



# The Evolution and Biology of SARS-CoV-2 Variants

Amalio Telenti,<sup>1,2</sup> Emma B. Hodcroft,<sup>3,4</sup> and David L. Robertson<sup>5</sup>

<sup>1</sup>Vir Biotechnology, San Francisco, California 94158, USA

<sup>2</sup>Department of Integrative Structural and Computational Biology, Scripps Research, La Jolla, California 92037, USA

<sup>3</sup>Institute of Social and Preventive Medicine, University of Bern, 3012 Bern, Switzerland

<sup>4</sup>Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

<sup>5</sup>MRC-University of Glasgow Centre for Virus Research, University of Glasgow, Glasgow G61 1QH, UK

Correspondence: atelenti@vir.bio; david.l.robertson@glasgow.ac.uk

Our understanding of the still unfolding severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic would have been extremely limited without the study of the genetics and evolution of this new human coronavirus. Large-scale genome-sequencing efforts have provided close to real-time tracking of the global spread and diversification of SARS-CoV-2 since its entry into the human population in late 2019. These data have underpinned analysis of its origins, epidemiology, and adaptations to the human population: principally immune evasion and increasing transmissibility. SARS-CoV-2, despite being a new human pathogen, was highly capable of human-to-human transmission. During its rapid spread in humans, SARS-CoV-2 has evolved independent new forms, the so-called “variants of concern,” that are better optimized for human-to-human transmission. The most important adaptation of the bat coronavirus progenitor of both SARS-CoV-1 and SARS-CoV-2 for human infection (and other mammals) is the use of the angiotensin-converting enzyme 2 (ACE2) receptor. Relaxed structural constraints provide plasticity to SARS-related coronavirus spike protein permitting it to accommodate significant amino acid replacements of antigenic consequence without compromising the ability to bind to ACE2. Although the bulk of research has justifiably concentrated on the viral spike protein as the main determinant of antigenic evolution and changes in transmissibility, there is accumulating evidence for the contribution of other regions of the viral proteome to virus–host interaction. Whereas levels of community transmission of recombinants compromising genetically distinct variants are at present low, when divergent variants cocirculate, recombination between SARS-CoV-2 clades is being detected, increasing the risk that viruses with new properties emerge. Applying computational and machine learning methods to genome sequence data sets to generate experimentally verifiable predictions will serve as an early warning system for novel variant surveillance and will be important in future vaccine planning. Omicron, the latest SARS-CoV-2 variant of concern, has focused attention on step change antigenic events, “shift,” as opposed to incremental “drift” changes in antigenicity. Both an increase in transmissibility and antigenic shift in Omicron led to it readily causing infections in the fully vaccinated and/or previously infected.

---

Editors: Richard Sever, Angela Rasmussen, and Stephen Goldstein  
Additional Perspectives on COVID-19 available at [www.cshperspectives.org](http://www.cshperspectives.org)

Copyright © 2022 Cold Spring Harbor Laboratory Press; all rights reserved; doi: 10.1101/cshperspect.a041390  
Cite this article as *Cold Spring Harb Perspect Med* 2022;12:a041390

A. Telenti et al.

Omicron’s virulence, while reduced relative to the variant of concern it replaced, Delta, is very much premised on the past immune exposure of individuals with a clear signal that boosted vaccination protects from severe disease. Currently, SARS-CoV-2 has proven itself to be a dangerous new human respiratory pathogen with an unpredictable evolutionary capacity, leading to a risk of future variants too great not to ensure all regions of the world are screened by viral genome sequencing, protected through available and affordable vaccines, and have non-punitive strategies in place for detecting and responding to novel variants of concern.

The new human pathogen severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first identified as the cause of coronavirus disease 2019 (COVID-19) in December 2019 when severe pneumonia-hospitalized cases in Wuhan city, China, were connected through epidemiological links to the Huanan Seafood Market in Wuhan (Worobey 2021; see key dates of the pandemic in Table 1). Despite some recording of direct human-to-human transmissions, the lack of known contacts between most of the cases detected early (not surprising in hindsight given the frequency of mild symptoms and asymptomatic cases) led to an incorrect assumption of exclusively animal-to-human transmission. The animal–market link (a known risk factor from the first SARS virus

spillovers 20 years ago) resulted in the clearing and closure of the Huanan market on January 1, 2020. Recognition of the extent of human-to-human transmission was slow, necessitating a lockdown on an unprecedented scale in Wuhan and other Hubei cities to control rising case numbers. Despite this well-reported outbreak, the knowledge by mid-January that a SARS-like virus had returned and the World Health Organization (WHO) reporting a public health emergency of international concern on January 30, 2020, most countries responded with a lack of urgency, and a pandemic was declared by the WHO on March 11, 2020. The nature of the response to the COVID-19 threat varied dramatically from country to country. SARS-CoV-2 is now estimated to have killed over six

**Table 1.** Some key dates in the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic

Key dates	Events
December 2019	First pneumonia cases identified in Wuhan city, Hubei Province, China
Late December 2019	Pneumonia cases of unknown etiology linked to Huanan market
31 December 2019	World Health Organization (WHO) notified of outbreak of severe pneumonia cases
1 January 2020	Huanan market closed, cleared of animals and sanitized
10 January 2020	First SARS-CoV-2 genome, Wuhan-Hu-1, released on Virological.org and 2 days later on GenBank, accession number MN908947
30 January 2020	WHO declares public health emergency of international concern (PHEIC)
11 March 2020	WHO declares the novel coronavirus outbreak a pandemic
March 2020	First companion animals (dogs) found to be infected with SARS-CoV-2 from owners
April 2020	Second companion animals (cats) found to be infected with SARS-CoV-2 from owners
April 2020	SARS-CoV-2 infections on mink farms reported with transmission back to humans
December 2020	First variant of concern (VOC) identified, Alpha; first detected in the United Kingdom
December 2020	Second VOC identified, Beta; first detected in South Africa
January 2021	Third VOC identified, Gamma; first detected in Japan in traveler from Brazil
May 2021	Fourth VOC identified, Delta; first detected in India
August 2021	Wild deer infections identified and possible transmission back to humans
November 2021	Fifth VOC identified, Omicron; first detected in South Africa

million people (covid19.who.int) and is firmly established in the human population as a new pathogen.

Classically, and in retrospect, naively, it is often assumed that novel human viruses will not initially be particularly successful as there has to be some adaptation/fine-tuning to a new host to achieve efficient spread. This has not been the case for SARS-CoV-2. Despite being descended from a bat coronavirus and most likely transmitted to humans via an intermediate host, it was immediately very successful in humans. This is very similar to the first SARS virus identified (SARS-CoV, referred to as SARS-CoV-1 for clarity), both in the likely route of emergence (live animal markets) and in its high transmissibility, although SARS-CoV-1 is not to quite the same extent as SARS-CoV-2. Worryingly, this host-species promiscuity/generalist property has led to infections in many other animals, as well as establishment of the virus in wild species and spillback to humans (Oreshkova et al. 2020; Pickering et al. 2022).

Hospital treatments such as intensive care and/or oxygen for the very sick, and non-pharmaceutical interventions—including mask mandates, shielding, and national lockdowns—protected healthcare workers and societies from COVID-19 to the degree that control measures were implemented and adopted. Vaccines have been developed rapidly, including use of new mRNA technologies, and more recently effective pharmaceuticals have begun to become available. These treatments work effectively because acute viral infections are (usually) rapidly cleared by the immune response, with chronic infections limited to immunocompromised individuals, and so are relatively rare. However, being new to the human population, SARS-CoV-2 has much potential to adapt to its new host species, in particular by increasing transmissibility, a property first exemplified by the spike amino acid replacement D614G (Korber et al. 2020). The first variants of concern (VOCs: Alpha, Beta, Gamma, and Delta; see key dates in Table 1) can be considered a second phase of the pandemic: the evolution of distinct variants capable of spreading more efficiently despite increasing

levels of host immunity (either vaccine or virus elicited by past infection). Surprisingly, given the initially relatively slow rate of evolution of SARS-CoV-2 (MacLean et al. 2021), each VOC has displayed a significant amount of evolutionary change resulting in rapid replacement of previous variants. Whether such step changes in SARS-CoV-2 biology continue or its evolution becomes more incremental, it is inevitable that further evolution and novel variants will emerge with altered properties of antigenic escape and potentially antigenic shift on top of high transmissibility, as shown by SARS-CoV-2 Omicron (Cameroni et al. 2022). However, in the context of increasing immunity from vaccinations and past infections (and better treatments), future variants will hopefully not cause severe disease in large numbers of infected people.

Another major concern is reverse zoonoses, the zoonothroponotic spread to animal populations, a direct result of the generalist nature of SARS-related viruses. Since 2020, SARS-CoV-2 was found to infect companion animals (Table 1) and has become established in wild animal species, most notably white-tailed deer in the United States (Hale et al. 2022). SARS-CoV-2 in animals could, thus, establish new reservoirs for future spillover. Coupled with coronaviruses with SARS-virus-like properties already being distributed over wide geographical areas in their host bat species, future human infections from animals are very likely and such variants could even recombine with circulating SARS-CoV-2. Here, we review current knowledge of the biological properties of amino acid replacements in SARS-CoV-2 proteins, SARS-CoV-2 lineages and VOCs, and variants of interest/being monitored (VOIs/VBMs). Key terms are listed in Box 1.

## PHYLOGENY OF THE CORONAVIRIDAE

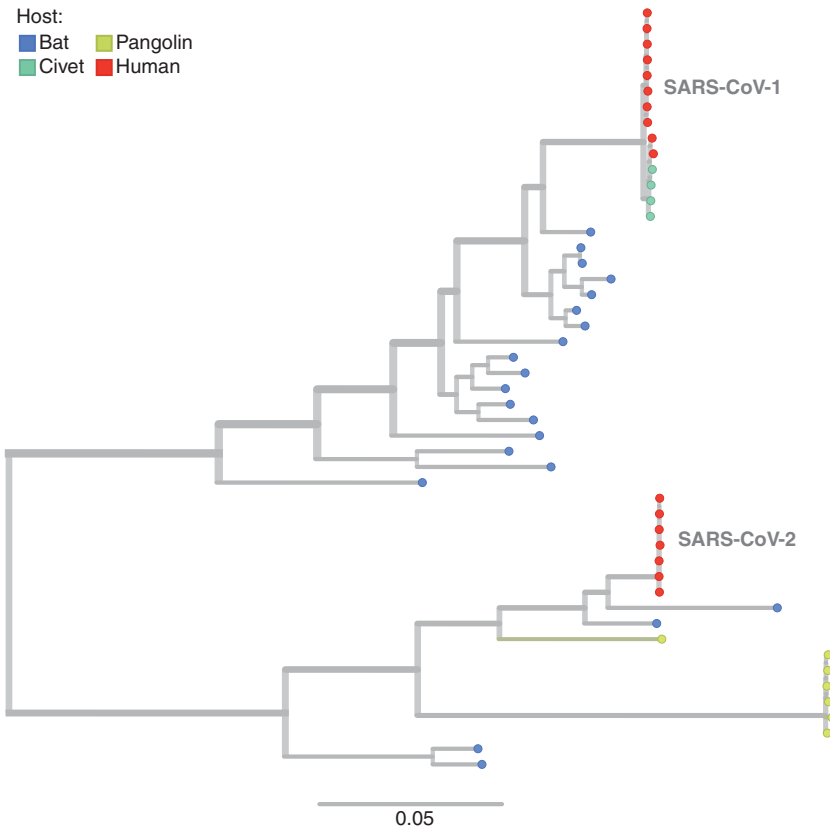
The Coronaviridae include two subfamilies, five genera, and 26 subgenera (ictvonline.org). SARS-CoV-2 and related bat sarbecoviruses are a sister clade to SARS-CoV-1 and various other bat sarbecoviruses (Fig. 1). Collectively, these form the species *severe acute respiratory syndrome-related coronavirus* (SARSr-CoV), the sole member of

**BOX 1. GLOSSARY**

Variant	A variant is a distinct viral genome that contains one or more mutations relative to other viral genomes.
Strain	A novel variant with significantly altered phenotypic properties compared to previous variants. Not used to date in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) nomenclature but arguably Omicron's phenotypic properties relative to the other SARS-CoV-2 variants might warrant its use.
Lineage	A group of closely related viruses with a shared common ancestor. Tends to be used interchangeably with clade, a monophyletic grouping in a phylogenetic tree.
Variant of concern (VOC)	Public health agencies, for example, the Centers for Disease Control and Prevention (CDC) or the UK Health Security Agency (UKHSA), define a VOC as a specific variant for which there is evidence of an increase in transmissibility, more severe disease, significant reduction in neutralization, reduced effectiveness of treatments or vaccines, or diagnostic detection failures.
Variant of interest (VOI) or variant under investigation (VUI)	Variant with specific genetic markers that have been associated with changes to receptor binding, reduced neutralization, reduced efficacy of treatments, potential diagnostic impact, or predicted increase in transmissibility or disease severity.
Variant being monitored (VBM)	Variants where data indicate there is a potential or clear impact on approved or authorized medical countermeasures or that have been associated with more severe disease or increased transmission but are no longer detected, or are circulating at very low levels.
Nextstrain	An open-source project and phylogenetic toolset to harness the scientific and public health potential of pathogen genome data ( <a href="http://nextstrain.org/">nextstrain.org/</a> ).
GISAID	The Global Initiative on Sharing All Influenza Data (GISAID) promotes the rapid sharing of data from all influenza viruses and the coronavirus causing coronavirus disease 2019 (COVID-19). This includes genetic sequence and related clinical and epidemiological data ( <a href="http://www.gisaid.org/">www.gisaid.org/</a> ).
Pango nomenclature	Phylogenetic assignment of named global outbreak (Pango) is a dynamic nomenclature to classify genetic lineages for SARS-CoV-2 ( <a href="http://www.pango.network/">www.pango.network/</a> ).
Spike	A viral surface protein that has a key role in viral entry and immune recognition. SARS-CoV-2 spike is involved in receptor recognition and cell membrane fusion.
Antigenic drift	The progressive accumulation of mutations within antigenically important sites in the virus genome/proteome.
Antigenic shift	Step change in the antigenicity of a virus conferring marked escape from previously acquired immunity. Can arise by repeated antigenic drift or recombination.
(h)ACE2	(Human) angiotensin-converting enzyme 2, the cellular receptor for SARS-CoV-2.

the subgenus *Sarbecovirus*, genus *Betacoronavirus*, of the Orthocoronaviridae subfamily (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses 2020). Following emergence, SARS-CoV-2 spread with unexpected ability among humans owing to efficient human-to-human transmission ([www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-](http://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-)

precautions). This immediate “success” in the spread of SARS-CoV-2 is also associated with the relatively promiscuous nature of the virus, evidenced by frequent transmission to mammals such as minks, cats, gorillas, white-tailed deer, and many other species (Telenti et al. 2021). This is presumably linked to the high diversity of bat species (and subspecies) in the horseshoe bat genus, with bat SARS-related coronaviruses be-



**Figure 1.** Phylogenetic tree showing severe acute respiratory syndrome coronavirus 1 and 2 (SARS-CoV-1 and -2) emerging from divergent SARS-related coronavirus lineages. The *Sarbecovirus* subgenus includes, in addition to the SARS-CoV-2 clade, related SARS-like viruses sampled from horseshoe bats, pangolins, civets, and the first SARS-virus clade, SARS-CoV-1 (see key). The tree-branching structure depicts most recent common ancestry relative to the genome sequences at the tips on the *right* (colored by host species). The scale bar shows nucleotide substitutions per site and refers to the horizontal branch lengths. Placing of some of these relationships is approximate owing to recombination in the evolutionary history of these viruses (Boni et al. 2020). (Analysis and images reprinted from Nextstrain.org [nextstrain.org/groups/bla/sars-like-cov] under an Attribution 4.0 International (CC-BY-4.0) license.)

ing unusually genus specialists (Zhou et al. 2020; Holmes et al. 2021). This generalist host-tropism has permitted other mammals to be infected without the need for much, if any, evolutionary change. As is sometimes the case in evolution, adaption to one circumstance has brought with it (blindly) a broader success (i.e., ability to infect many mammal species of which one, humans, because of a highly interconnected world that is near perfectly optimized with high population densities for transmission, permitted efficient virus spread).

To understand SARS-CoV-2’s origin, it has been imperative to identify the closest animal viral genomes. Initially only genome fragments in previously reported recombinant SARS-related coronaviruses (ZXC21, ZC45, and Longquan140) were identified but quickly a relatively close bat virus was reported, RaTG13 (Zhou et al. 2020), although with an unusual receptor-binding domain (RBD) in the spike protein, which was inferred to be due to recombination (Boni et al. 2020). Subsequently more recombinants genetically closer than RaTG13 to SARS-

CoV-2 were identified in Yunnan (RmYN02, RpYN06, and PrC31), outside China in Cambodia and Thailand (RshSTT182, RshSTT200, and RacCS203), and most recently in Laos: BANAL-52, -103, -116, -236 (Temmam et al. 2022). Although genetically close in at least parts of their genomes, all are too divergent from SARS-CoV-2, sharing a most recent ancestor decades ago, to be the direct progenitor (Lytras et al. 2022). Interestingly, all of these recombinant bat coronaviruses have non-SARS-CoV-2-like spike proteins except for one of the Laos viruses, BANAL-20-52. Importantly, BANAL-20-52 demonstrates definitively that the exact SARS-CoV-2 sequence conferring the ability to bind to and use angiotensin-converting enzyme 2 (ACE2) as a receptor exists naturally in bat viruses. The broader ACE2 usage is an ancestral state in the SARS-related coronaviruses (Starr et al. 2022) and was known before the COVID-19 pandemic (Menachery et al. 2015).

The use of the ACE2 receptor is one of the most important adaptations of both SARS-CoV-1 and SARS-CoV-2 for human infection (and other mammals). Whereas ACE2 is also used by other coronaviruses, other receptors support coronavirus entry of host cells (for review, see Millet et al. 2021). The spike proteins of several members of the Alphacoronavirus genus are known to bind to amino peptidase N (APN, CD13) molecules of their respective host species. Murine hepatitis virus (MHV) of the Betacoronavirus genus uses murine carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1). MERS-CoV uses dipeptidyl peptidase 4 (DPP4). Worryingly, a recent MERS-CoV-related virus, a merbecovirus, was shown to be able to use ACE2 (Xiong et al. 2022), emphasizing the need for surveillance for potential spillovers.

The only unique feature of SARS-CoV-2 yet to be identified in a closely related SARS-like virus is the furin cleavage site in spike that is important for virus entry. This was probably acquired by nonhomologous recombination between SARS-related coronaviruses (Lytras 2020) or could be derived from host transcripts (Peacock et al. 2021a). Presence of a furin cleavage site at the S1/S2 junction is not uncommon in human coronaviruses; half of human seasonal

coronaviruses as well as MERS-CoV contain furin cleavage sites (Coutard et al. 2020; Peacock et al. 2021c). The furin cleavage site does appear to be a crucial adaptation of the SARS-CoV-2 to nonbat species as loss of this affects SARS-CoV-2's fitness (Johnson et al. 2021). Collectively, the evidence for the origins of SARS-CoV-2 points conclusively to an animal source via a transient intermediate species, much like the first SARS virus (Holmes et al. 2021; Lytras et al. 2021) and centered on the Huanan market (Worobey et al. 2022).

The existence of SARS-related coronaviruses in horseshoe bats that could use hACE2 was known before the pandemic (Menachery et al. 2015). This research followed the first SARS outbreak in 2002 and again in 2003, which although connected to civets, ferret badgers, and raccoon dogs in wet markets selling live animals, was ultimately shown to be due to spillover of horseshoe bat SARS-related coronaviruses to humans via the intermediate species. Infected civets on farms were also identified in Hubei province (Hu et al. 2005; Shi and Hu 2008), explaining the large geographic distance between the most closely related bat viruses (i.e., animal trading introducing infected animals into city markets) (Lytras et al. 2021). Such a link to an intermediate species has not been made for SARS-CoV-2, because the market was cleared rapidly (Table 1) and none of the animal species reported to be susceptible that were present at the market were tested (Worobey et al. 2022). Serology studies have also demonstrated that SARS-related coronavirus spillovers happen fairly regularly in rural China (Sánchez et al. 2021). That both SARS-CoV-1 and -2 emerged in large urban areas, despite these infections in rural settings, shows the importance of high human population densities to sustain large chains of transmission.

SARS-CoV-2 infection relies on the ACE2 receptor. However, ACE2 is found at relatively low levels in the respiratory tract, which suggests that there may be additional mechanisms facilitating infection; C-type lectin receptors, DC-SIGN, L-SIGN, and the sialic-acid-binding immunoglobulin-like lectin 1 (SIGLEC1) function as attachment receptors to facilitate *trans*-infection, a mechanism of enhancing infection of



susceptible cells by virus captured on myeloid cells (Lempp et al. 2021). In addition to the high-affinity specific binding of the viral spike carboxy-terminal domain (CTD) to protein receptors, coronaviruses can bind with lower affinity to carbohydrates via their spike amino-terminal domain (NTD). This mode of interaction may be a lower affinity, nonspecialized binding module that many coronaviruses retain (Millet et al. 2021). These authors speculate that the NTD may play a critical role when viruses cross the species barrier by allowing an emerging coronavirus to adapt to a new host environment and maintain a minimal level of binding that would allow infection of new host cells via sialic acids, while the CTD gains adaptive mutations that optimize binding to a new host receptor (Millet et al. 2021).

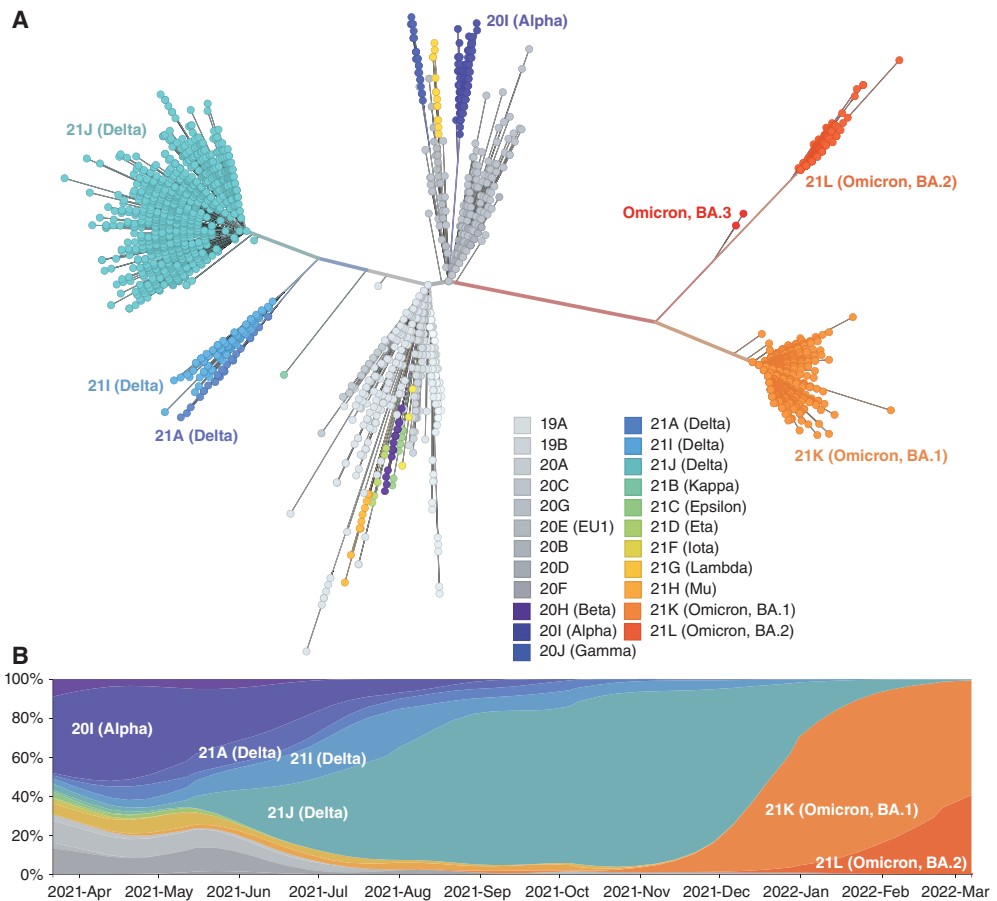
### EVOLUTION OF SARS-CoV-2

The evolution of a novel pathogen involves unpredictable changes in transmissibility and virulence (the degree of harm caused by infection). In the case of SARS-CoV-2, this unpredictability is magnified by it being a new human pathogen with, in most of the world, multiple periods of uncontrolled spread providing ongoing opportunity to acquire adaptations for more efficient transmission in the human population (Martin et al. 2021). Importantly, the oft-repeated statement that viruses inevitably evolve to lower virulence is generally incorrect. Whereas some viruses that have infected humans for many years may seem less virulent, this can simply be because the susceptible individuals died amid an evolutionary process (this is why we can see signatures of virus adaptations in the human genome) (McLaren et al. 2015; Enard et al. 2016) and/or survivors of infection acquired some or complete immunity to infection (i.e., this is a change in the human population not the intrinsic properties of the virus). Generally, there is a tradeoff between transmission and virulence, such that as long as transmission is high, virulence is relatively unconstrained.

SARS-CoV-2 is perceived to be unusual because new human pathogens observed in recent decades often have higher virulence as a result of

a lack of host immunity. However, whereas SARS-CoV-2's virulence is heavily skewed with a strong age, comorbidity, and stochastic effect, the majority of infections—and not discounting the rare but high disease burden of long COVID cases—are associated with relatively mild symptoms or asymptomatic infections. As SARS-CoV-2's transmission rate was already high, there was assumed to be no selective advantage for any decrease in virulence. Nonetheless in the case of Omicron there has been, unpredictably and thankfully given its high rate of transmissibility, a marked drop in the severity of infections owing to a change in preference for upper respiratory infection (Peacock et al. 2021b; McMahan et al. 2022; Meng et al. 2022; Willett et al. 2022). As the VOCs have evolved independently (Fig. 2) in the human population, each discovering different trajectories to greater success relative to early 2020 variants, we cannot assume that a new VOC will not emerge with different virulence properties once again. Whether post-Omicron evolution will be based on this new “milder” form of SARS-CoV-2 or another novel VOC will emerge with novel properties is thus unknown because of the unpredictable nature of SARS-CoV-2 evolution.

The main driver of virus evolution (i.e., reproductive success in terms of onward transmissions) is an interplay of chance epidemiological effects, mutations that impact immune escape, and adaptation to the new host, all occurring in a changing host landscape marked by past exposures, vaccinations, and boosters. An additional wildcard is that viral evolution involves occasional fixation of low-frequency mutations during transmission (i.e., there are strong transmission bottlenecks that dominate viral evolution even during superspreading events) (Hannon et al. 2022). SARS-CoV-2 has repeatedly acquired enhanced transmissibility and immune-evasion properties to enable it to persist in this changing landscape. The first of these was the amino acid replacement D614G in the spike protein that conferred a measurable transmission advantage (Korber et al. 2020). This was the defining amino acid replacement of the B.1 Pango/20A Nextstrain lineage (Box 2). Soon after, the characterization of N439K in the spike receptor-



**Figure 2.** The evolutionary and temporal relationships of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) clades. Since its emergence in late 2019, SARS-CoV-2’s two initial lineages (*A,B*) have continued to diversify, forming novel genetic variants with a subset of “variants of concern” dominating at different times in the pandemic. (A) Unrooted phylogenetic tree showing the main SARS-CoV-2 clades, which are defined by different nomenclature conventions (see Box 2). The tree-branching structure depicts most recent common ancestry relative to the genome sequences at internal nodes or the tips of each branch (colored by Nextstrain and WHO clade designations with Omicron also labeled with its Pango sublineages; see key). Branch lengths correspond to genetic change. (B) Frequencies of SARS-CoV-2 variants colored by clade (see key) from late March 2021 to March 2022. A subset of variants of concern have dominated in the pandemic (Alpha, Delta, and Omicron) with Omicron clades constituting the majority of infections as of March 2022. This visualization is based on a subsample of available full genome data (~600 genomes per continental region with ~400 from the previous 4 mo and ~200 from before this). This subsampling results in a more equitable global sequence distribution. (Analysis and images reprinted from Nextstrain.org [nextstrain.org/ncov/gisaid/global] under an Attribution 4.0 International (CC-BY-4.0) license.)

binding motif (RBM) provided conclusive evidence for a spike protein with enhanced binding affinity for the human ACE2 receptor, while simultaneously reducing the activity of some polyclonal sera from people who had been infected (Thomson et al. 2021). N439K viruses presented

similar in vitro replication fitness and cause infections with similar clinical outcomes compared with wild-type. N439K was the first RBM amino acid replacement, relative to the original genotype SARS-CoV-2 variant used in vaccine preparations, to increase to high frequency and

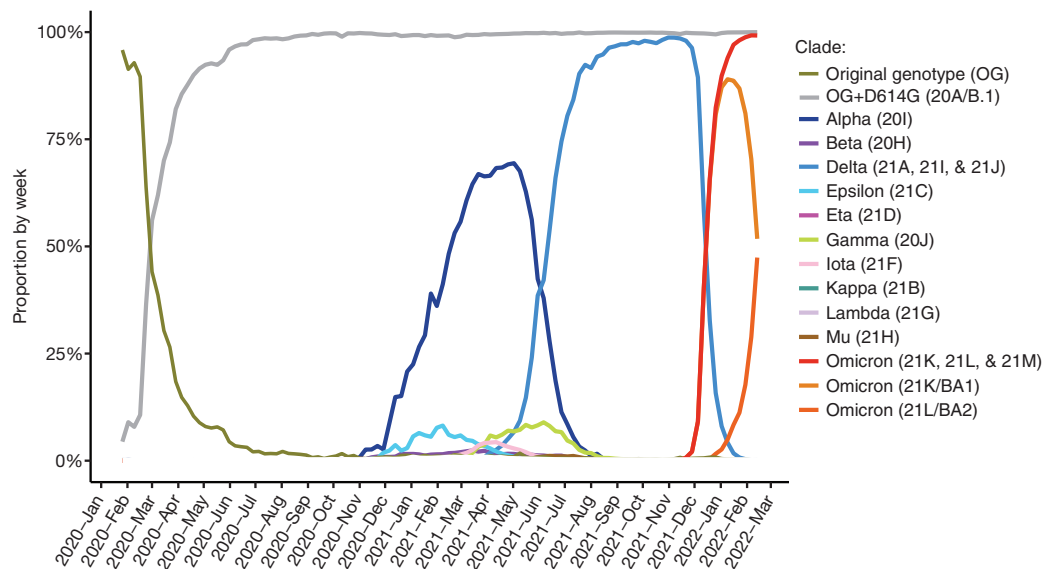


so was a portend for future immune-evasion evolution, clearly demonstrating the malleability of the RBD-ACE2 interaction (Telenti et al. 2021; Thomson et al. 2021).

For most of 2020, SARS-CoV-2 evolved relatively slowly for an RNA virus, which suggested vaccines would easily control infections (Dearlove et al. 2020), but by December 2020 the first VOC was detected (Fig. 3). This more heavily mutated lineage, Pango B.1.1.7/Nextstrain 20I, was subsequently named Alpha by the WHO. Alpha was first identified in the United Kingdom and demonstrated an unpredicted step change in SARS-CoV-2 evolution (Hill et al. 2022). A high number of mutations were apparently generated over a shorter period of time and were thought to be derived from chronic infection of immunocompromised hosts (Corey et al. 2021). Other VOCs (e.g., Beta and Gamma) seem to have evolved in areas with high infection rates, which is consistent with increasing host immunity creating a selective environment for their emergence (Martin et al. 2021). Experimental neutralization studies confirmed these variants would, to some extent, challenge vac-

cines based on the early Wuhan-Hu-1 sequence or other original genotype sequences (Fig. 4). Concomitant with the emergence of these VOCs in late 2020, there was a detectable increase in the positive selection signals in the viral genome that was inferred to be due to increasing host immunity (i.e., SARS-CoV-2 now circulating in a less susceptible human population because of immunity acquired from past infections) (Martin et al. 2021; Maher et al. 2022). The relative influences of this changing host environment/immune landscape versus the intrinsic advantage of increased transmissibility conferred by virus biology are hard to disentangle.

Alpha quickly seeded a second variant-driven global wave of SARS-CoV-2. However, rather than it accumulating further antigenic change and persisting, it was displaced by another even more transmissible variant, 21A/B.1.617.2, named Delta by the WHO. Delta, which emerged in India in late 2020, then seeded a third wave in the pandemic in 2021 (Mlcochova et al. 2021). Even in areas such as South America where variants such as the Mu VOC were spreading, Delta successfully invaded against the Mu background.



**Figure 3.** Dynamics of replacement of variants of concern (VOCs) during the first 2 years of the pandemic. Data are based on over eight million sequences from GISAID EpiCoV 2022\_02\_22. The emergence of D614G is included for completeness. BA.1 (orange) and BA.2 (dark orange) are subvariants of Omicron (dark blue) and are highlighted individually for completeness.



Omicron, characterized by a set of previously observed and novel mutations (Table 2) conferring both a transmission and antigenic shift, has now driven a fourth global wave of SARS-CoV-2. The Omicron wave, initially driven in most regions by the sublineage BA.1 (and BA.1.1, BA.1 with the addition of spike R346K) is now being replaced by the even more transmissible BA.2 sublineage (Yamasoba et al. 2022), expected to become the dominant lineage soon. Omicron is a representative of variants with hypermutated spike proteins and has accommodated some significant evolutionary changes (Martin et al. 2022). A possible source of such heavily mutated viruses is at least one, possibly several or more, infections of immunocompromised individuals (Table 2) or animal spillbacks (see below).

#### DRIVERS OF VIRUS CHANGE

Similarly to other RNA viruses, SARS-CoV-2 uses an RNA-dependent RNA-polymerase but unusually has an additional error-correction enzyme (Smith et al. 2014). This results in a relatively high replication fidelity compared with that of other RNA viruses. As SARS-CoV-2 diversified in 2020, the accumulation of mutations at a rate of about two per month led to the expectation that its evolution would be limited and vaccine-elicited immunity would be relatively long-lasting (Dearlove et al. 2020). It has now become clear SARS-CoV-2 can achieve what appears to be rapid acceleration in its evolutionary rate leading to more heavily mutated variants, some of which have led to the emergence of the VOCs as first observed with Alpha (Hill et al. 2022) and Beta (Martin et al. 2021). It is important to appreciate that all VOCs to date have emerged from the original SARS-CoV-2 2020 genotype, and not one from another (Fig. 2A). Whether this is true episodic evolution or incremental evolution coupled with epidemiological “right place right time” events is hard to discriminate. The potential link to persistent chronic infections, as discussed above for Alpha’s origin (Hill et al. 2022), would be consistent with an incremental process in a single chronic infection that appears episodic once the novel variant seeds a novel lineage/variant.

In the case of Gamma and Delta, it seems more likely they evolved in populations that had experienced high past SARS-CoV-2 infection rates (i.e., mutations presumably accumulated in the context of chains of infections), a process in its outcome indistinguishable from a chronic infection. In all likelihood, both processes are ongoing, for example, generating the divergent Omicron sublineages, BA.1, BA.2, and BA.3 (i.e., community transmission involving multiple immunocompromised individuals may be taking place as noted above).

Relaxed structural constraints provide plasticity to SARS-CoV-2’s spike protein and its ability to accommodate significant amino acid replacements (see Fig. 5; Thomson et al. 2021; McCullum et al. 2022). However, significant mutations also occur in other parts of SARS-CoV-2’s genome that have less well understood functional and antigenic consequences. In particular, the SARS-CoV-2 genome encodes a number of accessory proteins whose coding sequences are interspersed between the structural genes. These proteins are dispensable for virus replication but contribute to immune antagonism and pathogenesis (Lowery et al. 2021). Many of these genes are known to antagonize or evade the interferon response, which may contribute to the delayed interferon expression seen in COVID-19 patients. These proteins act both upstream of interferon production, through evasion and antagonism of pattern recognition receptor (PRR) recognition and signaling, and downstream by directly antagonizing the interferon signaling pathway, facilitating early rapid virus replication (Lowery et al. 2021; Thorne et al. 2022). There is, however, limited understanding of the role of variation in these genes, including for the Delta and Omicron VOCs.

In addition to RNA polymerase errors, recombination between divergent variants is frequent among bat coronaviruses (Goldstein et al. 2021). Specifically, genetic analysis indicates that recombination often occurs in betacoronaviruses (Lai et al. 1985; Keck et al. 1988; Lai and Cavanagh 1997), including natural populations of MERS-CoV (Corman et al. 2014; Dudas and Rambaut 2016), SARS-like coronaviruses (Hon et al. 2008; Boni et al. 2020), and SARS-CoV-2

**Table 2.** Mutational profile of Omicron versus sarbecoviruses clades, variants of interest (VOIs)/variants of concern (VOCs), and mutation profiles in immunocompromised hosts

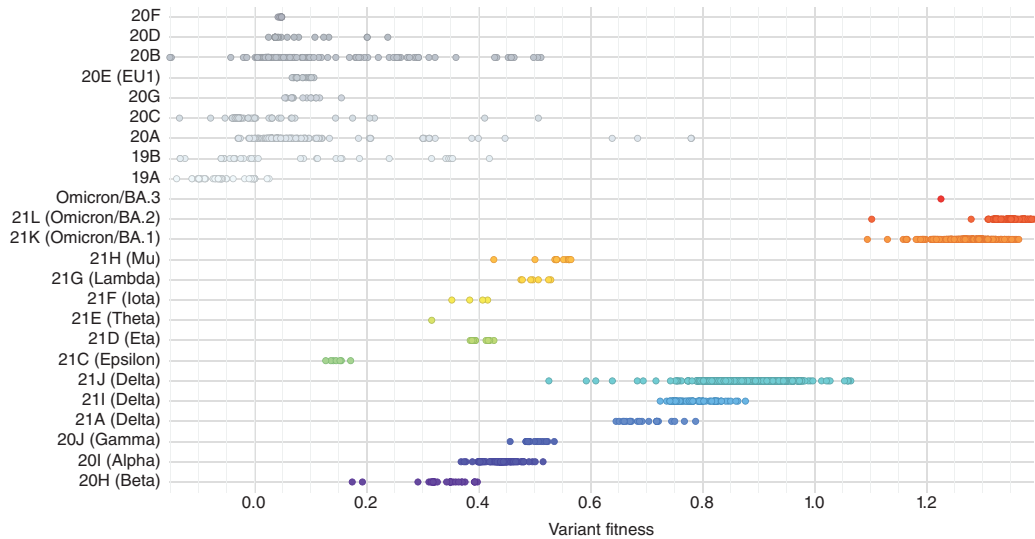
Omicron (BA.1)	Sarbecovirus				VOCs				Immunocompromised host				
	Clade 1	Clade 2	Clade 3	Alpha	Beta	Gamma	Delta	VOIs	Cele et al. (2022)	Kemp et al. (2021)	Choi et al. (2020)	Riddell et al. (2022)	
A67V	x							x					
H69del	x		x					x		x			
V70del	x		x					x		x			
T95I		x						x				(x)	(x)
G142D		x						G142V			x	(x)	(x)
V143del		x								x			
Y144del	x			x							x		
Y145del	x										x		
N211del		x											
L212I													
ins214EPE													
G339D													
S371L													
S373P													
S375F													
K417N								x					
N440K									x				
G446S	x												
S477N													
T478K	x							x					
E484A													
Q493R	x												
G496S													
Q498R													
N501Y													
Y505H	x												
T547K													

*Continued*

**Table 2. Continued**

	Sarbecovirus				VOCs				Immunocompromised host					
	Clade 1	Clade 2	Clade 3	Alpha	Beta	Gamma	Delta	VOIs	Cele et al. (2022)	Kemp et al. (2021)	Choi et al. (2020)	Riddell et al. (2022)		
Omicron (BA.1)				x	x	x	x	x				x	x	x
D614G														
H655Y		x												
N679K														
P681H				x				x				x	x	x
N764K														
D796Y	x								x	D796H/Y				
N856K														
Q954H														
N969K														
L981F														
Additional in immunocompromised									R190K	W64G	F486I	N30T	A570D	(F791F)
									L244del	Y200H	S494P	N450D	T716I	S254SF
									D427Y	T240I	I870V	A475V	S982A	A570D
									A475V	P330S	A1020S	F490Y	D1118H	T716I
									F490S		S494P	S982A	A570D	F486
									R682W		D337Y	D1118H	T716I	
									A1078V		F888L	S982A	D1118H	
														S982A
														D1118H

23 out of 37 Omicron 21M/BA.1 spike amino acid replacements have been observed in VOCs/VOIs or in the spike proteins of other sarbecoviruses (Cameroni et al. 2022). The mutational profile reported in a limited number of immunocompromised hosts with chronic infection is generally distinct (Choi et al. 2020; Kemp et al. 2021; Cele et al. 2022; Maponga et al. 2022; Riddell et al. 2022). Representative sequences of sarbecovirus clades are as described in Lu et al. (2020).

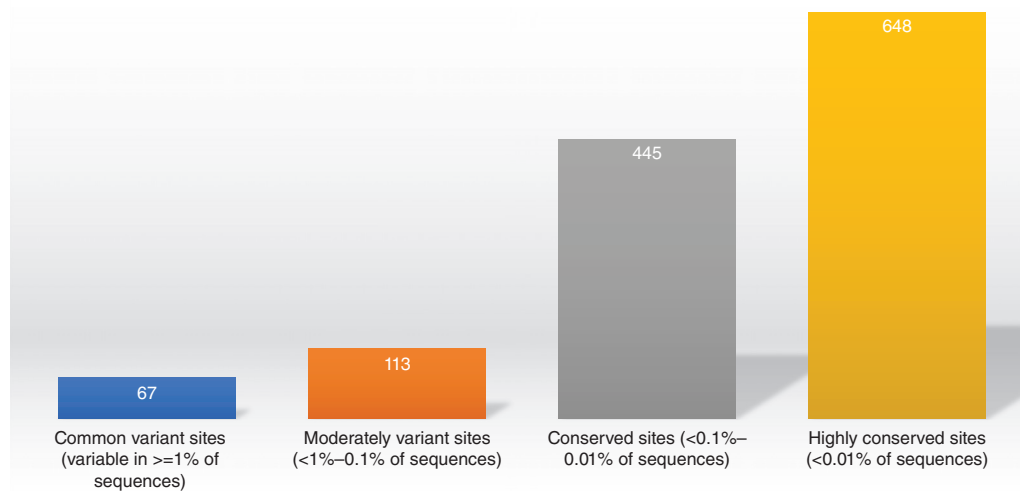


**Figure 4.** Variant fitness of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants. Shown is “mutational”/variant-level fitness measured by the increase in average transmission for different SARS-CoV-2 variants colored by clade relative to an early 2020 genotype: Wuhan-Hu-1 (Obermeyer et al. 2021). (Analysis and image reprinted from Nextstrain.org [nextstrain.org/ncov/gisaid/global] under an Attribution 4.0 International (CC-BY-4.0) license.)

(Jackson et al. 2021). Studies of SARS-CoV-2 viruses sampled from late 2020 and early 2021 in the United Kingdom identified genomes that carry mutations characteristic of the Alpha VOC but lack the full complement of lineage-defining mutations. Instead, the remainder of their genomes share contiguous genetic variation with non-Alpha viruses circulating in the same geographic area at the same time as the recombinants (Jackson et al. 2021). Specifically, four of eight described recombinants led to onward transmission. The transmitted recombinants inherited the more transmissible Alpha spike gene (Jackson et al. 2021). Similarly, recombination has been observed in New York (Wertheim et al. 2022). There are currently incidences of coinfection with SARS-COV-2 Omicron and Delta variants (Rockett et al. 2022) as well as potential recombinants (Roemer 2022). Currently, there is particular attention to several recombinant lineages being detected involving Omicron variants. Some are combinations of Delta and Omicron, for example, XD and XF (designations from the Pango nomenclature). XD has an Omicron/BA.1 S gene incorporated into a Delta

genome and is present in several European countries. XF caused a small cluster in the UK but has not been detected since February 15. XE is an Omicron/BA.1 and BA.2 recombinant, with the majority of the genome, including the S gene, belonging to BA.2. XE shows evidence of community transmission within England, and recent estimates (March 2022) suggest that it has a growth rate of 10% above that of BA.2 ([https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1063424/Tech-Briefing-39-25March2022\\_FINAL.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1063424/Tech-Briefing-39-25March2022_FINAL.pdf)). More generally, recombination breakpoints are observed disproportionately in or beside the spike protein region (Turkahia et al. 2021). Whereas levels of community transmission of recombinants are at present low relative to the emergence of new mutations, as there are more genetically divergent variants (in humans or other animals), recombination will make new combinations that pose risk of new properties. Whereas most mutations are deleterious and selected against, or are neutral, some will enhance virus biology and can be selected for (positive selection). These adaptive evolution forces are driven by and translate into





**Figure 5.** Conservation of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike protein. 180 of 1273 (14%) amino acid positions of the SARS-CoV-2 spike protein are identified as moderately variant or common sites of variation, while 86% remain conserved or highly conserved. Data from analysis of over nine million sequences from GISAID on March 2022.

changes in transmissibility, virulence, and immune escape.

### Transmissibility

Until late 2021, it had been mainly transmissibility that underpinned SARS-CoV-2's success. Transmission success was the key to original SARS-CoV-2 genotype, B.1 (defined by S: D614G), Alpha, and Delta's spread and dominance at different points in the pandemic (Fig. 4). However, Omicron has both a transmission advantage and a step change, a "shift," in its antigenicity. Most research emphasis has been on the capacity of some of the variants to modulate the affinity of interaction with ACE2. For example, deep mutagenesis scanning of the RBD of the SARS-CoV-2 spike glycoprotein experimentally measured how all amino acid replacements affect synthesis and stability of folded protein and its affinity for ACE2 (Starr et al. 2020). A substantial number of mutations are well tolerated or even enhance ACE2 binding, including at ACE2-interface residues that vary across SARS-related coronaviruses (Starr et al. 2020). Omicron's RBD binds to human ACE2

with enhanced affinity relative to the Wuhan-Hu-1 RBD and has acquired the ability to bind to mouse ACE2 (Cameroni et al. 2022). However, other mechanisms may contribute to transmissibility, for example, by increasing the viral load in the respiratory tract of infected individuals. The mechanism may involve viral replication and burst size, as well as escape from innate immunity. There is, as described in the previous sections, much still to understand about these cellular and molecular drivers of infection.

### Virulence

Epidemiological evidence indicates decoupling of dominant drivers of spread (younger people) and virulence (degree of harm caused in older people). This translates into higher disease severity in those who transmit the least. More generally, in respiratory viruses, severe disease is associated with the invasion of and replication in the lower respiratory tract. In the case of SARS-CoV-2, mutations that further optimize the use of human ACE2 (present in both the upper and lower respiratory tract) or alter the capacity of coreceptors to influence tropism and

**Table 3.** Clinical significance of variants of concern (VOCs)

VOC/variant being monitored (VBM)	Transmissibility	Disease burden/virulence	Neutralization by monoclonal antibodies	Vaccine efficacy
Alpha	Increased ~50%	Increased	Susceptible	
Beta	Increased ~25%	Increased	Susceptible	Reduced
Delta	Increased ~100%	Increased	AY.1 and AY.2 lineages are not susceptible to some clinically used monoclonal antibodies	Reduced
Epsilon		Unknown		
Gamma	Increased ~30%	Possibly increased	Not susceptible to some clinically used monoclonal antibodies	Modest reduction
Kappa	Increased	Unknown	Susceptible	Reduced
Mu	Increased	Unknown	Susceptible	Reduced
Omicron	Increased >100%	Decreased for BA.1 and BA.2	Not susceptible to most, but not all clinically used monoclonal antibodies	Greatly reduced

Some of the estimates are confounded by VOCs/variants under investigation (VUIs) entering a human population that has changing levels of immune competence (Campbell et al. 2021; Coutinho et al. 2021; Davies et al. 2021; Earnest et al. 2021; McCallum et al. 2021; Radvak et al. 2021; Uriu et al. 2021; US Centers for Disease Control and Prevention 2021; Walker et al. 2021; European Centre for Disease Prevention and Control 2022; Suzuki et al. 2022; Twohig et al. 2022).

infection (Lempp et al. 2021) are likely to increase both transmission and virulence (Telenti et al. 2021). Interferon antagonism/evasion is a well-known virulence determinant, as are interactions between virus and host that modulate the host response. This is not exclusively mediated by Spike-ACE2 interactions and likely is heavily influenced by other viral proteins. In the discussion of virulence, it is important to separate the intrinsic virulence of the virus, and the effective virulence in the population that has varying levels of immune competence (previous infection and vaccination). Current virulence of Omicron reflects to a large extent the protective effect of acquired immunity, although there is consensus that Omicron also displays decreased intrinsic virulence (Table 3).

### Immune Pressure

Antigenic change, in particular in the immunodominant RBD and NTD domains of the spike protein are the result of a humoral response that is dominant at those sites (Harvey et al. 2021). The humoral response is also the main driver of vaccine efficacy, as measured by Gilbert et al.

(2021), who reported that neutralizing antibodies mediate about two-thirds of an mRNA vaccine efficacy (Gilbert et al. 2021). There is less information about the role of CD4 and CD8 T-cell responses (Grifoni et al. 2021). Many such T-cell epitopes have been mapped experimentally, but there has been limited information on escape mutations and corresponding functional outcomes (Geers et al. 2021). Specifically, VOCs may not escape T-cell responses in COVID-19 convalescent donors and vaccine recipients (Geers et al. 2021). However, a recent study identified viral mutations in HLA-I-restricted epitopes leading to reduced recognition by CD8 T cells but reported evidence of these mutations being transmitted (Dolton et al. 2021). Dolton et al. (2021) also described a single amino acid replacement (spike P272L) that is not recognized by CD8 T cells targeting the most prevalent epitope in spike restricted by the most common HLA-I across the population (Dolton et al. 2021). Whereas these authors did not attribute the emergence and propagation of the spike P272L mutation to escape, they posited that mutations could evade immunodominant T-cell responses through population-frequent

## BOX 2. SEVERE ACUTE RESPIRATORY SYNDROME CORONAVIRUS 2 (SARS-COV-2) NOMENCLATURE

Increasing diversification characterizes the evolution of viruses and novel variants have been a major driver of virus spread in the coronavirus disease 2019 (COVID-19) pandemic. A coherent system for naming the growing number of phylogenetic lineages that compose the population diversity of SARS-CoV-2 was thus required (Rambaut et al. 2020). Based on a maximum likelihood tree, Rambaut and colleagues (2020) established lineage definitions based on phylogenetic evidence of emergence from an ancestral lineage into a geographically distinct population, suggesting substantial onward transmission in that population or at later times. This approach is the basis of the designation and naming of Phylogenetic Assignment of Named Global Outbreak LINEages (Pango lineage; [www.pango.network/](http://www.pango.network/)). An alternative nomenclature adopted by Nextstrain implemented a strategy of “year-letter” names borne out of work with seasonal influenza (Bedford et al. 2021). Another nomenclature was proposed for the tracking of ongoing convergent evolution of N501Y lineages (Martin et al. 2021). The Global Initiative on Sharing All Influenza Data (GISAID) introduced a nomenclature system for major clades based on marker mutations within eight high-level phylogenetic groupings (GISAID 2021). To reconcile these nomenclatures and bring focus to naming “variants of concern” (VOCs) and of interest (VOIs), and to assist with public discussions of variants, a COVID-19 expert group convened by the World Health Organization (WHO) recommended using letters of the Greek alphabet, which are easier and nonstigmatizing labels for VOCs and other letters for VOIs or those under monitoring (WHO 2022).

HLA alleles, with possible consequences for vaccine design. Recent studies assessed how Omicron mutations affect SARS-CoV-2 spike T-cell responses induced upon vaccination or infection (Keeton et al. 2021). They found that 70%–80% of the CD4 and CD8 T-cell response to spike was maintained across study groups. Moreover, the magnitude of Omicron cross-reactive T cells was similar to that of the Beta and Delta variants, despite Omicron harboring considerably more mutations. More recent work supports the idea that T cells may protect from severe disease (De Marco et al. 2021; GeurtsvanKessel et al. 2022; Liu et al. 2022; Tarke et al. 2022). As described earlier, there is also emerging data regarding the genetic and functional hallmarks of escape from innate immunity/interferon response (Thorne et al. 2022).

### CLINICAL SIGNIFICANCE

Until November 2021 and Omicron’s emergence, SARS-CoV-2 had not shown any changes consistent with attenuation but rather well-documented increase in transmissibility coupled with indications for increased disease severity (Table 3). Owing to the importance of vaccina-

tions as the primary control measure, there has been extensive focus on the consequences of new variants for immune control, in particular a loss of neutralization activity of plasma from vaccinated individuals. In particular, Omicron shows immune escape from the neutralization activity of polyclonal sera, a large number of therapeutic monoclonal antibodies, and vaccination (Cameroni et al. 2022; Carreño et al. 2022). In vitro and animal model studies indicate reduced viral load, lower respiratory tree/lung infectivity, and inflammation caused by Omicron (Abdelnabi et al. 2021; Bentley et al. 2021; Diamond et al. 2021; McMahan et al. 2022).

As of today, most models of disease burden use a number of assumptions regarding the proportion of infected individuals that will develop symptomatic disease (generally estimated at 20%), and the proportion of symptomatic infected individuals that will be hospitalized (generally estimated at 10%). However, the probability that a COVID-19 infection is symptomatic is difficult to estimate, and a wide range of values have been suggested (Subramanian et al. 2021, and references therein). Estimates from cruise ship outbreaks, Wuhan evacuees, long-term

care facilities, and contact tracing of index cases may not be representative of the general population (Subramanian et al. 2021). Mortality rates are highly dependent on age (O’Driscoll et al. 2021). For the vaccinated population, the proportions are estimated on the basis of the same rates multiplied by the estimated rates of vaccine efficacy against infection and by the rates of vaccine efficacy against hospitalization, both of which depend on vaccine type and time since vaccination.

Antigenic drift from one year to the next has been central to decision-making in vaccine strain selection and updating for influenza (Boni 2008). Transmission of influenza is associated with the capacity to undergo antigenic drift and periodic antigenic shift associated with reassortment. The degree of antigenic variation or “antigenic distance” between the hemagglutinin and neuraminidase proteins of influenza is the basis for needing to update the composition of influenza vaccines frequently (Gupta et al. 2006). For influenza, whenever the titers of antisera generated against vaccine strains and tested against circulating strains change over 8- to 10-fold in the hemagglutination inhibition assay, it typically indicates the need to update the vaccine composition. As indicated in Table 2 for SARS-CoV-2, VOCs are associated with a reduced neutralizing titer of antisera from vaccinated donors comparable to the extent of antigenic drift in influenza viruses that typically requires a change in the viruses selected for vaccine production (Telenti et al. 2021). Omicron has led to a discussion of antigenic “drift” (Yewdell 2021) versus “shift” in SARS-CoV-2 (Cameroni et al. 2022). Antigenic shift in influenza is due to reassortment in the context of influenza segmented genome, a novel antigenically distinct hemagglutinin segment appearing from an animal reservoir. Adapting the concept of antigenic shift to SARS-CoV-2 requires integrating other mechanisms, such as evolution in an immunocompromised host, recombination or human–animal transmission, which may result in considerable sequence evolution and effective antigenic shift when this novel variant spreads in the human population. Such antigenic shift in Omicron led to this VOC

readily causing reinfections in the fully vaccinated/boosted (Cameroni et al. 2022).

## EVOLUTION IN ANIMAL RESERVOIRS

The host range of SARS-CoV-2 extends to a variety of mammalian species (Mallapaty 2021; Prince et al. 2021). A recent review (Sharun et al. 2021) indicates that cats, ferrets, raccoon dogs, cynomolgus macaques, rhesus macaques, white-tailed deer, rabbits, Egyptian fruit bats, and Syrian hamsters are susceptible to experimental SARS-CoV-2 infection, and that cat-to-cat and ferret-to-ferret transmission can take place via contact and air. SARS-CoV-2 transmission from animals-to-humans has been reported from mink to humans in mink farms (Oude Munnink et al. 2021), from hamsters to humans (Yen et al. 2022), and, more recently, in epidemiologically linked white-tailed deer and human cases from the same geographic region and sampling period (Pickering et al. 2022). Viral evolution can occur in animal hosts, generating a suite of genomic changes in addition to those seen during human-to-human transfer (Telenti et al. 2021). There are two risks for the human population in relation to animal reservoirs. First, establishment of SARS-CoV-2 in other species (reverse zoonosis) could provide a refuge for the virus to reemerge in human populations in an evolutionarily distinct form—for example, upon waning of vaccine coverage or diminished natural- or vaccine-induced immunity that occurs over time (Telenti et al. 2021). Second, animal viruses may also diverge antigenically owing to different adaptive evolution and immune pressure. This could lead to a population of individuals born in postpandemic years who are susceptible to infection with animal SARS-CoV-2 viruses that are antigenically related to the original SARS-CoV-2 pandemic strain or extensively divergent from the original strains (Telenti et al. 2021).

Two recent reports propose that viral strains with high numbers of mutations may reflect infection of humans from an animal reservoir previously established via reverse zoonosis. Wei et al. (Wei et al. 2021) impute that Omicron may have originated from mice. Pickering et al. (Pickering



A. Telenti et al.

et al. 2022) propose that a new complex virus in humans originated in white-tailed deer. Both works use analysis of mutational spectra to capture rates of A > I, or C > U editing that are a reflection of the species' cellular environment and in particular the differences in the activity of viral restriction systems such as that provided by the ADAR (Pfaller et al. 2021) and APOBEC (Harris and Dudley 2015) families of proteins (Di Giorgio et al. 2020). We cannot exclude potential recombination events in the future between SARS-CoV-2 and other human coronaviruses (Telenti et al. 2021). Unfortunately, the number of sequences of SARS-CoV-2 from nonhuman animal sources is limited. Based on the metadata from GISAID (February 22, 2022), there were 1785 sequences from 38 nonhuman hosts. Most ( $n = 1212$ ) correspond to mink, followed by white-tailed deer ( $n = 145$ ), various felines ( $n = 251$ ), dog ( $n = 74$ ), non-human primates ( $n = 19$ ), and low numbers in other species (mice, hamsters, bats, pangolin, ferrets, and others). Thus, there is a need for more work on reverse zoonosis and reentry into the human population.

### PREDICTED EVOLUTION OF THE PANDEMIC

The SARS-CoV-2 pandemic has been the strongest use case for massive scale sequencing efforts and controlled access to pathogen data ever, with data centralized in GISAID (Elbe and Buckland-Merrett 2017). Such data sets can be explored by a machine learning method to generate experimentally verifiable predictions of potential practical value for public health. In addition, these resources and tools can be made central to the process of vaccine and therapy development, increasing the resilience to target variability and escape, and thus durability of efficacy for drugs against viral targets. Several groups have independently demonstrated the feasibility of models that anticipate viral evolution. Maher et al. used machine learning to analyze different streams of epidemiological, viral, immunological, and experimental data (Maher et al. 2022), Hie et al. used natural language processing applied to protein sequences (Hie et al. 2021), Obermeyer et al. used a hierarchical

Bayesian multinomial logistic regression model (Obermeyer et al. 2021, 2022), Rodriguez-Rivas et al. used epistatic models (Rodriguez-Rivas et al. 2022), and Genovese et al. applied ab initio quantum mechanical modeling (Genovese et al. 2021). As a proof-of-concept of using these models practically, Beguir et al. reported the implementation of Hie et al. natural language processing as an early warning system for vaccine planning at BioNTech (Beguir et al. 2021). Whereas it is early to assess the performance of forecasting models, the above reports identified a limited number of functionally important mutations that were eventually shared by convergent evolution by several VOCs. Models were generally capable of predicting future mutations over the ensuing 2–4 months with high accuracy. However, Omicron, presumably because of its differences from previous VOCs, was less predictable. With more data and ongoing iteration of experimental studies, and sophistication of machine learning models, predicting virus evolution can only improve.

Whereas most attention is dedicated to the evolution of viral proteins, there is also an interest in understanding the constraints to viral evolution and fitness imposed by RNA properties. This field has made considerable progress in experimental definition of 2D (and 3D) RNA structures (Smyth et al. 2018; Cao et al. 2021) with particular application for viral RNAs (Watts et al. 2009; Snoeck et al. 2011; Madden et al. 2020; Manfredonia et al. 2020). There is as yet no conclusive evidence to associate mutation in constrained, conserved RNA structures and phenotypic consequences, such as viral fitness of epidemiological advantage in SARS-CoV-2 (Maher et al. 2022).

### CONCLUSION

Whereas SARS-CoV-2 caught the world unaware, the current global situation is characterized by a process of ongoing adaptation to humans by a virus that has exhibited ever higher step changes in transmissibility, coupled with the capacity for antigenic drift (Yewdell 2021), and now, with Omicron, what can only be described as antigenic shift, a step change in its



immune evasion abilities (Cameroni et al. 2022; Willett et al. 2022). This process has led to virus spread in the case of Omicron BA.2/21L that is more or less uncontrolled despite high levels of individual and population-level immunity, and now challenges those countries that have to date maintained zero-COVID policies or who have low vaccination rates in susceptible populations. Whereas it is feasible an equilibrium will be reached with SARS-CoV-2 becoming a more seasonal virus, with a manageable burden of disease in an increasingly immune-competent human population, there is still the concern that SARS-CoV-2 may evolve novel virulence factors, for example, by changing cell/tissue tropism, evading innate immunity, or establishing longer infectious periods (e.g., in diverse tissue compartments) (Deinhardt-Emmer et al. 2021). This, particularly while SARS-CoV-2 has unexplored variability to exploit, may potentially lead to changes in virus biology and, as already observed with the VOCs, unpredictable disease severity (e.g., in the adult strata of the population as opposed to being mostly confined to older age groups). Future vaccine development and deployment of improved therapeutic agents will, if used widely and strategically, allow us to control this new, often deadly, human pathogen. Whereas it is human nature to be optimistic about the short- and longer-term evolution of this new human pathogen, it would be wise to continue to minimize infections wherever possible with available tools (behavioral, medical, and therapeutic) and prepare for future VOCs: Preparedness is preparation and readiness for worst-case scenarios based on lessons learned from 2 years of an ongoing pandemic.

#### COMPETING INTEREST STATEMENT

A.T. is an employee and shareholder of Vir Biotechnology.

#### ACKNOWLEDGMENTS

The authors thank all of the researchers who have shared genome data openly via the Global Initiative on Sharing All Influenza Data (GISAID). We thank J. di Iulio and L. Soriaga for

providing the plots and data in Figures 3 and 5. D.L.R. acknowledges support of the Medical Research Council (MC\_UU\_12014/12), the Wellcome Trust (220977/Z/20/Z), and G2P-UK National Virology Consortium (MR/W005611/1). We thank Herbert (Skip) Virgin, Richard Sever, Stephen Goldstein, and Angela Rasmussen for helpful comments on the manuscript.

#### REFERENCES

- Abdelnabi R, Foo CS, Zhang X, Lemmens V, Maes P, Slechten B, Raymenants J, André E, Weynand B, Dalmier K, et al. 2021. The omicron (B.1.1.529) SARS-CoV-2 variant of concern does not readily infect Syrian hamsters. *bioRxiv* doi:10.1101/2021.12.24.474086
- Bedford T, Hodcroft EB, Neher RA. 2021. Updated Nextstrain SARS-CoV-2 clade naming strategy. *Nextstrain Blog*, January 6, 2021. <https://nextstrain.org/blog/2021-01-06-updated-SARS-CoV-2-clade-naming> [Accessed February 14, 2022].
- Beguir K, Skwark MJ, Fu Y, Pierrot T, Lopez Carranza N, Laterre A, Kadri I, Lui BG, Sängler B, Liu Y, et al. 2021. Early computational detection of potential high risk SARS-CoV-2 variants. *bioRxiv* doi:10.1101/2021.12.24.474095
- Bentley EG, Kirby A, Sharma P, Kipar A, Mega DF, Bramwell C, Penrice-Randal R, Prince T, Brown JC, Zhou J, et al. 2021. SARS-CoV-2 Omicron-B.1.1.529 variant leads to less severe disease than Pango B and Delta variants strains in a mouse model of severe COVID-19. *bioRxiv* doi:10.1101/2021.12.26.474085
- Boni MF. 2008. Vaccination and antigenic drift in influenza. *Vaccine* **26** (Suppl 3): C8–C14. doi:10.1016/j.vaccine.2008.04.011
- Boni MF, Lemey P, Jiang X, Lam TT, Perry BW, Castoe TA, Rambaut A, Robertson DL. 2020. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat Microbiol* **5**: 1408–1417. doi:10.1038/s41564-020-0771-4
- Cameroni E, Bowen JE, Rosen LE, Saliba C, Zepeda SK, Culap K, Pinto D, VanBlargan LA, De Marco A, di Iulio J, et al. 2022. Broadly neutralizing antibodies overcome SARS-CoV-2 Omicron antigenic shift. *Nature* **602**: 664–670. doi:10.1038/s41586-021-04386-2
- Campbell F, Archer B, Laurenson-Schafer H, Jinnai Y, Kornings F, Batra N, Pavlin B, Vandemaale K, Van Kerkhove MD, Jombart T, et al. 2021. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. *Euro Surveill* **26**: 2100509. doi:10.2807/1560-7917.ES.2021.26.24.2100509
- Cao C, Cai Z, Xiao X, Rao J, Chen J, Hu N, Yang M, Xing X, Wang Y, Li M, et al. 2021. The architecture of the SARS-CoV-2 RNA genome inside virion. *Nat Commun* **12**: 3917. doi:10.1038/s41467-021-22785-x
- Carreño JM, Alshammay H, Tcheou J, Singh G, Raskin A, Kawabata H, Sominsky L, Clark J, Adelsberg DC, Bielak D, et al. 2022. Activity of convalescent and vaccine serum



A. Telenti et al.

- against SARS-CoV-2 Omicron. *Nature* **602**: 682–688. doi:10.1038/s41586-022-04399-5
- Cele S, Karim F, Lustig G, San JE, Hermanus T, Tegally H, Snyman J, Moyo-Gwete T, Wilkinson E, Bernstein M, et al. 2022. SARS-CoV-2 prolonged infection during advanced HIV disease evolves extensive immune escape. *Cell Host Microbe* **30**: 154–162.e5. doi:10.1016/j.chom.2022.01.005
- Choi B, Choudhary MC, Regan J, Sparks JA, Padera RF, Qiu X, Solomon IH, Kuo HH, Boucau J, Bowman K, et al. 2020. Persistence and evolution of SARS-CoV-2 in an immunocompromised host. *N Engl J Med* **383**: 2291–2293. doi:10.1056/NEJMc2031364
- Corey L, Beyrer C, Cohen MS, Michael NL, Bedford T, Roland M. 2021. SARS-CoV-2 variants in patients with immunosuppression. *N Engl J Med* **385**: 562–566. doi:10.1056/NEJMs2104756
- Corman VM, Ithete NL, Richards LR, Schoeman MC, Preiser W, Drosten C, Drexler JF. 2014. Rooting the phylogenetic tree of Middle East respiratory syndrome coronavirus by characterization of a conspecific virus from an African bat. *J Virol* **88**: 11297–11303. doi:10.1128/JVI.01498-14
- Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. 2020. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* **5**: 536–544. doi:10.1038/s41564-020-0695-z
- Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG, Decroly E. 2020. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res* **176**: 104742. doi:10.1016/j.antiviral.2020.104742
- Coutinho RM, Marquitti FMD, Ferreira LS, Borges ME, da Silva RLP, Canton O, Portella TP, Poloni S, Franco C, Plucinski MM, et al. 2021. Model-based estimation of transmissibility and reinfection of SARS-CoV-2 P.1 variant. *Commun Med* **1**: 48. doi:10.1038/s43856-021-00048-6
- Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, Pearson CAB, Russell TW, Tully DC, Washburne AD, et al. 2021. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **372**: eabg3055. doi:10.1126/science.abg3055
- Dearlove B, Lewitus E, Bai H, Li Y, Reeves DB, Joyce MG, Scott PT, Amare MF, Vasan S, Michael NL, et al. 2020. A SARS-CoV-2 vaccine candidate would likely match all currently circulating variants. *Proc Natl Acad Sci* **117**: 23652–23662. doi:10.1073/pnas.2008281117
- Deinhardt-Emmer S, Wittschieber D, Sanft J, Kleemann S, Elschner S, Haupt KF, Vau V, Häring C, Rödel J, Henke A, et al. 2021. Early postmortem mapping of SARS-CoV-2 RNA in patients with COVID-19 and the correlation with tissue damage. *eLife* **10**: e60361. doi:10.7554/eLife.60361
- De Marco L, D’Orso S, Pirronello M, Verdiani A, Termine A, Fabrizio C, Capone A, Sabatini A, Guerrera G, Placido R, et al. 2021. Preserved T cell reactivity to the SARS-CoV-2 Omicron variant indicates continued protection in vaccinated individuals. bioRxiv doi:10.1101/2021.12.30.474453
- Diamond M, Halfmann P, Maemura T, Iwatsuki-Horimoto K, Iida S, Kiso M, Scheaffer S, Darling T, Joshi A, Loeber S, et al. 2021. The SARS-CoV-2 B.1.1.529 Omicron virus causes attenuated infection and disease in mice and hamsters. *Res Sq* doi:10.21203/rs.3.rs-1211792/v1
- Di Giorgio S, Martignano F, Torcia MG, Mattiuz G, Conticello SG. 2020. Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. *Sci Adv* **6**: eabb5813. doi:10.1126/sciadv.abb5813
- Dolton G, Rius C, Hasan MS, Wall A, Szomolay B, Behiry E, Whalley T, Southgate J, Fuller A, The COVID-19 Genomics UK consortium, et al. 2021. Emergence of immune escape at dominant SARS-CoV-2 killer T-cell epitope. medRxiv doi:10.1101/2021.06.21.21259010.
- Dudas G, Rambaut A. 2016. MERS-CoV recombination: implications about the reservoir and potential for adaptation. *Virus Evol* **2**: vev023. doi:10.1093/ve/vev023
- Earnest R, Uddin R, Matluk N, Renzette N, Siddle KJ, Loreth C, Adams G, Tomkins-Tinch CH, Petrone ME, Rothman JE, et al. 2021. Comparative transmissibility of SARS-CoV-2 variants Delta and Alpha in New England, USA. medRxiv doi:10.1101/2021.10.06.21264641
- Elbe S, Buckland-Merrett G. 2017. Data, disease and diplomacy: GISAID’s innovative contribution to global health. *Glob Chall* **1**: 33–46. doi:10.1002/gch2.1018
- Enard D, Cai L, Gwennap C, Petrov DA. 2016. Viruses are a dominant driver of protein adaptation in mammals. *eLife* **5**: e12469. doi:10.7554/eLife.12469
- European Centre for Disease Prevention and Control. 2022. SARS-CoV-2 variants of concern as of 24 March 2022. ECDC, February 10, 2022. Last modified March 25, 2022. <https://www.ecdc.europa.eu/en/covid-19/variants-concern> [Accessed March 31, 2022].
- Geers D, Shamier MC, Bogers S, den Hartog G, Gommers L, Nieuwkoop NN, Schmitz KS, Rijsbergen LC, van Osch JAT, Dijkhuizen E, et al. 2021. SARS-CoV-2 variants of concern partially escape humoral but not T-cell responses in COVID-19 convalescent donors and vaccine recipients. *Sci Immunol* **6**: eabj1750. doi:10.1126/sciimmunol.abj1750
- Genovese L, Zaccaria M, Farzan M, Johnson W, Momeni B. 2021. Investigating the mutational landscape of the SARS-CoV-2 Omicron variant via ab initio quantum mechanical modeling. bioRxiv doi:10.1101/2021.12.01.470748
- GeurtsvanKessel CH, Geers D, Schmitz KS, Mykytyn AZ, Lamers MM, Bogers S, Scherbeijn S, Gommers L, Sabellelles RSG, Nieuwkoop NN, et al. 2022. Divergent SARS CoV-2 Omicron-reactive T and B cell responses in COVID-19 vaccine recipients. *Sci Immunol* **7**: eabo2202. doi:10.1126/sciimmunol.abo2202
- Gilbert PB, Montefiori DC, McDermott A, Fong Y, Benkeser D, Deng W, Zhou H, Houchens CR, Martins K, Jayashankar L, et al. 2021. Immune correlates analysis of the mRNA-1273 COVID-19 vaccine efficacy trial. medRxiv doi:10.1101/2021.08.09.21261290
- GISAID. 2021. Clade and lineage nomenclature. GISAID, March 2, 2021. <https://www.gisaid.org/resources/statements-clarifications/clade-and-lineage-nomenclature-aids-in-genomic-epidemiology-of-active-hcov-19-viruses/> [Accessed February 14, 2022].
- Goldstein SA, Brown J, Pedersen BS, Quinlan AR, Elde NC. 2021. Extensive recombination-driven coronavirus diver-

- sification expands the pool of potential pandemic pathogens. *bioRxiv* doi:10.1101/2021.02.03.429646
- Grifoni A, Sidney J, Vita R, Peters B, Crotty S, Weiskopf D, Sette A. 2021. SARS-CoV-2 human T cell epitopes: adaptive immune response against COVID-19. *Cell Host Microbe* **29**: 1076–1092. doi:10.1016/j.chom.2021.05.010
- Gupta V, Earl DJ, Deem MW. 2006. Quantifying influenza vaccine efficacy and antigenic distance. *Vaccine* **24**: 3881–3888. doi:10.1016/j.vaccine.2006.01.010
- Hale VL, Dennis PM, McBride DS, Nolting JM, Madden C, Huey D, Ehrlich M, Grieser J, Winston J, Lombardi D, et al. 2022. SARS-CoV-2 infection in free-ranging white-tailed deer. *Nature* **602**: 481–486. doi:10.1038/s41586-021-04353-x
- Hannon WW, Roychoudhury P, Xie H, Shrestha L, Addetia A, Jerome KR, Greninger AL, Bloom JD. 2022. Narrow transmission bottlenecks and limited within-host viral diversity during a SARS-CoV-2 outbreak on a fishing boat. *bioRxiv* doi:10.1101/2022.02.09.479546
- Harris RS, Dudley JP. 2015. APOBECs and virus restriction. *Virology* **479–480**: 131–145. doi:10.1016/j.virol.2015.03.012
- Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, Ludden C, Reeve R, Rambaut A, Consortium CGU, et al. 2021. SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol* **19**: 409–424. doi:10.1038/s41579-021-00573-0
- Hie B, Zhong ED, Berger B, Bryson B. 2021. Learning the language of viral evolution and escape. *Science* **371**: 284–288. doi:10.1126/science.abd7331
- Hill V, Du Plessis L, Peacock TP, Aggarwal D, Colquhoun R, Carabelli AM, Ellaby N, Gallagher E, Groves N, Jackson B, et al. 2022. The origins and molecular evolution of SARS-CoV-2 lineage B.1.1.7 in the UK. *bioRxiv* doi:10.1101/2022.03.08.481609
- Holmes EC, Goldstein SA, Rasmussen AL, Robertson DL, Crits-Christoph A, Wertheim JO, Anthony SJ, Barclay WS, Boni MF, Doherty PC, et al. 2021. The origins of SARS-CoV-2: a critical review. *Cell* **184**: 4848–4856. doi:10.1016/j.cell.2021.08.017
- Hon CC, Lam TY, Shi ZL, Drummond AJ, Yip CW, Zeng F, Lam PY, Leung FC. 2008. Evidence of the recombinant origin of a bat severe acute respiratory syndrome (SARS)-like coronavirus and its implications on the direct ancestor of SARS coronavirus. *J Virol* **82**: 1819–1826. doi:10.1128/JVI.01926-07
- Hu W, Bai B, Hu Z, Chen Z, An X, Tang L, Yang J, Wang H, Wang H. 2005. Development and evaluation of a multi-target real-time Taqman reverse transcription-PCR assay for detection of the severe acute respiratory syndrome-associated coronavirus and surveillance for an apparently related coronavirus found in masked palm civets. *J Clin Microbiol* **43**: 2041–2046. doi:10.1128/JCM.43.5.2041-2046.2005
- Jackson B, Boni MF, Bull MJ, Collier A, Colquhoun RM, Darby AC, Haldenby S, Hill V, Lucaci A, McCrone JT, et al. 2021. Generation and transmission of interlineage recombinants in the SARS-CoV-2 pandemic. *Cell* **184**: 5179–5188.e8. doi:10.1016/j.cell.2021.08.014
- Johnson BA, Xie X, Bailey AL, Kalveram B, Lokugamage KG, Muruato A, Zou J, Zhang X, Juelich T, Smith JK, et al. 2021. Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature* **591**: 293–299. doi:10.1038/s41586-021-03237-4
- Keck JG, Soe LH, Makino S, Stohman SA, Lai MM. 1988. RNA recombination of murine coronaviruses: recombination between fusion-positive mouse hepatitis virus A59 and fusion-negative mouse hepatitis virus 2. *J Virol* **62**: 1989–1998. doi:10.1128/jvi.62.6.1989-1998.1988
- Keeton R, Tincho MB, Ngomti A, Baguma R, Benede N, Suzuki A, Khan K, Cele S, Bernstein M, Karim F, et al. 2021. SARS-CoV-2 spike T cell responses induced upon vaccination or infection remain robust against Omicron. *medRxiv* doi:10.1038/s41586-022-04460-3
- Kemp SA, Collier DA, Datir RP, Ferreira I, Gayed S, Jahun A, Hosmillo M, Rees-Spear C, Mlcochova P, Lumb IU, et al. 2021. SARS-CoV-2 evolution during treatment of chronic infection. *Nature* **592**: 277–282. doi:10.1038/s41586-021-03291-y
- Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi EE, Bhattacharya T, Foley B, et al. 2020. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* **182**: 812–827.e19. doi:10.1016/j.cell.2020.06.043
- Lai MM, Cavanagh D. 1997. The molecular biology of coronaviruses. *Adv Virus Res* **48**: 1–100.
- Lai MM, Baric RS, Makino S, Keck JG, Egbert J, Leibowitz JL, Stohman SA. 1985. Recombination between nonsegmented RNA genomes of murine coronaviruses. *J Virol* **56**: 449–456. doi:10.1128/jvi.56.2.449-456.1985
- Lempp FA, Soriaga LB, Montiel-Ruiz M, Benigni F, Noack J, Park YJ, Bianchi S, Walls AC, Bowen JE, Zhou J, et al. 2021. Lectins enhance SARS-CoV-2 infection and influence neutralizing antibodies. *Nature* **598**: 342–347. doi:10.1038/s41586-021-03925-1
- Liu J, Chandrashekar A, Sellers D, Barrett J, Jacob-Dolan C, Lifton M, McMahan K, Sciacca M, VanWyk H, Wu C, et al. 2022. Vaccines elicit highly conserved cellular immunity to SARS-CoV-2 Omicron. *Nature* **603**: 493–496. doi:10.1038/s41586-022-04465-y
- Lowery SA, Sariol A, Perlman S. 2021. Innate immune and inflammatory responses to SARS-CoV-2: implications for COVID-19. *Cell Host Microbe* **29**: 1052–1062. doi:10.1016/j.chom.2021.05.004
- Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, et al. 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**: 565–574. doi:10.1016/S0140-6736(20)30251-8
- Lytras S. 2020. The *Sarbecovirus* origin of SARS-CoV-2's furin cleavage site. *Virological.org Discussion Forum*, August 2, 2020. <https://virological.org/t/the-sarbecovirus-origin-of-sars-cov-2-s-furin-cleavage-site/536> [Accessed February 14, 2022].
- Lytras S, Xia W, Hughes J, Jiang X, Robertson DL. 2021. The animal origin of SARS-CoV-2. *Science* **373**: 968–970. doi:10.1126/science.abh0117
- Lytras S, Hughes J, Martin D, Swanepoel P, de Klerk A, Lourens R, Kosakovsky Pond SL, Xia W, Jiang X, Robertson DL. 2022. Exploring the natural origins of SARS-CoV-2 in the light of recombination. *Genome Biol Evol* **14**: evac018. doi:10.1093/gbe/evac018





A. Telenti et al.

- MacLean OA, Lytras S, Weaver S, Singer JB, Boni MF, Lemey P, Kosakovsky Pond SL, Robertson DL. 2021. Natural selection in the evolution of SARS-CoV-2 in bats created a generalist virus and highly capable human pathogen. *PLoS Biol* **19**: e3001115. doi:10.1371/journal.pbio.3001115
- Madden EA, Plante KS, Morrison CR, Kutchko KM, Sanders W, Long KM, Taft-Benz S, Cruz Cisneros MC, White AM, Sarkar S, et al. 2020. Using SHAPE-MaP to model RNA secondary structure and identify 3'UTR variation in Chikungunya virus. *J Virol* **94**: e00701-20. doi:10.1128/JVI.00701-20
- Maher MC, Bartha I, Weaver S, di Iulio J, Ferri E, Soriaga L, Lempp FA, Hie BL, Bryson B, Berger B, et al. 2022. Predicting the mutational drivers of future SARS-CoV-2 variants of concern. *Sci Transl Med* **14**: eabk3445. doi:10.1126/scitranslmed.abk3445
- Mallapaty S. 2021. The search for animals harbouring coronavirus—and why it matters. *Nature* **591**: 26–28. doi:10.1038/d41586-021-00531-z
- Manfredonia I, Nithin C, Ponce-Salvatierra A, Ghosh P, Wirecki TK, Marinus T, Ogando NS, Snijder EJ, van Hemert MJ, Bujnicki JM, et al. 2020. Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. *Nucleic Acids Res* **48**: 12436–12452. doi:10.1093/nar/gkaa1053
- Maponga TG, Jeffries M, Tegally H, Sutherland AD, Wilkinson E, Lessells R, Msomi N, van Zyl G, de Oliveira T, Preiser W. 2022. Persistent SARS-CoV-2 infection with accumulation of mutations in a patient with poorly controlled HIV infection. *SSRN*, January 28, 2022. doi:10.2139/ssrn.4014499. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4014499](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4014499)
- Martin DP, Weaver S, Tegally H, San JE, Shank SD, Wilkinson E, Lucaci AG, Giandhari J, Naidoo S, Pillay Y, et al. 2021. The emergence and ongoing convergent evolution of the SARS-CoV-2 N501Y lineages. *Cell* **184**: 5189–5200. e7. doi:10.1016/j.cell.2021.09.003
- Martin DP, Lytras S, Lucaci AG, Maier W, Grüning B, Shank SD, Weaver S, MacLean OA, Orton RJ, Lemey P, et al. 2022. Selection analysis identifies unusual clustered mutational changes in Omicron lineage BA.1 that likely impact spike function. bioRxiv doi:10.1101/2022.01.14.476382
- McCallum M, Walls AC, Sprouse KR, Bowen JE, Rosen LE, Dang HV, De Marco A, Franko N, Tilles SW, Logue J, et al. 2021. Molecular basis of immune evasion by the Delta and Kappa SARS-CoV-2 variants. *Science* **374**: 1621–1626. doi:10.1126/science.abl8506
- McCallum M, Czudnochowski N, Rosen LE, Zepeda SK, Bowen JE, Walls AC, Hauser K, Joshi A, Stewart C, Veessler D, et al. 2022. Structural basis of SARS-CoV-2 Omicron immune evasion and receptor engagement. *Science* **375**: 864–868 doi:10.1126/science.abn8652
- McLaren PJ, Gawanbacht A, Pyndiah N, Krapp C, Hotter D, Kluge SF, Götz N, Heilmann J, Mack K, Sauter D, et al. 2015. Identification of potential HIV restriction factors by combining evolutionary genomic signatures with functional analyses. *Retrovirology* **12**: 41. doi:10.1186/s12977-015-0165-5
- McMahan K, Giffin V, Tostanoski LH, Chung B, Siamatu M, Suthar MS, Halfmann P, Kawaoka Y, Piedra-Mora C, Martinot AJ, et al. 2022. Reduced pathogenicity of the SARS-CoV-2 Omicron variant in hamsters. bioRxiv doi:10.1101/2022.01.02.474743
- Menachery VD, Yount BL Jr, Debink K, Agnihothram S, Gralinski LE, Plante JA, Graham RL, Scobey T, Ge XY, Donaldson EF, et al. 2015. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat Med* **21**: 1508–1513. doi:10.1038/nm.3985
- Meng B, Abdullahi A, Ferreira I, Goonawardane N, Saito A, Kimura I, Yamasoba D, Gerber PP, Fatihi S, Rathore S, et al. 2022. Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and fusogenicity. *Nature* **603**: 706–714. doi:10.1038/s41586-022-04474-x
- Millet JK, Jaimes JA, Whittaker GR. 2021. Molecular diversity of coronavirus host cell entry receptors. *FEMS Microbiol Rev* **45**: fuaa057. doi:10.1093/femsre/fuaa057
- Mlcochova P, Kemp SA, Dhar MS, Papa G, Meng B, Ferreira I, Dahir R, Collier DA, Albecka A, Singh S, et al. 2021. SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature* **599**: 114–119. doi:10.1038/s41586-021-03944-y
- Obermeyer F, Schaffner SF, Jankowiak M, Barkas N, Pyle JD, Park DJ, MacInnis BL, Luban J, Sabeti PC, Lemieux JE. 2021. Analysis of 2.1 million SARS-CoV-2 genomes identifies mutations associated with transmissibility. medRxiv doi:10.1101/2021.09.07.21263228
- Obermeyer F, Jankowiak M, Barkas N, Schaffner SF, Pyle JD, Yurkovetskiy L, Bosso M, Park DJ, Babadi M, MacInnis BL, et al. 2022. Analysis of 6.4 million SARS-CoV-2 genomes identifies mutations associated with fitness. medRxiv doi:10.1101/2021.09.07.21263228
- O'Driscoll M, Ribeiro Dos Santos G, Wang L, Cummings DAT, Azman AS, Paireau J, Fontanet A, Cauchemez S, Salje H. 2021. Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature* **590**: 140–145. doi:10.1038/s41586-020-2918-0
- Oreshkova N, Molenaar RJ, Vreman S, Harders F, Oude Munnink BB, Hakze-van der Honing RW, Gerhards N, Tolsma P, Bouwstra R, Sikkema RS, et al. 2020. SARS-CoV-2 infection in farmed minks, the Netherlands, April and May 2020. *Euro Surveill* **25**: 2001005. doi:10.2807/1560-7917.ES.2020.25.23.2001005
- Oude Munnink BB, Sikkema RS, Nieuwenhuijse DF, Molenaar RJ, Munger E, Molenkamp R, van der Spek A, Tolsma P, Rietveld A, Brouwer M, et al. 2021. Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science* **371**: 172–177. doi:10.1126/science.abe5901
- Peacock TP, Bauer DLV, Barclay WS. 2021a. Putative host origins of RNA insertions in SARS-CoV-2 genomes. *Virological.org Discussion Forum*, October 11, 2021. <https://virological.org/t/putative-host-origins-of-rna-insertions-in-sars-cov-2-genomes/761> [Accessed February 14, 2022].
- Peacock TP, Brown JC, Zhou J, Thakur N, Newman J, Kugathasan R, Sukhova K, Kaforou M, Bailey D, Barclay WS. 2021b. The SARS-CoV-2 variant, Omicron, shows rapid replication in human primary nasal epithelial cultures and efficiently uses the endosomal route of entry. bioRxiv doi:10.1101/2021.12.31.474653
- Peacock TP, Goldhill DH, Zhou J, Baillon L, Frise R, Swann OC, Kugathasan R, Penn R, Brown JC, Sanchez-David



- RY, et al. 2021c. The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. *Nat Microbiol* **6**: 899–909. doi:10.1038/s41564-021-00908-w
- Pfaller CK, George CX, Samuel CE. 2021. Adenosine deaminases acting on RNA (ADARs) and viral infections. *Annu Rev Virol* **8**: 239–264. doi:10.1146/annurev-virology-091919-065320
- Pickering B, Lung O, Maguire F, Kruczkiewicz P, Kotwa JD, Buchanan T, Gagnier M, Guthrie JL, Jardine CM, Marchand-Austin A, et al. 2022. Highly divergent white-tailed deer SARS-CoV-2 with potential deer-to-human transmission. bioRxiv doi:10.1101/2022.02.22.481551
- Prince T, Smith SL, Radford AD, Solomon T, Hughes GL, Patterson EI. 2021. SARS-CoV-2 infections in animals: reservoirs for reverse zoonosis and models for study. *Viruses* **13**: 494. doi:10.3390/v13030494
- Radvak P, Kwon HJ, Kosikova M, Ortega-Rodriguez U, Xiang R, Phue JN, Shen RF, Rozzelle J, Kapoor N, Rabara T, et al. 2021. SARS-CoV-2 B.1.1.7 (alpha) and B.1.351 (beta) variants induce pathogenic patterns in K18-hACE2 transgenic mice distinct from early strains. *Nat Commun* **12**: 6559. doi:10.1038/s41467-021-26803-w
- Rambaut A, Holmes EC, O'Toole A, Hill V, McCrone JT, Ruis C, du Plessis L, Pybus OG. 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* **5**: 1403–1407. doi:10.1038/s41564-020-0770-5
- Riddell AC, Kele B, Harris K, Bible J, Murphy M, Dakshina S, Storey N, Owoyemi D, Pade C, Gibbons JM, et al. 2022. Generation of novel SARS-CoV-2 variants on B.1.1.7 lineage in three patients with advanced HIV disease. medRxiv doi:10.1101/2022.01.14.21267836
- Rockett RJ, Draper J, Gall M, Sim EM, Arnott A, Agius JE, Johnson-Mackinnon J, Martinez E, Drew AP, Lee C, et al. 2022. Co-infection with SARS-COV-2 Omicron and Delta variants revealed by genomic surveillance. medRxiv doi:10.1101/2022.02.13.22270755
- Rodriguez-Rivas J, Croce G, Muscat M, Weigt M. 2022. Epistatic models predict mutable sites in SARS-CoV-2 proteins and epitopes. *Proc Natl Acad Sci* **119**: e2113118119. doi:10.1073/pnas.2113118119
- Roemer C. 2022. Potential US East Coast BA.1.1/AY.119\* recombinant (7 seq). *GitHub*, February 14, 2022. <https://github.com/cov-lineages/pango-designation/issues/439> [Accessed February 17, 2022].
- Sánchez CA, Li H, Phelps KL, Zambrana-Torrel C, Wang LF, Olival KJ, Daszak P. 2021. A strategy to assess spillover risk of bat SARS-related coronaviruses in Southeast Asia. medRxiv doi:10.1101/2021.09.09.21263359
- Sharun K, Dhama K, Pawde AM, Gortázar C, Tiwari R, Bonilla-Aldana DK, Rodriguez-Morales AJ, de la Fuente J, Michalak I, Attia YA. 2021. SARS-CoV-2 in animals: potential for unknown reservoir hosts and public health implications. *Vet Q* **41**: 181–201. doi:10.1080/01652176.2021.1921311
- Shi Z, Hu Z. 2008. A review of studies on animal reservoirs of the SARS coronavirus. *Virus Res* **133**: 74–87. doi:10.1016/j.virusres.2007.03.012
- Smith EC, Sexton NR, Denison MR. 2014. Thinking outside the triangle: replication fidelity of the largest RNA viruses. *Annu Rev Virol* **1**: 111–132. doi:10.1146/annurev-virology-031413-085507
- Smyth RP, Negroni M, Lever AM, Mak J, Kenyon JC. 2018. RNA structure—a neglected puppet master for the evolution of virus and host immunity. *Front Immunol* **9**: 2097. doi:10.3389/fimmu.2018.02097
- Snoeck J, Fellay J, Bartha I, Douek DC, Telenti A. 2011. Mapping of positive selection sites in the HIV-1 genome in the context of RNA and protein structural constraints. *Retrovirology* **8**: 87. doi:10.1186/1742-4690-8-87
- Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, Navarro MJ, Bowen JE, Tortorici MA, Walls AC, et al. 2020. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* **182**: 1295–1310.e20. doi:10.1016/j.cell.2020.08.012
- Starr TN, Zepeda SK, Walls AC, Greaney AJ, Alkhovsky S, Veelsler D, Bloom JD. 2022. ACE2 binding is an ancestral and evolvable trait of sarbecoviruses. *Nature* **603**: 913–918. doi:10.1038/s41586-022-04464-z
- Subramanian R, He Q, Pascual M. 2021. Quantifying asymptomatic infection and transmission of COVID-19 in New York City using observed cases, serology, and testing capacity. *Proc Natl Acad Sci* **118**: e2019716118. doi:10.1073/pnas.2019716118
- Suzuki R, Yamasoba D, Kimura I, Wang L, Kishimoto M, Ito J, Morioka Y, Nao N, Nasser H, Uriu K, et al. 2022. Attenuated fusogenicity and pathogenicity of SARS-CoV-2 Omicron variant. *Nature* **603**: 700–705. doi:10.1038/s41586-022-04462-1
- Tarke A, Coelho CH, Zhang Z, Dan JM, Yu ED, Methot N, Bloom NI, Goodwin B, Phillips E, Mallal S, et al. 2022. SARS-CoV-2 vaccination induces immunological T-cell memory able to cross-recognize variants from Alpha to Omicron. *Cell* **185**: 847–859. doi:10.1016/j.cell.2022.01.015
- Telenti A, Arvin A, Corey L, Corti D, Diamond MS, García-Sastre A, Garry RF, Holmes EC, Pang PS, Virgin HW. 2021. After the pandemic: perspectives on the future trajectory of COVID-19. *Nature* **596**: 495–504. doi:10.1038/s41586-021-03792-w
- Temmam S, Vongphayloth K, Salazar EB, Munier S, Bonomi M, Regnault B, Douangboubpha B, Karami Y, Chrétien D, Sanamxay D, et al. 2022. Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature* doi:10.1038/s41586-022-04532-4
- Thomson EC, Rosen LE, Shepherd JG, Spreafico R, da Silva Filipe A, Wojcechowskyj JA, Davis C, Piccoli L, Pascall DJ, Dillen J, et al. 2021. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell* **184**: 1171–1187.e20. doi:10.1016/j.cell.2021.01.037
- Thorne LG, Bouhaddou M, Reuschl AK, Zuliani-Alvarez L, Polacco B, Pelin A, Batra J, Whelan MVX, Hosmillo M, Fossati A, et al. 2022. Evolution of enhanced innate immune evasion by SARS-CoV-2. *Nature* **602**: 487–495. doi:10.1038/s41586-021-04352-y
- Turkahia Y, Thornlow B, Hinrichs A, McBroome J, Ayala N, Ye C, De Maio N, Haussler D, Lanfear R, Corbett-Detig R. 2021. Pandemic-scale phylogenomics reveals elevated recombination rates in the SARS-CoV-2 spike region. bioRxiv doi:10.1101/2021.08.04.455157
- Twohig KA, Nyberg T, Zaidi A, Thelwall S, Sinnathamby MA, Aliabadi S, Seaman SR, Harris RJ, Hope R, Lopez-





A. Telenti et al.

- Bernal J, et al. 2022. Hospital admission and emergency care attendance risk for SARS-CoV-2 delta (B.1.617.2) compared with alpha (B.1.1.7) variants of concern: a cohort study. *Lancet Infect Dis* **22**: 35–42. doi:10.1016/S1473-3099(21)00475-8
- Uriu K, Kimura I, Shirakawa K, Takaori-Kondo A, Nakada TA, Kaneda A, Nakagawa S, Sato K, Genotype to Phenotype Japan C. 2021. Neutralization of the SARS-CoV-2 Mu variant by convalescent and vaccine serum. *N Engl J Med* **385**: 2397–2399. doi:10.1056/NEJMc2114706
- US Centers for Disease Control and Prevention. 2021. SARS-CoV-2 variant classifications and definitions. CDC, December 1, 2021. <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html/> [Accessed February 14, 2022].
- Walker AS, Vihta KD, Gethings O, Pritchard E, Jones J, House T, Bell I, Bell JI, Newton JN, Farrar J, et al. 2021. Tracking the emergence of SARS-CoV-2 Alpha variant in the United Kingdom. *N Engl J Med* **385**: 2582–2585. doi:10.1056/NEJMc2103227
- Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW Jr, Swanstrom R, Burch CL, Weeks KM. 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**: 711–716. doi:10.1038/nature08237
- Wei C, Shan KJ, Wang W, Zhang S, Huan Q, Qian W. 2021. Evidence for a mouse origin of the SARS-CoV-2 Omicron variant. *J Genet Genomics* **48**: 1111–1121. doi:10.1016/j.jgg.2021.12.003
- Wertheim JO, Wang JC, Leelawong M, Martin DP, Havens JL, Chowdhury MA, Pekar J, Amin H, Arroyo A, Awandare GA, et al. 2022. Capturing intrahost recombination of SARS-CoV-2 during superinfection with Alpha and Epsilon variants in New York City. medRxiv doi:10.1101/2022.01.18.22269300
- WHO. 2022. Tracking SARS-CoV-2 variants. *World Health Organization*, February 3, 2022. <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/> [Accessed February 14, 2022].
- Willett BJ, Grove J, MacLean OA, Wilkie C, Logan N, Lorenzo GD, Furnon W, Scott S, Manali M, Szemiel A, et al. 2022. The hyper-transmissible SARS-CoV-2 Omicron variant exhibits significant antigenic change, vaccine escape and a switch in cell entry mechanism. medRxiv doi:10.1101/2022.01.03.21268111
- Worobey M. 2021. Dissecting the early COVID-19 cases in Wuhan. *Science* **374**: 1202–1204. doi:10.1126/science.abm4454
- Worobey M, Levy JL, Malpica Serrano LM, Crits-Christoph A, Pekar JE, Goldstein SA, Rasmussen AL, Kraemer MUG, Newman C, Koopmans MPG, et al. 2022. The Huanan market was the epicenter of SARS-CoV-2 emergence. *Zenodo*, February 26, 2022. <https://zenodo.org/record/6299600>
- Xiong Q, Cao L, Ma C, Liu C, Si J, Liu P, Gu M, Wang C, Shi L, Tong F, et al. 2022. Close relatives of MERS-CoV in bats use ACE2 as their functional receptors. bioRxiv doi:10.1101/2022.01.24.477490
- Yamasoba D, Kimura I, Nasser H, Morioka Y, Nao N, Ito J, Uriu K, Tsuda M, Zahradnik J, Shirakawa K, et al. 2022. Virological characteristics of SARS-CoV-2 BA.2 variant. bioRxiv doi:10.1101/2022.02.14.480335
- Yen HL, Sit THC, Brackman CJ, Chuk SSY, Gu H, Tam KWS, Law PYT, Leung GM, Peiris M, Poon LLM, et al. 2022. Transmission of SARS-CoV-2 delta variant (AY.127) from pet hamsters to humans, leading to onward human-to-human transmission: a case study. *Lancet* **399**: 1070–1078. doi:10.1016/S0140-6736(22)00326-9
- Yewdell JW. 2021. Antigenic drift: understanding COVID-19. *Immunity* **54**: 2681–2687. doi:10.1016/j.immuni.2021.11.016
- Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, et al. 2020. Addendum: A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **588**: E6. doi:10.1038/s41586-020-2951-z