



HAL
open science

In vitro and computational analysis of the putative furin cleavage site (RRARS) in the divergent spike protein of the rodent coronavirus AcCoV-JC34 (sub-genus luchacovirus)

Annette Choi, Deandria Singleton, Alison Stout, Jean K. Millet, Gary Whittaker

► To cite this version:

Annette Choi, Deandria Singleton, Alison Stout, Jean K. Millet, Gary Whittaker. In vitro and computational analysis of the putative furin cleavage site (RRARS) in the divergent spike protein of the rodent coronavirus AcCoV-JC34 (sub-genus luchacovirus). 2024. hal-04494649

HAL Id: hal-04494649

<https://hal.uvsq.fr/hal-04494649>

Preprint submitted on 7 Mar 2024

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

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

In vitro and computational analysis of the putative furin cleavage site (RRARS) in the divergent spike protein of the rodent coronavirus AcCoV-JC34 (sub-genus luchacovirus)

Annette Choi ^{1*}, Deandria T. Singleton ^{1*}, Alison E. Stout ¹, Jean K. Millet ²
and Gary R. Whittaker ^{1, 3#}

Departments of Microbiology & Immunology ¹ and Public & Ecosystem Health ³,
College of Veterinary Medicine, Cornell University, Ithaca, NY, 14853, USA.

Université Paris-Saclay, INRAE, UVSQ, Virologie et Immunologie Moléculaires, Jouy-
en-Josas, France ²

- corresponding author

grw7@cornell.edu

Abstract

The *Coronaviridae* is a highly diverse virus family, with reservoir hosts in a variety of wildlife species that encompass bats, birds and small mammals, including rodents. Within the taxonomic group alphacoronavirus, certain sub-genera (including the luchacoviruses) have phylogenetically distinct spike proteins, which remain essentially uncharacterized. Using in vitro and computational techniques, we analyzed the spike protein of the rodent coronavirus AcCoV-JC34 from the sub-genus luchacovirus, previously identified in *Apodemus chevrieri* (Chevrier's field mouse). We show that AcCoV-JC34—unlike the other luchacoviruses—has a putative furin cleavage site (FCS) within its spike S1 domain, close to the S1/S2 interface. The pattern of basic amino acids within the AcCoV-JC34 FCS (-RR-R-) is identical to that found in “pre-variant” SARS-CoV-2—which is in itself atypical for an FCS, and suboptimal for furin cleavage. Our analysis shows that, while containing an -RR-R- motif, the AcCoV-JC34 spike “FCS” is not cleaved by furin (unlike for SARS-CoV-2), suggesting the possible presence of a progenitor sequence for viral emergence from a distinct wildlife host.

Introduction

The animal reservoirs for pandemic potential viruses (including coronaviruses) are focused on the breadth of bat species (order Chiroptera) that exist around the world [1-3]. However certain coronaviruses, notably the sub-genus embecovirus (genus betacoronavirus) currently have no bat-origin examples and have a putative reservoir in animal species within the order Rodentia, which is the most diverse mammalian order on the planet and is well-documented as an important reservoir host for human diseases [4, 5].

While rodents are generally appreciated as an important reservoir for RNA viruses, surveillance and detection of coronaviruses is currently relatively limited. Following the initial discovery of what is now the prototype luchacovirus (Lucheng Rn rat coronavirus, or LRNV), along with two *Betacoronavirus* species [6], a study from Ge *et al.* examined 177 intestinal samples from three species of rodents in Yunnan Province, China and detected both alphacoronaviruses and betacoronviruses in three animal species (*Apodemus chevrieri*, *Eothenomys fidelis* and *Apodemis ilex*) [7]. Their study reported the full-length genome of a coronavirus (AcCoV-JC34) from *A. chevrieri* (Chevrier's field mouse) that was designated an alphacoronavirus (sub-genus luchacovirus) based on its genome structure and multiple sequence alignments, which included analysis of the whole genome and the ORF1a/b genes. However, Ge *et al.* noted that both AcCoV-JC34 and LRNV may represent a novel alphacoronavirus species. In particular, they noted that the luchacovirus S gene formed a distinct genetic lineage with low sequence identity (<25%) compared to other well characterized coronaviruses. Ge *et al.* also noted that AcCoV-JC34 S contained two predicted

proteolytic cleavage sites, one at residue 508 at the S1/S2 interface, and the other at residue 674 (the fusion peptide-proximal S2' position).

More recently, a more comprehensive sampling of rodents and other small mammals has identified a diverse range of coronaviruses in such animal reservoirs [8]. To determine the evolutionary history of rodent alphacoronaviruses in more detail, Tsoleridis *et al.* also reported sequence data from viruses sampled from European rodents, to define a single common ancestor for all rodent alphacoronaviruses with a shared recombinant betacoronavirus spike gene—also shared with batCoV HKU2, swine acute diarrhea syndrome (SADS) coronavirus and two shrew coronaviruses [9]. According to Tsoleridis *et al.*, the luchacoviruses (including AcCoV-JC34) comprised a distinct lineage within the “recombinant” viruses. In summary, it can be argued that coronaviruses of small mammals, including rodents, are still poorly understood.

We have previously reported that rodent coronavirus AcCoV-JC34 has a weakly predicted furin cleavage site (FCS) in its spike protein [10]. Here, we further analyze the AcCoV-JC34 spike and its “FCS” along with the other luchacoviruses, taking an *in vitro* and computational perspective.

Results

Phylogenetic analysis of luchacoviruses

To understand the relationship of AcCoV-JC34 and the other known luchacoviruses, we first constructed a phylogenetic tree of these viruses in comparison to representatives of the diverse coronavirus family, based on spike protein sequences (Figure 1). In agreement with Ge *et al.*,

luchacoviruses formed a monophyletic group with 100% bootstrap support, indicating a common ancestor origin outside of the established alphacoronavirus branch. Luchacoviruses clustered with rhinacoviruses, which include swine acute diarrhea syndrome coronavirus, Rhinolophus bat coronavirus HKU2, and porcine enteric alpha coronavirus (Figure 1).

Geographical distribution of sampled luchacoviruses

The geographical location, dates and rodent species sampled for the currently identified luchacoviruses are summarized in Figure 2 and Table 1. The luchacoviruses sampled to date are from a range of rodent hosts and are from the United Kingdom and several provinces in China (Figure 2), indicating a widespread distribution. Despite being sampled in these distinct locations, as mentioned above, luchacoviruses form a monophyletic group suggesting they have been associated with rodents for an extended period of time.

Multiple sequence alignment and structural analysis of AcCoV-JC34 spike

A multiple sequence alignment of spike proteins was performed on AcCoV-JC34 spike in comparison to the prototype luchacovirus Lucheng Rn rat CoV (LRNV), as well as SARS-CoV-2, SARS-CoV, HCoV-HKU1, HCoV-OC43 and MERS-CoV. This alignment revealed that the -RR-R- motif present in AcCoV-JC34 does not align precisely with the S1/S2 motif of most coronavirus spikes (Figure 3). However, it aligned with a potential secondary MERS-CoV furin cleavage site (RSTRS).

To investigate the structural location of AcCoV-JC34 furin cleavage site, the AcCoV-JC34 spike protein structure was structurally modeled (Figure 4). We used SADS-CoV spike for our modeling due to its available structure in the RCSB protein data bank and relatively high

identity with JC34 (41.5%). In our JC34 model, the potential furin cleavage site (-RR-R-) is located in an exposed loop of the protein which is predicted to increase its accessibility to proteases. However, the potential AcCoV-JC34 furin cleavage site was within a loop upstream of the typical S1/S2 furin cleavage site found in other CoVs (see Figure 3). In SARS-CoV-2, this upstream region aligned with a DQLTP sequence upstream of the expected S1/S2 cleavage site.

Bioinformatic and biochemical analysis of potential AcCoV-JC34 spike cleavage site

To determine whether furin processes the -RR-R- motif in AcCoV-JC34, we first utilized the PiTou and ProP furin cleavage prediction tools (Figure 5). A positive score for Pitou or a score above 0.5 for ProP indicates the likelihood of furin cleavage. AcCoV-JC34 has a weakly predicted furin cleavage site based on the PiTou score (see also [10]). Although bioinformatic tools are useful for prediction, these may not represent biologically relevant cleavage events, which need to be addressed experimentally.

To directly test whether furin cleaves this site *in vitro*, we performed peptide cleavage assays using furin, along with trypsin as a control. The peptide sequences used were TFMTKARARTTF (Lucheng Rn rat CoV, LRNV), TFSRRRARARTL (AcCoV-JC34), and TNSPRRARSVA (SARS-CoV-2). Trypsin cleaved all three peptides with varying efficiency. Furin, as expected from previous studies, cleaved the SARS-CoV-2 peptide; however, it did not cleave the LRNV or JC34 peptides (Figure 6). These data indicate that although AcCoV-JC34 has a minimal furin cleavage sequence (R-X-X-R) it is not able to be cleaved by furin when tested experimentally.

Discussion

The “furin cleavage site” or FCS of SARS-CoV-2 has been at the center of the many discussions on the origin of the COVID-19 pandemic; see [11] for a recent summary. Despite being interpreted as “highly unusual”, an FCS is—to the contrary—very common among the *Coronaviridae* [12], with sarbecoviruses and most alphacoronaviruses being the exception rather than the rule in lacking this important regulatory sequence. In fact, many zoonotic coronaviruses and those in reservoir hosts appear to contain sequences and structural loops at the S1/S2 interface that are sub-threshold for furin-mediated cleavage [13-16] and may be “poised” for spillover events. Examples include “pre-variant” SARS-CoV-2, as well as the sarbecoviruses RmYN02, RacCS203, BANAL-20-116, BANAL-20-246 that have potential phylogenetic homology to the SARS-CoV-2 FCS [17]—and may include the luchacovirus AcCoV-JC34 analyzed here. It is noteworthy that AcCoV-JC34 is the only luchacovirus containing this -R-RR- motif.

While containing an -RR-R- motif, as found in SARS-CoV-2, the data presented here show that this AcCoV-JC34 sequence is not cleaved by furin. The reasons for this are currently unclear. One possibility is that the upstream proline found in SARS-CoV-2, as well as in other spike cleavage site sequences, may promote cleavage by creating a structural turn beneficial for furin activity. It is also possible that the additional downstream arginine residue in AcCoV-JC34 spike may be inhibitory for the tight active site binding pocket present in furin [18]. Alternatively, the structural loop present in AcCoV-JC34 spike may be cleaved by other proprotein convertases of the furin family that have less stringent cleavage requirements, or

by trypsin-like enzymes or cathepsins. Notably, the -RR-R- motif is rare in furin substrates, and only other known example of this sequence motif in FurinDB (a database of furin substrates) is found in proaerolysin, a bacterial toxin [19].

One notable aspect of the -RR-R- motif in AcCoV-JC34 is that it does not align precisely with the S1/S2 motif of most coronavirus spikes (see Figure 3) and is a structurally exposed location above the typical S1/S2 loop (see Figure 4). Analysis of the MERS-CoV spike also shows an additional putative FCS in the MERS-CoV spike (SRSTRS); while this contains a minimal furin motif this sequence shows low scores for furin cleavage with both Pitou and ProP, and FRET-based peptides were not cleaved by furin in biochemical cleavage assays—in contrast to the PRSVRS motif at the expected S1/S2 junction (J. K. Millet, unpublished results). Nevertheless, it is possible that, as with AcCoV-Jc34, this “secondary” MERS-CoV sequence comprises a “blocked” FCS due to flanking hydrophobic and charged residues in the downstream C-terminal positions (i.e., SRSTRSMLKRRDS). This putative secondary cleavage site also lacks an upstream proline/proline-rich region, as with many other S1/S2 regions that are known to be cleaved by furin.

For SARS-CoV-2, it is clear that selection is occurring to up-regulate the spike FCS, as seen with several of the highly transmissible variants that have emerged [20-24]. The FCS can also be readily down regulated upon Vero cell adaptation; for examples see refs [25, 26]. Likewise, some coronaviruses in animal reservoirs may be “poised” for proteolytic cleavage-activation at S1/S2, with selection occurring along with modifications to their receptor binding domain. One interesting example of this may exemplified by the MERS-like bat-CoVs HKU-4 and

HKU-5, with HKU-4 binding human DPP4, but having no identifiable FCS, and with HKU-5 not able to bind hDPP4 and having a robust FCS [27].

Our studies highlight the possible presence of a distinct proteolytic cleavage loop in the coronavirus spike protein and the specific features of the luchacovirus spike—which along with that found in the rhinacoviruses (e.g., SADS-CoV) appears to represent an evolutionary disparate spike protein with apparent similarities to a betacoronavirus spike protein (see Figure 1), despite the taxonomic designation of these viruses as alphacoronaviruses.

References

1. Ruiz-Aravena, M., et al., Ecology, evolution and spillover of coronaviruses from bats. *Nature Reviews Microbiology*, 2021.
2. Plowright, R.K., et al., Pathways to zoonotic spillover. *Nat Rev Microbiol*, 2017. 15(8): p. 502-510.
3. Stout, A.E., et al., Viral and Host Attributes Underlying the Origins of Zoonotic Coronaviruses in Bats. *Comparative Medicine*, 2021. 71: p. 442-450.
4. Luis, A.D., et al., A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? *Proc Biol Sci*, 2013. 280(1756): p. 20122753.
5. Keeling, M.J. and C.A. Gilligan, Metapopulation dynamics of bubonic plague. *Nature*, 2000. 407(6806): p. 903-906.
6. Wang, W., et al., Discovery, diversity and evolution of novel coronaviruses sampled from rodents in China. *Virology*, 2015. 474: p. 19-27.
7. Ge, X.Y., et al., Detection of alpha- and betacoronaviruses in rodents from Yunnan, China. *Virol J*, 2017. 14(1): p. 98.
8. Wu, Z., et al., Comparative analysis of rodent and small mammal viromes to better understand the wildlife origin of emerging infectious diseases. *Microbiome*, 2018. 6(1): p. 178.
9. Tsoleridis, T., et al., Shared Common Ancestry of Rodent Alphacoronaviruses Sampled Globally. *Viruses*, 2019. 11(2).
10. Stout, A.E., et al., Furin cleavage sites in the spike proteins of bat and rodent coronaviruses: Implications for virus evolution and zoonotic transfer from rodent species. *One Health*, 2021. 13: p. 100282.
11. Chan, Y.A. and S.H. Zhan, The emergence of the spike furin cleavage site in SARS-CoV-2. *Molecular Biology and Evolution*, 2021.
12. Whittaker, G.R., SARS-CoV-2 spike and its adaptable furin cleavage site. *Lancet Microbe*, 2021. 2(10): p. e488-e489.
13. Jaimes, J.A., et al., Phylogenetic Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop. *J Mol Biol*, 2020. 432(10): p. 3309-3325.
14. Jaimes, J.A., J.K. Millet, and G.R. Whittaker, Proteolytic Cleavage of the SARS-CoV-2 Spike Protein and the Role of the Novel S1/S2 Site. *iScience*, 2020. 23(6): p. 101212.
15. Tang, T., et al., Proteolytic Activation of SARS-CoV-2 Spike at the S1/S2 Boundary: Potential Role of Proteases beyond Furin. *ACS Infect Dis*, 2021. 7(2): p. 264-272.
16. Millet, J.K. and G.R. Whittaker, Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. *Proc Natl Acad Sci U S A*, 2014. 111(42): p. 15214-9.

17. Lytras, S. <https://virological.org/t/the-sarbecovirus-origin-of-sars-cov-2-s-furin-cleavage-site/536>.
18. Henrich, S., et al., The crystal structure of the proprotein processing proteinase furin explains its stringent specificity. *Nat Struct Biol*, 2003. 10(7): p. 520-6.
19. Abrami, L., et al., The pore-forming toxin proaerolysin is activated by furin. *J Biol Chem*, 1998. 273(49): p. 32656-61.
20. Bushman, M., et al., Population impact of SARS-CoV-2 variants with enhanced transmissibility and/or partial immune escape. *Cell*, 2021.
21. Lubinski, B., et al., Spike protein cleavage-activation mediated by the SARS-CoV-2 P681R mutation: a case-study from its first appearance in variant of interest (VOI) A.23.1 identified in Uganda. *bioRxiv*, 2021: p. 2021.06.30.450632.
22. Lubinski, B., et al., Functional evaluation of the P681H mutation on the proteolytic activation the SARS-CoV-2 variant B.1.1.7 (Alpha) spike. *iScience* 2021 <https://doi.org/10.1016/j.isci.2021.103589>
23. Grubaugh, N.D., et al., Public health actions to control new SARS-CoV-2 variants. *Cell*, 2021. 184(5): p. 1127-1132.
24. Luring, A.S. and E.B. Hodcroft, Genetic Variants of SARS-CoV-2-What Do They Mean? *JAMA*, 2021. 325(6): p. 529-531.
25. Sasaki, M., et al., SARS-CoV-2 variants with mutations at the S1/S2 cleavage site are generated in vitro during propagation in TMPRSS2-deficient cells. *PLoS Pathog*, 2021. 17(1): p. e1009233.
26. Lamers, M.M., et al., Human airway cells prevent SARS-CoV-2 multibasic cleavage site cell culture adaptation. *Elife*, 2021. 10.
27. Yang, Y., et al., Receptor usage and cell entry of bat coronavirus HKU4 provide insight into bat-to-human transmission of MERS coronavirus. *Proc Natl Acad Sci U S A*, 2014. 111(34): p. 12516-21.

Methods

Furin prediction calculations. Prop: CoV sequences were analyzed using the ProP 1.0 Server hosted at: cbs.dtu.dk/services/ProP/. PiTou: CoV sequences were analyzed using the PiTou V3 software hosted at: <http://www.nuolan.net/reference.html>.

Amino acid alignments and phylogenetic trees

Multiple sequence alignment was performed on coronavirus spike protein using Geneious Prime ® (v.2019.2.3. Biomatters Ltd.). A maximum likelihood phylogenetic tree was constructed using MegaX, 100 boot strap replicates based on the spike protein. Amino acid sequences of S were obtained from NCBI GenBank. Accession numbers are: AcCoV-Jc34 (YP_009380521), Asian leopard cat CoV (EF584908.1), Bat_Hp/Zhejiang2013(YP_009072440), Bat-Rm/Yunnan/YN02/2019 (QPD89843.1), Bat-SL-CoV_ZC45 (AVP78031.1), BCoV (P15777), Bottlenose dolphin CoV-HKU22 (AHB63508), BtRf-AlphaCoV/YN2012 (YP_009200735), CCoV (AY436637.1), ECoV-NC99 (AAQ67205.1), FCoV-Black (EU186072.1), Ferret-CoV (NC_030292.1), FIPV 79-1146 (DQ010921.1), HCoV-229E (NC_002645.1), HCoV-HKU1 (NC_006577), HCoV-NL63 (NC_005831.2), HCoVOC43 (NC_006213.1), HeCoV (MK679660.1), HKU4 (YP_001039953), HKU5 (YP_001039962), HKU23 (QEY10673), HKU24 (QOE77327), IBV (NC_001451.1), Longquan R1 rat CoV (QOE77336.1), Lucheng Rn rat CoV (QOE77268.1), MERS-CoV (AFS88936.1), MHV-1 (ACN89742), PDCoV (MN942260.1), PEDV (NC_003436.1), PHEV (QTF73995.1), Porcine enteric alphacoronavirus GDS04 (ASK51717.1), Rabbit CoV-HKU14 (AFE48827), RaTG13 (QHR63300), Rhinolophus bat coronavirus HKU2 (YP_001552236.1), Rhinolophus bat CoV-BTKY72 (APO40579.1), Rhinolophus bat CoV-HKU32 (QCX35178), Rhinolophus bat CoV-HKU2 (YP_001552236.1), Rousett bat CoV-229E related (QHA24665), Rousettus bat CoV-GCCDC1 (QKF94914), RtClan-CoV/GZ2015 (), RtMurf-CoV-1/JL 2014 (ATP66738), RtRl-CoV/FJ2015 (KY370050), SARS-CoV (AAT74874.1), SARS-CoV-2 Wuhan-Hu-1 (YP_009724390.1), Sc-

BatCoV-512 (ABG47078), Swine acute diarrhea syndrome coronavirus(AVM41569.1), Swine acute diarrhea syndrome related coronavirus (AVM80500.1), TGEV(P07946), Turkey-CoV (QRR19172), UkMa1(QBG64648), UKRn3(QBG64657).

Spike structural modelling

Pairwise amino acid alignment between AcCoV-Jc34 (YP_009380521) and SADS-CoV (AVM80500) was performed using Geneious Prime ® (v.2019.2.3. Biomatters Ltd.). S protein models were built based on the SADS-CoV structure obtained from RCSB (PDB: 6M39), using UCSF Chimera (v.1.14, University of California) through the modeler homology tool of the Modeller extension (v.9.23, University of California).

Fluorogenic peptide cleavage assays

Fluorogenic peptide cleavage assays were performed as described previously [14]. Each reaction was performed in a 100 μ L volume consisting of buffer, protease, and AcCoV-Jc34 (TFSRRARARTL) or Lucheng Rn rat CoV (TFMTKARARTTF) or SARS-CoV-2 S1/S2 WT (TNSPRRARSVA) fluorogenic peptide in an opaque 96-well plate. For trypsin catalyzed reactions, 0.8 nM/well TPCK trypsin was diluted in PBS buffer. For furin catalyzed reactions, 1 U/well recombinant furin was diluted in buffer consisting of 20 mM HEPES, 0.2 mM CaCl₂, and 0.2 mM β -mercaptoethanol, at pH 7.0. Fluorescence emission was measured once per minute for 60 minutes using a SpectraMax fluorometer (Molecular Devices) at 30 °C with an excitation wavelength of 330 nm and an emission wavelength of 390 nm. V_{max} was calculated by fitting the linear rise in fluorescence to the equation of a line.

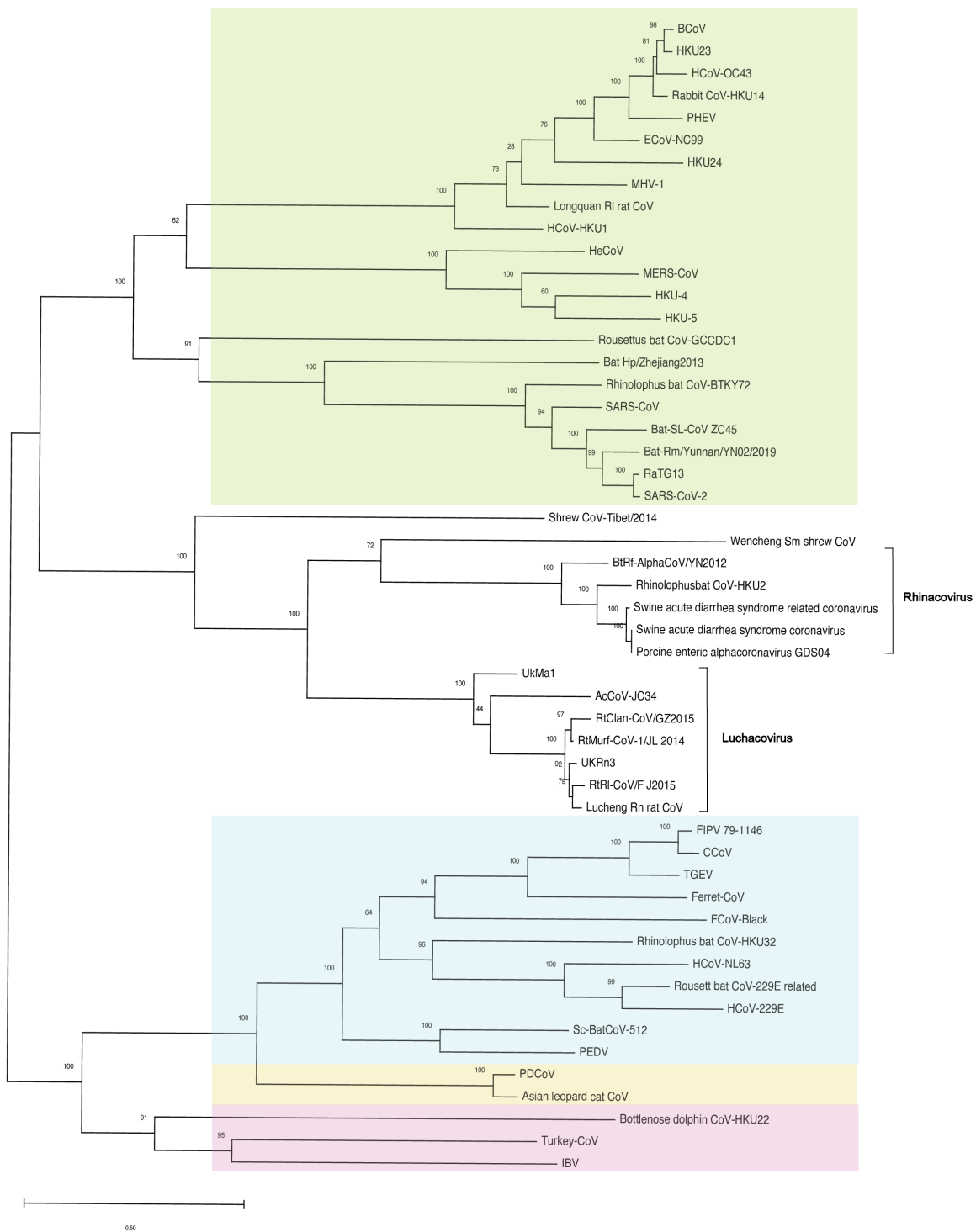


Figure 1. Phylogenetic tree of spike protein sequences.

The maximum likelihood phylogenetic tree was constructed using MegaX, 100 bootstraps, from a multiple sequence alignment of the spike sequences. Betacoronavirus spikes are shaded green, alphacoronavirus spikes are shaded blue, deltacoronavirus spikes are shaded yellow and gammacoronavirus spikes are shaded pink



Figure 2. Geographical distribution of currently identified luchacoviruses.

Luchacoviruses have been identified from surveillance studies in United Kingdom (East Midlands region) and China (Yunnan, Zhejiang, Fujian, Jilin, and Guizhou provinces).

Luchacovirus	Rodent species	Location	Time period
			2011.8
AcCoV-JC34	<i>Apodemus chevrieri</i>	China (Yuanna)	
			2011-2013
Lucheng Rn rat CoV	<i>Rattus norvegicus</i>	China (Lucheng)	
			2015.5
RtRI-CoV/FJ2015	<i>Rattus losea</i>	China (Fujian)	
			2014.8
RtMruf-CoV-1/JL2014	<i>Myodes rufocanus</i>	China (Jilin)	
			2015.12
RtClan-CoV/GZ2015	<i>Eothenomys melanogaster</i>	China (Guizhou)	
			2008-2015
UKRn3	<i>Rattus norvegicus</i>	United Kingdom	
			2008-2015
UKMa1	<i>Microtus agrestis</i>	United Kingdom	

Table 1. Rodent luchacoviruses identified from surveillance studies.

Figure 3

AcCoV-JC34	-----MALIFVLMILITLYRCPFVLCNFCVQCTDQLRQQ-----	EVYL-PNAVS	41
Lucheng Rn rat CoV	-----MYFFLLLLFVSADAAIQTCPAPGNVNLDISK-----	LYYG-TQASI	40
SARS-CoV-2	-----MFVFLVLLPLVSSQCVNLTTRTQ--LPPAYTN-----	SFTRGVVYP-DKVFR	44
SARS-CoV	-----MAIFLLFLTLTSGSDLRCTTFDDVQAPNYTQ-----	HTSSMRGVVYP-DEIFR	48
HKU 1	-----MLLIIFILPTTLAVIGDFNCTNFAINDLNTTV-----	PRISEYVVDVSYGLGTYIILDRVYL	57
OC43	----MFLILLISLPTAFAVIGDLNCTLDPRCLKGSFNN--	RDTGPPSISIDTVDVNTGLGTYIVLDRVYL	63
MERS-CoV	MIHSVFLMLFLTPTESYVDVGPDSVKSACIEVDIQQTFFDKTWPR----	PIDVSKADGIIYPQGRTRY	65
AcCoV-JC34	NATYIATA--VFPT-----RHWHCSSSNSTPGDNKFNGL-----	GVFVHRFNHPNW-----	86
Lucheng Rn rat CoV	NATFQVVQ--VLPQ-----VPWKCNSYSNG--PSNKFNGI-----	GVFVDLASAQHS-----	83
SARS-CoV-2	SSVLHSTQDLFLPF---FSNVTWFHAIHVSNGTNGTKRFDNP---	VLPFNDGVYFASTEKSN-----	100
SARS-CoV	SDTLYLTDLFLPF---YSNVTGFHTINHT-----FDNP---	VIPFKDGIYFAATEKSNV-----	97
HKU 1	NTTILFTG--YFPK---SGANFRDLSLKGTTYLSTLWYQKPF---	LSDFNNGIFSRVKNTKLYVNKTYL	118
OC43	NTTLFLNG--YYPT---SGSTYRNMALKGTDLSTLWFKPPF---	LSDFINGIFAKVKNTKVKFDGVMY	124
MERS-CoV	NITITYQG--LFPYQGDHGMVYVSAGHATGTTQKLFVANYSQDVKQFANGFVVRIGAAANSTGTVII		132
AcCoV-JC34	-----WHFAAYPAAPTNTK-----WILFWWHR-----	ATQAS-----TVFQV	118
Lucheng Rn rat CoV	-----WHLFVYPSMPTNKT-----WILSWADT-----	HTSFEHGSVISYVQI	120
SARS-CoV-2)	-----IRGWIFGTTLDLSDKT-----QSLLVNN-----	ATNVV--IKVCEFQF	135
SARS-CoV	-----VRGWVFGSTMNKS-----QSVIIINN-----	STNVV--IRACNFEL	132
HKU 1	SE-----FSTIVIGSVFINNS-----YTIVVQPH-----	NGVLE--ITACQYTM	155
OC43	SE-----FPAITIGSTFVNTS-----YSVVVQPRINSTQDGVNKLQGLLE--	VSVCQYNM	173
MERS-CoV	SPSTSATIRKIYPAFMLGSSVGNFSDGKMGRRFNHTLVLLP-----	DGCGT--LLRAFYCI	186
AcCoV-JC34	CKYQQPAVDVLTSGFR---CE----APTNPCCSA--IVNMAI-----	ECLINSTFRPTTFSF-	166
Lucheng Rn rat CoV	CKYPSNVVTDITNGNG---CHTNMAGPGATTCDV--ILSSPL-----	ECVLNRTYSQQYAGV-	172
SARS-CoV-2	CNDPFLGVYYHKNNKSWMESEFRVYS-SANNCTF-EYVSQLMDLEGKQGNFKNREFVFKNIDG---		199
SARS-CoV	CDNPFFAVSKPMGTQ---THTMIFD-NAFNCTF-EYISDAFSLDVSEKSGNFKHLREFVFKNKDGF-		194
HKU 1	CEYPHTICKSK-GSSR---NESWHFDKSEPLCLF-----	KKNFYTNVSTDWL-	198
OC43	CEYPHTICHPNLGNHF---KELWHYDYGVVSLY-----	KRNFTYDVNATYL-	217
MERS-CoV	LE-PRSGNHCPAGNSY---TSFATYHTPATDCSDGNYNRNASLNSFKEYFNLRNCTFMITYNITEDEIL		251
AcCoV-JC34	SYVTWYNSEINAVIH-----GQRFSE--VYEGFLWSNASAFCYDA-----	NGCI	208
Lucheng Rn rat CoV	SYITWYNDHIIASIQ-----GEVFTF-DIGEVLQWSNFSAFCTG-----	NKCG	215
SARS-CoV-2	-YFKIYSKHTPINLV--RDLPGQFSALEPLVDLPIGINITRFQTLALHRSYLTGPDSSSGWTAGAAA		265
SARS-CoV	-YV--YKGYQPIDV--RDLPSGFNTLKPIFKPLGINITNFRAILTAFSPAQDT-----	WGTSAAAY	252
HKU 1	-YHFHYQERGTFFAY---YADSGMPTTFLF-SLYLGTLLSHYYVPLTNCNAISSNTDN---E---	TLQY	256
OC43	-YHFHYQEGGTFFAY---FTDTGFVTKFLF-NVYLGMLSHYYVMP LTC--IRRPKDG---	F---SLEY	273
MERS-CoV	EFWGITQTAQGVHLFSSRYVDLYGGNMFQFATLPVYDTIKYYSIIPHSIRSIQSDRKA---W---A-AF		313
AcCoV-JC34	FNIPNTTSDWLVDSTSSGAINNFVDCGFDYESQLKCKNLVFELEPAVYHGAALPIE--SSVYYVANDLA		275
Lucheng Rn rat CoV	FSYATTLSEWLVRTDNDGTVIDYVICDTEFESQLKCKNMVFELTPAVYSGSAVELQ--SAIYYVSNELP		282
SARS-CoV-2	YVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLSFTVEKGIYQTSNFRVQPTESIV-RFPNIT		333
SARS-CoV	FVGYLKPTTFMLKYDENGITITDAVDCSQNPLAELKCSVKSFEIDKGIYQTSNFRVVPSPGDVV-RFPNIT		320
HKU 1	WVTPLSKRQYLLKFDNRGVITNAVDCSSSFFSEIQCKTKSLLPNTGVYDLSGFTVKVPATVHRRIPDLP		325
OC43	WVTPLTPRQYLLAFNQDGIIFNAVDCMSDFMSEIKCKTQSIAPPTGVYELNGYTVQPVADVRRKPDLP		342
MERS-CoV	YVYKQLPTFLLDVSDGYIRRAIDCGFNDLSQLHCSYESFDVESGVYVSSSFEAKPSGSVQAEGVE		382
AcCoV-JC34	D-CAFS--FADIFSDGTGNYGGLRRHVFTNCWVNYTSW-----	FLCDDGLACIIFNAIFSEVV---	330
Lucheng Rn rat CoV	D-CDFS--FADMFDGTGNFEGLRHVFSNCWVNYTAW-----	FACADDYSCIIIFNAIFAVER---	337
SARS-CoV-2	NLCPFGEVFNATRFASVYAWN---RKRSNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADS		399
SARS-CoV	NLCPFGEVFNATKFPVSYAVE---RKKISNCVADYSVLYNSTFFSTFKCYGVSATKLNLCFNSVYADS		386
HKU 1	D-CDIDKWLNNFNVPSPLNWE---RKIFSNCNFLLTLRLVHTDSFCNNFDESKIYGCFCFSIVLKD		390
OC43	N-CNIEAWLNDKSVSPLNWE---RKTFSNCNFMSLSMFIQADSFCTCNNIDAAKIYGMCFSSITIDK		407
MERS-CoV	--CDFSPLLSGTP-PQVYNFK---RLVFTNCNLYNLTKLLSLFSVNDFTCSQISPAAIASNCYSSLILDY		445
AcCoV-JC34	-----YNLTQPDGLINPFLRCNGLDVYTVKGCES-AGYVLRYLQFDSDG-----		372
Lucheng Rn rat CoV	-----YKLSQPDGLVNPFIKCNGLDLYSITKGCES-SGFVLRYLQYANG-----		379
SARS-CoV-2	FVIRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVG-----	NYNYLYRLFRRKS-----	459
SARS-CoV	FVVKGDVDRQIAPGQTGVIADYNYKLPDDFMGCVLAWNTRNIDATSTG----	NYNYKYRYLKHG-----	446
HKU 1	FAIPNSRRSDLQLGSSGFLQSSNYKIDTSSSSCQLYYSLPAINVTINNYNPSSWNRYYGFNNFN----		454
OC43	FAIPNRRKVDLQLGNLGYLQSSNYRIDTTATSCQLYNYLPAANVSVSRFNPSTWNRKRFIEDSVFVQP		476
MERS-CoV	FSYPLSMKSDLVSSAGPISQFNKQSFNPSTCLILATVPHNLTITK-----	PLKYSYINKC-----	503

AcCoV-JC34 -----QFNPDAYTPDYM--EC----- 386
 Lucheng Rn rat CoV -----SFDVNSYTPDYM--EC----- 393
 SARS-CoV-2 ---NLKPFERDISTEIQAGST--PC----- 480
 SARS-CoV ---KLRPFERDISNVFSPDGK--PC----- 467
 HKU 1 ----LSSHVVYSRYCFVSNNTFCPCAKPSFASSCKSHKPPSASCPIGTNYRSCESTTVLDHTDWCRCSS 519
 OC43 PTGVFTNHSHVVYAQHCFKAPKNFCPC-----SSCPGKNGIGTGPAGTNSLTCNDLCTLD----- 531
 MERS-CoV -SRLSDDRTEVPQLVNAVQYS--PC-----VS 528

AcCoV-JC34 -----FGYFQI-----DRSYIVYNAKFIT----- 405
 Lucheng Rn rat CoV -----FGYFAL-----YNGYVIYNAKFVS----- 412
 SARS-CoV-2 ---NGVEGFNCYF-----PLQS-----YGFQPTN-----GVGYQPYRVVVL----- 514
 SARS-CoV ---TPPALNCYW-----PLND-----YGFYTTT-----GIGYQPYRVVVL----- 500
 HKU 1 CLPDPITAYDPRSCSQKSLVGVGEHCAGFGVDEEKCGVLDGSYNVSCLCSTDAFLGWSYDTCVSNNRC 588
 OC43 --PITLKAPDTYKCPQSKSLVIGIGEHC SGLAVKSDYCG-----NNSCTCQPQAF LGWSAD SCLQGD KC 592
 MERS-CoV IVPSTVWEDGDYRQLSPLEG-----GGWLVAS-----GSTVAMTEQLQMG----- 570

AcCoV-JC34 -----EGLTVCVLQPVEPE-----LGVCKQYIDGVTFQGI LHTSNAN-ITTFHNL LLY--YG 454
 Lucheng Rn rat CoV -----KGLTVCVVQPVEPE-----LDVCKSYTIDGVTFQGI LRNTTHQ-IDSFHNILY--YG 461
 SARS-CoV-2 --FELLHAPATVCGPKKSTNLVKNK-----CVNFNFNGLTGTGVLTESNKK-FLPFQFGRDIAD 571
 SARS-CoV --FELLNAPATVCGPKLSTDLIKNQ-----CVNFNFNGLTGTGVLTPSSKR-FQPFQFGRDVSD 557
 HKU 1 NIFSNFILNGINSGTCSNDLLQPNTTEV--FTDVCVDYDLYGITGQGFKEVSAVYNSWQNL LLYDSNG 655
 OC43 NIFANFILHVDVNNGLTCSTDLQKANTEI--ELGVCVNYDLYGISGQGFVEVNATYNSWQNL LLYDSNG 659
 MERS-CoV --FGITVQYGTDTNSVCPKLEFANDTKIASQLGNCVEYSLYGVSGRGVFNQCTAV-GVRQRFVYDAYQ 636

(JC34 -RR-R- motif)

AcCoV-JC34 DMVSH--VRIKGI VYSVEPCNSFYYSV----LKTRSEVGYLYSGANCNSSDVVT--FSRRARARTLTD S 515
 Lucheng Rn rat CoV DMVSY--VRIRGVVYAVESCNRFYYSV----FKTL SAIGLYSGATCDSTDVTT--FMTKAKATTFVDS 522
 SARS-CoV-2 TDAVRDPQTLEI-LDITPCSFGGVSVITPGTNTSNQAVLYQDVNCTEVPVAI--HADQLTPTWRVYS 637
 SARS-CoV FTDSVRDPKTSEI-LDITPCSFGGVSVITPGTNASSEVAVLYQDVNCTDVSTAI--HADQLTPAWRIYS 623
 HKU 1 NIIGFKDFVTNKT-YNIFPCYAGRVSAAF--HQNASL LALLYRNLKCSYV LNNI-----SLTTQPYFDS 716
 OC43 NLYGFRDYITNRT-FMIHSCYSGRVSAAY--HANSSEPA L LFRNIKCNYVFNNS--LTRQLQPINYSFD 723
 MERS-CoV NLVGY--YSDDGNYCYLRACVSVPVSVIY--DKETKTHATLFGSVACEHISSTMSQYSRSTRSMLKRRD 701

S1 / S2

AcCoV-JC34 S-----LGCLVDVQLTNNN-YTTC SYPIGNGLCADVNVTG-----MPVVGNIYI-----QPH 561
 Lucheng Rn rat CoV S-----LGC FIDVSVTDGN-YTDCLNPIGNGFCVDVNVTG-----QPVVGNIFI-----QTH 568
 SARS-CoV-2 TGSNVFQTRAGCLIGAEHVNN--SYECDIPIGAGICASYQTQT-NSPRRARSVASQSIIAYTMSLGAEN 703
 SARS-CoV TGNNVFQTAGCLIGAEHVD--SYECDIPIGAGICASYHTVSL-----RSTSQKSIVAYTMSLGAEN 685
 HKU 1 Y-----LGC VFNADNLTDYSVSSCALRMGSGFCVDYNSPSSSSSRKRRSISASRYFVTF---EPF 774
 OC43 S-----YLGCVVNAYNSTAISVQTCDLTVGSGYCVDYFKNR---RSRRRAITTYRFTNF---EPF 777
 MERS-CoV STYGPLQTPVGCVLGLVNSSLF-VEDCKLPLGQSLCALPDTPTLTPRSVRSVPGEMRLASIAFNHPI- 768

AcCoV-JC34 DTDYARPILSPQ----IVSLPLDHVINVKEQFVQTSAPKFDVDCERYICDVS VQCKELLAKYGGYCPKI 626
 Lucheng Rn rat CoV DTDYARPILTAQ----QIELPIDHYVSVKEQFIQTSTPKFDVDCERYICDVS SDCRELLVKYGGYCSKI 633
 SARS-CoV-2 SVAYSNN S-----IAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECANLLLQYGSFCTQL 763
 SARS-CoV SIAYSNN T-----IAIPTNFSISITTEVMPVSMKTSVDCNMYICGDSTECANLLLQYGSFCTQL 745
 HKU 1 NVS FVNDSIESVGGLYEIKIPTNFTIVGQEEFIQTNSPKVTIDCSLFVCSNYAACHDLLSEYGTFCDNI 843
 OC43 TVNSVND SLEPVGGLYEIQIPSEFTIGNMEEFIQTSSPKVTIDCAAFVCGDYAACKLQLVEYGSFCDNI 846
 MERS-CoV QVDQLNSS-----YFKLSIPTNFSFGVTQEYIQTTIQKVTVDCQYVCGNFQKCEQLLREYGFCSKI 831

S2' (FP)

AcCoV-JC34 VSDIKSSSLQLDYQVQGIYKTLNVDVKVPDV-DFG-AFNF-----SMFESEPNGRSFIEDIL 681
 Lucheng Rn rat CoV LADIKSSSIQLDYQILGLYKTLAVDFKVPDI-DFG-DFNF-----SMYMEANGRSFIEDIL 688
 SARS-CoV-2 NRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFG-GFNF-----SQILPDPSKPSKRSFIEDLL 822
 SARS-CoV NRALSGIAAEQDRNTREVFQVKQMYKTPTLKYFG-GFNF-----SQILPDPLKPTKRSFIEDLL 804
 HKU 1 NSILDEVNGLLDTTQLHVADTLMQGVTLSSNLNTNLHFVDNINFKSLVGLGPHCGS--SSRSFFEDLL 911
 OC43 NAILTEVNELLDTTQLQVANSLMNGVTLSTKLDKGVNFVDDINFSPVLGCLGSECSKASSRSIAIEDLL 915
 MERS-CoV NQALHGANLRQDDSVRNLFASVKSSQSSPIIPGGGDFNL-----TLLEPVSI STGSRSAIAIEDLL 894

	fusion peptide (FP)	
AcCoV-JC34	FDKIVTTGPGFYQDYDCK--KMNLEDITCKQY	YNGIMVIPPVMDLTLITFWSSA-VAGSMTAGLFGG 746
Lucheng Rn rat CoV	FDKIVTTGPGFYQDYDCK--KMNLDLTKQY	YNGIMVIPPVMDLTLITFWSSA-VAGSMTAGLFGG 753
SARS-CoV-2	FNKVTLADAGFIKQYGDCL--GDIAARDLICAQK	FNGLTVLPLLTDemiaQYTSALLAGTITSGWTFG 889
SARS-CoV	FNKVTLADAGFMKQYGECL--GDINARDLICAQK	FNGLTVLPLLTDemiaAAYTAALVSGTATAGWTFG 871
HKU 1	FDKVKLSDVGFVEAYNCT--GGSEIRDLLCVQS	FNGIKVLPILLSQISGYTTAATVAAMFPPWS-- 976
OC43	FDKVKLSDVGFVEAYNCT--GGAEIRDLLCVQS	YKGIKVLPILLSQISGYTTAATSASLFPWPWT-- 980
MERS-CoV	FDKVTIADPGYMQYDDCMQQGPASARDLICAQY	VAGYKVLPLMDVNMEAYTSSLLGSIAGVGTAG 963
AcCoV-JC34	--QAGMVSWSIAGRLNALGVMQDALVNDVNKLANGFNNT	TQYVADGFKTTSQALSAIQSVVNNNAQQ 813
Lucheng Rn rat CoV	--QAGMVSWTVALAGRLNALGVMQDALVEDVNKLANGFNNT	TQYVSDGFKTTSQALSTIQAVVNNNAQQ 820
SARS-CoV-2	AGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFN	SAIGKIQDLSSTASALGKLDVNVQNAQA 958
SARS-CoV	AGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFN	KAISQIQESLTTTSTALGKLDVNVQNAQA 940
HKU 1	--AAAGIPFSLNVQYRINGLGVTMDVNLKNQKLIATAFN	NALLSIQNGFSATNSALAKIQSVVNSNAQA 1043
OC43	--AAAGVPFYLNVQYRINGLGVTMDVLSQNQKLIANAF	NALHAIQQGFATNSALVKIQAVVNAEA 1047
MERS-CoV	LSSFAAIPFAQSIFYRLNGVGITQQVLSQKLIANKFN	QALGAMQTGFTTTNEAFQKVQDAVNNNAQA 1032
AcCoV-JC34	ISQLVQGLSENFGAISNNFLVIAERLERLEAQMQMDRL	INGRMNILQNFVTNYKLSISELKSQQILAQS 882
Lucheng Rn rat CoV	VSQLVQGLSENFGAISNNFALIAERLERIEAAMQMDRL	INGRMNILQNFVTNYKLSISELKSQQALAQ 889
SARS-CoV-2	LNTLVKQLSSNFGAISSVLDILSRLDKVEAEVQIDRL	ITGRLQSLQTYVTTQQLIRAAEIRASANLAAT 1027
SARS-CoV	LNTLVKQLSSNFGAISSVLDVLSRLDKVEAEVQIDRL	ITGRLQSLQTYVTTQQLIRAAEIRASANLAAT 1009
HKU 1	LNSLLQQLFNKFGAISSLQEILSRLDALEAQVQIDRL	INGRLTALNAYVSQQLSDISLVKFGAALAME 1112
OC43	LNNLLQQLSNRFGAISASLQEILSRLDALEAEQIDRL	INGRLTALNAYVSQQLSDSTLVKFSAAQAME 1116
MERS-CoV	LSKLASELSNTFGAISASIGDIIQRLDVLQDAQIDRL	INGRLTTLNAFVAQQLVRSESAALSAQLAKD 1101
AcCoV-JC34	LVNECVYAQSNRNGFCGDGLHLFSLMQRAPDGI	FMFFHYTLVPNNTIDVETTPGLCLGN--VCIAPRDG 949
Lucheng Rn rat CoV	LINECVYAQSSRNGFCGDGLHLFSLMQRAPDGI	FMFFHYTLKPNNTIIVETTPGLCLSN--VCIAPKDG 956
SARS-CoV-2	KMSECVLGQSKRVDFCGKGYHLMSFPQSA	PHGVVFLHVTYVPAQEKNFTTAPAICHGDK---AHFPREG 1093
SARS-CoV	KMSECVLGQSKRVDFCGKGYHLMSFPQA	AAPHGVVFLHVTYVPSQERNFTTAPAICHEGK---AYFPREG 1075
HKU 1	KVNECVKSQSPRINFCGNGNHILSLVQNAPYGLL	FMHF SYKPI SFKTVLVSPGLCISGD--VGIAPKQG 1179
OC43	KVNECVKSQSSRINFCGNGNHIIISLVQNAPYGLY	IFHFNYVPTKYVTAKVSPGLCIAGN--RGIAPKSG 1183
MERS-CoV	KVNECVKAQSKRSGFCGQGTHTIVSFVNAPNGLY	FMHVGYYPSNHIEVVSAYGLCDAANPTNCIAPVNG 1170
AcCoV-JC34	LFVKTNIKRSV--WHFTTRNLNYPQAITVNSVIVNG	-GVNFTSLNQTIEGIEPPTIPS--FDEEFEDL 1013
Lucheng Rn rat CoV	LFVRLSTARDD--WHFTTRNRYSPFITVNTLTISG	-GVNFTVVNSTIDGIEPPANPS--FDEEFAEL 1021
SARS-CoV-2	VFVSNQTH----WFVTQRNFYEPQIITDNTFVSGN	CDVVIGIVNNTVYDPLQPELDS--FKEELDKY 1155
SARS-CoV	VFVFNQTS----WFVTQRNFYEPQIITDNTFVSGN	CDVVIGIINNNTVYDPLQPELDS--FKEELDKY 1137
HKU 1	YFIKHNH----WMFTGSSYYYPEPISDKNVVFMN	TCSVNFTKAPLVYLNHNSVPLSD--FESELSHW 1241
OC43	YFVNVNNT----WMTGSGYYYPEPITENNVVVM	STCAVNYTKAPYVMLNLSIPNLPD--FKEELDQW 1245
MERS-CoV	YFIKTNTRIVDEWSTGSSFYAEPITSLNTKYVAP	-QVTYQNIISTNLPPPLLGNSTGIDFQDELDEF 1238
AcCoV-JC34	YKNITLLEQLK-NIT-FDPELLNLTYYIDRLDEL	STNVSQHLVDISEFNKYVQYIKWPWYVWLAIFLV 1080
Lucheng Rn rat CoV	YKNVTLEQLK-NIS-FDPEMLNLTYYIDRLDEL	ATNVSQHLVDVSEFNKFVQYIKWPWYVWLAIFLV 1088
SARS-CoV-2	FKNHTSPDVLG-DISGINASVNIQKEIDRLNEVAKN	LESIDLQELGKYEYQYIKWPWYIWLGF IAG 1223
SARS-CoV	FKNHTSPDVLG-DISGINASVNIQKEIDRLNEVAKN	LESIDLQELGKYEYQYIKWPWYVWLGF IAG 1205
HKU 1	FKNQTSIAPNLTNLHTINATFLDLYEMNLIQES	IKSLNNSYINLKDIGTYEYVYKWPWYVWLLISFS 1310
OC43	FKNQTSVAPDL--SLDYINVTFLDLQVEMNRLQEA	IKVNLNHSYINLKDIGTYEYVYKWPWYVWLLICLA 1312
MERS-CoV	FKNVSTSI PNFG-SLTQINTTLLDLTYEMLSLQ	VVQKALNESYIDLKELGNYTYYNKWPWYIWLGF IAG 1306
AcCoV-JC34	LVLFSFMLLWCCCATGCCGCCGMLGSACNGCCT	KPQ---TVEFEKVHVQ-- 1126
Lucheng Rn rat CoV	LVLFSFMLLWCCCATGCCGCCGCGAACNGCCT	KPQ---PIEFEKVHVQ*- 1135
SARS-CoV-2	LIAIVMVTIMLCCMTSCCSCLKGC-CSCGSCCK	FDEDDSEPVLLKGVKLHYT 1273
SARS-CoV	LIAIVMVTILLCCMTSCCSCLKGA-CSCGSCCK	FDEDDSEPVLLKGVKLHYT 1255
HKU 1	FIIFLVLLFFICCTGCGSACF---SKCHNCCDE	YGGHDFVIKTSHDD-- 1356
OC43	GVAMLVLLFFICCTGCGTSCF---KKCGGCCDDY	TGYQELVIKTSHDD-- 1358
MERS-CoV	LVALALCVFFILCCTGCGTNMGK-LKCNRCDDRYE	EY-DLEPHKVHVH-- 1353

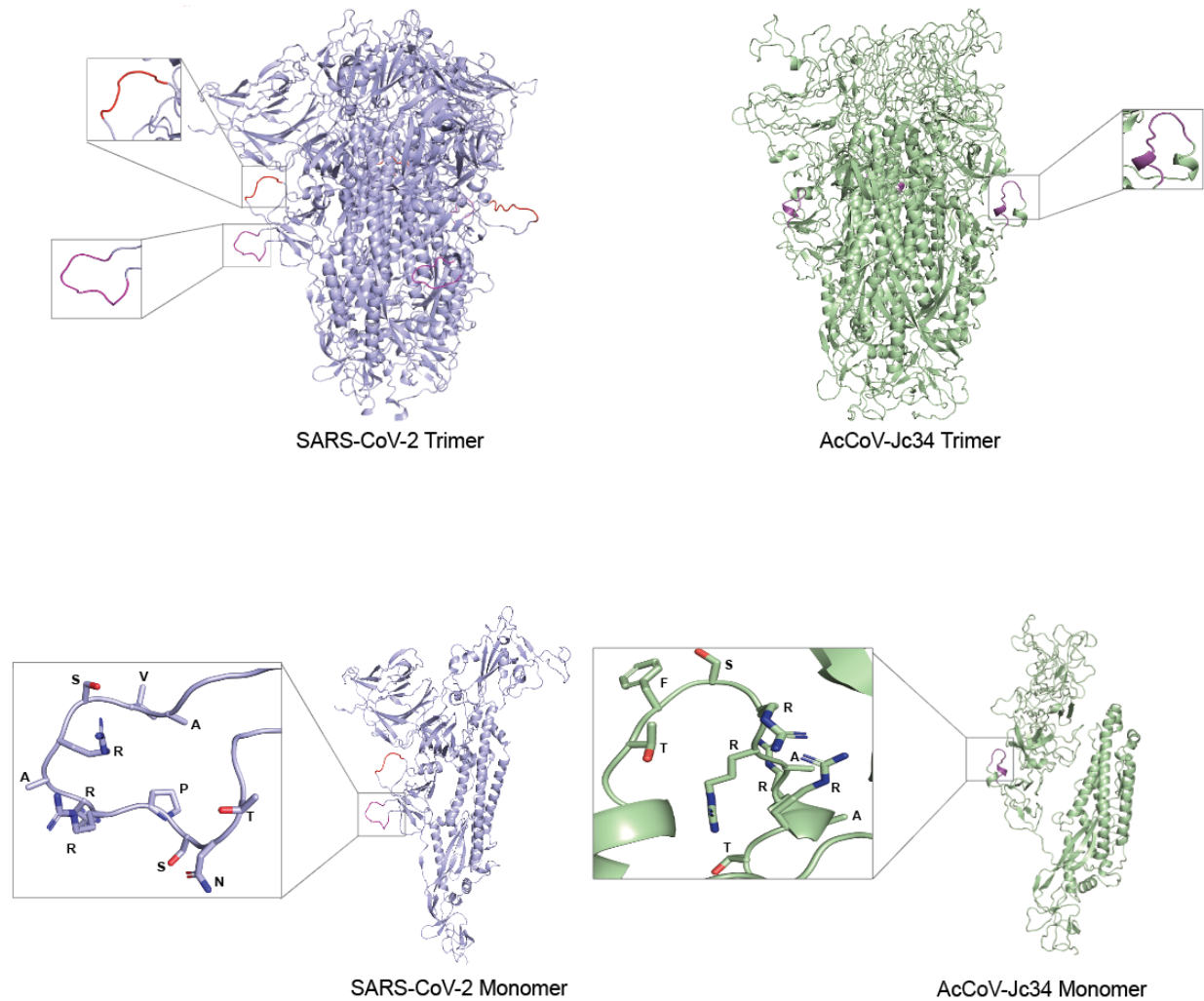


Figure 4. Predicted structure of AcCoV-JC34 spike protein.

The AcCoV-JC34 spike protein was modelled based on SARS-CoV spike. In SARS-CoV-2, the pink highlight indicates where the furin cleavage sequence (PRRAR) is located. The red highlight is the location that aligns with AcCoV-JC34 potential furin cleavage site. In the AcCoV-JC34 structural model, the pink highlight indicates location of the potential furin cleavage site (SRRAR).

Virus	Predicted S1/S2 sequence	PiTou	ProP
AcCoV-JC34	495 - CNSSDVVTFS RRAR ARTLTD - 514	+0.15	0.279
Lucheng Rn rat CoV	502 - CDSTDVTTFMT KAR ATTFVD - 521	-4.577	0.132
MERS-CoV	681 - HISSTMSQYS RSTR SML KRR - 699	-2.662	0.497
SARS-CoV-2	672 - ASYQTQTNS PRRAR SVASQS - 691	+9.196	0.62
SARS-CoV-1	654 - AGICASYHTVSL LR STS QKS - 673	-5.167	0.123
HCoV - HKU1	747 - YNSPSSSS RRKRR SISASY - 766	+14.634	0.88
OC43 (clinical)	750 - GYCVDYF KNRRSRR AITTYG - 769	+10.1	0.753

Figure 5. Furin cleavage analysis of CoV S1/S2 cleavage site

CoV S sequences were analyzed using the ProP 1.0 and PiTou 3.0 furin prediction algorithm, generating a score with bold numbers indicating predicted furin cleavage. (|) denotes the position of the predicted S1/S2 cleavage site. Basic residues, arginine (R) and lysine (K), are highlighted in blue.

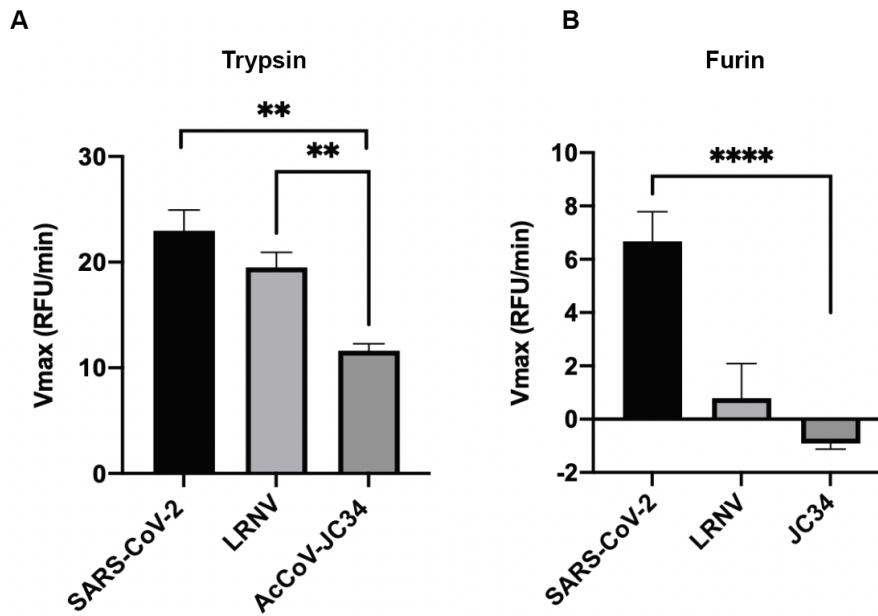


Figure 6. Fluorogenic peptide cleavage assays of the predicted AcCoV-JC34 furin cleavage site.

Peptides mimicking the S1/S2 site of the SARS-CoV-2 WT, AcCoV-JC34, and Lucheng Rn rat CoV (LRNV) were evaluated for in vitro cleavage with A) trypsin and B) furin proteases under pH 7.4 (trypsin), and 7.5 (furin) conditions. Trypsin cleaved all three peptides, while furin only cleaved SARS-CoV-2.

Acknowledgements:

DTS is supported by the Howard Hughes Medical Institute-Cornell University Transfer (HHMI-CURT) program

AES was supported by NIH Comparative Medicine Training Program T32OD011000.

Work in the author's lab is funded in part by the National Institute of Health research grant R01AI35270 (to GW).

We thank the members of the Whittaker Lab, past and present, for their helpful discussions during the preparation of this manuscript.