P429

The Origin of SARS-CoV-2 and its Furin Cleavage Site, a New and Parsimonious Evolutionary Model

by Jonathan Latham, PhD and Allison Wilson, PhD
The Bioscience Resource Project, Ithaca, NY, USA, jrlatham@biosciencesresource.org

The pathway of SARS-CoV-2's emergence from its bat reservoir species is still unresolved. Any proficient origin hypothesis should explain the following:

1) a phylogeographic origin in N. Laos/S. Yunnan
2) the outbreak location in Wuhan
3) a single spillover
4) the human and lung adaptedness of SARS-CoV-2
5) the absence of both intermediate viruses and an infected intermediate host.

We have been examining, as a key to unlock all the rest, a sixth unsolved puzzle: **the presence of a novel twelve nucleotide insertion that introduces the four amino acids PRRA.** This insertion created a furin cleavage site (protein motif RRAR) at the S1/S2 junction of the viral Spike protein and it distinguishes SARS-CoV-2 from all its close relatives (see **Figure 1**). This motif likely was a key factor in the switch from a faecal/oral life-history in bats to the human respiratory transmission seen in SARS-CoV-2 (Jackson *et al*., 2022; Temmam *et al*., 2023).

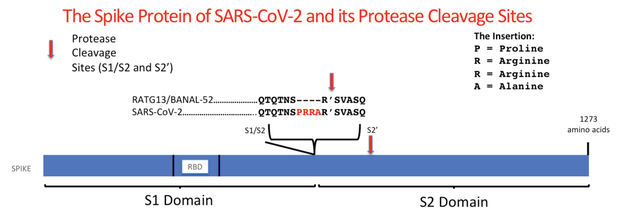


Figure 1: The Spike protein of SARS-CoV-2.
The Spike has two protease cleavage sites. S1/S2 is mostly targeted by furin and S2' is mostly cleaved by TMPRSS2. RBD = Receptor Binding Domain. RaTG13 and BANAL-52 are the presumptive ancestors of SARS-CoV-2 in the Spike region.

The Origin of the Furin Cleavage Site is a Genetic Puzzle

Extended nucleotide insertions in viruses are not created *de novo*. Typically, they arise through recombination, usually with a related virus. However, for SARS-CoV-2, no source for the PRRA has been identified. Candidate origin pathways from known bat virus sequences have nevertheless been hypothesised (Gallaher, 2020; Lytras, 2020). However, these schemes all require at least three indels and multiple substitutions, making them highly unparsimonious. Similar difficulties apply to suggestions the insertion is human-made (Segreto and Deigin, 2021; Harrison and Sachs, 2022). Multiple research groups *have* added furin cleavage sites to viruses, including coronaviruses, but none of them resembles the insertion in SARS-CoV-2. This is in part because RRAR is an unorthodox furin cleavage motif. *In short, neither conventional evolution nor human genetic manipulation well explain the PRRA insertion into SARS CoV-2.*

Was the Human gene ENaC-α the Donor for a Recombination Event?

The human mRNA that codes for the alpha subunit of the epithelial sodium channel (ENaC-α) contains a run of 8 amino acids (RRARSVAS) that exactly match the furin site of SARS-CoV-2 at S1/S2 (see **Figure 2 Part A**). The ENaC-α mRNA represents an attractive potential donor in the statistical sense since its extended homology permits recombination with precursors of SARS-CoV-2 without disrupting the downstream amino acid sequence of the receiving virus, and it is expressed in lung epithelial cells where SARS-CoV-2 is mostly found. However, ENaC-α does not supply the Proline of the PRRA insertion. But if, as outlined in **Figure 2 Part B** below, ENaC-α's mRNA supplied the full insertion length (of 12 nucleotides), and then a Proline substitution ocurred subsequently, this does give a parsimonious solution.

A

human ENaC-α 236	HGARRARSVASS
human SARS-CoV-2 682	TNQTNSPRRARSVASQSI
Bat RaTG13/BANAL-52	TNQTNS....RSVASQSI

B

human ENaC-α	705	cac ggg gcc cgu cga gcc cgu agc gug gcc ucc agc H G A R R A R S V A S
RaTG13/BANAL-52 Spike	23546	aau uc- - - - - - - - - - - a cgu agu gug gcc agu caa N S R S V A S
Intermediate 1	23546	aau ucg gcc cgu cga gca cgu agu gug gcc agu caa N S A R R A R S V A S
Intermediate 2	23546	aau ucg gcu cgu cga gca cgu agu gug gcc agu caa N S A R R A R S V A S
Intermediate 3	23546	aau ucg ccu cgu cga gca cgu agu gug gcc agu caa N S P R R A R S V A S
Intermediate 4=SARS-CoV-2	23546	aau ucu ccu cgg cgg gca cgu agu gua gcu agu caa N S P R R A R S V A S
SARS-CoV-2 Spike	23597	aau ucu ccu cgg cgg gca cgu agu gua gcu agu caa N S P R R A R S V A S

Figure 2. A proposed evolutionary pathway for a recombination event with ENaC-α.
Part A: The amino acid sequences of human ENaC-α, SARS-CoV-2, and the two closest bat coronaviruses, RaTG13 and BANAL-52. These latter are identical to each other in both nucleotide and amino acid sequences (in this region). Highlighted in red is the proposed donor sequence.
Part B: A molecular scheme showing how recombination with human ENaC-α can create SARS-CoV-2 from known ancestor viruses in a stepwise progression requiring only one further non-synonymous change (the A to P transition between intermediates 2 and 3). On each line, nucleotide and amino acid *differences* compared to the line above are shown in red. Between intermediates three and four the changes are all synonymous. For this figure, because synonymous changes are biologically inconsequential, these mutations are collapsed into the final transition. In this scheme, the two recombination breakpoints are presented as having both occurred immediately after nucleotide 23550 of the ancestor virus. However, there is latitude for other breakpoint positions, especially at the 3' end of the insertion where the amino acids of the viruses and ENaC-α are identical.

The Origin of the Furin Cleavage Site is also a Virus-Host Pathology Puzzle

Is there evidence to support this order of mutations? Yes. Based on extensive cell culture and whole animal experiments with SARS-CoV-2 mutants, SARS-CoV, and the ancestral bat viruses, others have proposed that adaptation of SARS-CoV-2 to human lungs was a progression of cell entry pathway adjustments whose order was significantly constrained (Khan *et al*., 2021; Jackson *et al*., 2022). Some of these constraints, which are listed below, are surprisingly complex:

Constraint 1) Acquisition of a furin site at S1/S2 was the initial step.

Constraint 2) Only subsequently was efficient TMPRSS2 cleavage (at the S2' site) achieved.

Constraint 3) Step two depended on achieving step one (i.e. step one *could not* be bypassed).

Constraint 4) The TMPRSS2 cleavage at S2' was achieved without altering the S2' site itself (since the S2' motif is unchanged in SARS-CoV-2 versus its precursor viruses).

Strikingly, the mutations set out in **Figure 2** precisely match, both in their order and in their biological effects, these constraints. That is to say, the recombination with ENaC-α generates a functional furin site (Constraint 1). Subsequently, substitution of A for P greatly enhances cleavage by TMPRSS2 at S2' giving efficient access to lung cells via the early pathway (Zhang *et al*., 2021; Vu *et al*., 2022). By being dependent on the prior insertion, this substitution satisfies both Constraints 2 and 3. Last, by not altering the S2' site itself, the A to P substitution also fulfills Constraint 4.

Figure 3 (below) integrates these genetic changes into a general spillover model that is also consistent with the wider body of evidence: For example, the absence of an intermediate host is consistent with the direct jump to humans implied by the model.

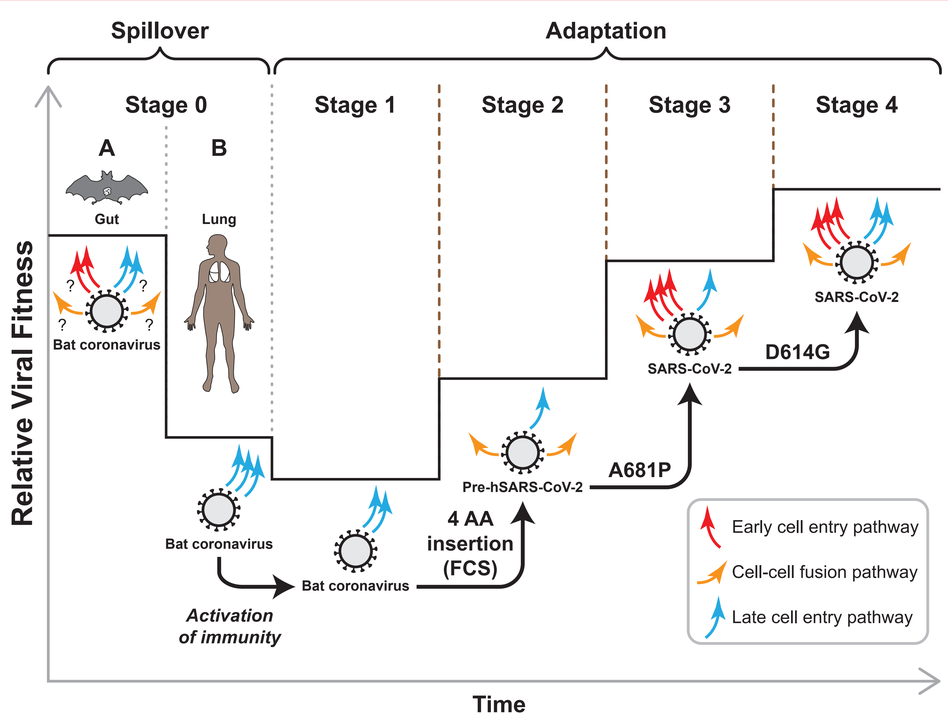
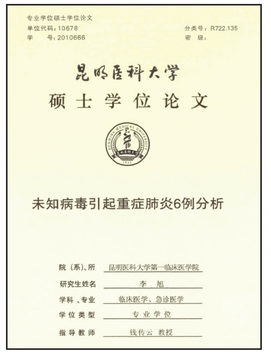


Figure 3: Model for the evolution of the SARS-CoV-2 furin cleavage site in a human lung from known bat Spike sequences.
This model depicts a stepwise evolution of SARS-CoV-2 starting from its bat progenitors. The focus of the model is on viral fitness (see y-axis) as it is affected by cleavage site modifications and their effects on cell entry. The coloured arrows represent the varied cell entry opportunities and their use by the virus as it gains fitness over time. The three entry routes are the 'early' pathway via the plasma membrane (red), the 'late' pathway via endosomes (blue), and cell-cell fusion (orange). Note that, in this figure, Pre-hSARS-CoV-2 is Intermediate 1 in figure 2 B.

The model

At Stage 0 of the model, the spillover stage, a single human inhales a precursor bat virus (equivalent to BANAL-52 or RaTG13) able to efficiently bind human ACE2. As an enteric bat pathogen the virus has low fitness in the human lung, however, and enters cells only through the late pathway. Inhalation rapidly triggers the innate and acquired immune systems. These defences restrict late pathway entry and viral spread and so viral fitness declines still further. This is the situation represented by Stage 1, the ground state of the model. The first mutation to elevate fitness is the acquisition of a furin site from the human mRNA of ENaC-α. This change gives the virus access to the cell-cell fusion pathway (note it also *hinders* late pathway entry). Then, through substitution of Ala by Pro at Spike position 681 (see **Figure 2**), the virus gains efficient entry via the early (TMPRSS2-dependent) pathway. D614G, the last mutation, arises only after human-human transmission has begun. Mechanistically, D614G is a partial intragenic suppressor of the defects associated with furin cleavage. Thus D614G restores the late pathway entry partially disabled at Stage 2.


专业学位硕士学位论文
单位代码：10678　　分类号：R722.135
学　　号：2010668　　密　　级：

昆明医科大学
硕士学位论文

未知病毒引起重症肺炎6例分析

院（系），所　昆明医科大学第一临床医学院
研究生姓名　李　旭
学科、专业　临床医学、急诊医学
学位类型　专业学位
指导教师　钱传云 教授

Is there a candidate for patient zero?

There is. In 2012, six miners were hospitalised with a mystery infection in Yunnan, China. As detailed in a Masters thesis published by their doctor (Li Xu, **Figure 4**) and translated by our non-profit, all six had severe COVID-19-like symptoms. Three died. These miners had been clearing bat guano at the Mojiang mine in S. Yunnan. The mine was heavily populated by bats with coronaviruses, including RaTG13. The miners were provisionally diagnosed at the time as victims of a *novel* coronavirus. Three of the miners were hospitalised on a longterm basis (5 months) and the Wuhan Institute of Virology (WIV) received samples taken from them at intervals for its investigation. According to the WIV (in 2020), their coronavirus tests were negative. However, other researchers, including the head of China's CDC, have explicitly contradicted this.

Figure 4: The Title Page of the 2013 Masters Thesis of Kunming Doctor, Li Xu: "*The analysis of 6 patients with severe pneumonia caused by unknown viruses*".

Since the Mojiang mine is inside the predicted phylogeographic source region of SARS-CoV-2 and since local bat viruses there had RBDs matching SARS-CoV-2 both in sequence and in human ACE2 binding ability (Temmam *et al*., 2023), we propose that SARS-CoV-2 evolved in one of the miners, probably patient 4 (see **Figures 2 and 3**). Just as Omicron and other COVID-19 variants are thought to have evolved extremely rapidly inside single individuals, the miners' lungs would have been well suited for a SARS-related coronavirus of the bat GI tract to become lung and human adapted. Thus, we suggest, the outbreak in Wuhan was an accidental leak from the WIV where this by now highly human-adapted virus was being isolated/studied.

Such a hypothesis can explain *all the salient features* of SARS-CoV-2.

References

Gallaher, W. R. (2020). A palindromic RNA sequence as a common breakpoint contributor to copy-choice recombination in SARS-COV-2. *Archives of Virology, 165*(10), 2341-2348.
Harrison, N. L., & Sachs, J. D. (2022). A call for an independent inquiry into the origin of the SARS-CoV-2 virus. *Proceedings of the National Academy of Sciences, 119*(21), e2202769119.
Lytras, S. (2020) The Sarbecovirus origin of SARS-CoV-2's furin cleavage site (https://virological.org/t/the-sarbecovirus-origin-of-sars-cov-2-s-furin-cleavage-site/536)
Jackson, C. B., Farzan, M., Chen, B., & Choe, H. (2022). Mechanisms of SARS-CoV-2 entry into cells. *Nature reviews Molecular cell biology, 23*(1), 3-20.
Khan, H., Winstone, H., Jimenez-Guardeño, J. M., Graham, C., Doores, K. J., Goujon, C., ... & Malim, M. H. (2021). TMPRSS2 promotes SARS-CoV-2 evasion from NCOA7-mediated restriction. *Plos Pathogens, 17*(11), e1009820.
Segreto, R., & Deigin, Y. (2021). The genetic structure of SARS-CoV-2 does not rule out a laboratory origin: SARS-COV-2 chimeric structure and furin cleavage site might be the result of genetic manipulation. *Bioessays, 43*(3), 2000240.
Temmam, S., Montagutelli, X., Herate, C., Donati, F., Regnault, B., Attia, M., ... & Eloit, M. (2023). SARS-CoV-2-related bat virus behavior in human-relevant models sheds light on the origin of COVID-19. *EMBO reports, 24*(4), e56055.
Vu, M. N., Lokugamage, K. G., Plante, J. A., Scharton, D., Bailey, A. O., Sotcheff, S., ... & Menachery, V. D. (2022). QTQTN motif upstream of the furin-cleavage site plays a key role in SARS-CoV-2 infection and pathogenesis. *Proceedings of the National Academy of Sciences, 119*(32), e2205690119.
Li Xu, (2013) The analysis of 6 patients with severe pneumonia caused by unknown virus.