Review

One year into the pandemic: Short-term evolution of SARS-CoV-2 and emergence of new lineages

Fernando González-Candelas [a], Marie-Anne Shaw [b], Tung Phan [c], Urmila Kulkarni-Kale [d], Dimitrios Paraskevis [e], Fabio Luciani [f], Hirokazu Kimura [g], Manuela Sironi [h],*

[a] Joint Research Unit Infection and Public Health FISABIO-University of Valencia, Institute for Integrative Systems Biology (I2SysBio) and CIBER in Epidemiology and Public Health, Valencia, Spain
[b] Leeds Institute of Medical Research at St James's, School of Medicine, University of Leeds, Leeds, United Kingdom
[c] Division of Clinical Microbiology, University of Pittsburgh and University of Pittsburgh Medical Center, Pittsburgh, PA, USA
[d] Bioinformatics Centre, Savitribai Phule Pune University, Ganeshkhind, Pune 411007, Maharashtra, India
[e] Department of Hygiene Epidemiology and Medical Statistics, Medical School, National and Kapodistrian University of Athens, 11527 Athens, Greece
[f] University of New South Wales, Sydney 2052, New South Wales, Australia
[g] Department of Health Science, Gunma Paz University Graduate School, Takasaki, Gunma 370-0006, Japan
[h] Bioinformatics Unit, Scientific Institute IRCCS E. MEDEA, Bosisio Parini (LC), Italy

ARTICLE INFO

ABSTRACT

The COVID-19 pandemic was officially declared on March 11th, 2020. Since the very beginning, the spread of the virus has been tracked nearly in real-time by worldwide genome sequencing efforts. As of March 2021, more than 830,000 SARS-CoV-2 genomes have been uploaded in GISAID and this wealth of data allowed researchers to study the evolution of SARS-CoV-2 during this first pandemic year. In parallel, nomenclatures systems, often with poor consistency among each other, have been developed to designate emerging viral lineages. Despite general fears that the virus might mutate to become more virulent or transmissible, SARS-CoV-2 genetic diversity has remained relatively low during the first ~ 8 months of sustained human-to-human transmission. At the end of 2020/beginning of 2021, though, some alarming events started to raise concerns of possible changes in the evolutionary trajectory of the virus. Specifically, three new viral variants associated with extensive transmission have been described as variants of concern (VOC). These variants were first reported in the UK (B.1.1.7), South Africa (B.1.351) and Brazil (P.1). Their designation as VOCs was determined by an increase of local cases and by the high number of amino acid substitutions harboured by these lineages. This latter feature is reminiscent of viral sequences isolated from immunocompromised patients with long-term infection, suggesting a possible causal link. Here we review the events that led to the identification of these lineages, as well as emerging data concerning their possible implications for viral phenotypes, reinfection risk, vaccine efficiency and epidemic potential. Most of the available evidence is, to date, provisional, but still represents a starting point to uncover the potential threat posed by the VOCs. We also stress that genomic surveillance must be strengthened, especially in the wake of the vaccination campaigns.

List of abbreviations

ACE2	angiotensin-converting enzyme 2
ARDS	Acute Respiratory Distress Syndrome
COG-UK	COVID-19 Genomics UK
CP	convalescent plasma
FDA	Food and Drug Administration
GISAID	Global Initiative on Sharing Avian Influenza Data

(continued on next column)

(continued)

NGS	next generation sequencing
NTD	N-terminal domain
PANGO	Phylogenetic Assignment of Named Global Outbreak Lineages
RBD	receptor binding domain
tMRCA	time to the most recent common ancestor
VOC	variant of concern
VUI	variant under investigation
WHO	World Health Organization.

* Corresponding author.
E-mail address: manuela.sironi@bp.LNF.it (M. Sironi).

1. Introduction

The small outbreak of "pneumonia-like illness" reported in Wuhan, China, during December 2019 has outgrown (Zhou et al., 2020b; Li et al., 2020a) to be the most devastating pandemic of the 21st century. The thirteen months of SARS-CoV-2 pandemic have also witnessed synergistic efforts of the global scientific community, not only to demystify the emergence of the virus, but also to develop diagnostics and vaccines. Like never before, the spread of the virus has been tracked nearly in real-time by genome sequencing efforts throughout the world. This wealth of data has allowed researchers to study the evolution of SARS-CoV-2, map the emergence of new viral variants and, in the past few months, track the emergence of novel lineages, which are generating concerns at multiple levels. Focusing on the recent events associated with the detection of novel viral lineages in the UK, South Africa, and Brazil, we provide a summary of the lessons learned during the first year of COVID-19 pandemic.

2. Origin of SARS-CoV-2

The most fundamental questions that needed to be addressed at the beginning of the pandemic were focused on the understanding of the disease and the origin of the virus. The disease of unknown aetiology was termed COVID-19 (Coronavirus disease 2019) by the World Health Organisation (WHO) (Novel Coronavirus (2019-nCoV) Situation Report 1, 21 January 2020, World Health Organization, 2020) and it became the first example of "Disease X", caused by an unknown infectious agent. The virus was referred to as a novel coronavirus (2019-nCoV) by the WHO and subsequently named SARS-CoV-2 by the International Committee on Taxonomy of Viruses (ICTV) (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020).

The outbreak of the SARS-CoV-2 at the animal market in Wuhan was investigated from two perspectives with respect to transmission of the virus. The virus could have been either transmitted to the human population from an animal source or it could have been introduced in the market by individual(s) already infected with the virus. Investigation of initial cases revealed that not all the patients had a direct link to the animal market. Further, information on the appearance of their symptoms around December 1, 2019 indicated that they might have acquired the infection from potentially undetected cases, prior to December 2019.

The initial investigation for the characterisation and identification of the viral pathogen was carried out using whole genome sequencing and the first SARS-CoV-2 genome sequence was deposited in the public domain (Virological.org) on January 10th, 2020 (Wu et al., 2020). This facilitated the comparison of the genomic sequence of isolate "Wuhan-Hu-1" (GenBank accession number NC_045512) with the sequences of known viruses, which provided insight into the evolution of SARS-CoV-2. These sequence-based searches revealed that the Wuhan-Hu-1 isolate of SARS-CoV-2 shared >96% identity with RaTG13, a bat coronavirus belonging to subgenus *Sarbecovirus* (genus *Betacoronavirus)* that was isolated in 2013 from *Rhinolophus affinis*, a horseshoe bat in Yunnan, China. Based on the genetic proximity between RaTG13 and SARS-CoV-2, a bat origin for the latter was proposed (Zhou et al., 2020a). However, viruses similar to SARS-CoV-2 were also found in pangolins (Lam et al., 2020; Xiao et al., 2020). An independent evolutionary analysis using the RNA directed RNA polymerase gene (RdRp), a stable genetic marker present in all RNA viruses, revealed that SARS-CoV-2 isolates are a homogeneous population devoid of recombination and cluster independently of SARS-CoV. This study further revealed that SARS-CoV-2 shares a common ancestor with Bat-CoV (RaTG13) and Pangolin-CoVs, hinting at the presence of additional host(s) that might have also contributed to the evolution of SARS-CoV-2 to infect humans (Kasibhatla et al., 2020). Another novel bat coronavirus, named RmYN02, subsequently

discovered using metagenomic analysis of bat viruses from the Yunnan province of China showed, in some parts of the genome, higher sequence similarity to SARS-CoV-2 as compared to its closest relative RaTG13 (Zhou et al., 2020b). Furthermore, both SARS-CoV-2 and RmYN02 were observed to harbour multiple amino acid insertions at the junction of the S1 and S2 subunits of the spike (S) protein. However, relatively lower sequence identity between the spike proteins of RmYN02 and SARS-CoV-2, coupled with the key differences in their RBDs (receptor binding domains), revealed that RmYN02 might not bind the ACE2 (angiotensin-converting enzyme 2) receptor. These studies also indicated that insertion events in spike proteins could naturally and independently occur amongst betacoronaviruses infecting animals and humans (Zhou et al., 2020b). An independent study conducted by Boni et al. to decipher the origin of the pandemic suggested that, though frequent recombination events might have been responsible for shaping genetic diversity and evolution of sarbecoviruses in general, SARS-CoV-2, in particular, is not a recombinant of any sarbecovirus reported so far (Boni et al., 2020).

Members of the family *Coronaviridae* have large genome sizes (~30,000 nt), with positive sense single-stranded RNA, and have a wide host-range that includes mammals, birds, and fishes (Cui et al., 2019; Forni et al., 2017). Members of the genus *Betacoronavirus* have been responsible for three major outbreaks in nearly two decades of the 21st century: the SARS, MERS and COVID-19 epidemics in 2002–2003, 2012 and 2019–2021, respectively (Chan et al., 2020; Drosten et al., 2003; Zaki et al., 2012). The other two members of this genus that infect humans, namely HCoV-HKU1 and HCoV-OC43, have been associated with mild diseases (Cui et al., 2019; Forni et al., 2017). However, based on the estimate that HCoV-OC43 emerged around 1890s from a related bovine coronavirus, HCoV-OC43 was proposed as the potential causative agent of the "Russian flu" pandemic (that occurred during 1889–1890), which was speculated to be caused by an influenza A virus (Vijgen et al., 2005). For all betacoronaviruses, bats are reported as reservoirs; however, these viruses also infect different intermediate hosts such as palm-civets (SARS-CoV) and dromedary camels (MERS-CoV) (Forni et al., 2017; Cui et al., 2019).

Since the genomic sequence of SARS-CoV-2 was obtained, the spike gene remained the focus of much attention, as it encodes the surface protein that harbours the RBD. Sequence comparisons revealed that it is one of the hyper-variable regions in coronavirus genomes and 5 out of 6 residues involved in receptor binding differ between SARS-CoV-2 and SARS-CoV. Structural and biochemical studies further explained the role of these mutations in binding to the ACE2 receptor, as well as in host switching (Andersen et al., 2020; Letko et al., 2020; Wrapp et al., 2020). Owing to the high sequence similarity between the ACE2 receptor-binding sites of SARS-CoV-2 and the pangolin-CoVs, a potential role of pangolins as intermediate hosts was also deliberated extensively. A more recent analysis positions pangolins as species that might have had a role in the transmission of SARS-CoV-2 to humans, but dismissed the possibility of pangolins contributing to the process of adaptation of SARS-CoV-2 to humans (Boni et al., 2020). Similarly, the presence of a polybasic furin cleavage site in the spike protein of SARS-CoV-2 and its absence in other betacoronaviruses triggered discussion regarding the emergence of furin cleave site due to recombination, which was ruled out. It was suggested that recombination events might have played a role in the evolution of sarbecoviruses prior to the diversification of SARS-CoV-2 (Boni et al., 2020). These analyses have also helped to dismiss the possibility of SARS-CoV-2 being a synthetic construct (Andersen et al., 2020) and underlined a need for systematic surveillance of potential reservoir species, not only to connect the missing dots and decipher the zoonotic spill-over of SARS-CoV-2, but also as a systematic plan for pandemic preparedness. More recent genomic analyses, however, indicate the possibility of recombination in SARS-CoV-2 due to coinfections, which is deliberated in the succeeding sections.

3. COVID-19: from an outbreak to the pandemic

Deciphering the emergence of disease, tracing its source, mapping associated events, identifying the potential mechanisms and/or routes of spread and releasing related information for consumption to governments, health officials and people at large are critical for management of infectious diseases in general and to arrest the spread of the infectious agent in particular. The WHO has played a significant role in monitoring the spread of SARS-CoV-2 through various stages. The WHO, in collaboration with other international agencies, has published several situation reports, guidelines and advisories to monitor the spread of the virus.

The first public release regarding the onset of "pneumonia-like" illness was made based on the detection of a "cluster of cases" by the Wuhan Municipal Health Commission on December 31, 2019. The first case of the disease outside China was reported in Thailand on January 13, 2020. On January 30, when the number of confirmed cases reached 98 and no deaths were reported outside of China, the WHO declared the outbreak of coronavirus as a public health emergency of international concern. The advisory urged nations to use the "window of opportunity" to prevent widespread transmission of the virus and to develop preparedness to combat the disease at every possible level. Subsequently, COVID-19 epidemic announcements were made by various nations. The epidemic threat of SARS-CoV-2 and its implications on global health are reviewed by Sheahan and Frieman (2020).

The WHO, in the interest of the health of the global population, declared the SARS-CoV-2 pandemic on March 11, 2020 (*Coronavirus Disease 2019 (COVID-19) Situation Report – 51*, World Health Organization, 2020), when the cases crossed over 118,000 and were spread over 110 countries. This action also had the purpose to inform people at large that the virus was circulating in several countries due to people with travel history and it was expected to spread far and beyond.

4. SARS-CoV-2 genetic diversity

Similarly to other viruses with an RNA genome, coronaviruses are prone to mutate during their replication and they do so at a much higher rate than viruses and cellular organisms with a DNA genome. In the case of viruses, mutation rates indicate the rate at which errors are made during genome replication, whereas substitution rates indicate the rate at which evolution proceeds at the molecular level, replacing pre-existing alleles by new mutations. For example, substitution rates for Influenza A viruses (negative sense ssRNA) are estimated to be around 1.8×10^{-3} substitutions per site per year (s/s/y) and that of human enterovirus 71 (positive sense ssRNA) is estimated to be 3.4 x 10^{-3} s/s/y (Jenkins et al., 2002). Substitutions rates can also vary across genomic regions of the same virus.

Substitution rates for SARS-CoV-2 derived using complete genome sequences vary slightly based on the molecular clock model used viz., strict and relaxed models. Values are 1.29×10^{-3} s/s/y (95% HPD 5.35×10^{-4}, 2.15×10^{-3}) and 1.23×10^{-4} s/s/y (95% HPD 5.63×10^{-4}, 1.98×10^{-3}) for relaxed and strict clock models, respectively (Duchene et al., 2020). As compared to other positive sense RNA viruses, the substitution rate of SARS-CoV-2 is not exceptional. It is, however, placed at the lower end of the distribution, because of the exonuclease proofreading activity of nsp14 (Eckerle et al., 2010, 2007; Smith et al., 2013; van Dorp et al., 2020a).

The time to the most recent common ancestor (tMRCA) has been calculated for SARS-CoV-2 based on complete genome sequences. Most studies reported a temporal signal with a clock-like pattern of molecular evolution. So far, all analyses have indicated late November 2019 (~13 to 19 Nov 2019, 95% CI: August 2019 to December 2019) as the tMRCA using log-normal relaxed molecular clock model(s). It is interesting to note that the tMRCA did not vary much, irrespective of the number of genomes analysed (Gómez-Carballa et al., 2020; Chaw et al., 2020; Ladner et al., 2020). This finding corroborates the fact that the first reported isolate dated to December 1, 2019 from Wuhan, and the

incubation time for this virus is about two weeks. Strict clock models show a slightly earlier date for the tMRCA (November 7, 2019). The likely progression of the SARS-CoV-2 lineages (as described by the Pangolin tool) are A, B and B1 (Gómez-Carballa et al., 2020).

Analyses of members of the *Sarbecovirus* subgenus (including Pangolin-CoV, Bat-CoV, and, SARS-CoV) revealed that the divergence time estimate for SARS-CoV-2 and RaTG13 is year 1969 (Boni et al., 2020). This hints at the fact that viruses closely related to SARS-CoV-2 have been circulating in horseshoe bats for many decades.

4.1. SARS-CoV-2 lineage nomenclature

One important problem in Virology is how to establish a nomenclature system below the species level that is useful, informative, and widely accepted by all those who need to name variants (Rambaut et al., 2020a). Categories below the species level in viruses have usually relied on immunogenic features (e.g., serotypes as in Dengue virus) or distinct and stable genetic differences (e.g., genotypes in hepatitis C virus or subtypes in influenza A virus). However, there are no common rules and additional levels, which usually are of interest for epidemiology, are difficult to accommodate consistently. This is especially so in the initial stages of spread of a new virus, when there has not been enough time to observe clear genetic differences that can be used at least for the major divisions within the species. In the case of SARS-CoV-2, there have been up to three different naming systems coexisting during the first year of the pandemic (Alm et al., 2020).

Most complete genome sequences of SARS-CoV-2 are being uploaded to GISAID (https://www.gisaid.org) since the start of the pandemic. GISAID initiated a naming system based on large clades defined by marker variants from the reference genome (WIV04, MN996528) (Table 1). This scheme is simple but it is not consistent and it fits poorly the evolutionary characteristics of this virus. The high mutation rates of RNA viruses, despite the lower value of Coronaviruses, imply frequent **parallel** and **backward mutations**, impacting reconstruction of evolutionary relationships and thereby introducing inconsistencies in assignment of lineages to viral isolates. If a lineage is defined solely by a few mutations, recurrent mutations make it possible that the same combination appears in a different lineage or a **backward mutation** may make a member of a lineage loose one of the defining mutations.

SARS-CoV-2 sequences deposited in GISAID are readily analysed in different platforms among which NextStrain (www.nextstrain.org) has gained a well-deserved popularity. In this platform, a consistent pipeline of phylogenetic and phylodynamic analysis (Hadfield et al., 2018) is applied to specific subsets of sequences. This is intended to optimize the balance between information and time used to produce it by reducing the total number of sequences used in each "build" (a new global build is constructed on a daily basis from around 4000 sequences). The nomenclature used in NextStrain builds is based on the inferred

Table 1
GISAID clade nomenclature

Clade	Marker variants (nt)	Marker variants (aa)	Number of sequences (210126)
L	WIV04-reference		4675
S	C8782T, T28144C	NS8:L84S	8036
V	G11083T, G26144T	NSP6:L37F, NS3:G251V	5754
G	C241T, C3037T, A23403G	S:D614G	57179
GH	C241T, C3037T, A23403G, G25563T	S:D614G, NS3: Q57H	79275
GR	C241T, C3037T, A23403G, G28882A	S:D614G, N: G204R	127622
GV	C241T, C3037T, A23403G, C22227T	S:D614G, S: A222V	85729
Other minor clades			5535

phylogenetic relationships, ultimately depending on the accumulation of mutations. Evolutionary stable lineages receive a name based on the year of emergence and a successive letter (e.g., 19A corresponds to the first lineage from 2019 and 20B to the second lineage from 2020). Some particular sublineages are identified with additional information (e.g., 20E/EU1, 20I/501Y.V1) but with no systematic rule.

To alleviate these problems, Rambaut et al. (2020a) proposed a new method of nomenclature for within viral species categories, based on evolutionary relationships and epidemiological relevance, denoted PANGO (Phylogenetic Assignment of Named Global Outbreak Lineages). This is a dynamic scheme that accommodates the expanding phylogenetic diversification of SARS-CoV-2 lineages by constraining the number and depth of hierarchical levels. The names start with a letter (A and B, for the earliest lineages) and up to 4 hierarchical levels, each defined as descendant from a preceding level given four conditions (van Dorp et al., 2020c): (a) one or more shared nucleotide differences from the ancestor lineage; (b) comprising at least five genomes with >95% of the genome sequenced; (c) genomes within a lineage exhibit at least one shared nucleotide change among them; and (d) a bootstrap support >70% for the lineage-defining node. The resulting names (e.g., A, B.1, B.2.5, etc.) are informative, consistent, and coherent and facilitate the use of neutral tags, with no negative connotations. Continuous supervision of active lineages is facilitated by easily accessible tools such as Pangolin (https://github.com/hCoV-2019/pangolin). Viral lineages with no observations are assumed to be inactive and are delabelled. The different names that a particular genome can receive are provided in the Nextstrain builds (https://nextstrain.org/ncov). Ideally, it would be desirable to have a "rule of equivalences" that would "translate" the name of a clade/lineage in a nomenclature system to the corresponding name in any other. However, this is not possible. Table 2 summarizes the distribution of most (96.6%) sequences deposited in GISAID as of February 4, 2021, according to the three naming systems. All clades included in GISAID are represented in the table, but only a few of the lineages included in NextStrain and PANGO are considered. The reduction is especially relevant in the PANGO system, because there are 18 sublineages derived from lineage A, and 815 lineages derived from B. It is evident that, although many combinations never occur, it is not possible to establish an equivalence between the clade/lineage a sequence corresponds to in a system and those in the other systems.

Recently, some recombinant lineages have been found in the UK (see below) and an appropriate naming system has been proposed for them within the PANGO framework (Pybus, 2021). The general rules for designation of PANGO lineages still apply, but new recombinant lineages of the highest level are preceded by an X, so we will have XA.1, XB.1.1. These names do not contain information about the putative parental lineages.

4.2. Emergence and spread of novel variants

The epidemic growth of an organism provides ideal conditions for rapid diversification. This is especially so in fast mutating viruses. SARS-CoV-2, as other coronaviruses, has a low mutation rate compared to other RNA viruses as described above (Sironi et al., 2020). The fate of these mutations is governed by the usual processes operating in evolving populations: genetic drift, natural selection, founder effects, and, in some cases, recombination. Most mutations appearing within an infected patient are deleterious, and hence usually eliminated by selection, or neutral, and their fate is dictated by stochastic events (i.e., genetic drift). Although some mutations might have a selective advantage at the within-individual level, their destiny at the population level will depend on factors such as their effects on transmissibility, which might increase or decrease their chances of survival during an epidemic. Fear of mutations was common in the early stages of the COVID-19 pandemic, although they did not represent any significant change in the virulence, transmissibility, lethality, or other relevant features of the infectious viruses (Grubaugh et al., 2020b). Furthermore, some of the mutations defining variants of concern also appeared in earlier lineages without provoking noticeable effects. Apart from the inherent difficulties in demonstrating a clear effect by a single mutation (MacLean et al., 2020; Tang et al., 2020), epidemiological processes may confound the genetic effects on the phenotype when evaluated in real populations.

Some of the early mutations were already present in individuals who originated large clusters of transmission. As a consequence, many of the viruses sampled for genomic surveillance will present shared mutations originating in their common ancestors, eventually giving rise to new variants, i.e., organisms that share a common set of mutations usually associated with successful transmission or particular features. The descendants of these variants gave rise to a new lineage or sublineage, many of which might be present at a given time during the pandemics while others emerged and disappeared some time after. Hence, it is important to differentiate between mutations (particular genetic changes), variants (sets of organisms that share some mutations), and lineages or clades (the set of individuals descending from a given variant). In Virology, a variant becomes a strain when there is a significant change in its transmissibility, pathogenicity, immunogenicity or lethality characteristics compared to other such variants (Kuhn et al., 2013; Van Regenmortel, 2007).

Before the current variants of concern and interest (analysed in detail below), there have been a few notable variants in the COVID-19 pandemic. The first such variant is known as D614G. This name corresponds to an A to G transition at position 23,403 of the SARS-CoV-2 genome (relative to the Wuhan reference sequence), that results in the replacement of an aspartic by a glycine at position 614 of the S or spike gene. The D614G mutation is usually accompanied by three additional mutations: a C-to-T mutation in the 5' UTR position 241, a silent C-to-T mutation at position 3,037, and a C-to-T mutation at position 14,408 that results in an amino acid change in the RNA-dependent RNA polymerase (RdRp:P323L). These mutations defined a clade (G clade under the GISAID nomenclature system) that became dominant worldwide by the end of March 2020. Although several cases of G614 were detected in

Table 2

Distribution of human SARS-CoV-2 sequences of human origin deposited at GISAID (on February 4, 2021) in the major clades/lineages according to the three nomenclature systems: GISAID (clades V, S, O, L, G, GH, GR, GV), NextStrain (only lineages 19A, 19B, 20A, 20A.EU2, 20B, 20C, 20D, 20E(EU1), 20G, 20H/501Y.V2, and 20I/501Y.V1), and PANGO (lineages A*, including 18 sublineages, and B*, including 815 sublineages). In total, 96.6% of the total number of sequences at GISAID are included in this analysis.

	19A		19B		20A		20A.EU2	20B		20C		20D	20E(EU1)		20G	20H/501Y.V2	201/501Y.V1
	A	B	A	B	A	B	B	A	B	A	B	B	A	B	B	B	B
V	3	6169	1														
S	18	1	8824	244		2		1									
O	45	4514	73	42	1	777	31		403	1	223	9		69	78	2	63
L		5138															
G	14	264	1	1	15	66615		13	840	4	284	38		1808			293
GH		55		1	1	21239	12446				49127			37	16968	853	
GR						505	3		104337			4007		6			44381
GV		1				119	2		2				1	99356			

China and Germany by the end of January 2020, they likely represented independent mutations, and the first representative of the D614G variant was sampled on 20 February 2020 in Italy from where it rapidly spread to Europe and the rest of the world (Korber et al., 2020).

The reasons for the rapid spread of this variant have been controversial, because its emergence was coincident with the start of the rapid diffusion of the virus outside of Asia, first in Europe and later in the Americas and rest of the world. The main point of discussion was whether the rapid spread was due to an intrinsic advantage of this variant compared to others or the main causes were related to epidemiological reasons and not to the intrinsic features of the variant (Grubaugh et al., 2020a). Structural and *in vitro* analyses, along with experiments in animal models, of the G614 form of the spike protein point at an increased infectivity and higher viral load than that of the D614 alternative, thus providing support for intrinsic viral properties as responsible for the observed replacement (Korber et al., 2020; Plante et al., 2020). A large study of more than 25,000 genomes from the UK tested the hypothesis of **positive selection** of G614 in this population (Volz et al., 2021). The results were not conclusive concerning the action of natural selection on this variant but they found a significant association with viral load and younger age of patients. Using an even larger sample and a new method that takes into account the recurrence of mutations, van Dorp et al. (2020b) found no evidence that D614G increases viral transmissibility.

Is selective advantage a necessary condition for a variant to increase its frequency? Population genetics theory shows that a mutant allele can increase its frequency, even reach fixation, without the action of selective processes. Genetic drift operates with intensity inversely proportional to population size, which seems to exclude it from acting in viral populations, although with increased overdispersion of transmission the more genetic drift would be expected. Additionally, the epidemic spread of SARS-CoV-2 has been shown to depend strongly on "super-spreading" events (Adam et al., 2020; Liu et al., 2020; Majra et al., 2021), the epidemiological equivalent of population bottlenecks followed by exponential growth, which can increase rapidly the frequency of a neutral, even slightly deleterious, mutation. Such events have been observed during the COVID-19 pandemic. For instance, the 20E.EU1 (or lineage B.1.177) variant was detected at the beginning of the 2020 summer in the North-East of Spain, linked to agricultural workers (Hodcroft et al., 2020). In a few weeks, it became the dominant variant in Spain, which opened its borders to international travel. At the beginning of fall, this variant had become the most prevalent in several European countries, including Switzerland, the UK, and Denmark. The most noticeable variant-defining mutation, A222V again on the spike protein, does not show significant effects on the ability of the protein to mediate viral entry. Hence, its spread seems to be better explained by chance and opportunity than by any selective advantage.

New, adaptive variants may arise at any time during a pandemic. However, benefitting from being advantageous depends on the environment, including external factors and genomic context, and on chance, because population size (i.e., genetic drift) is still important in medium sized populations and in those undergoing frequent bottlenecks. The same D614G mutation that became dominant in the COVID-19 pandemic was detected in other lineages before it originated the highly successful clade 20A. So far, there has been very little immune selection on SARS-CoV-2 from the human population, but this is likely to change as more people develop immune defences as a result of natural infection and vaccination. An excellent opportunity to learn how SARS-CoV-2 adapts to changing environmental conditions and the basis of how it might have jumped to infect humans is given by the changes observed upon jumping to another host species.

SARS-CoV-2 has been found in many domestic and captive animals because of infections from humans or experimental infections (Abdel-Moneim and Abdelwhab, 2020). One particular case has received quite attention, the jump to minks which have resulted in serious outbreaks in mink farms and even transmissions from minks to humans (Oude Munnink et al., 2021). Outbreaks in mink farms have been detected in several countries, including Denmark, the Netherlands, Sweden, USA, and Spain (van Dorp et al., 2020c). Apart from the economic losses for the fur industry, the possibility of adaptation of SARS-CoV-2 to a new host with a widespread distribution is cause of much concern for the additional difficulties in controlling a pandemic with new natural reservoirs.

A detailed genomic analysis of outbreaks in several Dutch mink farms revealed that they were caused by different lineages (Oude Munnink et al., 2021) and showed a clear relationship between the sequences obtained from minks and those from the corresponding farm workers. In one case, transmission from minks to workers of the farm was observed. However, although some mutations were repeatedly observed among the sequences obtained in the different farms, none was consistently found in all of them. Van Dorp et al. (2020c) identified up to 23 recurrent mutations including three non-synonymous mutations in the RBD that appeared independently in at least four occasions (Fig. 1A). These observations might indicate that the virus has explored several ways to adapt to a new host (MacLean et al., 2021). This might have also been the case during the phase of adaptation to humans.

4.3. Recently emerged SARS-CoV-2 lineages

As mentioned above, since the initial phases of the COVID-19 epidemic, a major source of concern was the possibility that SARS-CoV-2 might mutate to acquire novel phenotypes and, possibly, increased virulence or transmissibility. Worries about the rapid spread of the 20E.EU1 lineage were soon tempered, but emphasized the relevance of genomic surveillance during the pandemic. This became even clearer at the beginning of December 2020, when routine epidemiological investigation for increasing incidence of COVID-19 in Kent (England), together with analysis of sequences obtained by the COVID-19 Genomics UK (COG-UK) consortium, revealed the presence of a large **monophyletic cluster** highly divergent from genomes sampled in the UK and worldwide. Almost half of the genomes sampled in Kent belonged to this new cluster (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/947048/Technical_Briefing_VOC_SH_NJL2_SH2.pdf). Inspection of sampling dates indicated that the earlier genomes in the cluster were collected in September. The long branch (Fig. 2) separating these sequences from the others facilitated the identification of the new lineage, which was initially designated VUI-202012/01 (where VUI stands for variant under investigation) by Public Health England and then renamed VOC-202012/01 (here VOC is variant of concern) on December 18th (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/947048/Technical_Briefing_VOC_SH_NJL2_SH2.pdf). VOC-202012/01 is also referred to as B.1.1.7 (PANGO nomenclature) and 20B/501Y.V1 (Nextstrain nomenclature). As the long branch in the phylogenetic tree indicates, B.1.1.7 is characterized by an unusually large number of nucleotide substitutions, many of which are nonsynonymous changes or deletions in the spike protein (Rambaut et al., 2020b) (see below and Table 3). The surge in cases associated with B.1.1.7, as well as its biological features, prompted the UK Government to enforce strict control measures. Epidemiological investigations during this time of high social distancing suggested that B.1.1.7 is more transmissible than pre-existing lineages with estimated ratios of **reproduction numbers** varying between 1.4 and 1.8 (Leung et al., 2021; Volz et al., 2020; Vöhringer et al., 2020; Davies et al., 2021a). Alarmingly, B.1.1.7 also seems to be associated with higher viral loads and increased disease severity (Borges et al., 2021a; Davies et al., 2021b; Kidd et al., 2021). By the end March 2021, the B.1.1.7 lineage had spread in 94 countries, although the coverage of genomic surveillance varies greatly across the globe (https://www.GISAID.org) (Fig. 3). Importantly, the proportion of B.1.1.7 increased significantly from 1 to 70% over the course of the recent SARS-CoV-2 outbreak in Portugal (Borges et al., 2021b). Data from the USA, where B.1.1.7 was introduced
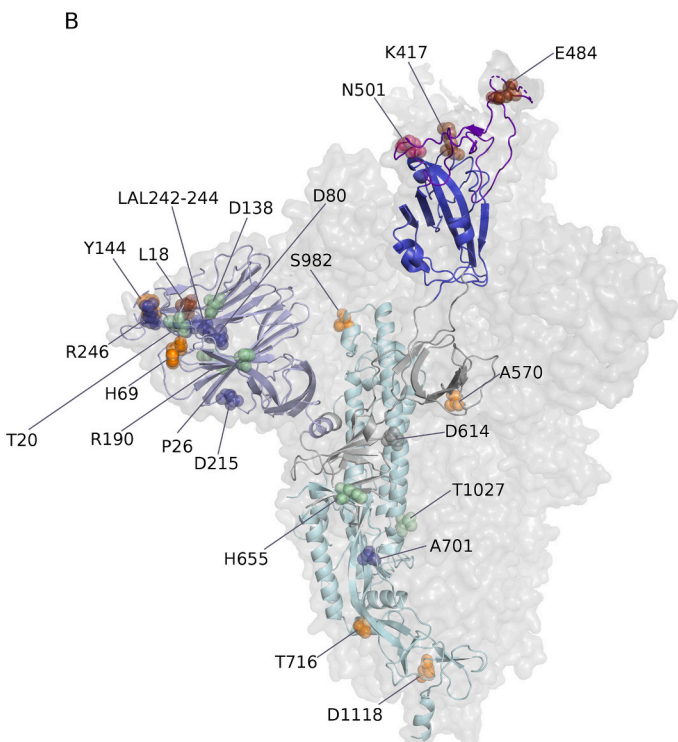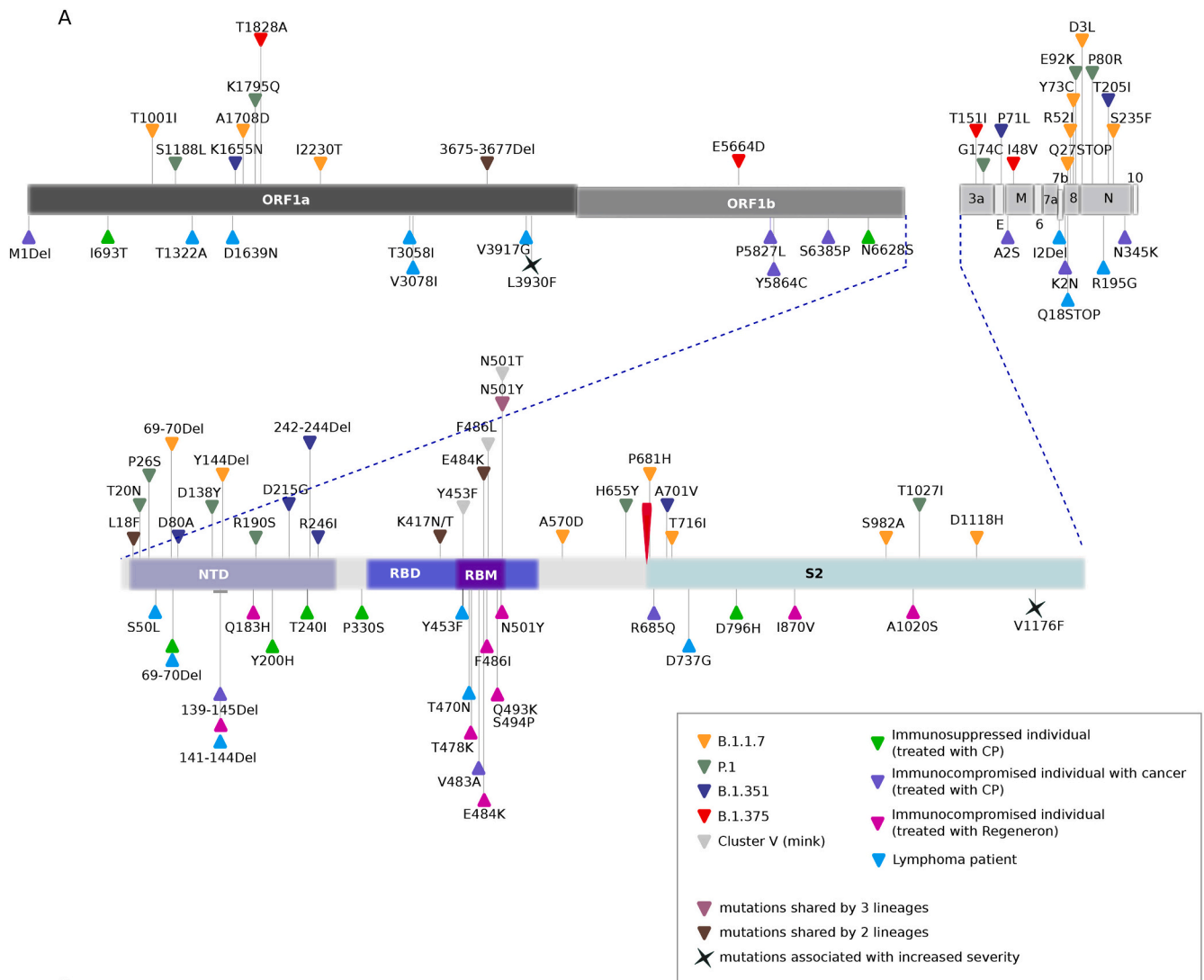
(caption on next page)

Fig. 1. Mutations observed in emerging SARS-CoV-2 lineages, in mink clusters, and in immunocompromised patients with long-term infection. (A) Schematic representation of the coding regions of the SARS-CoV-2 genome. The furin cleavage site in the S protein is represented with an elongated red triangle. Mutations are represented with colored triangles, as per legend. Variants possibly associated with increased mortality were derived from a previous work (Hahn et al., 2020). Subjects with long-term infection were described in the following works: immunosuppressed individual treated with CP (Kemp et al., 2020), Immunocompromised individual with cancer and treated with CP (Avanzato et al., 2020), Immunocompromised individual treated with Regeneron (Choi et al., 2020), lymphoma patient (Bazykin et al., 2021). (B) Mutations found in lineages B.1.1.7 (orange), B.1.351 (blue), and P.1 (green) are mapped onto the three dimensional structure of the spike protein. Mutations shared by two lineages are in chocolate, those shared by three lineages in dark red. The 3D structure corresponds to Swissmodel P0DTC2, which includes amino acids that are disordered in the spike crystallographic structures. For clarity, mutations are mapped on one monomer only.

in October-November, however indicated that this lineage has spread at an unremarkable pace in California, whereas its diffusion in Florida was definitely faster (Larsen and Worobey, 2021). The reasons why B.1.1.7 displays different epidemiological characteristics depending on the geographic area are presently unknown, but may relate to the specific control measures that are in place in distinct regions.

B.1.1.7 carries an in-frame deletion in the N-terminal domain (NTD) of the spike protein (HV69-70Del). An apparently unrelated lineage (B.1.375) with the same deletion was detected in multiple locations in the United States (Moreno et al., 2021; Larsen and Worobey, 2020, 2021). B.1.357 most likely originated in mid-September 2020 and, in addition to the deletion, carries fewer mutations compared to lineage B.1.1.7. (Fig. 1A, Table 3). At present, there is no indication that B.1.357 has peculiar characteristics in terms of disease severity or transmissibility (Larsen and Worobey, 2020; Moreno et al., 2021. Notably, the 69–70 deletion is found in yet another lineage, B.1.258, which has now been reported at considerable frequencies in several locations, especially in Europe (https://cov-lineages.org/lineages/lineage_B.1
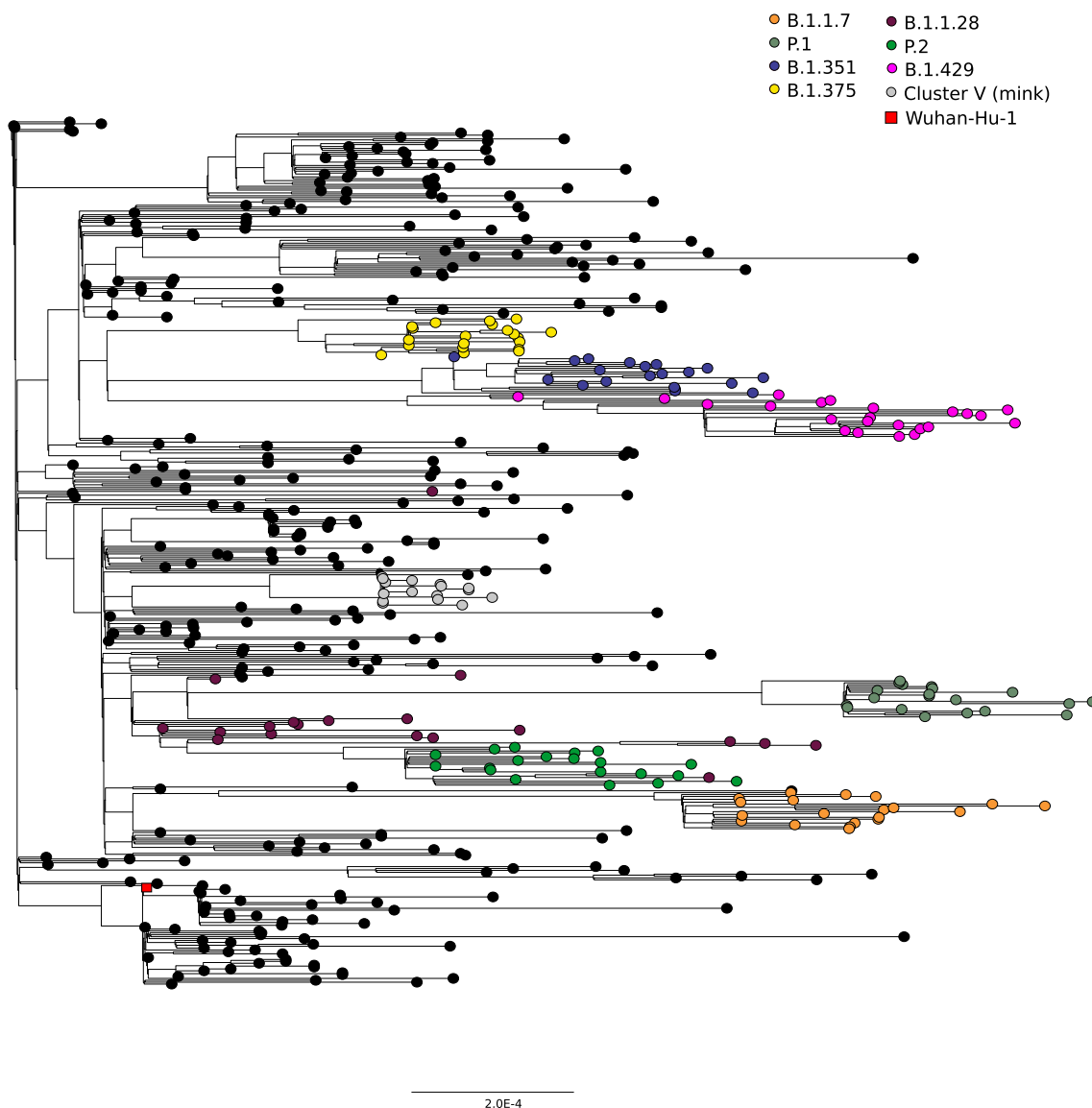
Fig. 2. Maximum likelihood phylogeny of 457 SARS-CoV-2 genomes. Twenty sequences belonging to lineages B.1.1.7, B.1.351, P.1, P.2, B.1.1.28, B.1.375, and B.1.429 were downloaded from GSAID. The complete genomes of 16 cluster V viruses were also obtained, together with 300 randomly selected SARS-CoV-2 genomes collected from January 2020 to March 2021 (Supplementary Acknowledgement Table). In particular, 20 sequences/month were included. Sequence alignments were generated using MAFFT (v7.427) (Katoh et al., 2019; Polack et al., 2020; Walsh et al., 2020) The phylogenetic tree was constructed using RAxMLversion 8.2.12 (Stamatakis, 2014) and visualized with FigTree (http://tree.bio.ed.ac.uk/).

Table 3

Mutations defining emerging SARS-CoV-2 lineages (B.1.1.7, B.1.351, P.1, and B.1.375). Gamage et al., 2020; Liu et al., 2021; Starr et al., 2021; Young et al., 2020.

Mutation	Protein product	Lineage	Functional effect	References
T1001I	ORF1a (nsp3)	B.1.1.7	Unknown	
S1188L	ORF1a (nsp3)	P.1	Unknown	
K1655N	ORF1a (nsp3)	B.1.351	Unknown	
A1708D	ORF1a (nsp3)	B.1.1.7	Unknown	
K1795Q	ORF1a (nsp3)	P.1	Unknown	
T1828A	ORF1a (nsp3)	B.1.375	Unknown	
I2230T	ORF1a (nsp3)	B.1.1.7	Unknown	
SGF3675-3677Del	ORF1a (nsp6)	B.1.1.7, P.1	Unknown	
E5664D	ORF1b (Hel)	B.1.375	Unknown	
L18F	S (NTD)	P.1, B.1.351	May increase resistance to CP; decreases binding of some mAbs *in vitro*	Wibmer et al., 2021; McCallum et al., 2021
T20N	S (NTD)	P.1	Adds a glycosylation sequon	
P26S	S (NTD)	P.1	Unknown	
HV69-70Del	S (NTD)	B.1.1.7, B.1.375	Recurrent mutation; emerged in a chronically infected patient treated with CP; two-fold higher infectivity *in vitro*; may decrease mAb binding	Kemp et al., 2020
D80A	S (NTD)	B.1.351	May increase resistance to CP; decreases binding of some mAbs *in vitro*	Wibmer et al., 2021; McCallum et al., 2021
D138Y	S (NTD)	P.1	Unknown	
Y144Del	S (NTD)	B.1.1.7	A larger deletion (142-144) detected in a chronically infected patient treated with Regeneron; two larger deletions encompassing position 140 observed in an immunocompromised patient treated with CP; decreases binding of some mAbs *in vitro*	Avanzato et al., 2020; Choi et al., 2020; McCallum et al., 2021
R190S	S (NTD)	P.1	Unknown	
D215G	S (NTD)	B.1.351	May increase resistance to CP	Wibmer et al., 2021
LAL242-242Del	S (NTD)	B.1.351	May increase resistance to CP	Wibmer et al., 2021
R246I	S (NTD)	B.1.351	May increase resistance to CP; decreases binding of some mAbs *in vitro*	Wibmer et al., 2021; McCallum et al., 2021
K417N/T	S (RBD)	B.1.351, P.1	Escape from mAbs *in vitro*; less susceptible to neutralization by vaccinee plasma[a]	Starr et al., 2021; Wang et al., 2021;
E484K	S (RBM)	B.1.351, P.1	Detected in a chronically infected patient treated with Regeneron; emerged *in vitro* when the virus was cultured with CP; less susceptible to neutralization by vaccinee plasma; escape from mAbs *in vitro*; increased ACE2 binding affinity	Wang et al., 2021; Choi et al., 2020; Weisblum et al., 2020; Z. Liu et al., 2021; Greaney et al., 2020
N501Y	S (RBM)	B.1.1.7, B.1.351, P.1	Increases ACE2 binding affinity; increased infectivity and virulence in mouse models; increased transmission (based on epidemiological data); detected in a chronically infected patient treated with Regeneron; unsure whether it is less susceptible to neutralization by vaccinee plasma	Wang et al., 2021; Choi et al., 2020; Starr et al., 2021; Leung et al., 2021; Gu et al. 2020, Xie et al. 2021; Hoffmann et al., 2021
A570D	S (S1)	B.1.1.7	Unknown	
H655Y	S (S1)	P.1	Unknown	
P681H	S (S1/S2 boundary)	B.1.1.7	Within the furin cleavage site	
A701V	S (S2)	B.1.351	Unknown	
T716I	S (S2)	B.1.1.7	Unknown	
S982A	S (S2)	B.1.1.7	Unknown	
T1027I	S (S2)	P.1	Unknown	
D1118H	S (S2)	B.1.1.7	Unknown	
G174C	ORF3a	P.1	Unknown	
T151I	ORF3a	B.1.375	Unknown	
P71L	E	B.1.351	Unknown	
I48V	M	B.1.375	Unknown	
Q27STOP	ORF8	B.1.1.7	Deletion of ORF8 seems to result in a milder phenotype and modest effect on virus replication	Gamage et al., 2020; Young et al., 20202
R52I	ORF8	B.1.1.7		
Y73C	ORF8	B.1.1.7		
E92K	ORF8	P.1		
D3L	N	B.1.1.7	Unknown	
P80R	N	P.1	Unknown	
T205I	N	B.1.351	Unknown	
S235F	N	B.1.1.7	Unknown	

Notes: CP, convalescent plasma; NTD, N-terminal domain; RBD, receptor binding domain; RBM, receptor binding motif; mAb, monoclonal antibody. Mutations shared by more than one lineage are shaded in gray. [a] Subjects were immunized with the Moderna or Pfizer-BioNTech vaccines.
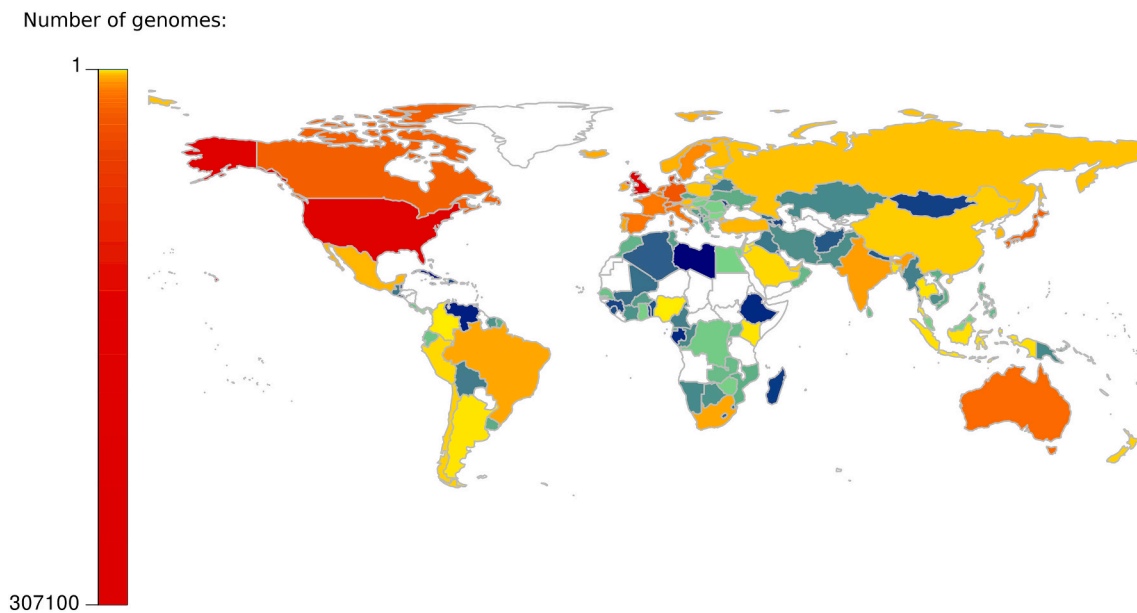
Number of genomes:

Fig. 3. Number of complete SARS-CoV-2 genomes sequenced in different countries. Data on 836,000 high-quality, complete genomes were retrieved from GISAID (https://www.gisaid.org/). Countries are colored according to the number of sequenced genomes, as per legend.

.258.html). Lineage B.1.258 is characterized by an additional mutation (N439K) in the RBD of the S protein. Although this mutation increases ACE2 binding affinity, B.1.258 does not seem to spread faster than other lineages. However, the N439K has been recently reported to confer resistance against several neutralizing monoclonal antibodies and to reduce virus sensitivity to convalescent plasma (CP) (Thomson et al., 2021).

Unfortunately, these are not the only lineages to generate concern in this phase of the pandemic, as the Network for Genomic Surveillance in South Africa reported the emergence of a distinctive SARS-CoV-2 lineage (lineage 501Y.V2 or B.1.351) in October 2020. Just like B.1.1.7, this lineage is defined by a substantial number of mutations, only a minority of which are shared with the B.1.1.7 lineage (Fig. 1, Table 3). Epidemiological evidence suggests that B.1.351, which emerged in early August in Nelson Mandela Bay, has been displacing other lineages in several provinces in South Africa (Tegally et al., 2021). B.1.351 has been isolated from people infected with SARS-CoV-2 from more than 50 different countries across the globe (https://www.GISAID.org). Very preliminary data from the Centre of Mathematical Modelling of Infectious Diseases (CMMID COVID-19 working group, London School of Hygiene and Tropical Medicine) indicated that B.1.351 is possibly more transmissible or less susceptible to cross-protection from previous exposure (or both) (https://cmmid.github.io/topics/covid19/sa-novel-variant.html).

Even more recently - i.e., in December 2020- a resurgence of COVID-19 cases in Manaus, Brazil, prompted a genome sequencing program that led to the identification of another novel lineage (P.1) (Sabino et al., 2021; Faria et al., 2021a). Specifically, Manaus experienced a high attack rate, which was estimated to be approximately 75% by October 2020 (Faria et al., 2021a). The P.1 lineage was absent in samples collected until November 2020, but its prevalence raised to 41% during December 2020. P.1, a descendant of lineage B.1.1.28, is phylogenetically distinct from pre-existing strains circulating in Brazil and elsewhere (Fig. 2). In analogy to the two novel lineages described above, it carries a number of mutations, particularly in the spike protein. Some of these are shared with B.1.1.7 and B.1.351 (Fig. 1, Table 3) (Faria et al., 2021a). The progenitor of P.1, lineage B.1.1.28, has been circulating in Brazil since the early pandemic phase (February–March 2020) (Resende et al., 2020). In addition to the P.1 lineage, B.1.1.28 also originated another independent sub-lineage (P.2) (Fig. 2), which shares one

mutation (E484K in the RBD of the S protein) with P.1 (Naveca et al., 2021). Both P.1 and P.2 have been associated with cases of re-infections in Brazil (Naveca et al., 2021; Vasques Nonaka et al., 2021; Resende et al., 2021). Notably, a recent update from Public Health England reported the detection of the E484K mutation in a small subset of sequences belonging to the B.1.1.7 lineage, suggesting multiple independent acquisitions of this change (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/957504/Variant_of_Concern_VOC_202012_01_Technical_Briefing_5_England.pdf).

Finally, on January 19th, 2021, the California Department of Public Health dispatched a note about a SARS-CoV-2 variant carrying the L452R mutation. This lineage, now referred to as B.1.429, caused multiple large COVID-19 outbreaks in Santa Clara County and other regions (https://www.cdph.ca.gov/Programs/OPA/Pages/NR21-020.aspx) and is now detected as majority lineage in California and Nevada (https://outbreak.info). B.1.429 is defined by four non-synonymous substitutions, three of which in the spike protein (S13I, W152C, L452R). Among these, the L452R mutation is located within the RBD. Epidemiological and in vitro analyses have suggested that B.1.429 has increased infectivity and transmissibility, as well as the ability to escape neutralization by CP and vaccine-induced antibodies (Deng et al., 2021; Li et al., 2020b). At the end of January 2021, a descendant lineage of B.1.429 carrying additional variants including Q677H (S protein) was detected in Colorado (Tomkins-Tinch et al., 2021). Substitutions at position 677 are notable because they seem to have arisen independently in multiple lineages, either as Q677H or as Q677P. The proximity of this position to the polybasic cleavage site might be consistent with a functional relevance for the proteolytic processing of the spike protein (Hodcroft et al., 2021).

4.4. Genomic features and possible origin of emerging SARS-CoV-2 lineages

Clearly, one of the most notable features of the emerging lineages, especially B.1.1.7, B.1.351, and P.1, is the large number of nucleotide substitutions they carry (Table 3, Fig. 1). This is readily visualized by the long phylogenetic branches separating these clusters from the other SARS-CoV-2 variants (Fig. 2). As mentioned above, the estimated substitution rate for SARS-CoV-2 is around 10^{-3} substitutions per site per

year (Duchene et al., 2020; van Dorp et al., 2020a; Ghafari et al., 2020). Thus, circulating viruses accumulate on average ~2 substitutions per month. The large number of changes on the emerging lineages is thus highly unexpected and represents a new twist in the evolutionary trajectory of SARS-CoV-2. Moreover, the nature of these changes and their location are strongly suggestive that some form of selective pressure underlies the origin of the novel lineages. In fact, evolutionary analyses indicated that the emergence of the three lineages carrying the N501Y substitution (B.1.1.7, B.1.351 and P.1) was accompanied by a shift in the strength of natural selection and all of them carry a number of sites that show evidence of ongoing adaptation (Martin et al., 2021). Most substitutions in these lineages are either missense, nonsense or indels (insertions/deletions) that alter protein sequences (Table 3), and the majority of them are located in the spike protein, which accounts for only ~13% of the coding capacity of SARS-CoV-2 (Fig. 1). Importantly, some of the spike protein changes have been associated with increased infectivity (e.g., N501Y, HV69-70Del), escape from immune responses (e.g., E484K, K417N/T, HV69-70Del), or spillover to mink farms (N501T) (Table 3). Several mutations in the S protein are also predicted to affect conformational epitopes (Fig. 4). It is also clear that substitutions at positions N501, K417, and E484 arose independently on multiple lineages, suggesting **convergent evolution** or recurrent mutation (Faria et al., 2021a). Indeed, recent observations indicate that a substantial fraction of mutations that define the emerging lineages occur in protein regions of remarkable evolutionary plasticity in sarbecovirus genomes (Garry et al., 2021). In fact, a spike protein alignment of representative sarbecoviruses detected regions that, along the evolutionary history of this virus subgenus, have accumulated micro-deletions and insertions. Such regions include the polybasic furin cleavage site, where P681H is located, but also exposed loop regions, which are most likely tolerant to change and subject to immune selection (e.g., the one where K484 lies). Overall, these observations were taken to imply that the selective pressures that have shaped the evolution of spike proteins in sarbecoviruses are driving the ongoing evolution of SARS-CoV-2 in humans, or at least contributed to the emergence of the highly divergent lineages (Garry et al., 2021). This hypothesis is also in line with preliminary analyses of B.1.351 sequences indicating that both the entire S gene and several mutated sites show evidence of **positive selection** (Kosakovsky Pond et al., 2020; Tegally et al., 2020).

Whereas all these lines of evidence suggest an underlying selective pressure, the question remains as to which factors or circumstances prompted SARS-CoV-2 to accrue mutations and originated the emerging lineages. At present, the most widely accepted hypothesis involves patients with chronic or long-standing SARS-CoV-2 infection. This is because highly divergent SARS-CoV-2 genomes that carry multiple mutations have been sequenced from immunocompromised subjects with long-term COVID-19, either treated or not with CP or therapeutic antibodies (Bazykin et al., 2021; Avanzato et al., 2020; Choi et al., 2020;
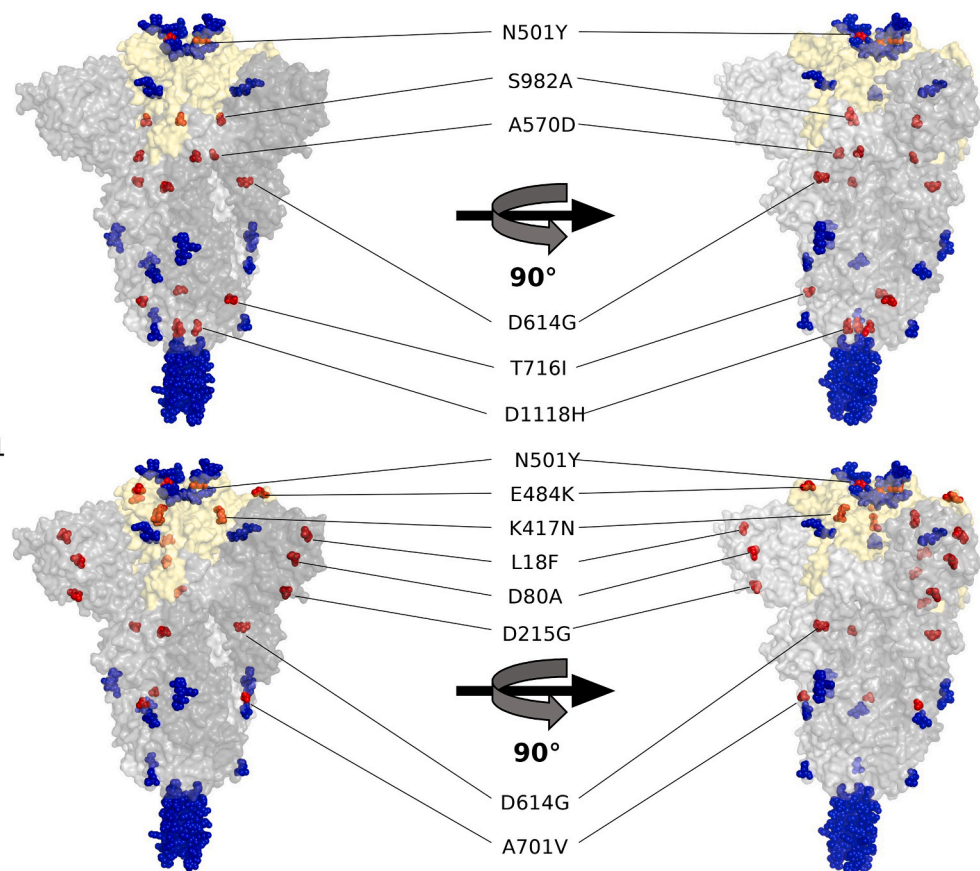
Fig. 4. Relationships between amino acid substitutions and conformational epitopes in the SARS-CoV-2 spike proteins. Conformational epitope may induce the neutralizing antibody against various viruses (Aso et al., 2019). Therefore, to estimate the vaccine's efficacy, the relationships between conformational epitopes and amino acid substitutions of the S protein in the B.1.1.7 and B.1.351 lineages were examined. Detailed procedures of these examinations were made as previously described (Aso et al., 2019). The changes detected in the two lineages (red) affected partially overlapping epitopes (blue). Particularly, amino acid substitutions (E484K and N501Y) in conformational epitopes were found in the RBD of the S protein (yellow). Reports suggested that amino acid substitutions of the conformational epitopes lead to viral reinfection and changes of vaccine efficacy (Russi et al., 2018). Thus, our *in silico* study suggests that the mRNA vaccine efficacy partially changes against the variants (Russi et al., 2018; Sahin et al., 2020), although this should also be examined *in vitro* and *in vivo*.

Kemp et al., 2020). Moreover, some of the mutations observed in these samples correspond to those identified in the B.1.1.7, B.1.351, and P.1 lineages (Table 3, Fig. 1). As of March 2021, four cases of long-term infections (up to four months) have been described in immunocompromised COVID-19 cases, two with haematological malignancies and two receiving immunosuppressant therapies (Bazykin et al., 2021; Avanzato et al., 2020; Choi et al., 2020; Kemp et al., 2020). In all patients, highly divergent viral lineages emerged, irrespective of symptom severity and disease outcome (two deceased and two recovered) (Fig. 1). Three of these patients were treated at least once with either CP or antibody cocktails (Regeneron). Longitudinal sequencing revealed a progressive accumulation of mutations, as well as the appearance of dynamic viral populations, indicative of intra-host evolution of SARS-CoV-2 (Avanzato et al., 2020; Choi et al., 2020; Kemp et al., 2020). In particular, Kemp and coworkers noted that a viral population carrying the VH69-70Del and D769H mutations emerged after the first CP administration, fell to low frequency in a few days, and then increased again after a second CP treatment. This observation suggests competition among intra-host viral populations and supports the idea that CP treatment exerts a selective pressure that favours specific mutants. Indeed, using *in vitro* assays they found that the VH69-70Del+D769H variant was less sensitive to neutralization by CP (Kemp et al., 2020).

Overall, these data raise the possibility that, in immunocompromised hosts, long-term infection and reduced immune control allow intra-host virus evolution, and that treatment with CP or therapeutic antibodies selects for specific mutations. It is however worth mentioning that a recent report on a patient with lymphoma and long-term COVID-19 also described the emergence of a highly divergent lineage carrying, among others, the VH69-70Del mutation. The patient did not receive CP or antibodies and did not develop neutralizing responses against SARS-CoV-2 (Bazykin et al., 2021). It is thus possible that long-term viral replication in the context of an immunocompromised host, irrespective of the treatment regime, creates the conditions to generate high viral diversity and novel lineages. This has previously been suggested for other viruses, including norovirus and influenza A virus (McMinn et al., 1999; Karst and Baric, 2015; Memoli et al., 2010; Rogers et al., 2015), although in the latter case the interpretation is complicated by the administration of antivirals to chronically infected patients. In this respect, it is also interesting to notice that long-term infections with SARS-CoV-2 (more than 2 months) were previously shown to occur in immunocompetent hosts and, when genome sequencing was performed, limited viral genetic diversity was observed (Li et al., 2020c; Abu-Raddad et al., 2020).

More abundant data on short-term infections indicate that, within the host, SARS-CoV-2 accumulates mutations at a pace consistent with its estimated substitution rate (i.e., ~1 mutation per genome every two weeks) (Tonkin-Hill et al., 2020). Thus, diversity remains low and viral evolution is mainly shaped by purifying selection (Abu-Raddad et al., 2020; Valesano et al., 2021; Popa et al., 2020; Tonkin-Hill et al., 2020; Lythgoe et al., 2021). Nonetheless, variability in the number of detectable viral variants was observed among patients with short-term infections (Popa et al., 2020), and still limited data indicate that SARS-CoV-2 intra-host diversity might increase with age (Al Khatib et al., 2020) and in cancer patients (Al Khatib et al., 2020; Siqueira et al., 2020). Intense monitoring of viruses transmitted by these (and other) patient categories will thus be required to assess whether they can transmit genetically diverse viral genomes.

In summary, given the limited evidence available to date, the most likely explanation for the emergence of the new, highly divergent lineages is that their evolution was accelerated by some specific circumstances (possibly infection of an immunocompromised host) and onward transmission introduced them back in the human population. This is also in line with an analysis of B.1.1.7 genomes (collected up to November 30th, 2020), which indicated that, since its detection in the UK, this lineage has been evolving with a substitution rate similar to that of other SARS-CoV-2 lineages (Rambaut et al., 2020b).

4.5. Recombination in SARS-CoV-2

New evidence points at the occurrence of recombination in SARS-CoV-2 (Jackson et al., 2021; Latinne et al., 2020; VanInsberghe et al., 2021). This is not a novelty in coronaviruses (Gribble et al., 2021) and previous analyses have shown that recombination has played an important role in the evolution of SARS-CoV-2 from its ancestors (Andersen et al., 2020; Boni et al., 2020; Kirtipal et al., 2020; MacLean et al., 2021; Wells et al., 2021). Some recombinant sequences detected in the UK involve a breakpoint near the 5' end of the spike gene from B.1.1.7 variants (Jackson et al., 2021) but there has been no indication of changes in their phenotypic properties nor an increase in their frequency above the threshold for gaining the assignment of a new PANGO lineage. Nevertheless, in light of the possible epistatic effects of mutations in different portions of the SARS-CoV-2 genome (McCallum et al., 2021a), it will certainly be necessary to closely watch the emergence and possible spread of recombinant genomes of this virus.

5. Possible impact of SARS-CoV-2 variants on the performance of molecular testing methods

Since COVID-19 has reached pandemic status causing a serious global health threat, a widespread availability of diagnostic testing is crucial to detect SARS-CoV-2 in a variety of specimen types collected from both symptomatic and asymptomatic patients. Many academic and commercial clinical microbiology laboratories and companies have worked around the clock to develop molecular tests that are fast, highly accurate, and inexpensive to meet testing demands. In addition, more accessible and scalable testing is a critical component in managing the COVID-19 pandemic. Molecular tests detect genetic material and they are sensitive enough to be able to pick up very small amounts of viral RNA very early in an infection. Molecular tests are considered the gold standard diagnostic test for SARS-CoV-2 detection. Molecular tests use two major techniques known as reverse transcription polymerase chain reaction (RT-PCR) and isothermal amplification. Another emerging molecular technology is the clustered regularly interspaced short palindromic repeats (CRISPR). The Sherlock™ CRISPR SARS-CoV-2 kit is the first CRISPR-based diagnostic test receiving an EUA by the FDA. This assay is intended for the qualitative detection of the SARS-CoV-2 virus in upper respiratory tract specimens from patients suspected of having COVID-19.

The COVID-19 molecular tests have been designed and developed based on genomic information of SARS-CoV-2. Although molecular tests can provide rapid and accurate diagnosis on the infection, there is a considerable risk of misdiagnosis due to genomic variations, which may have a critical impact on the test performance. Molecular tests often use different primer/probe sets targeting different regions of the SARS-CoV-2 genome. To safeguard against potential mutational drift in the SARS-CoV-2 genome, many molecular tests were developed to amplify and detect at least two conserved regions. Molecular tests designed to detect multiple SARS-CoV-2 genetic targets are less susceptible to the effects of genetic variation than tests designed to detect a single genetic target. The Xpert Xpress SARS-CoV-2/Flu/RSV test targets the N2 and E genes of SARS-CoV-2. The BioFire COVID-19 test consists of three independent and non-overlapping assays targeting the ORF1ab and ORF8 sequences. In the Simplexa COVID-19 Direct assay, two different regions of the SARS-CoV-2 genome, ORF1ab and S gene, are amplified. The Panther Fusion SARS-CoV-2 test detects two conserved regions of the ORF1ab gene, the two regions are not differentiated and amplification of either or both regions leads to a fluorescence signal. A recent study performed the *in-silico* reassessment of the previously published primers and probes for COVID-19 diagnosis using a total of 17,026 SARS-CoV-2 sequences. Mutations or mismatches on primer/probe binding regions were discovered in seven out of 27 molecular assays. While the US-CDC-N-1 probe in the US CDC 2019-nCoV Real-Time RT-PCR Diagnostic Panel showed only one mismatch with 1.6% viral sequences, the CN-CDC-N

forward primer (developed by China CDC, China) had three mismatches with 18.8% of viral sequences. The reverse primer of NIID-JP-N (developed by National Institute of Infectious Diseases, Japan) also showed one mismatch with all the sequences (Khan and Cheung, 2020).

Since single nucleotide polymorphisms (SNPs) of the SARS-CoV-2 genome are now a regular occurrence, with more discovered every day, it is unrealistic to avoid all SNPs on different primer/probe binding sites. However, many molecular tests can tolerate a few single nucleotide mismatches, which can have little to no impact at all on their performance. Lefever at el. discovered that single mismatches located >5-bp from the 3' end have a moderate effect on the target amplification and can be tolerated. In addition, four mismatches in a single primer block amplification almost completely, whereas three mismatches in one of the primers must be combined with at least two mismatches in the other primer to achieve the loss of target hybridization (Lefever et al., 2013). Recently, Ziegler and co-authors reported a case tested with the Xpert Xpress SARS-CoV-2 assay with a cycle threshold (CT) value of 22.7 for the E gene, but a negative result for the N2 gene. However, other platforms including the Allplex SARS-CoV-2 assay, the Charité protocol (Charité - Universitätsmedizin Berlin Institute of Virology, Germany) and the US CDC 2019-nCoV Real-Time RT-PCR Diagnostic Panel revealed positive results for all designed targets, including the N gene with CT values of 26.2-27. Sanger sequencing of the two independent PCR amplicons revealed three SNPs compared with the SARS-CoV-2 strain Wuhan-Hu-1 reference genome (Ziegler et al., 2020).

Given the high frequency of SNPs occurrence, the FDA alerts clinical laboratories that false negative results may occur with any molecular test for the detection of SARS-CoV-2 if a mutation occurs in the part of the virus' genome assessed by that test. Since primer/probe sequences of most commercial assays are not revealed, the FDA monitors closely the potential negative impact of genetic variation in molecular tests that have received Emergency Use Authorization (EUA). Based on the FDA's analysis to date, the Accula SARS-CoV-2 test performance may be impacted when a SARS-CoV-2 strain having a genetic variant at position 28,881 (GGG to AAC) is tested. Other two molecular tests, the TaqPath COVID-19 Combo Kit (which may also be labelled as the TaqPath COVID-19 Combo Kit Advanced) and the Linea COVID-19 Assay Kit, have significantly reduced sensitivity due to certain mutations, including one of the mutations in the recently identified VOC 202112/01 variant (lineage B.1.1.7 or 20I/501Y.V1). Since this test is designed to detect multiple genetic targets, the overall test sensitivity should not be impacted. However, the pattern of detection when certain mutations are present may help with early identification of new variants in patients to reduce further spread of infection. Be aware of the pattern of detection associated with certain mutations, including the B.1.1.7 variant, specifically a pattern of 2/3 positive targets showing the S-gene drop out (reduced sensitivity with the S-gene target, also denoted as SGTF, S gene target failure), when using the TaqPath COVID-19 Combo Kit, and a pattern of 1/2 positive targets showing the S-gene drop out when using the Linea COVID-19 Assay Kit. Recently, the FDA reports that the Cepheid tests are impacted by a single point mutation in the target area of the test. Two independent single point mutations reduce the test's sensitivity for detecting the N2 target. This observation is unexpected, and the FDA's analysis suggests that the impact of a single point mutation on the test performance is associated with the unique chemistry of the Cepheid tests. The E target is still detected when enough virus is present, leading to a "presumptive positive" result in the Xpert Xpress SARS-CoV-2 and Xpert Xpress SARS-CoV-2 DoD tests. Detection of the E target without detecting the N2 target will be reported as "positive" in the Xpert Omni SARS-CoV-2. The FDA also recommends considering repeat testing with a different test (with different genetic targets) if COVID-19 is still suspected after receiving a negative test result. (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; https://www.fda.gov/medical-devices/coronavirus-covid-19-and-medical-devices/sars-cov-2-viral-muta

tions-impact-covid-19-tests).

Viruses often mutate, and SARS-CoV-2 is no exception. As a consequence, there is the urgent need for continued surveillance of viral evolution and for fast, accurate and sensitive detection methods. Viral metagenomics has emerged to be a powerful method to detect SARS-CoV-2 mutants. The Illumina COVIDSeq test (Illumina, Inc.) is the first next-generation sequencing (NGS) test approved for use under the EUA. This amplicon-based NGS test can amplify up to 98 targets on SARS-CoV-2 genome for highly accurate detection. Other NGS *in vitro* diagnostic tests under EUA include Clear Dx SARS-CoV-2 test on the Oxford Nanopore GridION Sequencer (Clear Labs, Inc.), Guardant-19 on the Illumina NextSeq 500 & NextSeq 550 Sequencing Systems (Guardant Health, Inc.), and Helix COVID-19 NGS test on the Illumina NovaSeq 6000 Sequencing System (Helix OpCo LLC) https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/vitro-diagnostics-euas#individual-molecular). These high-throughput assays should be considered when clinical laboratories want to further characterize the clinical specimen with genetic sequencing when the above-mentioned pattern of detection associated with certain mutations is identified.

6. Implications of new variants for immunity and vaccination

6.1. Characterising the immune response to SARS-CoV-2 and the determinants of immunopathology

To understand the likely impact of new variants and facilitate vaccine design, we first need to understand the natural immune response and the way the immune responses differ between asymptomatic, mildly and severely infected individuals. Typically for viral infections, IgG and IgM antibodies are seen after symptom onset, and neutralising antibodies are correlated with resolution of disease (Long et al., 2020). Asymptomatic individuals show long periods of viral shedding and have low levels of antibody-mediated immunity (Long et al., 2020). The magnitude of the neutralising antibody response is positively correlated with disease severity, as is the slow decline of antibody levels (Seow et al., 2020). The major target of these neutralising antibodies is the S protein, both the S1 (containing the RBD) and S2 domains (Jeyanathan et al., 2020). The roles of T cells are key, including for memory responses, with CD4+ T helper cells stimulating the production of antibodies and CD8+ cytotoxic T cells (Grifoni et al., 2020; Sharma et al., 2020). Cytotoxic CD4+ and CD8+ T cells are also known to be associated with disease outcome (Sette and Crotty, 2021). CD4+ T cells are generated in high amounts in COVID-19: mild disease and the acute phase of infection have been associated with high SARS-CoV-2 specific CD4+ T cells, while these have been absent where the outcome has been fatal (Rydyznski Moderbacher et al., 2020).

Early expression of Type 1 interferons will induce an anti-viral state and promote a pro-inflammatory response through inflammatory cytokines and chemokines. The SARS-CoV-2 virus is associated with delayed Type 1 IFN production, the innate immune response is suppressed, and this delayed response causes an inability to control viral replication, leading to immunopathology, cellular damage of airway epithelia and the lung parenchyma, sometimes resulting in Acute Respiratory Distress Syndrome (ARDS) and an eventual lethal inflammatory cytokine storm (Jeyanathan et al., 2020; Varghese et al., 2020).

COVID-19 tends to be most severe in elderly patients, those with underlying health complications and South Asian and Black people have a higher chance of COVID-19 related death than white people, only partly accounted for by comorbidities (Williamson et al., 2020). With age, both innate and adaptive immunity undergo cellular and functional changes, known as 'immunosenescence', characterised by higher baseline inflammatory responses, so that even healthy elderly have a continual low-grade inflammation known as 'inflammaging'. They have low numbers of naïve T cells, and although they have good memory T cell populations, the diversity of the T cell repertoire is reduced. The

efficient adaptive immune response seen in the young, associated with a good prognosis to COVID-19, is not seen in the aging population, where the poor adaptive response fails to control viral replication (Rydyznski Moderbacher et al., 2020). Heightened innate immunity results in release of cytokines and an inflammatory storm, which may cause tissue damage, ARDS, and ultimately prove lethal (Cunha et al., 2020).

6.2. Approaches to vaccination: efficacy, protection and severity

Hundreds of vaccines are under development, and a thorough review of these is beyond the scope of this article. For most all vaccines under development, the major protein target is the S-protein with its direct involvement in infection, and as a major target of the immune response. The development of vaccines against COVID-19 has seen unprecedented use of a large number of platforms for design and development, which are already well reviewed, together with antigen selection, route of delivery and regimens (Jeyanathan et al., 2020; Grifoni et al., 2020; Sharma et al., 2020; Flanagan et al., 2020).

Examples of well progressed vaccines, that have been through clinical trials and are now being delivered to the general population, include the Pfizer/BioNTech, Moderna, and Oxford/AstraZeneca vaccines (Jeyanathan et al., 2020). It is from these first vaccinations that we are beginning to collect empirical data to understand immunity and vaccine efficacy, including with respect to new variants. The Pfizer/BioNTech and Moderna vaccines are both based on mRNA synthesis. These are lipid nanoparticle mRNA vaccines, coding for the RBD of the S-protein, showing very high efficacy and almost identical results in phase III trials. Reports of trials recorded high titres of neutralising antibodies, CD4+ and CD8+ T cell responses (Laczkó et al., 2020; Mulligan et al., 2020; Sahin et al., 2020.). Notably both these vaccines appeared to confer similar protection in both young and old age groups.

The Oxford/AstraZeneca vaccine is a non-replicating recombinant viral vector (ChAd) expressing the S protein. Phase I – III clinical trials have taken place in the UK, South Africa, the USA and Brazil with an overall efficacy of 71% and induction of neutralising antibodies and T cell responses (Folegatti et al., 2020; van Doremalen et al., 2020; Ziegler et al., 2020). It is notable that whichever vaccine is employed, and the reported vaccine efficacy, no severe cases have been found amongst vaccinated people to date. Most recently, The AstraZeneca US phase III trial (astrazeneca.com, 22nd March 2021) with 32,499 participants and 141 cases of symptomatic COVID-19 reported 79% vaccine efficacy at preventing symptomatic COVID-19, 100% efficacy against severe or critical disease and hospitalisation and comparable efficacy across ethnicity and age.

6.3. Predicting the immune response and new SARS-CoV-2 variants

As of March 2021, there is a paucity of empirical evidence in the peer-reviewed literature concerning the relevance of new variants on the ability to mount an effective immune response, to correlate with immunopathology or to reduce the protective immunity conferred by vaccination. However, the lack of information is rapidly changing, with preprints appearing relevant to the variants of concern (lineages B.1.1.7, B.1.351, and P.1) in 2021. It is envisaged that during the next few months, data-driven analysis will be performed to review current vaccines and potential modifications to account for loss of efficacy due to new escape variants.

The vast literature that exists on the analysis of viral variants, arising during the pandemic, is covered elsewhere in this article. Based on the very large databases of viral sequences, along with time of origin and details of sampling, predictive tools have been proposed to assess the likely effects of new variants: not only which variants will be selected for and spread through their increased transmissibility and those which result in increased virulence, but also for understanding effects relevant to immunity and vaccination (Tomaszewski et al., 2020). Interest in predicting the likely development and impact of new variants and

combinations of variants has often, but not exclusively, focussed on the S protein and the variable RBD. For instance, bioinformatics and machine learning methods have been applied to predict epitope targets for CD4+ and CD8+ T cell response (Sette and Crotty, 2021) and also to select for T and B cell epitopes in vaccine design (Kalita et al., 2020; Kiyotani et al., 2020). These predictions have been helpful to identify experimentally the T cell targets and hence to investigate the effects of virus mutations on the immune response. Such predictions, on both T-cell and B-cell epitopes, have been used together with the information pertaining to the likelihood of viral variants undergoing **antigenic drift** (Koyama et al., 2020).

A major hurdle to develop vaccines that specifically induce T cell responses is the design of a heterogeneous set of epitopes that can be used to cover the vast heterogeneity of MHC genotypes in humans. In the context of other viral infections, such as DENV, Influenza and HIV, several strategies have been developed to achieve an optimal set of epitopes, and these approaches are likely to be underway in COVID-19 vaccine development as well. The use of animal models in our understanding of new SARS-CoV-2 variants and their relevance to vaccine development can be difficult and they are certainly not essential (Flanagan et al., 2020). Interestingly, and potentially benefiting protection through cross reaction, some epitopes from S and N proteins map identically in SARS-CoV-1 and SARS-CoV-2 (Ahmed et al., 2020). Identification of SARS-CoV-2 specific T cell epitopes, including those encompassing the new variants, as well as epitopes recognised by T cells potentially cross reactive with related viruses will be key for understanding immunity and immunopathological sequelae, which will impact on vaccine design (Flanagan et al., 2020; Ahmed et al., 2020; Olvera et al., 2020).

6.4. Empirical data: immunity, immunopathology, vaccine efficacy and new variants?

Considering the widespread variants, SARS-CoV-2 D614G became dominant globally in early 2020, but despite greater infectivity and transmission, it is not thought to result in greater severity of disease (WHO.int, 2020). This suggests no marked change in overall immunity to this new variant.

Early information on SARS-CoV-2 lineage B.1.1.7, originated by VOC 202012/01, suggested that this variant shows increased transmissibility but again no greater severity of disease, as measured by hospitalisation and 28-day fatality, in a study matching 1769 variant cases with 1769 wild-type control cases. Similarly, reinfection rates did not appear to differ between the 2 groups with 2 reinfections in the variant case group compared to 3 in the wild-type case group (Public Health England, Investigation of novel SARS-CoV-2 variant, Variant of Concern 202012/01 Technical briefing 2- 28 December 2020. PHE: London; 2020). Conversely, the most recent data for the B.1.1.7 lineage suggests that it is associated with increased disease severity. In a large study of COVID-19 cases in England, Davies et al. (2021b) considering 4,945 deaths, used S gene target failure as a proxy for infection with B.1.1.7. After controlling for confounding variables, they found an increased mortality rate associated with B.1.1.7: for men in the age range of 55 to 69, the risk of death from COVID-19 increases from 0.6% to 0.9%, and overall there is a 61% higher hazard of death (Davies et al., 2021b). This VOC has also been associated with increased viral loads in respiratory samples, as well as with longer duration of infection and of viral shedding (Calistri et al., 2021; Kidd et al., 2021; Kissler et al., 2021; Borges et al., 2021a), these latter findings helping to explain its increased transmissibility.

Since the substitution N501Y in the RBD of the S-protein is common to the three rapidly spreading B.1.1.7, B.1.351 and P.1 lineages, and is a substitution shown to infect mice more efficiently (Gu et al., 2020) (Table 3), efforts are underway to test the substitution specific neutralisation activity of sera from recovered patients and vaccinated people to assess the likely impact on vaccine performance (who.int, 2020). BNT162b2 (Pfizer Inc. and BioNTech SE) is an mRNA-based vaccine

encoding the full length prefusion S protein (Polack et al., 2020; Walsh et al., 2020). The vaccine has been shown to elicit virus neutralising titres similar to those from recovered patients (Walsh et al., 2020). Isogenic N501 and Y501 SARS-CoV-2 strains were generated and tested against sera from 20 participants of the vaccine trial. The sera had equivalent titres of neutralising antibodies to the 2 viruses (Gu et al., 2020; Xie et al., 2021). No reduction in neutralising activity against the virus carrying the Y501 variant was noted to be consistent with preserved neutralisation of 15 pseudoviruses with other mutations found to be circulating (Sahin et al., 2020.). However, another study (Wang et al., 2021), which used a similar *in vitro* approach, found that the activity against the N501Y variant of plasma from vaccinees with either BNT162b2 or mRNA-1273 (Moderna) vaccines was slightly but significantly reduced; and two further studies on sera from BNT162b2 vaccinated individuals detected only a minor reduction in efficacy against the B.1.1.7 variant compared to the prototypic strain (Muik et al., 2021; Hoffmann et al., 2021; Muik et al., 2021). Finally, an *in vivo* study using hamsters as model organisms indicated that the ChAdOx1 vaccine (AZD1222, Oxford/AstraZeneca) is protective against the B.1.1.7 lineage (Fischer et al., 2021). Overall, these results suggest that most available vaccines should retain an acceptable efficacy against the B.1.1.7 variant.

Patients with lineage B.1.351, variant 20H/501Y.V2 originally detected in South Africa, which is now spreading rapidly albeit not yet globally, show a greater viral load which is in line with its increased transmissibility and greater disease burden. However, there is no clear evidence of greater disease severity in patients infected with this variant (who.int, SARS-CoV-2 Variants-Disease Outbreak News, 31st December 2020). Information is needed on cross neutralisation, where antibodies to an infecting strain might provide protection to a different strain. Faulkner et al. (2021) reported that antibodies elicited during B.1.1.7 infection had reduced recognition and neutralising ability of parental strains or B.1.351, and that this drop in cross reactivity was greater following infection with B.1.1.7 than parental strains (Faulkner et al., 2021). The rapid spread of B.1.351, containing multiple S protein mutations, is a cause for concern, not only because of its increased infectiousness but also because there is a suggestion that it could compromise vaccination (Sahin et al., 2020; Tegally et al., 2020). In particular, the E484K substitution is found in B.1.351 and also in P.1 lineages (Tegally et al., 2020) (Table 3, Fig. 1). This mutation is in the RBD recognised by neutralising antibodies. Greaney et al. (2021) longitudinally sampled polyclonal convalescent plasmas, and looking at the 3 main epitopes of the RBD (443-450, RB Motif, and 494-501) found binding of 11/11 samples was reduced at F456 and 9/11 reduced at E484. The most important site was E484 where neutralisation by some plasma was reduced >10 fold (Greaney et al., 2021). However, the impact of epitope variation on responsiveness varied amongst individuals and temporally in the same individual. E484K was predicted to be an immune escape variant by Andreano et al. (2020) and has been described to be present after reinfection (Resende et al., 2021; Vasques Nonaka et al., 2021). Experimentally SARS-CoV-2 was grown in the serum from a convalescing patient, selecting for mutations that were avoiding the antibody repertoire. Three mutations were picked up, one of which was E484. This mutation was missed by the patient's neutralising antibodies. We might expect the E484K mutation to have a greatly reduced susceptibility to neutralisation in some individuals (Tegally et al., 2020).

Additional studies further emphasise the need for vigilance with implications for immunity and vaccination against strains carrying combinations of new variants, particularly since there is conflicting data appearing in the literature. Wibmer et al. considering the mutations in SARS-CoV-2 B.1.351 demonstrated the complete escape of this lineage from therapeutically relevant monoclonal antibodies (Wibmer et al., 2021). Each mutation K417N and E484K, both in the ACE2 RBD, failed to bind 3 such antibodies, and the same antibodies were unable to neutralise respective pseudovirus. This lineage also showed substantial or complete escape from neutralising antibodies in COVID-19 plasma. In

their study of antibody and memory B cell responses of 20 volunteers who had received either of the mRNA vaccines, described above, Wang et al. (2021) showed high IgM and IgG anti-S, RBD titres after a second vaccination and RBD neutralising antibodies and memory B cell responses similar to those seen after natural infection, but RBD neutralising activity against E484K- and N501Y-carrying variants, or the SARS-CoV-2 B.1.351 combination K417N:E484K:N501Y, showed a small but significant decrease (Wang et al., 2021; Wibmer et al., 2021). In further monoclonal antibody studies, examining B cell memory, 26% of the antibodies showed at least a 5-fold decrease in binding to at least one of the RBD mutants tested, which included E484K. The K417N-E484K-N501Y combination is also found in the Brazilian P.1 lineages, reported by Faria et al. (2021b) as ~1.4-2.2 fold more transmissible and 25-61% more likely to evade protective immunity elicited by previous infection with non-P.1 lineages (Faria et al., 2021b). Recently, Wu et al. reported the neutralising capacity of sera of humans and non-human primates after vaccination with the Moderna BNT162b2 vaccine (Wu et al., 2021). When testing against the lineage B.1.1.7 S protein, including VH69-70del and N501Y, in a pseudovirus system, there was no significant impact on neutralising activity. However, there was a significant reduction in neutralising activity when testing against the B.1.351 lineage in pseudoviruses containing K417N-E484K-N501Y. Similar results were reached by Hoffmann and co-workers, who showed that the B.1.351 and P.1 variants are resistant to therapeutic monoclonal antibodies and to the neutralization activity of plasma from BNT162b2 vaccinated individuals (Hoffmann et al., 2021; Muik et al., 2021). Finally, a double-blind, randomized, controlled trial with ChAdOx1 in South Africa showed that the vaccine offered limited protection against mild and moderate COVID-19 caused by B.1.351 (Madhi et al., 2021).

Research on neutralising antibodies has focussed initially on the RBD. However, antibodies binding outside this region will also play a part in control of infection. McCallum et al. (2021b) generated monoclonal antibodies from memory B cells recognising the N-terminal domain. A subset of these antibodies had potent neutralising effects. Using a Syrian hamster model, they mapped neutralisation escape mutants and produced an NTD antigenic map. Similarly, although neutralising antibodies, including those recognising the RBD and NTD, will control infection *per se*, T cell responses will determine later stages of infection, determining disease severity, and are particularly relevant to observations made on cross strain protection and vaccine efficacy. Reports on T cell responses appearing in the literature include studies on model organisms and using cells from convalescent subjects. The Oxford/AstraZeneca vaccine efficacy against B.1.1.7 and B.1.351 was tested in a Syrian hamster model (Fischer et al., 2021). Whilst neutralising antibodies were much reduced in vaccinated hamsters against B.1.351 compared to B.1.1.1.7, nevertheless the vaccine was effective against clinical disease caused by either variant. This is likely due to protection against severe COVID-19 being mediated by T cells, since T cell epitopes are not affected by the substitutions in these lineages, though mild infection, prior to T cell activity, still occurs. Tarke et al. (2021) reported on CD4+ and CD8+ T cell responses in convalescent subjects, and Moderna and Pfizer/BionTech vaccinees, recognising the ancestral strain compared to variant lineages including B.1.1.7, B.1.351 and P.1; and concluding that T cell responses in convalescent subjects are not substantially affected by mutations in SARS-CoV-2 variants (Tarke et al., 2021).

6.5. Ongoing investigations concerning host immunity

Since variant strains are rapidly occurring and their novel genetic composition includes new substitutions as well as indels, emphasis in the coming months will need to be placed on identifying specific genomic alterations that significantly impact viral evolution, transmission, and, notably, immune responses, which are critical for successful vaccines. Effort is needed to associate particular mutations, and also their

combinations, with prevalence and transmissibility, for instance by identifying within novel variants changes in antigenic epitopes. While the majority of antigen epitopes lie within the S protein, a broad set of epitopes for T cells have also been discovered along the entire length of the genome (Sette and Crotty, 2021). It seems likely that mutations outside the S region may impact the immune response. Population data so far are encouraging as it seems that mutations are not notably impacting the immune response, and disease severity has not been markedly different in people infected with different strains to date. Consortium groups will be required to study the threats from new SARS-CoV-2 variants, such as the one at Imperial College in London, announced in January 2021.

However, recent laboratory findings suggest caution is needed and the complex problems associated with 'immune escape', particularly for vaccination and therapeutics are real (Kupferschmidt, 2021). More investigations are required to test the effect of novel variants on immune recognition and response, for instance identification of viral epitopes recognised in both recovered patients and vaccinees, which may shed some light on the extent of immune escape. Monitoring will also be required to understand whether novel variants may affect reinfection or vaccine failure. It is possible that vaccines will need to be adjusted annually, as is the case for vaccinating against influenza. Preparations for providing boosters adjusted for new variants are already underway (Kupferschmidt, 2021). Developers of new vaccines are reporting on efficacy against new variants within the approval phase; e.g., data is just appearing for the Novavax S-protein vaccine. Although it is currently thought that immunity to SARS-CoV-2 will last for months after the initial infection at a minimum, insufficient individuals have been studied and time elapsed to firmly establish the duration and nature of immunity to infection or current vaccines.

7. Importance of large-scale sequencing and surveillance

Since the early identification of SARS-CoV-2 a large effort has been made to characterize the global evolving genetic diversity of the virus. As of March 2021, more than 830 thousand full-genomic sequences of SARS-CoV-2 have been made available on public databases. Genomic characterization and analysis provide valuable tools for the global effort to control the pandemic. Specifically, sequence analysis has been used for phylogenetic classification into clades or lineages using slightly different approaches, such as the GISAID clades (https://www.gisaid.org/) and the PANGO lineages (https://cov-lineages.org/index.html; https://pangolin.cog-uk.io/) (Rambaut et al., 2020a), as detailed above. PANGO classification of newly sequenced viruses can be performed online through the pangolin website (https://pangolin.cog-uk.io/).

Genomic sequencing and analysis have been used for many purposes as for example to identify the geographic origin of SARS-CoV-2 transmission, to uncover the patterns of virus dispersal, to quantify the levels of virus importation, to infer transmission dynamics, and to estimate whether contacts identified by phylogenetic analysis were in accordance with those identified by contact tracing. Specifically, Worobey *et al.* investigated the temporal and geographic origin of SARS-CoV-2 infections at the early stage of the pandemic in Europe and North America suggesting that the origin of the virus in Washington state and Italy was from two independent sources from China (Worobey et al., 2020). In a study in Brazil, Candido *et al.* suggested that the majority of viruses were introduced from Europe during late February – beginning of March and that during the early phase of the pandemic virus spread within the country (Candido et al., 2020). Thereafter, several transmissions from large urban centres were detected across the country, which coincided with the increase in national air travel (Candido et al., 2020). In one of the earliest studies in Iceland, Gudbjartsson *et al* found that the founder strains in Iceland were separated from the original haplotype in Wuhan by 5 mutations (Gudbjartsson et al., 2020). They identified the number of distinct viral clades among those who had travelled abroad, and

provide a lower bound of virus importation events, and the putative geographic origin of the infections. Importantly, they also found that the contacts identified through molecular analyses and contact tracing were highly concordant, thus suggesting that the latter approach can be used for the accurate identification of contacts at highest risk and prevention of SARS-CoV-2 transmission (Gudbjartsson et al., 2020). Multiple molecular epidemiology studies from Europe (i.e. Austria, Scotland, Italy, Romania, Austria, the UK, the Netherlands, Greece, France) (da Silva Filipe et al., 2021; Di Giallonardo et al., 2020; du Plessis et al., 2021; Oude Munnink et al., 2020; Popa et al., 2020; Spanakis et al., 2021; Surleac et al., 2020), the Americas (Franco et al., 2021; Lemieux et al., 2021), Asia (Ko et al., 2021), or globally (Mastriani et al., 2021) investigated the origin of SARS-CoV-2 transmission during the first pandemic wave, the patterns of viral dispersal, and the dynamics of SARS-CoV-2 lineages (Dellicour et al., 2020).

Genomic surveillance is of particular importance for SARS-CoV-2 for many reasons, including the detection of new variants with biological importance, as is the case of the novel lineages, to monitor their local or global spread; this issue becomes of higher importance during vaccination and afterwards, when the emergence of immune escape variants is highly likely. Moreover, as shown by previous studies, genomic analysis is of relevance to better understand the spatiotemporal characteristics of SARS-CoV-2 transmission, thus providing an added value to traditional epidemiology for the control of the pandemic. In this respect, genomic epidemiology can be regarded as a tool to inform public-health decisions and containment strategies.

8. Conclusions and perspectives

The COVID-19 pandemic has triggered an unprecedented international effort at multiple levels. From the perspective of viral genomics, since its initial phases, the spread of SARS-CoV-2 has been described in real-time, through the phylodynamic analysis of viral genomes. The speed and scale of sequencing programs has been extraordinary: as of March 23rd 2021, more than 836,000 high-quality, complete genomes have been sequenced and deposited in public databases (Fig. 3). These efforts have proved to be pivotal for the early identification of novel variants and emerging lineages. However, genomic surveillance strategies and resources are highly heterogeneous across geographic areas, with the overwhelming majority of sequences coming from a few countries (Fig. 3) (The Lancet, 2021). Moreover, the top 10 countries (UK, USA, Denmark, Germany, Canada, Japan, Switzerland, Australia, Netherlands, Italy) produced 83.5% of sequences despite only having 35% of worldwide cases (data from the WHO, https://covid19.who.int/table). This implies that a large portion of the genetic diversity of SARS-CoV-2 remains unsampled and that, should new lineages emerge in regions where surveillance is leaky, they may remain undetected for a long time, with clear consequences on their spread and possibility of control (or lack thereof). Thus, it is essential to raise the bar on genomic surveillance, both in already active areas and in regions where sequencing efforts have lagged behind. Overall, there is a need to improve genomic sequencing capabilities to characterize the genetic diversity of SARS-CoV-2 isolates that are circulating in populations worldwide, in line with WHO guidelines (https://www.who.int/publications/i/item/WHO-2019-nCoV-genomic_sequencing-2021.1). This capability shall enhance diagnostic, therapeutic and preventive strategies. Active surveillance will be even more important in the coming months, when the deployment of large-scale vaccination campaigns and the use of monoclonal antibodies may subject the virus to novel selective pressures and possibly favour **antigenic drift**.

Finally, as noted elsewhere (Cyranoski, 2021), genomic epidemiology has transformed into an essential tool for adopting informed public health measures quickly, but its full potential will only be realized when embedded in surveillance programs that are widespread, standardized and incorporated into national programmes for pandemic-prevention.

Glossary

Antigenic drift: accumulation of mutations in viral protein regions that are recognized by the immune system (antigens).

Backward mutation: a mutation that acts on a previously mutated site restoring the original nucleotide (C → A → C).

Convergent evolution: in general, the evolution of similar traits in two or more distantly related organisms. In this context, the independent acquisition of the same mutation by different viral lineages.

Monophyletic cluster: in this context, a group of sequences that descend from a common ancestor.

Parallel mutations: Mutations that appear repeatedly and independently in different lineages.

Positive selection: the increase in frequency of a beneficial mutation in a population. In coding sequences, positive selection is often detected by searching for genes or sites that show an excess of non-synonymous changes compared to synonyms substitutions.

Reproduction number: the expected number of secondary cases produced by a single infected individual.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2021.104869.

References

Abdel-Moneim, A.S., Abdelwhab, E.M., 2020. Evidence for SARS-CoV-2 Infection of Animal Hosts. Pathogens 9. https://doi.org/10.3390/pathogens9070529.

Abu-Raddad, L.J., Chemaitelly, H., Malek, J.A., Ahmed, A.A., Mohamoud, Y.A., Younuskunju, S., Al Kanaani, Z., Al Khal, A., Al Kuwari, E., Butt, A.A., Coyle, P., Jeremijenko, A., Kaleeckal, A.H., Latif, A.N., Shaik, R.M., Abdul Rahim, H.F., Yassine, H.M., Al Kuwari, M.G., Al Romaihi, H.E., Al-Thani, M.H., Bertollini, R., 2020. Two prolonged viremic SARS-CoV-2 infections with conserved viral genome for two months. Infect. Genet. Evol. 88, 104684.

Adam, D.C., Wu, P., Wong, J.Y., Lau, E.H.Y., Tsang, T.K., Cauchemez, S., Leung, G.M., Cowling, B.J., 2020. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. Nat. Med. 26, 1714–1719.

Ahmed, S.F., Quadeer, A.A., McKay, M.R., 2020. Preliminary identification of potential vaccine targets for the COVID-19 Coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. Viruses 12. https://doi.org/10.3390/v12030254.

Al Khatib, H.A., Benslimane, F.M., Elbashir, I.E., Coyle, P.V., Al Maslamani, M.A., Al-Khal, A., Al Thani, A.A., Yassine, H.M., 2020. Within-host diversity of SARS-CoV-2 in COVID-19 patients with variable disease severities. Front. Cell. Infect. Microbiol. 10, 575613.

Alm, E., Broberg, E.K., Connor, T., Hodcroft, E.B., Komissarov, A.B., Maurer-Stroh, S., Melidou, A., Neher, R.A., O'Toole, Á., Pereyaslov, D., WHO European Region sequencing laboratories and GISAID EpiCoV group, WHO European Region sequencing laboratories and GISAID EpiCoV group*, 2020. Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. Euro. Surveill. 25. https://doi.org/10.2807/1560-7917.ES.2020.25.32.2001410.

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

Aso, J., Kimura, H., Ishii, H., Saraya, T., Kurai, D., Matsushima, Y., Nagasawa, K., Ryo, A., Takizawa, H., 2019. Molecular evolution of the fusion protein gene in human respirovirus 3. Front. Microbiol. 10, 3054.

Avanzato, V.A., Matson, M.J., Seifert, S.N., Pryce, R., Williamson, B.N., Anzick, S.L., Barbian, K., Judson, S.D., Fischer, E.R., Martens, C., Bowden, T.A., de Wit, E., Riedo, F.X., Munster, V.J., 2020. Case study: prolonged infectious SARS-CoV-2 shedding from an asymptomatic immunocompromised individual with cancer. Cell 183, 1901–1912 e9.

Bazykin, G.A., Stanevich, O., Danilenko, D., Fadeev, A., Komissarova, K., Ivanova, A., Sergeeva, M., Safina, K., Nabieva, E., Klink, G., Garushyants, S., Zabutova, J., Kholodnaia, A., Skorokhod, I., Ryabchikova, V.V., Komissarov, A., Lioznov, D., 2021. Emergence of Y453F and Δ69-70HV mutations in a lymphoma patient with long-term COVID-19. Virological.org. https://virological.org/t/emergence-of-y453f-and-69-70hv-mutations-in-a-lymphoma-patient-with-long-term-covid-19/580.

Boni, M.F., Lemey, P., Jiang, X., Lam, T.T.-Y., Perry, B.W., Castoe, T.A., Rambaut, A., Robertson, D.L., 2020. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nat. Microbiol. 5, 1408–1417.

Borges, V., Sousa, C., Menezes, L., Gonçalves, A.M., Picão, M., Almeida, J.P., Vieita, M., Santos, R., Silva, A.R., Costa, M., Carneiro, L., Casaca, P., Pinto-Leite, P., Peralta-Santos, A., Isidro, J., Duarte, S., Vieira, L., Guiomar, R., Silva, S., Nunes, B., Gomes, J.P., 2021a. Tracking SARS-CoV-2 lineage B.1.1.7 dissemination: insights from nationwide spike gene target failure (SGTF) and spike gene late detection (SGTL) data, Portugal, week 49 2020 to week 3 2021. Euro. Surveill. 26 https://doi.org/10.2807/1560-7917.ES.2021.26.10.2100130.

Borges, V., Sousa, C., Menezes, L., Gonçalves, A.M., Picão, M., Almeida, J.P., Vieita, M., Santos, R., Silva, A.R., Costa, M., Carneiro, L., Isidro, J., Duarte, S., Vieira, L., Guiomar, R., Silva, S., Nunes, B., Gomes, J.P., 2021b. Tracking SARS-CoV-2 VOC 202012/01 (lineage B.1.1.7) dissemination in Portugal: insights from nationwide RT-PCR Spike gene drop out data. Virological.org. https://virological.org/t/tracking-sars-cov-2-voc-202012-01-lineage-b-1-1-7-dissemination-in-portugal-insights-from-nationwide-rt-pcr-spike-gene-drop-out-data/600.

Calistri, P., Amato, L., Puglia, I., Cito, F., Di Giuseppe, A., Danzetta, M.L., Morelli, D., Di Domenico, M., Caporale, M., Scialabba, S., Portanti, O., Curini, V., Perletta, F., Cammà, C., Ancora, M., Savini, G., Migliorati, G., D'Alterio, N., Lorusso, A., 2021. Infection sustained by lineage B.1.1.7 of SARS-CoV-2 is characterised by longer persistence and higher viral RNA loads in nasopharyngeal swabs. Int. J. Infect. Dis. https://doi.org/10.1016/j.ijid.2021.03.005.

Candido, D.S., Claro, I.M., de Jesus, J.G., Souza, W.M., Moreira, F.R.R., Dellicour, S., Mellan, T.A., du Plessis, L., Pereira, R.H.M., Sales, F.C.S., Manuli, E.R., Thézé, J., Almeida, L., Menezes, M.T., Voloch, C.M., Fumagalli, M.J., Coletti, T.M., da Silva, C. A.M., Ramundo, M.S., Amorim, M.R., Hoeltgebaum, H.H., Mishra, S., Gill, M.S., Carvalho, L.M., Buss, L.F., Prete Jr., C.A., Ashworth, J., Nakaya, H.I., Peixoto, P.S., Brady, O.J., Nicholls, S.M., Tanuri, A., Rossi, Á.D., Braga, C.K.V., Gerber, A.L., de C Guimarães, A.P., Gaburo Jr., N., Alencar, C.S., Ferreira, A.C.S., Lima, C.X., Levi, J.E., Granato, C., Ferreira, G.M., Francisco Jr., R.S., Granja, F., Garcia, M.T., Moretti, M. L., Perroud Jr., M.W., Castiñeiras, T.M.P.P., Lazari, C.S., Hill, S.C., de Souza Santos, A.A., Simeoni, C.L., Forato, J., Sposito, A.C., Schreiber, A.Z., Santos, M.N.N., de Sá, C.Z., Souza, R.P., Resende-Moreira, L.C., Teixeira, M.M., Hubner, J., Leme, P. A.F., Moreira, R.G., Nogueira, M.L., Brazil-UK Centre for Arbovirus Discovery, Diagnosis, Genomics and Epidemiology (CADDE) Genomic Network, Ferguson, N. M., Costa, S.F., Proenca-Modena, J.L., Vasconcelos, A.T.R., Bhatt, S., Lemey, P., Wu, C.-H., Rambaut, A., Loman, N.J., Aguiar, R.S., Pybus, O.G., Sabino, E.C., Faria, N.R., 2020. Evolution and epidemic spread of SARS-CoV-2 in Brazil. Science 369, 1255–1260.

Chan, J.F.-W., Kok, K.-H., Zhu, Z., Chu, H., To, K.K.-W., Yuan, S., Yuen, K.-Y., 2020. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg. Microbes Infect. 9, 221–236.

Chaw, S.-M., Tai, J.-H., Chen, S.-L., Hsieh, C.-H., Chang, S.-Y., Yeh, S.-H., Yang, W.-S., Chen, P.-J., Wang, H.-Y., 2020. The origin and underlying driving forces of the SARS-CoV-2 outbreak. J. Biomed. Sci. 27, 73.

Choi, B., Choudhary, M.C., Regan, J., Sparks, J.A., Padera, R.F., Qiu, X., Solomon, I.H., Kuo, H.-H., Boucau, J., Bowman, K., Adhikari, U.D., Winkler, M.L., Mueller, A.A., Hsu, T.Y.-T., Desjardins, M., Baden, L.R., Chan, B.T., Walker, B.D., Lichterfeld, M., Brigl, M., Kwon, D.S., Kanjilal, S., Richardson, E.T., Jonsson, A.H., Alter, G., Barczak, A.K., Hanage, W.P., Yu, X.G., Gaiha, G.D., Seaman, M.S., Cernadas, M., Li, J.Z., 2020. Persistence and evolution of SARS-CoV-2 in an immunocompromised host. N. Engl. J. Med. 383, 2291–2293.

Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat. Microbiol. 5, 536–544.

Cui, J., Li, F., Shi, Z.-L., 2019. Origin and evolution of pathogenic coronaviruses. Nat. Rev. Microbiol. 17, 181–192.

Cunha, L.L., Perazzio, S.F., Azzi, J., Cravedi, P., Riella, L.V., 2020. Remodeling of the immune response with aging: immunosenescence and its potential impact on COVID-19 immune response. Front. Immunol. 11, 1748.

Cyranoski, D., 2021. Alarming COVID variants show vital role of genomic surveillance. Nature 589, 337–338.

da Silva Filipe, A., Shepherd, J.G., Williams, T., Hughes, J., Aranday-Cortes, E., Asamaphan, P., Ashraf, S., Balcazar, C., Brunker, K., Campbell, A., Carmichael, S., Davis, C., Dewar, R., Gallagher, M.D., Gunson, R., Hill, V., Ho, A., Jackson, B., James, E., Jesudason, N., Johnson, N., McWilliam Leitch, E.C., Li, K., MacLean, A., Mair, D., McAllister, D.A., McCrone, J.T., McDonald, S.E., McHugh, M.P., Morris, A. K., Nichols, J., Niebel, M., Nomikou, K., Orton, R.J., O'Toole, Á., Palmarini, M., Parcell, B.J., Parr, Y.A., Rambaut, A., Rooke, S., Shaaban, S., Shah, R., Singer, J.B., Smollett, K., Starinskij, I., Tong, L., Sreenu, V.B., Wastnedge, E., COVID-19 Genomics UK (COG-UK) Consortium, Holden, M.T.G., Robertson, D.L., Templeton, K., Thomson, E.C., 2021. Genomic epidemiology reveals multiple introductions of SARS-CoV-2 from mainland Europe into Scotland. Nat. Microbiol. 6, 112–122.

Davies, N.G., Abbott, S., Barnard, R.C., Jarvis, C.I., Kucharski, A.J., Munday, J.D., Pearson, C.A.B., Russell, T.W., Tully, D.C., Washburne, A.D., Wenseleers, T., Gimma, A., Waites, W., Wong, K.L.M., van Zandvoort, K., Silverman, J.D., CMMID COVID-19 Working Group; COVID-19 Genomics UK (COG-UK) Consortium, Diaz-Ordaz, K., Keogh, R., Eggo, R.M., Funk, S., Jit, M., Atkins, K.E., Edmunds, W.J., 2021a. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. Science 372 (6538), eabg3055. https://doi.org/10.1126/science.abg3055.

Davies, N.G., Jarvis, C.I., CMMID COVID-19 Working Group, 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. https://doi.org/10.1038/s41586-021-03426-1.

Dellicour, S., Durkin, K., Hong, S.L., Vanmechelen, B., Martí-Carreras, J., Gill, M.S., Meex, C., Bontems, S., André, E., Gilbert, M., Walker, C., De Maio, N., Faria, N.R., Hadfield, J., Hayette, M.-P., Bours, V., Wawina-Bokalanga, T., Artesi, M., Baele, G., Maes, P., 2020. A phylodynamic workflow to rapidly gain insights into the dispersal history and dynamics of SARS-CoV-2 lineages. Mol. Biol. Evol. https://doi.org/10.1093/molbev/msaa284.

Deng, X., Garcia-Knight, M.A., Khalid, M.M., Servellita, V., Wang, C., Morris, M.K., Sotomayor-Gonzalez, A., Glasner, D.R., Reyes, K.R., Gliwa, A.S., Reddy, N.P., Sanchez San Martin, C., Federman, S., Cheng, J., Balcerek, J., Taylor, J., Streithorst, J.A., Miller, S., Kumar, G.R., Sreekumar, B., Chen, P.-Y., Schulze-Gahmen, U., Taha, T.Y., Hayashi, J.M., Siomoneau, C.R., McMahon, S., Lidsky, P.V., Xiao, Y., Green, N.M., Hemarajata, P., Espinosa, A., Kath, C., Haw, M., Bell, J., Hacker, J.K., Hanson, C., Wadford, D.A., Anaya, C., Ferguson, D., Lareau, L.F., Frankino, P.A., Shivram, H., Wyman, S.K., Ott, M., Andino, R., Chiu, C.Y., 2021. Transmission, infectivity, and antibody neutralization of an emerging SARS-CoV-2 variant in California carrying a L452R spike protein mutation. bioRxiv. https://doi.org/10.1101/2021.03.07.21252647.

Di Giallonardo, F., Duchene, S., Puglia, I., Curini, V., Profeta, F., Cammà, C., Marcacci, M., Calistri, P., Holmes, E.C., Lorusso, A., 2020. Genomic epidemiology of the first wave of SARS-CoV-2 in Italy. Viruses 12. https://doi.org/10.3390/v12121438.

Drosten, C., Günther, S., Preiser, W., van der Werf, S., Brodt, H.-R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R.A.M., Berger, A., Burguière, A.-M., Cinatl, J., Eickmann, M., Escriou, N., Grywna, K., Kramme, S., Manuguerra, J.-C., Müller, S., Rickerts, V., Stürmer, M., Vieth, S., Klenk, H.-D., Osterhaus, A.D.M.E., Schmitz, H., Doerr, H.W., 2003. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N. Engl. J. Med. 348, 1967–1976.

du Plessis, L., McCrone, J.T., Zarebski, A.E., Hill, V., Ruis, C., Gutierrez, B., Raghwani, J., Ashworth, J., Colquhoun, R., Connor, T.R., Faria, N.R., Jackson, B., Loman, N.J., O'Toole, Á., Nicholls, S.M., Parag, K.V., Scher, E., Vasylyeva, T.I., Volz, E.M., Watts, A., Bogoch, I.I., Khan, K., COVID-19 Genomics UK (COG-UK) Consortium, Aanensen, D.M., Kraemer, M.U.G., Rambaut, A., Pybus, O.G., 2021. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. Science. https://doi.org/10.1126/science.abf2946.

Duchene, S., Featherstone, L., Haritopoulou-Sinanidou, M., Rambaut, A., Lemey, P., Baele, G., 2020. Temporal signal and the phylodynamic threshold of SARS-CoV-2. Virus Evol. 6, veaa061.

Eckerle, L.D., Lu, X., Sperry, S.M., Choi, L., Denison, M.R., 2007. High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. J. Virol. 81, 12135–12144.

Eckerle, L.D., Becker, M.M., Halpin, R.A., Li, K., Venter, E., Lu, X., Scherbakova, S., Graham, R.L., Baric, R.S., Stockwell, T.B., Spiro, D.J., Denison, M.R., 2010. Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. PLoS Pathog. 6, e1000896.

Faria, N.R., Morales Claro, I., Candido, D., Moyses Franco, L.A., Andrade, P.S., Coletti, T. M., Silva, C.A.M., Sales, F.C., Manuli, E.R., Aguiar, R.S., Gaburo, N., da C. Camilo, C., Fraiji, N.A., Crispim, M.A.E., Carvalho, M.S.S., Rambaut, A., Loman, N., Pybus, O.G., Sabino, E.C., CADDE Genomic Network, 2021a. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings SARS-CoV-2 coronavirus. Virological.org. https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586.

Faria, N.R., Mellan, T.A., Whittaker, C., Claro, I.M., Candido, D. da S., Mishra, S., Crispim, M.A.E., Sales, F.C., Hawryluk, I., McCrone, J.T., Hulswit, R.J.G., Franco, L. A.M., Ramundo, M.S., de Jesus, J.G., Andrade, P.S., Coletti, T.M., Ferreira, G.M., Silva, C.A.M., Manuli, E.R., Pereira, R.H.M., Peixoto, P.S., Kraemer, M.U., Gaburo, N., Camilo, C. da C., Hoeltgebaum, H., Souza, W.M., Rocha, E.C., de Souza, L.M., de Pinho, M.C., Araujo, L.J.T., Malta, F.S.V., de Lima, A.B., Silva, J. do P., Zauli, D.A.G., Ferreira, A.C. de S., Schnekenberg, R.P., Laydon, D.J., Walker, P.G. T., Schlüter, H.M., Dos Santos, A.L.P., Vidal, M.S., Del Caro, V.S., Filho, R.M.F., Dos Santos, H.M., Aguiar, R.S., Modena, J.L.P., Nelson, B., Hay, J.A., Monod, M., Miscouridou, X., Coupland, H., Sonabend, R., Vollmer, M., Gandy, A., Suchard, M.A.,

Bowden, T.A., Pond, S.L.K., Wu, C.-H., Ratmann, O., Ferguson, N.M., Dye, C., Loman, N.J., Lemey, P., Rambaut, A., Fraiji, N.A., Carvalho, M. do P.S.S., Pybus, O. G., Flaxman, S., Bhatt, S., Sabino, E.C., 2021b. Genomics and epidemiology of a novel SARS-CoV-2 lineage in Manaus, Brazil. medRxiv. https://doi.org/10.1101/2021.02.26.21252554.

Faulkner, N., Ng, K.W., Wu, M., Harvey, R., Hussain, S., Greco, M., Bolland, W., Warchal, S., Margaritis, M., Paraskevopoulou, S., Heaney, J., Rickman, H., Houlihan, C., Spyer, M., Frampton, D., Byott, M., de Oliveira, T., Sigal, A., Kjaer, S., Swanton, C., Gandhi, S., Beale, R., Gamblin, S., Mccauley, J., Daniels, R., Howell, M., Bauer, D., Nastouli, E., Kassiotis, G., 2021. Reduced antibody cross-reactivity following infection with B.1.1.7 than with parental SARS-CoV-2 strains. bioRxiv. https://doi.org/10.1101/2021.03.01.433314.

Fischer, R.J., van Doremalen, N., Adney, D.R., Yinda, C.K., Port, J.R., Holbrook, M.G., Schulz, J.E., Williamson, B.N., Thomas, T., Barbian, K., Anzick, S.L., Ricklefs, S., Smith, B.J., Long, D., Martens, C., Saturday, G., de Wit, E., Gilbert, S.C., Lambe, T., Munster, V.J., 2021. ChAdOx1 nCoV-19 (AZD1222) protects hamsters against SARS-CoV-2 B.1.351 and B.1.1.7 disease. bioRxiv. https://doi.org/10.1101/2021.03.11.435000.

Flanagan, K.L., Best, E., Crawford, N.W., Giles, M., Koirala, A., Macartney, K., Russell, F., Teh, B.W., Wen, S.C., 2020. Progress and pitfalls in the quest for effective SARS-CoV-2 (COVID-19) vaccines. Front. Immunol. 11, 579250.

Folegatti, P.M., Ewer, K.J., Aley, P.K., Angus, B., Becker, S., Belij-Rammerstorfer, S., Bellamy, D., Bibi, S., Bittaye, M., Clutterbuck, E.A., Dold, C., Faust, S.N., Finn, A., Flaxman, A.L., Hallis, B., Heath, P., Jenkin, D., Lazarus, R., Makinson, R., Minassian, A.M., Pollock, K.M., Ramasamy, M., Robinson, H., Snape, M., Tarrant, R., Voysey, M., Green, C., Douglas, A.D., Hill, A.V.S., Lambe, T., Gilbert, S.C., Pollard, A. J., Oxford COVID Vaccine Trial Group, 2020. Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. Lancet 396, 467–478.

Forni, D., Cagliani, R., Clerici, M., Sironi, M., 2017. Molecular evolution of human coronavirus genomes. Trends Microbiol. 25, 35–48.

Franco, D., Gonzalez, C., Abrego, L.E., Carrera, J.-P., Diaz, Y., Caicedo, Y., Moreno, A., Chavarria, O., Gondola, J., Castillo, M., Valdespino, E., Gaitán, M., Martínez-Mandiche, J., Hayer, L., Gonzalez, P., Lange, C., Molto, Y., Mojica, D., Ramos, R., Mastelari, M., Cerezo, L., Moreno, L., Donnelly, C.A., Pascale, J.M., Faria, N.R., Lopez-Verges, S., Martinez, A.A., Gorgas COVID19 team and Panama COVID19 Laboratory Network, 2021. Early transmission dynamics, spread, and genomic characterization of SARS-CoV-2 in Panama. Emerg. Infect. Dis. 27, 612–615.

Gamage, A.M., Tan, K.S., Chan, W.O.Y., Liu, J., Tan, C.W., Ong, Y.K., Thong, M., Andiappan, A.K., Anderson, D.E., Wang, Y., Wang, L.F., 2020. Infection of human nasal epithelial cells with SARS-CoV-2 and a 382-nt deletion isolate lacking ORF8 reveals similar viral kinetics and host transcriptional profiles. PLoS Pathog. 16 (12), e1009130 https://doi.org/10.1371/journal.ppat.1009130.

Garry, R.F., Andersen, K.G., Gallaher, W.R., Lam, T.T.-Y., Gangaparapu, K., Latif, A.A., Beddingfield, B.J., Rambaut, A., Holmes, E.C., 2021. Spike protein mutations in novel SARS-CoV-2 'variants of concern' commonly occur in or near indels. Virological.org. https://virological.org/t/spike-protein-mutations-in-novel-sars-cov-2-variants-of-concern-commonly-occur-in-or-near-indels/605.

Ghafari, M., du Plessis, L., Pybus, O.G., Katzourakis, A., 2020. Time dependence of SARS-CoV-2 substitution rates. Virological.org. https://virological.org/t/time-dependence-of-sars-cov-2-substitution-rates/542.

Gómez-Carballa, A., Bello, X., Pardo-Seco, J., Martinón-Torres, F., Salas, A., 2020. Mapping genome variation of SARS-CoV-2 worldwide highlights the impact of COVID-19 super-spreaders. Genome Res. 30, 1434–1448.

Greaney, A.J., Loes, A.N., Crawford, K.H.D., Starr, T.N., Malone, K.D., Chu, H.Y., Bloom, J.D., 2021. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. Cell Host Microbe 29, 463–476 e6.

Gribble, J., Stevens, L.J., Agostini, M.L., Anderson-Daniels, J., Chappell, J.D., Lu, X., Pruijssers, A.J., Routh, A.L., Denison, M.R., 2021. The coronavirus proofreading exoribonuclease mediates extensive viral recombination. PLoS Pathog. 17, e1009226.

Grifoni, A., Weiskopf, D., Ramirez, S.I., Mateus, J., Dan, J.M., Moderbacher, C.R., Rawlings, S.A., Sutherland, A., Premkumar, L., Jadi, R.S., Marrama, D., de Silva, A. M., Frazier, A., Carlin, A.F., Greenbaum, J.A., Peters, B., Krammer, F., Smith, D.M., Crotty, S., Sette, A., 2020. Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. Cell 181, 1489–1501 e15.

Grubaugh, N.D., Hanage, W.P., Rasmussen, A.L., 2020a. Making sense of mutation: what D614G means for the COVID-19 pandemic remains unclear. Cell. 182 (4), 812–827.

Grubaugh, N.D., Petrone, M.E., Holmes, E.C., 2020b. We shouldn't worry when a virus mutates during disease outbreaks. Nat. Microbiol. 5, 529–530.

Gu, H., Chen, Q., Yang, G., He, L., Fan, H., Deng, Y.-Q., Wang, Y., Teng, Y., Zhao, Z., Cui, Y., Li, Y., Li, X.-F., Li, J., Zhang, N.-N., Yang, X., Chen, S., Guo, Y., Zhao, G., Wang, X., Luo, D.-Y., Wang, H., Yang, X., Li, Y., Han, G., He, Y., Zhou, X., Geng, S., Sheng, X., Jiang, S., Sun, S., Qin, C.-F., Zhou, Y., 2020. Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. Science 369, 1603–1607.

Gudbjartsson, D.F., Helgason, A., Jonsson, H., Magnusson, O.T., Melsted, P., Norddahl, G.L., Saemundsdottir, J., Sigurdsson, A., Sulem, P., Agustsdottir, A.B., Eiriksdottir, B., Fridriksdottir, R., Gardarsdottir, E.E., Georgsson, G., Gretarsdottir, O.S., Gudmundsson, K.R., Gunnarsdottir, T.R., Gylfason, A., Holm, H., Jensson, B.O., Jonasdottir, A., Jonsson, F., Josefsdottir, K.S., Kristjansson, T., Magnusdottir, D.N., le Roux, L., Sigmundsdottir, G., Sveinbjornsson, G., Sveinsdottir, K.E., Sveinsdottir, M., Thorarensen, E.A., Thorbjornsson, B., Löve, A., Masson, G., Jonsdottir, I., Möller, A.D., Gudnason, T., Kristinsson, K.G.,

Thorsteinsdottir, U., Stefansson, K., 2020. Spread of SARS-CoV-2 in the Icelandic population. N. Engl. J. Med. 382, 2302–2315.

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.

Hahn, G., Wu, C.M., Lee, S., Hecker, J., Lutz, S.M., Haneuse, S., Qiao, D., DeMeo, D., Choudhary, M.C., Etemad, B., Mohammadi, A., Esmaeilzadeh, E., Cho, M.H., Tanzi, R.E., Li, J.Z., Randolph, A.G., Laird, N.M., Weiss, S.T., Silverman, E.K., Ribbeck, K., Lange, C., 2020. Two mutations in the SARS-CoV-2 spike protein and RNA polymerase complex are associated with COVID-19 mortality risk. Cold Spring Harbor Laboratory. https://doi.org/10.1101/2020.11.17.386714.

Hodcroft, E.B., Zuber, M., Nadeau, S., Crawford, K.H.D., Bloom, J.D., Veesler, D., Vaughan, T.G., Comas, I., Candelas, F.G., Stadler, T., Neher, R.A., 2020. Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. medRxiv. https://doi.org/10.1101/2020.10.25.20219063.

Hodcroft, E.B., Domman, D.B., Snyder, D.J., Oguntuyo, K., Van Diest, M., Densmore, K. H., Schwalm, K.C., Femling, J., Carroll, J.L., Scott, R.S., Whyte, M.M., Edwards, M. D., Hull, N.C., Kevil, C.G., Vanchiere, J.A., Lee, B., Dinwiddie, D.L., Cooper, V.S., Kamil, J.P., 2021. Emergence in late 2020 of multiple lineages of SARS-CoV-2 Spike protein variants affecting amino acid position 677. medRxiv. https://doi.org/10.1101/2021.02.12.21251658.

Hoffmann, M., Arora, P., Groß, R., Seidel, A., Hörnich, B.F., Hahn, A.S., Krüger, N., Graichen, L., Hofmann-Winkler, H., Kempf, A., Winkler, M.S., Schulz, S., Jäck, H.-M., Jahrsdörfer, B., Schrezenmeier, H., Müller, M., Kleger, A., Münch, J., Pöhlmann, S., 2021. SARS-CoV-2 variants B.1.351 and P.1 escape from neutralizing antibodies. Cell. https://doi.org/10.1016/j.cell.2021.03.036.

Jackson, B., Rambaut, A., Pybus, O.G., Robertson, D.L., Connor, T., Loman, N.J., The COVID-19 Genomics UK (COG-UK) consortium, 2021. Recombinant SARS-CoV-2 genomes involving lineage B.1.1.7 in the UK. Virological.org. https://virological.org/t/recombinant-sars-cov-2-genomes-involving-lineage-b-1-1-7-in-the-uk/658.

Jenkins, G.M., Rambaut, A., Pybus, O.G., Holmes, E.C., 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. J. Mol. Evol. 54, 156–165.

Jeyanathan, M., Afkhami, S., Smaill, F., Miller, M.S., Lichty, B.D., Xing, Z., 2020. Immunological considerations for COVID-19 vaccine strategies. Nat. Rev. Immunol. 20, 615–632.

Kalita, P., Padhi, A.K., Zhang, K.Y.J., Tripathi, T., 2020. Design of a peptide-based subunit vaccine against novel coronavirus SARS-CoV-2. Microb. Pathog. 145, 104236.

Karst, S.M., Baric, R.S., 2015. What is the reservoir of emergent human norovirus strains? J. Virol. 89, 5756–5759.

Kasibhatla, S.M., Kinikar, M., Limaye, S., Kale, M.M., Kulkarni-Kale, U., 2020. Understanding evolution of SARS-CoV-2: A perspective from analysis of genetic diversity of RdRp gene. J. Med. Virol. 92, 1932–1937.

Katoh, K., Rozewicki, J., Yamada, K.D., 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform. 20, 1160–1166.

Kemp, S.A., Collier, D.A., Datir, R., Ferreira, I., Gayed, S., Jahun, A., Hosmillo, M., Rees-Spear, C., Mlcochova, P., Lumb, I.U., Roberts, D.J., Chandra, A., Temperton, N., , The CITIID-NIHR BioResource COVID-19 Collaboration, The COVID-19 Genomics UK (COG-UK) Consortium, Sharrocks, K., Blane, E., Briggs, J.A.G., van Gils, M.J., Smith, K.G.C., Bradley, J.R., Smith, C., Doffinger, R., Ceron-Gutierrez, L., Barcenas-Morales, G., Pollock, D.D., Goldstein, R.A., Smielewska, A., Skittrall, J.P., Gouliouris, T., Goodfellow, I.G., Gkrania-Klotsas, E., Illingworth, C.J.R., McCoy, L.E., Gupta, R.K., 2020. Neutralising antibodies in Spike mediated SARS-CoV-2 adaptation. medRxiv, 2020.12.05.20241927, still in press.

Khan, K.A., Cheung, P., 2020. Presence of mismatches between diagnostic PCR assays and coronavirus SARS-CoV-2 genome. R. Soc. Open Sci. 7, 200636.

Kidd, M., Richter, A., Best, A., Cumley, N., Mirza, J., Percival, B., Mayhew, M., Megram, O., Ashford, F., White, T., Moles-Garcia, E., Crawford, L., Bosworth, A., Atabani, S.F., Plant, T., McNally, A., 2021. S-variant SARS-CoV-2 lineage B1.1.7 is associated with significantly higher viral loads in samples tested by ThermoFisher TaqPath RT-qPCR. J. Infect. Dis. https://doi.org/10.1093/infdis/jiab082.

Kirtipal, N., Bharadwaj, S., Kang, S.G., 2020. From SARS to SARS-CoV-2, insights on structure, pathogenicity and immunity aspects of pandemic human coronaviruses. Infect. Genet. Evol. 85, 104502.

Kissler, S.M., Fauver, J.R., Mack, C., Tai, C.G., Breban, M.I., Watkins, A.E., Samant, R.M., Anderson, D.J., Ho, D.D., Grubaugh, N.D., Grad, Y.H., 2021. Densely sampled viral trajectories suggest longer duration of acute infection with B.1.1.7 variant relative to non-B.1.1.7 SARS-CoV-2. bioRxiv. https://doi.org/10.1101/2021.02.16.21251535.

Kiyotani, K., Toyoshima, Y., Nemoto, K., Nakamura, Y., 2020. Bioinformatic prediction of potential T cell epitopes for SARS-Cov-2. J. Hum. Genet. 65, 569–575.

Ko, K., Nagashima, S., E B, Ouoba, S., Akita, T., Sugiyama, A., Ohisa, M., Sakaguchi, T., Tahara, H., Ohge, H., Ohdan, H., Kubo, T., Kishita, E., Kuwabara, M., Takahashi, K., Tanaka, J., 2021. Molecular characterization and the mutation pattern of SARS-CoV-2 during first and second wave outbreaks in Hiroshima, Japan. PLoS One 16, e0246383.

Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., Hastie, K.M., Parker, M.D., Partridge, D.G., Evans, C.M., Freeman, T.M., de Silva, T.I., Sheffield COVID-19 Genomics Group, McDanal, C., Perez, L.G., Tang, H., Moon-Walker, A., Whelan, S.P., LaBranche, C.C., Saphire, E.O., Montefiori, D.C., 2020. Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell 182, 812–827 e19.

Kosakovsky Pond, S.L., Wilkison, E., Weaver, S., James, S.E., Tegally, H., de Oliveira, T., Martin, D., 2020. A preliminary selection analysis of the South African V501.V2

SARS-CoV-2 clade. Virological.org. https://virological.org/t/a-preliminary-selection-analysis-of-the-south-african-v501-v2-sars-cov-2-clade/573.

Koyama, T., Weeraratne, D., Snowdon, J.L., Parida, L., 2020. Emergence of drift variants that may affect COVID-19 vaccine development and antibody treatment. Pathogens 9. https://doi.org/10.3390/pathogens9050324.

Kuhn, J.H., Bao, Y., Bavari, S., Becker, S., Bradfute, S., Brister, J.R., Bukreyev, A.A., Chandran, K., Davey, R.A., Dolnik, O., Dye, J.M., Enterlein, S., Hensley, L.E., Honko, A.N., Jahrling, P.B., Johnson, K.M., Kobinger, G., Leroy, E.M., Lever, M.S., Mühlberger, E., Netesov, S.V., Olinger, G.G., Palacios, G., Patterson, J.L., Paweska, J. T., Pitt, L., Radoshitzky, S.R., Saphire, E.O., Smither, S.J., Swanepoel, R., Towner, J. S., van der Groen, G., Volchkov, V.E., Wahl-Jensen, V., Warren, T.K., Weidmann, M., Nichol, S.T., 2013. Virus nomenclature below the species level: a standardized nomenclature for natural variants of viruses assigned to the family Filoviridae. Arch. Virol. 158, 301–311.

Kupferschmidt, K., 2021. New mutations raise specter of "immune escape.". Science 371, 329–330.

Laczkó, D., Hogan, M.J., Toulmin, S.A., Hicks, P., Lederer, K., Gaudette, B.T., Castaño, D., Amanat, F., Muramatsu, H., Oguin 3rd, T.H., Ojha, A., Zhang, L., Mu, Z., Parks, R., Manzoni, T.B., Roper, B., Strohmeier, S., Tombácz, I., Arwood, L., Nachbagauer, R., Karikó, K., Greenhouse, J., Pessaint, L., Porto, M., Putman-Taylor, T., Strasbaugh, A., Campbell, T.-A., Lin, P.J.C., Tam, Y.K., Sempowski, G.D., Farzan, M., Choe, H., Saunders, K.O., Haynes, B.F., Andersen, H., Eisenlohr, L.C., Weissman, D., Krammer, F., Bates, P., Allman, D., Locci, M., Pardi, N., 2020. A single immunization with nucleoside-modified mRNA vaccines elicits strong cellular and humoral immune responses against SARS-CoV-2 in mice. Immunity 53, 724–732 e7.

Ladner, J.T., Larsen, B.B., Bowers, J.R., Hepp, C.M., Bolyen, E., Folkerts, M., Sheridan, K., Pfeiffer, A., Yaglom, H., Lemmer, D., Sahl, J.W., Kaelin, E.A., Maqsood, R., Bokulich, N.A., Quirk, G., Watts, T.D., Komatsu, K.K., Waddell, V., Lim, E.S., Caporaso, J.G., Engelthaler, D.M., Worobey, M., Keim, P., 2020. An early pandemic analysis of SARS-CoV-2 population structure and dynamics in Arizona. MBio 11. https://doi.org/10.1128/mBio.02107-20.

Lam, T.T.-Y., Jia, N., Zhang, Y.-W., Shum, M.H.-H., Jiang, J.-F., Zhu, H.-C., Tong, Y.-G., Shi, Y.-X., Ni, X.-B., Liao, Y.-S., Li, W.-J., Jiang, B.-G., Wei, W., Yuan, T.-T., Zheng, K., Cui, X.-M., Li, J., Pei, G.-Q., Qiang, X., Cheung, W.Y.-M., Li, L.-F., Sun, F.-F., Qin, S., Huang, J.-C., Leung, G.M., Holmes, E.C., Hu, Y.-L., Guan, Y., Cao, W.-C., 2020. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. Nature. https://doi.org/10.1038/s41586-020-2169-0.

Larsen, B.B., Worobey, M., 2020. Identification of a novel SARS-CoV-2 Spike 69-70 deletion lineage circulating in the United States. Virological.org. https://virological.org/t/identification-of-a-novel-sars-cov-2-spike-69-70-deletion-lineage-circulating-in-the-united-states/577 (accessed 2.4.21).

Larsen, B.B., Worobey, M., 2021. Phylogenetic evidence that B.1.1.7 has been circulating in the United States since early- to mid-November. Virological.org. https://virological.org/t/phylogenetic-evidence-that-b-1-1-7-has-been-circulating-in-the-united-states-since-early-to-mid-november/598.

Latinne, A., Hu, B., Olival, K.J., Zhu, G., Zhang, L., Li, H., Chmura, A.A., Field, H.E., Zambrana-Torrelio, C., Epstein, J.H., Li, B., Zhang, W., Wang, L.-F., Shi, Z.-L., Daszak, P., 2020. Origin and cross-species transmission of bat coronaviruses in China. Nat. Commun. 11, 4235.

Lefever, S., Pattyn, F., Hellemans, J., Vandesompele, J., 2013. Single-nucleotide polymorphisms and other mismatches reduce performance of quantitative PCR assays. Clin. Chem. 59, 1470–1480.

Lemieux, J.E., Siddle, K.J., Shaw, B.M., Loreth, C., Schaffner, S.F., Gladden-Young, A., Adams, G., Fink, T., Tomkins-Tinch, C.H., Krasilnikova, L.A., DeRuff, K.C., Rudy, M., Bauer, M.R., Lagerborg, K.A., Normandin, E., Chapman, S.B., Reilly, S.K., Anahtar, M.N., Lin, A.E., Carter, A., Myhrvold, C., Kemball, M.E., Chaluvadi, S., Cusick, C., Flowers, K., Neumann, A., Cerrato, F., Farhat, M., Slater, D., Harris, J.B., Branda, J.A., Hooper, D., Gaeta, J.M., Baggett, T.P., O'Connell, J., Gnirke, A., Lieberman, T.D., Philippakis, A., Burns, M., Brown, C.M., Luban, J., Ryan, E.T., Turbett, S.E., LaRocque, R.C., Hanage, W.P., Gallagher, G.R., Madoff, L.C., Smole, S., Pierce, V.M., Rosenberg, E., Sabeti, P.C., Park, D.J., MacInnis, B.L., 2021. Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. Science 371. https://doi.org/10.1126/science.abe3261.

Letko, M., Marzi, A., Munster, V., 2020. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nat. Microbiol. 5, 562–569.

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. Euro. Surveill. 26. https://doi.org/10.2807/1560-7917.ES.2020.26.1.2002106.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S.M., Lau, E.H.Y., Wong, J.Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., Tu, W., Chen, C., Jin, L., Yang, R., Wang, Q., Zhou, S., Wang, R., Liu, H., Luo, Y., Liu, Y., Shao, G., Li, H., Tao, Z., Yang, Y., Deng, Z., Liu, B., Ma, Z., Zhang, Y., Shi, G., Lam, T.T.Y., Wu, J.T., Gao, G.F., Cowling, B.J., Yang, B., Leung, G.M., Feng, Z., 2020a. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N. Engl. J. Med. 382, 1199–1207.

Li, Q., Wu, J., Nie, J., Zhang, L., Hao, H., Liu, S., Zhao, C., Zhang, Q., Liu, H., Nie, L., Qin, H., Wang, M., Lu, Q., Li, X., Sun, Q., Liu, J., Zhang, L., Li, X., Huang, W., Wang, Y., 2020b. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. Cell 182, 1284–1294 e9.

Li, Q., Zheng, X.-S., Shen, X.-R., Si, H.-R., Wang, X., Wang, Q., Li, B., Zhang, W., Zhu, Y., Jiang, R.-D., Zhao, K., Wang, H., Shi, Z.-L., Zhang, H.-L., Du, R.-H., Zhou, P., 2020c. Prolonged shedding of severe acute respiratory syndrome coronavirus 2 in patients with COVID-19. Emerg. Microbes Infect. 9, 2571–2577.

Liu, Y., Eggo, R.M., Kucharski, A.J., 2020. Secondary attack rate and superspreading events for SARS-CoV-2. Lancet 395, e47.

Liu, Z., VanBlargan, L.A., Bloyet, L.-M., Rothlauf, P.W., Chen, R.E., Stumpf, S., Zhao, H., Errico, J.M., Theel, E.S., Liebeskind, M.J., Alford, B., Buchser, W.J., Ellebedy, A.H., Fremont, D.H., Diamond, M.S., Whelan, S.P.J., 2021. 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.

Long, Q.-X., Liu, B.-Z., Deng, H.-J., Wu, G.-C., Deng, K., Chen, Y.-K., Liao, P., Qiu, J.-F., Lin, Y., Cai, X.-F., Wang, D.-Q., Hu, Y., Ren, J.-H., Tang, N., Xu, Y.-Y., Yu, L.-H., Mo, Z., Gong, F., Zhang, X.-L., Tian, W.-G., Hu, L., Zhang, X.-X., Xiang, J.-L., Du, H.-X., Liu, H.-W., Lang, C.-H., Luo, X.-H., Wu, S.-B., Cui, X.-P., Zhou, Z., Zhu, M.-M., Wang, J., Xue, C.-J., Li, X.-F., Wang, L., Li, Z.-J., Wang, K., Niu, C.-C., Yang, Q.-J., Tang, X.-J., Zhang, Y., Liu, X.-M., Li, J.-J., Zhang, D.-C., Zhang, F., Liu, P., Yuan, J., Li, Q., Hu, J.-L., Chen, J., Huang, A.-L., 2020. Antibody responses to SARS-CoV-2 in patients with COVID-19. Nat. Med. 26, 845–848.

Lythgoe, K.A., Hall, M., Ferretti, L., de Cesare, M., MacIntyre-Cockett, G., Trebes, A., Andersson, M., Otecko, N., Wise, E.L., Moore, N., Lynch, J., Kidd, S., Cortes, N., Mori, M., Williams, R., Vernet, G., Justice, A., Green, A., Nicholls, S.M., Ansari, M.A., Abeler-Dörner, L., Moore, C.E., Peto, T.E.A., Eyre, D.W., Shaw, R., Simmonds, P., Buck, D., Todd, J.A., Oxford Virus Sequencing Analysis Group (OVSG), Connor, T.R., Ashraf, S., da Silva Filipe, A., Shepherd, J., Thomson, E.C., COVID-19 Genomics UK (COG-UK) Consortium, Bonsall, D., Fraser, C., Golubchik, T., 2021. SARS-CoV-2 within-host diversity and transmission. Science. https://doi.org/10.1126/science.abg0821.

MacLean, O.A., Orton, R.J., Singer, J.B., Robertson, D.L., 2020. No evidence for distinct types in the evolution of SARS-CoV-2. Virus Evol. 6, veaa034.

MacLean, O.A., Lytras, S., Weaver, S., Singer, J.B., Boni, M.F., Lemey, P., Kosakovsky Pond, S.L., Robertson, D.L., 2021. Natural selection in the evolution of SARS-CoV-2 in bats created a generalist virus and highly capable human pathogen. PLoS Biol. 19, e3001115.

Madhi, S.A., Baillie, V., Cutland, C.L., Voysey, M., Koen, A.L., Fairlie, L., Padayachee, S.D., Dheda, K., Barnabas, S.L., Bhorat, Q.E., Briner, C., Kwatra, G., Ahmed, K., Aley, P., Bhikha, S., Bhiman, J.N., Bhorat, A.E., du Plessis, J., Esmail, A., Groenewald, M., Horne, E., Hwa, S.-H., Jose, A., Lambe, T., Laubscher, M., Malahleha, M., Masenya, M., Masilela, M., McKenzie, S., Molapo, K., Moultrie, A., Oelofse, S., Patel, F., Pillay, S., Rhead, S., Rodel, H., Rossouw, L., Taoushanis, C., Tegally, H., Thombrayil, A., van Eck, S., Wibmer, C.K., Durham, N.M., Kelly, E.J., Villafana, T.L., Gilbert, S., Pollard, A.J., de Oliveira, T., Moore, P.L., Sigal, A., Izu, A., NGS-SA Group Wits–VIDA COVID Group, 2021. Efficacy of the ChAdOx1 nCoV-19 Covid-19 vaccine against the B.1.351 variant. N. Engl. J. Med. https://doi.org/10.1056/NEJMoa2102214.

Majra, D., Benson, J., Pitts, J., Stebbing, J., 2021. SARS-CoV-2 (COVID-19) superspreader events. J. Infect. 82, 36–40.

Martin, D.P., Weaver, S., Tegally, H., San, E.J., Shank, S.D., Wilkinson, E., Giandhari, J., Naidoo, S., Pillay, Y., Singh, L., Lessells, R.J., NGS-SA, COVID-19 Genomics UK (COG-UK), Gupta, R.K., Wertheim, J.O., Nekturenko, A., Murrell, B., Harkins, G.W., Lemey, P., MacLean, O.A., Robertson, D.L., de Oliveira, T., Pond, S.L.K., 2021. The emergence and ongoing convergent evolution of the N501Y lineages coincides with a major global shift in the SARS-CoV-2 selective landscape. medRxiv. https://doi.org/10.1101/2021.02.23.21252268.

Mastriani, E., Rakov, A.V., Liu, S.-L., 2021. Isolating SARS-CoV-2 strains from countries in the same meridian: genome evolutionary analysis. JMIR Bioinform. Biotech. 2, e25995.

McCallum, M., Bassi, J., Marco, A.D., Chen, A., Walls, A.C., Iulio, J.D., Tortorici, M.A., Navarro, M.-J., Silacci-Fregni, C., Saliba, C., Agostini, M., Pinto, D., Culap, K., Bianchi, S., Jaconi, S., Cameroni, E., Bowen, J.E., Tilles, S.W., Pizzuto, M.S., Guastalla, S.B., Bona, G., Pellanda, A.F., Garzoni, C., Van Voorhis, W.C., Rosen, L.E., Snell, G., Telenti, A., Virgin, H.W., Piccoli, L., Corti, D., Veesler, D., 2021a. SARS-CoV-2 immune evasion by variant B.1.427/B.1.429. bioRxiv. https://doi.org/10.1101/2021.03.31.437925.

McCallum, M., De Marco, A., Lempp, F.A., Tortorici, M.A., Pinto, D., Walls, A.C., Beltramello, M., Chen, A., Liu, Z., Zatta, F., Zepeda, S., di Iulio, J., Bowen, J.E., Montiel-Ruiz, M., Zhou, J., Rosen, L.E., Bianchi, S., Guarino, B., Fregni, C.S., Abdelnabi, R., Foo, S.-Y.C., Rothlauf, P.W., Bloyet, L.-M., Benigni, F., Cameroni, E., Neyts, J., Riva, A., Snell, G., Telenti, A., Whelan, S.P.J., Virgin, H.W., Corti, D., Pizzuto, M.S., Veesler, D., 2021b. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. Cell. https://doi.org/10.1016/j.cell.2021.03.028.

McMinn, P., Carrello, A., Cole, C., Baker, D., Hampson, A., 1999. Antigenic drift in influenza A (H3N2) virus in a persistently infected immunocompromised host is similar to that occurring in the community. Clin. Infect. Dis. 29, 456–458.

Memoli, M.J., Hrabal, R.J., Hassantoufighi, A., Jagger, B.W., Sheng, Z.-M., Eichelberger, M.C., Taubenberger, J.K., 2010. Rapid selection of a transmissible multidrug-resistant influenza A/H3N2 virus in an immunocompromised host. J. Infect. Dis. 201, 1397–1403.

Moreno, G., Braun, K., Larsen, B.B., Alpert, T., Worobey, M., Grubaugh, N., Friedrich, T., O'Connor, D., Fauver, J., Brito, A., 2021. Detection of non-B.1.1.7 Spike Δ69/70 sequences (B.1.375) in the United States. Virological.org. https://virological.org/t/detection-of-non-b-1-1-7-spike-69-70-sequences-b-1-375-in-the-united-states/587.

Muik, A., Wallisch, A.-K., Sänger, B., Swanson, K.A., Mühl, J., Chen, W., Cai, H., Maurus, D., Sarkar, R., Türeci, Ö., Dormitzer, P.R., Şahin, U., 2021. Neutralization of SARS-CoV-2 lineage B.1.1.7 pseudovirus by BNT162b2 vaccine-elicited human sera. Science 371, 1152–1153.

Mulligan, M.J., Lyke, K.E., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Neuzil, K., Raabe, V., Bailey, R., Swanson, K.A., Li, P., Koury, K., Kalina, W., Cooper, D., Fontes-Garfias, C., Shi, P.-Y., Türeci, Ö., Tompkins, K.R., Walsh, E.E., Frenck, R., Falsey, A.R., Dormitzer, P.R., Gruber, W.C., Şahin, U., Jansen, K.U., 2020. Phase I/II study of COVID-19 RNA vaccine BNT162b1 in adults. Nature 586, 589–593.

Naveca, F., da Costa, C., Nascimento, V., Souza, V., Corado, A., Nascimento, F., Costa, A., Duarte, D., Silva, G., Mejía, M., Pessoa, K., Gonçalves, L., Brandão, M.J., Jesus, M., Pinto, R., Silva, M., Mattos, T., Abdalla, L., Santos, J.H., Costa-Filho, R., Wallau, G.L., Siqueira, M.M., Delatorre, E., Gräf, T., Bello, G., Resende, P.C., 2021. SARS-CoV-2 reinfection by the new Variant of Concern (VOC) P.1 in Amazonas, Brazil. Virological.org. https://virological.org/t/sars-cov-2-reinfection-by-the-new-variant-of-concern-voc-p-1-in-amazonas-brazil/596.

Olvera, A., Noguera-Julian, M., Kilpelainen, A., Romero-Martín, L., Prado, J.G., Brander, C., 2020. SARS-CoV-2 consensus-sequence and matching overlapping peptides design for COVID19 immune studies and vaccine development. Vaccines (Basel) 8. https://doi.org/10.3390/vaccines8030444.

Oude Munnink, B.B., Nieuwenhuijse, D.F., Stein, M., O'Toole, Á., Haverkate, M., Mollers, M., Kamga, S.K., Schapendonk, C., Pronk, M., Lexmond, P., van der Linden, A., Bestebroer, T., Chestakova, I., Overmars, R.J., van Nieuwkoop, S., Molenkamp, R., van der Eijk, A.A., GeurtsvanKessel, C., Vennema, H., Meijer, A., Rambaut, A., van Dissel, J., Sikkema, R.S., Timen, A., Koopmans, M., Dutch-Covid-19 response team, 2020. Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. Nat. Med. 26, 1405–1410.

Oude Munnink, B.B., Sikkema, R.S., Nieuwenhuijse, D.F., Molenaar, R.J., Munger, E., Molenkamp, R., van der Spek, A., Tolsma, P., Rietveld, A., Brouwer, M., Bouwmeester-Vincken, N., Harders, F., Hakze-van der Honing, R., Wegdam-Blans, M.C.A., Bouwstra, R.J., GeurtsvanKessel, C., van der Eijk, A.A., Velkers, F.C., Smit, L.A.M., Stegeman, A., van der Poel, W.H.M., Koopmans, M.P.G., 2021. Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. Science 371, 172–177.

Plante, J.A., Liu, Y., Liu, J., Xia, H., Johnson, B.A., Lokugamage, K.G., Zhang, X., Muruato, A.E., Zou, J., Fontes-Garfias, C.R., Mirchandani, D., Scharton, D., Bilello, J.P., Ku, Z., An, Z., Kalveram, B., Freiberg, A.N., Menachery, V.D., Xie, X., Plante, K.S., Weaver, S.C., Shi, P.-Y., 2020. Spike mutation D614G alters SARS-CoV-2 fitness. Nature. https://doi.org/10.1038/s41586-020-2895-3.

Polack, F.P., Thomas, S.J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J.L., Pérez Marc, G., Moreira, E.D., Zerbini, C., Bailey, R., Swanson, K.A., Roychoudhury, S., Koury, K., Li, P., Kalina, W.V., Cooper, D., Frenck Jr., R.W., Hammitt, L.L., Türeci, Ö., Nell, H., Schaefer, A., Ünal, S., Tresnan, D.B., Mather, S., Dormitzer, P.R., Şahin, U., Jansen, K.U., Gruber, W.C., C4591001 Clinical Trial Group, 2020. Safety and efficacy of the BNT162b2 mRNA covid-19 vaccine. N. Engl. J. Med. 383, 2603–2615.

Popa, A., Genger, J.-W., Nicholson, M.D., Penz, T., Schmid, D., Aberle, S.W., Agerer, B., Lercher, A., Endler, L., Colaço, H., Smyth, M., Schuster, M., Grau, M.L., Martínez-Jiménez, F., Pich, O., Borena, W., Pawelka, E., Keszei, Z., Senekowitsch, M., Laine, J., Aberle, J.H., Redlberger-Fritz, M., Karolyi, M., Zoufaly, A., Maritschnik, S., Borkovec, M., Hufnagl, P., Nairz, M., Weiss, G., Wolfinger, M.T., von Laer, D., Superti-Furga, G., Lopez-Bigas, N., Puchhammer-Stöckl, E., Allerberger, F., Michor, F., Bock, C., Bergthaler, A., 2020. Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. Sci. Transl. Med. 12 https://doi.org/10.1126/scitranslmed.abe2555.

Pybus, O.G., 2021. Pango Lineage Nomenclature: provisional rules for naming recombinant lineages. Virological.org. https://virological.org/t/pango-lineage-nomenclature-provisional-rules-for-naming-recombinant-lineages/657.

Rambaut, A., Holmes, E.C., O'Toole, Á., Hill, V., McCrone, J.T., Ruis, C., du Plessis, L., Pybus, O.G., 2020a. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat. Microbiol. 5, 1403–1407.

Rambaut, A., Loman, N.J., Pybus, O.G., Barclay, W., Barrett, J., Carabelli, A., Peacock, T., Robertson, D.L., Volz, E., COVID-19 Genomics Consortium UK (CoG-UK)., 2020b. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. Virological.org. 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.

Resende, P.C., Delatorre, E., Gräf, T., Mir, D., Motta, F.C., Appolinario, L.R., da Paixão, A.C.D., Mendonça, A.C. da F., Ogrzewalska, M., Caetano, B., Wallau, G.L., Docena, C., Dos Santos, M.C., de Almeida Ferreira, J., Sousa Junior, E.C., da Silva, S.P., Fernandes, S.B., Vianna, L.A., Souza, L. da C., Ferro, J.F.G., Nardy, V.B., Santos, C.A., Riediger, I., do Carmo Debur, M., Croda, J., Oliveira, W.K., Abreu, A., Bello, G., Siqueira, M.M., 2020. Evolutionary dynamics and dissemination pattern of the SARS-CoV-2 lineage B.1.1.33 during the early pandemic phase in Brazil. Front. Microbiol. 11, 615280.

Resende, P.C., Bezerra, J.P., Teixeira de Vasconcelos, R.H., Arantes, I., Appolinario, L., Mendonça, A.C., Paixao, A.C., Duarte Rodrigues, A.C., Silva, T., Sampaio Rocha, A., Pauvolid-Corrêa, A., Couto Motta, F., Teixeira, D.L.F., de Oliveira Carneiro, T.F., Freire Neto, F.P., Diniz Herbster, I., Leite, A.B., Riediger, I.N., do Carmo Debur, M., Naveca, F.G., Almeida, W., Livorati, M., Bello, G., Siqueira, M.M., 2021. Spike E484K mutation in the first SARS-CoV-2 reinfection case confirmed in Brazil. Virological.org. https://virological.org/t/spike-e484k-mutation-in-the-first-sars-cov-2-reinfection-case-confirmed-in-brazil-2020/584.

Rogers, M.B., Song, T., Sebra, R., Greenbaum, B.D., Hamelin, M.-E., Fitch, A., Twaddle, A., Cui, L., Holmes, E.C., Boivin, G., Ghedin, E., 2015. Intrahost dynamics of antiviral resistance in influenza A virus reflect complex patterns of segment linkage, reassortment, and natural selection. MBio 6. https://doi.org/10.1128/mBio.02464-14.

Russi, R.C., Bourdin, E., García, M.I., Veaute, C.M.I., 2018. In silico prediction of T- and B-cell epitopes in PmpD: First step towards the design of a Chlamydia trachomatis vaccine. Biomed. J. 41, 109–117.

Rydyznski Moderbacher, C., Ramirez, S.I., Dan, J.M., Grifoni, A., Hastie, K.M., Weiskopf, D., Belanger, S., Abbott, R.K., Kim, C., Choi, J., Kato, Y., Crotty, E.G., Kim, C., Rawlings, S.A., Mateus, J., Tse, L.P.V., Frazier, A., Baric, R., Peters, B., Greenbaum, J., Ollmann Saphire, E., Smith, D.M., Sette, A., Crotty, S., 2020. Antigen-specific adaptive immunity to SARS-CoV-2 in acute COVID-19 and associations with age and disease severity. Cell 183, 996–1012 e19.

Sabino, E.C., Buss, L.F., Carvalho, M.P.S., Prete Jr., C.A., Crispim, M.A.E., Fraiji, N.A., Pereira, R.H.M., Parag, K.V., da Silva Peixoto, P., Kraemer, M.U.G., Oikawa, M.K., Salomon, T., Cucunuba, Z.M., Castro, M.C., de Souza Santos, A.A., Nascimento, V.H., Pereira, H.S., Ferguson, N.M., Pybus, O.G., Kucharski, A., Busch, M.P., Dye, C., Faria, N.R., 2021. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. Lancet. https://doi.org/10.1016/S0140-6736(21)00183-5.

Sahin, U., Muik, A., Vogler, I., Derhovanessian, E., Kranz, L.M., Vormehr, M., Quandt, J., Bidmon, N., Ulges, A., Baum, A., Pascal, K., Maurus, D., Brachtendorf, S., Lörks, V., Sikorski, J., Koch, P., Hilker, R., Becker, D., Eller, A.-K., Grützner, J., Tonigold, M., Boesler, C., Rosenbaum, C., Heesen, L., Kühnle, M.-C., Poran, A., Dong, J.Z., Luxemburger, U., Kemmer-Brück, A., Langer, D., Bexon, M., Bolte, S., Palanche, T., Schultz, A., Baumann, S., Mahiny, A.J., Boros, G., Reinholz, J., Szabó, G.T., Karikó, K., Shi, P.-Y., Fontes-Garfias, C., Perez, J.L., Cutler, M., Cooper, D., Kyratsous, C.A., Dormitzer, P.R., Jansen, K.U., Türeci, Ö., 2020. COVID-19 vaccine BNT162b1 elicits human antibody and TH1 T cell responses. Nature. 586, 594–599.

Seow, J., Graham, C., Merrick, B., Acors, S., Pickering, S., Steel, K.J.A., Hemmings, O., O'Byrne, A., Kouphou, N., Galao, R.P., Betancor, G., Wilson, H.D., Signell, A.W., Winstone, H., Kerridge, C., Huettner, I., Jimenez-Guardeño, J.M., Lista, M.J., Temperton, N., Snell, L.B., Bisnauthsing, K., Moore, A., Green, A., Martinez, L., Stokes, B., Honey, J., Izquierdo-Barras, A., Arbane, G., Patel, A., Tan, M.K.I., O'Connell, L., O'Hara, G., MacMahon, E., Douthwaite, S., Nebbia, G., Batra, R., Martinez-Nunez, R., Shankar-Hari, M., Edgeworth, J.D., Neil, S.J.D., Malim, M.H., Doores, K.J., 2020. Longitudinal observation and decline of neutralizing antibody responses in the three months following SARS-CoV-2 infection in humans. Nat. Microbiol. 5, 1598–1607.

Sette, A., Crotty, S., 2021. Adaptive immunity to SARS-CoV-2 and COVID-19. Cell. https://doi.org/10.1016/j.cell.2021.01.007.

Sharma, O., Sultan, A.A., Ding, H., Triggle, C.R., 2020. A review of the progress and challenges of developing a vaccine for COVID-19. Front. Immunol. 11, 585354.

Sheahan, T.P., Frieman, M.B., 2020. The continued epidemic threat of SARS-CoV-2 and implications for the future of global public health. Curr. Opin. Virol. 40, 37–40.

Siqueira, J.D., Goes, L.R., Alves, B.M., de Carvalho, P.S., Cicala, C., Arthos, J., Viola, J.P. B., de Melo, A.C., Soares, M.A., 2020. SARS-CoV-2 genomic and quasispecies analyses in cancer patients reveal relaxed intrahost virus evolution. bioRxiv. https://doi.org/10.1101/2020.08.26.267831.

Sironi, M., Hasnain, S.E., Rosenthal, B., Phan, T., Luciani, F., Shaw, M.-A., Sallum, M.A., Mirhashemi, M.E., Morand, S., González-Candelas, F., Editors of Infection, Genetics and Evolution, 2020. SARS-CoV-2 and COVID-19: A genetic, epidemiological, and evolutionary perspective. Infect. Genet. Evol. 84, 104384.

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

Spanakis, N., Kassela, K., Dovrolis, N., Bampali, M., Gatzidou, E., Kafasi, A., Froukala, E., Stavropoulou, A., Lilakos, K., Veletza, S., Tsiodras, S., Tsakris, A., Karakasiliotis, I., 2021. A main event and multiple introductions of SARS-Cov2 initiated the COVID-19 epidemic in Greece. J. Med. Virol. https://doi.org/10.1002/jmv.26778.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Starr, T.N., Greaney, A.J., Addetia, A., Hannon, W.W., Choudhary, M.C., Dingens, A.S., Li, J.Z., Bloom, J.D., 2021. Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. Science. https://doi.org/10.1126/science.abf9302. Epub 2021 Jan 25.

Surleac, M., Banica, L., Casangiu, C., Cotic, M., Florea, D., Sandulescu, O., Milu, P., Streinu-Cercel, A., Vlaicu, O., Paraskevis, D., Paraschiv, S., Otelea, D., 2020. Molecular epidemiology analysis of SARS-CoV-2 strains circulating in Romania during the first months of the pandemic. Life 10. https://doi.org/10.3390/life10080152.

Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., Duan, Y., Zhang, H., Wang, Y., Qian, Z., Cui, J., Lu, J., 2020. On the origin and continuing evolution of SARS-CoV-2. Natl. Sci. Rev. https://doi.org/10.1093/nsr/nwaa036.

Tarke, A., Sidney, J., Methot, N., Zhang, Y., Dan, J.M., Goodwin, B., Rubiro, P., Sutherland, A., da Silva Antunes, R., Frazier, A., Rawlings, S.A., Smith, D.M., Peters, B., Scheuermann, R.H., Weiskopf, D., Crotty, S., Grifoni, A., Sette, A., 2021. Negligible impact of SARS-CoV-2 variants on CD4 + and CD8 + T cell reactivity in COVID-19 exposed donors and vaccinees. bioRxivorg. https://doi.org/10.1101/2021.02.27.433180.

Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., Doolabh, D., Pillay, S., San, E.J., Msomi, N., Mlisana, K., von Gottberg, A., Walaza, S., Allam, M., Ismail, A., Mohale, T., Glass, A.J., Engelbrecht, S., Van Zyl, G., Preiser, W., Petruccione, F., Sigal, A., Hardie, D., Marais, G., Hsiao, M., Korsman, S., Davies, M.-A., Tyers, L., Mudau, I., York, D., Maslo, C., Goedhals, D., Abrahams, S., Laguda-Akingba, O., Alisoltani-Dehkordi, A., Godzik, A., Wibmer, C.K., Sewell, B.T., Lourenço, J., Alcantara, L.C.J., Kosakovsky Pond, S.L., Weaver, S., Martin, D., Lessells, R.J., Bhiman, J.N., Williamson, C., de Oliveira, T., 2020. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. medRxiv, 2020.12.21.20248640, still in press.

Tegally, H., Wilkinson, E., Lessells, R.J., Giandhari, J., Pillay, S., Msomi, N., Mlisana, K., Bhiman, J.N., von Gottberg, A., Walaza, S., Fonseca, V., Allam, M., Ismail, A., Glass, A.J., Engelbrecht, S., Van Zyl, G., Preiser, W., Williamson, C., Petruccione, F.,

Sigal, A., Gazy, I., Hardie, D., Hsiao, N.-Y., Martin, D., York, D., Goedhals, D., San, E. J., Giovanetti, M., Lourenço, J., Alcantara, L.C.J., de Oliveira, T., 2021. Sixteen novel lineages of SARS-CoV-2 in South Africa. Nat. Med. https://doi.org/10.1038/s41591-021-01255-3.

The Lancet, 2021. Genomic sequencing in pandemics. Lancet 397, 445.

Thomson, E.C., Rosen, L.E., Shepherd, J.G., Spreafico, R., da Silva Filipe, A., Wojcechowskyj, J.A., Davis, C., Piccoli, L., Pascall, D.J., Dillen, J., Lytras, S., Czudnochowski, N., Shah, R., Meury, M., Jesudason, N., De Marco, A., Li, K., Bassi, J., O'Toole, A., Pinto, D., Colquhoun, R.M., Culap, K., Jackson, B., Zatta, F., Rambaut, A., Jaconi, S., Sreenu, V.B., Nix, J., Zhang, I., Jarrett, R.F., Glass, W.G., Beltramello, M., Nomikou, K., Pizzuto, M., Tong, L., Cameroni, E., Croll, T.I., Johnson, N., Di Iulio, J., Wickenhagen, A., Ceschi, A., Harbison, A.M., Mair, D., Ferrari, P., Smollett, K., Sallusto, F., Carmichael, S., Garzoni, C., Nichols, J., Galli, M., Hughes, J., Riva, A., Ho, A., Schiuma, M., Semple, M.G., Openshaw, P.J.M., Fadda, E., Baillie, J.K., Chodera, J.D., ISARIC4C Investigators, COVID-19 Genomics UK (COG-UK) Consortium, Rihn, S.J., Lycett, S.J., Virgin, H.W., Telenti, A., Corti, D., Robertson, D.L., Snell, G., 2021. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. Cell 184, 1171–1187 e20.

Tomaszewski, T., DeVries, R.S., Dong, M., Bhatia, G., Norsworthy, M.D., Zheng, X., Caetano-Anollés, G., 2020. New pathways of mutational change in SARS-CoV-2 proteomes involve regions of intrinsic disorder important for virus replication and release. Evol. Bioinform. Online 16, 1176934320965149.

Tomkins-Tinch, C.H., Silbert, J., DeRuff, K.C., Siddle, K.J., Gladden-Young, A., Bronson, A., Marshall, J., Bean, Z.O., Sholtes, K., McQuade, K., Kary, A., Krohn Bevill, K., King, C., Hughes, M., Leenerman, E., Parrie, E., Baron, E., Vodzak, M.E., Gionet, G., Adams, G., Pearlman, L., Sani, K., Specht, I., Colubri, A., Petros, B.A., Loreth, C., Lemieux, J.E., Schaffner, S.F., Luban, J., Schifferli, M., Cronan, P., Glennon, O., Oba, K., Fry, B., Doss, M., Park, D.J., MacInnis, B.L., Sabeti, P.C., 2021. Detection of the recurrent substitution Q677H in the spike protein of SARS-CoV-2 in cases descended from the lineage B.1.429. Virological.org. https://virological.org/t/detection-of-the-recurrent-substitution-q677h-in-the-spike-protein-of-sars-cov-2-in-cases-descended-from-the-lineage-b-1-429/660.

Tonkin-Hill, G., Martincorena, I., Amato, R., Lawson, A.R.J., Gerstung, M., Johnston, I., Jackson, D.K., Park, N.R., Lensing, S.V., Quail, M.A., Gonçalves, S., Ariani, C., Chapman, M.S., Hamilton, W.L., Meredith, L.W., Hall, G., Jahun, A.S., Chaudhry, Y., Hosmillo, M., Pinckert, M.L., Georgana, I., Yakovleva, A., Caller, L.G., Caddy, S.L., Feltwell, T., Khokhar, F.A., Houldcroft, C.J., Curran, M.D., Parmar, S., Alderton, A., Nelson, R., Harrison, E., Sillitoe, J., Bentley, S.D., Barrett, J.C., Torok, M.E., Goodfellow, I.G., Langford, C., Kwiatkowski, D., The COVID-19 Genomics UK (COG-UK) Consortium, Wellcome Sanger Institute COVID-19 Surveillance Team, 2020. Patterns of within-host genetic diversity in SARS-CoV-2. bioRxiv. https://doi.org/10.1101/2020.12.23.424229.

Valesano, A.L., Rumfelt, K.E., Dimcheff, D.E., Blair, C.N., Fitzsimmons, W.J., Petrie, J.G., Martin, E.T., Lauring, A.S., 2021. Temporal dynamics of SARS-CoV-2 mutation accumulation within and across infected hosts. bioRxiv. https://doi.org/10.1101/2021.01.19.427330.

van Doremalen, N., Lambe, T., Spencer, A., Belij-Rammerstorfer, S., Purushotham, J.N., Port, J.R., Avanzato, V.A., Bushmaker, T., Flaxman, A., Ulaszewska, M., Feldmann, F., Allen, E.R., Sharpe, H., Schulz, J., Holbrook, M., Okumura, A., Meade-White, K., Pérez-Pérez, L., Edwards, N.J., Wright, D., Bissett, C., Gilbride, C., Williamson, B.N., Rosenke, R., Long, D., Ishwarbhai, A., Kailath, R., Rose, L., Morris, S., Powers, C., Lovaglio, J., Hanley, P.W., Scott, D., Saturday, G., de Wit, E., Gilbert, S.C., Munster, V.J., 2020. ChAdOx1 nCoV-19 vaccine prevents SARS-CoV-2 pneumonia in rhesus macaques. Nature 586, 578–582.

van Dorp, L., Acman, M., Richard, D., Shaw, L.P., Ford, C.E., Ormond, L., Owen, C.J., Pang, J., Tan, C.C.S., Boshier, F.A.T., Ortiz, A.T., Balloux, F., 2020a. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. Infect. Genet. Evol. 83, 104351.

van Dorp, L., Richard, D., Tan, C.C.S., Shaw, L.P., Acman, M., Balloux, F., 2020b. No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. Nat. Commun. 11, 5986.

van Dorp, L., Tan, C.C.S., Lam, S.D., Richard, D., Owen, C., Berchtold, D., Orengo, C., Balloux, F., 2020c. Recurrent mutations in SARS-CoV-2 genomes isolated from mink point to rapid host-adaptation. Cold Spring Harbor Laboratory. https://doi.org/10.1101/2020.11.16.384743.

Van Regenmortel, M.H.V., 2007. Virus species and virus identification: past and current controversies. Infect. Genet. Evol. 7, 133–144.

VanInsberghe, D., Neish, A.S., Lowen, A.C., Koelle, K., 2021. Recombinant SARS-CoV-2 genomes are currently circulating at low levels. bioRxiv. https://doi.org/10.1101/2020.08.05.238386.

Varghese, P.M., Tsolaki, A.G., Yasmin, H., Shastri, A., Ferluga, J., Vatish, M., Madan, T., Kishore, U., 2020. Host-pathogen interaction in COVID-19: Pathogenesis, potential therapeutics and vaccination strategies. Immunobiology 225, 152008.

Vasques Nonaka, C.K., Miranda Franco, M., Gräf, T., Almeida Mendes, A.V., Santana de Aguiar, R., Giovanetti, M., Solano de Freitas Souza, B., 2021. Genomic evidence of a SARS-CoV-2 reinfection case with E484K spike mutation in Brazil. Emerging Infectious Diseases 27 (5), 1522–1524. https://doi.org/10.3201/eid2705.210191.

Vijgen, L., Keyaerts, E., Moës, E., Thoelen, I., Wollants, E., Lemey, P., Vandamme, A.-M., Van Ranst, M., 2005. Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. J. Virol. 79, 1595–1604.

Vöhringer, H., Sinnott, M., Amato, R., Martincorena, I., Kwiatkowski, D.P., Barrett, J.C., Gerstung, M., The COVID-19 Genomics UK (COG-UK) consortium, 2020. Lineage-specific growth of SARS-CoV-2 B.1.1.7 during the English national lockdown.

Virological.org. https://virological.org/t/lineage-specific-growth-of-sars-cov-2-b-1-1-7-during-the-english-national-lockdown/575.

Volz, E., Swapnil Mishra, S., Chand, M., Barrett, J.C., Johnson, R., Geidelberg, L., Hinsley, W.R., Laydon, D.J., Dabrera, G., O'Toole, A., Amato, R., Ragonnet-Cronin, M., Harrison, I., Jackson, B., Ariani, C.V., Boyd, O., Loman, N.J., McCrone, J. T., Gonçalves, S., Jorgensen, D., Myers, R., Hill, V., Jackson, D.J., Gaythorpe, K., Groves, N., Sillitoe, J., Kwiatkowski, D.P., The COVID-19 Genomics UK (COG-UK) consortium, Flaxman, S., Ratmann, O., Bhatt, S., Hopkins, S., Gandy, A., Rambaut, A., Ferguson, N.M., 2020. Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. Virological.org. https://virological.org/t/transmission-of-sars-cov-2-lineage-b-1-1-7-in-england-i nsights-from-linking-epidemiological-and-genetic-data/576.

Volz, E., Hill, V., McCrone, J.T., Price, A., Jorgensen, D., O'Toole, Á., Southgate, J., Johnson, R., Jackson, B., Nascimento, F.F., Rey, S.M., Nicholls, S.M., Colquhoun, R. M., da Silva Filipe, A., Shepherd, J., Pascall, D.J., Shah, R., Jesudason, N., Li, K., Jarrett, R., Pacchiarini, N., Bull, M., Geidelberg, L., Siveroni, I., COG-UK Consortium, Goodfellow, I., Loman, N.J., Pybus, O.G., Robertson, D.L., Thomson, E.C., Rambaut, A., Connor, T.R., 2021. Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. Cell 184, 64–75 e11.

Walsh, E.E., Frenck Jr., R.W., Falsey, A.R., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Neuzil, K., Mulligan, M.J., Bailey, R., Swanson, K.A., Li, P., Koury, K., Kalina, W., Cooper, D., Fontes-Garfias, C., Shi, P.-Y., Türeci, Ö., Tompkins, K.R., Lyke, K.E., Raabe, V., Dormitzer, P.R., Jansen, K.U., Şahin, U., Gruber, W.C., 2020. Safety and immunogenicity of two RNA-based covid-19 vaccine candidates. N. Engl. J. Med. 383, 2439–2450.

Wang, Z., Schmidt, F., Weisblum, Y., Muecksch, F., Barnes, C.O., Finkin, S., Schaefer-Babajew, D., Cipolla, M., Gaebler, C., Lieberman, J.A., Yang, Z., Abernathy, M.E., Huey-Tubman, K.E., Hurley, A., Turroja, M., West, K.A., Gordon, K., Millard, K.G., Ramos, V., Da Silva, J., Xu, J., Colbert, R.A., Patel, R., Dizon, J., Unson-O'Brien, C., Shimeliovich, I., Gazumyan, A., Caskey, M., Bjorkman, P.J., Casellas, R., Hatziioannou, T., Bieniasz, P.D., Nussenzweig, M.C., 2021. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. bioRxiv. https://doi.org/10.1101/2021.01.15.426911.

Wells, H.L., Letko, M., Lasso, G., Ssebide, B., Nziza, J., Byarugaba, D.K., Navarrete-Macias, I., Liang, E., Cranfield, M., Han, B.A., Tingley, M.W., Diuk-Wasser, M., Goldstein, T., Johnson, C.K., Mazet, J.A.K., Chandran, K., Munster, V.J., Gilardi, K., Anthony, S.J., 2021. The evolutionary history of ACE2 usage within the coronavirus subgenus. Virus Evol. 7, veab007.

Wibmer, C.K., Ayres, F., Hermanus, T., Madzivhandila, M., Kgagudi, P., Lambson, B.E., Vermeulen, M., van den Berg, K., Rossouw, T., Boswell, M., Ueckermann, V., Meiring, S., von Gottberg, A., Cohen, C., Morris, L., Bhiman, J.N., Moore, P.L., 2021. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. bioRxiv. https://doi.org/10.1101/2021.01.18.427166.

Williamson, E.J., Walker, A.J., Bhaskaran, K., Bacon, S., Bates, C., Morton, C.E., Curtis, H.J., Mehrkar, A., Evans, D., Inglesby, P., Cockburn, J., McDonald, H.I., MacKenna, B., Tomlinson, L., Douglas, I.J., Rentsch, C.T., Mathur, R., Wong, A.Y.S., Grieve, R., Harrison, D., Forbes, H., Schultze, A., Croker, R., Parry, J., Hester, F., Harper, S., Perera, R., Evans, S.J.W., Smeeth, L., Goldacre, B., 2020. Factors associated with COVID-19-related death using OpenSAFELY. Nature 584, 430–436.

Worobey, M., Pekar, J., Larsen, B.B., Nelson, M.I., Hill, V., Joy, J.B., Rambaut, A., Suchard, M.A., Wertheim, J.O., Lemey, P., 2020. The emergence of SARS-CoV-2 in Europe and North America. Science 370, 564–570.

Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.-L., Abiona, O., Graham, B. S., McLellan, J.S., 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 367, 1260–1263.

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E.C., Zhang, Y.-Z., 2020. A new coronavirus associated with human respiratory disease in China. Nature 579, 265–269.

Wu, K., Werner, A.P., Moliva, J.I., Koch, M., Choi, A., Stewart-Jones, G.B.E., Bennett, H., Boyoglu-Barnum, S., Shi, W., Graham, B.S., Carfi, A., Corbett, K.S., Seder, R.A., Edwards, D.K., 2021. mRNA-1273 vaccine induces neutralizing antibodies against spike mutants from global SARS-CoV-2 variants. bioRxiv. https://doi.org/10.1101/2021.01.25.427948.

Xiao, K., Zhai, J., Feng, Y., Zhou, N., Zhang, X., Zou, J.-J., Li, N., Guo, Y., Li, X., Shen, X., Zhang, Z., Shu, F., Huang, W., Li, Y., Zhang, Z., Chen, R.-A., Wu, Y.-J., Peng, S.-M., Huang, M., Xie, W.-J., Cai, Q.-H., Hou, F.-H., Chen, W., Xiao, L., Shen, Y., 2020. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. Nature 583, 286–289.

Xie, X., Zou, J., Fontes-Garfias, C.R., Xia, H., Swanson, K.A., Cutler, M., Cooper, D., Menachery, V.D., Weaver, S., Dormitzer, P.R., Shi, P.-Y., 2021. Neutralization of N501Y mutant SARS-CoV-2 by BNT162b2 vaccine-elicited sera. bioRxiv. https://doi.org/10.1101/2021.01.07.425740.

Young, B.E., Fong, S.W., Chan, Y.H., Mak, T.M., Ang, L.W., Anderson, D.E., Lee, C.Y., Amrun, S.N., Lee, B., Goh, Y.S., Su, Y.C.F., Wei, W.E., Kalimuddin, S., Chai, L.Y.A., Pada, S., Tan, S.Y., Sun, L., Parthasarathy, P., Chen, Y.Y.C., Barkham, T., Lin, R.T.P., Maurer-Stroh, S., Leo, Y.S., Wang, L.F., Renia, L., Lee, V.J., Smith, G.J.D., Lye, D.C., Ng, L.F.P., 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 (10251), 603–611. https://doi.org/10.1016/S0140-6736(20)31757-8.

Zaki, A.M., van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D.M.E., Fouchier, R.A.M., 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N. Engl. J. Med. 367, 1814–1820.

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., Zheng, X.-S., Zhao, K., Chen, Q.-J., Deng, F., Liu, L.-L., Yan, B., Zhan, F.-X., Wang, Y.-Y., Xiao, G.-F., Shi, Z.-L., 2020a. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579, 270–273.

Zhou, H., Chen, X., Hu, T., Li, J., Song, H., Liu, Y., Wang, P., Liu, D., Yang, J., Holmes, E. C., Hughes, A.C., Bi, Y., Shi, W., 2020b. 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, 3896.

Ziegler, K., Steininger, P., Ziegler, R., Steinmann, J., Korn, K., Ensser, A., 2020. SARS-CoV-2 samples may escape detection because of a single point mutation in the N gene. Euro Surveill. 25 https://doi.org/10.2807/1560-7917.ES.2020.25.39.2001650.