

Natural and unnatural history of the coronavirus: The uncertain path to the pandemic

P. Balaram

There is currently an ongoing debate on the origins of the SARS-CoV-2 coronavirus, the causative agent of the COVID-19 pandemic. The hypothesis that the virus emerged as a result of natural zoonotic transfer has been controversially questioned, on the basis of unusual sequence signatures in the spike protein and the absence of evidence for an intermediate animal host. This article briefly surveys the background to coronavirus research and addresses the development of the methods of gain of function research, which have been suggested to be a potential public health threat in the event of laboratory leakage.

Keywords: Coronavirus, gain of function research, SARS-CoV-2, spike protein sequences.

What immortal hand or eye,

Could frame thy fearful symmetry?

– William Blake (1794) *The Tyger*

Prologue

IN the late 18th century, the poet William Blake reflected on the origins of the tiger and the lamb, a subject that could well be a topic of debate amongst evolutionary biologists and theologians, today. In considering the origins of the coronavirus, SARS-CoV-2, whose menacingly symmetrical images are now imprinted in our imagination, we can ask as Blake did over two centuries ago: ‘*Did he who make the lamb make thee?*’.

The coronavirus first made its appearance in the scientific literature in 1968, in the summary of a letter written to the journal *Nature*¹, where a group of virologists suggested this new name on the basis of the appearance of the virus in the first electron micrograph published in 1967 (ref. 2). That historical image was obtained of a viral isolate 229E, characterized from the nasal secretions of medical students at the University of Chicago by Dorothy Hamre³. In the years that followed coronaviruses were shown to be responsible for a significant number of relatively mild upper respiratory tract infections, which present symptoms of the common cold. Similar viruses had already been characterized in veterinary infections of chickens and pigs. In 1990, David Tyrrell, well known for his extensive and ultimately unfruitful search for a vaccine against the common cold⁴, concluded the first phase of coronavirus research by saying: ‘Coronaviruses

cause acute, mild upper respiratory infection (common cold)’⁵. Tyrrell could not have been more wrong. Beginning in late 2002, a mysterious new ‘flu’ had begun to emerge in the Guangdong province in China. The first sign of the emergence of a new disease came when a call went out to the WHO about a severely ill patient in Hanoi. Carlo Urbani, a heroic doctor, whose name is now indelibly linked to the discovery of the Severe Acute Respiratory Syndrome (SARS), identified and raised the alarm on 28 February 2003, about a new, most dangerous and easily transmissible human infectious disease. The disease was notified by the WHO on 12 March 2003, a sharp contrast to the organization’s reactions in 2020. A few weeks later Urbani sadly died of the infection in a hospital in Bangkok, on 29 March 2003 (ref. 6). The SARS outbreak of 2003 died down quickly but the statistics were alarming. Of the approximately 8000 persons infected, about 800 died, an unacceptable fatality rate⁷. In reflecting on SARS in 2008, Baric underscored ‘the critical need for maintaining active basic science research not only on the medically relevant human pathogens, but also on virus families that are associated with limited or benign disease outcomes in humans’⁸.

Their finest hour

The remarkable response of virologists, immunologists and structural biologists to the emergence of a new viral pathogen led to an extremely productive phase of research in the years immediately following the SARS-1 outbreak. The causative agent, the SARS-CoV-1 virus, was isolated and its sequence determined in record time with the full genomic sequence appearing in May 2003 (refs 9, 10). By October 2003 the virus had been detected in a bat species¹¹. In less than a year the stage was set for

P. Balaram is in the National Centre for Biological Sciences, Bengaluru 560 065, India.
e-mail: pb@iisc.ac.in

a remarkable decade or more of research on the coronavirus, with almost every aspect of the structure and function of its protein components established. Bats were identified as the natural reservoir of the virus, with initial transmission to an intermediate animal host, the Himalayan palm civet, found in the live animal market of Guangdong, China¹². Natural zoonotic transfer had been established beyond doubt. Science indeed rose to the occasion in 2003, one of virology's finest hours⁸. But the pandemic was localized and died down quickly. The interest in diagnostics, therapeutics and vaccines was now largely academic. The outbreak in the Middle East (MERS) in 2007 sparked interest, with fairly rapid identification of the camel as the intermediate host¹³. Quick containment resulted in a rapid disappearance of the coronavirus from public consciousness, until the dramatic and breathtakingly rapid developments of the period between December 2019 and March 2020. No intermediate animal host has been shown for the present virus. The suggestion that pangolins from the Wuhan wet market might have been the intermediate host has not been supported by an analysis of live animal sales from the Wuhan market during the period November 2019 and March 2020; no record of pangolins or bats being sold was found¹⁴.

The needle of suspicion

'We are always paid for our suspicion by finding what we suspect.'

– Henry David Thoreau

In January 2020, the group at the Wuhan Institute of Virology (WIV) published the genomic sequence of the new virus, SARS-CoV-2 (ref. 15). The pandemic was not yet officially declared, but the writing seemed to be on the wall as infections spread in Wuhan and surfaced with alarming rapidity in Italy and very soon all over the world. An infectious agent, surprisingly transmissible, but curiously leaving a vast majority of human hosts asymptomatic had begun to prey on the unsuspecting population of the world. Inevitably, the sequence of the coronavirus SARS-CoV-2 was subject to intense scrutiny by both dedicated sequence analysts and many amateur sleuths, with the intention of uncovering the origins of the virus^{16–18}. Zoonotic transfer would imply that there might be an intermediate host through which the bat virus would pass to its human destination. None was found, but such a negative result could have many interpretations, with the preferred one being if we wait long enough the intermediate host will be found. Direct virus transfer from bats to humans is another possibility but no compelling evidence is available. Can an evolutionary precursor sequence be identified and a model of recombination events be visualized, which will allow the transformation of a bat virus into a human pathogen?

Very early on after the release of the sequence of SARS-CoV-2, features of the spike protein attracted the attention of sequence watchers. This possibility that the virus may have the imprint of laboratory manipulation was indeed suggested in postings on bioRxiv, the repository for preprints. This raised the question: Can a bat genome sequence be identified, which serves as the template for engineering an efficient, infectious agent targeting human hosts? Initial doubts on the possible laboratory origin of the virus were dismissed as 'conspiracy theories', which are undoubtedly very popular amongst those who see an evil hand at every turn. As the death toll in the United States mounted, politics took over. Conspiracy theories assumed a decidedly Trumpian flavour. The proponents of natural evolution could easily dismiss and indeed silence any dissent, without the need to provide detailed scientific counter arguments^{19,20}, to rule out deliberate engineering of a pre-existing bat virus. Both evolution and laboratory manipulation might leave tell-tale signatures in the sequences of the SARS-CoV-2 virus, although protocols might be found to obscure the traces of deliberate engineering. Of the many protein components of the coronavirus, the most investigated has been the 'spike protein', which constitutes the characteristic antenna-like projections that dot the spherical exterior of the virus. It is here that much attention has been focused with one feature standing out, the furin cleavage loop^{16–18}, that I detailed in a previous Commentary²¹. If modified to enhance one of the key steps of viral-host membrane fusion, proteolysis of the spike protein into two pieces, one might anticipate an increase in infectivity. The choice of sequence to be inserted might be rationalized in a speculative thought experiment as outlined earlier²¹. To engineer a genome there must be a template genome and the appropriate protocols for effecting the desired manipulation of a large viral RNA genome (approximately 29000 nucleotides) and an animal model for characterizing the infectivity of the recombinant, engineered virus. These hurdles need to be surmounted by a great deal of research. Was all this done at the Wuhan Institute of Virology by Chinese scientists, under a cloak of secrecy which left the Western scientific establishment completely unaware of the progress made? Or was every potential protocol already well developed and shared in a collaborative 'gain of function' programme, designed to construct a more transmissible virus?

The trail of research

'Keep your friends close. Keep your enemies closer.'

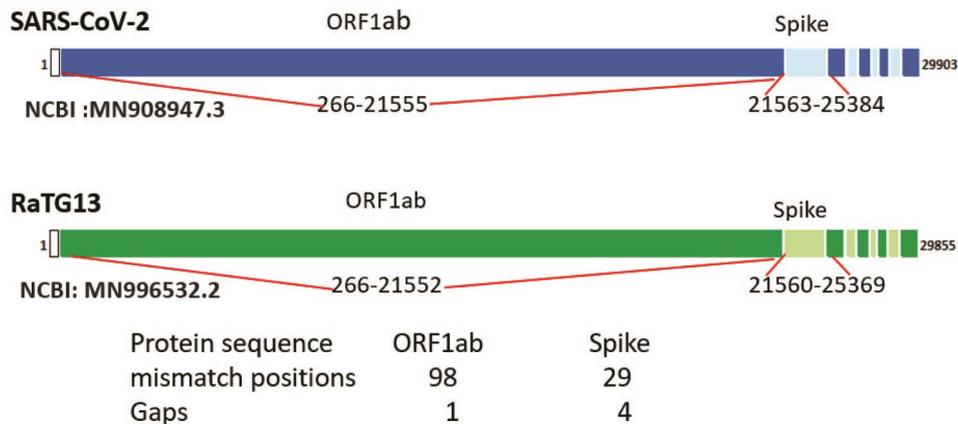
– Michael Corleone in *The Godfather II*

Virologists in the years following the SARS outbreak of 2003 focused their attention on this new, invisible and dangerous enemy. In a *tour de force* of genomic

technologies, Ralph Baric at the University of North Carolina (UNC) assembled the entire RNA genome of the 2003 SARS-CoV-1 virus, as early as August 2003 and demonstrated that expression, in human host cells in culture, yielded infective viral particles. The rationale for undertaking this large project was clearly outlined: ‘*The availability of a full-length cDNA of the SARS genome should allow for genetic manipulation of the replicase gene providing new insights into the role of specific proteolytic cleavages and replicase proteins during viral replication*’²². The stage was set to dissect the mechanisms of viral-host fusion *in vitro*. But to go beyond host cells in culture, an animal model was needed. The receptor for the SARS-CoV-1 virus had been identified very quickly after the pathogens emergence as the angiotensin converting enzyme (ACE2), a surface membrane bound protein on human host cells²³. It was now necessary to develop a transgenic mouse model, in which the mouse carried the human ACE2 receptor on its cell surfaces, a procedure that ‘humanizes’ mice making them surrogate hosts for viruses that infect humans. In 2005, this was achieved with transgenic mice proliferating the coronavirus strain 229E (ref. 24). In a commentary highlighting this advance, Baric notes that ‘*transgenic animals may well serve as a more robust model for SARS-CoV*’²⁵. The ‘humanized mice’ model quickly became available for SARS research. In 2015, the collaborative programme between the Baric laboratory at UNC Chapel Hill and the Zhengli Shi laboratory at the Wuhan Institute of Virology resulted in a paper in *Nature Medicine*²⁶, whose abstract merits reproduction. ‘*Here we examine the disease potential of a SARS-like virus, SHC014-CoV, which is currently circulating in Chinese horseshoe bat populations. Using the SARS-CoV reverse genetics system, we generated and characterized a chimeric virus expressing the spike of bat coronavirus SHC014 in a mouse-adapted SARS-CoV backbone. The results indicate that group 2b viruses encoding the SHC014 spike in a wild-type backbone can efficiently use multiple orthologs of the SARS receptor human angiotensin converting enzyme II (ACE2), replicate efficiently in primary human airway cells and achieve in vitro titers equivalent to epidemic strains of SARS-CoV. Additionally, in vivo experiments demonstrate replication of the chimeric virus in mouse lung with notable pathogenesis. Evaluation of available SARS-based immune-therapeutic and prophylactic modalities revealed poor efficacy; both monoclonal antibody and vaccine approaches failed to neutralize and protect from infection with CoVs using the novel spike protein. On the basis of these findings, we synthetically re-derived an infectious full-length SHC014 recombinant virus and demonstrate robust viral replication both in vitro and in vivo*’. A careful reading of this abstract, even by lay readers (and I count myself amongst them) might begin to raise flags. Those who doubt whether SARS viruses can be engineered would do well to read the paper and the fine print.

Genomic technologies are far advanced and laboratory manipulation of sequences is accessible in many centres in the West and presumably in China. The potential human pathogens engineered and studied in the 2015 report, were undoubtedly constructed under stringent safety protocols approved in the United States. The last sentence declares the ability to generate synthetic, recombinant viruses capable of infection. This American–Chinese collaboration presumably set the stage for outsourcing to Wuhan controversial ‘gain of function’ experiments, once the ban on such experiments came into effect in the USA. Curiously, a corrigendum to this article²⁷ declares the source of Wuhan funding: ‘*In the version of this article initially published online, the authors omitted to acknowledge a funding source, USAID-EPT-PREDICT funding from EcoHealth Alliance, to Z.-L.S.*’. On 30 March 2020, five years after publication of this paper, *Nature Medicine* put out the following Editors’ note: ‘*We are aware that this article is being used as the basis for unverified theories that the novel coronavirus causing COVID-19 was engineered. There is no evidence that this is true; scientists believe that an animal is the most likely source of the coronavirus.*’ As whispers of ‘conspiracy theories’ grew louder in the back alleys of the Internet, a curious corrigendum²⁸ appeared in the journal, on 20 May 2020, when the pandemic was in full swing: ‘*In the version of this article initially published, the sequence of the mouse adapted SHC015-MA15 virus had not been deposited in GenBank. The sequence has now been deposited in GenBank under accession number MT308984*’. This can only be interpreted as an action compelled by demands that the sequence be available to examine its relationship to SARS-CoV-2. This was a powerful and promiscuous pathogen as the authors note: ‘*Similarly to SARS, SHC014-MA15 also required a functional ACE2 molecule for entry and could use human, civet and bat ACE2 orthologs*’. Another addendum appeared from the Chinese group late in 2020. A bat sequence RaTG13 had been mentioned as many as 11 times in the original *Nature* article¹⁵, its sequence being the one closest to the new virus 2019-nCoV, now renamed as SARS-CoV-2. The origins of RaTG13 were not mentioned. In an addendum published online on 17 November 2020, presumably to bolster the zoonotic origin of the agent of the exploding pandemic, the Wuhan group divulged the origins of the sample, as collected in mid-2012, from patients suffering from severe respiratory illness in the Yunnan province, after cleaning an abandoned mine shaft of bat faeces^{29,30}. It is to RaTG13 that we must now turn. Box 1 shows schematic alignment of the RaTG13 and SARS-CoV-2 genomes. The identities are striking. Even more striking is the close sequence identity of the coronavirus spike protein.

The virus RaTG13 was presumably not cultured in the laboratory, but a sequence was available. Was this now the template for virus engineering? A synthetic construct

Box 1. Schematic comparison of two coronavirus genomes.

The genome lengths are SARS-CoV-2, 29903 nucleotides and RaTG13, 29885 nucleotides. The protein coding regions (ORFs) are shown as alternating dark and light-coloured segments. The number of positions of amino acid mismatch and gaps in the two major ORFs are indicated. There is an overall 96.1% identity in the two virus genomes.

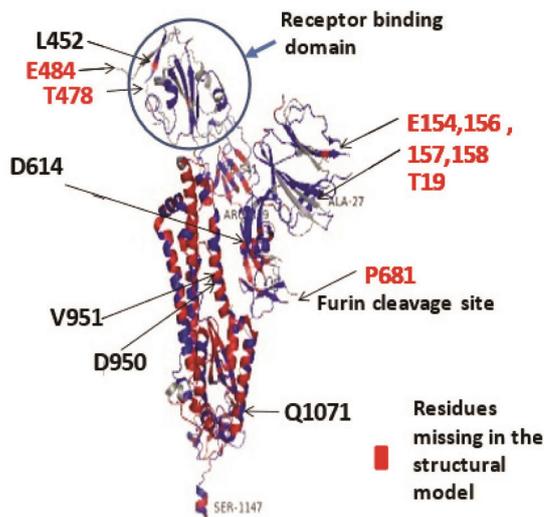
could easily have been generated, as the papers of 2003 (ref. 22) and 2015 (ref. 26) have shown. The genomes share an identity of 96.1%, which establishes that 3.9% of the nucleic acid sequence is different, a variation that some might say is large enough to cast doubt on the hypothesis that RaTG13 may have served as a starting point for laboratory manipulation. As discussed in the earlier commentary²¹, an attractive site for ‘gain of function’ studies would be to enhance the cleavability of the site for furin-mediated proteolysis. It has long been known that trypsin treatment of the influenza virus enhances infectivity in cell-based assays³¹. This simple pre-treatment procedure has been extended to coronaviruses leading to the conclusion that ‘trypsin treatment can unlock the zoonotic barrier’. This study indeed suggests that ‘proteolytic cleavage of the spike, not receptor binding, is the primary infection barrier’³². It is a reasonable conjecture that if a bat sequence was to be used as a template for ‘gain of function’ research, engineering the highly susceptible furin proteolytic site would be a good option. It is a matter of interest that the more transmissible mutant virus, now labelled as the *delta* strain, carries a mutation at the furin site, making it more susceptible to proteolysis, as demonstrated in a recent report³³. The *delta* strain appears to be the more widely circulating pathogen in the brutal second phase of the pandemic in India (see Box 2).

The gathering storm

The tools and technologies for engineering a virus are available with the collaborating laboratories in the scenario

outlined. Concerns about gain of function research have been often expressed. The funding by US agencies of research at Wuhan may well have been done to circumvent the regulatory restrictions in the US³⁴. The NIH imposed a moratorium on gain of function research in October 2014, which was lifted in December 2017. In a news report in *Nature*³⁵, the editors found it fit to highlight the fact that such studies ‘risk creating an accidental pandemic’. Prominent virologists, however, thought otherwise. Indeed, in a strongly argued editorial in 2018, Sheehan and Baric ask: ‘Is regulation preventing the development of therapeutics that may prevent future coronavirus pandemics?’ They conclude: ‘Despite the best intentions, one of the greatest risks of biosafety policy is unintended over-reach that limits our understanding of emerging infectious disease biology and prevents countermeasure development that could limit or prevent future pandemics’³⁶. Ironically, the proponents of the ‘lab leak’ hypothesis of the COVID-19 pandemic point to inadequate safety measures in the Wuhan laboratory, from where the infectious agent, natural or unnatural, may have escaped. In searching the PUBMED database for papers from the Wuhan and North Carolina laboratories, I came across a striking fact. Ralph Baric was listed as an author in as many as 114 papers in the period January 2020 to June 2021. Zhengli Shi was less prolific but her tally of 34 papers, in the same pandemic-dominated period, must be considered impressive, by any standard of scientific productivity. The virus has certainly provided ample fodder for virologists. Even while American politicians and media point fingers at Wuhan, it might be necessary to ask some hard questions about ‘gain of

Box 2. Coronavirus Spike Protein Mutations and the *Delta* Variant.



MN908947.3	<i>cct</i>	<i>cgg</i>	<i>cgg</i>	<i>gca</i>	<i>cgt</i>
PODTC2.1 <i>China</i>	P	R	R	A	R
MZ336033.1	<i>cct</i>	<i>cgg</i>	<i>cgg</i>	<i>gca</i>	<i>cgt</i>
QWE36170.1 <i>Bengaluru</i>	P	R	R	A	R
MZ350188.1	<i>cgt</i>	<i>cgg</i>	<i>cgg</i>	<i>gca</i>	<i>cgt</i>
QWF08991.1 <i>Gujarat</i>	R	R	R	A	R
MZ342815.1	<i>cat</i>	<i>cgg</i>	<i>cgg</i>	<i>gca</i>	<i>cgt</i>
QWE77830.1 <i>Gujarat</i>	H	R	R	A	R

The WHO classification of SARS-CoV-2 mutations prevalent in India is as follows: B.1.617: L452R, E484Q, D614G; B.1. 617.1 (*kappa*): T951I, G142D, E154K, L452R, E484Q, D614G, P681R, Q1071H; B1.617.2 (*delta*): T19R, G142D, 156 del, 157 del, R158G, L452R, T478K, D614G, P681R, D950N (See CDC site).

The positions of mutational variation are marked on the structure of the spike protein. The receptor binding domain (blue circle) and the furin cleavage site are marked. Many sites of mutation lie in the flexible, disordered regions of the structure. The colour coding for the ribbon representation of the protein is: blue, regions of high mutational variability and red, relatively conserved segments determined from an analysis of over 3000 spike proteins sequences infecting a wide range of animal hosts. Furin cleavage breaks the protein into two domains S1 (amino terminus) and S2 (carboxy terminus). Significantly greater sequence variability is observed in domain S1, which also harbours the receptor binding site.

The table compares a representative set of furin site sequences for samples sequenced in India, with the original sequence deposited in the NCBI database from China in January 2020 at the start of the pandemic. Accession numbers for both nucleic acid and protein sequences are listed. Furin site cleavage mutants which may lead to greater transmissibility are highlighted in red. The Bengaluru sample (Bangalore Medical College and Research Institute) is dated 2020 (month unavailable). The Gujarat samples (Gujarat Biotechnology Research Centre) are dated 2 April 2021 and 16 March 2021.

function' virus research, in American laboratories and possibly elsewhere.

A possible path to virus engineering based on published literature has been outlined above, an imaginary scenario to account for the appearance of an infectious agent that has devastated millions of lives. There is an alternative. Scenarios invoking coinfection of a host by two viral strains and subsequent recombination events could be invoked. Tell-tale signatures in the new genome may provide evidence. A Chinese–American collaborative study which includes both Zhengli Shi and Peter Daszak as authors reports an analysis of over 600 bat coronavirus sequences³⁷. A very recent paper, that has appeared at the time of writing, in the journal *Cell* reports an extensive phylogenetic analysis of bat sequences, which may presumably help in the search for the closest relatives of SARS-CoV-2 (ref. 38). Thus far, no compelling scientific evidence has been forthcoming, that per-

mits a distinction between the two hypotheses for the evolution of the virus, natural and unnatural. The credibility of the letters by groups of influential scientists in journals like *Lancet*¹⁹ and *Nature Medicine*²⁰, that seek to discredit those who may raise doubts about the natural origin of the virus, has been eroded by some signatories now supporting an investigation. The reported originator of the *Lancet* letter, Peter Daszak, was a collaborator of the Wuhan Laboratory and strangely enough a member of the team sent by WHO to Wuhan to investigate the origins of the pandemic. A video interview with Daszak on the sidelines of a conference on Nipah viruses in December 2019 is worth viewing, although bat virus sequencing appears only around 27 minutes into the recording³⁹. Curiously, a recent letter in *Science*, calling for an investigation, has Ralph Baric, one of the most prominent advocates of 'gain of function' research and a collaborator of the Wuhan laboratory, as a signatory⁴⁰. As recently as July

2020, a report from the Wuhan group with Baric as a co-author reports a study of the pathogenicity of the pandemic virus SARS-CoV-2, in mice carrying the human ACE 2 receptor⁴¹. The paper concludes optimistically even as the global crisis deepened: ‘*The successful establishment of an animal model for COVID-19 pathogenesis will be valuable for evaluating vaccines and therapeutics to combat SARS-CoV-2*’. The debates on ‘gain of function’ research on viruses are not new. In 2005, the 1918 influenza virus genome was pieced together and artificially constructed in the laboratory. In the run-up to the publication of the paper in *Science*⁴², the National Institutes of Health and the US government were involved in deciding that such a paper need not be classified as a threat to national security. The discussions were held in the aftermath of the 9/11 terrorist attack and the concerns about bioterrorism. In such discussions there are, of course, arguments on weighing the costs versus the benefits of such research. Revisiting those debates may be instructive in charting a course for the future, in the light of the lessons learnt during the COVID-19 pandemic^{43–46}.

Epilogue

The virus has revealed a number of fault lines in the structures of science. The premier scientific journals have silenced debate and discussion. In consigning a scientific debate to the dark alleys of the Internet and the parallel world of social media, the major journals which project themselves as gatekeepers for the imagined purity of the biological and biomedical literature have not enhanced their reputations, even if their ‘impact factors’ remain on an upward trend. Labelling plausible scenarios for origins of the virus as ‘conspiracy theories’ and denying early doubters a place to present their, albeit, meagre evidence, does a great disservice to a world that has been affected as never before in living memory. In managing the COVID-19 disaster, natural or man-made, many world leaders have demonstrated that arrogance, ignorance and hubris are common qualities in powerful leaders. Many prominent scientists have not been far behind in displaying these same qualities, in suppressing discussions on the origins of the virus and in vigorously pursuing gain of function research, despite a temporary ban on such research a few years ago. A thoughtful and concerned reader suggested, after reading my earlier commentary on this issue²¹, that such discussions may damage institutionalized science and destroy public faith in genetic engineering. We must remember the successes of recombinant DNA technologies and the spectacular contributions made by modern biology to medicine. If discussions of contentious topics are avoided then there is little difference between science and religion. The doubters must be content to accept received wisdom, unquestioningly. Seventy years ago, nuclear weapons attracted fierce

debate. The most enduring image of the revolution in physics in the 20th century was the mushroom cloud over the New Mexico desert, after the first atomic explosion in 1945. The man who led the Manhattan Project, Robert Oppenheimer, famously quoted from the Bhagavad Gita: ‘*I am become death, the destroyer of worlds*’. While the accuracy of the translation from Sanskrit has been debated, Oppenheimer’s imagery captured the moment for posterity. If the laboratory origins of the virus are established, the defining image of the revolution in biology, over the last few decades, may well be the coronavirus, which has truly brought death to all our doorsteps.

The term ‘engineering’ is a misnomer when applied to biology. A natural virus removed from its animal reservoir and treated or engineered to break the zoonotic barrier and infect human hosts, when leaked from a laboratory, will eventually become a part of nature, subject to random mutation, variation and natural selection. Irrespective of its origins, mutant strains of the SARS-CoV-2 virus are already circulating. A virus inhabits the shadowy no-man’s land between chemistry and biology. ‘Engineering’ individual genes and proteins have provided great benefit. It is not clear that ‘engineering’ viral genomes will provide similar benefit. Recombinant viruses will become a part of nature. We must remind ourselves that there may be no such thing as a benign virus. The German poet Wolfgang von Goethe’s words, probably written in 1780, famously translated from German and reproduced by Thomas Huxley, in the first editorial launching the journal *Nature* in 1869, may well be worth recalling: ‘*Nature! We are surrounded and embraced by her: powerless to separate ourselves from her, yet powerless to penetrate beyond her. She is ever shaping new forms: what is, has never yet been; what has been, comes not again. Everything is new, yet not but the old.*’

1. News and Views, Virology: Coronaviruses. *Nature*, 1968, **220**, 650; doi:10.1038/220650b0
2. Almeida, J. D. and Tyrrell, D. A. J., The morphology of three previously uncharacterised respiratory viruses that grow in organ culture. *J. Gen. Virol.*, 1967, **1**, 175–178.
3. Hamre, D. and Procknow, J. J., A new virus isolated from the human respiratory tract. *Proc. Soc. Exptl. Biol. Med.*, 1966, **121**, 190–193; doi:10.3181/00379727-121-30734.
4. Tyrrell, D. A. J. and Myint, S. H., Coronaviruses. In *Medical Microbiology* (ed. Baron, S.), University of Texas, Galveston, 1996, 4th edn, ch. 60.
5. Tyrrell, D. A. J. and Fiedler, M., *Cold Wars: The Fight Against the Common Cold*, Oxford University Press, 2002, p. 268.
6. Reilley, B. *et al.*, SARS and Carlo Urbani. *N. Engl. J. Med.*, 2003, **348**, 1951–1952; doi:10.1056/NEJMp030080.
7. Cherry, J. D. and Krogstad, P., SARS: the first pandemic of the 21st century. *Paediatric Res.*, 2004, **56**, 1–5; doi:10.1203/01.PDR.0000129184.87042.FC.
8. Baric, R. S., Lessons for global health. *Virus Res.*, 2008, **133**, 1–3.
9. Marra, M. A. *et al.*, The genome sequence of the SARS-associated coronavirus. *Science*, 2003, **300**, 1399–1404.
10. Rota, P. A. *et al.*, Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science*, 2003, **300**, 1394–1398.

11. Susanna, K. P. *et al.*, Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 14040–14045; <https://doi.org/10.1073/pnas.0506735102>
12. Guan, Y. *et al.*, Isolation and characterisation of viruses related to the SARS coronavirus from animals in Southern China. *Science*, 2003, **302**, 276–278.
13. Xiao, X. *et al.*, Animal sales from Wuhan wet markets immediately prior to the COVID19 pandemic. *Scientif. Rep.*, 2021, **11**, 11898; <https://doi.org/10.1038/s41598-021-91470-2>
14. Cui, J., Li, F. and Shi, Z.-L., Origin and evolution of pathogenic coronaviruses. *Nature Revs. Microbiol.*, 2019, **17**, 181–192.
15. Zhou, P., A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 2020, **579**, 270–273.
16. Segreto, R. *et al.*, Should we discount the laboratory origin of COVID-19? *Environ. Chem. Lett.*, 2021; <https://doi.org/10.1007/s10311-021-01211-0>
17. Wade, N., The origin of COVID: did people or nature open Pandora's box at Wuhan? *Bulletin of the Atomic Scientists*, 5 May 2021; <https://thebulletin.org/2021/05/the-origin-of-covid-did-people-or-nature-open-pandoras-box-at-wuhan>
18. Jacobsen, R., How amateur sleuths broke the Wuhan Lab Story and embarrassed the media. *Newsweek*, 2 June 2021; <https://www.newsweek.com/exclusive-how-amateur-sleuths-broke-wuhan-lab-story-embarrassed-media-1596958>
19. Calisher, C. *et al.*, Statement in support of the scientists, public health professionals, and medical professionals of China combating COVID-19. *Lancet*, 2020, **395**, e42–e43.
20. Andersen, K. G. *et al.*, The proximal origin of SARS-CoV-2. *Nature Med.*, 2020, **26**, 450–455.
21. Balaram, P., The murky origins of the coronavirus SARS-CoV-2, the causative agent of the COVID-19 pandemic. *Curr. Sci.*, 2021, **120**(11), 1663–1666.
22. Yount, B. *et al.*, Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus. *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 12995–13000.
23. Li, W. *et al.*, Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*, 2003, **426**, 450–454; doi:10.1038/nature02145.
24. Lassnig, C. *et al.*, Development of a transgenic mouse model susceptible to human coronavirus 229E. *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 8275–8280.
25. Baric, R. S. and Sims, A. C., Humanized mice develop coronavirus respiratory disease. *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 8073–8074.
26. Menachery, V. D. *et al.*, A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nature Med.*, 2015, **21**, 1508–1513.
27. Menachery, V. D. *et al.*, Corrigendum. *Nature Med.*, doi:10.1038/nm.3985 (corrected 20 November 2015).
28. Menachery, V. D. *et al.*, Author correction: A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nature Med.*, Published online: 22 May 2020; <https://doi.org/10.1038/s41591-020-0924-2>.
29. Zhou, P. *et al.*, Addendum: A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 2020, **588**, E6; <https://doi.org/10.1038/s41586-020-2951-z> (17 November 2020).
30. Rahalkar, M. C. and Bahulikar, R. A., Lethal pneumonia cases in Mojhiang miners (2012) and the mineshaft could provide important clues to the origin of SARS-CoV-2. *Frontiers in Public Health*, October 2020; <https://www.frontiersin.org/articles/10.3389/fpubh.2020.581569/full>
31. Klenk, H.-D. *et al.*, Activation of influenza A viruses by trypsin treatment. *Virology*, 1975, **68**, 426–439.
32. Menachery, V. D. *et al.*, Trypsin treatment unlocks barrier for Zoonotic Bat Coronavirus Infection. *J. Virology*, 2020, **94**(5), e01774-19.
33. 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*; <https://doi.org/10.1101/2021.05.28.446163>
34. Lin, C., Why US outsourced bat virus research to Wuhan. *ISPSW Strategy Series: Focus on Defense and International Security*, April 2020, issue 689; https://www.ispsw.com/wpcontent/uploads/2020/04/689_Lin.pdf
35. Reardon, S., Ban on pathogen studies lifted. *Nature*, 2018, **552**, 11; <https://media.nature.com/original/magazine-assets/d41586-017-08837-7/d41586-017-08837-7.pdf>
36. Sheahan, T. P. and Baric, R. S., Is regulation preventing the development of therapeutics that may prevent future coronavirus pandemics? *Future Virol.*, 2018, **13**, 143–146.
37. Latinne, A., Origin and cross-species transmission of bat coronaviruses in China. *Nature Commun.*, 2020, **11**, 4235; <https://doi.org/10.1038/s41467-020-17687-3>
38. Zhou, H. *et al.*, Identification of novel bat coronaviruses sheds light on the evolutionary origins of SARS-CoV-2 and related viruses. *Cell*, 2021 (in press); <https://doi.org/10.1016/j.cell.2021.06.008>
39. Daszak, P., *Interview*, December 2019; https://www.youtube.com/watch?v=IdYDL_RK--w
40. Bloom, J. D. *et al.*, Investigate the origins of COVID-19. *Science*, 2021, **372**, 694.
41. Jiang, R.-D. *et al.*, Pathogenesis of SARS-CoV-2 in transgenic mice expressing human angiotensin-converting enzyme 2. *Cell*, 2020, **182**, 50–58.
42. Tumpey, T. M. *et al.*, Characterization of the reconstructed 1918 Spanish influenza pandemic virus. *Science*, 2005, **310**, 77–80.
43. Sharp, P. A., Editorial. 1918 Flu and responsible science. *Science*, 2005, **310**, 17.
44. Kennedy, D., Better never than late. *Science*, 2005, **310**, 195.
45. Imperiale, M. J., Howard, D. and Casadevall, A., The silver lining in gain-of-function experiments with pathogens of pandemic potential. In *Influenza Virus: Methods and Protocols, Methods in Molecular Biology* (ed. Yamauchi, Y.), 2018, vol. 1836, ch. 28, pp. 575–587; https://doi.org/10.1007/978-1-4939-8678-1_28
46. Taubenberger, J. K. *et al.*, Reconstruction of the 1918 influenza virus: unexpected rewards from the past. *mBio*, 2012, **3**, e00201–e00212.

ACKNOWLEDGEMENTS. I am indebted to M. Vijayarathy and N. V. Joshi for help with the Figures and providing the results of preliminary analysis of data on coronavirus genomes. I am grateful to Sahil Lall, Sunil Laxman and N. V. Joshi for their critiques of this manuscript at different stages of its evolution.

Received and accepted 15 June 2021

doi: 10.18520/cs/v120/i12/1820-1826