environmental microbiology

Environmental Microbiology (2020) **22**(6), 2001–2006 doi:10.1111/1462-2920.15053

Opinion

SARS-CoV-2 variants: Relevance for symptom granularity, epidemiology, immunity (herd, vaccines), virus origin and containment?

Antoine Danchin [1,2,3*] **and Kenneth Timmis**[4]

[1]*Kodikos Labs, 24 rue Jean Baldassini, Lyon 69007, France.*

[2]*Institut Cochin, Paris 75013, France.*

[3]*School of Biomedical Sciences, Li Kashing Faculty of Medicine, University of Hong Kong, 21 Sassoon Road, SAR Hong Kong, China.*

[4]*Institute of Microbiology, Technical University of Braunschweig, Braunschweig, Germany.*

Summary

The origin of the SARS-CoV-2 virus remains enigmatic. It is likely to be a continuum resulting from inevitable mutations and recombination events. These genetic changes keep developing in the present epidemic. Mutations tending to deplete the genome in its cytosine content will progressively lead to attenuation as a consequence of Muller's ratchet, but this is counteracted by recombination when different mutants co-infect the same host, in particular, in clusters of infection. Monitoring as a function of time the genome sequences in closely related cases is critical to anticipate the future of SARS-CoV-2 and hence of COVID-19.

We keep hearing or reading about the 'missing' index patient—Patient zero—who would explain everything about the origin of the contemporary scourge, the SARS-CoV-2 virus. Yet the earliest studies failed to identify any indisputable origin of the virus, after a first official notice from Wuhan authorities on December 31, 2019 (http://wjw.wuhan.gov.cn/front/web/showDetail/2019123108989), informed the public that a local epidemic of atypical pneumonia was developing. A major reason for the failure to know when or where the virus first appeared seems to have been the widespread occurrence of asymptomatic cases of COVID-19. This precluded identification of authentic first cases (Cowling and Leung, 2020).

The recurrent preoccupation with origins results from an omnipresent trend of human thought, which seeks a single origin of everything: the Origin of the Universe with a Big Bang—intended as a joke by Fred Hoyle—the Origin of the first cells with a LUCA—despite the fact that Freeman Dyson demonstrated that the emergence of the first cells required at least two origins (Dyson, 1985)—and of course, the Origin of Man. We are 'adamists' we keep searching for a single event as a cause of everything. This preconceived attitude is so strongly built in our minds—we love two-dimensional decision trees—that the way we construct phylogenies is systematically using dichotomies: evolving would mean choosing between two paths, and then two paths, ad infinitum, and this permits us to look for an origin, THE origin. But there is often more than one solution to an evolutionary advance, so why not a three-dimensional mesh-like source of living things?

This general trend would be just anecdotal if it did not have, in the present context, serious consequences. We look for the origin of SARS-CoV-2, because we see it as a single entity against which we could act using a series of established approaches, since the virus would follow a predictable route of contagion. The epidemic would resolve after 'herd immunity' had been created. Yet, SARS-CoV-2 is not a single entity (e.g., see Forster *et al*., 2020) and may not have a single origin: contemplating herd immunity with a heterogeneous population of viruses may be very misleading, at best. The same belief may be true when we think that we will soon have an efficacious vaccine soon—of course, we could be lucky, but being lucky is not the most probable outcome with evolving viruses. To progress beyond routine strategy, we must take the point of view of the virus and of its

Received 29 April, 2020; revised , ; accepted 29 April, 2020. *For correspondence.* E-mail antoine.danchin@normalesup.org, Tel. +331 4441 2551; Fax +331 4441 2559.

evolution and adapt our strategy to the corresponding insights gained, possibly even using the virus itself as a weapon against severe viral infections. This asks for some reflection about scenarios of origins. We must subsequently try and understand how a population of viruses evolves.

Scenarios for multiple origins

To understand the past and anticipate the future of SARS-CoV-2, we must think out of the box. A detour will help us see how wrong it is to ask for the necessity of a unique origin. This will allow us to evade the unfortunate consequences of contemplating only one virus, whereas there is a continuum of variants, likely to display a variety of disease phenotypes (e.g., see Yao *et al*., 2020). As an illustration, here is an answer to the straightforward question, that of the origin of *Homo sapiens*. Just a few days ago yet another study, based on the analysis of the genomes of Icelanders, established that this population stemmed from *H. sapiens* ancestors who had hybridized with another *Homo* species, *Homo neanderthalensis*, and that, directly or possibly indirectly, their mongrel ancestors had also some *Homo denisova* genes, and perhaps genes from yet another descent (Skov *et al*., 2020). But the real situation is even worse, and it necessarily makes our origins extremely fuzzy. We have 46 chromosomes. Our ape cousins have 48. We know that our chromosome 2 is a head-to-head fusion of two separate chromosomes in apes (Stankiewicz, 2016). It is rather unlikely that this fusion event happened simultaneously in the matching chromosome pairs of a fertilized egg. It must have been a single rare event in a gamete of a member of an ancestral ape group, producing a single gamete (hence with half the chromosome complement of most cells) with 23 chromosomes instead of 24. Following mating, the parent animal produced an offspring with 47 chromosomes, an unbalanced number, which generated gametes with an uneven number of chromosomes. Assuming that this did not affect its fertility, through mating the mutant ape produced a progeny, half of which with the right number of chromosomes in its normal parent, 48, and half with 47, again. In a social group, this led to a small but significant cluster of individuals with 47 chromosomes, of both sexes, mixing up with 'normal' animals. Perhaps because of some tendency for homogamy—after all, the animals with 47 chromosome may have displayed a recognizable phenotype (Morris, 1999)—a couple with 47 chromosomes happened to mate. Assuming Mendelian standard inheritance, this initiated a progeny that carried chromosomes in the proportions 1/4 (48), 1/2 (47) and 1/4 (46). Now, the individuals carrying 46 chromosomes could mate with any of those members, and, after several generations, a stable colony of members with 46 chromosomes could start developing, progressively adapting to this new genome, presumably across many generations. The origins of these ancestral hominins cannot be ascribed to one particular individual. It is a continuum and this led to the *Homo* genus, which evolved and further split into individual species, that, as we know now, interbred, further expanding the palette of the continuum.

The same is much likely to be true when we look into the origins of the present SARS-CoV-2 population of viruses. It comes from a continuum of viruses with related genomes. Besides bats (Joffrin *et al*., 2020), a variety of animals have been suggested to be carriers of the immediate ancestor of the present virus, including snakes (Zhang *et al*., 2020a), pangolins (Zhang *et al*., 2020b) and even bovines (Luan *et al*., 2020) or dogs (Xia, 2020)! Why is this so important? Obtaining biological insight into the phenotypes associated with each viral genotype is the most interesting corollary of a multiple origin, but it may also answer a socio-political question that repeats itself over the years, namely that which considers that the virus could have escaped, deliberately or accidentally, from a laboratory, despite stringent biosafety/biosecurity precautions.

Rumours and facts

The idea that the virus escaped from a laboratory is not new. In fact, back in 2003, quite a few people in China were afraid that the virus had been constructed and released in their country by some rogue American foe. The rumour was so widespread that it even reached the institute created in Hong Kong as a joint venture between the Institut Pasteur in Paris and the University of Hong Kong, the HKU-Pasteur Research Centre Ltd (http://www.normalesup.org/~adanchin/archives-HKUPRC/Expo2001.html). To obtain evidence relevant to the rumour, those working on the 'in silico' floor of the laboratory were asked to study whether the virus genome displayed scars of recombination patterns. To their great surprise, they found that these families of viruses are likely to have suffered a considerable number of recombination events, notwithstanding convergent evolution (Zhang *et al*., 2005). This made the possibility that they were created by human manipulation extremely unlikely. The exact same question about SARS-CoV-2 has been raised, and the answer is again: the virus is certainly of natural origin (Andersen *et al*., 2020). Importantly, there appear to be observations that suggest that a virus causing severe cases of atypical pneumonia already existed in mid-December 2019 in China (Guo *et al*., 2020), possibly even in Italy, in France, and in the United States in January 2020 not initially ascribed to COVID-19. It could well be that early ancestors of SARS-CoV-2, producing

asymptomatic infections had already spread before causing clinical—COVID-19—disease (Forster *et al*., 2020).

Another confusing phenotype of SARS-CoV-2 viruses is that they propagate the disease according to quite different scenarios: random person to person transmission (Riou and Althaus, 2020), transmission within clusters (Shim *et al*., 2020) and 'superspreader' transmission (Hodcroft, 2020). These different modes may be characterized by differences in the way they channel evolution of each virus variant. In particular, we may anticipate two major scenarios of evolution: competing forms of the disease between patients hosting different virus mutants and co-infection of a single patient with two or more mutants of the virus, allowing it to generate recombinants.

Competing diseases between patients infected by different variant viruses

In Wuhan hospitals, the healthcare personnel observed that the clinical characteristics of COVID-19 patients at the onset of the disease became progressively different, possibly as new coronavirus variants emerged and spread over time and generations. Among a variety of other studies, a preprint published on March 2 by investigators working in Wuhan showed that it was as if there had been competition between different forms of the disease, assumed by the authors to result from infection by different strains of the virus (two major strains) infecting different people in Wuhan (Chen *et al*., 2020). This work established that SARS-CoV-2 was evolving gradually into two major forms, one causing flu-like symptoms and another causing subclinical infections and some asymptomatic carriers. This was likened to a somewhat similar course of the SARS outbreak in 2003 (Ng *et al*., 2003), with apparently two epidemic courses with different transmission routes occurring in Wuhan (Danchin *et al*., 2020). Such a course of the epidemic might possibly reflect a progressive attenuation of the virus, but this interpretation requires investigation because a discrepancy in the way patients were officially recorded could have had the same effect. Alternatively, the evolution of the disease might have been the opposite, first following an asymptomatic course, succeeded by cases with more severe outcomes. The fact is if there is competition between virus variants at the population level, the viruses might create population mosaics of disease characteristics—infection fatality rates, transmission, immune status, etc. The apparent discrepancy between situations in different countries or regions may reflect virus evolution and viral variants, so it is crucial to obtain a fine picture of the virus genome variations across countries, eventually correlated with associated disease phenotypes. It is also important, when possible, to get time series of the virus evolution within individual patients over the course of infections.

Of course, human polymorphism also plays a role in the spread and severity of the disease with HLA involved in antigen presentation (Rubino *et al*., 2020) and carbohydrate protein tagging associated with the level of transmission [(Le Mercier *et al*., 2019), the blood group antigen H variants, type O being perhaps less prone to be infected (Zhao *et al*., 2020)]. And what about the Lewis secretion system, found to be important for MERS-CoV but not yet explored for SARS-CoV (Park *et al*., 2019) and perhaps even the gut microbiome (Gou *et al*., 2020)? In the context of human polymorphism, competition between viruses would be more prevalent in populations sharing a common environment, such as hospitals, dormitories or care homes, resulting in some form of mosaicism of the disease phenotypes associating genetic traits both of the virus and of the hosts. If this type of scenario is responsible for disease mosaicism, it could be essential to understand if we wish to design an effective vaccine.

At a finer level of granularity, as the virus evolves within patients—at a rate that is not insignificant (Li *et al*., 2020)—people in families may be infected by different variants, creating family mosaicism. A local competition could thus be initiated between virus variants exhibiting different phenotypes such as one causing more severe symptoms and another causing milder symptoms but having enhanced transmission. Here, it is essential to remember that, from the point of view of a virus, success is not measured by severity of morbidity/rapidity of mortality, but by the ability to produce and transmit a sizeable progeny for a long time. The sampling and viral sequencing of all members of households with an infected member is thus an important endeavour. This is especially important as, in the long term, we can expect the virus to progressively attenuate as it slowly sheds its cytosine complement, for metabolic reasons discussed elsewhere (Danchin and Marlière, 2020). This hopeful situation would, however, be drastically altered if the organization of social groups, clusters in particular, led to increase its propensity to recombine.

Muller's ratchet: Recombination against attenuation

When the level of infection of a population becomes significant—either through natural transmission or when authorities try to enforce herd immunity, and also in cases happening in situations of sizeable case clusters—some individuals become multiply-infected with different variants of the virus. In this condition, there may be competition between the variants within an infected individual, resulting in mosaicism of pathology within an individual. Most mutations have either no effect or deleterious

effects affecting the fitness of the virus, with very few promoting sustained or increased virus survival and propagation; only the latter will survive. However, these will, in the long term, tend to attenuate the multiplication/virulence of the virus. There is, nevertheless, one family of mutations which may, transiently, during the course of the epidemic, improve the fitness of the virus, namely 'stealth' mutations that result in loss of a character of the virus metabolism recognized by innate immunity. Cases in point are mutations that alter cytosine metabolism, when loss of cytosines in the genome of the virus will make it somewhat more resistant to the antiviral, interferon-induced, protein viperin (Danchin and Marlière, 2020). This evolution which may be beneficial for the virus at some point cannot last over many generations, and attenuation of the symptoms of the disease is likely to happen sooner or later as highly evolved variants replaced the initial population.

Attenuated viruses have traditionally been exploited for vaccination. However, while the generation of attenuated strains that could be deployed as vaccines was observed very early on in the case of Yellow Fever Virus (Theiler and Smith, 1937), this took decades to be replicated for other viruses (Sabin and Boulger, 1973). Determination of variant 'success' over time could be instructive in this respect.

Already in 1935, Muller remarked that spontaneous mutations in a species would progressively lead to decrease its fitness and its evolutionary landscape, ending up in its demise (Muller, 1964). Thus, when viruses evolve as separate entities, without contact with other sources of genes, they are doomed to suffer the constraints of Muller's ratchet (e.g., see Fig. 1). However, it is known that coronaviruses are prone to recombine when a cell is co-infected by two or more viruses of different genome type (Lai, 1995; Graham and Baric, 2010). This has also been demonstrated with bacterial RNA viruses (Chao, 1990). Thus, co-infection of the same cell with different variants of a virus, or even different viruses of the same family (e.g., positive-sense RNA viruses) may lead to recombination, possibly restoring an ancestral phenotype. This is a critical process to counteract progressive loss of fitness as mutations accumulate.

As discussed at the beginning of this article, it is likely that recombination plays a key role in evolution of coronaviruses, i.e., the SARS-CoV-2 virus is a mosaic that has recruited/is recruiting a variety of functions via integration of pieces of genomes from different sources. This feature may prove to be a significant challenge in the search for the origin of the SARS-CoV-2 virus, especially in hosts such as bats, which host many different types of coronaviruses (Joffrin *et al*., 2020). Viruses multiply extremely rapidly. Wild animals are co-infected by many viruses simultaneously, and this creates an enormous evolutionary landscape for viruses to explore. This is

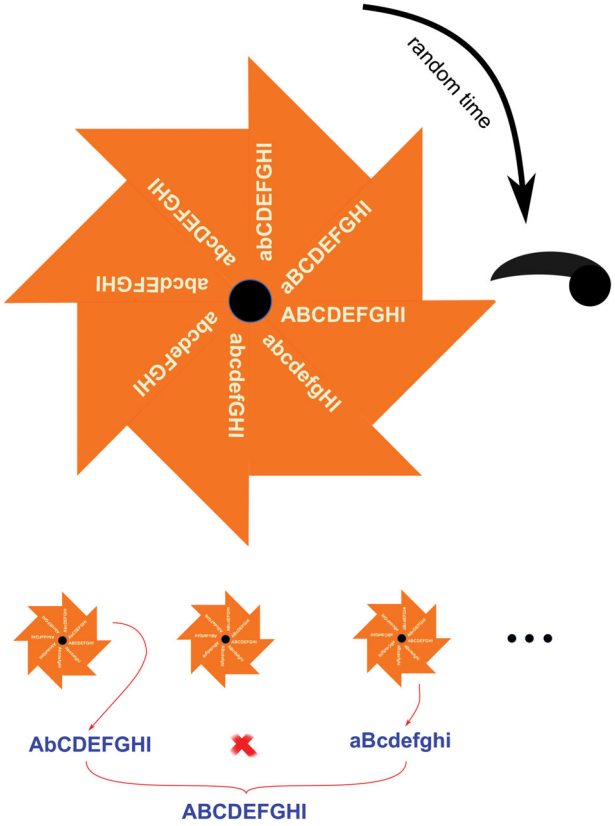

Fig 1 Muller's ratchet and recombination.
Genes A, B, C, D, E, F, G, H, I, in an arbitrary order are mutated in a form shown as the lower case counterpart, at random as time passes by. This behaves in a ratchet-like way because the probability of exact reversion is very low. After some time, all genes have been mutated and the virus has lost much of its initial virulence. The same process occurs independently for viruses of different descents. However, if some of the progeny of different genealogies are present in the same cell, they can recombine, and this allows them to recreate the ancestral form of the virus.

consistent with the idea, discussed in the case of the origin of *H. sapiens*, and especially because we do not have a clear 'index' patient, that there is not one origin of the SARS-CoV-2, but a continuum. And this raises the question: Has early SARS-CoV-2, or a precursor of it, been a member, even transiently and even possibly a long time ago, of the microbiomes of some members of the community, especially workers in wet markets, prior to mutating to greater pathogenicity and causing disease? Have attenuated variants of SARS-CoV-2 now become members of the microbiomes of some people? Only the testing of large numbers of asymptomatic members of society will answer this question.

A perspective

We have tried to summarize here reasons to hope for a positive outcome of the outbreak of COVID-2, namely via

progressive attenuation of the virus (Armengaud *et al*., 2020). There is precedence, in the epidemic that killed scores of pigs back in the 1980s. The infective coronavirus—transmissible gastroenteritis coronavirus, TGEV—has both enteric and respiratory tropism and causes an extremely severe disease. A mutation with essentially lung tropism appeared in some herds (Laude *et al*., 1984; Pensaert *et al*., 1986). After a few years of evolution, this variant of the virulent TGEV, with a deletion in the tropism-determining region of the spike protein that is used to bind specific receptors of host cells and trigger membrane fusion, had lost its virulence while propagating worldwide and essentially acting as a natural vaccine (Schwegmann-Wessels and Herrler, 2006). This makes that TGEV is no longer of major concern for pig breeding, despite the recent emergence of yet another swine acute diarrhoea syndrome (SADS) coronavirus coronavirus (Zhou etal. 2019). Thus, besides this optimistic exit from the COVID-19 epidemic, we should be concerned by the high recombination potential of the virus, which may erase the mutations that allowed it to evolve towards attenuation. Because infection clusters allow for high levels of virus recombination, it is critical to prevent, as far as possible, social activities that lead to creation of clusters of infection (Yong *et al*., 2020). Indeed, at this very moment, it seems clear that much of the difficulty to control the disease comes from the failure to break down clusters.

To investigate some of the scenarios we have sketched above, we consider it will be useful to trace the sequence evolution of the SARS-CoV-2 virus from its earliest stages, and associate genotype with phenotype and disease granularity. For this, it would be very useful to obtain information on virus sequence and viral loads in:

1. Pre-outbreak clinical samples from asymptomatic individuals in Wuhan, where the outbreak occurred. How to obtain these retrospectively? Potential sources might be stored buccal swab samples from patients presenting with pharyngitis/laryngitis, stored clinical samples taken from patients presenting for other conditions, stored samples from post-mortem examinations (here, viral loads in various organs would inform about tropism, etc.), especially from people working in wet markets and their families, if they exist
2. Pre-outbreak clinical samples from asymptomatic individuals in countries where the outbreak is delayed, focusing particularly on family members and friends of individuals showing symptoms
3. Longitudinal, repeat sampling of symptomatic individuals over the course of the infection and, of their household members
4. Sampling of different organs of individuals having succumbed to COVID-19.

References

Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., and Garry, R.F. (2020) The proximal origin of SARS-CoV-2. *Nat Med* **26**: 450–452.

Armengaud, J., Delaunay-Moisan, A., Thuret, J.-Y., van Anken, E., Acosta-Alvear, D., Aragón, T., *et al*. (2020) The importance of naturally attenuated Sars-Cov-2 in the fight against Covid-19. *Environ Microbiol*. in press, https://sfamjournals.onlinelibrary.wiley.com/doi/abs/10.1111/1462-2920.15039.

Chao, L. (1990) Fitness of RNA virus decreased by Muller's ratchet. *Nature* **348**: 454–455.

Chen, Z., Hu, J., Zhang, Z., Jiang, S., Wang, T., Shi, Z., and Zhang, Z. (2020) Caution: clinical characteristics of COVID-19 patients are changing at admission. *SSRN J*, e-publication, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3546044.

Cowling, B.J., and Leung, G.M. (2020) Epidemiological research priorities for public health control of the ongoing global novel coronavirus (2019-nCoV) outbreak. *Euro Surveill* **25** (in press). https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.6.2000110.

Danchin, A., and Marlière, P. (2020) Cytosine drives evolution of SARS-CoV-2. *Environ Microbiol* (in press). https://sfamjournals.onlinelibrary.wiley.com/doi/full/10.1111/1462-2920.15025.

Danchin, A., Ng, T.W.P., and Turinici, G. (2020) A new transmission route for the propagation of the SARS-CoV-2 coronavirus. *medRxiv*. (e-publication), https://www.medrxiv.org/content/10.1101/2020.02.14.20022939v1.

Dyson, F.J. (1985) *Origins of Life*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511546303.

Forster, P., Forster, L., Renfrew, C., and Forster, M. (2020) Phylogenetic network analysis of SARS-CoV-2 genomes. *Proc Natl Acad Sci USA*, **117**(17): 9241–9243. https://www.pnas.org/content/117/17/9241.

Gou, W., Fu, Y., Yue, L., Chen, G., Cai, X., Shuai, M., *et al*. (2020). Gut microbiota may underlie the predisposition of healthy individuals to COVID-19. *medRxiv*: **e-publication**, https://www.medrxiv.org/content/10.1101/2020.04.22.20076091v1.

Graham, R.L., and Baric, R.S. (2010) Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *J Virol* **84**: 3134–3146.

Guo, Y.-R., Cao, Q.-D., Hong, Z.-S., Tan, Y.-Y., Chen, S.-D., Jin, H.-J., *et al*. (2020) The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak—an update on the status. *Mil Med Res* **7**: 11.

Hodcroft, E.B. (2020) Preliminary case report on the SARS-CoV-2 cluster in the UK, France, and Spain. *Swiss Med Wkly* **2020**: w20212. https://smw.ch/article/doi/smw.2020.20212.

Joffrin, L., Goodman, S.M., Wilkinson, D.A., Ramasindrazana, B., Lagadec, E., Gomard, Y., *et al*. (2020) Bat coronavirus phylogeography in the Western Indian Ocean. *Sci Rep* **10**: 6873.

Lai, M.M.C. (1995) Transcription, replication, recombination, and engineering of coronavirus genes. In *Corona- and Related Viruses. Advances in Experimental Medicine and*

Biology, Talbot, P.J., and Levy, G.A. (eds). Boston, MA: Springer, US, pp. 463–471.

Laude, H., Charley, B., and Gelfi, J. (1984) Replication of transmissible gastroenteritis coronavirus (TGEV) in swine alveolar macrophages. *J Gen Virol* **65**: 327–332.

Le Mercier, P., Mariethoz, J., Lascano-Maillard, J., Bonnardel, F., Imberty, A., Ricard-Blum, S., and Lisacek, F. (2019) A bioinformatics view of glycan¯virus interactions. *Viruses* **10**(11): 636. https://doi.org/10.3390/v11040374.

Li, X., Wang, W., Zhao, X., Zai, J., Zhao, Q., Li, Y., and Chaillon, A. (2020) Transmission dynamics and evolutionary history of 2019-nCoV. *J Med Virol* **92**: 501–511.

Luan, J., Jin, X., Lu, Y., and Zhang, L. (2020) SARS-CoV-2 spike protein favors ACE2 from Bovidae and Cricetidae. *J Med Virol* (in press). https://onlinelibrary.wiley.com/doi/10.1002/jmv.25817.

Morris, D. (1999) *The Naked Ape*. New York: Random House.

Muller, H.J. (1964) The relation of recombination to mutational advance. *Mutat Res* **106**: 2–9.

Ng, T.W., Turinici, G., and Danchin, A. (2003) A double epidemic model for the SARS propagation. *BMC Infect Dis* **3**: 19.

Park, Y.-J., Walls, A.C., Wang, Z., Sauer, M.M., Li, W., Tortorici, M.A., *et al*. (2019) Structures of MERS-CoV spike glycoprotein in complex with sialoside attachment receptors. *Nat Struct Mol Biol* **26**: 1151–1157.

Pensaert, M., Callebaut, P., and Vergote, J. (1986) Isolation of a porcine respiratory, non-enteric coronavirus related to transmissible gastroenteritis. *Vet Q* **8**: 257–261.

Riou, J., and Althaus, C.L. (2020) Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Euro Surveill* **25**(4): pii:2000058. https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.4.2000058.

Rubino, S., Kelvin, N., Bermejo-Martin, J.F., and Kelvin, D. (2020) As COVID-19 cases, deaths and fatality rates surge in Italy, underlying causes require investigation. *J Infect Dev Ctries* **14**: 265–267.

Sabin, A.B., and Boulger, L.R. (1973) History of Sabin attenuated poliovirus oral live vaccine strains. *J Biol Stand* **1**: 115–118.

Schwegmann-Wessels, C., and Herrler, G. (2006) Transmissible gastroenteritis virus infection: a vanishing specter. *Dtsch Tierarztl Wochenschr* **113**: 157–159.

Shim, E., Tariq, A., Choi, W., Lee, Y., and Chowell, G. (2020) Transmission potential and severity of COVID-19 in South Korea. *Int J Infect Dis* **93**: 339–344.

Skov, L., Coll Macià, M., Sveinbjörnsson, G., Mafessoni, F., Lucotte, E.A., Einarsdóttir, M.S., *et al*. (2020) The nature of Neanderthal introgression revealed by 27,566 Icelandic genomes. *Nature* (in press). https://www.nature.com/articles/s41586-020-2225-9.

Stankiewicz, P. (2016) One pedigree we all may have come from - did Adam and eve have the chromosome 2 fusion? *Mol Cytogenet* **9**: 72.

Theiler, M., and Smith, H.H. (1937) The use of YELLOW FEVER virus modified by in vitro cultivation for human immunization. *J Exp Med* **65**: 787–800.

Xia, X. (2020) Extreme genomic CpG deficiency in SARS-CoV-2 and evasion of host antiviral defense. *Mol Biol Evol* (in press). https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msaa094/5819559.

Yao, H., Lu, X., Chen, Q., Xu, K., Chen, Y., Cheng, L., et al. (2020) Patient-derived mutations impact pathogenicity of SARS-CoV-2, infectious diseases (except HIV/AIDS).

Yong, S.E.F., Anderson, D.E., Wei, W.E., Pang, J., Chia, W. N., Tan, C.W., *et al*. (2020) Connecting clusters of COVID-19: an epidemiological and serological investigation. *Lancet Infect Dis*. (in press). https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30273-5/fulltext.

Zhang, X.W., Yap, Y.L., and Danchin, A. (2005) Testing the hypothesis of a recombinant origin of the SARS-associated coronavirus. *Arch Virol* **150**: 1–20.

Zhang, C., Zheng, W., Huang, X., Bell, E.W., Zhou, X., and Zhang, Y. (2020a) Protein structure and sequence reanalysis of 2019-nCoV genome refutes snakes as its intermediate host and the unique similarity between its spike protein insertions and HIV-1. *J Proteome Res* **19**: 1351–1360.

Zhang, T., Wu, Q., and Zhang, Z. (2020b) Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Curr Biol* **30**: 1346–1351.e2.

Zhao, J., Yang, Y., Huang, H., Li, D., Gu, D., Lu, X., *et al*. (2020) Relationship between the ABO blood group and the COVID-19 susceptibility. *medRxiv* (e-publication). https://www.medrxiv.org/content/10.1101/2020.03.11.20031096v2.

Zhou, Ling, Li, Qian N, Su, Jiang N, Chen, Gui H, Wu, Zi X, Luo, Ying, Wu, Rui T, Sun, Yuan, Lan, Tian, & Ma, Jing Y (2019). The re-emerging of SADS-CoV infection in pig herds in Southern China. *Transbound Emerg Dis*, **66**(5), 2180–2185. https://onlinelibrary.wiley.com/doi/full/10.1111/tbed.13270.