The Origins of SARS-CoV-2: A Critical Review

Edward C. Holmes[1], Stephen A. Goldstein[2], Angela L. Rasmussen[3], David L. Robertson[4], Alexander Crits-Christoph[5], Joel O. Wertheim[6], Simon J. Anthony[7], Wendy S. Barclay[8], Maciej F. Boni[9], Peter C. Doherty[10], Jeremy Farrar[11], Jemma L. Geoghegan[12], Xiaowei Jiang[13], Julian L. Leibowitz[14], Stuart J. D. Neil[15], Tim Skern[16], Susan R. Weiss[17], Michael Worobey[18], Kristian G. Andersen[19], Robert F. Garry[20,21], Andrew Rambaut[22].

[1]Marie Bashir Institute for Infectious Diseases and Biosecurity, School of Life and Environmental Sciences and School of Medical Sciences, The University of Sydney, Sydney, NSW 2006, Australia.

[2]Department of Human Genetics, University of Utah, Salt Lake City, UT 84112, USA.

[3]Vaccine and Infectious Disease Organization, University of Saskatchewan, Saskatoon, SK, S7N 5E3, Canada.

[4]MRC-University of Glasgow Centre for Virus Research, Glasgow, G61 1QH, UK.

[5]Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, CA 94704, USA.

[6]Department of Medicine, University of California San Diego, La Jolla, CA 92093, USA.

[7]Department of Pathology, Microbiology, and Immunology, University of California Davis School of Veterinary Medicine, Davis, CA 95616, USA.

[8]Department of Infectious Disease, Imperial College London, W2 1PG, UK.

[9]Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, University Park, PA 16802, USA.

[10]Department of Microbiology and Immunology, The University of Melbourne at the Doherty Institute, 792 Elizabeth St, Melbourne, VIC 3000, Australia.

[11]The Wellcome Trust, London, NW1 2BE, UK.

[12]Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand. Institute of Environmental Science and Research, Wellington, New Zealand.

[13]Department of Biological Sciences, Xi'an Jiaotong-Liverpool University (XJTLU), Suzhou, China.

[14]Department of Microbial Pathogenesis and Immunology, Texas A&M University, College Station, TX 77807, USA.

[15]Department of Infectious Diseases, King's College London, Guy's Hospital, London SE1 9RT, UK.
[16]Max Perutz Labs, Medical University of Vienna, Vienna Biocenter, Dr. Bohr-Gasse 9/3, A-1030 Vienna, Austria.
[17]Perelman School of Medicine, University of Pennsylvania. Philadelphia, PA 19104, USA.
[18]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA.
[19]Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA 92037, USA.
[20]Department of Microbiology and Immunology, Tulane University School of Medicine, New Orleans, LA 70112, USA.
[21]Zalgen Labs, Germantown, MD 20876, USA.
[22]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, EH9 3FL, UK.

Since the first reports of a novel SARS-like coronavirus in December 2019 in Wuhan, China, there has been intense interest in understanding how SARS-CoV-2 emerged in the human population. Recent debate has coalesced around two competing ideas: a "laboratory escape" scenario and zoonotic emergence. Here, we critically review the current scientific evidence that may help clarify the origin of SARS-CoV-2.

Evidence supporting a zoonotic origin of SARS-CoV-2

Coronaviruses have long been known to present pandemic risks. SARS-CoV-2 is the ninth documented coronavirus that infects humans and the seventh identified in the last 20 years[1,2]. All previous human coronaviruses have zoonotic origins, as have the vast majority of human viruses. The emergence of SARS-CoV-2 bears several signatures of these prior zoonotic events. It displays clear similarities to SARS-CoV that spilled over into humans in Foshan, Guangdong province, China in November 2002, and again in Guangzhou, Guangdong province in 2003[3]. Both these SARS-CoV emergence events were associated with markets selling live animals and involved species, particularly civets and raccoon dogs[4], that were also sold live in Wuhan markets in 2019[5] and are known to be susceptible to SARS-CoV-2 infection[6]. Animal traders working in 2003, without a SARS diagnosis, were documented to have high levels of IgG to SARS-CoV (13% overall and >50% for traders specializing in civets[7]). Subsequent serological surveys found ~3% positivity rates to SARS-CoV related (SARSr-CoV) viruses in residents of Yunnan province living close to bat caves[8], demonstrating regular exposure in rural locations. The closest known relatives to both SARS-CoV and SARS-CoV-2 are viruses from bats in Yunnan, although animals from this province have been preferentially sampled. For both SARS-CoV and SARS-CoV-2, there is a considerable geographic gap between Yunnan and the location of the first human cases, highlighting the difficulty in identifying the exact pathway of virus emergence and the importance of sampling beyond Yunnan.

SARS-CoV-2 also shows similarities to the four endemic human coronaviruses: HCoV-OC43, HCoV-HKU1, HCoV-229E, and HCoV-NL63. These viruses have zoonotic origins and the circumstances of their emergence are unclear. In direct parallel to SARS-CoV-2, HCoV-HKU1, which was first described in a large Chinese city (Shenzhen, Guangdong) in the winter of 2004, has an unknown animal origin, contains a furin cleavage site in its spike protein, and was originally identified in a case of human pneumonia[9].

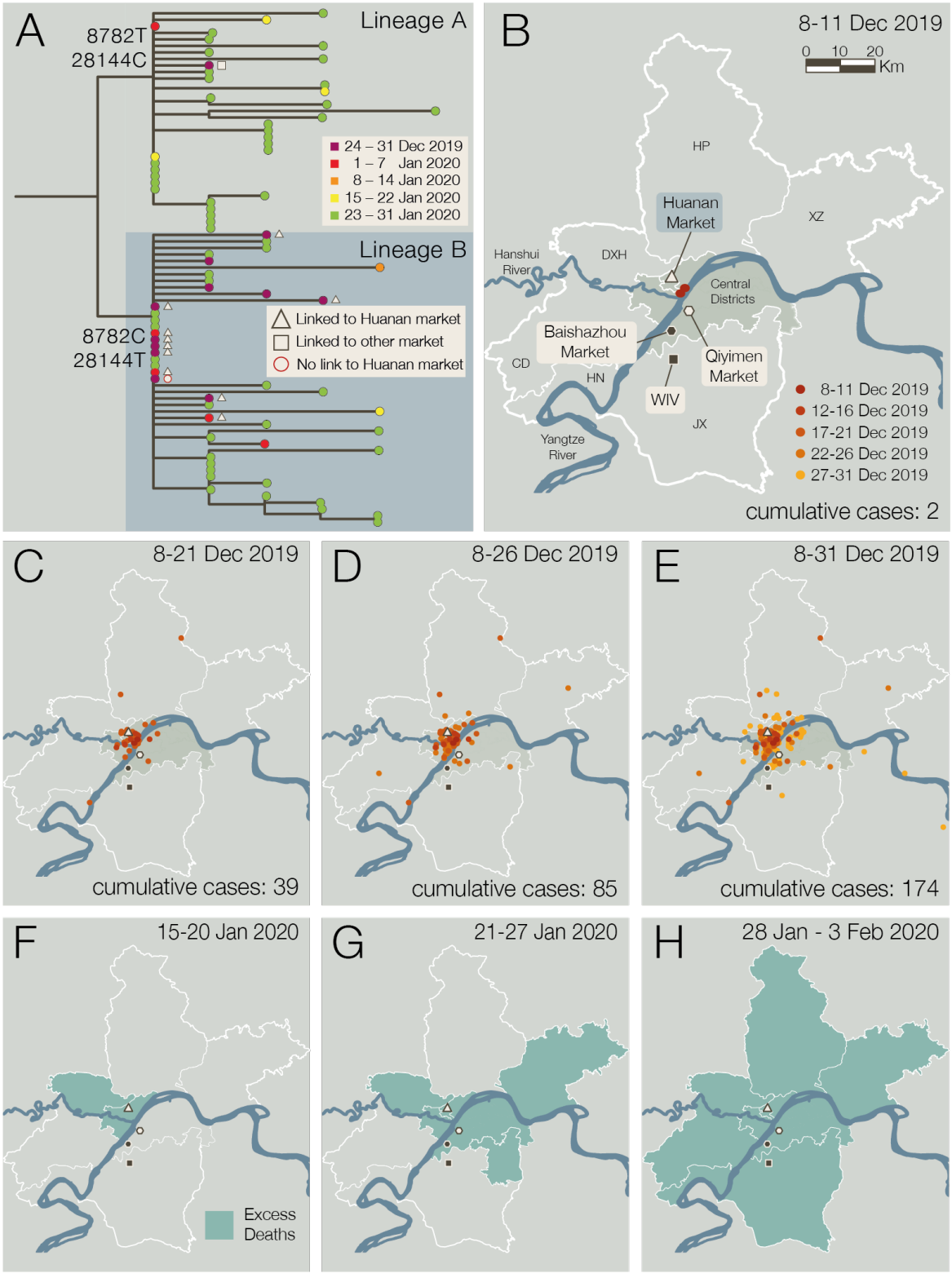

Figure 1 | Phylogenetic and epidemiological data on the early COVID-19 pandemic in Wuhan. (**a**) Phylogenetic tree of early SARS-CoV-2 genomes sampled from Wuhan during December 2019-January 2020. The split between lineages A and B is labelled with the coordinates and base of the

two differentiating nucleotide mutations. Cases with a known association to the Huanan or other markets are denoted by symbols (reported in ref. 10). (**b**) Map of districts of Wuhan showing the location of markets, the BSL-4 campus of the Wuhan Institute of Virology (where the coronavirus work of Dr. Shi Zhengli is performed) and the earliest known cases. (**c-e**) Location of recorded COVID-19 cases in Wuhan from 8th December to 31st December 2019. Cases with a home address outside of Wuhan city are not shown. (**f-h**) Map of districts of Wuhan indicating the first record of excess deaths due to pneumonia (shaded green) from 15th January 2020. Case and excess death data were extracted and redrawn from figures provided in ref 10. For more details see **supplementary information**. Map data copyright ©OpenStreetMap contributors.

Based on epidemiological data, the Huanan market in Wuhan was an early and major epicenter of SARS-CoV-2 infection. Two of the three earliest documented COVID-19 cases were directly linked to this market selling wild animals, as were 28% of all cases reported in December 2019[10]. Overall, 55% of cases during December 2019 had an exposure to either the Huanan or other markets in Wuhan, with these cases more prevalent in the first half of that month[10]. Examination of the locations of early cases shows that most cluster around the Huanan market, located north of the Yangtze river (**Fig. 1a-e**). These districts were also the first to exhibit excess pneumonia deaths in January 2020 (**Fig. 1f-h**). There is no epidemiological link to any other locality in Wuhan, including the BSL-4 campus of the Wuhan Institute of Virology (WIV) located south of the Yangtze and the subject of considerable speculation. Although some early cases do not have a direct epidemiological link to a market[10], this is expected given high rates of asymptomatic transmission and undocumented secondary transmission events, and was similarly observed in early SARS-CoV cases in Foshan[3].

During 2019, markets in Wuhan – including the Huanan market – traded many thousands of live wild animals including high-risk species such as civets and raccoon dogs[5]. Following its closure, SARS-CoV-2 was detected in environmental samples at the Huanan market, primarily in the western section that traded in wildlife and domestic animal products, as well as in associated drainage areas[10]. While animal carcasses retrospectively tested negative for SARS-CoV-2, these were unrepresentative of the live animal species sold, and specifically did not include raccoon dogs and other animals known to be susceptible to SARS-CoV-2[5].

The earliest split in the SARS-CoV-2 phylogeny defines two lineages - denoted A and B[11] - that likely circulated contemporaneously (**Fig. 1a**). Lineage B, which became dominant globally, was observed in early cases linked to the Huanan market and environmental samples taken there, while lineage A contains a case with exposure to other markets (**Fig. 1a,b**) as well as with later cases in Wuhan and other parts of China[10]. This phylogenetic pattern is consistent with the emergence of SARS-CoV-2 involving one or more contacts with infected animals and/or traders, including multiple spill-over events, as potentially infected or susceptible animals were moved into or between Wuhan markets via shared supply chains and sold for human consumption[5]. The potential emergence of SARS-CoV-2 across multiple markets again mirrors SARS-CoV in which high levels of infection, seroprevalence and genetic diversity in animals were documented at both the Dongmen market in Shenzhen[4,12] and the Xinyuan market in Guangzhou[13,14].

Viruses closely related to SARS-CoV-2 have been documented in bats and pangolins in multiple localities in South-East Asia, including in China, Thailand, Cambodia, and Japan[15,16], with serological evidence for viral infection in pangolins for more than a decade[17]. However, a significant evolutionary gap exists between SARS-CoV-2 and the closest related animal viruses: their genetic distances of approximately 4% (~1,150 mutations) equates to decades of evolutionary divergence[18]. Widespread genomic recombination also complicates the assignment of which viruses are closest to SARS-CoV-2. Although *Rhinolophus* bat virus RaTG13 collected in Yunnan has the highest average genetic similarity to SARS-CoV-2, a history of recombination means that three other bat viruses – RmYN02, RpYN06 and PrC31 – are closer in most of the virus genome (particularly ORF1ab) and thus share a more recent common ancestor with SARS-CoV-2[15,16,19]. None of these closer viruses were collected by the WIV. This demonstrates beyond reasonable doubt that RaTG13 is not the progenitor of SARS-CoV-2, with or without laboratory manipulation or experimental mutagenesis.

Although no bat reservoir nor intermediate animal host for SARS-CoV-2 has been identified to date, initial cross-species transmission events are very likely to go undetected. Most SARS-CoV-2 index case infections are unlikely to have resulted in sustained onward transmission[20] and only a very small subset of spillover events from animals to humans result in major outbreaks. Indeed, the animal origins of many well-known human pathogens, including Ebola virus, Hepatitis C virus, poliovirus, and the coronaviruses HCoV-HKU1 and HCoV-NL63, are yet to be identified,

while it took over a decade to discover bat viruses with >95% similarity to SARS-CoV and able to use hACE-2 as a receptor[21].

Could SARS-CoV-2 have escaped from a laboratory?

There are precedents for laboratory incidents leading to isolated infections and transient transmission chains, including SARS-CoV[22]. Aside from the 1977 A/H1N1 influenza pandemic that likely originated from a large-scale vaccine challenge trial[23], there are no documented examples of human epidemics or pandemics resulting from research activity.

The emergence of SARS-CoV-2 differs markedly from documented laboratory escapes that, with the exception of Marburg virus[24], have been of readily identifiable viruses capable of human infection and associated with sustained work in high titer cultures[25–27]. No previous epidemic has been caused by the escape of a novel virus and there is no data to suggest that the WIV—or any other laboratory—were working on SARS-CoV-2, or any virus close enough to be the progenitor, prior to the COVID-19 pandemic. Viral genomic sequencing without cell culture, which was routinely performed at the WIV, represents a negligible risk as viruses are inactivated during RNA extraction[28] and no case of laboratory escape has been documented following the sequencing of viral samples.

Known laboratory outbreaks have been traced to both workplace and family contacts of index cases and to the laboratory of origin[25–27,24]. Despite extensive contact tracing of early cases during the COVID-19 pandemic, there have been no reported cases related to any laboratory staff at the WIV and all staff in the laboratory of Dr. Shi Zhengli were reported to be seronegative for SARS-CoV-2 when tested in March 2020[10]. During a period of high influenza transmission and other respiratory virus circulation[29] reports of illnesses would need to be confirmed as caused by SARS-CoV-2 to be relevant. Epidemiological modeling suggests that the number of hypothetical cases needed to result in multiple hospitalized COVID-19 patients prior to December 2019 is incompatible with observed clinical, genomic, and epidemiological data[20].

The WIV possesses an extensive catalogue of samples derived from bats and has reportedly successfully cultured three SARSr-CoVs from bats, all of which are genetically distinct from SARS-CoV-2[30–32]. These viruses were isolated from fecal samples through serial amplification in

VeroE6 cells, a process that consistently results in the loss of the SARS-CoV-2 furin cleavage site[33–39]. It is therefore highly unlikely that these techniques would result in the isolation of a SARS-CoV-2 progenitor with an intact furin cleavage site. No published work indicates that other methods, including the generation of novel reverse genetics systems, were used at the WIV to propagate infectious SARSr-CoVs based on sequence data from bats. Gain-of-function research would be expected to utilize an established SARSr-CoV genomic backbone, or at a minimum a virus previously identified via sequencing. However, past experimental research using recombinant coronaviruses at the WIV has used a genetic backbone (WIV1) unrelated to SARS-CoV-2[32] and SARS-CoV-2 carries no evidence of genetic markers one might expect from laboratory experiments[40]. There is no rational experimental reason why a new genetic system would be developed using an unknown and unpublished virus, with no evidence nor mention of a SARS-CoV-2-like virus in any prior publication or study from the WIV[32,41,42], no evidence that the WIV sequenced a virus that is closer to SARS-CoV-2 than RaTG13, and no reason to hide research on a SARS-CoV-2-like virus prior to the COVID-19 pandemic. Under any laboratory escape scenario SARS-CoV-2 would have to have been present in a laboratory prior to the pandemic, yet no evidence exists to support such a notion and no sequence has been identified that could have served as a precursor.

A specific laboratory escape scenario involves accidental infection in the course of serial passage of a SARSr-CoV in common laboratory animals such as mice. However, early SARS-CoV-2 isolates were unable to infect wild-type mice[43]. While murine models are useful for studying infection *in vivo* and testing vaccines, they often result in mild or atypical disease[44–48]. These findings are inconsistent with a virus selected for increased pathogenicity and transmissibility through serial passage through rodents. Although SARS-CoV-2 has since been engineered[49] and adapted by serial passage[50–52], specific mutations in the spike protein, including N501Y, are necessary for such adaptation in mice[51,52]. Notably, N501Y has arisen convergently in multiple SARS-CoV-2 variants of concern in the human population, presumably being selected to increase ACE2 binding affinity[53–56]. If SARS-CoV-2 resulted from attempts to adapt a SARSr-CoV for study in animal models, it would likely have acquired mutations like N501Y for efficient replication in that model, yet there is no evidence to suggest such mutations existed early in the pandemic. Both the low pathogenicity in commonly used laboratory animals and the absence of genomic markers associated with rodent adaptation indicate that SARS-

CoV-2 is highly unlikely to have been acquired by laboratory workers in the course of viral pathogenesis or gain-of-function experiments.

Evidence from genomic structure and ongoing evolution of SARS-CoV-2

Considerable attention has been devoted to claims that SARS-CoV-2 was genetically engineered or adapted in cell culture or "humanized" animal models to promote human transmission[57]. Yet, since its emergence, SARS-CoV-2 has experienced repeated sweeps of mutations that have increased viral fitness[58,59]. The first clear adaptive mutation, the D614G substitution in the spike protein, occurred early in the pandemic[60,61]. Recurring mutations in the receptor binding domain of the spike protein, including N501Y, K417N/T, L452R, and E484K/Q—constituent mutations of the variants of concern—similarly enhance viral infectivity[54,55,62] and ACE2 binding[53,63], refuting claims that the SARS-CoV-2 spike protein was optimized for binding to human ACE2 upon its emergence[56]. Further, some pangolin-derived coronaviruses have receptor binding domains that are near-identical to SARS-CoV-2 at the amino acid level[40,64] and bind to human ACE2 even more strongly than SARS-CoV-2, showing that there is capacity for further human adaptation[65]. SARS-CoV-2 is also notable for being a host generalist virus[66], capable of efficient transmission in multiple mammalian species, including mink, tigers, cats, gorillas, dogs, raccoon dogs, ferrets, and large outbreaks have been documented in mink with spill-back to humans[67] and to other animals[68]. Combined, these findings show that no specific human "pre" adaptation was required for the emergence or early spread of SARS-CoV-2, and the claim that the virus was already highly adapted to the human host[57], or somehow optimized for binding to human ACE2, is without validity.

The genesis of the polybasic (furin) cleavage site in the spike protein of SARS-CoV-2 has been subject to recurrent speculation. Although the furin cleavage site is absent from the closest known relatives of SARS-CoV-2[40], this is unsurprising as the lineage leading to this virus is poorly sampled and the closest bat viruses have divergent spike proteins due to recombination[15,16,18]. Furin cleavage sites are commonplace in other coronavirus spike proteins, including some feline alphacoronaviruses, MERS-CoV, most but not all strains of mouse hepatitis virus, as well as in endemic human betacoronaviruses such as HCoV-OC43 and HCoV-HKU1[69–71]. A near identical nucleotide sequence is found in the spike gene of the bat coronavirus HKU9-1[72], and both SARS-CoV-2 and HKU9-1 contain short palindromic sequences immediately upstream of this sequence that are indicative of natural recombination break-points via template switching[72]. Hence, simple

evolutionary mechanisms can readily explain the evolution of an out-of-frame insertion of a furin cleavage site in SARS-CoV-2 (**Fig. 2**).

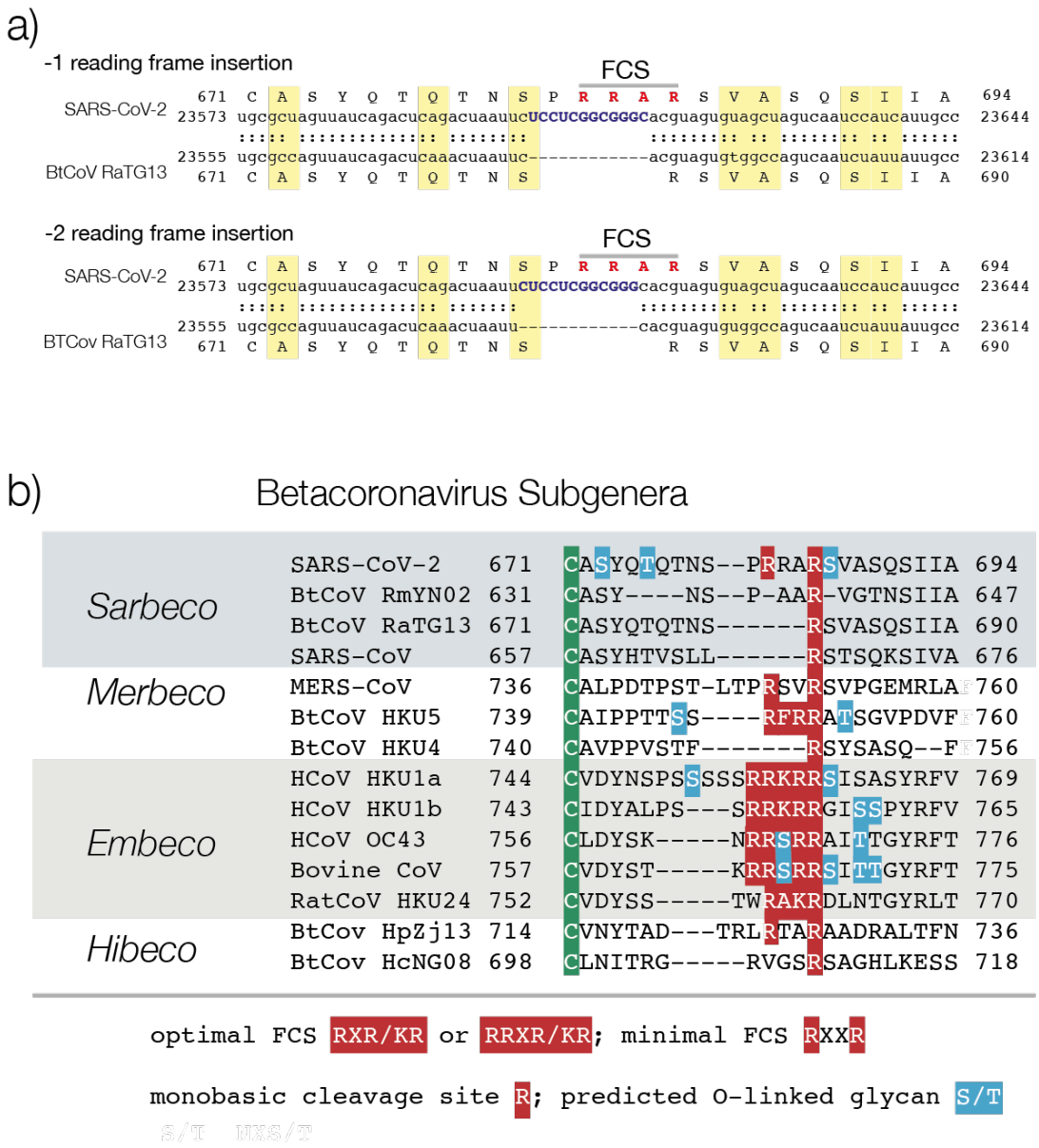

Figure 2 | Evolution of the furin cleavage site (FCS) in the spike protein of betacoronaviruses. (**a**) Sequence alignment of the region around the FCS in SARS-CoV-2 (NCBI accession MN908947) and bat coronavirus RaTG13 (NCBI accession MN996532) showing that the former was the result of an out-of-frame nucleotide sequence insertion. (**b**) Amino acid sequence alignment of the FCS region in representative members of the different subgenera of betacoronaviruses, highlighting the evolutionary volatility of this site and that the relevant amino acid motif (RRAR) in SARS-CoV-2 is functionally suboptimal. The residues predicted to be O-linked glycans are also marked. For more details see **supplementary information**.

The SARS-CoV-2 furin cleavage site (containing the amino acid motif RRAR) does not match its canonical form (R-X-R/K-R), is suboptimal compared to those of HCoV-HKU1 and HCoV-OC43, lacks either a P1 or P2 arginine (depending on the alignment), and was caused by an out-of-frame insertion (**Fig. 2**). The RRAR and RRSR S1/S2 cleavage sites in feline coronaviruses (FCoV) and cell-culture adapted HCoV-OC43, respectively, are not cleaved by furin[69]. There is no logical reason why an engineered virus would utilize such a poor furin cleavage site, which would entail such an unusual and needlessly complex feat of genetic engineering. The only previous studies of artificial insertion of a furin cleavage site at the S1/S2 boundary in the SARS-CoV spike protein utilized an optimal 'RRSRR' sequence in pseudotype systems[73,74]. Further, there is no evidence of prior research at the WIV involving the artificial insertion of complete furin cleavage sites into coronaviruses.

The recurring P681H/R substitution in the proline (P) residue preceding the SARS-CoV-2 furin cleavage site improves cleavage of the spike protein and is another signature of ongoing human adaptation of the virus[75]. The SARS-CoV-2 furin site is also lost under standard cell culture conditions[34,76], as is true of HCoV-OC43[73]. The presence of two CGG codons for arginines in the SARS-CoV-2 furin cleavage site is similarly not indicative of genetic engineering[77]. Although the CGG codon is rare in coronaviruses, it is observed in SARS-CoV, SARS-CoV-2 and other human coronaviruses at comparable frequencies[77]. Further, if low-fitness codons had been artificially inserted into the virus genome they would have been quickly selected against during SARS-CoV-2 evolution, yet both CGG codons are more than 99.8% conserved among the >1,800,000 near-complete SARS-CoV-2 genomes sequenced to date, indicative of strong functional constraints (**supplementary information, Table S1**).

Conclusions

As for the vast majority of human viruses, the most parsimonious explanation for the origin of SARS-CoV-2 is a zoonotic event. The documented epidemiological history of the virus is comparable to previous animal market-associated outbreaks of coronaviruses with a simple route for human exposure. The contact tracing of SARS-CoV-2 to markets in Wuhan exhibits striking similarities to the early spread of SARS-CoV to markets in Guangdong, where humans infected early in the epidemic lived near or worked in animal markets. Zoonotic spillover by definition selects for viruses able to infect humans. The laboratory escapes documented to date

have almost exclusively involved viruses brought into laboratories specifically because of their known human infectivity.

There is currently no evidence that SARS-CoV-2 has a laboratory origin. There is no evidence that any early cases had any connection to the WIV, in contrast to the clear epidemiological links to animal markets in Wuhan, nor evidence that the WIV possessed or worked on a progenitor of SARS-CoV-2 prior to the pandemic. The suspicion that SARS-CoV-2 might have a laboratory origin stems from the coincidence that it was first detected in a city that houses a major virological laboratory that studies coronaviruses. Wuhan is the largest city in central China with multiple animal markets and is a major hub for travel and commerce, well connected to other areas both within China and internationally. The link to Wuhan therefore more likely reflects the fact that pathogens often require heavily populated areas to become established[20].

We contend that there is substantial body of scientific evidence supporting a zoonotic origin for SARS-CoV-2. While the possibility of a laboratory accident cannot be entirely dismissed, and may be near impossible to falsify, this conduit for emergence is highly unlikely relative to the numerous and repeated human-animal contacts that occur routinely in the wildlife trade. Failure to comprehensively investigate the zoonotic origin through collaborative and carefully coordinated studies would leave the world vulnerable to future pandemics arising from the same human activities that have repeatedly put us on a collision course with novel viruses.

Acknowledgements

ECH is supported by an Australian Research Council Australian Laureate Fellowship (FL170100022). SAG is supported by the National Institutes of Health F32AI152341. JOW acknowledges support from the National Institutes of Health (AI135992). ALR acknowledges that VIDO receives operational funding from the Canada Foundation for Innovation - Major Science Initiatives Fund and from the Government of Saskatchewan through Innovation Saskatchewan and the Ministry of Agriculture. DLR acknowledges support of the Medical Research Council (MC_UU_12014/12) and the Wellcome Trust (220977/Z/20/Z). SJA acknowledges funding from the National Institute of Allergy and Infectious Diseases (R01AI149693). WB receives support from the Wellcome Trust (Z/205100 and Z/200187), BBSRC (BB/S008292) and MRC (MR/W005611/1). MFB acknowledges funding from the Bill and Melinda Gates Foundation (INV-005517). JLG is supported by a New Zealand Royal Society Rutherford Discovery Fellowship

(RDF-20-UOO-007). JLL is supported by the National Institutes of Health (R01AI141607, R21AI139738) and the National Science Foundation (Grant no. 2029949). SJN is supported by a Wellcome Trust Senior Fellowship (WT098049AIA), the Medical Research Council, and the Huo Family Charitable Foundation 2. TS was funded by the Austrian Science Fund (FWF) grant number P 28183. SRW is supported by the National Institutes of Health (R01AI140442, R01AI104887, R21AI138564, R21AI157147), as well as the Penn Center for Research on Coronaviruses and Other Emerging Pathogens. MW is supported by Bill and Melinda Gates Foundation INV004212 and the Arizona Board of Regents. KGA acknowledges support from the National Institutes of Health (U19AI135995, U01AI151812, and UL1TR002550). RFG is supported by the National Institutes of Health (R01AI132223, R01AI132244, U19AI135995, U54HG007480, U19AI142790, U01AI151812), the Coalition for Epidemic Preparedness Innovations (INTU1901 and ESEP1904) and the European & Developing Countries Clinical Trials Partnership (RIA2019LV-3053). AR acknowledges the support of the Wellcome Trust (Collaborators Award 206298/Z/17/Z – ARTIC network) and the European Research Council (grant agreement no. 725422 – ReservoirDOCS). We gratefully acknowledge the authors and the laboratories responsible for the genome sequence data shared via the GISAID Initiative, and provide a complete acknowledgement table for the data used here as a supplementary document.

Disclosures

JOW receives funding from the U.S. Centers for Disease Control and Prevention (ongoing) via grants and contracts to his institution unrelated to this research. SRW consults for Immunome and Ocugen. AR, ALR, MFB, SAG, and KGA have received consulting fees and compensated expert testimony on SARS-CoV-2 and the COVID-19 pandemic. RFG is co-founder of Zalgen Labs.

References

1. Vlasova, A. N. *et al.* Novel canine coronavirus isolated from a hospitalized pneumonia patient, east Malaysia. *Clin. Infect. Dis.* (2021) doi:10.1093/cid/ciab456.
2. Lednicky, J. A. *et al.* Emergence of porcine delta-coronavirus pathogenic infections among children in Haiti through independent zoonoses and convergent evolution. *medRxiv* (2021) doi:10.1101/2021.03.19.21253391.
3. Xu, R.-H. *et al.* Epidemiologic clues to SARS origin in China. *Emerg. Infect. Dis.* **10**, 1030–1037 (2004).
4. Guan, Y. *et al.* Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* **302**, 276–278 (2003).
5. Xiao, X., Newman, C., Buesching, C. D., Macdonald, D. W. & Zhou, Z.-M. Animal sales from Wuhan wet markets immediately prior to the COVID-19 pandemic. *Sci. Rep.* **11**, 1–7 (2021).
6. Freuling, C. M. *et al.* Susceptibility of raccoon dogs for experimental SARS-CoV-2 infection. *Emerg. Infect. Dis.* **26**, 2982–2985 (2020).
7. Centers for Disease Control and Prevention (CDC). Prevalence of IgG antibody to SARS-associated coronavirus in animal traders--Guangdong Province, China, 2003. *MMWR Morb. Mortal. Wkly. Rep.* **52**, 986–987 (2003).
8. Wang, N. *et al.* Serological evidence of bat SARS-related coronavirus infection in humans, China. *Virol. Sin.* **33**, 104–107 (2018).
9. Woo, P. C. Y. *et al.* Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J. Virol.* **79**, 884–895 (2005).
10. WHO. WHO-convened global study of origins of SARS-CoV-2: China Part. https://www.who.int/publications/i/item/who-convened-global-study-of-origins-of-sars-cov-2-china-part (2021).
11. Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* **5**, 1403–1407 (2020).
12. Al, H. E. Y. et. Surveillance of SARS coronavirus among wild animal sold in Dongmen market in Shenzhen city. *jbjc* **19**, 287–291 (2004).
13. Tu, C. *et al.* Antibodies to SARS coronavirus in civets. *Emerg. Infect. Dis.* **10**, 2244–2248 (2004).

14. Wang M. *et al.* Surveillance on severe acute respiratory syndrome associated coronavirus in animals at a live animal market of Guangzhou in 2004. *Zhonghua Liu Xing Bing Xue Za Zhi* **26**, 84–87 (2005).
15. 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.
16. Lytras, S. *et al.* Exploring the natural origins of SARS-CoV-2 in the light of recombination. doi:10.1101/2021.01.22.427830.
17. Wacharapluesadee, S. *et al.* Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. *Nat. Commun.* **12**, 1–9 (2021).
18. Boni, M. F. *et al.* Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat Microbiol* **5**, 1408–1417 (2020).
19. Li, L.-L. *et al.* A novel SARS-CoV-2 related virus with complex recombination isolated from bats in Yunnan province, China. *bioRxiv* 2021.03.17.435823 (2021) doi:10.1101/2021.03.17.435823.
20. Pekar, J., Worobey, M., Moshiri, N., Scheffler, K. & Wertheim, J. O. Timing the SARS-CoV-2 index case in Hubei province. *Science* **372**, 412–417 (2021).
21. Cyranoski, D. Bat cave solves mystery of deadly SARS virus - and suggests new outbreak could occur. *Nature* **552**, 15–16 (2017).
22. Parry, J. Breaches of safety regulations are probable cause of recent SARS outbreak, WHO says. *BMJ* **328**, 1222 (2004).
23. Rozo, M. & Gronvall, G. K. The reemergent 1977 H1N1 strain and the gain-of-function debate. *MBio* **6**, e01013–15 (2015).
24. Ristanović, E. S., Kokoškov, N. S., Crozier, I., Kuhn, J. H. & Gligić, A. S. A Forgotten episode of Marburg virus disease: Belgrade, Yugoslavia, 1967. *Microbiol. Mol. Biol. Rev.* **84**, e00095–19 (2020).
25. Senio, K. Recent Singapore SARS case a laboratory accident. *Lancet Infect. Dis.* **3**, 679 (2003).
26. Lim, P. L. *et al.* Laboratory-acquired severe acute respiratory syndrome. *N. Engl. J. Med.* **350**, 1740–1745 (2004).
27. Geddes, A. M. The history of smallpox. *Clin. Dermatol.* **24**, 152–157 (2006).
28. Blow, J. A., Dohm, D. J., Negley, D. L. & Mores, C. N. Virus inactivation by nucleic acid extraction reagents. *J. Virol. Methods* **119**, 195–198 (2004).

29. Liu, M., Deng, L., Wang, D. & Jiang, T. Influenza activity during the outbreak of coronavirus disease 2019 in Chinese mainland. *Biosaf Health* **2**, 206–209 (2020).
30. Ge, X.-Y. *et al.* Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **503**, 535–538 (2013).
31. Yang, X.-L. *et al.* Isolation and characterization of a novel bat coronavirus closely related to the direct progenitor of Severe Acute Respiratory Syndrome coronavirus. *J. Virol.* **90**, 3253–3256 (2015).
32. 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).
33. Liu, Z. *et al.* Identification of Common Deletions in the Spike Protein of Severe Acute Respiratory Syndrome Coronavirus 2. *J. Virol.* **94**, (2020).
34. Ogando, N. S. *et al.* SARS-coronavirus-2 replication in Vero E6 cells: replication kinetics, rapid adaptation and cytopathology. *J. Gen. Virol.* **101**, 925–940 (2020).
35. Wong, Y. C. *et al.* Natural transmission of bat-like SARS-CoV-2ΔPRRA variants in COVID-19 patients. *Clin. Infect. Dis.* (2020) doi:10.1093/cid/ciaa953.
36. Klimstra, W. B. *et al.* SARS-CoV-2 growth, furin-cleavage-site adaptation and neutralization using serum from acutely infected hospitalized COVID-19 patients. *J. Gen. Virol.* **101**, 1156–1169 (2020).
37. Davidson, A. D. *et al.* Characterisation of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein. *Genome Med.* **12**, 68 (2020).
38. 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.* **17**, e1009233 (2021).
39. Zhu, Y. *et al.* A genome-wide CRISPR screen identifies host factors that regulate SARS-CoV-2 entry. *Nat. Commun.* **12**, 961 (2021).
40. 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).
41. Menachery, V. D. *et al.* A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat. Med.* **21**, 1508–1513 (2015).
42. Ge, X. *et al.* Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. *J. Virol.* **86**, 4620–4630 (2012).

43. Wan, Y., Shang, J., Graham, R., Baric, R. S. & Li, F. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *J. Virol.* **94**, e00127–20 (2020).
44. Bao, L. *et al.* The pathogenicity of SARS-CoV-2 in hACE2 transgenic mice. *Nature* **583**, 830–833 (2020).
45. Israelow, B. *et al.* Mouse model of SARS-CoV-2 reveals inflammatory role of type I interferon signaling. *J. Exp. Med.* **217**, e20201241 (2020).
46. Rathnasinghe, R. *et al.* Comparison of transgenic and adenovirus hACE2 mouse models for SARS-CoV-2 infection. *Emerg. Microbes Infect.* **9**, 2433–2445 (2020).
47. Hassan, A. O. *et al.* A SARS-CoV-2 infection model in mice demonstrates protection by neutralizing antibodies. *Cell* **182**, 744–753.e4 (2020).
48. Sun, S.-H. *et al.* A mouse model of SARS-CoV-2 infection and pathogenesis. *Cell Host Microbe* **28**, 124–133.e4 (2020).
49. Dinnon, K. H., 3rd *et al.* A mouse-adapted model of SARS-CoV-2 to test COVID-19 countermeasures. *Nature* **586**, 560–566 (2020).
50. Leist, S. R. *et al.* A mouse-adapted SARS-CoV-2 induces acute lung injury and mortality in standard laboratory mice. *Cell* **183**, 1070–1085.e12 (2020).
51. Gu, H. *et al.* Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science* **369**, 1603–1607 (2020).
52. Sun, S. *et al.* Characterization and structural basis of a lethal mouse-adapted SARS-CoV-2. *bioRxiv* 2020.11.10.377333 (2020) doi:10.1101/2020.11.10.377333.
53. Liu, Y. *et al.* The N501Y spike substitution enhances SARS-CoV-2 transmission. *bioRxiv* (2021) doi:10.1101/2021.03.08.434499.
54. Khan, A. *et al.* Higher infectivity of the SARS-CoV-2 new variants is associated with K417N/T, E484K, and N501Y mutants: An insight from structural data. *J. Cell. Physiol.* (2021) doi:10.1002/jcp.30367.
55. Kuzmina, A. *et al.* SARS-CoV-2 spike variants exhibit differential infectivity and neutralization resistance to convalescent or post-vaccination sera. *Cell Host Microbe* **29**, 522–528.e2 (2021).
56. Starr, T. N. *et al.* Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* **182**, 1295–1310.e20 (2020).
57. Zhan, S. H., Deverman, B. E. & Chan, Y. A. SARS-CoV-2 is well adapted for humans. What

does this mean for re-emergence? *bioRxiv* 2020.05.01.073262 (2020) doi:10.1101/2020.05.01.073262.

58. Otto, S. P. *et al.* The origins and potential future of SARS-CoV-2 variants of concern in the evolving COVID-19 pandemic. *Curr. Biol.* (2021) doi:10.1016/j.cub.2021.06.049.
59. Simmonds, P. Rampant C→U hypermutation in the genomes of SARS-CoV-2 and other coronaviruses: causes and consequences for their short- and long-term evolutionary trajectories. *mSphere* **5**, e00408–20 (2020).
60. Korber, B. *et al.* Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* **182**, 812–827.e19 (2020).
61. Volz, E. *et al.* Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell* **184**, 64–75.e11 (2021).
62. Cai, Y. *et al.* Structural basis for enhanced infectivity and immune evasion of SARS-CoV-2 variants. *Science* (2021) doi:10.1126/science.abi9745.
63. Zhu, X. *et al.* Cryo-electron microscopy structures of the N501Y SARS-CoV-2 spike protein in complex with ACE2 and 2 potent neutralizing antibodies. *PLoS Biol.* **19**, e3001237 (2021).
64. Xiao, K. *et al.* Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature* **583**, 286–289 (2020).
65. Dicken, S. J. *et al.* Characterisation of B.1.1.7 and Pangolin coronavirus spike provides insights on the evolutionary trajectory of SARS-CoV-2. *bioRxiv* (2021) doi:10.1101/2021.03.22.436468.
66. Conceicao, C. *et al.* The SARS-CoV-2 Spike protein has a broad tropism for mammalian ACE2 proteins. *PLoS Biol.* **18**, e3001016 (2020).
67. Oude Munnink, B. B. *et al.* Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science* **371**, 172–177 (2021).
68. van Aart, A. E. *et al.* SARS-CoV-2 infection in cats and dogs in infected mink farms. *Transbound. Emerg. Dis.* (2021) doi:10.1111/tbed.14173.
69. de Haan, C. A. M. *et al.* Cleavage of group 1 coronavirus spike proteins: how furin cleavage is traded off against heparan sulfate binding upon cell culture adaptation. *J. Virol.* **82**, 6078–6083 (2008).
70. Gombold, J. L., Hingley, S. T. & Weiss, S. R. Fusion-defective mutants of mouse hepatitis virus A59 contain a mutation in the spike protein cleavage signal. *J. Virol.* **67**, 4504–4512

(1993).
71. Kirchdoerfer, R. N. *et al.* Pre-fusion structure of a human coronavirus spike protein. *Nature* **531**, 118–121 (2016).
72. Gallaher, W. R. A palindromic RNA sequence as a common breakpoint contributor to copy-choice recombination in SARS-COV-2. *Arch. Virol.* **165**, 2341–2348 (2020).
73. Follis, K. E., York, J. & Nunberg, J. H. Furin cleavage of the SARS coronavirus spike glycoprotein enhances cell-cell fusion but does not affect virion entry. *Virology* **350**, 358–369 (2006).
74. Belouzard, S., Chu, V. C. & Whittaker, G. R. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc. Natl. Acad. Sci. USA* **106**, 5871–5876 (2009).
75. Peacock, T. P. *et al.* The SARS-CoV-2 variants associated with infections in India, B.1.617, show enhanced spike cleavage by furin. *bioRxiv* 2021.05.28.446163 (2021) doi:10.1101/2021.05.28.446163.
76. Peacock, T. P. *et al.* The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. *Nat Microbiol* (2021) doi:10.1038/s41564-021-00908-w.
77. Maxmen, A. & Mallapaty, S. The COVID lab-leak hypothesis: what scientists do and don't know. *Nature* **594**, 313–315 (2021).