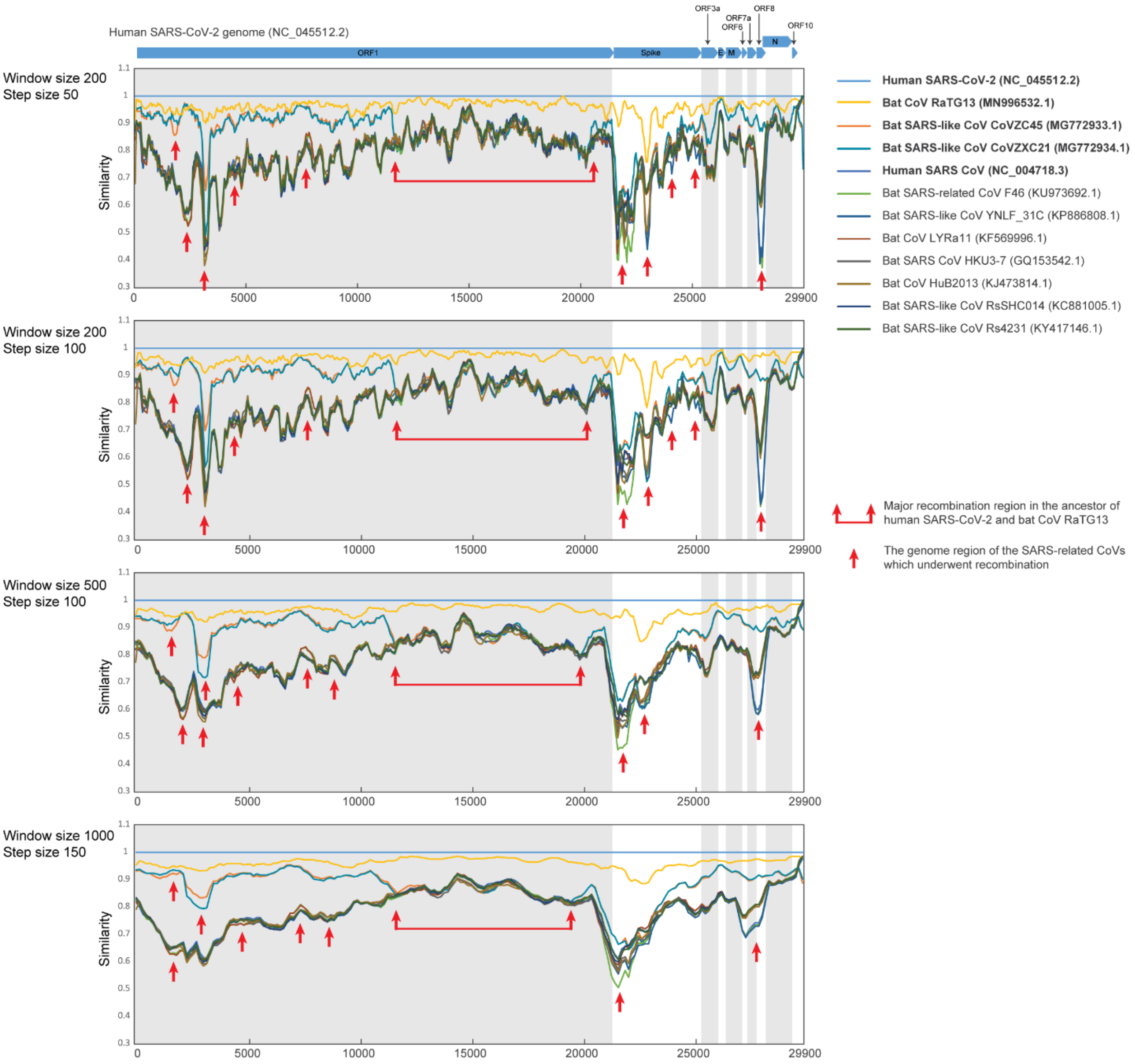
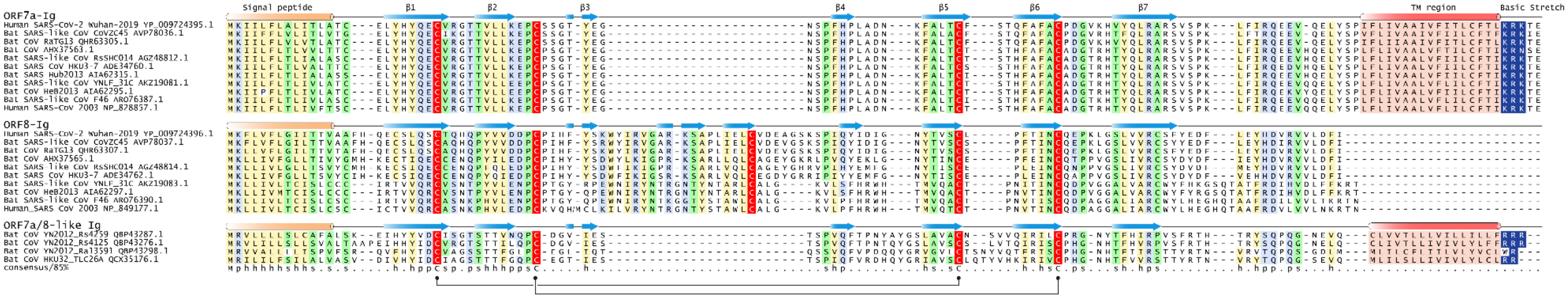
SUPPLEMENTAL INFORMATION

Supplementary Figure S1. Genome comparison analysis of SARS-related CoVs.

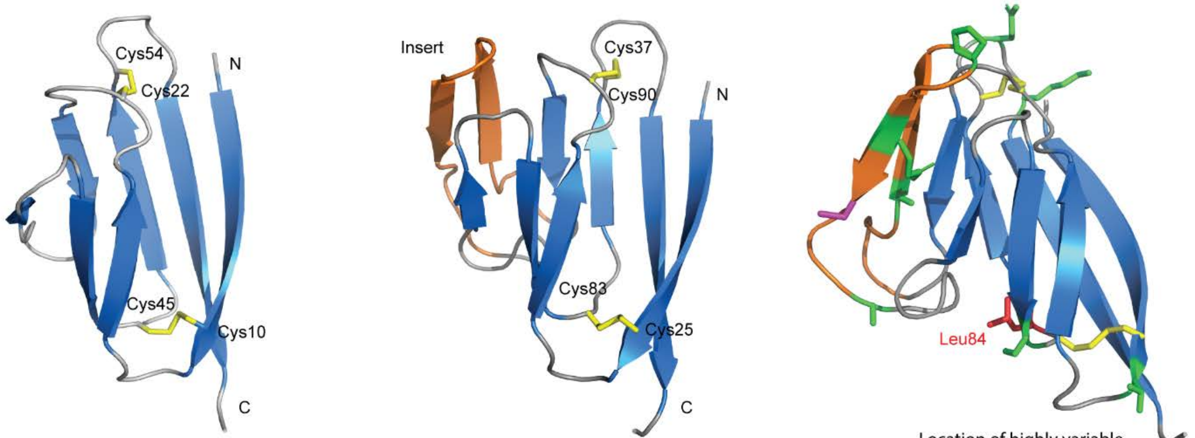


Similarity Plot of SARS-related CoVs against human SARS-CoV-2 Wuhan-Hu-1 genome (NC_045512.2) based on a multiple sequence alignment of the whole genomes. Each point represents a different slicing window size from the alignment with a different step size between each point. For each plot, the window size and step size are shown in the top left. Horizontal bars above the top plot correspond to the different open reading frames of the SARS-CoV-2 genome (NC_045512.2). Each different colored line corresponds to the nucleotide similarity between the human SARS-CoV-2 genome (NC_045512.2) and the respective SARS-related CoV genome. The red arrows and solid lines surround regions which display recombination within the SARS-related CoV genomes. The single red arrows point to specific regions of recombination.

Supplementary Figure S2. Full length multiple sequence alignment of ORF7a, ORF8-Ig and ORF7a/8-like proteins.
Each sequence in the MSA was labelled by its species abbreviation followed by its isolation and NCBI accession number. The predicted secondary structure is shown above the alignment and the consensus is shown below the alignment, where h stands for hydrophobic residues, s for small residues, and p for polar residues. The characteristic signal peptide, TM region and a stretch of basic residues are also labeled.



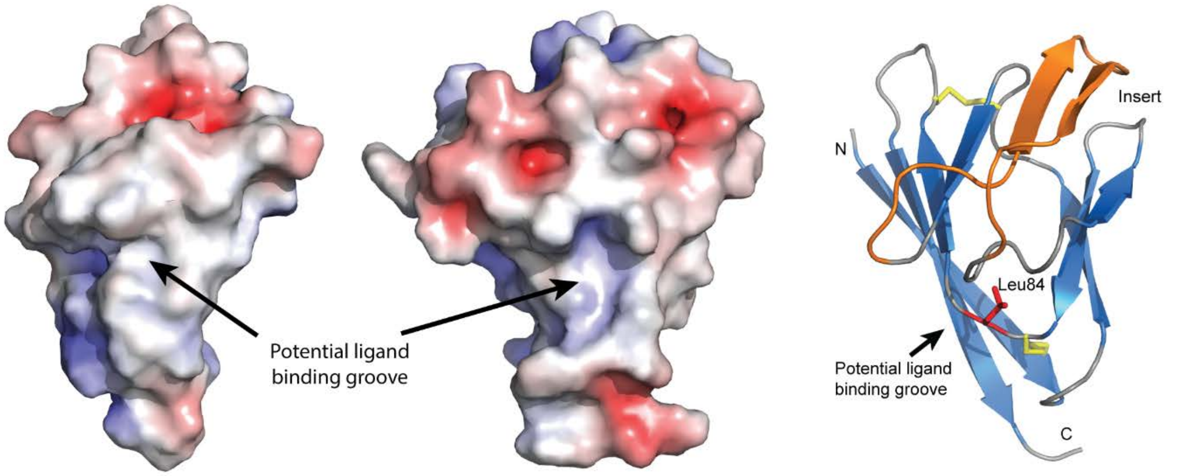
Supplementary Figure S3. Structural analysis of CoV Ig domains.



ORF7a-Ig domain
(PDB: 1xak_A)

Model of ORF8-Ig domain

Location of highly variable
residues of ORF8-Ig domain
(sticks in green and red)

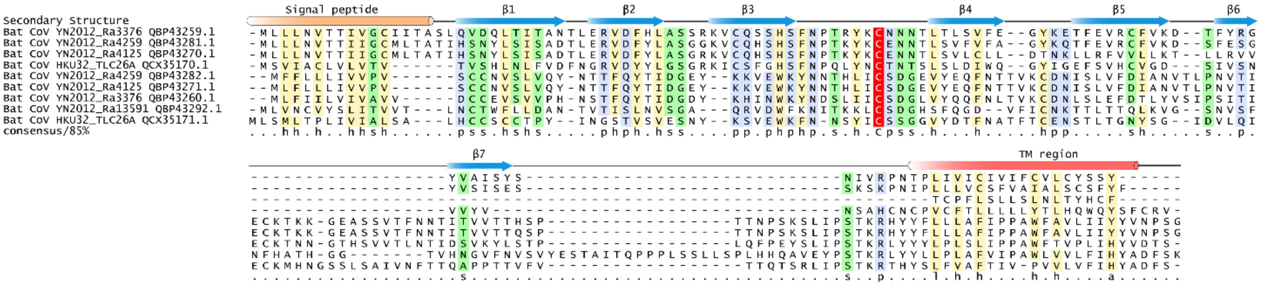


Surface view of ORF7a-Ig domain
(PDB: 1xak_A)

Surface view of ORF8-Ig domain

Location of hypervariable residue Leu84
on substrate-binding groove of the ORF8-ig domain
(stick in red)

Supplementary Figure S4. Multiple sequence alignment of alphacoronavirus E3-CR1-like Ig domain proteins and their related Ig domains identified by profile-profile searches.



```
>PF02440.15 ; Adeno_E3_CR1 ; Adenovirus E3 region protein CR1
Probab=90.89  E-value=2.8  Score=29.56  Aligned_cols=38  Identities=18%  Similarity=0.352  Sum_probs=0.0  Template_Neff=6.800

Q QBP43282.1       24 QYTIDGE-YKKVEWK----YNNTHLICSDGEVYEQFNTTVKCDNISLVF   67 (114)
Q Consensus        24 s~svd~~-y~rVdw~-----~n~s~kICs~ghsfn~f~~~~kC~N~TLt~   67 (114)
                      .++..+. +..|.|+    -..+.++|.....+      |+|++.+|++
T Consensus        17 NvTL~g~~~~~v~Wyr~~~~~~~~~LC~~~~~~~-------~~C~~~nLtL   59 (88)
T E3TMR2_ADE16/1   17 TCTLQGPQEGHVTWWRIYDNGGFARPCDQPGTK------FSCNGRDLTI   59 (88)
T ss_pred             cEEEeCCCCCCeEEEEeccCCCcccccccCCCe------eEeCCCceEE
```

```
>6JXR_n T-cell surface glycoprotein CD3 zeta; IMMUNE SYSTEM Homo sapiens
Probability: 62.4%, E-value: 84, Score: 22.12, Aligned cols: 126, Identities: 11%, Similarity: 0.052,

Q Q_9601142         19  VQYNTTFQYTIDGEYK---KVEWKYNNTHLICSDGEVYEQFN------TTVKCDNISLVFDIANVTLPNVTIECKTKKGE   89 (148)
Q Consensus         19  ~~~n~t~~~tv~g~yk---~VeWk~N~s~~iCSdG~vyq~fn-------t~~Cdn~tL~~~~~~vs~p~vtieck~~~G~   89 (148)
                        ...+.+++-.+.+.+.   .+.|.++.....-+.........      ..-..-..+|++.....+-..-.+.|.+.+..
T Consensus        135 ~g~~~~l~C~~~~~~~~~~~~~~~W~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~l~i~~~~~~d~~g~y~C~~~~~~~   214 (291)
T 6JXR_n          135 HTQKATLVCLATGFYPDHVELSWWVNGKEVHSGVSTDPQPLKEQPALNDSRYCLSSRLRVSATFWQNPRNHFRCQVQFYG   214 (291)

Q Q_9601142         90  ASSVTFNNTITVVTTHSPTTNPSKSLIPSTKRH----------YYFLLLAFIPPAWFAVLIIYYV   144 (148)
Q Consensus         90  ~~~~~vnn~~~~~tt~~p~~~p~~sLiPSTKR~----------yY~L~lafi~paw~~V~iihYv   144 (148)
                        .................+.......--+.....          .+.+.++++.-..++++++-.+
T Consensus        215 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~i~~~~   279 (291)
T 6JXR_n  215           LSENDEWTQDRAKPVTQIVSAEAWGRADCGFTSESYQQGVLSATILYEILLGKATLYAVLVSALV   279 (291)
```

```
>3OQ3_B Interferon alpha-5, IFN-alpha/beta binding protein; Ectromelia, Mousepox Virus, Moscow strain; Mus musculus
Probability: 51.52%, E-value: 150, Score: 21.59, Aligned cols: 82, Identities: 12%, Similarity: 0.143,

Q Q_9601142         19  VQYNTTFQYTIDGEYK-----KVEWKYNNTHLICSDGEVYEQFNTTVKCDNISLVFDIANVTLPNVTIECKTKKGEASSV   93 (148)
Q Consensus         19  ~~~n~t~~~tv~g~yk-----~VeWk~N~s~~iCSdG~vyq~fn~t~~Cdn~tL~~~~~~vs~p~vtieck~~~G~~~~~   93 (148)
                        .....+++-.+.|...     .+.|.+++...+-.+...+........-.+.+|++.-....--.-...|.+.|+.....
T Consensus        242 ~g~~~~l~C~~~~~~~~~~~~~~~~~~~~W~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~l~i~~~~~~d~G~~Y~C~a~n~~g~~~   321 (329)
T 3OQ3_B          242 IGEPANITCTAVSTSLLVDDVLIDWENPSGWIIGLDFGVYSILTSSGGITEATLYFENVTEEYIGNTYTCRGHNYYFDKT   321 (329)

Q Q_9601142         94  TFNNTIT   100 (148)
Q Consensus         94  ~vnn~~~   100 (148)
                        .-.++..
T Consensus        322 ~~~~l~v   328 (329)
T 3OQ3_B          322 LTTTVVL   328 (329)
```

Supplementary Table S1. Detailed information of Human SARS-CoV-2 Wuhan-Hu-1 genome and other SARS-related genomes that were used in this study (Figure 1 & Figure S1).

Organism	Host	NCBI ID	Year	Citation
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) isolate Wuhan-Hu-1	Human	NC_045512.2	2020	A novel coronavirus associated with a respiratory disease in Wuhan of Hubei province, China (Unpublished)
Bat SARS-like coronavirus isolate bat-SL-CoVZC45	Bat	MG772933.1	2018	(*1*)
Bat SARS-like coronavirus isolate bat-SL-CoVZXC21	Bat	MG772934.1	2018	(*1*)
Bat coronavirus isolate RaTG13	Bat	MN996532.1	2020	Not Available
Severe acute respiratory syndrome-related coronavirus	Human	NC_004718.3	2003	(*2*)
Severe acute respiratory syndrome-related coronavirus isolate F46	Bat	KU973692.1	2017	Identification of a new intermediate virus between bat-CoVs and SARS-CoVs from least horseshoe bats in China (Unpublished)
Bat SARS-like coronavirus YNLF_31C	Bat	KP886808.1	2015	Not Available
Rhinolophus affinis coronavirus isolate LYRa11	Bat	KF569996.1	2014	(*3*)
Bat SARS coronavirus HKU3-7	Bat	GQ153542.1	2010	(*4*)
BtRs-BetaCoV/HuB2013	Bat	KJ473814.1	2015	(*5*)
Bat SARS-like coronavirus RsSHC014	Bat	KC881005.1	2013	(*6*)
Bat SARS-like coronavirus isolate Rs4231	Bat	KY417146.1	2017	(*7*)

Supplementary Table S2. Summary of representatives of viral Ig domain proteins which were identified in this study.

Family	Genus	Organism	NCBI ID	pfam ID	Domain Famliy	Presence of Signal Peptide (Y/N)*	Number of Ig-like domain	Number of TM region**	Distinct Relative PDB Structure#
Coronaviridae	Beta-coronavirus	SARS-CoV-2	YP_009724395.1	PF08779	SARS_X4	Y	1	1	1XAK_A
			YP_009724396.1	PF12093	Corona_NS8	Y	1	0	1XAK_A
		SARS-CoV	NP_828857.1	PF08779	SARS_X4	Y	1	1	1XAK_A
			NP_828876.1 NP_828877.1	PF08779	Corona_NS8	Y	1	0	1XAK_A
	Alpha-coronavirus	Bat coronavirus	QBP43259.1	n/a	Adeno_E3_CR1-like	Y	1	1	5XMZ_A
			QBP43265.1	PF08779	SARS_X4	Y	1	1	1XAK_A
Adenoviridae	Mast-adenovirus	Human adenovirus 7d	AAF14132.1	PF02440	Adeno_E3_CR1	Y	1	1	6JXR_d
		Human adenovirus 23	AFK92306.1	PF02440	Adeno_E3_CR1	Y	3	1	3J8F_7
		Human adenovirus 21	AAW33363.1	PF04881	Adeno_GP19K	Y	1	1	5IRO_P
Herpesviridae	Mardivirus	Gallid alphaherpesvirus 2	YP_001034013.1	PF02480	Herpes_gE	Y	1	1	2GJ7_F
			YP_001034012.1	PF01688	Herpes_gI	Y	1	1	5OR7_C
			YP_001033973.1	PF02124	Marek_A	Y	3	1	3J8F_7
	Simplexvirus	Macacine alpha-herpesvirus 1	NP_851925.1	PF01537	Herpes_glycop_D	Y	1	1	4MYV_A
	Rhadinovirus	Human gamma-herpesvirus 8	YP_001129350.1	PF02960	K1	Y	2	1	5D6D_C
	Cytomegalo-virus	Panine beta-herpesvirus 2	NP_612760.1	PF16758	UL141	Y	1	1	4JM0_B
			NP_612778.1	PF05963	Cytomega_US3	Y	1	1	1IM3_P
		Human beta-herpesvirus 5	ABV71546.1	PF17622	UL16	Y	1	1	2WY3_B
		Aotine beta-herpesvirus 1	YP_004940175.1	PF08001	CMV_US	Y	1	2	1IM3_P
Poxviridae	Orthopoxvirus	Variola virus	NP_042191.1	PF08204	V-set_CD47	Y	1	5	5OR7_C
		Ectromelia virus	3OQ3_B	PF13895	ig	Y	3	0	3OQ3_B
Phenuiviridae	Goukovirus	Cumuto virus	YP_009664616.1	PF07245	Phlebovirus_G2	Y	4	1	6F8P_A 6EGU_B

* Signal Peptide Prediction was conducted by SignalP-5.0 program (*8*).

** Transmembrane (TM) region predictions were conducted by TMHMM Server (*9*).

\# The PDB structures which display similarity with the respective viral Ig domains identified by pofile-profile comparisons (*10*).

REFERENCES

1. D. Hu *et al.*, Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. *Emerging microbes & infections* **7**, 1-10 (2018).
2. M. A. Marra *et al.*, The genome sequence of the SARS-associated coronavirus. *Science* **300**, 1399-1404 (2003).
3. B. He *et al.*, Identification of diverse alphacoronaviruses and genomic characterization of a novel severe acute respiratory syndrome-like coronavirus from bats in China. *Journal of virology* **88**, 7070-7082 (2014).
4. S. K. Lau *et al.*, Ecoepidemiology and complete genome comparison of different strains of severe acute respiratory syndrome-related Rhinolophus bat coronavirus in China reveal bats as a reservoir for acute, self-limiting infection that allows recombination events. *Journal of virology* **84**, 2808-2819 (2010).
5. Z. Wu *et al.*, Deciphering the bat virome catalog to better understand the ecological diversity of bat viruses and the bat origin of emerging infectious diseases. *The ISME journal* **10**, 609-620 (2016).
6. X.-Y. Ge *et al.*, Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **503**, 535-538 (2013).
7. B. Hu *et al.*, Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS pathogens* **13**, (2017).
8. J. J. A. Armenteros *et al.*, SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature biotechnology* **37**, 420-423 (2019).
9. A. Krogh, B. Larsson, G. Von Heijne, E. L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology* **305**, 567-580 (2001).
10. J. Söding, A. Biegert, A. N. Lupas, The HHpred interactive server for protein homology detection and structure prediction. *Nucleic acids research* **33**, W244-W248 (2005).