blue) at 10^3 (dashed lines) or 10^4 PFU (solid lines). Naïve K18-hACE2 mice infected with 10^4 PFU of SARS-CoV-2 Wuhan-D614G served as reference (black line). The mean body weight (+/- SEM) variations are presented.

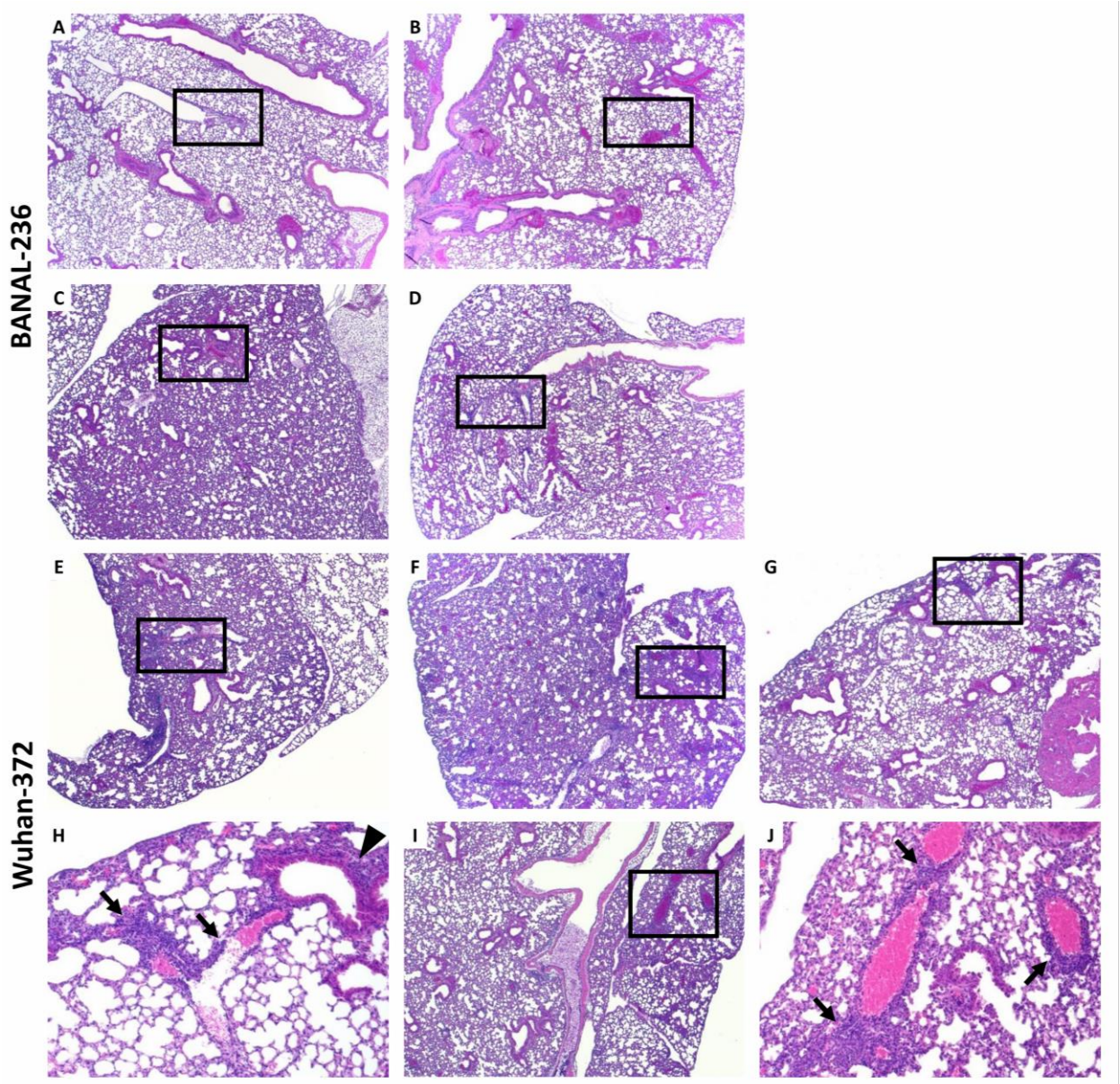

Fig. S3. Hematoxylin-eosin-stained histological section of lung lobes from infected mice. A-D: mice infected with 10^4 PFU of BANAL-236. Rectangles delineate high magnification fields presented in Fig. 1F-a to 1F-d. **E-J**: mice infected with 10^4 PFU of Wuhan-372. E-F: rectangles delineate high magnification fields presented in Fig. 1F-e and 1F-f. G & I: rectangles delineate high magnification fields presented in Fig. S3-H and S3-J, respectively. H & J: interstitial inflammation was generally centered on blood vessels (arrows) or bronchi/bronchioles (arrowheads).

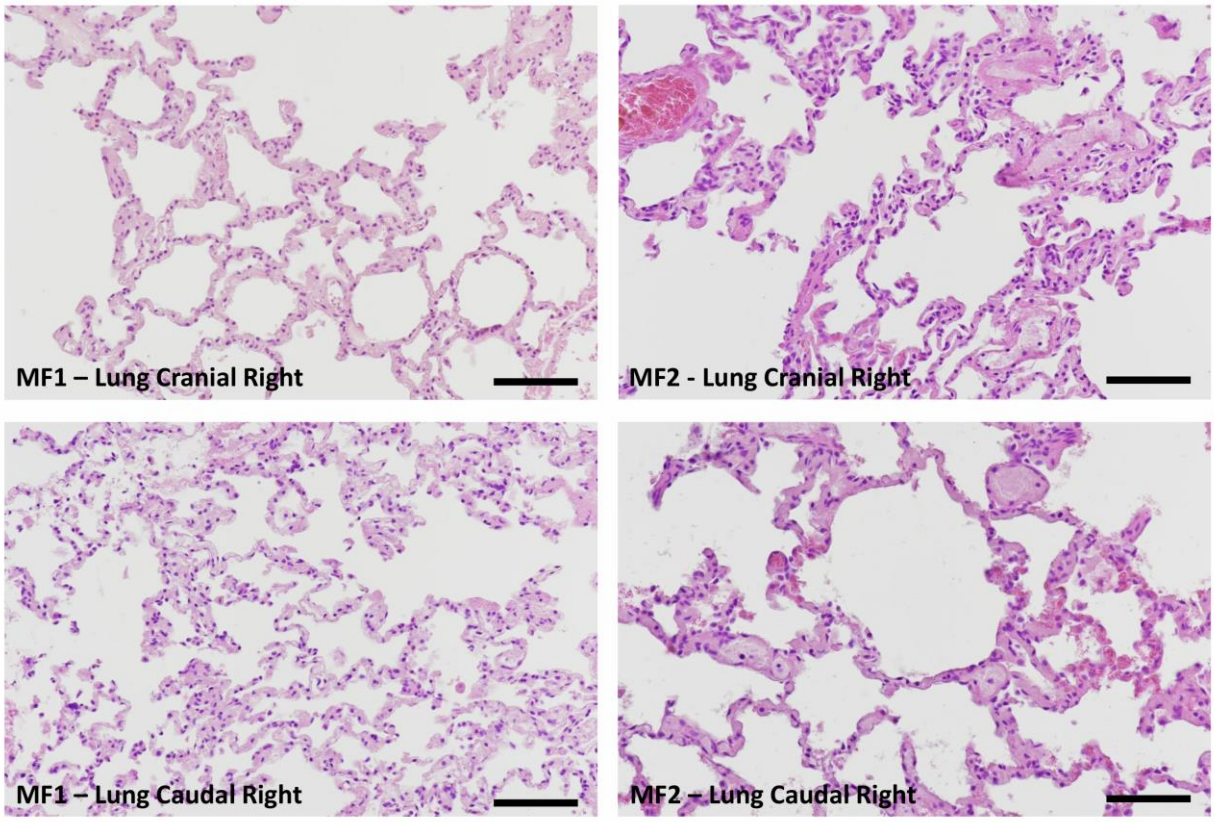

Fig. S4. Hematoxylin-eosin staining of the lungs in infected macaques. For both animals, no tissular changes was observed in both lobes of the lungs. Scale bar = 200µm.

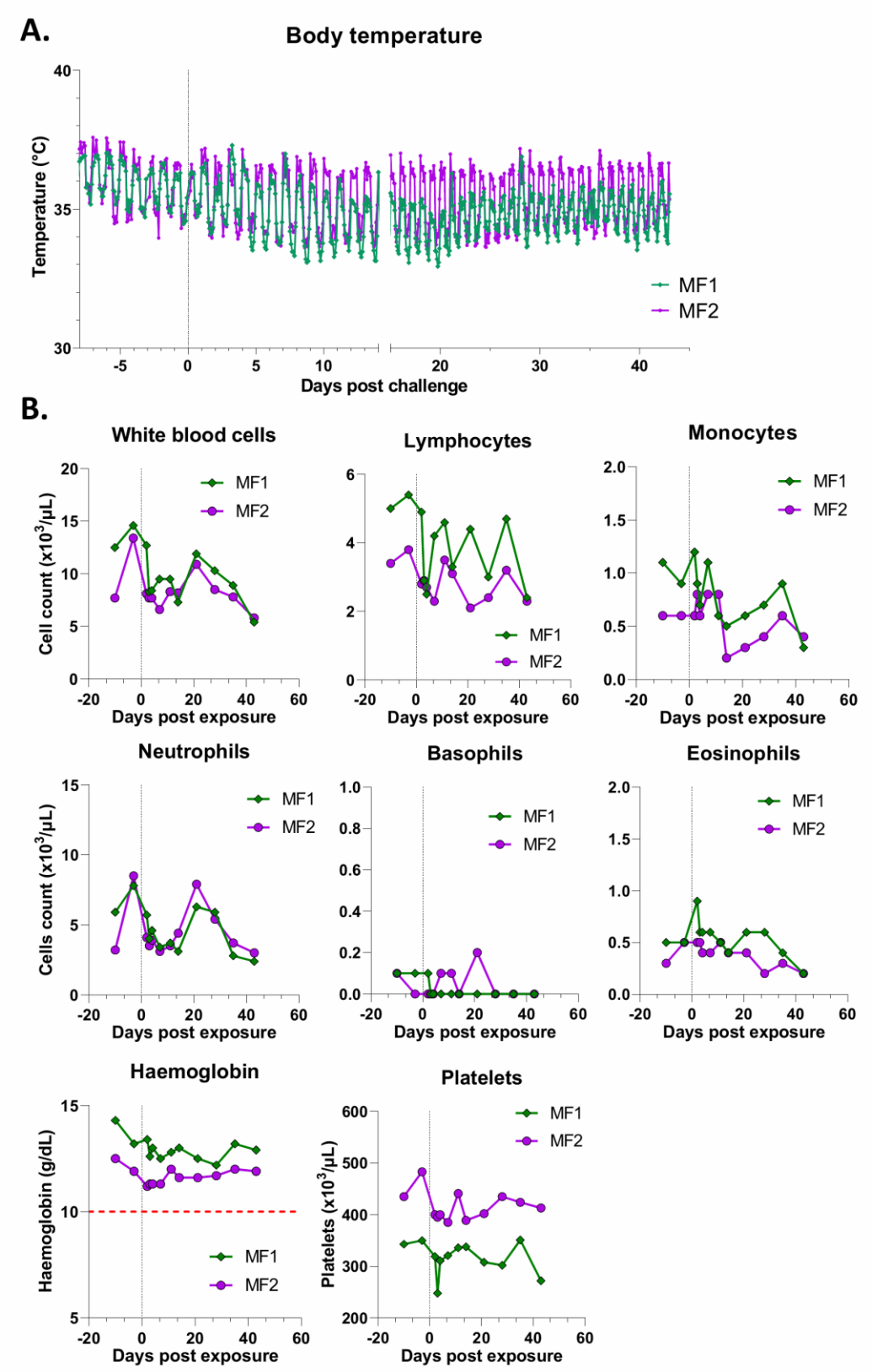

Fig. S5. Macaques body temperature (A) and hematological parameters (B) during BANAL-236 infection.

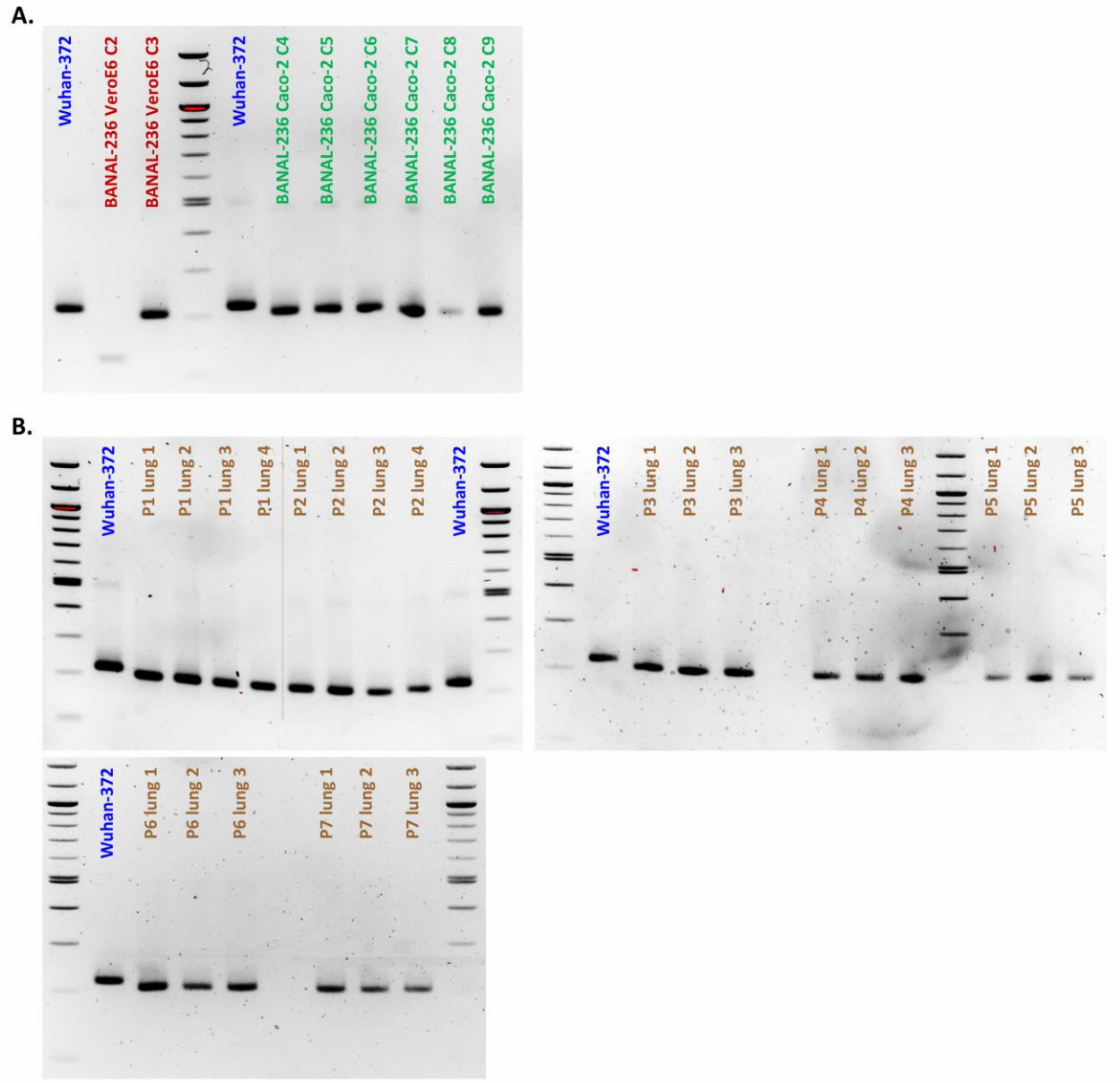

Fig. S6. Agarose gel electrophoresis of the PCR surrounding the furin site. A set of primers common to BANAL-236 and SARS-CoV-2 was used to verify the absence of the 12-nt long furin site either in the Vero E6 and Caco-2 (A) or in the K18-hACE2 mice (B) serial passages. SARS-CoV-2 Wuhan-372 RNA served as positive control (blue), and a 100-bp DNA ladder was used.

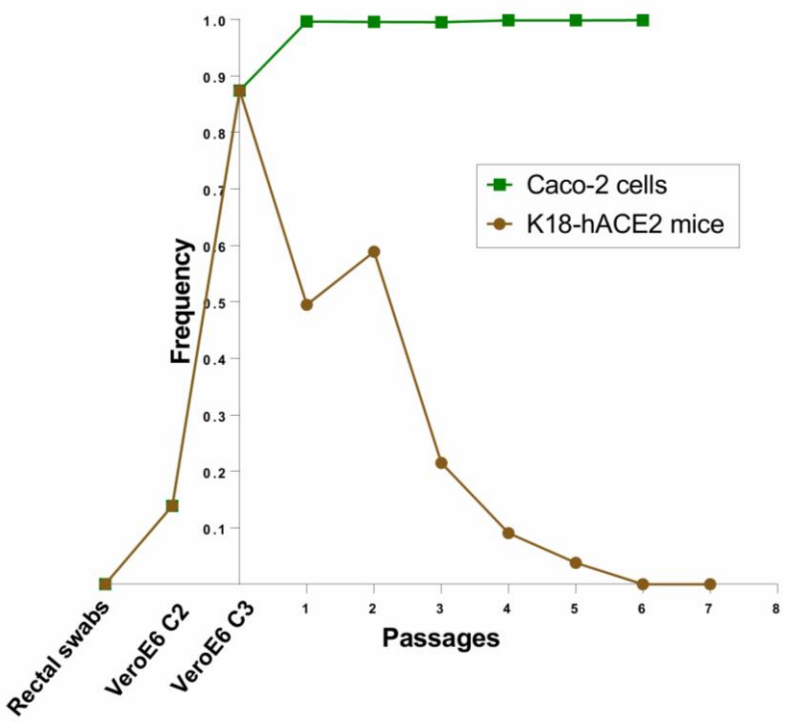

Fig. S7. Evolution of the frequency of the spike V391I mutation during the passages of BANAL-236 in transgenic mice or on Caco-2 cells.

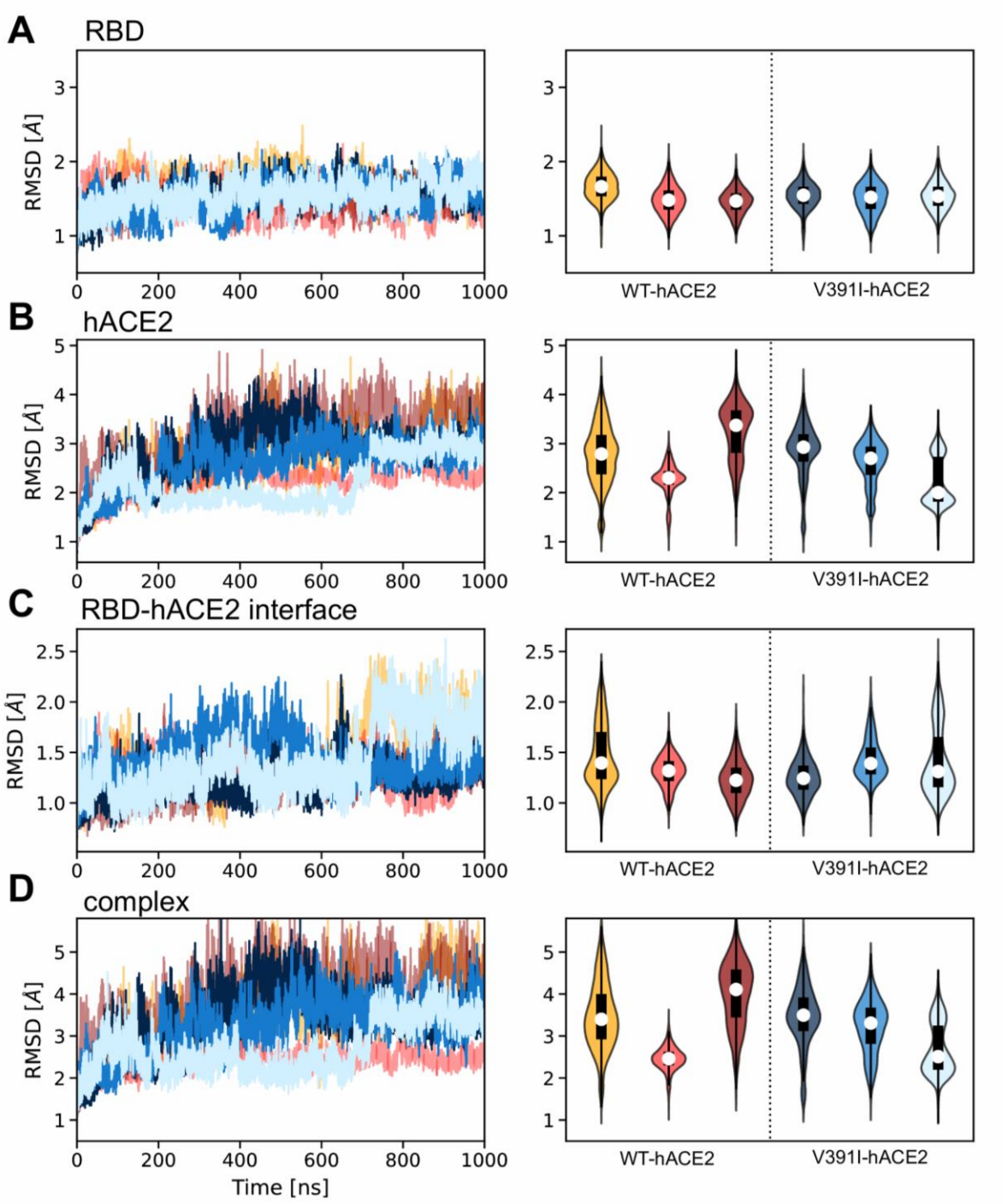

Fig. S8. Analysis of the stability of RBD–hACE2 complexes. Time series (left column) and violin plots (right column) of backbone Root Mean Square Deviation (RMSD) from the initial, energy-minimized model calculated on the residues in RBD (**A**), hACE2 (**B**), at the RBD-hACE2 interface (**C**), and on the entire complex (**D**). In the violin plots, the white circle corresponds to the median value, the black rectangle extends from the first to the third quantiles, and the thin black line represents the 95% confidence intervals. The analysis was performed for 6 different MD simulations (three replicates for each complex) of hACE2 in complex with RBDs from wild type BANAL-236 (WT-hACE2, shades of red) and V391I mutant (V391I-hACE2, shades of blue).

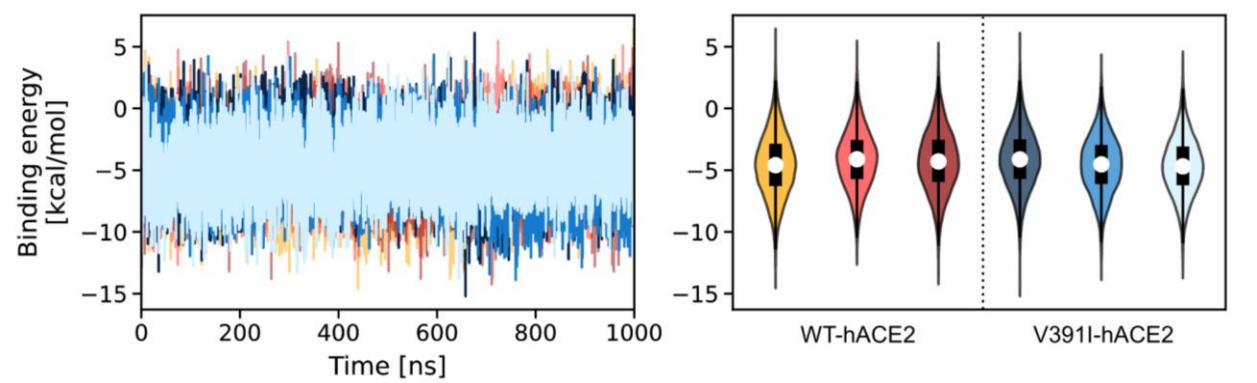

Fig. S9. Estimation of RBD–hACE2 binding energy. Time series (left column) and violin plots (right column) of the RBD–hACE2 binding energy estimated using FoldX. In the violin plots, the white circle corresponds to the median value, the black rectangle extends from the first to the third quantiles, and the thin black line represents the 95% confidence intervals. The analysis was performed for 6 different MD simulations (three replicates for each complex) of hACE2 in complex with RBDs from wild type BANAL-236 (WT-hACE2, shades of red) and V391I mutant (V391I-hACE2, shades of blue).

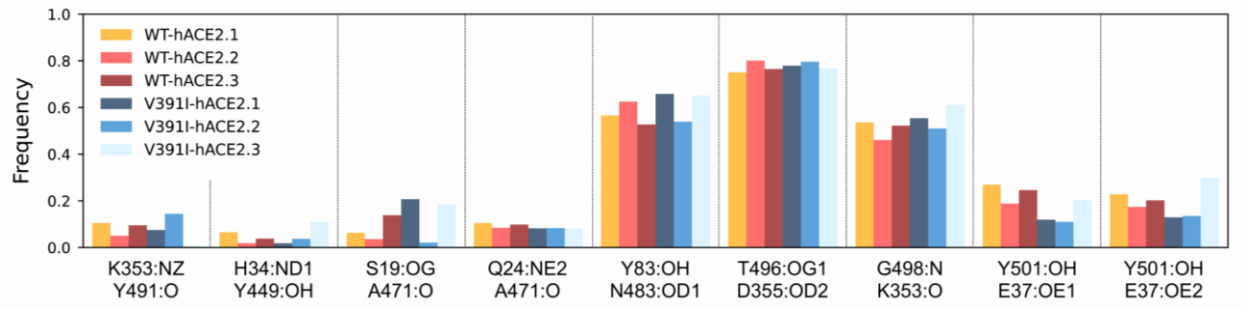

Fig. S10. Analysis of the inter-subunits hydrogen bonds at the interface of RBD and hACE2. Frequency of formation of hydrogen bonds at the RBD–hACE2 interface during the course of the MD simulations. The analysis was performed for 6 different MD simulations (three replicates for each complex) of hACE2 in complex with RBDs from the wild type BANAL-236 (WT-hACE2.1, WT-hACE2.2, and WT-hACE2.3, shades of red) and V391I mutant (V391I-hACE2.1, V391I-hACE2.2 and V391I-hACE2.3, shades of blue).

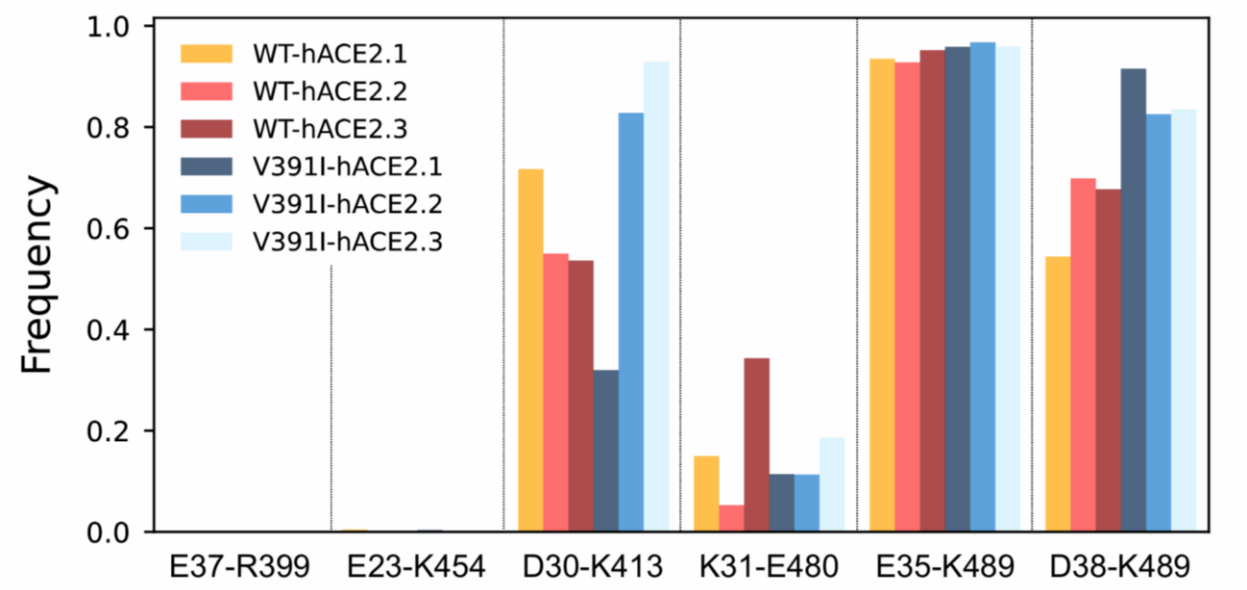

Fig. S11. Analysis of the inter-subunits salt bridges at the interface of RBD and hACE2. Frequency of formation of salt bridges at the RBD–hACE2 interface (from left to right: E37-R399, E23-K454, D30-K413, K31-E480, E35-K489, and D38-K489) during the course of the MD simulations. The analysis was performed for 6 different MD simulations (three replicates for each complex) of hACE2 in complex with RBDs from the wild type BANAL-236 (WT-hACE2.1, WT-hACE2.2, and WT-hACE2.3, shades of red) and V391I mutant (V391I-hACE2.1, V391I-hACE2.2 and V391I-hACE2.3, shades of blue).

Supplementary tables

Table S1. Sequencing depth of the furin cleavage site locus of BANAL-236 during passages or during macaques infection.

Sample	Median coverage	Average coverage	Standard deviation
BANAL-236 rectal swabs	9.00	22.09	21.10
VeroE6 C2	43.00	195.88	187.30
VeroE6 C3	35.00	254.36	266.66
Caco-2 C4	14.00	43.13	32.97
Caco-2 C5	94.00	203.32	158.06
Caco-2 C6	61.00	118.51	87.42
Caco-2 C7	44.00	55.70	29.34
Caco-2 C8	75.00	219.26	195.11
Caco-2 C9	67.00	137.35	107.46
Macaque MF2 rectal D7	176.00	145.62	55.18
Macaque MF2 rectal D11	150.00	137.06	57.85
Macaque MF2 rectal D14	1.00	1.80	0.91
K18-hACE2 lung P 1	59.00	56.37	26.20
K18-hACE2 lung P 2	23.00	19.79	7.93
K18-hACE2 lung P 3	6.00	5.39	2.05
K18-hACE2 lung P 4	20.00	19.46	9.18
K18-hACE2 lung P 5	24.00	27.99	14.54
K18-hACE2 lung P 6	24.00	20.35	8.89
K18-hACE2 lung P 7	101.00	88.88	38.37

Table S2. Sequencing depth along the whole genome during BANAL-236 passages or during the course of macaques infection.

Sample	Median coverage	Average coverage	Standard deviation
BANAL-236 rectal swabs	1,490.50	2,150.66	1,927.94
VeroE6 C2	71,641.50	97,462.57	90,208.34
VeroE6 C3	94,108.00	125,135.35	104,301.67
Caco-2 C4	184.00	306.04	430.68
Caco-2 C5	54,159.00	76,207.35	71,986.96
Caco-2 C6	30,000.50	44,656.44	47,303.14
Caco-2 C7	17,223.00	24,881.32	24,821.99
Caco-2 C8	47,569.50	68,089.55	68,566.49
Caco-2 C9	46,557.00	62,084.65	54,894.94
Macaque MF2 rectal D7	12,293.50	16,091.94	13,871.78
Macaque MF2 rectal D11	14,169.50	17,967.92	15,442.87
Macaque MF2 rectal D14	118.00	142.30	94.54
K18-hACE2 lung P 1	2,943.50	4,390.90	5,183.74
K18-hACE2 lung P 2	981.00	1,108.02	587.44
K18-hACE2 lung P 3	184.00	225.77	237.14
K18-hACE2 lung P 4	806.00	1,209.74	1,987.64
K18-hACE2 lung P 5	1,216.00	1,615.23	2,080.82
K18-hACE2 lung P 6	851.00	1,238.54	1,978.27
K18-hACE2 lung P 7	4,316.00	4,788.51	4,788.51

Table S3. Correspondence of BANAL-236 mutations onto the reference genome of SARS-CoV-2.

CDS (protein)	BANAL-236 MZ937003.2	SARS-CoV-2 NC_045512.2
ORF1ab (NSP3)	E1215D	(E1223)
ORF1ab (NSP10)	N4350K	(N4358)
ORF1ab (NSP12 RdRp)	N4984N	(N4992)
ORF1ab (NSP13 helicase)	S5575S	(S5583)
ORF1ab (NSP15 endoRNAse)	N6448T	(N6456)
S1 NTD	T108T	(T108)
S1 RBD	V391I	(V395)
S1	P627L	(P631)
ORF3a (NS3a)	T9A	(T9)
	V13A	(V13)
	L101R	(L101)
	T248T	(T248)
M	T8I	(T9)
ORF6 (NS6)	A12T	(A12)
N	T135T	(T135)

Table S4. List of the MD simulations performed.

ID	Construct	Initial model	# K/Cl ions	# waters	Total # atoms	Production time [μs]
WT-hACE2.1	WT BANAL-236 RBD / hACE2	X-ray	113/91	27990	96937	1
WT-hACE2.2	WT BANAL-236 RBD / hACE2	X-ray	113/91	27990	96937	1
WT-hACE2.3	WT BANAL-236 RBD / hACE2	X-ray	113/91	27990	96937	1
V391I-hACE2.1	V391I BANAL-236 RBD / hACE2	Homology model	113/91	27955	96835	1
V391I-hACE2.2	V391I BANAL-236 RBD / hACE2	Homology model	113/91	27955	96835	1
V391I-hACE2.3	V391I BANAL-236 RBD / hACE2	Homology model	113/91	27955	96835	1

Table S5. BANAL-236 substitutions observed in SARS-CoV-2 sequences. BANAL-236 nucleotide substitutions found in the GISAID EpiCoV^{TM} database as of June 15, 2022 which contains 11,403,565 viral genomes. The whole database, as well as the VOC Omicron, Alpha, Beta and Gamma, and VOI Lambda and MU sub-bases are queried. The first value corresponds to the amount of sequences for which a substitution is observed at the same position and the second value to the amount of genomes in which the same substitution as in BANAL-236 is found.

Sample	BANAL-236 MZ937003.2	SARS-CoV-2 NC_045512.2	Total viruses	VOC Omicron	VOC Delta	VOC Alpha	VOC Beta	VOC Gamma	VOI Lambda	VOI MU
ORF1ab (NSP3)	E1215D	(E1223)	0							
ORF1ab (NSP10)	N4350K	(N4358)	0							
ORF1ab (NSP12 RdRp)	N4984N	(N4992)	0							
ORF1ab (NSP13 helicase)	S5575S	(S5583)	0							
ORF1ab (NSP15 endoRNAse)	N6448T	(N6456)	0							
S1 NTD	T108T	(T108)	1,832	660	874	77	1	59	0	0
S1 RBD	V391I	(V395)	605 / 155	186 / 53	167 / 42	116 / 15	17 / 13	6 / 3	0	1 / 0
S1	P627L	(P631)	3,168	546	974	782	7	41	6	1
ORF3a (NS3a)	T9A	(T9)	10,172 / 252	8,262 / 102	1,367 / 131	168 / 5	3 / 0	28 / 2	1 / 0	3 / 0
	V13A	(V13)	17,264 / 581	3,944 / 216	6,418 / 204	1,180 / 124	40 / 0	228 / 2	41 / 1	112 / 1
	L101R	(L101)	13,758 / 137	1,341 / 40	9,808 / 55	1,100 / 55	75 / 2	96 / 1	3 / 0	15 / 0
	T248T	(T248)	573	207	277	27	3	4	0	0
M	T8I	(T9)	0							
ORF6 (NS6)	A12T	(A12)	6,238	1,914	1,237	2,557	20	123	12	4
N	T135T	(T135)	39,175	8,071	18,569	5,671	123	414	18	8