

Chapter 5

The Evolution of SARS-CoV-2



Susan F. Bailey and Mahfuza Akter

Abstract Species evolve over time and viruses are no exception. Extensive genome sequencing of SARS-CoV-2 throughout the global pandemic has allowed for highly detailed and invaluable characterization of the molecular evolution in a virus population, to an extent that has never been seen before. Tracking of the molecular evolutionary changes in SARS-CoV-2 allows for (1) inference of important local and global transmission routes by tracking the distribution of specific genotypes, (2) identification of adaptive evolutionary changes with potential human health implications, and (3) generation of expectations/predictions for future evolutionary changes to better tailor detection, mitigation, and treatment strategies. In this chapter, we begin by outlining the key processes driving evolution in viruses, namely, random genetic drift and natural selection. We summarize the evolutionary history of SARS-CoV-2 within the context of other coronavirus species. Then we explore how ongoing genomic and epidemiological patterns have been used to identify the extent to which natural selection has played a role in the evolution of SAR-CoV-2 throughout the global pandemic. Next, an outline is provided for the World Health Organization's criteria for identifying evolved Variants of Concern (VOC), along with a discussion about the impact of these evolved VOCs on human health. Finally, mechanisms are identified for extensive and rapid adaptive evolution in SARS-CoV-2 which suggest the need for closer monitoring. In addition, the possibilities for future evolution in SARS-CoV-2 are mentioned.

5.1 Evolution in Viruses

Viruses are parasites of cells, containing transmissible deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) as their genetic material. They are a large, extremely diverse group, capable of rapid evolution as genetic variation is continually gen-

S. F. Bailey (✉) · M. Akter
Department of Biology, Clarkson University, Potsdam, NY, USA
e-mail: sbailey@clarkson.edu; akterm@clarkson.edu

erated through random mutations and continually removed from the population through the processes of natural selection and genetic drift. These are the processes that drive evolution—changes in the inherited characteristics of a population from one generation to the next. In viruses, new genetic variation arises rapidly due to high mutation rates, typically through errors in replication which are then passed on to subsequent generations. However, in some RNA viruses, recombination and reassortment can also play an important role. For example, reassortment is quite common among influenza virus strains and may occur when multiple strains infect the same host cell and exchange segments of their genomes to produce hybrid viral progeny [1]. RNA viruses typically have higher mutation rates than DNA viruses. This is because they lack the proofreading activity of RNA polymerases, so new genetic variants constantly arise, allowing for evolution in changing environments [2, 3].

The processes of genetic drift and natural selection impact evolving populations to different degrees, depending on population size. Genetic drift drives random fluctuations in the frequency of different genetic variants in populations simply due to chance events. Under genetic drift conditions, new mutations with a whole range of effects (good, bad, neutral) may rise in frequency in a population simply by chance [4]. Evolution by means of genetic drift occurs in small sized populations, for example, a population that has undergone a severe population “bottleneck.” However, for large populations, evolution is typically driven by natural selection. Natural selection will drive a mutation to increase in frequency in a population if it confers a fitness benefit. On the other hand, it will drive a decrease in frequency if the mutation has a deleterious effect on fitness [5]. Evolutionary change driven by natural selection is referred to as “adaptive evolution.”

Viral life cycles are characterized by massive fluctuations in population sizes across multiple scales (see Fig. 5.1), and these fluctuations result in shifts in the balance between natural selection and genetic drift in driving evolutionary dynamics. In some virus species, infection of a new host can be initiated with as little as a single viral particle or virion (e.g., HIV; [7]). The data are still preliminary, but the transmission bottleneck in SARS-CoV-2 appears to be very small as well [8]. Once a new host is infected, the viral population typically grows rapidly, reaching population sizes that are easily over a billion virions within a matter of days (e.g., SAR-CoV-2: 1–100 billion virions at peak infection; [9]). With some infections, there are additional population fluctuations within the host, for example, dengue viruses experience a large bottleneck in their mosquito hosts as they move into the mosquito salivary gland [10]. Once a viral population has reached sufficient size in its host, it may then be transmitted onto the next susceptible host. At this stage, the size of transmission bottleneck can also depend on the specific mode of transmission, for example, with influenza, fewer virions are passed on through aerosol transmission versus contact transmission [11].

At the inter-host level, viruses can also experience extreme population fluctuations as epidemics initiate, spread exponentially, decrease (e.g., due to the implementation of public health practices such as quarantining), and then repeat, potentially going through multiple epidemic waves. Finally, a virus spreading

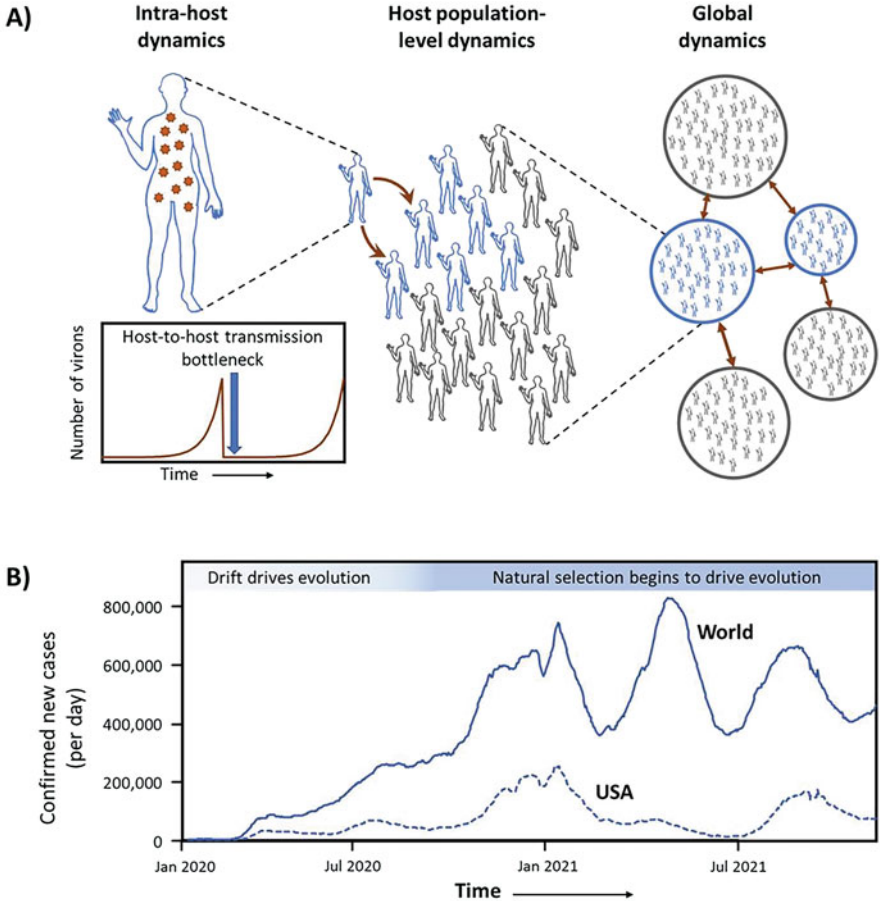


Fig. 5.1 (a) Overview of the different scales at which viral populations fluctuate in size from intra-host to local populations of hosts to global networks of populations. Different selection pressures may be important at each of these levels, and genetic drift will dominate evolutionary dynamics at small population sizes. (b) Fluctuating global population dynamics of SARS-CoV-2 showing the shifting importance of genetic drift versus natural selection in driving evolutionary dynamics in the viral population. Case count data was retrieved from the COVID-19 Content Portal <https://systems.jhu.edu/research/public-health/ncov/> [6]

through a host population may infect another geographically distinct host population, usually via host migration, resulting in complex and even asynchronous viral population fluctuations across an interconnected network of host populations. Figure 5.1 gives an overview of these different scales at which population fluctuations are expected in a virus population, such as that of SARS-CoV-2 (case numbers shown in Fig. 5.1b). Thus, virus populations follow a complex pattern of extreme changes in size which can have important implications for the relative impacts of genetic drift versus natural selection over the course of a pandemic.

Two key characteristics or traits of viruses that can impact their fitness are *transmission rate* and *virulence*. Transmission rate is the rate at which a virus moves from one infected host to a new susceptible host, while virulence can be defined as the harm that a pathogen inflicts on its host and results from the pathogen using the host resources for replication [12]. Much theoretical and empirical work on pathogen evolution has centered on the hypothesis that a trade-off between virulence and transmission drives pathogen fitness [13]. The assumption is that while increasing virulence might initially increase transmission rates (because there are more virions to transmit), increasing virulence may eventually result in increased host mortality rates, which typically slows the transmission rate because the host's infectious period is cut short—deceased hosts don't transmit the virus. Thus, the transmission rate is expected to be highest at intermediate levels of virulence, balancing the costs of replication and infectious period length [13]. Although empirical evidence of this trade-off is not clear-cut, there are many pathogens that appear to have evolved to maintain intermediate virulence because of this trade-off. Some examples are the Zika virus [14], HIV-1 [15], dengue virus [16], and the influenza virus [17]. It remains to be seen whether SARS-CoV will also follow this pattern.

5.2 Evolutionary History of SARS-CoV-2 Within the Coronavirus Group

Coronaviruses (CoVs) are a widely distributed group of RNA viruses, typically highly specialized at infecting humans, birds, and a range of mammals, causing mild to severe disease with both respiratory and gastrointestinal symptoms, depending on the species. This group of viruses is considered one of the zoonotic viruses posing great challenges for the global health community [18, 19]. Coronaviruses belong to the subfamily *Coronaviridae* which is subdivided in four genera: *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, and *Deltacoronavirus* in family *Coronaviridae* (not to be confused with the Alpha, Beta, Gamma, and Delta SARS-CoV-2 variants), categorized in the order *Nidovirales* [20]. Among the four genera, alpha- and betacoronaviruses infect only mammals, while gamma- and delta-coronaviruses infect primarily avian species along with few mammals [21, 22]. Alpha- and beta-CoVs normally cause respiratory illness to humans and gastroenteritis to other animals; however, there are also reports of hepatic and neurologic syndromes due to infection [23]. Although bats are considered as the major natural host shaping the evolutionary dynamics of CoVs, these viruses also circulate among wildlife and domestic livestock. These secondary animal species can play important roles as intermediate hosts before infecting humans [21, 24]. There are already seven reported instances of CoVs being transmitted from animals to humans that have led to the emergence of human CoVs with a wide range of virulence and transmission rates [25]. There are also some CoVs circulating in bats that are reported to have the capability of infecting humans without an intermediate host [20]. Table 5.1 outlines

Table 5.1 Species of coronavirus reported in human hosts [19, 21, 26]

Species in human host	Year of isolation in humans	Possible reservoir	Intermediate host
HCoV-229E	1965	<i>Hipposideros caffer ruber</i> (Noack's leaf-nosed bat)	Alpacas
HCoV-OC43	1967	Unconfirmed rodent	Cattle
SARS-CoV	2003	Horseshoe bat	Civet cat
HCoV-HKU1	2004	Unconfirmed rodent	Unknown
HCoV-NL63	2004	<i>Triaenops afer</i> (African trident bat)	Unknown
MERS-CoV	2012	<i>Taphozous perforatus</i> (Egyptian tomb bat), <i>Pipistrellus</i> spp. (Common pipistrelle), <i>Neoromicia</i> spp. (Vesper bat)	Dromedary camels
SARS-CoV-2	2019	<i>Rhinolophus</i> spp. (Horseshoe bat)	Pangolin (unconfirmed)

the seven human coronaviruses (HCoVs), and their animal origins. Figure 5.2 shows the evolutionary relationships between those HCoVs and a few other animal hosted coronaviruses within the alpha and beta groups.

Among the seven reported human coronaviruses, four of them include HCoV-229E, HCoV-OC43, HCoV-NL63, and HCoV-HKU1 that represent endemic and low virulence HCoVs causing upper respiratory tract disease, responsible for up to 15–30% of the common cold symptoms in adults. The earliest, HCoV-229E, was isolated in 1965 using standard tissue culture from a volunteer. In 1967, HCoV-OC43 was recovered from a tracheal and nasal organ culture from the Common Cold Unit in Salisbury, United Kingdom. Only these two HCoVs were widely reported and under study until the twenty-first century. In 2004, HCoV-NL63 was isolated from a 7-month-old child in the Netherlands, while in 2005, HCoV-HKU1 was isolated from a Hong Kong patient with pneumonia. Later both strains were identified in adults and infants, indicating that these two HCoVs can be considered as new agents responsible for respiratory infections [29, 30]. These four HCoVs generally demonstrate winter seasonality between the months of December and April and typically cause mild upper respiratory infections in humans with some exception of severe lower respiratory infection, for example, in the cases of elderly or immunocompromised individuals [31].

Apart from these four mild HCoVs, over the last two decades, the world has witnessed three major outbreaks of coronaviruses with increased morbidity and transmission rates. In 2002–2003, the outbreak of severe acute respiratory syndrome (SARS-CoV) occurred in Guangdong Province of China and quickly spread to over 30 countries, infecting 8000 people with a mortality rate of approximately 10%. A decade later, in 2012, MERS-CoV caused a severe respiratory disease that emerged in the 27 countries in the Middle East, Europe, North Africa, and Asia. MERS-CoV virus is still posing a potential threat in the Middle East where it is still

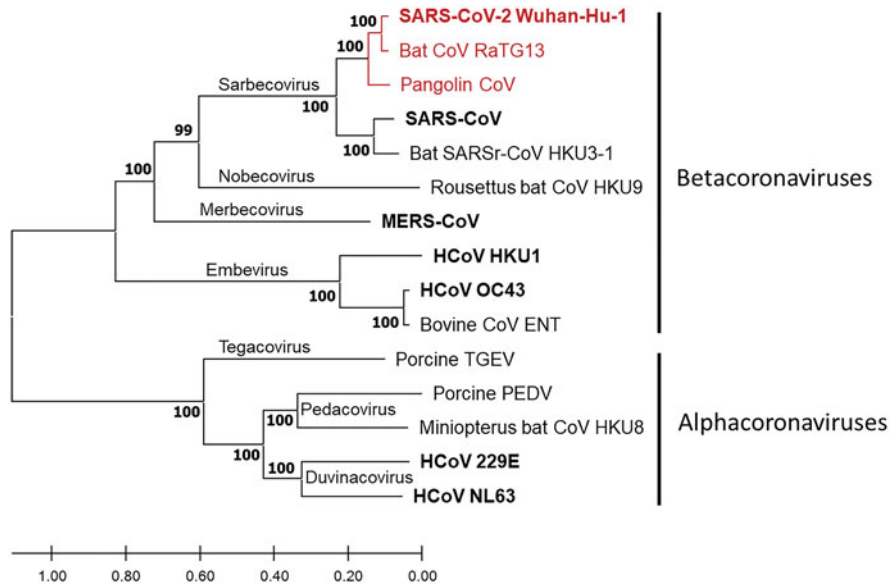


Fig. 5.2 Maximum likelihood phylogeny of a selection of alpha- and betacoronaviruses, including the seven coronaviruses known to infect humans (labeled in bold). This phylogeny was inferred using the GTR + G + I model [27] fit to ORF1ab nucleotide sequences and is drawn to scale with branch lengths measured in the number of substitutions per site (see scale bar). Evolutionary analyses were conducted in MEGA X [28], and bootstrap values were calculated from 500 replications. Red labels indicate coronaviruses in the SARS-CoV-2 group; bold text indicates species found in humans. Accession numbers for sequence data used in this figure include the following: NC_019843.3, MT797634.1, NC_006577.2, NC_006213.1, NC_004718.3, NC_045512.2, MN996532.2, MT121216.1, NC_003045.1, NC_003436.1, MG762674.1, DQ022305.2, NC_010438.1, NC_038861.1, and NC_005831.2

sporadically detected in the human population [32]. In December 2019, the novel coronavirus entitled severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was reported in Wuhan, Hubei, China. On March 11, 2020, the World Health Organization declared a global COVID-19 (the disease caused by SARS-CoV-2) pandemic due to the emergence of different variants and global spread [33, 34].

Bats are considered the most probable zoonotic origin for all three of these major outbreaks of coronavirus. SARS-CoV and MERS-CoV appear to have passed through intermediate hosts: civets and camels, respectively, before transmitting to humans. However, the details of these transmission events are unclear [19, 35, 36]. As for SARS-CoV-2, the highest average genetic similarity with other sequenced virus genomes to date is with CoV RaTG13 sampled from a *Rhinolophus affinis* bat, estimated to have shared a common ancestor with SARS-CoV-2 decades prior to the emergence of SARS-CoV-2 in humans [37]. So far, no viruses with a closer genetic similarity to SARS-CoV-2 have been collected, and so there remains some ambiguity about the immediate source of the virus [18]. Pangolins were also initially

considered as the possible SARS-CoV-2 reservoir due to high genomic similarity of Pangolin CoV with SARS-CoV-2. However, infected pangolins also exhibit clinical and histopathological changes when infected with CoV [32], and natural CoV hosts are expected to be asymptomatic due to their long coevolutionary history.

5.3 Evolution of SARS-CoV-2 in Human Hosts

As the SARS-CoV-2 epidemic began to emerge globally in early 2020, one key part of the global effort to characterize the virus and understand how it was spreading was to acquire whole genome sequence data. These sequence data quickly began to reveal mutational changes. The rate of change of the SARS-CoV-2 genome has been estimated at between $1-5 \times 10^{-6}$ nucleotide substitutions per site per day [38], or across the whole SARS-CoV-2 genome, approximately 20 genetic changes per year within a lineage [39]. In the early days of the pandemic, the observed mutations all appeared to be random and effectively neutral, but they provided a way of tracking transmission routes of specific genotypes, as the virus began to spread around the globe. For example, early on there was much discussion about what the initial transmission routes were for SAR-CoV-2 into the USA. Through sequencing and tracking of specific mutations, researchers were able to conclude that SARS-CoV-2 entered the USA via multiple independent sources from both Asia and Europe [40].

5.4 SARS-CoV-2 Whole Genome Sequence Data

The SARS-CoV-2 genome consists of ~29,903 nucleotides and has a gene composition and structure similar to that of other betacoronaviruses: ~70% of the genome comprises replicase orf1ab, and the rest of the genome consists of S (encoding the structural spike glycoprotein), ORF3a (ORF3a protein), E (structural envelope protein), M (structural membrane glycoprotein), ORF6 (ORF6 protein), ORF7a (ORF7a protein), ORF7b (ORF7b protein), ORF8 (ORF8 protein), N (structural nucleocapsid phosphoprotein), and ORF10 (ORF10 protein). We know much about SARS-CoV-2 diversity and evolution throughout the global pandemic from ongoing analysis of the unprecedented quantity of publicly available SARS-CoV-2 whole genome sequence data. This continually growing collection of sequence data has given researchers the opportunity to observe viral molecular evolution, essentially in real time. For decades, the research field of bioinformatics has supported and promoted the practice of open data sharing [41], and the SARS-CoV-2 sequence data has been no different. Importantly, free access to SARS-CoV-2 sequence data for all researchers means more teams can work in parallel and hopefully allow for more rapid characterization of patterns and even solutions. Data has been made available through databases, including NCBI as well as GISAID, a database

originally developed to organize sharing of influenza sequence data but now the main repository for global SAR-CoV-2 genome sequence data. As of late 2021, there were close to six million publicly available whole SARS-CoV-2 genome sequences on GSAID. Data are from all over the globe, but unsurprisingly the origin of sequence data is unevenly distributed, with the USA and the United Kingdom contributing the highest total numbers. However, on a per-case basis, the USA is far behind many other countries, having sequenced samples from less than 1% of patients diagnosed with COVID-19 [42]. Iceland has the best per-case sequencing rate, at close to 100% of diagnosed infections [43]. Figure 5.3 shows the inferred evolutionary relationships between a subset of available SARS-CoV-2 genomes ($N = 3643$) to give an idea of how the virus population has evolved and diversified over the course of the pandemic.

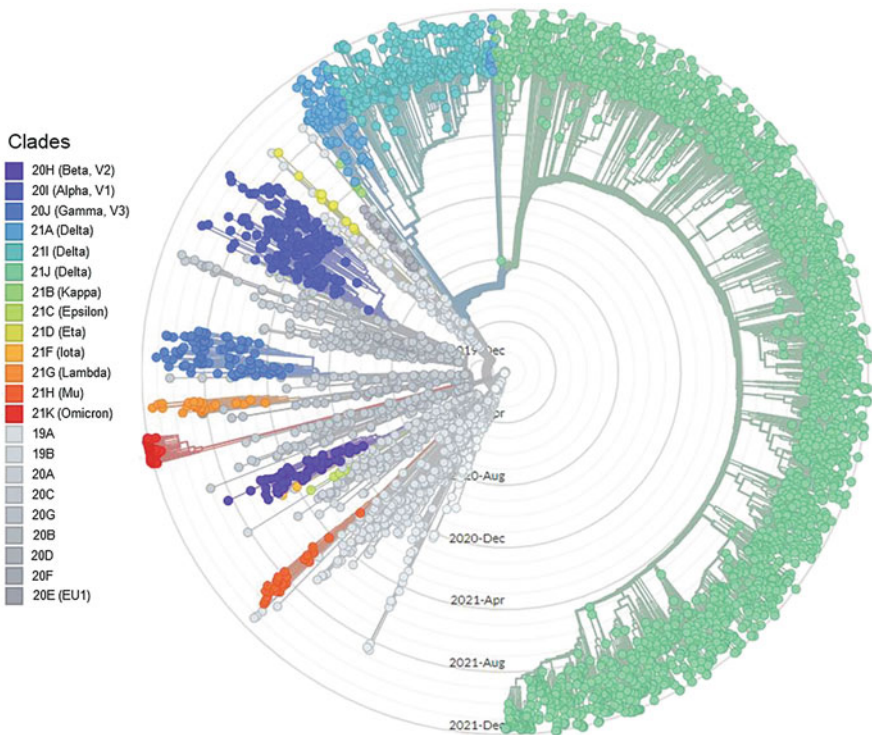


Fig. 5.3 Phylogenetic tree of 3643 genomes sampled globally between December 2019 and December 2021. Clades are named according to Nextstrain nomenclature, which distinguishes clades based on global frequency, year of emergence, and a unique letter. Visualization was performed by nextstrain.org [123] with data from GISAID

5.5 Natural Selection Begins to Drive Adaptive Evolution

By the fall of 2020, clear evidence was emerging to suggest that a few key mutations in the SARS-CoV-2 genome conferred adaptive impacts of human health concern, in particular increased transmission rates and virulence. These adaptations presented new challenges for governments and public health officials trying to reduce cases. Therefore, understanding and predicting the adaptive evolution of SARS-CoV-2 have become increasingly important. Here, we identify two broad categories of approaches that researchers have taken to identify regions in the SARS-CoV-2 genome where natural selection is driving evolution. The first type of approach aims to infer how specific mutations are likely to impact *protein structure and function at the molecular level* and so lead to impacts on virus fitness. The second type of approach analyzes variation in genome sequences collected during the SARS-CoV-2 pandemic to identify *epidemiological and phylogenetic patterns* that are unlikely to have occurred simply due to chance and are likely driven by natural selection. We discuss what we can learn from both types of approaches, as well as their drawbacks. Typically, multiple lines of evidence are required before a specific mutation is identified by consensus as adaptive.

5.6 Evidence of Selection from Models and Experimental Tests of Protein Function

At the molecular and functional level, predictive protein models and a range of in vitro experimental tests have allowed researchers to identify genes and even specific sites within genes on the SARS-CoV-2 genome where mutations are under selection and likely to have impacts on human health. The focus here is on exploring protein function and interactions with the human immune system and host cell. Using this kind of approach, earlier studies examining other coronaviruses had already identified the spike protein as one of crucial importance [44]. Soon after SARS-CoV-2 emerged, the spike protein was identified as important in this novel coronavirus as well [45]. The spike protein interacts with the angiotensin-converting enzyme 2 (ACE2) receptor on the surface of the human host cell. If this interaction is successful, then the coronavirus has access into the human host cell. Predictive models and experimental studies have identified specific regions and amino acid positions that are likely to be most important in impacting how the spike protein interfaces with ACE2 receptors. Of particular importance is the receptor-binding domain (RBD) of the spike protein [46]. This region varies between a closed/down position and an open/up position. To successfully bind with the human ACE2 receptor, it must be in the up position [47–49]. This is also the region where many neutralizing antibodies bind to prevent interaction with ACE2, effectively neutralizing the virus. Many in vitro experiments have been performed looking at the impacts of specific mutations in this region (e.g., [50, 51]). Other regions of

interest in the spike protein that have been identified using this approach include the furin cleavage site (e.g., [52]) and the N-terminal domain (e.g., [53]). There is also evidence that mutations in other genes of the SARS-CoV-2 genome may impact virus fitness (e.g., [54, 55]). However, the spike gene appears to have the largest predicted impact.

While studies exploring the impact of mutations on the structure and function of key proteins are useful for identifying genes and regions of potential importance, sometimes they give us only a hint at the true impacts of these mutations when *in vivo*. Furthermore, for logistical reasons, these studies typically examine the impacts of a single mutational change at a time. However, the impact of a mutation often depends on which other mutations are also present in a genome, a phenomenon known as epistasis [56]. There is now growing evidence of epistatic effects of mutations in SARS-CoV-2 [57]. Therefore, the impacts of multiple mutation combinations will often remain unclear until they play out in real human hosts.

5.7 Evidence of Selection from Phylogenetic and Epidemiological Data

This second type of approach uses observed SARS-CoV-2 sequence data to identify patterns that are unlikely to have occurred simply due to chance and so must be driven by natural selection. One way to do this is by tracking variant frequencies over time and looking for rapid increases in the frequency of virus variants. The assumption is that the genomes of those variants rapidly increasing in frequency must contain one or more mutations that increase viral fitness as compared to other variants circulating at that time. It is usually unclear from this type of data which mutation or mutations are the ones that are impacting fitness. Another drawback of this approach is that sometimes an observed rapid increase in variant frequency may occur simply due to random loss of other variants when a small number of viruses establish in a new host population. These are the potential results of evolution by means of genetic drift in a small population. Because of this potential for genetic drift to result in evolutionary patterns that are difficult to distinguish from the results of natural selection, the evidence for adaptive evolution is significantly strengthened when the same pattern of a rapidly rising variant frequency is observed in multiple independent host populations. This is the type of evidence that allowed for the identification of the spike protein amino acid change D614G as one that increases virus fitness and therefore is a potential human health concern. SARS-CoV-2 variants with this mutation were observed to rapidly increase in frequency across the globe and at multiple geographic levels: national, regional, and municipal, beginning in the spring of 2020 [58].

Another way that researchers have used observed SARS-CoV-2 sequence data to look for evidence of selection is by using phylogenetic techniques to fit models of molecular evolution and identify mutations that have evolved multiple times

across different independently evolving lineages of SARS-CoV-2. This pattern of repeated evolution is called “convergent evolution” and provides strong evidence for adaptive evolution, suggesting that the convergently evolved mutation must provide a fitness benefit of some kind to the virus. For example, Hodcroft et al. [59] showed evidence of convergent evolution at amino acid 677 in the spike protein. Their phylogenetic analyses of SARS-CoV-2 genome sequences from the USA (Sept.–Nov. 2020) revealed the independent evolution of spike mutations at position 677 in seven distinct lineages. The independent rise and spread of the mutation this many times is highly unlikely to have occurred by chance alone. It strongly suggests a fitness advantage in viruses that have it.

5.8 Adaptively Evolved Variants of Human Health Concern

In response to evidence that natural selection was driving the evolution of SARS-CoV-2 variants with important impacts on global human health, the Virus Evolution Working Group of the World Health Organization (WHO) established a name scheme for the variants to simplify and standardize communication about SARS-CoV-2 [60]. This naming scheme also has the added benefit of encouraging the media and the public to move away from location-based names such as the “UK variant” or the “South African variant” which can be stigmatizing to countries and their residents. Other nomenclature systems for naming and tracking SARS-CoV-2 genetic lineages have also been used [61–63]. However, the WHO naming scheme is specifically focused on identifying variants of potential interest for public health. In particular, as is outlined in detail on their website (www.who.int/en/activities/tracking-SARS-CoV-2-variants; [64]), they provide working definitions of what they have named variants under monitoring (VUMs), variants of interest (VOIs), and variants of concern (VOCs). Variants classified into these categories are more closely monitored and assessed through coordinated field and laboratory investigations by WHO member states and partners. Variants are placed on these lists if they seem to have increased transmissibility, virulence, immune escape, or undergone any other evolutionary changes with the potential to impact global public health. VOCs are those variants for which these evolutionary changes with global health impacts have been clearly demonstrated through the types of approaches described in the previous section of this chapter. Variants may be removed from these lists if “they have been conclusively demonstrated to no longer pose a major added risk to global public health compared to other circulating SARS-CoV-2 variants” [64]. Figure 5.4 shows the frequencies of five identified VOCs over time in four different representative countries.

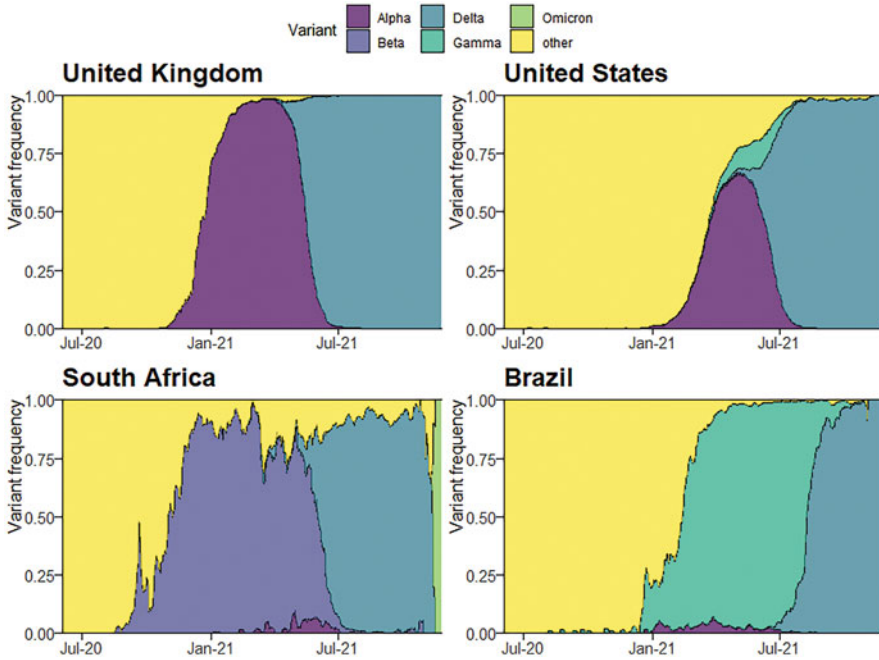


Fig. 5.4 Frequency of five SARS-CoV-2 variants of concern (VOCs) and pre-VOC variants over time, from June 2020 to Nov 2021, in the United Kingdom, the USA, South Africa, and Brazil. Variant frequency data retrieved from <https://cov-spectrum.ethz.ch>

5.8.1 *The First Clearly Identified Adaptively Evolved Variant: Alpha*

The Alpha variant (also known as B.1.1.7, or colloquially the UK variant) was the first SARS-CoV-2 variant for which there was clear evidence suggesting it had evolved impacts on human health. First detected in the United Kingdom in September 2020, it was officially designated as a VOC on December 18, 2020. The genome of the Alpha variant has 23 mutations compared to the Wuhan ancestor strain, but mutations of potential concern for human health all lie in the gene that codes for the spike protein [65]. A few lines of evidence confirmed that this variant evolved by means of natural selection and so has increased fitness compared to the other variants circulating at that time. The first line of evidence was its rapid increase in frequency, starting in Kent and Greater London in September 2020. It rapidly moved across the United Kingdom [65, 66], despite a country-wide lockdown, and was widespread across the United Kingdom by early December [67]. The Alpha variant also quickly rose in frequency in other countries as it began to spread globally. It accounted for most of the infections in the USA and many European countries by the second quarter of 2021 [38, 68, 69]. Epidemiological

models fit to case counts estimated from community-based COVID-19 testing, and secondary contact data suggested that the Alpha variant had a significant transmission advantage of 59–74% over other lineages circulating at that time [70–72]. Further evidence of adaptively evolved changes came from structural modeling of the amino acid changes in the spike protein. Modeling of the evolved amino acid changes predicted that the Alpha variant could bind more easily to host cells in the human respiratory tract [73]. Viral load (the number of viral particles in a host), measured in both hospitalized and walk-in test center patients, showed that patients with the Alpha variant had viral loads that were higher by a factor of 10 relative to non-Alpha variant patients [74, 75]. This suggested that high viral load might be driving the observed increase in transmission rates.

Adaptive evolution of higher transmission rates in SARS-CoV-2 variants is not surprising given that transmission is an important component of viral fitness [76]. Unfortunately, and less clearly linked to viral fitness, higher virulence has also evolved in the Alpha variant. Statistical models examining community test data and deaths found a significant increase in the mortality rate (64% higher; 95% CI: 32–104%) for patients testing positive for the Alpha variant compared to patients with other variants [77–79]. The Alpha variant does not appear to be associated with increased risk of reinfection [80] nor a significant decrease in vaccine efficacy [81–83]. Therefore, these initial adaptively evolved changes did not impact reinfection rates.

5.8.2 *Other Adaptively Evolved Variants Identified: Beta, Gamma, and Delta*

The Beta and Gamma variants emerged around roughly the same time as the Alpha variant but in different regions of the world. The Beta variant (also known as B.1.351) was first detected in South Africa in October 2020. However, subsequent phylogenetic analysis suggests that the variant first evolved in July or August 2020 [84]. The Gamma variant (also known as P.1) was first detected in Brazil in November/December 2020 [85].

Preliminary data suggested that the *Beta variant* had evolved to be more transmissible [84, 86] and resulted in higher rates of reinfection compared with earlier SARS-CoV-2 lineages [87, 88]. In South Africa, the second wave of the SARS-CoV-2 epidemic was larger than the first and was characterized by a more rapid increase in admissions to hospitals, along with increased in-hospital mortality. Some of the increased mortality in the second wave likely arose from higher numbers of cases in older individuals and increased pressure on the health system. However, some of the increased mortality may have also been due to higher virulence of the Beta variant compared to the variants that dominated the first epidemic wave in South Africa [89].

The *Gamma variant* spread rapidly through Brazil and other parts of South America, and its rate of spread suggested that it also had an increased transmission rate. Analysis of samples from SARS-CoV-2 patients in the Manaus region of Brazil showed that Gamma variant samples were significantly associated with higher viral loads. Epidemiological models for variant counts of cases in Manaus, Brazil, also suggest that the Gamma variant was between 1.7 and 2.4 times more transmissible than other non-Gamma variants as it emerged in Brazil [85]. The model further showed that SARS-CoV-2 infections were 1.2–1.9 times more likely to result in mortality after the emergence of Gamma in Brazil. However, this effect may have also been driven by increased stresses on the healthcare system at that time [85].

The *Delta variant* (also known as B.1.617.2) began to emerge a little later. It was first documented in India in October of 2020 and declared a VOC in May 2021. This variant has shown increased transmissibility compared to previous variants, spreading across much of the globe and rapidly replacing the Alpha variant in the United Kingdom [90, 91] and the USA [92]. Studies are ongoing to characterize this variant, but increased transmission is likely driven by higher viral loads, a shorter time to peak viral load, and a shorter incubation period [93–95], along with the ability to effectively resist antibodies [96].

5.8.3 *Emerging Variants of Concern: Omicron*

The first documented case of Omicron was in early November 2021. Since then, it has quickly swept through South Africa and was declared a VOC on November 26, 2021 [97]. Omicron is unusual as compared to other variants sequenced previously. It has more unique mutations than expected and so is quite distantly related to other sequenced SARS-CoV-2 variants. At the time of this writing, the impacts of these unique mutations are still emerging. However, the number and location of mutations on the gene that codes for the spike protein are a concern and suggest that transmission rate is likely to be further increased in this variant. Impacts of these mutations on virulence and vaccine efficacy are unclear.

5.8.4 *Evolved Impacts on Immunity and Vaccine Effectiveness*

Individuals who are infected with SARS-CoV-2 retain some level of immunity after recovery and so are less likely to become infected again. If they are reinfected, they tend to have milder symptoms. Similarly, individuals who have been fully vaccinated with one of the multiple available COVID-19 vaccines are also much less likely to become reinfected. No vaccine is 100% effective but at a roughly 80–95% effectiveness for preventing symptomatic pre-VOC SARS-CoV-2 infections, vaccines have made a significant impact [98]. One concern is that the virus is evolving to better evade human immune defenses, both those generated by past

Table 5.2 Estimates of vaccine efficacy at preventing symptomatic SARS-CoV-2 infection with pre-variant and evolved variants Alpha, Beta, and Delta

Variant	Vaccine type		
	Pfizer/BioNTech BNT162b	Moderna mRNA-1273	Janssen AD26.CoV2b
Pre-variant	95% [99]	93% [100]	72% [101]
Alpha	92% [102]	90% [100]	–
Beta	75% [81, 100]	88% [100]	64% [101]
Delta	83% [102]	73% [103]	–

Numbers shown are for people who are fully vaccinated following recommendations for that vaccine type. “–” indicates that no peer-reviewed estimates were available

infection and those generated through vaccination. There certainly have been significant shifts in vaccine efficacy due to some of the evolved VOCs, and these evolved shifts in vaccine efficacy differ between specific vaccine types. Table 5.2 summarizes some of these evolved shifts in vaccine efficacy for three of the available vaccines. The potential for future adaptive evolution of SARS-CoV-2 in response to vaccines remains to be seen.

5.9 Immunocompromised Patients as Hotspots for SARS-CoV-2 Adaptive Evolution

Typically, individuals infected with SARS-CoV-2 remain infectious for no longer than 10 days, while patients with severe-to-critical illness remain infectious for up to 20 days [\[104, 105\]](#). However, in immunocompromised patients, SARS-CoV-2 can sometimes successfully evade the immune system and persist for much longer [\[106–109\]](#), such as for over 6 months in one documented case [\[110\]](#). Over the course of a long-term infection, the virus population can evolve extensively within the patient (“intra-host evolution”) and accumulate an unusually high number of mutations. Immunocompromised patients are often treated over the course of their SARS-CoV-2 infection with the antiviral drug remdesivir or convalescent plasma (blood plasma derived from patients who have recovered from COVID-19), which may drive adaptive evolution in SARS-CoV-2 to better evade these treatments. For example, in Kemp et al. [\[109\]](#), the authors report on a long-term SARS-CoV-2 infection in an immunosuppressed patient treated with convalescent plasma. In this patient, the viral population evolved to better escape neutralizing antibodies, and eventually this infection was fatal. Thus, rapid evolution driven by selection in immunocompromised patients may have driven the adaptive evolution of some (if not all) of the VOCs [\[84, 106, 108\]](#).

The key piece of evidence suggesting that long-term intra-host evolution (due to chronic infections) has driven the adaptive evolution of many of the VOCs is that these variants all have more unique mutations than expected as compared to other co-circulating variants. For example, each of the first three identified VOCs had

about twice as many mutations in their genomes compared to other co-circulating lineages at the time they first emerged. The Alpha variant had 23 mutations compared to the Wuhan-Hu-1 reference sequence [65]. The Beta variant had 21 mutations [85], and the Gamma variant had 23 mutations [84]. These mutations are spread across the whole genome but tend to be biased toward non-synonymous mutations (those mutations that can affect protein structure) and focused on the S1 gene (the gene that codes for the spike protein) [38].

5.10 Potential Impacts of Cross-Species Spillover on Adaptive Evolution

Since SARS-CoV-2 is originally derived from a nonhuman animal host, we know it is certainly capable of cross-species transmission. Thus, reports of “spillover infections”—transmission of SARS-CoV-2 from humans to nonhuman animals—are not unexpected. Indeed, there have been quite a few species of nonhuman animals reported to have been infected with SARS-CoV-2 including pets and domesticated animals (e.g., cats, dogs, hamsters, rabbits, ferrets, and cows) and those commonly found in zoos (e.g., cougars, tigers, lions, gorillas), along with wildlife (e.g., white-tailed deer and skunks) [29, 111, 112]. Along with being a potential threat to the health of these nonhuman animals, repeated human to animal transmission increases the risk of adaptive evolution of SARS-CoV-2 occurring in a new animal host, followed by retransmission back to humans. The main concern with a high number of spillover cases is that new animal hosts may act as reservoirs, maintaining high numbers of infections and driving novel adaptive evolution of the virus. In 2020, this risk was realized in Denmark [113] and the Netherlands, where SARS-CoV-2 was transmitted from humans to farmed mink and then back to humans again [114, 115]. During its time in the mink population, the virus accumulated some new mutations including up to five in the spike protein, one of which was later found to confer partial escape from human antibodies [116]. In response to these events, the Danish government ordered the culling of all farmed mink in the country, estimated at approximately 17 million animals, and in the Netherlands, more than 2.7 million minks were culled [117]. At this time, it is not clear how big a risk mink farms are of becoming dangerous reservoirs of SARS-CoV-2, but at least for now, proactive measures have been taken to reduce this potential risk.

Other clear examples of transmission to other nonhuman animal hosts and then back to humans again have not been documented, but the potential for this occurring in the future is certainly there. Wild animal species of particular concern as SARS-CoV-2 reservoirs are those that are both abundant and live in close association with humans [111], for example, North American white-tailed deer. A study screening serum samples from wild deer in four US states in 2021 detected SARS-CoV-2 antibodies in 40% of the samples [118]. The high prevalence of antibodies in deer

sampled in this study across multiple states suggests a high likelihood of within-herd spread. A second study tested samples from free-living and captive deer in Iowa in 2020 for the presence of SARS-CoV-2 RNA—evidence of a current or recent infection. These researchers found that one-third of those deer sampled tested positive for SARS-CoV-2 [119]. These studies highlight the need for increased surveillance of deer and potentially other wildlife populations to better determine whether nonhuman animal populations are on their way to becoming long-term reservoirs for SARS-CoV-2, generating novel evolved variants that will spill back over into humans or even other animal hosts.

5.11 Future Evolution of SARS-CoV-2

What will the future evolution of SARS-CoV-2 look like? The evolutionary outcomes depend crucially on population size and spread. The larger the virus population, the greater the chance for rapid and continued adaptive evolution of SARS-CoV-2, with further implications for human health. The more we can work to reduce infection numbers through a range of public health practices including mask-wearing and vaccinations, the better the chance that random genetic drift will dominate the evolutionary dynamics of SARS-CoV-2 instead of natural selection. Without dramatic reductions in case numbers, natural selection will likely continue to drive adaptive evolution in SARS-CoV-2 in ways that are possible. It could be that the virulence-transmission trade-off hypothesis [13] will play out, and we will see a reduction in SARS-CoV-2 virulence moving forward, eventually shifting the disease to one that resembles the common cold. We may also see SARS-CoV-2 evolve in response to some of the biomedical interventions currently in use – either treatments or vaccinations or both. Certainly, we have already started to see SARS-CoV-2 evolve to better evade human host immune defenses generated by the current vaccines. It is very possible that the COVID-19 vaccines need periodic updates to counter ongoing evolutionary changes, much like the annually updated flu vaccine. The best way to prepare for these possibilities is through continued and globally coordinated surveillance, analysis, and modeling of the evolving SARS-CoV-2 genome (e.g., [120–122]).

References

1. McDonald, S. M., Nelson, M. I., Turner, P. E., & Patton, J. T. (2016). Reassortment in segmented RNA viruses: Mechanisms and outcomes. *Nature Reviews. Microbiology*, 14, 448–460.
2. Abbasi, J. (2021). Researchers tie severe immunosuppression to chronic COVID-19 and virus variants. *JAMA*, 325, 2033–2035.
3. Elena, S. F., & Sanjuán, R. (2005). Adaptive value of high mutation rates of RNA viruses: Separating causes from consequences. *Journal of Virology*, 79, 11555–11558.

4. Chao, L. (1990). Fitness of RNA virus decreased by Muller's ratchet. *Nature*, *348*, 454–455.
5. Sanjuán, R., & Domingo-Calap, P. (2021). Genetic diversity and evolution of viral populations. In *Encyclopedia of virology* (pp. 53–61). Elsevier.
6. Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, *20*, 533–534.
7. Zanini, F., Brodin, J., Thebo, L., Lanz, C., Bratt, G., Albert, J., et al. (2015). Population genomics of inpatient HIV-1 evolution. *eLife*, *4*, e11282.
8. Braun, K. M., Moreno, G. K., Wagner, C., Accola, M. A., Rehrauer, W. M., Baker, D. A., et al. (2021). Acute SARS-CoV-2 infections harbor limited within-host diversity and transmit via tight transmission bottlenecks. *PLoS Pathogens*, *17*, e1009849.
9. Sender, R., Bar-On, Y. M., Gleizer, S., Bernshtein, B., Flamholz, A., Phillips, R., et al. (2021). The total number and mass of SARS-CoV-2 virions. *Proceedings of the National Academy of Sciences of the United States of America*, *118*, e2024815118.
10. Sim, S., Aw, P. P. K., Wilm, A., Teoh, G., Hue, K. D. T., Nguyen, N. M., et al. (2015). Tracking dengue virus intra-host genetic diversity during human-to-mosquito transmission. *PLoS Neglected Tropical Diseases*, *9*, e0004052.
11. Varble, A., Albrecht, R. A., Backes, S., Crumiller, M., Bouvier, N. M., Sachs, D., et al. (2014). Influenza A virus transmission bottlenecks are defined by infection route and recipient host. *Cell Host & Microbe*, *16*, 691–700.
12. Bull, J. J. (1994). Virulence. *Evolution*, *48*, 1423–1437.
13. Anderson, R. M., & May, R. M. (1992). *Infectious diseases of humans: Dynamics and control* (Revised ed.). Oxford University Press.
14. Hanley, K. A., Azar, S. R., Campos, R. K., Vasilakis, N., & Rossi, S. L. (2019). Support for the transmission-clearance trade-off hypothesis from a study of Zika virus delivered by mosquito bite to mice. *Viruses*, *11*, 1072.
15. Blanquart, F., Grabowski, M. K., Herbeck, J., Nalugoda, F., Serwadda, D., Eller, M. A., et al. (2016). A transmission-virulence evolutionary trade-off explains attenuation of HIV-1 in Uganda. *eLife*, *5*, e20492.
16. Ben-Shachar, R., & Koelle, K. (2018). Transmission-clearance trade-offs indicate that dengue virulence evolution depends on epidemiological context. *Nature Communications*, *9*, 2355.
17. McKay, B., Ebell, M., Dale, A. P., Shen, Y., & Handel, A. (2020). Virulence-mediated infectiousness and activity trade-offs and their impact on transmission potential of influenza patients. *Proceedings of the Royal Society B: Biological Sciences*, *287*, 20200496.
18. Holmes, E. C., Goldstein, S. A., Rasmussen, A. L., Robertson, D. L., Crits-Christoph, A., Wertheim, J. O., et al. (2021). The origins of SARS-CoV-2: A critical review. *Cell*, *184*, 4848–4856.
19. Singh, D., & Yi, S. V. (2021). On the origin and evolution of SARS-CoV-2. *Experimental & Molecular Medicine*, *53*, 537–547.
20. Rehman, S., Shafique, L., Ihsan, A., & Liu, Q. (2020). Evolutionary trajectory for the emergence of novel coronavirus SARS-CoV-2. *Pathogens*, *9*, 240.
21. Cui, J., Li, F., & Shi, Z.-L. (2019). Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology*, *17*, 181–192.
22. Woo, P. C. Y., Lau, S. K. P., Lam, C. S. F., Lau, C. C. Y., Tsang, A. K. L., Lau, J. H. N., et al. (2012). Discovery of seven novel mammalian and avian coronaviruses in the genus Deltacoronavirus supports bat coronaviruses as the gene source of Alphacoronavirus and Betacoronavirus and avian coronaviruses as the gene source of Gammacoronavirus and Deltacoronavirus. *Journal of Virology*, *86*, 3995–4008.
23. Weiss, S. R., & Leibowitz, J. L. (2011). Coronavirus pathogenesis. In *Advances in virus research* (pp. 85–164). Elsevier.
24. Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., et al. (2020). On the origin and continuing evolution of SARS-CoV-2. *National Science Review*, *7*, 1012–1023.
25. Chen, B., Tian, E.-K., He, B., Tian, L., Han, R., Wang, S., et al. (2020). Overview of lethal human coronaviruses. *Signal Transduction and Targeted Therapy*, *5*, 89.

26. de Wit, E., van Doremalen, N., Falzarano, D., & Munster, V. J. (2016). SARS and MERS: Recent insights into emerging coronaviruses. *Nature Reviews. Microbiology*, *14*, 523–534.
27. Nei, M., & Kumar, S. (2000). *Molecular evolution and phylogenetics*. Oxford University Press.
28. Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*, 1547–1549.
29. Liu, Y., Hu, G., Wang, Y., Ren, W., Zhao, X., Ji, F., et al. (2021). Functional and genetic analysis of viral receptor ACE2 orthologs reveals a broad potential host range of SARS-CoV-2. *Proceedings of the National Academy of Sciences*, *118*, 2025373118.
30. Vijgen, L., Keyaerts, E., Moës, E., Thoelen, I., Wollants, E., Lemey, P., et al. (2005). Complete genomic sequence of human coronavirus OC43: Molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. *Journal of Virology*, *79*, 1595–1604.
31. Li, X., Luk, H. K. H., Lau, S. K. P., & Woo, P. C. Y. (2019). Human coronaviruses: General features. In *Reference module in biomedical sciences* (p. B9780128012383957000). Elsevier.
32. Schoeman, D., Gordon, B., & Fielding, B. C. (2021). Pathogenic human coronaviruses. In *Reference module in biomedical sciences* (p. B9780128187319001000). Elsevier.
33. Sharma, A., Ahmad Farouk, I., & Lal, S. K. (2021). COVID-19: A review on the novel coronavirus disease evolution, transmission, detection, control and prevention. *Viruses*, *13*, 202.
34. WHO. (2020b). *WHO Director-General's opening remarks at the media briefing on COVID-19—11 March 2020*. WHO.
35. Guan, Y., Zheng, B. J., He, Y. Q., Liu, X. L., Zhuang, Z. X., Cheung, C. L., et al. (2003). Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science*, *302*, 276–278.
36. Kahn, J. S., & McIntosh, K. (2005). History and recent advances in coronavirus discovery. *The Pediatric Infectious Disease Journal*, *24*, S223–S227.
37. Boni, M. F., Lemey, P., Jiang, X., Lam, T. T.-Y., Perry, B. W., Castoe, T. A., et al. (2020). Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nature Microbiology*, *5*, 1408–1417.
38. Tao, K., Tzou, P. L., Nouhin, J., Gupta, R. K., de Oliveira, T., Kosakovsky Pond, S. L., et al. (2021). The biological and clinical significance of emerging SARS-CoV-2 variants. *Nature Reviews. Genetics*, *22*, 757–773.
39. Otto, S. P., Day, T., Arino, J., Colijn, C., Dushoff, J., Li, M., et al. (2021). The origins and potential future of SARS-CoV-2 variants of concern in the evolving COVID-19 pandemic. *Current Biology*, *31*, R918–R929. <https://doi.org/10.1016/j.cub.2021.06.049>
40. Zhang, W., Govindavari, J. P., Davis, B. D., Chen, S. S., Kim, J. T., Song, J., et al. (2020). Analysis of genomic characteristics and transmission routes of patients with confirmed SARS-CoV-2 in Southern California during the early Stage of the US COVID-19 pandemic. *JAMA Network Open*, *3*, e2024191.
41. EMBL-EBI. (2021). *Data sharing collaborations*. EMBL-EBI.
42. Maxmen, A. (2021). Why US coronavirus tracking can't keep up with concerning variants. *Nature*, *592*, 336–337.
43. Furuse, Y. (2021). Genomic sequencing effort for SARS-CoV-2 by country during the pandemic. *International Journal of Infectious Diseases*, *103*, 305–307.
44. Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., et al. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*, *426*, 450–454.
45. Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, *579*, 270–273.
46. Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., et al. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, *367*, 1260–1263.

47. Carvalho, T., Krammer, F., & Iwasaki, A. (2021). The first 12 months of COVID-19: A timeline of immunological insights. *Nature Reviews. Immunology*, *21*, 245–256.
48. Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., et al. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, *581*, 215–220.
49. Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., et al. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature*, *581*, 221–224.
50. Baum, A., Fulton, B. O., Wloga, E., Copin, R., Pascal, K. E., Russo, V., et al. (2020). Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science*, *369*, 1014–1018.
51. Schmidt, F., Weisblum, Y., Muecksch, F., Hoffmann, H.-H., Michailidis, E., Lorenzi, J. C. C., et al. (2020). Measuring SARS-CoV-2 neutralizing antibody activity using pseudotyped and chimeric viruses SARS-CoV-2 neutralizing antibody activity. *The Journal of Experimental Medicine*, *217*, e20201181.
52. Johnson, B. A., Xie, X., Bailey, A. L., Kalveram, B., Lokugamage, K. G., Muruato, A., et al. (2021). Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature*, *591*, 293–299.
53. McCallum, M., De Marco, A., Lempp, F. A., Tortorici, M. A., Pinto, D., Walls, A. C., et al. (2021). N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell*, *184*, 2332–2347.e16.
54. Jiang, H., Zhang, H., Meng, Q., Xie, J., Li, Y., Chen, H., et al. (2020). SARS-CoV-2 Orf9b suppresses type I interferon responses by targeting TOM70. *Cellular & Molecular Immunology*, *17*, 998–1000.
55. Xia, H., Cao, Z., Xie, X., Zhang, X., Chen, J. Y.-C., Wang, H., et al. (2020). Evasion of type I interferon by SARS-CoV-2. *Cell Reports*, *33*, 108234.
56. Phillips, P. C. (2008). Epistasis—The essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews. Genetics*, *9*, 855–867.
57. Rochman, N. D., Faure, G., Wolf, Y. I., Freddolino, P. L., Zhang, F. & Koonin, E. V. (2021). Epistasis at the SARS-CoV-2 RBD Interface and the propitiously boring implications for vaccine escape. In *bioRxiv* 2021.08.30.458225.
58. 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.
59. Hodcroft, E. B., Domman, D. B., Snyder, D. J., Oguntuyo, K. Y., Van Diest, M., Densmore, K. H., et al. (2021). Emergence in late 2020 of multiple lineages of SARS-CoV-2 spike protein variants affecting amino acid position 677. In *medRxiv* 2021.02.12.21251658.
60. Konings, F., Perkins, M. D., Kuhn, J. H., Pallen, M. J., Alm, E. J., Archer, B. N., et al. (2021). SARS-CoV-2 variants of interest and concern naming scheme conducive for global discourse. *Nature Microbiology*, *6*, 821–823.
61. Bedford, T., Hodcroft, E. B. & Neher, R. A. (2021). *Updated Nextstrain SARS-CoV-2 clade naming strategy*. <https://nextstrain.org/blog/2021-01-06-updated-sars-cov-2-clade-naming>
62. GISAID. (2021). *Clade and lineage nomenclature aids in genomic epidemiology studies of active hCoV-19 viruses*. GISAID.
63. Rambaut, A., Holmes, E. C., O’Toole, Á., Hill, V., McCrone, J. T., Ruis, C., et al. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology*, *5*, 1403–1407.
64. WHO. (2021). *Tracking SARS-CoV-2 variants*. WHO.
65. Rambout, A., Loman, N., Pybus, O., Barclay, W., Barrett, J., Carabelli, A., et al. (2020). *Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel*. <https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563>
66. Chand, M., Hopkins, S., Achison, C., Anderson, C., Allen, H., Blomquist, P., et al. (2020). Investigation of novel SARS-CoV-2 variant: Variant of concern 202012/01. Public Health England Technical Briefing 2. <https://assets.publishing.service.gov.uk/government/>

[uploads/system/uploads/attachment_data/file/959361/Technical_Briefing_VOC202012-2_Briefing_2.pdf](#)

67. Kraemer, M. U. G., Hill, V., Ruis, C., Dellicour, S., Bajaj, S., McCrone, J. T., et al. (2021). Spatiotemporal invasion dynamics of SARS-CoV-2 lineage B.1.1.7 emergence. *Science*, 373, 889–895.
68. Alpert, T., Brito, A. F., Lasek-Nesselquist, E., Rothman, J., Valesano, A. L., MacKay, M. J., et al. (2021). Early introductions and transmission of SARS-CoV-2 variant B.1.1.7 in the United States. *Cell*, 184, 2595–2604.e13.
69. O’Toole, Á., Hill, V., Pybus, O. G., Watts, A., Bogoch, I. I., Khan, K., et al. (2021). *Tracking the international spread of SARS-CoV-2 lineages B.1.1.7 and B.1.351/501Y-V2 with grinch*. Wellcome Open Research.
70. Davies, N. G., Abbott, S., Barnard, R. C., Jarvis, C. I., Kucharski, A. J., Munday, J. D., et al. (2021a). Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*, 372, eabg3055.
71. Leung, K., Shum, M. H., Leung, G. M., Lam, T. T., & Wu, J. T. (2021). Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Eurosurveillance*, 26, 2002106.
72. Volz, E., Mishra, S., Chand, M., Barrett, J. C., Johnson, R., Geidelberg, L., et al. (2021). Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature*, 593, 266–269.
73. Teruel, N., Mailhot, O., & Najmanovich, R. J. (2021). Modelling conformational state dynamics and its role on infection for SARS-CoV-2 spike protein variants. *PLoS Computational Biology*, 17, e1009286.
74. Jones, T. C., Biele, G., Mühlemann, B., Veith, T., Schneider, J., Beheim-Schwarzbach, J., et al. (2021). Estimating infectiousness throughout SARS-CoV-2 infection course. *Science*, 373, eabi5273.
75. Kidd, M., Richter, A., Best, A., Cumley, N., Mirza, J., Percival, B., et al. (2021). S-variant SARS-CoV-2 lineage B.1.1.7 is associated with significantly higher viral load in samples tested by TaqPath polymerase chain reaction. *The Journal of Infectious Diseases*, 223, 1666–1670.
76. Wargo, A. R., & Kurath, G. (2012). Viral fitness: Definitions, measurement, and current insights. *Current Opinion in Virology*, 2, 538–545.
77. Challen, R., Brooks-Pollock, E., Read, J. M., Dyson, L., Tsaneva-Atanasova, K., & Danon, L. (2021a). Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: Matched cohort study. *British Medical Journal*, 372, n579.
78. Davies, N. G., Jarvis, C. I., Edmunds, W. J., Jewell, N. P., Diaz-Ordaz, K., & Keogh, R. H. (2021b). Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature*, 593, 270–274.
79. Grint, D. J., Wing, K., Williamson, E., McDonald, H. I., Bhaskaran, K., Evans, D., et al. (2021). Case fatality risk of the SARS-CoV-2 variant of concern B.1.1.7 in England, 16 November to 5 February. *Eurosurveillance*, 26, 2100256.
80. Graham, M. S., Sudre, C. H., May, A., Antonelli, M., Murray, B., Varsavsky, T., et al. (2021). Changes in symptomatology, reinfection, and transmissibility associated with the SARS-CoV-2 variant B.1.1.7: An ecological study. *The Lancet Public Health*, 6, e335–e345.
81. Abu-Raddad, L. J., Chemaitelly, H., & Butt, A. A. (2021). Effectiveness of the BNT162b2 Covid-19 vaccine against the B.1.1.7 and B.1.351 variants. *The New England Journal of Medicine*, 385, 187–189.
82. Emary, K. R. W., Golubchik, T., Aley, P. K., Ariani, C. V., Angus, B., Bibi, S., et al. (2021). Efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine against SARS-CoV-2 variant of concern 202012/01 (B.1.1.7): An exploratory analysis of a randomised controlled trial. *The Lancet*, 397, 1351–1362.
83. Heath, P. T., Galiza, E. P., Baxter, D. N., Boffito, M., Browne, D., Burns, F., et al. (2021). Safety and efficacy of NVX-CoV2373 Covid-19 vaccine. *The New England Journal of Medicine*, 385, 1172–1183.

84. Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., et al. (2021). Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature*, 592, 438–443.
85. Faria, N. R., Mellan, T. A., Whittaker, C., Claro, I. M., Candido, D., Mishra, S., et al. (2021). Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science*, 372, 815–821.
86. Pearson, C. A., Russell, T. W., Davies, N. G., Kucharski, A. J., CMMID COVID-19 Working Group, Edmunds, W. J., et al. (2021). *Estimates of severity and transmissibility of novel SARS-CoV-2 variant 501Y.V2 in South Africa*. UpToDate.
87. Cele, S., Gazy, I., Jackson, L., Hwa, S.-H., Tegally, H., Lustig, G., et al. (2021). Escape of SARS-CoV-2 501Y.V2 from neutralization by convalescent plasma. *Nature*, 593, 142–146.
88. Wibmer, C. K., Ayres, F., Hermanus, T., Madzivhandila, M., Kgagudi, P., Oosthuysen, B., et al. (2021). SARS-CoV-2 501Y.V2 escapes neutralization by south African COVID-19 donor plasma. *Nature Medicine*, 27, 622–625.
89. Jassat, W., Mudara, C., Ozougwu, L., Tempia, S., Blumberg, L., Davies, M.-A., et al. (2021). Difference in mortality among individuals admitted to hospital with COVID-19 during the first and second waves in South Africa: A cohort study. *The Lancet Global Health*, 9, e1216–e1225.
90. Challen, R., Dyson, L., Overton, C. E., Guzman-Rincon, L. M., Hill, E. M., Stage, H. B., et al. (2021b). Early epidemiological signatures of novel SARS-CoV-2 variants: Establishme.06.05.21258365.
91. Riley, S., Wang, H., Eales, O., Haw, D., Walters, C. E., Ainslie, K. E. C., et al. (2021). REACT-1 round 12 report: Resurgence of SARS-CoV-2 infections in England associated with increased frequency of the Delta variant. In *MedRxiv* 2021.06.17.21259103.
92. Bolze, A., Cirulli, E. T., Luo, S., White, S., Wyman, D., Rossi, A. D., et al. (2021). SARS-CoV-2 variant Delta rapidly displaced variant alpha in the United States and led to higher viral loads. In *MedRxiv* 2021.06.20.21259195.
93. Grant, R., Charmet, T., Schaeffer, L., Galmiche, S., Madec, Y., Platen, C. V., et al. (2021). Impact of SARS-CoV-2 Delta variant on incubation, transmission settings and vaccine effectiveness: Results from a nationwide case-control study in France. *Lancet Regional Health—Europe*, 13, 100278.
94. Li, B., Deng, A., Li, K., Hu, Y., Li, Z., Xiong, Q., et al. (2021). *Viral infection and transmission in a large well-traced outbreak caused by the Delta SARS-CoV-2 variant—SARS-CoV-2 coronavirus/nCoV-2019 genomic epidemiology*. Virological.
95. Ong, S. W. X., Chiew, C. J., Ang, L. W., Mak, T.-M., Cui, L., Toh, M. P. H., et al. (2021). *Clinical and virological features of SARS-CoV-2 variants of concern: A retrospective cohort study comparing B.1.1.7 (alpha), B.1.315 (Beta), and B.1.617.2 (Delta)*. Social Science Research Network.
96. Planas, D., Veyer, D., Baidaliuk, A., Staropoli, I., Guivel-Benhassine, F., Rajah, M. M., et al. (2021). Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature*, 596, 276–280.
97. TAG-VE. (2021). *Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern*.
98. Shah, A. S. V., Gribben, C., Bishop, J., Hanlon, P., Caldwell, D., Wood, R., et al. (2021). Effect of vaccination on transmission of SARS-CoV-2. *The New England Journal of Medicine*, 385, 1718–1720.
99. Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., et al. (2020). Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *The New England Journal of Medicine*, 383, 2603–2615.
100. Chung, H., He, S., Nasreen, S., Sundaram, M. E., Buchan, S. A., Wilson, S. E., et al. (2021). Effectiveness of BNT162b2 and mRNA-1273 covid-19 vaccines against symptomatic SARS-CoV-2 infection and severe covid-19 outcomes in Ontario, Canada: Test negative design study. *BMJ*, 374, n1943.

101. Sadoff, J., Gray, G., Vandebosch, A., Cárdenas, V., Shukarev, G., Grinsztejn, B., et al. (2021). Safety and efficacy of single-dose Ad26.COV2.S vaccine against Covid-19. *The New England Journal of Medicine*, 384, 2187–2201.
102. Sheikh, A., McMenamin, J., Taylor, B., & Robertson, C. (2021). SARS-CoV-2 Delta VOC in Scotland: Demographics, risk of hospital admission, and vaccine effectiveness. *The Lancet*, 397, 2461–2462.
103. Tang, P., Hasan, M. R., Chemaitelly, H., Yassine, H. M., Benslimane, F. M., Al Khatib, H. A., et al. (2021). BNT162b2 and mRNA-1273 COVID-19 vaccine effectiveness against the SARS-CoV-2 Delta variant in Qatar. *Nature Medicine*, 1–8.
104. Liu, W.-D., Chang, S.-Y., Wang, J.-T., Tsai, M.-J., Hung, C.-C., Hsu, C.-L., et al. (2020). Prolonged virus shedding even after seroconversion in a patient with COVID-19. *The Journal of Infection*, 81, 318–356.
105. 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.
106. 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.e9.
107. Bazykin, G. A., Stanevich, O., Danilenko, D., Fadeev, A., Komissarova, K., Ivanova, A., et al. (2021). Emergence of Y453F and Δ69-70HV mutations in a lymphoma patient with long-term COVID-19—SARS-CoV-2 coronavirus/nCoV-2019 genomic epidemiology. *Virological*.
108. 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. *The New England Journal of Medicine*, 383, 2291–2293. <https://doi.org/10.1056/NEJMc2031364>
109. 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*, 592, 277–282.
110. Borges, V., Isidro, J., Cunha, M., Cochicho, D., Martins, L., Banha, L., et al. (2021). Long-term evolution of SARS-CoV-2 in an immunocompromised patient with non-Hodgkin Lymphoma. *mSphere*, 6, e0024421.
111. Bosco-Lauth, A. M., Root, J. J., Porter, S. M., Walker, A. E., Guilbert, L., Hawvermale, D., et al. (2021). Peridomestic mammal susceptibility to severe acute respiratory syndrome coronavirus 2 infection. *Emerging Infectious Diseases*, 27, 2073–2080.
112. Fischhoff, I. R., Castellanos, A. A., Rodrigues, J. P. G. L. M., Varsani, A., & Han, B. A. (2021). Predicting the zoonotic capacity of mammals to transmit SARS-CoV-2. *Proceedings of the Royal Society B: Biological Sciences*, 288, 20211651.
113. WHO. (2020a). SARS-CoV-2 mink-associated variant strain—Denmark. WHO.
114. Bas, B., Munnink, O., Sikkema, R. S., Nieuwenhuijse, D. F., Molenaar, R. J., Munger, E., et al. (2021). Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science*, 371, 172–177.
115. Larsen, H. D., Fonager, J., Lomholt, F. K., Dalby, T., Benedetti, G., Kristensen, B., et al. (2021). Preliminary report of an outbreak of SARS-CoV-2 in mink and mink farmers associated with community spread, Denmark, June to November 2020. *Eurosurveillance*, 26, 2100009.
116. Hoffmann, M., Zhang, L., Krüger, N., Graichen, L., Kleine-Weber, H., Hofmann-Winkler, H., et al. (2021). SARS-CoV-2 mutations acquired in mink reduce antibody-mediated neutralization. *Cell Reports*, 35, 109017.
117. Koopmans, M. (2021). SARS-CoV-2 and the human-animal interface: Outbreaks on mink farms. *The Lancet Infectious Diseases*, 21, 18–19.
118. Chandler, J. C., Bevins, S. N., Ellis, J. W., Linder, T. J., Tell, R. M., Jenkins-Moore, M., et al. (2021). SARS-CoV-2 exposure in wild white-tailed deer (*Odocoileus virginianus*). *Proceedings of the National Academy of Sciences*, 118, e2114828118.
119. Kuchipudi, S. V., Surendran-Nair, M., Ruden, R. M., Yon, M., Nissly, R. H., Nelli, R. K., et al. (2021). Multiple spillovers and onward transmission of SARS-CoV-2 in free-living and captive white-tailed deer. *Proceedings of the National Academy of Sciences of the United States of America*, 119, e2121644119.

120. Egeren, D. V., Novokhodko, A., Stoddard, M., Tran, U., Zetter, B., Rogers, M., et al. (2021). Risk of rapid evolutionary escape from biomedical interventions targeting SARS-CoV-2 spike protein. *PLoS One*, *16*, e0250780.
121. Rella, S. A., Kulikova, Y. A., Dermitzakis, E. T., & Kondrashov, F. A. (2021). Rates of SARS-CoV-2 transmission and vaccination impact the fate of vaccine-resistant strains. *Scientific Reports*, *11*, 15729.
122. Starr, T. N., Greaney, A. J., Addetia, A., Hannon, W. W., Choudhary, M. C., Dingens, A. S., et al. (2021). Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science*, *371*, 850–854.
123. Hadfield, J., Megill, C., Bell, S. M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., Neher, R. A., (2018). *Nextstrain: real-time tracking of pathogen evolution*, *Bioinformatics*, *34*, 4121–4123, <https://doi.org/10.1093/bioinformatics/bty407>.