

SARS-CoV-2 one year on: evidence for ongoing viral adaptation

Thomas P. Peacock¹, Rebekah Penrice-Randal², Julian A. Hiscox^{2,3} and Wendy S. Barclay^{1,*}

Abstract

SARS-CoV-2 is thought to have originated in the human population from a zoonotic spillover event. Infection in humans results in a variety of outcomes ranging from asymptomatic cases to the disease COVID-19, which can have significant morbidity and mortality, with over two million confirmed deaths worldwide as of January 2021. Over a year into the pandemic, sequencing analysis has shown that variants of SARS-CoV-2 are being selected as the virus continues to circulate widely within the human population. The predominant drivers of genetic variation within SARS-CoV-2 are single nucleotide polymorphisms (SNPs) caused by polymerase error, potential host factor driven RNA modification, and insertion/deletions (indels) resulting from the discontinuous nature of viral RNA synthesis. While many mutations represent neutral 'genetic drift' or have quickly died out, a subset may be affecting viral traits such as transmissibility, pathogenicity, host range, and antigenicity of the virus. In this review, we summarise the current extent of genetic change in SARS-CoV-2, particularly recently emerging variants of concern, and consider the phenotypic consequences of this viral evolution that may impact the future trajectory of the pandemic.

INTRODUCTION

Towards the end of 2019, reports began of an unknown respiratory illness in the Chinese city of Wuhan. Within several weeks, it became clear these infections were being caused by a SARS-like coronavirus, which was termed SARS-CoV-2, with the associated disease called COVID-19. In severe cases this results in extensive immunopathology in the lungs [1]. By early March 2020, the virus had entered many countries across the world and the WHO declared a pandemic on 11 March [2]. In the months since, different countries across the world have enacted different pandemic response plans that vary from recurrent lockdowns, mask mandates, social distancing rules, or uncontrolled circulation in a hope to acquire herd immunity. In areas with elevated SARS-CoV-2 prevalence, high levels of morbidity and excess mortality, particularly in the elderly, has resulted. As of 13 March 2021, there have been an estimated 120 million confirmed cases of COVID-19 globally with over 2.6 million confirmed deaths [3].

SARS-CoV-2 is a betacoronavirus, containing a ~30 kb positive-sense RNA genome, among the largest of any RNA

virus (Fig. 1). Coronaviruses, such as SARS-CoV-2, avoid error catastrophe by encoding an exoribonuclease (nsp14) that confers a unique proofreading mechanism during viral RNA synthesis [4, 5]. Genome sequencing of SARS-CoV-2 throughout the course of the outbreak, has revealed a nucleotide substitution rate of $\sim 1 \times 10^{-3}$ substitutions per year [6]. This is comparable to the substitution rate observed for Ebola virus (1.42×10^{-3}) during the 2013–2016 West African outbreak [7]. However, SNPs are not the only genetic variation seen commonly in coronaviruses.

Replication of the coronavirus genome and transcription of viral subgenomic mRNAs (sgmRNAs) are complex processes. The genome is roughly organised into two regions. The first two thirds of the genome is immediately translated and proteolytically processed in the host cell cytoplasm to generate the viral polymerase/transcriptase complex and other viral proteins. The remaining one third of the genome is expressed and translated through a nested set of sgmRNAs, this includes the spike glycoprotein and other structural and accessory proteins. These sgmRNAs are 5' and 3' co-terminal with the genome; the 5' end contains a leader sequence that is present on the 5' end of the genome. Along the genome,

Received 24 January 2021; Accepted 16 March 2021; Published 15 April 2021

Author affiliations: ¹Department of Infectious Diseases, St Marys Medical School, Imperial College London, UK; ²Institute of Infection, Veterinary and Ecology Sciences, University of Liverpool, UK; ³A*STAR Infectious Diseases Laboratories (A*STAR ID Labs), Agency for Science, Technology and Research (A*STAR), Singapore.

***Correspondence:** Wendy S. Barclay, w.barclay@imperial.ac.uk

Keywords: SARS-CoV-2; COVID-19; coronavirus; mutant; adaptation; pandemic.

Abbreviations: ACE2, angiotensin-converting enzyme 2; ADAR, adenosine deaminases acting on RNA; APOBEC, apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like; COVID-19, coronavirus disease 2019; E, envelope; IFITM, interferon-induced transmembrane; NSP, non-structural protein; NTD, N-terminal domain; ORF, open reading frame; RBD, receptor binding domain; RdRp, RNA-dependent RNA polymerase; TMPRSS2, transmembrane serine protease 2; UTR, untranslated region; VOC, variant of concern; VUI, variant under investigation.

001584 © 2021 The Authors

 This is an open-access article distributed under the terms of the Creative Commons Attribution License. This article was made open access via a Publish and Read agreement between the Microbiology Society and the corresponding author's institution.

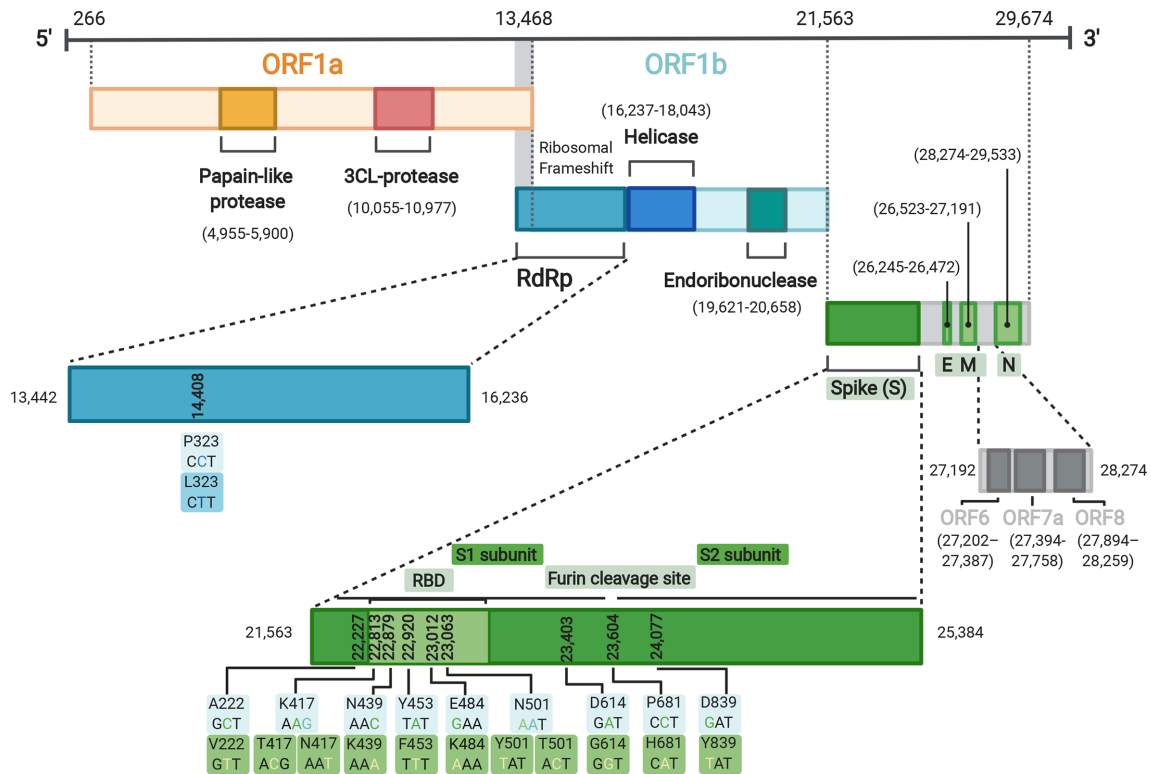


Fig. 1. Genome organisation of SARS-CoV-2 with regions of interest annotated. Mutations of interest (for example those found in B.1.1.7) shown as both nucleotide and amino acid changes. Figure made using Biorender (<https://biorender.com/>).

proceeding each ORF is a transcription regulatory sequence (TRS). The prevailing thought is that an integral part of the transcription mechanism in coronaviruses for the synthesis of viral sgRNAs involves a discontinuous step. The easiest way to visualise this, is that the polymerase/transcriptase complex binds to the 3' end of the positive strand and proceeds along the genome in a 3' to 5' direction synthesizing a negative strand. When the polymerase/transcriptase complex reaches a TRS, the newly synthesized negative strand can translocate to the 5' leader sequence of the genome where it is then copied. This forms a negative sense sgRNA that is then copied into the positive sense sgRNA [8]. This discontinuous nature has the consequence of a high degree of recombination resulting in the insertion of viral and non-viral sequences into - or frequent deletions of viral sequence from - the genome. This can result in the formation of viable genomes as well as defective interfering RNAs. Therefore, both SNPs and indels are likely to be the major processes allowing coronaviruses to rapidly switch host range or change their pathogenicity and/or virulence. For example, in cats infected by feline enteric coronavirus (FECV), variants can be generated within an infected animal by deletion of a key furin cleavage site in the spike protein. This results in feline peritonitis virus (FIPV) that causes a systemic fatal disease [9].

Recombination between different coronaviruses has been hypothesised to have given rise to both the genetically divergent receptor binding domain of SARS-CoV-2 spike

[10, 11], as well as the insertion of the S1/S2 furin (polybasic) cleavage site [12]. MERS-CoV is also thought to have had a major recombination event in recent evolutionary history [13]. Furthermore, deletions in the genome of the porcine coronavirus transmissible gastroenteritis virus (TGEV) gave rise to a new virus called porcine respiratory coronavirus (PRCV) [14]. Human seasonal coronavirus HCoV-OC43 and -HKU1 are thought to have acquired a hemagglutinin esterase (HE) gene following recombination between a progenitor coronavirus and influenza C-like virus [15]. Variants of OC43 and HKU1 HE have been shown to lose their sialic acid binding activity through progressive deletions in their lectin domains [16]. Finally, the N-terminal domain (NTD) of coronavirus spike proteins shares a number of structural similarities to eukaryotic galectins, leading to some to hypothesise the precursor to coronaviruses may have incorporated a portion of the host gene in the distant past [17]. Host RNA may also be a source for the polybasic cleavage site, similar to the proposed mechanism for generating highly pathogenic avian influenza viruses [18]. Studies on a recombinant attenuated SARS-CoV lacking the envelope (E) gene or the PDZ-binding motif (generated as a potential vaccine candidate) showed the virus could revert to virulence by partially duplicating a viral sequence (from ORF8a) which restored E function [19].

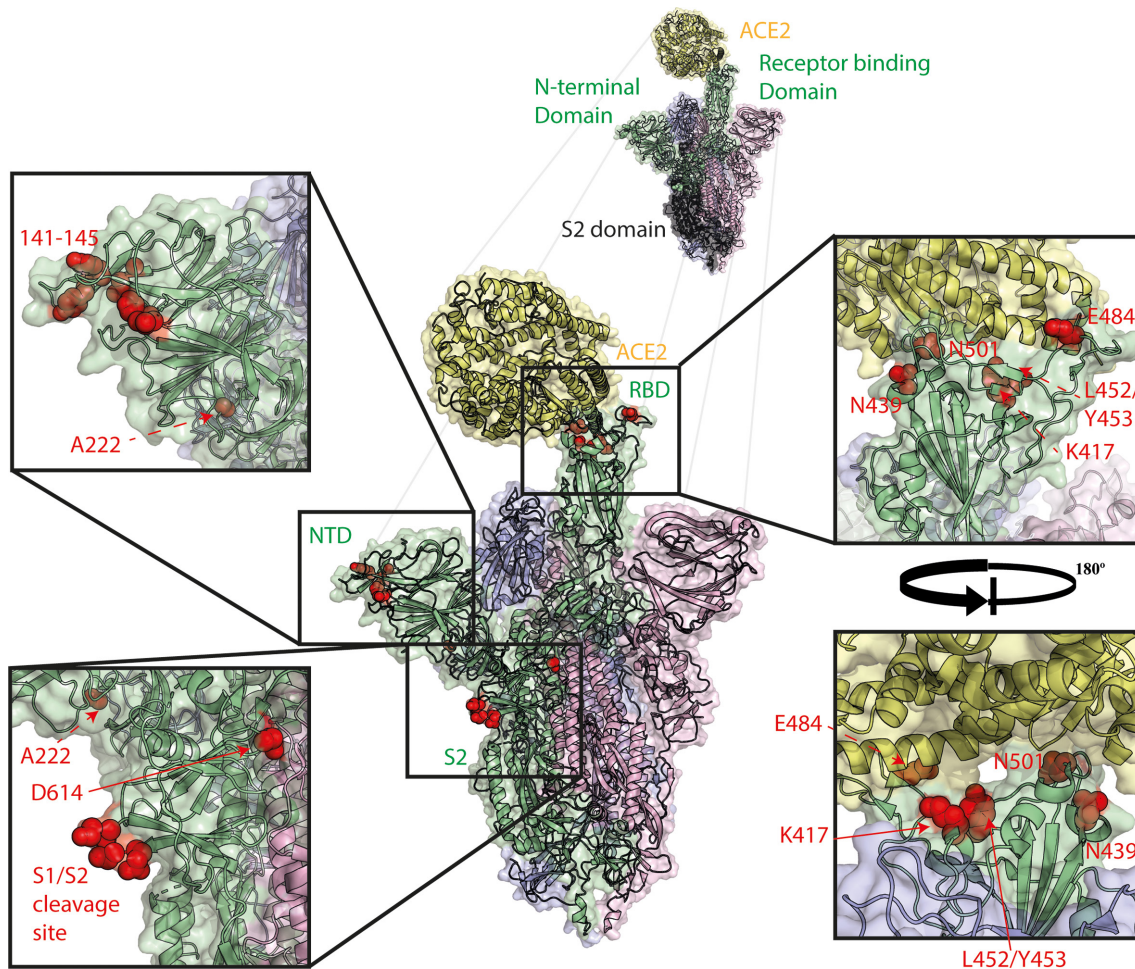


Fig. 2. Spike mutations of interest mapped to the spike trimer. Mutations shown in red, ACE2 shown in yellow, spike monomer in RBD 'up' conformation shown in green, spike monomers in RBD 'down' conformation shown in pink and blue. Structure made using PyMOL using PDBID 7A94 [24].

The spike protein is the major entry protein and antigen of SARS-CoV-2

Spike is the major glycoprotein responsible for SARS-CoV-2 entry, as well as the primary antigen and target of most SARS-CoV-2 vaccines currently in use and future development (Fig. 2). SARS-CoV-2 virions contain approximately 23 spike trimers on their surfaces [20]. The SARS-CoV-2 spike glycoprotein is synthesised as a single precursor polypeptide that forms trimers. Spike is subsequently cleaved into two major subunits, S1 and S2, by endogenous cellular furin [21]. The S1 subunit is composed of two further subdomains – an N-terminal domain (NTD), whose function is poorly described for SARS-CoV-2 but can act as a receptor binding domain in some coronaviruses and a potential glycan shield against antibody-mediated immunity, and a C-terminal receptor binding domain (RBD). The RBD of SARS-CoV-2 (as with SARS-CoV and seasonal HCoV-NL63) binds human angiotensin-converting enzyme 2 (ACE2), as its cognate cell surface receptor [22]. Spike glycoprotein shifts between two separate conformations – an 'open' or 'up' conformation able

to effectively bind ACE2, and a 'closed' or 'down' conformation, with its receptor binding interface packed down into the top of the spike trimer [23, 24]. Different trimers may have one, two or three spike glycoproteins in either conformation. It has been suggested that the closed conformation may allow for viral escape from RBD-binding neutralising antibodies. The S2 subunit contains the spike fusion peptide, a transmembrane domain and a short cytoplasmic tail. This short cytoplasmic tail contains a signal sequence that retains the spike in the endoplasmic reticulum from where, after particle assembly, virions are able to bud into the endoplasmic reticulum-Golgi intermediate compartment (ERGIC) [25]. Immediately adjacent to the fusion peptide is a second protease cleavage site termed the S2' cleavage site. Upon both S1/S2 cleavage and receptor binding by the RBD, the S1 subunit dissociates from S2 exposing the S2' site and enabling its cleavage by cellular proteases such as TMPRSS2 or Cathepsin L [26–28]. S2' cleavage results in immediate activation of the fusion peptide and subsequent spike-mediated membrane fusion [26].

Due to the spike glycoprotein being the major viral antigen and since the RBD/ACE2 interaction is a major host range determinant, there is considerable selection pressure placed on this region of the viral genome. Generally, the S1 subunit is thought to be the major inducer of a protective antibody response and variation in this region can result in antigenic drift either against previous infection by other variants or induced by vaccination [29].

D614G is now found in the majority of SARS-CoV-2 isolates and enhances virus infectivity and transmissibility in humans

The best characterised of the polymorphisms seen in SARS-CoV-2 since its emergence is the spike glycoprotein mutation D614G. Viruses with D614G were first detected in February 2020 and by May, around 80% of sequences globally were found to contain this mutation [30–32]. The rapid replacement of previously circulating SARS-CoV-2 strains is likely due to this virus being slightly more transmissible than the previous strains combined with a strong founder effect as the virus exponentially expanded in a first pandemic wave across Europe and the Americas [30, 33]. Notably, the major clade containing D614G (Pango lineage B.1 and its sub-lineages) also contained several other genetically linked mutations, including one in the main polymerase subunit NSP12, P323L, that may also have contributed to its dominance by exerting a fitness advantage. On the other hand, there are several examples of independent acquisition of D614G (but not P323L), such as the A.19 and A.2.4 lineages, which continue to circulate [34]. D614G has now been shown by a multitude of independent studies to enhance entry into human ACE2 expressing cells in pseudovirus based assays *in vitro* [30, 32, 35–37]. In addition, several studies linked D614G containing viruses to lower Ct values in clinical SARS-CoV-2 diagnostic PCR tests, indicating the virus replicated more efficiently in the human respiratory tract, although without any link to higher pathogenicity or severe clinical outcomes [30, 34, 38–40]. Several groups have also shown that D614G containing viruses (either recombinant or naturally occurring strains) had enhanced growth in primary human airway cells and replicated with greater efficiency in animal models such as hamster, ferret or the human ACE2-expressing mouse, and transmitted more efficiently in a hamster model, suggesting D614G alone is sufficient to confer this advantage in the absence of P323L [41–45].

Several non-mutually exclusive mechanisms have been proposed to explain how D614G enhances entry, replication, and transmission. The best described of which is that the polymorphism weakens an interaction at the trimer interface leading to a greater proportion of spike RBD in the ‘open’ conformation and subsequently allows enhanced ACE2 binding [24, 46–50]. An alternative proposed mechanism, though not necessarily incompatible with that previously described, is that D614G stabilises the pre-fusion structure of the spike trimer, preventing premature shedding of the S1 subdomain which can occur after furin cleavage [32, 47, 49, 51]. One additional proposed mechanism, again

not necessarily incompatible with the others, is that D614G results in changes in the conformation of the S1/S2 cleavage site loop allowing more efficient access by furin and therefore more efficient S1/S2 cleavage [42, 49, 51, 52].

Once it became clear D614G containing variants were rapidly expanding globally, a major concern was that this might affect the efficacy of vaccines that were being developed which universally contained spike antigens with the ancestral D614. This concern has been allayed by the repeated finding that the D614G variants are equally, if not more readily, neutralised by antisera raised against D614 containing virus or vaccines, as well as by therapeutic monoclonal antibodies [30, 32, 37, 41, 44, 46, 48].

At present, global sequencing surveillance suggests that viruses without D614G are almost non-existent. One exception is a lineage of viruses identified in Uganda (A.23) that, as of March 2021, continued to contain D614. This lineage, however, does contain a nearby spike mutation, Q613H. Although Q613H is currently uncharacterised, it may play a similar role to the D614G substitution, allowing this variant to continue co-circulating [53, 54].

Y453F is a mink adaptation that allows partial escape from antisera

Ferrets, which are members of the Mustelidae family, have traditionally been models for influenza virus transmission and infection and were quickly utilised in a similar manner for SARS-CoV-2 research, showing a dose dependent response to SARS-CoV-2 and protection from reinfection [55, 56]. Mink are closely related to ferrets and are farmed in many countries for their fur. It became apparent by the middle of 2020 that mink, like ferrets, were highly susceptible to reverse-zoonotic SARS-CoV-2 infection [57, 58]. Although mink (and ferrets) could be readily infected, several spike glycoprotein mutations rapidly and repeatedly arose in these hosts, both in the field and under laboratory conditions, most commonly Y453F and N501T in the RBD (Figs 1 and 2) [59–61]. Both Y453F and N501T have been shown to allow stronger binding of the spike RBD to human ACE2. Moreover, from analysis of the interaction between the spike glycoprotein and ACE2, it is apparent that the Y453F mutation may optimise an interaction with Y34 present in mink and ferret ACE2 [62–64]. Further alarm was raised when a large cluster of human cases were detected in Denmark and the Netherlands which harboured these mutations. In particular, the Y453F mutation was detected in Northern Denmark in combination with several other spike mutations including the NTD deletion $\Delta 69-70$ [60, 61]. This virus variant, known as ‘Cluster 5’, was shown to partially escape neutralisation by convalescent antisera [61, 65, 66]. This led to the culling of nearly 17 million mink [67], with several other countries making plans to close their own mink farms or carry out mass culling as a precaution.

It is possible that by adapting its suboptimal interaction with ACE2 protein found in mustelids, the virus may have inadvertently selected for stronger receptor binding to human ACE2 and this may account for loss of neutralisation. It is well

described for both human and avian influenza viruses that increases in receptor binding can allow non-specific antibody escape. A stronger interaction between the virus glycoprotein and host receptor may better outcompete weaker competitive binding between antibody and virus glycoprotein [68–70].

Additionally, Y453F has been reported in a single case from an immunocompromised patient, potentially as an adaptation to human ACE2 (see Table 1) [62, 71].

Deletions of the S1/S2 furin cleavage site emerge in cell culture as well as *in vivo* and attenuate the virus in airway cells and animal models of transmission

Vero E6 cells have been widely used for isolation and growth of SARS-CoV-2 stocks as they are readily available, easy to use and highly permissive to the virus [72]. However, during propagation of SARS-CoV-2 isolates in Vero cells, deletions incorporating, or flanking, the furin cleavage site between S1 and S2 spike subunits are often reported [43, 73–78]. Similar deletions have been detected in clinical samples at very low frequency, including from human autopsy samples [79–81]. The deletion of the furin cleavage site adapts the virus to higher replication in cells lacking TMPRSS2, such as Vero cells, but attenuates the virus in TMPRSS2-expressing cells such as primary human airway cells [51, 80, 82–85]. Furthermore, furin cleavage site deletions result in lower pathogenicity in animal models and attenuated virus transmission in hamster and ferret models [75, 80, 81, 84, 86]. This has implications for source virus used in infection and challenge studies and virus stocks should be sequenced prior to use to ensure the furin cleavage site is intact. We and others have suggested the furin cleavage site allows rapid TMPRSS2-dependent cell entry at the cell surface or early endosome allowing the virus to evade highly restrictive endosomal IFITM proteins (such as IFITM2 or IFITM3). In contrast, in the absence of TMPRSS2 the virus must enter via the endosome/lysosome to be activated by cathepsins. In the harsh conditions of the acidifying endosome or the lysosome, having a pre-cleaved S1/S2 site may be disadvantageous as it results in instability of the spike glycoprotein and premature S1 shedding [80, 85].

The globally emerging strain first found in the UK, B.1.1.7, contains the spike glycoprotein mutation N501Y and shows signs of having emerged from an immunocompromised host

In December 2020, a cluster of COVID-19 cases (known variously as B.1.1.7, 20B/501Y.V1, or VOC/202012/01) was detected in South East England [87]. This cluster showed evidence of higher transmissibility in the community compared to contemporary strains [88], significantly higher case-fatality rates [89, 90], and, depending on the study, lower Ct values from diagnostic PCR tests on clinical swabs [91–94]. In the UK, B.1.1.7 is now the predominant lineage, accounting for >90% of infections [95]. In many countries, initially imported B.1.1.7 is rapidly outcompeting local strains and becoming the major circulating strain [96–99]. B.1.1.7 contains seven non-synonymous mutations in the spike glycoprotein and a

Table 1. Summary of substitutions seen in SARS-CoV-2 isolates from immunocompromised patients

Protein	Mutation	References
Spike- RBD	N440K	Truong <i>et al.</i>
	Y453F	Bazykin <i>et al.</i>
	T470N	Bazykin <i>et al.</i>
	S477N	Khatamzas <i>et al.</i>
	T478K	Choi <i>et al.</i>
	V483A*	Avanzato <i>et al.</i> , Truong <i>et al.</i>
	E484K/A/Q	Choi <i>et al.</i> , Khatamzas <i>et al.</i> , Truong <i>et al.</i>
	F486I	Choi <i>et al.</i>
	Y489H	Choi <i>et al.</i>
	Q493K	Choi <i>et al.</i>
S494P	Choi <i>et al.</i>	
N501Y	Choi <i>et al.</i>	
Spike-NTD	V3G	Borges <i>et al.</i>
	P9L	Choi <i>et al.</i>
	Δ12–18	Choi <i>et al.</i>
	S13I	Truong <i>et al.</i>
	Δ18–30	Borges <i>et al.</i>
	T22I	Truong <i>et al.</i>
	S50L	Bazykin <i>et al.</i> , Borges <i>et al.</i>
	W64G	Kemp <i>et al.</i>
	Δ69–70	Bazykin <i>et al.</i> , Kemp <i>et al.</i>
	V70F	Truong <i>et al.</i>
	N87S	Borges <i>et al.</i>
	T95I	Truong <i>et al.</i>
	K97M	Truong <i>et al.</i>
	Deletions in 140 region	
Q183H	Choi <i>et al.</i>	
R190K	Truong <i>et al.</i>	
I197T	Truong <i>et al.</i>	
Y200H	Kemp <i>et al.</i>	
Δ211/L212I	Truong <i>et al.</i>	
N211K	Truong <i>et al.</i>	

Continued

Table 1. Continued

Protein	Mutation	References
	A222V	Borges <i>et al.</i>
	T240I	Kemp <i>et al.</i>
	P330S	Kemp <i>et al.</i>
Spike-other	R685Q (furin site)	Avanzato <i>et al.</i>
	D737G	Bazykin <i>et al.</i>
	D796H	Kemp <i>et al.</i>
	I870V	Choi <i>et al.</i>
	I1020S	Choi <i>et al.</i>
	N1108T	Truong <i>et al.</i>
NSP1	T78I	Truong <i>et al.</i>
	G82S	Truong <i>et al.</i>
	Δ85	Avanzato <i>et al.</i>
	V86I	Truong <i>et al.</i>
	G98V	Khatamzas <i>et al.</i>
	R124C	Truong <i>et al.</i>
NSP2	I21V	Choi <i>et al.</i>
	K489E	Khatamzas <i>et al.</i>
	I513T	Kemp <i>et al.</i>
	C540F	Truong <i>et al.</i>
	E617A	Truong <i>et al.</i>
NSP3	P74A	Truong <i>et al.</i>
	G282V	Khatamzas <i>et al.</i>
	T504A/I	Bazykin <i>et al.</i> , Khatamzas <i>et al.</i>
	I508V	Khatamzas <i>et al.</i>
	H682Y	Truong <i>et al.</i>
	D821N	Bazykin <i>et al.</i>
	Δ1267–9	Borges <i>et al.</i>
	S1375F	Borges <i>et al.</i>
L1870F	Truong <i>et al.</i>	
NSP4	T204I	Truong <i>et al.</i>
	T295I	Bazykin <i>et al.</i> , Borges <i>et al.</i> - lineage defining???
	A307V	Choi <i>et al.</i>
	V315I	Bazykin <i>et al.</i>
	S386F	Truong <i>et al.</i>
	N396S	Khatamzas <i>et al.</i>
	T492I	Truong <i>et al.</i>

Continued

Table 1. Continued

Protein	Mutation	References
NSP5	T21I	Khatamzas <i>et al.</i>
	L89F	Truong <i>et al.</i>
NSP6	L37F	Truong <i>et al.</i>
	M86I	Truong <i>et al.</i>
	M100V	Truong <i>et al.</i>
	Q160R	Khatamzas <i>et al.</i>
NSP7	V58G	Bazykin <i>et al.</i>
NSP9	T24I	Khatamzas <i>et al.</i>
	V157L	Kemp <i>et al.</i>
NSP12	R457C	Truong <i>et al.</i>
	E796D	Truong <i>et al.</i>
NSP13	P77L	Truong <i>et al.</i>
	T115I	Choi <i>et al.</i>
	T214I	Truong <i>et al.</i>
	P238L	Truong <i>et al.</i>
	V349L	Truong <i>et al.</i>
	P504L	Avanzato <i>et al.</i>
	Y541C	Avanzato <i>et al.</i>
NSP14	E453D	Truong <i>et al.</i>
	S461P	Avanzato <i>et al.</i>
NSP15	N177S	Kemp <i>et al.</i>
	A255V	Truong <i>et al.</i>
NSP16	A34V	Borges <i>et al.</i>
	S166A	Choi <i>et al.</i>
ORF3a	Q57H	Choi <i>et al.</i>
	S171L	Truong <i>et al.</i>
	T30I	Choi <i>et al.</i> , Borges <i>et al.</i> , Truong <i>et al.</i>
E	N48D	Truong <i>et al.</i>
	S50I	Truong <i>et al.</i>
M	A2S/V	Avanzato <i>et al.</i> , Choi <i>et al.</i>
	R42K	Borges <i>et al.</i>
	H125S/Y	Choi <i>et al.</i> , Truong <i>et al.</i>
	S197T	Truong <i>et al.</i>
ORF6	V5I	Khatamzas <i>et al.</i>
	T39S	Borges <i>et al.</i>
	S83L	Choi <i>et al.</i>

Continued

Table 1. Continued

Protein	Mutation	References
ORF7a	Q94K	Choi <i>et al.</i>
	A105V	Truong <i>et al.</i>
ORF7b	$\Delta 2$	Bazykin <i>et al.</i>
ORF8	K2N	Avanzato <i>et al.</i>
	T11I	Truong <i>et al.</i>
	Q18STOP	Bazykin <i>et al.</i>
N	L84S	Avanzato <i>et al.</i>
	T148A	Truong <i>et al.</i>
	R195G	Bazykin <i>et al.</i>
	A208S	Choi <i>et al.</i>
	T325K	Khatamzas <i>et al.</i>
	N345K	Avanzato <i>et al.</i>

*Mutations in bold have been found to have arisen in multiple isolates.

total of 23 mutations across the whole genome (see Table 2). Notable mutations in the spike glycoprotein include N501Y, $\Delta 69-70$, $\Delta 144$, and P681H in S1. N501Y lies in the RBD and has been described as increasing human ACE2 binding as well as enabling binding to mouse ACE2 [62, 100–102]. The P681H polymorphism lies adjacent to the S1/S2 furin cleavage site and we have shown this mutation alone, or in the B.1.1.7 spike, enhances its efficiency of furin cleavage [52]. The $\Delta 69-70$ and $\Delta 144$ deletions lie in the NTD region of spike and may modulate antigenicity [103–105]. Additionally, the B.1.1.7 lineage contains a premature stop codon in the accessory protein ORF8 and a three amino acid deletion in NSP6 (both described in more detail later in this review). The apparent long phylogenetic branch length and pattern of mutations has led to the hypothesis that this virus may have emerged from long-term infection in an immunocompromised patient, before spilling back into the general population [87]. There is growing evidence that B.1.1.7 viruses are only minimally less well neutralised by pre-B.1.1.7 convalescent or post-vaccine antisera [52, 66, 106–112]. Indeed vaccine effectiveness in the UK, where B.1.1.7 predominates, remains high [113].

Persistent infection in immunocompromised patients may allow viruses to rapidly generate diversity under prolonged selection pressures that are absent in typical SARS-CoV-2 infections that transmit within days and resolve within weeks. Such infections have been proposed to be a potential mechanism for rapid antigenic evolution in influenza [114]. Various NTD deletions in the SARS-CoV-2 spike glycoprotein are commonly observed in immune-suppressed patients with a long-term infection, supporting the idea of intra-host evolution (see Table 1) [71, 100, 103, 104, 115–118]. A recurrent NTD deletion in the 140–145 region has been found in nine separate chronically

infected patients, indicating this may be a signature mutation of these long-term infections [71, 100, 104, 115–118]. This mutation is also found in the B.1.1.7 lineage (as well the similar B.1.318, B.1.525, and P.3 lineages, discussed later). These 140-loop deletions have additionally been suggested to allow escape from NTD-targeting antibodies and encompasses part of a large neutralising epitope known as the NTD antigenic ‘supersite’ [104, 105, 119].

Furthermore, the $\Delta 69-70$ deletion in the NTD has also arisen multiple times independently, both in healthy and immunocompromised humans, as well as in mink, and often in combination with RBD interface mutants as described earlier in this review [61, 71, 103, 104, 120]. It has been hypothesised that this deletion could act as a ‘permissive’ mutation, somehow allowing or compensating for receptor binding mutations (such as N439K, Y453F or N501Y) that alone may be deleterious to virus fitness, due to a currently undescribed effect on spike stability or similar [120]. A mechanism for this relationship between NTD and RBD is unclear at present as residues 69–70 of the NTD are distal to the RBD in both the open and closed spike trimer conformations [24].

The 69–70 deletion removes six nucleotides that are part of the probe target sequence in one of the commonly used RT-PCR tests used to screen swabs for diagnosis with COVID-19. The resulting ‘S gene target failure’ has been a fortuitous way to easily monitor the growth of lineages carrying this deletion, such as the UK B.1.1.7 variant of concern [91]. Fortunately, the diagnosis of cases has not been compromised because of the redundancy built into the diagnosis platforms that use several different primer-probe sets across the SARS-CoV-2 genome.

Emerging variants first described in South Africa and Brazil contain the RBD mutation E484K which allows significant escape from human convalescent antisera

In recent months, several independent lineages of viruses containing the spike glycoprotein mutation E484K have been detected worldwide – once in South Africa (B.1.351 or 20B/501Y.V2) and at least twice independently in Brazil (P.1 or 20B/501Y.V3, and P.2) [121–123]. This mutation is of particular concern as independent studies have suggested E484K is a bona fide escape mutant to many convalescent antisera [108, 119, 124]. Both Brazil (particularly the Amazonas region) and South Africa experienced high disease burdens in 2020 and likely have high seroprevalence which may have driven emergence of these antigenic variants [125, 126]. This is further reinforced by several case studies showing E484K containing variants reinfecting healthcare workers in Brazil and a high rate of reinfection of seropositive individuals in the placebo arm of a vaccine trial in South Africa [122, 126–129]. Concerningly, recent evidence suggests that these E484K variants likely partially or fully escape vaccine- or naturally immunity-derived antisera [66, 108, 111, 130].

These E484K lineages represent three independent emergences of the same mutation: E484K alone in the P.2 lineage, or remarkably, together with K417T/N and N501Y in both the B.1.351 and P.1 lineages. Multiple independent lineages

Table 2. Substitutions and deletions seen in currently circulating variants of concern and variant of concern-like viruses

Variant names and aliases	Spike mutations	Other non-synonymous mutations in genome
B.1.1.7 VOC-20DEC-01 aka 20B/501Y.V1 (UK)	L18F* , †, Δ69–70 , Δ144 , N501Y , A570D, <i>D614G</i> ‡, P681H , T716I, S982A, D1118H	NSP3 – T183I, A890D, I1412T; NSP6 – Δ106–108 ; NSP12 – <i>P323L</i> ; ORF8 – Q27STOP, R52I, Y73C; N – D3L, S235F
B.1.351 aka 20B/501Y.V2 aka VOC-20DEC-02 (South Africa)	L18F* , D80A, D215G*, Δ242–244* , R246I*, K417N* , E484K , N501Y , <i>D614G</i> , A701V	NSP2 – T85I; NSP3 – K837N; NSP5 – K90R; NSP6 – Δ106–108* ; NSP12 – <i>P323L</i> ; ORF3a – <i>Q57H</i> , S171L; E – P71L; N – T205I
P.1 aka 20B/501Y.V3 aka VOC-21JAN-02 (Japan ex Brazil)	L18F , T20N§, P26S, D138Y, R190S*, K417T , E484K , N501Y , <i>D614G</i> , H655Y, T1027I	NSP3 – S370L, K977Q; NSP6 – Δ106–108 ; NSP12 – <i>P323L</i> ; NSP13 – E341D; Orf3a – S253P*; ORF8 – E92K; N – P80R <i>R203K</i> , <i>G204R</i>
A.23.1/E484K aka VUI-21FEB-01 (UK)	R102I, F157L, V367F, E484K , Q613H, P681R	NSP3 – L741F; NSP6 – M86I, L98F, M183I; ORF8 – <i>L84S</i> , E92K; N – S202N
B.1.525 aka VUI-21FEB-03 (UK ex West Africa)	Q52R, A67V*, Δ69–70 , Δ144 , E484K , <i>D614G</i> , Q677H , F888L	NSP3 – T1189I; NSP6 – Δ106–108 ; NSP12 – P323F; E – L21F; M – I82T; ORF6 – Δ2; N – Δ2/D3Y, A12G, <i>T205I</i>
B.1.1.318 aka VUI-21FEB-04 (UK ex West Africa)	T95I, Δ144 , E484K , <i>D614G</i> , P681H , D796H	NSP1 – Δ85 ; NSP3 – S126L*, E378V, K1693N; NSP4 – T173I, A446V; NSP5 – T21I; NSP6 Δ106–108 ; NSP12 – <i>P323L</i> ; NSP15 – V320M; NSP16 – A116S*; M – I82T; ORF7b/ORF8 fusion +ORF8 E106STOP; N – <i>R203K</i> , <i>G204R</i> , Δ208/R209G
B.1.324.1/E484K aka VUI-21MAR-01 (UK ex Antigua)	E484K , S494P, N501Y , <i>D614G</i> , P681H , E1111K	NSP2 – T85I; NSP3 – T1378P*; NSP4 – T189I, T439M; NSP6 – H111Q; NSP9 – P57S; NSP12 – S6L, <i>P323L</i> ; ORF3a – <i>Q57H</i> ; ORF8 – 35 bp deletion and frameshift; N – M234I
P.3 aka VUI-21MAR-02 (Philippines)	Δ141–143 , Δ243–244* , Y265C, E484K , N501Y , <i>D614G</i> , P681H , E1092K, H1101Y, V1176F	NSP3 – D736G, S1807F*; NSP4 – D217N*, L438P; NSP6 – D112E; NSP7 – L71F; NSP12 – <i>P323L</i> ; NSP13 – L280F*, A368V; ORF8 – K2Q; N – <i>R203K</i> , <i>G204R</i>
B.1.526 (New York)	L5F, T95I, D253G, E484K* or S477N* , <i>D614G</i> , A701V ,	NSP2 – T85I; NSP3 – V1139I; NSP4 – L438P; NSP6 – L37F, Δ106–108 , NSP12 – <i>P323L</i> ; NSP13 – Q88H; ORF3a – P42L, <i>Q57H</i> ; ORF8 – T11I; N – P199L, M234I
A.27 (Mayotte)	L18F* , L452R* , N501Y , A653V, H655Y , Q677H* , D796Y , G1219V	NSP2 – P106L; NSP4 – D217G*; NSP6 – N82S; NSP13 – P77L; ORF3a – V50A, 8nt deletion and frameshift; ORF8 – <i>L84S</i> , del119–120; N – S202N
Cluster 5 (Danish Mink)	Δ69–70 , Y453F, <i>D614G</i> , I692V, M1229I	NSP1 – Δ85 ; NSP3 – Δ1264; NSP12 – <i>P323L</i> , T739I; NSP15 – T112I; ORF3a – H182Y; N – S194L, <i>R203K</i> , <i>G204R</i>

*Indicates mutation found in some, but not all variants of this lineage.

†Residues in bold indicate mutations found in multiple variants.

‡Residues in italics indicate substitutions likely present in the ancestral viruses.

§Indicates mutations predicted to result in addition of a new N-linked glycosylation site.

gaining the same or similar patterns of mutations and rapidly increasing in frequency strongly suggests positive selection and parallel, or convergent evolution. The E484K mutation combined with the N501Y mutation has been suggested to synergistically enhance binding of the spike glycoprotein to human ACE2 [62, 102]. Furthermore, the B.1.1.7 lineage, the P.1 lineage, and several isolates from the B.1.351 lineage all contain an identical deletion of amino acids 106–108 in NSP6, known as the ‘SGF deletion’ (Table 2). NSP6 is a multi-pass transmembrane protein that is thought to be involved in autophagy and antagonism of innate immune responses, but it remains unclear what influence this deletion has on virus phenotype [131]. Finally, several clusters of the B.1.1.7 variant in the UK contain E484K (currently known as VOC 202102/02) and this single mutation allows B.1.1.7 to escape from neutralising antisera raised after vaccination or infection with first wave strains of virus [112, 132].

Repeated, independent emergence of SARS-CoV-2 variants of concern show signs of convergent evolution on a global scale

As well as the three main variants of concern described above (B.1.1.7, P.1, and B.1.351), in recent weeks further variants have been detected with similar amino acid substitutions throughout their genomes. The rapidly expanding list includes the B.1.526 lineage, first isolated in New York [133], the A.23.1 (E484K) lineage found in the UK (and related to the A.23 and A.23.1 lineages found in Uganda) [53, 54], the B.1.525 and B.1.1.318 lineages, both seen in multiple countries but linked with travel to West Africa [134], the A.27 lineage, associated with a cluster of cases on the island of Mayotte [135], the B.1.341.1 lineage [136], seen in the UK and related to travel to Antigua, and finally the P.3 lineage described first in the Philippines [137, 138]. Although there is not currently

strong evidence of more rapid transmission of most of these variants, they share a number of molecular characteristics with the major variants of concern, such as combinations of receptor avidity-enhancing mutations (N501Y, E484K and/or S477N), furin cleavage site adjacent mutations (P681R/H, Q677H, H655Y) and genomic deletions (spike $\sim\Delta 140$ deletions, $\sim\Delta 243$ deletions, and NSP6 $\Delta 106-108$, see Table 2). The similarity between many of these variants suggests a remarkable degree of convergent evolution.

N439K and L452R are antigenic variants that spread rapidly in 2020

Since March 2020, the spike mutation N439K has arisen multiple times (all alongside D614G), independently in Europe and the USA. N439K lies directly within the RBD/ACE2 binding interface. Subsequent binding studies have shown this variant shows a modest increase in ACE2 binding and clinical data indicates marginally lower Ct values in clinical diagnostic PCR tests, indicative of higher replication [40]. Furthermore, N439K moderately alters antigenicity with some human convalescent antisera and monoclonal antibodies less able to bind and neutralise the variant virus or pseudovirus, although these <fourfold differences for polyclonal antisera are unlikely to have a major impact on vaccine effectiveness [35, 40]. The predominant lineage containing N439K (B.1.258) is now mostly found in combination with the NTD deletion $\Delta 69-70$, although the significance of this is currently unknown [120].

In July 2020 a pair of sister lineages of SARS-CoV-2 (B.1.427 and B.1.429) were first detected in California containing the RBD mutation L452R (as well as a pair of mutations in the NTD). Subsequently these lineages have become widespread across the USA [139]. Like N439K, L452R has been shown by several studies to reduce convalescent antisera binding [124, 140], however others have since shown this reduction only has a minimal effect on vaccine-induced neutralising antibody titres, particularly when compared to the E484K-containing variants of concern [66]. L452R is also present in several emerging variants, such as the A.27 lineage (Table 2).

A222V and D839Y have emerged multiple times in the field and require further investigation

Although many spike glycoprotein variants have increased in frequency, phylogenetic analysis can show they are often the products of founder effects and it is therefore not clear whether they truly represent ongoing evolution and adaptation or are merely genetically neutral [33]. Two mutations outside of the RBD that have arisen multiple times and spread more widely than would be expected for neutral mutations are A222V, in the NTD, and D839Y, in the fusion peptide [39, 141]. There is speculation the main A222V-containing lineage (B.1.177) likely spread widely, particularly in the UK, due to a founder effect from tourists visiting Spain during the summer [141]. Unlike D614G, there was no evidence D839Y led to lower Ct values from swabs, suggesting this mutation doesn't have a large impact of virus replication within the host [39].

SARS-COV-2 MUTATIONS OUTSIDE THE SPIKE GLYCOPROTEIN

Although the majority of focus on polymorphisms in SARS-CoV-2 has been on the spike glycoprotein and its importance as the major antigen of the virus, a number of other interesting polymorphisms have been described throughout the rest of the genome, some of which have been phenotypically characterised.

Deletions in ORF8 have been associated with milder clinical disease

Like other coronaviruses, SARS-CoV-2 expresses several small accessory proteins. Whilst historically these proteins were not essential for cell culture propagation of coronaviruses, subsequently many of these proteins were found to have immune-modulating functions *in vivo* that may be host specific, therefore variation in such genes might be expected as SARS-CoV-2 adapts to optimal replication and transmission in humans [131].

ORF8 encodes an accessory protein of 121 amino acids that has been proposed to suppress the immune response during infection, potentially by downregulating host MHC class I expression [142]. Several independently emerging SARS-CoV-2 variants containing deletions in ORF8 have been described. One large cluster from Singapore in the Spring of 2020 contained a 382-nucleotide deletion resulting in a truncated ORF7b and complete ablation of ORF8 expression [143, 144]. This strain was associated with milder infection and improved disease outcomes [145], but subsequently became extinct as Singapore implemented strict biosecurity measures. In human primary nasal epithelial cells the deletion of ORF8 resulted in a modestly delayed replication kinetics at early time points but little or no difference in transcriptional profile compared to wild-type virus [146].

Interestingly, the aforementioned UK B.1.1.7 lineage, assumed to have arisen from an immunocompromised individual, contains a premature stop codon at position 27 of ORF8 – it is highly likely this results in loss of function of the accessory protein, similar to the Singapore cluster [87, 145, 146]. However, unlike the Singapore cluster, B.1.1.7 has been found to have a higher case fatality rate than other circulating lineages [89, 90], potentially due to the complex set of mutations in this variant of concern. Additionally, a single case report from an immunocompromised patient also included a premature stop codon in ORF8, similar to the B.1.1.7 lineage [71].

At the equivalent genome position of SARS-CoV-2 ORF8, SARS-CoV contains open reading frames for a pair of small accessory proteins, ORF8a and ORF8b. SARS-CoV animal isolates, as well as early human isolates, contained a single open reading frame (ORF8) at this position but later human isolates encoded ORF8a and ORF8b after a 29nt out-of-frame deletion in the gene [147, 148]. It is unclear if there is a link between this apparent SARS-CoV human adaptation and the recurrent SARS-CoV-2 ORF8 deletions seen, as the ORF8

proteins of the two viruses are highly divergent but this serves as an example of how accessory proteins might evolve during human adaptation.

Deletions and truncations of ORF7a have arisen multiple times independently

ORF7a is another accessory protein of SARS-CoV-2 and is thought, like its SARS-CoV orthologue, to interfere with surface expression of host restriction factor tetherin and to modulate host translation [149, 150], although others have recently suggested this tetherin inhibition may be a non-specific side effect of ORF7a-mediated Golgi fragmentation [151]. Several independent studies have detected SARS-CoV-2 isolates with unique in- and out-of-frame deletions in this protein resulting in heavily truncated versions being expressed [152–156]. It has been hypothesised that due to possible redundancy between ORF7a and ORF6 of SARS-CoV-2, which also is thought to also inhibit host translation, deletions in ORF7a may come at a low fitness cost *in vivo* [154]. Presently, no work has been published investigating what impact, if any, these deletions have on virus fitness.

The B.1.318 variant described previously, through a combination of nonsense mutations and a deletion of the intergenic region, is predicted to express a 146 amino acid ORF7b/ORF8 fusion protein – composed of the entire ORF7b protein fused to amino acids 3–105 of ORF8. It is unclear at present what the function or significance of this fusion protein might be, however it is by mechanisms such as this that we might expect novel accessory proteins to arise in coronaviruses.

Viruses with truncated ORF6 proteins result in a more pro-inflammatory innate immune response

ORF6 is another accessory protein involved in blocking the innate immune response. ORF6 of both SARS-CoV and SARS-CoV-2 can block nuclear import of STAT [157]. Multiple SARS-CoV-2 isolates have been found with truncations in ORF6, including in a nosocomial cluster, suggesting the viruses are able to transmit, at least in a hospital setting [158–160]. Virus replication kinetics of the variants with ORF6 truncations are equivalent to that of closely related isolates with intact ORF6 (160). However, the truncation does appear to result in an increase in NF- κ B related innate inflammatory responses [159].

The P323L mutation of the viral polymerase; an important human adaptation or a genetic hitchhiker?

The major catalytic component of the SARS-CoV-2 RNA-dependent RNA polymerase (RdRp) is the NSP12 subunit. NSP12 is a typical viral RdRp containing an serine-aspartic acid-aspartic acid (SDD) catalytic site and is the catalytic subunit in the replication/transcription complex responsible for viral RNA synthesis. The P323L polymorphism co-arose in the same virus cluster as the now dominant D614G mutation in the spike glycoprotein (lineage B.1 and its progeny lineages), along with a pair of non-coding/synonymous changes in the

5' UTR and NSP3 (30). Position 323 lies distal to the NSP12 catalytic core and is therefore unlikely to directly influence polymerase enzymatic activity. Rather this residue is located at the NSP12 surface, proximal to one of the binding sites for NSP8, a small polymerase cofactor. It is possible this mutation could be modulating NSP8 interaction or interaction with a yet unknown viral or host factor. Although reverse genetics experiments have shown D614G alone provides greater entry into human cells and greater viral fitness in animal models [41, 44, 45], it remains unclear if this mutation alone entirely recapitulates the phenotype seen with the whole virus isolates that also contain P323L. Further experiments are needed to determine whether P323L may also be playing a role in the success of this virus genotype or whether it is simply a genetic hitchhiker propelled by the success of the spike glycoprotein D614G substitution.

The B.1.525 lineage, which has many variants of concern-like properties and has been associated with travel to West Africa, contains the further NSP12 mutation, L323F. The significance of a further onward mutation at this position is currently unclear; a direct proline to phenylalanine substitution at this position would require a minimum of two nucleotide changes which is unusual, even in RNA viruses. Therefore, it is possible this mutation signifies fine tuning of this position to further enhance viral fitness, or that this is a partial reversion of a detrimental hitchhiker mutation. Either way emphasises position 323 plays an important, but undefined, role in polymerase function. *In vitro* assays with constituted polymerase could be of use to decipher this question, as might studies with SARS-CoV-2 replicons [131, 161].

Other rarer deletions of attenuating or unknown phenotype

As well as the mutations described above, several less common but notable deletions have been observed in SARS-CoV-2 sequences, most of which are currently phenotypically uncharacterised.

From analysis of 17 928 genome sequences, a nine-nucleotide deletion (amino acids Δ 241–243) was identified in NSP1 and was found in multiple different geographical locations. The role of NSP1 is to dampen the immune responses through host shut-off by inhibiting host ribosomal translation [162]. This deletion is hypothesised to destabilise the C-terminal domain, potentially impacting viral replication and leading to a less pathogenic phenotype [163]. Furthermore, a single amino acid deletion in NSP1 (Δ 85) has been described in several variants of concern-like viruses or immunocompromised patient sequences (Tables 1 and 2) [61, 115].

Viral genomes harbouring deletions can be found alongside wild-type genomes within a single patient. A 12-nucleotide deletion has been identified in the SARS-CoV-2 E in a clinical isolate; however, the mutant was rare in clinical samples and emerged *in vitro* during cell culture passage. E is a small transmembrane, structural protein involved viral envelope formation and maturation. Both the mutant and wild-type were able to infect Vero cells and produced comparable viral

titres, but the mutant virus isolate had higher spike glycoprotein content [164]. As the mutant was not identified in clinical samples, it is possible the deletion event is a result of cell culture adaptation. However, there are other reports of clusters of clinical isolates from India containing comparable deletions in E [165]. It has been shown previously that SARS-CoV artificially engineered to lack E gene expression is highly attenuated *in vitro* and *in vivo* [166].

Analysis of recombination at the minor variant level in a cohort of patients from the UK, show evidence of deletion events throughout the viral genome, particularly in ORF3a and ORF7a; accessory proteins that are both involved in interferon responses. These deletions may act as defective genomes [167]. Such deletion variants may contribute to the transient, within host emergence of SARS-CoV-2 isolates associated with milder disease [145].

The potential roles of host-mediated RNA-editing in SARS-CoV-2 evolution

Mutations in virus genomes are usually considered to be a result of error-prone viral polymerases; however, host factors can also play a role in the editing of the viral genome. The two main host candidates for endogenous mutagenesis are the RNA-editing enzymes of the APOBEC and ADAR families [168]. APOBEC proteins are interferon-stimulated and responsible for the deamination of cytosine to uracil after replication and before packaging of virus particles [169–171], whereas ADAR proteins are responsible for the deamination of adenine to inosine driving A →G changes during replication [172]. For influenza virus, we have previously shown that apparent RNA editing can result in the rapid emergence of antigenic variants with multiple concurrent amino acid changes [173]. Studies have shown a bias in C →U mutations within the SARS-CoV-2 genome, suggesting editing by APOBEC [168, 171, 174–179]. RNA editing is enriched in putative RNA loop regions, presumably due to being more exposed than other parts of the genome [177]. Due to observations of a bias in C →U mutations in the SARS-CoV-2 genome, it has been suggested that cytidine rich regions should be avoided during the design of diagnostic tests [178].

Furthermore, it has also been shown that virus-derived RNA sequences enriched for Uracil correlate with enhanced production of pro-inflammatory cytokines when comparing to the sequence of a reference virus. Based upon previous studies showing U-rich ssRNA stimulating the innate immune response through TLR7 signalling [180, 181], Kosuge *et al.* (2020), investigated the impact of C →U point mutations on the host response, showing an increase in TNF-α and IL-6 production in immune cell lines [176].

CONCLUSIONS

Although SARS-CoV-2 has only circulated in humans for a little over a year, an unprecedented sequencing effort has led to the description of many variants. Until recently, the only robust evidence for a genotypic change that had a strong

phenotype was the spike mutation D614G that has strongly been selected for in the human population. D614G has been shown to enhance virus entry and replication in the human respiratory tract. It is now necessary to understand which adaptive mutations are enhancing transmission and driving the increase of new variant of concern lineages such as B.1.1.7, now predominant in the UK and spreading worldwide, and whether the E484K mutation observed in multiple independent lineages from areas across the world that have encountered multiple waves of high virus circulation, including South Africa and Brazil, as well as a subset of the B.1.1.7 lineage, may allow antigenic escape and jeopardise vaccine efforts [119, 121, 123, 124, 127, 132]. The recent rise of new virus lineages underscores the importance of continued sequencing efforts to rapidly identify new variants. At present, a subset of countries, such as Denmark and the UK, are contributing a disproportionate amount of the total SARS-CoV-2 genome sequences, thanks largely to pre-existing sequencing capacity and recent public funding. Consequently, these countries have often been the first to describe these new, imported variants, however in many areas it is very unclear what strains and lineages are circulating. Ideally viruses must be sequenced from as broad a possible range of countries and regions and the results made available in a timely manner for surveillance of new variants to have significant policy impacts.

Several studies have examined potential antigenic variants using either ‘reverse genetics’ approaches - mostly generating libraries of pseudotypes with naturally occurring changes or changes in predicted antigenic regions - or ‘forward genetics’ approaches - using authentic virus, replication competent chimeric viruses (generally vesicular stomatitis virus with its native glycoprotein replaced with SARS-CoV-2 spike), or phage/yeast display screens and selecting with antibodies to drive the emergence of variants in either naturally occurring or mutagenesis-derived quasi-species or mutant libraries [35, 119, 124, 182, 183]. These approaches are a key part of our ability to predict the antigenic effect of mutations. However, although a variety of mutants have been shown to escape neutralisation or binding from monoclonal antibodies, and occasionally convalescent antisera, it is still unclear whether these mutations would come with a fitness costs in the context of infectious viruses that would make them less likely to arise in the field. Furthermore, although the majority of approved vaccines specifically target the humoral immune response against the spike protein, further work is needed to understand the potential role of cellular immunity in natural infection and how this could be used to optimise future vaccines.

As the level of natural immunity increases and global mass vaccination intensifies, it becomes ever more important to continuously sample, sequence and antigenically characterise novel virus variants, particularly from reinfections or from those who have been vaccinated [126]. This will allow for rapid detection of antigenic variants that could lead to potential vaccine failure, and for rapid vaccine updates

where required, in a similar manner to seasonal influenza. The possibility of antigenic drift is something vaccine designers, regulators and manufacturers should prepare for in the coming months and years. A likely scenario based on the development of animal coronavirus vaccines is that future SARS-CoV-2 vaccines may have to be multivalent to protect against multiple circulating antigenic variants, similar to vaccines against influenza or the avian gamma-coronavirus infectious bronchitis virus in poultry [184].

While changes in the spike protein are most important antigenically, genome alterations that change expression of viral accessory proteins are also expected and may influence transmission and pathogenicity. During its relatively brief human circulation, SARS-CoV ORF8 quickly gained a deletion leading to the creation of ORF8a and ORF8b [147]. Multiple SARS-CoV-2 isolates with accessory protein deletions and truncations are already described, these alterations can also occasionally lead to the expression of 'fusion ORFs' encompassing the N-terminus of one protein and the C-terminus of another. These variants remain rare and none has yet spread rapidly in the human population; most have quickly died out with the exception of the truncated ORF8 seen in the globally emerging B.1.1.7 lineage. However, this clearly remains an area that should be closely monitored during surveillance and performing whole genome sequencing (rather than just spike sequencing) remains imperative, since this type of genetic change might have an impact on transmission or clinical outcomes.

In addition, surveillance for reverse zoonoses and for mutations in chronically ill people with COVID-19 should be reinforced and intensified. Certain companion and farmed animals are clearly highly susceptible to SARS-CoV-2 and their infections could drive the selection of variant viruses with different receptor binding or antigenic properties that could cross back into humans. Similarly, the rapidly spreading B.1.1.7 lineage in the UK, which is hypothesised to have gained multiple mutations in a chronically ill patient, is only one of several variants reported from such individuals displaying multiple markers of potential altered antigenic and receptor binding properties, as well as a higher infectivity than previously circulating strains [100, 103]. Constant monitoring and enhanced biosecurity measures in these groups are essential to avoid novel virus variants from emerging.

It is important to note that, whilst consensus genomes are reported in global databases, individuals are infected with and have a population of virus within them [167], manifesting as a consensus genome and minor variants, and all of these genomes are subject to selection pressure. Individual variations can be selected if they are advantageous or through founder effect or a mixture of both. In viruses with significant deletions in proteins, or the presence of stop codons, reported at a consensus sequence level, these may be balanced within the virus population with the presence of functional proteins at a minor variant level, similar to that observed in Ebola virus infection in humans [185].

Deep sequencing of clinical isolates can reveal important cooperative interactions between members of the virus population within a single host, and a better understanding of transmission bottlenecks will allow us to understand whether such interactions are perpetuated in transmission chains.

To summarise, there is an urgent need to continue to perform in-depth surveillance and sequencing of SARS-CoV-2 isolates over the coming months and years coupled with detailed downstream phenotypic analysis of the impact of mutations in near real time. This analysis should include both traditional techniques with mutant virus isolates and closely related controls which can be performed fairly rapidly, as well as more modern, but slower, techniques such as reverse genetics [186], which are vital to disentangle the phenotypes of mutations that occur across multiple genes.

Funding information

TPP was supported by BBSRC (<https://bbsrc.ukri.org/>) grant BB/R013071/1. RP-R was supported by a PhD studentship from the MRC Discovery Medicine North (DiMeN) Doctoral Training Partnership (MR/N013840/1). JAH was funded by the US Food and Drug Administration contract number 75F40120C00085 'Characterization of severe coronavirus infection in humans and model systems for medical countermeasure development and evaluation' to support work on SARS-CoV-2 and MERS-CoV. WSB is supported by Wellcome Trust (<https://wellcome.org/>) grant 205100, BBSRC grants BB/K002465/1, BB/K002465/1 and BB/S008292/1. WB and JH are PIs in the UKRI G2P-UK consortium.

Acknowledgements

The authors would like to acknowledge Daniel Goldhill, Niluka Goonawardane, Rebecca Frise, Carol Sheppard, and Maya Moshe for their useful comments and help proofreading this review.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

1. Dorward DA, Russell CD, UM IH, Elshani M, Armstrong SD *et al*. Tissue-specific immunopathology in fatal COVID-19. *Am J Respir Crit Care Med* 2020.
2. News B. 2020. Coronavirus confirmed as pandemic by World Health Organization [bbc.co.uk: BBC. https://www.bbc.co.uk/news/world-51839944](https://www.bbc.co.uk/news/world-51839944)
3. World Health Organisation. WHO Coronavirus Disease (COVID-19) Dashboard. <https://covid19.who.int/2021>
4. Minskaia E, Hertzog T, Gorbatenya AE, Campanacci V, Cambillau C *et al*. Discovery of an RNA virus 3'->5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc Natl Acad Sci U S A* 2006;103:5108-5113.
5. Ogando NS, Zevenhoven-Dobbe JC, van der Meer Y, Bredenbeek PJ, Posthuma CC *et al*. The enzymatic activity of the nsp14 exoribonuclease is critical for replication of MERS-CoV and SARS-CoV-2. *J Virol* 2020;94:e01246-20.
6. Duchene S, Featherstone L, Haritopoulou-Sinaniidou M, Rambaut A, Lemey P *et al*. Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evol* 2020;6:veaa061.
7. Carroll MW, Matthews DA, Hiscox JA, Elmore MJ, Pollakis G *et al*. Temporal and spatial analysis of the 2014-2015 Ebola virus outbreak in West Africa. *Nature* 2015;524:97-101.
8. V'kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol* 2021;19:155-170.

9. Licitra BN, Millet JK, Regan AD, Hamilton BS, Rinaldi VD *et al.* Mutation in spike protein cleavage site and pathogenesis of feline coronavirus. *Emerg Infect Dis* 2013;19:1066–1073.
10. Lam TT-Y, Jia N, Zhang Y-W, Shum MH-H, Jiang J-F *et al.* Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* 2020;583:282–285.
11. Li X, Giorgi EE, Marichannegowda MH, Foley B, Xiao C *et al.* Emergence of SARS-CoV-2 through recombination and strong purifying selection. *Sci Adv* 2020;6:eabb9153.
12. Zhou H, Chen X, Hu T, Li J, Song H *et al.* 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* 2020;30:2196–2203.
13. Zhang Z, Shen L, Gu X. Evolutionary dynamics of MERS-CoV: potential recombination, positive selection and transmission. *Sci Rep* 2016;6:25049.
14. Chen F, Knutson TP, Rossow S, Saif LJ, Marthaler DG. Decline of transmissible gastroenteritis virus and its complex evolutionary relationship with porcine respiratory coronavirus in the United States. *Sci Rep* 2019;9:3953.
15. Zhang XM, Kousoulas KG, Storz J. The hemagglutinin/esterase gene of human coronavirus strain OC43: phylogenetic relationships to bovine and murine coronaviruses and influenza C virus. *Virology* 1992;186:318–323.
16. Bakkers MJG, Lang Y, Feitsma LJ, Hulswit RJG, de Poot SAH *et al.* Betacoronavirus adaptation to humans involved progressive loss of hemagglutinin-esterase lectin activity. *Cell Host Microbe* 2017;21:356–366.
17. Li F. Receptor recognition mechanisms of coronaviruses: a decade of structural studies. *J Virol* 2015;89:1954–1964.
18. Nao N, Yamagishi J, Miyamoto H, Igarashi M, Manzoor R *et al.* Genetic predisposition to acquire a polybasic cleavage site for highly pathogenic avian influenza virus hemagglutinin. *mBio* 2017;8:e02298–16.
19. Jimenez-Guardeño JM, Regla-Nava JA, Nieto-Torres JL, DeDiego ML, Castaño-Rodríguez C *et al.* Identification of the mechanisms causing reversion to virulence in an attenuated SARS-CoV for the design of a genetically stable vaccine. *PLoS Pathog* 2015;11:e1005215.
20. Ke Z, Oton J, Qu K, Cortese M, Zila V *et al.* Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature* 2020;588:498–502.
21. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT *et al.* Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 2020;181:281–292.
22. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020;579:270–273.
23. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 2020;367:1260–1263.
24. Benton DJ, Wrobel AG, Xu P, Roustan C, Martin SR *et al.* Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* 2020;588:327–330.
25. Lontok E, Corse E, Machamer CE. Intracellular targeting signals contribute to localization of coronavirus spike proteins near the virus assembly site. *J Virol* 2004;78:5913–5922.
26. Madu IG, Roth SL, Belouzard S, Whittaker GR. Characterization of a highly conserved domain within the severe acute respiratory syndrome coronavirus spike protein S2 domain with characteristics of a viral fusion peptide. *J Virol* 2009;83:7411–7421.
27. Millet JK, Whittaker GR. Host cell entry of middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. *Proc Natl Acad Sci U S A* 2014;111:15214–15219.
28. Belouzard S, Chu VC, Whittaker GR. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc Natl Acad Sci U S A* 2009;106:5871–5876.
29. Eguia R, Crawford KHD, Stevens-Ayers T, Kelnhofer-Millevolte L, Greninger AL *et al.* A human coronavirus evolves antigenically to escape antibody immunity. *bioRxiv* 2020.
30. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J *et al.* Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 2020;182:812–827.
31. Yang H-C, Chen C-H, Wang J-H, Liao H-C, Yang C-T *et al.* Analysis of genomic distributions of SARS-CoV-2 reveals a dominant strain type with strong allelic associations. *Proc Natl Acad Sci U S A* 2020;117:30679–30686.
32. Zhang L, Jackson CB, Mou H, Ojha A, Peng H *et al.* SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nat Commun* 2020;11:6013.
33. van Dorp L, Richard D, Tan CCS, Shaw LP, Acman M *et al.* No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. *Nat Commun* 2020;11:5986.
34. Volz E, Hill V, McCrone JT, Price A, Jorgensen D *et al.* Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell* 2021;184:64–75.
35. Li Q, Wu J, Nie J, Zhang L, Hao H *et al.* The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell* 2020;182:1284–1294.
36. Daniloski Z, Jordan TX, Ilmain JK, Guo X, Bhabha G *et al.* The spike D614G mutation increases SARS-CoV-2 infection of multiple human cell types. *elife* 2021;10:e65365 [Epub ahead of print 11 Oct 2021].
37. Ozono S, Zhang Y, Ode H, Sano K, Tan TS *et al.* SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. *Nat Commun* 2021;12:848.
38. Lorenzo-Redondo R, Nam HH, Roberts SC, Simons LM, Jennings LJ *et al.* A clade of SARS-CoV-2 viruses associated with lower viral loads in patient upper airways. *EBioMedicine* 2020;62:103112.
39. Borges V, Isidro J, Cortes-Martins H, Duarte S, Vieira L *et al.* Massive dissemination of a SARS-CoV-2 spike Y839 variant in Portugal. *Emerg Microbes Infect* 2020;9:2488–2496.
40. Thomson EC, Rosen LE, Shepherd JG, Spreafico R, da Silva Filipe A *et al.* Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell* 2021;184:1171–1187.
41. Hou YJ, Chiba S, Halfmann P, Ehre C, Kuroda M *et al.* SARS-CoV-2 D614G variant exhibits efficient replication *ex vivo* and transmission *in vivo*. *Science* 2020;370:1464–1468.
42. BW-Y M, Cremin CJ, Lau S-Y, Deng S, Chen P. SARS-CoV-2 spike D614G variant exhibits highly efficient replication and transmission in hamsters. *bioRxiv* 2020.
43. Pohl MO, Busnadiego I, Kufner V, Schmutz S, Zaheri M *et al.* Distinct phenotypes of SARS-CoV-2 isolates reveal viral traits critical for replication in primary human respiratory cells. *bioRxiv* 2020.
44. Plante JA, Liu Y, Liu J, Xia H, Johnson BA *et al.* Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* 2020 [Epub ahead of print 26 Oct 2020].
45. Zhou B, Thi Nhu Thao T, Hoffmann D, Taddeo A, Ebert N *et al.* SARS-CoV-2 spike D614G change enhances replication and transmission. *Nature* 2021 [Epub ahead of print 26 Feb 2021].
46. Yurkovetskiy L, Wang X, Pascal KE, Tomkins-Tinch C, Nyalile TP *et al.* Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant. *Cell* 2020;183:739–751.
47. Zhang J, Cai Y, Xiao T, Lu J, Peng H *et al.* Structural impact on SARS-CoV-2 spike protein by D614G substitution. *bioRxiv* 2020:2020.10.13.337980 [Epub ahead of print 13 Oct 2020].

48. Weissman D, Alameh M-G, de Silva T, Collini P, Hornsby H *et al.* D614G spike mutation increases SARS-CoV-2 susceptibility to neutralization. *Cell Host Microbe* 2021;29:23–31.
49. Gobeil SMC, Janowska K, McDowell S, Mansouri K, Parks R *et al.* D614G mutation alters SARS-CoV-2 spike conformation and enhances protease cleavage at the S1/S2 junction. *Cell Reports*.
50. Benton DJ, Wrobel AG, Roustan C, Borg A, Xu P *et al.* The effect of the D614G substitution on the structure of the spike glycoprotein of SARS-CoV-2. *Proc Natl Acad Sci U S A* 2021;118 [Epub ahead of print 02 03 2021].
51. Nguyen HT, Zhang S, Wang Q, Anang S, Wang J *et al.* Spike glycoprotein and host cell determinants of SARS-CoV-2 entry and cytopathic effects. *J Virol* 2020;JV1.02304-20.
52. Brown JC, Goldhill DH, Zhou J, Peacock TP, Frise R *et al.* Increased transmission of SARS-CoV-2 lineage B.1.1.7 (VOC 202012/01) is not accounted for by a replicative advantage in primary airway cells or antibody escape. *bioRxiv* 2021;2021.02.24.432576.
53. MRC/UVRI & LSHTM Uganda Research Unit, Entebbe SARS-CoV-2 Sequencing Group. SARS-CoV-2 diversity in Uganda, December 2020 virological.org; 2020.
54. Bugembe DL, Phan MVT, Ssewanyana I, Semanda P, Nansumba H *et al.* A SARS-CoV-2 lineage a variant (A.23.1) with altered spike has emerged and is dominating the current Uganda epidemic. *medRxiv* 2021.
55. Kim Y-I, Kim S-G, Kim S-M, Kim E-H, Park S-J *et al.* Infection and rapid transmission of SARS-CoV-2 in ferrets. *Cell Host Microbe* 2020;27:704–709.
56. Ryan KA, Bewley KR, Fotheringham SA, Slack GS, Brown P *et al.* Dose-dependent response to infection with SARS-CoV-2 in the ferret model and evidence of protective immunity. *Nat Commun* 2021;12:81.
57. Enserink M. Coronavirus rips through Dutch mink farms, triggering culls. *Science* 2020;368:1169.
58. Rabalski L, Kosinski M, Smura T, Aaltonen K, Kant R *et al.* Detection and molecular characterisation of SARS-CoV-2 in farmed mink (*Neovision vision*) in Poland. *bioRxiv* 2020.
59. Richard M, Kok A, de Meulder D, Bestebroer TM, Lamers MM *et al.* SARS-CoV-2 is transmitted via contact and via the air between ferrets. *Nat Commun* 2020;11:3496.
60. Oude Munnink BB, Sikkema RS, Nieuwenhuijse DF, Moleenaar RJ, Munger E *et al.* Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science* 2021;371:172–177.
61. Lassaunière R, Fonager J, Rasmussen M, Frische A, Polacek Strandh C *et al.* SARS-CoV-2 spike mutations arising in Danish mink and their spread to humans 2020.
62. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD *et al.* Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* 2020;182:1295–1310.
63. Welkers MRA, Han AX, Reusken C, Eggink D. Possible host-adaptation of SARS-CoV-2 due to improved ACE2 receptor binding in mink. *Virus Evolution* 2020.
64. Conceicao C, Thakur N, Human S, Kelly JT, Logan L *et al.* The SARS-CoV-2 spike protein has a broad tropism for mammalian ACE2 proteins. *PLoS Biol* 2020;18:e3001016.
65. Hoffmann M, Zhang L, Krüger N, Graichen L, Kleine-Weber H *et al.* SARS-CoV-2 mutations acquired in mink reduce antibody-mediated neutralization. *bioRxiv* 2021.
66. Garcia-Beltran WF, Lam EC, Denis KS, Nitido AD, Garcia ZH *et al.* Circulating SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *medRxiv* 2021.
67. Ministry of Environment and Food of Denmark. 2020. COVID-19: all mink in Denmark must be culled. <https://en.mfvmdk/news/news/nyhed/covid-19-all-mink-in-denmark-must-be-culled/>
68. Sealy JE, Yaqub T, Peacock TP, Chang P, Ermetal B *et al.* Association of increased receptor-binding avidity of influenza A(H9N2) viruses with escape from antibody-based immunity and enhanced zoonotic potential. *Emerg Infect Dis* 2018;25:63–72.
69. Hensley SE, Das SR, Bailey AL, Schmidt LM, Hickman HD *et al.* Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. *Science* 2009;326:734–736.
70. Peacock TP, Harvey WT, Sadeyen J-R, Reeve R, Iqbal M. The molecular basis of antigenic variation among A(H9N2) avian influenza viruses. *Emerg Microbes Infect* 2018;7:176.
71. Bazykin GA, Stanevich O, Danilenko D, Fadeev A, Komissarova K *et al.* Emergence of Y453F and Δ69-70HV mutations in a lymphoma patient with long-term COVID-19. *virological.org* 2021.
72. Chu H, Chan JF-W, Yuen TT-T, Shuai H, Yuan S *et al.* Comparative tropism, replication kinetics, and cell damage profiling of SARS-CoV-2 and SARS-CoV with implications for clinical manifestations, transmissibility, and laboratory studies of COVID-19: an observational study. *Lancet Microbe* 2020;1:e14–e23.
73. Davidson AD, Williamson MK, Lewis S, Shoemark D, Carroll MW *et al.* Characterisation of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein. *Genome Med* 2020;12:68.
74. Liu Z, Zheng H, Lin H, Li M, Yuan R *et al.* Identification of common deletions in the spike protein of SARS-CoV-2. *J Virol* 2020.
75. Lau S-Y, Wang P, Mok BW-Y, Zhang AJ, Chu H *et al.* Attenuated SARS-CoV-2 variants with deletions at the S1/S2 junction. *Emerg Microbes Infect* 2020;9:837–842.
76. Ogando NS, Dalebout TJ, Zevenhoven-Dobbe JC, Limpens RWAL, van der Meer Y *et al.* SARS-coronavirus-2 replication in Vero E6 cells: replication kinetics, rapid adaptation and cytopathology. *J Gen Virol* 2020;101:925–940.
77. Andrés C, Garcia-Cehic D, Gregori J, Piñana M, Rodriguez-Frias F *et al.* Naturally occurring SARS-CoV-2 gene deletions close to the spike S1/S2 cleavage site in the viral quasispecies of COVID19 patients. *Emerg Microbes Infect* 2020;9:1900–1911.
78. Klimstra WB, Tilston-Lunel NL, Nambulli S, Boslett J, McMillen CM *et al.* SARS-CoV-2 growth, furin-cleavage-site adaptation and neutralization using serum from acutely infected hospitalized COVID-19 patients. *J Gen Virol* 2020;101:1156–1169.
79. Liu Z, Zheng H, Lin H, Li M, Yuan R *et al.* Identification of common deletions in the spike protein of severe acute respiratory syndrome coronavirus 2. *J Virol* 2020;94 [Epub ahead of print 17 08 2020].
80. Peacock TP, Goldhill DH, Zhou J, Bailion L, Frise R *et al.* The furin cleavage site of SARS-CoV-2 spike protein is a key determinant for transmission due to enhanced replication in airway cells. *bioRxiv* 2020.
81. Wong YC, Lau SY, Wang To KK, Mok BWY, Li X *et al.* Natural transmission of bat-like SARS-CoV-2 variants in COVID-19 patients. *Clin Infect Dis* 2020 [Epub ahead of print 10 Jul 2020].
82. Mykityn AZ, Breugem TI, Riesebosch S, Schipper D, van den Doel PB *et al.* SARS-CoV-2 entry into human airway organoids is serine protease-mediated and facilitated by the multibasic cleavage site. *Elife* 2021;10:e64508 [Epub ahead of print 04 01 2021].
83. Sasaki M, Uemura K, Sato A, Toba S, Sanaki T *et al.* SARS-CoV-2 variants with mutations at the S1/S2 cleavage site are generated *in vitro* during propagation in TMPRSS2-deficient cells. *PLoS Pathog* 2021;17:e1009233.
84. Zhu Y, Feng F, Hu G, Wang Y, Yu Y *et al.* A genome-wide CRISPR screen identifies host factors that regulate SARS-CoV-2 entry. *Nat Commun* 2021;12:961.
85. Winstone H, Lista MJ, Reid AC, Bouton C, Pickering S *et al.* The polybasic cleavage site in the SARS-CoV-2 spike modulates viral sensitivity to Type I interferon and IFITM2. *J Virol* 2021 [Epub ahead of print 09 Feb 2021].

86. Johnson BA, Xie X, Bailey AL, Kalveram B, Lokugamage KG *et al.* Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature* 2021;591:293–299.
87. Rambaut A, Loman N, Pybus O, Barclay W, Barrett J *et al.* Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. *virological.org* 2020.
88. Volz E, Mishra S, Chand M, Barrett JC, Johnson R *et al.* Transmission of SARS-CoV-2 lineage B.1.1.7 in England: insights from linking epidemiological and genetic data. *medRxiv* 2021;2020.12.30.20249034.
89. Challen R, Brooks-Pollock E, Read JM, Dyson L, Tsaneva-Atanasova K *et al.* Increased hazard of mortality in cases compatible with SARS-CoV-2 variant of concern 202012/1 – a matched cohort study. *medRxiv* 2021;2021.02.09.21250937.
90. Davies NG, Jarvis CI, Edmunds WJ, Jewell NP, Diaz-Ordaz K *et al.* Increased hazard of death in community-tested cases of SARS-CoV-2 variant of concern 202012/01. *medRxiv* 2021;2021.02.01.21250959 [Epub ahead of print 03 Feb 2021].
91. Kidd M, Richter A, Best A, Mirza J, Percival B *et al.* S-variant SARS-CoV-2 is associated with significantly higher viral loads in samples tested by ThermoFisher TaqPath RT-PCR. *medRxiv* 2020.
92. Davies NG, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD *et al.* Estimated transmissibility and severity of novel SARS-CoV-2 variant of concern 202012/01 in England. *medRxiv* 2020.
93. Golubchik T, Lythgoe KA, Hall M, Ferretti L, Fryer HR *et al.* Early analysis of a potential link between viral load and the N501Y mutation in the SARS-CoV-2 spike protein. *medRxiv* 2021.
94. Walker AS, Vihta K-D, Gethings O, Pritchard E, Jones J *et al.* Increased infections, but not viral burden, with a new SARS-CoV-2 variant. *medRxiv* 2021.
95. Office for National Statistics. Coronavirus (COVID-19) infection survey, UK: 26 February 2021;2021.
96. Younes M, Hamze K, Nassar H, Makki M, Ghadar M *et al.* Emergence and fast spread of B.1.1.7 lineage in Lebanon. *medRxiv* 2021.
97. Haim-Boukobza S, Roquebert B, Trombert-Paolantoni S, Lecorche E, Verdurme L *et al.* Rapid SARS-CoV-2 variants spread detected in France using specific RT-PCR testing. *medRxiv* 2021.
98. Washington NL, Gangavarapu K, Zeller M, Bolze A, Cirulli ET *et al.* Genomic epidemiology identifies emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the United States. *medRxiv* 2021;2021.02.06.21251159 [Epub ahead of print 07 Feb 2021].
99. Brown KA, Gubbay J, Hopkins J, Patel S, Buchan SA *et al.* Rapid rise of S-gene target failure and the UK variant B.1.1.7 among COVID-19 isolates in the greater Toronto area, Canada. *medRxiv* 2021.
100. Choi B, Choudhary MC, Regan J, Sparks JA, Padera RF *et al.* Persistence and evolution of SARS-CoV-2 in an immunocompromised host. *N Engl J Med* 2020;383:2291–2293.
101. Gu H, Chen Q, Yang G, He L, Fan H *et al.* Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science* 2020;369:1603–1607.
102. Zahradník J, Marciano S, Shemesh M, Zoler E, Chiaravalli J *et al.* SARS-CoV-2 RBD *in vitro* evolution follows contagious mutation spread, yet generates an able infection inhibitor. *bioRxiv* 2021.
103. Kemp SA, Collier DA, Datir RP, Ferreira IATM, Gayed S *et al.* SARS-CoV-2 evolution during treatment of chronic infection. *Nature* 2021 [Epub ahead of print 05 Feb 2021].
104. McCarthy KR, Rennick LJ, Nambulli S, Robinson-McCarthy LR, Bain WG *et al.* Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *Science* 2021;371:eabf6950.
105. McCallum M, Marco AD, Lempp F, Tortorici MA, Pinto D *et al.* N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *bioRxiv* 2021;2021.01.14.426475 [Epub ahead of print 14 Jan 2021].
106. Xie X, Zou J, Fontes-Garfias CR, Xia H, Swanson KA *et al.* Neutralization of N501Y mutant SARS-CoV-2 by BNT162b2 vaccine-elicited sera. *bioRxiv* 2021.
107. Rees-Spear C, Muir L, Griffith S, Heaney J, Aldon Y *et al.* The impact of spike mutations on SARS-CoV-2 neutralization. *bioRxiv* 2021.
108. Wang Z, Schmidt F, Weisblum Y, Muecksch F, Barnes CO *et al.* mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *bioRxiv* 2021.
109. Muik A, Wallisch A-K, Sängler B, Swanson KA, Mühl J *et al.* Neutralization of SARS-CoV-2 lineage B.1.1.7 pseudovirus by BNT162b2 vaccine-elicited human sera. *bioRxiv* 2021.
110. Rathnasinghe R, Jangra S, Cupic A, Martínez-Romero C, Mulder LCF *et al.* The N501Y mutation in SARS-CoV-2 spike leads to morbidity in obese and aged mice and is neutralized by convalescent and post-vaccination human sera. *medRxiv* 2021;2021.01.19.21249592 [Epub ahead of print 20 Jan 2021].
111. Hu J, Peng P, Wang K, Fang L, Luo F-Y *et al.* Emerging SARS-CoV-2 variants reduce neutralization sensitivity to convalescent sera and monoclonal antibodies. *Cell Mol Immunol* 2021 [Epub ahead of print 25 Feb 2021].
112. Collier DA, De Marco A, Ferreira I, Meng B, Datir R *et al.* Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature* 2021.
113. Public Health England. PHE monitoring of the effectiveness of COVID-19 vaccination 2021 22 February 2021.
114. Xue KS, Moncla LH, Bedford T, Bloom JD. Within-Host evolution of human influenza virus. *Trends Microbiol* 2018;26:781–793.
115. Avanzato VA, Matson MJ, Seifert SN, Pryce R, Williamson BN *et al.* Case study: prolonged infectious SARS-CoV-2 shedding from an asymptomatic immunocompromised individual with cancer. *Cell* 2020;183:1901–1912.
116. Khatamzas E, Rehn A, Muenchhoff M, Hellmuth J, Gaitzsch E. Emergence of multiple SARS-CoV-2 mutations in an immunocompromised host. *medRxiv* 2021.
117. Borges V, Isidro J, Cunha M, Cochicho D, Martins L *et al.* Long-term evolution of SARS-CoV-2 in an immunocompromised patient with non-Hodgkin lymphoma. *virological.org* 2021.
118. Truong TT, Ryutov A, Pandey U, Yee R, Goldberg L *et al.* Persistent SARS-CoV-2 infection and increasing viral variants in children and young adults with impaired humoral immunity. *medRxiv* 2021;2021.02.27.21252099 [Epub ahead of print 02 Mar 2021].
119. Andreano E, Piccini G, Licastro D, Casalino L, Johnson NV *et al.* SARS-CoV-2 escape *in vitro* from a highly neutralizing COVID-19 convalescent plasma. *bioRxiv* 2020;2020.12.28.424451 [Epub ahead of print 28 Dec 2020].
120. Kemp S, Datir R, Collier D, Ferreira I, Carabelli A *et al.* Recurrent emergence and transmission of a SARS-CoV-2 spike deletion Δ H69/V70. *bioRxiv* 2020.
121. Naveca F, Nascimento V, Souza V, Corado A, Nascimento F *et al.* Phylogenetic relationship of SARS-CoV-2 sequences from Amazonas with emerging Brazilian variants harboring mutations E484K and N501Y in the spike protein. *virological.org* 2021.
122. Faria NR, Claro IM, Candido D, Moyses Franco LA, Andrade PS *et al.* Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. *virological.org* 2021.
123. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V *et al.* Emergence of a SARS-CoV-2 variant of concern with mutations in spike glycoprotein. *Nature* 2021 [Epub ahead of print 09 Mar 2021].
124. Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD *et al.* Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* 2021;29:463–476.
125. Buss LF, Prete CA, Abraham CMM, Mendrone A, Salomon T *et al.* Three-quarters attack rate of SARS-CoV-2 in the Brazilian

- Amazon during a largely unmitigated epidemic. *Science* 2021;371:eabe9728.
126. Shinde V, Bhikha S, Hoosain Z, Archary M, Bhorat Q et al. Preliminary efficacy of the NVX-CoV2373 Covid-19 vaccine against the B.1.351 variant. *medRxiv* 2021.
 127. Resende PC, Bezerra JF, Teixeira de Vasconcelos RH, Arantes I, Appolinario L. Spike E484K mutation in the first SARS-CoV-2 reinfection case confirmed in Brazil. *2020* 2021.
 128. Naveca F, da Costa C, Nascimento V, Souza V, Corado A et al. SARS-CoV-2 reinfection by the new variant of concern (VOC) P.1 in Amazonas, Brazil. *virological.org* 2021.
 129. Nonaka CKV, Franco MM, Gräf T, de Lorenzo Barcia CA, de Ávila Mendonça RN et al. Genomic evidence of SARS-CoV-2 reinfection involving E484K spike mutation, Brazil. *Emerg Infect Dis* 2021;27 [Epub ahead of print 19 Feb 2021].
 130. Wibmer CK, Ayres F, Hermanus T, Madzivhandila M, Kgagudi P et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *bioRxiv* 2021.
 131. Xia H, Cao Z, Xie X, Zhang X, Chen JY-C et al. Evasion of type I interferon by SARS-CoV-2. *Cell Rep* 2020;33:108234.
 132. Wise J. Covid-19: the E484K mutation and the risks it poses. *BMJ* 2021;372:n359.
 133. Annavajhala MK, Mohri H, Zucker JE, Sheng Z, Wang P et al. A novel SARS-CoV-2 variant of concern, B.1.526, identified in New York. *medRxiv* 2021:2021.02.23.21252259 [Epub ahead of print 25 Feb 2021].
 134. Pango Lineages. 2021. B.1.525. https://cov-lineages.org/global_report_B.1.525.html
 135. Wagner W, Hodcroft EB, Bell SM, Neher RA, Bedford T. Resurgence of SARS-CoV-2 19B clade corresponds with possible convergent evolution. *virological.org* 2021.
 136. . Proposal for lineage within B.1.324 with N501Y, P681H and others.
 137. Tablizo FA, Kim KM, Lapid CM, Castro MJR, Yangzon MSL et al. Genome sequencing and analysis of an emergent SARS-CoV-2 variant characterized by multiple spike protein mutations detected from the central Visayas region of the Philippines. *medRxiv* 2021.
 138. Public Health England. 2021. Variants of concern or under investigation: data up to 10 March 2021. <https://www.gov.uk/government/publications/covid-19-variants-genomically-confirmed-case-numbers/variants-distribution-of-cases-data>
 139. Zhang W, Davis BD, Chen SS, Sincuir Martinez JM, Plummer JT et al. Emergence of a novel SARS-CoV-2 variant in Southern California. *JAMA* 2021 [Epub ahead of print 11 Feb 2021].
 140. Liu Z, VanBlargan LA, Bloyet L-M, Rothlauf PW, Chen RE et al. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *Cell Host Microbe* 2021;29:477–488.
 141. Hodcroft EB, Zuber M, Nadeau S, Crawford KHD, Bloom JD et al. Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *medRxiv* 2020:2020.10.25.20219063 [Epub ahead of print 27 Nov 2020].
 142. Zhang Y, Zhang J, Chen Y, Luo B, Yuan Y et al. The ORF8 protein of SARS-CoV-2 mediates immune evasion through potently down-regulating MHC-I. *bioRxiv*. 2020.
 143. Gong Y-N, Tsao K-C, Hsiao M-J, Huang C-G, Huang P-N et al. SARS-CoV-2 genomic surveillance in Taiwan revealed novel ORF8-deletion mutant and clade possibly associated with infections in Middle East. *Emerg Microbes Infect* 2020;9:1457–1466.
 144. Su YCF, Anderson DE, Young BE, Linster M, Zhu F et al. Discovery and genomic characterization of a 382-nucleotide deletion in ORF7b and ORF8 during the early evolution of SARS-CoV-2. *mBio* 2020;11:e01610–01620.
 145. Young BE, Fong S-W, Chan Y-H, Mak T-M, Ang LW et al. 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* 2020;396:603–611.
 146. Gamage AM, Tan KS, Chan WOY, Liu J, Tan CW et al. 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* 2020;16:e1009130.
 147. Oostra M, de Haan CAM, Rottier PJM. The 29-nucleotide deletion present in human but not in animal severe acute respiratory syndrome coronaviruses disrupts the functional expression of open reading frame 8. *J Virol* 2007;81:13876–13888.
 148. Muth D, Corman VM, Roth H, Binger T, Dijkman R et al. Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. *Sci Rep* 2018;8:15177.
 149. Kopecky-Bromberg SA, Martinez-Sobrido L, Palese P. 7A protein of severe acute respiratory syndrome coronavirus inhibits cellular protein synthesis and activates p38 mitogen-activated protein kinase. *J Virol* 2006;80:785–793.
 150. Taylor JK, Coleman CM, Postel S, Sisk JM, Bernbaum JG et al. Severe acute respiratory syndrome coronavirus ORF7a inhibits bone marrow stromal antigen 2 virion tethering through a novel mechanism of glycosylation interference. *J Virol* 2015;89:11820–11833.
 151. Stewart H, Johansen KH, McGovern N, Palmulli R, Carnell GW et al. SARS-CoV-2 spike downregulates tetherin to enhance viral spread. *bioRxiv* 2021:2021.01.06.425396 [Epub ahead of print 06 Jan 2021].
 152. Holland LA, Kaelin EA, Maqsood R, Estifanos B, Wu LI, et al. An 81-Nucleotide deletion in SARS-CoV-2 ORF7a identified from sentinel surveillance in Arizona (January to March 2020). *J Virol* 2020;94:e00711–00720.
 153. Joonlasak K, Batty EM, Kochakarn T, Panthan B, Kümpornsin K et al. Genomic surveillance of SARS-CoV-2 in Thailand reveals mixed imported populations, a local lineage expansion and a virus with truncated ORF7a. *Virus Res* 2021;292:198233.
 154. Addetia A, Xie H, Roychoudhury P, Shrestha L, Loprieno M et al. Identification of multiple large deletions in ORF7a resulting in in-frame gene fusions in clinical SARS-CoV-2 isolates. *J Clin Virol* 2020;129:104523.
 155. Rosenthal SH, Kagan RM, Gerasimova A, Anderson B, grover D et al. Identification of eight SARS-CoV-2 ORF7a deletion variants in 2,726 clinical specimens. *bioRxiv* 2020.
 156. Nemudryi A, Nemudraia A, Wiegand T, Nichols J, Snyder DT et al. SARS-CoV-2 genomic surveillance identifies naturally occurring truncations of ORF7a that limit immune suppression. *medRxiv* 2021:2021.02.22.21252253 [Epub ahead of print 24 Feb 2021].
 157. Miorin L, Kehrler T, Sanchez-Aparicio MT, Zhang K, Cohen P et al. SARS-CoV-2 Orf6 hijacks Nup98 to block STAT nuclear import and antagonize interferon signaling. *Proc Natl Acad Sci U S A* 2020;117:28344–28354.
 158. Riojas MA, Frank AM, Puthuveetil NP, Flores B, Parker M et al. A rare deletion in SARS-CoV-2 Orf6 dramatically alters the predicted three-dimensional structure of the resultant protein. *bioRxiv* 2020:2020.06.09.134460 [Epub ahead of print 10 Jun 2020].
 159. Quéromès G, Destras G, Bal A, Regue H, Burfin G et al. Characterization of SARS-CoV-2 ORF6 deletion variants detected in a nosocomial cluster during routine genomic surveillance, Lyon, France. *Emerg Microbes Infect* 2021:1–56.
 160. Delbue S, D'Alessandro S, Signorini L, Dolci M, Pariani E et al. Isolation of SARS-CoV-2 strains carrying a nucleotide mutation, leading to a stop codon in the ORF 6 protein. *Emerg Microbes Infect* 2021;10:252–255.
 161. Walker AP, Fan H, Keown JR, Margitich V, Grimes JM et al. Enisamium is a small molecule inhibitor of the influenza A virus and SARS-CoV-2 RNA polymerases. *bioRxiv* 2020:2020.04.21.053017 [Epub ahead of print 06 Jan 2020].
 162. Schubert K, Karousis ED, Jomaa A, Scaiola A, Echeverria B et al. SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. *Nat Struct Mol Biol* 2020;27:959–966.

163. Benedetti F, Snyder GA, Giovanetti M, Angeletti S, Gallo RC *et al.* Emerging of a SARS-CoV-2 viral strain with a deletion in NSP1. *J Transl Med* 2020;18:329.
164. Sun Y-S, Xu F, An Q, Chen C, Yang Z-N *et al.* A SARS-CoV-2 variant with the 12-bp deletion at E gene. *Emerg Microbes Infect* 2020;9:2361–2367.
165. Kumar BK, Rohit A, Prithvisagar KS, Rai P, Karunasagar I *et al.* Deletion in the C-terminal region of the envelope glycoprotein in some of the Indian SARS-CoV-2 genome. *Virus Res* 2021;291:198222.
166. DeDiego ML, Álvarez E, Almazán F, Rejas MT, Lamirande E *et al.* A severe acute respiratory syndrome coronavirus that lacks the E gene is attenuated *in vitro* and *in vivo*. *J Virol* 2007;81:1701–1713.
167. Moore SC, Penrice-Randal R, Alruwaili M, Randle N, Armstrong S *et al.* Amplicon-Based detection and sequencing of SARS-CoV-2 in nasopharyngeal swabs from patients with COVID-19 and identification of deletions in the viral genome that encode proteins involved in interferon antagonism. *Viruses* 2020;12:1164.
168. Mourier T, Sadykov M, Carr MJ, Gonzalez G, Hall WW *et al.* Host-directed editing of the SARS-CoV-2 genome. *Biochem Biophys Res Commun* 2021;538:35–39.
169. Salter JD, Bennett RP, Smith HC. The APOBEC protein family: United by structure, divergent in function. *Trends Biochem Sci* 2016;41:578–594.
170. Salter JD, Smith HC. Modeling the embrace of a mutator: APOBEC selection of nucleic acid ligands. *Trends Biochem Sci* 2018;43:606–622.
171. Di Giorgio S, Martignano F, Torcia MG, Mattiuz G, Conticello SG. Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. *Sci Adv* 2020;6:eabb5813.
172. Placido D, Brown BA, Lowenhaupt K, Rich A, Athanasiadis A. A left-handed RNA double helix bound by the Z alpha domain of the RNA-editing enzyme ADAR1. *Structure* 2007;15:395–404.
173. Peacock TP, Benton DJ, James J, Sadeyen J-R, Chang P *et al.* Immune escape variants of H9N2 influenza viruses containing deletions at the hemagglutinin receptor binding site retain fitness *in vivo* and display enhanced zoonotic characteristics. *J Virol* 2017;91 [Epub ahead of print 15 07 2017].
174. Wang R, Hozumi Y, Zheng YH, Yin C, Wei GW. Host immune response driving SARS-CoV-2 evolution. *Viruses* 2020;12.
175. Simmonds P. Rampant C→U hypermutation in the genomes of SARS-CoV-2 and other coronaviruses: causes and consequences for their short- and long-term evolutionary trajectories. *mSphere* 2020;5:e00408-20 [Epub ahead of print 24 06 2020].
176. Kosuge M, Furusawa-Nishii E, Ito K, Saito Y, Ogasawara K. Point mutation bias in SARS-CoV-2 variants results in increased ability to stimulate inflammatory responses. *Sci Rep* 2020;10:17766.
177. Klimczak LJ, Randall TA, Saini N, Li J-L, Gordenin DA. Similarity between mutation spectra in hypermutated genomes of rubella virus and in SARS-CoV-2 genomes accumulated during the COVID-19 pandemic. *PLoS One* 2020;15:e0237689.
178. Wang R, Hozumi Y, Yin C, Wei G-W. Mutations on COVID-19 diagnostic targets. *Genomics* 2020;112:5204–5213.
179. Graudenzi A, Maspero D, Angaroni F, Piazza R, Ramazzotti D. Mutational signatures and heterogeneous host response revealed via large-scale characterization of SARS-CoV-2 genomic diversity. *iScience* 2021;24:102116.
180. Takeda K, Akira S. Toll-like receptors in innate immunity. *Int Immunol* 2005;17:1–14.
181. Heil F, Hemmi H, Hochrein H, Ampenberger F, Kirschning C *et al.* Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. *Science* 2004;303:1526–1529.
182. Baum A, Fulton BO, Wloga E, Copin R, Pascal KE *et al.* Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science* 2020;369:1014–1018.
183. Starr TN, Greaney AJ, Addetia A, Hannon WW, Choudhary MC *et al.* Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science* 2021;371:850–854.
184. Jordan B. Vaccination against infectious bronchitis virus: a continuous challenge. *Vet Microbiol* 2017;206:137–143.
185. Dong X, Munoz-Basagoiti J, Rickett NY, Pollakis G, Paxton WA *et al.* Variation around the dominant viral genome sequence contributes to viral load and outcome in patients with Ebola virus disease. *Genome Biol* 2020;21:238.
186. Thi Nhu Thao T, Labrousseau F, Ebert N, V'kovski P, Stalder H *et al.* Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature* 2020;582:561–565.

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.