microbial biotechnology

Lilliput

COVID-19: emergence and mutational diversification of SARS-CoV-2

Harald Brüssow* (iD)

Laboratory of Gene Technology, Department of Biosystems, KU Leuven, Leuven, Belgium.

Summary

The origin of the SARS-CoV-2 virus is not yet defined, but a viral zoonosis from bats – with or without an alternative animal as an intermediate host – is still the most likely hypothesis. The intensive virological and epidemiological research combined with massive sequencing efforts of whole viral genomes allowed an unprecedented analysis of an unfolding pandemic at the level of viral evolution with the documentation of extinction events, prevalence increases and rise to dominance for different viral lineages that provide not only fundamental insights into mechanisms of viral evolution, but influence also public health measures to contain the virus.

Introduction

Processes that shape the evolution of viruses comprise mutation, natural selection, genetic drift, recombination, reassortment (for viruses with segmented genomes) and migration (Geoghegan and Holmes, 2021). Mutation is the primary generator of diversity in viral genomes. Already at this level, there are large differences between different viruses. RNA viruses show a higher mutation rate than DNA viruses, reflecting the higher replication fidelity of DNA-dependent DNA polymerases over that of RNA-dependent RNA polymerases. Within both RNA and DNA viruses, viruses with single-stranded genomes have higher mutation rates than viruses with double-stranded genomes, possibly reflecting the availability of a complementary strand in the latter. Within single-stranded RNA viruses, positive-strand RNA viruses (where the viral genome encodes the protein information) have higher mutation rates than negative-strand RNA viruses (where the genome sequence is complementary to the protein-encoding sequence). Positive-strand RNA viruses display a mutation rate up to 10^{-4} substitutions per nucleotide per cell infection (s/n/c), a rate only surpassed by retrotranscribing viruses and viroids. The mutation rate in the viral world is not only a function of the Baltimore classification of viruses that groups them according to the chemical characteristics of the viral genome, but it is also inversely proportional to the genome size. This relationship is easily rationalized because viruses with high mutation rates, such as RNA viruses, would rapidly accumulate so many deleterious mutations in larger genomes that viral fitness would then be rapidly compromised ('error catastrophe'). Therefore, RNA viruses and single-strand DNA viruses tend to have small genomes. Coronaviruses, which are positive-sense single-stranded RNA viruses, with a 30-kb RNA genome are exceptionally large for an RNA virus; this is only possible because they encode a proofreading 3′-to-5′ exoribonuclease (ExoN, nsp14) that corrects errors made by the viral RNA-dependent RNA polymerases (RdRp) during replication. Without that enzyme, coronaviruses would not be able to protect their long genome against lethal mutagenesis (Smith *et al.,* 2013). Their mutation rate is, with 4×10^{-6} s/n/c, 100-fold lower than for RNA viruses without proof-reading enzymes (Sanjuán *et al.,* 2010), but still a 100-fold higher than found in double-stranded DNA viruses.

The emergence of SARS-CoV-2

When Chinese researchers sequenced the first SARS-CoV-2 genomes recovered from eight different patients living in Wuhan, the genomes showed a sequence identity of 99.98%, i.e. only four nucleotides differed over 30 000 nucleotides (Lu *et al.,* 2020). It was immediately clear that such a level of viral genomic identity could not come from an RNA virus that has been circulating for some time in the human population. It was concluded that SARS-CoV-2 must be the result of a recent spillover of an animal coronavirus into humans. Epidemiological

data quickly pointed to the wet food market in Wuhan as an infection source for the starting epidemic. This hypothesis is not farfetched since live animal markets have been the source for previous outbreaks of viral zoonosis in China (live chicken markets for epidemic influenza virus outbreaks and bats with palm civets as intermediate host for SARS-CoV, a related coronavirus). Unfortunately, animal samples were not taken from the wet market in Wuhan at that time, so it remains unknown whether the zoonosis took place there or whether it was only the place of a first super-spreading event. Four nucleotide changes could mean that SARS-CoV-2 possibly derives from a single viral spillover that occurred about 2 months before the first cases were described associated with the wet food market in Wuhan. A few indirect data (a child testing positive from Huangshi outside of Wuhan early during the outbreak; blood samples with SARS-CoV-2-specific antibodies in the months before the Wuhan outbreak) indicated that some 'cryptic transmission' might have occurred shortly before the outbreak was noted. Until now, the zoonotic origin for SARS-CoV-2 has not been identified. The search for the origin of SARS-CoV-2 has been overshadowed by political quarrelling between China and the USA, creating unhelpful and unproven hypotheses of the virus being accidentally introduced from a research institute working with bat coronaviruses in Wuhan (judged as highly unlikely by a WHO commission) or by deliberate genetic manipulation, fuelling wild conspiracy theories reminiscent of the well poisoning stories during the medieval pestilence waves (Winkler, 2007).

Bat viruses

Bats belong to the usual suspects for zoonosis, and indeed, a bat virus that shared 96% sequence identity with SARS-CoV-2 was isolated in Yunnan /China in 2013. However, a 4% sequence difference (>1000 bp) would indicate 20 to 50 years of separation from SARS-CoV-2, making this bat isolate an unlikely direct source for the nascent epidemic. Chinese researchers explored tissue and faecal samples from 227 bats representing 20 species living in China, collected between May and October 2019 and analysed them by metagenome sequencing. This investigation found that the closest relative of SARS-CoV-2 in this sample set shared 93.3% sequence identity over the entire genome, less than the bat coronavirus isolated in 2013 from the same province, Yunnan (Zhou *et al.,* 2020). The bat virus isolates differed over the spike gene, making binding to the human ACE-2 virus receptor unlikely. However, they showed a three amino acid insertion where all SARS-CoV-2 isolates showed a polybasic amino acid insertion,

corresponding to a furin cleavage site. The insertion differs between the human and bat virus isolates, demonstrating that amino acid insertions at this position occur naturally and support the dismissal of hypotheses of deliberate laboratory manipulation of the SARS-CoV-2 genome.

Intermediate host

Virologists then postulated an intermediate host in which a bat virus adapted for the spillover into the human population by modifying the spike protein allowing recognition of the human ACE-2 viral receptor. Various animals were discussed as intermediate hosts. A coronavirus was isolated from the lungs of two diseased pangolins, which shared 91% nucleotide sequence identity with SARS-CoV-2 (Zhang *et al.,* 2020a,b). Subsequently, Malayan pangolins were intercepted from smugglers, which yielded coronaviruses that shared up to 92% sequence identity with the Wuhan SARS-CoV-2 isolate. These pangolin viruses were more closely related to the human isolates over the receptor-binding domain (RBD) than bat coronaviruses and recombination events between bat and pangolin viruses have been discussed for the origin of SARS-CoV-2 (Lam *et al.,* 2020; Xiao *et al.,* 2020).

The search to identify an animal source for a spillover into humans is now difficult. Chinese scientists have inoculated a number of animals with both an environmental virus isolate from the wet food market in Wuhan and an isolate from an early patient from Wuhan. Ferrets could be infected and subsequently developed an upper respiratory tract infection with fever. Outbred domestic cats could also be infected and developed specific antibodies (Bosco-Lauth *et al.,* 2020). Cats could transmit infection via the airborne (droplet) route to other cats. Hamsters were also later added to the list of susceptible animals (Chan *et al.,* 2020; Sia *et al.,* 2020). Dogs showed reduced susceptibility, but seroconversion, while pigs, chicken and ducks (usual suspects for influenza virus reassortments) could not be infected with SARS-CoV-2. SARS-CoV-2 has thus a rather broad host range, which should not be surprising since coronaviruses are widespread veterinary viral pathogens (Shi *et al.,* 2020).

A number of wild and farmed animals were sold at the Wuhan market, namely foxes, raccoons and sika deer. However, isolating a close relative of SARS-CoV-2 from an animal is problematic since SARS-CoV-2 infection is now so widely distributed among humans that it will be difficult to exclude human-to-animal cross-infection as an origin for such animal isolates (Patterson *et al.,* 2020, Hosnie *et al.,* 2020). The case is well illustrated for minks.

In two fur animal farms in the Netherlands, mink have shown infections with viruses closely related to SARS-CoV-2. However, COVID-19-like symptoms were present in individuals working on the farms before pathological signs were seen in mink. This temporal sequence suggests human-to-mink viral transmissions. Subsequently, infections of minks with SARS-CoV-2-like viruses were notified in various countries. Denmark reported not only infected minks at more than 200 mink farms (Denmark is the world's greatest mink fur producer), but also 214 human COVID-19 cases infected with SARS-CoV-2 virus variants initially identified in virus isolates from mink. After finding a mutant virus with a Y453F substitution in the viral spike protein in both mink and humans from Denmark, the government ordered the culling of 12 million minks. With possible two-way cross-infection occurring between animals and humans, it will be difficult if not impossible to identify the animal host that originally introduced SARS-CoV-2 into the human population. The chances for obtaining valuable hints about the origin of SARS-CoV-2 are now probably greater with archived frozen samples of animals collected by surveys of wildlife for viruses with pandemic potential over decades, such as the US-funded PREDICTS project. Researchers are now looking for evidence of SARS-CoV-2 in samples from bats stored in laboratory freezers from many East and South-East Asian countries (Koopmans, 2020; Oude Munnink *et al.*, 2020; ECDC, 2020a,b).

The conversion of natural habitats to agricultural and urban ecosystems is a globally important mediator of infection risk and disease emergence in humans. These changes increase contact between humans and wildlife, which influence transmission dynamics and pathogen spillover risk to humans. Animals that are phylogenetically more closely related to humans, such as mammals, are more likely to contribute new infections to humans. British ecologists discovered that effective zoonosis sources are animal species persisting in ecosystems disturbed by humans, which could explain why bats and rodents are frequently linked to emerging infections. Expansion of agricultural and urban land will create growing hazardous interfaces for zoonotic pathogen exposure in the future. Zoonosis is not only determined by the richness of the virus pool in given animal groups (high in bats, rodents, primates, ungulates), but also by the opportunity for contact, which is very high between humans and bats, both present with large populations living in crowded societies. Many species of bats exist (they represent a fifth of all mammalian species), and they are the only flying mammals allowing – like air traffic in human societies – a quick geographical dispersal of viruses (Gibb *et al.*, 2020; Mollentze and Streicker, 2020; Olival *et al.*, 2020; Streicker and Gilbert, 2020).

The early evolution of SARS-CoV-2

Since the mutation rate of coronaviruses, which has been estimated by geneticists as 2 substitutions per genome per month for the early phase of the epidemic, is slow compared with its transmission rate (about 5 days), meaning that many identical genomes were initially spreading. Consequently, the entire global population of SARS-CoV-2 viruses sampled through March 2020 differed by a maximum of 12 nucleotide substitutions compared with the inferred Chinese ancestor virus of the whole pandemic (Worobey *et al.*, 2020). A few single-nucleotide polymorphisms (SNPs) allowed the grouping of the first 160 complete SARS-CoV-2 genome sequences into three clusters: A (closest to a bat coronavirus used as an outgroup), B (widely distributed in East Asia distinguished by 2 SNPs from A) and C (absent in China and differing from B by one further SNP) (Forster *et al.*, 2020).

Chinese virologists led the first systematic sequencing efforts by analysing 112 SARS-CoV-2 genomes from COVID-19 patients diagnosed between January 20 and February 25 (Zhang *et al.*, 2020a,b). Compared with the first-released genome (Wuhan-Hu-1), they identified 66 synonymous (nucleotide substitutions that do not change the encoded amino acid (aa) due to codon redundancy) and 103 nonsynonymous (nucleotide substitutions changing the encoded aa) variants in nine protein-coding regions. Substitution rates were similar over most genome regions and calculated as 3.5×10^{-4} substitutions per site per year. Only the nucleocapsid gene showed a 30-fold higher substitution rate. When a phylogenetic tree was constructed with these viral sequences, plus 221 sequences from the GISAID database, two major clades were identified. Clade I included several subclades, such as subclade V characterized by a substitution in the non-structural gene ORF3a (G251V) and subclade G (defined by the spike protein substitution D614G). Clade II is distinguished from clade I by two linked variations – a non-synonymous mutation in ORF8 and a synonymous mutation in ORF1ab. These two major haplotypes represented two lineages derived from a common ancestor that evolved independently in early December 2019 in Wuhan. According to these data, the zoonotic spillover event must have occurred between June and late November 2019 in Wuhan, but possibly outside of the food market. No statistical difference in disease severity, lymphocyte count, CD3 T-cell count, C-reactive protein level or D-dimer level, or in the duration of virus shedding was observed between Chinese patients infected with these two different viral clades.

Sequencing efforts by virologists from Singapore added another interesting early mutant: a 382-nucleotide deletion (Δ382) in the open reading frame 8 (ORF8),

which eliminates its transcription (Young *et al.*, 2020). ORF8 targets host proteins in the endoplasmic reticulum; it is apparently important for viral adaptation to humans and ORF8 is strongly immunogenic. Antibodies are produced against ORF8 early during SARS-CoV-2 infection. This deletion mutant virus emerged in Wuhan early in the pandemic and was exported to Singapore and Taiwan. In Singapore, it was transmitted as a co-infection with the wild-type virus, but infections consisting only of the deletion mutant were then also observed. Clinical outcomes in patients only infected with the deletion mutant were considerably better: fewer patients required supplemental oxygen and they showed an upregulation of T-cell activation associated with cytokines (IFN-γ, TNF-α, IL-2, and IL-5) that might explain the better clinical course in them compared with people infected with the wild type (Young *et al.*, 2020). Independent deletions of various lengths have been reported in ORF8 from patients in Bangladesh, Australia and Spain.

In the early phase of the pandemic, viral genome sequencing was mainly used to trace the geographic spread of the infection. For example, US scientists analysed 453 viral genomes collected between 20 February and 15 March 2020 from infected patients in Washington state, USA: 84% of these viruses fall into a closely related clade (called WA-2). WA-2 is characterized by 5 nucleotide polymorphisms, suggesting a single introduction event from China followed by local amplification and modest diversification (Bedford *et al.*, 2020). A further 9% of viral isolates fell into a separate, smaller clade and derive from viruses circulating in Europe. Epidemiologists sequenced the viral genomes of the first 9 COVID-19 cases at the US East coast observed in mid-March. The viral sequences from seven patients clustered with the US clade known from Washington state, documenting a rapid west-to-east national spread of the epidemic in the United States (Fauver *et al.*, 2020). The viral genomes differed from the ancestor virus in Wuhan by less than 10 mutations. In the early phase of the New York City epidemic, viral genomes from 84 patients were sequenced: 87% of the sequences clustered with the A2a clade, therefore suggesting introduction by travel from Europe. Two mutations were specific for New York City isolates and demonstrated local spread and diversification of the virus. A smaller number of clade B viral strains were of domestic origin and could be traced to Washington state (Gonzalez-Reiche *et al.*, 2020).

The first sustained spread of SARS-CoV-2 in Europe occurred in Italy seeded directly from China or less likely from China via Germany (Nadeau *et al.*, 2021). By 15 March, 189 SARS-CoV-2 viruses from the Netherlands were also sequenced. Multiple co-circulating sequence types were identified, strikingly also observed in cases with similar travel histories, suggesting that viral sequence diversity was already present in the country of origin, notably in Dutch tourists returning from Italy (Oude Munnink *et al.*, 2020). The average difference between the viruses was 7 nucleotides. An international consortium investigated the link between local outbreaks in Austria and the global pandemic by sequencing 345 SARS-CoV-2 genomes from Austrian cases (Popa *et al.*, 2020). The genome mutation profiles confirmed contact tracing data. The largest phylogenetic cluster included resident and travel-associated cases linked to the Austrian ski resort Ischgl. Cluster Tyrol-1 coincided with a local outbreak in France and was then identified in Iceland by tourists with a travel history to Austria. One week after the occurrence of viral strains with the mutation profile typical for Ischgl, an increasing number of related strains were found in New York City. Ischgl may have played a critical role as international transmission hub, representing a community-level superspreading event.

The first infection wave in UK

With the unfolding of the pandemic, virus genome sequencing increased worldwide with > 100 000 SARS-CoV-2 genomes by 1st of October 2020; half of them were obtained from the UK, which allowed a detailed molecular reconstitution of the epidemic in this country. During the first infection wave in UK (until end of June), most transmission lineages were small and short-lived (du Plessis *et al.*, 2021). Eight larger lineages comprising > 25% of the sequenced genomes were observed for longer time periods. Some of the major lineages showed a nationwide distribution, while others remained geographically restricted. The UK's first epidemic wave resulted from the concurrent growth of many hundreds of independently introduced transmission lineages. Their mutational pattern showed that 33% of UK transmission lineages stemmed from arrivals from Spain, 29% from France, 12% from Italy. Few initial restrictions on international arrivals led to the establishment and co-circulation of >1000 identifiable UK transmission lineages. The delay between importation and onward within-UK transmission was shown to be 8 days. The diversity of transmission lineages correlated with commuter travel to and within London. Lineage diversity peaked in late March and declined after the UK national lockdown, leading to extinction of lineages in a size-dependent manner. A subset of lineages then took hold across the UK.

A paradigm of SARS-CoV-2 evolution – the D614G spike protein mutation

Next-generation sequencing permitted nearly real-time detection of genetic variants that appeared in the viral population during the unfolding of the pandemic. RNA

viruses accumulate genomic mutations as they transmit. Evolutionary theory predicts that most new viral mutations are deleterious and short-lived, whereas mutations that persist and grow in observed frequency may be selectively neutral or advantageous to viral fitness. Discriminating between neutrality and positive selection is challenging. That a new mutation is increasing in prevalence or geographic range is insufficient to prove its selective advantage since it can be generated by neutral processes, such as genetic bottlenecks following founder events and range expansions. It needs a detailed investigation to analyse the causes for frequency increases of variant viruses (Geoghegan and Holmes, 2021).

The case is well illustrated by an early viral spike mutation. When US researchers compared 18 000 genome sequences in May 2020 from viral isolates sampled from all continents, they found limited diversity across SARS-CoV-2 genomes: only 11 sites showed polymorphisms in > 5% of sequences; yet two mutations, including the D614G mutation in the spike, had already become the consensus. They detected evidence of purifying selection, but little evidence of diversifying selection, with substitution rates comparable across structural versus non-structural genes (Dearlove *et al.*, 2020).

In Europe, the 614G variant was first observed in genomes sampled on 28 January 2020 in a small outbreak in Bavaria, Germany, which was initiated by a visitor from Shanghai (Böhmer *et al.*, 2020) and subsequently controlled through public health efforts. Instead of representing a founder effect, it is more likely that the D614G mutation evolved in China and was introduced on multiple occasions to European countries (Volz *et al.*, 2020). In early March, that variant was still rare globally but was gaining prominence in Europe. The transition from D614 to G614 occurred asynchronously in different regions throughout the world, beginning in Europe, followed by North America and Oceania, then Asia. The G614 variant increased in frequency also in regions where D614 was initially the clearly dominant form. The D614G change is accompanied by three other mutations, including an amino acid change in RNA-dependent RNA polymerase. This haplotype comprising the four genetically linked mutations was the globally dominant form after the first infection wave: Prior to March 1, it was found in 10% of 997 global sequences; after March 1, it represented 70% of 26 000 sequences (Korber *et al.*, 2020). By now, D614G has gone to near fixation worldwide, raising the question about its role in infection.

US researchers determined that D614G caused striking conformational changes in the spike protein by weakening of the interprotomer contacts, thus causing a dramatic change in the ratio of open to closed S protein trimers where a segment of the spike protein stretches out finger-like into the open conformation for increased interaction with the cell receptor protein ACE-2 (Yurkovetskiy *et al.*, 2020). G614 spike trimers populated a two-open conformation (39%) and an all-open state (20%), which was not detected for D614. In its closed conformation, the receptor-binding domain (RBD) is physically blocked for interaction with ACE-2. G614 spikes have thus an easier access to the cell receptor than D614 spikes. Changes at the spike position 614 might also confer a greater transmission capacity of the virus for another reason. D614 is a surface residue in the vicinity of the furin cleavage site, activating the viral spike protein for cell fusion (Gobeil *et al.*, 2020). The fusion site is located away from the tip of the spike containing the receptor-binding domain (RBD). Position 614 is located closer to the viral membrane and is mechanistically needed for the fusion process between the viral and cellular membrane during virus entry into the cell. Proteolysis experiments with the mutant spike protein revealed enhanced furin cleavage efficiency of the G614 variant over D614 spikes. However, the authors speculated that there might be a fitness trade-off with the D614G spike mutation, where more open spike protein structures make the virus more accessible to antibodies *in vivo* and thus less prone to evade immune responses. Indeed, when sera from spike-immunized mice, nonhuman primates and humans naturally infected with either 614 variants of spike protein were evaluated for neutralization of pseudoviruses bearing either D614 or G614 spike, the G614 pseudovirus was more susceptible to neutralization than D614 pseudovirus (Weissman *et al.*, 2021). The G614 pseudovirus was also more susceptible to neutralization by receptor-binding domain monoclonal antibodies than D614 pseudovirus. Negative stain electron microscopy showed in these studies a RBD "up" conformation of 82% for the G614 spike protein compared with only 46% for the D614 spike, suggesting increased epitope exposure as a mechanism for the enhanced neutralization.

Texan virologists used the infectious cDNA clone for SARS-CoV-2 and generated the D614 and G614 substitutions in the USA-WA1/2020 strain that had started the COVID-19 epidemic in the USA. The D614G mutation did not affect viral replication or virion infectivity in Vero E6 kidney cells, but improved SARS-CoV-2 replication on human lung epithelial cells Calu-3 through increased virion infectivity. Viral titres from nasal washes, trachea and various lobes of the lung were consistently higher in the G614- than D614-infected hamsters. When hamsters were infected intranasally with equal amounts of the two viruses, G614 virus had a consistent advantage over D614 virus. In a primary human airway tissue model displaying human tracheal/bronchial epithelial cells as multilayers at an air–liquid interface showing micro-ciliary responses, infectious viral titres of G614 were two- to

ninefold higher than those of D614. In this system, the G614 virus could rapidly outcompete the D614 virus, even if the latter was given at higher initial dose. The G614 virus retained higher infectivity over time and across different temperatures compared to D614, thus indicating a higher stability. Sera from D614 virus-infected hamsters consistently exhibit higher neutralization titres against G614 than against D614 virus (Plante *et al.,* 2020).

US scientists generated isogenic variants differing at the 614 aa position, as well as containing in addition a nanoLuciferase (nLuc) gene in place of an accessory coronavirus gene, which allowed an easy quantitative and histopathological comparison of both variants (Hou *et al.,* 2020). The G614 variant showed an up to eightfold higher transgene expression in different cell lines. In *ex vivo* primary human nasal and airway epithelia, both viruses infected mainly ciliated cells with higher titres for the variant virus. Indeed, the variant became dominant after 3 passages in competition experiments. In mice and hamsters, both viral types induced comparable pathological lesions and similar viral titres, but hamsters infected with the variant virus showed a greater weight loss than those infected with the wild-type virus. In competition experiments using hamsters infected with 1000 viruses containing a 1:1 ratio of both types, the variant virus became dominant after the first passage. When the researchers placed naïve hamster adjacent to a cage with an infected animal, both viruses were transmitted efficiently to naïve hamsters, resulting in comparable viral titres as tested after 4 days. However, five of eight hamsters exposed to the D614G infected group showed infection and detectable viral shedding already at day 2, whereas those exposed to the WT-infected group showed no viral shedding, indicating that the variant transmits significantly faster between hamsters through aerosol and droplets. The D614G virus was more sensitive to SARS-CoV-2 neutralizing antibodies than the wild-type virus.

The increased transmission of the variant virus between hamsters might reflect increased or accelerated replication in the nasal epithelium, a lower minimum infectious dose or subtle variations in virion stability in small and large droplets. Only further biological experiments can differentiate why transmission was more efficient with the variant virus.

In clinics, patients with the G614 virus developed higher levels of viral RNA in nasopharyngeal swabs than those infected with the D614 virus, but they did not develop more severe disease (Korber *et al.,* 2020). British scientists applied classic population genetic models to investigate the impact of the mutation by analysing the COVID-19 Genomics UK data set containing > 40 000 SARS-CoV-2 viral genome sequences with epidemiological and medical metadata (Volz *et al.,* 2020). In the UK, the 614G clusters appeared 16 days later than the 614D clusters, but 614G became the dominant form in late March and this trend has continued. The models did not prove a higher transmission rate for the 614G variant; the observed changes over time could still represent stochastic processes. Next, they investigated associations between the D614G polymorphism and virulence by linking virus genome sequence data with clinical data on patient outcomes. Patients with the G614 variant show reduced odds of death, but this effect disappeared after controlling for other known risk factors for severe COVID-19 outcomes. The researchers did not find any association with clinical severity indicated by the requirement for oxygenation or respiratory support and aa position 614 polymorphism, but an association with younger age. The data set showed a very slight, but significant difference for G614 being associated with lower C_t values in PCR tests, i.e. higher viral load in the nasopharynx. The researchers also found reversions back to D614, indicating that the D614 version is still relatively fit within individual hosts. Further mutations in the vicinity of the 614 position were found, e.g. a variant 615I, which was largely constrained to Wales. The clinical D614G mutant data do not indicate an altered trend towards attenuation or increased virulence. Despite the availability of a large dataset for the UK well represented by both spike 614 variants, not all analytical approaches have shown a conclusive signal of positive selection. The D614G mutation is associated with the B.1 lineage of SARS-CoV-2, which now dominates the global pandemic.

Molecular epidemiology of the second wave in the UK: the B.1.7.7 variant VOC

British scientists extended their genomic analysis though the second infection wave in the UK until end of December, 2020 (Volz *et al.,* 2021). They noted that a variant called B.1.1.7 or *virus of concern* (VOC) is rapidly expanding its geographic range and frequency in England with a focus on London and the South East. In parallel, the UK has faced a rapid increase in COVID-19 cases most pronounced in South East England, with a fourfold case increase in the last weeks of 2020. Currently, 60% of recent COVID-19 infections in London are associated with the VOC variant. The variant displayed a large number of non-synonymous substitutions of potential immunological significance. Overall, the variant displays 29 mutations with respect to the original Wuhan SARS-CoV-2 isolate. Points of concern (ECDC, 2020a, b) are that it carries multiple spike protein mutations (deletion 69-70, deletion 144, and the amino acid replacements N501Y, A570D, D614G, P681H, T716I,

S982A, D1118H), which affect the furin S1/S2 cleavage site (activating the virus for cell entry, mutation P681H) and the RBD (N501Y) that could thus change the viral phenotype. It also contains mutations in another viral gene, introducing a stop codon in ORF8. Within the B.1.1.7 lineage, this VOC has acquired 17 mutations all at once, which has not been observed before. Since non-synonymous mutations dominate over synonymous mutations, VOC might represent the result of selection. However, selection did not occur stepwise in the population since VOC is not linked by transitions to other isolates on the phylogenetic tree. Current hypotheses for its origin are selection in animals infected by human SARS-CoV-2 and crossing back into humans (as discussed for the mink transmission) since the N501Y RBD mutation occurred spontaneously in ferrets upon experimental infection with a wild-type human isolate. However, many of the individual VOC mutations were already observed in COVID-19 patients – including the N501Y RBD mutation recently identified in a variant from South Africa, which there, however, is associated with a distinct mutation pattern. At the moment, no data support an animal origin of VOC, more likely is their generation in an immunosuppressed patient (see below). In animal models, the N501Y replacement in the spike protein has been suggested to increase ACE-2 receptor binding and cell infectivity. The variant also possessed a deletion in the spike protein (Δ69-70), which provided a useful biomarker because it was associated with diagnostic test failure for a probe targeting the spike gene (FDA, 2020). Notably, in regions where the second lockdown led to a decrease in case numbers, the previously circulating viruses dominated. In contrast, in regions where case numbers increased directly after the end of the lockdown, the deletion mutant indicative of the B.1.7.7 VOC variant dominated and remained stable during lockdown. From this observation, the scientists calculated an approximate 50% transmission advantage for B.1.1.7 over prior isolates and observed an association between B.1.1.7 prevalence and the time-varying reproduction number, R_t, as well as a shift to younger patients. However, this association analysis cannot prove causality; the mechanism underlying increased transmission is not yet clear. Previous studies have hinted that the N501Y change allows the virus to attach to cells more strongly, making infection easier. Several mutations to the spike protein have enabled viruses to reach greater concentrations in animals, such as hamsters' upper airways, which could underlie greater transmission.

Through mathematical modelling, four alternative hypotheses were considered why the new variant might be spreading more efficiently, ranging from increased infectiousness; immune escape; increased susceptibility among children; and shorter viral generation time. The model with a 50% increase of transmission of B.1.7.7 gave the best fit and concurs with observations of lower C_t values (i.e. higher viral load) for B.1.7.7 (Davies *et al.,* 2020a,b). In addition, epidemiological data indicate that 15% of the contacts of people infected with B.1.1.7 in England became positive themselves, compared with 10% of contacts of those infected with other variants. The British data do not suggest evidence for a difference in the odds of hospitalization or relative risk of death.

British statisticians when reanalysing the data observed even that the 60-day survival assessed by Kaplan–Meier curves was higher in the subjects infected with the variant virus, but they showed that this observation is confounded by the younger average age of the cases infected with the variant virus. In fact, 80% of the cases infected with the variant virus were younger than 54 years old and thus at a lower mortality risk. When the subjects were stratified by age, subjects older than 70 years infected with the variant virus showed a lower survival rate after diagnosis than subjects infected with the previously circulating viruses. For males aged 70–84 years, the risk of death within 28 days increased from 4.7% to 6.1% (Davies *et al.*, 2020a,b).

Variant virus from South Africa: variant 501Y.V2

South Africa is the country in Africa that has been the most severely affected by COVID-19, with over 56 000 excess deaths. The second SARS-CoV-2 epidemic wave in South Africa began around October, just weeks after a low in infection numbers. When analysing 2500 SARS-CoV-2 whole genomes, a new monophyletic cluster emerged in September and rapidly became the dominant lineage, superseding the three previously main South African lineages (B.1.1.54, B.1.1.56 and C.1) that were circulating during the prior wave (Tegally *et al.,* 2021). The South Africa variant 501Y.V2 displayed nine changes in the spike protein divided into two subsets: one cluster in the N-terminal domain (NTD) that included four substitutions and a deletion, and another cluster of three substitutions in RBD (K417N, E484K and N501Y). The substitutions D614G and N501Y were shared with the VOC isolate from UK, while the other mutations were distinct from the VOC constellation. The scientists speculated that the high rate of immunosuppression by HIV-1-infected subjects in South Africa might lead to longer viral persistence and could thus be responsible for the generation of complex mutants such as variant 501Y.V2. However, no experimental or clinical data have so far indicated a distinct COVID-19 infection course in HIV-1 patients. Researchers tested the sera from 44 COVID-19 patients, half of them with neutralization titres > 400 (mostly from severe cases) against the previous variant D614G as reference virus, the South African variant

501Y.V2, and a pseudotype virus displaying only the 3 RBD substitutions from 501Y.V2. The majority of sera with low neutralizing antibody titres against the D614G reference virus failed to neutralize the South African variant, while this was only the case in 23% of the high-titred sera, but nearly all sera showed titre decreases against the variant. Failure was greater against the 501Y.V2 variant than against the RBD mutant pseudovirus, thus suggesting that polyclonal sera from COVID-19 patients also direct neutralizing antibodies against NTB epitopes (Wibmer *et al.,* 2020; Cele *et al.,* 2021).

Variant virus P.1 from Brazil

SARS-CoV-2 was introduced into Brazil at beginning of March 2020. Sequencing of a geographically representative 430 viral genome data set revealed more than 100 international virus introductions into Brazil, first predominantly from Europe, representing three viral clades circulating in Europe, and later also from the US (Candido *et al.,* 2020). Viruses circulated first locally within federal state boundaries, then also country-wide by multiple exportations from large urban centres. The viruses mutated slowly during their spread by accumulating an estimated 33 mutations per year over the genome. In Manaus, a 2-million inhabitant city in the Amazonian region and a hotspot of the Brazilian COVID-19 epidemic during late spring 2020, Brazilian virologists sequenced 31 viral genomes isolated in December 2020; 65% of the sequences belonged to the B.1.1.28 lineage. They saw several distinct branches of the B.1.1.28 global phylogeny tree, suggesting multiple introductions of this lineage into Manaus. Notably, 13 out of 31 of the mid/late-December genomes defined a new P.1 lineage which was related to, but distinct from, the B.1.1.28 lineage. The P.1 variant was not detected during the first wave in Manaus. P.1 shares the spike N501Y mutation and a deletion in ORF1b with the UK B.1.1.7 variant VOC. It shares three spike protein mutations (K417N/T, E484K, N501Y) and the ORF1b deletion with the South Africa variant 501Y.V2. The E484K mutation was observed in many previously sequenced Brazilian viral genomes, while the N501Y had not previously been detected in Brazil. Their distinct mutation patterns suggests that the UK, South Africa and Brazil variants have arisen independently. A common characteristic of the three variants is that they are associated with a rapid increase in cases in locations where previous attack rates have been very high. In Manaus, the P.1 lineage was identified in 42% of RT-PCR-positive samples collected between December 15 to 23, while it was absent in Manaus between March and November 2020 (CDC, 2021; Faria *et al.,* 2021).

Immune escape?

So far, 2737 mutations have been identified in the S gene of SARS-CoV-2 isolated from humans, giving rise to 1133 aa changes, including 171 substitutions in the RBD (Liu *et al.,* 2020). Mutations will occur inevitably and with constant rate. As these are stochastic processes, they will be more frequent in areas where many people are infected with viruses (hotspots) because here the virus is replicating much more frequently than in non-hotspot areas. One should therefore expect mutants in hotspot regions (UK, South Africa and Brazil) for COVID-19. There is also an observation bias since viral genome sequencing is done with markedly different speed in different countries.

When an increasing part of a population gets infected or vaccinated, the number of people with neutralizing antibodies against the circulating viruses will increase. This will become a strong selection pressure for immune escape mutants and could be the reason for the emergence and prevalence increase of the E484 spike variants. When researchers grew SARS-CoV-2 in the presence of low levels of the convalescent serum of a recovered COVID-19 patient, the virus had picked up three mutations that made it resistant to the patient's serum within 90 days (Andreano *et al.,* 2021). One was the E484K spike protein mutation; the other two were N-terminal domain changes in the spike protein, changes also found in the South African variant virus. The reinfection of two Brazilian patients with E484 spike variants in case studies would also fit this scenario. However, it is still controversially discussed whether places like Manaus have really reached herd immunity levels for selecting escape mutants (Hallal *et al.,* 2020; Buss *et al.,* 2021).

Intra-host variation and bottlenecks

So far, we have mostly considered the evolution of SARS-CoV-2 at the community level. Although the data are more limited, there is also evidence that SARS-CoV-2 can evolve within a single patient. When eight COVID-19 patients were investigated with deep RNA sequencing of lung washes, a median of four viral variants, characterized by single-nucleotide changes, were observed per patient. There was a single outlier patient with 51 variants in a single sample. The rate of non-synonymous over synonymous mutations was significantly smaller than 1, suggesting purifying selection. The number of variants did not increase with days after disease symptom onset. Among all observed 84 intra-host variants, only three were found to be polymorphic in the population data set from the same region, therefore suggesting strong bottleneck effects for the transmission of variant

viruses from an infected individual (Shen *et al.*, 2020). In a study from Austria, most patient samples showed a small number of stable low-frequency mutations, but a minority of patients exhibited higher variability, including fixation and loss of individual mutations. Depending on the viral load detected, variant numbers ranged from 30 to 100 variants per patient. Fixation of mutations within a patient depended on intra-host evolutionary dynamics, while between infector-infected pairs they depended on inter-host bottlenecks. Bottlenecks for transmissions are high for variants. The researchers estimated the number of virions that start an infection in a new host at around 10 −1000 infectious viruses, limiting the transmission of less frequent variants and their fixation in the population (Popa *et al.*, 2020). A large study from UK showed that most individuals harboured variant sites under strong within-host evolutionary constraints over the entire genome, including the spike protein. The researchers suggest that the transmission bottleneck is wide enough to permit co-transmission of multiple genotypes in some instances, but small enough that multiple variants do not persist after a few rounds of subsequent transmissions (Lythgoe *et al.*, 2020). One study documented infections with two distinct viral populations co-occurring in the same COVID-19 patient but at different parts of the body. One variant was recovered from a throat swab, and another variant was recovered from a sputum sample of the same patient. Only the virus found in the throat swab was transmitted to a contact person (Wölfel *et al.*, 2020).

SARS-CoV-2 evolution in immunosuppressed patients

Many patients are maintained under medically induced immunosuppression. Such patients might be unable to eliminate the virus replication, leading to a persistent infection and thus to prolonged *in vivo* evolution of the virus. There is some speculation that the complex variants observed in several countries might have acquired their mutations in immunosuppressed patients. Some data support such speculations. An immunosuppressed autoimmune patient showed a persistent infection with SARS-CoV-2 over 154 days despite extensive treatment. The virus grew to high titres in the patient in the nasopharynx, lung, spleen and blood, and the patient excreted infectious virus over the observation period. The patient was infected with a single viral type that showed *in vivo* an accelerated evolution with deletions and point mutations. An excess of non-synonymous over synonymous mutations suggested *in vivo* selection of the virus variants. Notably, most genome changes were seen over the viral spike protein (Choi *et al.*, 2020). Another immunosuppressed leukaemia patient showed a

persistent viral infection over 105 days and shedding of infectious virus over 70 days from the naso-, but not the oropharynx. The researchers observed a marked within-host genomic evolution of SARS-CoV-2 with continuous turnover of dominant viral variants, while the replication kinetics of the excreted virus in cell culture was not affected. Several single-nucleotide substitutions were detected within ORF1ab, spike, M and ORF8 genes. Two in-frame deletions were transiently observed in the N-terminal domain (NTD) of the spike protein (Avanzato *et al.*, 2020). An immunosuppressed lymphoma patient also demonstrated a persistent SARS-CoV-2 infection for at least 119 days. Nasopharynx and sputum samples showed variants differing by two substitutions in ORF1a. Sequential sequencing of nine viruses from the patient demonstrated within-host evolution, with several non-synonymous mutations in ORF1a becoming fixed (Baang *et al.*, 2021).

A lymphoma patient suffering from B cell- and T cell-combined immunodeficiency died from COVID-19 after a persistent infection of 102 days. During the first 40 days of persistent infection without antiviral treatment, only one mutation in ORF7 was observed. However, when the patient was treated with transfusions of convalescent plasma, a complex pattern of mutants emerged, waxing and waning in prevalence. Directly after plasma treatment, a dominant viral strain bearing spike protein mutations (substitution D796H and deletion ΔH69/ ΔV70 in the N-terminal domain NTD) was observed. However, after waning of the transfused antibodies, this variant was replaced by variants showing Spike mutations Y200H and T240I accompanied by mutations in two non-structural proteins (NSP2, NSP15) and in the RNA-dependent RNA polymerase RdRp. Patterns in the variant frequencies suggest competition between virus populations carrying different mutations. After another infusion of plasma, re-emergence of the initial mutant viral population (D796H + ΔH69/ΔV70) was observed. This double mutation showed an increased infectivity in cell culture and was less sensitive to neutralization by the plasma. Despite the fluctuations in the viral population, the overall viral load remained high and constant leading to hyperinflammation, multiorgan failure and death. The authors do not think that the described processes occur in immune-competent patients treated with plasma but recommend to treat immune-deficient patients under enhanced infection control precautions to avoid creating mutants that could propagate in the general population and compromise vaccine protection (Kemp *et al.*, 2021). However, not all immunosuppressed COVID-19 patients show in-host evolution of the virus. US clinicians investigated 20 SARS-CoV-2-infected patients who were immunosuppressed for stem cell transplantation. In cell culture, they detected excretion of infectious virus for up

to 61 days. Each individual maintained a virus with the same genome sequence, indicating a persistent infection.

Coronaviruses slowly acquire substitutions due to a proofreading RNA-dependent RNA polymerase slowing its evolution during an epidemic. US scientists wondered whether coronaviruses have methods to develop diversity that escape control from proofreading and proposed deletions as a diversity creating mechanism. Within 146 800 SARS-CoV-2 genome sequences, they identified 1100 viruses with deletions in the S gene, 90% occupied four discrete sites within the N-terminal domain which they named recurrent deletion regions (RDRs 1-4). The deletions were found in replication competent viruses. RDR 1 and 3 overlap epitopes recognized by monoclonal antibodies, but human convalescent polyclonal antisera neutralized efficiently the deletion mutants, suggesting that many more changes would be required to generate serologically distinct SARS-CoV-2 variants. Fitness of RDR variants is evident by their representation in persistent infections of immunosuppressed patients. Defining recurrent and convergent patterns of adaptation in viral genomes can provide predictions of future evolutionary trajectories of viruses in the pandemic. It is significant that the variants of global concern are RDR mutants and include a Mink variant with Δ69-70, the UK variant B.1.1.7 with Δ69-70 and Δ144/145 and the South African variant B.1.351 with Δ242-244 (McCarthy *et al.,* 2021).

Outlook

Due to its practical importance, coronavirus genome sequencing is being conducted at an unprecedented quantitative scale. Since it follows the unfolding of a pandemic with a previously unknown temporal and geographical resolution, this research field opens up new research opportunities for the understanding of viral evolution. Central subjects of viral evolution theory can be studied with this material, as alluded to by key terms used in this review, such as intra-host and inter-host evolution, mutation rate, error catastrophe, selection versus genetic drift, migration and transmission bottlenecks. The sheer amount of sequence data with their spatiotemporal resolution and increasing combination with clinical and epidemiological metadata provide the raw material for powerful statistical tests of evolutionary theories whose relevance goes far beyond general virology. The absence of data on some evolutionary processes or concepts is also of note: Recombination as a process for creation of beneficial genotypes or for purging of deleterious mutations is under-investigated with SARS-CoV-2, despite the fact that for coronaviruses, such as murine hepatitis virus, up to 25% of the progeny

genomes of coinfected cells might be recombinant genomes. The relative lack of data on processes, such as intra-host evolution over time and along different organs within the same patient, is also striking. Does SARS-CoV-2 evolve according to different evolutionary trajectories in the nasopharynx, in the lung and in the gut and can we exploit such differences? It will also be important to include more discussion on the concept of the RNA virus quasi-species. An important hypothesis in this context is the quasi-species effect, which is also described as the concept of the 'survival of the flattest'. This concept describes the situation where a population of mutants that have a similar mean fitness can outcompete a population with a lower average fitness despite containing variants of higher individual fitness. If not a specific viral mutant, but a viral population is the unit of selection, we should not tacitly anticipate the 'survival of the fittest' concept (e.g. a variant virus that escapes a neutralizing antibody) as the determinant of viral evolutionary success. It is probably simplistic to consider that nucleotide sites evolve independently. Not only do amino acids located distantly in the primary structure of a given protein interact, different proteins of a virus also interact intimately. Therefore, interactions of mutant sites in different viral genes will also determine the fitness of a variant, introducing the concept of antagonistic and synergistic epistatic interactions for the SARS-CoV-2 genome analysis. Suddenly, very theoretical concepts of viral evolution get a practical public health importance, e.g. for the risk assessment of the variant SARS-CoV-2 viruses for the future evolution of the pandemic, and what they mean for vaccination efforts or whether SARS-CoV-2 will evolve towards higher transmission, higher virulence or towards attenuation to seasonal coronavirus infections or even getting extinct. A fruitful crosstalk between theoretical evolutionary biologists, virologists and public health experts can be anticipated from analysing the SARS-CoV-2 sequence space.

Conflict of interest

None declared.

References

Andreano, E., Piccini, G., Licastro, D., Casalino, L., Johnson, N.V., Paciello, I., *et al.* (2021) SARS-CoV-2 escape in vitro from a highly neutralizing COVID-19 convalescent plasma. *bioRxiv.* https://doi.org/10.1101/2020.12.28.424451

Avanzato, V.A., Matson, M.J., Seifert, S.N., Pryce, R., Williamson, B.N., Anzick, S.L., *et al.* (2020) Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an Asymptomatic Immunocompromised Individual with Cancer. *Cell* **183:** 1901–1912.

Baang, J.H., Smith, C., Mirabelli, C., Valesano, A.L., Manthei, D.M., Bachman, M.A., *et al.* (2021) (2021) Prolonged Severe Acute Respiratory Syndrome Coronavirus 2 replication in an immunocompromised patient. *J Infect Dis* **223:** 23–27.

Bedford, T., Greninger, A.L., Roychoudhury, P., Starita, L.M., Famulare, M., Huang, M.L., *et al.* (2020) Cryptic transmission of SARS-CoV-2 in Washington State. *Science* **370:** 571–575.

Böhmer, M.M., Buchholz, U., Corman, V.M., Hoch, M., Katz, K., Marosevic, D.V., *et al.* (2020) Outbreak of COVID-19 in Germany resulting from a Single Travel-Associated Primary Case. *Lancet Infect Dis* **20:** 920–928.

Bosco-Lauth, A.M., Hartwig, A.E., Porter, S.M., Gordy, P.W., Nehring, M., Byas, A.D., *et al.* (2020) Experimental infection of domestic dogs and cats with SARS-CoV-2: Pathogenesis, transmission, and response to reexposure in cats. *Proc Natl Acad Sci USA* **117:** 26382–26388.

Buss, L.F., Prete, C.A. Jr, Abrahim, C.M.M., Mendrone, A. Jr, Salomon, T., de Almeida-Neto, C., *et al.* (2021) Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. *Science* **371:** 288–292.

Candido, D.S., Claro, I.M., de Jesus, J.G., Souza, W.M., Moreira, F.R.R., Dellicour, S., *et al.* (2020) Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science* **369:** 1255–1260.

CDC Centers for Disease Control and Prevention (2021) *New COVID-19 variants*. [WWW document]. URL https://www.cdc.gov/coronavirus/2019-ncov/transmission/variant.html.

Cele, S., Gazy, I., Jackson, L., Hwa, S.-H., Tegally, H., Lustig, G., *et al.* (2021) *Escape of SARS-CoV-2 501Y.V2 variants from neutralization by convalescent plasma.* medRxiv preprint. [WWW document]. URL https://www.medrxiv.org/content/10.1101/2021.01.26.21250224v1.

Chan, J.F., Zhang, A.J., Yuan, S., Poon, V.K., Chan, C.C., Lee, A.C., *et al.* (2020) Simulation of the clinical and pathological manifestations of Coronavirus Disease 2019 (COVID-19) in golden Syrian hamster model: implications for disease pathogenesis and transmissibility. *Clin Infect Dis* **71:** 2428–2446.

Choi, B., Choudhary, M.C., Regan, J., Sparks, J.A., Padera, R.F., Qiu, X., *et al.* (2020) Persistence and evolution of SARS-CoV-2 in an immunocompromised host. *N Engl J Med* **383:** 2291–2293.

Davies, N.G., Barnard, R.C., Jarvis, C.I., Kucharski, A., Munday, J., Pearson, C.A.B., *et al.* (2020a) *Estimated transmissibility and severity of novel SARS-CoV-2 Variant of Concern 202012/01 in England.* medRxiv preprint. [WWW document]. URL https://www.medrxiv.org/content/10.1101/2020.12.24.20248822v1.

Davies, N.G., Jarvis, C.I., Edmunds, W.J., Jewell, N.P., Diaz-Ordaz, K., Keogh, R.H., *et al.* (2020b) *Increased hazard of death in community-tested cases of SARS-CoV-2 Variant of Concern 202012/01.* medRxiv. https://

doi.org/10.1101/2021.02.01.21250959. Preprint 2021. medRxiv preprint. [WWW document]. URL https://www.medrxiv.org/content/10.1101/2021.02.01.21250959v2.

Dearlove, B., Lewitus, E., Bai, H., Li, Y., Reeves, D.B., Joyce, M.G., *et al.* (2020) A SARS-CoV-2 vaccine candidate would likely match all currently circulating variants. *Proc Natl Acad Sci USA* **117:** 23652–23662.

ECDC European Centre for Disease Prevention and Control (2020a) *Rapid increase of a SARS-CoV-2 variant with multiple spike protein mutations observed in the United Kingdom*. [WWW document]. URL https://www.ecdc.europa.eu/sites/default/files/documents/SARS-CoV-2-variant-multiple-spike-protein-mutations-United-Kingdom.pdf.

ECDC European Centre for Disease Prevention and Control (2020b) *Rapid Risk Assessment: Detection of new SARS-CoV-2 variants related to mink*. [WWW document]. URL https://www.ecdc.europa.eu/en/publications-data/detection-new-sars-cov-2-variants-mink.

Faria, N.R., Claro, I.M., Candido, D., Moyses Franco, L.A., Andrade, P.S., Coletti, T.M., *et al.* (2021) *Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings*. Virological Forum preprint. [WWW document]. URL https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586.

Fauver, J.R., Petrone, M.E., Hodcroft, E.B., Shioda, K., Ehrlich, H.Y., Watts, A.G., *et al.* (2020) Coast-to-coast spread of SARS-CoV-2 during the early epidemic in the United States. *Cell* **181:** 990–996.e5.

FDA (2020) *Genetic Variants of SARS-CoV-2 May Lead to False Negative Results with Molecular Tests for Detection of SARS-CoV-2 - Letter to Clinical Laboratory Staff and Health Care Providers*. [WWW document]. URL https://www.fda.gov/medical-devices/letters-health-care-providers/genetic-variants-sars-cov-2-may-lead-false-negative-results-molecular-tests-detection-sars-cov-2.

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

Geoghegan, J.L., and Holmes, E.C. (2021) Virus evolution. In *Fields Virology Vol. 1 Emerging viruses.* Howley, P.M. (ed.). Philadelphia, Wolters Kluwer.

Gibb, R., Redding, D.W., Chin, K.Q., Donnelly, C.A., Blackburn, T.M., and Newbold, T., *et al.* (2020) Zoonotic host diversity increases in human-dominated ecosystems. *Nature* **584:** 398–402.

Gobeil, S., Janowska, K., McDowell, S., Mansouri, K., Parks, R., Manne, K., *et al.* (2020) D614G mutation alters SARS-CoV-2 spike conformational dynamics and protease cleavage susceptibility at the S1/S2 junction. *bioRxiv.* https://doi.org/10.1101/2020.10.11.335299. Preprint. [WWW document]. URL https://www.biorxiv.org/content/10.1101/2020.10.11.335299v1.

Gonzalez-Reiche, A.S., Hernandez, M.M., Sullivan, M.J., Ciferri, B., Alshammary, H., Obla, A., *et al.* (2020) Introductions and early spread of SARS-CoV-2 in the New York City area. *Science* **369:** 297–301.

Hallal, P.C., Hartwig, F.P., Horta, B.L., Silveira, M.F., Struchiner, C.J., Vidaletti, L.P., *et al.* (2020) SARS-CoV-2 antibody prevalence in Brazil: results from two successive

nationwide serological household surveys. *Lancet Glob Health* **8:** e1390–e1398.

Hosnie, M.J., Epifano, I., Herder, V., Orton, R.J., Stevenson, A., Johnson, N., *et al.* (2020) Respiratory disease in cats associated with human-to-cat transmission of SARS-CoV-2 in the UK. *bioRxiv*. Preprint. [WWW document]. URL https://www.biorxiv.org/content/10.1101/2020.09.23.309948v1.

Hou, Y.J., Chiba, S., Halfmann, P., Ehre, C., Kuroda, M., Dinnon, K.H. 3rd, *et al.* (2020) SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* **370:** 1464–1468.

Kemp, S.A., Collier, D.A., Datir, R.P., Ferreira, I.A.T.M., Gayed, S., Jahun, A., *et al.* (2021) SARS-CoV-2 evolution during treatment of chronic infection. *Nature*. https://doi.org/10.1038/s41586-021-03291-y. Online ahead of print.

Koopmans, M. (2020) SARS-CoV-2 and the human-animal interface: outbreaks on mink farms. *Lancet Infect Dis* **21:** 18–19.

Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., *et al.* (2020) Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* **182:** 812–827.e19.

Lam, T.T., Jia, N., Zhang, Y.W., Shum, M.H., Jiang, J.F., Zhu, H.C., *et al.* (2020) Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature* **583:** 282–285.

Liu, Z., VanBlargan, L.A., Bloyet, L.M., Rothlauf, P.W., Chen, R.E., and Stumpf, S., *et al.* (2020) Landscape analysis of escape variants identifies SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *bioRxiv*. https://doi.org/10.1101/2020.11.06.372037. Preprint. https://www.biorxiv.org/content/10.1101/2020.11.06.372037v2

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., *et al.* (2020) Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395:** 565–574.

Lythgoe, K.A., Hall, M., Ferretti, L., de Cesare, M., MacIntyre-Cockett, G., Trebes, A., *et al.* (2020) Within-host genomics of SARS-CoV-2. bioRxiv preprint. [WWW document]. URL https://www.biorxiv.org/content/10.1101/2020.05.28.118992v4.

McCarthy, K.R., Rennick, L.J., Nambulli, S., Robinson-McCarthy, L.R., Bain, W.G., Haidar, G., and Duprex, W.P. (2021) Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *Science* eabf6950. https://doi.org/10.1126/science.abf6950. Online ahead of print.

Mollentze, N., and Streicker, D.G. (2020) Viral zoonotic risk is homogenous among taxonomic orders of mammalian and avian reservoir hosts. *Proc Natl Acad Sci USA* **117:** 9423–9430.

Nadeau, S.A., Vaughan, T.G., Scire, J., Huisman, J.S., and Stadler, T. (2021) The origin and early spread of SARS-CoV-2 in Europe. *Proc Natl Acad Sci USA* **118:** e2012008118.

Olival, K.J., Hosseini, P.R., Zambrana-Torrelio, C., Ross, N., Bogich, T.L., and Daszak, P. (2020) Host and viral traits predict zoonotic spillover from mammals. *Nature* **546:** 646–650.

Oude Munnink, B.B., Nieuwenhuijse, D.F., Stein, M., O'Toole, Á., Haverkate, M., Mollers, M., *et al.* (2020) Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat Med* **26:** 1405–1410.

Patterson, E.I., Elia, G., Grassi, A., Giordano, A., Desario, C., Medardo, M., *et al.* (2020) Evidence of exposure to SARS-CoV-2 in cats and dogs from households in Italy. *Nat Commun* **11:** 6231.

Plante, J.A., Liu, Y., Liu, J., Xia, H., Johnson, B.A., Lokugamage, K.G., *et al.* (2020) Spike mutation D614G alters SARS-CoV-2 fitness and neutralization susceptibility. *bioRxiv*. https://www.biorxiv.org/content/10.1101/2020.09.01.278689v1.

du Plessis, L., McCrone, J.T., Zarebski, A.E., Hill, V., Ruis, C., Gutierrez, B., *et al.* (2021) Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science* **371:** 708–712.

Popa, A., Genger, J.W., Nicholson, M.D., Penz, T., Schmid, D., Aberle, S.W., *et al.* (2020) Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. *Sci Transl Med* **12:** eabe2555.

Sanjuán, R., Nebot, M.R., Chirico, N., Mansky, L.M., and Belshaw, R. (2010) Viral mutation rates. *J Virol* **84:** 9733–9748.

Shen, Z., Xiao, Y., Kang, L., Ma, W., Shi, L., Zhang, L., *et al.* (2020) Genomic diversity in of SARS-CoV-2 in Coronavirus disease 2019 patients. *Clin Infect Dis* **71:** 713–720.

Shi, J., Wen, Z., Zhong, G., Yang, H., Wang, C., Huang, B., *et al.* (2020) Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS–coronavirus 2. *Science* **368:** 1016–1020.

Sia, S.F., Yan, L.M., Chin, A.W.H., Fung, K., Choy, K.T., Wong, A.Y.L., *et al.* (2020) Pathogenesis and transmission of SARS-CoV-2 in golden hamsters. *Nature* **583:** 834–838.

Smith, E.C., Blanc, H., Surdel, M.C., Vignuzzi, M., and Denison, M.R. (2013) Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: evidence for proofreading and potential therapeutics. *PLoS Pathog* **9:** e1003565.

Streicker, D.G., and Gilbert, A.T. (2020) Contextualizing bats as viral reservoirs. *Science* **370:** 172–173.

Tegally, H., Wilkinson, E., Lessells, R.J., Giandhari, J., Pillay, S., Msomi, N., *et al.* (2021) Sixteen novel lineages of SARS-CoV-2 in South Africa. *Nat Med*. https://doi.org/10.1038/s41591-021-01255-3. Online ahead of print.

Volz, E., Hill, V., McCrone, J.T., Price, A., Jorgensen, D., O'Toole, Á., *et al.* (2020) Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell* **184:** 64–75.e11.

Volz, E., Mishra, S., Chand, M., Barrett, J.C., Johnson, R., Geidelberg, L., *et al.* (2021) *Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data*. medRxiv preprint. [WWW document]. URL https://www.medrxiv.org/content/10.1101/2020.12.30.20249034v2

Weissman, D., Alameh, M.G., de Silva, T., Collini, P., Hornsby, H., Brown, R., *et al.* (2021) D614G Spike mutation increases SARS CoV-2 susceptibility to neutralization. *Cell Host Microbe* **29:** 23–31.e4.

Wibmer, C.K., Ayres, F., Hermanus, T., Madzivhandila, M., Kgagudi, P., Lambson, B.E., *et al.* (2020) SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *bioRxiv.* https://doi.org/10.1101/2021.01.18.427166. [WWW document]. URL https://www.biorxiv.org/content/10.1101/2021.01.18.427166v1

Winkler, A. (2007) The approach of the Black Death in Switzerland and the Persecution of Jews, 1348–1349. *Swiss Am Historic Soc Rev* **43:** 2.

Wölfel, R., Corman, V.M., Guggemos, W., Seilmaier, M., Zange, S., Müller, M.A., *et al.* (2020) Virological assessment of hospitalized patients with COVID-2019. *Nature* **581:** 465–469.

Worobey, M., Pekar, J., Larsen, B.B., Nelson, M.I., Hill, V., Joy, J.B., *et al.* (2020) The emergence of SARS-CoV-2 in Europe and North America. *Science* **370:** 564–570.

Xiao, K., Zhai, J., Feng, Y., Zhou, N., Zhang, X., Zou, J.J., *et al.* (2020) Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature* **583:** 286–289.

Young, B.E., Fong, S.W., Chan, Y.H., Mak, T.M., Ang, L.W., Anderson, D.E., *et al.* (2020) Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *Lancet* **396:** 603–611.

Yurkovetskiy, L., Wang, X., Pascal, K.E., Tomkins-Tinch, C., Nyalile, T.P., Wang, Y., *et al.* (2020) Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell* **183:** 739–751.e8.

Zhang, T., Wu, Q., and Zhang, Z. (2020a) Probable Pangolin Origin of SARS-CoV-2 associated with the COVID-19 Outbreak. *Curr Biol* **30:** 1578.

Zhang, X., Tan, Y., Ling, Y., Lu, G., Liu, F., Yi, Z., *et al.* (2020b) Viral and host factors related to the clinical outcome of COVID-19. *Nature* **583:** 437–440.

Zhou, H., Chen, X., Hu, T., Li, J., Song, H., Liu, Y., *et al.* (2020) A Novel Bat Coronavirus Closely Related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein. *Curr Biol* **30:** 2196–2203.e3.